ViFLAVOR: NHẬN DẠNG ẨM THỰC VIỆT TỪ CẤP ĐỘ ƯU TIÊN CỦA NGUYÊN LIỆU

Mã Trường Thành, Huỳnh Gia Khương, Đổ Hiếu Nghĩa, Đỗ Thanh Nghị

Trường Công nghệ thông tin và Truyền thông, Trường Đại học Cần Thơ, Cần Thơ, Việt Nam

*{mtthanh, hgkhuong, dtnghi}@ctu.edu.vn, nghiab2016985@student.ctu.edu.vn*

**TÓM TẮT**— Ẩm thực Việt là một phần không thể thiếu trong văn hóa và đời sống của người Việt Nam. Với sự phong phú, đa dạng về các thành phần nguyên liệu và hương vị đặc trưng, ẩm thực Việt Nam đã thu hút sự quan tâm, yêu thích và trở thành nguồn cảm hứng vô tận cho các đầu bếp và nhà nghiên cứu ẩm thực trên khắp thế giới. Tuy nhiên, đối với những người ngoại quốc hoặc những người không tìm hiểu nhiều về ẩm thực địa phương thì việc nhận dạng các món ăn đặc sản của Việt Nam sẽ có thể gặp khó khăn vì đặc điểm và cách chế biến của chúng thường mang những đặc trưng riêng biệt cho từng vùng miền, chẳng hạn, thành phần bún riêu cua của các miền là khác nhau. Vì vậy, một hệ thống trí tuệ nhân tạo (AI) có khả năng nhận biết món ăn đặc sản của Việt Nam thông qua việc xác định các thành phần nguyên liệu thì hoàn toàn mong đợi. Hệ thống sẽ giúp cho việc tìm hiểu và trải nghiệm ẩm thực địa phương trở nên dễ dàng hơn, đặc biệt là đối với du khách quốc tế. Nghiên cứu này sẽ tập trung phát hiện và nhận dạng các thành phần nguyên liệu trên ảnh chụp về món ăn. Từ đó, dựa trên các thành phần nguyên liệu được tìm thấy, giải thuật D2D đã đã được giới thiệu để xác định tên món ăn. Kết quả phân loại món ăn của mô hình đề xuất đạt hiệu quả mong đợi trên 94%. Cuối cùng, chúng tôi đã cài đặt một Website để triển khai ViFLAVOR và cung cấp mã nguồn trên Github.

**Từ khóa**— Món ăn Việt Nam, Thị giác máy tính, Phát hiện đối tượng, Giải thuật máy học, CNN

# GIỚI THIỆU

Ẩm thực Việt Nam được xem là một nét đặc trưng không thể thiếu trong văn hóa dân tộc. Mỗi vùng miền khác nhau sẽ có những món ăn đặc sản riêng. Bên cạnh sự đa dạng về các món ăn thì sự đa dạng và phong phú giữa các thành phần nguyên liệu từ món ăn cũng đóng vai trò quan trọng trong việc thể hiện được giá trị văn hóa, đặc trưng rất riêng của ẩm thực ở từng địa phương. Chính hương vị độc đáo đã để lại ấn tượng khó quên trong lòng du khách mỗi khi đặt chân đến và thưởng thức. Ẩm thực còn được xem là cầu nối giao thương văn hóa, kinh tế với các quốc gia trên thế giới. Nó đóng vai trò quan trọng trong việc lan tỏa hình ảnh văn hóa của đất nước ra thế giới, thu hút du khách và đóng góp vào sự phát triển kinh tế, du lịch của nước nhà. Nền ẩm thực nước ta có văn hóa, địa lý rộng lớn, thường xuyên giao thoa với các quốc gia láng giềng, cộng với lịch sử phát triển lâu đời, thường xuyên cập nhật về các nguyên liệu, quá trình chế biến, tên gọi cho nên có thể gây ra nhầm lẫn với du khách hoặc người Việt không quen thuộc với các món ăn Việt Nam. Trong quá trình du lịch và khám phá ẩm thực Việt Nam, không hiếm trường hợp du khách quốc tế hoặc người Việt không biết rõ về tên gọi chính xác của các món ăn đặc sản của địa phương mà họ thưởng thức. Những nhầm lẫn này thường xuyên xảy ra, khiến cho việc trải nghiệm ẩm thực trở nên khó khăn hơn. Tuy nhiên, điều mà mỗi du khách vẫn mong muốn đó là biết rõ về các món ăn, biết thêm về một nét đặc sắc riêng của vùng miền họ đang khám phá trên tinh thần học hỏi và tôn trọng văn hóa đất nước mình ghé thăm.

Trong những năm gần đây, những nghiên cứu nổi trội về ẩm thực có ứng dụng trí tuệ nhân tạo vào với các mục đích khác nhau nhưng đều có điểm chung là mong muốn giúp cho mọi người hiểu rõ về các món ăn mà họ được thưởng thức để biết được sự độc đáo của chúng. Một trong những nghiên cứu đầu tiên mà chúng tôi muốn đề cập liên quan đến việc phân loại được các loại thức ăn nhanh xem chúng có tốt cho sức khỏe của người thưởng thức hay không. Nghiên cứu của Yian et al. [10] tạo ra một tập dữ liệu gồm các mẫu dữ liệu ảnh chụp từ 20 loại thức ăn không tốt cho sức khỏe (i.e., burger, hotdog, pizza, bánh phô mai…) đã xây dựng mô hình có khả năng nhận dạng thực phẩm không lành mạnh. Với một số kĩ thuật tăng cường dữ liệu kèm sử dụng mô hình Yolov3 đem lại kết quả xấp xỉ 98.05%. Nghiên cứu góp phần vào việc giúp người dùng có thể nhận diện và hạn chế sử dụng các loại thức ăn gây hại cho cơ thể. Bên cạnh nghiên cứu trên, nhóm tác giả tại trường Đại học Cần Thơ đã đề xuất một hướng tiếp cận khác liên quan đến ẩm thực. Nghiên cứu của Thành et al. [8] đã đề xuất sử dụng những mô hình CNN và ontology với mục đích là bảo tồn các món ăn đặc trưng Việt Nam. Ý tưởng này xuất phát từ việc nhận thấy hiện nay có rất ít nghiên cứu và ứng dụng trí tuệ nhân tạo (AI) tập trung vào lĩnh vực bảo tồn các giá trị truyền thống văn hóa ẩm thực, hầu hết những nghiên cứu chỉ tập trung vào phân lớp hình ảnh và thiếu thông tin toàn diện của từng món ăn. Nhóm tác giả đã xây dựng một bản thể luận (ontology) các món ăn đặc sản Việt Nam để lưu trữ thông tin liên quan và mô hình phân lớp hình ảnh các món ăn đặc sản ở vùng đồng bằng sông Cửu Long với độ hiệu quả lên đến 96%. Một nghiên cứu khác tập trung vào việc phân tích các thành phần dinh dưỡng của một món ăn để xây dựng một lối sống ăn uống lành mạnh. Nghiên cứu của Konstantakopoulos et al. [7] tận dụng hai bài toán phổ biến trong thị giác máy tính gồm: phân đoạn đối tượng (object segmentation) và phân lớp đối tượng (object classification). Nhóm tác giả đã tiến hành phân đoạn hình ảnh món ăn, đồng thời ước tính thể tích của món ăn. Từ đó ước lượng được thành phần dinh dưỡng có trong bữa ăn để tính được lượng calories của bữa ăn.

Tiếp nối sự thành công của các nghiên cứu trên, nhóm nghiên cứu mong muốn xây dựng một hệ thống nhận dạng món ăn Việt Nam dựa trên các thành phần nguyên liệu (gọi tắt là NL). Lấy cảm hứng từ những tô (bát) bún mà đã được “xáo trộn” trong khi ăn để đề xuất ý tưởng. Nhìn chung hầu hết các nghiên cứu trước đó chỉ tập trung vào các ảnh món ăn được trình bày “đẹp và ban đầu” mà không quan tâm đến những ảnh mà món ăn đã bị xáo trộn (hoặc đang ăn) sẽ không còn đầy đủ các thành phần *(như ban đầu)*. Hơn nữa, câu hỏi được đặt ra là từ các thành phần “nhìn thấy được” tại thời điểm chụp liệu chúng ta có nhận ra món ăn đó là gì không? Một cách suy luận tự nhiên, chúng ta sẽ đoán bằng cách liệt kê các món ăn mà có thể được tạo thành từ các nguyên liệu. Ý tưởng chung của hướng tiếp cận này là kết hợp phát hiện đối tượng (OD, Object Detection) và ý tưởng của giải thuật KNN (ở đây sử dụng 1-NN), tuy nhiên nghiên cứu sẽ không sử dụng đặc trưng ảnh (i.e., SIFT, GIST, HOG, …) mà sử dụng đặc trưng là các NL phát hiện được từ OD; hơn nữa, những đặc trưng này sẽ được kết hợp với “khoảng cách của độ ưu tiên phân tầng” để giải quyết bài toán phân lớp món ăn. Từ ý tưởng đó, những đóng góp chính của bài báo như sau:

