STUDY OF USING ATR-FTIR SPECTROSCOPY AND MULTIVARIATE STATISTICAL (MSA) TECHNIQUES TO CLASSIFY SESAME PRODUCTS IN VIETNAM

NGHIÊN CỨU SỬ DỤNG KỸ THUẬT QUANG PHỔ ATR-FTIR VÀ THỐNG KÊ ĐA BIẾN NHẰM PHÂN LOẠI SẢN PHẨM VỪNG TẠI VIỆT NAM

Nguyễn Quang Trung(*), Bùi Quang Minh, Tạ Thị Thanh Huyền, Trương Ngọc Minh

Trung tâm Nghiên cứu và Chuyển giao Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam
Email: nqt79@yahoo.com Điện thoại: +84912141580

ABSTRACT:

The research objective of this paper is to find a fast and reliable method for classification of sesame products, with the aim of trademarks protection. The Fourier-transform infrared spectroscopy (FTIR) and the multivariate statistical techniques including Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) were used. The spectral band data were simplified by selecting peak coordinates prior to being processed by the multivariate statistical techniques. This method has classified 15 samples of whole-seed sesame products in Vietnam markets, including 8 white sesame samples and 7 black sesame samples collected from supermarkets, local markets and food suppliers in Hanoi. Research results showed that the task of evaluating and classifying sesame products on the market of each sesame brand has been successful. Two principle component analysis (PCA) and linear discriminant analysis (LDA) methods supported each other in this task.

Keywords: Sesame, FTIR, PCA, LDA

1. MỞ ĐẦU

Hạt vừng (*Sesamum indicum L.*) với hương thơm đặc trưng cùng vị béo bùi, mang giá trị dinh dưỡng cao là "điểm nhấn hương sắc" cho các món ăn nói chung [1] và đối với người dân Việt Nam là thành phần không thể thiếu để tạo nét đặc sắc trong việc dung để trang trí các món ăn từ thời xa xưa.

Được coi là loại hạt cho dầu lâu đời nhất của nhân loại. Cây vừng có nhiều loài (với khoảng 30 loài), và những loài hoang dã được khai thác lâu đời nhất ở Châu Phi và Ấn Độ. Hồ sơ từ Babylon và Assyria, có niên đại khoảng 4.000 năm trước đã đề cập đến vừng [2]. Ở các nước châu Á trong vài nghìn năm hạt vừng đã được sử dụng như một loại thực phẩm sức khỏe để phòng chống bệnh tật. Chúng làm tăng đáng kể g-tocopherol trong huyết tương và tăng cường hoạt động của vitamin E rất tốt cho sức khỏe [4]. Nhiều loại khách nhau nhưng hiện nay được trồng phổ biến nhất là vừng đen và vừng trắng. Vừng đen còn gọi là hồ ma, hắc chỉ ma, cự thắng, cự thắng tử, ô ma, ô ma tử, du ma, giao ma, tiểu hổ ma. Vừng trắng còn gọi là bạch du ma, bạch hổ ma [3].

Trong hạt vừng có chứa đủ các thành phần dinh dưỡng như protid, lipid, glucid, xơ, vitamin B_1, B_2, PP, E, các chất khoáng như: Ca, P, K, Na, Mg, Fe, Zn, Se, Cu, Mn… Các nghiên cứu đã chỉ ra rằng dầu vừng có thể ức chế sự phát triển ung thư ở người, giảm huyết áp, giảm quá trình peroxy hóa lipid và tăng tình trạng chống oxy hóa ở bệnh nhân cao huyết áp [2]. Theo Y học cổ truyền, hạt vừng có vị ngọt, béo, tính bình, quy 4 kinh phế, tỳ, can, thận. Có tác dụng nhuận tràng, bổ khí huyết, bổ ngũ tạng, ích khí lực, bổ não tủy, mạnh gân cốt, sáng tai mắt, ích lão trường thọ. Nên vừng có thể được dùng đơn độc hoặc kết hợp cùng các vị thuốc khác chữa trị bệnh. [4].

