

**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**ĐC \* ĐC**



**NHẬP MÔN THỊ GIÁC MÁY TÍNH**

**CS231.P11**

**PHÂN LOẠI THỂ LOẠI PHIM DỰA VÀO  
POSTER PHIM**

**Giảng viên hướng dẫn:** Mai Tiến Dũng

**Sinh viên thực hiện:**

STT	Họ tên	MSSV	Ngành
1	Trần Khôi Nguyên	22520987	KHDL

**TP. HỒ CHÍ MINH – 14/1/2025**

## **LỜI CẢM ƠN**

Chúng em xin trân trọng gửi lời cảm ơn sâu sắc đến TS. Mai Tiến Dũng – Giảng viên Khoa Khoa học Máy tính, Trường Đại học Công nghệ thông tin, Đại học Quốc gia Thành phố Hồ Chí Minh, cũng là giảng viên giảng dạy lớp CS231.P11 – Môn Nhập môn Thị giác máy tính. Cảm ơn thầy đã dành thời gian, tận tâm hướng dẫn và định hướng cho chúng em trong suốt quá trình thực hiện và hoàn thành đồ án này.

Mặc dù chúng em đã nỗ lực hết sức để hoàn thiện đồ án một cách tốt nhất, song vẫn không thể tránh khỏi những thiếu sót ngoài ý muốn. Chúng em rất mong nhận được những nhận xét và góp ý chân thành từ thầy và các bạn, để chương trình ngày càng được hoàn thiện. Mỗi ý kiến đóng góp đều là nguồn động lực lớn lao giúp chúng em cải tiến và phát triển đồ án, từ đó khắc phục những điểm yếu và nâng cao những ưu điểm của chương trình.

## MỤC LỤC

### Nội dung

<b>I. GIỚI THIỆU VÀ TỔNG QUAN ĐỀ TÀI .....</b>	<b>2</b>
I.1. Giới thiệu đề tài .....	2
I.2. Input/ Output của bài toán .....	3
<b>II. PHƯƠNG PHÁP TIẾP CẬN .....</b>	<b>5</b>
II.1. Quy trình thực hiện.....	5
II.2. Giới thiệu các phương pháp rút trích đặc trưng .....	6
II.3. Giới thiệu các mô hình phân loại .....	9
II.4. Giới thiệu các phương pháp phân loại.....	11
<b>III. BỘ DỮ LIỆU.....</b>	<b>15</b>
III.1. Quy trình thu thập dữ liệu .....	15
III.2. Xử lý dữ liệu.....	16
<b>IV. ĐÁNH GIÁ .....</b>	<b>19</b>
IV.1. Thang đo.....	19
IV.2. Kết quả đánh giá.....	20
IV.3. Nhận xét.....	22
<b>PHỤ LỤC PHÂN CÔNG NHIỆM VỤ .....</b>	<b>0</b>

## **I. GIỚI THIỆU VÀ TỔNG QUAN ĐỀ TÀI**

### **I.1. Giới thiệu đề tài**

Đồ án "Phân loại thể loại phim dựa vào poster" được thực hiện trong phạm vi môn học Nhập môn Thị giác máy tính, với mục tiêu áp dụng các kỹ thuật học máy và thị giác máy tính vào việc phân loại tự động thể loại phim dựa trên hình ảnh poster. Với sự phát triển mạnh mẽ của ngành công nghiệp điện ảnh và sự gia tăng số lượng phim được sản xuất mỗi năm, việc phân loại tự động giúp người dùng tiết kiệm thời gian tìm kiếm và lựa chọn phim theo sở thích, đồng thời cũng hỗ trợ các nhà phát hành phim trong việc quản lý và phân phối phim hiệu quả hơn.

Ý tưởng thực hiện đồ án này xuất phát từ nhu cầu thực tiễn của người xem khi phải đối mặt với vô vàn lựa chọn phim mà không biết thể loại hoặc nội dung của mỗi bộ phim. Các poster phim, là yếu tố quan trọng giúp người xem hình dung về nội dung và thể loại của phim, đóng vai trò then chốt trong việc quyết định liệu họ có nên xem bộ phim đó hay không. Bằng cách phân tích hình ảnh từ poster, nhóm chúng em muốn phát triển một hệ thống có thể tự động nhận diện và phân loại các thể loại phim phổ biến như hành động, hài hước, khoa học viễn tưởng, kinh dị, ... giúp nâng cao trải nghiệm người dùng trong việc tìm kiếm phim, cho trải nghiệm xem phim thêm phần thú vị.

Động lực phát triển đồ án này xuất phát từ sự kết hợp giữa niềm đam mê với công nghệ và sự yêu thích điện ảnh. Thị giác máy tính và học sâu ngày nay đã có thể thực hiện các tác vụ phân loại hình ảnh với độ chính xác cao, và chúng em muốn ứng dụng những công nghệ này vào một lĩnh vực cụ thể như điện ảnh, nơi mà việc phân loại và tìm kiếm thông tin rất quan trọng. Hệ thống phân loại tự động sẽ giúp người dùng dễ dàng tìm kiếm thể loại phim yêu thích mà không cần phải xem nhiều trailer (video giới thiệu) hay đọc qua các mô tả dài dòng.

Ứng dụng có thể được ứng dụng nhiều trong các hệ thống khác nhau như:

- Hệ thống gợi ý phim (Recommendation System): Hệ thống sẽ tự động gợi ý các bộ phim liên quan dựa vào những bộ phim yêu thích của người dùng
- Tạo danh sách phim tự động: Các nền tảng có thể sử dụng phân loại thể loại phim để tạo các danh sách phim tự động, như "Top phim hành động," "Phim

hài hước," hoặc "Phim cho gia đình," giúp người dùng dễ dàng chọn lựa mà không phải tốn thời gian tìm kiếm.

- Xây dựng các hệ thống phân loại phim tự động cho các dịch vụ lưu trữ: Các hệ thống lưu trữ và quản lý phim có thể áp dụng công nghệ phân loại để tự động phân loại hàng nghìn bộ phim vào các thể loại khác nhau, giúp dễ dàng quản lý và truy cập dữ liệu.

## **I.2. Input/ Output của bài toán**

Bài toán “Phân loại thể loại phim dựa vào poster phim” được xác định là bài toán phân loại nhiều nhãn (*Multilabel classification*) khi một bộ phim có thể có một hoặc nhiều thể loại kèm theo.

Bài toán được phát biểu với đầu vào và đầu ra như sau:

### **a. Đầu vào**

Gồm một bộ dữ liệu gồm nhiều điểm dữ liệu. Mỗi điểm dữ liệu là một cặp gồm ảnh poster của một bộ phim (được định dạng theo dạng *.png*) và thể loại phim tương ứng với bộ phim đó

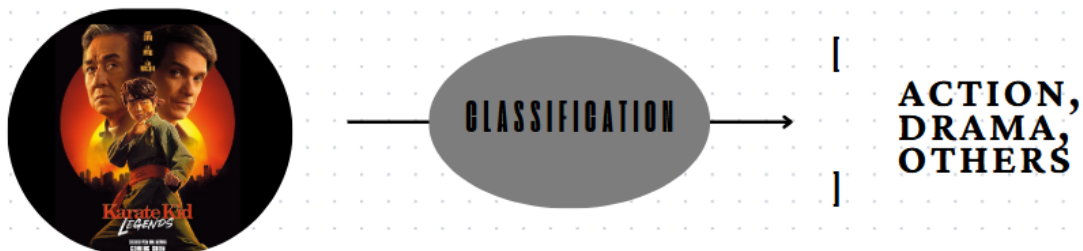
- Thể loại phim được phân bố vào 13 thể loại sau:
  - + Hành động (*Action*): Phim tập trung vào những cảnh hành động mạnh mẽ, thường là chiến đấu, đua xe, hay các cuộc chiến đầy căng thẳng.
  - + Phiêu lưu (*Adventure*): Phim kể về hành trình khám phá những vùng đất mới, trải nghiệm mạo hiểm và giải quyết các thử thách.
  - + Hài (*Comedy*): Phim có nội dung chủ yếu nhằm tạo ra tiếng cười, với các tình huống hài hước và nhân vật gây cười.
  - + Tội phạm (*Crime*): Phim xoay quanh các hoạt động phạm tội, như điều tra vụ án, hành động của các nhân vật tội phạm hoặc lực lượng thi hành pháp luật.
  - + Tài liệu (*Documentary*): Phim mang tính chất ghi lại sự kiện, câu chuyện hoặc cuộc sống thực tế, thường có yếu tố giáo dục và thông tin.
  - + Chính kịch (*Drama*): Phim tập trung vào các mối quan hệ giữa con người, cảm xúc sâu sắc và các tình huống thực tế hoặc giả tưởng.

