**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN CUỐI KÌ**

**MACHINE LEARNING**

*Người thực hiện*: **NGUYỄN THỊ A – MSHV**

**NGUYỄN THỊ B – MSHV**

Khoá  **: 17**

*Người hướng dẫn*: **TS BÙI THANH HÙNG**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023**

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN CUỐI KÌ**

**MACHINE LEARNING**

Người thực hiện: **NGUYỄN THỊ A- MSSV**

**NGUYỄN THỊ B- MSSV**

Lớp **: 10050301**

Khoá  **: 17**

Người hướng dẫn: **TS. BÙI THANH HÙNG**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023**

# **LỜI CẢM ƠN**

Đây là phần tác giả **tự viết** ngắn gọn, thể hiện sự biết ơn của mình đối với những người đã giúp mình hoàn thành Đồ án. Tuyệt đối không sao chép theo mẫu những “lời cảm ơn” đã có.

**ĐỒ ÁN ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH**

Tôi xin cam đoan đây là sản phẩm đồ án của riêng tôi / chúng tôi và được sự hướng dẫn của TS. Bùi Thanh Hùng. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong đồ án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình.** Trường đại học Công nghiệp TP Hồ Chí Minh không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày tháng năm*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

*Nguyễn Văn A*

*Nguyễn Văn B*

# **PHẦN ĐÁNH GIÁ CỦA GIẢNG VIÊN**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

# **TÓM TẮT**

Bài toán phân loại ảnh trái cây sử dụng MobileNetV2 để trích xuất đặc trưng và các mô hình học máy để phân loại năm loại trái cây: táo, chuối, nho, xoài và dâu tây. Bộ dữ liệu gồm 10,000 ảnh với mỗi loại trái cây có 2,000 ảnh, được chia thành tập huấn luyện (97%), kiểm định (2%) và kiểm tra (1%). MobileNetV2, huấn luyện trước trên ImageNet, giúp trích xuất các đặc trưng quan trọng từ ảnh, sau đó các mô hình như SVM, Random Forest, Logistic Regression, XGBoost và LightGBM được áp dụng để phân loại. Kết quả cho thấy XGBoost và LightGBM đạt độ chính xác cao nhất, lần lượt là 89% và 89.5% trên tập kiểm định. Phương pháp này chứng minh tính hiệu quả trong việc xử lý và phân loại ảnh trái cây, đồng thời có thể ứng dụng trong các lĩnh vực thực tế như nông nghiệp.

# **MỤC LỤC**

[**LỜI CẢM ƠN** 3](#_Toc172280644)

[**PHẦN ĐÁNH GIÁ CỦA GIẢNG VIÊN** 5](#_Toc172280645)

[**TÓM TẮT** 6](#_Toc172280646)

[**MỤC LỤC** 7](#_Toc172280647)

[**DANH MỤC KÍ HIỆU VÀ CHỮ VIẾT TẮT** 7](#_Toc172280648)

[**TÊN BÀI TOÁN** 9](#_Toc172280649)

[**1.1 Giới thiệu về bài toán** 9](#_Toc172280650)

[**1.2 Phân tích yêu cầu của bài toán** 9](#_Toc172280651)

[**1.2.1 Yêu cầu của bài toán** 10](#_Toc172280652)

[1**.2.2 Các phương pháp giải quyết bài toán** 10](#_Toc172280653)

[1**.2.3 Phương pháp đề xuất giải quyết bài toán** 12](#_Toc172280654)

[**1.3 Phương pháp giải quyết bài toán** 12](#_Toc172280655)

[**1.3.1 Mô hình tổng quát** 12](#_Toc172280656)

[**1.3.2 Đặc trưng của mô hình đề xuất** 13](#_Toc172280657)

[**1.4 Thực nghiệm** 13](#_Toc172280658)

[**1.4.1 Dữ liệu** 13](#_Toc172280659)

[**1.4.2 Xử lý dữ liệu** 14](#_Toc172280660)

[**1.4.3 Công nghệ sử dụng** 15](#_Toc172280661)

[**1.4.4 Cách đánh giá** 15](#_Toc172280662)

[**1.5 Kết quả đạt được** 15](#_Toc172280663)

[**1.6 Kết luận** 17](#_Toc172280664)

