ĐẠI HỌC QUỐC GIA HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA MẠNG MÁY TÍNH VÀ TRUYỀN THÔNG

VƯƠNG ĐINH THANH NGÂN
LÊ MINH NHÃ

ĐỒ ÁN CHUYÊN NGÀNH

MÔ HÌNH KẾT HỢP CHUỖI KHỐI VÀ HỌC LIÊN KẾT CHO HỆ THỐNG Y TẾ THÔNG MINH PHI TẬP TRUNG

AN APPROACH OF FEDERATED LEARNING AND BLOCKCHAIN FOR DECENTRALIZED SMART HEALTHCARE

KỸ SƯ NGÀNH AN TOÀN THÔNG TIN

TP. Hồ Chí Minh, 2023

ĐẠI HỌC QUỐC GIA HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA MẠNG MÁY TÍNH VÀ TRUYỀN THÔNG

VƯƠNG ĐINH THANH NGÂN - 20521649
LÊ MINH NHÃ - 20521690

ĐỒ ÁN CHUYÊN NGÀNH

MÔ HÌNH KẾT HỢP CHUỖI KHỐI VÀ HỌC LIÊN KẾT CHO HỆ THỐNG Y TẾ THÔNG MINH PHI TẬP TRUNG

AN APPROACH OF FEDERATED LEARNING AND BLOCKCHAIN FOR DECENTRALIZED SMART HEALTHCARE

KỸ SƯ NGÀNH AN TOÀN THÔNG TIN

GIẢNG VIÊN HƯỚNG DẪN:
ThS. Phan Thế Duy

TP.Hồ Chí Minh - 2023

LỜI CẢM ƠN

Trong quá trình nghiên cứu và hoàn thành đồ án chuyên ngành, nhóm đã nhận được sự định hướng, giúp đỡ, các ý kiến đóng góp quý báu và những lời động viên của các giáo viên hướng dẫn và giáo viên bộ môn. Nhóm xin bày tỏ lời cảm ơn tới thầy Phan Thế Duy đã tận tình trực tiếp hướng dẫn, giúp đỡ trong quá trình nghiên cứu.

Nhóm cũng chân thành cảm ơn các quý thầy cô trường Đại học Công nghệ Thông tin - ĐHQG TP.HCM, đặc biệt là các thầy cô khoa Mạng máy tính và Truyền thông, các thầy cô thuộc bộ môn An toàn Thông tin đã giúp đỡ nhóm.

Vương Đinh Thanh Ngân **Lê Minh Nhã**

MỤC LỤC

DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

SL	Swarm Learning
FL	Federated Learning
BC	Blockchain
GETH	Go-Ethereum
SWOP	Swarm Operations
SWCI	Swarm Learning Command Interface
SN	Swarm Network
APLS	Autopass License Server
SPIRE	SPIFFE Runtime Environment
SVID	SPIFFE Verifiable Identity
SPIFFE	Secure Production Identity Framework for Everyone

DANH MỤC CÁC HÌNH VẼ

TÓM TẮT ĐỒ ÁN CHUYÊN NGÀNH

Tính cấp thiết của đề tài nghiên cứu:

Trong những năm gần đây, đã có sự phát triển liên quan đến hai khái niệm quan trọng trong lĩnh vực học máy bảo mật: Federated Learning (Học tập Liên Kết) và Swarm Learning. Cả hai phương pháp này đều nhằm tăng cường tính riêng tư và bảo mật khi huấn luyện mô hình trên dữ liệu phân tán.

Trong Federated Learning, mô hình được huấn luyện trên nhiều thiết bị hoặc khách hàng phân tán, mà không cần truyền dữ liệu đến một máy chủ trung tâm. Thay vào đó, các khách hàng tính toán gradient trên dữ liệu cục bộ và gửi nó đến máy chủ trung tâm, nơi gradient được tổng hợp và áp dụng vào mô hình chung. Phương pháp này giúp bảo vệ sự riêng tư của dữ liệu, vì không có dữ liệu gốc được chuyển đến máy chủ trung tâm. Swarm Learning tương tự nhưng trong đó các khách hàng học tập liên kết với nhau và chia sẻ thông tin gradient một cách phi tập trung. Quá trình huấn luyện diễn ra tại các thiết bị cá nhân mà không yêu cầu truyền dữ liệu đến một máy chủ trung tâm.

Tuy nhiên, cả Federated Learning và Swarm Learning đều đối mặt với nguy cơ tấn công Poisoning (nhiễu độc). Trong tấn công này, kẻ tấn công cố gắng thay đổi hoặc xâm nhập vào dữ liệu huấn luyện để làm sai lệch mô hình. Điều này có thể dẫn đến việc mô hình học được các quy tắc sai lệch hoặc dự đoán không chính xác. Tuy nhiên, việc tấn công trong Federated Learning và Swarm Learning khó khăn hơn so với huấn luyện truyền thống, vì kẻ tấn công không thể truy cập hoặc thay đổi toàn bộ dữ liệu huấn luyện.

Tóm lại, Federated Learning và Swarm Learning là hai phương pháp quan trọng trong việc tăng cường tính riêng tư và bảo mật trong huấn luyện mô hình học máy. Tuy nhiên, cả hai phương pháp này vẫn đối mặt với nguy cơ tấn công Poisoning, mặc dù khó khăn hơn so với huấn luyện truyền thống.

CHƯƠNG 1. TỔNG QUAN

1.1. Giới thiệu vấn đề

Trong nhiều tình huống thực tế như trong lĩnh vực y tế, thiết bị IoT và nhiều lĩnh vực khác, dữ liệu được phân phối một cách phi tập trung và khối lượng dữ liệu cục bộ không đủ để đào tạo các mô hình mạnh mẽ và đáng tin cậy. Phương pháp Học tập liên kết (Federated Learning - FL) đã trở nên phổ biến hơn, giảm thiểu một số lo ngại trên. FL cho phép dữ liệu được lưu trữ cục bộ và giải quyết các vấn đề về bảo mật cục bộ. Tuy nhiên, FL vẫn sử dụng một trung tâm giám sát để giữ các tham số mô hình, có thể dễ bị tấn công để lộ thông tin về danh tính và sở thích của người dùng, ngay cả khi sử dụng kỹ thuật Học sâu bảo vệ quyền riêng tư. Ngoài ra, kiến trúc hình sao của FL cũng làm mất khả năng chịu lỗi.

Để giải quyết các vấn đề về đào tạo mô hình máy học trên dữ liệu phân tán có hai phương pháp được đề ra:

- Phương pháp thứ nhất: Đào tạo trên thiết bị mà không cần tải dữ liệu hoặc kết quả trung gian lên máy chủ trung tâm. Tuy nhiên, phương pháp này gây ra sự không khớp về dữ liệu vì mỗi thiết bị có thể có dữ liệu và đặc trưng riêng.

- Phương pháp thứ hai: Federated Learning (FL) phi tập trung, đây là một mô hình FL mới sử dụng một chuỗi khối để điều phối việc cập nhật mô hình và tham số. Một mô hình FL phi tập trung tiên tiến và tiêu biểu nhất, kết hợp hạ tầng phần cứng phi tập trung và học máy phân tán với chuỗi khối để an toàn và tự động bầu chọn người dẫn đầu và hợp nhất các tham số mô hình được gọi là Swarm Learning (SL)

Trong nghiên cứu này, nhóm em đã sử dụng phương pháp thứ hai, sự phi tập trung của Swarm Learning để áp dụng dữ liệu vào các mô hình học máy. Nhóm em chỉ di chuyển những dữ liệu đã học được, thay vì di chuyển toàn bộ dữ liệu. Để đảm bảo tính bảo mật và độ tin cậy, nhóm em sử dụng công nghệ chuỗi khối để tạo ra một "Swarm" gồm nhiều vị trí cạnh, trong đó các thành viên có thể chia sẻ thông tin với nhau một cách an toàn và ngăn chặn những hành động xấu từ việc truy cập trái phép vào Swarm Learning.

1.2. Giới thiệu những nghiên cứu liên quan

1.2.1. Những tiềm năng đối với mô hình học Swarm Learning

Swarm Learning (hay còn gọi là Học tập bày đàn), là một giải pháp học máy phi tập trung, được xây dựng dựa trên tính toán cạnh và công nghệ blockchain, nhằm khuyến khích sự cộng tác ngang hàng. Trong nghiên cứu này, nhóm em tận dụng Swarm Learning để giải quyết vấn đề lớn trong lĩnh vực học máy, đó là sự phụ thuộc vào dữ liệu tập trung. Thay vì chia sẻ dữ liệu chính thức, Swarm Learning cho phép các bên liên quan chia sẻ thông tin quan trọng về dữ liệu mà không tiết lộ nội dung thực tế. Điều này đảm bảo sự bảo mật và riêng tư của dữ liệu trong khi vẫn cho phép tất cả các thành viên đóng góp trong mạng lưới cùng hưởng lợi từ những kiến thức tổng hợp được. Công nghệ blockchain được sử dụng để xây dựng một môi trường tin cậy, ngăn chặn các hành động xấu từ việc truy cập trái phép vào hệ thống và đảm bảo tính toàn vẹn của dữ liệu. Bằng cách kết hợp tính toán cạnh và blockchain, Swarm Learning mang lại một giải pháp mạnh mẽ cho việc cộng tác trong lĩnh vực học máy, mở ra nhiều cơ hội cho việc nghiên cứu và phát triển trong tương lai.

1.2.2. Tấn công Poisoning trong Swarm Learning

Tấn công Poisoning là một hình thức tấn công trong lĩnh vực học máy, mục tiêu của nó là thay đổi hoặc nhiễu độc dữ liệu huấn luyện để làm sai lệch hoặc

phá hủy hiệu suất của mô hình học máy. Trong bối cảnh Swarm Learning, một hệ thống học máy phi tập trung và dựa trên cộng tác ngang hàng, tấn công poisoning có thể gây ảnh hưởng nghiêm trọng đến hiệu suất của các mô hình học máy trong mạng lưới Swarm.

Tấn công poisoning trong Swarm Learning có thể xảy ra khi các bên tham gia vào hệ thống cố tình gửi thông tin sai lệch hoặc dữ liệu bị nhiễu độc. Điều này dẫn đến việc xây dựng các mô hình học máy không đáng tin cậy và có khả năng đưa ra các kết quả không chính xác. Điểm đặc biệt của Swarm Learning là việc chia sẻ thông tin về dữ liệu mà không cần chia sẻ dữ liệu thực tế, tạo điều kiện thuận lợi cho tấn công này.

1.3. Tính ứng dụng mới và sáng tạo

Có thể FL và BC đang là công nghệ phổ biến cho nhiều lĩnh vực như: Tiền điện tử và thanh toán, Internet of Things, Chuỗi cung ứng và quản lý chuỗi cung ứng,..Tuy nhiên, cả FL và BC đang mang trong mình nhưng khuyết điểm bù trừ lẫn nhau và từ đó thì chúng ta dần hình thành nên SL.

