

Discrimination model for cultivation origin of paper mulberry bast fiber and Hanji based on NIR and MIR spectral data combined with PLS-DA

Kyung-Ju Jang, So-Yoon Jung, In-Hee Go, and Seon-Hwa Jeong[★]

*Korea Restoration Technology Division, National Research Institute of Cultural Heritage,
132 Munji-ro, Yuseong-gu, Daejeon 34122, Korea*

(Received October 15, 2018; Revised December 11, 2018; Accepted January 22, 2019)

닥나무 인피섬유와 한지의 원산지 판별모델 개발을 위한 NIR 및 MIR 스펙트럼 데이터의 PLS-DA 적용

장경주 · 정소윤 · 고인희 · 정선화[★]

국립문화재연구소, 복원기술연구실

(2018. 10. 15. 접수, 2018. 12. 11. 수정, 2019. 1. 22. 승인)

Abstract: The objective of this study was the development of a discrimination model for the cultivational origin of paper mulberry bast fiber and Hanji using near infrared (NIR) and mid infrared (MIR) spectroscopy combined with partial least squares discriminant analysis (PLS-DA). Paper mulberry bast fiber was purchased in 10 different regions of Korea, and used to make Hanji. PLS-DA was performed using pre-treated FT-NIR and FT-MIR spectral data for paper mulberry bast fiber and Hanji. PLS-DA of paper mulberry bast fiber and Hanji samples, using FT-NIR spectral data, showed 100% performance in cross validation and the confusion matrix (accuracy, sensitivity, and specificity). The discrimination models showed four regional groups which demonstrated clearer separation and much superior score plots in the NIR spectral data-based model than in the MIR spectral data-based model. Furthermore, the discrimination model based on the NIR spectral data of paper mulberry bast fiber had highly similar score morphology to that of the discrimination model based on the NIR spectral data of Hanji.

요약: 본 연구에서는 닥나무 인피섬유와 이를 이용하여 제조한 한지의 FT-NIR 및 FT-MIR 스펙트럼 데이터를 각각 PLS-DA에 적용하여 닥나무 인피섬유 및 한지의 원산지 판별 모델을 개발하고자 하였다. 본 연구를 위하여 서로 다른 원산지의 국내산 닥나무 인피섬유 10점을 채취하여 한지로 제조하였다. 상기 시료의 FT-NIR 및 FT-IR 스펙트럼 데이터는 데이터 전처리 과정을 거쳐 PLS-DA를 수행하였다. 모델링 결과, 닥나무 인피섬유와 한지의 NIR 스펙트럼 데이터가 판별모델의 교차 검증결과 및 성능평가(정확도, 민감도, 특이도)에서 모두 100%로 MIR 스펙트럼 데이터보다 우수한 판별 성능을 나타냈다. 또한 지역별로 4개의 그룹을 형성하는 것을 확인 할 수 있었으며, 닥나무 인피섬유와 한지의 원산지 판별 모델 간 score 형태가 유사하게 나타내는 것을 확인하였다.

Key words: Paper mulberry bast fiber, Hanji, Near infrared spectroscopy, Mid infrared spectroscopy, Partial least squares discriminant analysis

[★] Corresponding author

Phone : +82-(0)42-860-9344 Fax : +82-(0)42-861-5168

E-mail : jeongsh0707@korea.kr

This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서 론

우리나라 고유의 종이인 한지를 연구하는데 있어 주원료인 닥나무(Paper mulberry: *Broussonetia kazinoki* Siebold) 인피섬유는 가장 기본적이며 중요한 요소이다. 닥나무는 섬유가 길고 강하며 윤기와 탄력이 있는 특징이 있어, 이를 이용하여 제조한 한지는 목재 섬유로 제조된 기계지에 비해 종이의 기계적 강도 및 보존성이 우수하다고 알려져 있다.^{1,3} 닥나무는 같은 종류라고 할지라도 원산지의 기후, 풍토 조건, 일교차 정도 등에 따라 닥나무 인피섬유의 길이, 폭 등이 달라져 한지의 품질 및 보존성에 영향을 줄 수 있다.⁴ 그러나 수작업으로 제조되는 한지의 제조공정상 시간을 단축시키기 위해 인피섬유 형태로 가공되어 유통되는 경우가 많아 육안으로 원산지를 판별하는 것은 한계가 있다. 더 나아가, 닥나무 인피섬유의 정보가 정확하지 않은 상태로 유통될 경우 품질뿐만 아니라 소비자의 신뢰도가 저하될 수 있는 문제점이 있다. 따라서 국내 한지산업의 발전과 과학적인 원산지 관리를 목적으로 원산지 판별 기법을 개발하는 것은 매우 시급한 실정이다.

원산지를 판별하는 방법에는 Nuclear magnetic resonance spectroscopy (NMR), Gas chromatography/mass spectrometer (GC/MS), Fourier-transform near infrared spectroscopy (FT-NIR), Fourier-transform mid infrared spectroscopy (FT-MIR) 등의 분석장비를 이용하여 스펙트럼을 조사하고 다변량 통계분석(multivariate data analysis)법을 통해 원산지 판별 모델을 개발하는 방법이 있다.⁵⁻⁷ 특히 NIR 및 MIR 분석의 경우 다른 이화학적 분석장비와 달리 비파괴적으로 신속하게 다수의 시료로부터 스펙트럼을 얻을 수 있다는 장점이 있다. 따라서 본 연구에서는 원산지 판별을 위한 여러 분석법 중 FT-NIR 및 FT-MIR 분광 분석법을 이용한 원산

지 판별을 시도하였다.