* ***Xây dựng được tập dữ liệu các món ăn đặc sản Việt Nam***. Để có được bộ dữ liệu này, chúng tôi đã thực hiện thu thập những hình ảnh về món ăn đặc sản Việt Nam. Nguồn ảnh được thu thập từ 02 nguồn: (1) Google Images, (2) tự chụp ảnh món ăn ngoài thực tế. Tập dữ liệu trong nghiên cứu này với khoảng trên 3400 ảnh cho 14 lớp;
* ***Huấn luyện dữ liệu với ba mô hình phát hiện*** (object detection) thành phần nguyên liệu món ăn gồm: Yolov8 [3], RT-DETR [11], Yolov9 [1]. Mỗi mô hình đều được tinh chỉnh các siêu tham số sao cho mô hình đạt được độ chính xác cao nhưng hạn chế overfitting. Nhìn chung, kết quả thực nghiệm của hầu hết các mô hình đều có độ chính xác cao và kích thước phù hợp nhằm đáp ứng tốt yêu cầu về *“thời gian thực”;*
* ***Giải thuật D2D*** (Distance to Decide) đã được đề xuất để xác định món ăn dựa trên khoảng cách;
* ***Triển khai framework đề xuất trên ứng dụng web*** *(sử dụng Vuejs và Flask framework)* để đưa mô hình vào ứng dụng thực tế: Các mô hình AI được xây dựng, nghiên cứu, phát triển đều cần phải đáp ứng được tiêu chí ứng dụng được vào cuộc sống thực tế, hướng tới đối tượng người sử dụng. Giao diện website cũng được thiết kế tối giản, giúp người dùng dễ dàng sử dụng hệ thống dù là lần đầu tiên tiếp xúc. Ở đây, ViFLAVOR sử dụng Vuejs để trình bày (hiển thị) thông tin trong khi Flask sẽ xử lý các mô hình máy học.

Trong bài báo này, chúng tôi khám phá mô hình phân lớp với hai hướng tiếp cận để so sánh, đánh giá và xác định tên của món ăn: toàn cục và cục bộ. (1) Toàn cục - hướng tiếp cận thứ nhất gắn liền với việc phân lớp các ảnh món ăn dựa vào các mô hình học sâu, hướng tiếp cận đang rất phổ biến hiện nay trong việc phân lớp hình ảnh. Các ảnh được đưa vào các mô hình học sâu để huấn luyện và kiểm tra kết quả. Tuy nhiên do sự giống nhau khá rõ từ các lớp ảnh món ăn này nên kết quả phân lớp nhận được qua nhiều mô hình không cao, kết quả này cũng sẽ được trình bày các phần tiếp theo. (2) Cục bộ - hướng tiếp cận thứ hai sẽ quan tâm đến các chi tiết cục bộ trên các ảnh món ăn này, mà cụ thể đó chính là các thành phần của mỗi món ăn. Hướng tiếp cận này sẽ cho ra được một tập thành phần nguyên liệu của một món ăn, sau đó một giải thuật được đề xuất để xác định tên của món ăn cần được nhận dạng. Hướng tiếp cận thứ hai này cho kết quả thực nghiệm cao hơn nhiều so với hướng tiếp cận thứ nhất nên chúng tôi sẽ tích hợp các mô hình trong hướng tiếp cận này vào ứng dụng web cuối cùng để triển khai thực tế.

Lưu ý rằng, hướng tiếp cận cục bộ (2) đã được đề xuất trong bài báo này được hiểu như là một KNN cải tiến *(với “khoảng cách hai chiều” do chúng tôi định nghĩa và xem xét)* cho bài toán phân lớp. Từ ý tưởng này, ViFLAVOR đang hướng tới *giải quyết bài toán dữ liệu ít (few-shot learning)*. Do vậy, tập dữ liệu cho bài báo này sẽ không nhiều “mà chỉ đủ” để các mô hình CNN *(hướng tiếp cận toàn cục (1))* không bị học vẹt (overfitting). Một điểm lưu ý về đóng góp của bài báo này là chúng tôi không tập trung vào ứng dụng/hệ thống mà tập trung chính là đề xuẩt một ý tưởng sáng tạo về việc phân lớp các món ăn dựa trên các nguyên liệu được phát hiện trên ảnh (thị giác máy tính). Có thể nói, đây là bài toán cải tiến cho KNN từ góc nhìn *“đặc trưng thật”,* nó chính là các nguyên liệu trong ảnh (i.e., tôm, gà, bún.) thay vì đặc trưng là CNN, SIFT, GIFT hay HOG (đặc trưng ảnh). Những đặc trưng “thật” này sẽ giúp dễ dàng cho việc giải thích kết quả phân lớp (Explainable AI, XAI). Hơn nữa, trường hợp ***dữ liệu ít hoặc món ăn bị “xáo trộn lên”*** thì CNN hay các phương pháp máy học truyền thống sẽ kém hiệu quả (kết quả đã được kiểm chứng tại phần thực nghiệm). Ý tưởng phân tầng đặc trưng “thật” với gán trọng số cho bài toán món ăn thì chưa được nghiên cứu nào thực hiện trước đó.

Bài báo này được tổ chức như sau: Phần II sẽ giới thiệu những cơ sở lý thuyết được sử dụng trong nghiên cứu. Tiếp đến, phần III sẽ mô tả chi tiết mô hình đề xuất, bao gồm cả các thành phần và quy trình thực hiện. Trong phần IV, chúng tôi sẽ trình bày các thí nghiệm được thực hiện trên hệ thống, đồng thời thảo luận về kết quả đạt được khi áp dụng mô hình trên nền tảng Web. Cuối cùng, phần V sẽ đưa ra những kết luận quan trọng từ nghiên cứu này và đề xuất hướng phát triển cho các nghiên cứu trong tương lai.

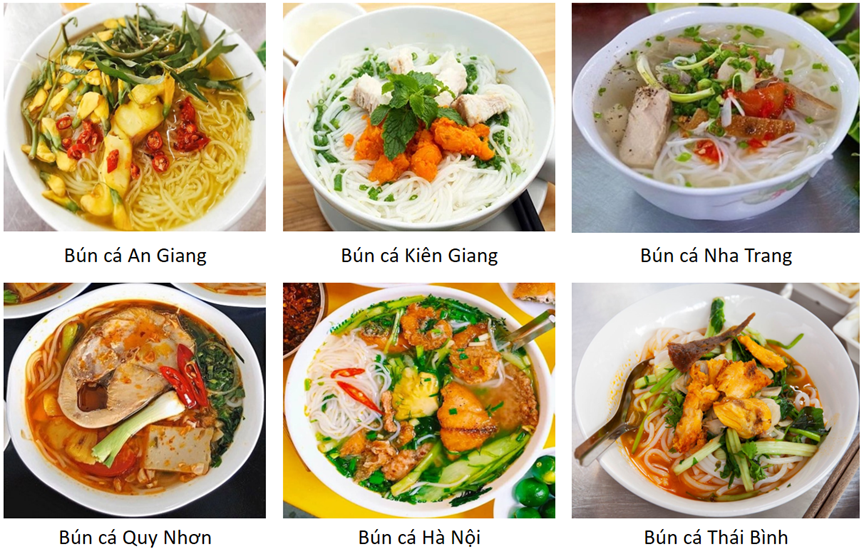
# CƠ SỞ LÝ THUYẾT

Trong mục này, những đặc trưng về ẩm thực Việt và những kỹ thuật trí tuệ nhân tạo được vận dụng trong nghiên cứu này sẽ được giới thiệu.

