

AI Course

# Capstone Project Final Report

©2023 SAMSUNG. All rights reserved.

Samsung Electronics Corporate Citizenship Office holds the copyright of this document.

This document is a literary property protected by copyright law so reprint and reproduction without permission are prohibited.

To use this document other than the curriculum of Samsung Innovation Campus, you must receive written consent from copyright holder

# Mô hình học không giám sát cho bài toán phân loại rác thải

20/07/2025

R4BBISH

Nguyễn Trung Kiên

Trần Nam Hải

Vũ Văn Đông

Vũ Quang Khánh

# ***Nội dung***

## **1. Mô tả bài toán**

- 1.1. Đặt vấn đề & Mục tiêu**
- 1.2. Dữ liệu đầu vào**
- 1.3. Dữ liệu đầu ra**

## **2. Giải quyết vấn đề**

- 2.1. Cơ sở lý thuyết**
- 2.2. Mô hình sử dụng**
- 2.3. Kiến trúc hệ thống**

## **3. Thực nghiệm, đánh giá**

- 3.1. Thực nghiệm**
- 3.2. Đánh giá**
- 3.3. Kết quả**

## **4. Kết luận và hướng phát triển**

- 4.1. Kết luận**
- 4.2. Tổng kết**
- 4.3. Hướng phát triển**

## **5. Đánh giá và Nhận xét của Thành viên Nhóm**

## **6. Đánh giá và Nhận xét của Giảng viên Hướng dẫn**

# 1. Mô tả bài toán

## 1.1. Đặt vấn đề & Mục tiêu

### 1.1.1. Đặt vấn đề

Với tốc độ đô thị hóa và công nghiệp hóa nhanh chóng, lượng rác thải sinh hoạt và công nghiệp ngày càng tăng cao, gây áp lực lớn lên hệ thống quản lý và xử lý chất thải. Việc phân loại rác tại nguồn là một trong những giải pháp quan trọng giúp tái chế hiệu quả, giảm ô nhiễm môi trường và tiết kiệm tài nguyên. Tuy nhiên, công tác này đang đối mặt với nhiều thách thức:

- **Thiếu hiệu quả trong phân loại thủ công:** Việc phân loại rác hiện nay chủ yếu dựa vào con người, dễ xảy ra sai sót, tốn thời gian và không nhất quán.
- **Khó khăn trong thu thập dữ liệu nhãn:** Việc gán nhãn cho từng loại rác (hữu cơ, vô cơ, tái chế...) đòi hỏi nguồn lực lớn, dẫn đến thiếu dữ liệu được gán nhãn đầy đủ để huấn luyện các mô hình học máy có giám sát.
- **Đa dạng hình dạng, màu sắc, chất liệu rác thải:** Rác có đặc điểm thị giác đa dạng, làm cho việc xây dựng mô hình phân loại trở nên phức tạp

