

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC BÁCH KHOA
KHOA KHOA HỌC ỨNG DỤNG



ĐỒ ÁN TỐT NGHIỆP

**Nhận diện hành động hút dịch nội
khí quản: Phương pháp trích xuất đặc
trưng và phân loại dựa trên dữ liệu
khung xương**

Phan Hoàng Khang

TP. HỒ CHÍ MINH, 12/2024

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC BÁCH KHOA
KHOA KHOA HỌC ỨNG DỤNG



ĐỒ ÁN TỐT NGHIỆP

**Nhận diện hành động hút dịch nội
khí quản: Phương pháp trích xuất đặc
trưng và phân loại dựa trên dữ liệu
khung xương**

Ngành: Vật Lý Kỹ Thuật (7520401)

Sinh Viên: Phan Hoàng Khang

MSSV: 2111460

GVHD: Lê Nhật Tân

**ĐẠI HỌC QUỐC GIA TP.HCM
TRƯỜNG ĐẠI HỌC BÁCH KHOA
KHOA KHOA HỌC ỨNG DỤNG
BỘ MÔN: VẬT LÝ KỸ THUẬT**

**CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập – Tự do – Hạnh phúc
TP.Hồ Chí Minh, Ngày 2 tháng 4 năm 2025**

NHIỆM VỤ LUẬN VĂN TỐT NGHIỆP

**HỌ VÀ TÊN: Phan Hoàng Khang
NGÀNH: VẬT LÝ KỸ THUẬT**

**MSSV: 2111460
LÓP: KU21VLV2**

1. Đầu đề luận văn:

Nhận diện hành động hút dịch nội khí quản: Phương pháp trích xuất đặc trưng và phân loại dựa trên dữ liệu khung xương

2. Nhiệm vụ:

- Tìm hiểu tổng quan về nhận diện hành động
- Tìm hiểu kỹ thuật Trích xuất đặc trưng
- Tìm hiểu kỹ thuật Tạo sinh dữ liệu bằng học sâu
- Đưa ra phương pháp nhận diện hành động của kỹ thuật viên trong hành động hút dịch nội khí quản bằng học máy
- Đưa ra điểm khác biệt giữa hành động của nhân viên y tế và sinh viên ngành y

3. Ngày giao nhiệm vụ luận văn:

4. Ngày hoàn thành nhiệm vụ:

4. Họ và tên người hướng dẫn:

ThS. Lê Nhật Tân

Tên đề tài và nội dung LVTN đã được thông qua Bộ môn.

NGƯỜI HƯỚNG DẪN
(Ký và ghi rõ họ tên)

CHỦ NHIỆM BỘ MÔN
(Ký và ghi rõ họ tên)

Mục lục

1 Mở Đầu	1
1.1 Cơ sở khoa học và thực tiễn của đề tài	1
1.2 Ý nghĩa đề tài	2
1.3 Mục tiêu đề tài	2
2 Cơ sở lý thuyết và các nghiên cứu liên quan	3
2.1 Tổng quan về nhận diện hành vi con người	3
2.2 Tổng quan của khí quản	3
2.2.1 Giải phẫu và sinh lý của khí quản	3
2.2.2 Dịch nội khí quản và nguyên nhân gây ra dịch nội khí quản	4
2.3 Tổng quan về hành động hút dịch nội khí quản	5
2.3.1 Sự nguy hiểm của hút dịch nội khí quản	6
2.3.2 Quy trình thực hiện thủ thuật hút dịch nội khí quản	6
2.4 Tổng quan về học máy và trí tuệ nhân tạo	7
2.4.1 Lịch sử học máy	7
2.4.2 Học máy có giám sát	8
2.4.3 Học máy không giám sát	10
2.4.4 Học máy tăng cường	10
2.4.5 Giảm chiêu dữ liệu	10
2.5 Học sâu	12
2.5.1 Các toán tử trong học sâu	13
2.5.1.a Toán tử tích chập	13
2.5.1.b Toán tử Pool	14
2.5.1.c Toán tử Padding	15
2.5.1.d Toán tử chuẩn hoá	17
2.5.1.e Hàm kích hoạt	18
2.5.1.f Toán tử kết nối toàn diện	19
2.6 Tạo sinh dữ liệu	20
2.6.1 Tạo sinh dữ liệu bằng công cụ thống kê	20
2.6.2 Tạo sinh dữ liệu bằng công cụ học sâu	21
2.7 Các nghiên cứu liên quan	22
2.7.1 Các nghiên cứu về tính khả thi	22
2.7.2 Các nghiên cứu tăng cường kết quả bằng tạo sinh dữ liệu	23
2.7.3 Các nghiên cứu tăng cường kết quả bằng trích xuất đặc trưng	24

3 Phương pháp nghiên cứu	26
3.1 Bộ dữ liệu	26
3.2 Làm mịn dữ liệu	27
3.3 Chuẩn hoá dữ liệu	28
3.4 Trích xuất đặc trưng	28
3.4.1 Đặc trưng về vị trí	30
3.4.2 Đặc trưng vị trí tương đối và di chuyển	31
3.4.3 Đặc trưng về góc	32
3.4.4 Đặc trưng liên quan đến thời gian	32
3.4.5 Đặc trưng tuần tự	33
3.5 Lựa chọn đặc trưng	33
3.6 Tạo sinh dữ liệu	34
3.7 Mô hình phân loại	34
3.8 Quy trình phân loại	35
3.9 Phương pháp đánh giá mô hình	36
3.10 Phương pháp đánh giá sự khác biệt giữa nhân viên y tế và sinh viên	37
4 Kết quả và bàn luận	38
4.1 Kết quả mô hình phân loại	38
4.2 Bàn luận	39
4.2.1 Về nhận diện sự khác nhau về hành động của học viên y khoa và nhân viên y tế có kinh nghiệm bằng XGB	39
4.2.2 Các ứng dụng tiềm năng của đề tài	40
5 Tổng kết	43
A Các công bố khoa học cho đồ án này	49
B Danh mục hình ảnh hành động hút dịch nội khí quản	50

Danh sách hình vẽ

2.1 Giải phẫu đường hô hấp trên ¹	4
2.2 Một số nguyên nhân phổ biến cho viêm khí quản ²	5
2.3 Hình ảnh minh họa của hành động hút dịch nội khí quản ³	5
2.4 Sự phát triển của học máy cổ điển. Nhờ các bước phát triển học máy (machine learning) đầu tiên đạt kết quả rực rỡ, những ứng dụng phát triển về các hệ thống phức tạp hơn đã được nghiên cứu và phát triển.	7
2.5 Sự phát triển của mô hình học sâu. Học sâu đã xuất hiện từ giữa thế kỷ XX với các mô hình có sự xuất hiện của perceptron. nhưng sớm gặp các thách thức và có khoảng dừng 10 năm. Đầu thế kỷ XXI, học sâu trở lại mạnh mẽ khi máy tính có sự phát triển khả năng tính toán mạnh, tạo tiền đề cho những mô hình lớn được khai sinh (Alexnet, Resnet, YOLO,...). Thế nhưng, DL đã thành công rực rỡ khi Transformer xuất hiện, khai mở khả năng tính toán song song và giảm khối lượng tính toán, từ đó mở ra kỷ nguyên AI như hiện tại.	8
2.6 Hình ảnh minh họa cách phân cây quyết định của XGBoost (Nguồn NVIDIA, [15])	9
2.7 Hình ảnh minh họa cách phân cây quyết định của Random Forest (Nguồn NVIDIA, [17])	9
2.8 Hình ảnh minh họa mô hình GMM	11
2.9 Hình ảnh minh họa mô hình K-Means	11
2.10 Hình ảnh minh họa của PCA. Ta có thể thấy dữ liệu cho ban đầu có 3 chiều dẫn đến chi phí tính toán cao. bằng cách phần tích thành phần và giữ lại thành phần chính (vector xanh dương và vector đỏ), ta vẫn có thể phân loại 4 nhóm phân loại trên.	13
2.11 Hình minh họa của SHAP. SHAP ở đây sẽ hoạt động bằng cách lây bên phải và truy xuất về từng robot làm được bao nhiêu tiền. Trong lĩnh vực học máy, SHAP sẽ đóng vai trò truy xuất sự đóng góp của từng biến đến kết quả cuối cùng theo nguyên lý trò chơi (game theory), từ đó người kỹ sư có thể nhìn vào thông số đó để biết đóng góp của từng biến nhất định để chọn các biến phù hợp.	14
2.12 Hình minh họa của trực quan hóa tác động của một thành phần bằng SHAP.	15
2.13 Hình ảnh minh họa mạng thần kinh nhân tạo (NN). Đây là nền tảng cho sự phát triển cho trí tuệ nhân tạo tại thời điểm hiện tại. NN được lý giải là sự tối ưu hoá của các node thần kinh ở mỗi lớp (layer) qua nguyên lý đạo hàm và nâng cấp (update) trọng số. Ngoài ra một vấn đề quan trọng trong NN là việc chọn chính xác hàm kích hoạt (activation function), và hệ số học (learning rate hay ϵ) để việc tối ưu hoá được tối ưu.	16

2.14 Hình ảnh minh họa mạng thần kinh VGG 16 (thuộc nhóm CNN)	17
2.15 Hình ảnh minh họa mạng thần kinh BiLSTM (thuộc nhóm RNN)	17
2.16 Cấu trúc của mô hình VAE (một ví dụ điển hình trong hệ mô hình AE hay Auto Encoder). VAE tận dụng hiệu ứng nút cổ chai (bottle neck effect) nhằm đưa dữ liệu về một không gian đặc trưng nhỏ hơn nhằm cô đọng thông tin (thông tin được chuyển về còn $Z, \sigma(z), \bar{z}$) ở bước mã hoá. Sau đó mô hình sẽ tái lập hình ảnh thông qua mô hình giải mã để tạo ra hình ảnh. quy trình này sẽ được huấn luyện để tạo ra hình ảnh có sự tương đồng về ngữ nghĩa với thông tin cô đọng từ bước mã hoá.	21
2.17 Cấu trúc của mô hình GAN. Mô hình GAN được ứng dụng rộng rãi trong tạo sinh dữ liệu do độ tin cậy mà mạng D mang lại. Do có mạng D, mô hình tạo sinh dữ liệu tự đánh giá chất lượng hình ảnh mà nó tạo ra trước khi trả về dữ liệu cho người dùng. Vì thế, GAN thường tạo ra các dữ liệu có chất lượng cao nhưng cũng vì vậy độ phủ của GAN thường khá thấp.	22
2.18 Cấu trúc của mô hình khuếch tán. Mô hình khuếch tán là mô hình hiện đại nhất tại thời điểm hiện tại của lĩnh vực học sâu trong tạo ra dữ liệu mới. Mô hình này được lấy ý tưởng từ kết hợp hai điểm mạnh của AE và GAN để tạo nên một công cụ tạo sinh dữ liệu mạnh mẽ. Việc tạo thông tin này bắt đầu bằng việc khuếch tán dữ liệu gốc nhiều lần cho đến khi đạt được sự hội tụ. Sau đó, mô hình sẽ học cách tái dụng nhiều này thành hình ảnh gốc bằng các hàm học sâu. Từ khả năng tạo ảnh từ hình ảnh khuếch tán này, một nhiễu ngẫu nhiên \mathcal{N} có thể được mã hoá lại thành hình ảnh \mathcal{I}	23
3.1 Hình ảnh minh họa của dữ liệu khung xương (chấm xanh lá) theo hình ảnh video thực hiện thủ thuật. Đặc biệt là bộ dữ liệu này có một khuyết điểm là chỉ xuất hiện 1 nửa thân trái của người thực hiện thủ thuật gây nên sự thiếu dữ liệu về nửa còn lại của cơ thể trong phần lớn thời lượng hành động.	26
3.2 Hình ảnh minh họa xử lý dữ liệu khuyết (giá trị bằng 0) tại vị trí vai trái theo phương x. Ta có thể thấy phần lớn dữ liệu khuyết đã được giải quyết bằng phương pháp làm mịn này.	27
3.3 Hình ảnh minh họa xử lý dữ liệu khuyết (giá trị bằng 0) tại vị trí vai trái theo phương y. Ta có thể thấy phần lớn dữ liệu khuyết đã được giải quyết bằng phương pháp làm mịn này.	28
3.4 Hình ảnh minh họa xử lý dữ liệu khuyết (giá trị bằng 0) tại vị trí vai phải theo phương x. Ta có thể thấy phần lớn dữ liệu khuyết vẫn được giữ nguyên, nguyên nhân là do vai phải thường là bộ phận bị mất đi vì chỉ ghi hình được nửa cơ thể bên trái trong bộ dữ liệu.	29
3.5 Hình ảnh minh họa xử lý dữ liệu khuyết (giá trị bằng 0) tại vị trí vai phải theo phương y. Ta có thể thấy phần lớn dữ liệu khuyết vẫn được giữ nguyên, nguyên nhân là do vai phải thường là bộ phận bị mất đi vì chỉ ghi hình được nửa cơ thể bên trái trong bộ dữ liệu.	30
3.6 Lưu đồ trích xuất tháo tạm thời đường thở nhân tạo đã xong (is 1 done)	33

3.7	Lưu đồ trích xuất khử trùng ống thông đã xong (is 4 done)	33
3.8	Lưu đồ của quy trình phân loại	35
3.9	Hình ảnh trực quan các giá trị trong ma trận nhầm lẫn	36
4.1	Biểu đồ hộp của F-score theo 5 nhóm đặc trưng	40
4.2	Hình ảnh về sự khác nhau của động tác trong hành động hút đờm. Có thể thấy hành động của sinh viên (hình 4.2a), sinh viên đứng thẳng so với nhân viên y tế (hình 4.2b) hơi khom người để nhìn thấy được rõ hơn dịch đờm. Đây là lý do tiên quyết giúp mô hình nhận diện hành động do bác sĩ hay sinh viên ra quyết định. (hình 4.3 và 4.4)	40
4.3	Biểu đồ SHAP về độ quan trọng của đặc trưng trong nhận biết hành động do học sinh hay y tá thực hiện ở hành động Tháo tạm đường thở nhân tạo	41
4.4	Biểu đồ SHAP về độ quan trọng của đặc trưng trong nhận biết hành động do học sinh hay y tá thực hiện ở hành động Hút đờm	41
4.5	Quy trình minh họa cho ứng dụng sửa lỗi sai. Việc kiểm có thể được thực hiện liên tục, thông suốt quá trình thực hiện thủ thuật. Ngoài ra có thể phát triển thêm hệ thống cho phép thay đổi quy trình trong trường hợp ngoại lệ.	42
B.1	Hình ảnh minh họa hành động chuẩn bị ống thông	50
B.2	Hình ảnh minh họa hành động tháo tạm thời đường thở nhân tạo	51
B.3	Hình ảnh minh họa hành động hút đờm	51
B.4	Hình ảnh minh họa hành động lắp lại đường thở nhân tạo	52
B.5	Hình ảnh minh họa hành động khử trùng ống thông	52
B.6	Hình ảnh minh họa hành động thải bỏ găng tay	53
B.7	Hình ảnh minh họa hành động định vị	53
B.8	Hình ảnh minh họa hành động nghe nhịp tim	54

Danh sách bảng

2.1	Bảng biểu về tổng hợp các nghiên cứu liên quan	25
3.1	Bảng về loại đặc trưng và số lượng.	29
4.1	Kết quả phân loại tổng quát (đơn vị phần trăm)	38
4.2	Kết quả mô hình phân loại theo từng người (đơn vị phần trăm)	39
4.3	Hiệu suất Phân loại giữa Y tá và Sinh viên theo ID Hoạt động bằng XGB	39

Chương 1

Mở Đầu

1.1 Cơ sở khoa học và thực tiễn của đề tài

Trong thế giới phát triển nhanh chóng như hiện nay, ngành y tế đóng vai trò then chốt trong việc duy trì và nâng cao chất lượng cuộc sống của con người. Từ việc phòng ngừa bệnh tật đến điều trị các bệnh lý phức tạp, ngành y tế không ngừng nỗ lực để mang lại sức khỏe và hạnh phúc cho cộng đồng. Tuy nhiên, dẫu sau những thành tựu đáng kể, vẫn còn tồn tại những thách thức cần được giải quyết, một trong số đó là việc nhận diện chính xác các hành động của nhân viên y tế.

Việc nhận diện hành động của bác sĩ và y tá không chỉ đơn thuần là việc quan sát các thao tác, mà còn là quá trình phân tích và đánh giá sự chính xác, hiệu quả và an toàn của các hành động đó. Trong môi trường y tế, nơi mà mỗi quyết định và hành động đều có thể ảnh hưởng trực tiếp đến tính mạng của bệnh nhân, việc nhận diện sai sót hoặc thiếu sót có thể dẫn đến những hậu quả nghiêm trọng. Do đó, việc nghiên cứu và phát triển các phương pháp nhận diện hành động hiệu quả là vô cùng quan trọng.

Hơn nữa, việc hiểu rõ hơn về các hành động của kỹ thuật viên y tế cũng đóng vai trò quan trọng trong việc nâng cao chất lượng đào tạo và nghiên cứu y học. Kỹ thuật viên y tế là những người trực tiếp thực hiện các xét nghiệm, phân tích và các thủ thuật kỹ thuật, đóng góp không nhỏ vào quá trình chẩn đoán và điều trị bệnh. Việc so sánh và phân tích sự khác biệt giữa hành động của sinh viên y khoa và kỹ thuật viên chuyên nghiệp sẽ giúp các nhà giáo dục y khoa có cái nhìn sâu sắc hơn về quá trình học tập và phát triển kỹ năng của sinh viên, từ đó đưa ra những phương pháp giảng dạy và đánh giá phù hợp.

Thế nhưng, tồn tại một số khó khăn trong thu thập dữ liệu. Ví dụ như việc lấy dữ liệu thực tế có thể gây nên những rủi ro an toàn thông tin của tình nguyện viên hay chi phí cao để thu thập dữ liệu. Vì vậy, trong thực tế, nhu cầu phải phát triển thêm những hệ thống để tạo sinh dữ liệu giúp nâng cao hiệu suất mô hình đang hiện hữu và được quan tâm nhiều [1, 2, 3, 4].

Vì thế đề tài này chọn việc nhận diện hành động của kỹ thuật viên thông qua dữ liệu từ video cho hành động hút dịch nội khí quản (ES) để đánh giá các kỹ thuật về tạo sinh dữ liệu đến việc nhận diện hành động của kỹ thuật viên cũng như so sánh về sự sai khác của nhân viên y tế có kinh nghiệm và sinh viên y khoa.

1.2 Ý nghĩa đề tài

Với các vấn đề trình bày ở trên, đề tài sẽ nghiên cứu và phát triển các phương pháp nhận diện hành động của nhân viên y tế, đặc biệt là trong bối cảnh khó khăn về thu thập dữ liệu thực tế cho hành động ES, từ đó góp phần nâng cao chất lượng đào tạo, nghiên cứu và thực hành y tế, hướng tới mục tiêu cuối cùng là cải thiện chất lượng chăm sóc sức khỏe cho cộng đồng.

1.3 Mục tiêu đề tài

Để giải quyết các vấn đề nêu trên, bài nghiên cứu này sẽ phân tích về hành động của nhân viên y tế và học viên y khoa trong hành động hút nội khí quản. Mục tiêu của đề tài này bao gồm:

1. Trích xuất một tập hợp các đặc trưng toàn diện từ dữ liệu ước tính tư thế, nhằm nắm bắt những đặc điểm của các hoạt động ES. Những đặc trưng này sẽ hỗ trợ đánh giá được hiệu quả và độ thành thạo trong các hành động của y tá, từ đó cho phép đánh giá một cách mạnh mẽ về hiệu suất và tính hiệu quả của họ.
2. Đề xuất quy trình học máy kết hợp (ensemble learning) phù hợp để tối ưu hóa hiệu suất nhận dạng các hoạt động ES, từ đó cho phép phân loại chính xác và hiểu sâu hơn về các đặc điểm của quy trình.
3. Tăng cường độ chính xác phân loại và giải quyết tình trạng mất cân bằng lớp bằng cách sử dụng các kỹ thuật tăng cường dữ liệu thông qua mô hình tạo sinh dữ liệu (generative modeling).
4. Đánh giá mối tương quan hiệu suất giữa y tá có ba năm kinh nghiệm và sinh viên điều dưỡng là điều cần thiết để làm rõ ảnh hưởng của kinh nghiệm đối với hiệu quả của quy trình. Những phát hiện này có tiềm năng làm sâu sắc thêm sự hiểu biết của chúng ta về cách mức độ kinh nghiệm ảnh hưởng đến việc thực hiện các quy trình.

Chương 2

Cơ sở lý thuyết và các nghiên cứu liên quan

2.1 Tổng quan về nhận diện hành vi con người

Nhận diện hành vi con người là quá trình phân tích dữ liệu thu thập được từ các cảm biến (như gia tốc, con quay hồi chuyển, camera,...) để xác định các hoạt động mà con người đang thực hiện [5, 6, 7, 8, 9, 10, 11, 11, 12]. Các hoạt động này có thể đơn giản như đi bộ, chạy, ngồi, hoặc phức tạp hơn như thực hiện các thao tác trong công việc, chơi thể thao, hoặc tương tác xã hội. Phạm vi của HAR bao gồm việc thu thập dữ liệu, tiền xử lý dữ liệu, trích xuất đặc trưng, và phân loại hành vi.

Trong lĩnh vực y tế, HAR có thể được sử dụng để theo dõi hoạt động của bệnh nhân/kỹ thuật viên, phát hiện các bất thường trong hành vi (như ngã, co giật), và đánh giá hiệu quả của các liệu pháp điều trị. Ngoài ra HAR còn có thể ứng dụng trong việc quan sát hành động của các bác sĩ hay kỹ thuật viên. Ví dụ, HAR có thể giúp người nhà có thể hiểu được các hành động của bác sĩ trong phòng mổ, giúp người nhà bệnh nhân yên tâm về ca mổ của người thân mà không cần một người truyền tải thông tin.

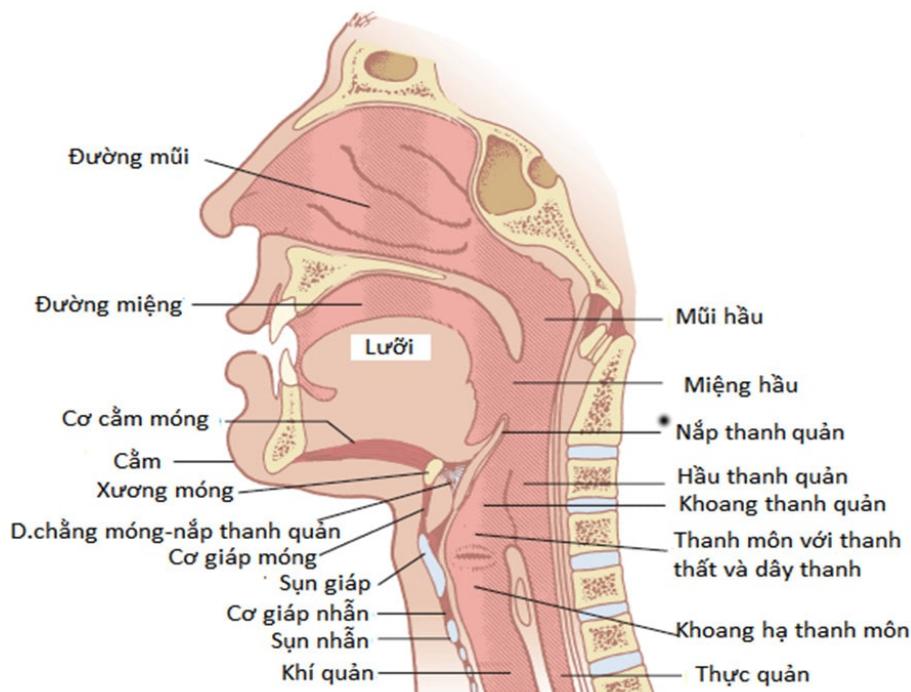
2.2 Tổng quan của khí quản

2.2.1 Giải phẫu và sinh lý của khí quản

Khí quản là một phần quan trọng của hệ hô hấp, đóng vai trò chính trong việc dẫn khí từ miệng và mũi đến phổi. Cấu trúc của khí quản giúp duy trì luồng không khí và bảo vệ phổi khỏi các tác nhân gây hại từ môi trường bên ngoài. Cụ thể hơn, khí quản là một ống dẫn khí dài khoảng 10–12 cm và có đường kính trung bình khoảng 2–2,5 cm ở người trưởng thành [13]. Thành khí quản được tạo thành từ nhiều lớp khác nhau, bao gồm lớp niêm mạc (biểu mô trụ giả tầng có lông chuyển giúp di chuyển các hạt bụi và dịch nhầy ra khỏi đường hô hấp), lớp dưới niêm mạc (chứa các tuyến tiết nhầy giúp bôi trơn và bảo vệ đường hô hấp) và lớp sụn cùng cơ trơn (các vòng sụn hình chữ C giúp duy trì độ mở của khí quản, trong khi phần cơ trơn nằm ở phía sau có thể co giãn để điều chỉnh lưu lượng không khí).

