Auto Classification of Biomass through Characterization of their Pyrolysis Behaviors by using Thermogravimetric Analysis with Support Vector Machine Algorithm: Case Study for Tobacco

Chao Yin
Xiamen University
Xiaohua Deng
China Tobacco Fujian Industrial Co Ltd
Zhiqiang Yu
China Tobacco Fujian Industrial Co Ltd
Ruting Chen
Xiamen University
Hongxiang Zhong
China Tobacco Fujian Industrial Co Ltd
Zechun Liu
China Tobacco Fujian Industrial Co Ltd
Guohua Cai
China Tobacco Fujian Industrial Co Ltd
Quanxing Zheng
China Tobacco Fujian Industrial Co Ltd
Xiucai Liu
China Tobacco Fujian Industrial Co Ltd
Jiawei Zhong
China Tobacco Fujian Industrial Co Ltd
Pengfei Ma
China Tobacco Fujian Industrial Co Ltd
Wei He
China Tobacco Fujian Industrial Co Ltd
Kai Lin
China Tobacco Fujian Industrial Co Ltd
Qiaoling Li (lql10684@fjtic.cn)
China Tobacco Fujian Industrial Co Ltd https://orcid.org/0000-0003-4995-3614
Anan Wu
Xiamen University

Research

Keywords: Thermogravimetric analysis, Machine learning, SVM algorithm, Tobacco

Posted Date: November 23rd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-112676/v1

Abstract

Background: During the biomass-to-bio-oil conversion process, many researches focus on the study of the association between the biomass and the bio-products by using near infrared spectra (NIR) and chemical analysis method. However, the characterization of biomass pyrolysis behaviors by using thermogravimetric analysis (TGA) with support vector machine (SVM) algorithm has not been reported. In this study, tobacco was chosen as the object for biomass, because the cigarette smoke (including water, tar and gases) released by tobacco pyrolysis reactions decide the sensory quality, which is similar to the use of biomass as a renewable resource through the pyrolysis process.

Results: Support vector machine (SVM) has been employed to automatically classify the planting area and growing position of tobacco leaves by using thermogravimetric analysis data as the information source for the first time. 88 single-grade tobacco samples belonging to 4 grades and 8 categories were split into the training, validation and blind testing set. Our model showed excellent performances in both the training and validation set as well as in the blind test, with accuracy over 91.67%. Throughout the whole dataset of 88 samples, our model not only provides precise results on the planting area of tobacco leave, but also accurately distinguishes the major grades among the upper, lower and middle positions. Error only occurs in the classification of subgrades of the middle position.

Conclusions: Our results not only validated the feasibility of using thermogravimetric analysis with SVM algorithm as an objective and rapid method for automatic classification of tobacco planting area and growing position, but also showed this new analysis method would be a promising way to exploring bio-oil quality prior to biomass pyrolysis production.

Background

Pyrolysis of biomass is a promising method to produce a variety of gases, liquids (bio-oil), or solid materials (bio-char) that can then be used for fuel production. The product compositions depend largely on the variability of different proportions of protein, triglycerides, hemicellulose, cellulose, lignin etc., in the original biomass [1, 2]. Therefore, many researches focus on the study of the association between the biomass and the bio-products [3–5][6]. In this study, tobacco was chosen as the object for biomass. As a commercial product, the cigarette smoke (including water, tar and gases) released by tobacco pyrolysis reactions can satisfy the consumer's demand, not the tobacco itself, which is similar to the use of biomass as a renewable resource through the pyrolysis process.

Tobacco leaves cultivated in different areas have different styles, and their grades are based on the positions they grow on the stalk. The classification of tobacco style and grade is great important in the processes of tobacco blend design and product maintenance of cigarette [7]. Current evaluation of tobacco style and grade mainly relies on artificial sensory analysis, which is subjective and relatively unstable [8]. It is therefore necessary and urgent in tobacco industry to develop a new rapid and convenient method to automatically evaluate the tobacco style and grade.

Artificial intelligence has opened a new page in the field of data analysis. Many efforts have been devoted to develop automatic evaluation methods by using the advance of the machine learning (ML) algorithm with the data from the tobacco leaves and smoke. Early works mainly focused on the classification of tobacco cultivation area and growing position using near infrared spectra (NIR) due to its high efficiency and non-destructive characteristic. Hana *et al.* employed artificial neural networks (ANNs) to classify whether the burley tobacco grows in the USA or outside USA, and obtained high prediction accuracy [9]. For the classification of tobacco style and grade, Ni *et al.* developed an improved and simplified K-nearest neighbors algorithm (IS-KNN) to discriminate more than 1000 Chinese flue-cured tobacco leaf samples with moderate accuracy [10]. Their results suggest that it is better to establish classification model of tobacco grade from the same cultivation fields to get better classification results. By applying a combined random-forest (CRF) based on gas chromatography (GC) fingerprinting, Lin *et al.* managed to classify three different grades of “Furong” series cigarettes with accuracy up to 93.74% [11]. Based on image processing on tobacco color, texture and shape, Zhang and Zhang implemented a two-level fuzzy comprehensive evaluation (FCE) and classified the tobacco leaves into three grades, but accuracy is achieved just 72% for the non-trained tobacco leaves [12]. Recently, Gu *et al.* successfully built a relationship between chemical compounds and the aromatic quality of flue-cured tobacco leaves, by using support vector machine (SVM) algorithm with 22 chemical compounds selected by Relief-F-particle swarm optimization (R-PSO), and obtained high accuracy of 90.95% [13].Very recently, Wang *et al.* employed genetic algorithm (GA) to optimize the performance of SVM for data analysis of NIR spectroscopy sensors. They demonstrated that the GA could indeed improve the performance of SVM for tobacco classification based on NIR spectra although the accuracy is just 83% [14]. All previous works have focused on the relationship of tobacco style and grade with either the components of the reactant (tobacco) or the product (smoke). In this study we choose to pay attention on the tobacco pyrolysis reaction process, which can be visually expressed by the thermogravimetric analysis (TGA). To the best of our knowledge, the automatic classification of tobacco planting area and growing position based on thermogravimetric analysis has not yet been reported.