SL đảm bảo quá trình học tập sẽ được phân tán hơn FL, dữ liệu cũng sẽ được bảo mật hơn thông qua việc loại bỏ máy chủ trung tâm và thông qua sử dụng thuật toán đồng thuận, smart contract.

SL hiện tại là một công nghệ mới, chưa được ứng dụng quá rộng rãi nhưng không đồng nghĩa với việc nó thiếu tính ứng dụng. SL có thể thay thế các hệ thống đang sử dụng BC kết hợp với FL, góp mặt vào các quy trình tự động, tiền điện tử, y tế,... như FL với BC bởi những tính năng mới có thể bao quát FL và BC.

1.4. Mục tiêu, đối tượng, và phạm vi nghiên cứu

1.4.1. Mục tiêu nghiên cứu

Tìm hiểu sự kết hợp giữa mạng chuỗi khối (blockchain) và học liên kết (FL) phi tập trung tạo ra một mô hình mới Swarm Learning, hiểu được cách thức hoạt động của SL, cũng như khả năng tấn công Poisoning trên mô hình FL và SL trên các tệp dữ liệu riêng biệt trong ngữ cảnh chăm sóc sức khỏe và gian lận thẻ tín dụng.

1.4.2. Đối tượng nghiên cứu

Những đối tượng nghiên cứu trong đồ án của nhóm em bao gồm:

- Mạng chuỗi khối (blockchain)
- Mô hình học liên kết (Federated Learning)
- Mô hình Swarm Learning
- Cơ chế tấn công Poisoning

1.4.3. Cấu trúc đồ án chuyên ngành

Nhóm em xin trình bày nội dung của Đồ án chuyên ngành theo cấu trúc như sau:

- Chương 1: Giới thiệu tổng quan về đề tài của đồ án chuyên ngành và những nghiên cứu liên quan.
- Chương 2: Trình bày cơ sở lý thuyết và kiến thức nền tảng liên quan đến đề tài.
- Chương 3: Trình bày phương pháp luận và thiết kế hệ thống.

- Chương 4: Trình bày kết quả thực nghiệm, phân tích và đánh giá.
- Chương 5: Kết luận và hướng phát triển của đề tài.

CHƯƠNG 2. CƠ SỞ LÝ THUYẾT

Chương này trình bày cơ sở lý thuyết của nghiên cứu: Blockchain, mô hình học máy, mô hình học liên kết, mô hình Swarm Learning

2.1. Blockchain

Trong những năm gần đây, nghiên cứu về công nghệ blockchain và sổ cái thông minh (smart ledger) đã trở nên phổ biến hơn do sự xuất hiện của các loại tiền điện tử như Bitcoin và Ethereum. Blockchain được sử dụng để lưu trữ và chia sẻ dữ liệu một cách phân tán, đáng tin cậy và không thể thay đổi, loại bỏ sự phụ thuộc vào các bên trung gian và không yêu cầu sự can thiệp từ một trung tâm kiểm soát. Sự minh bạch trong blockchain cung cấp một phương pháp đơn giản để truy cập vào các giao dịch được lưu trữ trên sổ cái qua mạng, nó kết nối với sức mạnh tính toán từ nhiều nút (node) trong mạng blockchain, giúp tăng tốc độ tính toán một cách đáng kể. Công nghệ blockchain bao gồm các kỹ thuật và dịch vụ quan trọng như Giao thức Đồng thuận (*Consensus Protocol*), Mã hóa Hash (*Hash Cryptography*), Sổ cái không thể thay đổi (*Immutable Ledger*), Mạng phân tán P2P (*Distributed P2P Networking*) và quá trình khai thác (*mining*).

Công nghệ blockchain sử dụng một mạng phân tán để lưu trữ dữ liệu theo hình thức không thể xâm phạm. Các giao dịch trên blockchain chỉ được cập nhật hoặc thêm mới thông qua việc tạo ra các giá trị mã hash mới, do đó, các giao dịch hiện có không thể bị thay đổi. Các đặc điểm độc đáo của công nghệ blockchain có thể được mô tả như sau:

- **Sổ cái phân tán *(Distributed Ledger)***: Các giao dịch được thêm vào hệ thống phân tán trên mạng, điều này đảm bảo khả năng phục hồi hệ thống

mà không phụ thuộc vào một điểm thất bại duy nhất hay một thực thể trung tâm.

- **Cơ chế đồng thuận (*Consensus Mechanism*)**: Các giao dịch chỉ được cập nhật khi tất cả các người dùng đã được xác minh trong mạng đồng ý với điều kiện của giao dịch.
- **Nguyên gốc (*Provenance*)**: Lịch sử đầy đủ của dữ liệu hoặc tài sản có sẵn trên mạng blockchain.
- **Tính bất biến (*Immutability*)**: Các bản ghi trên mạng không thể bị thay đổi hoặc xâm phạm, từ đó đảm bảo tính an toàn và tin cậy của thông tin.
- **Tính chất dứt khoát (*Finality*)**: Khi một giao dịch được xác nhận trên blockchain, nó không thể bị thay đổi hoặc đảo ngược.
- **Hợp đồng thông minh (*Smart Contract*)**: Mã lệnh được tạo ra trên mạng blockchain, và các máy tính và nút mạng thực thi mã lệnh khi có sự kiện kích hoạt xảy ra. Điều này cho phép mã lệnh tự động thực thi trong một khoảng thời gian nhất định.

Công nghệ blockchain mang lại những lợi ích đáng kể trong việc xây dựng một hệ thống lưu trữ và xử lý dữ liệu an toàn, minh bạch và phi tập trung. Việc hiểu rõ về các tính năng và tiềm năng của công nghệ này có thể cung cấp một cơ sở vững chắc cho việc nghiên cứu và ứng dụng trong nhiều lĩnh vực khác nhau.

***Smart Contract**

Hợp đồng thông minh (Smart Contract) là một thỏa thuận được xây dựng và thi hành tự động thông qua mã lập trình. Chúng có khả năng thực hiện các quy định và điều kiện được xác định trước một cách tự động, không cần sự can thiệp của bên thứ ba.

Sức mạnh tính toán của hợp đồng thông minh phụ thuộc vào quy tắc của hệ thống mà nó hoạt động. Có những hệ thống chỉ hỗ trợ các chức năng cơ bản của

hợp đồng thông minh, ví dụ như việc xác minh chữ ký trong Bitcoin. Trong khi đó, có những hệ thống như Ethereum cho phép thực hiện các chức năng hoàn chỉnh Turing (Turing-complete), có khả năng tính toán mạnh mẽ và linh hoạt.

Hợp đồng thông minh mang lại nhiều lợi ích trong việc tăng cường tính tự động và đáng tin cậy của các giao dịch. Chúng giúp loại bỏ sự phụ thuộc vào sự tin tưởng và can thiệp của bên thứ ba, đồng thời giảm thiểu sự mâu thuẫn và tranh chấp trong quá trình thi hành thỏa thuận. Sự phát triển của công nghệ hợp đồng thông minh đang mở ra nhiều cơ hội mới trong lĩnh vực tài chính, giao dịch điện tử và quản lý hợp đồng trong nhiều ngành công nghiệp khác nhau.

2.2. Mô hình học máy

Mô hình học máy (Machine Learning) là một lĩnh vực trong trí tuệ nhân tạo (Artificial Intelligence) mà các máy tính được lập trình để tự động học hỏi từ dữ liệu và cải thiện hiệu suất mà không cần phải được lập trình rõ ràng cho từng nhiệm vụ cụ thể. Thay vì chỉ thực hiện các tác vụ theo cách mà chúng được lập trình trước đó, máy tính sử dụng các thuật toán và mô hình thống kê để phân tích và rút ra các mẫu, kết luận hoặc dự đoán từ dữ liệu. Máy học đang trở thành một lĩnh vực nghiên cứu quan trọng và đóng vai trò ngày càng quan trọng trong nhiều lĩnh vực ứng dụng, bao gồm nhận dạng hình ảnh, xử lý ngôn ngữ tự nhiên, dự đoán thị trường tài chính, y học và nhiều lĩnh vực khác. Các phương pháp máy học đã đạt được sự chú ý lớn và thành công đáng kể trong thời gian gần đây, nhờ vào sự phát triển của phần cứng mạnh mẽ, khả năng tính toán và sự sẵn có của dữ liệu lớn.

Máy học có thể được chia thành các loại chính như học có giám sát (supervised learning), học không giám sát (unsupervised learning) và học bán giám sát (semi-supervised learning).

- Học máy giám sát (*Supervised learning*): được xác định bởi việc sử dụng các tập dữ liệu được gán nhãn để huấn luyện các thuật toán để phân loại dữ

liệu hoặc dự đoán kết quả một cách chính xác. Khi dữ liệu đầu vào được đưa vào mô hình, mô hình điều chỉnh các trọng số của nó cho đến khi nó được điều chỉnh phù hợp. Điều này xảy ra như một phần của quá trình xác thực chéo để đảm bảo rằng mô hình tránh tình trạng quá khớp hoặc thiếu khớp. Học máy giám sát giúp các tổ chức giải quyết một loạt các vấn đề thực tế với quy mô lớn, chẳng hạn như phân loại thư rác vào một thư mục riêng biệt khỏi hộp thư đến của bạn. Một số phương pháp được sử dụng trong học máy giám sát bao gồm mạng thần kinh (neural networks), naïve bayes, hồi quy tuyến tính (linear regression), hồi quy logistic (logistic regression), random forest và support vector machine (SVM).

- Học máy không giám sát (*Unsupervised learning*): sử dụng các thuật toán học máy để phân tích và nhóm các tập dữ liệu không được gán nhãn. Những thuật toán này khám phá các mẫu ẩn hoặc nhóm dữ liệu mà không cần sự can thiệp của con người. Khả năng của phương pháp này trong việc khám phá sự tương đồng và khác biệt trong thông tin làm cho nó lý tưởng cho việc phân tích dữ liệu khám phá, chiến lược bán hàng chéo, phân đoạn khách hàng, và nhận dạng hình ảnh và mẫu. Nó cũng được sử dụng để giảm số lượng đặc trưng trong một mô hình thông qua quá trình giảm chiều dữ liệu. Phân tích thành phần chính (PCA-Principal Component Analysis) và phân tích giá trị đơn (SVD-Singular Value Decomposition) là hai phương pháp thông thường được sử dụng cho điều này. Các thuật toán khác được sử dụng trong học máy không giám sát bao gồm mạng thần kinh (neural networks), phân cụm k-means (k-means cluster) và các phương pháp phân cụm xác suất.

- Học máy bán giám sát (*Semi-supervised learning*): cung cấp một sự kết hợp lý tưởng giữa học máy giám sát và học máy không giám sát. Trong quá trình huấn luyện, nó sử dụng một tập dữ liệu nhãn nhỏ hơn để hướng dẫn phân loại và trích xuất đặc trưng từ một tập dữ liệu không được gán nhãn

lớn hơn. Học máy bán giám sát có thể giải quyết vấn đề không có đủ dữ liệu được gán nhãn cho thuật toán học máy giám sát. Nó cũng hữu ích khi việc gán nhãn đủ dữ liệu quá tốn kém.