Về mặt chức năng sinh lý, khí quản đóng vai trò dẫn khí, giúp không khí đi từ thanh quản xuống các phế quản và vào phổi (xem Hình 2.1). Nhờ vào cấu trúc vững chắc của vòng sụn, khí quản luôn duy trì được độ mở cần thiết để đảm bảo dòng khí thông suốt. Ngoài ra, khí quản

còn có chức năng lọc và bảo vệ đường hô hấp nhờ hệ thống lông chuyển giúp đẩy bụi bẩn và vi khuẩn ra ngoài, chất nhầy giữ lại các hạt bụi và vi sinh vật, cùng với phản xạ ho giúp loại bỏ các tác nhân kích thích ra khỏi đường thở [13]. Bên cạnh đó, khí quản có khả năng điều chỉnh lưu lượng khí nhờ vào các cơ trơn và sự kiểm soát của hệ thần kinh tự chủ. Khi cơ thể cần tăng cường thông khí, khí quản có thể giãn nở để giúp không khí đi vào phổi dễ dàng hơn. Hệ thần kinh giao cảm có tác dụng làm giãn cơ trơn khí quản, tăng lưu lượng khí vào phổi, trong khi hệ thần kinh đối giao cảm gây co thắt khí quản, làm giảm lưu lượng khí khi không cần thiết.



Hình 2.1: Giải phẫu đường hô hấp trên¹.

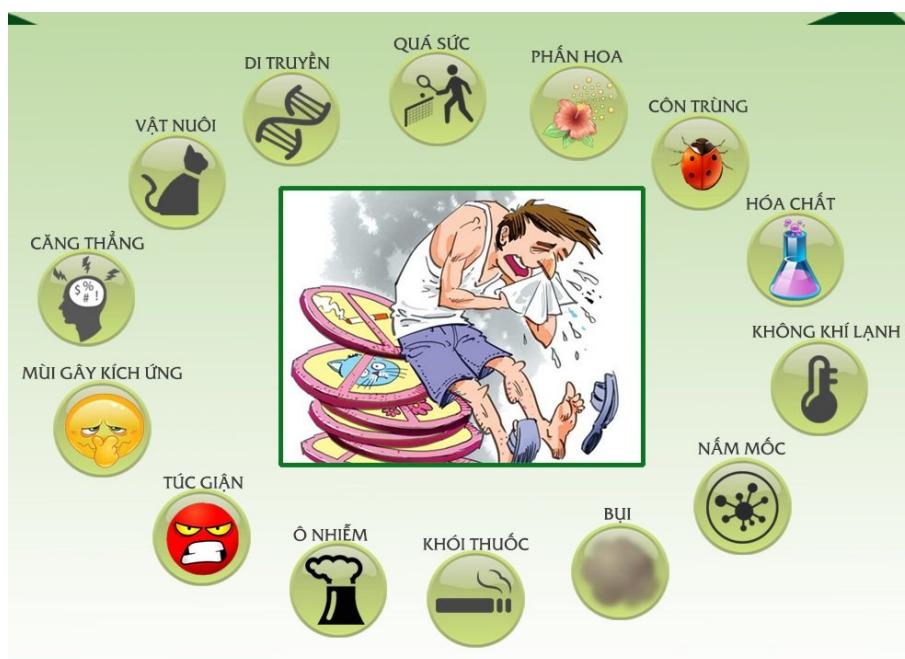
2.2.2 Dịch nội khí quản và nguyên nhân gây ra dịch nội khí quản

Dịch nội khí quản là sự tích tụ chất lỏng trong lòng khí quản, có thể gây cản trở quá trình thông khí và làm tăng nguy cơ nhiễm trùng. Nguyên nhân chính gây ra tình trạng này bao gồm nhiễm trùng đường hô hấp, phản ứng viêm do kích thích từ ống nội khí quản và rối loạn trong cơ chế tiết dịch của đường hô hấp.

Một trong những nguyên nhân phổ biến nhất là nhiễm trùng đường hô hấp, đặc biệt là viêm phổi và viêm khí quản. Khi vi khuẩn hoặc virus xâm nhập vào đường hô hấp, cơ thể phản ứng bằng cách tiết ra nhiều dịch nhầy hơn để loại bỏ tác nhân gây bệnh. Tuy nhiên, nếu hệ thống lông chuyển hoạt động kém hiệu quả hoặc bệnh nhân bị suy giảm miễn dịch, dịch nhầy có thể bị ứ đọng, gây tích tụ trong khí quản.

Ngoài ra, đặt ống nội khí quản trong thời gian dài có thể gây kích thích niêm mạc khí quản, dẫn đến phản ứng viêm và tăng tiết dịch. Điều này thường xảy ra ở bệnh nhân phải thở máy hoặc những người trải qua phẫu thuật cần gây mê kéo dài. Áp lực từ ống nội khí quản có thể làm tổn thương niêm mạc, làm suy giảm cơ chế bảo vệ tự nhiên của khí quản và tạo điều kiện thuận lợi

¹Truy cập từ <https://www.hoihohaptphcm.org/index.php/chuyende/benh-phoi/398-k-thuat-dat-noi-khi-quan-cap-cuu>



Hình 2.2: Một số nguyên nhân phổ biến cho viêm khí quản².
cho vi khuẩn phát triển.

Rối loạn trong cơ chế tiết dịch cũng là một yếu tố quan trọng gây ra dịch nội khí quản. Ở những bệnh nhân có bệnh lý nền như bệnh phổi tắc nghẽn mãn tính (COPD) hoặc suy tim sung huyết, sự mất cân bằng giữa sản xuất và đào thải dịch có thể dẫn đến tích tụ dịch trong khí quản. Bên cạnh đó, mất nước hoặc sử dụng thuốc ức chế tiết dịch cũng có thể làm thay đổi độ nhớt của dịch nhầy, khiến nó khó được đào thải ra ngoài hơn.

2.3 Tổng quan về hành động hút dịch nội khí quản



Hình 2.3: Hình ảnh minh họa của hành động hút dịch nội khí quản³.

²Hình ảnh từ <https://bidophar.com/chi-tiet-tin/nguyen-nhan-gay-viem-phe-quan-cach-dieu-tri>

³Hình ảnh lấy từ <https://merinco.com.vn/blogs/bai-viet-nghien-cuu/hut-dich-ong-mo-khi-quan-quy-trinh-hut-nong-va-sau>

Hút dịch nội khí quản là một thủ thuật y tế xâm lấn, phức tạp, đòi hỏi sự chính xác và cẩn thận cao. Mục đích chính của thủ thuật này là loại bỏ chất tiết tích tụ trong đường thở của bệnh nhân đang được đặt nội khí quản, giúp duy trì đường thông thoáng và cải thiện khả năng hô hấp.

2.3.1 Sự nguy hiểm của hút dịch nội khí quản

Như trình bày ở phần trên, hút dịch nội khí quản là một thủ thuật quan trọng giúp làm sạch đường thở cho bệnh nhân thở máy hoặc suy hô hấp. Tuy nhiên, nếu không thực hiện đúng kỹ thuật, thủ thuật này có thể gây ra nhiều biến chứng nguy hiểm.

Một trong những rủi ro phổ biến là tổn thương niêm mạc khí quản do lực hút mạnh hoặc thực hiện quá thường xuyên, dẫn đến chảy máu, phù nề và tăng nguy cơ nhiễm trùng. Ngoài ra, tụt oxy máu có thể xảy ra khi hút dịch kéo dài, làm giảm nồng độ oxy trong máu, gây suy hô hấp cấp.

Hút dịch không vô khuẩn cũng có thể dẫn đến nhiễm trùng hô hấp, đặc biệt là có thể gây viêm phổi. Một nguy cơ khác là co thắt phế quản hoặc phản xạ cường phế vị, gây tụt huyết áp, nhịp tim chậm, thậm chí ngừng tim. Bên cạnh đó, hút dịch không đúng kỹ thuật còn có thể đẩy dịch sâu hơn vào phổi, làm tình trạng bệnh nhân xấu đi. Vì vậy, để đảm bảo an toàn, thủ thuật này cần được thực hiện bởi người có chuyên môn, sử dụng kỹ thuật vô khuẩn và theo dõi sát tình trạng bệnh nhân trong suốt quá trình hút dịch.

2.3.2 Quy trình thực hiện thủ thuật hút dịch nội khí quản

Trước khi tiến hành thủ thuật, việc chuẩn bị kỹ lưỡng là vô cùng quan trọng. Nhân viên y tế cần đảm bảo đầy đủ các dụng cụ cần thiết, bao gồm ống hút, máy hút, nước muối sinh lý vô khuẩn và găng tay vô khuẩn.

Trong quá trình thực hiện thủ thuật, việc đưa ống hút vào nội khí quản cần được thực hiện một cách nhẹ nhàng và cẩn thận, tránh gây tổn thương cho niêm mạc đường thở. Lực hút cũng cần được điều chỉnh phù hợp, không quá mạnh để tránh gây tổn thương cho phổi. Việc xoay ống hút trong quá trình hút dịch cũng giúp loại bỏ chất tiết hiệu quả hơn.

Sau khi hút dịch, nhân viên y tế bắt buộc phải xem xét tình trạng của bệnh nhân. Nhân viên y tế cần quan sát tình trạng hô hấp, màu sắc da và các dấu hiệu sinh tồn khác để đảm bảo bệnh nhân không gặp phải biến chứng nào. Việc xử lý dụng cụ sau khi sử dụng cũng cần được thực hiện đúng quy trình để đảm bảo vệ sinh và tránh lây nhiễm.

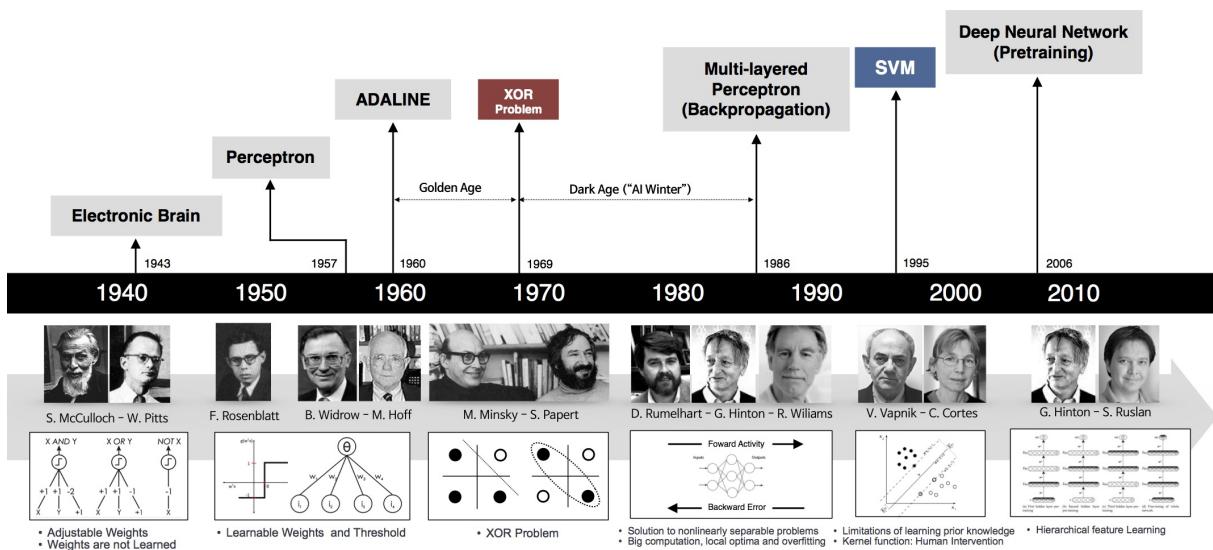
Vì vậy, việc nhận diện chính xác các bước trong hành động hút dịch nội khí quản là rất quan trọng để đảm bảo an toàn và hiệu quả cho người bệnh. Việc này giúp cho các nhân viên y tế có thể tối ưu hóa quá trình làm việc, cũng như giảm thiểu rủi ro cho người bệnh.

Trong bối cảnh công nghệ phát triển, việc ứng dụng các công nghệ nhận diện hành vi vào quá trình đào tạo và đánh giá kỹ năng của nhân viên y tế là một xu hướng tất yếu. Điều này giúp nâng cao chất lượng đào tạo, đảm bảo nhân viên y tế có đủ năng lực để thực hiện các thủ thuật y tế một cách an toàn và hiệu quả.

2.4 Tổng quan về học máy và trí tuệ nhân tạo

Trong cuộc sống hiện đại, trí tuệ nhân tạo ngày càng được phát triển và ứng dụng trong thực tế. Theo sự phát triển của thế giới, trí tuệ nhân tạo ngày càng được tối ưu và có được khả năng hiểu và phân tích một cách logic. Sự phát triển này của trí tuệ nhân tạo ngày càng thúc đẩy sự phát triển của con người dẫn tới một viễn cảnh con người và máy có thể kết hợp thực hiện các việc con người chưa từng nghĩ là có thể.

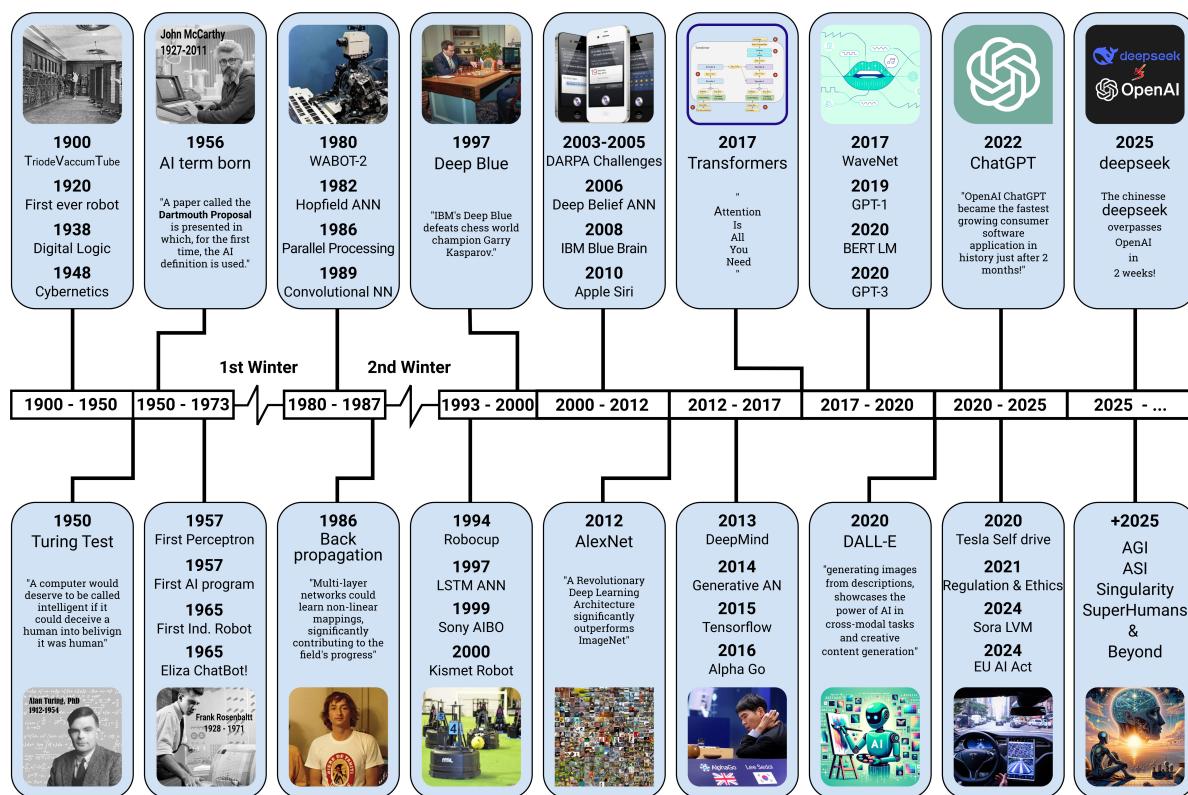
2.4.1 Lịch sử học máy



Hình 2.4: Sự phát triển của học máy cổ điển. Nhờ các bước phát triển học máy (machine learning) đầu tiên đạt kết quả rực rỡ, những ứng dụng phát triển về các hệ thống phức tạp hơn đã được nghiên cứu và phát triển.

Học máy, một nhánh của trí tuệ nhân tạo, ra đời với khát vọng tạo ra những hệ thống có khả năng học hỏi và tự cải thiện từ dữ liệu, giúp con người giải quyết được những vấn đề học búa. Ý tưởng này bắt nguồn từ mong muốn mô phỏng quá trình học tập của con người, nơi chúng ta rút ra bài học từ kinh nghiệm để đưa ra quyết định. Những năm 1950 đánh dấu bước khởi đầu khi các nhà khoa học bắt đầu nghiên cứu các thuật toán đơn giản (SVM, cây quyết định) cho phép máy tính nhận biết các mẫu trong dữ liệu. Tuy nhiên, do hạn chế về công suất tính toán và dung lượng dữ liệu, sự phát triển của học máy thời kỳ này còn khá khiêm tốn.

Sau một giai đoạn trầm lắng, học máy chứng kiến sự bùng nổ mạnh mẽ vào đầu thế kỷ 21. Sự gia tăng dữ liệu khổng lồ từ Internet và các thiết bị di động, cùng với sự phát triển của sức mạnh và tài nguyên tính toán, đã tạo điều kiện thuận lợi cho những đột phá trong lĩnh vực này. Các thuật toán học sâu (deep learning) dựa trên mạng thần kinh nhân tạo đã đạt được những thành công vượt trội trong nhiều lĩnh vực, từ nhận dạng hình ảnh đến xử lý ngôn ngữ tự nhiên, và đến giải thích được các sự vật hiện tượng trong cuộc sống của con người. Sự thành công của các ứng dụng học máy trong cuộc sống hàng ngày, như trợ lý ảo, xe tự lái, đã thúc đẩy sự đầu tư mạnh mẽ vào nghiên cứu và phát triển.



Hình 2.5: Sự phát triển của mô hình học sâu. Học sâu đã xuất hiện từ giữa thế kỷ XX với các mô hình có sự xuất hiện của perceptron. nhưng sớm gặp các thách thức và có khoảng dừng 10 năm. Đầu thế kỷ XXI, học sâu trở lại mạnh mẽ khi máy tính có sự phát triển khả năng tính toán mạnh, tạo tiền đề cho những mô hình lớn được khai sinh (Alexnet, Resnet, YOLO,...). Thế nhưng, DL đã thành công rực rỡ khi Transformer xuất hiện, khai mở khả năng tính toán song song và giảm khối lượng tính toán, từ đó mở ra kỷ nguyên AI như hiện tại.

Do những điểm ưu việt của học máy đem lại, con người đã bước vào một kỷ nguyên mới, nơi mà con người có thể thu thập những phản hồi và gợi ý từ các hệ thống thông minh. những gợi ý và báo cáo này không những giúp con người kiểm soát được cuộc sống mình tốt hơn mà còn mở ra nhiều vùng đất mới cho con người khám phá về bản thân (như nhu cầu hiểu về bản thân, quản lý năng suất của bản thân,...)

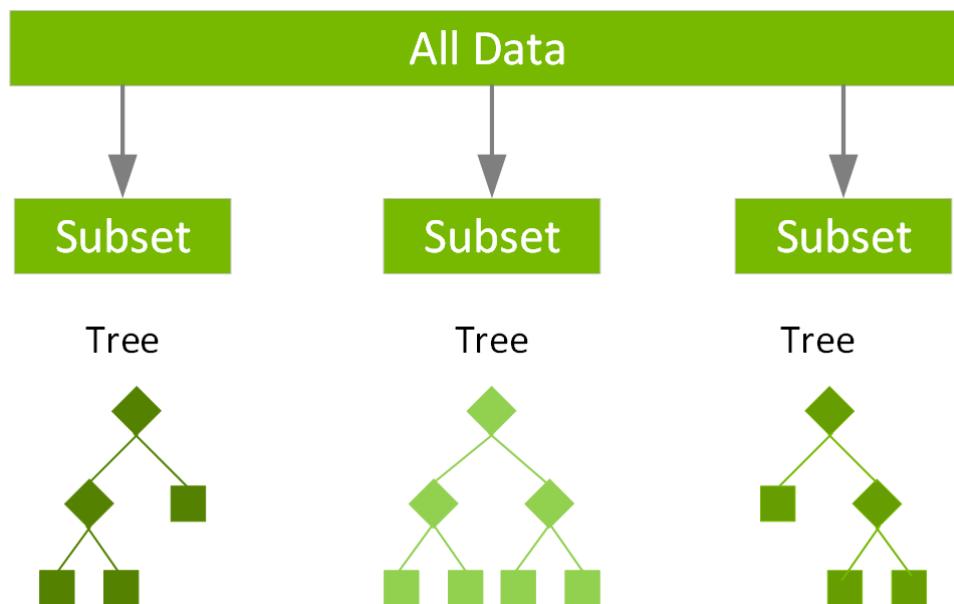
2.4.2 Học máy có giám sát

Học máy có giám sát là một nhánh của trí tuệ nhân tạo, nơi các mô hình máy học được huấn luyện trên một tập dữ liệu lớn, mỗi dữ liệu đều được gắn một nhãn cụ thể. Giống như một đứa trẻ học hỏi từ những ví dụ thực tế, các mô hình này sẽ dần rút ra các quy luật, mối quan hệ giữa dữ liệu và nhãn để có thể đưa ra dự đoán chính xác cho dữ liệu mới.

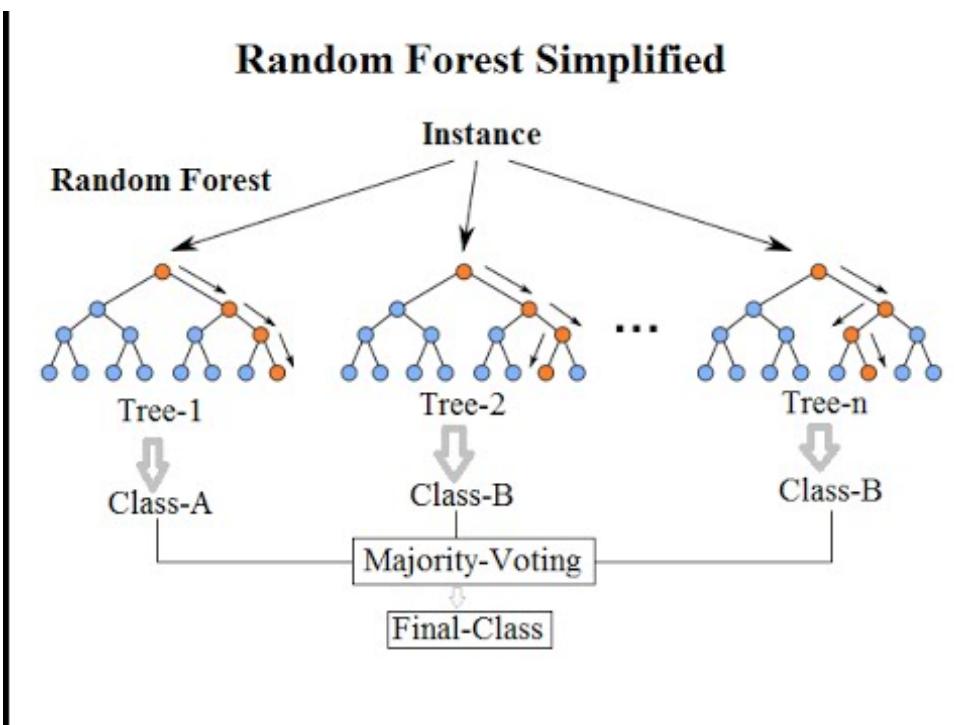
Để thực hiện quá trình học và dự đoán, nhiều thuật toán khác nhau đã được phát triển có thể kể đến như: XGBoost, Random Forest, hay HGBC.