Tiêu thụ hạt vừng trên toàn thế giới là 6559,0 triệu USD vào năm 2018 và sẽ đạt 7244,9 triệu USD vào năm 2024, với tốc độ CAGR (tốc độ tăng trưởng kép hàng năm) là 1,7%. Tiêu thụ vừng trên toàn cầu đang tăng đều đặn chủ yếu do người tiêu dùng thay đổi cách tiêu dùng và nâng

cao nhận thức về sức khỏe. Khoảng 70% hạt vừng trên thế giới được sử dụng để sản xuất dầu và bột. Tổng lượng tiêu thụ dầu và thực phẩm hàng năm lần lượt là khoảng 65% và 35% [7]. Tại Việt Nam năm 2019 giá vừng đen 40.000-45.000 đ/kg, đến vụ vừng năm nay (2020) tăng lên 50.000-52.000 đ/kg. Tính toán chi phí sản xuất theo thời giá hiện nay, nông dân trồng vừng đạt năng suất bình quân trên 1,4 tấn/ha lãi khoảng 60-70 triệu đồng/ha. Năng suất vừng tăng lên, có thị trường tiêu thụ tốt nên mang lại giá trị kinh tế lớn cho người dân [5].

Khảo sát thị trường giá vừng hiện giao động từ 40.000 vnđ đến hơn 50.000 vnđ đối với 1kg vừng, nhưng với một vài gian hàng chỉ bán khoảng 20.000vnđ/1kg và cũng có nhưng gian hàng để giá lên đến 480.000vnđ/ 1kg [6]. Chính vì sự đa dạng về chủng loại và giá trị kinh tế mà Hạt vừng mang lại, nên người ta thường làm giả thương hiệu, chủng loại để có lợi nhuận cao hơn. Nghiên cứu này được thực hiện nhằm sử dụng quang phổ hồng ngoại (ATR-FTIR) kết hợp với các thuật toán thống kê đa biến để phân biệt các loại vừng Việt Nam bán trên thị trường. Theo đó, các kết quả nghiên cứu có thể xây dựng thành phương pháp phân tích hóa học để phân biệt chủng loại, thương hiệu vừng, nhằm bảo vệ thương hiệu sản phẩm.

2. THỰC NGHIỆM

2.1. Thu thập và xử lý mẫu

Các mẫu được thu thập tại các chợ địa phương và các siêu thị trong phạm vi quận Cầu Giấy và quận Ba Đình, thành phố Hà Nội gồm15 mẫu hạt vừng (cả trắng và đen) được chia làm 9 thương hiệu khác nhau. Trong đó, các sản phẩm được mua tại siêu thị được đóng gói và có thương hiệu rõ ràng và một vài sản phẩm được lấy tên địa điểm bán hàng, coi là một thương hiệu riêng biệt.

Bảng 1: Thông tin 15 mẫu hạt vừng

Thương hiệu	Các loại sản phẩm	
	Trắng	Đen
PMT	VT1	VD1
Thu Dung	VT2	VD2
VANA	VT3	-
Chợ Nghĩa Tân Q1	VT4	VD4
Chợ Nghĩa Tân Q3	VT5	VD5
Chợ 800A	VT6	-
Danavi	VT7	VD7
	VT8	-
Đức Trung Tín	-	VD9
3D	-	VD10

Mẫu vừng thu thập ở dạng hạt được nghiền nhỏ và đồng hóa bằng máy xay mẫu thực phẩm khô (Sunhouse), được sấy khô tại nhiệt độ 45°C trong vòng 2h đồng hồ trong tủ sấy (Memmert UN110). Sau đó mẫu được đo trực tiếp trên thiết bị quang phổ.

2.2. Phân tích quang phổ hồng ngoại

Các mẫu sau công đoạn xay nhỏ và sấy khô thành dạng bột được phân tích trực tiếp trên thiết bị ATR-FTIR (Nicolet™ iS50, Thermo Scientific, Mỹ). Sơ lược các bước tiến hành: Lấy khoảng 2 mg bột vừng được đặt trên bề mặt tinh thể kim cương-ZnSe. Công cụ nén mẫu tích hợp trên thiết bị được sử dụng để tăng diện tích tiếp xúc giữa mẫu và cảm biến cũng như đảm bảo thiết bị cho tín hiệu dải phổ tốt nhất bằng cách áp dụng một lực nhất định lên trên nền mẫu khô. Kết quả là tín hiệu phổ được thu thập trên phần mềm OMNIC đã cài trong máy tính. Các giải phổ thu thập được có bước sóng nằm trong khoảng từ 4000-400 cm^{-1}, bao gồm 70 lần quét mẫu và có độ phân giải 4 cm^{-1}. Mỗi mẫu được phân tích lặp đi lặp lại 5 lần nhằm mang lại kết quả chính xác nhất cần đạt.

