


# THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo: <https://youtu.be/hmZpbOSwJJ4>
- Link slides: [https://github.com/vinh0406/CS519.Q11/blob/main/CS519\\_Slide.pdf](https://github.com/vinh0406/CS519.Q11/blob/main/CS519_Slide.pdf)

<ul style="list-style-type: none"><li>• Họ và Tên: Đặng Quang Vinh</li><li>• MSSV: 23521786</li></ul> 	<ul style="list-style-type: none"><li>• Lớp: <a href="#">CS519.Q11.KHTN</a></li><li>• Tự đánh giá (điểm tổng kết môn): 9/10</li><li>• Số buổi vắng: 0</li><li>• Số câu hỏi QT cá nhân: 9</li><li>• Số câu hỏi QT của cả nhóm: 1</li><li>• Link Github: <a href="https://github.com/vinh0406/CS519.Q11">https://github.com/vinh0406/CS519.Q11</a></li><li>• Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:<ul style="list-style-type: none"><li>○ Viết đề cương</li><li>○ Viết slide phần nội dung, phương pháp và kết quả dự kiến</li><li>○ Làm video YouTube phần nội dung, phương pháp và kết quả dự kiến</li><li>○ Làm poster phần mô tả chi tiết đề tài</li></ul></li></ul>
--	--

- Họ và Tên: Cao Lê Công Thành
- MSSV: 23521437



- Lớp: [CS519.Q11.KHTN](#)
- Tự đánh giá (điểm tổng kết môn): 9/10
- Số buổi vắng: 0
- Số câu hỏi QT cá nhân: 9
- Số câu hỏi QT của cả nhóm: 1
- Link Github:  
<https://github.com/ThanhCaoUIT24/CS519.Q11>
- Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:
  - Lên ý tưởng đề tài
  - Viết slide phần giới thiệu và mục tiêu
  - Làm video YouTube phần giới thiệu và mục tiêu
  - Làm poster phần tổng quan của đề tài

# ĐỀ CƯƠNG NGHIÊN CỨU

## TÊN ĐỀ TÀI (IN HOA)

ĐỊNH LƯỢNG Ý NGHĨA THỐNG KÊ CHO PHƯƠNG PHÁP PHÁT HIỆN BẤT THƯỜNG DỰA TRÊN MÔ HÌNH HỌC SÂU BÁN GIÁM SÁT THÔNG QUA SUY DIỄN CHỌN LỌC

## TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

QUANTIFYING STATISTICAL SIGNIFICANCE FOR DEEP SEMI-SUPERVISED ANOMALY DETECTION VIA SELECTIVE INFERENCE

## TÓM TẮT

Trong kỷ nguyên chuyển đổi số, bài toán Phát hiện bất thường (Anomaly Detection) đóng vai trò sống còn trong nhiều lĩnh vực thiết yếu, từ việc chẩn đoán khối u trong y tế đến phát hiện lỗi sản phẩm trên dây chuyền công nghiệp. Hiện nay, mô hình Deep Semi-Supervised Anomaly Detection (Deep SAD) đang được xem là phương pháp tiên tiến nhất nhờ khả năng tận dụng dữ liệu có nhãn để đạt độ chính xác cao. Tuy nhiên, rào cản lớn nhất ngăn cản việc ứng dụng rộng rãi mô hình này là tính chất "hộp đen": Deep SAD chỉ đưa ra kết quả nhị phân (Bình thường hoặc Bất thường) mà không cung cấp bất kỳ thước đo nào về độ tin cậy của dự đoán đó.

Việc thiếu vắng cơ sở định lượng này dẫn đến nguy cơ mất kiểm soát tỷ lệ "báo động giả" (False Positive). Trong y tế, một báo động giả có thể dẫn đến các xét nghiệm xâm lấn đau đớn không cần thiết; trong sản xuất, nó gây lãng phí thời gian và chi phí vận hành.

