

ỦY BAN NHÂN DÂN TP. HỒ CHÍ MINH

**TRƯỜNG ĐẠI HỌC SÀI GÒN**

**---------------o0o---------------**

**ĐỀ TÀI : NHẬN DIỆN CÁC LOẠI TRÁI CÂY BẰNG HÌNH ẢNH DÙNG PHƯƠNG PHÁP DEEP LEARNING**

**MÔN HỌC : PHƯƠNG PHÁP NGHIÊN CỨU KHOA HỌC**

**GIẢNG VIÊN HƯỚNG DẪN :** TS. ĐỖ NHƯ TÀI

**NHÓM SINH VIÊN THỰC HIỆN :**

1. TRẦN HOÀNG KHANH\_3121410257
2. DƯƠNG DUY KHANG\_3121410250
3. NGUYỄN NGỌC HẢI\_3121410175

**Thành phố Hồ Chí Minh,** **Tháng 5/2025**

**MỤC LỤC**

[LỜI MỞ ĐẦU 6](#_Toc198686245)

[Chương 1 : Tổng quan vấn đề 7](#_Toc198686246)

[1.1. Đặt vấn đề 7](#_Toc198686247)

[1.2. Lý do chọn đề tài 8](#_Toc198686248)

[1.3. Mục đích nghiên cứu 8](#_Toc198686249)

[1.4. Câu hỏi nghiên cứu 9](#_Toc198686250)

[1.5. Phạm vi nghiên cứu 9](#_Toc198686251)

[Chương 2 : Lược khảo tài liệu (Literature Review) 11](#_Toc198686252)

[2.1. Tổng hợp các tài liệu, nghiên cứu trước liên quan. 11](#_Toc198686253)

[2.2. Tập dữ liệu Fruits-360. 11](#_Toc198686254)

[2.3. Cơ sở lý thuyết cho nghiên cứu 12](#_Toc198686255)

[2.3.1. Thị giác máy tính (Computer Vision) 12](#_Toc198686256)

[2.3.2. Mạng nơ-ron tích chập (CNN) 12](#_Toc198686257)

[2.2.3. Thư viện TensorFlow 14](#_Toc198686258)

[Chương 3: Phương pháp nghiên cứu (Methodology) 16](#_Toc198686259)

[3.1. Cấu trúc của mạng nơ-ron được sử dụng trong các thí nghiệm 16](#_Toc198686260)

[3.2 Đối tượng và mẫu nghiên cứu 17](#_Toc198686261)

[3.3. Cách thu thập dữ liệu 23](#_Toc198686262)

[3.4. Phân tích khám phá dữ liệu 24](#_Toc198686263)

[3.5. Tiền xử lý dữ liệu. 25](#_Toc198686264)

[Chương 4. Thực nghiệm và Thảo luận 26](#_Toc198686265)

[4.1. Thí nghiệm số 26](#_Toc198686266)

[4.2. Huấn luyện trên các mô hình khác. 27](#_Toc198686267)

[Chương 5. Kết luận 28](#_Toc198686268)

[5.1. Tóm tắt kết quả đạt được và hướng nghiên cứu tương lai. 28](#_Toc198686269)

[5.2. Câu trả lời cho câu hỏi nghiên cứu 31](#_Toc198686270)

[TÀI LIỆU THAM KHẢO 32](#_Toc198686271)

**DANH MỤC BẢNG BIỂU, HÌNH ẢNH**

[Hình 1.Quy trình thông thường của quy trình sơ chế và phân loại hoa quả ở nước ta 6](#_Toc198686053)

[Hình 2.Bảng tổng hợp các nghiên cứu liên quan 10](#_Toc198686054)

[Hình 3. Bảng cấu trúc của mạng nơ-ron được sử dụng trong bài báo này. 16](#_Toc198686055)

[Hình 4. Biểu diễn đồ họa của mạng nơ-ron tích chập được sử dụng trong các thí nghiệm. 16](#_Toc198686056)

[Hình 5. Số lượng hình ảnh cho mỗi loại trái cây 22](#_Toc198686057)

[Hình 6. Bên trái: hình ảnh gốc. Bên phải: quả sau khi xóa phần nền và sau khi được thu nhỏ xuống 100x100 pixel. 23](#_Toc198686058)

[Hình 7. Tổng quan dữ liệu 23](#_Toc198686059)

[Hình 8. Số lượng ảnh của từng nhãn trái cây 24](#_Toc198686060)

[Hình 9. So sánh Accuracy của mô hình 26](#_Toc198686061)

[Hình 10.Biểu đồ Accuracy 27](#_Toc198686062)

[Hình 11.Biểu đồ Loss 28](#_Toc198686063)

[Hình 12. Ma trận nhầm lẫn 29](#_Toc198686064)

**DANH MỤC CÁC CHỮ VIẾT TẮT**

|  |  |
| --- | --- |
| **AI** | Trí tuệ nhân tạo |
| **CNN** | Convolutional Neural Networks/ Mạng nơ-ron tích chập |
| **SVM** | Support Vector Machine |
| **CPU** | Central Processing Unit/ Bộ xử lý trung tâm |
| **GPU** | Graphics Processing Unit/ Bộ xử lý đồ họa |

# LỜI MỞ ĐẦU

Trong bối cảnh tự động hóa ngày càng đóng vai trò quan trọng trong nông nghiệp và chuỗi cung ứng thực phẩm, nhu cầu xây dựng hệ thống nhận diện và phân loại trái cây thông qua hình ảnh ngày càng trở nên cấp thiết. Việc phân loại thủ công không chỉ tiêu tốn nhiều thời gian mà còn dễ xảy ra sai sót, trong khi công nghệ học sâu (deep learning), đặc biệt là mạng nơ-ron tích chập (CNN), đang cho thấy tiềm năng vượt trội trong xử lý và phân tích hình ảnh.

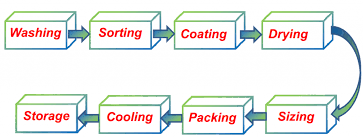
Nghiên cứu này hướng đến mục tiêu phát triển một mô hình học sâu có khả năng tự động nhận diện và phân loại chính xác các loại trái cây dựa trên hình ảnh đầu vào. Bên cạnh việc xây dựng kiến trúc mô hình, đề tài còn tập trung tối ưu thuật toán xử lý ảnh, đánh giá hiệu suất giữa các phương pháp như CNN, SVM, và Random Forest để tìm ra giải pháp phù hợp nhất. Dữ liệu đầu vào là hình ảnh trái cây trên nền đơn giản, đầu ra là tên loại trái cây tương ứng.

Kết quả cho thấy phương pháp học sâu không chỉ nâng cao độ chính xác và tốc độ nhận diện mà còn góp phần tối ưu hóa quy trình phân loại sản phẩm, từ đó mở ra hướng ứng dụng hiệu quả AI trong lĩnh vực nông nghiệp. Nghiên cứu tập trung chuyên sâu vào bài toán nhận diện trái cây nhằm đảm bảo hiệu quả và độ tin cậy cao nhất cho hệ thống được xây dựng.

