

THÔNG TIN VỀ LUẬN ÁN TIẾN SĨ

1. Họ và tên nghiên cứu sinh: NGUYỄN ĐỨC THANH
2. Giới tính: Nam
3. Ngày sinh: 28/04/1982
4. Nơi sinh: Quảng Ninh
5. Quyết định công nhận nghiên cứu sinh: số 3304/QĐ-ĐHKHTN ngày 11/11/2022 của Hiệu trưởng Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội.
6. Các thay đổi trong quá trình đào tạo (nếu có):
7. Tên đề tài luận án: Nghiên cứu phân biệt và phân loại các sản phẩm nước cam trên cơ sở phân tích thành phần hóa học, tín hiệu đo kết hợp với máy học.
8. Chuyên ngành: Hóa phân tích
9. Mã số: 9440118
10. Cán bộ hướng dẫn khoa học:
 1. PGS.TS. Tạ Thị Thảo - Trường Đại học KHTN - Đại học Quốc gia HN
 2. PGS.TS. Lê Thị Hồng Hảo - Viện KN An toàn vệ sinh Thực phẩm Quốc gia
11. Tóm tắt các kết quả mới của luận án:

Mục đích:

- Nghiên cứu phát triển phương pháp sắc ký lỏng để định lượng các nhóm chất hóa học chính có mặt trong nước cam
- Xây dựng các thuật toán học máy để phân biệt, phân loại các sản phẩm nước cam dựa trên các dữ liệu đầu vào như thành phần hóa học, dữ liệu phổ UV-Vis, dữ liệu điện hóa.

Đối tượng nghiên cứu:

Đối tượng nghiên cứu là quả cam với các giống khác nhau được trồng ở các tỉnh của Việt Nam và nước cam đang được bán trên thị trường Hà Nội.

Các phương pháp nghiên cứu đã sử dụng:

- Xây dựng và xác nhận giá trị sử dụng các phương pháp định lượng thành phần hóa học chính trong nước cam bao gồm nhóm acid hữu cơ, carotenoid và flavonoid bằng sắc ký lỏng hiệu năng cao
- Phân biệt, phân loại các sản phẩm nước cam dựa trên thành phần hóa học
- Phân biệt, phân loại các sản phẩm nước cam dựa trên dữ liệu phổ
- Phân biệt, phân loại các sản phẩm nước cam dựa trên dữ liệu điện hóa

Các kết quả chính:

- Đã phát triển phương pháp HPLC định lượng đồng thời năm acid hữu cơ gồm: acid oxalic, acid malic, acid ascorbic, acid citric, acid fumaric với RSD từ 1,3% - 3,5% và độ thu hồi của nằm trong khoảng 90% – 107%. Ba carotenoid gồm: lutein, lycopene và

β -caroten với RSD từ 0,3% - 2,4% và độ thu hồi nằm trong khoảng 80 - 110%. Bốn flavonoid gồm: naringin, hesperidin, neohesperidin và quercetin với RSD từ 1,35% - 3,32% và độ thu hồi nằm trong khoảng 80 - 110%.

- Đã xây dựng được phương pháp phân loại nước cam theo giống và theo nguồn gốc địa lý dựa trên thành phần hóa học kết hợp máy học. Để giảm thiểu sự mất cân bằng dữ liệu, bộ dữ liệu hàm lượng được tiền xử lý với SMOOTE. Với tiêu chí phân loại theo giống, thuật toán MLP cho khả năng phân loại đạt 93,1%. Với tiêu chí phân loại theo vùng địa lý, các thuật toán SVC và RF đạt độ chính xác phân loại tuyệt đối.

- Đã xây dựng được phương pháp xác thực nguồn gốc địa lý của nước cam dựa trên dữ liệu phổ UV - Vis kết hợp máy học. Bộ dữ liệu phổ được tiền xử lý bằng chuẩn hóa standard normal variate (SNV) kết hợp làm trơn Savitzky-Golay (SG). Khi sử dụng dải phổ hẹp 250 nm - 400 nm, LDA cho khả năng phân loại 100%, khi sử dụng toàn phổ (250 - 1000 nm) cả 3 thuật toán LDA, RF, ANN cho khả năng phân loại 91,3 - 95,6%. Tiền xử lý bằng SNV + SG kết hợp thêm đạo hàm bậc nhất, ANN đạt 100% trên cả tập huấn luyện và tập kiểm tra.