2.3. Kỹ thuật thống kê đa biến

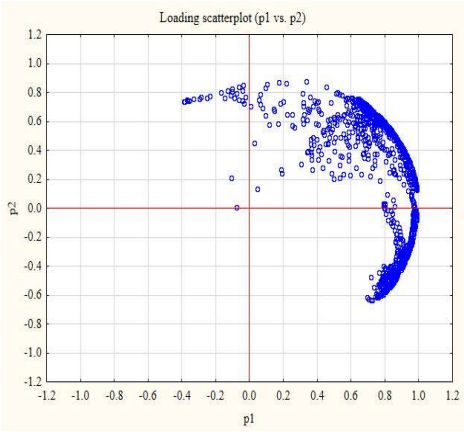
Công cụ The Unscrambler® X 10.4: công cụ giúp hiệu chỉnh đường cơ sở (baseline) và chuyển hóa thành dạng dự liệu trên Excel từ các dự liệu phổ thu được. Từ đó thông tin chỉnh trên dải phổ được sang lọc và rút gọn dựa trên các đỉnh peak khác nhau để tạo nên một tệp dữ liệu mới. Sử dụng nền tảng thống kê XLSTAT 2016.02.28451 thông qua các phép phân tích như PCA - Phân tích thành phần chính và LDA-

Phân tích phân biệt được tiến hành trên tệp dữ liệu mới tích hợp trên Microsoft Excel và STATISTICA 12.

3. KẾT QUẢ VÀ THẢO LUẬN

3.1 Khảo sát phân tán PCA của 15 mẫu vừng

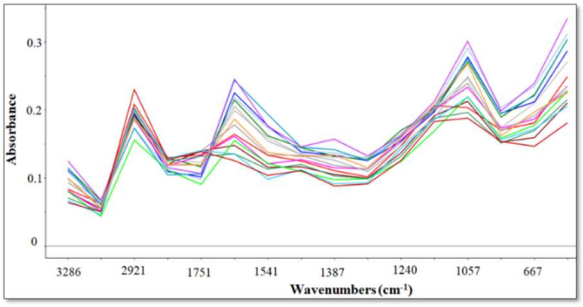
PCA được sử dụng để trích xuất thông tin quan trọng từ bảng dữ liệu đa biến và biểu thị thông tin này dưới dạng một tập hợp một vài biến mới được gọi là thành phần chính (PC) [7]. Từ đó xem xét tác động của nguồn gốc và tính phân biệt của 15 mẫu vừng. Các biến mới này tương ứng với sự kết hợp tuyến tính của các biến ban đầu, số lượng các thành phần chính nhỏ hơn hoặc bằng số lượng các biến ban đầu. PCA giảm kích thước của dữ liệu đa lượng biến thành hai hoặc ba thành phần chính, có thể được trực quan hóa bằng đồ thị, với sự mất mát thông tin tối thiểu. Kết quả thu về các số liệu sẽ được sắp xếp lại, các điểm thành phần chính đầu tiền sẽ được đại diện để mô tả thông tin, PCA giả định rằng các hướng có phương sai lớn nhất là "quan trọng" nhất [7]. Qua khảo sát phân tán PCA của 15 mẫu vừng ta thấy qua hình 1 các bước sóng phân tán tập trung nhiều ở 2 PC nhưng vẫn cần bớt lại 1 vài bước song để có thể thấy sự phân tán được rõ ràng hơn.



