

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



LƯU TIẾN THÀNH

NHẬN DẠNG THỰC PHẨM TƯƠI SỐNG BẰNG HÌNH ẢNH

CHUYÊN NGÀNH : KHOA HỌC MÁY TÍNH

MÃ SỐ: 60.48.01.01

TÓM TẮT LUẬN VĂN THẠC SĨ

(Theo định hướng ứng dụng)

NGƯỜI HƯỚNG DẪN KHOA HỌC: TS. PHẠM VĂN CƯỜNG

HÀ NỘI - 2016

Luận văn được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: TS. Phạm Văn Cường

Phản biện 1: TS. Nguyễn Duy Phương.....

Phản biện 2: PGS.TS. Nguyễn Ngọc Hóa.....

Luận văn sẽ được bảo vệ trước Hội đồng chấm luận văn thạc sĩ tại
Học viện Công nghệ Bưu chính Viễn thông

Vào lúc: ... giờ ngày ..20... tháng ...8... năm ..2016..

Có thể tìm hiểu luận văn tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông

MỞ ĐẦU

Tính cấp thiết của đề tài

Thực phẩm là vấn đề tất yếu của cuộc sống, chúng cung cấp chất dinh dưỡng cho sự phát triển của cơ thể. Thực phẩm với chế độ dinh dưỡng hợp lý sẽ đảm bảo sức khỏe con người. Ngược lại, chế độ dinh dưỡng bất hợp lý sẽ dẫn tới các bệnh nghiêm trọng như tim mạch, tiểu đường, béo phì... Đặc biệt trong những năm gần đây, bệnh béo phì đang gia tăng nhanh chóng ở mọi lứa tuổi, mọi tầng lớp và mọi quốc gia, tạo nên gánh nặng to lớn đối với sự tiến bộ của toàn nhân loại. Tại Việt Nam, theo Cục Y tế dự phòng, Bộ Y tế cho biết, tỷ lệ người trưởng thành bị thừa cân béo phì chiếm khoảng 25% dân số [1]. Trong xã hội hiện đại, tình trạng thừa cân, béo phì ở người trưởng thành đang có xu hướng ngày càng phổ biến và trở thành một trong những thách thức lớn đối với chương trình chăm sóc sức khỏe ở mọi quốc gia. Nguyên nhân căn bản của thừa cân, béo phì là do tình trạng mất cân bằng về năng lượng giữa lượng calo đưa vào cơ thể và lượng calo được sử dụng. Điều này cho thấy cần có một chế độ dinh dưỡng khoa học và phù hợp cho từng cá nhân để đảm bảo trọng lượng cơ thể ở mức cần thiết. Một hệ thống nhận dạng thực phẩm hàng ngày với chi phí thấp, độ chính xác cao là cần thiết. Dựa trên hệ thống này, chúng tôi sẽ thực hiện xây dựng ứng dụng quản lý chế độ dinh dưỡng trên nền tảng di động nhằm đảm bảo tính cá nhân, tự động và dễ sử dụng.

Mục đích nghiên cứu

Mục đích của nghiên cứu chính của luận văn nhằm xây dựng một hệ thống nhận dạng thực phẩm, dựa trên hệ thống này, chúng tôi xây dựng ứng dụng quản lý chế độ dinh dưỡng nhằm giúp cho người dùng quản lý và theo dõi chế độ dinh dưỡng hàng một cách hiệu quả nhất. Dữ liệu thu được từ ứng dụng cũng sẽ giúp các chuyên gia y tế tốt hơn trong việc theo dõi và đưa ra phác đồ điều trị hợp lý. Việc nhận dạng các loại thực phẩm dựa trên việc phân loại hình ảnh bằng các thuật toán học máy là rừng ngẫu nhiên (Random Forest) và máy hỗ trợ vector(SVM). Do đó, chúng tôi cần nghiên cứu và nắm vững các kiến thức về xử ảnh và học máy.

Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu chính của chúng tôi là các thuật toán xử lý ảnh, nhằm cung cấp tập dữ liệu tốt nhất cho việc nhận dạng thực phẩm. Cùng với đó chúng tôi cần nghiên cứu và hiểu rõ các thuật toán học máy là Random Forest và SVM nhằm đưa ra đánh giá khách

quan về tính hiệu quả của hai thuật toán trong việc nhận dạng thực phẩm từ các hình ảnh đã xử lý, từ đó đưa ra quyết định thuật toán nào sẽ được sử dụng cho hệ thống và ứng dụng.

Trên thực tế, thực phẩm hàng ngày rất đa dạng và phong phú, rất khó cho việc nhận dạng thậm chí bằng mắt thường vì đặc điểm bên ngoài của chúng có thể tương nhau như thịt gà và thịt vịt, thịt bò và thịt lợn. Vì vậy trong khuôn khổ luận văn này, chúng tôi chỉ tập trung vào phạm vi nhận dạng các loại thực phẩm tươi rau, củ, quả ở giai đoạn tiền chế biến.

Phương pháp nghiên cứu

Luận văn sử dụng thuật toán phân cụm K-mean để phân đoạn hình ảnh, sau đó thực hiện trích chọn các đặc trưng như màu sắc, SURF, kết cấu, hình dạng; rút trích các vector đặc trưng (feature extractor) nhằm cung cấp đầu vào cho thuật toán SVM và thuật toán Random Forest phục vụ cho việc phân loại hình ảnh. Cuối cùng, chúng tôi thực hiện so sánh, đánh giá hai thuật toán về mặt hiệu năng, chi phí, độ chính xác để chọn ra thuật toán tốt nhất và dựa trên đó để xây dựng hệ thống nhận dạng và ứng dụng.

CHƯƠNG 1: TỔNG QUAN VỀ BÀI TOÁN NHẬN DẠNG THỰC PHẨM

Trong chương này chúng tôi nói về bài toán nhận dạng thực phẩm và sự cần thiết phải xây dựng bài toán, một số các nghiên cứu trước đây và sự hạn chế của những bài toán nhận dạng thực phẩm đó nhằm tìm ra phương pháp tiếp cận phù hợp, chúng tôi cũng đưa ra các khía cạnh, giả định và giới hạn phạm vi nghiên cứu cho bài toán.

1.1 Bài toán nhận dạng thực phẩm

Chế độ dinh dưỡng có tác động đáng kể tới sức khỏe chúng ta. Các nghiên cứu trước đây [16] chỉ ra rằng nhiều bệnh như cao huyết áp, tiểu đường đau dạ dày, béo phì ... liên quan trực tiếp tới chế độ dinh dưỡng ko cân bằng. Theo báo cáo của Hiệp hội Tim mạch Hoa Kỳ [3], có tới hơn 154 triệu người tại Hoa Kỳ bị béo phì hoặc thừa cân (chỉ số cơ thể BMI vượt quá 25kg/m²). Chi phí ước tính cho việc quản lý và điều trị vào khoảng 254 tỷ USD trong năm 2013 và tăng lên tới 927 tỷ USD trong năm 2030, khoảng 18% các ca tử vong có liên quan tới béo phì [24] gây ảnh hưởng xấu tới nền kinh tế và y học. Không có ngoại lệ ở Việt Nam, theo một báo cáo thống kê [36] của tổ chức y tế thế giới WHO năm 2015, tỷ lệ

người trưởng thành bị béo phì chiếm khoảng 3.55% (2.3% ở nam giới và 4.8% ở nữ giới) với tổng chi phí điều trị hơn 600 triệu USD mỗi năm được Chính phủ chi ra.

Nguyên nhân căn bản của thừa cân, béo phì là do tình trạng mất cân bằng về năng lượng giữa lượng calo đưa vào cơ thể và lượng calo được sử dụng. Chìa khóa của vấn đề là đưa ra được chế độ dinh dưỡng phù hợp cho từng cá nhân và việc nắm rõ các thành phần thực phẩm được chuẩn bị cho bữa ăn là một trong những giải pháp quan trọng cho vấn đề này.

Một hệ thống quản lý chế độ dinh dưỡng với chi phí thấp, độ chính xác cao là cần thiết. Dựa trên hệ thống này, chúng tôi tiến hành xây dựng ứng dụng quản lý dinh dưỡng dựa trên nền tảng di động nhằm đảm bảo tính cá nhân, tự động và dễ sử dụng. Dữ liệu được hệ thống ghi lại cũng sẽ giúp các chuyên gia y tế đưa ra phác đồ điều trị chính xác hơn. Lợi ích của hệ thống không chỉ giúp các bệnh nhân béo phì mà còn hỗ trợ quản lý tất cả các vấn đề liên quan tới kiểm soát dinh dưỡng, trọng lượng.

1.2 Các nghiên cứu trước đây

Trước đây đã có một số nghiên cứu nhằm xây dựng các hệ thống quản lý dinh dưỡng. Đại học Carnegie Mellon đã đưa ra hệ thống quản lý chế độ dinh dưỡng của người bệnh tiểu đường dựa trên thực đơn nhằm ghi lại thực phẩm tiêu thụ hàng ngày gọi tắt là DIMA (Diabetes Management Assistant) [5][17]. Hệ thống cho phép chụp ảnh các bữa ăn và lưu vào cuốn nhật ký điện tử (ghi log) thay cho việc ghi chép thủ công.