- Đã xây dựng được phương pháp phân biệt nước cam tự nhiên với nước cam hóa học dựa trên dấu vân tay điện hóa kết hợp máy học. Các dữ liệu điện hóa được tiền xử lý bằng SVN kết hợp với SG, sau đó được giảm chiều bằng iPLS. Thuật toán RF cho khả năng phân loại nước cam tự nhiên bị pha trộn đạt 95%, tuy nhiên khả năng phân biệt hạn chế khi mức độ pha trộn ở mức tỉ lệ ngang nhau cỡ 45 - 50% pha trộn.

Các đóng góp mới:

Hiện tại chưa có nghiên cứu nào xây dựng bộ dữ liệu về thành phần hóa học trong Cam Việt Nam. Cho đến nay cũng chưa có công bố nào xác thực và phân loại nước cam Việt Nam dựa trên các thành phần hóa học hay các tín hiệu phân tích. Luận án góp phần giải quyết hai vấn đề trên.

- Đã xây dựng và thẩm định được phương pháp định lượng các thành phần hóa học chính có trong nước cam (acid hữu cơ, carotenoid, flavonoid). Đây là nghiên cứu toàn diện, đóng góp nâng cao chất lượng công tác kiểm nghiệm an toàn thực phẩm.

- Ứng dụng học máy xây dựng được các mô hình phân loại cam theo giống và theo nguồn gốc địa lý, phân biệt được nước cam tự nhiên và nước cam pha chế dựa trên thành phần hóa học và các tín hiệu phân tích (UV, Volt-ampere).

Ý nghĩa khoa học và thực tiễn:

Luận án đã xây dựng được các phương pháp định lượng 3 nhóm hợp chất chính có mặt trong nước cam bao gồm: acid hữu cơ, carotenoid, flavonoid, các phương pháp này đang được triển khai thường quy tại Viện Kiểm nghiệm an toàn vệ sinh thực phẩm Quốc gia. Trong đó có 02 phép thử (acid, carotenoid) đã được công nhận phù hợp theo tiêu chuẩn ISO 17025:2017. Số hiệu phòng thử nghiệm VILAS 203.

Kết quả luận án mở ra cơ hội ứng dụng cho các cơ sở kiểm nghiệm trong việc xây dựng được bộ dữ liệu nhận dạng, phân biệt các sản phẩm nước cam trên thị trường nhằm đánh giá tính xác thực các công bố của nhà sản xuất. Đồng thời, việc truy xuất nguồn gốc của các sản phẩm cam tươi có thể triển khai theo nhiều cách tùy thuộc điều kiện phòng thí nghiệm. Hơn nữa có thể mở rộng ứng dụng cho nhiều loại thực phẩm, dược liệu, dược phẩm..., góp phần giải quyết các vấn đề thực tế hiện nay tại Việt Nam.

Kết quả của luận án có thể áp dụng trên thực tế kiểm nghiệm ở nhiều đơn vị, do đề tài tiến hành thử nghiệm trên nhiều loại thiết bị như UV, HPLC hay cực phổ. Các cơ sở có thể lựa chọn phương pháp phù hợp với điều kiện thực tế tại đơn vị mình. Hơn nữa, các thuật toán được xây dựng trên phần mềm mở Python nên không bị giới hạn bởi bản quyền, kinh phí và dễ dàng thay đổi thuật toán cho người sử dụng.

12. Các hướng nghiên cứu tiếp theo:

Xây dựng bộ dữ liệu thành phần hóa học và hình ảnh, từ đó xây dựng được các mô hình phân loại cam theo chất lượng (độ ngọt, độ tươi...) dựa trên hình ảnh mà không cần phá hủy mẫu. Từ đó ứng dụng vào các dây chuyền trong sơ chế cam trước khi đưa ra thị trường.