Hình 1: Khảo sát phân tán PCA của 15 mẫu vừng

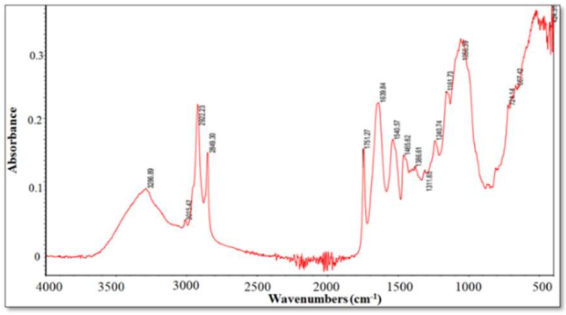
3.2 Phân tích phổ hồng ngoại

Dùng hạt vừng làm thực phẩm là cách phổ biến nhất ở tất cả các nước. Trong hạt vừng có chứa: 45 - 55% chất béo, 19 - 20% Protein, 8 - 11% đường, 5% nước, 4 - 6% chất khoáng. Thành phần chứa lignans pinoresinol và lariciresinol. Bên cạnh một số thành phần cấu tạo chính được xác định, có rất nhiều các hợp chất nhỏ lẻ khác chưa được làm rõ, nhưng chúng chắc chắn sẽ đóng góp vào dải phổ hồng ngoại [8, 14].



Hình 2: Tổng hợp dải phổ của 15 mẫu hạt vừng

Hình 2 mô tả đồ thị các đỉnh của 15 mẫu vừng. Kết quả của mỗi mẫu được đo lặp lại 5 lần để xác định sai số của phương pháp. Tuy nhiên trong đó không nhiều sự chênh lệch tại mỗi kết quả cho thấy độ lệch chuẩn tương đối. Sai số của phép đo được lấy trung bình tại mọi đỉnh của phổ hồng ngoại không nằm ngoài khoảng cho phép.

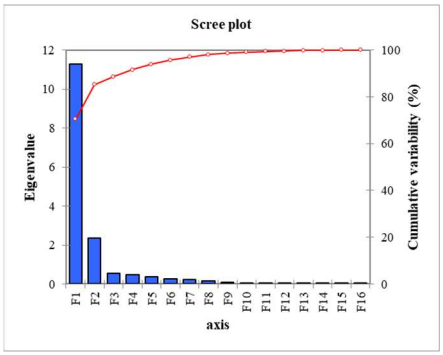


Hình 3: Đỉnh peak của 15 mẫu hạt vừng

Để giảm thiểu các dữ liệu không cần thiết, chỉ tổng hợp các đỉnh cực đại của phổ hồng ngoại của tất cả các mẫu. Hình 3 mô tả đồ thị 16 đỉnh của 15 mẫu vừng. từ bước sóng 3300 đến 2500 và từ bước sóng 1390 đến 1310 (dao động của nhóm O–H); từ bước sóng 1648 đến 1638 (dao động của nhóm C=C); từ bước sóng 1550 đến 1500 (dao động của nhóm N-O); từ bước sóng 1465 đến 1450 (dao động của nhóm C-H); từ

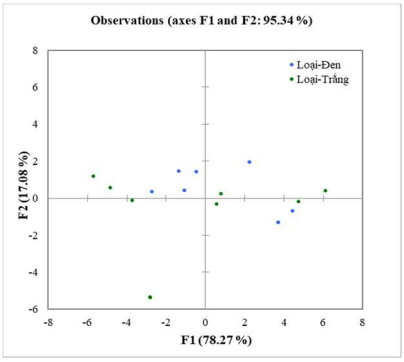
bước sóng 1000 đến 500 (dao động biến dạng) [9].

3.3 Ứng dụng phân tích thành phần chính



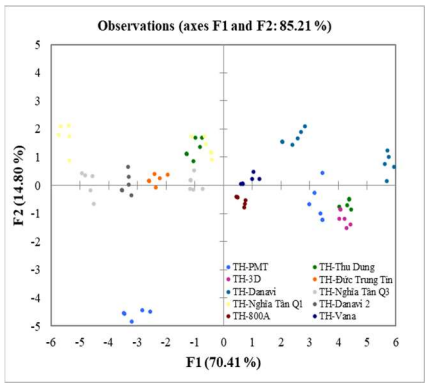
Hình 4: Sơ đồ sàng lọc của thành phần chính

Hình 4 cho thấy sơ đồ sàng lọc được sử dụng để minh họa sự đóng góp của các thành phần chính đối với giá trị riêng và vector riêng. Có thể thấy rằng 2 PC đầu tiên chiếm đến gần đến 90% lượng thông tin chính trên tổng số biến thiên của các mẫu. Do đó, có thể nói rằng những PC này mang thông tin chính của các biến.



