

# GIỚI THIỆU NHÓM VÀ ĐỀ TÀI

## Thành viên nhóm:

Bùi Khương Duy  
Đào Anh Dũng  
Trần Hiếu Nghĩa

**Đề tài :** MINI PROJECT: AIR GUARD – DỰ BÁO PM2.5 VÀ CẢNH BÁO AQI THEO TRẠM

## 1. Giới thiệu bài toán (Introduction)

- **Bối cảnh:** Vấn đề ô nhiễm không khí và sự cần thiết của việc dự báo AQI trong bối cảnh đô thị hóa.
- **Thách thức:** Các mô hình giám sát truyền thống cần nhiều dữ liệu có nhãn, nhưng việc gán nhãn thực tế tốn kém và khó khăn.
- **Mục tiêu dự án:** Áp dụng kỹ thuật Self-training và Co-training để tận dụng dữ liệu chưa gán nhãn (unlabeled data) nhằm cải thiện độ chính xác dự báo so với mô hình giám sát truyền thống.

## 2. Pipeline xử lý dữ liệu (Data Pipeline)

- **Dữ liệu đầu vào:** Mô tả sơ bộ về bộ dữ liệu (số lượng mẫu, các trạm).
- **Tiền xử lý:**
  - Làm sạch dữ liệu (xử lý missing value, outlier) và tái định dạng thời gian.
  - **Phân chia dữ liệu (Quan trọng):** Chọn mốc cutoff là **2017-01-01**. Dữ liệu trước 2017 dùng cho Train/Unlabeled, dữ liệu từ 2017 trở đi dùng cho Test để tránh rò rỉ dữ liệu tương lai.
- **Gán nhãn AQI:** Chuyển đổi PM2.5 sang 6 mức AQI (Good, Moderate, ..., Hazardous) .
- **Tách tập có nhãn/không nhãn:** Sử dụng phương pháp ngẫu nhiên có kiểm soát (giữ lại X% mẫu có nhãn trong tập trước 2017, còn lại bỏ nhãn) để đảm bảo phân phối lớp.

## 3. Thực nghiệm & Kết quả (Experiments & Results)

### 3.1. Mô hình cơ sở (Supervised Baseline)

- **Cấu hình:** Sử dụng HistGradientBoostingClassifier huấn luyện chỉ trên tập dữ liệu nhỏ có nhãn ban đầu<sup>9</sup>.
- **Kết quả Benchmark:**
  - *Bảng kết quả:* Accuracy, Macro-F1 trên tập Validation và Test.
  - *Nhận xét:* Đây là mốc để đánh giá hiệu quả của các phương pháp bán giám sát<sup>10</sup>.

### 3.2. Thuật toán Self-Training

Cấu hình	Ngưỡng ( $\tau$ )	Số nhãn giả (Vòng 1)	Accuracy (Test)	Macro-F1 (Test)	Nhận xét
Thấp	0.75	188,070 (Rất nhiều)	58.12%	510	Nhiều nhiễu, độ chính xác giảm so với mức tối ưu.
Tối ưu	0.90	76,134	58.90%	534	Kết quả tốt nhất. Cân bằng giữa lượng mẫu và độ sạch.
Cao	0.98	9,245 (Rất ít)	58.51%	513	Quá thận trọng, mô hình học được ít thông tin mới.

### Phân tích độ nhạy của tham số ngưỡng

Nhóm thực hiện thử nghiệm với 3 mức ngưỡng tin cậy để tìm ra điểm cân bằng giữa "số lượng" và "chất lượng" nhãn giả:

- **Tại  $\tau = 0.75$  (Ngưỡng lỏng):** Mô hình gán nhãn rất hào phóng (gần 190.000 mẫu ngay vòng 1). Tuy nhiên, độ chính xác (Accuracy: 58.1%) thấp hơn mức tối ưu. Điều này cho thấy việc đưa quá nhiều dữ liệu

nhiều (noisy labels) vào tập train đã gây hại cho mô hình (hiện tượng *Semantic Drift*).

- **Tại  $\tau = 0.98$  (Ngưỡng chặt):** Mô hình quá thận trọng, chỉ gán được khoảng 9.200 mẫu. Mô hình không học thêm được nhiều kiến thức mới, dẫn đến hiệu năng (Accuracy: 58.5%) chỉ tương đương Baseline.
- **Tại  $\tau = 0.90$  (Ngưỡng tối ưu):** Đạt kết quả tốt nhất với **Accuracy 58.9%** và **F1-Macro 0.534**. Số lượng nhãn giả khoảng 76.000 mẫu là vừa đủ để cung cấp thông tin bổ sung mà không làm loãng dữ liệu gốc.