Wellness Diary là một dự án nghiên cứu của Nokia Beta Labs [26][48] nhằm xây dựng một trong những ứng dụng di động đầu tiên cho phép theo dõi sức khỏe hàng ngày của người sử dụng được viết trên hệ điều hành Symbian Series 60 cho các thiết bị di động. Với ứng dụng, người dùng có thể giám sát một loạt các thông số như giấc ngủ, trọng lượng, chế độ tập thể dục và đặc biệt là chế độ ăn uống, tất cả đều được thực hiện qua việc ghi chép thủ công.

Hiện tại, MyFitnessPal [23] là một ứng dụng miễn phí cho phép theo dõi chế độ ăn uống, luyện tập để xác định lượng calor tối ưu và chất dinh dưỡng phù hợp đối với từng người sử dụng. Một điểm đáng khen của ứng dụng là việc tự động nhận biết được khối lượng calo của thực phẩm dựa trên tính năng quét mã BarCode.

Tuy nhiên, điểm hạn chế đáng tiếc của các hệ thống trên là việc bắt buộc người dùng phải tự ghi chép thủ công các bữa ăn và thực phẩm tiêu thụ hàng ngày gây bất tiện, tốn kém thời gian.

Việc nhận dạng thực phẩm tự động dựa trên hình ảnh là khá khó khăn, những thách thức hay gặp phải là việc thực phẩm bị biến dạng, chồng chéo lên nhau và các nhiễu bất lợi từ môi trường khi thu thập. Những nỗ lực nghiên cứu trước đây chỉ giới hạn trong một số loại thực phẩm cụ thể như cá, thịt, cam, quýt [20]. Các nỗ lực khác bị giới hạn bởi sự sắp xếp thực phẩm trên đĩa hoặc các khay ngăn chia cách chúng [33][12].

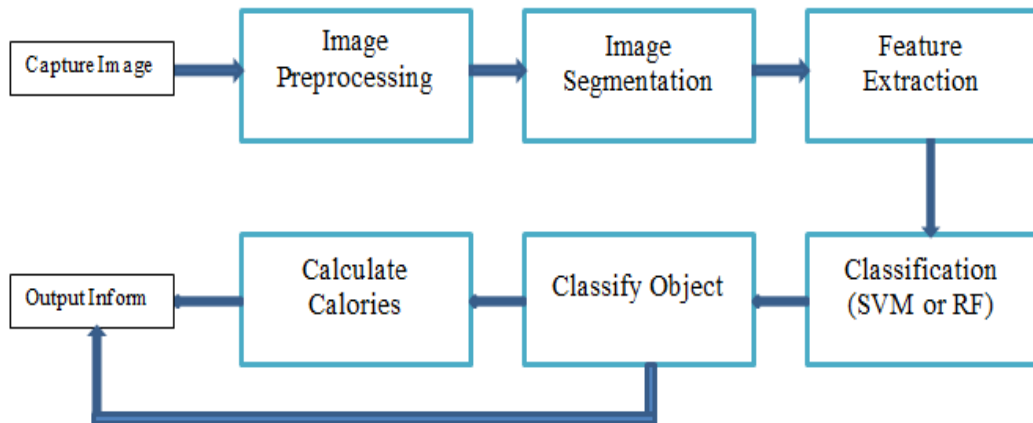
Một trong những thuật toán nhận dạng thực phẩm dựa trên việc nhận dạng hộp đựng thức ăn đã được phát triển nhằm mục đích điều chỉnh cài đặt lò vi sóng theo thức ăn trong hộp với các quy tắc được xác định trước về vị trí của các loại thực phẩm trong hộp, các loại thực phẩm được gán với các vị trí cụ thể trong hộp, gọi là bản đồ sắp xếp thực phẩm [33]. Tuy nhiên rất ít nỗ lực hướng tới việc sử dụng nhận dạng thực phẩm trong các lĩnh vực lớn như y tế và công nghệ hỗ trợ. Một nỗ lực được phát triển cho môi trường bệnh viện dựa trên phương pháp rút trích các đĩa thức ăn sử dụng Mạng Nơron dựa trên thuật toán nhận dạng ảnh cho phép đo lường thức ăn [12]. Tuy nhiên, ngay cả trong trường hợp này, thuật toán vẫn bị giới hạn bởi miền các loại thực phẩm phục vụ trong bệnh viện và vị trí của chúng trong khay.

Phần lớn các nghiên cứu trước đây chỉ tập trung vào nhận dạng thực phẩm nấu chín [33][19][18] hoặc thực phẩm ăn liền [19][35]. Trong khi đó việc nhận dạng thực phẩm tươi sống ít được quan tâm hơn. Vì vậy chúng tôi cố gắng tập trung vào việc nhận dạng thực phẩm ở giai đoạn tiền chế biến, chúng tôi tin rằng việc này sẽ giúp người dùng chủ động hơn trong việc lựa chọn các loại thực phẩm và chế biến chúng. Ngoài ra, các nghiên cứu và phương pháp trên cũng không đáp ứng được nhu cầu sử dụng hàng ngày vì chúng không dựa trên kiến trúc và nền tảng di động. Chúng tôi cố gắng để thu hẹp khoảng cách giữa sức khỏe cá nhân, công nghệ và người sử dụng bằng cách đưa việc giám sát dinh dưỡng lên thiết bị di động.

1.3 Phương pháp tiếp cận

Chúng tôi đề xuất một hệ thống quản lý dinh dưỡng để giải quyết các vấn đề trên. Hình ảnh thực phẩm được người dùng chụp bằng điện thoại di động với độ phân giải thấp.

Từ những hình ảnh này, hệ thống sẽ thực hiện phân loại thực phẩm qua các bước tiền xử lý ảnh, phân đoạn, trích chọn đặc trưng và phân loại.



Hình 1.4: Giải pháp quản lý dinh dưỡng.

1.4 Các vấn đề của bài toán nhận dạng ảnh

Việc nhận dạng thực phẩm dựa trên hình ảnh rất phức tạp và có nhiều khía cạnh cần được xem xét hoặc bỏ qua. Một số khía cạnh cần được xem xét như sau.

1.4.1 Các khía cạnh chung của bài toán nhận dạng ảnh

- *Sự co giãn hình ảnh:* Hình ảnh thực phẩm có thể được chụp từ các khoảng cách khác nhau tạo ra cảm nhận thiếu chính xác về kích thước thực sự của các thực phẩm.
- *Sự xoay ảnh:* Hình ảnh được chụp tại cùng khoảng cách với các góc xoay khác nhau cần được xem xét, nó tạo cảm giác sai lệch về hình dạng của thực phẩm.
- *Hướng chụp:* Khi thực phẩm được chụp ở các góc khác nhau, chúng có thể trông hoàn toàn khác nhau với các cách chụp từ trên xuống, trái qua phải Các tính chất như dây màu sắc, kết cấu, hình dạng và các dữ liệu khác có thể thay đổi đáng kể.
- *Chất lượng camera:* Không phải tất cả các máy di động đều cho cùng một chất lượng ảnh như nhau. Thậm chí cùng một thiết bị, tùy thuộc vào người sử dụng cũng sẽ có các thiết lập khác nhau.
- *Độ sáng khác nhau:* Đây là yếu tố ảnh hưởng lớn tới chất lượng ảnh. Không phải tất cả các ảnh đều được chụp trong điều kiện ánh sáng tương tự. Một sự thay đổi ánh sáng nhỏ có thể thay đổi hình ảnh đáng kể, màu tím chụp dưới điều kiện ánh sáng mạnh khó có thể phân biệt với màu đen. Đôi khi có thể xuất hiện bóng gây ra sự biến

dạng các đặc trưng, gây ảnh hưởng nghiêm trọng tới quá trình phân loại.

1.4.2 Các khía cạnh riêng của bài toán nhận dạng thực phẩm

- *Vị trí thực phẩm:* Thứ tự hình ảnh đứng trước, đứng sau, hay chỉ thấy được một góc của thực phẩm. Việc này làm tăng tính phức tạp trong việc xác nhận hình ảnh thực của mỗi loại thực phẩm có trong hình.
- *Sự đa dạng về hình dạng:* Không phải tất cả các loại thực phẩm sẽ có hình dạng giống nhau cho mỗi lần chụp. Mỗi bức ảnh tuy chụp cùng một khung cảnh nhưng sẽ cho các hình dạng khác nhau với mỗi loại thực phẩm. Thực tế cùng một loại thực phẩm cũng có nhiều hình dạng khác nhau.
- *Sự thay đổi trong kết cấu hình ảnh:* Điều kiện ánh sáng và điều kiện bảo quản có thể thay đổi kết cấu của thực phẩm.
- *Sự chồng chéo:* các loại thực phẩm bị sắp xếp chồng chéo đè lên nhau, một số bị che lấp gây khó khăn và sai lệch cho việc nhận dạng.
- *Bối cảnh lộn xộn:* thực phẩm có thể được chụp trong các bối cảnh khác nhau, trong siêu thị, trong bếp gây ra ảnh nền lộn xộn khiến việc phát hiện đối tượng, loại bỏ nhiễu nền trở nên khó khăn.