## Ẩm thực Việt Nam

Ẩm thực Việt Nam[[1]](#footnote-1) [15, 16] bao gồm rất nhiều món ăn đặc sắc với hương vị rất riêng cho từng vùng miền, nó phản ánh được sự đặc trưng trong nền văn hóa của nước ta. Ngoài sự đa dạng về các món ăn thì các nguyên liệu tạo ra món ăn cũng đa dạng và phong phú không kém. Ví dụ, trong món phở, sự kết hợp giữa nước dùng từ xương gà hoặc bò, thịt, đi kèm các loại gia vị như gừng, hành, hạt tiêu được ăn kèm rau sống như húng quế, ngò gai, rau mùi tùy thuộc vào vùng miền khác nhau. Bún cá thường được làm từ bún, cùng với nước dùng được nấu từ xương cá hoặc hải sản khác như tôm, cua, hay cá lóc. Cá được cắt thành từng miếng nhỏ và nấu chín mềm trong nước dùng, được phục vụ cùng với tôm và các loại rau như rau răm, lá chuối, lá lốt. Chỉ cần minh họa qua hai món ăn thì đã thấy sự đa dạng từ nguyên liệu cùng cách chế biến trong ẩm thực Việt Nam. Bên cạnh đó, mỗi vùng miền của Việt Nam lại có những nguyên liệu đặc trưng riêng, phản ánh điều kiện tự nhiên và văn hóa của địa phương đó. Ví dụ, miền Trung có nhiều món sử dụng các loại tôm, cá, sò điệp từ biển, trong khi miền Nam thường sử dụng nhiều loại rau cải và đậu, còn miền Bắc có nhiều món ăn sử dụng thịt lợn, gà, và các loại rau sống. Mỗi vùng miền, mỗi địa phương sẽ có xu hướng ăn uống riêng kết hợp với đa dạng nguồn nguyên liệu, dẫn đến nét riêng trong công đoạn sử dụng các thành phần này, góp phần tạo nên sự đa dạng cho cách thức chế biến món ăn. Hình 1 thể hiện rõ vấn đề tuy cùng một món bún cá nhưng khi mang về mỗi tỉnh thành lại có sự khác biệt trong tên gọi lẫn hương vị đặc trưng riêng của địa phương.

Nghiên cứu này sẽ đề cập đến 14 món ăn Việt tiêu biểu của ba miền đất nước (Bắc, Trung, Nam). Đáng chú ý rằng, một số món ăn có nét đặc trưng giống nhau cũng được xem xét trong nghiên cứu này. Chẳng hạn, món “Mì Quảng” và “Cao lầu” là hai món ăn mà gần như giống nhau về các thành phần chính. Một số thành phần phụ sẽ giúp phân biệt giữa hai đặc sản này. Hơn nữa, việc lựa chọn đa dạng các món ăn nước như phở, bún cá, bún nước lèo, hủ tiếu (gọi chung tiếng anh là rice noodles) sẽ là một trong những thách thức cần được quan tâm cho hướng tiếp cận trong nghiên cứu này. Để dễ dàng cho việc trình bày, một cách hình thức, chúng tôi sẽ gọi là tập các món ăn và từng món ăn sẽ chỉ định là . Khi đó, chúng ta sẽ có .



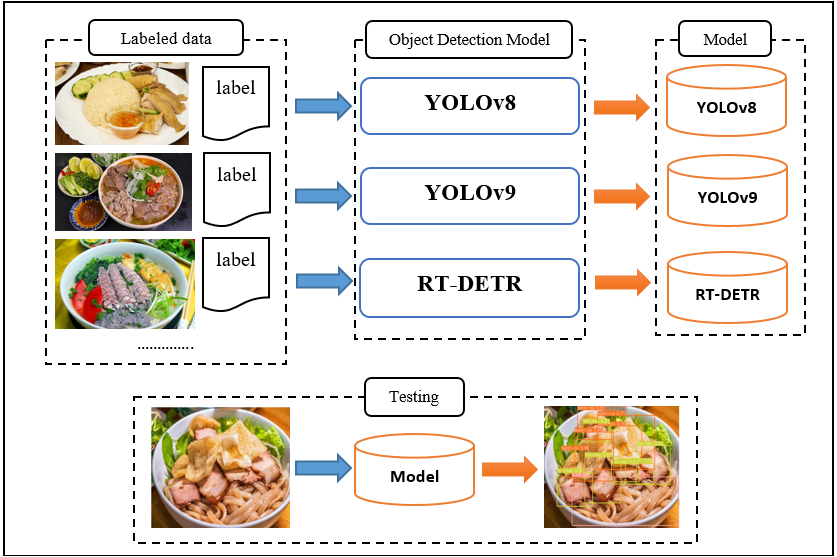
1. Các món ăn đặc sản Việt Nam

## Kế tiếp, phương pháp dò tìm nguyên liệu trên và phân lớp ảnh sẽ được giới thiệu trong những phần ngay sau đây. Lưu ý rằng, việc dò tìm đối tượng sẽ được sử dụng cho hướng tiếp cận cục bộ (local) (phân lớp từng nguyên liệu được dò tìm) và phân lớp với mạng tích chập học sâu sẽ được vận dụng trong hướng tiếp cận toàn cục (phân lớp với đầu vào là ảnh “nguyên chất” mà không xem xét từng đối tượng trên ảnh).

## Phát hiện đối tượng

Phát hiện đối tượng [14] (object detection, OD) là một kỹ thuật trong thị giác máy tính giúp nhận diện và xác định vị trí của các đối tượng có trong hình ảnh. Hệ thống phát hiện đối tượng sẽ trả về tọa độ của các đối tượng trong ảnh mà nó đã được huấn luyện để nhận dạng. Hệ thống cũng sẽ trả về mức độ tin cậy và các tọa độ điểm (vùng (Region of Interest, viết tắt là ROI) được dò tìm), chỉ số này thể hiện mức độ tin cậy của hệ thống đối với dự đoán là chính xác. Mục tiêu của việc phát hiện đối tượng là trả lời cho câu hỏi: Đối tượng là gì và ở đâu?

Phát hiện đối tượng phục vụ nhiều cho các công việc liên quan đến thị giác máy tính như phân đoạn đối tượng (object segmentation), chú thích hình ảnh (image captioning), theo dõi đối tượng (object tracking), … Trong những năm gần đây, từ sự phát triển nhanh chóng của các kiến trúc học sâu, đã thúc đẩy sự tiến bộ của các mô hình phát hiện đối tượng. Trong nghiên cứu này hệ thống sẽ tận dụng các mô hình kiến trúc YOLO (cụ thể là hai kiến trúc mới nhất Yolov8[[2]](#footnote-2) và Yolov9[[3]](#footnote-3)) cùng với mô hình RT-DETR[[4]](#footnote-4) để thử nghiệm và so sánh kết quả. Ở đây, các hình ảnh món ăn *(sau khi được thu thập)* sẽ được gán nhãn các thành phần nguyên liệu. Kế tiếp, mô hình phát hiện đối tượng sẽ huấn luyện. Roboflow đã được sử dụng cho việc gán nhãn và nền tảng Colaboratory của Google đã được sử dụng cho việc huấn luyện. Cuối cùng, mỗi kiến trúc khác nhau sẽ có kết quả mô hình khác nhau, và rồi được sử dụng để phát hiện các nguyên liệu có trong món ăn thông qua ảnh chụp. Trực quan hóa những mô hình trình bày cụ thể qua Hình 2.



1. Xây dựng các mô hình phát hiện đối tượng từ tập dữ liệu đã gán nhãn

Lưu ý rằng, nghiên cứu này không gộp chung cả ba mô hình cho việc “dò tìm và phát hiện” mà vận dụng riêng từng mô hình với giải thuật D2D. Một cách hình thức, chức năng / hàm để phát hiện đối tượng sẽ được chỉ định là , trong đó là chức năng (Detection) và là mô hình sử dụng (). Chẳng hạn, là chức năng dò tìm đối tượng sử dụng mô hình Yolov8.

## Mạng nơ-ron tích chập

Mạng nơ-ron tích chập, thường được gọi là CNN (Convolutional Neural Network), là một kiến trúc mạng nơ-ron nhân tạo được thiết kế đặc biệt để xử lý dữ liệu có cấu trúc lưới, chẳng hạn như hình ảnh. CNN đã trở thành công cụ quan trọng và phổ biến trong lĩnh vực học máy, đặc biệt là trong các ứng dụng liên quan đến thị giác máy tính, như nhận dạng hình ảnh, phát hiện đối tượng và phân loại hình ảnh. Ngày nay, việc sử dụng các mạng học sâu tích chập CNN trong việc phân lớp hình ảnh đã trở nên rất phổ biến. Ở nghiên cứu này, chúng tôi thử nghiệm sử dụng các mạng học sâu phổ biến để phân lớp cho 14 món ăn để kiểm tra hiệu quả phân lớp. Các mô hình học sâu mà được sử dụng trong nghiên cứu này bao gồm các kiến trúc sau: ***LeNet, MobileNet, InceptionV3, VGG16, ResNet101, DenseNet201***. Tập dữ liệu để thực nghiệm được lưu trữ theo cấu trúc thư mục riêng biệt cho từng lớp ảnh món ăn, kết quả sau khi huấn luyện sẽ cho ra được các mô hình phân lớp ảnh món ăn. Các kết quả độ chính xác phân lớp món ăn từ các mô hình được đánh giá bằng tập kiểm tra – tập được tách ra từ tập dữ liệu thực nghiệm ban đầu theo tỷ lệ 7:3 (70% cho huấn luyện và 30% cho kiểm tra). Các kết quả thực nghiệm về các mô hình này sẽ được trình bày trong các phần tiếp theo.

