Neural Network Analysis of Simulated Heroin Samples

Kar-Weng Chan[a]

[a] *Department of Chemistry Malaysia, Ministry of Science, Technology and Innovation (MOSTI), 46661 Petaling Jaya, Selangor*

ABSTRACT: Sample classification remains a popular concern in drug profiling. Various statistical techniques have been investigated but literature pertaining to the potential of neural network analysis (NNA) in the realm of drug intelligence is relatively scarce. This study employed seven links containing 216 samples (parent samples and post-cut samples) to assess the performance of NNA via Radial Basis Function Network (RBFN). Statistical validation of the pretreatment method for the target samples has been reported in a previous study. It was found that eleven quotients derived from five opium based alkaloids were important for grouping samples under the conditions specified in this study. By using the same quotients in this study, 70% of the samples were assigned to train RBFN in recognizing key patterns (or characteristics) inherent in the links, while 30% served to test the learned patterns. By allowing RBFN itself to choose any random samples into the training and testing sets (this method is known as 'selection with minimal bias'), ten rounds of analysis showed different outcomes. It had 50% chances of achieving 100% correctness in predicting sample groupings both in training and testing sets. The worst performance with only 77.8% correctness was displayed by the testing set in round 5 for group 1. Occasional failures observed in the classification indicated that RBFN was not able to segregate some of the sample units belonging to group 1 from group 2. This study however highlights that the established quotients remain practical for grouping samples of known relationships. When it is used for RBFN through the selection with minimal bias approach, correctness of grouping may vary due to random sample assignment to the training and testing sets.

Keywords: forensic science, simulated heroin, neural network analysis, radial basis function network

Introduction

The abuse of illicit heroin has been a major social menace in Malaysia. Heroin cases submitted to the enforcement laboratory are steadily on the rise (Fig. 1). Heroin constituted approximately 47% of all drug cases received by the laboratory in 2012 in Malaysia.

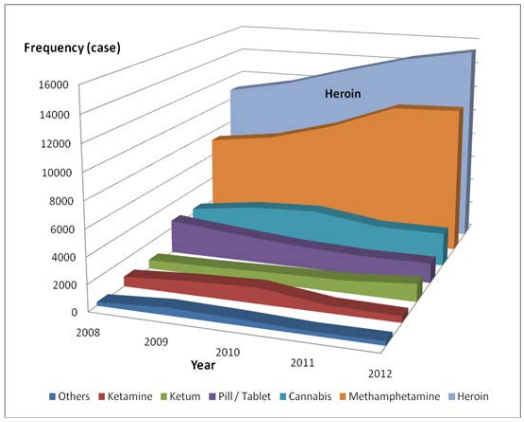

Figure 1: Prevalence of drug cases submitted to the Department of Chemistry Malaysia

Heroin is produced by attaching two acetyl groups to morphine, a major ingredient found in opium. This addictive drug is often used by drug addicts in small doses to create euphoria. It has been argued that heroin itself acts solely as a vehicle which travels in the bloodstream and crosses the blood-brain-barrier (BBB) before it is converted back to morphine to exert opiate-related effects in the brain. Therefore, heroin is characterized as a more potent drug than morphine because the former has a greater ability to cross the BBB. As far as opiate group of drugs is concerned, drug addicts tend to choose heroin rather than morphine for recreational activities.

Opium contains more than 35 alkaloids [1]. Illicit heroin samples seized in Malaysia often contain natural alkaloids such as codeine and morphine as well as by-products like acetylcodeine and monoacetylmorphine (MM) in detectable and quantifiable amounts [2-3]. Two types of MM, namely 3-MM and 6-MM are common. During processing, morphine is converted to heroin through the formation of 3-MM as an intermediate compound. Under improper storage conditions, heroin is likely

decomposed back to morphine via 6-MM [4-6]. As a result, samples from the same batch will display varying compositions of morphine, heroin and MM. Besides, after cutting is performed on a pre-cut heroin sample, the amount of the opium based alkaloids will become much lower. Prior to packing the drug for street sales, another cutting process may be performed. Therefore, the opium based alkaloids in the cut samples will exhibit a significant difference in the absolute compositions, resulting from two different cutting processes. Collectively, heroin samples coming from the same source eventually show different compositions because of the decomposition and cutting effects.