TGA has been proven to be a useful tool to study the pyrolysis behavior and kinetics of pyrolysis process since it provides precise measurement depending on temperature and other experimental conditions that are well-known and well-controlled [15–17]. Investigations on biomass have shown that the differences in pyrolytic characteristics are mainly caused by the differences in the constituent and physical structure [18–20]. Studies on the pyrolysis of tobacco have also demonstrated that the tobacco pyrolysis DTG curve can be divided into different Gaussian peaks representing the thermal decomposition of individual components [21, 22]. For instance, the mass loss below 373K represents the evaporation of water [23]; The peaks between 373-473K corresponds to the thermal decomposition of sugars, nicotine, pectin and some other volatile species [24, 25]; and in the temperature of 474-873K the mass loss would be attributed to the pyrolysis of hemicellulose, cellulose and lignin, respectively [26–28]. Moreover, Baker and Bishop have demonstrated that the thermogravimetric analysis spectra of tobacco pyrolysis is highly reproducible under well-defined conditions [29]. The thermogravimetric analysis data not only represent

the tobacco pyrolysis characteristics, but also supply the information of the tobacco composition. Hence, it can be taken as an important index to evaluate tobacco planting area and growing position.

Recently, we demonstrated that thermogravimetric analysis data in conjunction with the normalized root mean square error (NRMSE) can be used to quantitatively evaluate the pyrolysis difference between tobacco of different stalk positions, planting areas and crop years [30]. On this basis, we proposed a tobacco leaves substitute scheme in tobacco blend maintenance, and the results showed that this substitute scheme could achieve artificial substitute level [31]. In this work, we further extended previous investigations and introduced SVM to the thermogravimetric analysis for the first time. Using the thermogravimetric analysis data as the information source, we demonstrated that automatic classification of tobacco planting area and growing position can be achieved with high accuracy as well as high efficiency by applying SVM. In our opinions, during the biomass-to-bio-oil conversion process, this new analysis method would be a promising way to exploring bio-oil quality prior to biomass pyrolysis production.

Results And Discussion

Classification of tobacco leaves

88 tobacco leaves were collected from different growing positions in Fujian (FJ) and Yunnan (YN) provinces, which were shown in Table 1. 88 single-grade tobacco leaves were classified into 8 categories according to their planting areas and growing positions. Three positions are identified, namely B, X and C, corresponding to the upper, lower and middle position of tobacco stalk, respectively. The middle group is further divided into 2 subgrades as shown in Table 1. The notation FJ-C1 implies that the sample is at the first grade of the middle group from the Fujian province.

Table 1
Categories of 88 single-grade tobacco leaves

Categories	Type[1,2]	Sample code
1	FJ-B	1 ~ 7
2	FJ-X	8 ~ 10
3	FJ-C1	11 ~ 20
4	FJ-C2	21 ~ 35
5	YN-B	36 ~ 44
6	YN-X	45 ~ 50
7	YN-C1	51 ~ 64
8	YN-C2	65 ~ 88
[1]. FJ represents Fujian province and YN represents Yunnan province.		
[2]. B, X and C correspond to the upper, lower and middle portion of tobacco stalk, respectively.		

Although all these samples, planted in either Fujian or Yunnan provinces, have similar tobacco style (all belonging to the same light-flavor style), they can still be distinguished in artificial sensory analysis. This leads to the most stringent test for the automatic classification of tobacco style in order to verify the effectiveness and practicability of the SVM model in the analysis of thermogravimetric analysis data.

Analysis of thermogravimetric analysis data

For a better comparison, the thermogravimetric analysis data (namely DTG curves) of tobacco leaves belonging to the same category were averaged to obtain an averaged-DTG curve, which can represent the pyrolysis characteristics of the corresponding type of tobacco leaves, as shown in Fig. 1(a) and 1(b).

Close analysis of Fig. 1(a) and 1(b) reveals that the main differences in the DTG curves of tobacco leaves from the same planting area lie in the temperature range of 373-473K, which correspond to the thermal decompositions of sugar, nicotine, pectin and some other volatile species. While in the temperature range of 473 ~ 873 K (corresponding to the pyrolysis of hemicellulose, cellulose and lignin), the DTG curves are basically coincident. Figure 1(c)-1(f) present the comparisons of DTG curves of tobacco leaves from the same growing position but from different planting areas. It is found that the main differences fall in the temperature range of 473 ~ 873 K. Hence, we may infer, from the thermogravimetric analysis spectra point of view, that the physical structure characteristics of tobacco leave (hemicellulose, cellulose and lignin reflect the tobacco physical structure) is determined by the planting area. Namely, the tobacco leaves from the same planting area have similar physical structure characteristics while the tobacco leaves from different planting areas have different physical structure characteristics. We may also draw the conclusion that the grade of tobacco leaves qualitatively depends on the proportion of sugar, nicotine, pectin and some other volatile species, in which X < B < C2 < C1.

In summary, our preliminary analysis reveals that the growing position characteristics of tobacco, which is closely related to the content of sugar, nicotine, pectin and some other volatile species, are mainly reflected in the temperature range of 373 ~ 473 K, and the planting area characteristics of tobacco determined by the tobacco physical structure are mainly reflected in the temperature range of 473 ~ 873 K. These results are in line with how the traditional classification of tobacco leaves is performed in tobacco industry, namely the grade and style are discriminated separately.

