Determining the geographical origin of crude palm oil with the combined use of GC-IMS fingerprinting and chemometrics

K. A. Goggin[1], E. Brodrick[2], A. Wicaksono[3], J.A. Covington[3], A. N. Davies[1,4], D. J. Murphy[1*]

[1]Faculty of Computing, Engineering and Science, University of South Wales, Pontypridd, CF37 1DL (UK)

[2]IMSPEX Diagnostics Ltd, Abercynon CF45 4SN (UK)

[3]School of Engineering, University of Warwick, Warwick, CV4 7AL (UK)

[4]Nouryon, 7418 AJ Zutphenseweg 10, P.O. Box 10, 7400 AA, Deventer, The Netherlands

Email addresses: kirstie.goggin@southwales.ac.uk; emma@imspex.com; A.Wicaksono@warwick.ac.uk; J.A.Covington@warwick.ac.uk; tony.davies@nouryon.com;

*Corresponding author: E-mail- denis.murphy@southwales.ac.uk

ABSTRACT

Current administrative controls used to verify geographical provenance within palm oil supply chains require enhancement and strengthening by more robust analytical methods. In this study, the application of volatile organic compound fingerprinting, in combination with five different analytical classification models, has been used to verify the regional geographical provenance of crude palm oil samples. For this purpose, 108 crude palm oil samples were collected from two regions within Malaysia, namely Peninsular Malaysia (32) and Sabah (76). Samples were analysed by gas chromatography-ion mobility spectrometry (GC-IMS) and the five predictive models (Sparse Logistic Regression, Random Forests, Gaussian Processes, Support Vector Machines, and Artificial Neural Networks) were built and applied. Models were validated using 10-fold cross-validation. The Area Under Curve (AUC) measure was used as a summary indicator of the performance of each classifier. All models performed well (AUC $\geq$ 0.96) with the Sparse Logistic Regression model giving best performance (AUC = 0.98). This demonstrates that the verification of the geographical origin of crude palm oil is feasible by volatile organic compound fingerprinting, using GC-IMS supported by chemometric analysis.

KEYWORDS

Palm oil, fingerprinting, volatile organic compounds, GC-IMS, chemometrics, geographical origin

INTRODUCTION

Palm oil obtained from the fruit of the oil palm (*Elaeis guineensis*), is the most consumed vegetable oil globally. In 2018 it was estimated that 68 million tonnes were produced globally (Statista, 2018). The oil palm plant originates from West Africa, but now grows in wild, semi-wild and cultivated states right across the equatorial tropics, including Malaysia, Indonesia, Papua New Guinea, Western Africa and South and Central America (Corley &

Tinker, 2008). Two different oils are extracted from palm fruits, namely Crude Palm Oil (CPO) and Crude Palm Kernel Oil. CPO is the main oil of commercial interest, being semi-solid at room temperature and containing high proportions of both saturated and monounsaturated fatty acids. These properties make for a versatile oil that is predominantly used in foodstuffs as an ingredient in thousands of processed foods ranging from noodles to chocolate. Oil palm cultivation has grown rapidly in recent decades due to low production costs and high demands from the food industry, especially for CPO (Corley & Tinker, 2008; Paddison et al., 2014).

The initial rapid expansion of the oil palm industry in Malaysia occurred by conversion of land from other plantation crops, mainly rubber. However, after 2000 large tracts of primary and secondary rainforests, as well as peatlands, were also converted especially in Indonesia. This process often occurred in regions of high biodiversity and conservation value (Koh and Wilcove, 2008; E. Meijaard *et al*., 2018). As a reaction to this process, the Roundtable on Sustainable Palm Oil (RSPO) was established in 2004 to improve sustainability and traceability of the industry. However, to a great extent, current traceability methods are largely based upon potentially fallible audit trails. Therefore, it is increasingly important that there are alternative methodologies that can be applied reliably within supply chains and which enable authentication of geographical provenance, to facilitate current traceability measures.

Chemically based methods for the authentication of geographical provenance of vegetable oils has been well studied in the case of olive oil but less so for other oils (Janin *et al*., 2014; Ou *et al*., 2015; Portarena, Gavrichkova *et al*., 2014). All vegetable oils are complex natural mixtures comprising of many components. Fatty acid composition is the most studied component for vegetable oil authentication (Janin *et al*., 2014; Korifi *et al*., 2011; Tres *et al*., 2013). However, other important components can also be utilised for authentication including sterols, elemental isotope ratios, volatile organic compounds (VOCs) and tocopherols. With the exception of isotope ratios, fingerprinting techniques are the most common approach for

assessing such components, as they provide analytical information about a sample in a non-selective way (Ruiz-Samblás *et al*., 2013).