Cụ thể hơn, XGBoost (hay Gradient Boosting)(tham khảo [14, 15] và Hình 2.6) là một thuật toán mạnh mẽ, kết hợp nhiều mô hình đơn giản để tạo thành một mô hình phức tạp hơn. Nó có

khả năng xử lý dữ liệu lớn và đạt được độ chính xác cao. Random Forest (RF hay Rừng ngẫu nhiên) (tham khảo [16, 17] và Hình 2.7), cũng là một thuật toán tập hợp (hay Bagging), tạo ra nhiều cây quyết định khác nhau, sau đó kết hợp kết quả của chúng để đưa ra quyết định cuối cùng. Random Forest có khả năng giảm thiểu quá khứ và rất ổn định.



Hình 2.6: Hình ảnh minh họa cách phân cây quyết định của XGBoost (Nguồn NVIDIA, [15])



Hình 2.7: Hình ảnh minh họa cách phân cây quyết định của Random Forest (Nguồn NVIDIA, [17])

Với sự phát triển không ngừng của học máy, các mô hình kết hợp được nghiên cứu và phát triển mạnh mẽ hơn. Cụ thể là, Histogram-Based Gradient Boosting Classifier (HGBC) [18],

một thuật toán học máy dựa trên Gradient Boosting và RF, được tạo ra nhằm tối ưu hóa quá trình huấn luyện bằng cách sử dụng histogram để chia nhỏ dữ liệu. Thay vì xem xét từng giá trị riêng lẻ, thuật toán ánh xạ các giá trị đặc trưng thành các nhóm (bins), giúp giảm độ phức tạp giải thuật đạt $O(T.M.N.2^D)$ của mô hình RF giảm xuống $O(T.M.B.2^D)$ (với M: số đặc trưng, B: số khoảng trong HGBC, D: chiều sâu của cây quyết định, T: số lượng cây trong mô hình, N: là lượng dữ liệu). Cách tiếp cận này không chỉ tăng tốc độ huấn luyện mà còn giảm đáng kể mức tiêu thụ bộ nhớ, giúp mô hình hoạt động hiệu quả hơn trên tập dữ liệu lớn (do $B \ll N$). Một ưu điểm quan trọng khác của HGB Classifier là khả năng tự động điều chỉnh độ phức tạp của mô hình thông qua các siêu tham số như số lượng bins, learning rate, số lượng cây và độ sâu tối đa. So với các phương pháp boosting truyền thống như XGBoost hay LightGBM, HGBC có thể đạt hiệu suất tương đương hoặc tốt hơn trong một số trường hợp, đặc biệt là khi làm việc với dữ liệu có kích thước lớn. Nhờ những cải tiến về tốc độ và hiệu suất, HGBC ngày càng được ưa chuộng trong các bài toán phân loại sử dụng mô hình boosting.

2.4.3 Học máy không giám sát

Nếu học máy có giám sát kể trên là quá trình máy tính học cách gán nhãn cho những dữ liệu có nhãn do ta gán sẵn, thì học máy không giám sát lại là một câu chuyện ngược lại. Ở đây, máy tính sẽ tự mình khám phá, tìm kiếm các mẫu ẩn, các cấu trúc tiềm ẩn trong dữ liệu mà không cần bất kỳ nhãn nào.

Một trong những nhiệm vụ phổ biến nhất của học máy không giám sát là clustering, hay còn gọi là phân cụm. Clustering giúp chia dữ liệu thành các nhóm (cụm) sao cho các điểm dữ liệu trong cùng một cụm có sự tương đồng cao, trong khi các điểm dữ liệu thuộc các cụm khác nhau lại khác biệt. Việc phân cụm này giúp cho nhà nghiên cứu có thể quan sát rõ hơn cách hành vi của một hoặc nhiều nhóm dữ liệu có sự tương đồng từ đó hiểu hơn về các nhóm dữ liệu.

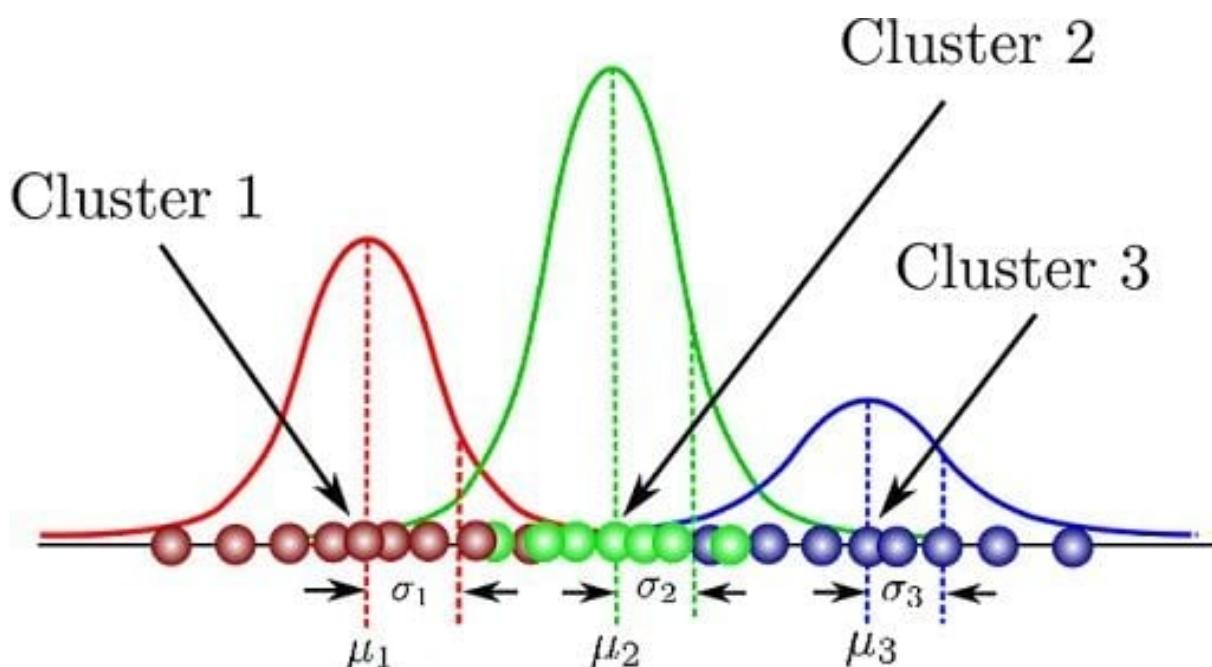
Về clustering, các thuật toán nổi trội có thể kể đến như kMEANS, DBSCAN, hoặc Gaussian Mixture Models (GMM). Các thuật toán này đã mở ra khả năng cho việc gom cụm dữ liệu phục vụ và phát triển cho ngành học máy nói chung.

2.4.4 Học máy tăng cường

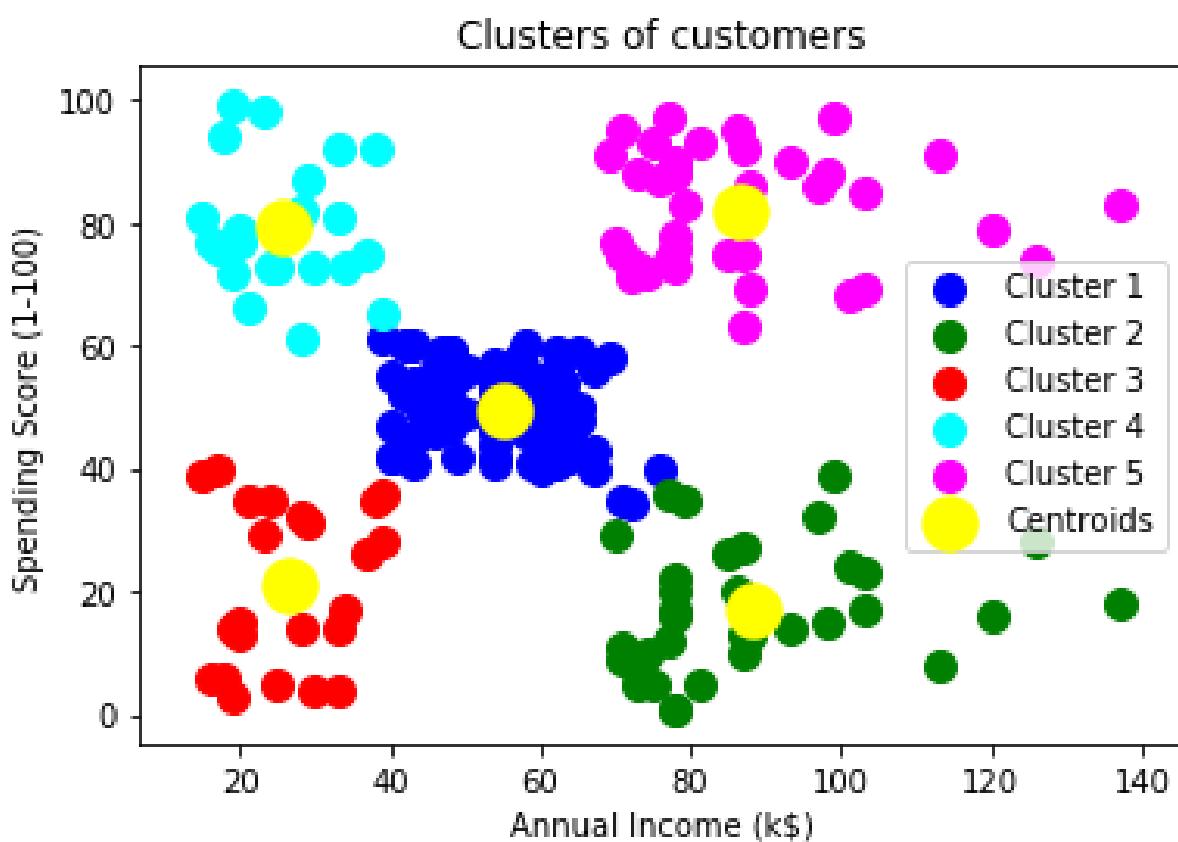
Nếu ví học giám sát là máy được học dữ liệu qua nhãn dán sẵn, học không giám sát là máy tự tìm ra đặc trưng và gom cụm các dữ liệu có tính tương đồng, thì học tăng cường lại giống như phép thử và làm lại. Nói cách khác, mô hình học tăng cường không chỉ đơn thuần phân loại hay tìm kiếm các mẫu ẩn mà còn tương tác trực tiếp với môi trường. Mô hình sẽ thực hiện các hành động và nhận lại phản thưởng hoặc hình phạt (qua hàm nhận và giá cả (gain and loss function)) dựa trên kết quả của hành động đó. Qua quá trình thử và sai liên tục, mô hình sẽ học cách đưa ra những quyết định tối ưu để đạt được mục tiêu đã đặt ra.

2.4.5 Giảm chiều dữ liệu

Trong quá trình học máy, việc quá nhiều chiều dữ liệu sẽ gây cho mô hình sinh ra nhiều sự kết hợp các đặc trưng để đưa ra kết quả, từ đó làm mô hình trở nên rối loạn và nguy hiểm hơn là không thể tạo ra được một mô hình phân loại tổng quát. Do đó, các kỹ sư đã tạo ra các



Hình 2.8: Hình ảnh minh họa mô hình GMM



Hình 2.9: Hình ảnh minh họa mô hình K-Means

công cụ để giảm số chiều dữ liệu như phân tích thành phần chính (PCA [19, 20, 21]), phân tích phân biệt tuyến tính (LDA [22]), các chỉ số thống kê (điểm F, điểm T, ANOVA [1]) và điểm SHAPley [23, 24].

Việc giảm số chiều dữ liệu không chỉ giúp cải thiện hiệu suất tính toán mà còn làm cho mô hình trở nên dễ hiểu hơn. Bằng cách áp dụng các kỹ thuật giảm chiều như PCA, các nhà nghiên

cứu có thể tách ra các thành phần chính, từ đó tập trung vào những biến số quan trọng nhất, loại bỏ nhiễu và giảm thiểu tác động của những đặc trưng không cần thiết. Tương tự, LDA cho phép tìm ra các đường biên tuyến tính tối ưu giữa các lớp dữ liệu, giúp cải thiện khả năng phân biệt giữa các nhóm đối tượng.

Các chỉ số thống kê như điểm F, điểm T và ANOVA cung cấp thông tin quý giá về mức độ phân biệt của từng đặc trưng, hỗ trợ quá trình lựa chọn các đặc trưng có ý nghĩa cho mô hình. Hơn nữa, việc sử dụng các giá trị SHAPley (2.1) mang lại khả năng giải thích rõ ràng về tác động của từng đặc trưng đối với dự đoán của mô hình, tạo ra một hệ thống phân loại minh bạch và dễ dàng được đánh giá.

$$\phi_i(x) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} \left[f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S) \right] \quad (2.1)$$

Trong đó:

- x : quan sát đầu vào.
- F : tập hợp tất cả các đặc trưng.
- f_S : mô hình đã được huấn luyện trên tập con các đặc trưng S .
- $f_{S \cup \{i\}}$: mô hình đã được huấn luyện trên tập con các đặc trưng $S \cup \{i\}$.
- x_S : đầu vào x được ràng buộc theo tập con các đặc trưng S .
- $x_{S \cup \{i\}}$: đầu vào x được ràng buộc theo tập con các đặc trưng $S \cup \{i\}$.

Nhờ vào những công cụ giảm chiều dữ liệu này, các nhà nghiên cứu và kỹ sư học máy có thể xây dựng các mô hình phân loại tổng quát hơn, vừa hiệu quả về mặt tính toán, vừa có khả năng giải thích và đưa ra các quyết định dựa trên dữ liệu một cách chính xác, góp phần nâng cao độ tin cậy của hệ thống trong các ứng dụng thực tiễn.

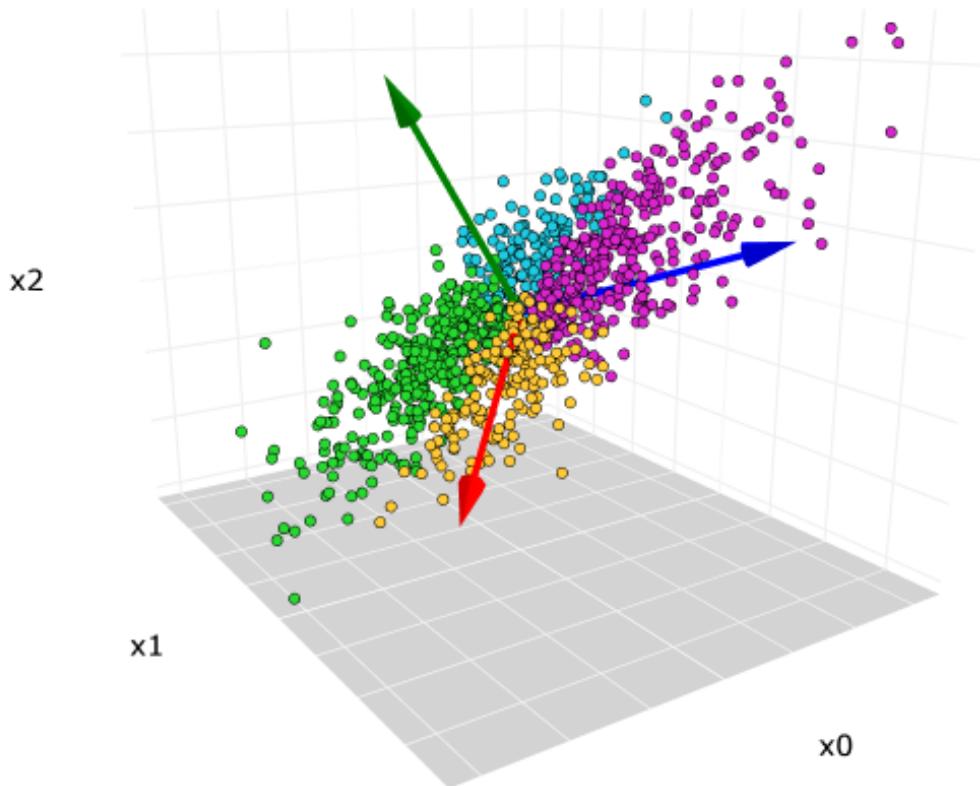
2.5 Học sâu

Khi nhắc tới học máy hiện đại, học sâu (hay Deep Learning) lây được sự chú ý trong thời điểm gần đây. Lấy cảm hứng từ cấu trúc và chức năng của não người, mạng học sâu đã tạo và sử dụng các mạng thần kinh nhân tạo với nhiều lớp để học từ dữ liệu, tìm ra các đặc trưng phức tạp và đưa ra quyết định thông qua điều chỉnh trọng số W và hàm mất mát L .

$$\mathbf{W}^{(l+1)} \leftarrow \mathbf{W}^{(l)} - \eta \frac{\partial L}{\partial \mathbf{W}^{(l)}} \quad (2.2)$$

Với học sâu, mạng này có thể tạo thực hiện những công việc từ phân loại (CNN, RNN), gom cụm (NN-clustering) và cũng có thể học tăng cường (Deep Q-Networks (DQN)). Ngoài ra, học sâu còn có một khả năng ưu việt khác là trích xuất dữ liệu - một điều các giải thuật học máy thông thường không thể.

Vì vậy học sâu đang ngày càng mở rộng tầm ảnh hưởng trong thực tế và là công cụ được tin dùng hàng đầu trong kỷ nguyên này.



Hình 2.10: Hình ảnh minh họa của PCA. Ta có thể thấy dữ liệu cho ban đầu có 3 chiều dẫn đến chi phí tính toán cao. bằng cách phân tích thành phần và giữ lại thành phần chính (vector xanh dương và vector đỏ), ta vẫn có thể phân loại 4 nhóm phân loại trên.

2.5.1 Các toán tử trong học sâu

2.5.1.a Toán tử tích chập

Toán tử tích chập là một công cụ toán học mạnh mẽ, không chỉ là nền tảng của xử lý tín hiệu mà còn là “trái tim” của các hệ thống học sâu hiện đại. Cụ thể, tích chập giữa hai hàm số $f(t)$ và $g(t)$ được định nghĩa bởi

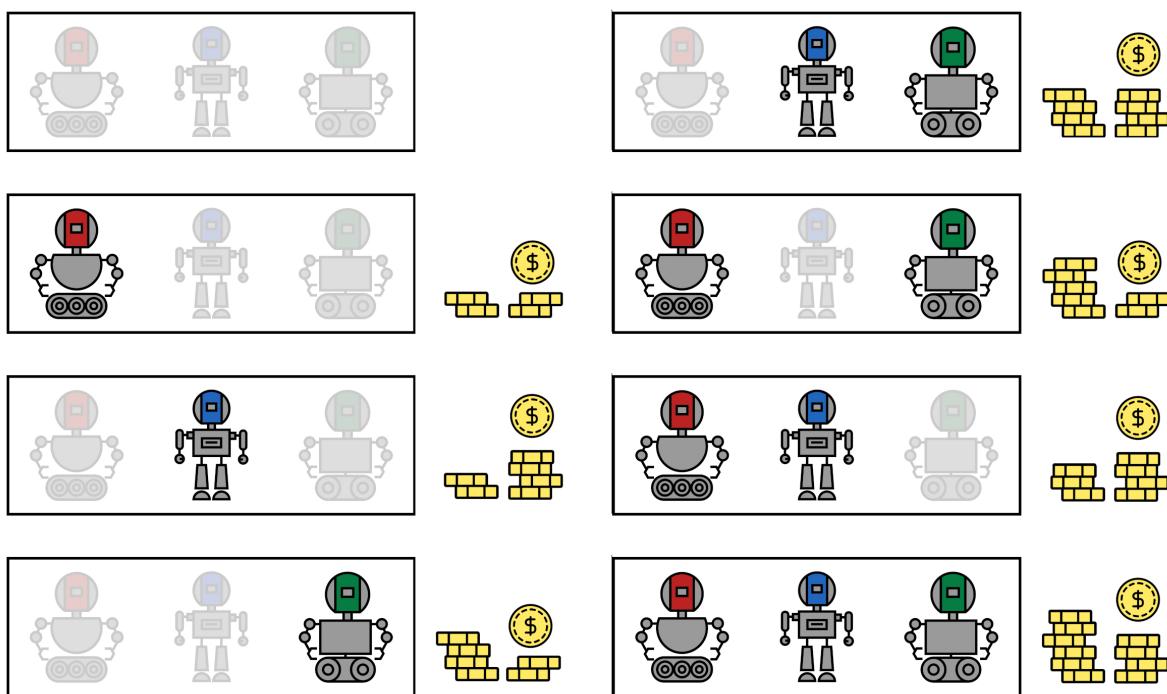
$$(f * g)(t) = \int_{-\infty}^{+\infty} f(\tau) g(t - \tau) d\tau.$$

Đây là quá trình “quét” một hàm qua hàm kia, thu nhận các đặc trưng cốt lõi của tín hiệu thông qua việc nhân và tích phân, từ đó phát hiện những mẫu thông tin ẩn giấu. Tính chất giao hoán của tích chập, $f * g = g * f$, giúp đơn giản hóa việc phân tích và thiết kế các bộ lọc tín hiệu phức tạp.