NIR 및 MIR 스펙트럼을 이용한 기존 원산지 판별 연구로는 Ilaria Santoni⁸ 외는 가문비 나무에 대해 MIR 분광분석법과 principal component analysis (PCA)을 이용하여 원산지 판별 연구를 발표하였으며, Diego C. Silva⁹ 외는 마호가니 나무에 대해 NIR 분광분석법과 partial least squares for discriminant analysis (PLS-DA) 및 soft independent modeling of class analogy (SIMCA)을 이용하여 원산지 판별 연구를 발표하였다. 이처럼 목재의 원산지 판별에 관한 선행연구들은 있으나 펄프 및 제지의 원산지 판별에 대한 연구는 거의 이루어 있지 않다.

또한 수집한 스펙트럼 데이터로 원산지 판별 모델을 개발하기 위해서는 위에서 언급한 PCA, PLS-DA, SIMCA 등과 같은 다변량 통계분석기법을 적용하여야 한다. PLS-DA는 차원 감소 및 판별 분석을 하나의 알고리즘으로 결합할 수 있는 방법으로, 적외선 스펙트럼 데이터와 같이 상호 상관성이 높은 고차원 독립 변수들을 이용한 판별 모델 개발 시 문제가 될 수 있는 노이즈, 다중공선성(multicollinearity) 등의 통계적 문제를 개선할 수 있는 기법이다.^{10,11} PLS-DA는 PCA를 기반으로 다수의 독립변수와 종속변수가 존재할 때, 독립변수뿐만 아니라 종속변수들의 상관성을 동시에 고려하여 새로운 변수를 찾고 이를 이용하여 회귀식을 추정하고 기존 변수의 차원을 줄일 수 있다(Fig. 1).¹² 따라서 PLS-DA를 적용할 경우 고차원 데이터에서 PCA보다 예측력이 높고, 노이즈 제거 효과 및 차원을 낮출 수 있는 다변량 통계분석법으로서 예측 신뢰도가 높은 모델을 얻을 수 있다.¹³

이에 본 연구에서는 국내에서 주로 자생하고 있는 주요 지역별 닥나무의 인피섬유 및 이를 이용하여 제조한 한지를 검출시간이 짧고 비 파괴분석이 가능한

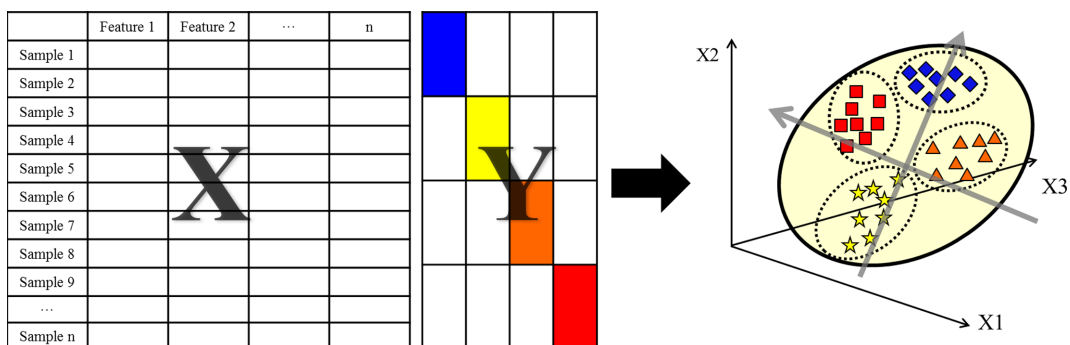


Fig. 1. Illustration of PLS-DA for models that includes four classes.^{24,25}

FT-IR 및 FT-NIR 분석방법으로 스펙트럼 데이터를 수집하였다. 또한, 상기 NIR 및 MIR 스펙트럼 데이터와 PLS-DA를 접목하여 다펀 인피섬유와 한지의 원산지 판별모델 개발 및 판별 모델 간의 유사성을 확인해 보고자 하였다.

2. 재료 및 방법

2.1. 공시재료

본 연구에 사용된 공시재료는 국내 한지 제조 공방 주변에 자생하고 있는 다펀 중에서 1년생 다펀을 이용해 제조한 다펀 백피를 수급하여 사용하였다. 수급한 공시재료는 총 10점으로 강원도 원주시(GW-WJ), 경기도 가평군(GG-GP), 충청도 단양군(CC-DY), 충청도 괴산군(CC-GS), 경상도 예천군(GS-YC), 경상도 문경시(GS-MG), 경상도 안동시(GS-AD), 경상도 의성군 신평면 증율리(GS-USJ), 경상도 의성군 신평면 덕봉리(GS-USD), 경상도 청송군(GS-CS)에서 수급하였으며, 이를 Table 1에 나타내었다. 분석에 이용되는 공시 재료는 상대습도 50.0±2.0%, 온도 23.0±1.0 °C의 조건에서 48시간동안 안정화한 후 분석을 실시하였다.

2.2. Lab. scale 한지제조

상기 다펀 인피섬유 10점을 이용하여 lab. scale 한지(이하 “한지”라고 한다)를 제조하였다. 괴산지역에서 수급한 콩대(농가부산물)를 불완전 연소하여 제조한 재 30 g에 80 °C의 증류수 300 mL를 가해 2회 여과하였다. 여과된 잿물 240 mL를 약 5 cm로 절단한 다펀 백피 30 g에 가한 뒤 1시간 동안 상온에서 방치하였다. 상기 혼합물을 autoclave를 이용하여 120 °C

에서 1시간 30분간 증해(蒸解)시킨 뒤, 표준헤리기를 사용하여 20,000 rpm으로 헤리 후 세척하였다. 세척된 시료에 증류수를 가해 약 1.0%~1.5%의 지료(w/v)를 제조하였다. 괴산 지역에서 공급받은 전건 황촉규 100 g을 증류수 3,000 mL에 침지시켜 추출하였다. 상기 황촉규 점질물 100 mL를 제조된 지료에 첨가하여 고르게 분산시켰다. 상기 혼합물을 표준 규격에 의거하여 평량 40±2 g/m²로 사각 수초지를 이용하여 초지하였다. 제작한 습지는 드럼 드라이어를 이용하여 건조한 뒤, 60 °C로 조절된 열풍 항온건조기에서 2시간 건조하였다. 초지한 종이는 TAPPI T402 om-88에 의거 상대습도 50.0±2.0%, 온도 23.0±1.0 °C에서 24시간 동안 조습한 뒤 분석에 이용하였다.