Nghiên cứu này giải quyết vấn đề trên bằng cách tích hợp lý thuyết Suy diễn chọn lọc (Selective Inference) vào kiến trúc Deep SAD. Mục tiêu của chúng tôi là chuyển đổi mô hình từ trạng thái "hộp đen" sang "minh bạch hóa". Cụ thể, hệ thống sẽ không chỉ đưa ra cảnh báo bất thường mà còn cung cấp kèm theo một giá trị thống kê ( $p$ -value), giúp định lượng chính xác xác suất mẫu dữ liệu đó là bất thường thực sự hay chỉ do nhiễu ngẫu nhiên.

Kết quả nghiên cứu hướng tới việc xây dựng một công cụ có khả năng kiểm soát chặt chẽ tỷ lệ báo động giả ở mức ý nghĩa thống kê mong muốn (ví dụ 5%), bất kể độ phức tạp của dữ liệu. Điều này cung cấp cơ sở khoa học vững chắc giúp người dùng (bác sĩ, kỹ sư vận hành,...) tự tin hơn khi ra quyết định can thiệp, nâng cao hiệu quả và độ an toàn trong ứng dụng thực tế.

## GIỚI THIỆU

Trong những năm gần đây, bài toán Phát hiện bất thường (Anomaly Detection - AD) đã trở thành tâm điểm nghiên cứu, đóng vai trò then chốt trong các hệ thống giám sát an toàn thuộc nhiều lĩnh vực trọng yếu như chẩn đoán y khoa, an ninh mạng và kiểm soát chất lượng công nghiệp. Mô hình Deep Semi-supervised Anomaly Detection là một trong những phương pháp tiên tiến nhất (State-of-the-art), nhờ khả năng vượt trội trong việc tổng hợp các biểu diễn đặc trưng phức tạp từ dữ liệu đa chiều và tận dụng hiệu quả lượng nhỏ thông tin có nhãn để cải thiện độ chính xác.

Tuy nhiên, việc ứng dụng Deep SAD vào thực tiễn đang gặp rào cản lớn về phương pháp luận: sự thiếu vắng các thước đo định lượng về ý nghĩa thống kê. Các mô hình hiện tại chủ yếu hoạt động như những "hộp đen", đưa ra quyết định dựa trên tối ưu hóa hàm mục tiêu mà không cung cấp cơ sở đánh giá độ tin cậy. Khi có cảnh báo bất thường, người vận hành không có công cụ để xác định liệu đó là bất thường thực sự hay chỉ là kết quả ngẫu nhiên do nhiễu dữ liệu và quá trình lựa chọn mô hình gây ra.

Khoảng trống này dẫn đến nguy cơ không kiểm soát được Tỷ lệ dương tính giả (False Positive Rate), gây lãng phí nguồn lực và giảm niềm tin người dùng trong các ứng dụng nhạy cảm. Đồng thời, các phương pháp kiểm định truyền thống như Naive p-value thường thất bại khi áp dụng cho mạng nơ-ron sâu do không tính đến tính chất thích nghi dữ liệu của quá trình huấn luyện, dẫn đến kết quả kiểm định bị sai lệch.

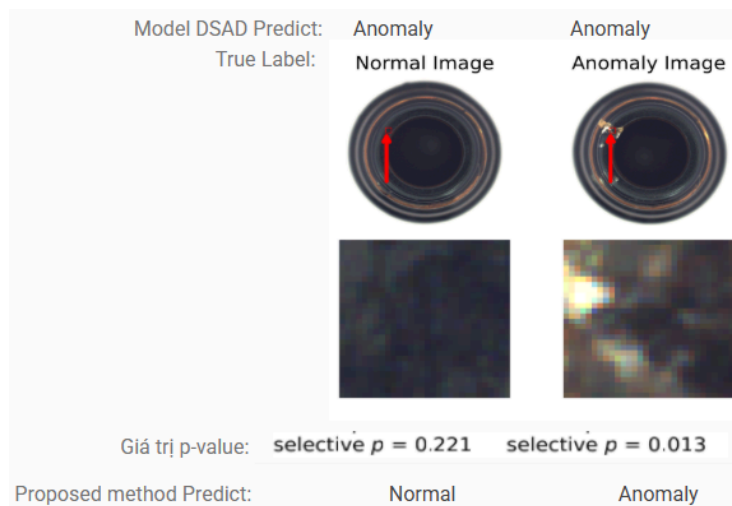
Để giải quyết nhu cầu cấp thiết về việc "minh bạch hóa" độ tin cậy của mô hình, đề tài tập trung nghiên cứu tích hợp lý thuyết Suy diễn chọn lọc (Selective Inference) vào kiến trúc Deep SAD. Đây là hướng tiếp cận mới giúp xây dựng khung lý thuyết chặt chẽ để tính toán giá trị p-value hợp lệ có xét đến quá trình lựa chọn mô hình, từ đó cung cấp cơ sở khoa học vững chắc để kiểm soát rủi ro và nâng cao tính khả thi của hệ thống trong thực tế.

