

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
ĐẠI HỌC CÔNG NGHỆ THÔNG TIN



Báo cáo đồ án

Môn học: Máy học-CS114.K21.KHTN

Năm học: Học kì II 2019-2020

Đề tài:

PHÂN LOẠI TRÁI CÂY QUA HÌNH ẢNH

GVHD:

PSG.TS: Lê Đình Duy

ThS: Phạm Nguyễn Trường An

Sinh viên thực hiện:

Phạm Xuân Trí 18521530

Hồ Chí Minh Ngày 15 Tháng 08 Năm 2020

Mục lục

I-Giới thiệu:	2
1-Lời mở đầu:.....	2
2- Vai trò của bài toán:.....	3
3-Mục tiêu đề tài :	3
II- Mô tả bài toán:	4
III- Cách thức thu thập và xử lí dữ liệu:	4
1- Cách thu thập bộ dữ liệu:.....	4
2- Tiền xử lí dữ liệu:	4
3- Rút trích đặc trưng :	5
IV-Huấn luyện và đánh giá mô hình:	7
V- Tổng kết:	11
1 Kết quả đạt được:.....	11
2 Kinh nghiệm đạt được:	11
VI- Tài liệu tham khảo	12

I-Giới thiệu:

1-Lời mở đầu:

Tại Israel chỉ với 2,5% dân số làm nông nghiệp trong điều kiện tự nhiên khắc nghiệt (2/3 diện tích lãnh thổ là sa mạc) nhưng nền nông nghiệp điện tử của nước này tạo ra giá trị sản lượng gần 23 tỉ USD/năm, xuất khẩu nông sản đạt trên 3 tỉ USD/năm. Đến nay mỗi nông dân Israel có thể nuôi được hơn 150 người khác. Tại sao với những khó khăn về điều kiện tự nhiên mà đất nước này vẫn phát triển nền nông nghiệp mạnh mẽ như vậy ? Đó chính là nhờ vào việc áp dụng các thành tựu CNTT vào nông nghiệp.



Nhìn chung việc áp dụng CNTT vào nông nghiệp có 2 hướng chính:

- + Hệ thống theo dõi, kiểm soát trồng trọt (74,62%)
- + Hệ thống theo dõi, kiểm soát chăn nuôi gia súc, gia cầm (18,28%).

Số liệu được lấy [từ đây](#)

Trong đồ án lần này tôi tìm hiểu về một bài toán nhỏ trong việc kiểm soát trồng trọt. Đó là bài toán nhận diện trái cây thông qua hình ảnh. Bài toán được thực hiện dựa trên các kiến thức đã học từ môn machine learning trong kì này

2- Vai trò của bài toán:

Nhận diện trái cây tuy là một bài toán nhỏ trong các bài toán lớn, nhưng nó cũng đóng một vai trò quan trọng. Các mô hình quản lí trái cây ở siêu thị, hay việc đếm trái cây tự động và nhiều bài toán khác đều cần bước cơ bản đầu tiên chính là **nhận diện được trái cây**.

3-Mục tiêu đề tài :

Hiện nay bài toán đã được thực hiện với việc áp dụng các kĩ thuật tiên tiến như: Deeplearning. Tuy nhiên trong đồ án này tôi sẽ không sử dụng

phương pháp này mà dùng các mô hình học máy truyền thống để phân loại trái cây. Việc chọn cách làm này nhằm mục đích áp dụng các kiến thức đã học vào đồ án cụ thể và thực tế. Cũng như được trải nghiệm được quy trình thực hiện đồ án máy học, sẽ gặp những thách thức và khó khăn như thế nào.