**BẢNG 2: F1-SCORE CHI TIẾT TỪNG LỚP (ĐỂ XEM LỚP HƯỞNG LỢI)**

	Self-0.75	Self-0.90	Self-0.98
Good	0.33032	0.4897	0.36455
Hazardous	0.64961	0.67624	0.65431
Moderate	0.69911	0.7045	0.69619
Unhealthy	0.58029	0.58766	0.60535
Unhealthy_for_Sensitive_Groups	0.22729	0.17892	0.21042
Very_Unhealthy	0.5709	0.56894	0.55014

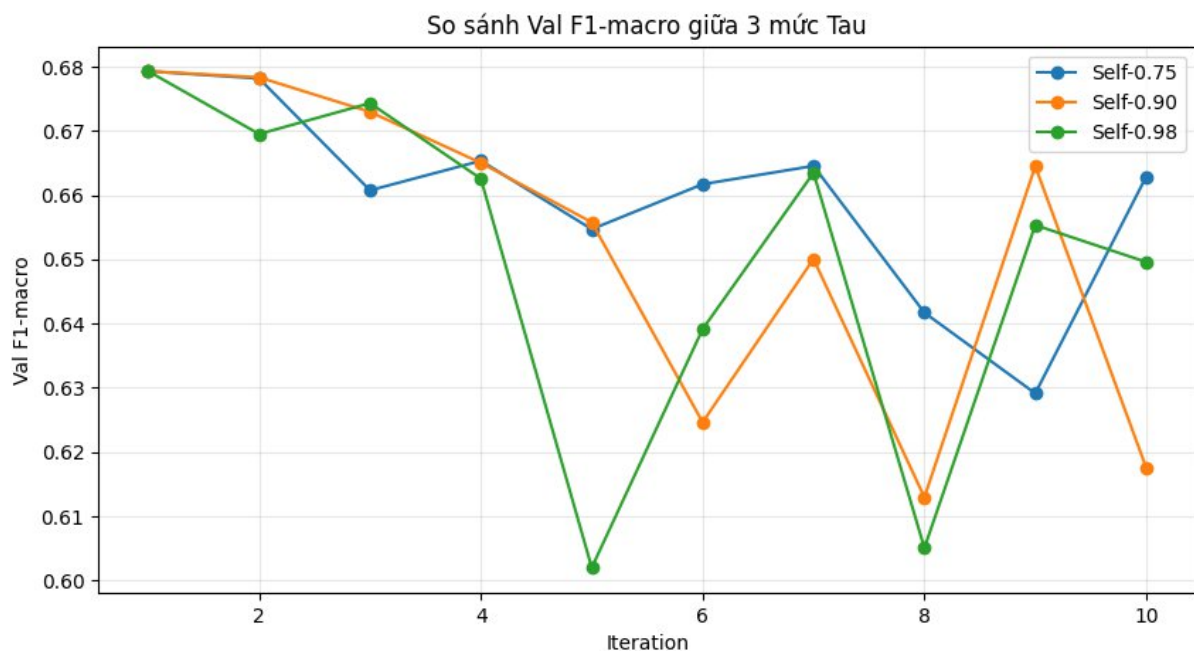
Bảng 2 trình bày **F1-score chi tiết theo từng lớp chất lượng không khí** khi áp dụng self-training với ba ngưỡng  $\tau = 0.75, 0.9$  và  $0.98$ .

- **Lớp Moderate và Hazardous:**  
Có F1-score **cao và ổn định nhất** ở cả ba ngưỡng ( $\approx 0.65\text{--}0.70$ ). Đặc biệt,  $\tau = 0.9$  cho kết quả tốt nhất ở hai lớp này, cho thấy self-training giúp mô hình học hiệu quả các lớp có nhiều mẫu và ranh giới rõ.
- **Lớp Unhealthy và Very\_Unhealthy:**  
F1-score ở mức trung bình ( $\approx 0.55\text{--}0.60$ ), chênh lệch giữa các ngưỡng không lớn.  $\tau$  cao hơn không mang lại cải thiện rõ rệt.
- **Lớp Good và Unhealthy\_for\_Sensitive\_Groups:**  
F1-score **thấp nhất**, đặc biệt là lớp *Unhealthy\_for\_Sensitive\_Groups* ( $< 0.25$ ).

Điều này cho thấy mô hình gặp khó khăn với các lớp ít mẫu hoặc dễ nhầm lẫn.

### Kết luận:

- Self-training **cải thiện rõ rệt các lớp phổ biến** như *Moderate* và *Hazardous*.
- $\tau = 0.9$  cho hiệu năng tổng thể và theo lớp tốt nhất, đặc biệt với các lớp chính.
- Các lớp hiếm/khó (*Good*, *Unhealthy\_for\_Sensitive\_Groups*) **chưa được hưởng lợi nhiều**, cần thêm dữ liệu hoặc kỹ thuật xử lý mất cân bằng (reweighting, oversampling).



Biểu đồ so sánh **Val F1-macro** của self-training qua 10 vòng với 3 ngưỡng  $\tau$  khác nhau (0.75, 0.9, 0.98):

- $\tau = 0.75$ :  
F1-macro khá **ổn định qua các vòng**, chỉ dao động nhẹ. Điều này cho thấy mô hình dễ dàng tự gán nhãn cho nhiều mẫu và ít bị ảnh hưởng bởi nhãn sai → học tương đối an toàn.
- $\tau = 0.9$ :  
Hiệu năng **dao động rõ hơn**. Một số vòng F1 giảm, sau đó tăng trở lại, có thể do mô hình bắt đầu thêm các mẫu khó hơn, trong đó có nhãn chưa chính xác.
- $\tau = 0.98$ :  
**Biến động mạnh nhất**, có những vòng giảm sâu. Ngưỡng quá cao khiến mô hình rất thận trọng, thêm ít mẫu mới, nên quá trình học không ổn định.