Nhắc lại rằng, đóng góp chính của nghiên cứu này “sẽ không là” bài toán CNN cho việc phân lớp mà chỉ sử dụng để so sánh với mô hình đề xuất. Ngay bây giờ, phương pháp đề xuất sẽ được trình bày trong phần kế tiếp.

# ĐỀ XUẤT PHƯƠNG PHÁP

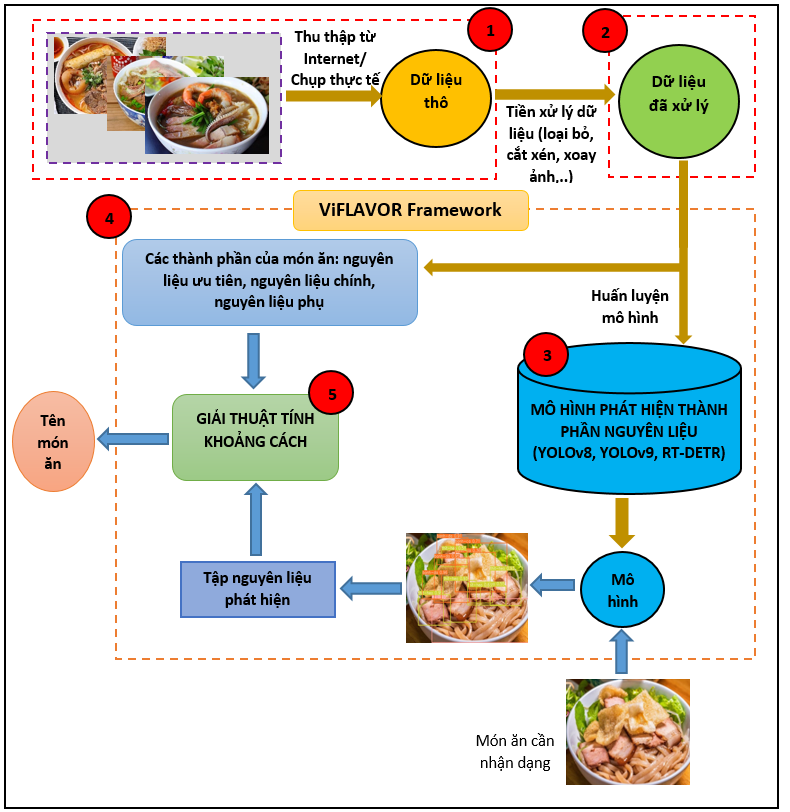
Để xác định các món ăn đặc sản của Việt Nam, phần này sẽ giới thiệu mô hình tổng quan của nghiên cứu. Ngoài ra, giải thuật nhận dạng và phương pháp xác định kết quả cũng sẽ được trình bày chi tiết. Hướng tiếp cận của hệ thống sẽ được triển khai theo cấu trúc như Hình 3.



1. Hướng tiếp cận chính của nghiên cứu

## Mô hình ViFLAVOR

Trong bài báo này, chúng tôi giới thiệu về mô hình ViFLAVOR để nhận dạng các món ăn đặc sản Việt Nam. Kiến trúc của nó gồm năm bước chính được mô tả chi tiết như sau (trong Hình 4):

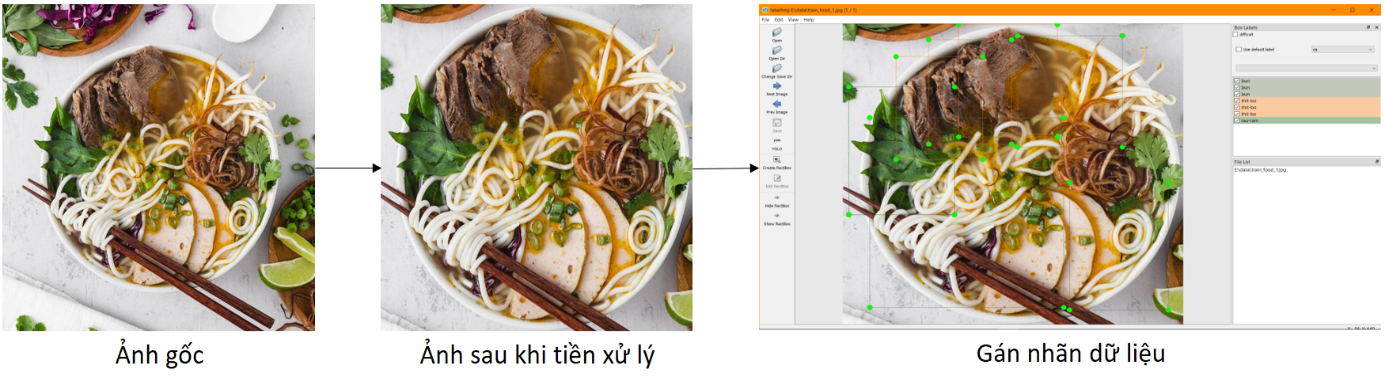


1. Mô hình ViFLAVOR đã đề xuất

Một mô tả chi tiết về hệ thống nhận dạng món ăn như sau:

### **Thu thập hình ảnh và thông tin của món ăn**: thực hiện thu thập hình ảnh của 14 món ăn đặc sản Việt Nam gồm: bún cá, hủ tiếu Mỹ Tho, bún nước lèo, cơm tấm Long Xuyên, bún hải sản bề bề, bánh hỏi heo quay, cơm gà, cao lầu, mì Quảng, bún bò Huế, phở Hà Nội, bún mực, bún mọc, bún đậu mắm tôm. Ngoài ra còn thu thập thêm hình ảnh của một số món ăn khác. Nguồn ảnh được thu thập từ Google Images, kết hợp với chụp ảnh món ăn ngoài thực tế bằng nhiều thiết bị chụp ảnh khác nhau – nhằm tạo nên sự đa dạng cho tập dữ liệu.

### **Tiền xử lý dữ liệu:** ảnh sau khi được thu thập sẽ có thể chứa các thành phần thừa không liên quan đến món ăn nên ViFLAVOR đã cắt ảnh để tập trung vào món ăn trong ảnh. Ngoài ra nghiên cứu còn áp dụng kỹ thuật tăng cường dữ liệu bằng cách xoay ảnh ở nhiều góc độ khác nhau để số lượng dữ liệu được đa dạng hơn. Thực hiện gán nhãn cho các thành phần nguyên liệu trên từng ảnh món ăn bằng công cụ LabelImg. Tổng số lượng nhãn là 27 nhãn ứng với 27 loại nguyên liệu từ các món ăn bao gồm: cá, bún, hủ tiếu, tôm, thịt heo, gan, thịt heo quay, thịt bò, trứng, thịt bằm, tôm tít, mực, viên mọc, bánh hỏi, dưa chua, phở, cơm, sườn, bì, thịt gà, rau răm, mì, bánh đa, chả cốm, dồi sụn, đậu hũ, chả giò. Hình 5 mô tả quá trình tiền xử lý dữ liệu cho một mẫu minh họa.



1. Tiền xử lý và gán nhãn dữ liệu

### **Huấn luyện các mô hình phát hiện đối tượng:** Sau khi thu thập ảnh và tiền xử lý, bước kế tiếp là huấn luyện các mô hình phát hiện đối tượng với Yolov8, Yolov9 và RT-DETR. Ở nội dung này, bài báo sẽ cung cấp, trình bày và so sánh kết quả của các mô hình bằng cách sử dụng chỉ số mAP (Mean Average Precision) để đánh giá hiệu suất của từng mô hình.

### **Phát triển ứng dụng nhận dạng món ăn ViFLAVOR:** Từ mô hình đã đạt được huấn luyện, chúng tôi triển khai mô hình ViFLAVOR trên nền tảng Web sử dụng Flask và Vuejs. Quá trình này còn bao gồm việc kiểm thử mô hình và tích hợp giải pháp vào ứng dụng Web thực tế.

### **Cài đặt giải thuật D2D tính khoảng cách để xác định kết quả**: sau khi sử dụng mô hình phát hiện được tập các thành phần nguyên liệu có trong món ăn thì bước cuối cùng là xác định được tên gọi của món ăn đó. Đây là điều khó khăn vì sự phức tạp trong nguyên liệu của các món ăn và như đề cập ban đầu, mỗi vùng miền lại sử dụng các thành phần đặc trưng riêng ở địa phương của họ. Ngoài ra việc kết hợp các thành phần ăn kèm với món ăn lại khác nhau và rất đa dạng. Vì vậy, một bảng mô tả các thành phần của các món ăn (cụ thể là 14 món ăn như đề cập ở bước 1) bao gồm nguyên liệu ưu tiên, nguyên liệu chính và nguyên liệu phụ đã được cung cấp. Từ bảng thành phần này, một giải thuật tính khoảng cách (giữa các thành phần nguyên liệu phát hiện được và bảng thành phần) đã được đề xuất. Dựa trên khoảng cách tìm được để đưa ra kết luật. Cụ, thể tìm các nguyên liệu trong ảnh mà gần nhất với món ăn trong bảng thành phần. Bảng thành phần và giải thuật này sẽ được trình bày ở mục tiếp theo.