Opium based alkaloids are crucial in determining the source level relationships of heroin samples. Many gas chromatographic and high performance liquid chromatographic methods have been developed to profile illicit heroin [1, 7-11]. Certainly, a valid analytical method is vital for providing high precision data for clustering. In addition, a valid statistical technique must also work symbiotically with the analytical method. To this end, researchers tended to find the best pretreatment method, distance measures, linkage methods and cut-off values for a given statistical technique, hoping that the chosen technique could give meaningful clustering results for decision making [1,3,12-13].

In this paper, data from a previous study [3] were re-evaluated using neural network analysis (NAA). In the previous study, clustering outcomes generated by principal component analysis (PCA), hirarchical cluster analysis (HCA) and discriminant analysis (DA) have been reported. The pretreatment method worked satisfactorily well with the above-cited statistical techniques. All of these statistical techniques function by reducing data variabilities existing in multiple dimensions into decipherable two or three dimensions so that the human brain can interpret the pattern. PCA and HCA are commonly used for unsupervised pattern recognition because they do not need to be trained upfront using data of known sources/relationships. PCA usually provides a two- or three-dimensional diagram to illustrate the relative distances (or relationships) of the sample units. HCA on the other hand presents a tree diagram (or dendogram) through which direct relationships of the sample units can be revealed. DA and NAA are better employed for supervised pattern recognition which must first acquire the inherent characteristics of known samples/groups before they can assign unknown samples into the known groups based on the learned qualities. To further explore the role of chemometrics, it is the focus of this paper to re-evaluate the chosen pretreatment method as well as NNA (especially Radial Basis Function Network, RBFN) which in this case automatically assigned samples into training and testing sets – an enhanced evaluation approach that gives results with minimal bias for decision making in sample classification. This relatively low bias manner is afforded because the user does not have the right to assign samples into the sets, as opposed to DA reported in the previous study where user-defined sample assignment is possible. RBFN operates like neurons in human central nervous system. In general, if signals coming from the input exceed a threshold level, they will determine the kind of output that corresponds the input. These signals are better represented by the characteristics of the sample units used in sample classification via RBFN. The hidden layer of RBFN receives non-linear transformation from the input layer and directs linear transformation to the output layer. One major limitation of RBFN is that it requires a large amount of sample units from known sources/groups to identify the characteristics of the groups.

Materials and methods

Standards and Solvents

Ten food-coloring agents (ammarant, tetrazine, green S, rhodamine BS, fast green, brilliant blue, sunset yellow, carmoisine, red 2G, and erythrosine BS) and acetylcodeine hydrochloride were obtained from the Department of Chemistry Malaysia. Paracetamol and dextromethorphan hydrobromide were supplied by Y.S.P. Industries. Codeine phosphate, morphine hydrochloride, and heroin hydrochloride were purchased from Johnson Matthey Macfarlan Smith. 6-Monoacetylmorphine hydrochloride was commercially obtained from Lipomed and 2,2,2 triphenyl acetophenone (internal standard [IS]), from Aldrich Chemical Company. Caffeine was purchased from Merck. HPLC grade methanol and analytical reagent grade chloroform were both purchased from Fisher Scientific.

Preparation of simulated links

The procedure involved in the preparation of simulated links of heroin samples has been described in [3]. Table 1 summarizes the details of seven links prepared from four unrelated case samples (parent samples marked A to D) and a mixture of chemical standards. These samples were mixed with specified standards or diluents to give first level lab-based parent samples. Subsequently, second level lab-based samples were prepared by mixing specified amounts of the first level lab-based sample and a selected mixture of diluents (X or Y) to give post-cut samples containing varying concentrations of opium based alkaloids (Table 1). Each sample was divided into two portions where the second portion was subjected to coloring. A total of 8 first level lab-based parent samples, 32 uncolored and 32 colored second level lab-based samples were investigated. From each of these samples, three random lab samples, each weighing 80-85 mg were taken and dissolved in 9:1 chloroform:methanol with the presence of 0.18 mg/mL IS. Finally, all samples were analyzed in triplicate by a gas chromatograph.