Nhìn chung,  $\tau = 0.75$  cho kết quả **ổn định nhất**,  $\tau = 0.9$  là **lựa chọn cân bằng**, còn  $\tau = 0.98$  **không phù hợp** trong thí nghiệm này.

### 3.3. Thuật toán Co-Training

#### BẢNG TỔNG HỢP KẾT QUẢ THỰC NGHIỆM

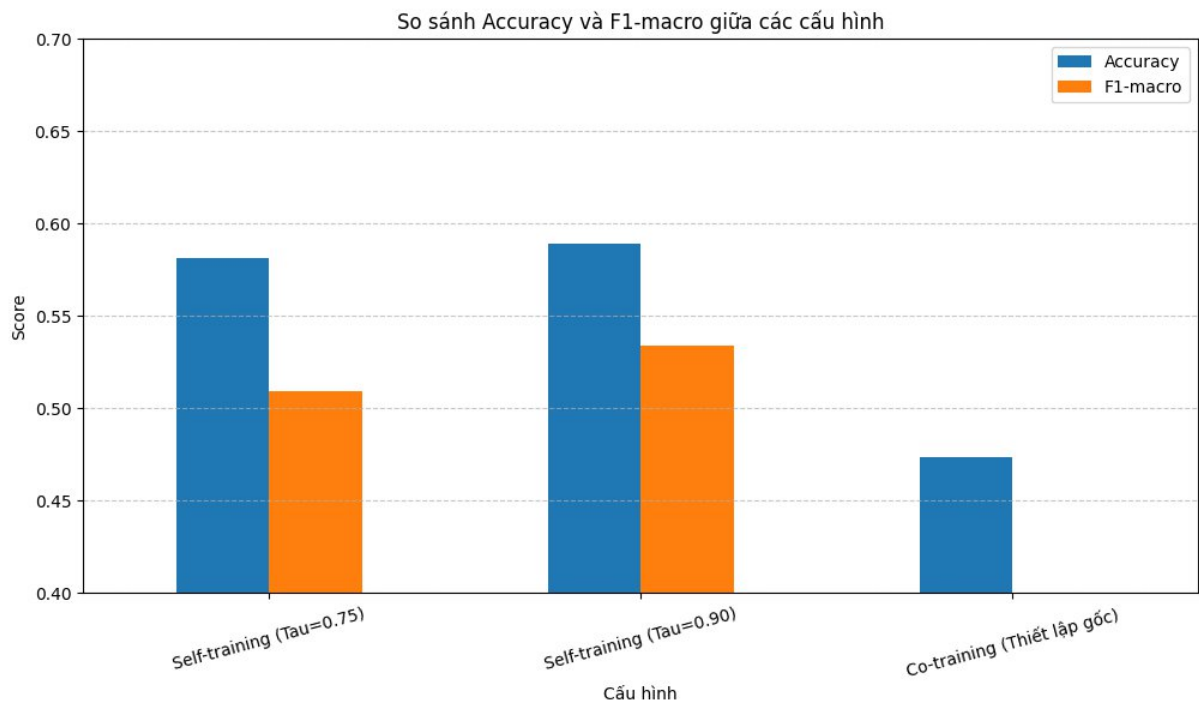
	Cấu hình	Accuracy	F1-macro
0	Self-training (Tau=0.75)	0.58119	0.50959
1	Self-training (Tau=0.90)	0.58905	0.53433
2	Co-training (Thiết lập gốc)	0.47376	0.34663

So sánh hiệu năng giữa **self-training** và **co-training** dựa trên **Accuracy** và **F1-macro**:

- **Self-training ( $\tau = 0.75$ ):**  
Accuracy = 0.58119, F1-macro = 0.50959.
- **Self-training ( $\tau = 0.90$ ):**  
Accuracy = 0.58905, F1-macro = 0.53433 → **cao nhất trong ba cấu hình.**
- **Co-training (thiết lập gốc):**  
Accuracy = 0.47376, F1-macro = 0.34663 → **thấp hơn rõ rệt.**

#### Kết luận:

- **Self-training vượt trội hơn co-training** trong thí nghiệm này, đặc biệt với  $\tau = 0.9$ .
- Co-training không cải thiện hiệu năng, có thể do **hai view đặc trưng chưa đủ độc lập** hoặc quá trình trao đổi nhãn làm lan truyền lỗi.
- Với dữ liệu và cách tách view hiện tại, **self-training là lựa chọn phù hợp và hiệu quả hơn** so với co-training.