# Chương 1 : Tổng quan vấn đề

## 1.1. Đặt vấn đề

Tại Việt Nam, nông nghiệp – đặc biệt là ngành trồng trọt và chế biến trái cây – đóng vai trò quan trọng trong phát triển kinh tế, góp phần lớn vào kim ngạch xuất khẩu. Tuy nhiên, một trong những khâu then chốt nhưng vẫn còn nhiều bất cập là việc phân loại trái cây sau thu hoạch. Phần lớn quy trình phân loại hiện nay vẫn phụ thuộc vào lao động thủ công, sử dụng mắt thường để đánh giá và phân biệt các loại trái cây dựa trên hình dáng, màu sắc hoặc kích thước. Phương pháp này không chỉ thiếu tính chính xác và nhất quán, mà còn tiêu tốn thời gian, nhân lực và rất dễ bị ảnh hưởng bởi yếu tố chủ quan của người thực hiện.



Hình 1.Quy trình thông thường của quy trình sơ chế và phân loại hoa quả ở nước ta

Hình trên mô tả quy trình xử lý sau thu hoạch của trái cây bao gồm các bước liên tiếp nhằm đảm bảo chất lượng và kéo dài thời gian bảo quản. Đầu tiên, trái cây được rửa sạch (Washing) để loại bỏ bụi bẩn và tạp chất, sau đó được phân loại (Sorting) theo kích thước, màu sắc hoặc chất lượng. Tiếp theo là bước phủ lớp bảo vệ (Coating) như sáp hoặc chất bảo quản để giữ độ tươi, rồi được làm khô (Drying) để tránh ẩm mốc. Sau khi làm khô, trái cây tiếp tục được phân cỡ (Sizing), đóng gói (Packing) và làm lạnh (Cooling) nhằm duy trì độ tươi và ổn định chất lượng. Cuối cùng, sản phẩm được đưa vào bảo quản (Storage) trong kho lạnh cho đến khi phân phối ra thị trường.

Trong khi đó, tại nhiều quốc gia có nền nông nghiệp phát triển như Mỹ, Nhật Bản hay Hà Lan, quá trình phân loại trái cây đã được tự động hóa cao, áp dụng các hệ thống nhận diện hình ảnh kết hợp với công nghệ học sâu (deep learning) để đạt được hiệu quả tối ưu. Những hệ thống này không chỉ giúp tăng tốc độ xử lý hàng loạt, mà còn đảm bảo độ chính xác cao, giảm thiểu tỉ lệ sai sót trong quá trình phân loại – từ đó nâng cao chất lượng sản phẩm và khả năng cạnh tranh trên thị trường quốc tế.

So với các hệ thống hiện đại đó, Việt Nam đang gặp phải nhiều thách thức: thiếu công nghệ phù hợp, chi phí đầu tư cao, và quan trọng nhất là chưa có mô hình nhận diện phù hợp với đặc thù trái cây nội địa. Đặc điểm của trái cây Việt Nam là đa dạng về chủng loại, hình dạng, màu sắc và thường không đồng nhất về điều kiện chụp ảnh (ánh sáng, phông nền, góc độ...) – điều này khiến việc áp dụng các mô hình chung từ nước ngoài gặp nhiều khó khăn.

## 1.2. Lý do chọn đề tài

Vấn đề đặt ra là làm sao để xây dựng được một hệ thống nhận diện và phân loại trái cây tự động, có thể hoạt động hiệu quả với dữ liệu hình ảnh thực tế tại Việt Nam, vừa đảm bảo tính chính xác cao, vừa có khả năng mở rộng ứng dụng trong các dây chuyền sản xuất, phân phối và xuất khẩu nông sản.

Trước yêu cầu đó, một giải pháp tiềm năng được đề xuất là ứng dụng học sâu, đặc biệt là các mô hình mạng nơ-ron tích chập (Convolutional Neural Networks – CNN) trong xử lý ảnh. Các mô hình này đã được chứng minh hiệu quả vượt trội trong nhiều bài toán phân loại hình ảnh phức tạp. Khi được huấn luyện đúng cách với dữ liệu phù hợp, mạng CNN có thể nhận diện đặc trưng hình ảnh của từng loại trái cây một cách chính xác, từ đó hỗ trợ phân loại nhanh chóng và ổn định.

Dựa trên những cơ sở lý thuyết và thực tiễn nói trên, nghiên cứu này được thực hiện nhằm xây dựng một hệ thống nhận diện và phân loại trái cây thông qua hình ảnh, ứng dụng phương pháp học sâu, với mục tiêu phù hợp với điều kiện thực tế tại Việt Nam và góp phần từng bước hiện đại hóa quy trình kiểm định, phân loại nông sản trong nước.

## 1.3. Mục đích nghiên cứu

Nghiên cứu này được thực hiện với mục đích chính là **xây dựng một hệ thống tự động nhận diện và phân loại trái cây dựa trên hình ảnh**, nhằm thay thế quy trình thủ công còn nhiều hạn chế hiện nay. Việc phát triển hệ thống này không chỉ giúp tăng độ chính xác và tốc độ xử lý trong khâu phân loại, mà còn góp phần **hiện đại hóa quy trình sản xuất, đóng gói và kiểm định chất lượng nông sản tại Việt Nam**.

Cụ thể, các mục tiêu hướng đến trong nghiên cứu gồm:

* **Ứng dụng phương pháp học sâu**, đặc biệt là mạng nơ-ron tích chập (CNN), để phát triển mô hình có khả năng nhận diện và phân loại các loại trái cây phổ biến dựa trên hình ảnh đầu vào.
* **Xây dựng quy trình huấn luyện mô hình** dựa trên tập dữ liệu hình ảnh thực tế, đảm bảo mô hình học được đặc trưng quan trọng của từng loại trái cây trong điều kiện chụp ảnh đa dạng.
* **Đánh giá hiệu suất mô hình** thông qua các chỉ số đo lường như độ chính xác, độ nhạy và tốc độ xử lý, đồng thời so sánh với các mô hình học sâu khác để chọn ra phương pháp tối ưu.

## 1.4. Câu hỏi nghiên cứu

Câu hỏi nghiên cứu tập trung vào nhận diện trái cây:

1. Mô hình học sâu nào hiệu quả nhất để nhận diện trái cây dựa trên tập dữ liệu Fruits-360?
2. Kỹ thuật tiền xử lý hình ảnh nào giúp tối ưu hóa mô hình nhận diện trái cây?
3. Độ chính xác của CNN khi nhận diện từng loại trái cây đạt mức nào, và có cần cải tiến thêm không?
4. Yếu tố nào ảnh hưởng đến khả năng nhận diện, ví dụ như kích thước, hình dạng, ánh sáng và nền ảnh?

Làm thế nào để đảm bảo mô hình có thể nhận diện chính xác các loại trái cây khác nhau trong điều kiện thực tế?

## 1.5. Phạm vi nghiên cứu

Phạm vi của nghiên cứu này tập trung vào việc nhận diện và phân loại hình ảnh các loại trái cây phổ biến tại Việt Nam thông qua phương pháp học sâu, cụ thể là sử dụng mạng nơ-ron tích chập (CNN). Đề tài không hướng đến tất cả các loại nông sản, mà chỉ chọn một **tập hợp các loại trái cây có hình dáng đặc trưng và thường gặp trong thực tiễn sản xuất, phân phối và tiêu dùng**.

**Đối tượng nghiên cứu**: Các thuật toán Machine Learning, đặc biệt là mạng nơ-ron tích chập (CNN), được sử dụng để phân loại trái cây dựa trên hình ảnh trong tập dữ liệu Fruits-360.

**Dữ liệu đầu vào**: Ảnh tĩnh, định dạng số, thu thập từ nguồn dữ liệu có sẵn (các bộ dataset công khai) hoặc tự tạo, đảm bảo tính đa dạng về góc chụp và kích thước nhưng vẫn kiểm soát được yếu tố nền để tập trung vào đặc trưng trái cây.