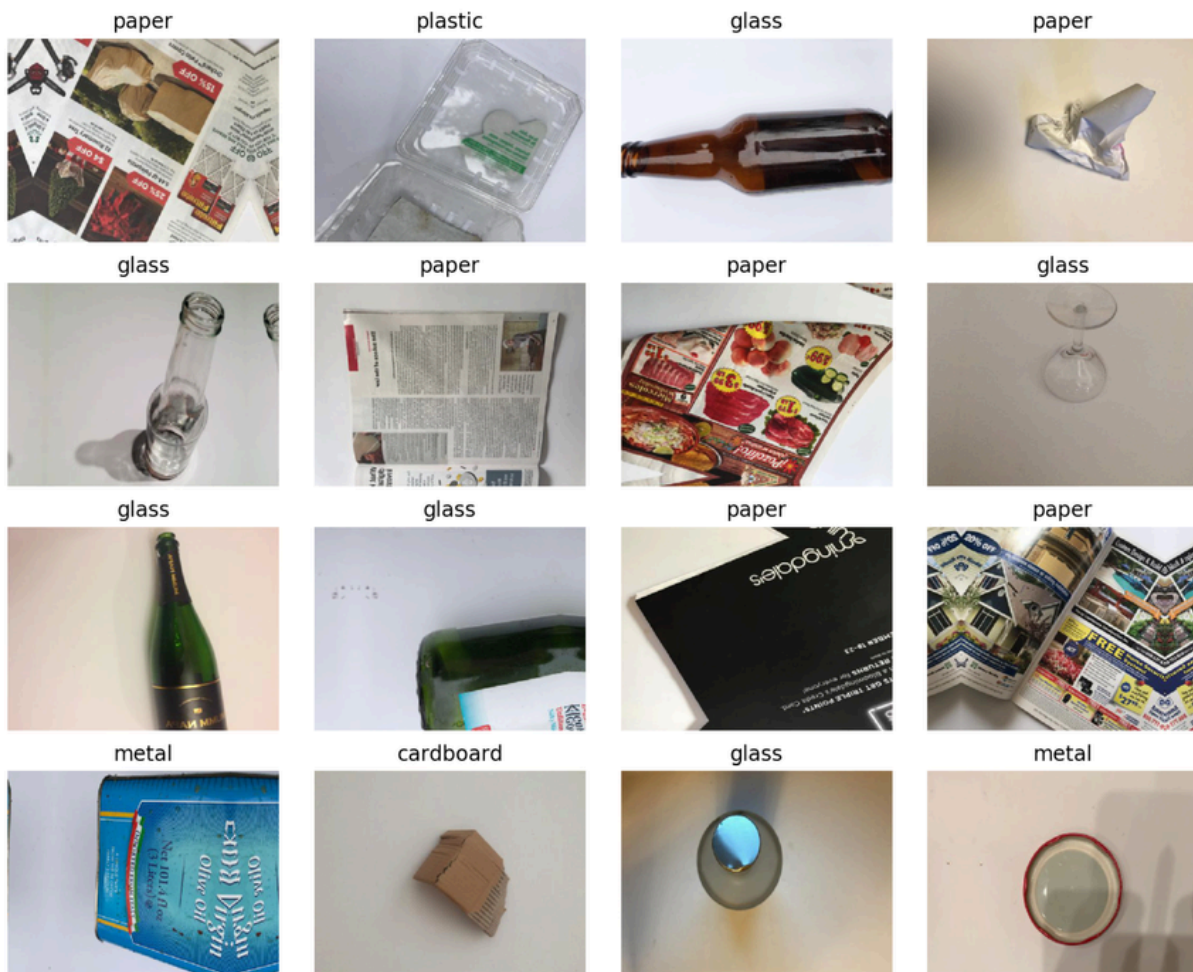
Vì vậy, cần một giải pháp thông minh có khả năng tự động học được cấu trúc dữ liệu và phát hiện các nhóm rác tương đồng mà không cần dựa vào nhãn — đó là lý do **mô hình học không giám sát** được đặt ra trong bài toán phân loại rác thải.

### 1.1.2. Mục tiêu

- **Xây dựng một mô hình học không giám sát** có khả năng **tự động phân nhóm rác thải** dựa trên đặc trưng thị giác (hình ảnh).
- **Tận dụng các phương pháp clustering (phân cụm)** như K-Means, DBSCAN, hoặc mô hình tự mã hóa (Autoencoder) để phát hiện cấu trúc ngầm của dữ liệu mà không cần nhãn.
- **Giảm phụ thuộc vào dữ liệu gán nhãn thủ công**, từ đó tăng tính khả thi và khả năng mở rộng của hệ thống phân loại rác.
- **Góp phần vào hệ thống phân loại và tái chế rác tự động**, hướng tới thành phố thông minh và phát triển bền vững.
- **Tạo tiền đề** cho việc phát triển các mô hình phân loại rác có giám sát bằng cách tạo ra nhãn tạm thời hoặc cụm dữ liệu chất lượng cao để phục vụ huấn luyện về sau.

## 1.2. Dữ liệu đầu vào

- Bộ dữ liệu TrashNet: bao gồm 2.527 ảnh tách thành 6 nhóm khác nhau
- Bộ dữ liệu VN-TRASHNET: bao gồm 12.091 ảnh tách thành 9 nhóm khác nhau



### 1.3. Dữ liệu đầu ra

File CSV bao gồm đường dẫn ảnh và nhãn phân cụm được gán bởi thuật toán học không giám sát

## 2. Giải quyết vấn đề

### 2.1. Cơ sở lý thuyết

#### 2.1.1. Học không giám sát và phân cụm (Clustering)

Học không giám sát là phương pháp học máy trong đó mô hình cố gắng rút ra cấu trúc tiềm ẩn từ dữ liệu **không có nhãn**. Trong bối cảnh phân loại rác thải, học không giám sát giúp hệ thống tự động **phân nhóm các hình ảnh rác** có đặc điểm thị giác tương đồng mà không cần nhãn thủ công, từ đó hỗ trợ việc phân loại rác hiệu quả hơn.