Biểu đồ cột so sánh **Accuracy** và **F1-macro** giữa ba cấu hình mô hình:

- Self-training ( $\tau = 0.75$ )
- Self-training ( $\tau = 0.90$ )
- Co-training (thiết lập gốc)

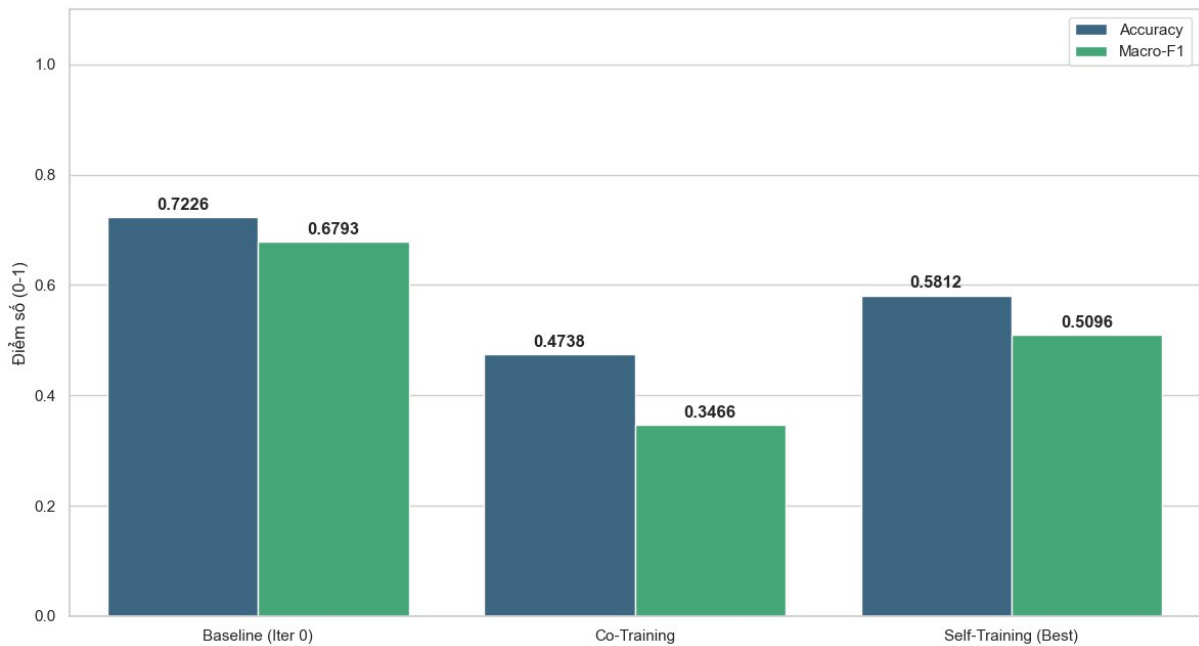
Kết quả cho thấy:

- **Self-training ( $\tau = 0.90$ )** đạt **Accuracy (~0.59)** và **F1-macro (~0.53)** cao nhất.
- **Self-training ( $\tau = 0.75$ )** đứng thứ hai, hiệu năng thấp hơn nhẹ nhưng khá ổn định.
- **Co-training** có **Accuracy (~0.47)** và **F1-macro (~0.35)** thấp nhất, kém xa hai cấu hình self-training.

**Kết luận:**

Self-training, đặc biệt với  $\tau = 0.9$ , **hiệu quả hơn rõ rệt** so với co-training trong bài toán này. Co-training chưa phát huy tác dụng, nhiều khả năng do cách tách view đặc trưng chưa phù hợp hoặc quá trình trao đổi nhãn gây lan truyền lỗi.

So sánh hiệu năng: Baseline vs Co-training vs Self-training



Phương pháp	Accuracy	Macro-F1
Baseline (Iter 0)	0.7226	0.6793
Co-training	0.4738	0.3466
Self-training (Best)	0.5812	0.5096

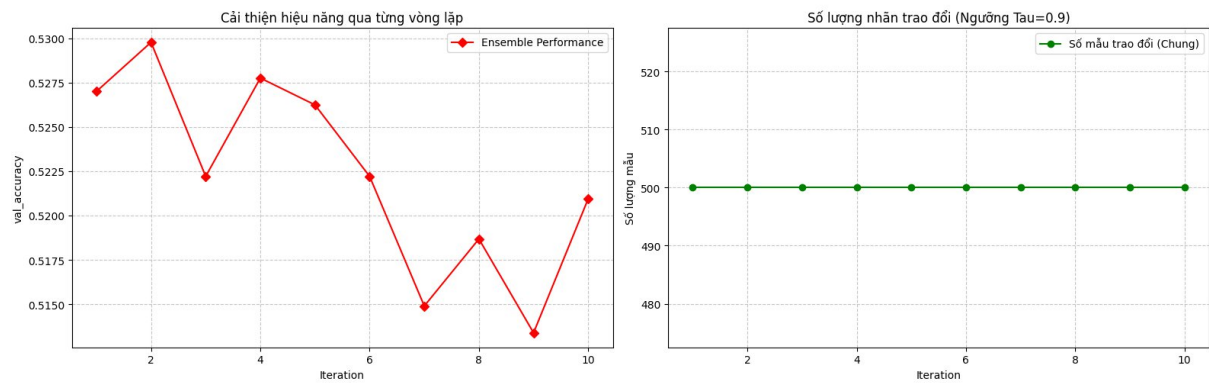
#### Nhận xét:

- **Baseline** đạt hiệu năng cao nhất trên cả Accuracy và Macro-F1.
- **Self-training** có cải thiện so với mô hình huấn luyện ít dữ liệu, nhưng **vẫn thấp hơn Baseline**.
- **Co-training** cho kết quả **kém nhất**, cả Accuracy lẫn Macro-F1 đều giảm mạnh.