- + Huyền bí (*Fantasy*): Phim có các yếu tố kỳ ảo, vượt ra ngoài giới hạn của thực tế, thường xuất hiện các sinh vật tưởng tượng và phép thuật.
  - + Kinh dị (*Horror*): Phim nhằm tạo ra cảm giác sợ hãi, hồi hộp, với các yếu tố ma quái, quái vật hoặc tình huống đáng sợ.
  - + Bí Ẩn (*Mystery*): Phim kể về các vụ án hoặc tình huống bí ẩn cần được giải quyết, thường xoay quanh quá trình khám phá sự thật.
  - + Giật gân (*Thriller*): Phim gây căng, kịch tính với các tình huống bất ngờ, mang lại cảm giác lo lắng và căng thẳng.
  - + Lãng mạn (*Romance*): Phim kể về các câu chuyện tình yêu, tình cảm lãng mạn giữa các nhân vật chính.
  - + Viễn tưởng (*Scifi*): Phim dựa trên các yếu tố khoa học giả tưởng, công nghệ tiên tiến, không gian hoặc các vũ trụ tưởng tượng.
  - + Các thể loại khác (*Others*): Các bộ phim không thuộc vào các thể loại không thuộc các thể loại nêu trên, khó phân loại
- Các điểm dữ liệu là các bộ phim không trùng nhau

Gồm một hoặc nhiều posters phim khác nhau, được lấy bất kì hoặc lấy từ bộ dữ liệu được đề cập ở phần trên. Đây là đầu vào của mô hình phân loại

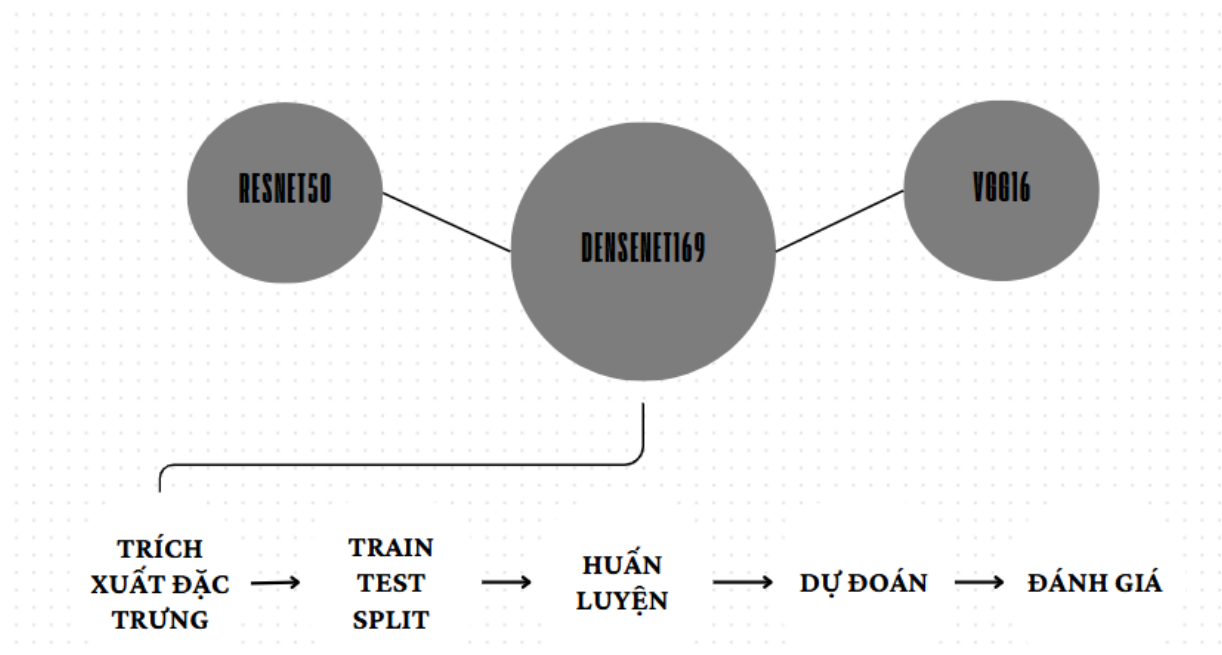
### b. Đầu ra

Đầu ra của mô hình phân loại là kết quả dự đoán của những poster phim đầu vào. Kết quả của mỗi điểm dữ liệu đầu vào là một hoặc nhiều thể loại, được xác định trong 13 thể loại được mô tả ở trên. Một phim có thể có một hoặc nhiều thể loại.



## II. PHƯƠNG PHÁP TIẾP CẬN

### II.1. Quy trình thực hiện



Quy trình thực hiện gồm 5 bước chính, đó là:

- Trích xuất đặc trưng (*Feature Extraction*): Là quá trình sử dụng các mô hình học sâu (Deep Learning) như Densenet168, Resnet50 và VGG16 để rút ra các đặc trưng quan trọng từ hình ảnh poster của phim. Các mô hình này đã được huấn luyện sẵn trên các tập dữ liệu lớn và có khả năng nhận diện các đặc điểm hình ảnh quan trọng, giúp hệ thống hiểu được nội dung và cấu trúc của poster, từ đó phân loại đúng thể loại phim. Áp dụng các mô hình trích xuất đặc trưng lên toàn bộ tập dữ liệu.
- Chia tập dữ liệu (*Data Splitting*): Là bước chia nhỏ tập dữ liệu sau khi rút trích đặc trưng thành các bộ dữ liệu riêng biệt. Bộ dữ liệu được chia thành hai tập: tập huấn luyện (Training Set), tập kiểm tra (Test Set).
- Huấn luyện (*Training*): Quá trình huấn luyện mô hình học máy (machine learning model) với tập dữ liệu huấn luyện. Trong giai đoạn này, mô hình sẽ học cách nhận diện các đặc trưng của hình ảnh (poster phim) và liên kết chúng với các nhãn thể loại phim tương ứng.

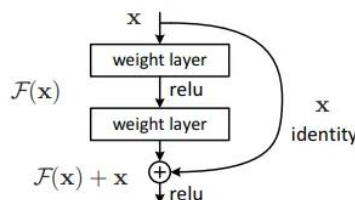
- Dự đoán (*Prediction*): Sau khi mô hình đã được huấn luyện, bước dự đoán sẽ được thực hiện, trong đó mô hình sẽ đưa ra kết quả dự đoán thể loại của các poster phim mới chưa từng thấy trước đó.
- Đánh giá bàn luận (*Evaluation and Discussion*): Đây là bước đánh giá hiệu quả của mô hình. Sau khi dự đoán, kết quả sẽ được so sánh với nhãn thực tế để tính toán độ chính xác của mô hình, thông qua các chỉ số như Accuracy, Precision, Recall, F1-Score, ... Đồng thời, trong bước này, các kết quả sẽ được thảo luận và phân tích để xác định các điểm mạnh, điểm yếu của mô hình, và những cải tiến có thể thực hiện để nâng cao hiệu suất phân loại trong tương lai.

## II.2. Giới thiệu các phương pháp rút trích đặc trưng

Đồ án sử dụng 3 phương pháp rút trích đặc trưng, được tham khảo và thực hiện dựa vào bài báo cáo cho môn học C230 vào mùa đông 2020 của trường đại học Stanford [1].

Các phương pháp trích xuất đặc trưng bao gồm 3 kiến trúc chính là Resnet50, Densenet169, VGG16, Các kiến trúc này đã cho ra các kết quả SOTA trên nhiều tác vụ phân loại hình ảnh hiện nay.