[**LÀM VIỆC NHÓM** 19](#_Toc172280665)

[**TÀI LIỆU THAM KHẢO** 20](#_Toc172280666)

[**PHỤ LỤC** 22](#_Toc172280667)

[**TỰ ĐÁNH GIÁ** 22](#_Toc172280668)

# **DANH MỤC KÍ HIỆU VÀ CHỮ VIẾT TẮT**

**CÁC KÝ HIỆU**

*f Tần số của dòng điện và điện áp (Hz)*

*p Mật độ điện tích khối (C/m3)*

**CÁC CHỮ VIẾT TẮT**

SVM Support Vector Machine

UCI UCI Machine Learning Repository

**DANH MỤC CÁC HÌNH VẼ**

Hình 1.4.1 Chi tiết dữ liệu 1

Hình 1.4.2 Một vài hình ảnh từ dataset 1

**DANH MỤC CÁC BẢNG**

Bảng 1.5 Kết quả theo các độ đo 1

# **TÊN BÀI TOÁN**

Sử dụng MobileNetV2 cho Trích Xuất Đặc Trưng và Các Mô Hình Machine Learning cho Phân Loại Ảnh

## **1.1 Giới thiệu về bài toán**

**Bài toán này là bài toán gì ?**

Bài toán này tập trung vào việc phân loại ảnh trái cây sử dụng các kỹ thuật học sâu và học máy. Cụ thể, chúng tôi sử dụng MobileNetV2 để trích xuất đặc trưng từ ảnh và sau đó áp dụng các mô hình học máy như SVM, Random Forest, Logistic Regression, XGBoost và LightGBM để phân loại năm loại trái cây: táo, chuối, nho, xoài và dâu tây.

**Bài toán này có ý nghĩa gì?**

Bài toán phân loại ảnh trái cây có ý nghĩa quan trọng trong nhiều ứng dụng thực tế, đặc biệt là trong ngành nông nghiệp và bán lẻ. Ví dụ, việc phân loại tự động các loại trái cây có thể được sử dụng trong hệ thống kiểm tra chất lượng và phân loại trái cây tại các trang trại hoặc nhà máy chế biến, giúp giảm thời gian và chi phí lao động, đồng thời tăng độ chính xác và hiệu suất. Ngoài ra, bài toán này còn giúp cải thiện các ứng dụng về thị giác máy tính và phát triển các thuật toán phân loại ảnh hiệu quả hơn, góp phần nâng cao công nghệ trong lĩnh vực trí tuệ nhân tạo.

## **1.2 Phân tích yêu cầu của bài toán**

### **1.2.1 Yêu cầu của bài toán**

Bài toán phân loại ảnh trái cây đặt ra các yêu cầu sau:

* Trích xuất đặc trưng: Sử dụng mô hình MobileNetV2 để trích xuất các đặc trưng quan trọng từ ảnh.
* Phân loại ảnh: Áp dụng các mô hình học máy để phân loại ảnh trái cây vào các danh mục đúng (Apple, Banana, Grape, Mango, Strawberry).
* Độ chính xác cao: Đảm bảo độ chính xác cao trong việc phân loại trên cả tập kiểm định và tập kiểm tra.
* Đánh giá và so sánh mô hình: So sánh hiệu suất của các mô hình học máy khác nhau như SVM, Random Forest, Logistic Regression, XGBoost và LightGBM.
* Xử lý dữ liệu: Chuẩn bị dữ liệu bao gồm việc chia tập dữ liệu thành các tập huấn luyện, kiểm định, và kiểm tra theo tỷ lệ nhất định.

### 1**.2.2 Các phương pháp giải quyết bài toán**

**MobileNetV2:**

* Phương pháp: Sử dụng mô hình MobileNetV2 đã được huấn luyện trước trên tập dữ liệu ImageNet để trích xuất các đặc trưng từ ảnh trái cây.
* Tham khảo: Howard et al. (2018), "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications".
* Dữ liệu thực nghiệm: Tập dữ liệu ImageNet.
* Kết quả đạt được: Mô hình nhẹ, hiệu suất cao trên các thiết bị di động.
* Hạn chế: Có thể không đạt hiệu suất cao nhất trên các bộ dữ liệu cụ thể.