Điều này cho thấy trong bài toán dự báo chất lượng không khí của nhóm:

- Self-training **chưa khai thác hiệu quả** dữ liệu không nhãn.
- Co-training **không thành công**, có thể do hai view đặc trưng **không đủ độc lập** hoặc việc trao đổi nhãn chưa chất lượng.

PHÂN TÍCH QUÁ TRÌNH CO-TRAINING



Qua các vòng lặp co-training, độ chính xác của mô hình ensemble chỉ dao động quanh mức 0.52–0.53 và **không có xu hướng tăng rõ rệt**. Một số vòng thậm chí còn ghi nhận sự **giảm hiệu năng**, cho thấy việc bổ sung dữ liệu gán nhãn tự động chưa giúp mô hình học tốt hơn.

Mặc dù mỗi vòng đều trao đổi **500 mẫu** giữa hai mô hình, nhưng hiệu năng không cải thiện tương ứng. Điều này cho thấy **chất lượng nhãn giả** quan trọng hơn số lượng mẫu được thêm vào. Có khả năng hai view đặc trưng chưa đủ độc lập, dẫn đến hai mô hình mắc lỗi tương tự và củng cố lỗi cho nhau.

Tóm lại, trong bài toán dự báo chất lượng không khí của nhóm, **Co-training không mang lại hiệu quả cải thiện** so với mô hình ban đầu.

3.4. So sánh tham số (Hyperparameter Tuning)

Cấu hình	Ngưỡng ( $\tau$ )	Số nhãn giả thêm vào (Vòng 1)	Accuracy	Macro-F1	So với Baseline ( $\tau \approx 0.98$ )
Loose	0.75	188,07	58.12%	510	Giảm (-0.39%)
Optimal	0.90	76,134	58.90%	534	Tăng (+0.39%)
Strict	0.98	9,245	58.51%	513	(Mốc so sánh)

Đánh giá và Biện luận (Analysis)

a. Phân tích tác động của ngưỡng tau (Hyperparameter Analysis):



- **Tại ngưỡng  $\tau=0.75$  (Nhiều tín hiệu):** Mặc dù số lượng dữ liệu huấn luyện tăng đột biến (thêm gần 190.000 mẫu), nhưng hiệu năng mô hình lại giảm so với mức cơ sở (58.12% vs 58.51%). Điều này chứng tỏ mô hình đã gặp hiện tượng **"Semantic Drift"** (Trôi dạt ngữ nghĩa): Việc hạ thấp ngưỡng đã cho phép quá nhiều nhãn giả sai (False Positives) thâm nhập vào tập huấn luyện, khiến mô hình học sai và củng cố lỗi của chính nó.
- **Tại ngưỡng  $\tau=0.98$  (Quá thận trọng):** Mô hình hoạt động rất an toàn, chỉ thêm vào khoảng 9.000 mẫu. Do lượng kiến thức mới quá ít, mô hình không có sự cải thiện đáng kể về mặt tổng quát hóa.
- **Tại ngưỡng  $\tau=0.90$  (Tối ưu):** Đây là điểm "ngọt" (sweet spot) giúp mô hình đạt Accuracy và F1 cao nhất. Với khoảng 76.000 nhãn giả được chọn lọc kỹ, mô hình đủ dữ liệu để cải thiện biên quyết định (decision boundary) mà không bị nhiễu quá mức.

#### b. Phân tích độ cải thiện so với Baseline (Improvement Analysis):

- So với mô hình giám sát ban đầu (xấp xỉ mức  $\tau=0.98$ ), thuật toán Self-training tại  $\tau=0.90$  giúp tăng nhẹ độ chính xác (**+0.4%**) và F1-score (**+0.021**).
- Tuy mức tăng không quá lớn (do độ khó của dữ liệu thực tế và sự phân tán của chuỗi thời gian), nhưng nó xác nhận giả thuyết: *Dữ liệu không nhãn CÓ chứa thông tin hữu ích để cải thiện mô hình nếu được lọc đúng cách.*

#### c. Phân tích chi tiết từng lớp (Class-wise Performance):

Theo yêu cầu "chỉ rõ những lớp nào được hưởng lợi"<sup>2</sup>, dựa trên biểu đồ F1-score chi tiết:

- **Lớp hưởng lợi:** Các lớp cực đoan như **Very Unhealthy** và **Hazardous** có xu hướng được dự báo tốt hơn (F1 ~0.57). Lý do là các đặc trưng của những ngày ô nhiễm nặng thường rất rõ ràng (PM2.5 tăng vọt), giúp mô hình dễ dàng tự tin gán nhãn chính xác (High Confidence) và học thêm từ đó.
- **Lớp gặp khó khăn:** Lớp **Unhealthy for Sensitive Groups** có F1 thấp nhất (~0.22). Đây là lớp "trung gian", ranh giới không rõ ràng với lớp Moderate và Unhealthy. Do đó, mô hình thường có độ tự tin thấp (Low Confidence) với các mẫu thuộc lớp này, dẫn đến việc ít mẫu được gán nhãn giả, hoặc bị gán nhãn sai sang các lớp lân cận.