### Input:

- Một mẫu dữ liệu đã được mô hình dự đoán là bất thường cần kiểm tra độ tin cậy.

### Output :

- Giá trị *p-value*: biểu thị xác suất quan sát được mẫu dữ liệu đang xét dưới giả thuyết rằng mẫu là bình thường, sau khi đã điều chỉnh các yếu tố do quá trình lựa chọn của mô hình gây ra.
- Nhãn dự đoán: *Bình thường* hoặc *Bất thường*.



**Hình 1.** Minh họa khả năng kiểm soát tỷ lệ báo động giả (False Positive) của phương pháp đề xuất. Trong khi Deep SAD dự đoán cả hai mẫu đều là "Bất thường" (Anomaly), phương pháp đề xuất tính toán giá trị p-value để kiểm định lại độ tin cậy. Ở ảnh bên trái (ảnh bình thường bị nhận diện sai), giá trị  $p = 0.221 > 0.05$  cho thấy không đủ bằng chứng thống kê để kết luận là bất thường, giúp loại bỏ báo động giả. Ngược lại, ở ảnh bên phải, giá trị  $p = 0.013 < 0.05$  khẳng định đây là một bất thường thực sự với độ tin cậy cao.

## MỤC TIÊU

- Phát triển một phương pháp suy diễn thống kê cho mô hình Deep SAD có khả năng định lượng được ý nghĩa thống kê chính xác cho các dự đoán bất thường từ mô hình.
- Đánh giá thực nghiệm chứng minh phương pháp đề xuất mang lại hiệu quả: Kiểm thử phương pháp trên các tập dữ liệu chuẩn như MVTec AD. Chứng minh tỷ lệ dương tính giả (False Positive Rate) của phương pháp đề xuất duy trì ổn định xấp xỉ mức ý nghĩa  $\alpha$  (khắc phục tình trạng sai lệch của phương pháp Naive p-value). So sánh sức mạnh kiểm định (Power - True Positive Rate) với các phương pháp hiệu chỉnh truyền thống (như Bonferroni correction) để khẳng định phương pháp đề xuất thể hiện sức mạnh kiểm định lớn hơn.

## NỘI DUNG VÀ PHƯƠNG PHÁP

### ❖ Nội dung 1: Tìm hiểu tổng quan đề tài và cơ sở lý thuyết

#### ➤ Phương pháp:

- Tìm hiểu về mô hình Deep SAD: Đọc và phân tích các bài báo khoa học liên quan đến Deep Semi-Supervised Anomaly Detection (Deep SAD) để hiểu cơ chế hoạt động, ưu điểm và hạn chế của nó trong việc

phát hiện bất thường

- Nghiên cứu về Suy diễn chọn lọc (Selective Inference): Tìm hiểu nguyên lý toán học của phương pháp này và cách nó được ứng dụng để kiểm định độ tin cậy trong các mô hình học máy khác.
- Khảo sát các phương pháp hiện có: Tìm hiểu các kỹ thuật kiểm định thống kê truyền thống (như Naive p-value, Bonferroni correction) để làm cơ sở so sánh với phương pháp mới của chúng tôi.