### a. ResNet50



layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				

The architecture of ResNet50 (He et al., 2015)

**Tổng quan:** ResNet (Residual Networks) là một trong những kiến trúc mạng nơ-ron sâu (deep neural networks) được phát triển để giải quyết vấn đề "biến mất gradient" khi huấn



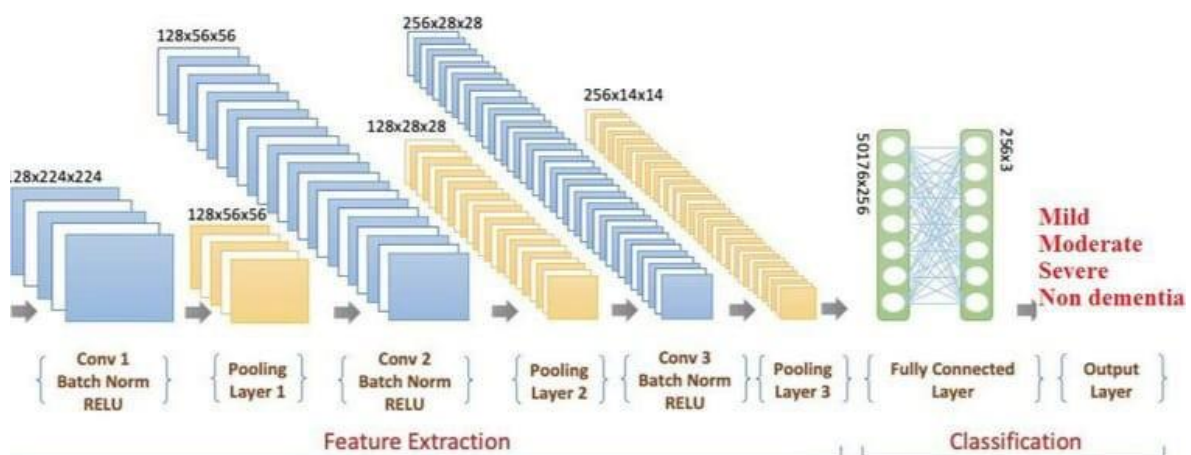
luyện các mô hình sâu. ResNet50 là phiên bản nhẹ của mạng ResNet, với 50 lớp, được tối ưu hóa để học các đặc trưng phức tạp từ dữ liệu hình ảnh.

**Cách thức hoạt động:** ResNet sử dụng các "kết nối dư thừa" (residual connections), cho phép tín hiệu đầu vào đi thẳng vào các lớp sâu mà không bị biến mất trong quá trình lan truyền ngược. Điều này giúp mô hình học tốt hơn và dễ dàng hơn khi được huấn luyện với các mạng sâu.

**Ứng dụng:** Mô hình này rất hiệu quả trong việc phân loại hình ảnh và rút trích đặc trưng. ResNet50 đã được huấn luyện sẵn trên các tập dữ liệu lớn (như ImageNet) và có thể rút ra các đặc trưng hình ảnh mạnh mẽ, giúp phân loại thể loại phim từ poster một cách chính xác.

**Kiến trúc trích xuất đặc trưng:** Đặc trưng hình ảnh được trích xuất từ mô hình Resnet50 sau khi loại bỏ phần classifier cuối cùng. Kết quả thu được là một vector đặc trưng có chiều dài 2048

#### b. DenseNet169



**Tổng quan:** DenseNet (Densely Connected Convolutional Networks) là một kiến trúc mạng nơ-ron sâu khác, trong đó mỗi lớp trong mạng kết nối trực tiếp với tất cả các lớp trước đó, tạo ra một mạng dày đặc. Mô hình DenseNet169 bao gồm 169 lớp và là một trong những phiên bản của DenseNet với độ sâu lớn.

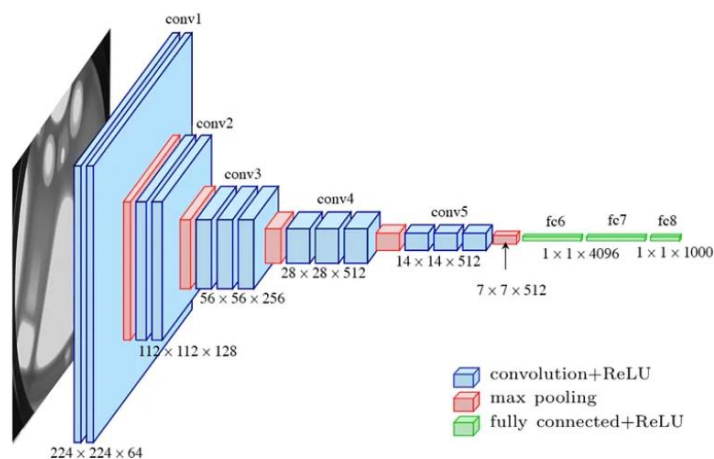
**Cách thức hoạt động:** Các lớp trong DenseNet không chỉ nhận đầu vào từ lớp liền trước mà còn nhận thông tin từ tất cả các lớp trước đó, giúp mô hình tối ưu hóa việc sử dụng

các đặc trưng đã học. Điều này làm cho DenseNet rất hiệu quả trong việc truyền đạt thông tin giữa các lớp, từ đó giảm thiểu vấn đề mất thông tin trong quá trình huấn luyện.

**Ứng dụng:** DenseNet169 thường được sử dụng trong các bài toán phân loại hình ảnh phức tạp, đặc biệt là trong những trường hợp đòi hỏi khả năng khai thác đặc trưng hình ảnh mạnh mẽ từ nhiều lớp sâu, như phân loại thể loại phim dựa vào hình ảnh poster.

**Kiến trúc trích xuất đặc trưng:** Đặc trưng hình ảnh được trích xuất từ mô hình DenseNet169 sau khi loại bỏ phần classifier cuối cùng. Kết quả thu được là một vector đặc trưng có chiều dài 1664.

### c. VGG16



The architecture of VGG16 (Simonyan and Zisserman, 2014)

**Tổng quan:** VGG16 là một kiến trúc mạng nơ-ron sâu được phát triển bởi nhóm nghiên cứu của Đại học Oxford, nổi bật với sự đơn giản và tính hiệu quả trong việc nhận diện hình ảnh. VGG16 có 16 lớp, bao gồm các lớp tích chập (convolutional layers) và lớp kết nối đầy đủ (fully connected layers).

**Cách thức hoạt động:** VGG16 sử dụng các bộ lọc (filters) có kích thước cố định  $3 \times 3$  trong các lớp tích chập, tạo ra một kiến trúc đơn giản nhưng rất hiệu quả trong việc trích xuất đặc trưng hình ảnh. Các lớp sau được kết nối đầy đủ giúp mô hình học các đặc trưng cao cấp từ dữ liệu đầu vào.

**Ứng dụng:** VGG16 đã được huấn luyện trên tập dữ liệu ImageNet và là một trong những mô hình phổ biến trong các bài toán phân loại hình ảnh. Nhờ cấu trúc đơn giản nhưng

mạnh mẽ, VGG16 có thể rút trích đặc trưng hình ảnh từ các poster phim, giúp phân loại thể loại phim một cách chính xác.

**Kiến trúc trích xuất đặc trưng:** Đặc trưng hình ảnh được trích xuất từ mô hình VGG16 sau khi loại bỏ phần classifier cuối cùng. Kết quả thu được là một vector đặc trưng có chiều dài 4096.

### II.3. Giới thiệu các mô hình phân loại

Các mô hình phân loại được áp dụng cho bài toán này bao gồm: Random Forest, Logistic Regression và SVC:

#### a. Random Forest

Random Forest là một mô hình học máy dựa trên kỹ thuật ensemble learning, trong đó nhiều cây quyết định (decision trees) được huấn luyện và kết hợp với nhau để cải thiện hiệu suất phân loại. Mỗi cây trong rừng quyết định dựa trên các mẫu dữ liệu ngẫu nhiên và đưa ra dự đoán, sau đó kết quả cuối cùng được xác định thông qua phương pháp "bỏ phiếu" (voting) hoặc lấy trung bình.