Algorithm

The above preliminary analysis has demonstrated that the thermogravimetric analysis data can reflect the planting area and growing position characteristics of tobacco leaves. To achieve automatic classification of tobacco leaves, machine learning is introduced to the analysis of the thermogravimetric analysis data.

Among numerical algorithms for machine learning, the traditional neural network algorithm requires a large amount of training data. However, due to sampling limitation, the number of samples (88) in this work cannot meet the requirements of neural networks for data training. Meanwhile, too many feature points (5890) in comparison to the number of samples (88) may also lead to dimensional disaster in neural network [32, 33]. For classification problems, the SVM algorithm [34] has been proven to be one of the best supervised learning algorithms, with faster speed and smaller sample size than other machine learning algorithms [35]. Therefore, we choose the SVM algorithm to perform automatic classification of tobacco quality and style. We would like to note that traditional SVM only supports 2 categories but our case involves 8 different categories. Hence, the one-against-one method is adopted [36].

Dataset sampling

Investigations on the generalization performance of SVM indicated that the sizes of training set, validation set and testing set are crucial for the estimated model performance [37]. Too many or too few samples in the training set may have a negative effect. Hence, it is necessary to have a good balance between the sizes of training set and validation set to have a reliable estimation of model performance. Typically, one can take around 70–80% of the data to use as a training set and split the remaining data as the validation and testing set. In this work, 88 samples were split into three sets: training set, validation set and testing set with ratio of 64/12/12, as shown in Table 2. Kennard-Stone-like algorithm [38] for data splitting was employed to maintain the generalization of the model. Namely, given n samples available in a category, the first m (with $0.6 < m/n$ and $m <= n$) samples with largest Euclidian distance in this category are used as the training set and the unselected samples are randomly split into the validation and testing set with ration of 1/1.

Table 2
The sample codes for tobacco leaves of eight categories

Categories	Type[1,2]	Sample code		
		Training set	Validation set	Testing set
1	FJ-B	1 ~ 3, 6, 7	4	5
2	FJ-X	8 ~ 10		
3	FJ-C1	11, 13 ~ 18, 20	19	12
4	FJ-C2	21, 22, 24 ~ 27, 29, 30, 32, 33, 35	23, 28,	31, 34
5	YN-B	36 ~ 39, 42 ~ 44	41	40
6	YN-X	46, 48 ~ 50	47	45
7	YN-C1	53 ~ 56, 58 ~ 61, 63, 64	51, 52	57, 62
8	YN-C2	65 ~ 68, 70, 72, 74, 76, 77, 81 ~ 87	78, 79, 80, 88	69, 71, 73, 75

1. FJ represents Fujian province and YN represents Yunnan province.

2. B, X and C correspond to the upper, lower and middle portion of tobacco stalk, respectively.

Model selection

Kernel function often plays an important role while classifying with SVM. Different kernel functions may have different application scopes. In the case where the number of feature points is much larger than the number of samples, the linear kernel has been proven to perform very well [39]. Hence, the linear kernel function was selected for training in this work. On the other hand, the penalty parameter C in the linear classification with SVM also plays a significant role in the training and prediction. Too small or too large C may have a negative effect on the prediction power of the model. To find an optimal C, the training set was used to build the model for each C and each trained model was tested with the validation set. As the samples in the validation set are not known to the model, therefore, the performance on the validation set can reflect the prediction power of the model. Based on the performance on the validation set, the optimal penalty parameter C was determined using the one with the highest accuracy. As shown in Fig. 2, the model has excellent performance in both the training and the validation set when the penalty parameter C equals to 1.66 (Log(C) = 0.22), with accuracy being 98.44% and 91.67%, respectively. Therefore, the penalty parameter C was chosen to be 1.66 in this work.

Classification accuracy

In the field of machine learning and the problem of classification with multiple categories, classification accuracy alone might be misleading. The confusion matrix can give a better idea of what the model is getting right and what types of errors it is making.

Detailed analysis of the performance of our optimal model on the training and validation set demonstrated that our optimal model performed remarkably well in the style classification, giving all correct results for the planting area, as shown in Fig. 3(a) and 3(b). In the case of growing position classification, our model also correctly identified the upper, lower and middle positions. Errors only occurs in the classification of the subgrades of middle, namely C1 and C2. For the training set, only one sample (sample code: 18) belonging to FJ-C1 was mis-assigned to FJ-C2. Similar incorrect prediction was also found in the validation set, in which the sample (sample code: 79) belonging to YN-C2 was predicted to be YN-C1 instead. As elucidated in Sect. 3, both C1 and C2 grades correspond to the middle position of tobacco stalk and the grade difference is relatively small in comparison to the grade difference between X/B and C. This might be the reason for the mis-assignment of samples in C1 and C2. We would like to note that none of previous investigations have ever tried to discriminate subgrades of the middle. Nonetheless, out optimal model showed excellent performance in both the training and validation set with overall accuracies being 98.44% and 91.67%, respectively.

Westerhuis *et al.* showed that the performance by cross-validation might be an over-optimistic one and it is of importance in having an additional blind test [40]. To verify the practicality and generalization capability of our model, we further applied the optimal model to the testing set, which is not used during the model training and selection. It is found that this model work very well in the testing set, with an overall accuracy of 91.67%. One out of 12 samples was misclassified. Detailed analysis of the confusion matrix for the testing set, as shown in Fig. 4, indicated that our model performed extremely well in the prediction of planting areas as well as in the prediction of major grades of the upper, lower and middle positions. None of the 12 samples was misclassified. Like in the training and validation set, error only occurs in the classification of subgrades of the middle while applying our model to the testing set. The sample (sample code: 34 belonging to FJ-C2 was misclassified to FJ-C1. Such a high accurate blind test indicates that our model has an excellent generalization capability. We also applied the traditional naive Bayes algorithm [41] to the same datasets. The accuracies for the training, validation and testing set were 84.38%, 58.33% and 58.33%, respectively. Obviously, in our case, the SVM algorithm is superior to the traditional naive Bayes algorithm.

