Pattern Recognition of Melaleuca cajuputi Powell Essential Oils Fingerprint in Terengganu, Malaysia Using Chemometric Approach

Mohd Hanif Abdullah1, Hafizan Juahir1*, Fathurahman Lananan1, Mohd Khairul Amri Kamarudin1, Adiana Ghazali1, Azierah Azemin1, Noor Jima Abd Wahab1, Siti Musliha Mat-Rasid1, Muhammad Hafiz Md Saad1

1East Coast Environmental Research Institute (ESERI), Universiti Sultan Zainal Abidin, Gong Badak Campus, 21300 Kuala Nerus, Terengganu, Malaysia
*Corresponding author E-mail: hafizanj@gmail.com

Abstract

Cajuputi essential oil is extracted from the leaves of Melaleuca cajuputi Powell. This study is performed to spatially classify the variation of Melaleuca cajuputi essential oil fingerprint based on different sampling location using chemometric technique along Terengganu coastal area. Discriminant Analysis (DA) successfully discriminate 10 fingerprint of essential oil into three different groups with three significant peaks in FTIR analysis. Hierarchical agglomerative cluster analysis (HACA) successfully grouped the 10 sampling stations into three groups (cluster A, B and C). Classification criteria is based on the intensity movement of functional group either bending or stretching of the essential oil compound Multiple linear regression (MLR) was used to develop an equation model that explains the prediction of species fingerprint in each cluster by different locations.

Keywords: Essential oils; fingerprint spectroscopy; chemometric technique; multiple linear regression (MLR).

1. Introduction

Melaleuca cajuputi Powell, a tree species inside the family "Myrtaceae", is a fundamental natural tree in east coast of Malaysia especially in the state of Terengganu. This genus mainly grows at the hot humid climatic zone, such as Malaysia. Mean yearly rainfall ranges from 600 to 4000 mm with a solid monsoonal design of up to months normally suitable for the development of this species. This species grows in a widespread variety of conditions yet most trees are found on low swampy coastal plains often on heavy-textured black soils that are subject to flooding [1]. While it is tolerant to and can exist alongside harsh and saline water [2] it prospers best in peat lands, waterlogged lowlands and other flooded areas, particularly swamps area. While drier surroundings are not a barrier to growth, Melaleuca cajuputi Powell have a tendency to advance evolve stunted and adopt curved structures. It had been discovered that Melaleuca cajuputi Powell trees also have been found at heights in abundance of 400 m above ocean level. While exceptional to certain geographic conditions, this occurrence of the species frequently has a decent possibility of survival under new conditions [3].

Cajuputi (likewise spelt 'cajuput' or 'cajeput') oil, or minyak kayu putih as it is brought in Indonesia and Malaysia, is gotten from the leaves of Melaleuca cajuputi and some firmly related species by steam refining. Melaleuca cajuputi can adapt to unfavourable local conditions better to different species, particularly in places where flooding is profound, where it will become established more readily than in shallow flooded areas. Therefore, it becomes exceptionally well, with high growth and yield, in the peat swamp area where most of the agricultural crops and forest trees cannot grow successfully. Malaysian verbalization 'cajuputi' is a result of the colour of the bark of the trees, meaning 'white wood'. Melaleuca cajuputi is also native to peninsular Malaysia [4, 5].