2.3. NIR 및 MIR 스펙트럼 데이터 수집

국내 10개 지역에서 수집된 다펀 인피섬유 및 한지는 통계분석을 위해 원산지 별로 5개의 시험편이 사용되었으며, 모든 시험편은 각기 다른 개체에서 채취하여 사용되었다.

NIR 스펙트럼 데이터는 Fourier-transform near infrared spectroscopy (FT-NIR, Bruker Optics Inc. FT-NIR MPA)를 이용하여 수집하였다. 각 50점의 다펀 인피섬유 및 한지 시료는 12,000~4,000 cm⁻¹ 범위에서 4 cm⁻¹ 간격으로 각 측정점당 32회 반복 측정된 평균 스펙트럼을 분석에 사용하였다.

MIR 스펙트럼 데이터는 Fourier-transform infrared spectroscopy (FT-IR, Bruker Optics Inc. VERTEX 70)를 Attenuated total reflection infrared spectroscopy (ATR) 모드로 수집하였다. 각 50점의 다펀 인피섬유 및 한지 시료는 4,000~600 cm⁻¹ 범위에서 4 cm⁻¹ 간격으

Table 1. The sample abbreviation and collection origin of Korean Paper mulberry

Cultivation region		Symbol
Major categories	Sub-categories	
Gangwon-do	Usan-dong, Wonju-si	GW-WJ
Gyeonggi-do	Sangcheon-ri, Cheongpyeong-myeon, Gapyeong-gun	GG-GP
Chungcheong-do	Yongbuwon-ri, Daegang-myeon, Danyang-gun	CC-DY
	Wonpung-ri, Yeonpung-myeon, Goesan-gun	CC-GS
Gyeongsang-do	Gaepo-myeon, Yecheon-gun	GS-YC
	Naeseo-ri, Nongam-myeon, Mungyeong-si	GS-MG
	Sosan-ri, Pungsan-eup, Andong-si	GS-AD
	Jungyul-ri, Sinpyeong-myeon, Uiseong-gun	GS-USJ
	Deokbong-ri, Shinpyeong-myeon, Uiseong-gun	GS-USD
	SongKang-ri, Pacheon-myeon, Cheongsong-gun	GS-CS

로 각 측정점당 32 회 반복 측정된 평균 스펙트럼을 분석에 사용하였다.

모든 스펙트럼 데이터는 control plate 스펙트럼(background)을 제외한 후 샘플의 스펙트럼을 나타냈다. 수집된 FT-NIR 및 FT-MIR 스펙트럼 데이터는 노이즈 제거 및 예측 정확도를 향상시키기 위해 전체 구간이 아닌 각각 8,000~4,000 cm^{-1} 및 2,000~600 cm^{-1} 파장 범위의 일부 최적구간을 선택하였다. 또한 일정 간격에서 얻은 흡광도간에는 상관관계가 높으므로 다중공선성의 영향을 감소시키기 위해 Savitzky-Golay 알고리즘으로 5 차 다항식에 의해 2 차 미분하는 것으로 전처리를 실시하였다.

2.4. 스펙트럼 데이터의 PLS-DA 분석

PLS-DA는 PLS기법을 이용하여 판별 분석을 수행하는 알고리즘으로, PLS-DA와 PLS는 동일한 분석방법이지만 PLS의 경우는 스펙트럼과 연속형 종속변수를 활용하여 회귀모델로 개발 및 예측하는 반면, PLS-DA의 경우에는 범주형 종속변수를 가지는 자료를 이용하여 회귀모델을 개발하고 예측한다는 차이가 있다.^{14,15} NIR 및 MIR 스펙트럼 데이터는 R 통계 분석 프로그램(ver 3.4.1, Aucklandm New Zealand)를 통해 분석하였다. 변수수가 p 이고 관측수가 n 일 때 X ($n \times p$) 행렬이고, Y 는 ($n \times 1$) 행렬이면, PLSR모형은 다음 식 (1)과 같다.

$$\begin{aligned} X &= TP^T + E = \sum_{i=1}^a (t_i p_i) + E \\ Y &= UQ^T + f = \sum_{i=1}^a u_i q_i + f \\ u_i &= b_i + h, i = 1, \dots, a \end{aligned} \quad (1)$$

여기서 a 는 새롭게 찾은 잠재변수(latent variable)의 수, t_i 와 u_i 는 각각 X 와 Y 의 i 번째 잠재변수들의 스코어 벡터이며 T 와 U 는 스코어행렬이다. P 와 Q 는 X 와 Y 의 로딩행렬이며 E, f, h 는 각 식의 오차항을 나타내는 행렬 및 벡터이다.

PLS-DA는 다음과 같은 과정을 통해 수행한다.¹⁶

1단계: 데이터의 종속변수를 로지스틱 회귀분석을 통해 확률 값으로 변환한다.

2단계: PLS 알고리즘에 적용하여 새로운 변수를 찾고 축소된 차원의 변환 값을 얻는다.

3단계: 변환 값을 이용하여 판별분석에 적용한다.

상기 PLS-DA를 통해 얻어진 score로 원산지 판별 예측력을 조사하기 위해 총 n 개의 스펙트럼 데이터 (50 개)에 대해 $n-1$ 개의 training set (49 개)과 1 개의 test set (1 개)로 나눈 다음 이를 학습 시키고 다른 49 개의 데이터에 대해서도 순차적으로 1 개씩 test set으로 설정한 뒤 상기 과정을 반복하여 cross validation하는 leave one-out cross validation (LOOCV)을 수행하였다.