Xây dựng bộ dữ liệu hình ảnh của cam được trồng ở các vùng khác nhau, xây dựng các thuật toán, kết hợp IoT để có thể nhanh chóng truy xuất nguồn gốc địa lý của quả cam bằng thiết bị điện thoại thông minh.

13. Các công trình công bố liên quan đến luận án:

1. Nguyễn Mạnh Sơn, Nguyễn Đức Phong, Bùi Xuân Thành, Tạ Thị Thảo, Lê Thị Hồng Hảo, **Nguyễn Đức Thanh** (2024), “Một số ứng dụng của học máy kết hợp với thị giác máy tính, mũi điện tử và phân tích không mục tiêu trong kiểm soát chất lượng thực phẩm”, *Kỷ yếu Hội nghị khoa học quốc tế về Kiểm nghiệm thực phẩm*, tr. 313-328.

2. Nguyen Manh Son, Nguyen Duc Phong, Nguyen Quoc Hieu, Nguyen Van Dai, Tran Khanh Du, Vu Tien Dung, Ta Thi Thao, Bui Xuan Thanh, Nguyen Trong Hieu, Vu Duong, Nguyen Thi Van Anh, **Nguyen Duc Thanh** (2024), “Predicting orange’s sweetness using deep learning coupled with computer vision”, Báo cáo poster, *Hội nghị khoa học quốc tế về Kiểm nghiệm thực phẩm*.

3. **Nguyễn Đức Thanh**, Dương Thị Dung, Nguyễn Thị Vân Anh, Vũ Thị Trang, Lê Thị Hồng Hảo, Tạ Thị Thảo (2024), “Nghiên cứu định lượng đồng thời một số carotenoid trong nước cam tự nhiên”, *Tạp chí Phân tích Hóa, Lý và Sinh học* Tập 30(1), tr. 8 - 17.

4. **Nguyễn Đức Thanh**, Nguyễn Thị Quyên, Nguyễn Thị Vân Anh, Lê Việt Ngân, Lê Thị Hồng Hảo, Tạ Thị Thảo (2024), “Nghiên cứu định lượng đồng thời một số flavonoid trong nước cam Việt Nam Bằng phương pháp sắc ký lỏng hiệu năng cao”, *Tạp*

chí Khoa học và Công nghệ Trường Đại học Công nghiệp Hà nội Tập 60(11) , tr. 245 - 253.

5. **Nguyen Duc Thanh**, Nguyen Manh Son, Nguyen Duc Phong, Pham Huu Vang, Ninh Duc Ha, Nguyen Thi Van Anh, Le Thi Hong Hao, Nguyen Manh Ha, Nguyen Thi Kim Thuong, Ta Thi Thao (2026), “Rapid Identification of Orange Juice Adulteration Using Voltammetric Profiling and Machine Learning”, *ASEAN Journal of Scientific and Technological Report* Vol. 29(2), e260129, DOI: <https://doi.org/10.55164/ajstr.v29i2.260129>, Q4, Scopus.

6. **Duc Thanh Nguyen**, Manh Son Nguyen, Huu Vang Pham, Manh Ha Nguyen, Duc Phong Nguyen, Thi Van Anh Nguyen, Thi Hong Hao Le, Ba Lich Pham, Thi Thao Ta (2026), “Enhancing geographical origin classification of Vietnamese oranges via hybrid preprocessing and supervised learning of UV-Vis spectral data”, *Talanta Open* (2026), DOI: <https://doi.org/10.1016/j.talo.2026.100625>, Q2, IF 3.7.