There are two major interconnected Melaleuca cajuputi forest in Terengganu which is mangrove and wetland. Both of them are essential ecology that support huge biodiversity for socio-economic activities and play major roles in maintaining ecological function [6-8]. Melaleuca cajuputi is a multipurpose tree that each part is usable and the local individuals have been acknowledging its usefulness for quite a while. Stems of Melaleuca cajuputi are utilized for structural post, fuel wood, charcoal production, fence, platform, fishing rod, agricultural pole and stake. Wood is utilized for construction and bark is for siding, roofing, and pontoon fixing material and so on. When the leaves of Melaleuca cajuputi were scrapped from the tree, it formed a “tea-tree smell” and that is the sources of cajuputi oil, a sort of essential oil [9]. Cajuputi is in all likelihood a defilement of the Indonesian name for the tree, kayu putih (kayu for wood and putih for white). It is fundamental oil utilized range of personal hygiene products as well as shampoo, fragrance and skin care, including skincare products. Besides that it also performs well in antioxidant [11], active cytotoxicity effect [12] and antibacterial [13, 14]. Before this, it has been shown that cajuputi oil is a strong disinfectant against micro bacteria organisms [15]. Furthermore, it and can be used as a mosquito and termite obstacle [16].

Melaleuca cajuputi remains generally common all through its wide geographic range and have genetic variation in many traits, given the range of environments where the species is found and large variation in morphology and oil characteristics. Careful selection of all well-adapted and rapid growing seed sources of the high-cineole...
2. Methodology

2.1. Sample Collection

Melaleuca cajuputi is a moderately fast growing tropical tree adapted to both waterlogged and well-drained soils. Mostly in Malaysia and Indonesia, it occurs in and bordering freshwater and brackish swamps, and on Ceram and Buru Islands (Indonesia) it occurs extensively in hilly terrain on dry, rocky and infertile soils. In southern Irian Jaya and Papua New Guinea it is found commonly in areas of seasonal flooding, but attains its best development on slightly elevated terrain with better drainage. Determination of this study sampling point is based on abundance of the peat swamp forest behind beaches and mangroves data using GIS and its favourite place to growth at the peat swamp area in Terengganu Coastal Area. Melaleuca cajuputi forest, even though in a small proportion in large areas, it is found abundance growing behind the mangroves in the north eastern part of Terengganu, a state that resides in the East Coast of Peninsular Malaysia [19]. 10 different location of Melaleuca cajuputi trees with 10 different villages had being identified along the coastal area of Terengganu and fresh leaves was being collected for extraction process inside the laboratory. Locations of all the fresh leaves being collected were at Kg. Gong Badak, Kg. Merabang Panjang, Kg. Beris Tok Ku, Kg. Mengrang, Kg. Pulai Baru, trees near to Merang River, Kg. Lembah Bidong, Kg. Telaga Papan, Kg. Sekeping and also Kg. Rhu Tapai. All those locations were located along Terengganu coastal area as seen in Figure 1.

![Fig. 1: Sampling Area of Melaleuca cajuputi Powell fresh leaves](image)

2.2. Sample Extraction

Steam distillation has been the extraction technique for Melaleuca cajuputi since the oil was first produced and the market accepts steam distilled oil as a standard normal tea tree oil. Steam distillation is an appropriate method of extraction for Melaleuca cajuputi Powell because of its nature that does neither altered nor changes the original composition of the essential oils. Cajuputi essential oils are volatile oils, with aromatic and therapeutic smell. Oil produce by another technique might be off to some extent have different chemical composition and therefore might not be acknowledged by the market as normal tea tree oil [20]. The yield of the essential oil normally ranges from 0.3% to 0.6% for each extraction. Samples from 10 stations were extracted from fresh leaves by steam distillation for period of 4 hours for each extraction [21]. By using steam at atmospheric pressure, the oil will be distilled from the plant material at a temperature slightly below that of boiling water. Due to Boyle’s Law the vapour pressure of a two or more phase liquid mixture in which the phases are not miscible, is equal to the sum of the partial pressures of the phases [22]. Extraction constantly for four hours has enabled to segregate and generate essential oil from subjected samples, made up with a Clevenger type apparatuses which were comprised with 3000ml round bottomed distillation flask, oil separator tube and condenser. Fresh leaves collected were grounded into small particles to maximize the essential oil production. Fresh leaves samples weight was recorded and then transferred into the distillation flask. Samples then mixed with distilled water and heated at 100˚ C. The samples were allowed to boil slowly until distillation process was completed. The mixture of oil and water was allowed to settle for four hours to ensure separation of water and oil layers. The water was slowly drawn off and the remained oil layer were slowly collected into a glass beaker and dried over anhydrous sodium sulphate. The remaining oil was pipette out into an amber colour bottle and final volume of essential oil was recorded and kept at 4 to 5 °C [23].