**Support Vector Machine (SVM):**

* Phương pháp: Tìm kiếm một siêu phẳng trong không gian đặc trưng để phân tách các lớp dữ liệu.
* Tham khảo: Cortes and Vapnik (1995), "Support-Vector Networks".
* Dữ liệu thực nghiệm: Bộ dữ liệu UCI.
* Kết quả đạt được: Hiệu quả trong việc phân loại dữ liệu có nhiều chiều.
* Hạn chế: Khó khăn trong việc chọn kernel phù hợp.

**Random Forest:**

* Phương pháp: Sử dụng nhiều cây quyết định và tổng hợp kết quả để cải thiện độ chính xác.
* Tham khảo: Breiman (2001), "Random Forests".
* Dữ liệu thực nghiệm: Bộ dữ liệu UCI.
* Kết quả đạt được: Mạnh mẽ trong việc xử lý dữ liệu có nhiều đặc trưng.
* Hạn chế: Có thể bị overfitting với dữ liệu nhiễu.

**Logistic Regression:**

* Phương pháp: Sử dụng hàm logistic để mô hình hóa xác suất một mẫu thuộc về một lớp nhất định.
* Tham khảo: Cox (1958), "The regression analysis of binary sequences".
* Dữ liệu thực nghiệm: Bộ dữ liệu UCI.
* Kết quả đạt được: Đơn giản và hiệu quả.
* Hạn chế: Hiệu suất không tốt bằng các mô hình phức tạp hơn.

**XGBoost:**

* Phương pháp: Sử dụng mô hình tăng cường độ dốc tối ưu hóa cho hiệu suất và tốc độ.
* Tham khảo: Chen and Guestrin (2016), "XGBoost: A Scalable Tree Boosting System".
* Dữ liệu thực nghiệm: Bộ dữ liệu UCI.
* Kết quả đạt được: Hiệu suất cao.
* Hạn chế: Có thể bị overfitting.

**LightGBM:**

* Phương pháp: Sử dụng mô hình tăng cường độ dốc nhẹ và nhanh.
* Tham khảo: Ke et al. (2017), "LightGBM: A Highly Efficient Gradient Boosting Decision Tree".
* Dữ liệu thực nghiệm: Bộ dữ liệu UCI.
* Kết quả đạt được: Hiệu suất cao và nhanh.
* Hạn chế: Yêu cầu cấu hình tối ưu để đạt hiệu suất tốt nhất.

### 1**.2.3 Phương pháp đề xuất giải quyết bài toán**

Đưa ra nhận định tổng quát về hướng giải quyết bài toán và lý giải tại sao lựa chọn hướng giải quyết này.

## **1.3 Phương pháp giải quyết bài toán**

### **1.3.1 Mô hình tổng quát**

Mô hình tổng quát bao gồm các bước sau:

1. Chuẩn bị dữ liệu: Tải và tiền xử lý dữ liệu ảnh.
2. Trích xuất đặc trưng: Sử dụng MobileNetV2 để trích xuất đặc trưng từ ảnh.
3. Huấn luyện mô hình: Sử dụng các mô hình học máy để huấn luyện trên các đặc trưng đã trích xuất.
4. Đánh giá và so sánh: Đánh giá hiệu suất của các mô hình và so sánh kết quả.