Hà Nội, ngày 20 tháng 6 năm 2026

TM. Tập thể hướng dẫn
(Ký và ghi rõ họ tên)

Nghiên cứu sinh
(Ký và ghi rõ họ tên)

PGS.TS. Tạ Thị Thảo

Nguyễn Đức Thanh

INFORMATION ON DOCTORAL THESIS

1. Full name: Nguyen Duc Thanh
2. Sex: Male
3. Date of birth: 28/04/1982
4. Place of birth: Quang Ninh
5. Admission decision number 3304/QĐ-ĐHKHTN dated 11/11/2022 by Rector of VNU University of Science, Ha Noi.
6. Changes in academic process:
7. Official thesis title: Authentication and classification of orange juice based on chemical composition and analytical signals coupled with machine learning.
8. Major: Analytical Chemistry
9. Code: 9440118
10. Supervisors:
 1. Assoc. Prof. Dr. Ta Thi Thao - VNU University of Science, Ha Noi
 2. Assoc. Prof. Dr. Lê Thị Hồng Hảo - National Institute for Food Control
11. Summary of the new findings of the thesis

Purpose:

- To research and develop liquid chromatography methods for the quantification of major chemical groups present in orange juice.
- To construct machine learning algorithms to authenticate and classify orange juice based on input data such as chemical composition, UV-Vis spectral data, and electrochemical data.

Objectives

The research subjects are various orange cultivars grown in different provinces of Vietnam and orange juice products currently being sold in the Hanoi market.

Research methods:

- Establishing and validating analytical procedures for the quantification of primary chemical constituents in orange juice, including organic acid, carotenoids, and flavonoids, using high-performance liquid chromatography.
- Authenticating and classifying orange juice based on chemical composition.
- Authenticating and classifying orange juice based on spectral data.
- Authenticating and classifying orange juice based on electrochemical data.

The major results:

- Development of HPLC methods for quantifying five organic acids: oxalic, malic, ascorbic, citric, and fumaric acids with RSD: 1.3% - 3.5% and recovery: 90% - 107%. Three carotenoids: lutein, lycopene, and beta-carotene with RSD: 0.3% - 2.4% and

recovery: 80% - 110%. Four flavonoids: naringin, hesperidin, neohesperidin, and quercetin with RSD: 1.35% - 3.32% and recovery: 80% - 110%.

- Authentication of orange juice by cultivar and geographical origin using chemical composition combined with machine learning, pre-processed data with SMOTE. MLP achieved 93.1% accuracy for cultivar classification; SVC and RF reached absolute accuracy for geographical origin.

- Authentication of geographical origin using UV-Vis spectroscopy combined with machine learning, pre-processed data with SNV and Savitzky-Golay (SG) smoothing. LDA gave 100% accuracy in 250 - 400 nm range. Full spectrum (250 - 1000 nm) with SNV+SG and first-order derivative via ANN reached 100% accuracy.

- Differentiation between natural and chemically adulterated orange juice using electrochemical fingerprints combined with machine learning. Pre-processed data by SNV+SG; dimensionality reduced using iPLS. RF algorithm achieved 95% accuracy.

Novel contributions:

To date, there has been a lack of comprehensive databases regarding the chemical composition of Vietnamese oranges. Furthermore, no published studies have addressed the authentication and classification of Vietnamese orange juice based on either chemical constituents or analytical profiles. This dissertation directly addresses these two critical gaps:

- Robust analytical methods were successfully developed and validated for the quantification of the primary chemical classes in orange juice, namely organic acids, carotenoids, and flavonoids. This comprehensive study provides a significant advancement in enhancing the efficacy of food safety control and quality assurance.

- Advanced machine learning algorithms were deployed to construct predictive models for classifying oranges based on cultivar and geographical origin. Additionally, these models successfully differentiated authentic (natural) orange juice from formulated products utilizing chemical compositions and analytical data (UV-Vis and voltammetry).

Scientific and practical significance:

This research successfully established quantitative methods for three major compound classes in orange juice (organic acids, carotenoids, and flavonoids). These methods are currently implemented for routine analysis at the Vietnam national institute for food control. Notably, two of these analytical assays (organic acids and carotenoids) have been formally accredited under the ISO/IEC 17025:2017 standard (laboratory code: VILAS 203).

The findings of this dissertation offer substantial opportunities for testing laboratories to build reference databases for identification and differentiation. This enables the

verification of commercial orange juice products against manufacturer claims. Concurrently, accurate geographical traceability of fresh oranges can be tailored and deployed via multiple analytical workflows depending on individual laboratory infrastructure. Furthermore, this methodological framework can be scaled and extended to various other matrices, including agricultural foods, medicinal herbs, and pharmaceuticals, thereby addressing urgent, real-world quality control challenges in Vietnam.