II- Mô tả bài toán:

Đây là bài toán phân loại 12 loại trái cây bao gồm: **Cam, chuối, ổi, thanh long, dưa hấu, mận, bơ, xoài, măng cầu, dưa, khế, táo**. Đầu vào của bài toán là một ảnh chứa một loại trái cây. Đầu ra của bài toán sẽ cho biết loại trái cây đó thuộc loại nào trong 12 loại đã nêu trên. Hướng mở rộng của bài toán sẽ phân loại nhiều loại trái cây hơn. Hoặc là phân loại trái cây quay video

III- Cách thức thu thập và xử lý dữ liệu:

1- Cách thu thập bộ dữ liệu:

Đồ án này thu thập dữ liệu từ 2 nguồn chính là internet và tự thu thập. Với 95% dữ liệu là tự thu thập cùng với 2 bạn cùng lớp, còn lại là 5% lấy từ nguồn internet. Ban đầu dữ liệu tự thu thập được thu thập bằng cách quay video sau đó tách frame. Phương pháp này giúp tiết kiệm thời gian thu thập dữ liệu nhưng lại cho chất lượng thấp. Vì vậy chúng tôi đã thu thập thêm dữ liệu bằng cách chụp trực tiếp từ điện thoại. Ảnh cho ra chất lượng cao, tuy nhiên phương pháp này đòi hỏi người thu thập phải cực khổ hơn phương pháp trước. Bộ dữ liệu trên Internet được tải trên kaggle[5]

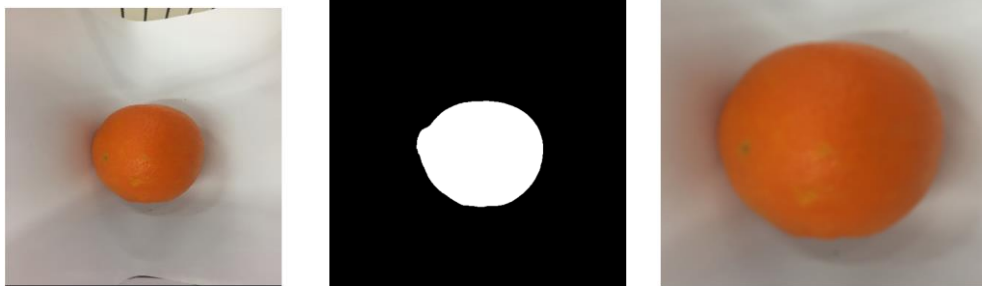
Dữ liệu sẽ được trộn với nhau. Bộ dữ liệu sau khi trộn lại sẽ chia làm 75% cho tập huấn luyện (train data) và 25% cho tập kiểm thử (test data). Dữ liệu được lưu trên drive, mỗi ảnh của trái cây sẽ nằm trong folder trái cây tương ứng.

2- Tiền xử lý dữ liệu:

Dữ liệu sau khi thu thập sẽ qua bước tiền xử lý. Ảnh sẽ được trích xuất ra vùng **ROI (region of interest)**. ROI trong đồ án này được hiểu là vùng nhỏ nhất của ảnh có thể chứa được trái cây. Việc trích xuất ROI làm cho các mô hình dự đoán được hiệu quả hơn, giảm không gian làm việc trên bức ảnh. Có 2 cách thực hiện việc này:

- Cách thức nhất dùng code để thực hiện việc này. Tôi sẽ xây dựng hàm trích xuất ROI thông qua 5 bước [1]:

- Bước 1: Chuyển đổi hệ màu từ RGB sang HSV
- Bước 2: Chuyển đổi ảnh sang ảnh trắng đen bằng cách chỉ lấy chanel màu S của hệ màu HSV
- Bước 3: Tiến hành khử nhiễu với opening và closing trong Opencv, đây là các kĩ thuật trong Morphological Transformations nhằm xử lí nhiễu đối với ảnh trắng đen
- Bước 4: Tìm vùng ROI tương ứng bằng OpenCV, kết quả đầu ra là thông tin của bouding box bao quanh đối tượng
- Bước 5: Tiến hành cắt ảnh theo bouding box và lưu lại ảnh mới



- Cách thứ hai thực hiện bằng phương pháp thủ công, đó là cắt tay các hình ảnh này.

Tuy cách thứ nhất đem lại hiệu quả về mặt thời gian tuy nhiên. Nó bị fail ở một số ảnh vì bị ảnh hưởng bởi ánh sáng.



Hàm này sẽ tiếp tục được cải thiện cập nhật và sau. Trong đồ án này chúng tôi sử dụng cách 2 để thực hiện nhằm đảm bảo bộ dữ liệu không bị hỏng. Đây là cách khá tốn thời gian tuy nhiên nó an toàn.