**Kỹ thuật áp dụng**: Chỉ giới hạn trong các phương pháp học sâu xử lý ảnh, trong đó tập trung chính vào mạng nơ-ron tích chập (CNN)

**Phạm vi ứng dụng**: Mô hình được thiết kế và kiểm thử trong môi trường mô phỏng hoặc trên tập dữ liệu chuẩn hóa.

Nghiên cứu không mở rộng sang các bài toán nhận diện phức tạp hơn như phân biệt trái cây cùng loại nhưng khác chất lượng (ví dụ: chín - xanh, hỏng - tươi), hay xử lý ảnh trong điều kiện môi trường không kiểm soát (nhiễu, nhiều vật thể, phông nền phức tạp). Điều này nhằm đảm bảo mô hình có thể hoạt động tốt và chính xác trong một phạm vi cụ thể, làm nền tảng cho các bước nghiên cứu chuyên sâu hơn trong tương lai.

# Chương 2 : Lược khảo tài liệu (Literature Review)

## 2.1. Tổng hợp các tài liệu, nghiên cứu trước liên quan.

|  |  |  |  |
| --- | --- | --- | --- |
| **Năm** | **Tác giả** | **Tên nghiên cứu** | **Nguồn công bố / Hội nghị** |
| 2018 | Horea Mureșan, Mihai Oltean | Fruit recognition from images using deep learning | Acta Univ. Sapientiae, Informatica, Vol. 10(1), pp. 26-42 |
| 2018 | Ma Linjuan, Fuquan Zhang, Lin Xu | Fruit Detection Using Faster R-CNN Based | Advances in Smart Vehicular Tech., Springer, Vol. 128 |
| 2018 | A. Kausar et al. | Pure-CNN: A Framework for Fruit Images Classification | IEEE CSCI, Las Vegas, USA, pp. 404-408 |
| 2018 | Mustaffa M. R. et al. | DURIAN RECOGNITION BASED ON MULTIPLE FEATURES... | Malaysian Journal of Computer Science, pp. 57–72 |
| 2019 | D.T.P. Chung, D.V. Tai | A fruits recognition system... deep learning technique | J. Phys.: Conf. Series, Vol. 1327 |
| 2022 | Horea Mureșan | An Automated Algorithm for Fruit Image Dataset Building | FedCSIS 2022, Sofia, pp. 103-107 |
| 2022 | Aherwadi N. et al. | Prediction of Fruit Maturity, Quality... | Electronics, Vol. 11(24), 4100 |
| 2024 | Yonis Gulzar et al. | Adaptability of deep learning: datasets and strategies in fruit classification | BIO Web Conf., Vol. 85, 01020 |
| 2024 | Xu Qinghui et al. | Research on fruit image processing... based on ResNet50 | Scientific Innovation in Asia, Vol. 2(1), pp. 1-8 |
| 2025 | Sathyadhas G.N. et al. | LwSANet: Lightweight Self-Attention Network for fruit recognition | Traitement du Signal, Vol. 42(1), pp. 183-200 |
| 2025 | Panhwar Ali Orangzeb et al. | A Scheme Based on Deep Learning for Fruit Classification | Mehran Univ. Res. J. Eng. & Tech., Vol. 44(1), pp. 8–19 |
| 2025 | Mansouri Merieme et al. | Estimating glycemic index in... Moroccan cuisine dataset | Journal of Intelligent Systems, Vol. 34(1), 20240122 |

Hình 2.Bảng tổng hợp các nghiên cứu liên quan

## 2.2. Tập dữ liệu Fruits-360.

Bộ dữ liệu Fruits-360 là bộ dữ liệu công khai và được chia sẻ rộng rãi cho mục đính nghiên cứu và học hỏi là chính, phiên bản mới nhất 2025.03.24.0 do tác giả là Mihai Oltean thực hiện. Thành phần: Hình ảnh của hơn 100 loại trái cây, rau củ, hạt và hạt giống khác nhau, bao gồm nhiều giống như táo (Crimson Snow, Golden, Granny Smith, Pink Lady, Red Delicious), chuối (vàng, đỏ, Lady Finger), nho (xanh, hồng, trắng), cà chua (nhiều loại), dưa hấu, dâu tây, cà rốt, bắp cải, hành tây, khoai tây, chanh, cam, ớt, v.v.

## 2.3. Cơ sở lý thuyết cho nghiên cứu

### 2.3.1. Thị giác máy tính (Computer Vision)

Thị giác máy tính giúp máy có khả năng hiểu và phân tích hình ảnh, đặc biệt trong bài toán nhận diện trái cây. **Xử lý ảnh số**: Kỹ thuật tiền xử lý giúp chuẩn hóa hình ảnh, cân bằng ánh sáng, tăng cường dữ liệu để cải thiện hiệu suất mô hình. **Trích xuất đặc trưng**: Xác định các yếu tố đặc trưng như màu sắc, hình dạng, kết cấu để mô hình dễ dàng phân loại trái cây.

### 2.3.2. Mạng nơ-ron tích chập (CNN)

Mạng nơ-ron tích chập (CNN) là một phần của các mô hình học sâu. Một mạng như vậy có thể bao gồm các lớp tích chập, lớp gộp, các lớp ReLU, lớp kết nối đầy đủ và lớp mất mát . Trong một kiến ​​trúc CNN điển hình, mỗi lớp tích chập được theo sau bởi một lớp Đơn vị tuyến tính chỉnh lưu (ReLU), sau đó là một lớp gộp, sau đó là một hoặc nhiều lớp tích chập và cuối cùng là một hoặc nhiều lớp kết nối đầy đủ. Một đặc điểm phân biệt CNN với mạng nơ-ron thông thường là tính đến cấu trúccủa hình ảnh trong khi xử lý chúng. Phần tiếp theo xin được mô tả chi tiết mỗi lớp trong mạng CNN.

#### 2.3.2.1. Lớp Convolutional\_lớp tích chập

Các lớp tích chập được đặt tên theo phép toán tích chập. Trong toán học, tích chập là phép toán trên hai hàm tạo ra một hàm thứ ba là phiên bản đã sửa đổi (tích chập) của một trong các hàm gốc. Hàm kết quả đưa ra tích phân của phép nhân từng điểm của hai hàm dưới dạng hàm của lượng mà một trong các hàm gốc được dịch chuyển.

Một lớp tích chập bao gồm các nhóm nơ-ron tạo nên các hạt nhân (kernels). Các hạt nhân này có kích thước nhỏ nhưng luôn có cùng độ sâu với đầu vào. Mỗi nơ-ron trong hạt nhân được kết nối với một vùng nhỏ của đầu vào, gọi là trường tiếp nhận (receptive field). Việc này được thực hiện vì sẽ rất không hiệu quả nếu liên kết tất cả các nơ-ron với toàn bộ đầu vào, đặc biệt khi đầu vào có kích thước lớn như hình ảnh. Ví dụ, một hình ảnh có kích thước 100 x 100 sẽ có 10.000 pixel, và nếu lớp đầu tiên có 100 nơ-ron thì sẽ cần đến 1.000.000 tham số – điều này là không thực tế. Thay vì để mỗi nơ-ron có trọng số cho toàn bộ chiều của đầu vào, mỗi nơ-ron trong lớp tích chập chỉ giữ trọng số tương ứng với kích thước của hạt nhân. Các hạt nhân này sẽ trượt qua chiều rộng và chiều cao của đầu vào, trích xuất các đặc trưng cấp cao và tạo ra một bản đồ kích hoạt hai chiều.