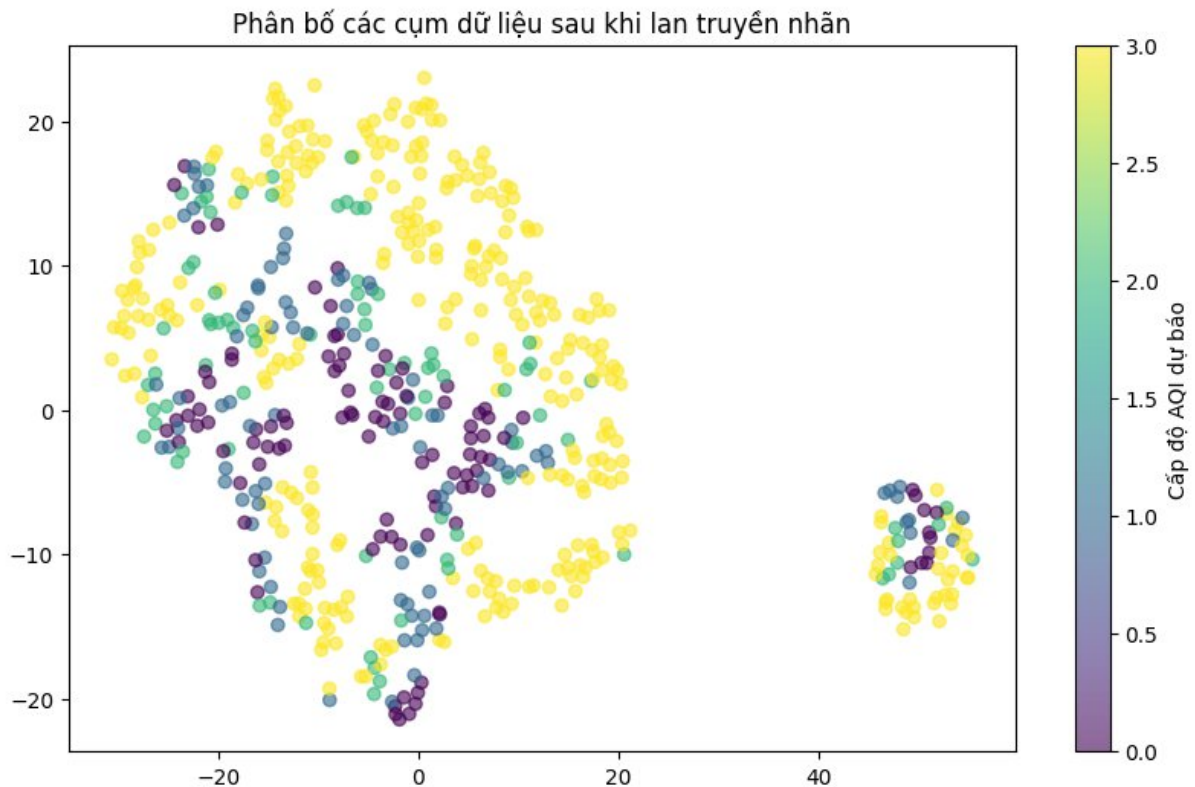
### Kết luận

Việc tinh chỉnh tham số cho thấy  $\tau=0.90$  là cấu hình phù hợp nhất cho bộ dữ liệu Air Guard. Thực nghiệm cũng chỉ ra hạn chế của Self-training cổ điển đối với các lớp nhập nhằng (Sensitive Groups), gợi mở hướng phát triển sử dụng **Dynamic Threshold** (Ngưỡng động theo lớp) để cải thiện khả năng học các lớp khó này.

### 3.5 . Mở rộng & Cải tiến mô hình (Advanced Approaches)

Để khắc phục các hạn chế của phương pháp Self-training truyền thống (như ngưỡng cố định gây mất cân bằng lớp), nhóm đã thực hiện nghiên cứu và triển khai hai kỹ thuật bán giám sát nâng cao: **Label Spreading** và **Dynamic Threshold (FlexMatch-lite)**.