Table 1: Preparation of 216 samples for data analysis

Link/group	Parent sample	Quantified heroin base (%)	Standards added[a]	First level lab-based parent sample	Cutting agent[b]	Second level lab-based sample: % of first level lab-based sample added[c] (coloring agent)[d]
1	A	47.27%	NM	A1	X	5 (AR), 12.5 (TZ), 25 (GS), 50 (RBS)
2			MP, MM	A2	Y	5 (FG), 5 (BB), 25 (SY), 50 (CS)
3	B	38.03%	NM	B1	X	5 (AR), 12.5 (TZ), 25 (GS), 50 (RBS)
4			DM, MP, AC, MM	B2	Y	5 (R2G), 12.5 (EBS), 25 (AR), 50 (TZ)
5[e]	C	23.05%	NM	C1	Y	5 (GS), 12.5 (RBS), 25 (FG), 25 (BB)
			PC	C2	Y	2.5 (R2G), 7.5 (EBS), 12.5 (AR), 25 (SY)
6	D	49.14%	NM	D1	X	2.5 (FG), 7.5 (BB), 12.5 (SY), 25 (CS)
7	Chemical standards (PC, CF, DM, CD, MP, AC, MM, HR)[f]	22.80%	NM	E1	X	2.5 (SY), 7.5 (R2G), 12.5 (EBS), 25 (CS)

[a]PC: Paracetamol, CF: Caffeine, DM: Dextromethorphan, CD: Codeine, MP: Morphine, AC: Acetylcodeine, MM: Monoacetylmorphines (both 3-MM and 6-MM), HR: Heroin, NM: No modification

[b]X contains PC: CF: DM (6:70:4); Y contains PC: CF: DM (9:70:1)

[c]Example for the first sample of link 1, 5% A1 was mixed with 95% mixture of cutting agents X

[d]Coloring agents are applicable to the second portions of the second level lab-based sample. AR: Ammarant, TZ: Tetrazine, GS: Green S, RBS: Rhodamine BS, FG: Fast green, BB: Brilliant blue, SY: Sunset yellow, CS: Carmoisine, R2G: Red 2G, EBS: Erythrosine BS

[e]Two links (C1 and C2) maintained similar alkaloidal ratios, hence they belong to a single link.

[f]Phenolphthalein was used as part of the parent sample of link 7

Gas chromatography-flame ionization detector

Quantitative analysis was achieved using a HP6890N GC–FID preinstalled with a J&W HP Ultra 2 (length 25 m, i.d. 0.20 mm, film thickness 0.33 µm) capillary column. Chromatographic separation was accomplished by holding the oven temperature at 240°C for 1 min and heating up to 270°C at the rate of 12°C/min. The oven was then held for 8 min at this temperature. The injector and detector temperatures were set at 290°C, and an injection volume of 1 µL with a split ratio of 40:1 was employed [3, 14].

Data analysis

Data of 216 samples (readings in mg/mL) were collected for evaluation. Basic data manipulation was accomplished with Excel spreadsheet through which calculation for the quotients (e.g. MM/HR) and data pretreatment such as standardization were carried out. NNA was performed with the aid of SPSS version 18. NNA using RBFN was chosen to classify the heroin samples. The rule of partition 7:3 for training:testing was used to analyze the data. A relatively larger sample size was assigned to the training set since this could enhance the reliability of the RBFN in

identifying the characteristics of different target groups/links. The number of neurons/units in the hidden layer was illustrated in each network diagram generated by the software program.

Results and discussion

Heroin is relatively stable if the conditions are unfavorable for decomposition. Extreme heat, moisture and acid are the main contributing factors that render the conversion of heroin to 6-MM. The total MM being quantified is therefore the mix of this 6-MM as well as the pre-existing 3-MM, a product from incomplete acetylation during processing. In this regard, data pretreatment is crucial to compensate for this effect. Some researchers were of the view that the use of quotient involving MM+HR or MP+MM+HR in calculation is sufficient to account for the possible errors arising from decomposition where source level classification is concerned. Such theoretical decision is feeble unless it is supported by empirical data.

In the previous study [3], suitable quotients were estimated using simulated heroin samples. The performance of the quotients was assessed with the aid of PCA, HCA and DA. All of these techniques suggested that the eleven variables derived from five opium based alkaloids were ideal. DA and HCA explicitly showed no errors in the classification outcomes. These 11 variables are as follows:

– AC/HR
– AC/(MM+HR)
– MM/HR
– HR/MM
– AC/MM
– AC/(MP+MM+HR)
– (CD+MP)/(MM+HR)
– (CD+AC)/(MP+MM+HR)
– (CD+MP+MM+HR)/AC
– HR/(CD+MP+AC+MM)
– (MP+MM+HR)/(CD+AC)

The success of these variables is largely due to their abilities in reducing the decomposition and cutting effects in each sample. Standardization was also used by dividing each variable by its standard deviation in order to equate the effects of all variables in the dataset. Hence, eleven pretreated variables are sufficient for clustering (source level classification) all the samples in their respective groups without errors on a dendogram (Fig. 2). Therefore, it was decided to use the aforesaid pretreatment (eleven standardized quotients) and HCA as the primary procedure for classifying samples of unknown sources for case samples seized in Malaysia.