2.5 데이터 성능 척도

개발된 판별모델의 성능 척도를 판단하기 위해서 정확도(Accuracy), 민감도(Sensitivity) 그리고 특이도(Specificity)를 이용하였다.

상기 값에 대한 식 (2)는 다음과 같다.

$$\begin{aligned} Accuracy &= \frac{TP+TN}{TP+FN+TN+FP} \\ Sensitivity &= \frac{TP}{TP+FN} \\ Specificity &= \frac{TN}{TN+FP} \end{aligned} \quad (2)$$

여기서 A와 B라는 레이블을 가진 데이터를 분류 시 관심범주를 A라고 가정할 때 TP, FN, FP, TN의 의미는 다음과 같다. True Positive (TP)는 A인 레이블을 A라고 구분한 경우로 관심 범주가 정확하게 분류되는 것을 의미한다. False Negative (FN)는 A인 레이블을 B라고 구분한 경우로 관심범주가 아닌 것으로 잘못 분류되는 것을 의미한다. False Positive (FP)는 B인 레이블을 A라고 구분한 경우로 관심범주라고 잘못 분류되는 것을 의미한다. True Negative (TN)는 B인 레이블을 B라고 하는 경우로 관심범주가 아닌 것을 정확하게 분류되는 것을 의미한다 (Table 2).

Table 2. Term define about TP, FN, FP, TN used for Confusion matrix

Classification	Condition	
	Positive	Negative
Prediction	Positive	True Positive (TP)
	Negative	False Negative (FN)
		False Positive (FP)
		True Negative (TN)

3. 결과 및 고찰

3.1. 닥나무 인피섬유 및 한지의 FT-NIR 및 FT-MIR 스펙트럼 데이터 분석

닥나무 인피섬유와 한지의 FT-NIR 및 이들의 2차 미분된 결과를 Fig. 2에 나타냈다. 닥나무 인피섬유 및 한지의 FT-NIR 스펙트럼 데이터에서 패턴 변화가 이루어진 부분은 ①의 7,325~7,315 cm^{-1} 구간에서 cellulose에서 기인된 first overtone C-H stretching vibration 및 C-H deformation vibration에 의한 peak,¹⁷ ②의 6,730~6,710 cm^{-1} 구간에서 cellulose에서 기인된 first overtone O-H stretching,¹⁸ ③의 5,610~5,590 cm^{-1} 구간에서 cellulose에서 기인된 first overtone C-H stretching vibration에 의한 peak,¹⁷ ④의 4,760~4,740 cm^{-1} 구간에서 cellulose에서 기인된 O-H stretching 및 C-H deformation vibration에 의한 peak,¹⁸ ⑤의 4,290~4,260 cm^{-1} 구간에서 cellulose에서 기인된 C-H stretching vibration 및 C-H deformation vibration에 의한 peak¹⁷ 를 확인하였다.

닥나무 인피섬유와 한지의 FT-MIR 스펙트럼과 이들의 2차미분된 결과를 Fig. 3에 나타냈다. 닥나무 인피섬유 및 한지의 FT-MIR 스펙트럼 데이터에서 패턴

변화가 이루어진 부분은 ①의 1,640~1,620 cm^{-1} 구간에서 cellulose 및 absorbed water에서 기인된 O-H bending,¹⁹ ②의 1,435~1,415 cm^{-1} 부근에서 lignin 및 hemicellulose에서 기인된 C-H deformation vibration에 의한 peak,^{8,20} ③의 1,330~1,310 cm^{-1} 구간에서 cellulose에서 기인된 C-H vibration 및 lignin에서 기인된 C-O vibration에 의한 peak,²¹ ④의 1,050~1,030 cm^{-1} 구간에서 cellulose 및 hemicellulose에서 기인된 C-O stretching vibration,^{20,22} ⑤의 910~890 cm^{-1} 구간에서 cellulose 및 hemicellulose에서 기인된 C-H deformation vibration^{22,23} 등으로 확인하였다.

상기 분석 결과를 통해 cellulose, hemicellulose, lignin 등의 정보를 반영하는 부분에서 스펙트럼의 패턴 변화가 이루어지고 있음을 알 수 있다. 또한 각각의 2차 미분 스펙트럼 데이터를 도 단위의 권역 별로 평균값을 산출한 결과에서(Fig. 4) 2차 미분 MIR 스펙트럼 데이터(우측)는 권역 별로 거의 동일한 패턴을 나타냈다. 반면 2차 미분 NIR 스펙트럼 데이터(좌측)에서는 ①의 7,500~6,750 cm^{-1} 구간 및 ②의 6,000~5,000 cm^{-1} 부근에서 원산지 별로 특정 패턴 변화를 나타내는 것을 확인하여 권역 별로 원산지 판별이 가능할 것으로 조사되었다.