2.3. Mô hình học liên kết

Mô hình học liên kết (FL-Federated Learning) là một mô hình tính toán mới trong lĩnh vực nghiên cứu khoa học [2]. Trong mô hình này, dữ liệu được lưu trữ trên các thiết bị của người dùng và không bao giờ được thu thập và tập trung lại ở một nơi duy nhất. Thay vào đó, các bản cập nhật mô hình tối thiểu và tập trung được truyền đến máy chủ để tiến hành quá trình huấn luyện.

Phương pháp FL cho phép chúng ta huấn luyện các mô hình mà không cần thu thập dữ liệu từ các thiết bị của người dùng. Điều này mang lại lợi ích lớn về bảo mật và quyền riêng tư, vì dữ liệu người dùng không phải rời khỏi thiết bị của họ. Thay vào đó, chỉ có các thông tin cập nhật nhỏ và tập trung được truyền về máy chủ để cập nhật mô hình. FL cũng có thể được kết hợp với các kỹ thuật bảo mật khác để tăng cường quyền riêng tư. Ví dụ, phương pháp tính toán đa bên an toàn (secure multi-party computation) có thể được áp dụng để đảm bảo tính bảo mật khi các thiết bị người dùng tham gia vào quá trình huấn luyện mô hình. Ngoài ra, kỹ thuật bảo mật khác biệt (differential privacy) cũng có thể được sử dụng để bảo vệ thông tin cá nhân khi tiến hành các phép tính và truyền dữ liệu.

Một điểm mạnh của FL là khả năng ổn định và hiệu quả khi xử lý dữ liệu không cân bằng và không đồng nhất (non-IID). Điều này có ý nghĩa rằng các mô hình huấn luyện bằng FL có thể đạt được hiệu suất tốt ngay cả khi dữ liệu từ các thiết bị người dùng không tuân theo sự phân phối đồng nhất. Một triển khai điển hình của FedSGD với $C = 1$ và một tỉ lệ học tập cố định η được thực hiện bằng cách mỗi khách hàng k tính toán $g_k = \bigtriangledown F_k(w_t)$, đạo hàm trung bình trên dữ liệu cục bộ của nó tại mô hình hiện tại w_t, và máy chủ trung tâm

tổng hợp những đạo hàm này và áp dụng cập nhật $w_{t+1} \longleftarrow w_t - \eta \sum_{k=1}^{K} \frac{n_k}{n} g_k$, vì $\sum_{k=1}^{K} \frac{n_k}{n} g_k = \bigtriangledown f(w_t)$. Một cập nhật tương đương được cho bởi $\forall k, w_{t+1} \longleftarrow w_t - \eta g_k$ và sau đó $w_{t+1} \longleftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$. Đó là, mỗi khách hàng cục bộ thực hiện một bước của tính toán gradient trên mô hình hiện tại bằng cách sử dụng dữ liệu cục bộ của mình, và sau đó máy chủ lấy trung bình có trọng số của các mô hình kết quả. Sau khi thuật toán được viết theo cách này, chúng ta có thể thêm tính toán hơn cho mỗi khách hàng bằng cách lặp lại cập nhật cục bộ $w^k \longleftarrow w^k - \eta \bigtriangledown F_k(w^k)$ nhiều lần trước bước trung bình hóa. Họ gọi phương pháp này là FederatedAveraging (hoặc FedAvg). Số lượng tính toán được điều khiển bởi ba tham số chính: C, tỷ lệ khách hàng thực hiện tính toán trong mỗi vòng lặp; E, số lần huấn luyện mỗi khách hàng thực hiện trên tập dữ liệu cục bộ của nó trong mỗi vòng lặp; và B, kích thước minibatch cục bộ được sử dụng cho các cập nhật khách hàng. Chúng tôi viết $B = \infty$ để chỉ rằng toàn bộ tập dữ liệu cục bộ được coi là một minibatch duy nhất. Do đó, ở một đầu của họ gia đình thuật toán này, chúng ta có thể lấy $B = \infty$ và $E = 1$ tương ứng với FedSGD chính xác. Đối với một khách hàng có nk ví dụ cục bộ, số lượng cập nhật cục bộ mỗi vòng lặp được cho bởi $u_k = E\frac{n_k}{B}$;

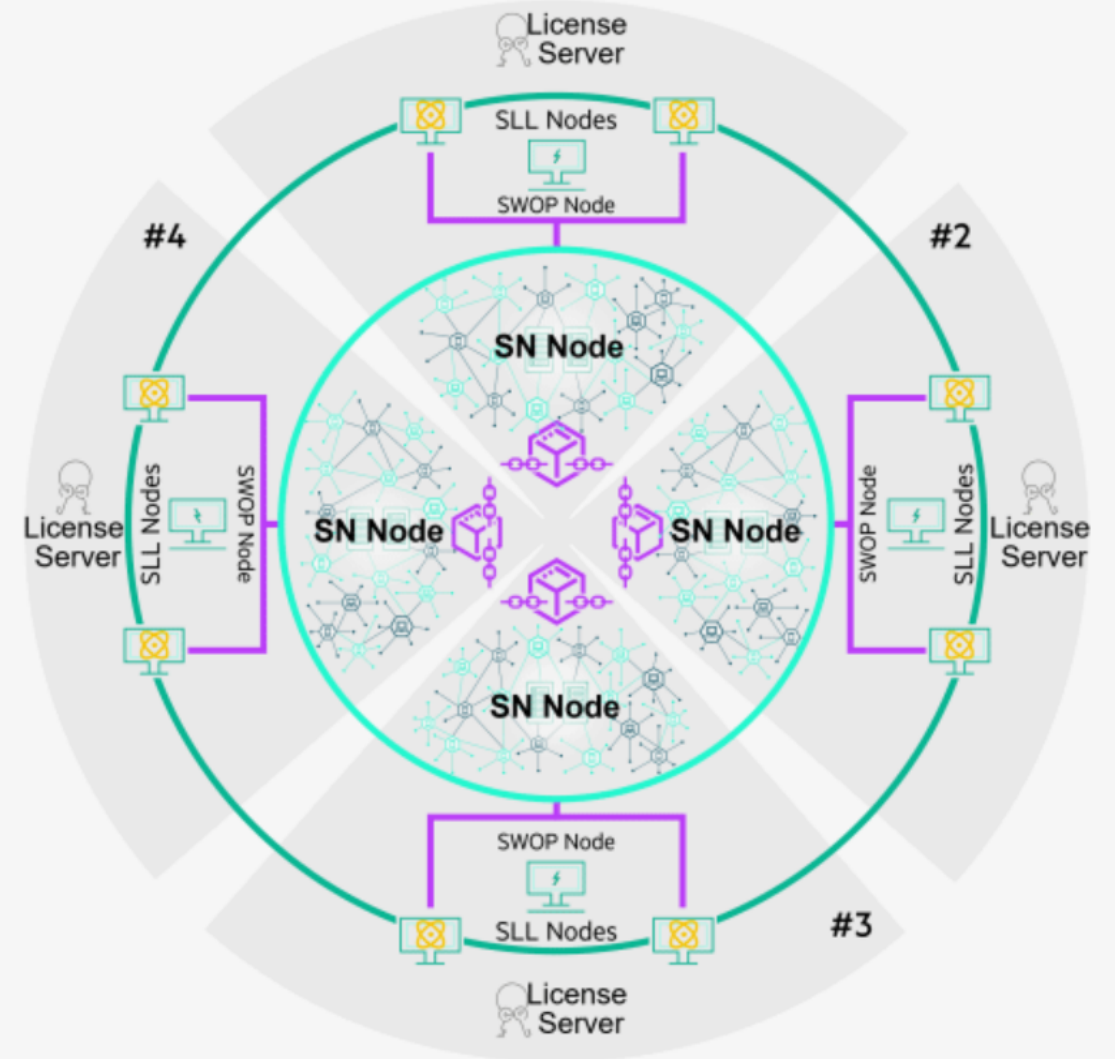
2.4. Mô hình Swarm Learning

2.4.1. Tổng quan

Mô hình Swarm Learning là một mô hình FL phi tập trung kết hợp hạ tầng phần cứng phi tập trung và học máy phân tán với chuỗi khối để an toàn và tự động bầu chọn người dẫn đầu và hợp nhất các tham số mô hình.

SL chia sẻ các tham số mô hình thông qua mạng Swarm và xây dựng các mô hình độc lập trên dữ liệu riêng tư tại các nút cạnh của Swarm mà không cần người giám sát trung tâm. SL đảm bảo quyền riêng tư, bảo mật và an toàn dữ liệu nhờ vào tính năng của chuỗi khối. Mỗi người tham gia được xác định rõ ràng và chỉ những người được ủy quyền mới có thể tham gia và thực hiện giao

dịch. Trong quá trình làm việc của SL, một nút cạnh mới đăng ký thông qua hợp đồng thông minh chuỗi khối, lấy mô hình và thực hiện đào tạo cục bộ cho đến khi đạt được thời gian đồng bộ hóa do người dùng xác định. Sau đó, các tham số mô hình cục bộ được trao đổi và hợp nhất để cập nhật mô hình toàn cầu trước vòng lặp huấn luyện tiếp theo. Các thông tin dưới đây được trích dẫn từ trang chủ của công ty **Hewlett Packard Enterprise (HPE)**, là một công ty đã thành công xây dựng và phát triển khung Swarm Learning.