Không chỉ vậy, tính kết hợp của tích chập, được thể hiện qua công thức

$$f * (g * h) = (f * g) * h,$$

giúp xây dựng các hệ thống xử lý đa tầng, nơi mỗi tầng có thể được tối ưu riêng biệt để nâng



Hình 2.11: Hình minh họa của SHAP. SHAP ở đây sẽ hoạt động bằng cách lấy bên phải và truy xuất về từng robot làm được bao nhiêu tiền. Trong lĩnh vực học máy, SHAP sẽ đóng vai trò truy xuất sự đóng góp của từng biến đến kết quả cuối cùng theo nguyên lý trò chơi (game theory), từ đó người kỹ sư có thể nhìn vào thông số đó để biết đóng góp của từng biến nhất định để chọn các biến phù hợp.

cao chất lượng trích xuất đặc trưng. Trong miền Fourier, tích chập chuyển hóa thành phép nhân:

$$\mathcal{F}\{f * g\} = \mathcal{F}\{f\} \cdot \mathcal{F}\{g\},$$

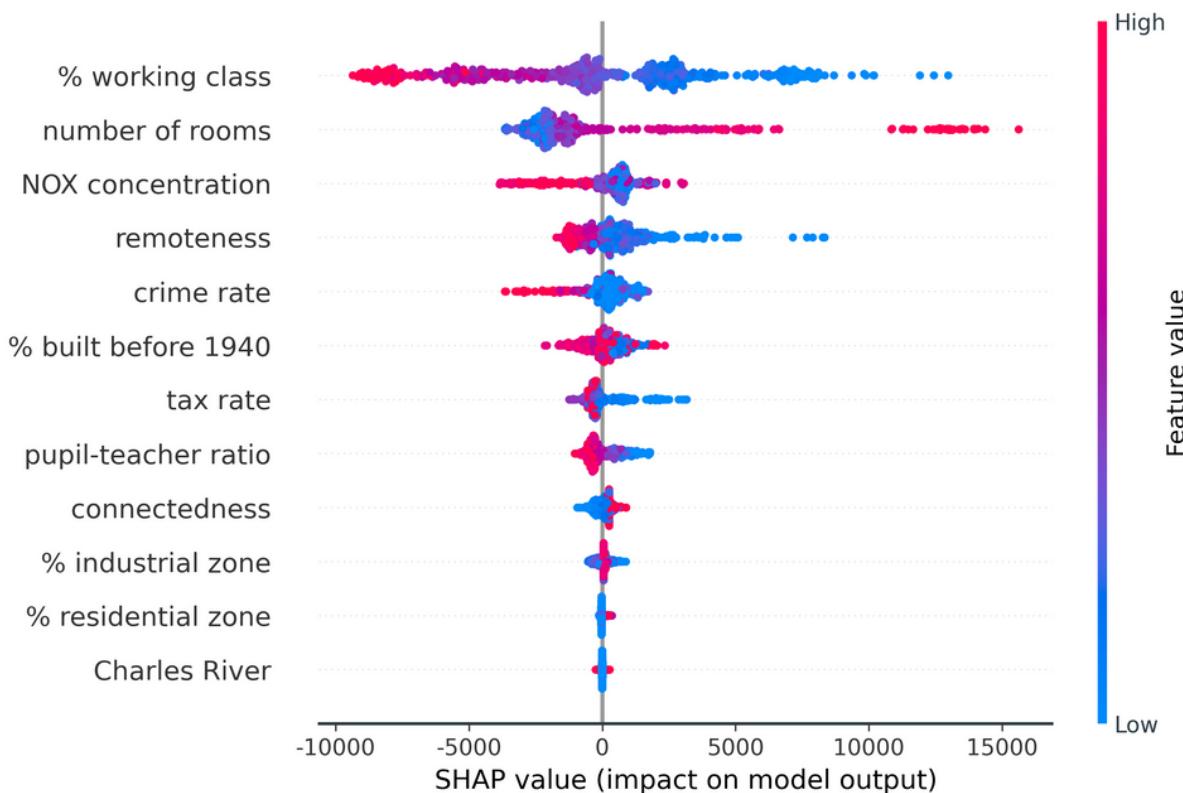
một mối liên hệ giúp tăng tốc độ tính toán và giảm thiểu độ phức tạp của các bài toán xử lý tín hiệu.

Trong lĩnh vực xử lý ảnh, toán tử tích chập là công cụ chủ đạo cho các phép lọc như làm mờ, phát hiện biên, và tăng cường độ tương phản, mang lại hiệu quả vượt trội trong việc cải thiện chất lượng hình ảnh. Đặc biệt, trong các mô hình mạng nơ-ron tích chập (CNN), lớp tích chập không chỉ giúp trích xuất đặc trưng tự động mà còn giảm thiểu số lượng tham số, tạo nên các hệ thống nhận dạng hình ảnh với độ chính xác và tốc độ ấn tượng.

Nhờ vào khả năng “tóm gọn” và khai thác tối đa thông tin từ dữ liệu, toán tử tích chập đã tạo ra bước đột phá trong các ứng dụng công nghệ cao, góp phần cách mạng hóa các lĩnh vực như trí tuệ nhân tạo, y tế số và các hệ thống tự động thông minh.

2.5.1.b Toán tử Pool

Trong học sâu, toán tử *pooling* được sử dụng để giảm kích thước không gian của các đặc trưng được trích xuất sau các lớp tích chập, từ đó giúp giảm số lượng tham số và kiểm soát overfitting. Hai phương pháp pooling phổ biến là *Max Pooling* và *Mean Pooling*.



Hình 2.12: Hình minh họa của trực quan hóa tác động của một thành phần bằng SHAP.

Max Pooling: Cho X là ma trận đầu vào (ví dụ, một ma trận đặc trưng) và $R_{i,j}$ là vùng lân cận được chia nhỏ, kết quả sau lớp max pooling được tính bằng:

$$Y_{i,j} = \max_{(p,q) \in R_{i,j}} X_{p,q}.$$

cho biết giá trị đầu ra tại vị trí (i, j) là giá trị lớn nhất trong vùng $R_{i,j}$.

Mean Pooling (Average Pooling): Tương tự, với mean pooling, giá trị đầu ra tại vị trí (i, j) được tính bằng trung bình cộng của các giá trị trong vùng $R_{i,j}$:

$$Y_{i,j} = \frac{1}{|R_{i,j}|} \sum_{(p,q) \in R_{i,j}} X_{p,q},$$

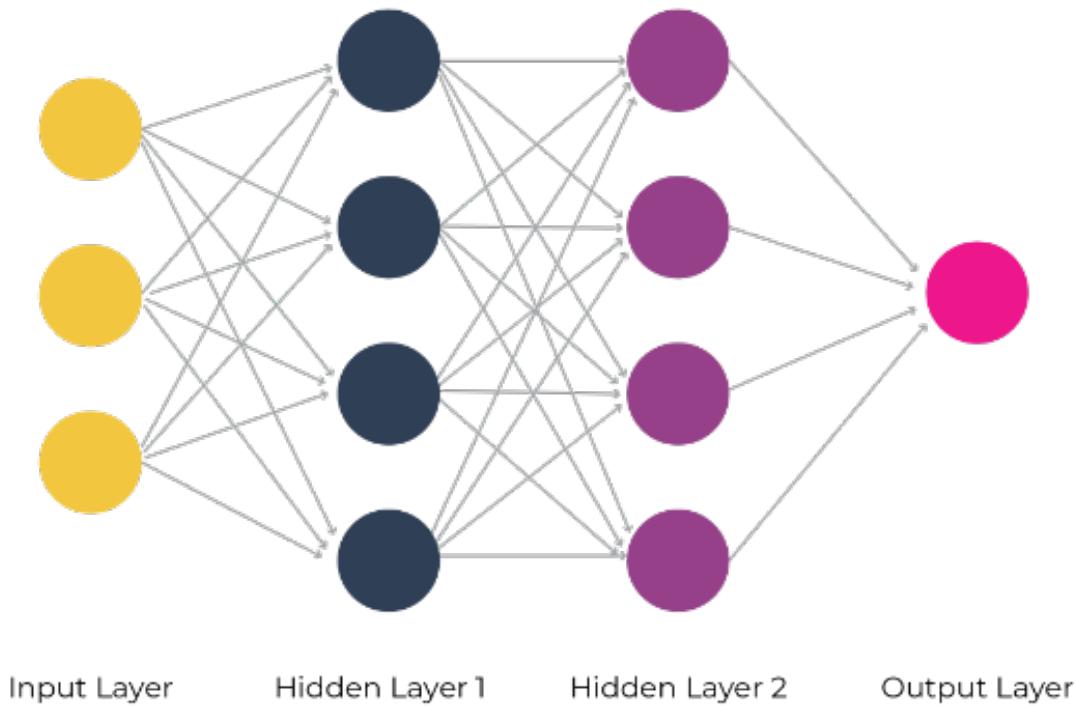
trong đó $|R_{i,j}|$ là số lượng phần tử trong vùng $R_{i,j}$.

Hai phương pháp này giúp giảm số chiều dữ liệu và giữ lại các đặc trưng quan trọng, góp phần tăng cường hiệu quả xử lý và giảm độ phức tạp của mô hình học sâu.

2.5.1.c Toán tử Padding

Trong các mô hình học sâu, toán tử *padding* (đệm) được sử dụng để mở rộng kích thước của đầu vào (ví dụ: hình ảnh hoặc ma trận đặc trưng) bằng cách thêm các giá trị (thường là 0) vào biên. Việc này giúp duy trì thông tin biên sau khi thực hiện các phép tính tích chập và kiểm soát kích thước của đầu ra sau khi áp dụng các bộ lọc.

Ví dụ, giả sử đầu vào ban đầu là một ma trận X kích thước $m \times n$. Nếu áp dụng padding với



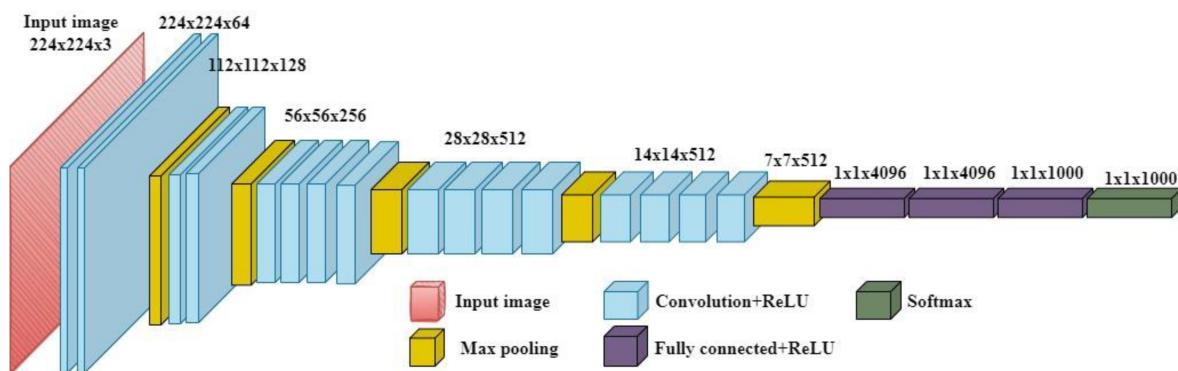
Hình 2.13: Hình ảnh minh họa mạng thần kinh nhân tạo (NN). Đây là nền tảng cho sự phát triển cho trí tuệ nhân tạo tại thời điểm hiện tại. NN được lý giải là sự tối ưu hoá của các node thần kinh ở mỗi lớp (layer) qua nguyên lý đạo hàm và nâng cấp (update) trọng số. Ngoài ra một vấn đề quan trọng trong NN là việc chọn chính xác hàm kích hoạt (activation function), và hệ số học (learning rate hay ϵ) để việc tối ưu hoá được tối ưu.

số hàng và số cột thêm vào là p ở trên và dưới, q ở trái và phải, thì đầu ra sau khi padding sẽ có kích thước:

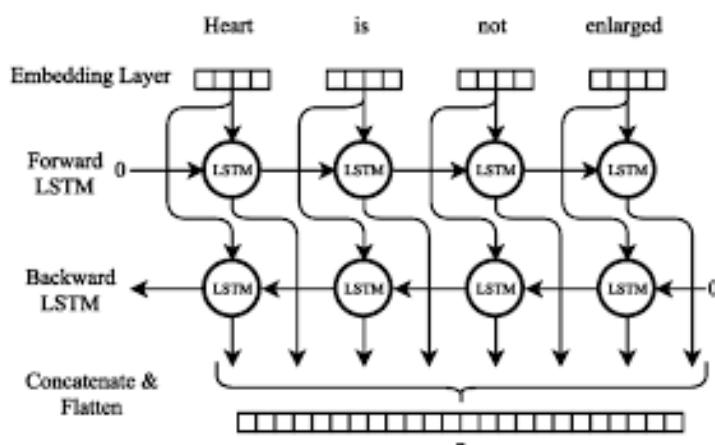
$$(m + 2p) \times (n + 2q).$$

Công thức biểu diễn dạng tổng quát khi sử dụng zero padding là:

$$X_{\text{pad}}(i, j) = \begin{cases} 0, & \text{nếu } i \leq p \text{ hoặc } i > m + p, \text{ hoặc } j \leq q \text{ hoặc } j > n + q, \\ X(i - p, j - q), & \text{không thì.} \end{cases}$$



Hình 2.14: Hình ảnh minh họa mạng thần kinh VGG 16 (thuộc nhóm CNN).



Hình 2.15: Hình ảnh minh họa mạng thần kinh BiLSTM (thuộc nhóm RNN).

Phương pháp padding không chỉ giúp duy trì kích thước của đặc trưng khi áp dụng các lớp tích chập mà còn có thể cải thiện hiệu suất học của mô hình bằng cách giữ lại các thông tin biên quan trọng.

2.5.1.d Toán tử chuẩn hoá

Trong học sâu, *toán tử chuẩn hoá* hay *Batch Normalization* là một kỹ thuật được sử dụng để giảm thiểu vấn đề biến đổi nội bộ (internal covariate shift) trong quá trình huấn luyện. Batch Normalization giúp mô hình học ổn định hơn và hội tụ nhanh hơn bằng cách chuẩn hoá các đầu vào của mỗi lớp theo từng lô dữ liệu.

Giả sử ta có một lô dữ liệu $B = \{x_1, x_2, \dots, x_m\}$ với m là số mẫu trong lô. Quá trình chuẩn hoá theo lô được thực hiện qua các bước sau:

1. Tính trung bình của lô dữ liệu:

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i.$$

2. Tính phương sai của lô dữ liệu:

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2.$$

3. Chuẩn hoá từng mẫu:

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}},$$

trong đó ϵ là hằng số vô cùng nhỏ nhằm đảm bảo không chia cho 0.

4. Cuối cùng, áp dụng phép biến đổi tuyến tính với các tham số có thể học được γ và β để phục hồi lại khả năng biểu diễn của mô hình:

$$y_i = \gamma \hat{x}_i + \beta.$$

Toán tử Batch Normalization không chỉ giúp cân bằng lại phân phối của đầu vào cho các lớp tiếp theo mà còn có tác dụng như một dạng regularization, giúp giảm hiện tượng overfitting. Kỹ thuật này đã được ứng dụng rộng rãi trong nhiều kiến trúc mạng nơ-ron, từ các mô hình CNN đến các mạng nơ-ron hồi tiếp (RNN), góp phần cải thiện đáng kể hiệu suất của các mô hình học sâu.

2.5.1.e Hàm kích hoạt

Trong mạng nơ-ron nhân tạo (Artificial Neural Network - ANN), *hàm kích hoạt* đóng vai trò quan trọng trong việc tạo ra tính phi tuyến tính, giúp mạng có khả năng học các quan hệ phức tạp giữa đầu vào và đầu ra. Nếu không có hàm kích hoạt, mô hình chỉ đơn giản trở thành một tổ hợp tuyến tính của các đầu vào, làm giảm đáng kể khả năng biểu diễn của mạng.

Hàm Sigmoid [25] là một trong những hàm kích hoạt cổ điển, có dạng:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.3)$$

Hàm này đưa đầu ra về khoảng $(0, 1)$, phù hợp với các bài toán phân loại. Tuy nhiên, nó có nhược điểm là làm giảm độ lớn của gradient khi giá trị tuyệt đối của đầu vào lớn, gây ra vấn đề vanishing gradient. Ngoài ra, đạo hàm của Sigmoid là:

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)) \quad (2.4)$$

Do giá trị của $\sigma(x)$ luôn nằm trong khoảng $(0, 1)$, đạo hàm này có thể rất nhỏ, gây khó khăn trong việc cập nhật trọng số khi huấn luyện mạng sâu theo công thức (2.2).

Hàm Tanh [26] là một biến thể khác của Sigmoid, với công thức:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.5)$$

Khác với Sigmoid, hàm Tanh có đầu ra nằm trong khoảng $(-1, 1)$, giúp giảm độ lệch trung bình

của dữ liệu về 0. Tuy nhiên, nó vẫn gặp vấn đề vanishing gradient như Sigmoid. Với đạo hàm là:

$$\tanh'(x) = 1 - \tanh^2(x) \quad (2.6)$$

Tức có gradient lớn hơn Sigmoid, nhưng khi đầu vào có giá trị lớn, đạo hàm vẫn tiến gần về 0.

Hàm ReLU (Rectified Linear Unit) [27] được sử dụng rộng rãi trong deep learning vì tính đơn giản và hiệu quả:

$$ReLU(x) = \max(0, x) \quad (2.7)$$

ReLU giúp giảm bớt vấn đề vanishing gradient, nhưng có thể gây ra vấn đề *dead neurons*, khi gradient trở thành 0 đối với các giá trị âm. Đạo hàm của ReLU là:

$$ReLU'(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (2.8)$$

Điều này có nghĩa là nếu một nơ-ron rơi vào vùng $x \leq 0$, nó sẽ không cập nhật trọng số trong quá trình huấn luyện.

Hàm Leaky ReLU [28] là một cải tiến của ReLU, cho phép một giá trị âm nhỏ khi đầu vào nhỏ hơn 0:

$$LReLU(x) = \max(\alpha x, x), \quad \text{với } \alpha > 0 \quad (2.9)$$

Đạo hàm của Leaky ReLU là:

$$LReLU'(x) = \begin{cases} 1, & x > 0 \\ \alpha, & x \leq 0 \end{cases} \quad (2.10)$$

Với α thường được chọn là một số nhỏ như 0.01, Leaky ReLU giúp hạn chế vấn đề dead neurons của ReLU.

Hàm Softmax [29] thường được sử dụng trong lớp đầu ra của mạng phân loại nhiều lớp:

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}} \quad (2.11)$$

Đạo hàm của Softmax không đơn giản như các hàm trên mà được tính theo ma trận Jacobian:

$$\frac{\partial \text{Softmax}(x_i)}{\partial x_j} = \text{Softmax}(x_i)(\delta_{ij} - \text{Softmax}(x_j)) \quad (2.12)$$

với δ_{ij} là hàm Kronecker. Điều này giúp Softmax tạo ra sự cạnh tranh giữa các lớp đầu ra.

Mỗi hàm kích hoạt có ưu nhược điểm riêng, và việc lựa chọn hàm phù hợp tùy thuộc vào bài toán cụ thể. Trong thực tế, ReLU và các biến thể của nó thường được ưa chuộng trong các mô hình deep learning hiện đại.

2.5.1.f Toán tử kết nối toàn diện

Trong mạng nơ-ron nhân tạo, *kết nối toàn diện* (Fully Connected - FC) là một lớp quan trọng, trong đó mỗi nơ-ron ở lớp trước được liên kết với mọi nơ-ron ở lớp sau. Điều này giúp mô hình

có khả năng học biểu diễn phức tạp bằng cách tổng hợp toàn bộ thông tin từ lớp trước. Công thức toán học mô tả toán tử FC có dạng:

$$\mathbf{y} = \mathbf{Wx} + \mathbf{b} \quad (2.13)$$

trong đó \mathbf{x} là đầu vào, \mathbf{W} là ma trận trọng số, \mathbf{b} là vector hệ số điều chỉnh (bias), và \mathbf{y} là đầu ra.

Việc nâng cấp hệ số trong FC chủ yếu tập trung vào tối ưu hóa trọng số \mathbf{W} và bias \mathbf{b} thông qua quá trình huấn luyện. Một số kỹ thuật phổ biến gồm:

Chuẩn hóa trọng số: Giúp duy trì ổn định độ lớn của trọng số, tránh hiện tượng bùng nổ hoặc tiêu biến gradient.

Dropout: Ngẫu nhiên loại bỏ một số nơ-ron trong quá trình huấn luyện để giảm overfitting.

Batch Normalization: Chuẩn hóa đầu vào của từng lớp giúp tăng tốc độ hội tụ và ổn định huấn luyện.

Dù có khả năng biểu diễn mạnh mẽ, lớp FC có số lượng tham số lớn, dẫn đến chi phí tính toán cao. Do đó, các mô hình hiện đại thường kết hợp FC với các lớp tích chập để tối ưu hóa hiệu suất và khả năng khai quát hóa của mạng.

2.6 Tạo sinh dữ liệu

2.6.1 Tạo sinh dữ liệu bằng công cụ thống kê

Trong lĩnh vực tạo sinh dữ liệu, các hàm toán học được ứng dụng triệt để nhằm ghi nhận và tái tạo lại dữ liệu. Cụ thể hơn, ta có thể dùng phương pháp Gaussian Copulas (2.14) để tạo sinh dữ liệu [30]. Phương pháp này sử dụng các hàm phân phối luỹ $\Phi_{\mathbf{R}}$ của các hàm phân phối đơn biến Φ^{-1} để tạo ra những giá trị của dữ liệu tạo sinh. Phương pháp này đặc biệt mạnh mẽ trong việc tạo sinh dữ liệu và được áp dụng trong các ứng dụng cần độ tin cậy cao

$$C_{\mathbf{R}}(u_1, \dots, u_d) = \Phi_{\mathbf{R}}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)) \quad (2.14)$$

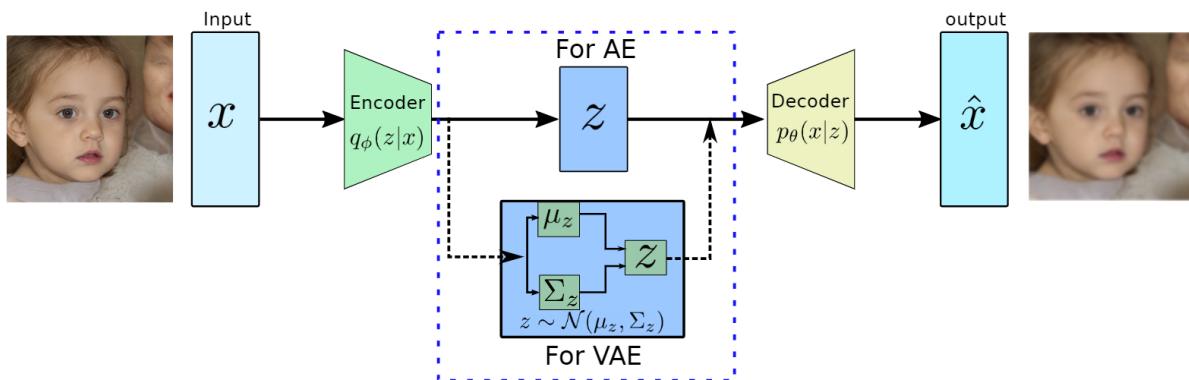
Ngoài phương pháp Gaussian Copulas, một phương pháp tạo sinh dữ liệu toán học khác cũng có thể kể tên như tạo sinh bằng biến đổi Fourier (T-Gen) (2.15). Thay vì trực tiếp tạo ra các giá trị trong miền thời gian, T-Gen tiếp cận vấn đề từ góc độ tần số, tận dụng sức mạnh của DFT để phân tích và tái tạo chuỗi thời gian. Cụ thể, chuỗi thời gian đầu vào được chuyển đổi sang miền tần số thông qua DFT, biểu diễn dưới dạng các hệ số tần số phức tạp, nắm bắt thông tin về biên độ và pha của các thành phần tần số. Sau đó, quá trình tạo dữ liệu diễn ra trong miền tần số, nơi các hệ số tần số được lấy mẫu từ một phân phối xác suất, cho phép kiểm soát linh hoạt các đặc tính tần số mong muốn. Cuối cùng, biến đổi Fourier rời rạc ngược (IDFT) được áp dụng để tái tạo chuỗi thời gian từ các hệ số tần số đã được tạo ra, mang lại một chuỗi thời gian mới với các đặc tính được điều chỉnh.

$$G = F^{-1}(Norm(F(x))) \quad (2.15)$$

với F là toán tử Fourier và F^{-1} là toán tử nghịch đảo Fourier

2.6.2 Tạo sinh dữ liệu bằng công cụ học sâu

Các công cụ học sâu đang lấn sân vào nhiều lĩnh vực tại thời điểm hiện tại. Ngoài lĩnh vực nhận biết đã được trình bày tại 2.4, công nghệ học sâu còn được áp dụng cho các ứng dụng tạo sinh dữ liệu như mô hình VAE, GAN, hay mô hình khuếch tán nhằm tạo nên những dữ liệu có độ chính xác cao và đa dạng.