However, fingerprinting usually generates issues of 'big data' analysis, which require the use of appropriate multivariate statistics to extract the most important information for characterizing a particular sample (Cumeras *et al*., 2015; Hauschild *et al*., 2012; Szymanska *et al*., 2014). Other studies have shown that VOC fingerprints can be useful for discerning vegetable oils by geographical origin because their quality and composition depend on several factors, including genetic variety, growing conditions, processing technologies and storage (Alba Tres *et al*., 2011). In the case of palm oils, the composition may be significantly affected by seasonal variation, fertilisation regime, oil processing techniques, etc.

In this work, gas chromatography-ion mobility spectrometry (GC-IMS) was used to generate VOC fingerprints. IMS was initially developed in the 1970s for detection of explosives and chemical warfare agents. It relies upon the separation of charged particles in an electric field, with separation depending upon mass, shape, size and collisional cross-sectional area (Eiceman *et al*., 2016) of each ion cluster. However, IMS spectrometers typically have low resolution due to overlapping signals resulting from ion-ion or ion-molecule reactions in the ionisation process (Garrido-Delgado *et al*., 2011). The method is therefore often coupled with other techniques for fast pre-separation, usually a standard GC column, as is the case in this study. GC-IMS is now increasingly applied in the environmental, biomedical and food and flavour industries due to its selectivity and sensitivity, time of analysis, small footprint, low cost and its ability for easy on-site implementation by relatively unskilled operatives, meaning it is potentially accessible to laboratories worldwide.

The present work is one of only a few studies that have sought to characterise CPO samples by geographical origin. Four previous studies have sought to do this on a continental scale (South-East Asia vs. South America vs. Africa) (Obisesan *et al*., 2017; Pérez-Castaño *et*

al., 2015; Ruiz-Samblás *et al*., 2013; Tres *et al*., 2013) and one study was on a regional scale (Central Malaysia vs. Northern Malaysia vs. East coast Malaysia vs. Southern Malaysia vs. East Malaysia) (Muhammad *et al*., 2017). While, GC-IMS has previously been used to distinguish different olive oil samples by grade (Garrido-Delgado *et al*., 2015; Garrido-Delgado *et al*., 2011), to our knowledge this study is the first time GC-IMS has been used for palm oil analysis. Here we describe the application of chemometrics to raw GC-IMS chromatograms to successfully establish models for the prediction of regional geographical provenance of CPO samples in Malaysia.

MATERIALS AND METHODS

Palm oil samples

A total of 108 palm oil samples were provided by Wageningen University of Research (WUR), Netherlands. These samples had been collected from various mills across Peninsular Malaysia and the State of Sabah in North Borneo. A total of 32 samples originated from Peninsular Malaysia whilst 76 originated from Sabah. Samples were stored at 4 °C until analysis.

Sample preparation

No pre-preparation or derivatisation of samples is required prior to GC-IMS analysis. CPO samples were melted at 45 °C to enable aliquoting of 1 g to a 20 mL glass headspace vial and vials were secured with a magnetic screw cap, sealed with a PTFE/silicon septum. Samples were pre-conditioned at 60 °C and 275 rpm for 15 minutes, via an integrated sample introduction system (SIS) unit (CTC-PAL, CTC Analytics AG, Zwingen, Switzerland) to ensure equilibration between the sample and headspace. 200 µL of sample headspace was directly injected into the GC-IMS system via a 2.5 mL Hamilton syringe with a 51 mm needle.

GC-IMS analysis

All CPO analyses were performed on a commercially available GC-IMS instrument (model, FlavourSpec©) from Gesellschaft für Analytische Sensorsysteme mbH (G.A.S., Dortmund, Germany). The headspace sample was injected via a heated splitless injector on to a low polarity GC column consisting of 9%-diphenyl – 95% dimethylpolysiloxane of 15 m length, an internal diameter of 0.53 mm and 1 µm of film thickness (FS-SE-54-CB-1 of CS-Chromatographie Service GmbH, Düren, Germany) facilitated by Nitrogen (6.0) carrier gas. The analytes enter the ionisation region and undergo soft ionisation via a cascade reaction by a Tritium H^3 radioactive ionisation source of 300 MBq.

Ion swarms are released into the drift region though a Bradbury Nielsen gate (grid pulse width of 100 µs and a sampling frequency of 150 kHz) when the electric field strength of the grid set of the shutter is weakened or eliminated. Ions travel towards the detector (Faraday plate) against an opposing drift gas (Nitrogen 6.0) and are separated based on mass, charge, size and cross-sectional collision surface area, due to the presence of an electric field. Subsequently, different ions reach the detector at different times, with each component having a specific IMS drift time.