Cách thức hoạt động: Mô hình Random Forest xây dựng nhiều cây quyết định từ các mẫu dữ liệu khác nhau và sử dụng đa dạng trong việc phân chia các đặc trưng. Khi áp dụng cho bài toán phân loại, mỗi cây quyết định sẽ đưa ra một dự đoán riêng, và mô hình sẽ kết hợp các kết quả này để đưa ra phân loại chính xác hơn.

Ưu điểm:

- Khả năng xử lý hiệu quả các vấn đề phân loại phức tạp và đa dạng.
- Giảm thiểu overfitting nhờ sự kết hợp của nhiều cây quyết định.
- Tự động xử lý các đặc trưng thiếu hoặc không đầy đủ trong dữ liệu.

Nhược điểm:

- Cần tài nguyên tính toán lớn, đặc biệt khi số lượng cây trong rừng tăng lên.
- Mô hình có thể phức tạp và khó giải thích khi số lượng cây quá lớn.

Áp dụng trong bài toán phân loại thể loại phim: Random Forest có thể được sử dụng để phân loại các thể loại phim từ dữ liệu đặc trưng hình ảnh, giúp xác định chính xác thể loại mà mỗi bộ phim thuộc về dựa trên các yếu tố đặc trưng học được.

### **b. Logistic Regression**

Logistic Regression là một mô hình phân loại tuyến tính, sử dụng hàm sigmoid để dự đoán xác suất của các lớp phân loại. Mặc dù có tên gọi là "regression" (hồi quy), Logistic Regression thực chất là một mô hình phân loại, vì nó dự đoán xác suất thuộc về một lớp cụ thể và chuyển đổi xác suất đó thành nhãn phân loại.

Cách thức hoạt động: Logistic Regression sử dụng một hàm sigmoid để tính toán xác suất các lớp phân loại. Trong bài toán multi-label classification, mô hình sẽ dự đoán xác suất thuộc về mỗi nhãn thể loại phim một cách độc lập.

Ưu điểm:

- Đơn giản, dễ triển khai và dễ hiểu.
- Hiệu quả với dữ liệu có mối quan hệ tuyến tính hoặc mô hình hóa đơn giản.

Nhược điểm:

- Không hiệu quả với các dữ liệu có mối quan hệ phi tuyến tính phức tạp.
- Khó khăn khi dữ liệu có các đặc trưng không độc lập hoặc quan hệ giữa các đặc trưng phức tạp.

Áp dụng trong bài toán phân loại thể loại phim: Logistic Regression có thể được sử dụng để phân loại các thể loại phim từ dữ liệu đặc trưng hình ảnh, giúp xác định thể loại phim dựa trên các đặc trưng học được, tuy nhiên, mô hình này có thể gặp khó khăn khi các thể loại phim có sự chồng lấp hoặc quan hệ phi tuyến tính giữa các đặc trưng.

### **c. Support Vector Classification (SVC)**

SVC (Support Vector Classification) là một mô hình học máy mạnh mẽ dựa trên thuật toán Support Vector Machines (SVM), nhằm tìm ra siêu phẳng (hyperplane) tối ưu trong không gian đặc trưng để phân chia các lớp dữ liệu. SVC tìm kiếm siêu phẳng sao cho khoảng cách giữa các điểm dữ liệu của các lớp là lớn nhất, giúp phân biệt các lớp hiệu quả hơn.

Cách thức hoạt động: SVC tìm kiếm một siêu phẳng (hoặc siêu đường trong không gian đặc trưng) tối ưu để phân chia các lớp dữ liệu. Trong bài toán phân loại nhiều nhãn (multi-label classification), SVC có thể được áp dụng theo phương pháp One-vs-Rest (OvR), trong đó một mô hình SVC được huấn luyện cho từng nhãn để phân biệt nhãn đó với các nhãn còn lại.

Ưu điểm:

- Mạnh mẽ trong việc phân loại các lớp có mối quan hệ phức tạp, kể cả khi các lớp không tuyến tính.
- Hiệu quả cao khi có không gian đặc trưng rõ ràng và phân biệt tốt giữa các lớp.
- Khả năng hoạt động tốt với các dữ liệu phức tạp và nhiều chiều.

Nhược điểm:

- Cần điều chỉnh tham số kernel phù hợp để đạt hiệu quả tối ưu.
- Tốn kém tài nguyên tính toán đối với các tập dữ liệu lớn.

Áp dụng trong bài toán phân loại thể loại phim: SVC có thể được áp dụng để phân loại các thể loại phim từ dữ liệu đặc trưng học sâu. SVC sẽ phân biệt các thể loại phim một cách chính xác bằng cách tìm ra các ranh giới phân lớp tối ưu giữa các thể loại, ngay cả khi có sự chồng lấp giữa chúng.

## II.4. Giới thiệu các phương pháp phân loại

Sau khi trích xuất được các đặc trưng học sâu (Deep Features) từ hình ảnh dựa vào ba kiến trúc học sâu nói trên, một số mô hình và phương pháp phân loại được áp dụng để thực hiện huấn luyện mô hình trả về kết quả và câu trả lời dựa trên những đặc trưng đã học được.

Các phương pháp được sử dụng để giải quyết bài toán gồm hai phương pháp chính là phân loại nhiều đầu ra (*Multi-output classifier*) và học chuyển giao (*Transfer learning*).

### a. Phân loại nhiều đầu ra (*Multi-output classifier*):

Multi-output classifier là một kỹ thuật cho phép xử lý các bài toán phân loại khi mỗi mẫu dữ liệu có thể có nhiều nhãn (labels). Trong bài toán multi-label classification, mỗi

mẫu có thể có một hoặc nhiều nhãn, vì vậy chúng ta cần mô hình có thể dự đoán nhiều nhãn cho một mẫu dữ liệu duy nhất.

Việc áp dụng Multi-output classifier là rất quan trọng vì các mô hình học máy (*Machine learning*) truyền thống như Random Forest, SVC hay Logistic Regression đều là những mô hình phân loại hai nhãn (*Binary Classification*) hoặc phân loại đa nhãn (*Multiclass Classification*). Vì vậy, để áp dụng các mô hình học máy truyền thống vào bài toán phân loại nhiều nhãn (*Multilabel Classification*), Multi-output classifier như cầu nối để kết nối các hướng tiếp cận này lại với nhau.

Phương pháp sử dụng trong Multi-output classifier là Binary Relevance (*BR*):

- Đây là phương pháp đơn giản nhất. Mỗi nhãn được xử lý như một bài toán phân loại nhị phân riêng biệt. Mỗi nhãn có một mô hình phân loại độc lập, do đó không có sự tương tác giữa các nhãn.
- Mỗi nhãn sẽ có một mô hình phân loại riêng biệt (chẳng hạn sử dụng mô hình học máy), và mỗi mô hình này sẽ trả về một giá trị nhãn (có hoặc không có nhãn).
- Ví dụ: Nếu một bức ảnh có các nhãn là ["hành động", "hài", "giật gân"], các mô hình phân loại sẽ dự đoán từng nhãn một cách độc lập và trả về một bộ kết quả như: [1, 1, 0] (hành động và hài, có mặt, nhưng giật gân thì không có).

Thực hiện phương pháp Multi-Output Classifier lên các mô hình Random Forest, Logistic Regression và SVC với các tham số huấn luyện mặc định của các mô hình:

- Random Forest:
  - + **n\_estimators=100**: Số lượng cây quyết định trong rừng. Mặc định là 100.
  - + **criterion='gini'**: Tiêu chí sử dụng để phân chia các nút trong cây quyết định (có thể là "gini" hoặc "entropy").
  - + **max\_depth=None**: Độ sâu tối đa của cây. Nếu là None, cây sẽ phát triển cho đến khi tất cả các lá đều là các nút thuần nhất hoặc chứa ít mẫu hơn min\_samples\_split.
  - + **min\_samples\_split=2**: Số lượng mẫu tối thiểu để phân chia một nút.
  - + **min\_samples\_leaf=1**: Số lượng mẫu tối thiểu ở mỗi lá của cây.