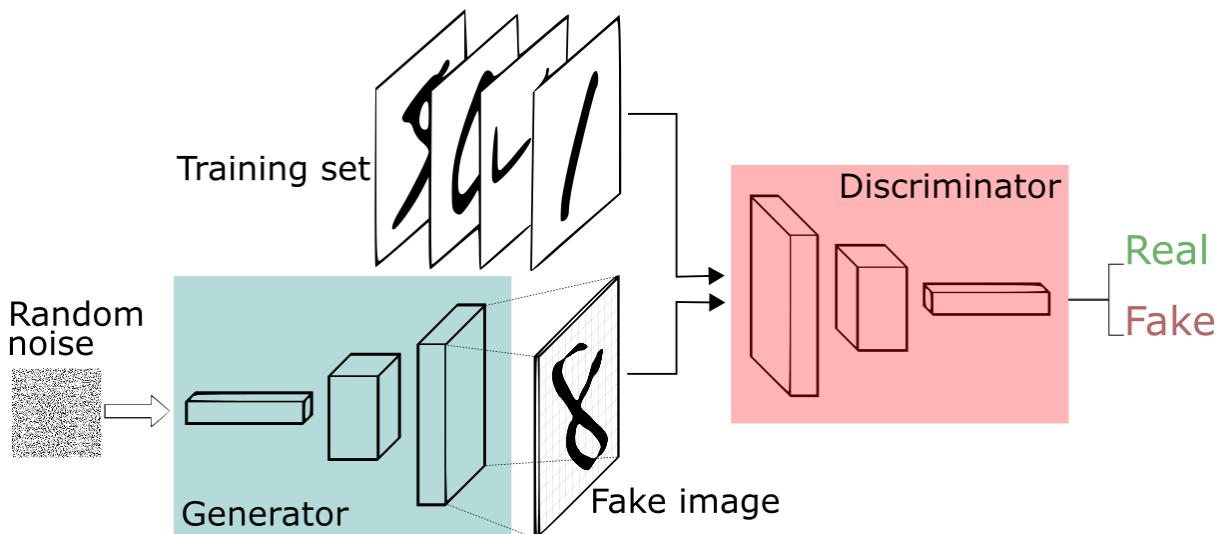
Hình 2.16: Cấu trúc của mô hình VAE (một ví dụ điển hình trong hệ mô hình AE hay Auto Encoder). VAE tận dụng hiệu ứng nút cổ chai (bottle neck effect) nhằm đưa dữ liệu về một không gian đặc trưng nhỏ hơn nhằm cô đọng thông tin (thông tin được chuyển về còn $Z, \sigma(z), \bar{z}$) ở bước mã hóa. Sau đó mô hình sẽ tái lập hình ảnh thông qua mô hình giải mã để tạo ra hình ảnh. quy trình này sẽ được huấn luyện để tạo ra hình ảnh có sự tương đồng về ngữ nghĩa với thông tin cô đọng từ bước mã hóa.

Một trong những mô hình đó - GAN - hay mạng tạo sinh đối ngẫu [31] tận dụng khả năng co hẹp và nhân rộng (Dense hay Fully connected) để tạo nên một mạng thần kinh tạo dữ liệu giống thật. Cụ thể hơn mạng thần kinh này sẽ sử dụng một mạng sinh dữ liệu (Generator hay G) để tạo dữ liệu có hàm mất mát \mathcal{L}_G , sau đó dữ liệu này sẽ được mạng kiểm dữ liệu (Discriminator hay D) để nhận diện dữ liệu tạo ra có thể bị phân biệt với dữ liệu thật có hàm mất mát \mathcal{L}_D . Nếu mạng D có thể phân biệt dữ liệu thật, tức giá trị hàm mất mát \mathcal{L}_D cao dẫn tới \mathcal{L}_G , mạng G sẽ bị phạt và phải học thêm cách tạo dữ liệu thực tế hơn

$$\mathcal{L}_G = -\log(D(G(z))) \quad (2.16)$$

$$\mathcal{L}_D = -[y \cdot \log(D(x)) + (1 - y) \cdot \log(1 - D(G(z)))] \quad (2.17)$$

với $D(x)$ là đầu ra của bộ phân biệt cho dữ liệu thật x và $D(G(z))$ là đầu ra của bộ phân biệt cho dữ liệu tạo sinh z. Điều này dẫn đến GAN sẽ tối ưu việc tạo sinh bằng cách hạn chế \mathcal{L}_G và tăng cao \mathcal{L}_D



Hình 2.17: Cấu trúc của mô hình GAN. Mô hình GAN được ứng dụng rộng rãi trong tạo sinh dữ liệu do độ tin cậy mà mạng D mang lại. Do có mạng D, mô hình tạo sinh dữ liệu tự đánh giá chất lượng hình ảnh mà nó tạo ra trước khi trả về dữ liệu cho người dùng. Vì thế, GAN thường tạo ra các dữ liệu có chất lượng cao nhưng cũng vì vậy độ phủ của GAN thường khá thấp.

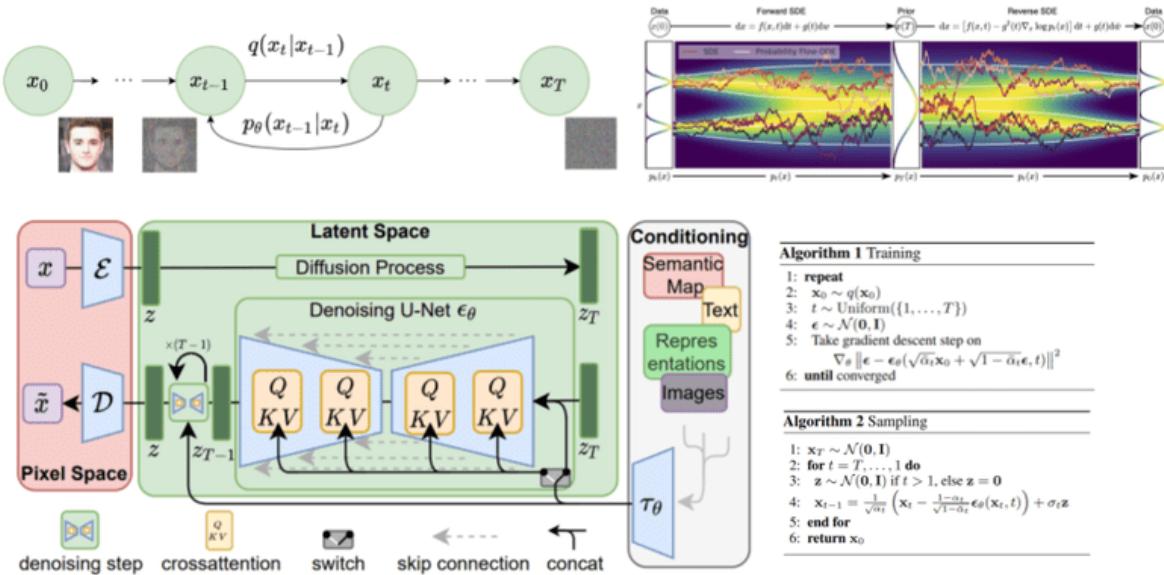
2.7 Các nghiên cứu liên quan

2.7.1 Các nghiên cứu về tính khả thi

Việc nhận diện hành động bằng camera đang thu hút được sự quan tâm, nhóm của Fortes [11] ứng dụng hình ảnh từ camera, chuyển hóa dữ liệu của từng người sang dạng 1 người chuẩn có chiều cao bằng 1m70 bằng SMPL. Từ đó, ứng dụng các giải thuật, nhóm của Fortes đã thành công trong việc tái tạo dễ dàng gia tốc và góc quay. Nhóm của Fortes ứng dụng công nghệ này vào hai bộ dữ liệu OpenPack (về hành động trong công nghiệp) [32, 33, 34] và MM-fit (về hành động trong tập luyện) [35]. Với phương pháp của mình, nhóm của Fortes đạt được F1 lên đến 63% và 93% lần lượt trong OpenPack và MM-fit.

Đối với hành động hút nội khí quản, là một thủ thuật phức tạp, vẫn đề nhận diện các hành động nhỏ trong thủ thuật này là một điều cần thiết và đang được phát triển mạnh mẽ. Nhóm nghiên cứu của Halim [36] đã trích xuất đặc trưng bằng cách tính toán tất cả các góc giữa bất kỳ ba khớp nào trên dữ liệu khung xương và tất cả các khoảng cách giữa mỗi cặp khớp, sắp xếp chúng thành các vectơ đặc trưng cho mỗi tư thế tĩnh. Các chuỗi vectơ đặc trưng này sau đó biểu diễn các hành động động phức tạp. Phương pháp này đã được áp dụng cho bộ dữ liệu cử chỉ Kinect MSRC-12. Bằng cách sử dụng rừng ngẫu nhiên cục bộ và toàn cục trên bộ dữ liệu, độ chính xác ẩn tượng 94,7% đã đạt được khi một phần ba bộ dữ liệu MSRC-12 được sử dụng để đào tạo và độ chính xác 95,8% đã đạt được khi việc đào tạo bao gồm hai phần ba bộ dữ liệu. Thế nhưng một vấn đề chưa được xem xét trong tình huống này là sự thiếu dữ liệu huấn luyện.

Nghiên cứu của Ngo và đồng nghiệp [7] thực hiện chỉ ra một sự giảm đáng kể về độ chính xác trong tình huống mất mát thông tin về một số vị trí và bộ dữ liệu nhỏ hơn. Cụ thể hơn, nghiên cứu của Ngo chỉ ra, với bộ dữ liệu của họ, độ chính xác của mô hình nhận dạng trong



Hình 2.18: Cấu trúc của mô hình khuếch tán. Mô hình khuếch tán là mô hình hiện đại nhất tại thời điểm hiện tại của lĩnh vực học sâu trong tạo ra dữ liệu mới. Mô hình này được lấy ý tưởng từ kết hợp hai điểm mạnh của AE và GAN để tạo nên một công cụ tạo sinh dữ liệu mạnh mẽ. Việc tạo thông tin này bắt đầu bằng việc khuếch tán dữ liệu gốc nhiều lần cho đến khi đạt được sự hội tụ. Sau đó, mô hình sẽ học cách tái dụng nhiều này thành hình ảnh gốc bằng các hàm học sâu. Từ khả năng tạo ảnh từ hình ảnh khuếch tán này, một nhiều ngẫu nhiên \mathcal{N} có thể được mã hóa lại thành hình ảnh \mathcal{I} .

tình huống góc quay hép chỉ đạt lên đến 54% và 46% MacroF1. Điều này đưa lên vấn đề cho việc phát triển hệ thống tạo sinh dữ liệu cho những tình huống thiếu thông tin.

2.7.2 Các nghiên cứu tăng cường kết quả bằng tạo sinh dữ liệu

Để vượt qua những thách thức liên quan đến sự đa dạng hành động và biến đổi xương, các phương pháp tăng cường dữ liệu cùng với các kỹ thuật tiền xử lý dữ liệu, chẳng hạn như định hình lại, đã được sử dụng rộng rãi để nâng cao hiệu suất của các mô hình phân loại xương. Mặc dù đã có những tiến bộ trong phân tích tư thế xương, các vấn đề liên quan đến độ chính xác và tốc độ xử lý vẫn còn tồn tại, đặc biệt là trong các tình huống liên quan đến nhiều đối tượng và sự che khuất. Để cải thiện khả năng tổng quát hóa của các mô hình nhận diện hoạt động dựa trên xương, Meng và cộng sự [37] đã thiết kế Mạng Hợp nhất Mẫu (SFN) kết hợp Mạng Bộ nhớ Dài Ngắn Hạn (LSTM) và bộ tự mã hóa cho mục đích tăng cường dữ liệu. Cách tiếp cận này đã nâng cao độ chính xác của mô hình nhận diện hoạt động thêm 10% (tức lên đến 90,75%) bằng cách áp dụng mô hình SFN.

Trong một chiến lược khác để tăng cường dữ liệu, Hyilym Ramirez và cộng sự [38] đã sử dụng Mạng Đối nghịch Sinh tạo Dạng Bảng (TABGAN) để tổng hợp các khung đặc trưng được

trích xuất từ dữ liệu xương. Sau đó, phân loại được thực hiện trên tập dữ liệu bao gồm cả dữ liệu thực và tổng hợp bằng cách sử dụng Biểu diễn Bộ mã hóa Hai chiều từ Transformers (BERT). Sự kết hợp giữa BERT và TABGAN đã thể hiện sự cải thiện hiệu suất đáng kể, đạt độ chính xác 99,5% và F1 87,2%, so với độ chính xác 99,14% và F1 80,95% với mô hình chỉ BERT. Những nghiên cứu này nhấn mạnh hiệu quả của việc tăng cường dữ liệu trong việc giải quyết các thách thức liên quan đến sự biến đổi giữa các cá nhân và sự đa dạng của tập dữ liệu. Do đó, việc lựa chọn một kỹ thuật tăng cường dữ liệu phù hợp là rất quan trọng đối với tập dữ liệu được sử dụng trong nghiên cứu này.

2.7.3 Các nghiên cứu tăng cường kết quả bằng trích xuất đặc trưng

Ngoài việc tạo dựng dữ liệu, việc trích xuất đặc trưng hiệu quả cũng là một yếu tố then chốt được các nhóm nghiên cứu [3, 4, 39, 40] đặc biệt chú trọng. Bởi vì, việc trích xuất các đặc trưng mang tính chuyển hóa từ dữ liệu gốc không chỉ giúp giảm thiểu độ phức tạp tính toán, mà còn nâng cao đáng kể khả năng phân biệt và nhận diện của mô hình.

Cụ thể hơn, nhóm nghiên cứu của Sakai và Lin [4, 40] đã khai thác triệt để các phương pháp toán học truyền thống, được đề xuất bởi mô hình ngôn ngữ lớn (LLM), vốn đã được chứng minh là hiệu quả trong lĩnh vực xử lý số hiệu. Các kỹ thuật như xê dịch thời gian (time warping), rắc nhiễu (noise injection), làm mịn (smoothing), chuẩn hóa (normalization) đã được áp dụng một cách sáng tạo vào bộ dữ liệu, nhằm mục đích tăng cường độ bền vững (robustness) của mô hình trước những biến động và nhiễu loạn không mong muốn trong dữ liệu thực tế.

Trong khi đó, một số nghiên cứu khác [3, 39] đã mở ra một hướng tiếp cận đầy hứa hẹn, đó là ứng dụng sức mạnh của trí tuệ nhân tạo (AI) thông qua các LLM như ChatGPT, Gemini, Claude,... để trích xuất những đặc trưng mang tính ngữ nghĩa và biểu tượng cao. Thay vì chỉ tập trung vào các đặc trưng thô như vị trí khớp hay vận tốc, các nghiên cứu này đã khai thác khả năng hiểu biết ngôn ngữ và suy luận logic của LLM để trích xuất các đặc trưng phức tạp hơn, chẳng hạn như góc hình của khớp, đặc trưng theo quy luật chuyển động, hoặc thậm chí là các đặc trưng mô tả ý định và mục đích của hành động.

Những phương pháp trích xuất đặc trưng tiên tiến này đã mang lại những kết quả ấn tượng, với mức tăng độ chính xác của mô hình dao động từ 10% đến 20% giúp mô hình phân loại đạt độ chính xác lên đến 85%. Điều này không chỉ chứng minh sự hiệu quả vượt trội của việc trích xuất và lựa chọn đặc trưng một cách thông minh, mà còn mở ra những tiềm năng to lớn cho việc ứng dụng AI trong lĩnh vực nhận diện hành vi con người.

Việc kết hợp giữa các phương pháp toán học truyền thống và các kỹ thuật AI hiện đại đã tạo ra một sức mạnh tổng hợp, cho phép các nhà nghiên cứu khai thác tối đa thông tin từ dữ liệu và xây dựng các mô hình nhận diện hành vi con người với độ chính xác và độ bền vững cao hơn bao giờ hết. Điều này có ý nghĩa đặc biệt quan trọng trong các ứng dụng thực tế, nơi mà mô hình phải hoạt động trong môi trường phức tạp và đa dạng, với sự biến động và nhiễu không thể tránh khỏi.

Bảng 2.1: Bảng biểu về tổng hợp các nghiên cứu liên quan

Tác giả chính (bài nghiên cứu)	Đóng góp chính	Kết quả nghiên cứu
Halim [36]	Trích xuất toàn bộ góc có được từ dữ liệu khung xương	Mô hình phân loại hành động đạt 95.8%
Ngo [7]	Phương pháp trích xuất điểm khung xương bằng YOLOv7 trong tình huống có người ở phía sau và người thực hiện được khi hình không đầy đủ	Kết quả mô hình phân loại giảm xuống 54% chính xác và 46% F1
Meng [37]	Thiết kế mạng tạo sinh dữ liệu SFN+LSTM	Độ chính xác phân loại tăng 10% lên hơn 90%
Ramirez [38]	Ứng dụng TABGAN và TABGAN+BERT vào tạo sinh dữ liệu	Nâng cao F1 của mô hình phân loại đạt đến 99.5%
Sakai và Lin [4,40]	Ứng dụng LLM để gợi ý phương pháp trích xuất đặc trưng theo thời gian	Độ chính xác của mô hình phân loại tăng lên đến 85%
Mamun và Jiang [3,39]	Ứng dụng LLM để gợi ý phương pháp trích xuất đặc trưng không gian	Mô hình phân loại tăng độ chính xác lên đến 82%

Chương 3

Phương pháp nghiên cứu

3.1 Bộ dữ liệu



Hình 3.1: Hình ảnh minh họa của dữ liệu khung xương (chấm xanh lá) theo hình ảnh video thực hiện thủ thuật. Đặc biệt là bộ dữ liệu này có một khuyết điểm là thường chỉ xuất hiện 1 nửa thân trái của người thực hiện thủ thuật gây nên sự thiếu dữ liệu về nửa còn lại của cơ thể trong phần lớn thời lượng hành động.

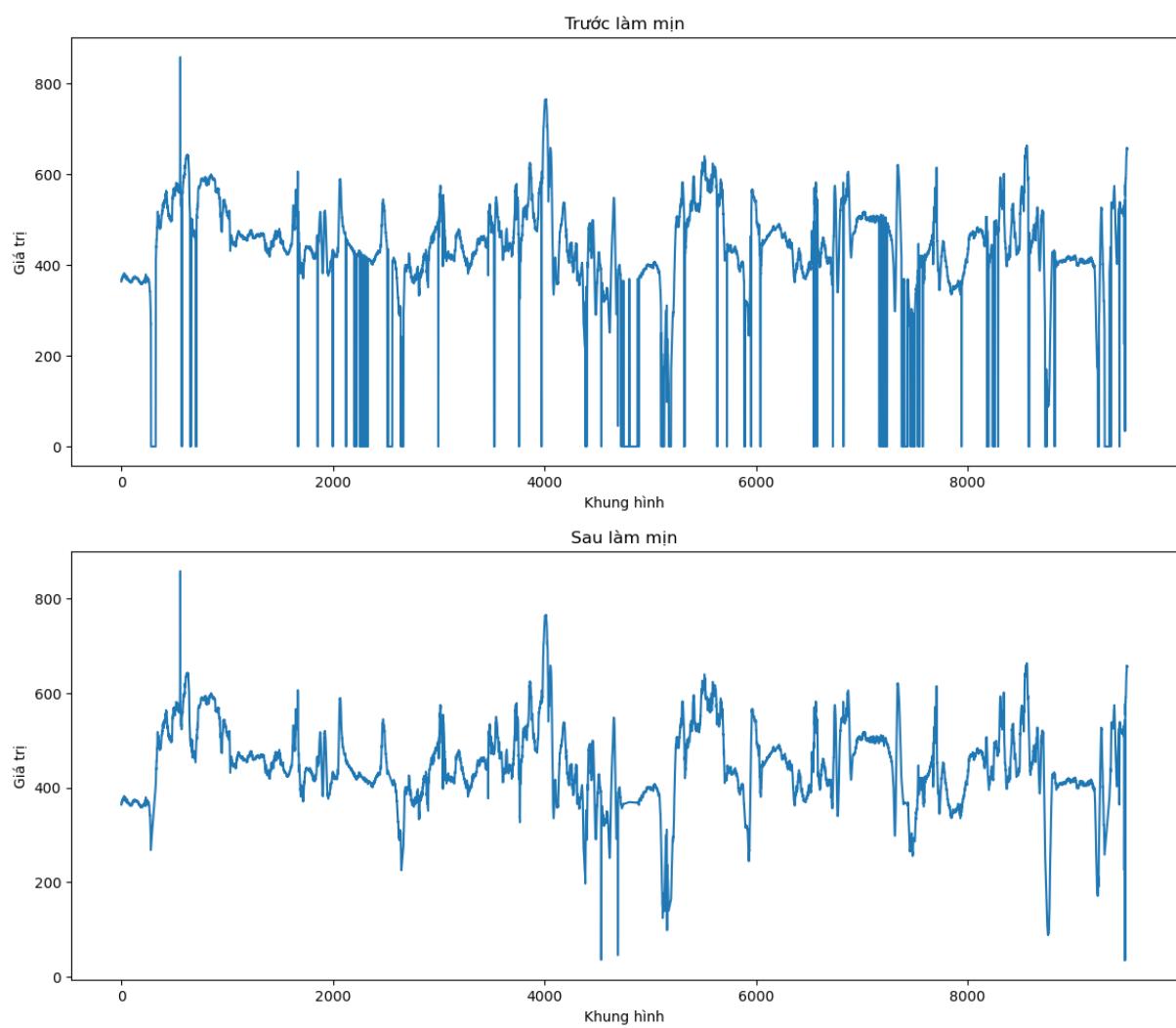
Tập dữ liệu được sử dụng trong nghiên cứu này xuất phát từ các video về hoạt động Hút Dịch Nội Khí Quản (ES) được thực hiện bởi mươi y tá có trên 3 năm kinh nghiệm lâm sàng về hút dịch, và mươi hai sinh viên điều dưỡng từ trường đại học Hokkaido [7]. Những người tham gia này thực hiện các quy trình ES bằng hệ thống mô phỏng ESTE-SIM. Tập dữ liệu bao gồm hai loại dữ liệu riêng biệt: các video được ghi lại từ phía trước của người tham gia (hình nền tại Hình 4.5), và dữ liệu Khung Xương (dạng tọa độ x, y theo từng bộ phận được vẽ thành các chấm xanh trên hình nền tại Hình 4.5), được trích xuất từ video bằng YOLOv7. Quy trình ES được chia thành tổng cộng chín hoạt động riêng biệt, với mỗi hoạt động được gán một nhãn duy nhất.