❖ **Nội dung 2:** Nghiên cứu và đề xuất phương pháp mới

➤ Phương pháp:

- Phân tích cấu trúc của mạng của mô hình Deep SAD để xác định các yếu tố ảnh hưởng đến kết quả dự đoán.
- Thiết kế quy trình tính toán độ tin cậy dựa trên lý thuyết Suy diễn chọn lọc, đảm bảo kết quả phản ánh đúng độ tin cậy của mô hình.

❖ **Nội dung 3:** Hiện thực hóa phương pháp đề xuất

➤ Phương pháp:

- Sử dụng ngôn ngữ Python và các thư viện học sâu (Deep Learning frameworks) để xây dựng module.
- Lập trình chức năng: Nhận vào hình ảnh → Mô hình phát hiện bất thường → Tính toán độ tin cậy → Trả kết quả cuối cùng.

❖ **Nội dung 4:** Thực nghiệm và đánh giá kết quả

➤ Phương pháp:

- Thu thập và xử lý dữ liệu: Sử dụng bộ dữ liệu chuẩn trong công nghiệp như MVTec AD (gồm ảnh các sản phẩm lỗi và không lỗi) để làm dữ liệu kiểm thử.
- Thiết lập kịch bản kiểm thử: Tiến hành chạy thực nghiệm các phương pháp kiểm định hiện có và phương pháp do chúng tôi đề xuất.
- Thống kê kết quả, so sánh hiệu năng và khả năng kiểm soát tỷ lệ báo động giả trong quá trình thực nghiệm giữa các phương pháp

## KẾT QUẢ MONG ĐỢI

- Báo cáo khoa học hoàn chỉnh về lý thuyết của phương pháp đề xuất và kiểm chứng thực nghiệm:

- + Về mặt lý thuyết: Tài liệu trình bày chi tiết cơ sở lý thuyết và chứng minh toán học của phương pháp đề xuất.
- + Về mặt thực nghiệm: Cung cấp các kết quả định lượng để minh chứng cho tính đúng đắn của phương pháp đề xuất và so sánh hiệu năng và khả năng kiểm soát tỷ lệ báo động giả của phương pháp đề xuất với các phương pháp kiểm định thống kê hiện có.
- Module Python:
  - + Module hiện thực hoá phương pháp Suy diễn chọn lọc cho mô hình Deep SAD với chức năng cung cấp nhãn dự đoán và kèm theo một thước đo độ tin cậy giúp người dùng đánh giá độ tin cậy của từng quyết định phát hiện bất thường từ mô hình Deep SAD.

## **TÀI LIỆU THAM KHẢO**

- [1]. Lukas Ruff, Robert A. Vandermeulen, Nico Görnitz, Alexander Binder, Emmanuel Müller, Klaus-Robert Müller, Marius Kloft: Deep Semi-Supervised Anomaly Detection. ICLR 2020
- [2]. Mizuki Niihori, Shuichi Nishino, Teruyuki Katsuoka, Tomohiro Shiraishi, Kouichi Taji, Ichiro Takeuchi: Quantifying Statistical Significance of Deep Nearest Neighbor Anomaly Detection via Selective Inference. NeurIPS 2025
- [3]. Vo Nguyen Le Duy, Ichiro Takeuchi: More Powerful Conditional Selective Inference for Generalized Lasso by Parametric Programming. JMLR. 23: 300:1-300:37 (2022)
- [4]. Vo Nguyen Le Duy, Shogo Iwazaki, Ichiro Takeuchi: Quantifying Statistical Significance of Neural Network-based Image Segmentation by Selective Inference. NeurIPS 2022: 31627–31639
- [5]. Daiki Miwa, Vo Nguyen Le Duy, Ichiro Takeuchi: Valid P-Value for Deep Learning-Driven Salient Region. ICLR 2023
- [6]. Vo Nguyen Le Duy, Hiroki Toda, Ryota Sugiyama, Ichiro Takeuchi: Computing Valid p-value for Optimal Changepoint by Selective Inference using Dynamic Programming. NeurIPS 2020: 11356-11367
- [7]. Lukas Ruff, Robert A. Vandermeulen, Nico Görnitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, Marius Kloft: Deep One-Class Classification. ICML 2018: 4390-4399