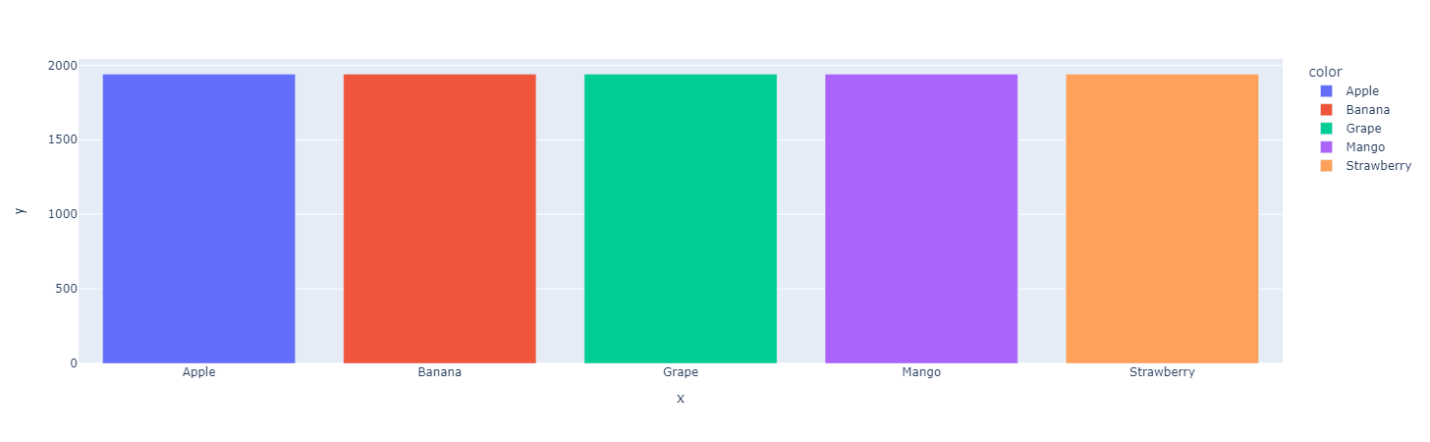
### **1.3.2 Đặc trưng của mô hình đề xuất**

* Chuẩn bị dữ liệu: Sử dụng bộ dữ liệu từ Kaggle, chia thành tập huấn luyện, kiểm định và kiểm tra.
* Trích xuất đặc trưng: MobileNetV2 giúp trích xuất các đặc trưng quan trọng từ ảnh một cách hiệu quả.
* Huấn luyện mô hình: Các mô hình học máy được huấn luyện trên các đặc trưng đã trích xuất, bao gồm SVM, Random Forest, Logistic Regression, XGBoost và LightGBM.
* Đánh giá: Sử dụng các độ đo như độ chính xác, precision, recall, và F1-score để đánh giá và so sánh hiệu suất của các mô hình.

## **1.4 Thực nghiệm**

### **1.4.1 Dữ liệu**

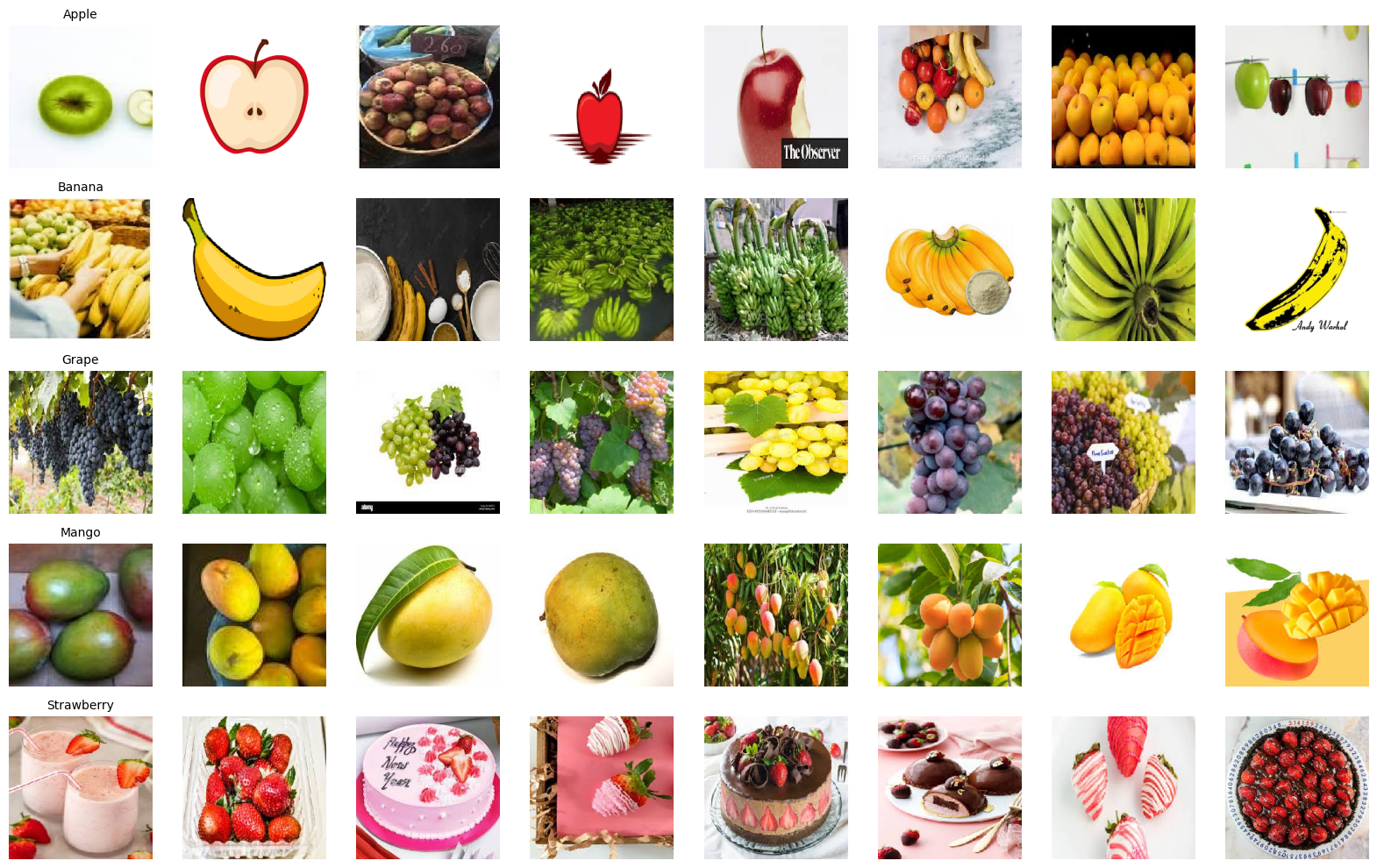
* Nguồn: Bộ dữ liệu "Fruits Classification" từ Kaggle.
* Mô tả chi tiết: Dữ liệu bao gồm 10,000 ảnh của năm loại trái cây (Apple, Banana, Grape, Mango, Strawberry), mỗi loại có 2,