1.5 Các giả định của bài toán

Dựa trên các vấn đề đã nêu ở trên, chúng tôi chỉ xem xét những hình ảnh cho bài toán nhận dạng thực phẩm với giả định như sau:

- *Không chạm vào đối tượng:* để hạn chế khả năng đối tượng bị chồng chéo, che lấp, dễ dàng hơn cho việc phân cụm.
- *Chụp được tổng thể đối tượng:* dễ dàng hơn trong nhận biết đối tượng dựa trên hình dạng, màu sắc, kích thước thực của chúng, tính toán được lượng calorie chuẩn xác.
- *Màu nền chung:* màu nền là đơn màu, thuận tiện hơn cho việc phân đoạn ảnh, chỉ cần tập trung vào những đối tượng được phát hiện.
- *Nền nhẹ hơn đối tượng:* nhằm dễ dàng áp dụng kỹ thuật loại bỏ nhiễu nền.
- *Hạn chế bóng tối và ánh sáng đèn flash:* Để tránh trường hợp các đặc trưng của thực phẩm bị tổn hại do xuất hiện bóng của đối tượng hoặc các điểm đèn flash.

1.6 Phạm vi nghiên cứu

Trong thực tế có rất nhiều loại thực phẩm với các đặc điểm về hình dáng, màu sắc, kích thước giống nhau như trứng gà và trứng vịt, các loại thịt. Để có một hệ thống nhận dạng chính xác và trực tiếp từ nhiều nguồn như vậy cần rất nhiều các trích chọn đặc trưng cho từng loại thực phẩm và tập dữ liệu rất lớn. Việc này đòi hỏi nhiều thời gian trong quá trình thu thập dữ liệu và nghiên cứu.

Trong luận văn này, chúng tôi chỉ dừng lại ở việc nhận dạng các loại rau, củ, quả trong giai đoạn tiền chế biến với số lượng từ 8 tới 12 loại. Tập dữ liệu dùng để nhận dạng là các hình ảnh được thu thập bằng điện thoại di động ở độ phân giải thấp với số lượng từ 2000 tới 3000 ảnh.

1.7 Kết luận chương 1

Trong chương này, chúng tôi trình bày tổng quan về bài toán nhận dạng thực phẩm, thực trạng của bệnh thừa cân, béo phì và các tác động xấu của nó tới Việt Nam và thế giới, từ đó chúng tôi đưa ra giải pháp xây dựng hệ thống quản lý dinh dưỡng trên nền tảng di động nhằm giúp đỡ các bệnh nhân béo phì trong việc kiểm soát dinh dưỡng, trọng lượng và đưa ra phác đồ điều trị thích hợp.

Chúng tôi cũng nêu ra các hạn chế trong các hệ thống quản lý dinh dưỡng trước đây, hầu hết các hệ thống thực hiện việc kiểm soát dinh dưỡng một cách thủ công thông qua việc chép tay các loại thực phẩm, bữa ăn hoặc tiến bộ hơn là thông qua việc quét mã barcode. Công việc nhận dạng thực phẩm tự động là khá khó khăn khi gặp phải các thách thức như thực phẩm bị biến dị, chồng chéo lên nhau, khay đựng, nhiễu từ môi trường thu thập Các nghiên cứu trước đây dường như chỉ tập trung vào thực phẩm nấu chín với kết quả hạn chế. Vì vậy chúng tôi cố gắng tập trung vào việc nhận dạng thực phẩm ở giai đoạn tiền chế biến nhằm giúp người dùng chủ động hơn trong việc lựa chọn thực phẩm.

Chúng tôi cố gắng thu hẹp khoảng cách giữa sức khỏe cá nhân, công nghệ và người sử dụng bằng hệ thống được đề xuất trong Hình 1.4. Hệ thống sẽ thực hiện tiền xử lý, phân đoạn, rút trích đặc trưng và phân loại ảnh nhằm tính toán lượng calor cần thiết cho người sử dụng.

Việc nhận dạng thực phẩm dựa trên hình ảnh rất phức tạp và có nhiều khía cạnh cần được xem xét hoặc bỏ qua. Chúng tôi trình bày các khía cạnh chung của bài toán nhận dạng ảnh, các khía cạnh riêng của bài toán nhận dạng thực phẩm; dựa trên các vấn đề này, chúng

tôi đưa ra các giả định cho bài toán nhận dạng thực phẩm trong nghiên cứu của chúng tôi. Chúng tôi cũng giới hạn phạm vi nghiên cứu của bài toán gồm các loại rau, củ, quả trong giai đoạn tiền chế biến với số lượng từ 8 tới 12 loại. Tập dữ liệu được thu thập bằng điện thoại di động ở độ phân giải thấp với số lượng từ 2000 tới 3000 ảnh.

CHƯƠNG 2: PHƯƠNG PHÁP NHẬN DẠNG THỰC PHẨM

Trong chương này, chúng tôi trình bày các bước và thuật toán chính được sử dụng trong hệ thống nhận dạng thực phẩm: thuật toán K-mean nhằm phân đoạn ảnh, các kỹ thuật trích chọn đặc trưng, thuật toán SVM, thuật toán Random Forest phục vụ cho việc phân loại ảnh. Phương pháp quản lý chế độ dinh dưỡng được đưa ra nhằm xây dựng một hệ thống nhận dạng thực phẩm và quản lý dinh dưỡng phù hợp.

2.1 Tiền xử lý ảnh

Trong quá trình thu thập, ảnh có thể bị nhiễu hoặc độ tương phản thấp nên cần đưa vào bộ tiền xử lý để nâng cao chất lượng, làm ảnh rõ nét hơn không bị mất hay bị biến dạng các tính chất về màu sắc, kết cấu. Sau khi điều chỉnh ảnh về kích thước mong muốn (300x300), chúng tôi tiến hành khử nhiễu và mờ bằng cách sử dụng bộ lọc trung vị (Median Filter). Sau đó, ảnh tiếp tục được xử lý để tăng độ tương phản. Việc điều chỉnh sẽ giúp ảnh sau khi phân đoạn thể hiện đầy đủ các tính chất về màu sắc kết cấu

2.2 Phân đoạn ảnh

Sau khi tiền xử lý, chúng tôi chọn thuật toán K-mean để tiến hành phân đoạn dựa trên đặc trưng màu sắc của ảnh. K-mean là thuật toán rất quan trọng và được sử dụng phổ biến trong kỹ thuật phân cụm với khả năng cài đặt dễ dàng và độ phức tạp tuyến tính. Tư tưởng chính của thuật toán K-mean là tìm cách phân nhóm các đối tượng đã cho vào K cụm (K là số các cụm được xác định trước, K nguyên dương) sao cho tổng bình phương khoảng cách giữa các đối tượng đến tâm nhóm là nhỏ nhất. Giải thuật K-mean được trình bày tóm tắt như sau:

1. Chọn ngẫu nhiên K tâm (center) của K cụm (cluster).
2. Tính toán khoảng cách từ tâm tới tất cả các đối tượng (thường dùng khoảng cách Euclid).
3. Nhóm đối tượng vào các nhóm gần nhất dựa trên khoảng cách.
4. Cập nhật lại các tâm của K cụm, mỗi tâm là giá trị trung bình (mean) của các phần tử trong cụm của nó.
5. Lặp lại bước 2, 3, 4 cho tới khi hội tụ.



Hình 2.3: Kết quả sau khi tiến hành phân đoạn K-mean ($K = 3$).

Giả sử màu sắc bề mặt của các đối tượng trong ảnh là một thuộc tính không đổi và được ánh xạ vào một không gian 3 chiều. Khi đó áp dụng giải thuật phân cụm K-mean cho việc xác định các cụm màu, mỗi cụm màu có tập các điểm ảnh tương tự nhau. Trong luận văn này chúng tôi chọn tham số $K = 3$, thuật toán sẽ tiến hành phân cụm ảnh và cho ra ba kết quả khác nhau, mỗi kết quả tương ứng với một màu trong ảnh. Kết quả thu được cuối cùng là sự kết hợp của cả ba cụm tạo ra một ảnh đã được loại bỏ phông nền.

2.3 Trích chọn đặc trưng

2.3.1 Đặc trưng màu sắc

Màu sắc là một đặc trưng nổi bật dựa trên bề mặt của các đối tượng trong ảnh. Mỗi một điểm ảnh (thông tin màu sắc) có thể được biểu diễn như một điểm trong không gian màu sắc ba chiều RGB. Chúng tôi tiến hành rút trích hai đặc trưng màu sắc là *moment màu* (color moment) và *biểu đồ màu* (color Histogram). Chúng không bị ảnh hưởng bởi sự dịch chuyển hay sự xoay của ảnh và rất ít bị ảnh hưởng bởi tỷ lệ và góc nhìn [32].