## Bảng thành phần nguyên liệu và giải thuật tính khoảng cách D2D (Distance to Decide)

Bảng 1 trình bày các thành phần nguyên liệu tương ứng với 14 món ăn đặc sản Việt Nam. Ở đây, chúng tôi đã chia ra ba nhóm thành phần chính: nguyên liệu ưu tiên, nguyên liệu chính, và nguyên liệu phụ. Trong Bảng 1 cũng đề cập đến tỉnh thành để phục vụ cho việc xác định vị trí của món ăn cho hệ thống “mở rộng”, tuy nhiên, trong bài báo này chúng tôi sẽ không sử dụng đặc trưng này. Một cách hình thức, tập các nguyên liệu ưu tiên, chính và phụ sẽ được chỉ định lần lượt là . Hơn nữa, chúng ta gọi là tập của chúng, khi đó . Các nguyên liệu sẽ được biểu diễn theo chỉ số như trong Bảng 1, chẳng hạn .

1. Tập các thành phần nguyên liệu (NL) có trong các món ăn đặc sản Việt Nam

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Món ăn (** | **NL ưu tiên** | **NL chính** | **NL phụ** | **Tỉnh thành** | **SL cho phát hiện** | **SL cho**  **Phân lớp** |
| 1 | Bún cá | Bún | Cá | Tôm, rau răm | An Giang | 147 | 91 |
| 2 | Hủ tiếu Mỹ Tho | Bủ tiếu | Thịt băm, thịt heo, gan | Trứng, tôm | Tiền Giang | 195 | 46 |
| 3 | Bún nước lèo | Bún | Cá, tôm, thịt heo quay | Rau răm | Sóc Trăng | 163 | 94 |
| 4 | Cơm tấm Long Xuyên | Cơm | Sườn, bì, trứng | Dưa chua | An Giang | 206 | 159 |
| 5 | Bún hải sản bề bề | Bún | Tôm tít | Tôm, mực | Vũng Tàu | 215 | 144 |
| 6 | Bánh hỏi heo quay | Bánh hỏi | Thịt heo quay |  | Cần Thơ | 140 | 81 |
| 7 | Cơm gà | Cơm | Thịt gà | Rau răm | Quảng Nam | 190 | 132 |
| 8 | Cao lầu | Mì, thịt heo | Bánh đa |  | Quảng Nam | 184 | 128 |
| 9 | Mì Quảng | Mì | Bánh đa, trứng | Thịt gà, thịt heo, cá, tôm | Quảng Nam, Đà Nẵng | 165 | 117 |
| 10 | Bún bò Huế | Bún | Thịt bò |  | Thừa Thiên Huế | 173 | 108 |
| 11 | Phở Hà Nội | Phở | Thịt bò |  | Hà Nội | 257 | 141 |
| 12 | Bún mực | Bún | Mực | Tôm | Phú Yên | 254 | 71 |
| 13 | Bún mọc | Bún | Viên mọc | Thịt heo | Hà Nội | 150 | 65 |
| 14 | Bún đậu mắm tôm | Bún | Chả cốm, dồi sụn, đậu hũ, thịt heo, chả giò |  | Hà Nội | 176 | 121 |
| 15 | Món khác (để lấy nguyên liệu) |  |  |  |  | 258 | 0 |
| Tổng cộng | | | | | | 2,873 | 1,498 |

Cụ thể, (1) *nguyên liệu ưu tiên* được hiểu là nguyên liệu có “*tính chất quyết định nhất*” đến tên món ăn. Chẳng hạn, các món được chế biến từ “Bún”, thì . Lý do chúng tôi chia ra và một cách cụ thể là vì giữa “Bún” và “Hủ tiếu” sẽ khác nhau hoàn toàn về cách cấu thành món ăn và đây là nguyên liệu “bắt buộc” phải có trong món ăn đó. Một vài trường hợp có thể cùng nhưng khác nhau thì sẽ có tên gọi khác nhau. Chẳng hạn, “Hủ tiếu” và “Bún nước lèo” sẽ có rất nhiều giống nhau, tuy nhiên đây là hai món ăn khác hoàn toàn về “chất liệu”. (2) Nguyên liệu chính là các thành phần nguyên liệu cũng có tính chất quyết định đến tên món ăn, nhưng thấp hơn . Ví dụ ở món bún cá sẽ có , ở món bún mực sẽ có thành phần chính là mực (). (3) Nguyên liệu phụ là các thành phần nguyên liệu thường được ăn kèm trong món ăn đó, cùng với . Ví dụ ở món bún cá và cơm gà, thực khách thường ăn kèm với rau răm. Các thành phần nguyên liệu phụ này cũng có tính chất quyết định đến món ăn, đóng vai trò nhận biết món ăn. Một cách hình thức gọi là trọng số ưu tiên và là mức độ ưu tiên. Khi đó, chúng ta sẽ có: .

Giả sử, cho là ảnh đầu vào để phân lớp, chúng ta áp dụng mô hình phát hiện đối tượng để tìm ra tập các thành phần nguyên liệu được phát hiện từ ảnh , được chỉ định là . Khi đó, chúng ta có được với . Sau khi thu thập được các nguyên liệu được tìm thấy trong ảnh, chúng ta sẽ tiến hành tính khoảng cách với ***giải thuật D2D*** được đề xuất.

Ý tưởng của giải thuật tính khoảng cách là dựa vào mức độ quan trọng của các thành phần nguyên liệu để quyết định được tên của món ăn cần nhận dạng. Ở đây, chúng ta sẽ sử dụng các thông tin từ Bảng 1 để tìm tên món ăn

**Giải thuật: Tính khoảng cách D2D**

**Đầu vào:** : danh sách các nguyên liệu mà phát hiện được, danh sách các món ăn, : danh sách các nguyên liệu

**Đầu ra**: Danh sách kết quả là tên các món ăn được xác định

2. :
3. | |
4. | | //
5. | |
6. | | //
7. | |
8. | |
9. | |
10. //
11. | //

Giải thuật được xem xét từ hai góc nhìn khác nhau: (1) và (2) . Giải thích cho giải thuật D2D cụ thể như sau: đầu vào sẽ bao gồm: (danh sách từ phát hiện đối tượng), (danh sách các món ăn), (danh sách các NL) và ba mức độ ưu tiên. Đầu tiên, tại dòng 3, giải thuật sẽ duyệt qua từng món ăn trong Bảng 1 để tìm món ăn. Ở đây, góc nhìn (1) được hiểu như là xét từng NL của món ăn với kết quả dò tìm được (dòng 4 đến dòng 6) trong khi góc nhìn (2) là kiểm tra kết quả dò tìm so với tất cả các NL của món ăn (dòng 7). Khoảng cách của cả hai trường hợp (1) + (2) sẽ được lưu trữ vào tại dòng 9. Tại dòng 8 sẽ là tính khoảng cách của góc nhìn (1). Chúng sẽ là tích của số lượng NL tìm được với trọng số ưu tiên ( , chẳng hạn . Một điểm lưu ý rằng mỗi NL ưu tiên và chính được đếm (tính khoảng cách) là 01 (dòng 4 và 5) nếu nó không thuộc , điều này được hiểu là không tìm thấy một vài NL ưu tiên và chính trong món ăn đó; hiểu một cách khác là kết quả dò tìm sẽ có khoảng cách xa nếu NL ưu tiên và chính không được tìm thấy. Mặt khác, NL phụ được hiểu là khoảng cách “kéo lại” so với NL ưu tiên và chính, điều này được hiểu là NL phụ chỉ để giúp phân biệt giữa các món ăn có NL ưu tiên và chính giống nhau. Lý do D2D thực hiện phép trừ cho NL phụ (tại dòng 8) là vì hầu hết các NL này sẽ tùy thuộc vào sự yêu thích của “người ăn” để cho vào món ăn. Một vài trường hợp họ cho vào rất nhiều NL phụ vào món ăn, khi đó, sẽ ảnh hưởng việc xem xét của NL chính và ưu tiên, ví dụ trường hợp mì quảng và cao lầu. Việc sử dụng khoảng cách “kéo lại” của nguyên liệu phụ nhằm mục đích sàn lọc những món ăn gần như giống nhau với khoảng cách gần hơn (do vậy thực hiện “trừ ra” thay vì “cộng vào”). Sau khi tính khoảng cách, một mảng đã được sử dụng để lưu trữ khoảng cách vừa tìm được và tên món ăn tương ứng (tại dòng 10). Dòng 11 sẽ tìm ra món ăn có khoảng cách nhỏ nhất () từ tập . Dựa trên , một danh sách các món ăn mà gần với các NL được dò tìm sẽ được tìm thấy. Danh sách này được hiểu như là kết quả phân lớp.