2.3. Attenuated Total Reflectance-Fourier Transform Infrared Spectroscopy (ATR-FTIR)

The FTIR spectra were obtained and recorded using a Perkin Elmer spectrum 400 infrared (IR) spectrophotometers coupled with air-cooled deuterated triglycine sulfate. The IR measurements were made at a resolution of 4 cm⁻¹, and 16 interferograms were co-added before the Fourier transformation. The background spectrum spectra were recorded prior to analysis of the samples. The Attenuated Total Reflectance (ATR) scan technique was used directly for all samples. The FD crude extract were analyzed by single-bounce Attenuated Total Reflection (ATR-FT-IR). The sticky and powdered extracts of FD were placed direct on the diamond crystal for data acquisition. The data were recorded at the mid-IR range of 4000-600 cm⁻¹ [24]. Spectroscopic information was gotten from ATR-FTIR analysis spectra information as it was a basic, fast and precise spectroscopy method that has been connected for multivariate analysis for plant extraction in deciding the vibration functional groups and polar compounds. Instrumental noise region of spectra (400-539 cm⁻¹) was then cut off based on the numerical data generated by the noise. As for plant studies, only near-IR and mid-IR were widely used.

In this study, FT-MIR (mid-IR) was employed for samples analysis. The spectra were baseline corrected by auto correction option using spectrum software in order to minimize the differences between spectra due to baseline shift. Later, the spectra files were saved and exported as a spectrum. SP file and were imported into The Unscrambler X 10.1 (CAMO, Trondheim, Norway) where the wave number of total spectra was transformed into mathematical data in transmittance and reflectance reading. The datasets of numerical were normalized using maximum normalization option in The Unscrambler X 10.1 (CAMO, Trondheim, Norway) software, in order to set the normality. The normalized datasets were then rotated using spectroscopic option in The Unscrambler X software on purpose to alter the transmittance and reflectance to absorbance reading.

2.4. Statistical Analysis
Chemometric analysis applied in this study was consisting of the pattern recognition technique for multivariate analysis. The Discriminant Analysis (DA) and Hierarchical Cluster Analysis (HACA) followed by data modelling which is Multi Linear Regression was performed towards spectroscopic data. In this study, chemometric analysis was performed on spectroscopic data in order to assess the 10 spatial variations of Melaleuca cajuputi Powell essential oil variations. Spectroscopy techniques have been applied for phytochemical identification and provide the significant information regarding the qualitative and quantitative essential oil composition as well as their pattern recognition by chemometrics [25, 26]. DA, CA and MLR were performed using XLSTAT 2012.

2.4.1. Discriminant Analysis

Discriminant analysis determines the variables that discriminate between two or more naturally occurring groups/clusters. It constructs a discriminant function (DF) for each group [27]. DFs are calculated using Equation (1):

\[ f(G_i) = k_i \sum_{j=1}^{n} w_{ij} P_{ij} \]  

(1)

where \( i \) is the number of groups (G), \( k_i \) is the constant inherent to each group, \( n \) is the number of parameters used to classify a set of data into a given group, and \( w_{ij} \) is the weight coefficient assigned by DF analysis (DFA) to a given parameter (pj) [28].

In this study, DA was applied to resolve whether the groups differ with regard to the mean of a variable and to use that variable to predict group membership. The DA was applied to the raw data using the standard, forward stepwise and backward stepwise modes. These were used to construct DFs to evaluate spatial variations of the essential oils fingerprint data. The stations (spatial) were the grouping (dependent) variables, while the entire measured functional group peaks constitute the independent variables. In the forward stepwise mode, variables are included step by step beginning with the most significant variable until no significant changes were obtained. In the backward stepwise mode, variables are removed step by step beginning with the less significant variable until no significant changes were obtained [29].