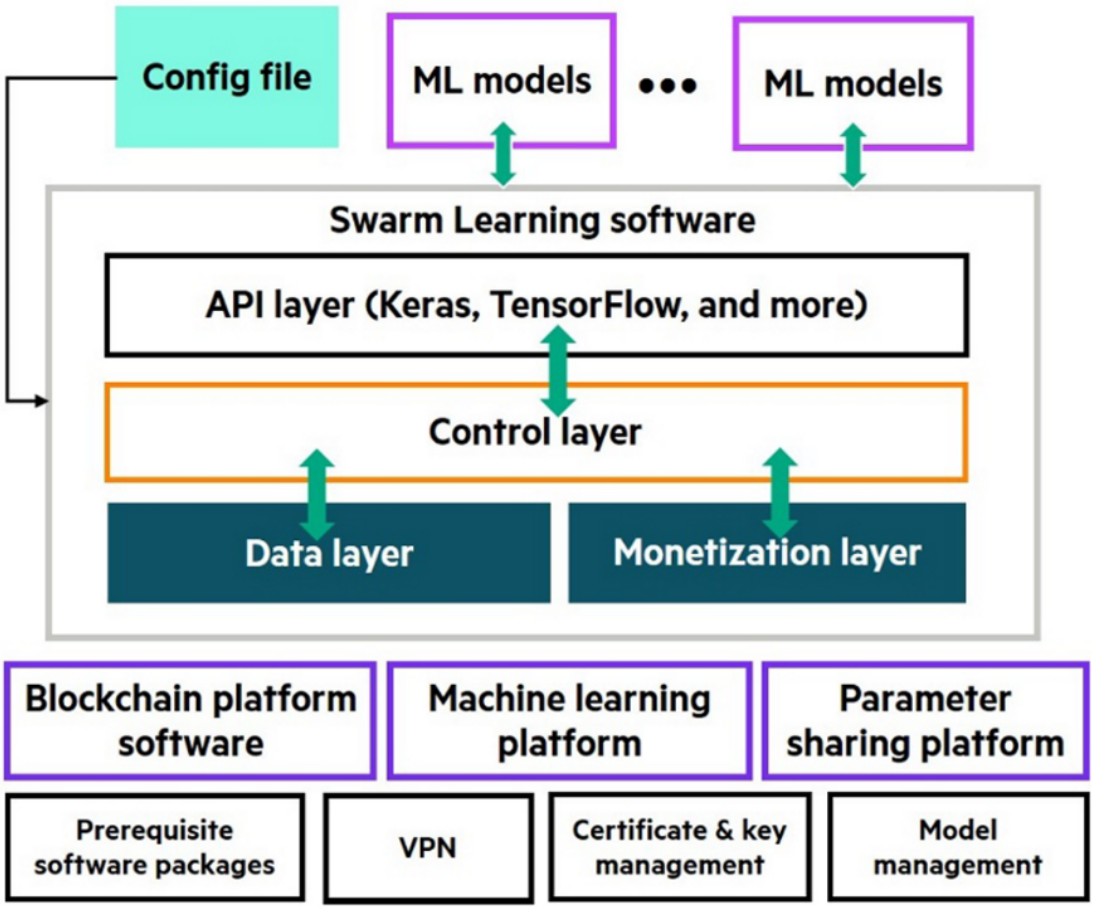


Hình 2.1: *Mô hình tổng quan về Swarm Learning*

Ngoài các nút cạnh trong Swarm, còn có các nút điều phối khác của Swarm có trách nhiệm duy trì siêu dữ liệu như trạng thái mô hình, tiến trình đào tạo

và giấy phép mà không bao gồm tham số mô hình.

Một mạng SL có thể chia thành 4 lớp: API, Control, Data, Monetization.



Hình 2.2: *Các lớp trong một mạng Swarm Learning*

- **Lớp API**: Swarm Learning được triển khai dưới dạng một thư viện API có sẵn cho nhiều khung công việc phổ biến như Keras, và các khung công việc tương tự. Các API này cung cấp một giao diện tương tự với các API huấn luyện trong các khung công việc gốc mà các nhà khoa học dữ liệu quen thuộc. Gọi các API này sẽ tự động chèn các *hook* cần thiết cho Swarm Learning để các nút tự động trao đổi tham số vào cuối mỗi vòng lặp huấn luyện mô hình và tiếp tục huấn luyện sau khi thiết lập lại mô hình cục bộ dựa trên tham số đã hợp nhất tại mô hình toàn cầu. Với một số thay đổi mã đơn giản, toàn bộ mạng học như một đơn vị với tất cả những phức tạp

về điều khiển và luồng dữ liệu xảy ra tự động bên dưới nền.

- **Lớp Control**: Trách nhiệm duy trì mạng Swarm phi tập trung trong trạng thái nhất quán toàn cầu nằm ở tầng điều khiển và nó được triển khai bằng công nghệ blockchain. Tầng điều khiển này đảm bảo rằng tất cả các hoạt động và trạng thái của mạng được thực hiện một cách đồng nhất và không thể phân mảnh. Trạng thái của mạng Swarm bao gồm thông tin về epoch hiện tại, các thành viên hiện tại của Swarm với địa chỉ IP và cổng mà họ đang sử dụng, và các URI (Uniform Resource Identifier) để truy cập các tệp tham số. Tập hợp các hoạt động bao gồm các quy tắc và luật logic để bầu chọn người đứng đầu (hay còn gọi là lãnh đạo) của Swarm vào cuối mỗi epoch, cơ chế để ngăn chặn lỗi và tự phục hồi, cùng với việc tín hiệu giữa các nút để bắt đầu và hoàn thành các giai đoạn khác nhau của quá trình.

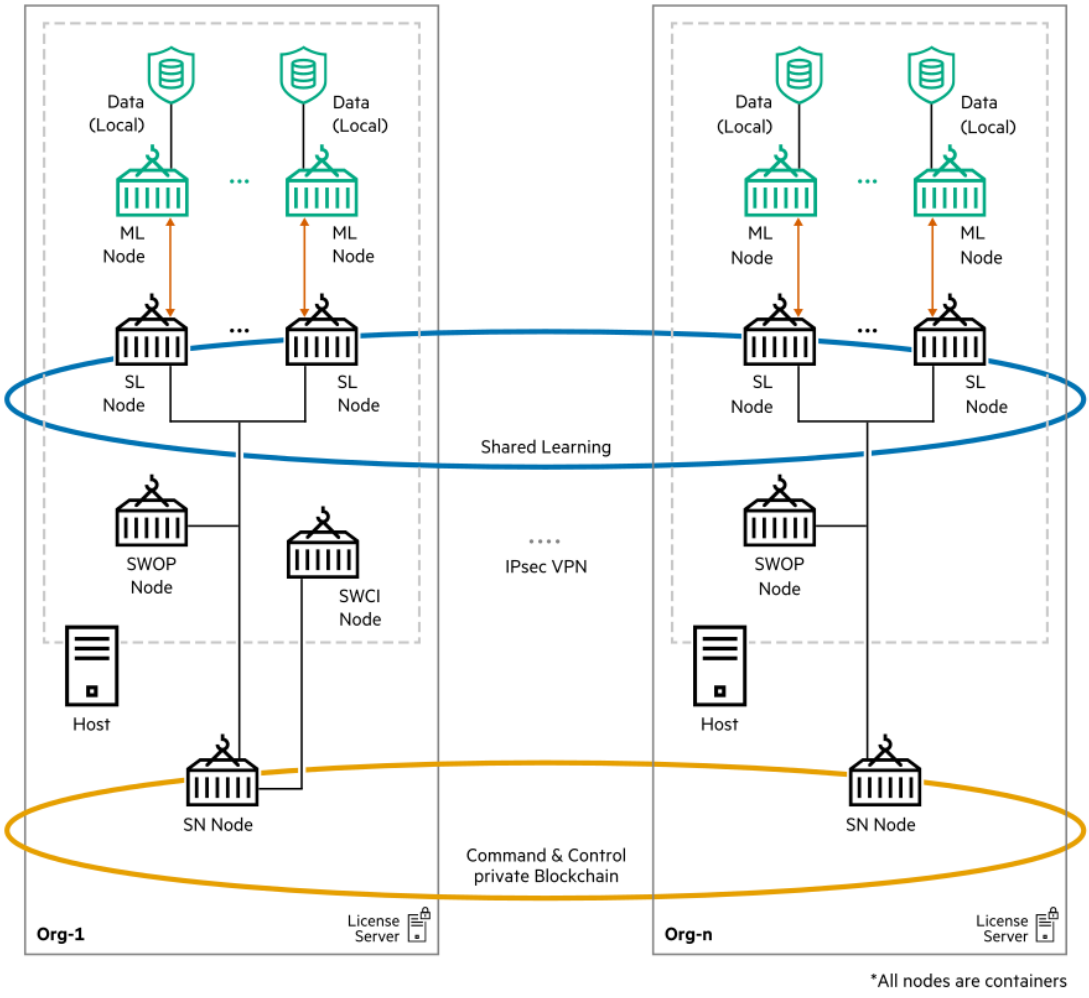
- **Lớp Data**: Tầng dữ liệu đảm nhận việc chia sẻ các tham số mô hình một cách đáng tin cậy và an toàn trên mạng Swarm. Tầng này là nơi các tham số của mô hình được truyền đi và nhận về giữa các nút trong mạng Swarm. Tầng dữ liệu có khả năng thay thế và hỗ trợ nhiều cơ chế khác nhau để chia sẻ các tệp tin, bao gồm HTTPS/TLS, IPFS và các cơ chế khác. Điều này cho phép sự linh hoạt và tương thích với nhiều giao thức và cơ chế chia sẻ dữ liệu khác nhau. Tầng này được điều khiển thông qua các hoạt động được gọi bởi tầng điều khiển, các hoạt động này đảm bảo việc truyền thông tin và duy trì trạng thái hoạt động của tầng dữ liệu.

- **Lớp Monetization**: Tầng dữ liệu này đo lường việc sử dụng dữ liệu và đóng góp trong quá trình huấn luyện mô hình để tính toán phần thưởng tiền tệ cho mỗi thành viên tham gia, phần thưởng sẽ được trao tại cuối quá trình huấn luyện. Để đảm bảo tính minh bạch và công bằng, tầng dữ liệu sử dụng hợp đồng thông minh không thể sửa đổi và tự xác minh của blockchain để theo dõi các đóng góp từng thành viên, và khung công nghệ tiền điện tử tích hợp sẵn để chuyển giao phần thưởng một cách hoàn toàn

tự động.

2.4.2. Cấu trúc mô hình Swarm Learning

Swarm Learning có cấu trúc mô hình chủ yếu với 5 thành phần chính: Swarm Network, Swarm Operator, Swarm Command Interface, License Server, Swarm Learning.[4][3][1]



Hình 2.3: *Cấu trúc mô hình Swarm Learning*

Các nút Swarm Learning là các nút cạnh Swarm, còn các nút Swarm Network, Licence Server, Swarm Operations, Swarm Command Interface là các nút điều phối khung SL.

- **License Server (LS)**: Máy chủ giấy phép là được tạo bởi nút LS dùng để lưu trữ cài đặt và quản lý giấy phép để chạy khung SL. Cổng API của LS được sử dụng bởi nút LS để chạy một máy chủ API dựa trên REST và một giao diện quản lý. Máy chủ API được sử dụng bởi các nút SN và nút SL để kết nối với nút LS và lấy giấy phép. Giao diện quản lý được sử dụng bởi các quản trị viên khung SL để kết nối với nút LS từ trình duyệt và quản lý các giấy phép. Các giấy phép được cấp phép từ máy chủ giấy phép APLS (Autopass Liecense Server).

- **Swarm Network (SN)**: là một mạng lưới blockchain được tạo thành bởi các nút SN. Trong mạng Swarm Learning, phiên bản hiện tại sử dụng một phiên bản mã nguồn mở của Ethereum để xây dựng nền tảng blockchain. Các nút SN tương tác với nhau thông qua nền tảng blockchain này để duy trì và theo dõi tiến trình của mạng. Các nút SN sử dụng thông tin trạng thái và tiến trình này để điều phối hoạt động của các thành phần Swarm Learning khác. Trong đó, **Sentinel Node** là một nút SN đặc biệt. Nút Sentinel chịu trách nhiệm khởi tạo mạng blockchain. Đây cũng là nút đầu tiên được khởi động trong khung SL.