Một trong những kỹ thuật phổ biến trong học không giám sát là **phân cụm (clustering)**. Mục tiêu của phân cụm là nhóm các đối tượng (ở đây là hình ảnh rác) thành các cụm sao cho các phần tử trong cùng một cụm tương đồng với nhau hơn so với phần tử ở cụm khác.

Các phương pháp phân cụm sử dụng trong mô hình bao gồm:

- **KMeans**: Chia dữ liệu thành K cụm dựa trên khoảng cách Euclidean đến các centroid.
- **BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)**: Phân cụm hiệu quả trên dữ liệu lớn bằng cách xây dựng cây CF (Clustering Feature).
- **Agglomerative Clustering**: Phân cụm phân cấp từ dưới lên (bottom-up), gộp từng phần tử thành cụm dựa trên độ tương đồng.

#### 2.1.2. Trích xuất đặc trưng bằng học sâu

Để mô hình phân cụm hoạt động hiệu quả, cần trích xuất đặc trưng chất lượng cao từ hình ảnh. Mô hình sử dụng:

- **ResNet50**: Mạng tích chập sâu với kiến trúc residual giúp học được các đặc trưng sâu và ổn định.
- **Vision Transformer (ViT)**: Mạng transformer áp dụng cho ảnh, có khả năng học các mối quan hệ không gian và ngữ cảnh hiệu quả.

#### 2.1.3. Contrastive Learning

Contrastive Learning là phương pháp học biểu diễn (representation learning) không giám sát, trong đó mô hình học cách kéo các mẫu "giống nhau" lại gần nhau trong không gian đặc trưng và đẩy các mẫu "khác nhau" ra xa nhau. Điều này giúp mô hình học được **biểu diễn đặc trưng mạnh mẽ, phân tách tốt**, hỗ trợ các thuật toán phân cụm phía sau đạt kết quả chính xác hơn.

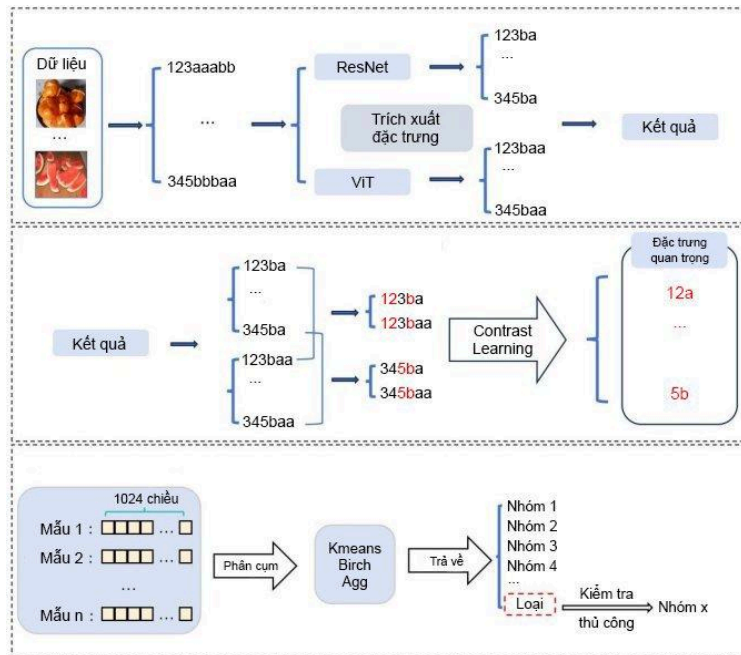
### 2.2. Mô hình sử dụng

Mô hình đề xuất gồm các thành phần chính sau:

- **Trích xuất đặc trưng hình ảnh**: Mỗi hình ảnh rác được đưa qua hai mạng học sâu là **ResNet50** và **Vision Transformer (ViT)** để thu được các vector đặc trưng. Các đặc trưng này biểu diễn nội dung và cấu trúc trực quan của ảnh.
- **Tăng cường đặc trưng với Contrastive Learning**: Vector đặc trưng sau đó được tối ưu thêm bằng **Contrastive Learning**, giúp cải thiện khả năng phân biệt giữa các loại rác khác nhau.
- **Phân cụm không giám sát**: Ba thuật toán phân cụm được áp dụng song song
  - KMeans
  - BIRCH

- Agglomerative Clustering
- **Voting tổ hợp kết quả phân cụm:**  
 Kết quả phân cụm từ ba mô hình được tổng hợp bằng **quy trình bỏ phiếu (voting)** để xác định cụm phân loại cuối cùng cho mỗi ảnh. Việc kết hợp này giúp giảm ảnh hưởng của sai số đơn lẻ từ từng thuật toán, tăng độ ổn định và độ tin cậy của mô hình.