2.4.2. Cluster Analysis

Hierarchical clustering analysis (HACA) method was the common approach method used [30, 31] and was the most popular clustering technique in the quality evaluation of medicinal plants [32]. Cluster analysis is a technique used to group the objects into clusters so that the objects (sampling stations) within a cluster are similar to each other while other objects located in other clusters are different characteristic from each other [33]. The similarity between two classes of this method can be expressed as following Equation (2):

\[ S_{ij} = 1 - \left( \frac{d_{ij}}{d_{max}} \right) \]  

(2)

where \( S_{ij} \): similarity between two objects/ clusters, \( d_{ij} \): distance between the two objects/ clusters \( d_{max} \): maximum distance between the two objects/ clusters.

Ward’s method using Euclidean distances as a measure of similarity [34-36] within HACA has proved to be a very efficient method. The result is illustrated by a dendogram, presenting the clusters and their proximity. The Euclidean distance (linkage distance) is reported as Dlink/Dmax, which represents the quotient between the linkages distances divided by the maximal distance. The quotient is usually multiplied by 100 as a way to standardize the linkage distance represented by the y-axis [37, 38].

2.4.3. Multi Linear Regression

Spectroscopy procedures have been connected with phytochemical recognition. It gives the significant data with respect to the qualitative and quantitative essential oil chemical fingerprint on top of their pattern recognition by chemometric analysis [25, 26]. Multiple Linear Regressions (MLR) multiple linear regression (MLR) is used in to predict the fingerprint data of the species based on the spatial location [33]. This technique was been used for investigating the relationship among various independent (spatial station) and dependent variables (fingerprint spectroscopy of essential oil) by fitting a linear equation to observed data [39, 40]. In this study, it was used to justify the relationship between the fingerprinting of the essential oil based on the spatial variations. The general multiple linear regression model has \( k \) independent variables and there are \( n \) observations. Thus the regression model can be written as in Equation (3).

\[ Y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \epsilon_i \]  

(3)

where \( i = 1, ..., n \) and \( \beta \) are the regression coefficients, \( x_1 \) means independent variables and \( \epsilon \) is the error associated with the regression. Coefficient of determination, \( R^2 \), adjusted \( R^2 \) and root mean square error (RMSE) are the values that need to be considered in the best fitting regression linear equation. \( R^2 \) is the fundamental measurement of the goodness of the fit of a linear model and is defined as the proportion of the variability in the dependent variable which is accounted for by the regression equation [41]. \( R^2 \) is calculated for all possible subset models. By using this technique, the model with the largest \( R^2 \) declared the best linear model [42].

3. Results and Discussion

The variation of spectroscopic fingerprint data of essential oil from 10 sampling locations in Terengganu coastal area can be seen as Figure 2. Through the general observation, 10 different locations of the same species appeared to have variations in their spectroscopic fingerprint data. It shows that even though all the essential oils were extracted from the same species of plant, still, essential oil fingerprint data shows variations in different locations of the sampling area. It can be observed that the mean of spectroscopic data in Kg. Gong Badak with Kg. Beris Tok Ku and Kg. Merabang Panjang have possibility value that is greater than 0.05. It shows that mean of two villages does not have any significant different with confidence interval of 95% to Kg. Gong Badak. As for mean spectroscopic data of Kg. Merang, Kg. Pulai Baru, Kg. Merang and Kg. Gong Badak with Kg. Beris Tok Ku and Kg. Gong Badak. As for mean spectroscopic data of Kg. Merang, Kg. Pulai Baru, Kg. Merang and Kg. Gong Badak also show that p-value greater than 0.05 with value of p=0.006 shows mean data of all the villagers mention does not have any significant different with each other. In addition, four essential oil fingerprint data from Kg. Telaga Papan, Kg. Sekeping, Kg. Rhu Tapai and Kg. Beris Tok Ku shows variations in different locations of the sampling area.