- + **max\_features='sqrt'**: Số lượng đặc trưng tối đa được lựa chọn tại mỗi phân chia (tính bằng căn bậc hai của tổng số đặc trưng).
- + **bootstrap=True**: Có sử dụng bootstrap (lấy mẫu với lại) để xây dựng cây quyết định hay không.
- Logistic Regression:
  - + **penalty='l2'**: Kiểu phạt L2 cho mô hình, dùng để giảm độ phức tạp và tránh overfitting.
  - + **dual=False**: Không sử dụng bài toán đối ngẫu cho các tham số tuyến tính.
  - + **C=1.0**: Tham số điều chỉnh độ mạnh của việc phạt (Regularization). Giá trị nhỏ giúp giảm độ phức tạp của mô hình, trong khi giá trị lớn sẽ làm mô hình phù hợp hơn với dữ liệu huấn luyện.
  - + **solver='lbfgs'**: Thuật toán tối ưu hóa sử dụng (L-BFGS là một thuật toán tối ưu hóa dựa trên gradient).
  - + **max\_iter=100**: Số lần lặp tối đa trong quá trình huấn luyện.
  - + **multi\_class='deprecated'**: Cách thức phân loại đa lớp. Lưu ý rằng đối với multi-label classification, chúng ta sẽ sử dụng **One-vs-Rest** (OvR).
- SVC:
  - + **C=1.0**: Tham số điều chỉnh độ cứng của siêu phẳng phân loại. Giá trị cao hơn làm mô hình ít lỗi hơn trong huấn luyện, nhưng có thể gây overfitting.
  - + **kernel='rbf'**: Sử dụng kernel Gaussian Radial Basis Function (RBF) để xử lý các mối quan hệ phi tuyến tính.
  - + **degree=3**: Bậc của kernel khi sử dụng kernel đa thức.
  - + **coef0=0.0**: Hệ số điều chỉnh trong kernel cho SVC (sử dụng khi kernel là polynomial hoặc sigmoid).
  - + **decision\_function\_shape='ovr'**: Cách phân loại cho nhiều nhãn, trong đó One-vs-Rest (OvR) phân loại mỗi nhãn độc lập.

*Các giá trị tham số được lấy từ trang thông tin về mô hình của sklearn*

## b. Học chuyển giao (*Transfer Learning*)

Transfer learning là một phương pháp trong học máy cho phép tận dụng các mô hình đã được huấn luyện trên một tập dữ liệu lớn (chẳng hạn như ImageNet) để áp dụng cho các bài toán khác, đặc biệt là những bài toán có ít dữ liệu hơn. Khi sử dụng phương pháp học chuyển giao, ta có thể tận dụng các đặc trưng đã được mô hình học từ tập dữ liệu lớn và điều chỉnh mô hình đó sao cho phù hợp với bài toán mới.

Phương pháp học chuyển giao trong bài báo cáo này được tham khảo từ tài liệu của môn học C230, trường Đại học Stanford, mùa đông 2020 [1], tuy nhiên đã có một số thay đổi. Theo tài liệu gốc [1], nhóm nghiên cứu đã thực hiện fine-tuning toàn bộ các tham số của các mô hình ResNet50, DenseNet169 và VGG16, dẫn đến tổng số tham số cần huấn luyện lần lượt là 25.764.871, 14.321.543 và 15.372.103. Tuy nhiên, do số lượng tham số quá lớn và tài nguyên huấn luyện có hạn, nhóm quyết định thay vì huấn luyện toàn bộ tham số, sẽ chỉ huấn luyện lớp classifier của ba mô hình này. Cấu trúc của lớp classifier bao gồm ba lớp fully connected với các kích thước lần lượt là 1024, 128 và 13, được tham khảo từ bài báo cáo [1]. Các hàm kích hoạt sử dụng cho các lớp này lần lượt là ReLU, ReLU và Sigmoid. Tổng số tham số cần huấn luyện cho ba mô hình này sau khi điều chỉnh là: 2.231.053 cho ResNet50, 1.837.837 cho DenseNet169 và 4.328.205 cho VGG16.

Mô hình được huấn luyện với số epochs=10, batch\_size=32, optimizer=Adam

Hàm mất mát được thiết kế dựa trên hàm mất mát được nêu trong bài báo *sigmoidF1: A Smooth F1 Score Surrogate Loss for Multilabel Classification* [2] được công bố vào năm 2021, có công thức là:

$$\text{Soft-F1} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN} + \epsilon}$$

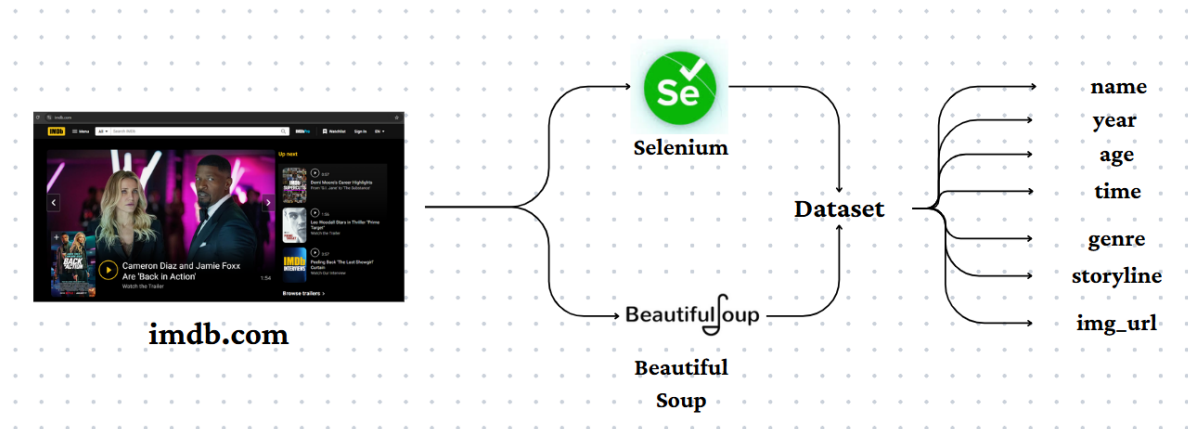
Where:

- True Positives (TP) =  $\sum y_{\text{hat}} \cdot y$  (overlap between predictions and ground truth)
- False Positives (FP) =  $\sum y_{\text{hat}} \cdot (1 - y)$  (predicted positive but not actually positive)
- False Negatives (FN) =  $\sum (1 - y_{\text{hat}}) \cdot y$  (actual positive but not predicted as such)



### III. BỘ DỮ LIỆU

#### III.1. Quy trình thu thập dữ liệu



Hình 1: Quy trình thu thập dữ liệu

Dữ liệu được thu thập từ trang web imdb.com theo 6 chủ đề chính là: Documentary, Drama, Horror, Mystery, Romance, Scifi, Thriller, được thu thập sử dụng Selenium và BeautifulSoup. Bộ dữ liệu cuối cùng gồm 7 thuộc tính chính là:

- Name: Tên bộ phim
- Year: Năm sản xuất
- Age: Giới hạn độ tuổi
- Time: Thời lượng phim
- Genre: Thể loại phim
- Storyline: Cốt truyện
- Img\_url: Link ảnh poster phim

Sau khi tải các ảnh cần thiết dựa vào img\_url, nhóm nhận thấy việc sử dụng toàn bộ thuộc tính là không cần thiết. Vì thế, chỉ 2 thuộc tính chính là genre và img\_url được sử dụng để đưa vào mô hình phân loại.

Bộ dữ liệu ban đầu gồm 19525 bộ phim trải dài trong 104 thể loại khác nhau (Mặc dù được thu thập trên 6 thể loại chính nhưng các bộ phim có nhiều hơn một thể loại ngoài thể loại chính có thể có những thể loại khác làm tăng số lượng thể loại của bộ dữ liệu).