Bước tiến mà một hạt nhân trượt được đưa ra như một tham số. Đầu ra của một lớp tích chập được tạo ra bằng cách xếp chồng các bản đồ kích hoạt kết quả, sau đó được sử dụng để xác định đầu vào của lớp tiếp theo.

Áp dụng một lớp tích chập lên hình ảnh có kích thước 32 x 32 sẽ tạo ra một bản đồ kích hoạt có kích thước 28 x 28. Nếu tiếp tục áp dụng nhiều lớp tích chập hơn, kích thước sẽ tiếp tục giảm, dẫn đến việc hình ảnh bị thu nhỏ đáng kể, gây ra hiện tượng mất thông tin và vấn đề độ dốc biến mất (vanishing gradient). Để khắc phục tình trạng này, chúng ta sử dụng đệm (padding). Đệm giúp tăng kích thước của đầu vào bằng cách điền thêm các hằng số xung quanh dữ liệu đầu vào. Trong hầu hết các trường hợp, hằng số này là 0, vì vậy thao tác này được gọi là đệm bằng không (zero padding).

Có hai loại đệm phổ biến:

* Đệm "Giống nhau" (Same padding): nghĩa là bản đồ đặc trưng đầu ra có cùng kích thước không gian với đầu vào. Thao tác này cố gắng đệm đều sang trái và phải, nhưng nếu số cột cần thêm là số lẻ, hệ thống sẽ thêm một cột vào phía bên phải.
* Đệm "Hợp lệ" (Valid padding): tương đương với không sử dụng đệm nào cả, dẫn đến việc giảm kích thước sau mỗi lớp tích chập.

Bước tiến (stride) là tham số điều khiển khoảng cách mà hạt nhân trượt trên ảnh đầu vào. Khi bước tiến > 1, một số pixel trong hình ảnh bị bỏ qua và không được tính vào đầu ra, giúp giảm kích thước kết quả và tăng tốc quá trình tính toán. Bước tiến xác định cách hoạt động của phép toán tích chập khi xử lý hình ảnh lớn hơn hoặc sử dụng hạt nhân phức tạp hơn.

Cuối cùng, lớp ReLU (Rectified Linear Unit) – hay còn gọi là lớp Đơn vị tuyến tính chỉnh lưu – áp dụng hàm kích hoạt max(0, x) cho mỗi giá trị đầu vào. Lớp này không làm thay đổi kích thước của dữ liệu, nhưng nó giúp tăng tính phi tuyến của mạng, từ đó cải thiện khả năng học các đặc trưng phức tạp.

#### 2.3.2.2. Lớp Pooling\_lớp gộp

Các lớp Pooling được sử dụng một mặt để giảm kích thước không gian của biểu diễn và để giảm lượng tính toán được thực hiện trong mạng. Một công dụng khác của các lớp Pooling là để kiểm soát tình trạng quá khớp. Lớp Pooling được sử dụng nhiều nhất có các bộ lọc có kích thước 2 x 2 với bước nhảy 2. Điều này thực sự giảm kích thước đầu vào xuống còn một phần tư kích thước ban đầu của nó.

#### 2.3.2.3. Lớp Fully connected

Các lớp Fully connected là các lớp từ một mạng nơ-ron thông thường. Mỗi nơ-ron từ một lớp được kết nối đầy đủ được liên kết với mỗi đầu ra của lớp trước đó. Các hoạt động đằng sau một lớp tích chập giống như trong một lớp được kết nối đầy đủ. Do đó, có thể chuyển đổi giữa hai lớp.

#### 2.3.2.4. Lớp mất mát (Loss)

Các lớp mất mát được sử dụng để phạt mạng vì lệch khỏi đầu ra mong đợi. Đây thường là lớp cuối cùng của mạng. Có nhiều hàm mất mát khác nhau: softmax được sử dụng để dự đoán một lớp từ nhiều lớp không giao nhau, sigmoid cross-entropy được sử dụng để dự đoán nhiều xác suất độc lập.

### 2.2.3. Thư viện TensorFlow

Với mục đích triển khai, đào tạo và thử nghiệm mạng được mô tả trong bài báo này, chúng tôi đã sử dụng thư viện TensorFlow. Đây là một khuôn khổ nguồn mở cho máy học do Google tạo ra để tính toán số bằng biểu đồ luồng dữ liệu. Các nút trong biểu đồ biểu diễn các phép toán, trong khi các cạnh biểu đồ biểu diễn các mảng dữ liệu đa chiều được gọi là tenxơ.

Các thành phần chính trong hệ thống TensorFlow là máy khách, sử dụng giao diện Phiên để giao tiếp với máy chủ và một hoặc nhiều quy trình công nhân, trong đó mỗi quy trình công nhân chịu trách nhiệm phân xử quyền truy cập vào một hoặc nhiều thiết bị tính toán (như lõi CPU hoặc thẻ GPU) và thực thi các nút biểu đồ trên các thiết bị đó theo hướng dẫn của máy chủ.

# Chương 3: Phương pháp nghiên cứu (Methodology)

## 3.1. Cấu trúc của mạng nơ-ron được sử dụng trong các thí nghiệm

Đối với dự án này, chúng tôi đã sử dụng mạng nơ-ron tích chập. Như đã mô tả trước đó, loại mạng này sử dụng các lớp tích chập, các lớp gộp, các lớp ReLU, các lớp được kết nối đầy đủ và các lớp mất mát. Trong một kiến ​​trúc CNN điển hình, mỗi lớp tích chập được theo sau bởi một lớp Đơn vị tuyến tính chỉnh lưu (ReLU), sau đó là một lớp Ghép, sau đó là một hoặc nhiều lớp tích chập và cuối cùng là một hoặc nhiều lớp được kết nối đầy đủ.

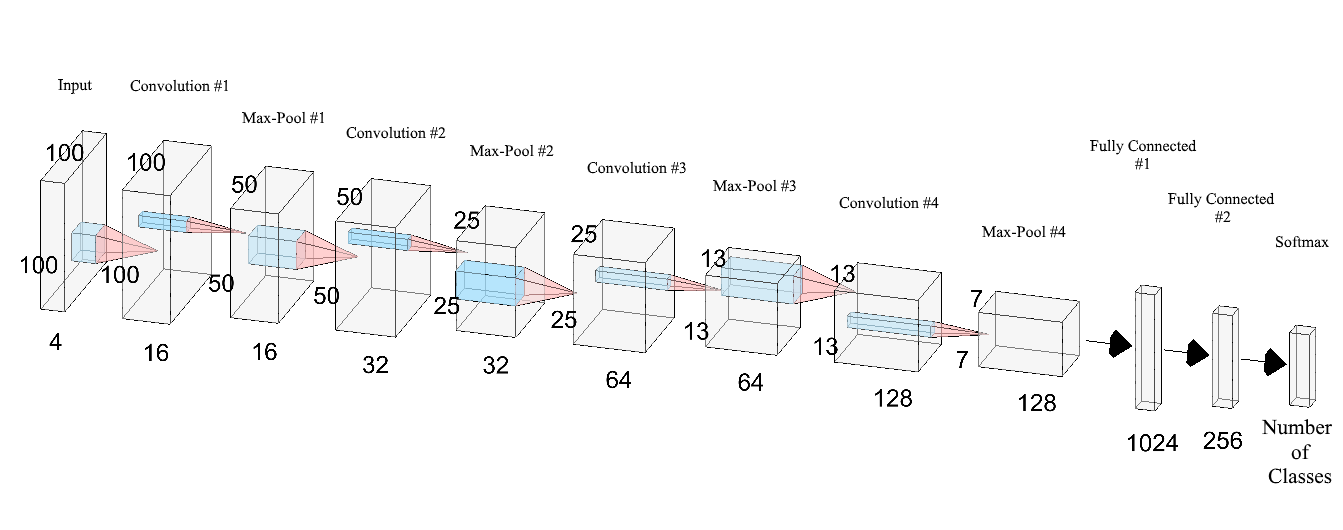
Đầu vào mà chúng tôi sử dụng bao gồm các hình ảnh RGB chuẩn có kích thước 100 x 100 pixel. Mạng nơ-ron mà chúng tôi sử dụng trong dự án này có cấu trúc được đưa ra trong Bảng 2.