Fig. 2: Spectroscopic fingerprint data distribution

3.1. Discriminant Analysis
In order to study the spatial variation among the different sampling locations, DA was applied on the spectroscopic raw data. 10 sampling locations in Terengganu coastal area were treated as dependent variables, while the spectroscopic fingerprint data of essential oils were treated as independent variables. DA was carried out via standard, forward stepwise, and backward stepwise methods. The accuracy of spatial classification using standard, forward stepwise, and backward stepwise mode DFA all shows that 100% correct classification with standard, forward stepwise and backward stepwise mode with seven, three and seven significant peaks respectively (Table 1). By using standard and backward stepwise mode, it shows that peak number 812.0 cm-1, 1398.4 cm-1, 2981.9 cm-1, 3552.9 cm-1, 3616.5 cm-1, and 3724.5 cm-1 were significant peak that shows variations of the essential oils fingerprint. Using forward stepwise method, peak frequency number 2981.9 cm-1, 3616.5 cm-1, 3724.5 cm-1 were found to be the significant peaks that differentiate all essential oils into three different groups (Figure 3). Furthermore, Figure 3 shows that essential oils from 10 sampling locations have been groups into three different groups. Group A represent essential oil fingerprint from Kg. Telaga Papan and Kg. Sekeping, group B in Figure 3 shows fingerprint spectroscopic from Kg. Rhu Tapai, Kg. Merang, Kg. Lembah Bidong, Kg. Merang and Kg. Pulai Baru.

Group C in discriminant analysis shows that essential oils from Kg. Merabang Panjang, Kg. Gong Badak and Kg. Beris To Ku. Significant peak illustrate by discriminate analysis shows different behaviour of functional groups either by stretching or bending behaviour of carbon compound inside the essential oils. All this essential oils variations shows different pattern in terms of the spatial distribution even though all essential oils were taken from the same species of Melaleuca cajuputi Powell.

3.2. Cluster Analysis

In addition, HACA in this study was employed to investigate the grouping of the sampling sites (spatial) and add more conformation of group based on CA analysis that had being conducted. Result in this cluster analysis was illustrated by a dendogram, presenting the clusters and their similarity of each sampling location characteristic. This analysis resulted in the grouping of sampling stations into three clusters/groups (Figure 4).

3.3. Multiple Linear Regressions for Spatial Modelling of Essential Oils Fingerprints (MLR)

MLR was applied in order to develop an explicit equation model with low level of complexity, identify the percentage contribution of each essential oils fingerprint sampling locations and in the cluster or group that had being discriminated.
The proposed equations as well as the coefficient of determination \((R^2)\), adjusted coefficient of determination \((\text{Adjusted } R^2)\) and the RMSE for all clusters shown in Equation (4)-(6):

i. Cluster A
Kg. Beris Tok Ku = 7.32016+0.73678*Kg. Gong Badak+0.17087*Kg. Merabang Panjang \((R^2 = 0.978, \text{Adjusted } R^2 = 0.978, \text{RMSE} = 0.802)\) (4)

ii. Cluster B
Kg. Merang = 9.38775+0.64984*Kg. Pulai Baru+0.06338*Sg. Merang+0.06203*Kg. Lembah Bidong+0.13971*Kg. Rhu Tapai \((R^2 = 0.934, \text{Adjusted } R^2 = 0.934, \text{RMSE} = 0.919)\) (5)

iii. Cluster C
Kg. Sekeping = 17.88827+0.80740*Kg. Telaga Papan \((R^2 = 0.985, \text{Adjusted } R^2 = 0.985, \text{RMSE} = 0.476)\) (6)