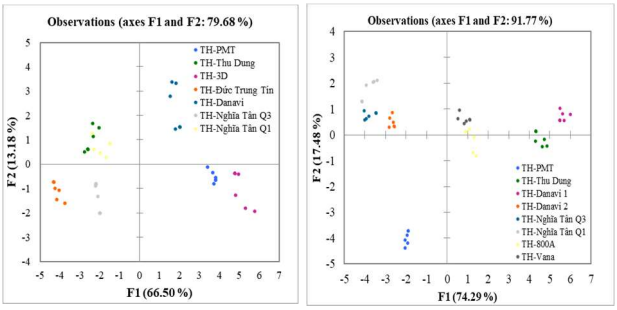
Hình 5: Áp dụng PCA phân biệt chủng loại

Kết quả PCA cho phân biệt chủng loại vừng chỉ ra rằng hai thành phần chính đầu tiên (F1 và F2) nắm giữ phần lớn các thông tin dữ liệu mẫu, chiếm 95,34%, trong đó 78,27% và 17,08% là tổng số biến thể lần lượt thuộc về thành phần chính đầu tiên (F1) và thứ hai (F2). Tuy nhiên, có thể thấy rằng không thể phân biệt được hai chủng loại vừng do không quan sát được sự phân nhóm giữa hai chủng loại và sự phân bố mẫu khá rời rạc và xen kẽ nhau (Hình 5). Có thể thấy được dấu hiệu tích cực hơn khi áp dụng PCA cho nhận biết thương hiệu trên cả hai loại vừng (không tách riêng từng loại). Đã có sự phân nhóm cho từng thương hiệu, nhưng lại quan sát được sự tách nhóm riêng trong mỗi thương hiệu được biểu thị trên biểu đồ PCA (Hình 6). Đây là bằng chứng cho sự khác nhau của hai chủng loại vừng đen và trắng. Do vậy, khảo sát sâu hơn được tiến hành trên từng chủng loại vừng nhằm quan sát rõ hơn sự khác biệt thương hiệu.



Hình 6: Áp dụng PCA phân biệt thương hiệu

Kết quả sử dụng mô hình PCA không thành công trong việc phân biệt chủng loại vừng, tuy nhiên lại khá hiệu quả khi áp dụng phân biệt thương hiệu trên từng chủng loại vừng. Các thương hiệu nhìn chung đã được phân nhóm rõ ràng hơn do các điểm thuộc từng thương hiệu được phân bố tập trung và tách biệt hơn. Hình 7a cho thấy với vừng đen, các mẫu của thương hiệu Thu Dung và Nghĩa Tân có vị trí khá sát nhau. Nguyên nhân giải thích điều này được cho rằng có liên quan tới sự tương đồng trong thành phần hóa học của mẫu vừng thuộc hai thương hiệu trên. Ngoài ra, sự phân tách giữa các nhóm thương hiệu còn lại được biểu thị khá rõ trên biểu đồ PCA.

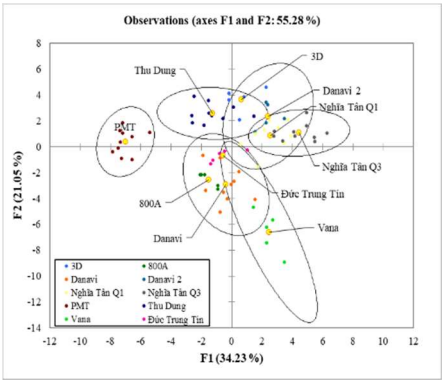


Hình 7: Áp dụng PCA cho phân biệt thương hiệu vừng: (a) – Vừng đen; (b) – Vừng trắng