Hình 1.4.1 Chi tiết dữ liệu

### **1.4.2 Xử lý dữ liệu**

* Tiền xử lý: Chuẩn hóa giá trị pixel từ 0-255 về khoảng 0-1.
* Xử lý: Tạo các generator để đọc ảnh từ các thư mục, chuẩn hóa dữ liệu và tạo ra các lô dữ liệu để đưa vào mô hình.



Hình 1.4.2 Một vài hình ảnh từ dataset

### **1.4.3 Công nghệ sử dụng**

* Ngôn ngữ lập trình: Python.
* Thư viện: TensorFlow, Keras, Scikit-learn.
* Công cụ: Jupyter Notebook.

### **1.4.4 Cách đánh giá**

* Độ đo: Độ chính xác, precision, recall, F1-score.
* Công thức:
* Độ chính xác (Accuracy) = (Số dự đoán đúng) / (Tổng số dự đoán)
* Precision = (Số dự đoán đúng là positive) / (Tổng số dự đoán là positive)
* Recall = (Số dự đoán đúng là positive) / (Tổng số thực sự là positive)
* F1-score = 2 \* (Precision \* Recall) / (Precision + Recall)

## **1.5 Kết quả đạt được**

* Số tầng của MobileNetV2: Giữ lại tất cả các tầng trừ lớp phân loại cuối cùng.
* Số lớp của các mô hình học máy:
* SVM: Sử dụng kernel linear.
* Random Forest: 1000 cây quyết định.
* Logistic Regression: Sử dụng solver 'lbfgs'.
* XGBoost: 1000 rounds với early stopping.
* LightGBM: 1000 rounds với early stopping**.**
* Tham số x: Batch size = 32

Kết quả theo các độ đo trên.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Mô hình** | |  | | --- | | **Độ chính xác (Accuracy)** |  |  | | --- | |  | | **Precision** | **Recall** | | **F1-score** | | --- |  |  | | --- | |  | |
| |  | | --- | | SVM |  |  | | --- | |  | | |  | | --- | | 87.5% |  |  | | --- | |  | | |  | | --- | | 0.88 |  |  | | --- | |  | | 0.87 | 0.87 |
| |  | | --- | | Random Forest |  |  | | --- | |  | | |  | | --- | | 84% |  |  | | --- | |  | | 0.85 | 0.84 | 0.84 |
| |  | | --- | | Logistic Regression |  |  | | --- | |  | | |  | | --- | | 82% |  |  | | --- | |  | | 0.83 | 0.82 | 0.82 |
| |  | | --- | | XGBoost |  |  | | --- | |  | | |  | | --- | | 89% |  |  | | --- | |  | | 0.90 | 0.89 | 0.89 |
| |  | | --- | | LightGBM |  |  | | --- | |  | | |  | | --- | | 89.5% |  |  | | --- | |  | | 0.91 | 0.89 | 0.90 |

