D. Comparative Analysis

On overall, MIR 1, 2 and 3 produced modelling accuracy ranged between 85-70%, with the former showed the highest mean accuracy at 80%. The distinguished difference between the three batches of SOM models are consistency accuracy over the six test sets with MIR 2 is the only one gave most robust models despite not able to produce the highest accuracy.

Based on a previous study [21], we noticed that PC 3 and 5 calculated from MIR 1 are the most discriminating variables for the three paper varieties. Therefore, we also built a series of SOM models using PC 3 and 5 constructed from untreated, mean-centered, variance scaled and autoscaled MIR 1 (Fig. 7). PCs derived from mean-centered IR spectra showed lowest accuracy contrast to PCs calculated from autoscaled IR spectra which showed over 90% accuracy. With that, we decided to use only PC 3 and 5 to build the final SOM model for data at hand.

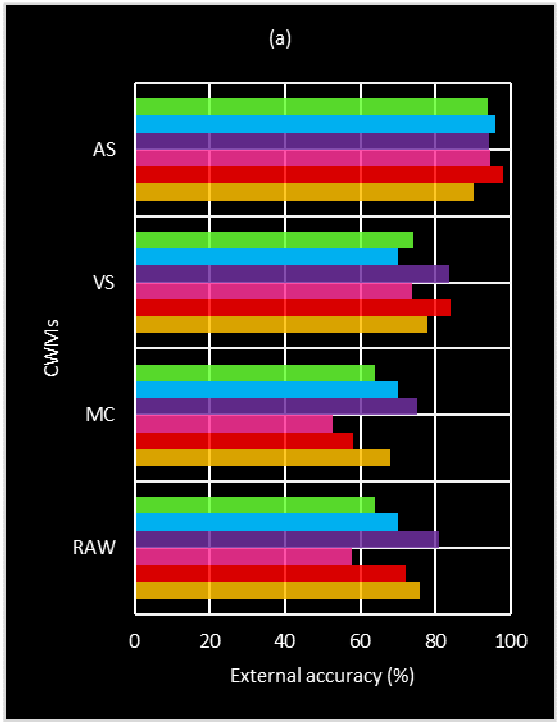

Fig. 7 External validation accuracy obtained with PC 3 and 5 calculated from raw, mean-centered, variance scaled and autoscaled MIR 1. The six test sets are respectively created with (1) Kennard-Stone, (2) Naes, (3) Select, (4) Puchwein, (5) Honigs and (6) DUPLEX.

E. SOM Model

In this study, a 3 x 4 map (i.e. in hexagonal network topology) is selected to map the 100 IR spectra of papers (i.e. the training set), and the weight given to the X map in the calculation of Tanimoto distance for updating Y is set at 0.5. This leads to the plots in Fig. 4. The network has captured most of the variability in the data as the three classes of paper are visible on the map with only two samples being misplaced. The plot indicates that all units composed of homogeneous samples except two units. IY showed the highest diversity in the patterns and is mapped with seven units.

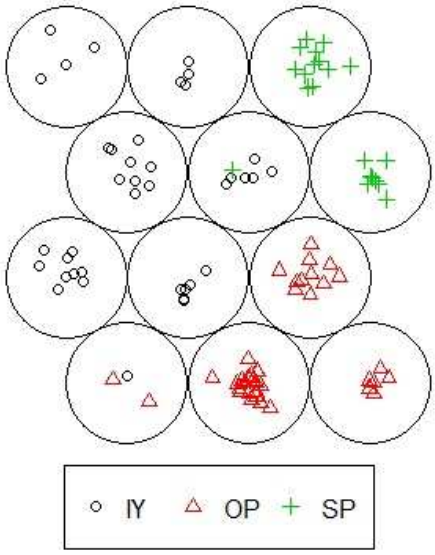

Fig. 8 Separations of 100 paper samples on a 3 x 4 topological regions

IV. Conclusion

In conclusions, the benefits of using SOM coupling with ATR-FTIR spectroscopy were demonstrated. Although it does not achieve 100% classification accuracy, major benefits of SOM is its ability to handle many complex data and is relevant to most the forensic science research that are targeting large databases in the future. Another advantage of SOMs that is relevant to forensics applications is that they can be updated easily because forensic analysis often accounted new data during investigations.