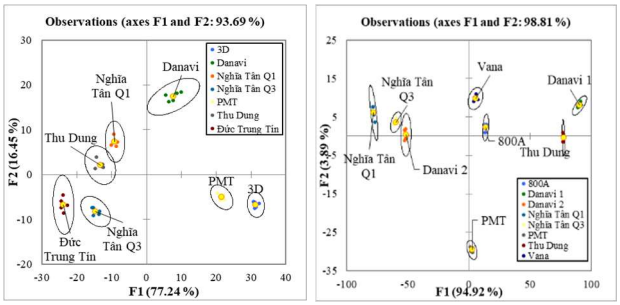
Sự phân nhóm cũng được thể hiện rõ ràng khi quan sát kết quả phân biệt thương hiệu trên vừng trắng (Hình 7b). Trong đó, sự khác biệt lớn nhất thấy được ở thương hiệu Thu Dung,Danavi 1 và PMT khi 3 nhóm này nằm ở vị trí riêng biệt và tách hẳn so với các nhóm còn lại. Trong khi nhóm *Danavi 2-Nghĩa Tân Q3-Nghĩa Tân Q1* và *Vana-800A* là hai nhóm có phân bố sát nhau. Tổng hợp các kết quả thu được cho thấy phương pháp phân tích thành phần chính – PCA mang lại hiệu quả khá rõ ràng đối với việc phân biệt thương hiệu. Chứng minh cho điều này là sự phân nhóm theo thương hiệu của các thông tin biểu thị trên biểu đồ PCA.

3.4 Ứng dụng phân tích phân biệt

LDA (Linear discriminant analysis) là một công cụ phân loại bao gồm tính toán tuyến tính kết hợp (hàm phân loại) của các biến ban đầu có thuộc tính, tối đa hóa sự khác biệt và giảm thiểu sự khác biệt trong các nhóm [10,11]. Áp dụng cho 15 mẫu vừng (bao gồm cả vừng trắng và vừng đen), có thể thấy được sự phân biệt thương hiệu của các mẫu vừng chưa hiệu quả (Hình 8), các khoảng tin cậy (Confidence ellipse) của các mẫu có sự trùng lặp, ngoại trừ mẫu thương hiệu PMT. Kết quả xử lý LDA cho thấy các thương hiệu được phân nhóm và được phân bố ở các tọa độ khá tách biệt với nhau. Duy chỉ đối với vừng đen (Hình 9a), thương hiệu *Nghĩa Tân Q1* và *Thu Dung* có sự trùng lặp tại khoảng tin cậy và đối với vừng trắng (Hình 9b), *Nghĩa Tân Q3* và *Danavi 2* là hai thương hiệu duy nhất phân bố sát nhau, thể hiện sự giống nhau trong thành phần hóa học. Ngoài ra, PMT là thương hiệu biểu thị sự phân bố xa nhất so với các nhóm thương hiệu còn lại (Hình 8b).



Hình 8: Kết quả áp dụng phân tích phân biệt thương hiệu



Hình 9: Kết quả áp dụng phân tích phân biệt (LDA) đối với vừng: (a) – Vừng đen; (b) – Vừng trắng

Vậy có thể thấy khi phân biệt thương hiệu áp dụng trên cả hai loại vừng trắng và đen thì phương pháp LDA không thể hiện rõ ràng như PCA (Hình 5). Tuy nhiên khi áp dụng phương pháp phân tích LDA để phân biệt thương hiệu đối với từng chủng loại cụ thể lại mang lại kết tốt hơn (Hình 9a, 9b).

4. KẾT LUẬN

Kết quả nghiên cứu cho thấy việc áp dụng Quang phổ hồng ngoại (ATR-FTIR) kết hợp với kỹ thuật thống kê đa biến bao gồm: Kỹ thuật phân tích thành phần chính (PCA); Thống kê phân tích phân biệt (LDA) có thể dùng được cho việc phân loại Vừng của 9 thương hiệu vừng trên thị trường. Quá trình chạy số liệu bao gồm PCA, theo sau là phương pháp LDA dựa trên các điểm PC là khả quan. Qua khảo sát và đánh giá kết quả

2 phương pháp thống kê đa biến khác nhau, có thể khẳng định rằng chúng bổ trợ được cho nhau trong nghiên cứu phân loại thương hiệu của các mẫu vừng trên thị trường. Cụ thể đối với PCA các thương hiệu vừng khi không phân tách chủng loại vừng được phân biệt tốt hơn LDA, nhưng khi phân tách riêng chủng loại vừng thì phương pháp LDA lại mang lại kết quả rõ ràng hơn so với PCA. Tổng kết nghiên cứu cho thấy nhiệm vụ đánh giá việc phân loại các sản phẩm vừng trên thị trường của từng thương hiệu vừng đã hoàn thành, song vẫn cần phải có những nghiên cứu xa hơn để có thể mang lại kết quả nhanh và hiệu quả hơn.