Trong giai đoạn thu thập dữ liệu, 22 người tham gia đã thực hiện quy trình Hút Dịch Nội Khí Quản (ES), bao gồm 9 hoạt động (Chuẩn bị ống thông, Tháo tạm thời đường thở nhân tạo, Hút đờm, Lắp lại đường thở nhân tạo, Khử trùng ống thông, Thải bỏ găng tay, Định vị, Nghe nhịp

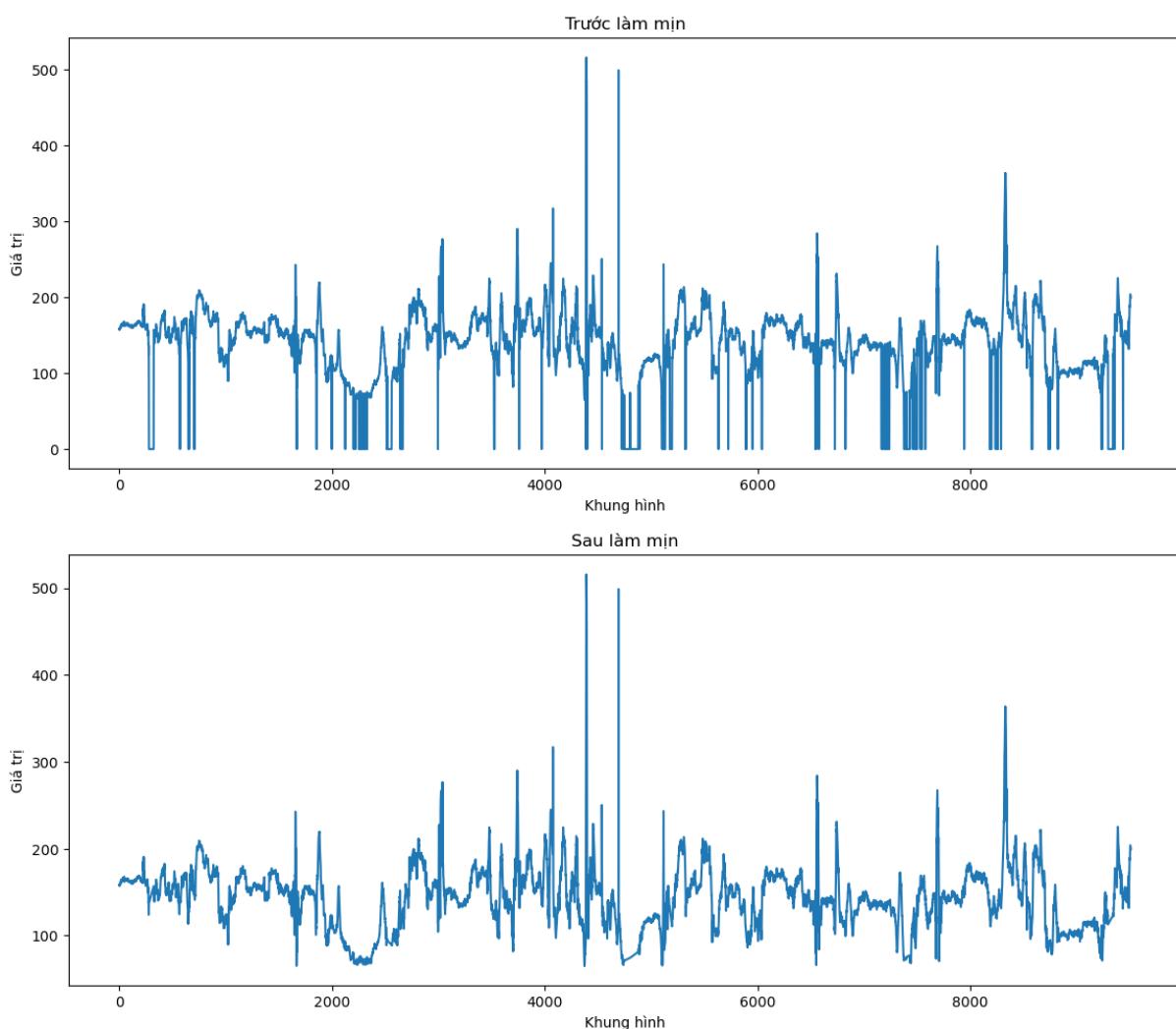
tim và Các thao tác khác) được trình bày tại phụ lục B, trong hai lần. Mỗi video được ghi lại có tốc độ khung hình 30 khung hình/giây và độ phân giải hình ảnh 1920x1080 pixel.

3.2 Làm mịn dữ liệu

Trong dữ liệu thực tế không thể tránh được việc dữ liệu bị khuyết. Giải quyết được các dữ liệu khuyết hợp lý là điều bắt buộc cho những kỹ sư học máy để nâng cao hiệu suất của mô hình và khả năng chống chịu nhiễu của mô hình. Nhìn nhận về bộ dữ liệu, việc thiếu về vị trí của một vị trí có thể do hai nguyên nhân chính (1) sự thất bại của YOLOv7 trong nhận diện và (2) bộ phận đó hoàn toàn không tồn tại trong video. Do lý do trên tôi phương pháp nội suy tuyến tính cho các dữ liệu khuyết nhưng có điều kiện rằng bộ phận bị khuyết quá 4 giây sẽ giữ nguyên giá trị khuyết của dữ liệu do [7] đề xuất để xử lý dữ liệu khuyết.



Hình 3.2: Hình ảnh minh họa xử lý dữ liệu khuyết (giá trị bằng 0) tại vị trí vai trái theo phương x. Ta có thể thấy phần lớn dữ liệu khuyết đã được giải quyết bằng phương pháp làm mịn này.



Hình 3.3: Hình ảnh minh họa xử lý dữ liệu khuyết (giá trị bằng 0) tại vị trí vai trái theo phương y. Ta có thể thấy phần lớn dữ liệu khuyết đã được giải quyết bằng phương pháp làm mịn này.

3.3 Chuẩn hoá dữ liệu

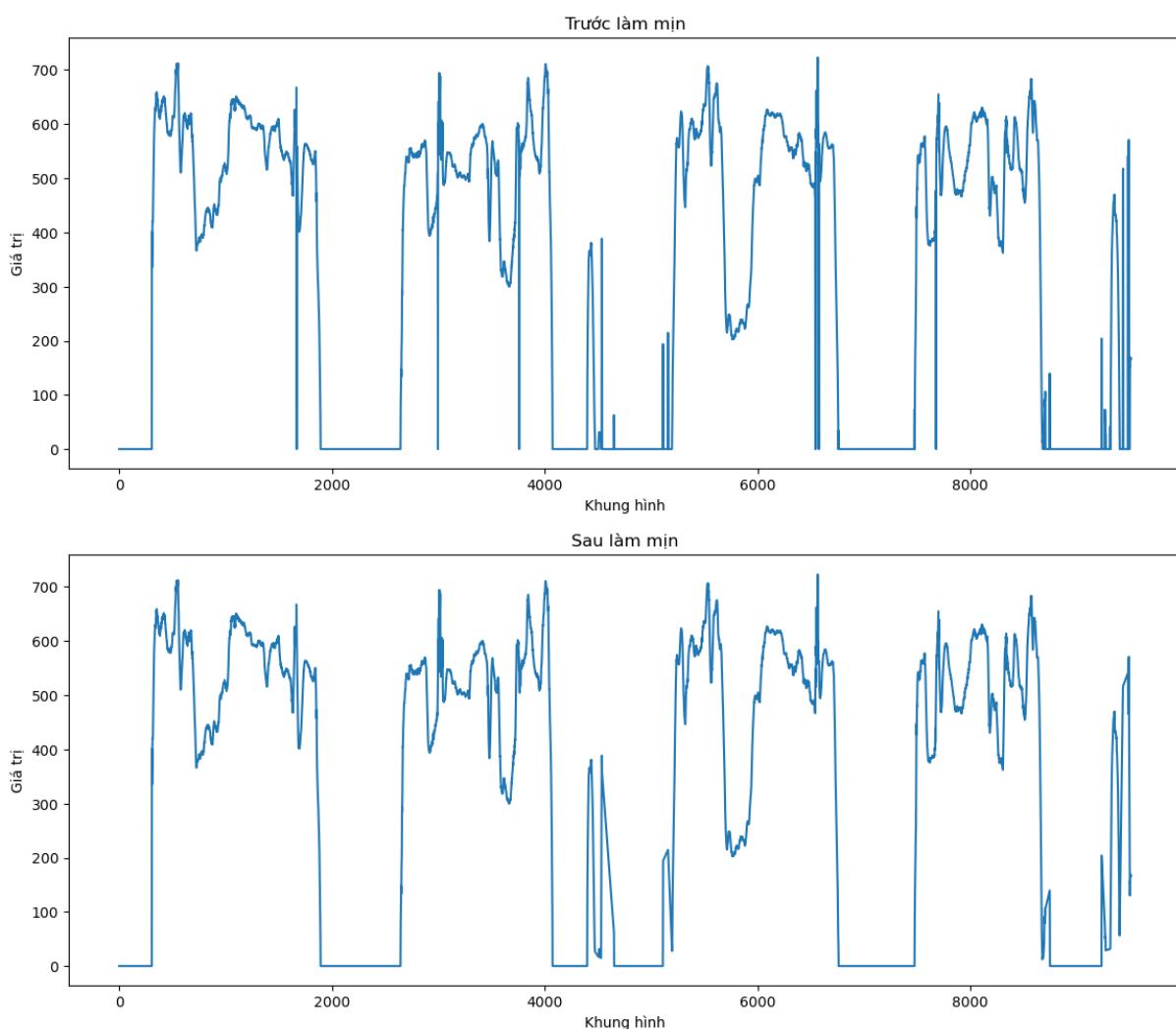
Vì đây là bộ dữ liệu thực tế, nên việc sự sai khác giữa người với người là rất lớn. Do đó, Việc chuẩn hoá dữ liệu là cần thiết. Ở nghiên cứu này, tôi chuẩn hoá dựa vào phương pháp chuẩn hoá chuẩn (Standard Scaler) (3.1) để chuẩn hoá dữ liệu.

$$scaled_data_m = \frac{original_data - \mu}{\sigma} \quad (3.1)$$

với μ và σ là giá trị trung bình, và độ lệch chuẩn của $original_data$

3.4 Trích xuất đặc trưng

Trong nghiên cứu này, chúng tôi đã sử dụng 13 trong số 17 vị trí cơ thể từ dữ liệu khung xương được trích xuất, bao gồm: mũi, mắt trái, mắt phải, tai trái, tai phải, vai trái, vai phải, hông trái, hông phải, cổ tay trái, cổ tay phải, khuỷu tay trái và khuỷu tay phải, mỗi vị trí được theo dõi theo cả chiều x và y.

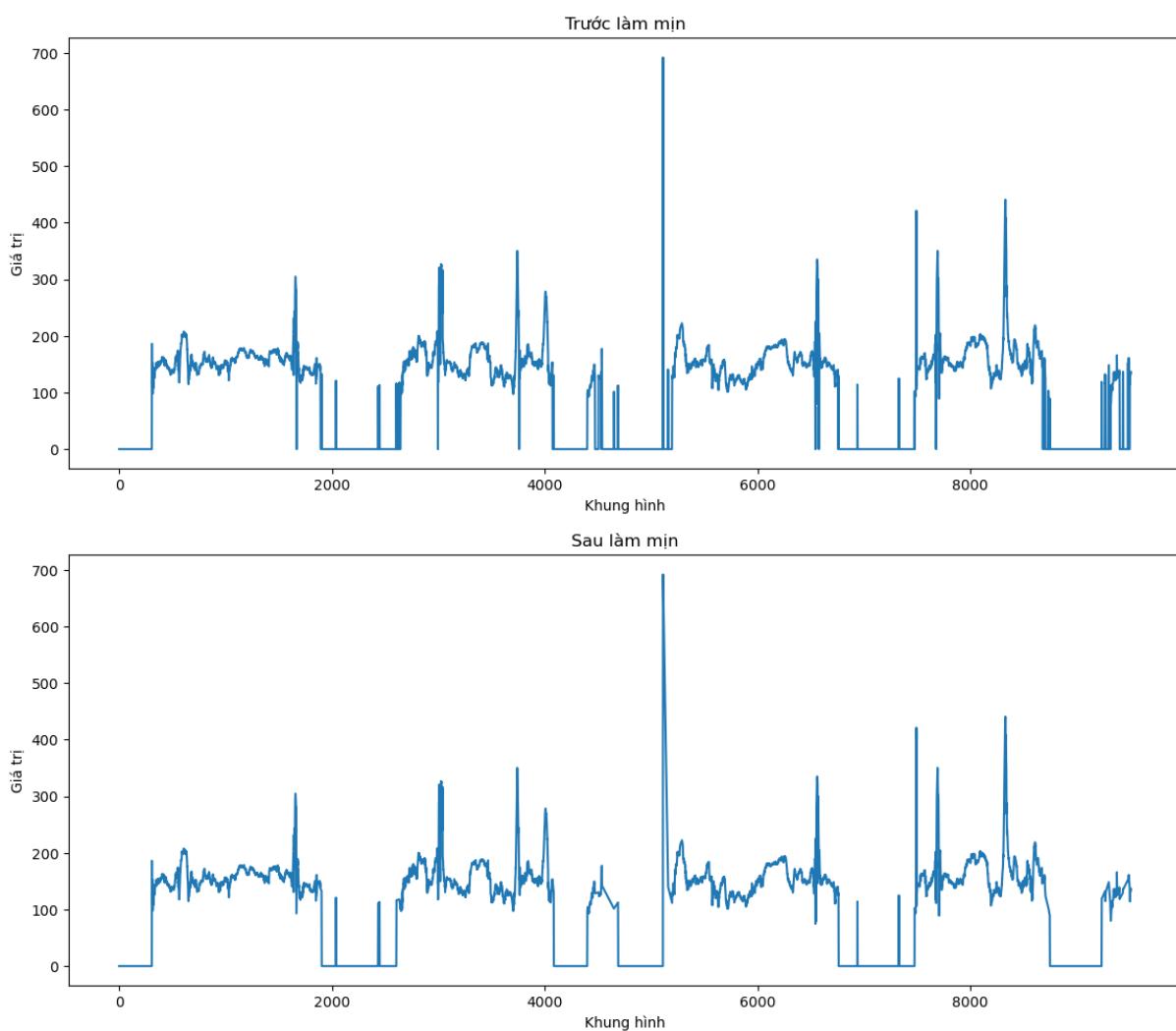


Hình 3.4: Hình ảnh minh họa xử lý dữ liệu khuyết (giá trị bằng 0) tại vị trí vai phải theo phương x. Ta có thể thấy phần lớn dữ liệu khuyết vẫn được giữ nguyên, nguyên nhân là do vai phải thường là bộ phận bị mất đi vì chỉ ghi hình được nửa cơ thể bên trái trong bộ dữ liệu.

Từ các điểm chính khung xương được trích xuất bởi YOLOv7, tổng cộng 8681 đặc trưng đã được trích xuất trong nghiên cứu này, được chia thành 4 nhóm: đặc trưng vị trí và chuyển động; đặc trưng vị trí và chuyển động tương đối; đặc trưng dựa trên góc; và đặc trưng độ trễ thời gian (như được hiển thị trong Bảng 3.1). Ngoài ra, 2 đặc trưng dựa trên trình tự đã được xác định trong công việc này dựa trên trật tự tự nhiên của hoạt động ES.

Bảng 3.1: Bảng về loại đặc trưng và số lượng.

Nhóm đặc trưng	Loại đặc trưng	Số lượng
Nhóm A	Vị trí và chuyển động	2240
Nhóm B	Vị trí và chuyển động tương đối	1920
Nhóm C	Dựa trên góc	180
Nhóm D	Độ trễ thời gian	4341
Nhóm E	Dựa trên trình tự	2
	Tổng	8683



Hình 3.5: Hình ảnh minh họa xử lý dữ liệu khuyết (giá trị bằng 0) tại vị trí vai phải theo phương y. Ta có thể thấy phần lớn dữ liệu khuyết vẫn được giữ nguyên, nguyên nhân là do vai phải thường là bộ phận bị mất đi vì chỉ ghi hình được nửa cơ thể bên trái trong bộ dữ liệu.

3.4.1 Đặc trưng về vị trí

Các đặc trưng vị trí và chuyển động bao gồm tọa độ của từng bộ phận cơ thể và các đặc điểm tư thế, bao gồm giá trị vận tốc, gia tốc và hướng. Những đặc trưng này cung cấp cái nhìn về cách đối tượng thực hiện hoạt động, ví dụ như vị trí của các bộ phận cơ thể, tốc độ di chuyển, gia tốc của việc di chuyển hoặc hướng di chuyển. Bằng cách phân tích những khía cạnh này, có thể đánh giá hiệu suất hoạt động.

Trong nghiên cứu này, chúng tôi xem xét các đặc điểm vị trí và chuyển động theo tọa độ x và y của 14 điểm đánh dấu trên khung xương (13 điểm đánh dấu cơ thể từ dữ liệu khung xương và 1 điểm trung tâm) trên bốn yếu tố chính: vị trí, vận tốc, gia tốc và hướng di chuyển. Để giải quyết thách thức về việc các bộ phận cơ thể phía dưới bị che khuất trong các buổi thực hành, chúng tôi đã tính toán một điểm trung tâm bằng cách lấy giá trị trung bình của tọa độ các điểm khung xương nằm phía trên đường eo của y tá. Điều này cung cấp trọng tâm của cơ thể theo trực x và y. Các đặc trưng vị trí được lấy từ dữ liệu khung xương được cung cấp bởi bộ dữ liệu,

trong khi vận tốc được tính bằng cách xác định sự khác biệt giữa các giá trị tọa độ liên tiếp theo thời gian, như được thể hiện trong công thức (3.2). Tương tự, gia tốc được tính là sự khác biệt giữa các giá trị vận tốc liên tiếp theo thời gian, như được thể hiện trong công thức (3.3). Khoảng thời gian giữa hai giá trị liên tiếp được cố định là 1/30 giây (do dữ liệu được ghi hình trong điều kiện 30 fps). Để xác định hướng di chuyển, chúng tôi đã sử dụng các số liệu nhị phân (tiến hoặc không), trong đó giá trị bằng 1 (tiến) nếu di chuyển về phía trước và 0 nếu di chuyển về phía sau hoặc không tịnh tiến lên trước. Việc xác định này dựa trên giá trị vận tốc, như được mô tả trong công thức (3.4).

$$v = \Delta x \div 1/30 \quad (3.2)$$

$$a = \Delta v \div 1/30 \quad (3.3)$$

$$\text{Hướng tịnh tiến} = \begin{cases} 1 & \text{nếu } v > 0 \\ 0 & \text{Trường hợp còn lại} \end{cases} \quad (3.4)$$

Sau khi nhận được các chuỗi tọa độ, vận tốc, gia tốc và hướng di chuyển cho 14 điểm đánh dấu cơ thể theo cả chiều x và y, các đặc trưng thống kê đã được tính toán để tạo ra tập hợp đặc trưng cuối cùng. Ban đầu, dữ liệu được tỉ lệ hóa bằng thuật toán Robust scaler. Sau đó, 20 đặc trưng dựa trên thống kê đã được trích xuất cho mỗi phân đoạn cửa sổ 3 giây. Những đặc trưng này bao gồm giá trị trung bình, độ lệch chuẩn, giá trị lớn nhất, giá trị nhỏ nhất, phương sai, trung vị, tổng, kurtosis, skewness, phân vị thứ 25 và 75, căn bậc hai trung bình bình phương, giá trị tuyệt đối trung bình, mode, lỗi căn bậc hai trung bình bình phương, lỗi tuyệt đối trung bình, lỗi lớn nhất, lỗi nhỏ nhất, lỗi chuẩn, lỗi phân vị thứ 25 và lỗi phân vị thứ 75. Do đó, tổng cộng 2240 đặc trưng đã được trích xuất trong nhóm đặc trưng này.

3.4.2 Đặc trưng vị trí tương đối và di chuyển

Do sự khác biệt trong các khung hình ghi và vị trí tương đối của đối tượng so với máy quay trong quá trình thu thập dữ liệu, các vị trí được ghi lại có thể thể hiện sự biến đổi, dẫn đến các đặc trưng không được chuẩn hóa được trích xuất từ các tọa độ này. Để giải quyết vấn đề này, tọa độ tương đối đã được sử dụng trong nghiên cứu này. Điểm vai phải đóng vai trò là điểm tham chiếu, thiết lập một hệ thống vectơ từ tham chiếu này đến 12 điểm đánh dấu cơ thể khác theo cả trục x và y. Sau đó, các đặc trưng vị trí tương đối được xác định lại từ các vectơ này. Ngoài ra, vận tốc tương đối, gia tốc và hướng di chuyển cũng được tính toán dựa trên vị trí tương đối bằng cách sử dụng công thức (3.2), (3.3) và (3.4), tương ứng.

Tương tự như việc trích xuất các đặc trưng vị trí và chuyển động, chúng tôi đã tính toán một tập hợp 20 đặc trưng thống kê cho các phân đoạn 3 phút của các chuỗi vị trí tương đối, vận tốc, gia tốc và hướng di chuyển tương đối. Quá trình này được thực hiện trên 2 trục của 12 điểm đánh dấu cơ thể tương đối và tạo ra tổng cộng 1920 đặc trưng trong danh mục này.

3.4.3 Đặc trưng về góc

Các góc khớp đóng một vai trò quan trọng trong việc phản ánh hoạt động cơ bản bằng cách nắm bắt tư thế và chuyển động tương đối giữa các bộ phận cơ thể. Việc bao gồm thông tin bổ sung này tăng cường khả năng của mô hình trong việc giải thích dữ liệu một cách chính xác và đưa ra dự đoán. Trong nghiên cứu này, mỗi góc khớp được tính toán bằng cách sử dụng ba điểm đánh dấu thông qua vectơ được hình thành bởi các tọa độ này (tham khảo Phương trình (3.5)).

$$\text{Angle} = \arccos\left(\frac{\overrightarrow{AB} \cdot \overrightarrow{AC}}{\|\overrightarrow{AB}\| \times \|\overrightarrow{AC}\|}\right) \quad (3.5)$$

trong đó A là tọa độ của điểm giữa, B và C là tọa độ của các điểm còn lại.

Trong nghiên cứu này, tận dụng vị trí tương đối từ phần 3.4.2 để giảm tài nguyên hao phí cho việc tính toán, trích xuất được tổng cộng 9 góc đã được xác định, bao gồm khuỷu tay trái-vai-hông, khuỷu tay phải-vai-hông, cổ tay trái-khuỷu tay-vai, cổ tay phải-khuỷu tay-vai, khuỷu tay phải-vai phải-vai trái, khuỷu tay trái-vai trái-vai phải, khuỷu tay-trung tâm-khuỷu tay, vai-trung tâm-vai và cổ tay-trung tâm-cổ tay. Cuối cùng, tương tự như quá trình trích xuất các đặc trưng vị trí và chuyển động, tổng cộng 20 đặc trưng thống kê đã được tính toán cho mỗi phân đoạn từ các chuỗi góc được xác định.

3.4.4 Đặc trưng liên quan đến thời gian

Các thành phần chuỗi thời gian thể hiện sự phụ thuộc tuần tự, cho thấy rằng giá trị tại một thời điểm cụ thể có thể được suy ra từ các thời điểm trước đó [41]. Điều này liên quan đến việc sử dụng các giá trị quá khứ của chuỗi mục tiêu làm đặc trưng, do đó căn chỉnh chúng với các giá trị đương thời mà chúng ta muốn dự đoán (tức là trong cùng ngữ cảnh thời gian). Cụ thể, bằng cách dịch chuyển các giá trị đặc trưng từ 3 giây trước của mỗi điểm trong chuỗi đặc trưng, chúng tôi bổ sung cho các mô hình của mình thông tin bổ sung để cải thiện hiệu suất phân loại. Ví dụ cho vấn đề này có thể được giải thích bởi hành động đưa ống để hút dịch màng phổi, việc trước đó tẩy của nhân viên thực hiện ở vị trí xa hay vị trí gần sẽ giúp ích cho việc nhận diện đang kéo ống vào cho hút đờm, hay việc rút ống ra cho lắp lại đường thở nhân tạo. Do vậy, các đặc trưng độ trễ thời gian tạo điều kiện thuận lợi cho việc xác định các mẫu và xu hướng trong dữ liệu chuỗi thời gian, đồng thời truyền các đặc tính tự hồi quy cho mô hình được chọn, từ đó mở rộng việc sử dụng các giá trị quá khứ để dự đoán các kết quả trong tương lai.

Trong nghiên cứu này, các đặc trưng độ trễ thời gian đã được kết hợp vào ba nhóm đặc trưng: 2240 đặc trưng vị trí và chuyển động, 1920 đặc trưng vị trí và chuyển động tương đối và 180 đặc trưng dựa trên góc. Đáng chú ý, các tập hợp đặc trưng này đã được sử dụng lại từ phân đoạn ngay trước đó.

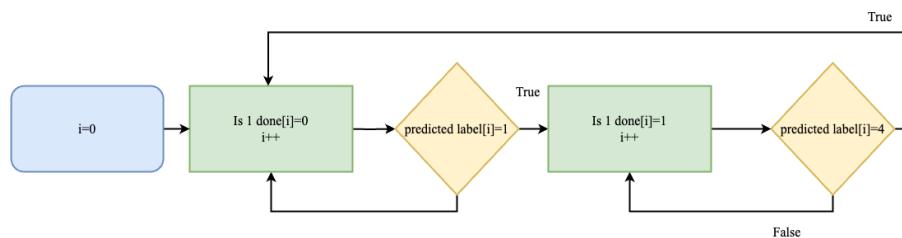
Hơn nữa, đặc trưng thời gian, đại diện cho khía cạnh thời gian của buổi thực hành, cũng được bao gồm. Đặc trưng này mô tả dấu thời gian mà phân đoạn này được thực hiện và được định nghĩa trong Phương trình (3.6).

$$time_i = real_time_i / real_time_{max} \quad (3.6)$$

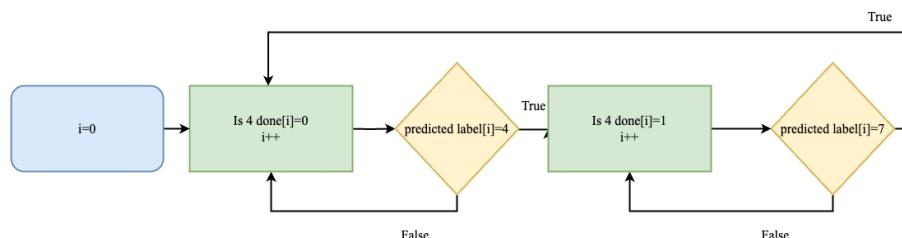
trong đó $real_time_i$ là thời gian lúc ghi và $real_time_{max}$ là tổng thời gian của lần thực hiện.