IMS data were acquired in positive mode using Laboratory Analytical Viewer (LAV) software (v.2.0.0) from G.A.S (G.A.S, 2018). Each spectrum had an average of 6 scans, obtained using a repetition rate of 30 ms. Instrumental and experimental parameters for CPO analysis are displayed in Table 1. Working principles of the FlavourSpec© are displayed in Fig. 1.

Data analysis

GC-IMS spectral data was exported into CSV format for data processing (typically 11,000,000 data points per file). The general workflow is summarised in Fig. 2 and was

developed in *R* (v 3.0.2). A number of pre-processing steps were undertaken prior to chemometric analysis. The first step cropped an area of interest, reducing data points by a factor of ten. All data were aligned in the x axis relative to the Reactant Ion Peak (RIP) position of the first file and thresholding to remove background was followed by x/y realignment, further reducing the number of data points to below 100,000. At this stage, a tenfold cross validation technique was applied. In each fold around 90% of the data was used as the training set. Within the training set, features were identified using a Wilcoxon rank-sum between the two groups (Sabah vs. Peninsular Malaysia). 100 features (data points) with the lowest *p*-values were retained and used to construct the models. This model was then applied to the remaining test set and this was repeated until each sample has a prediction as a test sample (Martinez-vernon *et al*., 2018). The five classification models used in this study are listed below:

- Sparse Logistic Regression
- Random forests
- Gaussian Processes
- Support Vector Machines
- Artificial Neural Networks

RESULTS AND DISCUSSION

Examples of the GC-IMS spectra obtained from Sabah and Peninsular CPO samples are shown in Fig. 3. GC-IMS analysis results in a three-dimensional topographic plot where the x-axis represents IMS drift time (ms), the y-axis represents GC retention time (s) and the z-axis represents peak height/intensity (V). Due to the three-dimensional nature of the data, each spectrum contains around 11 million data points making visual comparison of different samples arduous and inefficient. Furthermore, distinguishing less intense but perhaps important signals is not possible as they may not be readily apparent above the background noise. This is why the application of chemometrics was required in order to process data

automatically, to reduce dimensionality and size, and to build classification models for discerning CPO samples by geographical origin (Sabah vs. Peninsular Malaysia).

The five different classification models (Sparse Logistic Regression, Random Forests, Gaussian Processes, Support Vector Machines, Artificial Neural Networks) include both linear and non-linear methods. A single classification model was not selected for this study as the dataset was relatively small and until larger data sets can be tested, it is recommended that multiple classifier models should be used. In order to quantify the quality of classification results, several performance features were proposed as metrics. The estimation of such metrics is based upon the classifiers ability to distinguish classes correctly and to subsequently avoid classification failure (Martinez-vernon *et al*., 2018). The different quality metrics used in this paper for evaluating the classification results are shown below (Pérez-Castaño *et al*., 2015):

1. **Area under curve (AUC):** the area under the ROC (Receiver Operating Characteristic) curve is a measure of the quality of classification models that can summarise the performance of a classifier into a single metric. Its value varies between 0 and 1, although values should generally be greater than 0.5.
2. **Sensitivity:** also known as the true positive rate and measures the proportion of actual positives that are correctly identified as such. The range of values for this feature is 0 to 1.
3. **Specificity:** also known as the true negative rate and measures the proportion of actual negatives that are correctly identified as such. The range of values for this feature is between 0 and 1.
4. ***p*-value:** a measure to determine the significance of the results. $p \leq 0.05$ typically indicates strong evidence against the null hypothesis meaning the result is significant.

All five models produced strong results for discerning Sabah and Peninsular Malaysia CPO samples (AUC ≥ 0.96) meaning they could correctly distinguish between samples at least

96% of the time. However, the Sparse Logistic Regression method performed best (AUC 0.98) (Table 2 and Fig. 4). Since GC-IMS is a rapid, sensitive and selective, cost-effective and non-destructive technique, which can be readily implemented on-site, it could be proposed as an initial screening technique for the geographical origin of crude palm oil, prior to the utilisation of more costly and time-consuming targeted techniques.

Whilst the aim of this study was to assess the use of GC-IMS as a fingerprinting approach for discerning samples by origin, the pipeline used in this study allowed for feature extraction to identify significant data points involved in the classification. For example, several data points may have formed a single peak that was only present in one group of samples and not the other, therefore it might have been of interest to identify this peak using the NIST2014 database and IMS library. However, in this study, there was no correlation between specific features and individual spectral peaks. The features with the greatest variance were spread across the spectra and likely represented global changes in total profiles, making peak picking and subsequent compound identification difficult. Nevertheless, this study has shown GC-IMS combined with chemometrics to be a feasible fingerprinting approach for discerning between CPO samples from Sabah and Peninsular Malaysia. Further work should be conducted on a larger sample size to increase the likelihood of detection of a specific geographical marker using feature extraction, followed by compound identification using NIST2014 and IMS databases.