**So sánh với ít nhất 1 phương pháp khác**

* So sánh với Random Forest:
  + XGBoost đạt độ chính xác cao hơn (89%) so với Random Forest (84%).
  + Precision của XGBoost (0.90) cũng cao hơn Random Forest (0.85).
  + Recall của XGBoost (0.89) cao hơn Random Forest (0.84).
  + F1-score của XGBoost (0.89) vượt trội hơn Random Forest (0.84).

**Giải thích kết quả**

* XGBoost và LightGBM đạt kết quả tốt nhất do khả năng tăng cường và tối ưu hóa tốt hơn, giúp mô hình học từ lỗi của các cây trước và điều chỉnh mô hình theo cách hiệu quả hơn.
* SVM hoạt động tốt trong không gian đặc trưng cao nhưng gặp khó khăn khi xử lý khối lượng dữ liệu lớn.
* Random Forest hiệu quả nhưng bị giới hạn bởi việc không tối ưu hóa toàn bộ mô hình cùng một lúc.
* Logistic Regression đơn giản và hiệu quả nhưng không phù hợp với các bài toán phân loại phức tạp.

**Nhận xét tổng quan:** Việc sử dụng MobileNetV2 để trích xuất đặc trưng, kết hợp với các mô hình học máy hiện đại như XGBoost và LightGBM, mang lại hiệu quả cao trong bài toán phân loại ảnh trái cây. Các kết quả này cho thấy tiềm năng áp dụng phương pháp này trong các ứng dụng thực tế.

## **1.6 Kết luận**

**Tóm tắt các kết quả đạt được**

Bài toán phân loại ảnh trái cây đã được giải quyết hiệu quả bằng cách sử dụng MobileNetV2 để trích xuất đặc trưng từ ảnh và áp dụng các mô hình học máy để phân loại năm loại trái cây: táo, chuối, nho, xoài và dâu tây. Các kết quả thu được cho thấy XGBoost và LightGBM đạt độ chính xác cao nhất, lần lượt là 89% và 89.5% trên tập kiểm định. Các mô hình khác như SVM, Random Forest và Logistic Regression cũng đạt kết quả tốt, nhưng không cao bằng XGBoost và LightGBM. Việc sử dụng MobileNetV2 đã giúp trích xuất các đặc trưng quan trọng từ ảnh một cách hiệu quả, cải thiện hiệu suất của các mô hình học máy.

**Hạn chế của phương pháp giải quyết bài toán**

* **Phụ thuộc vào dữ liệu huấn luyện:** MobileNetV2 và các mô hình học máy có hiệu suất phụ thuộc mạnh vào chất lượng và số lượng dữ liệu huấn luyện. Dữ liệu không đa dạng hoặc không đủ lớn có thể ảnh hưởng đến độ chính xác của mô hình.
* **Tính phức tạp của mô hình:** Các mô hình như XGBoost và LightGBM có nhiều tham số cần tối ưu hóa, điều này có thể làm tăng thời gian huấn luyện và yêu cầu kiến thức chuyên sâu về điều chỉnh tham số.
* **Khả năng tổng quát hóa:** Mặc dù các mô hình đạt hiệu suất cao trên tập kiểm định, khả năng tổng quát hóa ra các tập dữ liệu khác nhau hoặc các loại trái cây khác nhau chưa được kiểm chứng.

