ABSTRACT

IMPLEMENTATION OF PATTERN RECOGNITION TECHNIQUES COMBINING WITH FEATURE EXTRACTION METHODS TO DISCRIMINATE BLACK TEAS ACCORDING TO THE QUALITIES MEASURED USING ELECTRONIC NOSE

By
KOMBO OTHMAN KOMBO
17/420697/PPA/05520

An electronic nose (E-nose) using array of eight metal oxide semiconductor (MOS) gas sensors has been developed and applied. The purpose of this research was to employ E-nose with the implementation of pattern recognition techniques combined with feature extraction methods to discriminate black teas according to their qualities. In tea factory, eight batches of black teas were initially justified and discriminated by the human panel testing as quality one (Q1), quality two (Q2) and quality three (Q3), were attempted in the experiment. To reduce and transform the original data into an understandable format, three signal pre-processing methods including maximum value (F1), the fractional difference (F2), and piecewise function (F3) were applied before subjecting the data to the pattern recognition techniques for analysis. The three pre-processing methods were applied to find out the appropriate one which could provide the best results with pattern recognition techniques. Several pattern recognition techniques such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), k-Nearest Neighbor (kNN), Support Vector Machine (SVM) with linear and radial kernels (SVM-Linear and SVM-Radial, respectively), and Random Forest (RF) were used in this study to get the best performance. PCA and LDA used for clustering observation. LDA with F3 pre-processing method allowed the possibility of grouping the tea samples from eight brands according to their quality with 100% of data variability showing the clear separation of each quality. Finally, five different linear or nonlinear discrimination models, which were QDA, kNN, SVM-Linear, SVM-Radial, and RF, were compared in developing the best discrimination model. For such purpose, the data first were split into two sets. One set used for training and internal-validation, and the other for testing (external-validation). To minimize possible overfitting issues and reducing the risk of over-optimistic results, 10-fold cross-validation procedure was used. The model parameters were also optimized for the best discrimination results. The experimental results showed that the performance of SVM-Radial model was superior to the other models. By using F3 pre-processing method, the optimum SVM-Radial model was achieved with a discrimination rate of 100% in the training set, and 98% in the external validation respectively. The overall results demonstrated that E-nose system coupled with SVM-Radial supervised model could be successfully used in the discrimination of black teas with different qualities. Finally, the results proved that E-nose combined with appropriate pattern recognition technique could be used by tea manufacturing industries as a promising low-cost, fast, and easy-to-use practical device for discrimination of black teas based on the qualities.

Keywords: black tea, electronic nose, feature extractions, pattern recognition technique, human panel testing

INTISARI

IMPLEMENTASI TEKNIK PENGENALAN POLA YANG DEKOMBINASIKAN DENGAN METODE EKSTRAKSI FITUR UNTUK
DISKRIMINASI TEH HITAM SESUAI DENGAN KUALITAS YANG DIUKUR DENGAN MENGGUNAKAN HIDUNG ELEKTRONIK

Oleh
KOMBO OTHMAN KOMBO
17/420697 / PPA / 05520

Hidung elektronik (E-nose) *menggunakan array delapan sensor gas semikonduktor oksida logam* (MOS) *telah dikembangkan dan diterapkan. Tujuan dari penelitian ini adalah menggunakan* E-nose *dengan penerapan teknik pengenalan pola yang dikombinasikan dengan metode ekstraksi fitur untuk membedakan teh hitam sesuai dengan kualitasnya. Di pabrik teh, delapan batch teh hitam awalnya dibedakan oleh pengujian panel manusia sebagai kualitas satu (Q1), kualitas dua (Q2), dan kualitas tiga (Q3) digunakan dalam percobaan. Untuk mengurangi dan mengubah data asli menjadi format yang dapat dimengerti, tiga metode pra-pemrosesan sinyal termasuk nilai maksimum (F1), perbedaan fraksional (F2), dan fungsi sambungan (F3) diterapkan sebelum memasukkan data ke teknik pengenalan pola untuk analisis. Tiga metode pra-pemrosesan diterapkan untuk menemukan metode yang sesuai yang dapat memberikan hasil terbaik dengan teknik pengenalan pola. Beberapa teknik pengenalan pola seperti* Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), k-Nearest Neighbor (kNN), Support Vector Machine (SVM) with linear and radial kernels (SVM-Linear and SVM-Radial, respectively), and Random Forest (RF) *digunakan dalam penelitian ini untuk mendapatkan kinerja terbaik. PCA dan LDA digunakan untuk observasi pengelompokan. LDA dengan metode pra-pemrosesan F3 memungkinkan kemungkinan pengelompokan sampel teh dari delapan merek sesuai dengan kualitasnya dengan 100% variabilitas data yang menunjukkan pemisahan yang jelas dari setiap kualitas. Akhirnya, lima model diskriminasi linear atau nonlinear yang berbeda, yaitu QDA, kNN, SVM-Linear, SVM-Radial, dan RF, dibandingkan dalam mengembangkan model diskriminasi terbaik. Untuk tujuan tersebut, data pertama dipecah menjadi dua set. Satu set digunakan untuk pelatihan dan validasi internal, dan yang lainnya untuk pengujian (validasi eksternal). Untuk meminimalkan kemungkinan masalah overfitting dan mengurangi risiko hasil yang terlalu optimis, prosedur validasi silang 10 kali digunakan. Parameter model juga dioptimalkan untuk hasil diskriminasi terbaik. Hasil percobaan menunjukkan bahwa kinerja model SVM-Radial lebih unggul daripada model lainnya. Dengan menggunakan metode pra-pemrosesan F3, model SVM-Radial optimal dicapai dengan tingkat diskriminasi 100% dalam pelatihan, dan 98% masing-masing dalam validasi eksternal. Hasil keseluruhan menunjukkan bahwa sistem E-nose ditambah dengan model yang diawasi SVM-Radial dapat berhasil digunakan dalam diskriminasi teh hitam dengan kualitas yang berbeda. Akhirnya, hasilnya membuktikan bahwa E-nose dikombinasikan dengan teknik pengenalan pola yang tepat dapat digunakan oleh industri pembuatan teh sebagai perangkat praktis berbiaya rendah, cepat, dan mudah digunakan untuk diskriminasi teh hitam berdasarkan kualitas.*

Kata kunci: *teh hitam, hidung elektronik, ekstraksi fitur, teknik pengenalan pola, pengujian panel manusia*