## 2.3. Kiến trúc hệ thống



### 2.3.1. Mô-đun Thu thập dữ liệu

- **Nguồn dữ liệu:** Tập hợp hình ảnh rác thải từ bộ dữ liệu
- **Tiền xử lý ảnh:** Cắt ảnh, thay đổi kích thước, chuẩn hóa để phù hợp với yêu cầu đầu vào của ResNet50 và ViT.

### 2.3.2. Mô-đun Trích xuất và Học biểu diễn

- **Trích xuất đặc trưng:**
  - Ảnh được đưa qua ResNet50 và ViT để lấy vector đặc trưng.
  - Các vector từ hai mô hình có thể được kết hợp để tăng độ biểu diễn.
- **Contrastive Learning:**
  - Áp dụng contrastive loss để cải thiện chất lượng biểu diễn.
  - Các cặp ảnh được chọn (giống hoặc khác) nhằm tăng độ tách biệt giữa các loại rác.

### 2.3.3. Mô-đun Phân cụm

- **Thực hiện phân cụm song song bằng 3 thuật toán:** KMeans, BIRCH và Agglomerative Clustering.
- **Voting kết quả:** Dựa trên nhãn phân cụm từ 3 mô hình, sử dụng cơ chế bỏ phiếu để xác định nhãn cuối cùng cho mỗi ảnh

### 3. Thực nghiệm, đánh giá & kết quả

#### 3.1. Thực nghiệm

Thực nghiệm tập trung vào việc kiểm thử quy trình pipeline và đánh giá mô hình phân cụm học không giám sát, được thiết kế nhằm đảm bảo tính **ổn định, chính xác và khả năng ứng dụng thực tế** trong bài toán **phân loại rác thải**. Quá trình thực nghiệm bao gồm hai phần chính:

##### 3.1.1. Tính toàn vẹn của Luồng dữ liệu (Data Pipeline)

Thực hiện kiểm tra toàn diện từng bước trong pipeline xử lý dữ liệu, từ khâu xử lý ảnh, trích xuất đặc trưng, học contrastive đến giai đoạn phân cụm, để đảm bảo rằng dữ liệu không bị sai lệch, thiếu hụt hay hỏng hóc ở bất kỳ bước nào.

- Dữ liệu đầu vào được chuẩn hóa về kích thước, màu sắc và định dạng ảnh để phù hợp với các mô hình học sâu như ResNet50 và ViT.
- Trong giai đoạn trích xuất đặc trưng và học biểu diễn (contrastive learning), các bộ kiểm thử nội bộ được sử dụng để đảm bảo không có lỗi xảy ra trong quá trình tạo vector đặc trưng.
- Dữ liệu sau khi được trích xuất và phân cụm được kiểm tra lại để đảm bảo định dạng đầu ra thống nhất và có thể phân tích được.

##### 3.1.2. Kiểm thử mô hình

Thực nghiệm đánh giá hiệu quả của **phương pháp học không giám sát** được so sánh với các mô hình **học giám sát CNN truyền thống**. Các mô hình học giám sát được huấn luyện bao gồm:

- VGG16, VGG19
- InceptionV2, InceptionV3
- ResNet50, ResNet101
- MobileNet
- DenseNet
- EfficientNetB7

Mô hình học không giám sát sử dụng:

- Trích xuất đặc trưng bằng **ResNet50 và ViT**
- **Contrastive Learning** để cải thiện biểu diễn
- Phân cụm bằng **KMeans, BIRCH, Agglomerative Clustering**
- **Voting** để xác định cụm cuối cùng

Các tiêu chí kiểm thử mô hình bao gồm:

- **Tính ổn định đầu ra:** Đầu ra của mô hình không giám sát phải ổn định khi dữ liệu được nhiễu nhẹ hoặc thay đổi thứ tự ảnh.
- **Tính tổng quát:** Khả năng phân cụm tốt với các loại rác chưa từng gặp trong bộ huấn luyện.
- **Khả năng tái sử dụng nhãn (pseudo-labels)** từ phân cụm làm đầu vào cho các mô hình học giám sát sau này.