It is worthwhile to note that the tobacco styles of the tobacco leaves from Fujian and Yunnan province are relatively close among the traditionally defined three major scent types, all belonging to the light flavor type. Our SVM model, based on the thermogravimetric analysis spectra, can achieve as high accuracy as 91.67% under such a stringent test, which has verified the feasibility and practicability of the automatic classification of tobacco style and grade. Unfortunately, we are unable to collect sufficient samples of other styles of tobacco at current stage. We will leave them for further investigation in the future study.

Conclusions

In this study, we conducted a thermogravimetric analysis over 88 single-grade tobacco leaves belonging to 4 grades and 8 categories. Preliminary analysis of the thermogravimetric analysis spectra reveals that

the tobacco leaves from the same planting area have similar physical structure characteristics while the tobacco leaves from different planting areas have different physical structure characteristics, which are reflected in the DTG curves in temperature range of 473-873K. Further analysis of the DTG curves also demonstrate that the growing position characteristic of tobacco leaves is mainly reflected in the temperature range of 373-473K. On this basis, we introduced the SVM algorithm to automatically classify the planting area and growing position of tobacco leave by using the thermogravimetric analysis spectra as the information source. Our SVM model shows excellent performances in both the training and validation set as well as in the blind test, with overall accuracies over 91.67%. Throughout the whole dataset of 88 samples, our model not only provides precise results on the planting areas of tobacco leaves, but also accurately distinguishes major grades of the upper, middle and lower parts of tobacco stalk. Error only occurs in the classification of the subgrades of the middle. In the blind test, the sample (sample code: 34) belonging to FJ-C2 was misclassified to FJ-C1. Such a high accuracy in the blind test indicates that our model has a very good generalization capability. Our results not only validated the feasibility of using thermogravimetric analysis with SVM algorithma as an objective and rapid method for automatic classification of tobacco style and grade, but also showed this new analysis method would be a promising way to exploring bio-oil quality prior to biomass pyrolysis production.

Methods

Materials

The tobacco samples were supplied by Fujian China Tobacco Industry Co., Ltd. For 48 hours prior to analysis, all tobacco samples were conditioned in a chamber at 22 ± 1 ^{o}C and with relative humidity of 60 ± 2%.

Thermogravimetric analysis experiment

To guarantee the reproducibility, tobacco samples were pulverized into powder using a coffee mill and then sifted through a 100-mesh sieve to remove big tobacco particles before the TGA test.

Pyrolysis of tobacco powder was performed in a TGA (STA 449 F3 TG–DTA/DSC Instruments, NETZSCH, Germany). 10 mg of tobacco powder was loaded evenly in an open ceramic pan and warmed up to 873 K from room temperature at a heating rate of 10K/min. Dry nitrogen at a flow rate of 100 mL/min was used as purge gas throughout the test. In order to reduce the influence of water, the thermogravimetric analysis data (DTG curve) of 373 ~ 873 K were selected for calculation and analysis. The number of feature points of each sample is 5890 which were obtained by recording 120 feature points per minute. The DTG curves of all 88 samples are given in the supporting information.

Declarations

Acknowledgments

The authors thank Pro. Xiaodong Chen in Soochow University for help in numerical simulation of cigarette burning process, from which we developed the idea in this study.

Authors' contributions

CY performed the research, data analysis and prepared the manuscript. XD and ZY collected the tobacco samples and performed artificial sensory analysis. RC prepared figures and cowrote the manuscript. HZ and ZL supervised the study. GC, QZ, XL, JZ, PM, WH and KL performed experiments and data analysis. QL and AW developed the idea for the study, set up the methodology and edited the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (Nos. 21773193) and the fundamental research Funds for the Central Universities (Grant No. 20720160031).

Availability of data and materials

All data generated and analyzed in this study are included in this published article.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1] Fujian Provincial Key Laboratory for Theoretical and Computational Chemistry, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, Fujian, China

[2] Technology Center, China Tobacco Fujian Industrial Co., Ltd., Xiamen 361021, Fujian, China