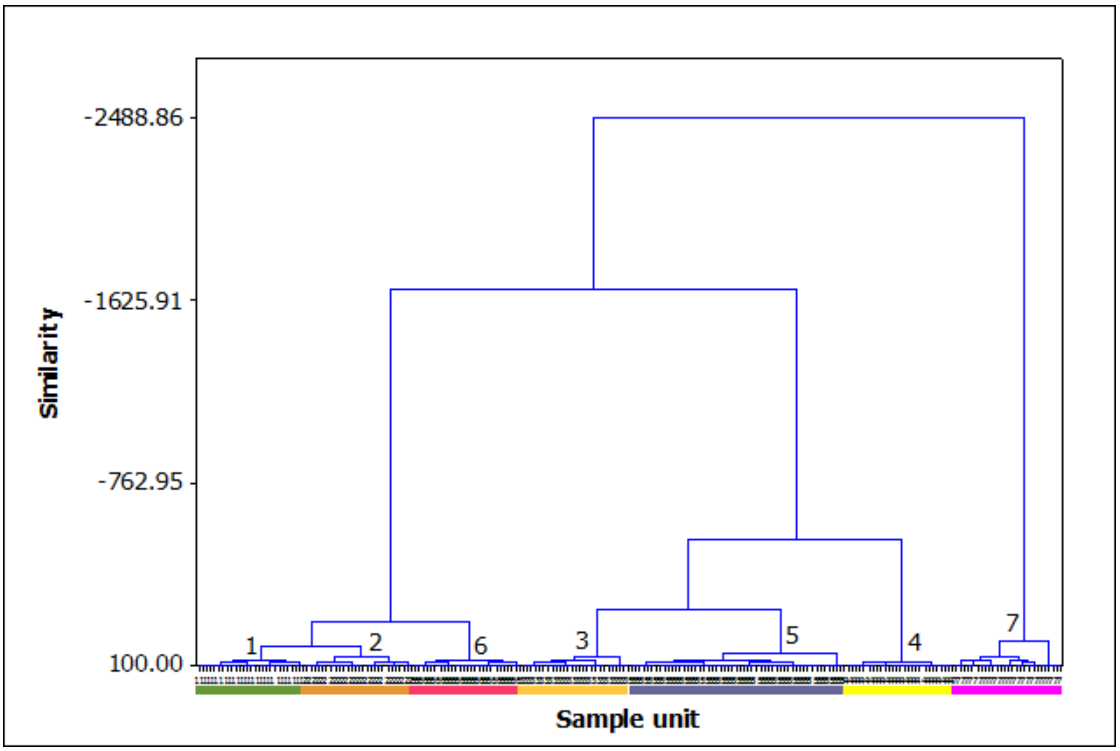

Figure 2: Dendogram for 216 samples analyzed by HCA using Manhattan distance and Ward linkage. Numerals 1 – 7 denote the group/link numbers (refer to Table 1); the x-axis color bars represent the sequences of samples in the respective groups.

This study aims to further evaluate the power of RBFN (also known as 'machine') in classifying the same samples using the established pretreatment. By using the same pretreated dataset for RBFN, the pretreated variables were linked to a hidden layer containing several neurons before the sample units were assigned to seven groups in the output layer (Fig. 3). In the network, mathematical rules are applied to decide if a sample unit belongs to a particular node or path that leads to its eventual group.

Unlike DA done in the previous study [3], where assignment of samples to training and testing sets was strictly user-defined. In contrast, RBFN in this study has the power to decide and put any random samples in either of the sets. Therefore, the correctness of grouping may vary depending on which sample units have been used for training each time. This in turn minimizes biasness in statistical analysis.