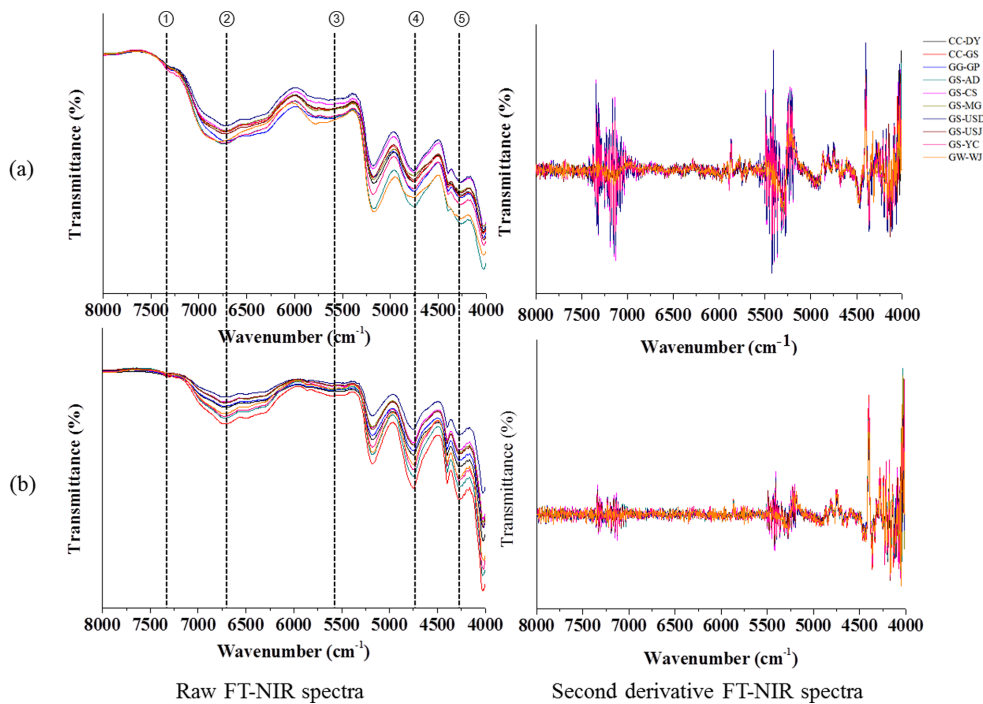


Fig. 2. FT-NIR raw spectra and second derivative spectra of (a) Paper mulberry bast fiber and (b) Hanji.

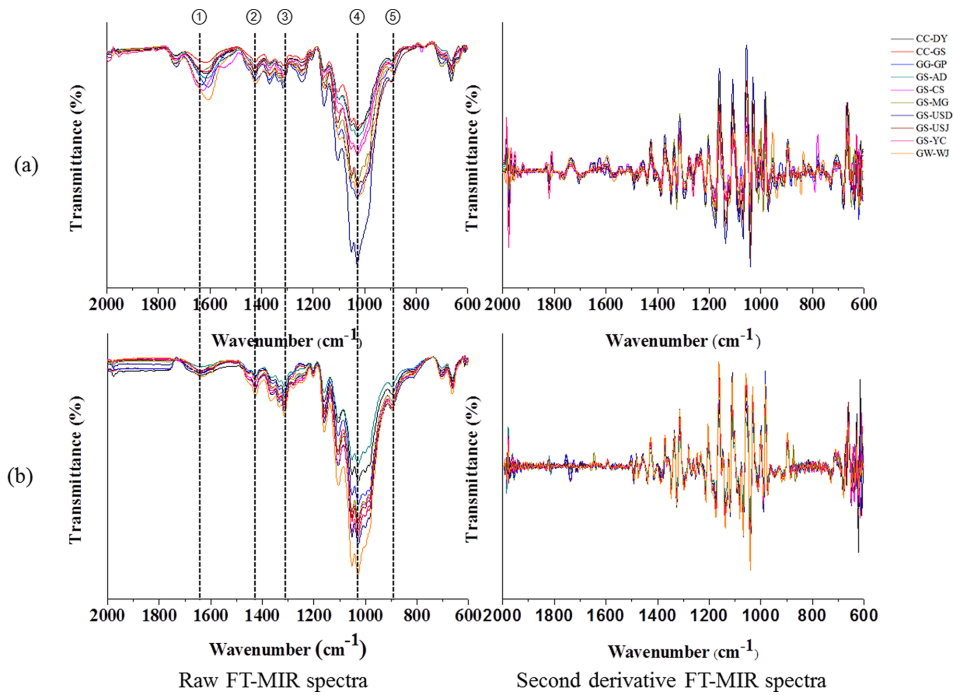


Fig. 3. FT-MIR raw spectra and second derivative spectra of (a) Paper mulberry bast fiber and (b) Hanji.

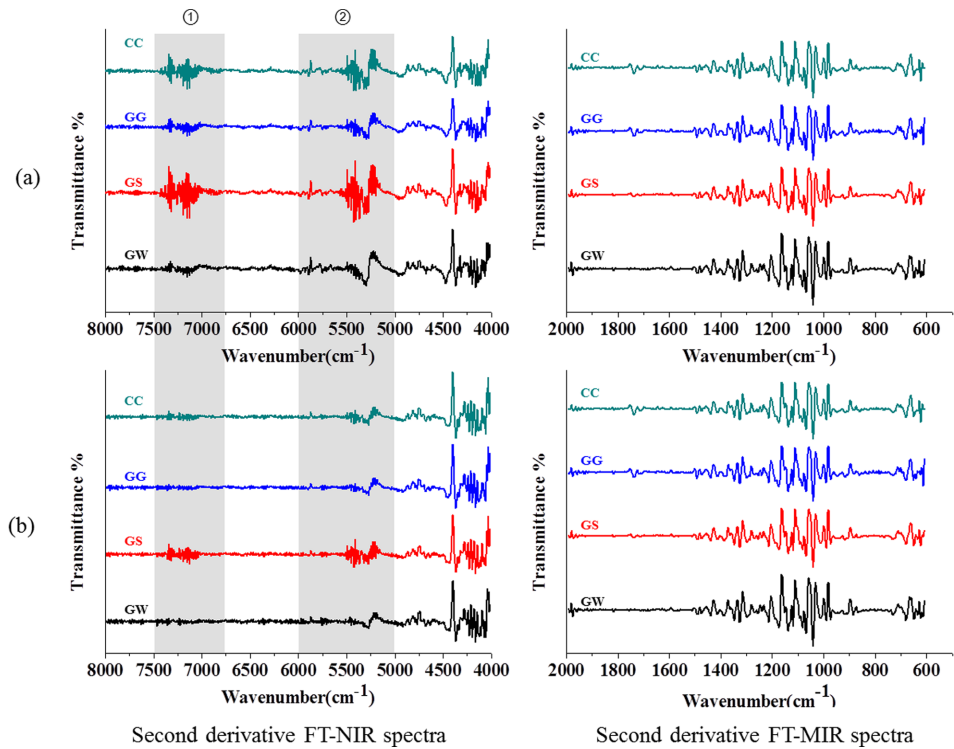


Fig. 4. Average second derivative FT-NIR and FT-MIR spectra of major cultivation region of (a) Paper mulberry bast fiber and (b) Hanji.