Sample and group size have been major limitations in all previously published studies in this area. Our study is only one of two which has successfully discerned CPO samples by region of origin and is the first to do so using a fingerprinting approach combined with chemometrics, on a much larger sample set. Nevertheless, even in this work, sample set is still a limitation because recommended minimum group size for chemometric analysis to be statistically significant is 30, which is close to that of the Peninsular Malaysia group (n=32).

Furthermore, due to the availability of samples, group size in this study is not optimally balanced, meaning bias may be introduced. Any further work should be conducted on larger and better balanced groups. Nonetheless, this study has demonstrated promising results using the provided sample set and has shown that increased spatial specificity can be obtained. Further work should be conducted using CPO from the same mills/regions, but further studies should also ensure that samples are collected and analysed continuously over long term periods of several months to years in order to capture as much variation as possible. In this way it will be possible to successfully validate such approach and train predictive models more effectively.

Analytical methods for verification of geographical provenance of palm oils will have positive implications within the industry and will support and strengthen the current administrative controls in place. Whilst VOC fingerprinting is a well-studied approach and has been successfully used for other vegetable oils, further work should be undertaken annually, using as many authentic samples as possible, to assess the impact of seasonal variation, changes in fertilisation regime, changes in processing etc.

CONCLUSIONS

Fingerprinting approaches combined with use of appropriate multivariate statistics (chemometrics) is common practice for authentication of foodstuffs. However, this is the first study of its kind that has shown that the application of chemometrics to raw chromatograms of GC-IMS data, is effective for discerning CPO samples by regional geographical provenance.

A single classification model was not selected for this study as the dataset is relatively small, alternatively five different models (linear and non-linear) were used and should continue to be used until a large enough dataset has been analysed. All models were successful in discerning CPO samples from Sabah and Peninsular Malaysia. Since GC-IMS is a rapid, sensitive and selective, cost-effective and non-destructive technique, which can be readily implemented on-site, it could be proposed as an initial screening technique for the geographical

origin of crude palm oil, prior to the utilisation of more costly and time-consuming targeted techniques.

Such analytical methods for verifying the geographical provenance of palm oils will have positive implications within the industry and will support and strengthen the administrative controls currently in place (Goggin & Murphy, 2018). This is only one of a few which have sought to distinguish CPOs by geographical origin and is only the second to do so on a regional level and the only one using GC-IMS. Whilst VOC fingerprinting is a well-studied approach and has been successfully used for other vegetable oils, further work should be undertaken annually to assess the impact of seasonal variation, changes in fertilisation regime, changes in processing etc. A larger sample set should also be studied to determine whether further spatial specificity can be obtained (i.e. at mill or plantation level).

ACKNOWLEDGEMENTS

This work was partially funded by the European Social Fund through the Welsh Government under a KESS2 studentship, University of South Wales and IMSPEX Diagnostics Ltd., awarded to K. A Goggin (Grant number: MAXI 20539). The authors are very grateful to Professor Saskia van Ruth at Wageningen University of Research (WUR) for providing the crude palm oil samples for the study.

REFERENCES

Corley, R. H. V., & Tinker, P. B. H. (2008). *The Oil Palm* (4th ed.). Wiley-Blackwell.

Cumeras, R., Figueras, E., Davis, C. E., Baumbach, J. I., & Gràcia, I. (2015). Review on Ion Mobility Spectrometry. Part 2: hyphenated methods and effects of experimental parameters. *The Analyst*, *140*, 1391–1410. https://doi.org/https://doi.org/10.1039/C4AN01101E

Eiceman, G. A., Karpas, Z., & Hill, H. H. (2016). *Ion Mobility Spectrometry* (3rd ed.). Boca

Raton: CRC Press.

G.A.S. (2018). *L.A.V v2.0.0*.

Garrido-Delgado, R., Dobao-Prieto, M. D. M., Arce, L., & Valcárcel, M. (2015). Determination of volatile compounds by GC-IMS to assign the quality of virgin olive oil. *Food Chemistry*, *187*, 572–579. https://doi.org/10.1016/j.foodchem.2015.04.082

Garrido-Delgado, R., Mercader-Trejo, F., Sielemann, S., de Bruyn, W., Arce, L., & Valcárcel, M. (2011). Direct classification of olive oils by using two types of ion mobility spectrometers. *Analytica Chimica Acta*, *696*(1–2), 108–115. https://doi.org/10.1016/j.aca.2011.03.007

Goggin, K., & Murphy, D. J. (2018). Monitoring the traceability, safety and authenticity of palm oils imported in Europe. *Oilseeds and Fats, Crops and Lipids*, *25*(6), 1–15. https://doi.org/https://doi.org/10.1051/ocl/2018059