- **Swarm Operator (SWOP)**: một nút đại diện có thể quản lý các hoạt động của Swarm Learning. SWOP chịu trách nhiệm thực hiện các nhiệm vụ được giao cho nó. Một nút SWOP chỉ có thể thực hiện một nhiệm vụ tại một thời điểm. SWOP giúp thực hiện các nhiệm vụ như khởi động và dừng chạy Swarm, xây dựng và nâng cấp các thùng chứa ML (ML container), và chia sẻ mô hình để huấn luyện.

- **Swarm Command Interface (SWCI)**: công cụ giao diện dòng lệnh dùng để tương tác và điều khiển các hoạt động của mạng Swarm. SWCI bằng cách sử dụng các ngữ cảnh (contexts) và hợp đồng (contracts) để xem trạng thái, điều khiển và quản lý khung công việc của Swarm Learning.

- **Swarm Learning (SL)**: Không hiểu nó là mô hình Swarm Learning mà

trong ngữ cảnh này, SL được hiểu là các nút cạnh của mô hình. Nút SL đề cập đến các thiết bị hoặc nút tính toán tham gia vào quá trình học tập SL. Các nút SL là các thành viên trong mạng Swarm và đóng vai trò quan trọng trong việc tích hợp dữ liệu cục bộ và chứa các nút ML cùng tham gia vào quá trình học tập trong mô hình SL.

Cách thức hoạt động của các nút trong mạng Swarm Learning trên sẽ được giải thích chi tiết trong Chương 3.

2.5. Cơ chế Poisoning

Cơ chế poisoning (hoặc còn được gọi là data poisoning) là một hình thức tấn công mà các bên thứ 3 cố gắng thay đổi hoặc làm sai lệch dữ liệu trong quá trình huấn luyện phân tán. Mục tiêu của tấn công poisoning là làm cho mô hình chung (global model) trở nên không đáng tin cậy hoặc gây ra lỗi trong quá trình dự đoán. Có rất nhiều loại tấn công poisoning tới FL và bao gồm SL:

- Tấn công dữ liệu gian lận (Data Poisoning Attack): bên tấn công chèn dữ liệu sai lệch vào dữ liệu huấn luyện trên các thiết bị địa phương. Điều này có thể làm sai lệch quá trình huấn luyện và làm mô hình chung không đáng tin cậy.

- Tấn công với mục tiêu (Targeted Attack): Trong loại tấn công này, bên tấn công tập trung vào một số lớp hoặc nhãn cụ thể trong mô hình. Họ có thể cố gắng thay đổi dữ liệu để làm sai lệch dự đoán của mô hình chung đối với các lớp hoặc nhãn này.

- Tấn công dự đoán gian lận (Model Poisoning Attack): Kẻ tấn công cố gắng thay đổi quá trình huấn luyện của mô hình chung bằng cách gửi dữ liệu gian lận từ các thiết bị địa phương, làm sai lệch mô hình chung và làm cho nó không đáng tin cậy.

- Tấn công phá hủy (Denial-of-Service Attack): Trong loại tấn công này, kẻ cấu cố gắng gây sự cố cho quá trình học liên kết bằng cách gửi dữ liệu sai hoặc gửi lượng lớn dữ liệu gây quá tải hệ thống.

2.6. Các công trình nghiên cứu liên quan

Khung học tập của Swarm Learning đã được đề xuất bởi **Hewlett Packard Enterprise (HPE)** - một công ty edge-to-cloud toàn cầu. Năm 2021, một nhóm tác giả đến từ nhiều trường đại học lớn như Đại học Bonn của Đức cùng với sự hợp tác của các kỹ sư cấp cao tại HPE, đã đề xuất một khung Swarm Learning có ứng dụng trong lĩnh vực học máy lâm sàn [4]. Khung Swarm Learning này được thử nghiệm và đạt được nhiều kết quả đáng chú ý:

- Dự đoán bệnh bạch cầu từ dữ liệu PBMCPBMC: Sử dụng dữ liệu biểu hiện gen để dự đoán bệnh bạch cầu, với sự phân chia dữ liệu huấn luyện và kiểm tra tại các nút Swarm. Họ đã thử nghiệm và so sánh nhiều kịch bản khác nhau, và thực hiện các thử nghiệm bổ sung với các trường hợp bệnh bạch cầu lympho cấp và mở rộng số nút và loại bạch cầu để đánh giá hiệu suất của SL.

- Xác định bệnh nhân mắc bệnh lao hoặc bệnh lý phổi: Swarm learning (SL) đã được sử dụng để xác định bệnh nhân mắc bệnh lao (TB) từ các biểu đồ chuyển đổi gen trong máu. SL đã được so sánh với các nút cá nhân và mô hình trung tâm trong việc phân loại các trường hợp TB và nhóm điều khiển. Kết quả cho thấy SL vượt trội hơn các nút cá nhân và hoạt động tốt hơn trong các điều kiện khó khăn khi số lượng mẫu huấn luyện bị giảm. SL cũng đã được áp dụng để dự đoán các kết quả hình ảnh từ tập dữ liệu X-quang ngực và cho thấy hiệu suất vượt trội so với các nút cá nhân trong việc phân loại các kết quả khác nhau

- Phát hiện bệnh nhân mắc COVID-19 trong cùng một kịch bản bùng phát:

SL sử dụng huyết đồ gen (blood transcriptomics) để đánh giá phản ứng miễn dịch của máy chủ và dự đoán bệnh COVID-19. SL đã được thử nghiệm trên nhiều tình huống và cho thấy hiệu suất vượt trội so với các nút huấn luyện và các phương pháp riêng lẻ khác. Ngoài ra, SL cũng có khả năng xử lý các sai lệch và phân biệt giữa các trường hợp COVID-19 nhẹ và nặng.

Và vào năm 2022, một nhóm tác giả đến từ Canada hợp tác cùng nhiều tác giả đến từ nhiều nơi khác, đã đề xuất tiếp một khung Swarm Learning về ngữ cảnh bệnh ung thư [3], họ cũng đưa ra được các kết quả khả quan của khung SL trong việc dự đoán bệnh:

- Khung SL cho phép đào tạo các mô hình Trí tuệ nhân tạo (AI) cho bệnh lý. Cụ thể, nhóm đã phát triển một đường ống AI sử dụng SL để phân loại phân tử các khối u rắn dựa trên hình ảnh mô bệnh học.
- Khung SL có khả năng dự đoán trạng thái đột biến BRAF, giúp xác định các biến thể di truyền của bệnh ung thư.
- Khung SL cũng có thể dự đoán về trạng thái không ổn định của microsatellite (MSI)/sự thiếu hụt sửa lỗi không phù hợp (dMMR), một yếu tố quan trọng trong đánh giá tính đáng tin cậy và điều trị ung thư.
- Khung SL cũng có thể dự đoán về trạng thái không ổn định của microsatellite (MSI)/sự thiếu hụt sửa lỗi không phù hợp (dMMR), một yếu tố quan trọng trong đánh giá tính đáng tin cậy và điều trị ung thư.
- Tính hiệu quả của khung SL được chứng minh qua việc hiệu quả trong việc thu thập và xử lý dữ liệu cũng như lấy các mẫu hợp lý liên quan đến nghiên cứu bệnh ung thư.

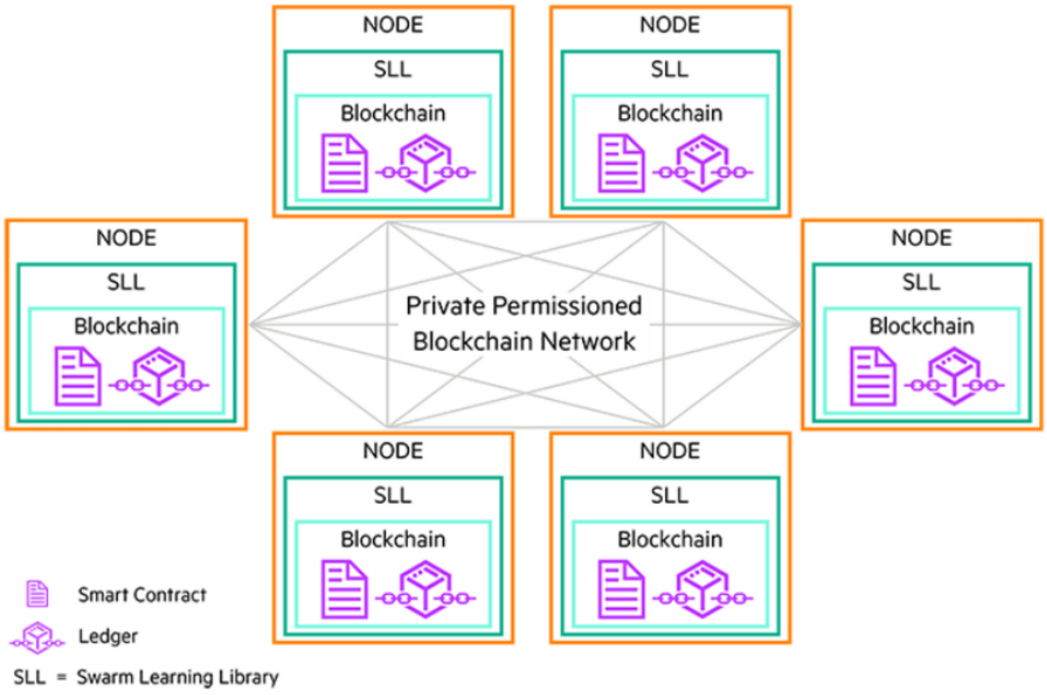
Những kết quả nổi bật này cho thấy tiềm năng và sự hứa hẹn của Swarm Learning trong ứng dụng y học, đặc biệt là trong việc hỗ trợ chuẩn đoán bệnh và điều trị ung thư.

CHƯƠNG 3. PHƯƠNG PHÁP LUẬN VÀ THIẾT KẾ HỆ THỐNG

3.1. Federated Learning phi tập trung kết hợp Blockchain tạo ra tính bảo mật cho dữ liệu

Mạng riêng tư có quyền truy cập được xây dựng dựa trên mô hình phân phối hoàn toàn phi tập trung, trong đó không có bộ điều khiển trung tâm. Các hoạt động trong mạng này được thực hiện thông qua hợp đồng thông minh hoạt động trên blockchain (*Hình 3.1*). Giao tiếp giữa các thành viên trong mạng sử dụng mô hình P2P, không yêu cầu phiên kết nối vĩnh viễn. Để đảm bảo tính an toàn, mTLS (Transport Layer Security) được sử dụng cho tất cả các giao tiếp trong mạng. Các container cũng được củng cố với các biện pháp bảo mật bao gồm quét mã độc, quét lỗ hổng và tính năng ký mã.