References

1. Jahirul MI, Rasul MG, Chowdhury AA, Ashwath N. Biofuels Production through Biomass Pyrolysis—A Technological Review. 2012;4952–5001.
2. Sharma A, Pareek V, Zhang D. Biomass pyrolysis - A review of modelling, process parameters and catalytic studies. Renew Sustain Energy Rev [Internet]. Elsevier; 2015;50:1081–96. Available from: http://dx.doi.org/10.1016/j.rser.2015.04.193
3. Lazzari E, Schena T, Marcelo MCA, Primaz CT, Silva AN, Ferrão MF, et al. Classification of biomass through their pyrolytic bio-oil composition using FTIR and PCA analysis. Ind Crops Prod [Internet]. Elsevier; 2018;111:856–64. Available from: http://dx.doi.org/10.1016/j.indcrop.2017.11.005
4. Schlund M, Scipal K, Davidson MWJ. Forest classification and impact of BIOMASS resolution on forest area and aboveground biomass estimation. Int J Appl Earth Obs Geoinf [Internet]. Elsevier B.V.; 2017;56:65–76. Available from: http://dx.doi.org/10.1016/j.jag.2016.12.001
5. Olatunji OO, Akinlabi S, Madushele N, Adedeji PA. Property-based biomass feedstock grading using k-Nearest Neighbour technique. Energy. Elsevier B.V.; 2020;190.
6. Li H, Zhao P. Improving the accuracy of tree-level aboveground biomass equations with height classification at a large regional scale. For Ecol Manage [Internet]. Elsevier B.V.; 2013;289:153–63. Available from: http://dx.doi.org/10.1016/j.foreco.2012.10.002
7. Thielen A, Klus H, Müller L. Tobacco smoke: Unraveling a controversial subject. Exp Toxicol Pathol. 2008;60:141–56.
8. Thruling.N. The aroma of flue-cured tobacco. Sensory testing for the discrimination of varieties. Aust J Exp Agric. 1964;4:367–70.
9. Hana M, McClure WF, Whitaker TB, White MW, Bahler DR. Applying artificial neural networks: Part II. Using near infrared data to classify tobacco types and identify native grown tobacco. J Near Infrared Spectrosc. 1997;5:19–25.
10. Ni LJ, Zhang LG, Xie J, Luo JQ. Pattern recognition of Chinese flue-cured tobaccos by an improved and simplified K-nearest neighbors classification algorithm on near infrared spectra. Anal Chim Acta. 2009;633:43–50.
11. Lin X, Sun L, Li Y, Guo Z, Li Y, Zhong K, et al. A random forest of combined features in the classification of cut tobacco based on gas chromatography fingerprinting. Talanta [Internet]. Elsevier B.V.; 2010;82:1571–5. Available from: http://dx.doi.org/10.1016/j.talanta.2010.07.053
12. Zhang F, Zhang X. Classification and quality evaluation of tobacco leaves based on image processing and fuzzy comprehensive evaluation. Sensors. 2011;11:2369–84.
13. Gu L, Xue LC, Song Q. Classification of the fragrant style and evaluation of the aromatic quality of flue-cured tobacco leaves by machine-learning methods. J Bioinform Comput Biol. 2016;14:1650033.
14. Wang D, Xie L, Yang SX, Tian F. Support vector machine optimized by genetic algorithm for data analysis of near-infrared spectroscopy sensors. Sensors (Switzerland). 2018;18.

15. Zhou L, Luo T, Huang Q. Co-pyrolysis characteristics and kinetics of coal and plastic blends. Energy Convers Manag [Internet]. Elsevier Ltd; 2009;50:705–10. Available from: http://dx.doi.org/10.1016/j.enconman.2008.10.007
16. Várhegyi G, Czégény Z, Jakab E, McAdam K, Liu C. Tobacco pyrolysis. Kinetic evaluation of thermogravimetric-mass spectrometric experiments. J Anal Appl Pyrolysis. 2009;86:310–22.
17. Várhegyi G, Antal MJ, Jakab E, Szabó P. Kinetic modeling of biomass pyrolysis. J Anal Appl Pyrolysis. 1997;42:73–87.
18. Saldarriaga JF, Aguado R, Pablos A, Amutio M, Olazar M, Bilbao J. Fast characterization of biomass fuels by thermogravimetric analysis (TGA). Fuel [Internet]. Elsevier Ltd; 2015;140:744–51. Available from: http://dx.doi.org/10.1016/j.fuel.2014.10.024
19. Vamvuka D, Kakaras E, Kastanaki E, Grammelis P. Pyrolysis characteristics and kinetics of biomass residuals mixtures with lignite. Fuel. 2003;82:1949–60.
20. Biagini E, Tognotti L. A generalized procedure for the devolatilization of biomass fuels based on the chemical components. Energy and Fuels. 2014;28:614–23.
21. Wang H, Xin H, Liao Z, Li J, Xie W, Zeng Q, et al. Study on the Effect of Cut Tobacco Drying on the Pyrolysis and Combustion Properties. Dry Technol. 2014;32:130–4.
22. Senneca O, Chirone R, Salatino P, Nappi L. Patterns and kinetics of pyrolysis of tobacco under inert and oxidative conditions. J Anal Appl Pyrolysis. 2007;79:227–33.
23. Jakab E, Faix O, Till F, Székely T. Thermogravimetry/mass spectrometry study of six lignins within the scope of an international round robin test. J Anal Appl Pyrolysis. 1995;35:167–79.
24. Sung YJ, Seo YB. Thermogravimetric study on stem biomass of Nicotiana tabacum. Thermochim Acta. 2009;486:1–4.
25. Oja V, Hajaligol MR, Waymack BE. The vaporization of semi-volatile compounds during tobacco pyrolysis. J Anal Appl Pyrolysis. 2006;76:117–23.
26. Guo G, Liu X, Li R, Li Q, Yu HB, Li MJ. Characterization of tobacco stalk lignin using nuclear magnetic resonance spectrometry and its pyrolysis behavior at different temperatures. J Anal Appl Pyrolysis [Internet]. Elsevier; 2019;142:104665. Available from: https://doi.org/10.1016/j.jaap.2019.104665
27. Wu W, Mei Y, Zhang L, Liu R, Cai J. Kinetics and reaction chemistry of pyrolysis and combustion of tobacco waste. Fuel [Internet]. Elsevier Ltd; 2015;156:71–80. Available from: http://dx.doi.org/10.1016/j.fuel.2015.04.016
28. Yang H, Yan R, Chen H, Lee DH, Zheng C. Characteristics of hemicellulose, cellulose and lignin pyrolysis. Fuel. 2007;86:1781–8.
29. Baker RR, Bishop LJ. The pyrolysis of tobacco ingredients. J Anal Appl Pyrolysis. 2004;71:223–311.
30. Li Q, Chen K, Liu Z, Deng X, Huang H, Huang C, et al. TGA-based analysis of pyrolysis differential between different tobacco samples. Tob Sci Technol. 2017;50.
31. Li Q, Chen K, Deng X, Guo S, Chen H, Zhong H, et al. Method of tobacco substitution based on differential analysis of tobacco pyrolysis. Tob Sci Technol. 2018;51.