These outcomes are transferable to diverse testing facilities, as the experimental workflows were validated across multiple instrumental platforms, including UV-Vis, HPLC, and polarography. Laboratories can thus select the most compatible approach aligned with their existing instrumentation. Moreover, for classification and discrimination tasks, the chemometric and machine learning algorithms were developed using the open-source Python environment. This eliminates licensing costs, lowers financial barriers, and ensures flexible algorithmic customization for end-users.

12. Further research directions

Non-destructive quality assessment: to construct an integrated database combining chemical profiles and digital imagery, subsequently training classification models to evaluate orange quality attributes (e.g., sweetness, freshness) via a non-destructive imaging approach. This will facilitate automated sorting in post-harvest processing lines prior to market distribution.

Smart traceability systems: to compile a comprehensive image database of oranges cultivated across diverse geographical regions. By integrating computer vision algorithms with IoT frameworks, this will enable rapid, smartphone-based geographical traceability of orange fruits directly in the field or marketplace.

13. Thesis-related publications:

1. Nguyen Manh Son, Nguyen Duc Phong, Bui Xuan Thanh, Ta Thi Thao, Le Thi Hong Hao, **Nguyen Duc Thanh** (2024), “Applications of machine learning combined with computer vision, electronic nose, and non-targeted analysis in food quality control”, *Proceedings of the International Conference on Food Control*, pp. 313 - 328. [in Vietnamese].

2. Nguyen Manh Son, Nguyen Duc Phong, Nguyen Quoc Hieu, Nguyen Van Dai, Tran Khanh Du, Vu Tien Dung, Ta Thi Thao, Bui Xuan Thanh, Nguyen Trong Hieu, Vu Duong, Nguyen Thi Van Anh, **Nguyen Duc Thanh** (2024), “Predicting orange’s sweetness using deep learning coupled with computer vision” [Poster presentation], *International Conference on Food Control*.

3. **Nguyen Duc Thanh**, Duong Thi Dung, Nguyen Thi Van Anh, Vu Thi Trang, Le Thi Hong Hao, Ta Thi Thao (2024), “Simultaneous quantification of selected

carotenoids in natural orange juice”, *Journal of Analytical Chemistry, Physics and Biology*, Vol. 30(1), pp. 8 - 17. [in Vietnamese].

4. **Nguyen Duc Thanh**, Nguyen Thi Quyen, Nguyen Thi Van Anh, Le Viet Ngan, Le Thi Hong Hao, Ta Thi Thao (2024), “Simultaneous determination of selected flavonoids in Vietnamese orange juice using high-performance liquid chromatography”, *Journal of Science and Technology - Hanoi University of Industry*, Vol. 60(11), 245 - 253. [in Vietnamese].

5. **Nguyen Duc Thanh**, Nguyen Manh Son, Nguyen Duc Phong, Pham Huu Vang, Ninh Duc Ha, Nguyen Thi Van Anh, Le Thi Hong Hao, Nguyen Manh Ha, Nguyen Thi Kim Thuong, Ta Thi Thao (2026), “Rapid Identification of Orange Juice Adulteration Using Voltammetric Profiling and Machine Learning”, *ASEAN Journal of Scientific and Technological Report* Vol. 29(2), e260129, DOI: <https://doi.org/10.55164/ajstr.v29i2.260129>, Q4, Scopus.

6. **Duc Thanh Nguyen**, Manh Son Nguyen, Huu Vang Pham, Manh Ha Nguyen, Duc Phong Nguyen, Thi Van Anh Nguyen, Thi Hong Hao Le, Ba Lich Pham, Thi Thao Ta (2026), “Enhancing geographical origin classification of Vietnamese oranges via hybrid preprocessing and supervised learning of UV-Vis spectral data”, *Talanta Open* (2026), DOI: <https://doi.org/10.1016/j.talo.2026.100625>, Q2, IF 3.7.

Date: 20/06/2026

On behalf of academic supervisors

PhD. Student

Assoc. Prof. Dr. Ta Thi Thao

Nguyen Duc Thanh