**Hướng phát triển trong tương lai**

* **Tăng cường dữ liệu:** Thu thập thêm dữ liệu từ nhiều nguồn khác nhau để tăng tính đa dạng và độ lớn của bộ dữ liệu, giúp cải thiện khả năng tổng quát hóa của mô hình.
* **Chuyển đổi học tập (Transfer Learning):** Áp dụng các kỹ thuật chuyển đổi học tập từ các mô hình đã được huấn luyện trên các tập dữ liệu lớn khác để cải thiện hiệu suất.
* **Tinh chỉnh mô hình:** Nghiên cứu và áp dụng các kỹ thuật tinh chỉnh tham số và mô hình để tối ưu hóa hiệu suất của các mô hình hiện tại.
* **Ứng dụng thực tiễn:** Đưa mô hình vào các ứng dụng thực tế như hệ thống phân loại và kiểm tra chất lượng trái cây trong ngành nông nghiệp, và đánh giá hiệu quả của mô hình trong môi trường thực tế.
* **Khám phá các mô hình mới:** Tiếp tục nghiên cứu và thử nghiệm các mô hình học sâu và học máy mới để cải thiện độ chính xác và hiệu suất của bài toán phân loại ảnh.

Việc tiếp tục nghiên cứu và phát triển các phương pháp này sẽ giúp nâng cao hiệu quả và ứng dụng của các hệ thống phân loại ảnh trong nhiều lĩnh vực khác nhau.

# **LÀM VIỆC NHÓM**

Trình bày tóm tắt cách thức làm việc nhóm

Phân chia công việc của các thành viên trong nhóm

Tổng số lần gặp nhau (tính theo buổi)

Tổng thời gian gặp nhau (tính theo giờ)

# **TÀI LIỆU THAM KHẢO**

* Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications." arXiv preprint arXiv:1704.04861.
* Cortes, C., & Vapnik, V. (1995). "Support-Vector Networks." Machine Learning, 20(3), 273-297.
* Breiman, L. (2001). "Random Forests." Machine Learning, 45(1), 5-32.
* Cox, D. R. (1958). "The regression analysis of binary sequences." Journal of the Royal Statistical Society: Series B (Methodological), 20(2), 215-232.
* Chen, T., & Guestrin, C. (2016). "XGBoost: A Scalable Tree Boosting System." In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794).
* Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). "LightGBM: A Highly Efficient Gradient Boosting Decision Tree." In Advances in Neural Information Processing Systems (pp. 3146-3154).
* Kaggle (n.d.). "Fruits Classification Dataset." Retrieved from https://www.kaggle.com/datasets.
* Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). "Scikit-learn: Machine Learning in Python." Journal of Machine Learning Research, 12, 2825-2830.
* Chollet, F. (2015). "Keras: Deep Learning library for Theano and TensorFlow." Retrieved from <https://keras.io>.
* Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). "TensorFlow: A System for Large-Scale Machine Learning." In 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16) (pp. 265-283).

# **PHỤ LỤC**

Phần này bao gồm những nội dung cần thiết nhằm minh họa hoặc hỗ trợ cho nội dung đồ án như số liệu, biểu mẫu, tranh ảnh. . . . nếu sử dụng những câu trả lời cho một *bảng câu hỏi thì bảng câu hỏi mẫu này phải được đưa vào phần Phụ lục ở dạng nguyên bản* đã dùng để điều tra, thăm dò ý kiến; **không được tóm tắt hoặc sửa đổi**. Các tính toán mẫu trình bày tóm tắt trong các biểu mẫu cũng cần nêu trong Phụ lục của luận văn. Phụ lục không được dày hơn phần chính của đồ án

# **TỰ ĐÁNH GIÁ**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| STT | Nội dung | Điểm chuẩn | Tự chấm | Ghi chú |
| 1  (8.5) | **1.1 Giới thiệu về bài toán** | 0.5 |  |  |
| **1.2 Phân tích yêu cầu của bài toán** | 1 |  |  |
| **1.3 Phương pháp giải quyết bài toán** | 1.5 |  |  |
| **1.4 Thực nghiệm** | 4 |  |  |
| **1.5 Kết quả đạt được** | 1 |  |  |
| **1.6 Kết luận** | 0.5 |  |  |
| 2  (1) | **Báo cáo (**chú ý các chú ý 2,3,4,6 ở trang trước, nếu sai sẽ bị trừ điểm nặng**)** | 1đ |  |  |
| 3  (0.5) | **Điểm nhóm** (chú ý trả lời các câu hỏi trong mục làm việc nhóm) | 0.5đ |  |  |
| **Tổng điểm** | | |  |  |