32. Prieto A, Cabestany J, Sandoval F. Computational intelligence and bioinspired systems. Neurocomputing. 2007;70:2701–3.
33. Bellman R. The Structure of Dynamic Programming Processes. Dyn Program. 1957;3:81–115.
34. Corinna C, Vladimir V. Support-Vector Networks. Mach Learn. 1995;20:273–97.
35. Mountrakis G, Im J, Ogole C. Support vector machines in remote sensing: A review. ISPRS J Photogramm Remote Sens [Internet]. Elsevier B.V.; 2011;66:247–59. Available from: http://dx.doi.org/10.1016/j.isprsjprs.2010.11.001
36. Hsu CW, Lin CJ. A comparison of methods for multiclass support vector machines. IEEE Trans Neural Networks. 2002;13:415–25.
37. Xu Y, Goodacre R. On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. J Anal Test [Internet]. Springer Singapore; 2018;2:249–62. Available from: https://doi.org/10.1007/s41664-018-0068-2
38. Taylor P, Kennard RW, Stone LA. Technometrics Computer Aided Design of Experiments. Technometric. 1969;11:137–48.
39. Yuan GX, Ho CH, Lin CJ. Recent advances of large-scale linear classification. Proc IEEE. 2012;100:2584–603.
40. Westerhuis JA, Hoefsloot HCJ, Smit S, Vis DJ, Smilde AK, Velzen EJJ, et al. Assessment of PLSDA cross validation. Metabolomics. 2008;4:81–9.
41. Borgelt C, Kruse R. Graphical Models: Methods for Data Analysis and Mining. United Kingdom: J. Wiley and Sons; 2002.

Figures

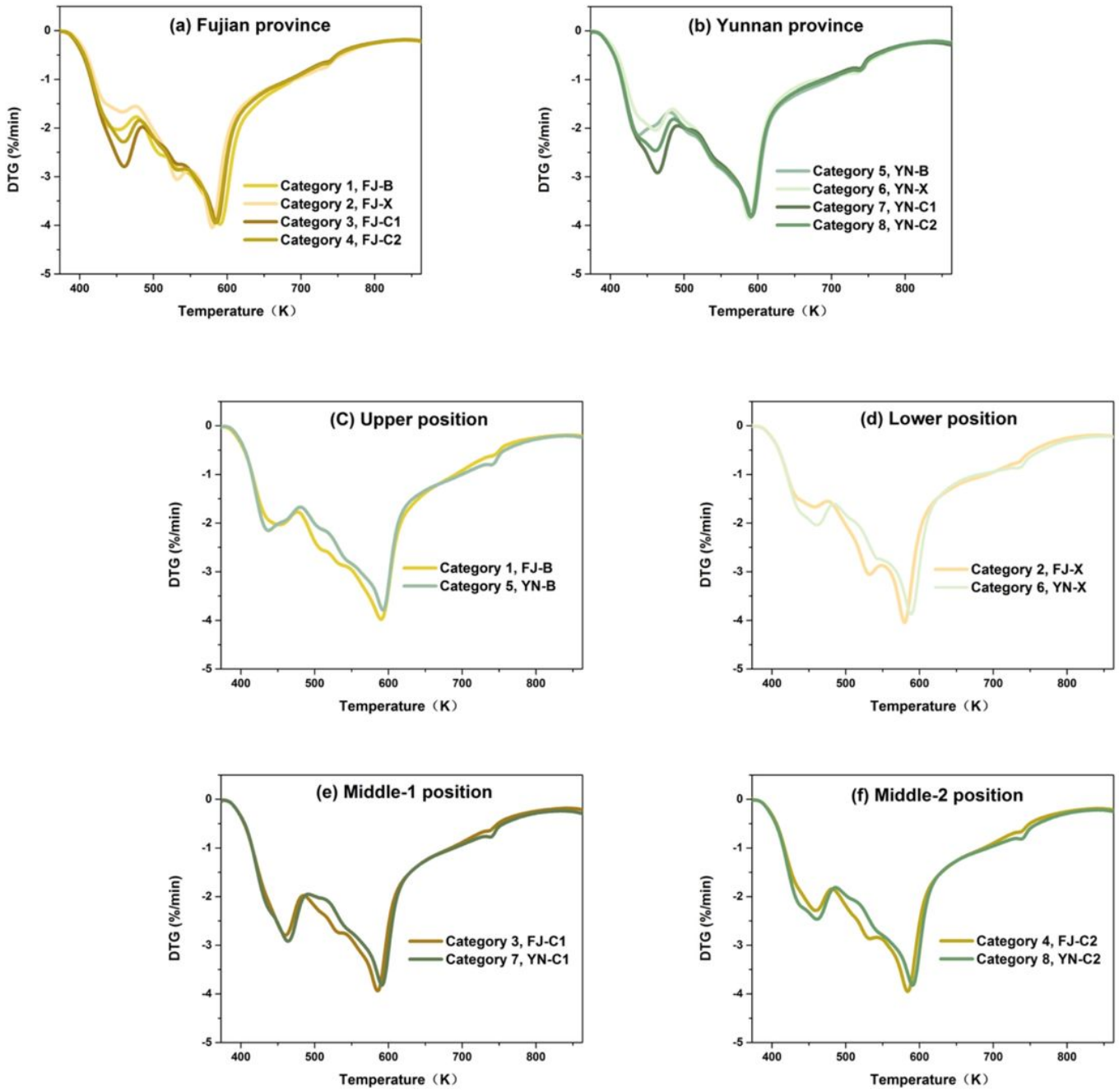

Figure 1

Comparison of the thermogravimetric analysis curves of tobacco leaves between eight categories. (a) the DTG curves of averaged four grades of Fujian province. (b) the DTG curves of averaged four grades of Yunnan province. (c) the DTG curves of averaged grade B from Fujian and Yunnan province. (d) the DTG curves of averaged grade X from Fujian and Yunnan province. (e) the DTG curves of averaged grade C1 from Fujian and Yunnan province. (f) the DTG curves of averaged grade C2 from Fujian and Yunnan province.