### 3.2. Đánh giá

Để đánh giá hiệu quả của mô hình học không giám sát, chúng tôi sử dụng bộ chỉ số **chuẩn mực và có thể so sánh được** với các mô hình học giám sát. Việc đánh giá dựa trên **một bộ dữ liệu có nhãn thực tế** để đối chiếu.

#### 3.2.1 Độ chính xác (Accuracy)

- Được tính bằng công thức:  $(\text{Số lượng ảnh phân loại đúng}) / (\text{Tổng số lượng ảnh})$
- Được dùng để đo hiệu quả tổng thể của việc phân cụm (sau khi ánh xạ cụm về nhãn thật bằng thuật toán như Hungarian Matching).
- Cho phép so sánh trực tiếp giữa mô hình không giám sát và các mô hình học giám sát CNN.

#### 3.2.2 Độ phủ (Recall)

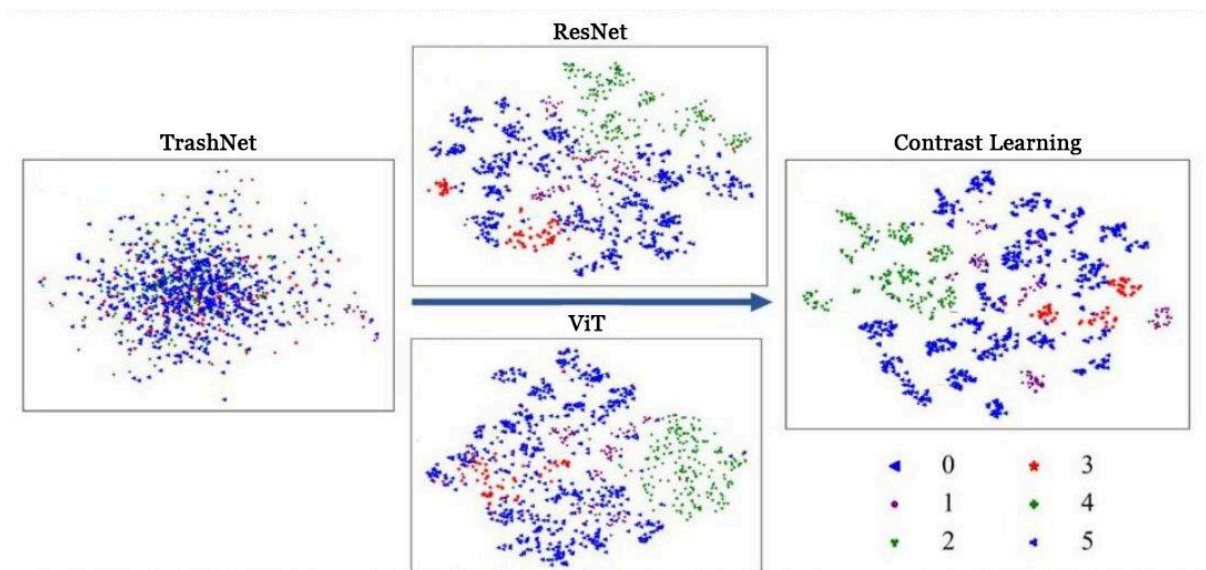
- Được tính bằng công thức:  $(\text{Số lượng ảnh phân loại đúng của một nhãn}) / (\text{Tổng số lượng ảnh của nhãn đó})$
- Giúp đo khả năng mô hình tìm được **đầy đủ** các đối tượng thuộc một lớp cụ thể (ví dụ: rác hữu cơ).
- Đặc biệt hữu ích khi một số loại rác xuất hiện ít hơn và dễ bị bỏ sót.

#### 3.2.3 Data Discarded

- Đây là chỉ số đặc biệt của học không giám sát: Số ảnh không cần gán nhãn mà vẫn được phân cụm hiệu quả
- Cho thấy khả năng **tiết kiệm công sức gán nhãn thủ công** và tăng tính ứng dụng thực tế của mô hình trong các hệ thống phân loại rác quy mô lớn.