2.3.2 Đặc trưng SURF

SURF (Speed Up Robust Features) là một trong những đặc trưng mạnh mẽ được sử dụng để nhận dạng và phân loại đối tượng trong lĩnh vực thị giác máy tính, SURF được giới thiệu lần đầu tiên bởi Herbert Bay và nhóm cộng sự tại hội nghị European Conference on Computer Vision năm 2006 [31]. Đặc trưng SURF là các tính năng được mô tả dựa trên máy dò Fast-Hessian và đã được chứng minh tính hiệu quả trong xử lý ảnh mờ (có khả năng rút trích các đặc trưng từ các hình ảnh bị mờ). Hơn nữa, đặc trưng SURF là bất biến với tính chất co giãn và xoay ảnh, những đặc điểm này rất quan trọng trong việc phân loại thực phẩm khi các ảnh thu thập có thể chụp với các góc xoay và khoảng cách khác nhau.



Hình 2.6: Các điểm nổi bật được phát hiện.

Thuật toán của SURF gồm hai phần chính là “phát hiện” và “mô tả”. Bộ phát hiện đặc trưng của SURF sử dụng phép xấp xỉ trên ma trận Hessian và ảnh tích hợp (Integral Image) [27] để nâng cao hiệu năng tính toán một cách đáng kể. Bộ mô tả đặc trưng sử dụng vector 64 chiều chứa thông tin biến thiên trên ảnh dựa trên sự phân phối bậc nhất Haar Wavelet tác động trên trục x và y, kết hợp với ảnh tích lũy làm tăng tốc độ tính toán.

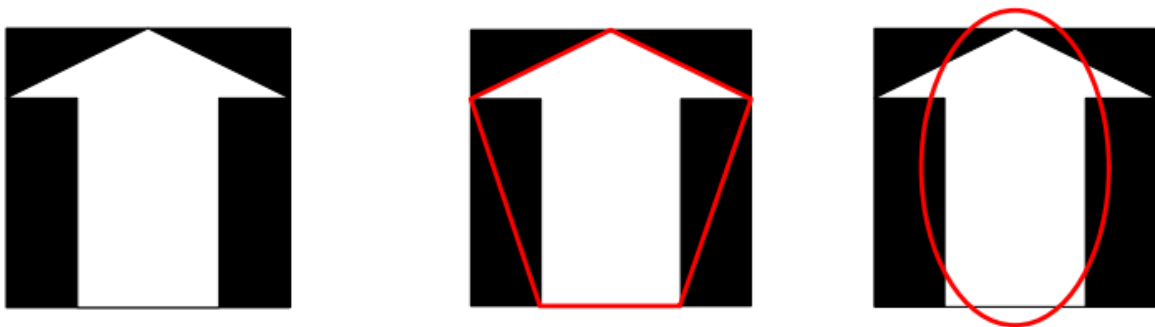
Trong hệ thống của chúng tôi, SURF-64 được lựa chọn để tiến hành rút trích đặc trưng. Mỗi ảnh chúng tôi chọn 32 điểm đặc trưng, mỗi điểm đặc trưng là một vector SURF 64 chiều. Như vậy chúng tôi thu được đặc trưng SURF là một vector $32 \times 64 = 2048$ phần tử, vector đặc trưng SURF sẽ được kết hợp cùng các vector màu sắc, hình dáng và kết cấu để tạo nên bộ vector đặc trưng kết hợp làm đầu vào cho các thuật toán học máy.

2.3.3 Đặc trưng hình dạng

Hình dạng (shape) được biết đến như một đặc trưng quan trọng nhận dạng và phân loại đối tượng trong thế giới thực, mà mục đích là để mã hóa các loại hình học đơn giản như đường thẳng theo các hướng khác nhau. Kỹ thuật trích xuất đặc trưng hình dạng có thể được phân loại thành hai nhóm chính là phương pháp dựa vào đường biên (boundary-based) và phương pháp dựa trên khu vực (region-based) [11].

Trong nghiên cứu này, chúng tôi lựa chọn tám đặc trưng là "*Area*", "*Perimeter*", "*Euler Number*", "*Convex Area*", "*Solidity*", "*Minor Length*", "*Major Length*" và "*Eccentricity*" chúng được chia vào 3 nhóm chính sau:

- "*Area*", "*Perimeter*", và "*Euler Number*" được trích xuất trực tiếp từ đối tượng.
- Tạo một vỏ lồi bằng phương pháp quét Graham [21], đó là đa giác lồi nhỏ nhất bao quanh đối tượng, từ đó trích xuất được đặc trưng "*Convex Area*" và "*Solidity*".
- Tạo một hình Elipse bao quanh đối tượng, từ đó rút trích được các đặc trưng "*Minor length*", "*Major length*" và "*Eccentricity*".



Hình 2.11: Ảnh gốc, đa giác lồi và Elipse bao quanh đối tượng.

2.3.4 Đặc trưng kết cấu

Đặc trưng kết cấu là một tập hợp các chỉ số được tính toán trong xử lý ảnh, nó được thiết kế để xác định số lượng kết cấu của một ảnh. Đặc trưng kết cấu cung cấp thông tin về cách tổ chức không gian của màu sắc hoặc cường độ sáng trong ảnh hoặc một vùng được chọn [32]. Để phân loại đặc trưng cấu trúc ảnh, người ta thường sử dụng hai phương pháp chính là *Phương pháp dựa trên cấu trúc (Structured)* và *Phương pháp dựa trên thống kê (Statistical)*.

Phương pháp dựa trên cấu trúc trình bày cấu trúc như một tập hợp các texel trong một mô hình thường xuyên lặp lại [4]. Phương pháp này hoạt động tốt khi phân tích các kết cấu nhân tạo, nó tỏ ra hữu ích cho việc tổng hợp hơn là phân tích ảnh. Phương pháp dựa trên thống kê trình bày các kết cấu gián tiếp bởi các tính chất không xác định quản lý việc phân phối và quan hệ giữa các mức xám trong ảnh [4], nó phù hợp với kết cấu tự nhiên và tính toán dễ dàng.

Trong luận văn này chúng tôi chọn *Phương pháp dựa trên thống kê* để trích xuất các đặc trưng kết cấu vì nó phù hợp với kết cấu tự nhiên của các thực phẩm tiền chế biến và dễ dàng tính toán, sử dụng. Chúng tôi thực hiện trích chọn bảy đặc trưng theo phương pháp này bao gồm ***Contrast, Homogeneity, Energy*** và ***Correlation*** dựa trên ma trận đồng hiện mức xám (Co-occurrence Maxtrix). Ngoài ra chúng tôi còn thực hiện rút trích các đặc trưng khác là ***Mean, Variance*** và ***Entropy*** dựa trên biểu đồ thống kê mức xám của ảnh.

2.4 Phân loại hình ảnh

Từ việc trích chọn đặc trưng, chúng tôi thu được các vector đặc trưng, mỗi vector gồm bốn loại đặc trưng là màu sắc, SURF, hình dạng và kết cấu. Từ các vector đặc trưng này, chúng tôi sử dụng độc lập hai thuật toán học máy là SVM và RandomForest để phân loại hình ảnh nhằm so sánh hiệu năng của chúng và chọn ra thuật toán tốt nhất cho hệ thống.

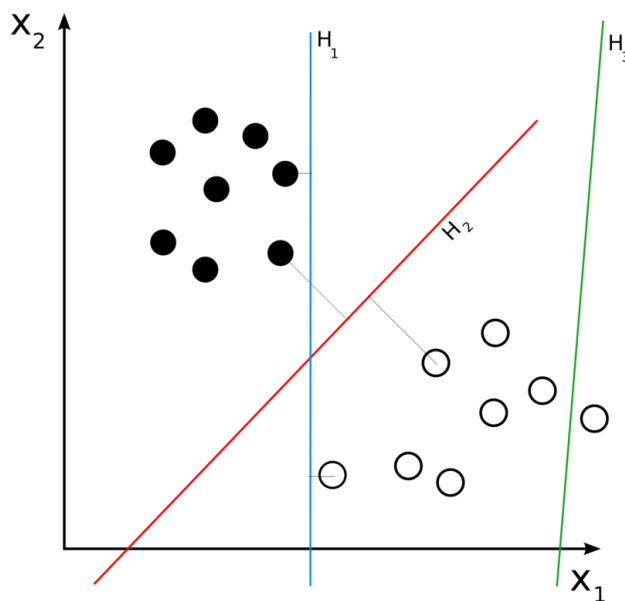
2.4.1 Bài toán phân lớp

Phân lớp là một tiến trình xử lý nhằm xếp các mẫu dữ liệu hay các đối tượng vào một trong các lớp đã được định nghĩa trước. Các mẫu dữ liệu hay các đối tượng được xếp vào các lớp dựa vào giá trị của các thuộc tính cho một mẫu dữ liệu hay đối tượng. Sau khi đã xếp tất cả các đối tượng đã biết trước vào các lớp tương ứng thì mỗi lớp được đặc trưng bởi tập các thuộc tính của các đối tượng chứa trong lớp đó. Quá trình phân lớp còn được gọi là

quá trình gán nhãn cho các tập dữ liệu. Nhiệm vụ của bài toán phân lớp dữ liệu là cần xây dựng mô hình (bộ) phân lớp để khi có một dữ liệu mới vào thì mô hình phân lớp sẽ cho biết dữ liệu đó thuộc lớp nào. Có nhiều cách để biểu diễn một mô hình phân lớp và có rất nhiều thuật toán giải quyết nó. Các thuật toán phân lớp tiêu biểu bao gồm như mạng nơron, cây quyết định, suy luận quy nạp, mạng Bayesian, Support Vector Machine (SVM), Random Forest... Trong các kỹ thuật trên, SVM và RandomForest là những giải thuật học nhanh, chịu đựng nhiễu tốt và phân lớp hiệu quả các tập dữ liệu có số chiều lớn.