Cảm ơn:

Nghiên cứu này do Viện Hàn lâm Khoa học và Công nghệ Việt Nam tài trợ, mã số: TĐNDTP.01 / 19-21 và số cấp: QTCZ01.01 / 20-21.

TÀI LIỆU THAM KHẢO

1. Báo Sức Khỏe và Đời Sống, Khám phá điều kỳ diệu trong hạt vừng nhỏ bé, **2019**. (https://suckhoedoisong.vn/kham-pha-dieu-ky-dieu-trong-hat-vung-nho-be-n138062.html)
2. Tạp chí Công Thương, Nghiên cứu sản xuất thực phẩm giàu hoạt chất sinh học từ vừng đen, **2020**. (http://tapchicongthuong.vn/bai-viet/nghien-cuu-san-xuat-thuc-pham-giau-hoat-chat-sinh-hoc-tu-vung-den-69108.htm)
3. Báo Nhân dân điện tử, Bài thuốc từ hạt vừng **2005**. (https://nhandan.com.vn/tin-tuc-y-te/bai-thuoc-tu-hat-vung-602701/)
4. Elleuch M.; Bedigian D.; Zitoun A. Sesame *(Sesamum indicum L)* Seeds in Food, Nutrition, and Health, **2011, 122**, 1029-1036.
5. Báo Nông Nghiệp Ciệt Nam, Chuyển đổi luân canh cây mè trên đất lúa **2020.** (https://nongnghiep.vn/chuyen-doi-luan-canh-cay-me-tren-dat-lua-d265751.html)
6. Cổng Thông Tin điện tử Tỉnh Gia Lai, Nông dân thu nhập khá từ cây mè, **2020.** (https://gialai.gov.vn/tin-tuc/nong-dan-thu-nhap-kha-tu-cay-me.65840.aspx)
7. Myint D.; Gilani S.A.; Kawase M.; Watanabe K.N. Sustainable Sesame *(Sesamum indicum L.)* Production through Improved Technology: An Overview of Production, Challenges, and Opportunities in Myanmar, **2020**.
8. Fuji Y.; Uchida A.; Fukahori K.; Chino M.; Ohtsuki T.; Matsufuji H. Chemical characterization and biological activity in young sesame leaves *(Sesamum indicum L.)* and changes in iridoid and polyphenol content at different growth stages, **2018**.
9. STHDA Statistical tools for high-throughput data analysis, **2017**.
10. IR Spectrum Table & Chart
11. XLSTAT- DISCRIMINANT ANALYSIS (DA)
12. K.V. Mardia. Mahalanobis distances and angles. In ***Multivariate Analysis*** (ed. P. R. Krishnaiah, North-Holland, Amsterdam, The Netherlands, **1977, 495**, 5-11.
13. I.T. Jolliffe**.** *Principal Components Analysis***.** SpringerVerlag New York, USA, **1986**.
14. Alpaslan, M., Boydak, E., and Demircim, M. Protein and oil composition of soybean and sesame seed grown in the Harran (GAP) area of Turkey. *Food Chem*., **2001**, 88, 23-25.
15. D. J. Lyman, R. Benck, S. Dell, S. Merle and M. W. Jacqueline. FTIR-ATR Analysis of Brewed Coffee: Effect of Roasting Conditions. *Journal of Agricultural and Food Chemistry*, **2003**, *51(11)*, 3268-3272
16. X. Wu, B. Wu, J. Sun, M. Li and H. Du. Discrimination of Apples Using Near Infrared Spectroscopy and Sorting Discriminant Analysis, *International Journal of Food Properties*, **2016**, *19(5)*, 1016-1028.
17. Dabrowski K.J.; Sosulski F.W Composition of free and hydrolyzable phenolic acids in defatted flours of ten oilseeds, *J Agrie Food Chem*, **1984**, 32:128–130
18. Jain S.C. Isolation of pedaliin from *Sesamum indicum* L. tissue culture. Agrie Biol Chem, **1981**, 45:2127.

Liên hệ: **Nguyễn Quang Trung**
Trung tâm Nghiên cứu và Chuyển giao Công nghệ
Tầng 6, Tòa nhà A28, Số 18B Hoàng Quốc Việt, Cầu Giấy, Hà Nội
Email: nqt79@yahoo.com
Điện thoại: +84912141580