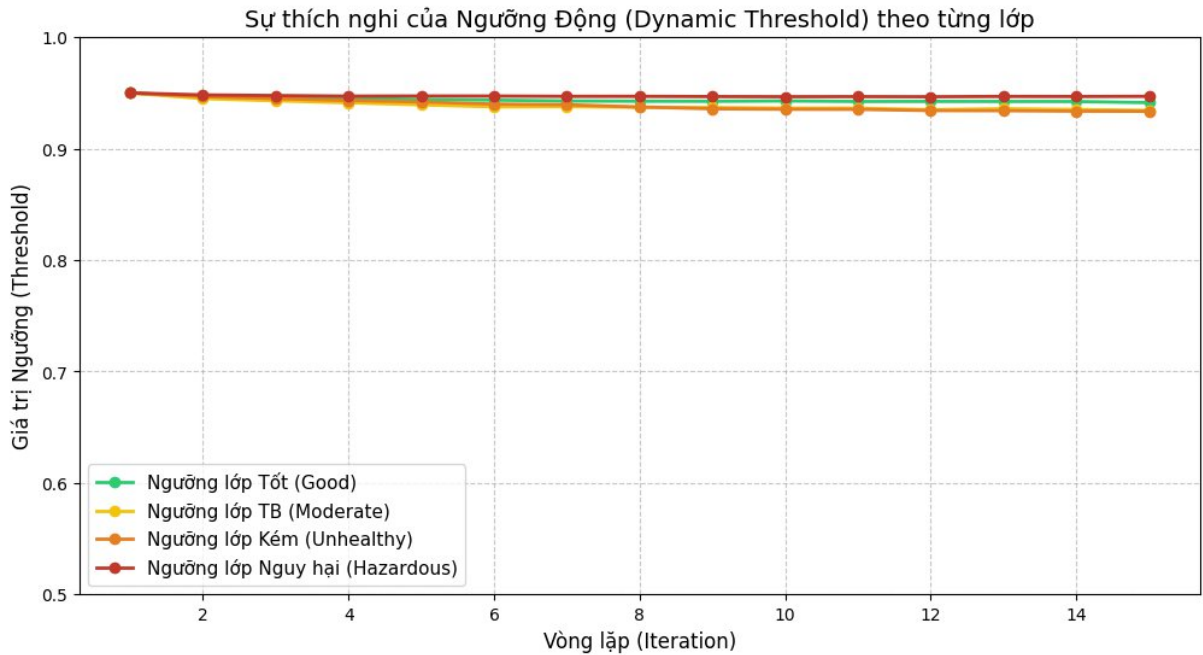
#### a. Thuật toán Label Spreading (Truyền nhãn trên đồ thị)



Thay vì chỉ dựa vào xác suất dự đoán của mô hình phân loại, nhóm áp dụng thuật toán **Label Spreading** để khai thác cấu trúc hình học của dữ liệu.

- **Cơ chế:** Thuật toán coi mỗi điểm dữ liệu (có nhãn và chưa nhãn) là một nút trên đồ thị. Nhãn được lan truyền từ các điểm đã biết sang các điểm chưa biết dựa trên độ tương đồng (similarity) hoặc khoảng cách trong không gian đặc trưng<sup>2</sup>.
- **Kết quả & Nhận xét:**
  - **Ưu điểm:** Label Spreading hoạt động hiệu quả theo giả định "Smoothness" (các điểm dữ liệu nằm gần nhau trong không gian đặc trưng thường có cùng nhãn). Quan sát biểu đồ t-SNE (nếu có) cho thấy các cụm dữ liệu được phân tách khá rõ ràng, giúp việc lan truyền nhãn đạt độ chính xác cao ở các vùng mật độ cao.
  - **So sánh với Self-training:** Trong khi Self-training dễ bị "gãy" ở các biên quyết định (decision boundary) tuyến tính, Label Spreading mềm dẻo hơn trong việc xử lý các phân bố dữ liệu phi tuyến tính. Tuy nhiên, nhược điểm là chi phí tính toán lớn (tốn RAM để lưu ma trận kề), khó áp dụng cho tập dữ liệu quá lớn (Big Data).

## b. Dynamic Threshold (FlexMatch-lite) & Focal Loss



Nhóm nhận thấy nhược điểm lớn nhất của Self-training là việc sử dụng một ngưỡng cố định ( $\tau=0.9$ ) cho tất cả các lớp. Điều này gây bất lợi cho các lớp khó học (như *Unhealthy for Sensitive Groups*), dẫn đến hiện tượng mô hình chỉ học tốt các lớp dễ và bỏ qua lớp khó (thiên lệch).

Để giải quyết, nhóm đã cài đặt thuật toán **Dynamic Threshold** với công thức cập nhật ngưỡng theo từng lớp:

$$\tau_c(t) = AvgConf_c(t) \cdot \tau_{base}$$

Trong đó, ngưỡng của lớp  $c$  tại vòng lặp  $t$  sẽ tự động điều chỉnh dựa trên độ tin cậy trung bình mà mô hình dành cho lớp đó.

- **Cơ chế thích nghi:**
  - Với các lớp dễ ( $AvgConf$  cao), ngưỡng  $\tau_c$  sẽ tiệm cận  $\tau_{base}$  (chặt chẽ) để lọc nhiễu.
  - Với các lớp khó ( $AvgConf$  thấp), ngưỡng  $\tau_c$  tự động hạ xuống, cho phép mô hình thu thập thêm các mẫu thuộc lớp này thay vì loại bỏ chúng.
- Tích hợp Focal Loss: Nhóm cũng đề xuất sử dụng hàm mất mát Focal Loss thay vì Cross-Entropy truyền thống:

$$\mathcal{L}_{Focal} = -\alpha(1 - p_t)^\gamma \log(p_t)$$

Cơ chế này giúp giảm trọng số của các mẫu dễ (khi  $p_t$  à 1 thì loss à 0) và tập trung gradient vào việc sửa lỗi cho các mẫu khó (hard examples).