Hauschild, A., Schneider, T., Pauling, J., Rupp, K., Jang, M., Baumbach, J. I., & Baumbach, J. (2012). Computational Methods for Metabolomic Data Analysis of Ion Mobility Spectrometry Data—Reviewing the State of the Art, 733–755. https://doi.org/10.3390/metabo2040733

Janin, M., Medini, S., & Técher, I. (2014). Methods for PDO olive oils traceability: state of art and discussion about the possible contribution of strontium isotopic tool. *European Food Research and Technology*, *239*(5), 745–754. https://doi.org/10.1007/s00217-014-2279-8

Koh, L. P., & Wilcove, D. S. (2008). Is oil palm agriculture really destroying tropical biodiversity? *Policy Perspective*, *1*, 60–64. https://doi.org/https://doi.org/10.1111/j.1755-263X.2008.00011.x

Korifi, R., Le Dréau, Y., Molinet, J., Artaud, J., & Dupuy, N. (2011). Composition and authentication of virgin olive oil from French PDO regions by chemometric treatment of Raman spectra. *Journal of Raman Spectroscopy*, *42*(7), 1540–1547.

https://doi.org/10.1002/jrs.2891

Martinez-vernon, S., Covington, J. A., Arasaradnam, R. P., Id, S. E., Connell, N. O., Kyrou, I., & Savage, R. S. (2018). An improved machine learning pipeline for urinary volatiles disease detection : Diagnosing diabetes, 1–20. https://doi.org/10.1371/journal.pone.0204425

Meijaard, E., Garcia-Ulloa, J., Sheil, D., Wich, S. A., Carlson, K. M., Juffe-Bignoli, D., & Brooks, T. M. (2018). *Oil Palm and Biodiversity. A situation analysis by the IUCN Oil Palm Task Force*.

Meijaard, Erik, & International, N. C. (2013). *Oil-Palm Plantations in the Context of Biodiversity Conservation*. *Encyclopedia of Biodiversity* (Vol. 5). Elsevier Ltd. https://doi.org/10.1016/B978-0-12-384719-5.00340-3

Muhammad, S. A., Seow, E., Omar, A. K. M., Rodhi, A. M., Hassan, H. M., Lalung, J., … Ibrahim, B. (2017). Variation of δ2H, δ18O & δ13C in crude palm oil from different regions in Malaysia: Potential of stable isotope signatures as a key traceability parameter. *Science and Justice*, *58*(1), 59–66.

Obisesan, K. A., Jiménez-Carvelo, A. M., Cuadros-Rodríguez, L., Ruisanchez, I., & Callao, M. P. (2017). HPLC-UV and HPLC-CAD Chromatographic Data Fusion for the Authentication of the Geographical Origin of Palm Oil. *Talanta*, 413–418.

Ou, G., Hu, R., Zhang, L., Li, P., Luo, X., & Zhang, Z. (2015). Advanced detection methods for traceability of origin and authenticity of olive oils Analytical Methods MINIREVIEW. *Anal. Methods*, *7*(7), 5731–5739. https://doi.org/10.1039/c5ay00048c

Paddison, L., Purt, J., Moulds, J., Balch, O., Riadi, Y., & Ifansasti, U. (2014). From rainforest to cupboard: the real story of palm oil - interactive. Retrieved January 17, 2019, from https://www.theguardian.com/sustainable-business/ng-interactive/2014/nov/10/palm-oil-rainforest-cupboard-interactive

Pérez-Castaño, E., Ruiz-Samblás, C., Medina-Rodríguez, S., Quirós-Rodríguez, V., Jiménez-Carvelo, A. M., Valverde-Som, L., … Cuadros-Rodríguez, L. (2015). Comparison of different analytical classification scenarios: application for the geographical origin of edible palm oil by sterolic (NP) HPLC fingerprinting. *Anal. Methods*, *7*(10), 4192–4201. https://doi.org/10.1039/C5AY00168D

Portarena, S., Gavrichkova, O., Lauteri, M., & Brugnoli, E. (2014). Authentication and traceability of Italian extra-virgin olive oils by means of stable isotopes techniques. *Food Chemistry*, *164*, 12–16. https://doi.org/10.1016/j.foodchem.2014.04.115

Ruiz-Samblás, C., Arrebola-Pascual, C., Tres, A., Van Ruth, S., & Cuadros-Rodríguez, L. (2013). Authentication of geographical origin of palm oil by chromatographic fingerprinting of triacylglycerols and partial least square-discriminant analysis. *Talanta*, *116*, 788–793. https://doi.org/10.1016/j.talanta.2013.07.054

Statista. (2018). Consumption of vegetable oils worldwide from 2013/14 to 2017/18, by oil type (in million metric tonnes). Retrieved March 12, 2018, from https://www.statista.com/statistics/263937/vegetable-oils-global-consumption/