Các kết quả về việc thử nghiệm này sẽ được trình bày ở phần tiếp theo của bài báo.

# THỰC NGHIỆM, ĐÁNH GIÁ VÀ THẢO LUẬN

## Thu thập dữ liệu và cài đặt môi trường

Trong bài báo này, ViFLAVOR sẽ thực nghiệm bước đầu với 14 lớp tương ứng với 14 món ăn, bao gồm: bún cá, hủ tiếu Mỹ Tho, bún nước lèo, cơm tấm Long Xuyên, bún hải sản bề bề, bánh hỏi heo quay, cơm gà, cao lầu, mì Quảng, bún bò Huế, phở Hà Nội, bún mực, bún mọc, bún đậu mắm tôm *(như đã giới thiệu ở mục II.A và Bảng 1).* Về số lượng ảnh thu thập, nghiên cứu này đã thực hiện với 2,873 cho bài toán phát hiện đối tượng và 1,498 cho bài toán phân lớp. Ở đây, dữ liệu sẽ được chia theo tỷ lệ 7:3 (cụ thể, huấn luyện sử dụng 70% và kiểm tra với 30%). Lưu ý rằng, chúng tôi đã thực nghiệm với tập dữ liệu ít *(số lượng mẫu đủ để mô hình phân lớp không bị học vẹt)*, hơn nữa, chúng tôi cũng đã tiến hành thu thập một số ảnh mà món ăn đã bị xáo trộn thay vì món ăn “đẹp” ban đầu để kiểm chứng ViFLAVOR. Cụ thể, số lượng ảnh đã thu thập trình bày trong Bảng 1. Chúng tôi đã sử dụng Roboflow để gán nhãn ảnh *(với 27 nguyên liệu đã được gán nhãn từ 2873 ảnh, mỗi ảnh sẽ có nhiều nguyên liệu. Tổng cộng có 22,148 nhãn đã được gán (annotations))*. Trong tập ảnh để huấn luyện OD, chúng tôi đã chọn lọc tập ảnh và ảnh xáo trộn để thưc hiện phân lớp. Cụ thể, chúng tôi giữ lại 20-50% tập ảnh cho OD và thêm vào các ảnh bị xáo trộn từ 15 ảnh trở lên cho bài toán phân lớp.

Hệ thống được phát triển sử dụng ngôn ngữ lập trình Python để cài đặt giải thuật, huấn luyện mô hình phát hiện đối tượng. Giao diện Web của ứng dụng được xây dựng bằng Vue.js và sử dụng Flask framework để xử lý phía máy chủ. Việc kết hợp Vue.js và Flask sẽ tận dụng điểm mạnh của cả hai framework, mang đến một front-end mạnh mẽ và một back-end linh hoạt. Vue.js cung cấp giao diện người dùng linh hoạt và phản ứng, trong khi Flask cho phép xử lý phía máy chủ hiệu quả và có thể mở rộng. Chúng tôi lựa chọn Flask vì framework này đơn giản và linh hoạt cho việc cài đặt cho bài toán thực tế. Để huấn luyện mô hình phát hiện đối tượng nghiên cứu đã sử dụng thư viện ultralytics[[5]](#footnote-5) (8.2.11), Keras (2.15.0). Cấu hình máy tính để triển khai ViFLAVOR thì như sau: hệ điều hành: windows 10; CPU: Intel Core i7 1165G7; RAM: 16GB; Card đồ hoạ: NVIDIA GEFORCE GTX 1050 4GB GDDR5. Một điều quan trọng cần nhấn mạnh là tập dữ liệu (được thu thập cùng) với các cài đặt đã được công bố trên nền tảng GitHub[[6]](#footnote-6).

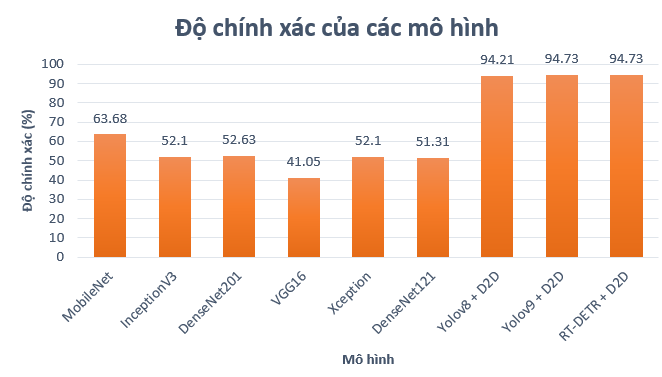
## So sánh và đánh giá mô hình

Từ hướng tiếp cận chính như đề cập bên trên, nhóm nghiên cứu đã tiến hành thực nghiệm và cho ra kết quả thống kê độ chính xác, dung lượng của các mô hình phát hiện thành phần nguyên liệu được thể hiện chi tiết ở Bảng 3.

1. Thống kê độ chính xác, dung lượng của ba mô hình phát hiện thành phần nguyên liệu

|  |  |  |  |
| --- | --- | --- | --- |
| **Mô hình** | **mAP50** | **mAP50 – 95** | **Kích thước (MB)** |
| Yolov8 | 0.862 | 0.71 | 130 |
| RT-DETR | 0.865 | 0.681 | 63.1 |
| Yolov9 | 0.85 | 0.677 | 49.2 |

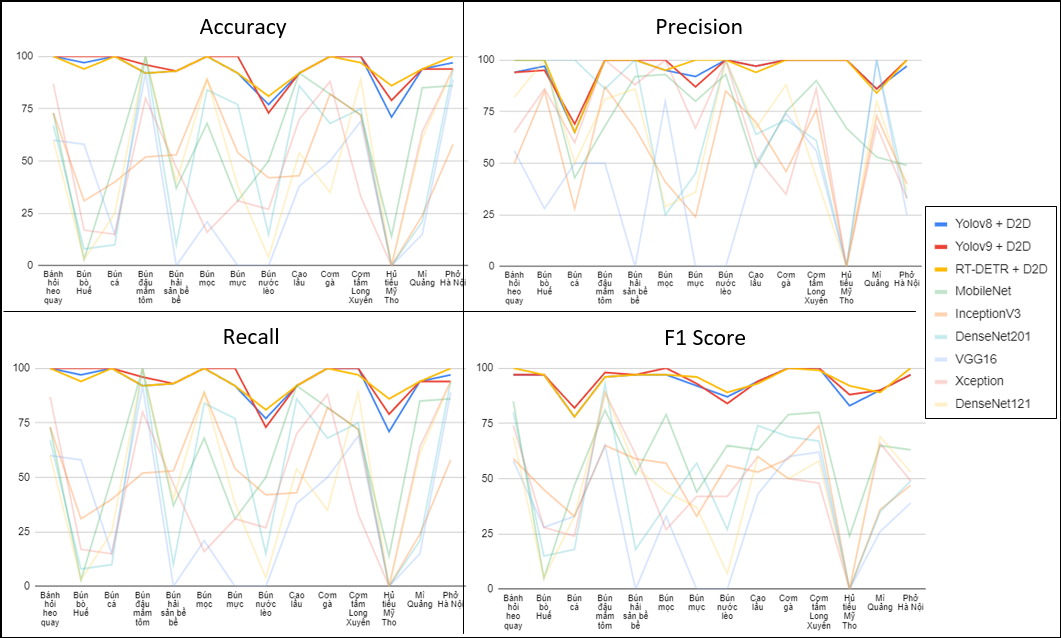
Bên cạnh đó, chúng tôi cũng thực nghiệm trên sáu mô hình học sâu bao gồm LeNet, MobileNet, InceptionV3, VGG16, ResNet101, DenseNet201. Chúng tôi đã sử dụng một tham số là batch size: 64, kích thước ảnh là: 64x64, sử dụng early-stopping với patience: 5. Kết quả độ chính xác đạt được trên các mô hình theo hai hướng tiếp cận đối với tập kiểm tra được thể hiện trong Hình 6.