3.4.5 Đặc trưng tuần tự

Việc nhận biết tuần tự giúp chúng ta có thể nhận diện các hành động giống nhau nhưng ở những khoảng thời gian khác nhau. Ví dụ như để tạo một hộp giấy có nắp ta thường dùng băng keo để dán và tạo khung hộp, và việc dán đóng hộp ta cũng có hành động tương tự. Việc nhận biết hành động nào là gì trên ví dụ trên có thể được nhận biết thông qua việc nhìn độ kín của hộp tổng quát, hoặc xem xét theo một quy trình cụ thể ví dụ tạo hộp, đóng gói đồ vào hộp, đóng hộp. Tận dụng tư duy trên và đặc điểm tuần tự của hoạt động y tế, tôi đề xuất 2 đặc trưng tuần tự tháo tạm thời đường thở nhân tạo đã xong (is 1 done) và khử trùng ống thông đã xong (is 4 done). Việc này sẽ giúp dễ dàng hơn trong việc nhận diện hai hành động tương tự nhau gồm nhóm chuẩn bị ống thông (gồm khử trùng ống thông, xem hình B.1) và khử trùng ống thông(xem hình B.5) và nhóm tháo tạm thời đường thở nhân tạo (gồm đeo găng tay, xem hình B.2) và thải bỏ găng tay(xem hình B.6).



Hình 3.6: Lưu đồ trích xuất tháo tạm thời đường thở nhân tạo đã xong (is 1 done)



Hình 3.7: Lưu đồ trích xuất khử trùng ống thông đã xong (is 4 done)

3.5 Lựa chọn đặc trưng

Trong nghiên cứu này, để sàng lọc và xác định các đặc điểm phù hợp và giàu thông tin nhất từ một cơ sở lớn các đặc trưng được trích xuất, điểm ANOVA f-test đã được sử dụng để tính toán và lựa chọn độ quan trọng của đặc trưng. Cụ thể, điểm kiểm tra F ANOVA đã được áp dụng cho 8683 đặc trưng được trích xuất trên 5 nhóm. Sau đó, đặc trưng thời gian và 2001 đặc trưng có điểm kiểm tra F cao nhất (p -value thấp nhất) đã được chọn cho giai đoạn phân loại. Quy trình này không chỉ giảm chiều dữ liệu mà còn tăng tốc tính toán mô hình, giảm tiêu thụ tài nguyên bằng cách loại trừ các đặc trưng ít ý nghĩa và giảm thiểu nguy cơ quá khớp trong phân loại.

3.6 Tạo sinh dữ liệu

Các nghiên cứu trước đó [37, 38] đã thực hiện các chiến thuật tạo sinh dữ liệu bằng các công nghệ học sâu hiện đại nhưng chưa có nghiên cứu tập trung vào phương pháp cổ điển như Copulas như trình bày tại phần 2.6.1, tại đồ án này tôi sử dụng Gausian Copulas để tạo sinh dữ liệu góp phần nâng cao hiệu suất mô hình.

Gaussian Copulas có khả năng mô phỏng các tập dữ liệu thực tế và đặc biệt hiệu quả với các tập dữ liệu bảng có quy mô lớn [42]. Phương pháp này tính toán các thống kê trên các bảng dữ liệu liên kết và áp dụng các kỹ thuật mô hình hóa đa biến tiên tiến. Thông qua cách tiếp cận lặp, Gaussian Copulas xây dựng một mô hình tổng quát của toàn bộ tập dữ liệu, giúp tránh hiện tượng quá khớp (overfitting) và cải thiện chất lượng dữ liệu sinh ra.

Sau khi mô hình được huấn luyện, nó có thể tạo ra dữ liệu tổng hợp một cách linh hoạt, cho phép lấy mẫu từ bất kỳ phân đoạn nào trong cơ sở dữ liệu. Nhờ đó, Gaussian Copulas mang lại tính linh hoạt và ứng dụng cao trong việc tạo ra các tập dữ liệu tổng hợp phục vụ cho nhiều mục đích phân tích khác nhau [43].

Trong nghiên cứu này, Gaussian Copulas (2.14) được áp dụng cho dữ liệu của những người tham gia. Trước tiên, 2001 đặc trưng quan trọng nhất được chọn lọc từ tập dữ liệu xương gốc dựa trên chỉ số ANOVA F-score. Sau đó, các đặc trưng này được đưa vào mô hình Gaussian Copulas để huấn luyện mô hình sinh dữ liệu. Khi quá trình huấn luyện hoàn tất, Gaussian Copulas được sử dụng để tạo ra dữ liệu tổng hợp, với số lượng mẫu tương đương với tập dữ liệu gốc. Cuối cùng, các đặc trưng theo thời gian (time-based features) được trích xuất từ dữ liệu tổng hợp để tạo ra tập dữ liệu mở rộng cuối cùng.

3.7 Mô hình phân loại

Để thực hiện phân loại hoạt động ES dựa trên các đặc trưng được trích xuất và lựa chọn, ba mô hình học tập kết hợp, Extreme Gradient Boosting Classifier (XGBoost) và Hist Gradient Boosting Classifier (HGBC), đã được sử dụng trong công việc này.

- Mô hình Extreme Gradient Boosting (XGBoost) là một mô hình học tập kết hợp boosting, giúp tăng cường hiệu suất phân loại bằng cách huấn luyện tuần tự các mô hình để sửa lỗi của các mô hình trước đó. XGBoost đã trở nên phổ biến vì hiệu quả của nó trong nhiều nhiệm vụ học máy khác nhau do khả năng xử lý các tập dữ liệu phức tạp và tạo ra các dự đoán có độ chính xác cao [14]. Trong nghiên cứu này, XGBoost được thiết lập với các chỉ số đánh giá là mlogloss.
- Mô hình Hist Gradient Boosting (HGBC): Tương tự như mô hình XGBoost, mô hình HGBC hoạt động trong khuôn khổ gradient boosting. Nó bao gồm việc xây dựng tuần tự các mô hình cây quyết định, với mỗi cây nhằm mục đích sửa lỗi của các cây trước đó. Tuy nhiên, HGBC tự phân biệt bằng cách giảm độ phức tạp huấn luyện thông qua việc phân nhóm các đặc trưng thành biểu đồ. Việc tóm tắt phân phối dữ liệu này đặc biệt có lợi cho việc xử lý các tập dữ liệu đặc trưng lớn. Trong nghiên cứu này, HGBC được cấu hình với trọng số lớp cân bằng, như được mô tả trong tài liệu [44].

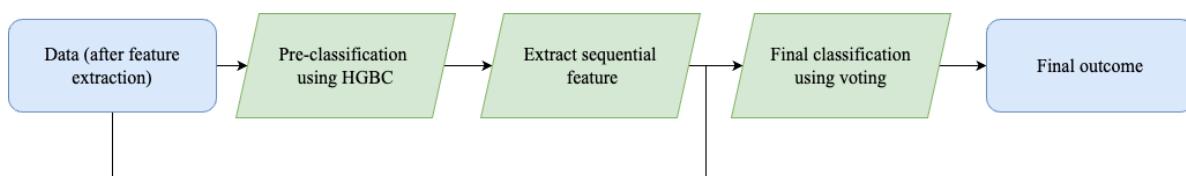
- Mô hình Voting: là một phương pháp tận dụng nhiều mô hình học máy để cùng nhau quyết định kết quả phân loại cuối cùng [45]. Trong công việc này, chúng tôi sử dụng mô hình voting mềm, là sự kết hợp của bốn mô hình học máy đã được huấn luyện (hai HGBC và hai XGB) để bỏ phiếu và trả về đối số cực đại của lớp có khả năng trung bình cao nhất để phân loại hoạt động.

Trong nghiên cứu này, chúng tôi đã sử dụng ba mô hình được mô tả trước đó trong nhiều kết hợp và tập dữ liệu khác nhau để đạt được hiệu suất phân loại tốt nhất. Bằng cách khám phá các kịch bản khác nhau về kết hợp mô hình và sử dụng tập dữ liệu, chúng tôi hướng tới việc tối ưu hóa độ chính xác phân loại và độ bền vững của hệ thống của mình.

3.8 Quy trình phân loại

Tập dữ liệu được chia theo từng cá nhân thành tập huấn luyện và tập xác thực như sau:

- Tập huấn luyện chứa 4 y tá (ID 1, 4, 7, 12) và 6 sinh viên (ID 1, 5, 7, 9, 10, 11)
- Tập xác thực chứa 3 y tá (ID 2, 6, 11) và 3 sinh viên (ID 2, 3, 8).



Hình 3.8: Lưu đồ của quy trình phân loại

Để thu được dữ liệu tuần tự, cần có các giả định về nhãn hoạt động. Trong nghiên cứu này, mô hình HGBC (mô hình tiền phân loại) đã được sử dụng để tiền huấn luyện trên tập đặc trưng gốc sau khi lấy mẫu quá mức bằng SMOTE. Các nhãn được xác định trước sau đó được sử dụng để tính toán hai đặc trưng tuần tự.

Để tạo ra các dự đoán cuối cùng, cả tập đặc trưng gốc và tập đặc trưng tổng hợp đều được sử dụng trong quá trình huấn luyện chính trên hai mô hình phân loại. Sau khi lấy mẫu quá mức lớp thiểu số bằng SMOTE, dữ liệu được đưa vào các mô hình trong hai kịch bản: chỉ sử dụng dữ liệu gốc và kết hợp cả dữ liệu gốc và dữ liệu tổng hợp. Sau đó, bốn mô hình đã được huấn luyện: XGBoost (XGB1, XGB2) và Histogram-Based Gradient Boosting Classifier (HGBC1, HGBC2)¹. Cuối cùng, hiệu suất phân loại được đánh giá trên ba kịch bản:

- Mô hình Voting: Sử dụng mô hình voting mềm của XGB1, XGB2, HGBC1 và HGBC2 để dự đoán dự đoán cuối cùng. Mô hình có thể tận dụng thế mạnh của mô hình chỉ được huấn luyện bởi dữ liệu thật và mô hình huấn luyện bỏ dữ liệu kết hợp, tạo nên tính ổn định của mô hình.
- XGB+SDV: Sử dụng mô hình XGB1 đã được huấn luyện để dự đoán dự đoán cuối cùng.
- XGB: Sử dụng mô hình XGB2 đã được huấn luyện để dự đoán dự đoán cuối cùng.

¹*1 mô hình biểu thị mô hình được huấn luyện bằng cách kết hợp tập đặc trưng gốc và tập đặc trưng được tạo bởi SDV, trong khi *2 mô hình biểu thị mô hình chỉ được huấn luyện với dữ liệu gốc.

3.9 Phương pháp đánh giá mô hình

$$Precision = \frac{TP}{TP + FP} \quad (3.7)$$

$$Recall = \frac{TP}{TP + FN} \quad (3.8)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (3.9)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3.10)$$

$$Weighted_F1 = \frac{\sum n_i \times F1_i}{\sum n_i} \quad (3.11)$$

với

- TP, FP, FN, TN được đề cập trong hình 3.9
- n_i là số lượng mẫu hỗ trợ lớp i
- $F1_i$ là hệ số F1 (xem công thức (3.10)) của lớp i

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Hình 3.9: Hình ảnh trực quan các giá trị trong ma trận nhầm lẫn

Trong học máy, để đánh giá hiệu suất của một mô hình, độ chính xác (accuracy) được sử dụng phổ biến. Giá trị này thể hiện tỷ lệ các dự đoán đúng so với tổng số dự đoán. Độ chính xác thường được chọn làm thước đo đầu tiên bởi tính đơn giản, trực quan và dễ tính toán của nó. Nó cung cấp một cái nhìn tổng quan về hiệu suất chung của mô hình, giúp các nhà nghiên cứu và kỹ sư có thể so sánh hiệu quả giữa các mô hình khác nhau.

Mặc dù vậy, độ chính xác lại có những hạn chế nhất định khi áp dụng cho các tập dữ liệu không cân bằng. Trong trường hợp này, một mô hình có thể đạt được độ chính xác cao đơn giản bằng cách dự đoán tất cả các mẫu vào lớp chiếm đa số. Để khắc phục vấn đề này, các nhà nghiên cứu đã đề xuất sử dụng các thước đo khác, điển hình là weighted F1-score.

Weighted F1-score (3.11) là một thước đo tổng hợp, kết hợp cả precision (độ đúng (3.7)) và recall (độ phủ (3.8)). Nó tính toán trung bình điều hòa (harmonic mean) của precision và recall, trong đó trọng số của mỗi lớp được điều chỉnh để phản ánh tầm quan trọng tương đối của chúng. Nhờ vậy, weighted F1-score cung cấp một cái nhìn toàn diện hơn về hiệu suất của mô hình, đặc biệt trong các bài toán phân loại có dữ liệu.

3.10 Phương pháp đánh giá sự khác biệt giữa nhân viên y tế và sinh viên

Trong nghiên cứu trước đó [1], tôi đã chỉ ra khả năng nhận diện hành động do nhân viên y tế hoặc sinh viên thực hiện thông qua mô hình phân loại. Điều này đã làm cho tôi nhận ra có một sự sai khác giữa y tá và sinh viên, điều này, nếu trong thực tế xảy ra, có thể gây nên các rủi ro ở bệnh nhân. Do đó ở nghiên cứu này, tôi khai thác sâu hơn về những nguyên nhân làm cho mô hình dễ nhận diện hành động của nhân viên y tế hay sinh viên, từ đó có thểm thông tin về hiện tượng thú vị này.

Cụ thể hơn, nhận diện khác biệt do hành động sinh viên và y tá, tôi đã sử dụng mô hình XGB để nhận diện hành động do y tá hay sinh viên thực hiện. Sau đó tôi ứng dụng SHAP để đưa ra top 10 đặc trưng để giải thích phần được mô hình sử dụng để nhận diện các hành động của đối tượng nào thực hiện. Sau cùng tôi sử dụng công cụ thống kê và trực quan bằng hình ảnh từ video để phân tích.

Chương 4

Kết quả và bàn luận

4.1 Kết quả mô hình phân loại

Bảng 4.1: Kết quả phân loại tổng quát (đơn vị phần trăm)

	Voting	XGB + SDV	XGB	Pre-classify
Weighted avg	89.9 ± 7.1	86.0 ± 8.6	89.9 ± 7.0	86.8 ± 5.8
F1-score trung bình				
Độ chính xác trung bình	90.4 ± 7.3	86.8 ± 8.5	90.3 ± 7.0	87.7 ± 5.9
Weighted avg	92.5	89.8	91.9	86.7
F1-score trung vị				
Độ chính xác trung vị	93.9	90.9	92.3	87.3

Kết quả phân loại, như thể hiện trong Bảng 4.1 và Bảng 4.2, cho thấy hiệu suất đáng kể trong việc phân loại hoạt động ES qua ba kịch bản mô hình học. Đặc biệt, mô hình Voting đã đạt được các chỉ số hiệu suất cao nhất, với Weighted F1-score và độ chính xác lần lượt là 89.9% và 90.4%. Hơn nữa, mô hình Voting có hiệu suất vượt trội nhẹ so với mô hình XGB về độ chính xác trung vị và Weighted F1-score, với mức cải thiện khoảng 1.5% và 0.6%. Tuy nhiên, cả hai mô hình Voting và XGB đều có độ chính xác và F1-score cao hơn nhẹ, khoảng 4%, so với mô hình XGB + SDV. Sự khác biệt này có thể cho thấy khả năng overfitting khi áp dụng cho các bộ dữ liệu lớn hoặc những bất cập trong các bản ghi tổng hợp được tạo ra.

Những phát hiện này nhấn mạnh tầm quan trọng của mô hình tổng hợp (ensemble) trong việc đạt được hiệu suất phân loại vượt trội. Các đánh giá cá nhân chi tiết trong Bảng 4.2 càng củng cố điều này, với bộ phân loại Voting cho ra Weighted F1-score và độ chính xác ấn tượng (phần lớn trên 90% và có 4 trường hợp lên đến khoảng 95%). Đáng chú ý, ở một số người tham gia (N06T1 và S03T1), hiệu suất của mô hình XGB trên cả dữ liệu gốc và dữ liệu tổng hợp có phần giảm nhẹ, thể hiện qua F1-score lần lượt là 69% và 77%.

Một số lượng lớn các đặc trưng được trích xuất đã được đánh giá bằng điểm ANOVA F-test, như minh họa trong Hình 4.1. Trong số này, các đặc trưng theo trình tự đề xuất (Nhóm E) cho

Bảng 4.2: Kết quả mô hình phân loại theo từng người (đơn vị phần trăm)

	Weighted avg F1 score			Độ chính xác		
	Voting	XGB + SDV	XGB	Voting	XGB + SDV	XGB
S08T1	90	88	92	90	89	92
S08T2	72	92	73	73	93	74
N11T1	94	95	95	96	97	96
N11T2	94	93	94	96	95	96
N06T1	85	69	79	84	66	79
N06T2	90	89	92	92	90	93
S02T2	88	85	89	89	86	89
S02T1	89	80	90	88	81	89
N02T2	92	87	93	94	89	94
N02T1	95	87	96	96	89	97
S03T1	95	77	94	95	77	94
S03T2	92	92	92	93	92	92

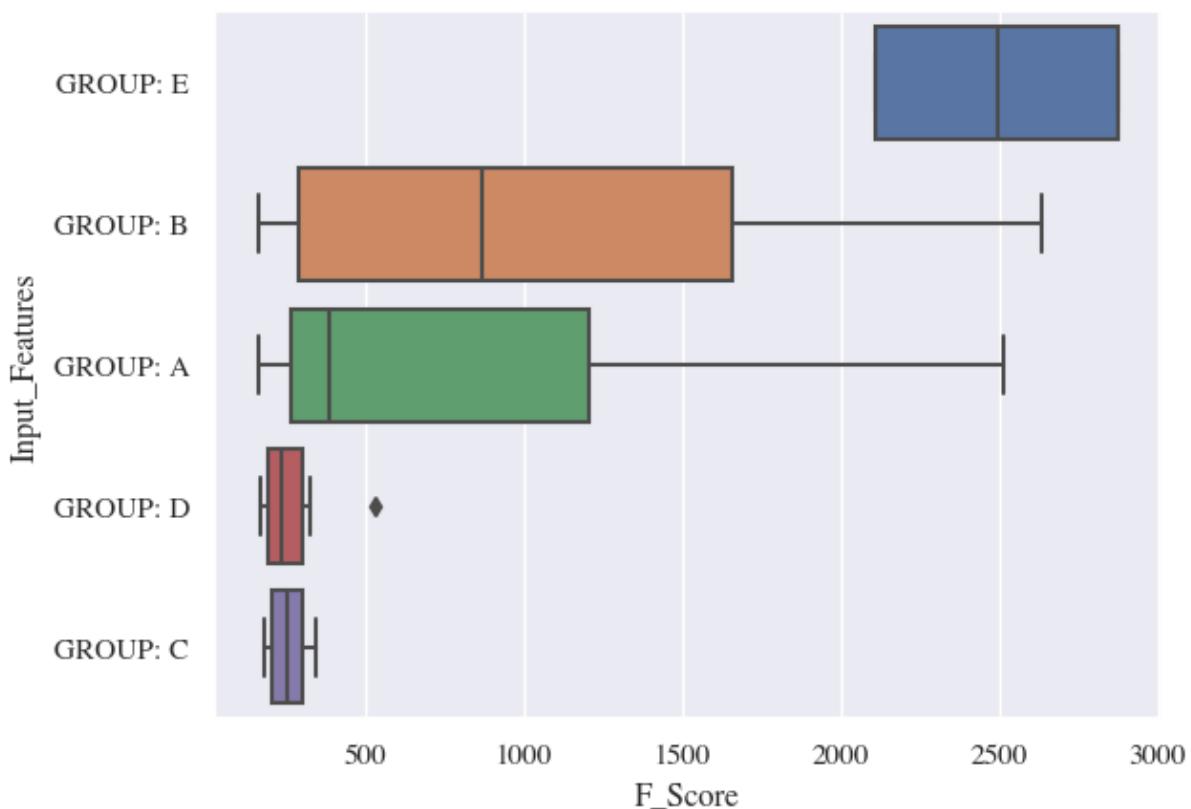
thấy F-score trung bình cao nhất, nhấn mạnh tầm quan trọng của thứ tự hoạt động. Cả các đặc trưng vị trí và chuyển động ban đầu (Nhóm A) và tương đối (Nhóm B) cũng cho thấy F-score cao, với khoảng giá trị phần tử từ khoảng 250 đến 1250 và 1750, tương ứng. Đáng chú ý, các đặc trưng vị trí và chuyển động tương đối đạt được điểm số cao hơn nhẹ nhàng nhờ việc điều chỉnh phương sai của khung ghi hình. Trong khi đó, các đặc trưng dựa trên độ trễ thời gian và góc cho ra F-score trong khoảng từ 100 đến 400.