3.4. Percentage Contribution of Individual Essential Oils Fingerprint in Each Cluster

MLR is a statistical technique that is used to predict the variability that exists between the dependent and independent variable (36, 40, and 41). The Multiple Linear Regression (MLR) analysis was carried out to predict the equation for the essential oil fingerprint plot for each cluster and also the percentage of the contribution of the essential oils in the cluster. In this study the dependent variable was each of the sampling cluster from cluster A, B and C while the independent variables were the rest of the sampling stations from the cluster. In this study, the standardized coefficient and the leave-one-out cross-validation method were used to model the linearity of essential oils fingerprint for each of the cluster. One of the sampling station in each cluster become dependent variable and other sampling station in each cluster become independent variable to predict linearity of spectroscopy fingerprint data of the essential oils. Figure 5 shows each contribution of the sampling station for each of the cluster. For cluster A, Kg. Beris Tok Ku become the dependent variable and Figure 5(a) shows table graph contribution of sampling station in Cluster A and essential oils from Kg. Gong Badak contribute more to the cluster compare to Kg. Merabang essential oils in the Cluster A. Furthermore, for Cluster B essential oils fingerprint, Kg. Merang essential oil fingerprint become the dependent variable to determine the \(R^2\) value and the linearity model. Figure 6.2 (a, b, c and d) shows the linearity model of actual and predicted fingerprint plot with \(R^2\) value of 0.934 indicate the accuracy of the model with 93% correctness. Moreover, as for Cluster C, there is only 2 sampling station in cluster C. Figure 6.3 (a) with \(R^2\) value of 0.985 give 98% accuracy of the essential oils from Kg. Telaga Papan with Kg. Sekeping fingerprint plot in Cluster C.

Using the proposed equations from Equation (4)-(6), the linearity of spectroscopy fingerprint data of the essential oils can be predicted. Based on Figure 6.1 (a) and (b), Kg. Beris Tok Ku act as a dependent variable in cluster A shows the actual and predicted fingerprint plot with Kg. Gong Badak and Kg. Merabang Panjang with value of \(R^2 = 0.978\) that can give indication of 97% accuracy of the essential oil fingerprint plot from Kg. Beris To Ku with the rest of other sampling station in Cluster A. As for Cluster B linearity of spectroscopy fingerprint of the essential oils, Kg. Merang act as dependent variable to determine the \(R^2\) value and the linearity model. Figure 6.2 (a, b, c and d) shows the linearity model of actual and predicted fingerprint plot with \(R^2\) value of 0.934 indicate the accuracy of the model with 93% correctness. Moreover, as for Cluster C, there is only 2 sampling station in cluster C. Figure 6.3 (a) with \(R^2\) value of 0.985 give 98% accuracy of the essential oils from Kg. Telaga Papan with Kg. Sekeping fingerprint plot in Cluster C.

4. Conclusion

This study has vindicated the effectiveness of chemometric technique in modelling of essential oils Melaleuca cajuputi Powell from different sampling locations with the same species of plant. For spatial variations, DA gives encouraging results in discriminating the 10 sampling station with three and seven significant peaks
assigning 100% cases correctly using forward and backward stepwise mode and also successfully discriminate variability of sampling location with conformation of HACA analysis. HACA analysis also successfully groups the monitoring stations into three clusters based on their level of similarities and dissimilarities of the compound behaviour and the character of the functional groups in the essential oils. This classification helps to eliminate unnecessary sampling sites which in turn will reduce the cost and time of monitoring. The explicit equation model developed by MLR show that cluster C has the highest \( R^2 = 0.985 \), followed by cluster A with a prediction value of about 0.978% and 0.934% for cluster B.

Acknowledgement

The author would like to thank UniSZA for providing financial support for this research and the East Coast Environmental Research Institute (ESERI), UniSZA for give permission to use research facilities and supporting in this research.

References