3- Rút trích đặc trưng :

Dữ liệu sau khi được crop xong sẽ tiến hành rút trích đặc trưng. Trong đó tôi chọn 2 hướng rút trích đặc trưng cho bộ dữ liệu:

- Cách 1 Kết hợp đặc trưng color, shape, texture với nhau.
- Cách 2: Chỉ dùng đặc trưng HOG.

Cách 1

- Shape: dùng đặc trưng Hu Moments [2] là một tập hợp 7 số được tính toán bằng cách sử dụng các central moments bất biến đối với các phép biến đổi hình ảnh. 6 moments đầu tiên đã được chứng minh là bất biến đối với phép dịch, tỷ lệ và xoay và phản xạ. Trong khi dấu hiệu của moments thứ 7 thay đổi để phản chiếu hình ảnh. Tác dụng của đặc trưng này nhằm lấy đặc trưng về hình dáng của trái cây trong bức ảnh
- Color Histogram: đây là đặc trưng dựa trên sự phân bố màu sắc có trong hình ảnh.
- Texture: sử dụng đặc trưng Haralick Texture [3] texture là đặc trưng mô tả sự thay đổi mức cường độ xám giữa các vùng trong hình ảnh. Haralick texture sử dụng gray level co-occurrence matrices (GLCM). Ma trận này dựa trên sự phân bố xác suất giữa 2 pixel trong ảnh

Cách đặc trưng này sẽ được kết hợp về một đặc trưng duy nhất để đưa vào mô hình huấn luyện.

Cách 2 :

Dùng đặc trưng HOG [4]: với đặc trưng HOG trích xuất sử dụng thông tin về sự phân bố của các cường độ gradient (intensity gradient) hoặc của hướng biên (edge directions) để mô tả các đối tượng cục bộ trong ảnh. Để thực hiện việc trích xuất đặc trưng HOG đầu tiên chia bức ảnh thành các cell và tính toán một histogram về các hướng của gradients cho các điểm nằm trong cell. Ghép các histogram lại với nhau ta sẽ có một biểu diễn cho bức ảnh ban đầu. Để tăng cường hiệu năng nhận dạng, các histogram cục bộ có thể được chuẩn hóa về độ tương phản bằng cách tính một ngưỡng cường độ trong một vùng lớn hơn cell, gọi là các khối (blocks) và sử dụng giá trị ngưỡng đó để chuẩn hóa tất cả các cell trong khối. Kết quả sau bước chuẩn hóa sẽ là một vector đặc trưng có tính bất biến cao hơn đối với các thay đổi về điều kiện ánh sáng.