Mạng blockchain riêng tư của Swarm được thiết kế để ngăn chặn đầu vào trái phép vào mô hình. Hợp đồng thông minh trên blockchain đảm bảo bảo vệ chống lại các thành viên không trung thực trong mạng. Mô hình chỉ cho phép truy cập trong các thùng chứa (container) an toàn, đồng thời loại bỏ máy chủ trung tâm hoặc người quản lý, tạo ra một mô hình phi tập trung và toàn cầu. Ngoài ra, mạng này cũng ngăn chặn truy cập white-box, tức là truy cập vào chi tiết nội dung của hệ thống. Mạng hỗ trợ các phương pháp ẩn danh và giả danh (anonymization, pseudonymization) để bảo vệ thông tin cá nhân. Định danh và kiểm soát quyền truy cập có thể được thực hiện thông qua SPIFFE OPA, đảm bảo rằng chỉ những người được ủy quyền mới có thể truy cập vào dữ liệu. Các thùng chứa (container) an toàn giúp ngăn chặn truy cập trái phép vào dữ liệu trên nút cục bộ, đảm bảo tính bảo mật. Hơn nữa, mạng cũng hỗ trợ các



Hình 3.1: *Blockchain trong Swarm Learning*

phương pháp biến đổi dữ liệu (data perturbation) và quyền riêng tư khác biệt (differential privacy) để bảo vệ thông tin và đảm bảo tính riêng tư của người dùng.

3.2. Quy trình huấn luyện Swarm Learning

Quy trình làm việc của Swarm Learning có thể được chia thành **ba giai đoạn** hoạt động chính:

3.2.1. Khởi tạo và tham gia

Quá trình tham gia là một quá trình ngoại tuyến mà liên quan đến việc các thực thể quan tâm đến việc sử dụng Swarm Learning trong ML hợp tác và đề ra các yêu cầu vận hành và pháp lý của hệ thống phi tập trung. Điều này bao gồm các khía cạnh như thỏa thuận chia sẻ dữ liệu (tham số), các thỏa thuận để đảm

bảo khả năng nhìn thấy các nút qua các ranh giới tổ chức của các thực thể và đạt được sự nhất trí về kết quả dự kiến từ quá trình huấn luyện mô hình. Các giá trị của các tham số có thể cấu hình do Swarm cung cấp, chẳng hạn như các nút phát hiện ngang hàng được cung cấp trong quá trình khởi động và tần suất đồng bộ hóa giữa các nút, cũng được hoàn thiện ở giai đoạn này. Cuối cùng, cần đồng ý về mô hình chung để huấn luyện và hệ thống thưởng (nếu có).

3.2.2. Cài đặt và cấu hình

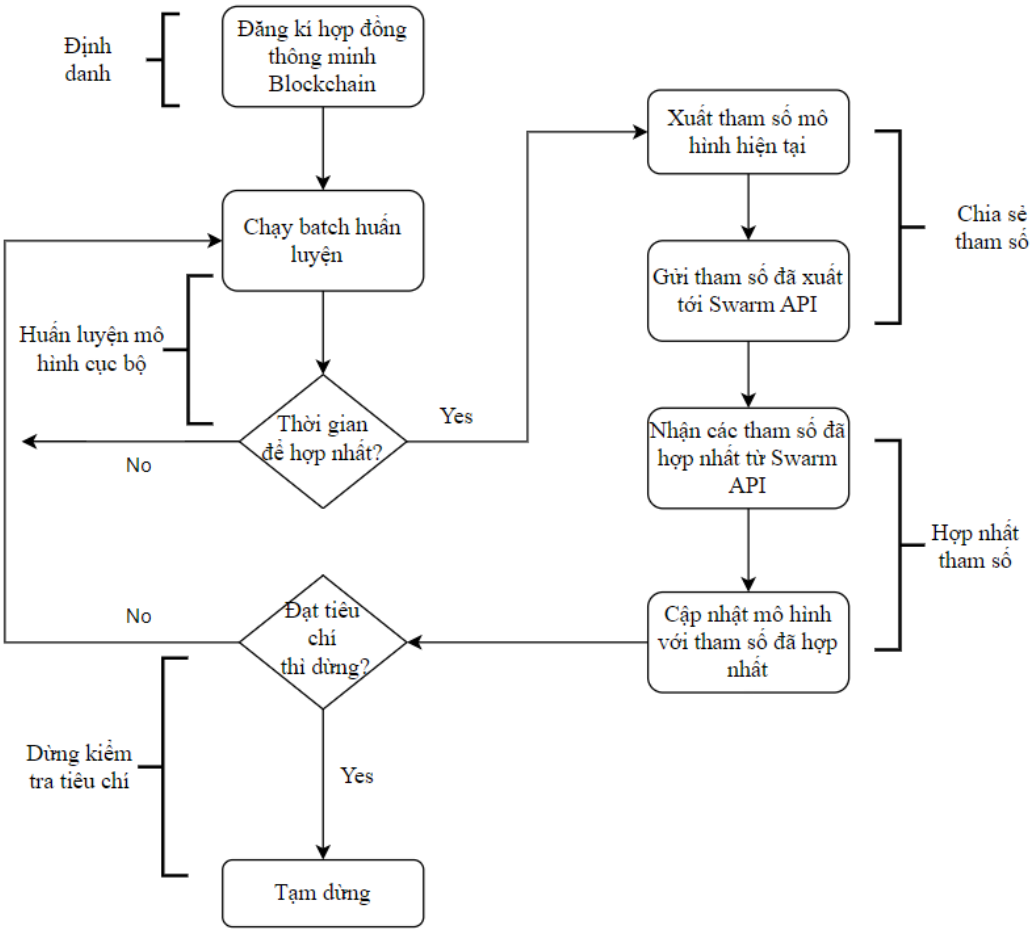
Sau khi quá trình tham gia hoàn thành, tất cả các thành viên trong nhóm hợp tác tải xuống và cài đặt nền tảng Swarm trên các máy chủ riêng của họ (nút), trong quá trình đó cũng được cung cấp cấu hình cho mạng Swarm Learning đã hoàn thiện trong giai đoạn khởi tạo và tham gia. Sau đó, nền tảng Swarm Learning được khởi động và khởi tạo kết nối của các nút với mạng Swarm, đó là một tầng lớp blockchain trên kết nối mạng cơ bản giữa các nút. Quá trình khởi động được thực hiện theo trật tự, trong đó tập hợp các nút tham gia được chỉ định là các nút phát hiện ngang hàng (trong giai đoạn khởi tạo) được khởi động trước, sau đó là các nút còn lại trong mạng.

3.2.3. Tích hợp và huấn luyện

Swarm Learning cung cấp một tập hợp các API đơn giản để cho phép tích hợp nhanh chóng với nhiều frameworks khác nhau. Những API này được tích hợp vào mã nguồn hiện có để nhanh chóng chuyển đổi một nút máy học độc lập thành một thành viên tham gia vào Swarm Learning. Quá trình huấn luyện mô hình có thể được chia thành các bước sau: (3.2)

- Bước 1: Định danh

  Quá trình Swarm Learning bắt đầu với việc đăng ký, hoặc đăng nhập, vào hợp đồng thông minh Swarm bởi mỗi nút. Đây là một quá trình chỉ thực hiện một lần. Sau đó, mỗi nút ghi lại các thuộc tính liên quan của nó trong



Hình 3.2: *Quá trình huấn luyện mô hình Swarm Learning*

hợp đồng, chẳng hạn như định danh tài nguyên thống nhất (URI) từ đó tập hợp các tham số đã được huấn luyện của chính nút đó có thể được tải xuống bởi các nút khác.

- Bước 2: Huấn luyện mô hình cục bộ

  Sau đó, các nút tiếp tục huấn luyện bản sao cục bộ của mô hình theo phương pháp lặp lại qua nhiều vòng, mỗi vòng gọi là một epoch. Trong mỗi epoch, mỗi nút huấn luyện mô hình cục bộ của mình bằng cách sử dụng một hoặc nhiều nhóm dữ liệu trong một số lần lặp cố định. Khi số lần lặp đã đạt đến ngưỡng, nút xuất giá trị tham số vào một tệp và tải lên hệ thống tệp chia sẻ để các nút khác có thể truy cập. Sau đó, nút thông báo cho các nút khác

biết rằng nó đã sẵn sàng cho bước chia sẻ tham số.

- Bước 3: Chia sẻ tham số

  Bước này bắt đầu khi số lượng nút đã sẵn sàng cho bước chia sẻ tham số đạt đến một ngưỡng tối thiểu cụ thể được chỉ định trong quá trình khởi tạo. Nó bắt đầu bằng quá trình bầu chọn người lãnh đạo epoch, người có nhiệm vụ hợp nhất các tham số được thu được sau quá trình huấn luyện cục bộ trên tất cả các nút. Quá trình này diễn ra rất nhanh chóng và xảy ra vào cuối mỗi epoch. Sử dụng thuật toán bầu chọn lãnh đạo đã xác định trước, một trong các nút sẽ trở thành lãnh đạo và sau đó sử dụng thông tin URI của tất cả các thành viên, tải xuống các tệp tham số từ mỗi nút để thực hiện bước hợp nhất tham số. Mô hình đồng tâm (star topology) được sử dụng, trong đó một lãnh đạo duy nhất thực hiện quá trình hợp nhất; các mô hình khác như hợp nhất k-way (k-way merge) trong đó quá trình hợp nhất được thực hiện bởi một tập hợp các nút cũng có thể được sử dụng và dễ dàng cấu hình.

- Bước 4: Hợp nhất tham số

  Sau đó, người lãnh đạo hợp nhất các tệp tham số đã tải xuống. Framework hỗ trợ nhiều thuật toán hợp nhất như trung bình, trung bình có trọng số, trung vị, và nhiều thuật toán khác. Sử dụng thuật toán hợp nhất đã chọn, người lãnh đạo kết hợp các giá trị tham số từ tất cả các nút để tạo ra một tệp mới chứa các tham số đã được hợp nhất và thông báo cho các nút khác biết rằng có một tệp mới đã có sẵn. Sau đó, mỗi nút tải xuống tệp từ người lãnh đạo và cập nhật mô hình cục bộ của mình với bộ tham số mới.

- Bước 5: Kiểm tra tiêu chí dừng

  Cuối cùng, các nút đánh giá mô hình với các giá trị tham số đã cập nhật bằng cách sử dụng dữ liệu cục bộ của mình để tính toán các chỉ số xác thực khác nhau. Các giá trị thu được từ bước này được chia sẻ bằng cách sử

dụng biến trạng thái của hợp đồng thông minh. Khi mỗi nút hoàn thành bước này, nó thông báo cho mạng rằng quá trình cập nhật và xác thực đã hoàn thành. Trong khi đó, người lãnh đạo tiếp tục kiểm tra tín hiệu hoàn thành cập nhật từ mỗi nút. Khi người lãnh đạo nhân thấy rằng tất cả các thành viên hợp nhất đã thông báo hoàn thành, người lãnh đạo hợp nhất các số liệu xác thực cục bộ để tính toán số liệu toàn cầu. Bước đồng bộ hóa sau đó được đánh dấu là hoàn thành.

3.3. Cách chuyển đổi mô hình từ học máy sang Swarm Learning

Giao diện máy khách Swarm Learning được cung cấp dưới dạng gói wheels (.whl) Python. Nó bao gồm hai API chính là Swarm Callback API và SWCI API.