|  |  |  |
| --- | --- | --- |
| Kiểu lớp | Kích thước | Đầu ra |
| Convolutional | 5 x 5 x 4 | 16 |
| Max pooling | 2 x 2 — Bước: 2 | - |
| Convolutional | 5 x 5 x 16 | 32 |
| Max pooling | 2 x 2 — Bước: 2 | - |
| Convolutional | 5 x 5 x 32 | 64 |
| Max pooling | 2 x 2 — Bước: 2 | - |
| Convolutional | 5 x 5 x 64 | 128 |
| Max pooling | 2 x 2 — Bước: 2 | - |
| Fully connected | 5 x 5 x 128 | 1024 |
| Fully connected | 1024 | 256 |
| Softmax | 256 | 60 |

Hình 3. Bảng cấu trúc của mạng nơ-ron được sử dụng trong bài báo này.

Biểu diễn trực quan của mạng nơ-ron được sử dụng được đưa ra trong Hình 3.

* Lớp đầu tiên (Convolution #1) là lớp tích chập áp dụng 16 bộ lọc 5 x 5. Trên lớp này, chúng tôi áp dụng max pooling với bộ lọc hình dạng 2 x 2 với bước 2, chỉ định rằng các vùng được gộp không chồng lên nhau (Max-Pool #1). Điều này cũng làm giảm chiều rộng và chiều cao xuống 50 pixel mỗi vùng.
* Lớp tích chập thứ hai (Convolution #2) áp dụng 32 bộ lọc 5 x 5, tạo ra 32 bản đồ kích hoạt. Chúng tôi áp dụng trên lớp này cùng loại max pooling (Max-Pool #2) như trên lớp đầu tiên, hình dạng 2 x 2 và bước 2.
* Lớp tích chập thứ ba (Convolution #3) áp dụng 64 bộ lọc 5 x 5. Sau đây là một lớp nhóm tối đa khác (Max-Pool #3) có hình dạng 2 x 2 và bước 2.
* Lớp tích chập thứ tư (Convolution #4) áp dụng 128 bộ lọc 5 x 5 sau đó chúng ta áp dụng lớp nhóm tối đa cuối cùng (Max-Pool #4).



Hình 4. Biểu diễn đồ họa của mạng nơ-ron tích chập được sử dụng trong các thí nghiệm.

## 3.2 Đối tượng và mẫu nghiên cứu

Các loại trái cây, hoa quả và hạt được thu thập và đánh nhãn như sau :

|  |  |  |
| --- | --- | --- |
| Label | Số lượng hình ảnh huấn luyện | Số lượng hình ảnh thử nghiệm |
| Apple Braeburn | 492 | 164 |
| Apple Golden 1 | 492 | 164 |
| Apple Golden 2 | 492 | 164 |
| Apple Golden 3 | 481 | 161 |
| Apple Granny Smith | 492 | 164 |
| Apple Red 1 | 492 | 164 |
| Apple Red 2 | 492 | 164 |
| Apple Red 3 | 429 | 144 |
| Apple Red Delicious | 490 | 166 |
| Apple Red Yellow 1 | 492 | 164 |
| Apple Red Yellow 2 | 672 | 219 |
| Apricot | 492 | 164 |
| Avocado | 427 | 143 |
| Avocado ripe | 491 | 166 |
| Banana | 490 | 166 |
| Banana Lady Finger | 450 | 152 |
| Banana Red | 490 | 166 |
| Cactus fruit | 490 | 166 |
| Cantaloupe 1 | 492 | 164 |

|  |  |  |
| --- | --- | --- |
| Cantaloupe 2 | 492 | 164 |
| Carambula | 490 | 166 |
| Cherry 1 | 492 | 164 |
| Cherry 2 | 738 | 246 |
| Cherry Rainier | 738 | 246 |
| Cherry Wax Black | 492 | 164 |
| Cherry Wax Red | 492 | 164 |
| Cherry Wax Yellow | 492 | 164 |
| Chestnut | 450 | 153 |
| Clementine | 490 | 166 |
| Cocos | 490 | 166 |
| Dates | 490 | 166 |
| Granadilla | 490 | 166 |
| Grape Blue | 984 | 328 |
| Grape Pink | 492 | 164 |
| Grape White | 490 | 166 |
| Grape White 2 | 490 | 166 |
| Grape White 3 | 492 | 164 |
| Grape White 4 | 471 | 158 |
| Grapefruit Pink | 490 | 166 |
| Grapefruit White | 492 | 164 |
| Guava | 490 | 166 |
| Huckleberry | 490 | 166 |
| Kaki | 490 | 166 |
| Kiwi | 466 | 156 |
| Kumquats | 490 | 166 |
| Lemon | 492 | 164 |
| Lemon Meyer | 490 | 166 |
| Limes | 490 | 166 |
| Lychee | 490 | 166 |
| Mandarine | 490 | 166 |
| Mango | 490 | 166 |
| Mangostan | 300 | 102 |
| Maracuja | 490 | 166 |

|  |  |  |
| --- | --- | --- |
| Melon Piel de Sapo | 738 | 246 |
| Mulberry | 492 | 164 |
| Nectarine | 492 | 164 |
| Orange | 479 | 160 |
| Papaya | 492 | 164 |
| Passion Fruit | 490 | 166 |
| Peach | 492 | 164 |
| Peach 2 | 738 | 246 |
| Peach Flat | 492 | 164 |
| Pear | 492 | 164 |
| Pear Abate | 490 | 166 |
| Pear Monster | 490 | 166 |
| Pear Williams | 490 | 166 |
| Pepino | 490 | 166 |
| Physalis | 492 | 164 |
| Physalis with Husk | 492 | 164 |
| Pineapple | 490 | 166 |
| Pineapple Mini | 493 | 163 |
| Pitahaya Red | 490 | 166 |
| Plum | 447 | 151 |
| Pomegranate | 492 | 164 |
| Quince | 490 | 166 |
| Rambutan | 492 | 164 |
| Raspberry | 490 | 166 |
| Redcurrant | 492 | 164 |
| Salak | 490 | 162 |
| Strawberry | 492 | 164 |
| Strawberry Wedge | 738 | 246 |
| Tamarillo | 490 | 166 |
| Tangelo | 490 | 166 |
| Tomato 1 | 738 | 246 |
| Tomato 2 | 672 | 225 |
| Tomato 3 | 738 | 246 |
| Tomato 4 | 479 | 160 |

Hình 5. Số lượng hình ảnh cho mỗi loại trái cây

## 3.3. Cách thu thập dữ liệu

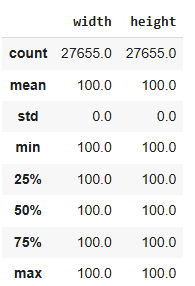
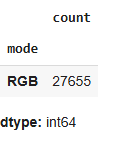
Trong phần này sẽ trình bày mô tả cách tạo bộ dữ liệu và nội dung của bộ dữ liệu. Hình ảnh được thu được bằng cách quay phim các loại trái cây trong khi chúng được một động cơ quay và sau đó trích xuất các khung hình.