Table 3. Summary of the PLS-DA classification result from NIR and MIR spectral data by LOOCV. (a) NIR spectral data from Paper mulberry bast fiber, (b) MIR spectral data from Paper mulberry bast fiber, (c) NIR spectral data from Hanji (d) MIR spectral data from Hanji

Prediction	Original															
	(a)				(b)				(c)				(d)			
	CC	GG	GS	GW	CC	GG	GS	GW	CC	GG	GS	GW	CC	GG	GS	GW
CC	10	0	0	0	5	1	1	1	10	0	0	0	10	0	2	0
GG	0	5	0	0	0	3	0	1	0	5	0	0	0	5	0	0
GS	0	0	30	0	5	0	27	2	0	0	30	0	0	0	28	1
GW	0	0	0	5	0	1	2	1	0	0	0	5	0	0	0	4
Accuracy	1				0.72				1				0.94			
Error rate	0				0.28				0				0.06			

3.2. LOOCV를 이용한 스펙트럼 데이터 기반 PLS-DA 판별식의 검증

다나무 인피섬유와 한지의 NIR 및 MIR 스펙트럼 데이터에 PLS-DA를 적용하여 확립된 판별식을 LOOCV 방법을 통해 검증하여 Table 3에 나타냈다. 다나무 인

피섬유와 한지의 NIR 스펙트럼 데이터는 두 시료 모두에서 100%의 원산지 판별 정확도를 보여주었다 (Table 3(a), (b)). MIR 스펙트럼 데이터의 경우 다나무 인피섬유 시료는 50 점 중 31 점의 원산지를 정확히 판정하여 72%의 정확도를 나타냈으며 (Table 3(b)), 한

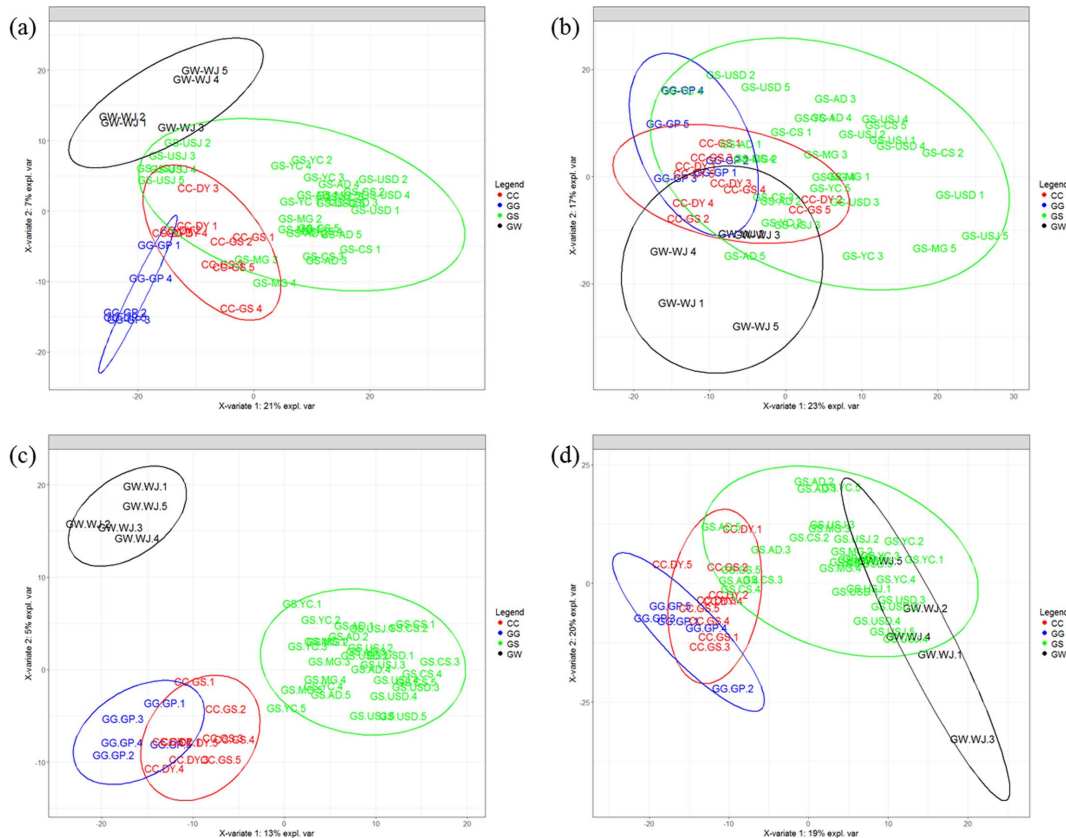


Fig. 5. A two-dimensional PLS-DA score of the (a) FT-NIR spectral data from Paper mulberry bast fiber, (b) FT-MIR spectral data from Paper mulberry bast fiber, (c) FT-NIR spectral data from Hanji and (d) FT-MIR spectral data from Hanji.

지 시료는 50 점 중 47 점의 원산지를 정확히 판정하여 94%의 원산지 판별 정확도를 나타냈다(Table 3(d)).