### 3.3. Kết quả

- Kết quả thực nghiệm cho thấy rằng việc **kết hợp giữa trích xuất đặc trưng từ các mô hình học sâu (ResNet50, ViT)** với **học tương phản (Contrastive Learning)** giúp cải thiện rõ rệt chất lượng biểu diễn dữ liệu, từ đó nâng cao hiệu quả phân cụm.



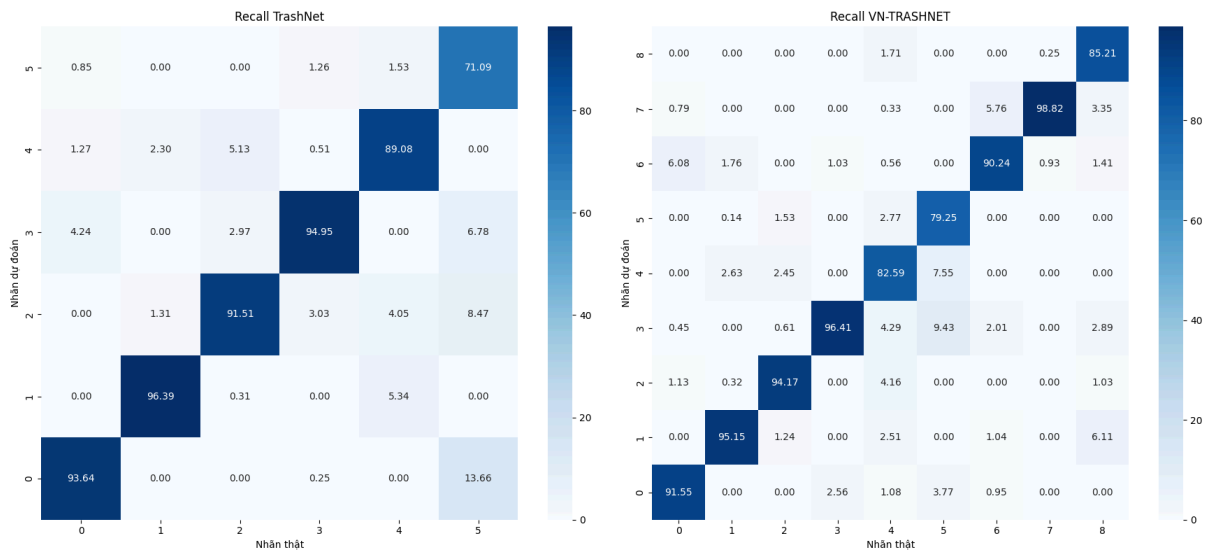
- Mô hình học không giám sát đạt **độ chính xác trên 90%** với bộ dữ liệu TrashNet, vượt qua hiệu suất của nhiều mô hình học giám sát CNN truyền thống như VGG16, MobileNet, ResNet50,...

Model	Train set	Test set	Test Accuracy	Data discarded
Our method	0	1263	91.68%	1084 (85.76%)
ResNet50	1264	1263	89.70%	0
DenseNet	1264	1263	87.80%	0
VGG16	1264	1263	87.33%	0
MobileNet	1264	1263	85.74%	0
InceptionV3	1264	1263	79.57%	0