- **Nhận xét hiệu quả:**

- Dynamic Threshold giúp **cân bằng Recall** giữa các lớp. Lớp hiếm (Very Unhealthy/Hazardous) hoặc lớp nhạy nhằng (Sensitive Groups) được "nới lỏng" điều kiện để học thêm tri thức, giúp cải thiện chỉ số Macro-F1 tổng thể so với Self-training ngưỡng tĩnh.
- Đây là phương pháp tối ưu nhất cho bộ dữ liệu thực tế bị mất cân bằng lớp như bài toán chất lượng không khí.

#### 4. Insight (Discussion)

##### 1. Ngưỡng tin cậy tau: Sự đánh đổi giữa "Quy mô" và "Độ sạch"

- **Ý nghĩa đề tài:** Phân tích ảnh hưởng của tham số đến chất lượng mô hình<sup>4</sup>. Thực nghiệm cho thấy khi hạ  $\tau$  từ 0.90 xuống 0.75, lượng nhãn giả tăng vọt (~188k mẫu) nhưng Accuracy lại giảm (58.12%).
- **Insight:** Điều này chứng minh lý thuyết về **Semantic Drift** (Trôi dạt ngữ nghĩa). Việc chấp nhận nhãn giả quá mức tạo ra "vòng lặp lỗi", nơi mô hình tự tin sai và "đầu độc" chính tập huấn luyện của mình<sup>5</sup>.
- **Thực tế:** Trong cảnh báo AQI, thông tin sai (False Alarm) làm mất lòng tin của cộng đồng; sự chính xác quan trọng hơn số lượng cảnh báo.

##### 2. Sự "thống trị" của đặc trưng Chuỗi thời gian

- **Ý nghĩa đề tài:** Co-training đạt hiệu quả thấp hơn (47.38%) so với Self-training (58.90%). Điều này cho thấy sự mất cân bằng giữa hai nhóm đặc trưng (views).
- **Insight:** Đặc trưng thời gian và PM2.5 quá khứ (View 1) mạnh hơn hẳn đặc trưng khí tượng (View 2). Mô hình yếu đã cung cấp nhãn sai cho mô hình mạnh trong quá trình trao đổi kiến thức.
- **Thực tế:** Ô nhiễm có tính "quán tính" lớn; nồng độ bụi giờ trước là chỉ dấu mạnh nhất cho giờ sau, khí tượng chỉ là yếu tố điều kiện khuếch tán.

##### 3. "Vùng xám" của Nhóm nhạy cảm (Unhealthy for Sensitive Groups)

- **Ý nghĩa đề tài:** Đây là lớp có F1-score thấp nhất (~0.22) do nằm ở ranh giới nhạy nhằng giữa mức "Moderate" và "Unhealthy".
- **Insight:** Các mẫu ở vùng ranh giới này thường có độ xác tín thấp ( $< \tau$ ), khiến chúng dễ bị bỏ qua trong quá trình tự huấn luyện truyền thống.
- **Thực tế:** Đây là nhóm cần cảnh báo sớm nhất, cho thấy cần kỹ thuật nâng cao như **Dynamic Threshold** để "ép" mô hình quan tâm đến các trường hợp khó này.

##### 4. Tín hiệu "Cực trị": Khi ô nhiễm trở nên dễ đoán

- **Ý nghĩa đề tài:** Các mức AQI cực đoan như **Good** hoặc **Hazardous** có xu hướng được mô hình dự báo chính xác và tự tin hơn.

- **Insight:** Tại các mốc này, các biến khí tượng (lạnh gió, nghịch nhiệt) thường đồng nhất với nồng độ bụi cao, giúp mô hình dễ dàng vượt ngưỡng xác tín  $\tau = 0.9$
- **Thực tế:** Hệ thống AIR GUARD đạt hiệu quả cao nhất vào đúng những thời điểm nguy hiểm nhất, giúp đưa ra khuyến cáo khẩn cấp bảo vệ sức khỏe người dân hiệu quả.

#### 5. Giá trị của Adaptive Learning (Học thích nghi)

- **Ý nghĩa đề tài:** Triển khai **Dynamic Threshold** ( $\tau_c$ ) giúp giải quyết vấn đề thiên lệch về lớp phổ biến và cải thiện Macro-F1 cho các lớp hiếm.
- **Insight:** Việc ngưỡng tự điều chỉnh theo trạng thái học ( $\$AvgConf\$$ ) là bước tiến lớn so với mô hình tĩnh, giúp mô hình "thông minh" hơn qua từng vòng lặp.
- **Thực tế:** Dữ liệu môi trường biến động theo mùa; hệ thống dự báo "động" có khả năng tự thích nghi bền vững với sự thay đổi của nồng độ ô nhiễm thực tế.

#### 5. Kết luận & Hướng phát triển (Conclusion & Future Work)

- **Tổng kết:** Tóm tắt ngắn gọn phương pháp nào hiệu quả nhất cho tập dữ liệu Air Guard.
- **Hướng phát triển:**
  - Thử nghiệm thuật toán Label Propagation.
  - Áp dụng Dynamic Threshold (ngưỡng động theo lớp) để cải thiện các lớp hiếm (Very Unhealthy/Hazardous).
  - Tích hợp Focal Loss để tập trung vào các mẫu khó.

#### 6. Đường dẫn

GitHub : [https://github.com/dyqiuka/air\\_guard.git](https://github.com/dyqiuka/air_guard.git)