Các loại trái cây được đặt trong trục của một động cơ tốc độ thấp (3 vòng/phút) và một bộ phim ngắn 20 giây đã được ghi lại. Đằng sau các loại trái cây, chúng tôi đặt một tờ giấy trắng làm nền.



Hình 6. Bên trái: hình ảnh gốc. Bên phải: quả sau khi xóa phần nền và sau khi được thu nhỏ xuống 100x100 pixel.

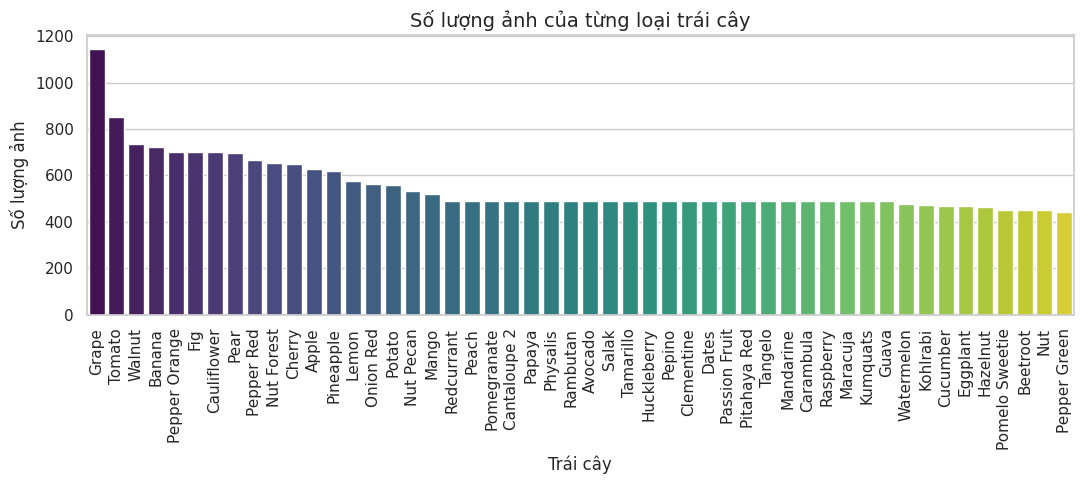
## 3.4. Phân tích khám phá dữ liệu

Hình 7. Tổng quan dữ liệu

Tất cả ảnh đều có kích thước đúng bằng 100×100 pixels và đều ở định dạng RGB.

Không có độ lệch chuẩn (std = 0), nên mọi ảnh đều có cùng kích thước.



Hình 8. Số lượng ảnh của từng nhãn trái cây

Có thể thấy, Class nhiều nhất (Grape: 1146 ảnh) gấp hơn 2.5 lần so với class ít nhất (Pepper Green: 444 ảnh).Nhãn từ 490 trở xuống khá đều, nhưng có một số class cao vượt trội (Grape, Tomato, Walnut,...). Dữ liệu mất cân bằng nhẹ đến vừa.

Dữ liệu bị mất cân bằng nhẹ tuy mức chênh lệch không quá nghiêm trọng nhưng có thể khiến mô hình thiên lệch học tốt các lớp phổ biến và bỏ qua các lớp ít hơn, dẫn đến độ chính xác tổng thể cao nhưng kém ở những lớp ít dữ liệu.

Cách giải quyết đề xuất: Thay vì làm tăng hoặc giảm số lượng mẫu, ta sẽ sử dụng class\_weight trong quá trình huấn luyện. Kỹ thuật này giúp mô hình ưu tiên học nhiều hơn từ các lớp có ít dữ liệu, bằng cách tăng độ quan trọng (trọng số) cho các lớp đó trong hàm mất mát. Đây là giải pháp đơn giản, hiệu quả, và không làm thay đổi dữ liệu gốc.

## 3.5. Tiền xử lý dữ liệu.

Do điều kiện ánh sáng thay đổi, nền không đồng nhất và chúng tôi đã viết một thuật toán chuyên dụng để trích xuất trái cây từ nền. Thuật toán này thuộc loại tô màu tràn: chúng tôi bắt đầu từ mỗi cạnh của hình ảnh và đánh dấu tất cả các pixel ở đó, sau đó chúng tôi đánh dấu tất cả các pixel được tìm thấy trong vùng lân cận của các pixel đã đánh dấu mà khoảng cách giữa các màu nhỏ hơn một giá trị quy định. chúng tôi lặp lại bước trước cho đến khi không thể đánh dấu thêm pixel nào nữa.

Tất cả các pixel được đánh dấu đều được coi là nền (sau đó được tô màu trắng) và các pixel còn lại được coi là thuộc về đối tượng. Giá trị tối đa cho khoảng cách giữa 2 pixel lân cận là tham số của thuật toán và được thiết lập (bằng cách thử và sai) cho mỗi bộ phim.

# Chương 4. Thực nghiệm và Thảo luận

## 4.1. Thí nghiệm số

Bộ dữ liệu được chia thành 2 phần: tập huấn luyện - bao gồm 27655 hình ảnh về trái cây và tập kiểm tra - bao gồm 10178 hình ảnh.