Szymanska, E., Brodrick, E., Williams, M., Antony, N., Manen, H. Van, & Buydens, L. M. C. (2014). Data size reduction strategy for the classification of breath and air samples using multi capillary column - ion mobility spectrometry (MCC-IMS). https://doi.org/10.1021/ac503857y

Tres, A., Ruiz-Samblas, C., Van Der Veer, G., & Van Ruth, S. M. (2013). Geographical provenance of palm oil by fatty acid and volatile compound fingerprinting techniques. *Food Chemistry*, *137*(1–4), 142–150. https://doi.org/10.1016/j.foodchem.2012.09.094

Tres, Alba, Van Der Veer, G., Alewijn, M., Kok, E., & Van Ruth, S. M. (2011). Palm Oil Authentication: Classical and State-of-the-Art Techniques. In S. A. Penna (Ed.), *Oil Palm: Cultivation, Production and Dietary Components* (pp. 1–44). New York: Nova

Science Publishers, Inc. https://doi.org/10.3213/1612-1651-10072

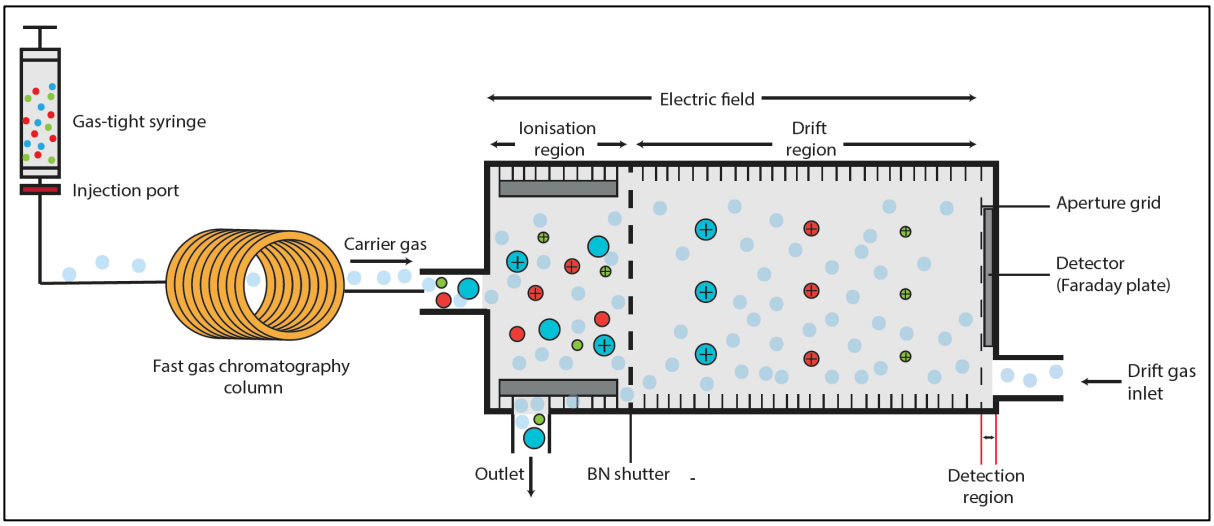

Figure 1. A diagrammatic overview of the working principles of GC-IMS.