2.4.2 Thuật toán máy hỗ trợ vector (SVM)

Thuật toán máy hỗ trợ vector (SVM) [29] xây dựng một siêu phẳng hoặc một tập hợp các siêu phẳng trong một không gian nhiều chiều hoặc vô hạn chiều, có thể được sử dụng cho phân loại, hồi quy, hoặc các nhiệm vụ khác. Mục tiêu của thuật toán là xác định xem một điểm dữ liệu *mới* sẽ thuộc về lớp nào. Mỗi điểm dữ liệu được biểu diễn dưới dạng một vector p - chiều, và ta muốn biết liệu có thể chia tách hai lớp dữ liệu bằng một siêu phẳng $p - 1$ chiều, đây gọi là phân loại tuyến tính. Có nhiều siêu phẳng có thể phân loại được dữ liệu, một lựa chọn hợp lý trong chúng là siêu phẳng có lề lớn nhất giữa hai lớp.



Hình 2.14: Đường thẳng tuyến tính H_1 (xanh lơ) và H_2 (đỏ) phân tách hai lớp dữ liệu.

- SVM cho bài toán phân lớp tuyến tính
- Xác định siêu phẳng với lề cực đại
- Kỹ thuật hàm nhân
- Thiết lập các tham số

Trong luận văn này, chúng tôi sử dụng thuật toán tối thiểu hóa tuần tự (Sequential Minimal Optimization – SMO) được phát minh bởi John Platt tại Microsoft Research [28]. SMO được sử dụng rộng rãi cho việc huấn luyện SVM và được thực thi bởi công cụ LIBSVM phổ biến [9][22]. Chúng tôi chọn tham số $C = 1$ và hàm nhân đa thức (polynomial kernel function) với $d = 1$:

$$K(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}' + 1)^d$$

Khi đó, hàm nhân đa thức là hàm tuyến tính và SVM là tuyến tính.

2.4.3 Thuật toán rừng ngẫu nhiên (Random Forest)

Rừng ngẫu nhiên là một phương pháp phân lớp thuộc tính dựa trên các cây quyết định được phát triển bởi Leo Breiman tại đại học California, Berkeley [7]. Rừng ngẫu nhiên được xây dựng dựa trên 3 thành phần chính là: (1) CART, (2) học toàn bộ, hội đồng các chuyên gia, kết hợp các mô hình, và (3) tổng hợp bootstrap (bagging).

Rừng ngẫu nhiên xây dựng cây phân lớp bằng cách lựa chọn ngẫu nhiên một nhóm nhỏ các thuộc tính tại mỗi nút của cây để phân chia cho mức tiếp theo của cây phân lớp, bằng cách chia nhỏ không gian tìm kiếm thành các cây nhỏ hơn như vậy cho phép thuật toán có thể phân loại nhanh dù không gian thuộc tính rất lớn. Ngoài ra tập mẫu của mỗi cây cũng được lựa chọn ngẫu nhiên bằng phương pháp bootstrap từ tập mẫu ban đầu. Số lượng các cây phân lớp trong rừng là không hạn chế và mỗi cây sẽ “bỏ phiếu” cho lớp phổ biến nhất. Giải thuật rừng cho bài toán phân lớp được trình bày như sau:

- Cho tập huấn luyện $D = \{(\mathbf{x}_i, y_i) : i = 1, 2, \dots, n\}$. Lấy ra k mẫu bootstrap từ tập huấn luyện, mỗi mẫu gồm m phần tử được chọn ngẫu nhiên từ D . Ta có tập mẫu $K = \{(D_i) : i = 1, 2, \dots, k\}$ với $D_i = \{(\mathbf{x}_j, y_j) : j = 1, 2, \dots, m\}$
- Đối với mỗi cây trong rừng k cây
 - Xây dựng cây phân lớp T_i với tập D_i được lấy bootstrap từ D
 - Tại mỗi nút của cây, chọn thuộc tính tốt nhất để phân chia cây từ tập thuộc tính F_i được chọn ngẫu nhiên trong tập thuộc tính F .
 - Mỗi cây được xây dựng đến độ sâu tối đa (không cắt nhánh)
- Đưa ra dự đoán bằng cách sử dụng kết quả số đông của các cây (bỏ phiếu).

➤ *Thiết lập các tham số*

Qua việc thực nghiệm, chúng tôi chọn số lượng cây là 300 cho quá trình huấn luyện. Số lượng này đảm bảo độ chính xác đủ tốt và thời gian thực thi đủ nhanh.

2.5 Quản lý chế độ dinh dưỡng

2.6 Kết luận chương 2

Trong chương này, chúng tôi trình bày bốn bước chính của hệ thống nhận dạng thực phẩm tươi bằng hình ảnh. Bước đầu tiên chúng tôi tiến hành tiền xử lý ảnh nhằm loại bỏ nhiễu không mong muốn và nâng cao độ tương phản nhằm mục đích sử dụng thuật toán K-mean để phân đoạn ảnh đầu vào một cách dễ dàng ở bước tiếp theo. Từ đó chúng tôi trích chọn được những đặc trưng của ảnh. Bước cuối cùng, chúng tôi sử dụng các vector đặc trưng bao gồm màu sắc, SURF, hình dạng, kích thước làm đầu vào cho hai thuật toán học máy rừng ngẫu nhiên (Random Forest) và máy hỗ trợ vector (SVM). Phương pháp quản lý chế độ dinh dưỡng được trình bày tạo cơ sở cho việc xây dựng hệ thống nhận dạng thực phẩm và quản lý dinh dưỡng phù hợp và hiệu quả.

CHƯƠNG 3: THỬ NGHIỆM VÀ ĐÁNH GIÁ

3.1 Tập dữ liệu huấn luyện

Tập dữ liệu được chúng tôi thu thập từ điện thoại di động (Microsoft Lumia 630, 8MP) ở độ phân giải thấp nhất là 640x480 điểm ảnh, ngoài ra chúng tôi cũng tiến hành thu thập một số hình ảnh trên mạng. Tập ảnh thu thập được gồm khoảng 2000-3000 ảnh của 10 loại thực phẩm khác nhau bao gồm: cà chua, cà rốt, dưa, dưa chuột, khoai tây, khổ qua, ớt đỏ, ớt vàng, ớt xanh và susu được chúng tôi phân loại bằng tay vào hai nhóm chính:

- *Hình ảnh đơn thực phẩm (SingleFood)*: mỗi hình ảnh chỉ có một loại thực phẩm duy nhất. Các hình ảnh đều đáp ứng được giả định của bài toán nêu ra trong phần 1.5. Số lượng rơi vào khoảng 2000 ảnh.
- *Hình ảnh đa thực phẩm (MultiFood)*: mỗi hình ảnh có từ 2 loại thực phẩm trở lên. Các hình ảnh đều đáp ứng được giả định của bài toán nêu ra trong phần 1.5. Số lượng khoảng 1000 ảnh.

Loại thực phẩm	Số lượng ảnh
Cà chua	160
Cà rốt	166
Dứa	195
Dưa chuột	171
Khoai tây	134
Khổ qua	284
Ớt đỏ	264
Ớt vàng	264
Ớt xanh	220
Susu	181
Tổng	2039

Bảng 3.1: Tập dữ liệu SingleFood.

Do thời gian hạn chế, chúng tôi chỉ tiến hành nghiên cứu và thử nghiệm trên tập dữ liệu SingleFood gồm 2039 ảnh, số lượng cụ thể mỗi loại được chúng tôi liệt kê chi tiết trong Bảng 3.1.

3.2 Thử nghiệm và đánh giá

Chúng tôi tiến hành thử nghiệm và đánh giá trên tập dữ liệu SingleFood bằng hai thuật toán SVM và Random Forest. Sau giai đoạn trích chọn đặc trưng, chúng tôi thu được tập các vector đặc trưng được sử dụng để huấn luyện cho hai bộ phân loại. Tập các vector đặc trưng được chúng tôi chia vào 3 nhóm huấn luyện chính như sau:

- **Nhóm thứ nhất:** gồm tập các vector đặc trưng được thu thập dựa trên 2 đặc trưng là màu sắc và SURF (Color + SURF).
- **Nhóm thứ hai:** gồm tập các vector đặc trưng được thu thập dựa trên 3 đặc trưng là màu sắc, SURF và hình dạng (Color + SURF + Shape).
- **Nhóm thứ ba:** gồm tập các vector đặc trưng được thu thập dựa trên 4 đặc trưng là màu sắc, SURF, hình dạng và kết cấu (Color + SURF + Shape + Texture).