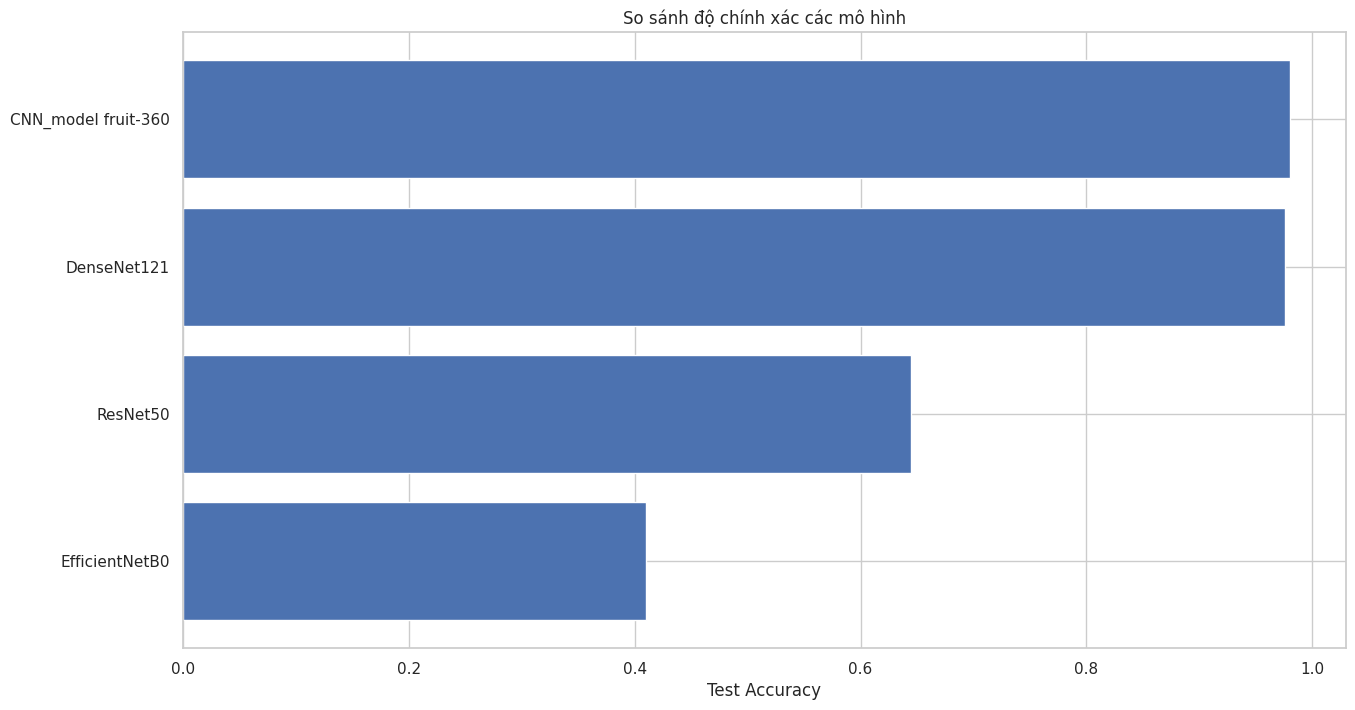
Dữ liệu được đóng gói thành tệp TFRecords (dành riêng cho TensorFlow). Đây là tệp nhị phân chứa bộ đệm giao thức có bản đồ đặc điểm. Trong bản đồ này có thể lưu trữ thông tin như chiều cao, chiều rộng, chiều sâu của hình ảnh và thậm chí cả hình ảnh thô. Sử dụng các tệp này, chúng ta có thể tạo hàng đợi để cung cấp dữ liệu cho mạng nơ-ron.

Chúng tôi đã thực nghiệm với các phương pháp sau để đánh giá trong đó mạng nơ-ron được đào tạo bằng cách sử dụng các mức tăng cường dữ liệu và tiền xử lý khác nhau:

* Chuyển đổi hình ảnh RGB đầu vào thành thang độ xám.
* Giữ nguyên hình ảnh đầu vào trong không gian màu RGB.
* Chuyển đổi hình ảnh RGB đầu vào thành không gian màu HSV.
* Chuyển đổi hình ảnh RGB đầu vào thành không gian màu HSV và thang độ xám và hợp nhất chúng.
* Áp dụng các thay đổi ngẫu nhiên về sắc độ và độ bão hòa trên hình ảnh RGB đầu vào, lật ngẫu nhiên chúng theo chiều ngang và chiều dọc, sau đó chuyển đổi chúng thành không gian màu HSV và thang độ xám rồi hợp nhất chúng.

## 4.2. Huấn luyện trên các mô hình khác.

Để xác định cấu hình mạng tốt nhất để phân loại hình ảnh trong tập dữ liệu của chúng tôi, chúng tôi đã lấy nhiều cấu hình, sử dụng tập huấn luyện để huấn luyện chúng và sau đó tính toán độ chính xác của chúng trên tập kiểm tra và tập huấn luyện.

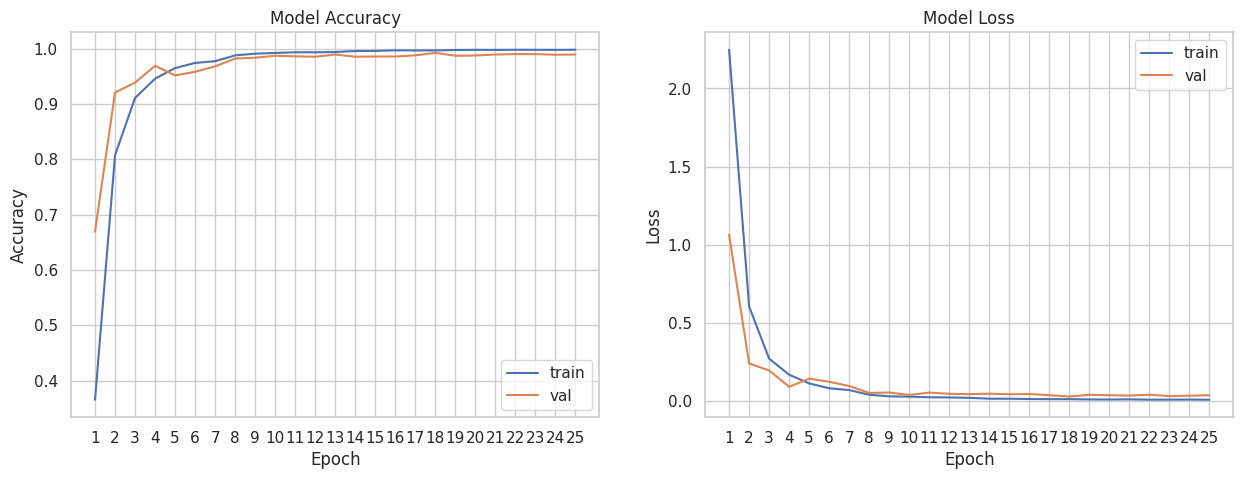


Hình 9. So sánh Accuracy của mô hình

* Mô hình CNN (fruit-360) đạt kết quả cao nhất, theo sau là DenseNet121.
* ResNet50 và EfficientNetB0 cho kết quả thấp hơn rõ rệt.

# Chương 5. Kết luận

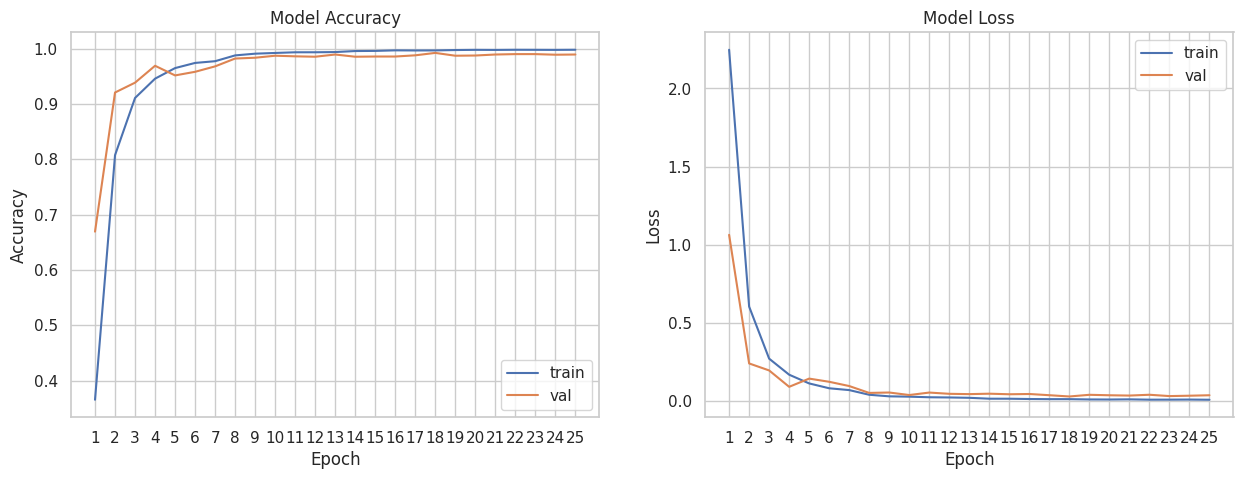
### 5.1. Tóm tắt kết quả đạt được và hướng nghiên cứu tương lai.