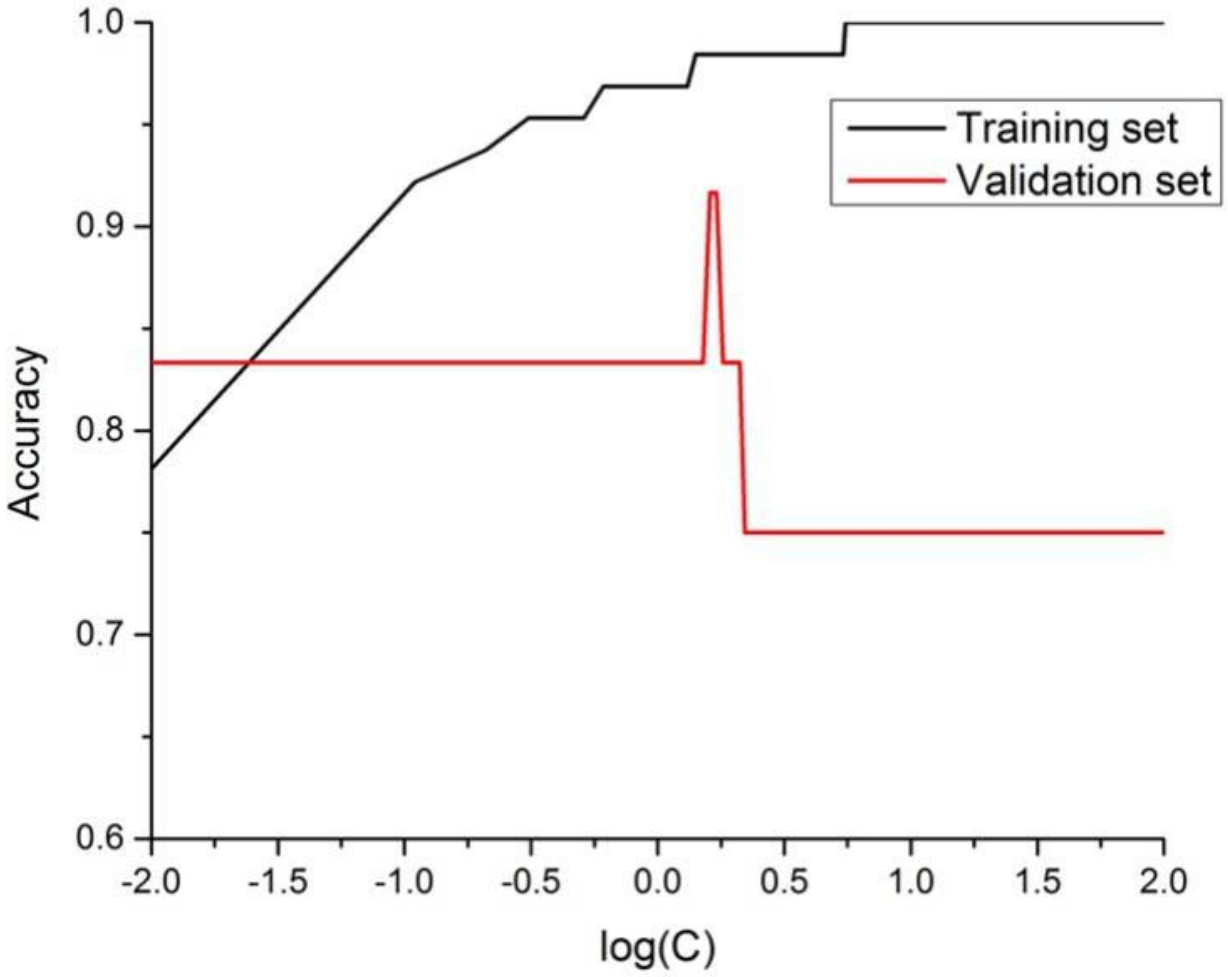

Figure 2

The influence of penalty parameter C on the accuracy of the training and validation set