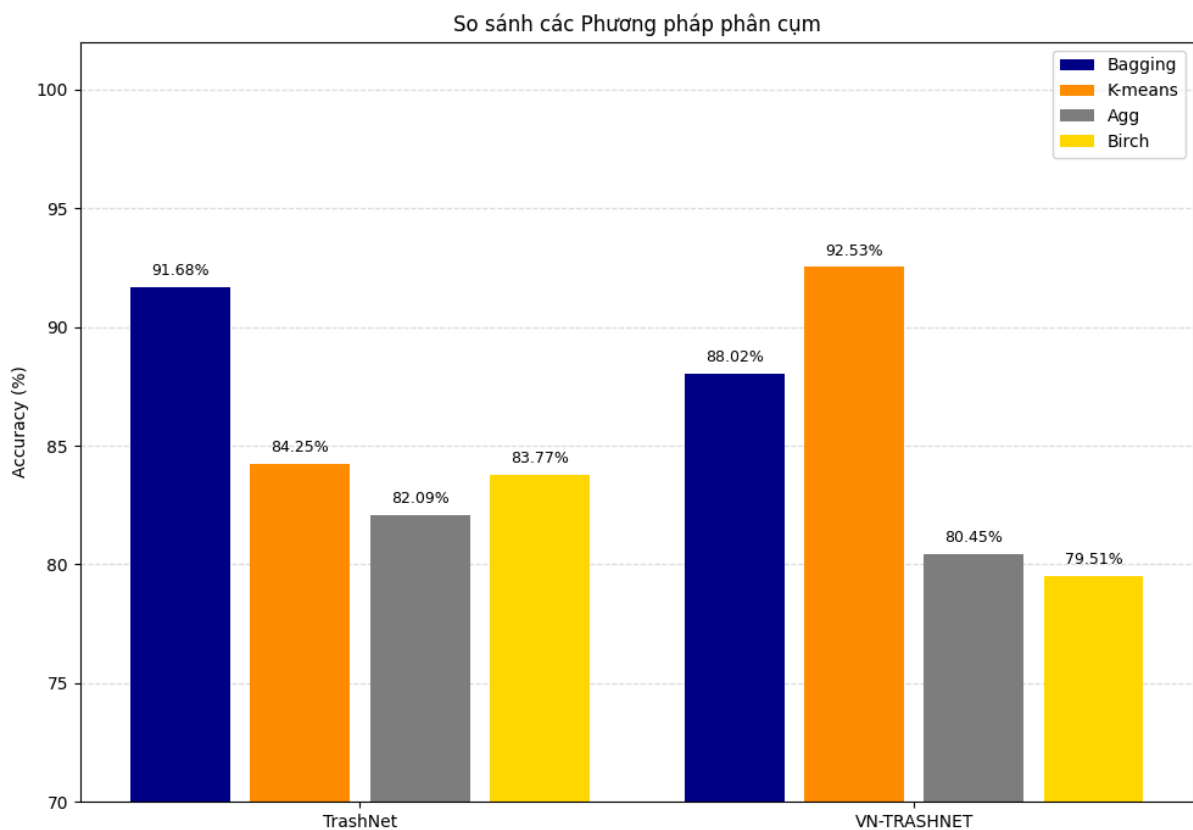
- Với VN-TRASHNET, mô hình không giám sát vẫn duy trì hiệu suất cao, đạt **độ chính xác cao thứ hai**, chỉ sau mô hình ResNet50
- Kết quả này cho thấy rằng phương pháp đề xuất hoàn toàn **cạnh tranh với các mô hình học giám sát**, đặc biệt khi áp dụng cho các bộ dữ liệu phức tạp hơn và có sự đa dạng trong đặc trưng hình ảnh.

Model	Train set	Test set	Test Accuracy	Data discarded
Our method	0	1263	91.68%	1084 (85.76%)
ResNet50	1264	1263	89.70%	0
DenseNet	1264	1263	87.80%	0
VGG16	1264	1263	87.33%	0
MobileNet	1264	1263	85.74%	0
InceptionV3	1264	1263	79.57%	0

- Độ phủ (Recall) của mô hình không giám sát **duy trì ở mức cao** trên hầu hết các nhãn.
- Tuy nhiên, một số nhãn có đặc trưng thị giác tương tự như **glass vs plastic**, hoặc **trash vs paper**, vẫn gây **nhầm lẫn nhẹ**, dẫn đến độ phủ giảm cục bộ. Đây là điều dễ hiểu do không có sự hỗ trợ từ nhãn huấn luyện.



- Việc kết hợp ba thuật toán phân cụm (KMeans, BIRCH, Agglomerative Clustering) cùng với **bỏ phiếu tổ hợp (majority voting)** giúp **tăng độ ổn định và nhất quán của nhãn phân cụm**.
- Cách tiếp cận tổ hợp này giúp **giảm thiểu sai lệch do từng thuật toán phân cụm riêng lẻ**, đồng thời cải thiện hiệu suất tổng thể của hệ thống.