## III.2. Xử lý dữ liệu

### a. Chuẩn hóa thể loại

Do số lượng thể loại ở bộ dữ liệu cuối cùng vô cùng lớn (104 thể loại), khiến cho việc phân loại chính xác 104 thể loại gần như là không thể, vì thế nhóm hướng đến ý tưởng đưa những thể loại có tần suất xuất hiện thấp thành một nhãn duy nhất (Others)

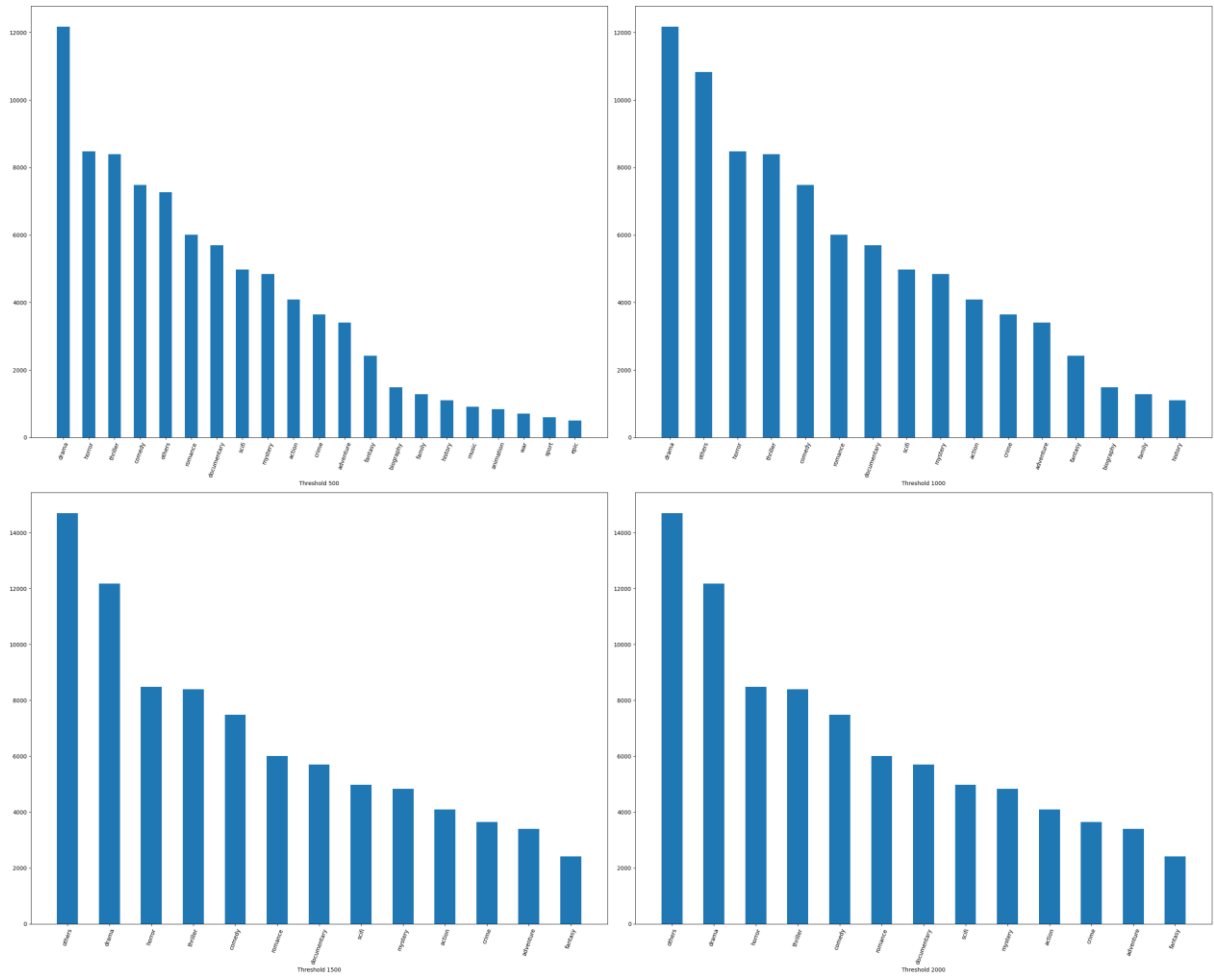
Trước hết, tất cả các thể loại đều được chuẩn hóa thành chữ thường (Lowercase) và loại bỏ hết các dấu phụ đặc biệt (Remove Punctuation) như : -, \_ , ...

Tiếp đến, qua quá trình khảo sát dữ liệu, nhóm nhận thấy một thể loại chính sẽ có 5-6 thể loại phụ, ví dụ như:

- Documentary: History Documentary, Art Documentary, ...
- History: War History, Musical History, ...
- ...

Nhóm thực hiện tìm và lọc các thể loại con như trên thành một thể loại chính duy nhất.

Sau đó, nhóm thử nghiệm trên từng threshold khác nhau. Threshold là tham số chỉ ngưỡng của tần suất xuất hiện của các nhãn. Nếu nhãn nào có tần suất xuất hiện bé hơn Threshold ( $x < thres$ ) thì sẽ được gán bằng nhãn Others, ngược lại vẫn giữ nguyên nhãn đó. Biểu đồ thể hiện phân bố các nhãn với các mức threshold khác nhau như sau



Ta có thể thấy, với mức threshold 1500, các nhãn được phân bố đồng đều và hợp lý nhất, vì thế mức threshold được chọn là 1500.

Tập dữ liệu sau khi được xử lý sẽ có 13 nhãn chính bao gồm : Hành động (*Action*), Phiêu lưu (*Adventure*), Hài (*Comedy*), Tội phạm (*Crime*), Tài liệu (*Documentary*), Chính kịch (*Drama*), Huyền bí (*Fantasy*), Kinh dị (*Horror*), Bí Ẩn (*Mystery*), Giật gân (*Thriller*), Lãng mạn (*Romance*), Viễn tưởng (*Sci-fi*), Các thể loại khác (*Others*)

### b. Train Test Split

Để đảm bảo sự phân bố đồng đều của các nhãn trong mỗi tập Train Set và Test Set, nhóm sử dụng hàm `iterative_train_test_split` của thư viện `sklearn`. Phương pháp này là một kỹ thuật phân chia dữ liệu giúp giữ tỷ lệ phân phối của các lớp (classes) trong các tập dữ liệu huấn luyện và kiểm tra giống nhau, đồng thời vẫn giữ được sự phân bố đa dạng giữa các lớp (class distribution) cho mỗi tập. Điều này đặc biệt hữu ích trong các bài toán phân loại với dữ liệu mất cân bằng (imbalanced datasets). Do vấn đề thiếu hụt

tài nguyên huấn luyện, tập Training Set và Testing Set có số lượng điểm dữ liệu lần lượt là: 4861 và 2511, tương ứng với lần lượt 35% và 20% tập dữ liệu ban đầu.

## IV. ĐÁNH GIÁ

### IV.1. Thang đo

Các thang đo sử dụng để đánh giá bài toán này bao gồm: F1 Macro, Precision Average Macro, Recall Average Macro, Hamming Loss

#### a. F1 Macro

$$F_1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

**F1 Score** là một chỉ số tổng hợp giữa độ chính xác (Precision) và độ phủ (Recall)

**F1 Macro** là một chỉ số hữu ích khi các lớp trong bài toán phân loại có tỷ lệ phân bố không đều (imbalanced). Khi dữ liệu bị mất cân bằng, F1 Macro sẽ cân nhắc hiệu suất của mô hình đối với tất cả các lớp mà không thiên về lớp lớn (thường gặp phải trong các thang đo như accuracy).

#### b. Precision Average Macro



The diagram illustrates the calculation of Precision and its macro-average. On the left, the formula for Precision for a specific class A is shown:  $\text{Precision}_{\text{Class A}} = \frac{TP_{\text{Class A}}}{TP_{\text{Class A}} + FP_{\text{Class A}}}$ . On the right, the formula for the Macro-average of Precision is shown:  $\text{Precision}_{\text{Macro-average}} = \frac{\text{Precision}_{\text{Class A}} + \text{Precision}_{\text{Class B}} + \dots + \text{Precision}_{\text{Class N}}}{N}$ .

**Precision Macro** cung cấp thông tin về mức độ chính xác của mô hình trong việc dự đoán đúng các nhãn (classes), và giống như F1 Macro, việc tính trung bình của tất cả các lớp giúp giảm thiểu sự thiên lệch đối với lớp chiếm ưu thế. Do đó, nó là một chỉ số hiệu quả khi đánh giá mô hình trong trường hợp dữ liệu mất cân bằng.