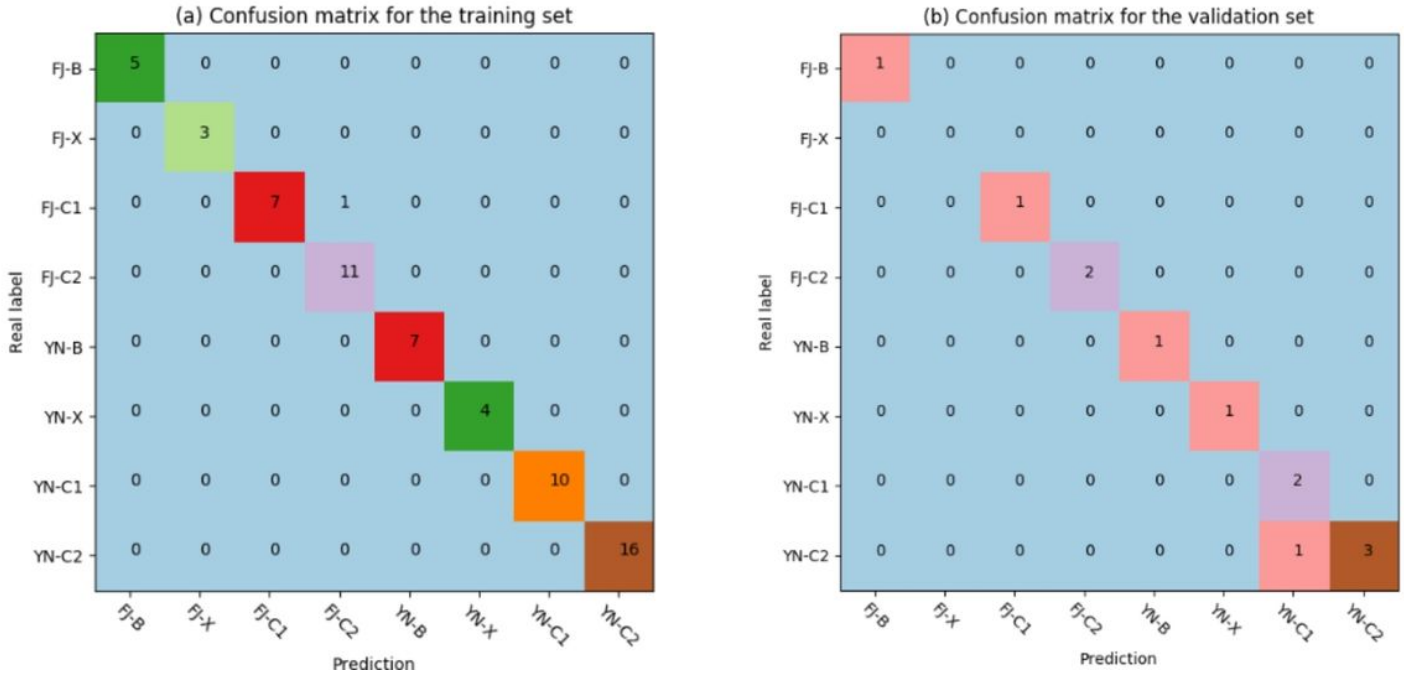

Figure 3

The confusion matrix for the training and validation set. The horizontal axis is the predicted label and the vertical axis is the real label.

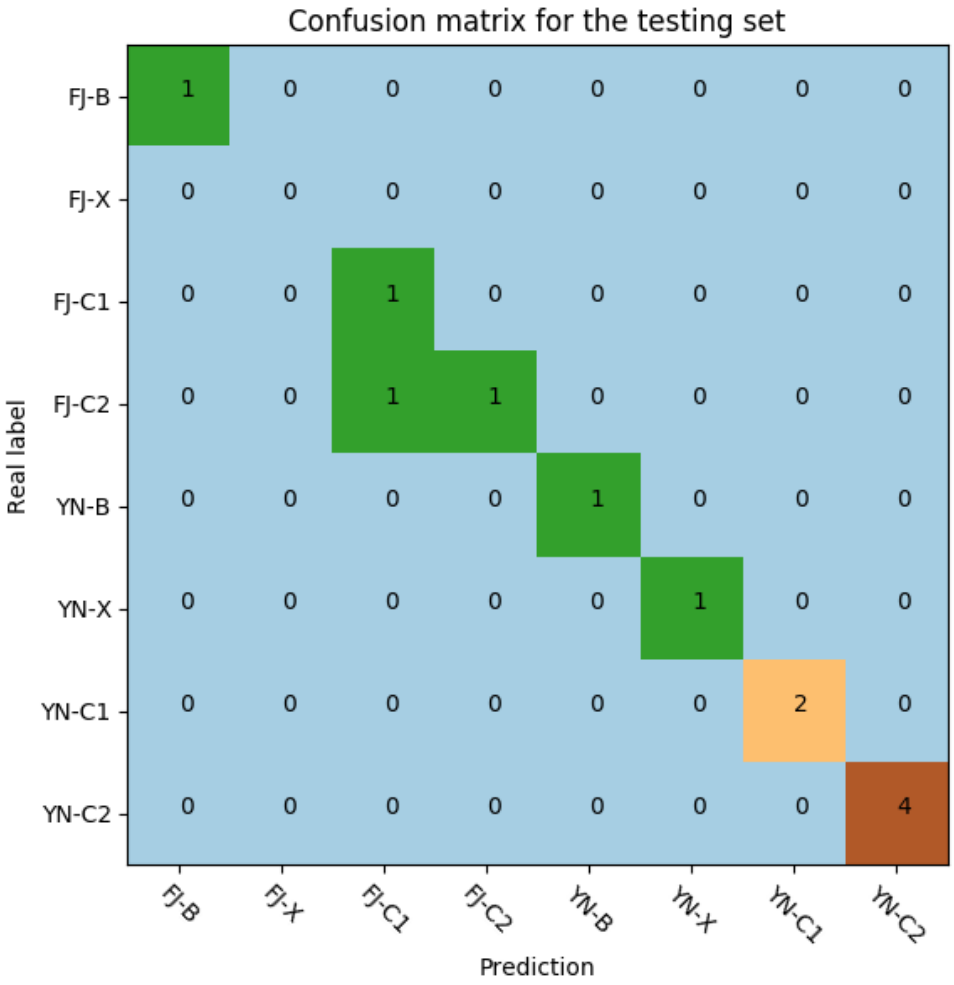

Figure 4

The confusion matrix for the testing set. The horizontal axis is the predicted label and the vertical axis is the real label.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupportingInformation.doc