Chúng tôi sử dụng phương pháp kiểm thử *k-fold cross validation*[13] với $k = 10$ để tránh việc trùng lặp giữa các tập kiểm thử.

Kết quả thử nghiệm gồm 2 độ đo là Precision (độ chính xác) và Recall (độ bao phủ) được tính toán theo công thức sau:

$$\text{Precision} = TP / (TP + FP)$$

$$\text{Recall} = TP / (TP + FN)$$

Trong đó:

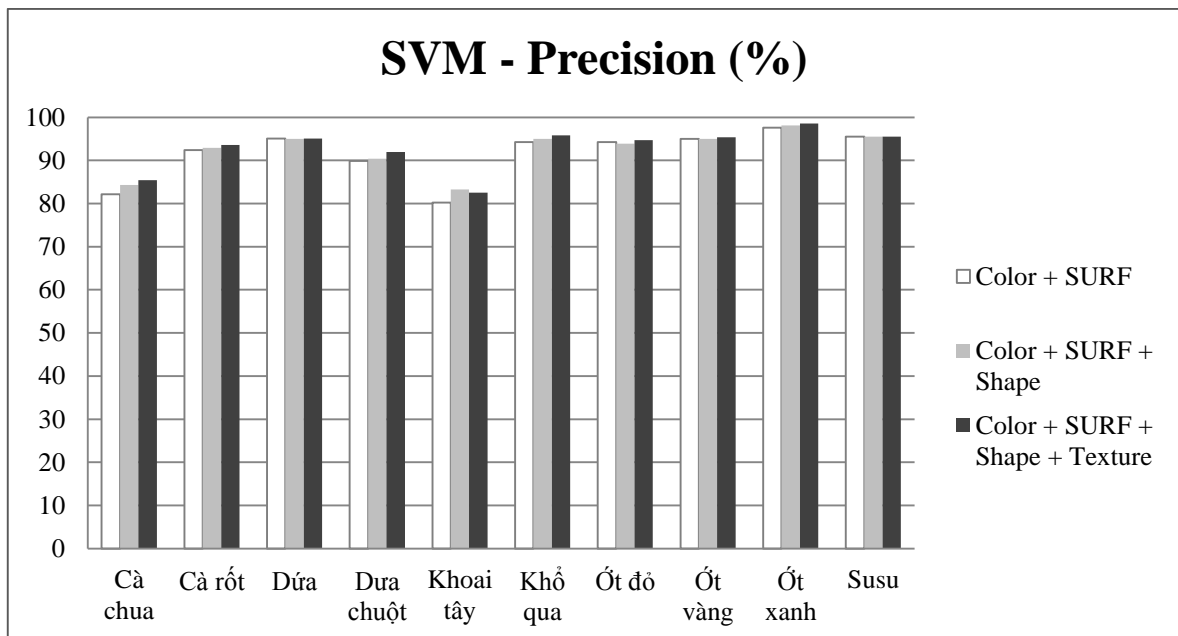
TP (True Positive): là số lượng các ví dụ thuộc lớp C_i được phân loại chính xác vào lớp C_i .

FP (False Positive): là số lượng các ví dụ không thuộc lớp C_i bị phân loại nhầm vào lớp C_i .

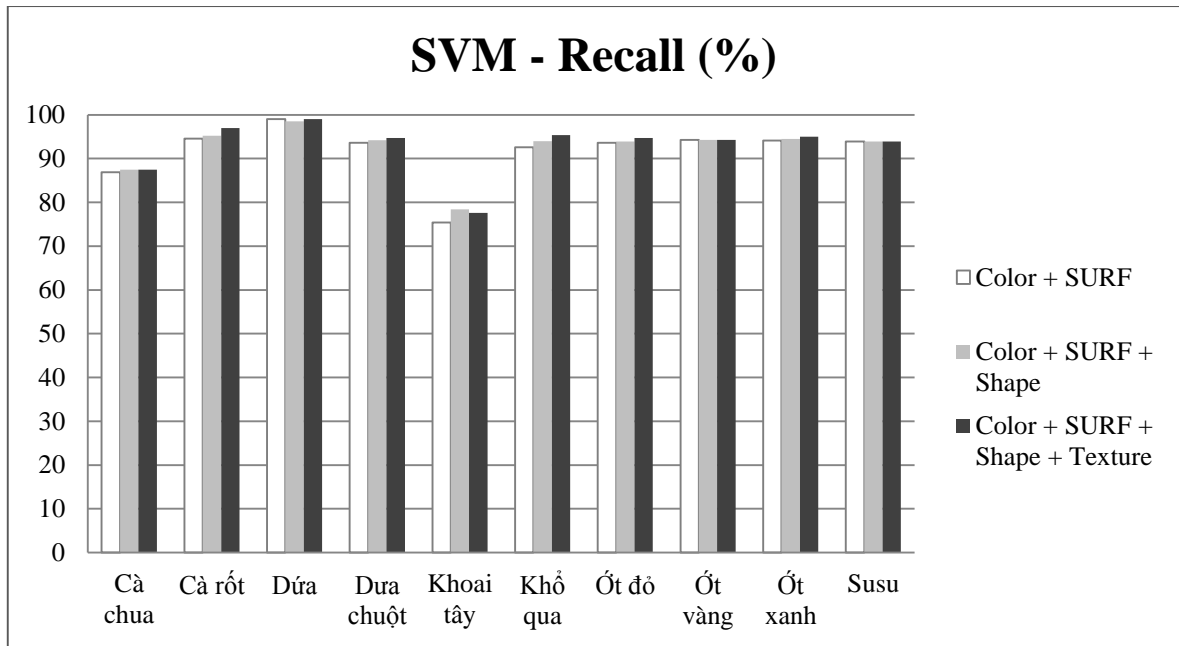
FN (False Negative): là số lượng các ví dụ thuộc lớp C_i bị phân loại nhầm vào các lớp khác C_i .

3.2.1 Thử nghiệm và đánh giá trên thuật toán SVM

Chúng tôi sử dụng thuật toán SMO [28] được thực thi bởi thư viện LibSVM [9][22] với tham số $C = 1$ và hàm nhân đa thức $K(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}' + 1)^d : (d = 1)$. Kết quả thu được như sau :



Hình 3.2: Kết quả Precision khi chạy thử nghiệm với SVM.

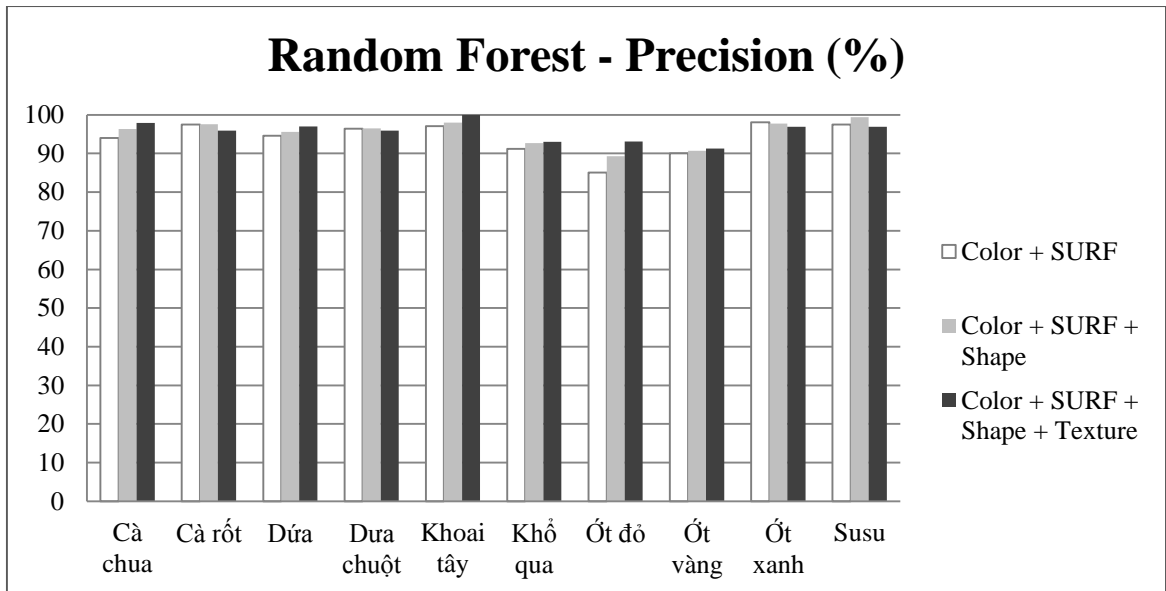


Hình 3.3: Kết quả Recall khi chạy thử nghiệm với SVM.