Acknowledgment

The authors would like to thank SAC Narenasagaran A/L Thangaveloo, Supt. Shaikh Abdul Adzis bin Shaikh Abdullah and Mr. Rizal of the Makmal Maktab PDRM, Cheras for their valuable help in acquiring the IR data. Financial support from the KPTM (grant no. FRGS/2/2013/ST06/UKM/02/1) is gratefully acknowledged.

References

[1] D. Ellen, *The Scientific Examination of Questioned Documents: Methods and Techniques*, London, UK: CRC Press, 1997.

[2] J. S. Kelly and B. S. Lindblom, *Scientific Examination of Questioned Documents*, Boca Raton: CRC Press, 2006.

[3] F. M. Udristioiu, I. G. Tanase, A. A. Bunaciu and H. Y. Aboul-Eneinods, "Paper analysis: non-destructive and destructive analytical methods," *Appl. Spectrosc. Rev.*, vol. 47, pp. 550-570, 2012.

[4] V. Causin, R. Casamassima, G. Marruncheddu, G. Lenzoni, G. Peluso and L. Ripani, "The discrimination potential of diffuse reflectance ultraviolet-visible-near infrared spectrophotometry for the forensic analysis of paper," *Forensic Sci. Int.* vol. 216, pp. 163-167, 2012.

[5] A. van Es, J. de Koeijer and G. van der Peijl, "Discrimination of document paper by XRF, LA-ICP-MS and IRMS using multivariate statistical techniques" *Sci. Justice.* vol. 49, pp. 120-126, 2009.

[6] M. Calcerrada and C. Garcia-Ruiz, "Review on the analysis of questioned documents," *Anal. Chimica Acta*, vol. 853, pp.143-166, 2015.