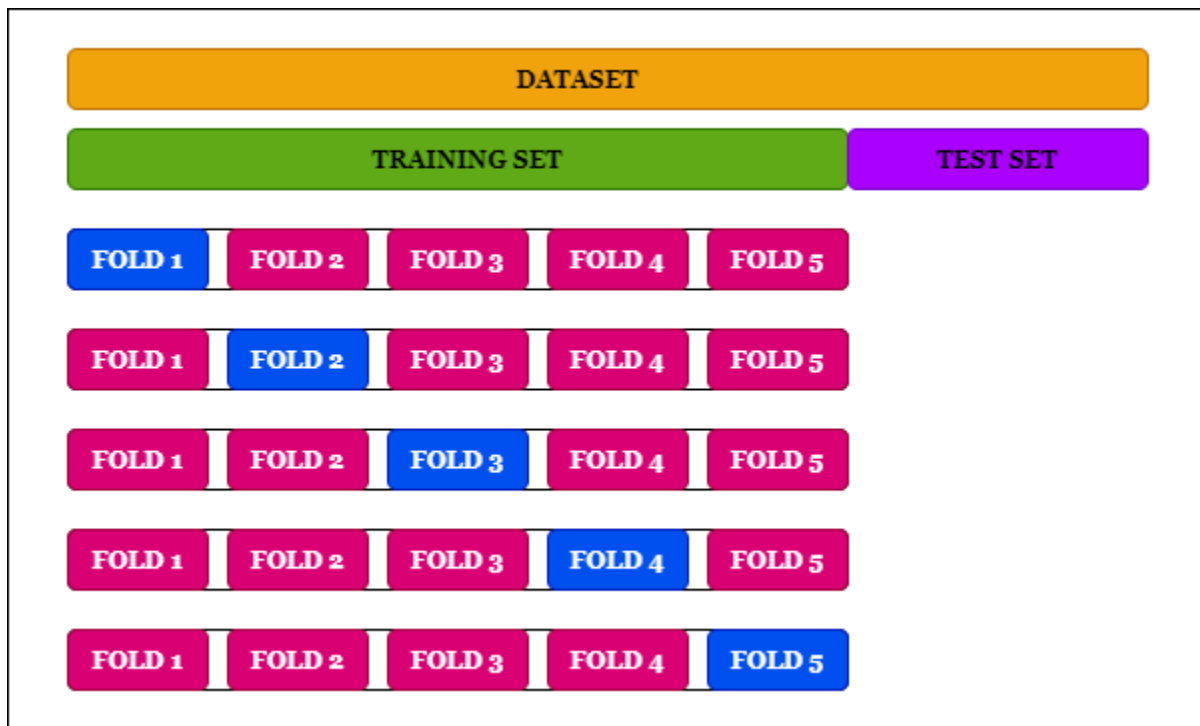


IV-Huấn luyện và đánh giá mô hình:

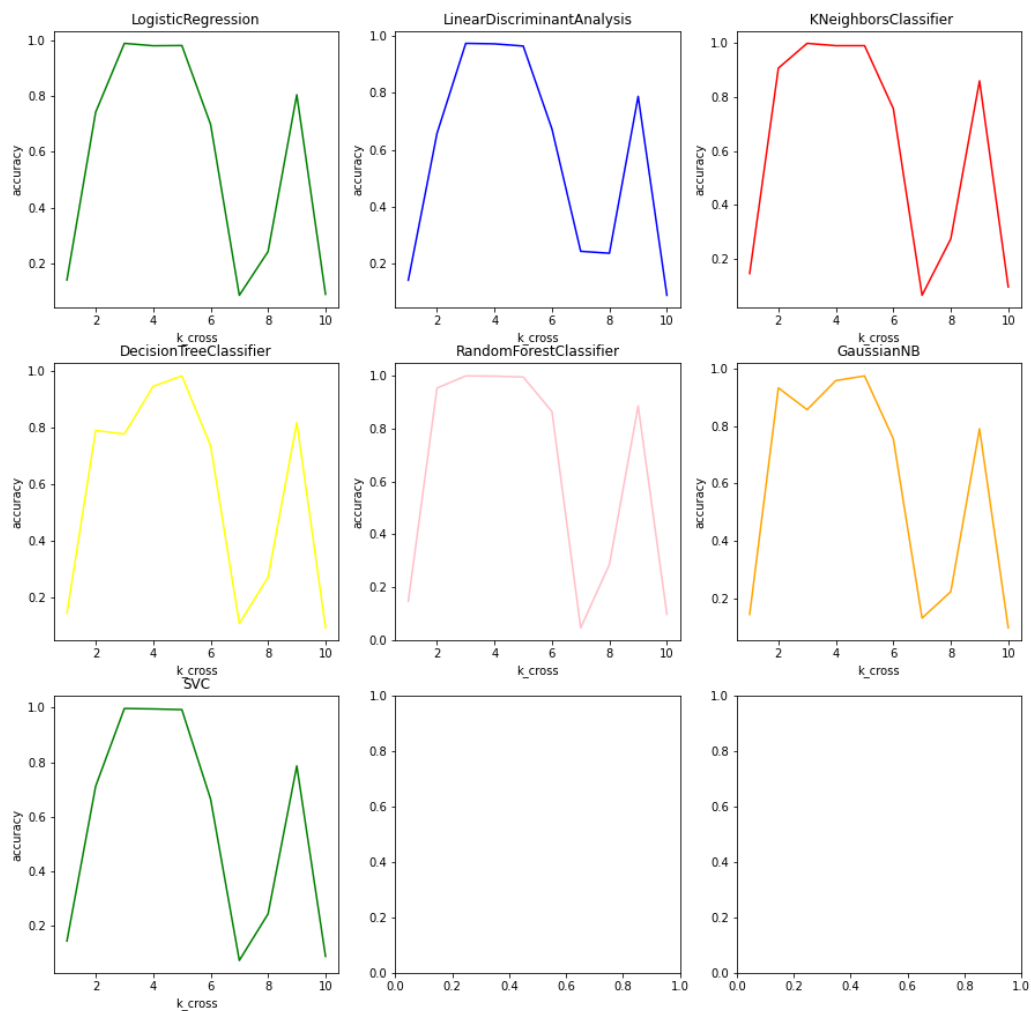
Các mô hình sử dụng để thử nghiệm gồm có 7 mô hình thường được dùng trong bài toán phân lớp bao gồm:

- + Logistic Regression
- + Linear Discriminant Analysis
- + KNN
- + Decision Tree Classifier
- + Random forest
- + GaussianNB
- + SVM

Các mô hình trên đã được xây dựng sẵn trên thư viện sklearn. Nên tôi không đi xây dựng lại mà áp dụng các mô hình có sẵn từ sklearn. Sau khi áp dụng các mô hình tôi sẽ đánh giá mô hình này bằng phương pháp **cross k-fold**. **K-fold** sẽ chia bộ dữ liệu thành k tập con từ bộ train data. Các mô hình sẽ được huấn luyện trên k-1 tập và 1 tập để đánh giá. Ta sẽ thực hiện việc đánh giá này với k lần và lấy kết quả trung bình cho k lần thực hiện.



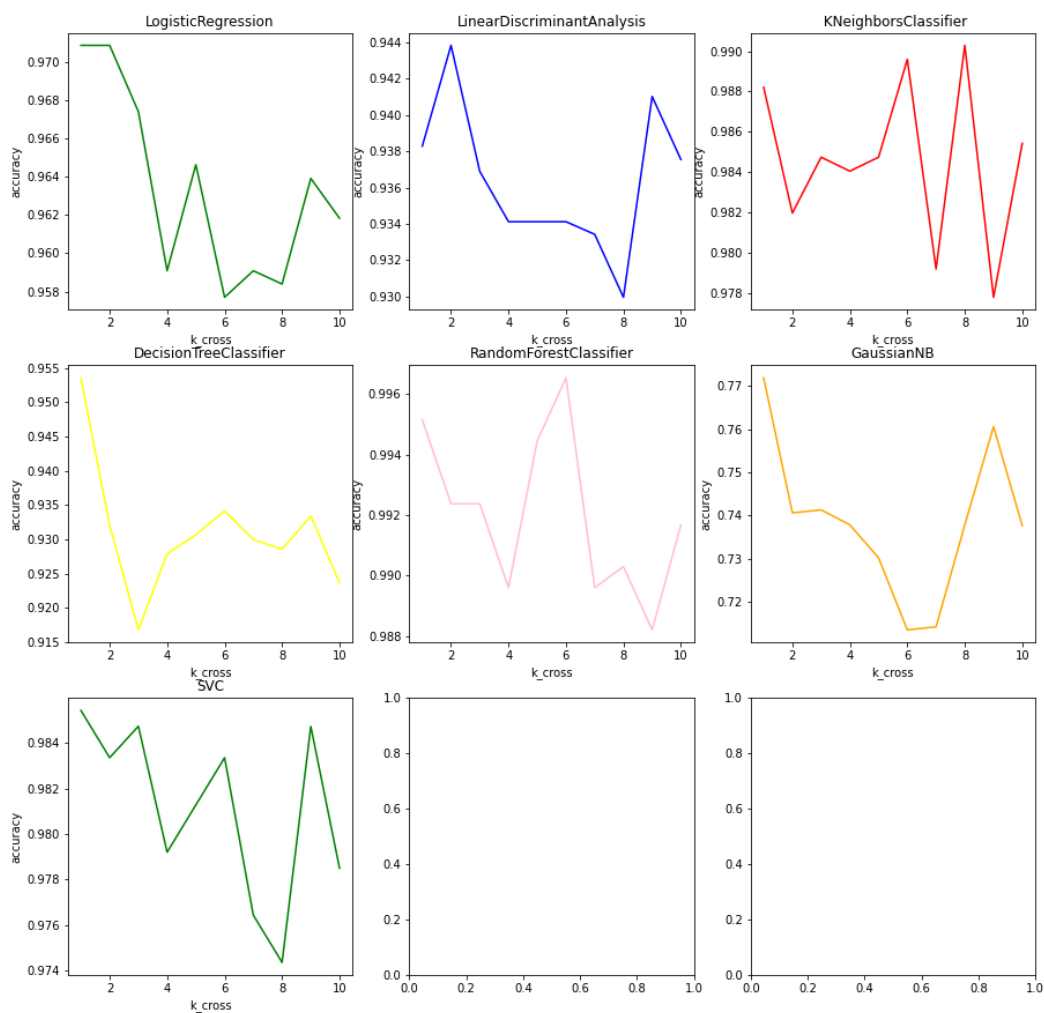
Ban đầu, tôi áp dụng phương pháp này với cách chia dữ liệu 0.75 – 0.25. Tập dữ liệu kiểm thử chiếm 25%, chỉ chứa các ảnh chụp trực tiếp từ điện thoại, đây là ảnh có độ phân giải cao. Tập dữ liệu huấn luyện chiếm 75%, tuy nhiên tập dữ liệu này đa số là các ảnh được tách frame từ video, nên chất lượng của ảnh thấp, và có nhiều hình ảnh tương đồng nhau. Dẫn tới kết quả đánh giá cho ra không cao.