API SWCI là một API SWCI Python3 để thực hiện các thao tác liên quan đến SWCI trong chương trình.

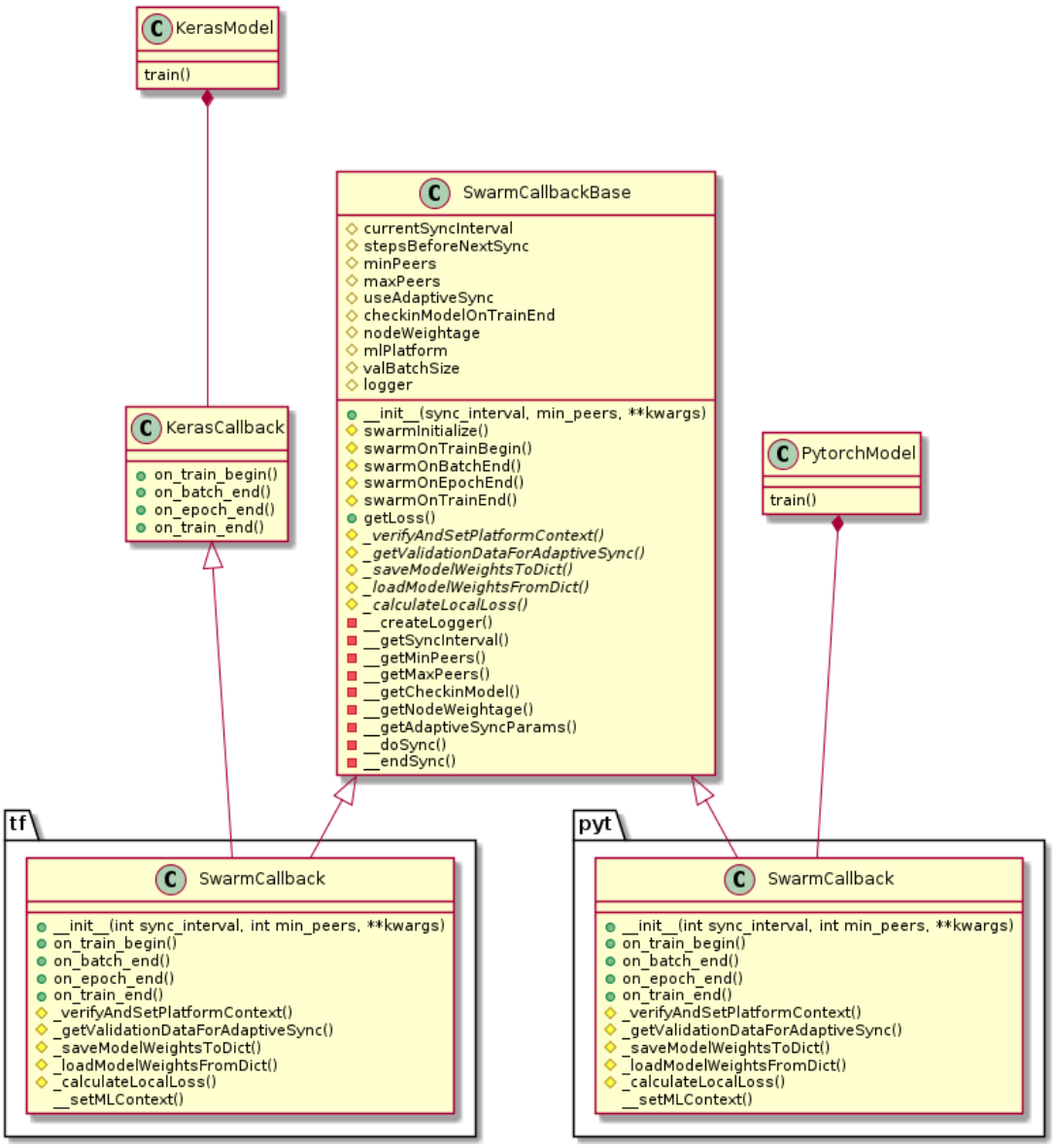
Swarm Callback API là một API Python3 được hỗ trợ cho các nền tảng ML như PyTorch và Keras (dựa trên TensorFlow 2). Swarm Callback là một lớp gọi lại (callback) tùy chỉnh, bao gồm một tập hợp các chức năng có thể được áp dụng trong các giai đoạn khác nhau của quá trình huấn luyện của Machine Learning. Và thư viện Swarm Callback (*Hình 3.3*) được khởi tạo sử dụng cho việc chuyển đổi mô hình từ học máy sang Swarm Learning.

- Lớp SwarmCallbackBase: lớp gọi lại cơ sở cho tất cả các lớp gọi lại đặc thù trong nền tảng mô hình được kế thừa từ đó. Lớp này chứa các biến và chức năng chung cho TensorFlow (TF), PyTorch (PYT) và các nền tảng Machine Learning (ML) khác.

- Lớp tf.SwarmCallback: một triển khai cụ thể cho nền tảng TensorFlow và Keras của Swarm Callback. Đây là điểm vào chính để sử dụng khung Swarm.

Người dùng cần khởi tạo lớp này và gọi các phương thức khác nhau trong quá trình huấn luyện để sử dụng các chức năng của .

- Lớp pyt.SwarmCallback: một triển khai Swarm Callback đặc thù cho nền tảng PyTorch. Đây cũng là điểm vào chính để sử dụng khung Swarm. Người dùng cần khởi tạo lớp này và gọi các phương thức khác nhau trong quá trình huấn luyện để sử dụng các chức năng của Swarm.



Hình 3.3: *Sơ đồ lớp của thư viện Swarm Callback*

CHƯƠNG 4. KẾT QUẢ THỰC NGHIỆM, PHÂN TÍCH VÀ ĐÁNH GIÁ

Ở chương này nhóm em đưa ra tài nguyên môi trường, cách cài đặt và đưa ra các tiêu chí đánh giá về mức độ hiệu quả của mô hình.

4.1. Môi trường thực nghiệm

4.1.1. Tài nguyên

Thực nghiệm được chia làm 2 phần: Khởi tạo khung Swarm Learning và Lật nhãn Poisoning.

Phần "Khởi tạo khung Swarm Learning" để thử nghiệm huấn luyện mô hình cục bộ, với mã nguồn mở do **Hewlett Packard Enterprise (HPE)** phụ trách và đề xuất tại https://github.com/HewlettPackard/swarm-learning. Hệ thống khung mô hình Swarm Learning được thực hiện trên môi trường hệ điều hành Ubuntu 20.4 với cấu hình Ram 32GB, dung lượng bộ nhớ 100GB, thiết lập Docker bản 20.10.5 sử dụng IPv4, có cài đặt thư viện Keras phiên bản 2.9.0 (TensorFlow 2 backend) và PyTorch phiên bản 1.5 dựa trên mô hình máy học sử dụng Python3.

Phần "Lật nhãn Poisoning" cho dữ liệu sử dụng trên công cụ có sẵn của Google là Google Colab.

4.1.2. Tập dữ liệu

Trong phần thực nghiệm này, nhóm em huấn luyện trên ba tập dữ liệu riêng biệt. Trong đó, có hai tập dữ liệu về kịch bản chăm sóc sức khỏe và một tập dữ liệu phát hiện gian lận thẻ tín dụng:

- Bộ dữ liệu COVID-19 Radiography Database: Một cơ sở dữ liệu hình ảnh X-quang ngực cho các trường hợp dương tính với COVID-19 cùng với hình ảnh viêm phổi bình thường và virus. Trong lần phát hành đầu tiên, tác giả đã phát hành 219 hình ảnh chụp X-quang (CXR) phổi COVID-19, 1341 bình thường và 1345 hình ảnh chụp X-quang phổi (CXR) viêm phổi do vi-rút. Trong bản cập nhật thứ 2, tác giả đã tăng cơ sở dữ liệu lên 3616 trường hợp dương tính với COVID-19 cùng với 10.192 bình thường, 6012 độ mờ phổi (nhiễm trùng phổi không phải COVID) và 1345 hình ảnh viêm phổi do vi-rút.

- Bộ dữ liệu ECG Heartbeat Categorization bao gồm tín hiệu nhịp tim được kết hợp từ hai bộ dữ liệu nổi tiếng trong phân loại nhịp tim - Bộ dữ liệu chứng loạn nhịp tim MIT-BIH và Cơ sở dữ liệu ECG chẩn đoán PTB. Tập dữ liệu này đã được sử dụng để phân loại nhịp tim bằng cách sử dụng kiến trúc mạng thần kinh sâu và quan sát một số khả năng học chuyển đổi trên đó. Các tín hiệu tương ứng với hình dạng điện tâm đồ (ECG) của nhịp tim đối với trường hợp bình thường và các trường hợp bị ảnh hưởng bởi rối loạn nhịp tim và nhồi máu cơ tim khác nhau. Các tín hiệu này được xử lý trước và phân đoạn, với mỗi phân đoạn tương ứng với nhịp tim.

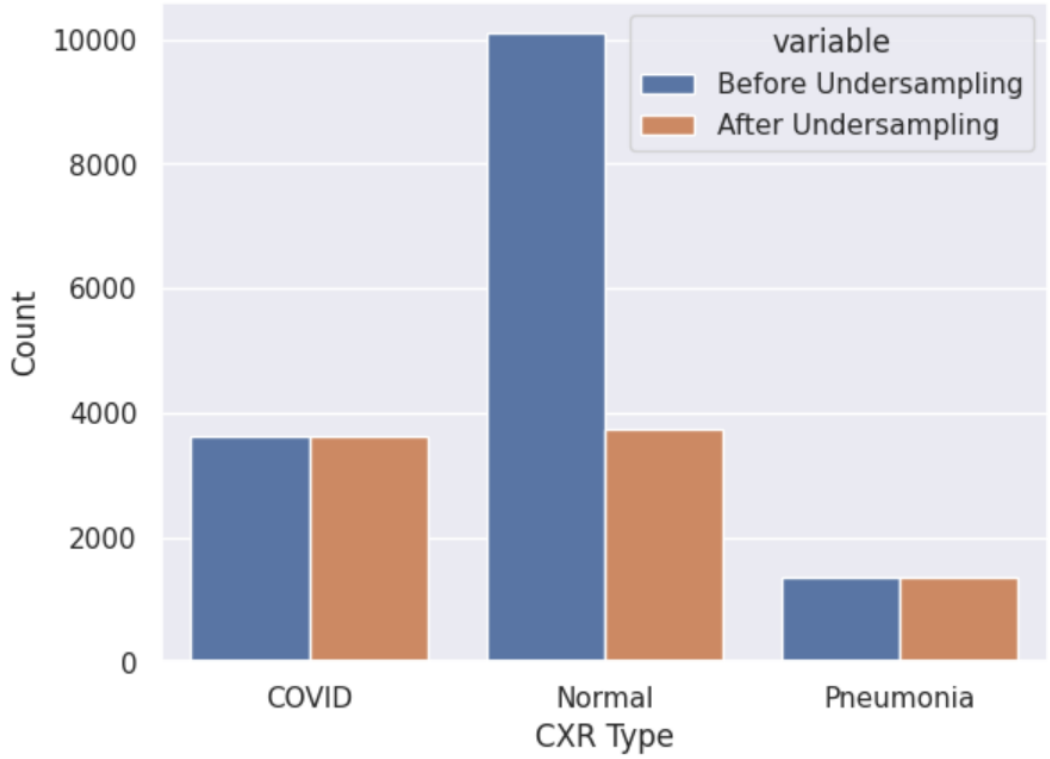
- Tập dữ liệu Credit Card Fraudulent Detection (Phát hiện gian lận thẻ tín dụng):

  Bộ dữ liệu này chứa thông tin về các giao dịch thẻ tín dụng được thực hiện bởi chủ thẻ châu Âu vào tháng 9 năm 2013. Trong hai ngày đó, có tổng cộng 284.807 giao dịch, trong đó có 492 trường hợp gian lận. Tuy nhiên, tỷ lệ gian lận chỉ chiếm 0,172% trong tổng số giao dịch, nghĩa là bộ dữ liệu rất mất cân bằng.