[7] C. K. Muro, K. C. Doty, J. Bueno, L. Halamkova and I. K. Lednev, "Vibrational spectroscopy: recent developments to revolutionize forensic science," *Anal. Chem.*, vol. 87, pp. 306-327, 2015.
[8] A. van Es, J. de Koeijer and G. van der Peijl, "Discrimination of document paper by XRF, LA-ICP-MS and IRMS using multivariate statistical techniques," *Sci. Justice*, vol. 49, pp.120-126, 2009.
[9] L. D. Spence, A. T. Baker and J. P. Byrne, "Characterization of document paper using elemental compositions determined by coupled plasma mass spectrometry," *J. Anal. Atom. Spectro*., vol. 15, pp. 813-819, 2000.
[10] A. Kher, S. Stewart, and M. Mulholland, "Forensic classification of paper with infrared spectroscopy and principal component analysis," *J. Near Infrared Spectrosc.*, vol. 13, pp. 225-229, 2005.
[11] A. Racz, K. Heberger and M. Fodor, "Quantitative determination and classification of energy drinks using near-infrared spectroscopy," *Anal. Bioanal. Chem.* vol. 408, pp. 6403-6411, 2016.
[12] K. Porker, M. Zerner and D. Cozzolino, "Classification and Authentication of Barley (*Hordeum vulgare*) malt varieties: Combining attenuated total reflectance mid-infrared spectroscopy with chemometrics," *Food Anal. Methods*. In. press
[13] C. A. Lima, V. P. Goulart, L. Correa, D. M. Zezell, "Using Fourier transform infrared spectroscopy to evaluate biological effects induced by photodynamic therapy," *Lasers in Surgery and Medicine.* vol. 48, pp. 538-545, 2016.
[14] G. Theophilou, K. M. G. Lima, P. L. Martin-Hirsh, H. F. Stringfellow and F. L. Martin, "ATR-FTIR spectroscopy coupled with chemometric analysis discriminates normal, borderline and malignant ovarian tissue: classifying subtypes of human cancer," *Analyst*, vol. 141, pp. 585-594, 2016.
[15] A. Yan, X. Nie, K. Wang and M. Wang, "Classification of Aurora kinase inhibitors by self-organizing map (SOM) and support vector machine (SVM)," *European Journal of Medicinal Chemistry*, vol. 61,pp. 73-83, 2013.
[16] W. N. S. Mat-Desa, D. Ismail and N. N. Daeid, "Classification and source determination of medium petroleum distillates by Chemometric and Artificial neural networks: A self-organizing feature approach," *Anal. Chem.* vol. 83, pp. 7745-7754, 2011.
[17] K. S. Mattingly, C. A. Ramseyer, J. J. Rosen, T. L. Mote and R. Muthyala, "Increasing water vapor transport to the Greenland Ice Sheet revealed using self-organizing maps," *Geophysical Research Letters*, vol. 43, pp. 9250-9258, 2016.
[18] S-G. Lee and E-Y Cha, "Style classification and visualization of art painting's genre using self-organizing maps," *Hum. Cent. Comput. Inf. Sci*. vol. 6, pp. 1-11, 2016.
[19] L. C. Lee, C-Y Liong, K. Osman, and A. A. Jemain, "Forensic differentiation of paper by ATR-FTIR spectroscopy technique and PLS-DA," in *AIP Conference Proceedings*, vol. 1750, 2016, p. 060016.
[20] C-Y Liong, L. C. Lee, K. Osman and A. A. Jemain, "Genetic algorithm for wavenumber selection in forensic differentiation of paper by LDA," in *AIP Conference Proceedings*, vol. 1750, 2016, p. 060017.
[21] L. C. Lee, C-Y Liong, K. Osman and A. A. Jemain, "Forensic classification of white paper based on ATR-FTIR spectrum and Multivariate statistical technique," in *7th Asian Forensic Sciences Network*, 2015, pp. 93.
[22] T. Kohonen, "Self-Organized Formation of Topologically correct feature maps," *Biological Cybernetics*, vol. 43, pp. 59-69, 1982.
[23] T. Kohonen, "Essentials of the self-organizing map," *Neural Networks*, vol. 37, pp. 52-65, 2013.
[24] S-L Shieh and I-En Liao, "A new approach for data clustering and visualization using self-organizing maps," *Expert Systs. Appl.* vol. 39, pp. 11924-11933, 2012.
[25] R. Bro, A. K. Smilde,"Principal component analysis," *Anal. Methods*, vol. 6, pp. 2812-2831, 2014.
[26] D. E. Honigs, G. M. Hieftje, H. L. Mark and T. B. Hirschfeld, "Unique-Sample Selection via Near-Infrared Spectral Subtraction," *Anal. Chem.*, vol. 57, pp. 2299-2303, 1985.
[27] R. W. Kennard and L. A. Stone, "Computer aided design of experiments," *Technometrics*, vol. 11, pp.137-148, 1969.
[28] T. Naes, "The design of calibration in near infra-red reflectance analysis by clustering," *J. Chemo.* vol. 1, pp. 121-134, 1987.
[29] G. Puchwein, "Selection of calibration samples for near-infrared spectrometry by factor analysis of spectra," *Anal. Chem.* vol. 60, pp. 569-573, 1988.
[30] J. S. Shenk and M. O. Westerhaus, "Population definition, sample selection and calibration procedures for Near Infrared Reflectance Spectroscopy," *Crop. Science*. vol. 31, pp. 469-474, 1991.
[31] R. D. Snee, "Validation of regression models: methods and examples," *Technometrics*, vol. 19, pp. 415-428, 1977.
[32] Wehrens, R. (2015, Sept 04). Package 'kohonen' [Online]. Available: https://cran.r-project.org/web/packages/kohonen/kohonen.pdf
[33] Kuhn, M. (2016, Apr 11). Package 'caret' [Online]. Available: https://cran.r-project.org/web/packages/caret/caret.pdf
[34] Kuhn, M. (2016, Apr 11). Package 'caret' [Online]. Available: https://cran.r-project.org/web/packages/caret/caret.pdf
[35] Stevens, A., Ramirez-Lopez, L (2015, Feb 20). Package 'prospectr' [Online]. Available: https://cran.r-project.org/web/packages/prospectr/prospectr.pdf