Acuracy của các model đều giao động mạnh từ 0.2-0.9. Đây là một kết quả không khả quan.

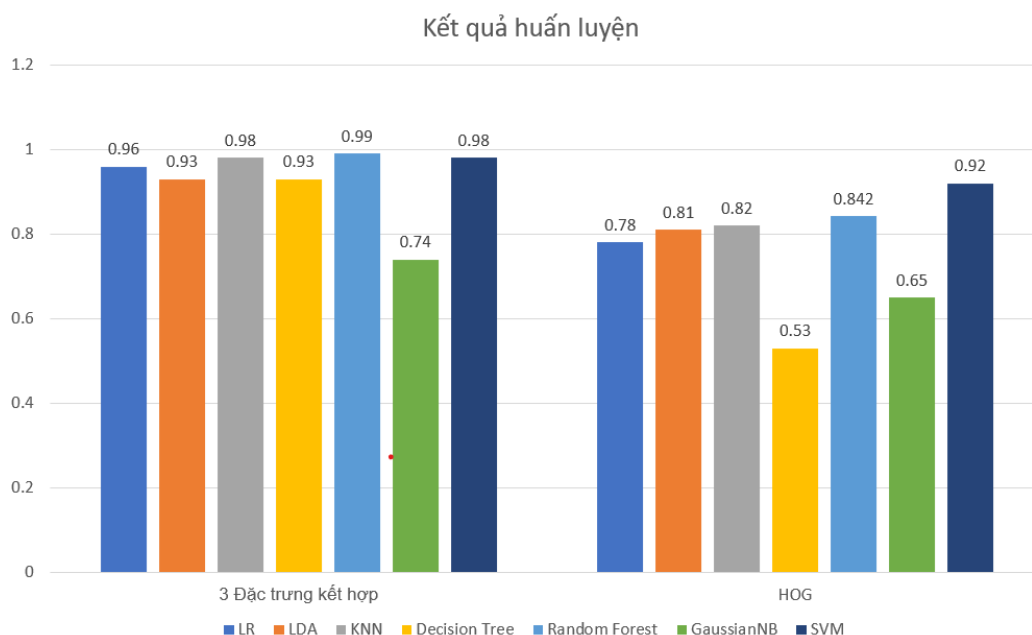
Sau đó tôi đã cải thiện bằng cách trộn 2 bộ dữ liệu này vào chia lại với tỉ lệ huấn luyện và kiểm thử là: 0.9-0.1

Kết quả cho ra đã được cải thiện rất nhiều



Với cách làm này accuracy dao động trong khoảng từ 0.7-0.99. Kết quả có được là nhờ tập huấn luyện được bổ sung thêm các ảnh chất lượng từ tập kiểm tra nhờ đó mô hình được cải thiện tốt hơn. Việc này cho thấy được các yếu tố về **độ đa dạng góc chụp, đối tượng, cũng như chất lượng hình ảnh tác động** nhiều tới hiệu quả của các mô hình máy học. Đây là một điều nên lưu ý khi thực hiện xây dựng, phát triển các mô hình, thuật toán máy học.