## 4. Kết luận và hướng phát triển

### 4.1. Kết luận

Trong nghiên cứu này, nhóm chúng em đã đề xuất một hệ thống phân loại rác thải sử dụng phương pháp **học không giám sát**, kết hợp các kỹ thuật hiện đại như:

- Trích xuất đặc trưng bằng ResNet50 và Vision Transformer (ViT)
- Tăng cường biểu diễn bằng Contrastive Learning
- Phân cụm với ba mô hình (KMeans, BIRCH, Agglomerative Clustering)
- Voting tổ hợp để tăng độ ổn định

Kết quả thực nghiệm cho thấy mô hình học không giám sát đạt **độ chính xác cao hơn mô hình học giám sát CNN truyền thống**, đặc biệt trên các bộ dữ liệu như TrashNet và VN-TRASHNET. Tuy nhiên:

- Mặc dù hiệu quả cao, mô hình **vẫn chưa thể cạnh tranh hoàn toàn với các phương pháp hiện đại có giám sát** trong điều kiện đầy đủ dữ liệu gán nhãn.
- Học không giám sát **phát huy rõ ưu thế khi số lượng ảnh có nhãn ít hoặc không có nhãn**, giúp giảm đáng kể công sức gán nhãn thủ công.
- Đây là lựa chọn **đặc biệt phù hợp cho các hệ thống thực tế quy mô lớn**, nơi dữ liệu thu thập nhanh chóng nhưng việc gán nhãn gặp nhiều trở ngại.

### 4.2. Tổng kết

Hệ thống đề xuất không chỉ giải quyết tốt bài toán phân loại rác, mà còn đưa ra **một khung giải pháp tổng quát** có thể mở rộng cho các bài toán phân loại ảnh không giám sát nói chung. Cụ thể:

- **Cấu trúc pipeline linh hoạt:** Từ trích xuất đặc trưng, tăng cường biểu diễn, đến phân cụm có thể được tái sử dụng cho các lĩnh vực khác như phân loại y tế, ảnh vệ tinh, hình ảnh công nghiệp,...
- **Hiệu quả về mặt chi phí và tài nguyên:** Không yêu cầu nhãn dữ liệu, phù hợp với các ứng dụng thực tế quy mô lớn hoặc dữ liệu liên tục cập nhật.

### 4.3. Hướng phát triển

Trong tương lai, mô hình có thể được mở rộng và nâng cấp theo một số hướng sau:

- **Kết hợp học không giám sát với học bán giám sát (semi-supervised learning):** Sử dụng kết quả phân cụm làm nhãn tạm để huấn luyện mô hình CNN hiện đại, tận dụng ưu điểm của cả hai hướng tiếp cận.
- **Áp dụng cho các miền dữ liệu khác:** Với nền tảng trích xuất đặc trưng mạnh mẽ và phương pháp tổ hợp phân cụm, hệ thống có thể được áp dụng cho các bài toán phân loại ảnh trong y học, công nghiệp, nông nghiệp...
- **Nâng cao contrastive learning bằng Augmentation nâng cao:** Kết hợp các kỹ thuật như SimCLR hoặc MoCo để cải thiện hơn nữa chất lượng biểu diễn.
- **Tối ưu hóa thời gian suy luận và tài nguyên:** Tinh gọn pipeline để triển khai thực tế trên các thiết bị edge như camera giám sát hoặc thiết bị IoT.

## 5. Đánh giá và Nhận xét của Thành viên Nhóm



Nguyễn  
Trung Kiên



Trần Nam  
Hải



Vũ Văn  
Đông



Vũ Quang  
Khánh

HỌ VÀ TÊN	ĐÁNH GIÁ CÔNG VIỆC
Nguyễn Trung Kiên	Thu thập dữ liệu Hoàn thành tốt việc chuẩn bị slide thuyết trình cũng như thiết kế giao diện cơ bản cho website
Trần Nam Hải	Thiết kế, hoàn thiện cấu trúc tổng thể pipeline Kiểm tra và đánh giá các mô hình tiền huấn luyện dùng để trích xuất đặc trưng ảnh
Vũ Văn Đông	Tìm hiểu và nghiên cứu về các mô hình phân cụm sử dụng trong phương pháp đề xuất Đóng góp tích cực trong việc hoàn thiện slide thuyết trình
Vũ Quang Khánh	Tìm hiểu và chuẩn bị các mô hình CNN truyền thống để sử dụng và so sánh với phương pháp đề xuất Góp phần hoàn thiện website và chuẩn bị kết quả (ảnh, biểu đồ) cho slide thuyết trình

## 6. Đánh giá và Nhận xét của Giảng viên Hướng dẫn

CATEGORY	SCORE	REVIEW and COMMENT
IDEA	___/10	
APPLICATION	___/30	
RESULT	___/30	
PROJECT MANAGEMENT	___/10	
PRESENTATION & REPORT	___/20	
TOTAL	___/100	