1. Độ chính xác tổng thể của các mô hình

Nhận thấy, kết quả phân loại trên các mô hình học sâu có độ chính xác không cao (đạt dưới 80%). So sánh giữa các mô hình được trực quan hóa trong Hình 6. Kết quả cho thấy, đề xuất của chúng tôi đạt kết quả cao hơn so với hướng tiếp cận CNN. Lưu ý rằng, với ba tham số thì có chiến lược chọn với các yêu cầu sau: (1) cần phải lớn hơn nhiều so với hai tham số còn lại để chỉ ra rằng món ăn đó không được nhầm lẫn, ví dụ bún, hủ tiếu thì cần sự rõ ràng. Trong một số trường hợp, sẽ tác động đến kết quả (vì cùng các nguyên liệu nhưng khác nhau về nguyên liệu bắt buộc). (2) và thì không cần cách nhau nhiều vì đây là các nguyên liệu cấu thành nên món ăn. Tuy nhiên, cần lớn hơn ít nhất 5 để đảm bảo sự cách biệt cần thiết. (3) . Chi tiết, là hệ số của “kéo lại”, do vậy hệ số này được hiểu là các thành phần đó đôi khi không xuất hiện, tuỳ theo sở thích của từng cá nhân. Trong thực nghiệm này, chúng tôi đã thực nghiệm với nhiều tham số khác nhau. Kết quả đạt độ chính xác cao là: .

Cuối cùng, chúng tôi đã thống kê chi tiết hiệu suất trên từng lớp món ăn của các mô hình. Kết quả thống kê được thể hiện trên biểu đồ như Hình 7. Nhìn chung, kết quả đạt được từ mô hình đề xuất thì hoàn toàn đầy mong đợi. Tất cả độ chính xác của mô hình đều có kết quả cao hơn so với hướng tiếp cận CNN ở từng nhãn món ăn.

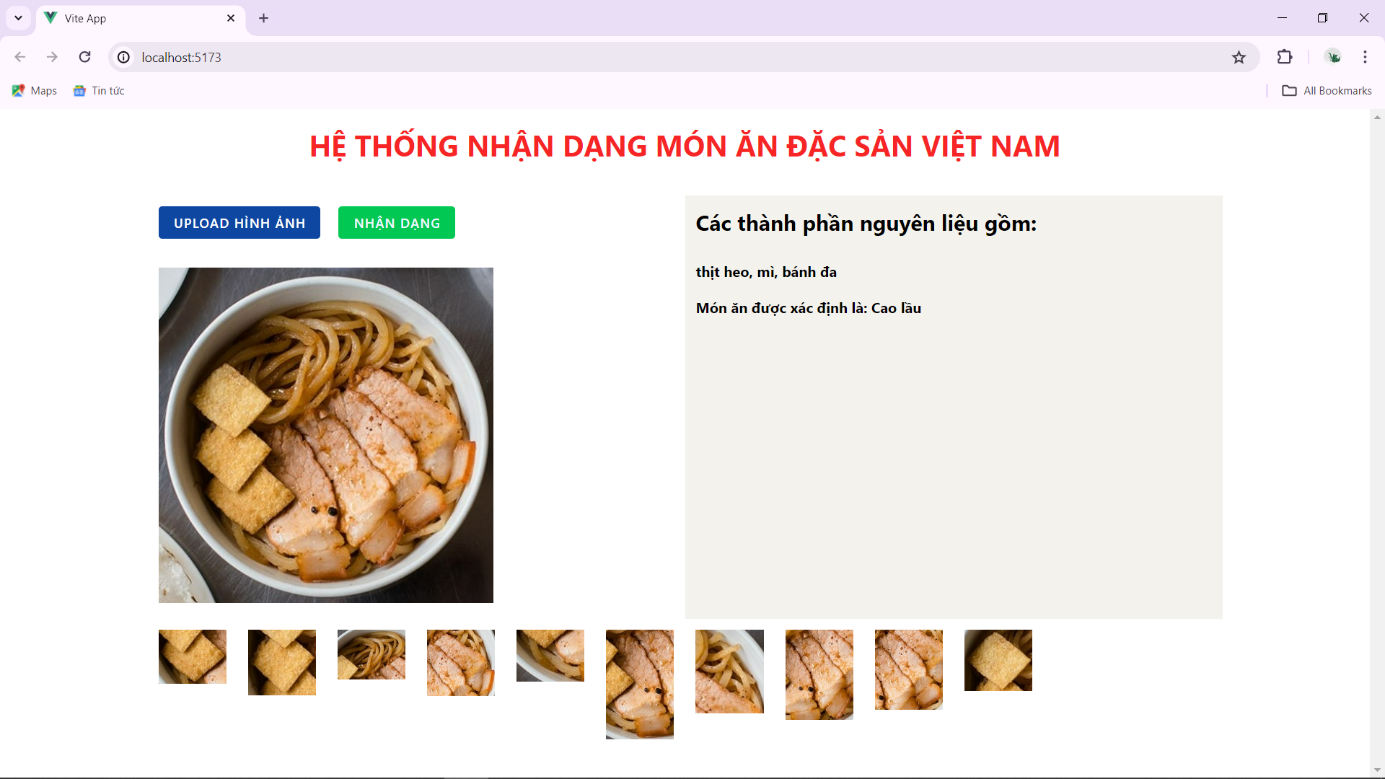
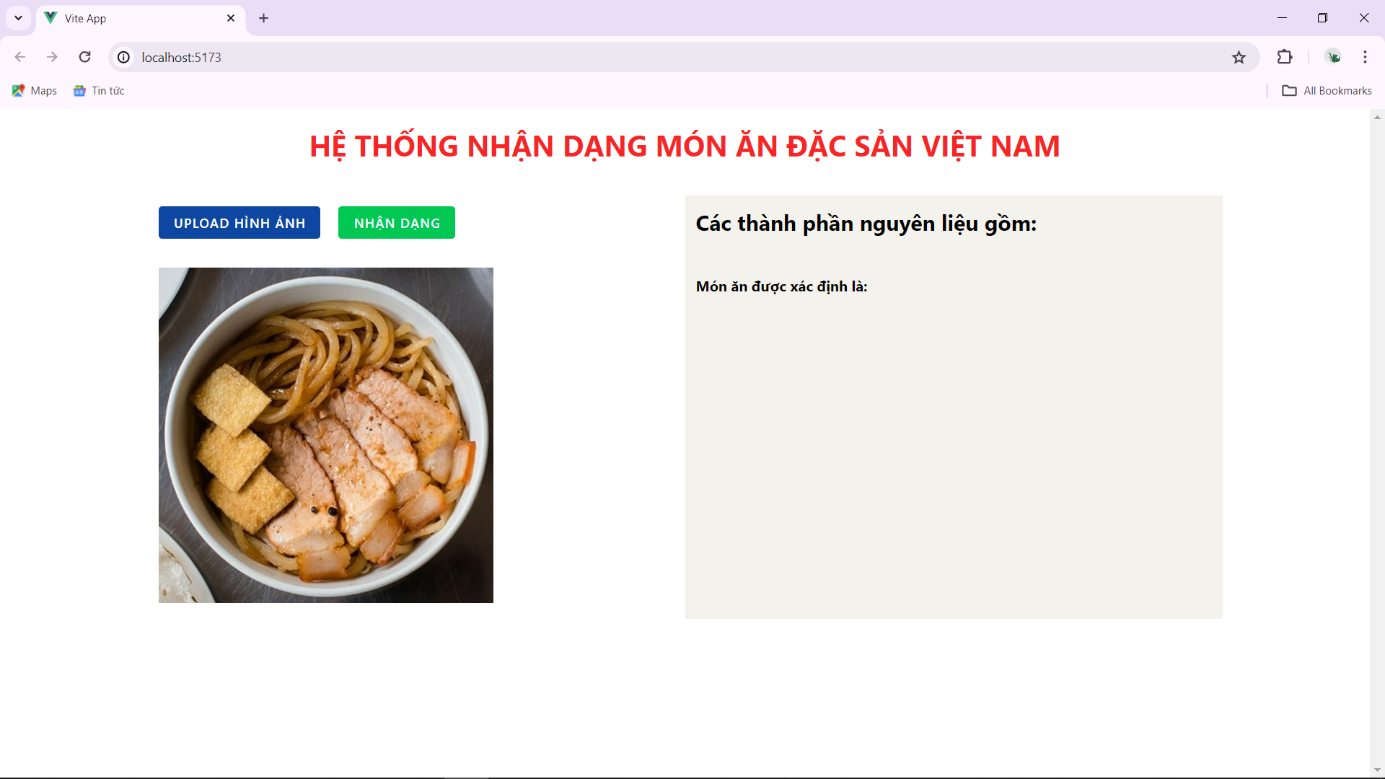


1. Độ chính xác của các mô hình trên các nhãn

Với kết quả này, chúng tôi kết luận được phương pháp chính của nghiên cứu này: **xác định món ăn đặc sản dựa trên các thành phần nguyên liệu, mang lại kết quả vượt trội hơn phương pháp phân lớp CNN** – vốn đang được sử dụng rộng rãi.