#### c. Recall Average Macro



The diagram illustrates the calculation of Recall and its macro-average. On the left, the formula for Recall for a specific class A is shown:  $\text{Recall}_{\text{Class A}} = \frac{TP_{\text{Class A}}}{TP_{\text{Class A}} + FN_{\text{Class A}}}$ . On the right, the formula for the Macro-average of Recall is shown:  $\text{Recall}_{\text{Macro-average}} = \frac{\text{Recall}_{\text{Class A}} + \text{Recall}_{\text{Class B}} + \dots + \text{Recall}_{\text{Class N}}}{N}$ .

**Recall Macro** giúp đánh giá mức độ mô hình có thể nhận diện được tất cả các nhãn trong dữ liệu, bao gồm các lớp ít xuất hiện. Khi dữ liệu mất cân bằng, mô

hình có thể dễ dàng bỏ sót các lớp ít xuất hiện, do đó Recall Macro giúp đảm bảo rằng các lớp thiểu số không bị bỏ qua.

#### d. Hamming Loss

$$\text{Hamming Loss} = \frac{1}{nL} \sum_{i=1}^n \sum_{j=1}^L I(y_i^j \neq \hat{y}_i^j)$$

**Hamming Loss** đo lường số lượng nhãn sai trong dự đoán của mô hình, tính bằng tỷ lệ số lượng nhãn dự đoán sai trên tổng số nhãn.

Hamming Loss tính toán mức độ sai sót trong dự đoán của mô hình ở mỗi nhãn, có thể áp dụng cho mỗi nhãn trong một bài toán phân loại đa nhãn.

## IV.2. Kết quả đánh giá

Kết quả đánh giá trên các thang đo bằng các phương pháp phân loại và classification report tương ứng cho mô hình cho kết quả tốt nhất của từng phương pháp là như sau:

	VGG				DENSENET169				RESNET			
Metrics	f1_score (avg)	precision (avg)	recall (avg)	hamming loss	f1_score (avg)	precision (avg)	recall (avg)	hamming loss	f1_score (avg)	precision (avg)	recall (avg)	hamming loss
RANDOM FOREST	<u>0.35</u>	0.31	0.44	0.32	0.27	0.23	0.37	0.38	0.29	0.23	0.42	0.4
LOGISTIC REGRESSION	0.18	0.36	0.13	0.23	0.1	0.34	0.07	0.23	0.08	0.35	0.05	0.23
SVC	0.1	0.36	0.07	0.22	0.07	0.32	0.05	0.22	0.07	0.45	0.05	0.22

*Kết quả đánh giá cho phương pháp MultiOutput Classifier*

		precision	recall	f1-score	support	
I	ACTION	0	0.33	0.27	0.30	475
	ADVENTURE	1	0.25	0.12	0.17	318
	COMEDY	2	0.43	0.55	0.48	640
	CRIME	3	0.23	0.13	0.17	375
	DOCUMENTARY	4	0.33	0.49	0.39	508
	DRAMA	5	0.43	0.95	0.59	1051
	FANTASY	6	0.07	0.02	0.03	226
	HORROR	7	0.33	0.58	0.42	569
	MYSTERY	8	0.20	0.14	0.17	463
	THRILLER	9	0.35	0.86	0.50	850
	ROMANCE	10	0.42	0.56	0.48	600
	SCIFI	11	0.38	0.35	0.36	525
	OTHERS	12	0.35	0.73	0.47	749
I	micro avg	0.37	0.55	0.44	7349	
	macro avg	0.31	0.44	0.35	7349	
	weighted avg	0.34	0.55	0.40	7349	
	samples avg	0.38	0.56	0.42	7349	

*Classification Report cho mô hình đạt kết quả tốt nhất của phương pháp MultiOutput Classifier*

	VGG				DENSENET169				RESNET			
Metrics	f1_score (avg)	precision (avg)	recall (avg)	hamming loss	f1_score (avg)	precision (avg)	recall (avg)	hamming loss	f1_score (avg)	precision (avg)	recall (avg)	hamming loss
thres = 0.25	0.4	0.36	0.53	0.32	0.38	0.27	0.76	0.54	0.37	0.26	0.86	0.63
thres = 0.5	0.37	0.4	0.43	0.29	0.37	0.29	0.71	0.5	0.37	0.26	0.84	0.63
thres = 0.75	0.32	0.43	0.34	0.27	0.36	0.31	0.64	0.47	0.37	0.27	0.83	0.62

*Kết quả đánh giá cho phương pháp Transfer Learning*

		precision	recall	f1-score	support	
I	ACTION	0	0.40	0.44	0.42	475
	ADVENTURE	1	0.34	0.21	0.26	318
	COMEDY	2	0.47	0.51	0.49	641
	CRIME	3	0.25	0.27	0.26	375
	DOCUMENTARY	4	0.37	0.41	0.39	508
	DRAMA	5	0.42	1.00	0.59	1051
	FANTASY	6	0.23	0.06	0.10	226
	HORROR	7	0.43	0.55	0.48	569
	MYSTERY	8	0.22	0.64	0.33	463
	THRILLER	9	0.34	1.00	0.51	850
	ROMANCE	10	0.46	0.56	0.51	600
	SCIFI	11	0.43	0.43	0.43	525
	OTHERS	12	0.36	0.78	0.50	750
I	micro avg	0.37	0.62	0.46	7351	
	macro avg	0.36	0.53	0.40	7351	
	weighted avg	0.38	0.62	0.45	7351	
	samples avg	0.37	0.62	0.44	7351	

*Classification Report cho mô hình đạt kết quả tốt nhất của phương Transfer Learning*

### IV.3. Nhận xét

Dựa trên bảng kết quả, có thể thấy rằng mô hình phân loại hiệu quả nhất là phương pháp **MultiOutput Classifier** kết hợp với kiến trúc **VGG16** và **Random Forest**. Đồng thời, phương pháp **Transfer Learning** với kiến trúc **VGG16**, sử dụng mức chấp nhận (threshold) là 0.25, cũng cho kết quả phân loại tốt nhất. Dưới đây là các nhận xét:

#### a. MultiOutput Classifier

**FANTASY, ADVENTURE, CRIME, MYSTERY:** Hiệu suất của các thể loại này khá thấp (F1-Score < 0.2) do dữ liệu huấn luyện cho bốn nhãn này ít, khiến cho mô hình không học được đầy đủ đặc trưng của các thể loại này. Điều này dẫn đến việc mô hình khó phân loại chính xác các thể loại ít xuất hiện trong dữ liệu.

**COMEDY, DRAMA, THRILLER, ROMANCE, OTHERS:** Các thể loại này có hiệu suất cao nhất trong tất cả các thể loại (với Recall cao), nhờ vào lượng dữ liệu huấn luyện phong phú. Điều này giúp mô hình học được nhiều đặc trưng của các thể loại này. Tuy nhiên, sự phân bố dữ liệu lệch về các thể loại này có thể dẫn đến hiện tượng bias, khiến mô hình có xu hướng thiên về các lớp chiếm ưu thế và bỏ sót các lớp thiểu số.



Sự thay đổi này giúp làm rõ hơn lý do tại sao các thể loại có ít dữ liệu học lại gặp khó khăn trong việc huấn luyện, đồng thời nêu bật sự thiên lệch có thể xảy ra khi phân bố dữ liệu không đồng đều.

### **b. Transfer Learning**

**FANTASY, ADVENTURE, CRIME:** Hiệu suất của các thể loại này rất thấp ( $F1\text{-Score} < 0.3$ ) do lượng dữ liệu huấn luyện cho ba nhãn này ít, khiến cho mô hình không thể học đầy đủ các đặc trưng của các thể loại này. Vì vậy, mô hình gặp khó khăn trong việc phân loại chính xác các thể loại ít xuất hiện.

**COMEDY, DRAMA, THRILLER, ROMANCE, HORROR, OTHERS:** Các thể loại này có hiệu suất cao nhất trong tất cả các thể loại, nhờ vào lượng dữ liệu huấn luyện phong phú. Điều này giúp mô hình học được nhiều đặc trưng của các thể loại này, với Recall cao. Tuy nhiên, sự phân bố lệch về các thể loại này có thể dẫn đến hiện tượng bias, khiến mô hình thiên về các lớp chiếm ưu thế.