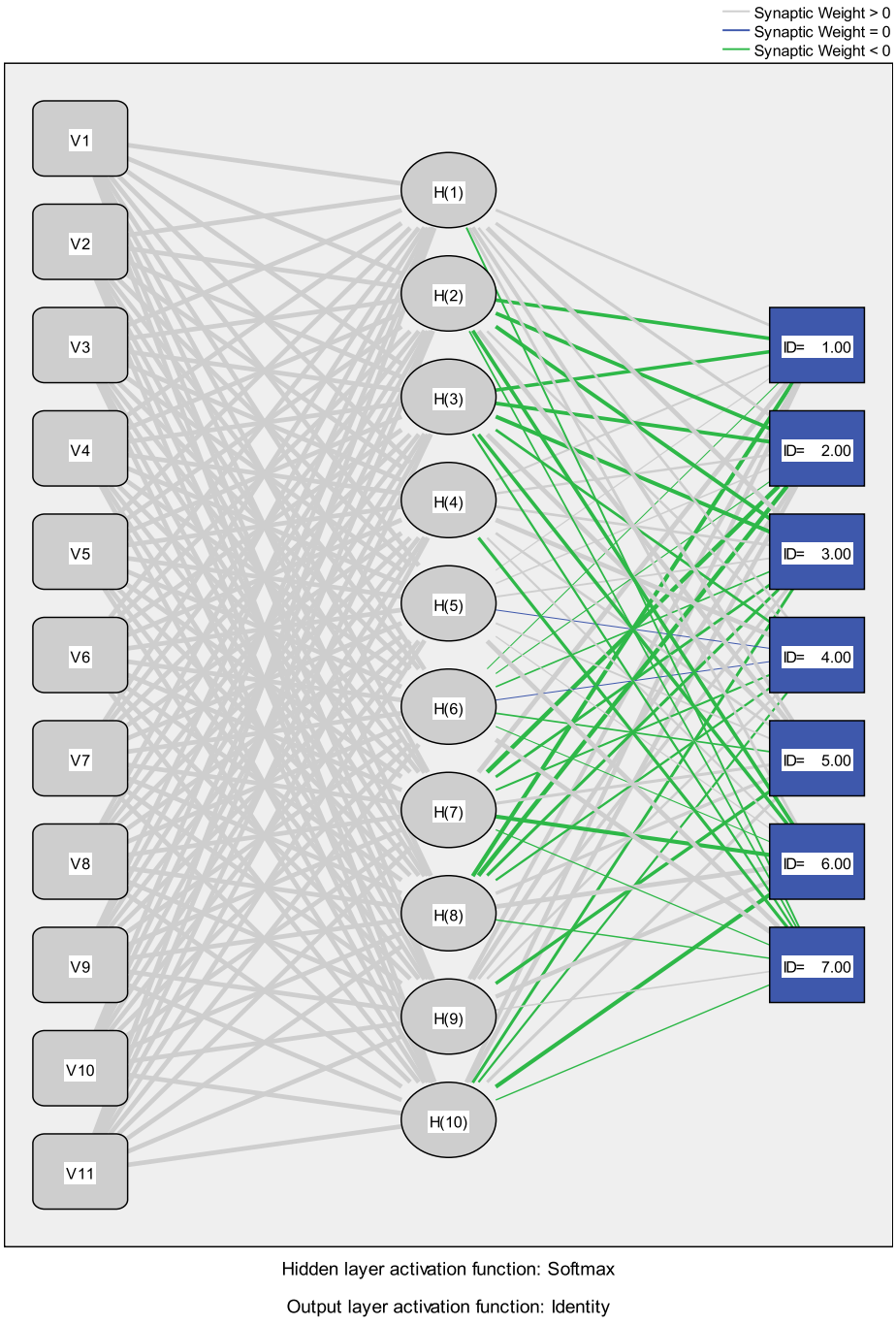

Figure 3: One of the ten random networks generated from eleven pretreated variables (V1 to V11) in the input layer before assigning sample units to seven known groups (ID = 1.00 to ID= 7.00)

By analyzing repeatedly the same pretreated data using the same specifications imposed on RBFN, different classification outcomes will be displayed. For example, in the ten trials, it required seven, nine or ten neurons in the hidden layer to make decisions. This may or may not affect the eventual decision for sample grouping depending on how good the training set is in providing inputs for analysis. The insignificant variation in the outcomes would signify that every time the machine selects random sample units from the dataset into training and testing sets respectively, the data are statistically robust to give more or less the same outcome.

From ten analyses, 50% of the trials were able to give 100% correctness for both training and testing sets (Table 2). Rounds 3 and 10 only displayed errors in either of the sets. The worst results (Table 3) were given in rounds 2, 5 and 7 where both training and testing sets showed errors. These errors came from group 1 where one to three samples units of this group were always mistakenly assigned to group 2. The random assignment of RBFN collaborates with the HCA performance reported in [3] where different distance measures and linkage methods would also give different degrees of correctness.