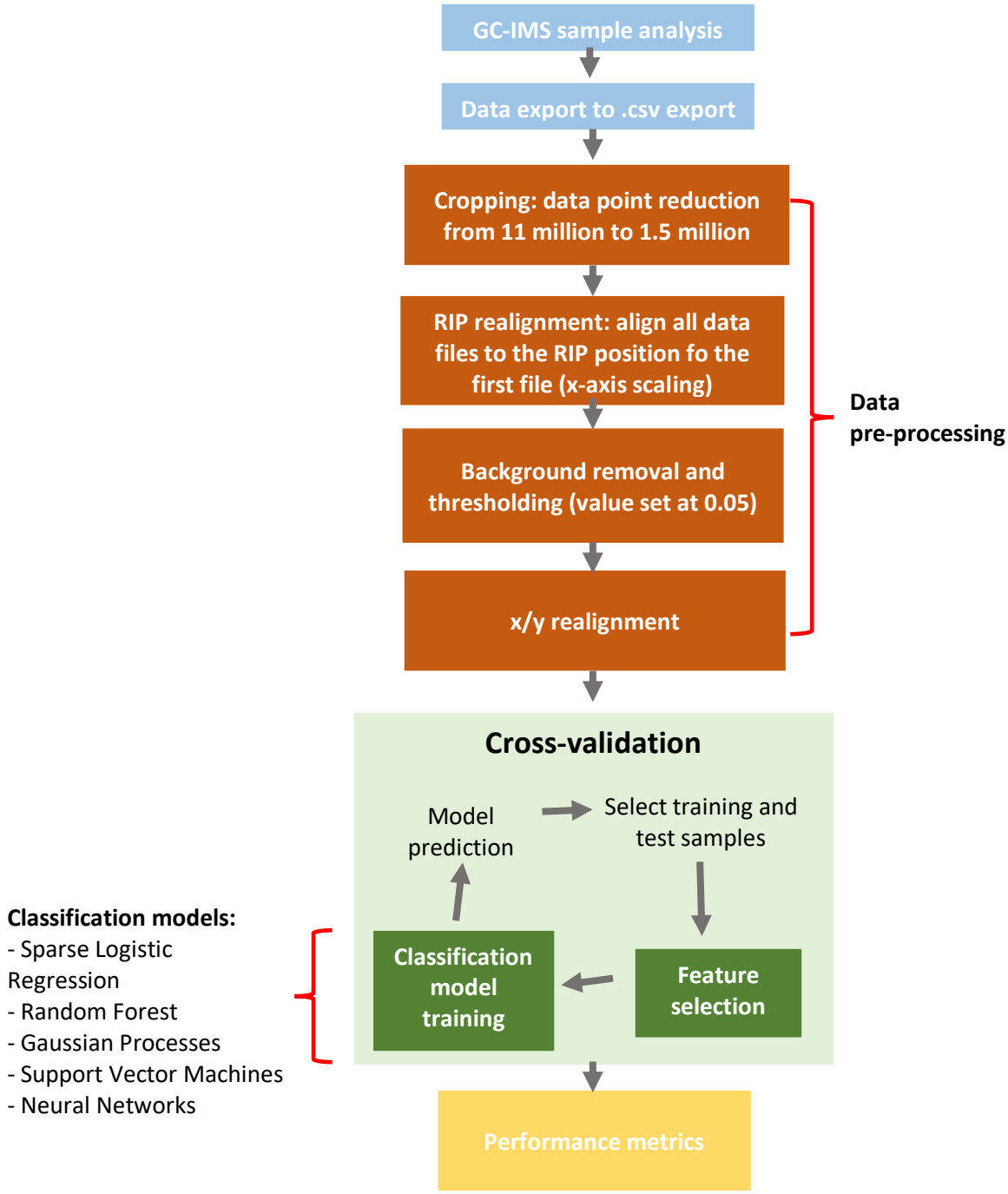

Figure 2. An overview of the general workflow used for classifying GC-IMS data into Sabah or Peninsular Malaysia classes.

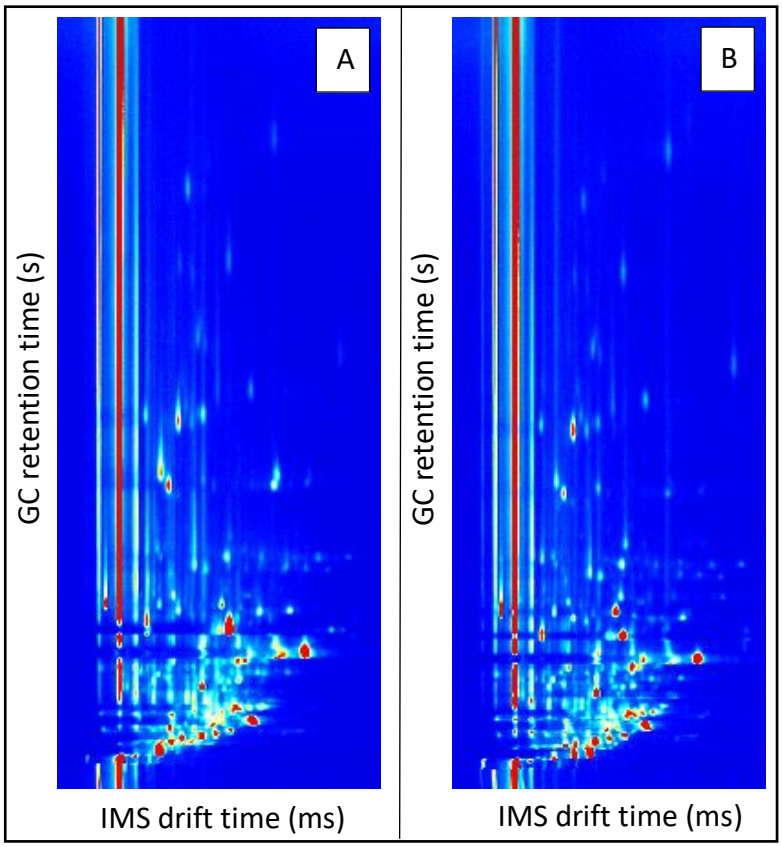

Figure 3. A side-by-side comparison of typical GC-IMS spectra from Sabah (A) and Peninsular Malaysia (B) crude palm oil samples.