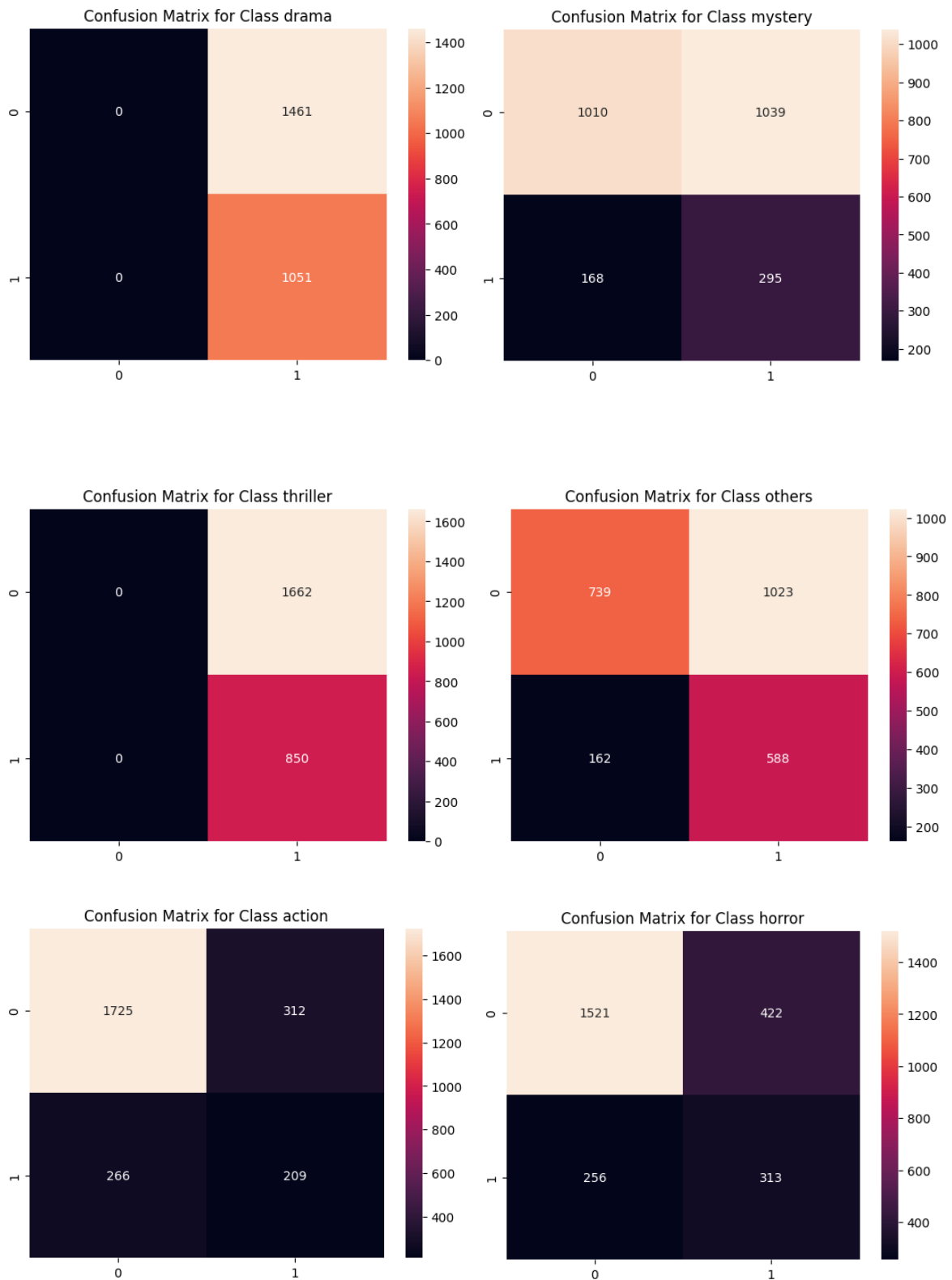
**SO SÁNH:** Mô hình học sâu cho thấy hiệu suất tốt hơn khi so với các mô hình máy học truyền thống, nhờ khả năng học và trích xuất đặc trưng sâu từ dữ liệu, đặc biệt là với các thể loại có lượng dữ liệu huấn luyện lớn.

Nhận xét đã được sửa để làm rõ hơn vấn đề phân bố dữ liệu và sự khác biệt về hiệu suất giữa các thể loại, cũng như so sánh với các mô hình truyền thống.

### **c. Nhận xét tổng quát**

Sau khi xem xét sâu hơn vào khả năng dự đoán và nhãn thực tế các mô hình dự đoán được mỗi mô hình, nhóm chúng em rút ra được những nhận xét tiêu biểu như sau:

Có thể thấy, ở cả hai mô hình tốt nhất của cả hai phương pháp, các nhãn Drama, Thriller, Others, mặc dù có tần suất xuất hiện của các nhãn cao hơn so với các nhãn còn lại, tuy nhiên mô hình vẫn không thể học được đặc trưng riêng biệt và phân loại đúng với các loại thể loại này. Điều này có thể do một số yếu tố sau:



*Confusion Matrix của mô hình VGG16 sử dụng kỹ thuật **Transfer Learning** với mức chấp nhận 0.25*



*Confusion Matrix của mô hình VGG16 sử dụng kỹ thuật **Multi Output Classifier** với **RandomForest***

- Việc kết hợp các thể loại ít xuất hiện thành nhãn *Others* dẫn đến việc các hình ảnh thuộc nhãn này không có sự đồng nhất, vì các thể loại trong nhãn này có

đặc điểm khác nhau. Khi các thể loại ít xuất hiện (như tài liệu, hoạt hình, vv.) bị gom chung vào nhãn *Others*, mô hình không thể nhận diện được đặc điểm riêng biệt của từng thể loại. Điều này khiến mô hình khó học được đặc trưng của nhãn *Others* vì sự pha trộn quá lớn giữa các thể loại khác nhau, từ đó ảnh hưởng đến khả năng phân loại chính xác.

- Đối với nhãn Thriller, hầu hết các điểm dữ liệu tồn tại hai nhãn này đều bị dự đoán sai. Phim Thriller có xu hướng đồng thời mang nhãn Action (vì nhiều cảnh hành động trong phim Thriller), dẫn đến mô hình phân loại bị thiên lệch. Mô hình sẽ có xu hướng phân loại các phim này vào nhãn Action nhiều hơn vì chúng có điểm tương đồng về hành động. Mặt khác, phim Thriller thường có phong cách thiết kế với tông màu tối chủ đạo, điều này dễ dẫn đến việc nhầm lẫn với thể loại Horror, vốn cũng có thiết kế tối màu. Sự trùng lặp trong cách thiết kế này làm cho mô hình phân loại không chính xác, vì các đặc điểm hình ảnh của Thriller bị đánh giá tương tự như của Horror.
- Đối với nhãn Drama, poster phim Drama thường sử dụng tông màu sáng và tươi, dẫn đến sự nhầm lẫn với thể loại Romance. Poster phim Drama có xu hướng sử dụng các tông màu sáng, nổi bật, dễ gây nhầm lẫn với thể loại Romance, vốn cũng hay sử dụng tông màu nhẹ nhàng và sáng. Sự tương đồng về thiết kế này khiến mô hình phân loại không phân biệt rõ ràng được giữa Drama và Romance, từ đó làm giảm độ chính xác của dự đoán.

## TÀI LIỆU THAM KHẢO

[http://cs230.stanford.edu/projects\\_winter\\_2020/reports/32643471.pdf](http://cs230.stanford.edu/projects_winter_2020/reports/32643471.pdf)

<https://arxiv.org/abs/2108.10566>

<https://scikit-learn.org/stable/>

<https://scikit-learn.org/1.5/modules/multiclass.html>

<https://github.com/d-misra/Multi-label-movie-poster-genre-classification>

## PHỤ LỤC PHÂN CÔNG NHIỆM VỤ

STT	Thành viên	Nhiệm vụ
1	Trần Khôi Nguyên	Làm hết