Table 2: Performance results generated from ten rounds of RBFN

Trial	Number of neurons in the hidden layer	Set	Number of samples used	Total correctness of prediction (%)
Round 1	10	Training	141 (65.3%)	100.0
		Testing	75 (34.7%)	100.0
Round 2	10	Training	154 (71.3%)	99.4 (93.8[1])
		Testing	62 (28.7%)	96.8 (81.8[1])
Round 3	9	Training	145 (67.1%)	100.0
		Testing	71 (32.9%)	98.6 (88.9[1])
Round 4	10	Training	140 (64.8%)	100.0
		Testing	76 (35.2%)	100.0
Round 5	9	Training	144 (66.7%)	99.3 (94.4[1])
		Testing	72 (33.3%)	97.2 (77.8[1])
Round 6	10	Training	161 (74.5%)	100.0
		Testing	55 (25.5%)	100.0
Round 7	7	Training	140 (64.8%)	98.6 (88.9[1])
		Testing	76 (35.2%)	98.7 (88.9[1])
Round 8	10	Training	146 (67.6%)	100.0
		Testing	70 (32.4%)	100.0
Round 9	9	Training	146 (67.6%)	100.0
		Testing	70 (32.4%)	100.0
Round 10	7	Training	147 (68.1%)	98.0 (84.2[1])
		Testing	69 (31.9%)	100.0

Note: Superscript indicates the real group number that shows the specified % correctness in bracket.

The performance is important to provide information on whether the data are statistically robust and stable for RBFN. Apparently, the performance did not highly rely on the number of samples used in the training set. For example, Table 2 shows that errors are still observed even though more samples were used to train RBFN in round 2.

Table 3: Classification results generated from rounds 2, 5 and 7

Classification									
Sample	Observed	Predicted							
		1.00	2.00	3.00	4.00	5.00	6.00	7.00	Percent Correct
Training	1.00	15	1	0	0	0	0	0	93.8%
		17	1	0	0	0	0	0	94.4%
		16	2	0	0	0	0	0	88.9%
	2.00	0	24	0	0	0	0	0	100.0%
		0	21	0	0	0	0	0	100.0%
		0	14	0	0	0	0	0	100.0%
	3.00	0	0	19	0	0	0	0	100.0%
		0	0	16	0	0	0	0	100.0%
		0	0	16	0	0	0	0	100.0%
	4.00	0	0	0	18	0	0	0	100.0%
		0	0	0	17	0	0	0	100.0%
		0	0	0	18	0	0	0	100.0%
	5.00	0	0	0	0	37	0	0	100.0%
		0	0	0	0	37	0	0	100.0%
		0	0	0	0	39	0	0	100.0%
	6.00	0	0	0	0	0	21	0	100.0%
		0	0	0	0	0	19	0	100.0%
		0	0	0	0	0	14	0	100.0%
	7.00	0	0	0	0	0	0	19	100.0%
		0	0	0	0	0	0	16	100.0%
		0	0	0	0	0	0	21	100.0%
	Overall Percent	9.7%	16.2%	12.3%	11.7%	24.0%	13.6%	12.3%	99.4%
		11.8%	15.3%	11.1%	11.8%	25.7%	13.2%	11.1%	99.3%
		11.4%	11.4%	11.4%	12.9%	27.9%	10.0%	15.0%	98.6%
Testing	1.00	9	2	0	0	0	0	0	81.8%
		7	2	0	0	0	0	0	77.8%
		8	1	0	0	0	0	0	88.9%
	2.00	0	3	0	0	0	0	0	100.0%
		0	6	0	0	0	0	0	100.0%
		0	13	0	0	0	0	0	100.0%
	3.00	0	0	8	0	0	0	0	100.0%
		0	0	11	0	0	0	0	100.0%
		0	0	11	0	0	0	0	100.0%
	4.00	0	0	0	9	0	0	0	100.0%
		0	0	0	10	0	0	0	100.0%
		0	0	0	9	0	0	0	100.0%
	5.00	0	0	0	0	17	0	0	100.0%
		0	0	0	0	17	0	0	100.0%
		0	0	0	0	15	0	0	100.0%
	6.00	0	0	0	0	0	6	0	100.0%
		0	0	0	0	0	8	0	100.0%
		0	0	0	0	0	13	0	100.0%
	7.00	0	0	0	0	0	0	8	100.0%
		0	0	0	0	0	0	11	100.0%
		0	0	0	0	0	0	6	100.0%
	Overall Percent	14.5%	8.1%	12.9%	14.5%	27.4%	9.7%	12.9%	96.8%
		9.7%	11.1%	15.3%	13.9%	23.6%	11.1%	15.3%	97.2%
		10.5%	18.4%	14.5%	11.8%	19.7%	17.1%	7.9%	98.7%

Note: Data presented in three rows for each observed group represent data from rounds 2, 5 and 7.