3.3. PLS-DA 판별 모델

닥나무 인피섬유와 한지의 NIR 및 MIR 스펙트럼 데이터에 PLS-DA를 적용해본 결과를 Fig. 5에 나타냈다. NIR 스펙트럼 데이터 기반 판별모델에서는 닥나무 인피섬유 시료(Fig. 5(a))에서 좌측 상단부에 GW(원산지: 강원도)의 시료가 위치하며, 중단부에 GS(원산지: 경상도)의 시료가 위치하고 있으며, 좌측 중·하단부에 CC(원산지: 충청도)의 시료가 위치하며, 좌측 하단부에 GG(원산지: 경기도)의 시료가 위치하고 있는 것을 확인 할 수 있었다. 상기 닥나무 인피섬유를 이용하여 제조한 한지(Fig. 5(c))의 경우도 이와 유사한 형태를 나타내며 분류가 확인해짐을 알 수 있다.

반면, MIR 스펙트럼 데이터 기반 판별모델에서는 닥나무 인피섬유(Fig. 5(b))와 한지(Fig. 5(d)) 모두 분류는 가능하나 원산지 별 score plot이 다소 혼재되어 있어 구별하기에 어려움이 있다. 따라서 닥나무 인피섬유 및 한지의 MIR 스펙트럼 데이터 기반 판별모델에 비해 NIR 스펙트럼 데이터 기반 판별모델이 닥나무 인피섬유 및 한지의 원산지를 분석하는데 있어 개선된 결과를 도출할 수 있음을 확인할 수 있었다.

닥나무 인피섬유 및 한지는 주 화학조성이 lignin, hemicellulose, cellulose로 이루어져있다. FT-NIR 분광분석법은 C-H, O-H 등의 작용기의 배움 및 결합음에 의한 에너지 흡수를 이용하기 때문에 FT-MIR에 비해 투과성이 뛰어나다. 따라서 Fig. 4와 같이 원산지 별로 특정 패턴변화를 나타내어 원산지 판별에 적합하였으며 본 연구에서 다른 닥나무 인피섬유와 한지의 경우 FT-MIR에 비해 FT-NIR 분광분석법의 원산지 구분이 명확하였다.

3.4. PLS-DA 원산지 판별모델의 성능 척도

원산지 판별 모델의 성능평가 시 정확도는 전체에서 정확히 판별한 비율, 민감도는 실제 참(true)인 것을 참(true)으로 분류한 비율 그리고 특이도는 실제 거짓(false)인 것을 거짓(false)으로 분류한 비율을 의미한다. Table 4에 나타낸 바와 같이 판별 모델들의 성능평가에서 FT-NIR 스펙트럼 기반 판별 모델들은 닥나무 인피섬유와 한지 모두에서 100%의 정확도, 민감도 및 특이도를 나타냈다(Table 4(a), (c)). 반면 FT-MIR 스펙트럼 기반 판별 모델들은 닥나무 인피섬유(Table 4)의 경우 4개의 원산지 모두 정확도가 80%

Table 4. Confusion matrix of PLS-DA model of the (a) NIR spectral data from Paper mulberry bast fiber, (b) MIR spectral data from Paper mulberry bast fiber, (c) NIR spectral data from Hanji and (d) MIR spectral data from Hanji

	Method	CC	GG	GS	GW
(a)	Accuracy	1	1	1	1
	Sensitivity	1	1	1	1
	Specificity	1	1	1	1
(b)	Accuracy	0.84	0.94	0.8	0.86
	Sensitivity	0.5	0.6	0.9	0.2
	Specificity	0.925	0.978	0.65	0.933
(c)	Accuracy	1	1	1	1
	Sensitivity	1	1	1	1
	Specificity	1	1	1	1
(d)	Accuracy	0.96	1	0.94	0.98
	Sensitivity	1	1	0.93	0.8
	Specificity	0.95	1	0.95	1

이상으로 나타났으며, 특이도 역시 경상도(65%)를 제외하고 92% 이상으로 높은 성능을 냈다. 반면, 민감도는 경상도에서 90%로 높은 성능을 나타내지만 충청도 50%, 경기도 60%, 강원도 20%로 상대적으로 낮은 성능을 나타냈다. 한지(Table 4(d))의 경우 FT-NIR 스펙트럼 기반 판별 모델에 비해 다소 낮지만 전체적으로 80% 이상의 정확도, 민감도 및 특이도를 나타내어 안정적인 판별 성능 결과를 나타냈다.

4. 결 론

본 연구에서는 서로 다른 원산지의 국내산 닥나무 인피섬유 및 이를 이용하여 제조한 한지를 대상으로, NIR 및 MIR 스펙트럼 데이터에 다변량 통계분석방법 중 하나인 PLS-DA를 적용하여 원산지 판단 모델을 설계하였다.

닥나무 인피섬유 및 한지의 FT-IR 및 FT-NIR 측정 결과 주로 lignin, cellulose 및 hemicellulose의 구조로 이루어져 있는 것을 확인하였다. 2 차 미분된 상기 스펙트럼 데이터를 기반으로 개발된 판별식의 성능은 LOOCV를 통해 닥나무 인피섬유 및 한지의 NIR 스펙트럼 데이터에서 모두 100%의 정확도를 나타냈으며, 닥나무 인피섬유 및 한지의 MIR 스펙트럼 데이터의 경우 각각 72%와 94%의 정확도를 나타내는 것을 확인하였다.

개발된 판별모델에서 닥나무 인피섬유 및 한지의

NIR 스펙트럼 데이터기반 모델은 구분 성능이 정확도, 민감도 및 특이도에서 100 %로 원산지 구분이 가능하였다. 반면, 닥나무 인피섬유 및 한지의 MIR 스펙트럼 데이터기반 모델은 구분 성능이 비교적 낮게 나타났다. 닥나무 인피섬유의 경우 정확도, 민감도 및 특이도의 평균값은 각각 86 %, 55 %, 87.15 %로 확인하였으며, 한지의 경우 정확도, 민감도 및 특이도의 평균값은 각각 97 %, 93.25 %, 97.5 %로 원산지 구분이 가능하였다.

