Feature extraction technique using Weighted Histogram Analysis Method (WHAM) for herbs discrimination based on gas chromatography signal

ABSTRACT

Herbs discrimination by investigating volatile compound using Gas Chromatography Mass Spectrometry (GCMS) is a common method adopted by botanists and scientists. Based on this common method, usually botanists and scientists would only focus on the major volatile compound in order to determine the species of the herbs. However, it is difficult to differentiate the herbs species of the same family group based on the pattern of chromatography signal since they may have almost similar physical features, characteristics, and aroma. In this case, the minor volatile compound needs to be considered in the herbs discrimination analysis. This study proposes the adoption of a Weighted Histogram Analysis Method (WHAM) that utilizes a combination histogram between two single feature histograms of peak area and peak height data in order to extract the new features based on minor and major volatile compound data (chemical properties) derived from chromatography signal patterns. From the results, it is found that WHAM technique results in better discrimination and classification between herbs species in same family group compared to the results without application of WHAM technique for feature extraction. The improvement in reducing the overlap between herbs group clustering can result in better classification as it will increase the classification accuracy.

Keyword: Weighted histogram analysis method; Gas chromatography signal; Herbs discrimination; Herbs classification; Volatile organic compounds; Feature extraction