Conclusion

As RBFN has reduced the biasness in sample assignment, the resulting performance is thus with minimal bias. The established quotients performed satisfactorily well with RBFN with a minimum total correctness down to 97.2% while 12 out of 20 correctness estimates gave 100%. This implies that the quotients are still sensitive enough to segregate unrelated sample units without jeopardizing the relationships between related sample units when NNA is chosen for sample classification.

References

1. Zelkowics, A., Magora, A., Ravreby M. D. and Levy, R. (2005), Analysis of a simulated heroin distribution chain by HPLC, *Journal of Forensic Sciences*, 50(4), 849-852.
2. Chan, K. W., Tan G. H. and Wong R. C. S. (2012), Gas chromatographic method validation for the analysis of major components in illicit heroin seized in Malaysia, *Science & Justice*. 52, 9-16.
3. Chan, K. W., Tan G. H. and Wong R. C. S. (2013), Statistical validation for the profiling of heroin by associating simulated post-cut samples with the

corresponding pre-cut sample, *Journal of Forensic Sciences*, 58, S199-S207.
4. O'Neil, P. J., Baker P. and Gough T. (1984), Illicitly imported heroin products: Some physical and chemical features indicative of their origin, *Journal of Forensic Sciences*, 29(3), 889-902.
5. Sibley, J. A. (1996), Formation of O^6acetylmorphine into the "homebake" preparation of heroin, *Forensic Science International*, 77, 159-167.
6. Moore, J. M. and Klein, M. (1978), Identification of O^3-monoacetylmorphine in illicit heroin using gas chromatography-electron-capture detection and mass spectrometry, *Journal of Chromatography* 154, 76-83.
7. Narayanaswami, K. (1985), Parameters for determining the origin of illicit heroin samples. *Bulletin on Narcotics* XXXVII(1), 49–62.
8. Kaa, E. and Bent K. (1986), Impurities, adulterants and diluents of illicit heroin in Denmark (Jutland and Funen). *Forensic Science International*, 31, 195–210.
9. Zhang, D., Shi, X., Yuan, Z. and Ju, H. (2004), Component analysis of illicit heroin samples with GC/MS and its application in source identification. *Journal of Forensic Sciences*, 49(1), 81–86.
10. Anastos, N., Lewis, S. W., Barnett, N. W., Pearson, J. R. and Kirkbride, K. P. (2005), The rapid analysis of heroin drug seizures using micellar electrokinetic chromatography with short-end injection. *Journal of Forensic Sciences*, 50(1), 37–42.
11. Dufey, V., Dujourdy, L., Besacier, F. and Chaudron, H. (2007), A quick and automated method for profiling heroin samples for tactical intelligence purposes. *Forensic Science International*, 168, 108–117.
12. Esseiva, P., Dujourdy, L., Anglada, F., Taroni, F. and Margot, P. (2003), A methodology for illicit heroin seizures comparison in a drug intelligence perspective using large databases, *Forensic Science International*, 132, 139-152.
13. Lociciro, S., Esseiva, P., Hayoz, P., Dujourdy, L., Besacier, F. and Margot, P. (2008), Cacaine profiling for strategic intelligence, a cross-border project between France and Switzerland Part II. Validation of the statistical methodology for the profiling of cocaine. *Forensic Science International*, 177, 199-206.
14. Chan, K. W., Tan, G. H., Wong, R. C. S. (2012) Gas chromatographic method validation for the analysis of major components in illicit heroin seized in Malaysia. *Science & Justice* 52(1), 9–16.

Additional information and reprint request:

Kar-Weng Chan
Email:chankarweng@yahoo.com
chankw@kimia.gov.my
Narcotics Section
Department of Chemistry Malaysia,
Jalan Sultan,
46661 Petaling Jaya, Selangor