4.2 Bàn luận

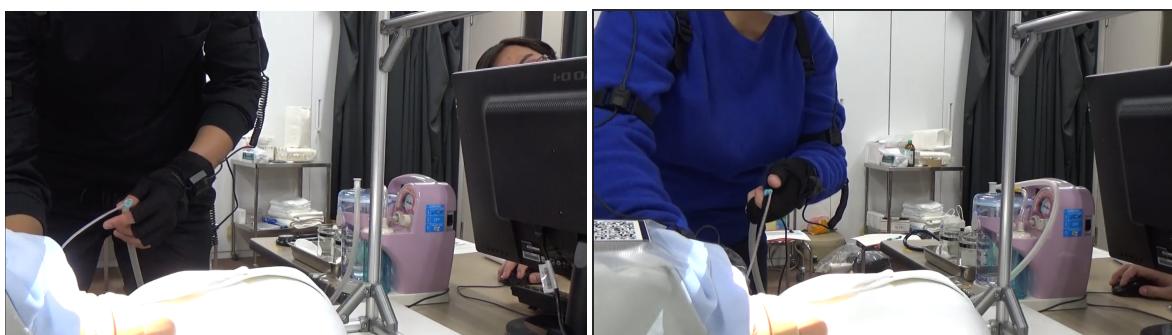
4.2.1 Về nhận diện sự khác nhau về hành động của học viên y khoa và nhân viên y tế có kinh nghiệm bằng XGB

Bảng 4.3: Hiệu suất Phân loại giữa Y tá và Sinh viên theo ID Hoạt động bằng XGB

Mã Hoạt động	Tên Hoạt động	Độ chính xác (%)
0	Chuẩn bị ống thông	63
1	Tháo tạm thời đường thở nhân tạo	88
2	Hút đờm	80
3	Lắp lại đường thở nhân tạo	82
4	Khử trùng ống thông	59
5	Thải bỏ găng tay	68
6	Định vị	80
7	Nghe nhịp tim	66
8	Khác	71



Hình 4.1: Biểu đồ hộp của F-score theo 5 nhóm đặc trưng



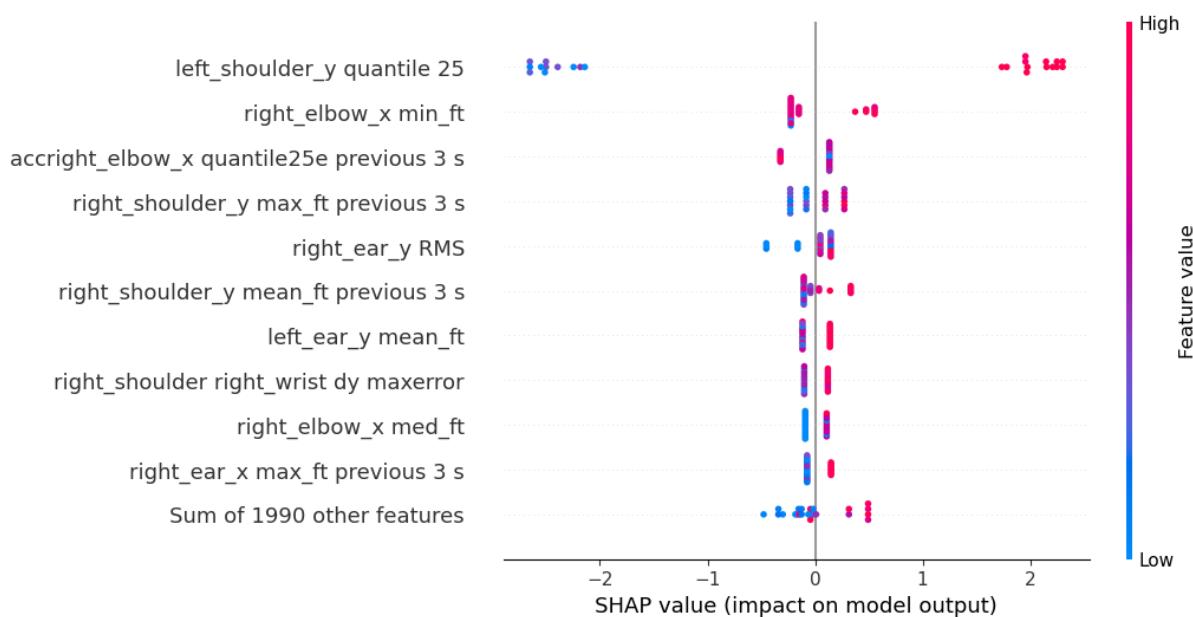
(a) Sinh viên

(b) Nhân viên y tế

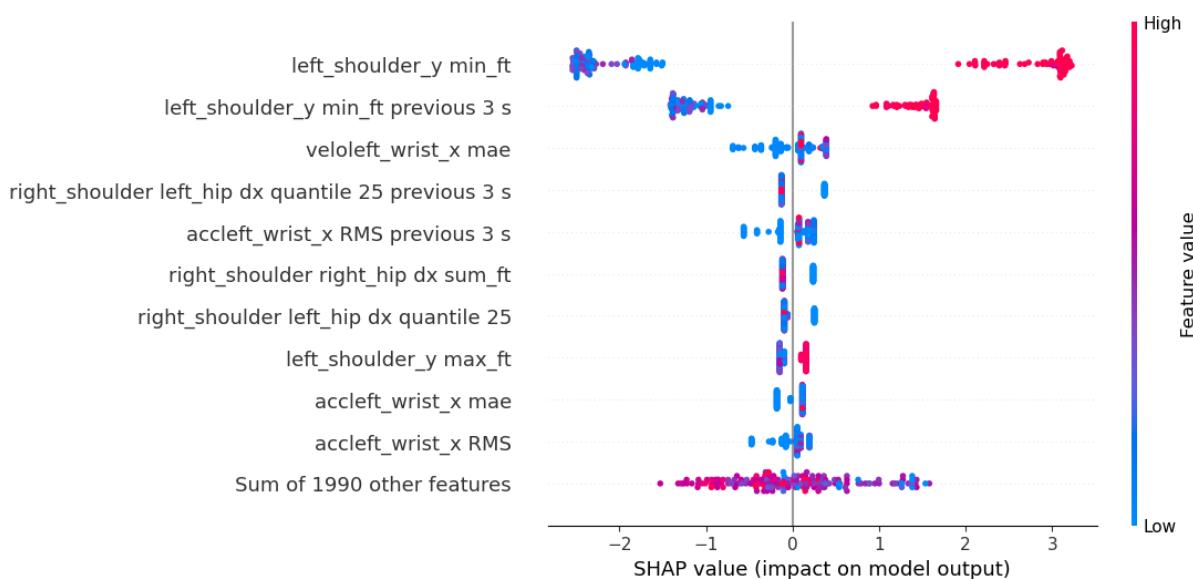
Hình 4.2: Hình ảnh về sự khác nhau của động tác trong hành động hút đờm. Có thể thấy hành động của sinh viên (hình 4.2a), sinh viên đứng thẳng so với nhân viên y tế (hình 4.2b) hơi khom người để nhìn thấy được rõ hơn dịch đờm. Đây là lý do tiên quyết giúp mô hình nhận diện hành động do bác sĩ hay sinh viên ra quyết định. (hình 4.3 và 4.4)

4.2.2 Các ứng dụng tiềm năng của đè tài

Về ứng dụng của đè tài này, đè tài này đã đề xuất phương án để nhận diện hành vi của người thực hiện hút dịch nội khí quản. Từ đây, một số ứng dụng về nhắc người thực hiện thủ thuật thực hiện quy trình (tương tự với đè xuất trong [46]) có thể được cân nhắc. Việc nhắc nhở này có thể giúp nhân viên thực hiện thủ thuật làm theo quy trình khi cần thiết, giảm thiểu các sai sót

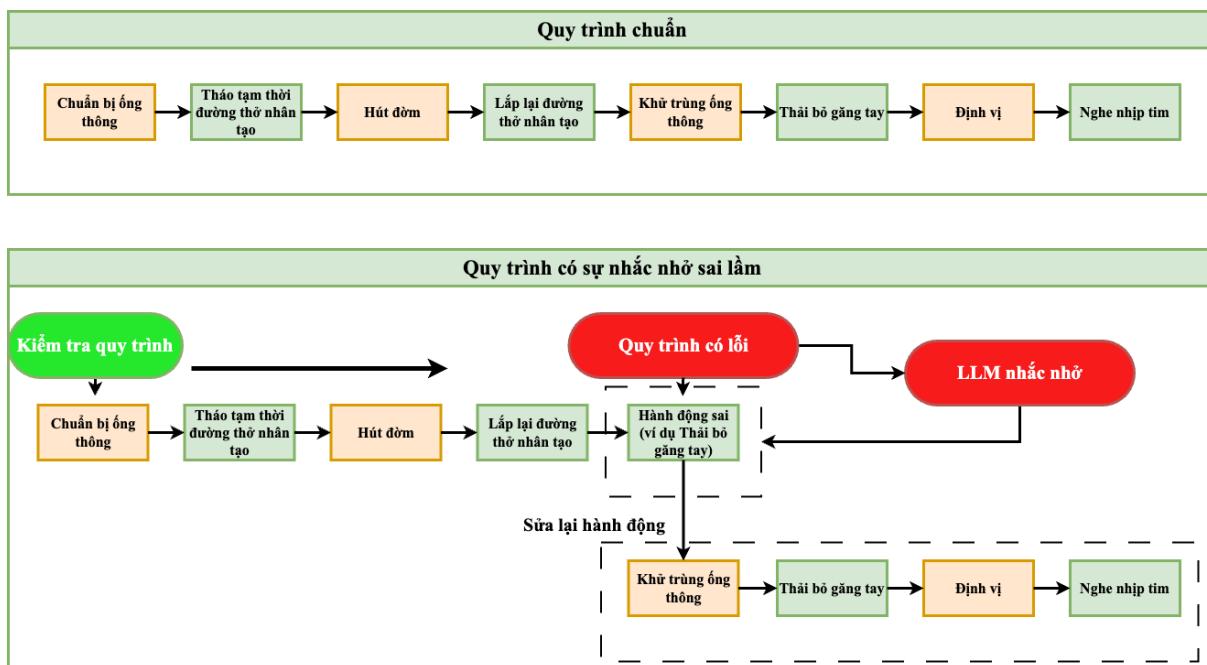


Hình 4.3: Biểu đồ SHAP về độ quan trọng của đặc trưng trong nhận biết hành động do học sinh hay y tá thực hiện ở hành động Tháo tạm đường thở nhân tạo



Hình 4.4: Biểu đồ SHAP về độ quan trọng của đặc trưng trong nhận biết hành động do học sinh hay y tá thực hiện ở hành động Hút đờm

không đáng có. Ngoài ra các hướng dẫn này có thể mở rộng hơn bao gồm nhận biết tốc độ hút, độ sâu của ống hút dịch nội khí quản từ đó ngăn ngừa những nguy cơ do những sai sót về kỹ thuật người thực hiện thủ thuật.



Hình 4.5: Quy trình minh họa cho ứng dụng sửa lỗi sai. Việc kiểm có thể được thực hiện liên tục, thông suốt quá trình thực hiện thủ thuật. Ngoài ra có thể phát triển thêm hệ thống cho phép thay đổi quy trình trong trường hợp ngoại lệ.

Một ứng dụng khác là ghi nhận quá trình thực hiện hành động (tương tự với đề xuất từ [47]). Việc ghi nhận từng quá trình cụ thể trong ca thực hiện thủ thuật là cần thiết cho việc truy xuất, kiểm tra khi có sự cố, hoặc nâng cao chất lượng của thủ thuật viên. Nhưng việc ghi nhận một video dài không thực sự là giải pháp tối ưu do dung lượng lớn cần ghi hình của một ca thực hiện thủ thuật dài trong khi tương tác giữa bác sĩ-dụng cụ-bệnh nhân mới là các phần quan trọng cần chú ý. Vì vậy ứng dụng đề tài của tôi mở ra một hướng đi mới trong việc ghi nhận và giảm chi phí lưu trữ cho các ứng dụng ghi hình trong quá trình thực hiện thủ thuật trong thực tế.

Hơn thế, một ứng dụng tiềm năng là việc thông báo quá trình thực hiện thủ thuật cho người nhà bệnh nhân tự động. Việc tự động nhận diện và gửi cho người nhà bệnh nhân người thân của mình đang được các y-bác sĩ đang thực hiện thủ thuật nào có thể giúp người nhà bệnh nhân yên tâm hơn, từ đó giúp gia tăng chất lượng chăm sóc bệnh nhân và người nhà của trung tâm y tế.

Chương 5

Tổng kết

Tài liệu tham khảo

- [1] H. K. Phan, T. N. K. Nguyen, T. V. Bui, K. C. D. Nguyen, T. P. Nguyen, and N. T. Le, “Recognition of endotracheal suctioning activities: A feature extraction and ensemble learning approach based on pose estimation data,” in *2024 International Conference on Activity and Behavior Computing (ABC)*, pp. 1–9, IEEE, 2024.
- [2] S. Islam, M. Z. Uddin, S. Hossain, and M. A. R. Ahad, “Enhancing nursing activity recognition during endotracheal suctioning through video-based pose estimation and machine learning,” *International Journal of Activity and Behavior Computing*, vol. 2024, no. 3, pp. 1–15, 2024.
- [3] P. Jiang, B. Dai, B. Lyu, Z. Fan, and G. Dogan, “Recognition of nurse activities in endotracheal suctioning procedures: A comparative analysis using lightgbm and other algorithms,” *International Journal of Activity and Behavior Computing*, vol. 2024, no. 3, pp. 1–18, 2024.
- [4] M. Sakai and M. Shuzo, “Hybrid method combining learning machine and rule-based algorithm for activity recognition of endotracheal suctioning; and attempt of data augmentation using generative ai,” *International Journal of Activity and Behavior Computing*, vol. 2024, no. 3, pp. 1–19, 2024.
- [5] H. K. Phan, T. N. K. Nguyen, K. C. D. Nguyen, N. P. Vo, A. T. Ha, and N. T. Le, “Enhanced transportation and locomotion mode recognition through difference and variance analysis in inertial sensing data,” in *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 585–590, 2024.
- [6] H. K. Phan *et al.*, “Student mobility behavior analysis for daily stress recognition.” Presented at the 2024 Activity and Behavior Computing, Fukuoka, Oita, Japan,, 2024.
- [7] H. A. V. Ngo, Q. N. P. Vu, N. Colley, S. Ninomiya, S. Kanai, S. Komizunai, A. Konno, M. Nakamura, and S. Inoue, “Toward recognizing nursing activity in endotracheal suctioning using video-based pose estimation,” *International Journal of Activity and Behavior Computing*, vol. 2024, no. 1, p. 1, 2024.
- [8] K. Xia, W. Li, S. Gan, and S. Lu, “Ts2act: Few-shot human activity sensing with cross-modal co-learning,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 7, no. 4, pp. 1–22, 2024.

- [9] Z. Leng, A. Bhattacharjee, H. Rajasekhar, L. Zhang, E. Bruda, H. Kwon, and T. Plötz, “Imugpt 2.0: Language-based cross modality transfer for sensor-based human activity recognition,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 8, no. 3, pp. 1–32, 2024.
- [10] Y. Zhou, H. Zhao, Y. Huang, T. Röddiger, M. Kurnaz, T. Riedel, and M. Beigl, “Autoaughar: automated data augmentation for sensor-based human activity recognition,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 8, no. 2, pp. 1–27, 2024.
- [11] V. Fortes Rey, L. S. S. Ray, Q. Xia, K. Wu, and P. Lukowicz, “Enhancing inertial hand based har through joint representation of language, pose and synthetic imus,” in *Proceedings of the 2024 ACM International Symposium on Wearable Computers*, pp. 25–31, 2024.
- [12] U. Dobhal, C. Garcia, and S. Inoue, “Synthetic skeleton data generation using large language model for nurse activity recognition,” in *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 493–499, 2024.
- [13] K. L. Moore and A. F. Dalley, *Clinically oriented anatomy*. Wolters kluwer india Pvt Ltd, 2018.
- [14] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’16*, p. 785–794, ACM, Aug. 2016.
- [15] NVIDIA, “Xgboost.”
- [16] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, p. 5–32, Oct. 2001.
- [17] NVIDIA, “Random forest.”
- [18] A. Guryanov, “Histogram-based algorithm for building gradient boosting ensembles of piecewise linear decision trees,” in *Analysis of Images, Social Networks and Texts: 8th International Conference, AIST 2019, Kazan, Russia, July 17–19, 2019, Revised Selected Papers 8*, pp. 39–50, Springer, 2019.
- [19] A. Maćkiewicz and W. Ratajczak, “Principal components analysis (pca),” *Computers & Geosciences*, vol. 19, no. 3, pp. 303–342, 1993.
- [20] S. Roweis, “Em algorithms for pca and spca,” *Advances in neural information processing systems*, vol. 10, 1997.
- [21] H. Abdi and L. J. Williams, “Principal component analysis,” *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433–459, 2010.
- [22] J. Ye, R. Janardan, and Q. Li, “Two-dimensional linear discriminant analysis,” *Advances in neural information processing systems*, vol. 17, 2004.

- [23] W. E. Marcílio and D. M. Eler, “From explanations to feature selection: assessing shape values as feature selection mechanism,” in *2020 33rd SIBGRAPI conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 340–347, Ieee, 2020.
- [24] M. T. Ta, H. K. Phan, and L. G. Tran, “Enhanced semg controlled prosthetic hands activity by xai.” Presented at the 10th International Conference in Vietnam on the Development of Biomedical Engineering (BME10), Phan Thiet, Vietnam, 2024.
- [25] S. Narayan, “The generalized sigmoid activation function: Competitive supervised learning,” *Information sciences*, vol. 99, no. 1-2, pp. 69–82, 1997.
- [26] B. L. Kalman and S. C. Kwasny, “Why tanh: choosing a sigmoidal function,” in *[Proceedings 1992] IJCNN International Joint Conference on Neural Networks*, vol. 4, pp. 578–581, IEEE, 1992.
- [27] A. F. Agarap, “Deep learning using rectified linear units (relu),” *arXiv preprint arXiv:1803.08375*, 2018.
- [28] J. Xu, Z. Li, B. Du, M. Zhang, and J. Liu, “Reluplex made more practical: Leaky relu,” in *2020 IEEE Symposium on Computers and communications (ISCC)*, pp. 1–7, IEEE, 2020.
- [29] X. Liang, X. Wang, Z. Lei, S. Liao, and S. Z. Li, “Soft-margin softmax for deep classification,” in *International Conference on Neural Information Processing*, pp. 413–421, Springer, 2017.
- [30] P. Xue-Kun Song, “Multivariate dispersion models generated from gaussian copula,” *Scandinavian Journal of Statistics*, vol. 27, no. 2, pp. 305–320, 2000.
- [31] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” *Advances in neural information processing systems*, vol. 27, 2014.
- [32] N. Yoshimura, J. Morales, T. Maekawa, and T. Hara, “Openpack: A large-scale dataset for recognizing packaging works in iot-enabled logistic environments,” in *2024 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 90–97, IEEE, 2024.
- [33] N. Yoshimura, J. Morales, and T. Maekawa, “Openpack: Public multi-modal dataset for packaging work recognition in logistics domain,” 2022.
- [34] T. Maekawa, N. Yoshimura, J. Morales, and T. Hara, “Brief introduction of the openpack dataset and lessons learned from organizing activity recognition challenge using the dataset,” in *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 116–120, 2024.
- [35] D. Strömbäck, S. Huang, and V. Radu, “Mm-fit: Multimodal deep learning for automatic exercise logging across sensing devices,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 4, pp. 1–22, 2020.

- [36] A. A. Halim, C. Dartigues-Pallez, F. Precioso, M. Riveill, A. Benslimane, and S. Ghoneim, “Human action recognition based on 3d skeleton part-based pose estimation and temporal multi-resolution analysis,” in *2016 IEEE international conference on image processing (ICIP)*, pp. 3041–3045, IEEE, 2016.
- [37] F. Meng, H. Liu, Y. Liang, J. Tu, and M. Liu, “Sample fusion network: An end-to-end data augmentation network for skeleton-based human action recognition,” *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5281–5295, 2019.
- [38] H. Ramirez, S. A. Velastin, S. Cuellar, E. Fabregas, and G. Farias, “Bert for activity recognition using sequences of skeleton features and data augmentation with gan,” *Sensors*, vol. 23, no. 3, p. 1400, 2023.
- [39] M. I. Mamun, S. Hossain, M. B. Islam, and M. A. R. Ahad, “Generative ai for recognizing nurse training activities in skeleton-based video data,” *International Journal of Activity and Behavior Computing*, vol. 2024, no. 3, pp. 1–20, 2024.
- [40] P. Lin, H. Li, Q. Xie, Y. Zhu, S. Li, and G. Dogan, “Nurse activity recognition using resnet-3 model,” *International Journal of Activity and Behavior Computing*, vol. 2024, no. 3, pp. 1–19, 2024.
- [41] O. Surakhi, M. A. Zaidan, P. L. Fung, N. Hossein Motlagh, S. Serhan, M. AlKhanafseh, R. M. Ghoniem, and T. Hussein, “Time-lag selection for time-series forecasting using neural network and heuristic algorithm,” *Electronics*, vol. 10, no. 20, p. 2518, 2021.
- [42] M. Endres, A. Mannarapotta Venugopal, and T. S. Tran, “Synthetic data generation: A comparative study,” in *Proceedings of the 26th international database engineered applications symposium*, pp. 94–102, 2022.
- [43] N. Patki, R. Wedge, and K. Veeramachaneni, “The synthetic data vault,” in *2016 IEEE international conference on data science and advanced analytics (DSAA)*, pp. 399–410, IEEE, 2016.
- [44] A. Guryanov, “Histogram-based algorithm for building gradient boosting ensembles of piecewise linear decision trees,” in *Analysis of Images, Social Networks and Texts: 8th International Conference, AIST 2019, Kazan, Russia, July 17–19, 2019, Revised Selected Papers 8*, pp. 39–50, Springer, 2019.
- [45] Y. Zhang, H. Zhang, J. Cai, and B. Yang, “A weighted voting classifier based on differential evolution,” in *Abstract and applied analysis*, vol. 2014, p. 376950, Wiley Online Library, 2014.
- [46] R. Arakawa and M. Goel, “Unified framework for procedural task assistants powered by human activity recognition,” in *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 513–518, 2024.
- [47] V. Parikh, S. Mahmud, D. Agarwal, K. Li, F. Guimbretière, and C. Zhang, “Echoguide: Active acoustic guidance for llm-based eating event analysis from egocentric videos,” in

Proceedings of the 2024 ACM International Symposium on Wearable Computers, pp. 40–47, 2024.

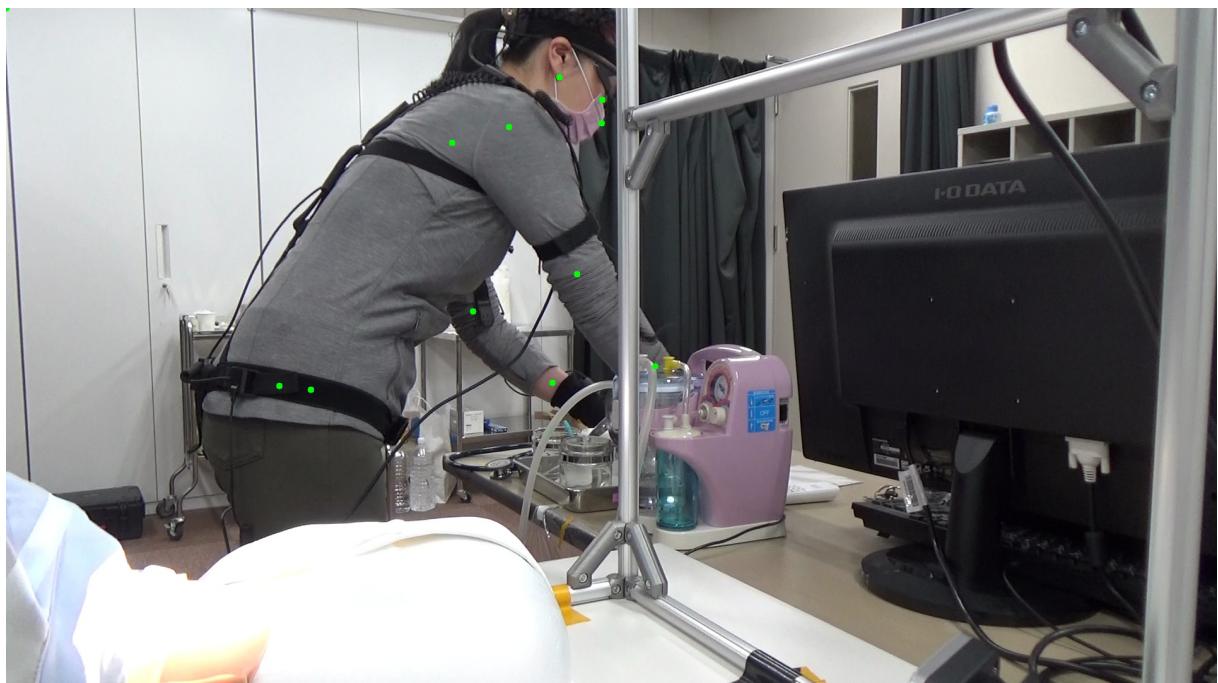
Phụ lục A

Các công bố khoa học cho đồ án này

Một phần đồ án này được trình bày và công bố tại hội nghị ABC 2024 (IEEE, Scopus-indexed) [1]

Phụ lục B

Danh mục hình ảnh hành động hút dịch nội khí quản



Hình B.1: Hình ảnh minh họa hành động chuẩn bị ống thông



Hình B.2: Hình ảnh minh họa hành động tháo tạm thời đường thở nhân tạo



Hình B.3: Hình ảnh minh họa hành động hút đờm



Hình B.4: Hình ảnh minh họa hành động lắp lại đường thở nhân tạo



Hình B.5: Hình ảnh minh họa hành động khử trùng ống thông



Hình B.6: Hình ảnh minh họa hành động thải bỏ găng tay



Hình B.7: Hình ảnh minh họa hành động định vị



Hình B.8: Hình ảnh minh họa hành động nghe nhịp tim