상기 결과를 통해 닥나무 인피섬유 및 한지의 원산지 판별모델은 MIR 스펙트럼 데이터보다 NIR 스펙트럼 데이터에 PLS-DA기법을 적용 하였을 때 전체적으로 개선된 결과를 확인할 수 있었다.

또한, 닥나무 인피섬유 및 한지의 원산지 판별 모델 간 score 형태가 유사한 것을 확인함에 따라 닥나무 인피섬유 및 한지의 원산지 추정 분석 가능성을 확인할 수 있었다. 따라서 FT-NIR 스펙트럼 데이터의 PLS-DA 분석방법은 닥나무 인피섬유와 한지의 원산지를 판별하는 평가 수단으로 활용이 가능할 것으로 기대된다.

향후 닥나무 인피섬유 및 한지의 원산지 판별에 본 방법이 이용 될 수 있도록 지속적인 연구가 필요하다. 따라서 국내뿐만 아니라 국외의 다양한 산지에서 닥나무 인피섬유 시료를 확보하여 본 연구 방법에 적용한다면 원산지 판별 모델의 정확도 및 활용도가 높은 연구가 될 것으로 기대된다.

감사의 글

본 연구는 문화재청 국립문화재연구소 문화유산조사연구(R&D)사업의 일환으로 이루어졌다. 또한 본 논문의 통계적 분석 및 해석은 충북대학교 통계분석센터의 컨설팅 및 통계 분석을 지원 받아 작성되었다.

References

- O. H. Kwon and H. C. Kim, *J. Korea TAPPI*, **43**(4), 59-66 (2011).
- J. G. Kwon, C. Seo, S. S. Hong, D. W. Seo, J. S. Oh, and J. K. Kim, *J. Korean Soc. Food Sci. Nutr.*, **45**(11), 1604-1609 (2016).
- J. H. Cho, K. J. Kim, S. B. Park, and T. J. Eom, *J. Korea TAPPI*, **41**(3), 42-48 (2009).
- H. C. Kim and U. Y. Kim, *Fiber Technol. Industry*, **13**(1), 13-20 (2009).
- S. I. Lee, W. H. Chung, W. S. Kwon, J. H. Lim, and S. C. Kim, *J. Health Info. Stat.*, **36**(2), 183-192 (2011).
- K. J. Kim and T. J. Eom, *J. Korea TAPPI*, **48**(1), 34-42 (2016).
- H. Lee, Y. Lee, C. H. Jun, and J. H. Hong, *KJAS*, **23**(2), 295-304 (2010).
- I. Santoni, E. Callone, A. Sandak, J. Sandak, and S. Dirè, *Carbohydr. Polym.*, **117**, 710-721 (2015).
- D. C. Silva, T. C. M. Pastore, L. F. Soares, F. A. S. de Barros, M. C. J. Bergo, V. T. H. Coradin, A. B. Gontijo, M. H. Sosa, C. B. Chacón, and J. W. B. Braga, *Holz-forschung*, **72**(7), 521-530 (2018).
- C. J. Lee, J. W. Ko, and G. B. Lee, *Korean Chem. Eng. Res.*, **48**(6), 717-724 (2010).
- S. N. Ismail, M. Maulidiani, M. T. Akhtar, F. Abas, I. S. Ismail, A. Khatib, N. A. M. Ali, and K. Shaari, *Molecules*, **22**(10), 1-13 (2017).
- J. Y. Leem, *Yakhak Hoeji*, **60**(1), 29-35 (2016).
- G. Jiang and J. Y. Leem, *Korean J. Medicinal Crop. Sci.*, **24**(2), 93-100 (2016).
- D. Y. Kim, C. Mo, J. S. Kang, and B. K. Cho, *J. Korean Soc. Nondestruct. Test.*, **35**(1), 1-11 (2015).
- M. S. Lee, H. M. Park, B. I. Kim, and T. Y. Heo, *J. Korea Industr. Inf. Syst. Res.*, **17**(5), 89-96 (2012).
- Y. D. Kim, C. H. Jun, and H. S. Lee, *JKDIS*, **22**(5), 931-940 (2011).
- M. Schwanninger, J. C. Rodrigues, and K. Fackler, *J. Near Infrared Spectrosc.*, **19**(5), 287-308 (2011).
- S. W. Hwang, W. H. Lee, Y. Horikawa, and J. Sugiyama, *J. Korean Wood Sci. Technol.*, **43**(6), 701-713 (2015).
- J. P. McLean, G. Jin, M. Brennan, M. K. Nieuwoudt, and P. J. Harris, *Can. J. For. Res.*, **44**(7), 820-830 (2014).
- S. Jiangtao, X. Dong, and L. Jian, *Energy Procedia*, **16**(B), 758-762 (2012).
- R. Rana, R. Langenfeld-Heyser, R. Finkeldey, and A. Polle, *Wood Sci. Technol.*, **44**(2), 225-242 (2010).
- S. Durmaz, Ö. Özgenç, İ. H. Boyacı, Ü. C. Yıldız, and E. Erişir, *Vib Spectroscopy*, **85**, 202-207 (2016).
- M. Traoré, J. Kaal, and A. M. Cortizas, *Wood Sci. Technol.*, **52**(2), 487-504 (2018).
- R. G. Brereton and G. R. Lloyd, *J. Chemometrics*, **28**(4), 213-225 (2014).
- J. Trygg, E. Holmes, and T. Lundstedt, *J. Proteome Res.*, **6**(2), 469-79 (2007).

Authors' Positions

Kyung-Ju Jang : Researcher
So-Yoon Jung : Researcher
In-Hee Go : Researcher
Seon-Hwa Jeong : Researcher