Hình 10.Biểu đồ Accuracy

Đường val (validation) bám sát hoặc thậm chí hơi cao hơn đường train giai đoạn đầu cho thấy mô hình chưa bị over-fit rõ ràng, và việc generalize ra dữ liệu unseen khá tốt.

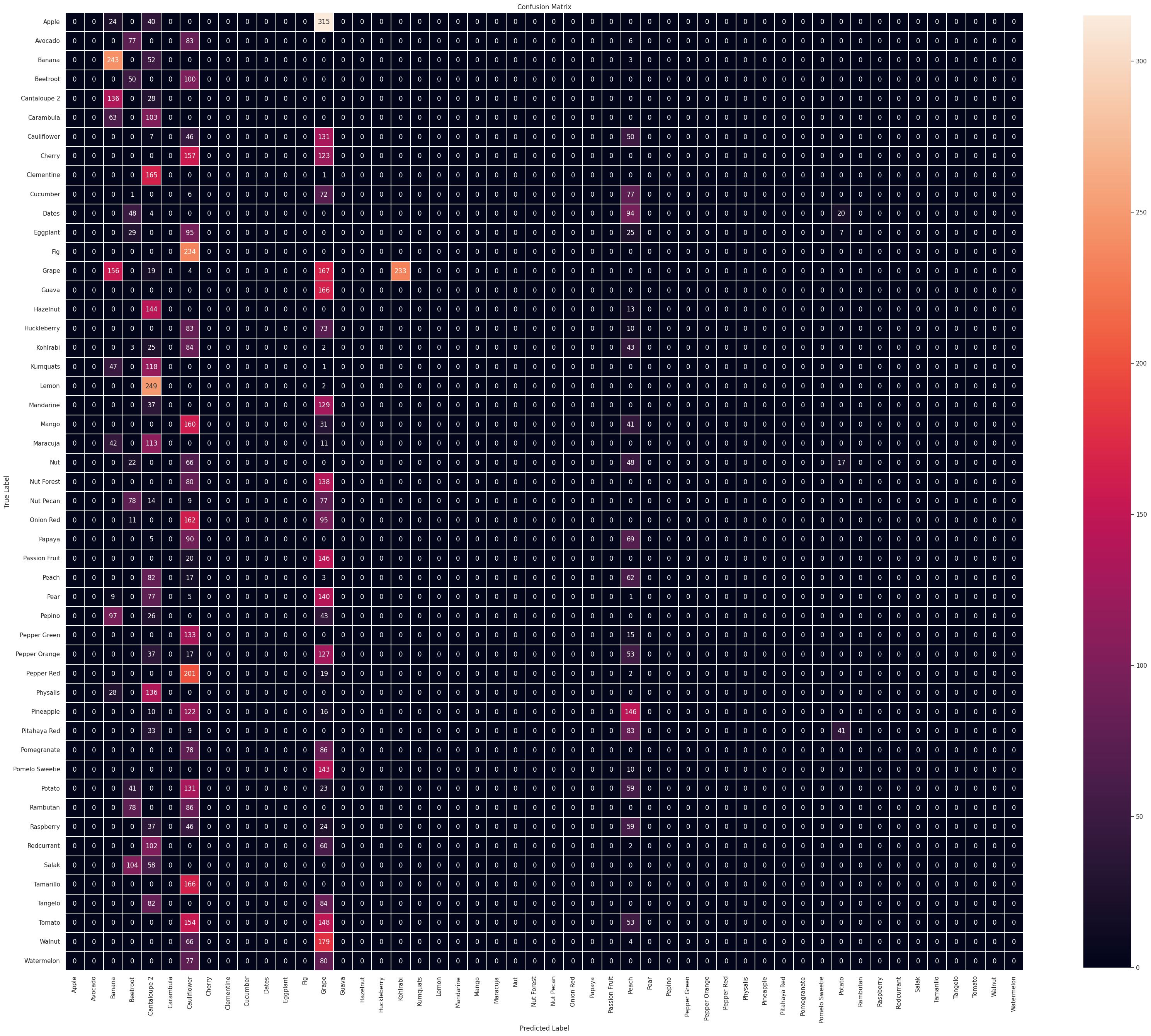
=> Kết luận: Mô hình học tốt, tổng quát hóa tốt trên tập validation.



Hình 11.Biểu đồ Loss

Train loss và validation loss đều giảm mạnh trong vài epoch đầu. Từ khoảng epoch 10 trở đi, cả hai loss gần như không thay đổi nhiều (rất thấp, < 0.1).

=> Kết luận: Mô hình hội tụ tốt, không bị underfitting hoặc overfitting nghiêm trọng.



Hình 12. Ma trận nhầm lẫn

Đa số các lớp có số lượng dự đoán đúng rất cao nằm trên đường chéo (ví dụ như **“Cantaloupe”, “Cucumber”, “Grapes”, “Pear”**, v.v.).

Nhiều ô ngoài đường chéo có giá trị bằng 0 hoặc rất nhỏ → chứng tỏ mô hình ít nhầm lẫn và phân biệt tốt giữa nhiều lớp.

Một số lớp có sự nhầm lẫn nhất định:

* Beets bị nhầm khá nhiều với Apple và Blackberry.
* Lemon có nhầm lẫn với Limes (tên và hình dạng tương đối giống nhau → cần thêm preprocessing hoặc augmentation).
* Plums có vẻ bị nhầm với Red Cherries và Pomegranate.
* Tomato, Peach, Passion Fruit và Papaya có nhầm lẫn nhẹ.

### 5.2. Câu trả lời cho câu hỏi nghiên cứu

*Liệu có thể xây dựng một hệ thống tự động xử lý và nhận dạng hình ảnh trái cây quay bằng thiết bị đơn giản và nền trắng không đồng đều hay không?*

***Câu trả lời:***

* Với việc áp dụng thuật toán định lượng (tách nền bằng flood fill) và mô hình CNN tự xây dựng, bài toán nhận dạng trái cây từ video thu được đã được giải quyết một cách hiệu quả.
* Quá trình tách nền hoạt động tốt trong hầu hết các trường hợp và mô hình học sâu đã khai thác tốt dữ liệu đầu vào sạch.

# TÀI LIỆU THAM KHẢO

[1] H. Mureșan and M. Oltean, “Fruit recognition from images using deep learning,” Acta Univ. Sapientiae, Informatica, vol. 10, no. 1, pp. 26–42, 2018. DOI: 10.2478/ausi-2018-0002.

[2] Frida Femling, Adam Olsson, Fernando Alonso-Fernandez, “Fruit and Vegetable Identification Using Machine Learning for Retail Applications”, arXiv:1810.09811v1 [cs.CV] 23 Oct 2018.

[3] J. Brownlee. (2016). Machine Learning Mastery With Python: Understand Your Data, Create Accurate Models and Work Projects End-To-End. Melbourne, VIC, Australia: Machine Learning Mastery.

[4] Cheng, H., Damerow, L., Sun, Y., and Blanke,M. Early yield prediction using image analysis of apple fruit and tree canopy features with neural networks. Journal of Imaging 3, 1 (2017).

[5] Cires¸an, D. C., Giusti, A., Gambardella, L. M., and Schmidhuber, J. Deep neural networks segment neuronal membranes in electron microscopy images. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2 (USA, 2012), NIPS’12, Curran Associates Inc., pp. 2843–2851.