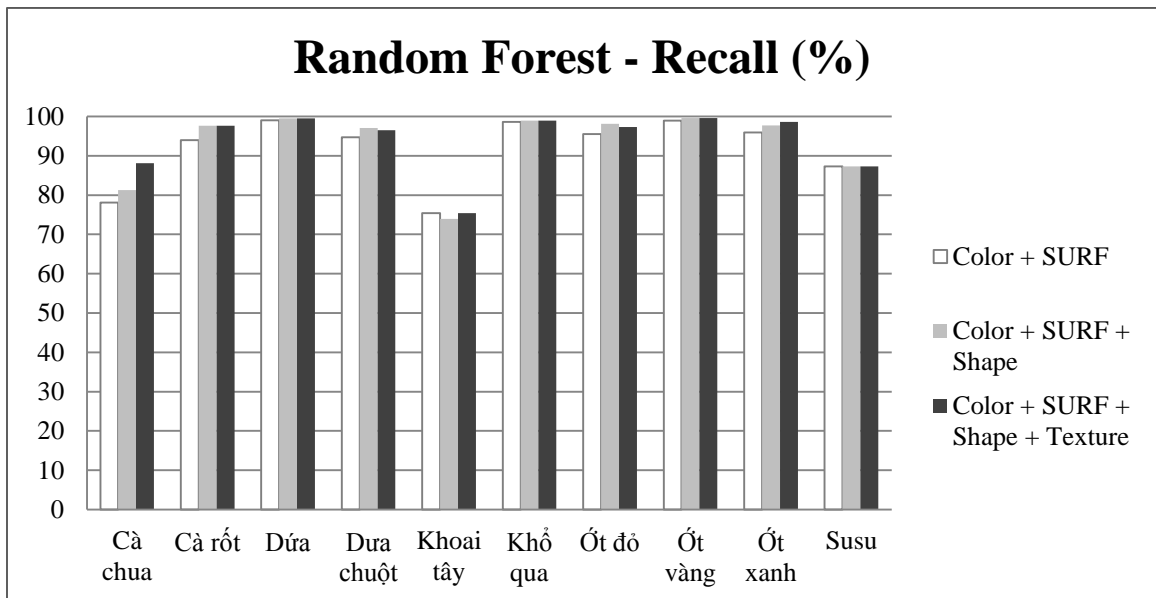
Từ kết quả trên, chúng tôi thấy rằng nhóm đặc trưng thứ ba bao gồm màu sắc, SURF, hình dạng và kết cấu cho kết quả phân loại tốt nhất với giá trị trung bình Precision và Recall đều đạt 93%. Trong số này, cà chua và khoai tây có kết quả phân loại kém nhất với Precision và Recall đạt khoảng 80%. Màu sắc củ cà chua tương tự như ớt đỏ, hình dạng tương tự như khoai tây và các loại ớt ở một số góc chụp khác nhau, quan trọng hơn cả là đặc trưng SURF của chúng không mạnh do không có đặc điểm riêng về hình dạng và có phần tương tự nhau, điều này khiến việc phân loại dễ bị nhầm lẫn. Số còn lại đều đạt kết quả cao với Precision và Recall đạt trên 90%, trong đó dứa cho kết quả phân loại cao nhất do chúng có đặc trưng rất riêng biệt về hình dạng khiến đặc trưng SURF của chúng mạnh hơn các loại thực phẩm khác.

3.2.2 Thử nghiệm và đánh giá trên thuật toán *Random Forest*

Chúng tôi chọn số lượng cây là 300 cho quá trình phân loại, giá trị này được chúng tôi chọn trong quá trình thực nghiệm bằng cách thay đổi liên tục các giá trị trong quá trình phân loại và chọn ra kết quả tốt nhất. Kết quả thu được như sau:



Hình 3.4: Kết quả Precision khi chạy thử nghiệm với Random Forest.



Hình 3.5: Kết quả Recall khi chạy thử nghiệm với Random Forest.

Thuật toán rừng ngẫu nhiên cho kết quả phân loại tốt nhất với nhóm đặc trưng thứ ba bao gồm màu sắc, SURF, hình dạng và kết cấu. Giá trị trung bình của Precision và Recall đều đạt 95%. Kết quả thấp nhất vẫn nằm ở hai loại cà chua và khoai tây, với giá trị Recall tương ứng là 88% và 75%; do đặc trưng SURF và hình dạng của chúng là tương tự nhau nên chúng dễ bị phân loại nhầm vào các lớp khác và ngược lại, tuy nhiên độ chính xác (Precision) của chúng đã được cải thiện đáng kể với kết quả đều đạt trên 90%.

3.2.3 So sánh kết quả SVM và Random Forest

Từ việc thử nghiệm hai thuật toán, chúng tôi thấy rằng nhóm đặc trưng thứ ba bao gồm màu sắc, SURF, hình dạng và kết cấu cho ra kết quả tốt nhất. Chúng tôi tiến hành so sánh hai thuật toán dựa trên kết quả của nhóm đặc trưng này. Kết quả so sánh được trình bày trong Bảng 3.2 sau:

Độ đo Loại thực phẩm	Precision		Recall	
	SVM	RF	SVM	RF
Cà chua	85.4	97.9	87.5	88.1
Dứa	95.1	97	99	99.5
Dưa chuột	92	95.9	94.7	96.5
Khoai tây	82.5	100	77.6	75.4
Khổ qua	95.8	93	95.4	98.9
Ớt đỏ	94.7	93.1	94.7	97.3
Ớt vàng	95.4	91.3	94.3	99.6
Ớt xanh	98.6	96.9	95	98.6
Susu	95.5	96.9	93.9	87.3
Tổng	93.6	95.3	93.6	95.1

Bảng 3.2: So sánh kết quả chạy thực nghiệm giữa SVM và Random Forest (%).

Từ bảng so sánh trên, chúng tôi thấy rằng thuật toán Random Forest cho kết quả phân loại tốt hơn, đặc biệt là việc nâng cao độ chính xác (Precision) của Cà chua và Khoai tây từ 85% tới trên 97%, đây là hai loại thực phẩm có bộ vector đặc trưng yếu do chúng không có nhiều đặc trưng SURF riêng và hình dạng là tương tự nhau, đặc biệt màu sắc và hình dạng của cà chua có thể tương tự như ớt đỏ ở một số góc chụp khác nhau. Tuy nhiên thuật toán Random Forest đã cải tiến đáng kể độ chính xác nhờ việc tiến hành dự đoán dựa trên kết quả số đông (bỏ phiếu) của các cây. Chúng tôi cũng thấy rằng đặc trưng SURF đóng vai trò chủ chốt trong việc tính toán độ chính xác và phản hồi của cả hai thuật toán. Điều này có thể nhìn thấy dễ dàng qua kết quả phân loại của Dứa, khổ qua và các loại ớt; chúng đều chứa các đặc trưng SURF rất mạnh.

Ngoài ra trong quá trình nghiên cứu, chúng tôi cũng tiến hành chạy thử nghiệm cả hai thuật toán trên hệ thống thực tế để so sánh hiệu năng thực sự. Chúng tôi sử dụng Matlab

R2016a (Image Processing Toolbox, Statistic Toolbox) để thực hiện quá trình xử lý ảnh, phân đoạn, rút trích đặc trưng, huấn luyện và nhận dạng. Kết quả huấn luyện được chúng tôi lưu vào các tập tin nhị phân, trong quá trình phân loại dữ liệu huấn luyện sẽ được đọc ra từ các tập tin nhị phân này nhằm giảm chi phí về mặt bộ nhớ thời gian. Bỏ qua khía cạnh bộ nhớ lưu trữ của hệ thống, chúng tôi chỉ xem xét hiệu năng về mặt thời gian. Thời gian chạy thực tế được tính từ lúc ảnh được thu thập cho tới khi có kết quả phân loại. Kết quả được chúng tôi thử nghiệm trên 10 loại thực phẩm khác nhau được trình bày trong Bảng 3.3 sau:

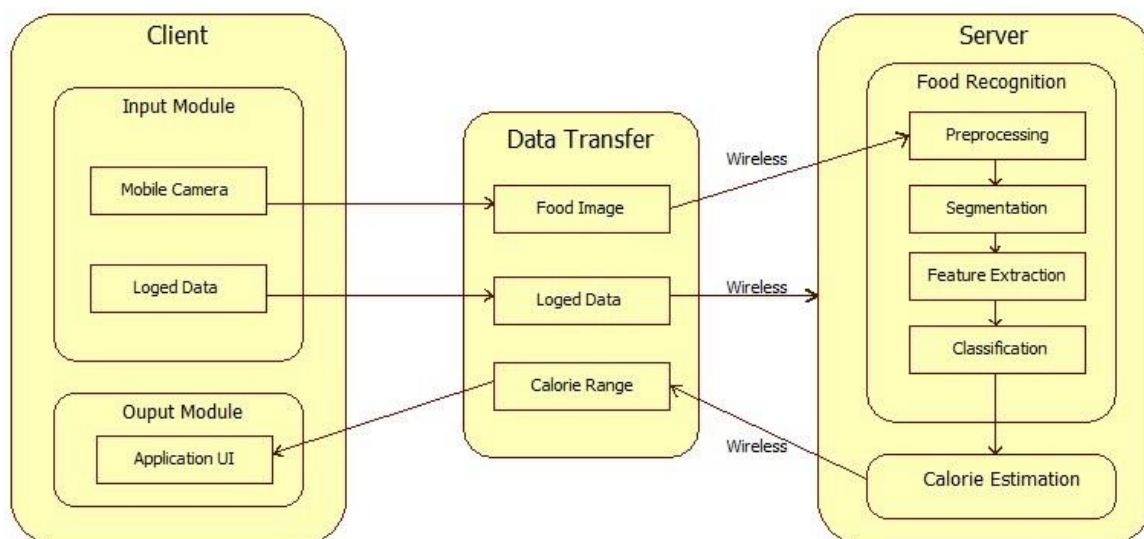
Thuật toán Lần chạy	SVM	RF
Lần 1	7.5	5.3
Lần 2	6.9	4.6
Lần 3	7.2	4.2
Lần 4	6.8	4.4
Lần 5	6.9	4.4
Lần 6	7.0	5.5
Lần 7	6.5	4.2
Lần 8	6.6	4.4
Lần 9	6.9	4.6
Lần 10	6.7	4.1
Trung bình	6.9	4.57

Bảng 3.3: So sánh thời chạy thực nghiệm giữa SVM và Random Forest (giây).