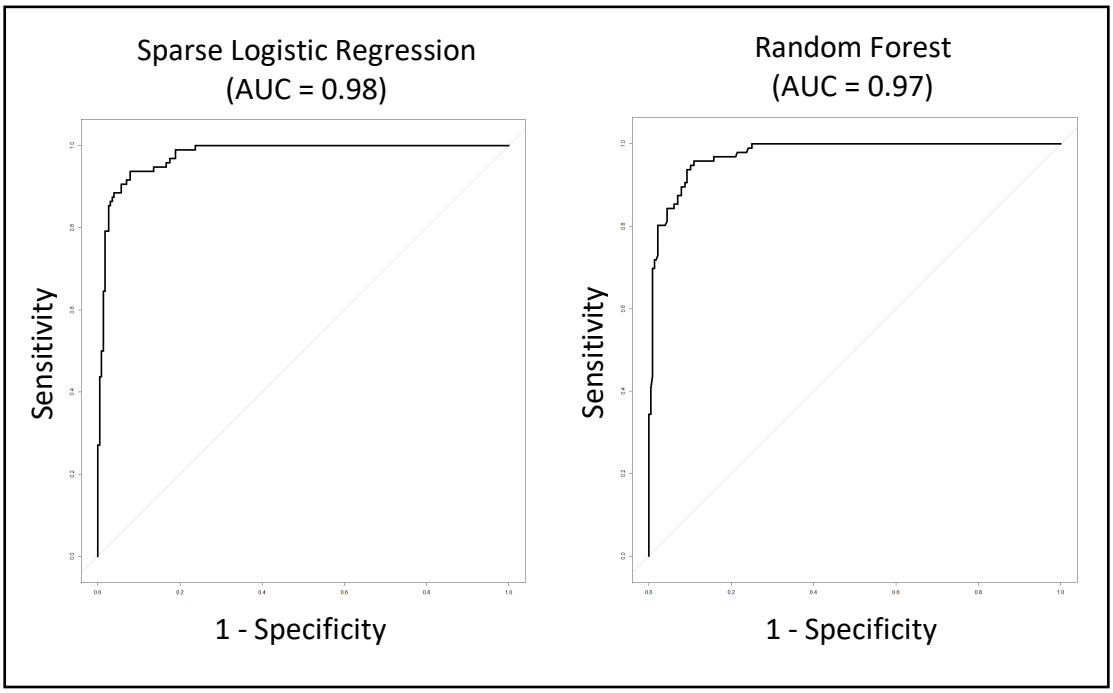

Figure 4. Examples of the ROC curves, summarising the performance of each model used in study

Table 1. Instrumental and experimental parameters for CPO analysis

Parameter	Values and units
SIS	
Sampling type/volume	Headspace (200 μL)
Agitation time	15 min
Incubation temperature	60 °C
Syringe temperature	80 °C
Column	
Injector temperature	80 °C
Capillary Column	SE-54 (low polar) ID 0.53 mm, 1 μm
Column Length	15 m
Column Temperature	40 °C
GC Run time	16 min
Carrier gas flow rate	T= 0-10 min: 2 mL min^{-1} to 50 mL min^{-1} T= 10-15 min: 50 mL min^{-1} to 150 mL min^{-1} T= 15-16 min: 150 mL min^{-1} (N_2 6.0)
IMS	
Ionization source	Tritium (30 MBq)
Voltage	Positive drift
Drift length	9.8 cm
Electric field strength	510 V cm^{-1}
Drift gas flow rate	150 mL min^{-1}
IMS temperature	45 °C

Table 2. Model performance comparison for 100 features.

Model	AUC	Sensitivity	Specificity	*p*-value
Sparse Logistic Regression	0.98 (95% CI: 0.96-0.99)	0.94 (95% CI: 0.87-0.98)	0.92 (95% CI: 0.88-0.95)	<0.01
Random Forest	0.97 (95% CI: 0.96-0.99)	0.96 (95% CI: 0.90-0.99)	0.84 (95% CI: 0.84-0.93)	<0.01
Gaussian Process	0.96 95% (CI: 0.95-0.98)	0.91 (95% CI: 0.83-0.96)	0.91 (95% CI: 0.87-0.95)	<0.01
Support Vector Machine	0.96 (95% CI: 0.93-0.99)	0.95 (95% CI: 0.88-0.98)	0.95 (95% CI: 0.92-0.98)	<0.01
Neural Net	0.97 (95% CI: 0.96-0.99)	0.95 (95% CI: 0.88-0.98)	0.95 (CI: 0.92-0.98)	<0.01