## Triển khai ViFLAVOR và thảo luận

Sau khi quan sát kết quả huấn luyện từ mô hình, chúng tôi quyết định chọn mô hình phát hiện đối tượng tốt nhất theo hướng tiếp cận thứ hai để xây dựng ứng dụng nhận dạng món ăn đặc sản Việt Nam. Cụ thể, chúng tôi đã lựa chọn RT-DETR +D2D để triển khai hệ thống, đảm bảo cân bằng giữa độ chính xác và dung lượng của mô hình. Chúng tôi đã thiết kế một giao diện trực quan với các thao tác đơn giản. Các chức năng chính bao gồm: (1) Người dùng chọn một tập tin hình ảnh cần nhận dạng từ giao diện, (2) Người dùng chọn “Nhận dạng” để hệ thống phát hiện các thành phần nguyên liệu trên món ăn và xác định tên món ăn. Giao diện của ViFLAVOR được minh họa trong Hình 8.



1. Giao diện kết quả từ ứng dụng nhận dạng món ăn đặc sản Việt Nam

Nhìn chung kết quả dự đoán của bài báo này đã đạt được như mong đợi. Ở đây, mặc dù chúng tôi đã lựa chọn phong phú và đa dạng cho các món ăn, tuy nhiên với sự đa dạng của ẩm thực Việt thì vẫn còn chưa nhiều. Ngoài đóng góp chính của nhóm nghiên cứu để xây dựng được ứng dụng ViFLAVOR, đề tài đã xoay một hướng nhìn mới về giải thuật KNN cho bài toán phân lớp. Một cách tổng quan, *mô hình ViFLAVOR* có thể được xem như một hướng mở rộng của KNN, hướng tiếp cận mới này dựa trên đặc trưng trên hình ảnh, khoảng cách được xác định theo hướng phân tầng và không có quá trình học. Hơn nữa, hướng tiếp cận này cũng giải quyết tốt cho bài toán học ít dữ liệu như đã đề cập trong phần giới thiệu. Từ kết quả Hình 7 cho thấy, nó làm việc tốt với dữ liệu ít và cho kết quả vượt trội và đồng đều cho từng lớp. Điểm hạn chế lớn nhất là phụ thuộc vào OD rất nhiều. Tuy rằng, giải thuật D2D là nhìn từ cả 2 phía, tuy nhiên kết quả từ OD sẽ là sự quan tâm lớn nhất cho hướng tiếp cận này. Để giải quyết nó, chúng tôi sẽ tìm đến việc trộn nhiều OD để có kết quả tốt hơn.

# KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Nghiên cứu này đã đóng góp một mô hình phân lớp các món ăn đặc sản Việt Nam. Hơn nữa, tập dữ liệu thực tế đã được thu thập cho bài toán này. Bằng cách áp dụng các mô hình phát hiện đối tượng cho các thành phần nguyên liệu của món ăn và kết hợp với giải thuật tính khoảng cách mới đề xuất, chúng tôi đã đạt được kết quả ấn tượng với độ chính xác trên 94%. Nhóm nghiên cứu cũng đã xây dựng một trang web thử nghiệm, cho phép người dùng dễ dàng nhận diện các món ăn đặc sản Việt Nam khác nhau.

Đối với định hướng nghiên cứu kế tiếp, chúng tôi sẽ mở rộng thêm các món ăn đặc trưng và phát triển thêm các ứng dụng di động thông minh. Hơn nữa, chúng tôi đang lên kế hoạch cải tiến giải thuật bằng cách áp dụng lý thuyết trộn niềm tin (belief merging) để nâng cao độ chính xác và hiệu quả của hệ thống.

# TÀI LIỆU THAM KHẢO

1. Chien-Yao Wang, I-Hau Yeh, Hong-Yuan Mark Liao. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information, arXiv:2402.13616v2.
2. Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna. Rethinking the inception architecture for computer vision, Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.
3. Dillon Reis, Jordan Kupec, Jacqueline Hong, Ahmad Daoudi. Real-Time Flying Object Detection with YOLOv8, arXiv:2305.09972.
4. El-Sawy, Ahmed, Hazem El-Bakry, and Mohamed Loey. CNN for handwritten arabic digits recognition based on LeNet-5, Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016 2. Springer International Publishing, 2017.
5. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.
6. Karen Simonyan, Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, 2014.
7. Konstantakopoulos, Fotios S., Eleni I. Georga, and Dimitrios I. Fotiadis. A review of image-based food recognition and volume estimation artificial intelligence systems, IEEE Reviews in Biomedical Engineering, 2023.
8. Mã Trường Thành, Châu Ngân Khánh, Thạch Minh Hớn, Phạm Xuân Hiền, Phan Bích Chung. Xác định món ăn đặc sản Việt Nam với sự kết hợp của mạng học sâu và bản thể học, Tạp chí Khoa học Đại học Cần Thơ, Vol. 59, pp. 93-101.
9. Phiphiphatphaisit, Sirawan, and Olarik Surinta. Food image classification with improved MobileNet architecture and data augmentation, Proceedings of the 3rd International Conference on Information Science and Systems, 2020.
10. Shifat, Sirajum Munira, Takitazwar Parthib, Sabikunnahar Talukder Pyaasa, Nila Maitra Chaity, Niloy Kumar, and Md Kishor Morol. A Real-time Junk Food Recognition System based on Machine Learning, International Conference on Bangabandhu and Digital Bangladesh.
11. Yian Zhao, Wenyu Lv, Shangliang Xu, Jinman Wei, Guanzhong Wang, Qingqing Dang, Yi Liu, Jie Chen. DETRs beat yolos on real-time object detection, arXiv preprint arXiv:2304.08069, 2023.
12. Hussain, Mahbub, Jordan J. Bird, and Diego R. Faria. “A study on CNN transfer learning for image  
    classification.” Advances in Computational Intelligence Systems: Contributions Presented at the 18th UK Workshop  
    on Computational Intelligence, September 5-7, 2018, Nottingham, UK. Springer International Publishing, 2019.
13. Li, Qing, et al. "Medical image classification with convolutional neural network." 2014 13th international conference  
    on control automation robotics & vision (ICARCV). IEEE, 2014.
14. Amit, Yali & Felzenszwalb, Pedro & Girshick, Ross. (2020). Object Detection. 10.1007/978-3-030-03243-2\_660-1.
15. Nguyễn Hương (2012). Ẩm thực Việt có giúp xây dựng thương hiệu du lịch Việt Nam. Bộ Văn hóa, Thể thao và Du lịch. Ngày đăng tải: 11/12/2012.
16. Phương Mai (2013). Văn hóa ẩm thực Việt Nam đối với ngành du lịch. Tin học kinh tế Sài Gòn. Ngày đăng tải: 10/01/2013.
17. Do, TN., Pham, TP., Nguyen, HH., Pham, NK. (2021). Visual Classification of Intangible Cultural Heritage Images  
    in the Mekong Delta. In: Belhi, A., Bouras, A., Al-Ali, A.K., Sadka, A.H. (eds) Data Analytics for Cultural Heritage.  
    Springer, Cham. <https://doi.org/10.1007/978-3-030-66777-1_4>
18. Cosovic, Marijana & Jankovic Babic, Radmila. (2020). CNN Classification of the Cultural Heritage Images.  
    10.1109/INFOTEH48170.2020.9066300.

ViFLAVOR: VIETNAMESE SPECIALTY CUISINE IDENTIFICATION VIA INGREDIENT PRIORITY ORDER

Truong-Thanh Ma, Gia-Khuong Huynh, Hieu-Nghia Do, Thanh-Nghi Do

**ABSTRACT** – Vietnamese cuisine is an integral component of Vietnamese culture and lifestyle. Vietnamese cuisine, with its richness and diversity of ingredients and distinct flavors, has captivated and inspired chefs and culinary scholars all over the world. However, for foreigners or those unfamiliar with local cuisine, identifying Vietnamese delicacies can be challenging because their qualities and preparation techniques frequently differ. Each location has its own unique qualities; for example, the components in crab vermicelli varies from one another. As a result, an artificial intelligence (AI) system capable of detecting Vietnamese delicacies by identifying components is entirely predictable. The approach will facilitate studying and tasting local food, particularly for overseas tourists. This research will focus on recognizing and identifying substances in food photographs. Based on the identified ingredients, a D2D algorithm was developed to identify the dish's name. The suggested model's food categorization results met the required efficiency of more than 94%. Finally, the technique has been deployed on the Web platform and the implementation source code has been released.

1. Ẩm thực Việt Nam bao hàm ý nghĩa khái quát nhất để chỉ tất cả các món ăn phổ biến của cộng đồng người Việt. [↑](#footnote-ref-1)
2. <https://docs.ultralytics.com/models/yolov8/> [↑](#footnote-ref-2)
3. <https://docs.ultralytics.com/models/yolov9/> [↑](#footnote-ref-3)
4. <https://docs.ultralytics.com/models/rtdetr> [↑](#footnote-ref-4)
5. <https://www.ultralytics.com> [↑](#footnote-ref-5)
6. <https://github.com/dohieunghia2002/Identify_Vietnamese_Dishes> [↑](#footnote-ref-6)