Qua việc so sánh độ chính xác và hiệu năng, chúng tôi thấy rằng thuật toán rừng ngẫu nhiên cho kết quả phân loại cao hơn và hiệu năng tốt hơn về mặt thời gian thực, vì vậy rừng ngẫu nhiên là thuật toán được chúng tôi lựa chọn để xây dựng hệ thống nhận dạng. Các đặc trưng được hệ thống sử dụng trong quá trình phân loại bao gồm màu sắc, SURF, hình dạng và kết cấu.

3.3 Hệ thống nhận dạng thực phẩm

Chúng tôi tiến hành xây dựng hệ thống FoodAutoManage nhằm quản lý dinh dưỡng cá nhân với kiến trúc được thể hiện trong Hình 3.6.



Hình 3.6: Kiến trúc hệ thống FoodAutoManage.

Hệ thống được chúng tôi chia làm hai phần chính bao gồm Client và Server.

- **Client:** được chúng tôi xây dựng trên nền tảng di động Windows 10 (Microsoft Lumia 650) dựa trên ngôn ngữ C#. Client bao gồm hai module chính là Input và Output. Input module cho phép người dùng thu thập ảnh từ camera hoặc hình ảnh có sẵn trong thiết bị, sau đó ảnh sẽ được gửi tới Server để tiến hành xử lý. Input module cũng đồng thời tiến hành thu thập dữ liệu nhằm ghi log các hoạt động của người dùng, điều này khiến việc quản lý dinh dưỡng dễ dàng hơn, các chuyên gia y tế cũng có thể dựa vào dữ liệu này để đưa ra các dự đoán và phác đồ điều trị dễ dàng hơn. Output module cho phép Client hiển thị các thông tin kết quả trả về từ Server, hiển thị các thực phẩm được người dùng lựa chọn, hiển thị danh sách log.
- **Server:** được chúng tôi xây dựng trên nền tảng PC Windows 10, ngôn ngữ C# và Matlab. Việc kết nối và truyền dữ liệu giữa Client và Server được thực hiện thông qua kết nối không dây TCP/IP. Module chính của Server là FoodRecognition được chúng tôi thực hiện bằng ngôn ngữ Matlab (Matlab R2016a - Image Processing Toolbox, Statistic Toolbox), nhằm thực hiện xử lý ảnh, phân đoạn (K-mean), rút trích đặc trưng, nhận dạng (Random Forest) và trả kết quả về Client.

Như chúng tôi đã trình bày trong phần 3.2.3, độ chính xác của hệ thống rơi vào khoảng 95%. Thời gian chạy trung bình của mỗi phiên (tính từ lúc ảnh được gửi từ phía

Client tới khi Server trả về kết quả) rơi vào khoảng 4 - 5 giây. Những kết quả này là khả thi để có thể áp dụng hệ thống vào thực tế.

3.4 Kết luận chương 3

Trong chương này, chúng tôi đã nêu ra cách thức thu thập dữ liệu huấn luyện sử dụng máy ảnh của điện thoại di động (Microsoft Lumia 650, 8MP) ở độ phân giải thấp (640x480 pixel). Tập ảnh thu được gồm khoảng 2000 tới 3000 ảnh của 10 loại thực phẩm khác nhau bao gồm: cà chua, cà rốt, dưa, dưa chuột, khoai tây, khổ qua, ớt đỏ, ớt vàng, ớt xanh và susu được chúng tôi phân loại bằng tay vào 2 nhóm chính là *hình ảnh đơn thực phẩm* (*SingleFood* – 2000 ảnh) và *hình ảnh đa thực phẩm* (*MultiFood* – 1000 ảnh). Do hạn chế về mặt thời gian, chúng tôi chỉ tiến hành nghiên cứu và thử nghiệm trên tập dữ liệu *SingleFood* gồm 2039 ảnh được liệt kê chi tiết trong Bảng 3.1.

Kết quả thử nghiệm và đánh giá được chúng tôi tiến hành bằng hai thuật toán SVM và Random Forest trên tập các vector đặc trưng được chúng tôi chia thành 3 nhóm huấn luyện chính (nhóm thứ nhất gồm Color + SURF, nhóm thứ hai gồm Color + SURF + Shape và nhóm thứ ba gồm Color + SURF + Share + Texture). Chúng tôi sử dụng phương pháp kiểm thử *10 – fold cross validation* để tránh việc trùng lặp giữa các tập kiểm thử. Kết quả thu được cho thấy Random Forest cho kết quả phân loại tốt hơn trên nhóm huấn luyện thứ ba với kết quả độ chính xác và bao phủ trung bình đạt trên 95%, thời gian phân loại trung bình từ 4-5 giây với số lượng cây được chúng tôi chọn là 300.

Hệ thống nhận dạng thực phẩm FoodAutoManage được chúng tôi xây dựng với kiến trúc được trình bày trong Hình 3.6 gồm hai phần chính là Client và Server dựa trên kết nối TC/IP. Client tiến hành thu thập ảnh và gửi về để Server tiến hành xử lý, phân đoạn, rút trích đặc trưng, phân loại ảnh và trả về kết quả để người dùng lựa chọn.

KẾT LUẬN

Luận văn tập trung vào việc xây dựng hệ thống nhận dạng thực phẩm tươi và ứng dụng quản lý chế độ dinh dưỡng trên nền tảng di động với chi phí thấp, độ chính xác cao, tự động, dễ dùng và mang tính cá nhân. Hệ thống giúp người dùng quản lý và theo dõi chế độ dinh dưỡng hàng ngày, thuận tiện cho quá trình kiểm soát trọng lượng cơ thể. Những dữ liệu do hệ thống ghi lại sẽ cung cấp thông tin hữu ích cho các chuyên gia y tế trong việc chăm sóc và đưa ra phác đồ điều trị hợp lý cho các bệnh nhân.

Việc xây dựng hệ thống trải qua các bước chính bao gồm tiền xử lý ảnh, phân đoạn ảnh bằng thuật toán K-mean, trích chọn các đặc trưng của ảnh về màu sắc, hình dạng, kết cấu, SURF và kết hợp chúng nhằm tạo ra vector đặc trưng để tiến hành phân loại. Chúng tôi tiến hành chạy thử nghiệm hai thuật toán học máy là SVM và Random Forest trên tập dữ liệu huấn luyện nhằm đưa ra các đánh giá, so sánh về hiệu năng, độ chính xác của chúng nhằm chọn ra thuật toán tốt nhất để tiến hành xây dựng hệ thống. Kết quả thực nghiệm trên tập dữ liệu SingleFood cho thấy Random Forest cho kết quả phân loại cao hơn với độ chính xác và bao phủ đều đạt tới 95%; thời gian phân loại nhanh hơn với tổng thời gian tính từ lúc ảnh được chụp tới khi có kết quả phân loại từ 4 – 5 giây. Kết quả là khả thi để áp dụng hệ thống vào thực tế.

Những công việc trên đòi hỏi chúng tôi có các nghiên cứu chuyên sâu về các lĩnh vực xử lý ảnh, học máy, phân tích, thiết kế và xây dựng hệ thống. Tuy nhiên do những hạn chế nhất định về mặt thời gian và kiến thức của bản thân, luận văn mới chỉ dừng lại ở mức thử nghiệm hệ thống trên tập dữ liệu SingleFood với số lượng ảnh khá khiêm tốn cho việc nhận dạng thực phẩm.

Trong thời gian tới, chúng tôi sẽ tiếp tục tiến hành nghiên cứu và thử nghiệm hệ thống với tập dữ liệu MultiFood. Chúng tôi cũng tiếp tục tăng số lượng ảnh huấn luyện và đưa thêm vào đặc trưng ngữ cảnh để nâng cao độ chính xác, đồng thời tìm cách giảm số chiều không gian của dữ liệu để nâng cao hiệu năng hướng tới xây dựng một hệ thống hoàn chỉnh.