Không dừng lại ở việc chỉ sử dụng 3 đặc trưng kết hợp. Như nêu ở trên tôi cũng thử các mô hình với việc sử dụng đặc trưng HOG. Kết quả được thể hiện ở biểu đồ bên dưới.



Kết quả trên cho thấy sự dụng đặt trưng bằng dùng 3 đặc trưng kết hợp vẫn tốt hơn so với việc chỉ sử dụng một đặc trưng là HOG. Việc kết hợp 3 đặc trưng sẽ giúp các mô hình được học nhiều hơn và từ đó cho ra kết quả dự đoán chính xác cao hơn

V- Tổng kết:

1 Kết quả đạt được:

Qua việc huấn luyện và đánh giá trên ta thấy mô hình Random Forest có hiệu quả cao nhất với việc sử dụng 3 đặc trưng kết hợp là Hu moment, Haralick Texture, Color Histogram. Mô hình sau khi được huấn luyện được tôi triển khai trên web local để thuận tiện cho việc upload hình ảnh. Đây là link video demo. (<https://www.youtube.com/watch?v=WAC08tQlxQ>)

2 Kinh nghiệm đạt được:

Việc resize ảnh tác động nhiều tới quá trình huấn luyện cho các mô hình. Resize nhỏ thuật toán chạy nhanh, tuy nhiên nếu quá nhỏ các đặc trưng của ảnh sẽ bị mất đi. Còn nếu quá lớn thì làm thuật toán chúng ta chạy chậm, tốn nhiều không gian lưu trữ. Thường resize kích thước về dạng 2^n để tiện cho việc lưu trữ của bộ nhớ. Việc resize các bức ảnh góp phần

tránh được lỗi mounting drive time out khi sử dụng Colab để huấn luyện cho mô hình. Cách chúng ta chọn các thuật toán cũng ảnh hưởng nhiều tới kết quả của model chẳng hạn với GaussianNB thì cho kết quả thấp hơn những model còn lại. Lí do vì model thường hiệu quả trên bài toán có giá trị đầu ra liên tục. Còn các model khác thì đều rất tốt cho bài toán phân lớp.

Bộ dataset cho việc huấn luyện được tách frame từ các video tự thu thập được nên ảnh trùng lặp khá nhiều dẫn tới các mô hình không học được nhiều đặc trưng làm hiệu quả giảm như đã nêu ở trên. Dữ liệu còn thiếu tính đa dạng nên việc dự đoán các ảnh **hoàn toàn mới** không nằm trong bộ dữ liệu thu thập **còn fail rất nhiều**. Hướng cải thiện là tăng cường dữ liệu về góc chụp, số lượng trái, ánh sáng, nhiễu,.....Đồng thời thử nghiệm với kĩ thuật Deeplearning.

Quá trình thực hiện đồ án tốn thời gian nhất chính là lúc thu thập data. Chúng tôi phải tự đi thu thập, phân loại, xử lí và tổ chức lưu trữ. Qua đây tôi đã thực sự trải nghiệm được phần nào quy trình làm một bài toán máy học. Biết được những khó khăn và thách thức khi đi giải quyết bài toán thực tế.

VI- Tài liệu tham khảo

- [1] R. S. S. N. L. S.Arivazhagan, "Fruit Recognition using Color and Texture Features," CIS, 2010.
- [2] M.-K. HU, "Visual Pattern Recognition by Moment Invariants", " IRE , 1962.
- [3] K. S. A. I. D. ROBERT M. HARALICK, "Textural Features for Image Classification," IEEE, 1973.
- [4] N. D. a. B. Triggs, "Histograms of Oriented Gradients for Human Detection," IEEE, 2005.
- [5] <https://www.kaggle.com/moltean/fruits>