  Dữ liệu chỉ cung cấp các biến số đã được biến đổi bằng phép PCA. Do vấn đề bảo mật, không có thông tin về các tính năng ban đầu và các thông tin cơ bản khác về dữ liệu. Các tính năng V1, V2, ..., V28 là các thành phần

chính được thu được bằng PCA, chỉ có hai tính năng không được chuyển đổi bằng PCA là 'Thời gian' và 'Số lượng'. Tính năng 'Thời gian' cho biết số giây trôi qua giữa mỗi giao dịch và giao dịch đầu tiên trong tập dữ liệu. Tính năng 'Số tiền' là số tiền trong giao dịch, có thể được sử dụng để phân loại dựa trên mức chi phí. Tính năng 'Lớp' là biến phản hồi, có giá trị 1 nếu giao dịch là gian lận và 0 nếu không.

4.1.3. Tiền xử lý dữ liệu



Hình 4.1: *Sự chênh lệch dữ liệu giữa các nhãn trong dữ liệu X-ray COVID19*

- Với bộ dữ liệu về Covid19, bộ dữ liệu khi tải về sẽ gồm có 4 nhãn là : COVID, Lung-Opacity, Viral Pneumonia, Normal. Nhóm em đã sử dụng dữ liệu của 3 nhãn là COVID, Viral Pneumonia, Normal, tuy nhiên dữ liệu sẽ có sự chênh lệch rất lớn giữa các nhãn. Do đó, nhóm em thực hiện đưa dữ liệu từ đa nhãn thành hai nhãn, nhãn COVID, Viral Pneumonia sẽ được

về một nhãn là bất thường (COVID), giữ nguyên nhãn Normal và dùng Undersampling để cắt bớt dữ liệu của nhãn Normal.

Đây là một bộ dữ liệu ảnh vì vậy nhóm em thực hiện chuyển ảnh (bao gồm 2500 ảnh normal và 1345 ảnh Covid) thành dạng số, để đưa về tệp csv. Tổng cộng dữ liệu có 4 tệp csv với 3 tệp cho Client (User) và 1 tệp train với kích thước 64.2MB bằng nhau (mỗi file csv đều chưa 625 nhãn Normal và 336 Covid19).

- Bộ dữ liệu ECG Heartbeat Categorization bao gồm 5 nhãn: 0 (N - Normal beat), 1 (S - Supraventricular premature beat), 2 (V - Premature ventricular contraction), 3 (F - Fusion of ventricular and normal beat), 4 (Q- Unclassifiable beat). Tuy nhiên trong đó nhãn 0 chiếm ưu thế nên nhóm em cũng đã quy từ đa nhãn về hai nhãn là nhịp tim bình thường (0) và nhịp tim bất thường (gom nhãn 1, 2, 3, 4 thành 1). Bộ dữ liệu có tập train và test, nhóm lấy Lấy khoảng 60% tập train, 60% tập test để sử dụng.

  Dữ liệu tập train được chia thành 3 tệp train (68MB) cho từng User (chia dữ liệu theo tỉ lệ 5/5 - tức là tỉ lệ nhãn 0 và 1 bằng nhau) và lấy 60% dữ liệu từ tệp test ban đầu (66.2MB) để làm tập test chung cho các User . Sau đó theo thứ tự, nhóm em thực hiện lật nhãn dữ liệu ngẫu nhiên với User1 lật 10%, User2 lật 20%, User3 lật 30%.

- Bộ dữ liệu nhịp tim thì tập dữ liệu Credit Card Fraudulent Detection đã được tác giả chia sẵn gồm tập test (379KB) và tập train cho từng User (127KB). Nhóm chúng em thực hiện thêm việc lật nhãn ngẫu nhiên trên bộ redit Card Fraudulent Detection với hai trường hợp:
  - TH1: User1 lật 10%, User2 lật 20%, User3 lật 30%
  - TH2: User1 lật 0%, User2 lật 100%, User3 lật 30%

Với ba bộ dữ liệu trên nhóm em sẽ chạy trên SL và FL để so sánh hiệu suất huấn luyện trên hai mô hình. Đối với hai bộ dữ liệu ECG Heartbeat Categorization và Credit Card Fraudulent Detection sẽ tiếp tục được thực hiện thử nghiệm

tấn công poisoning bằng cách lật nhãn dữ liệu. Từ đó đưa ra kết luận về việc tấn công poisoning trên SL và FL.

4.1.4. Phương pháp học máy

Trong mô hình SL và FL nhóm em đã sử dụng chính hai phương pháp học máy với các hàm kích hoạt là Softmax, Sigmoid, Relu:

- Logistic Regression (Hồi quy Logistic): là 1 thuật toán phân loại được dùng để gán các đối tượng cho 1 tập hợp giá trị rời rạc (như 0, 1, 2, ...). Thuật toán trên dùng hàm sigmoid logistic để đưa ra đánh giá theo xác suất.

- CNN (Convolutional Neural Network): là một loại mạng nơ-ron nhân tạo được thiết kế đặc biệt để xử lý dữ liệu không gian như hình ảnh và video. CNN sử dụng các lớp tích chập (convolutional layer) để trích xuất đặc trưng từ dữ liệu đầu vào. Các lớp tích chập giúp nhận diện các đặc trưng cục bộ trong hình ảnh bằng cách áp dụng các bộ lọc trượt qua toàn bộ không gian hình ảnh.

4.1.5. Huấn luyện dữ liệu trên mô hình Federated Learning

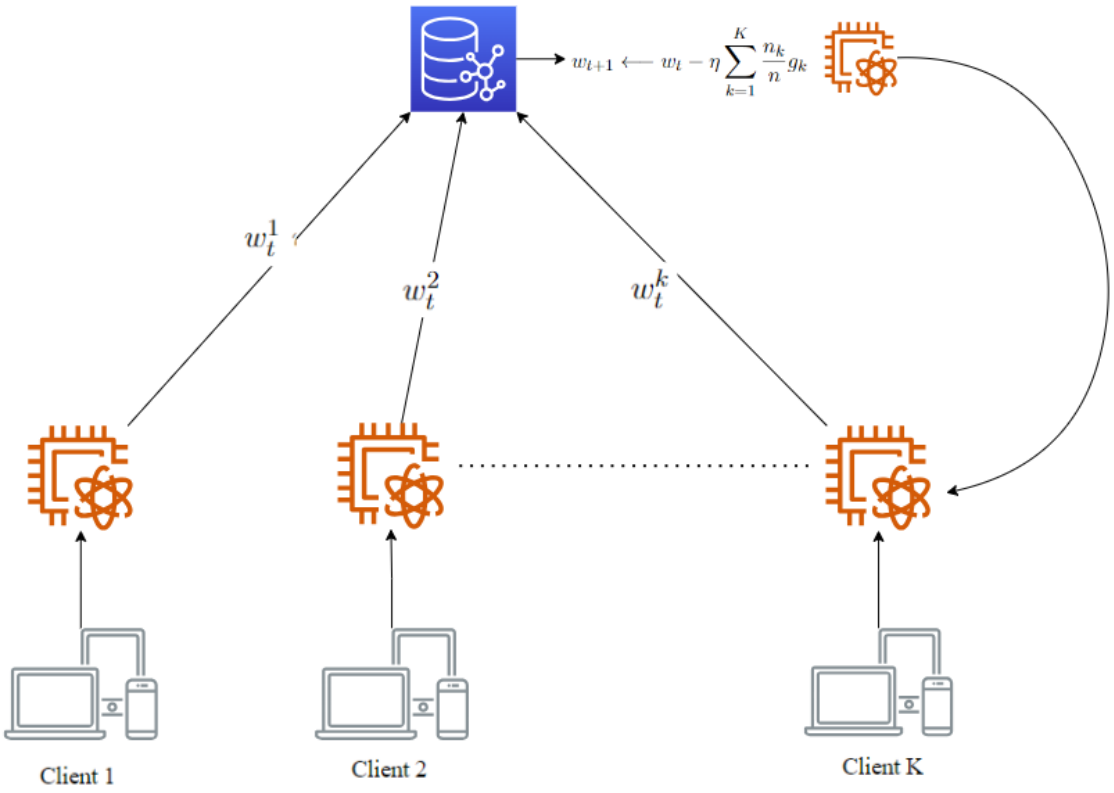
Thực nghiệm của nhóm em dựa theo thuật toán Fedavg và huấn luyện trên 3 client.

Thuật toán **Federated Averaging (FedAvg)** trong Federated Learning[2]

Đầu vào:

- C: Tỷ lệ máy khách thực hiện tính toán trên mỗi vòng lặp
- E: Số lần máy khách thực hiện qua dữ liệu cục bộ trên mỗi vòng lặp
- B: Kích thước minibatch cục bộ được sử dụng cho cập nhật của máy khách
- K, k: Số K máy khách được lập chỉ mục bởi k
- η: Tỷ lệ học tập (learning rate)

Các bước liên quan:



Hình 4.2: *Mô hình Federated Learning*

Bước 1: Khởi tạo mô hình toàn cầu w_t

Bước 2: Đối với mỗi vòng lặp $t = 1, 2, 3, ...$ thực hiện:

Chọn một phần trăm C của máy khách tham gia trong vòng lặp này

Đối với mỗi máy khách k trong tập đã chọn thực hiện:

- Nhận mô hình toàn cầu hiện tại w_t
- Khởi tạo mô hình cục bộ $w_t^k = w_t$
- Đối với mỗi cập nhật cục bộ $u = 1, 2, ..., u_k$ thực hiện:

+ Lấy một minibatch có kích thước B từ bộ dữ liệu cục bộ của máy khách k

+ Tính toán gradient cục bộ

+ Cập nhật mô hình cục bộ

Gửi mô hình cục bộ đã cập nhật w_t^k trở lại máy chủ

Tổng hợp các mô hình từ các máy khách tham gia:

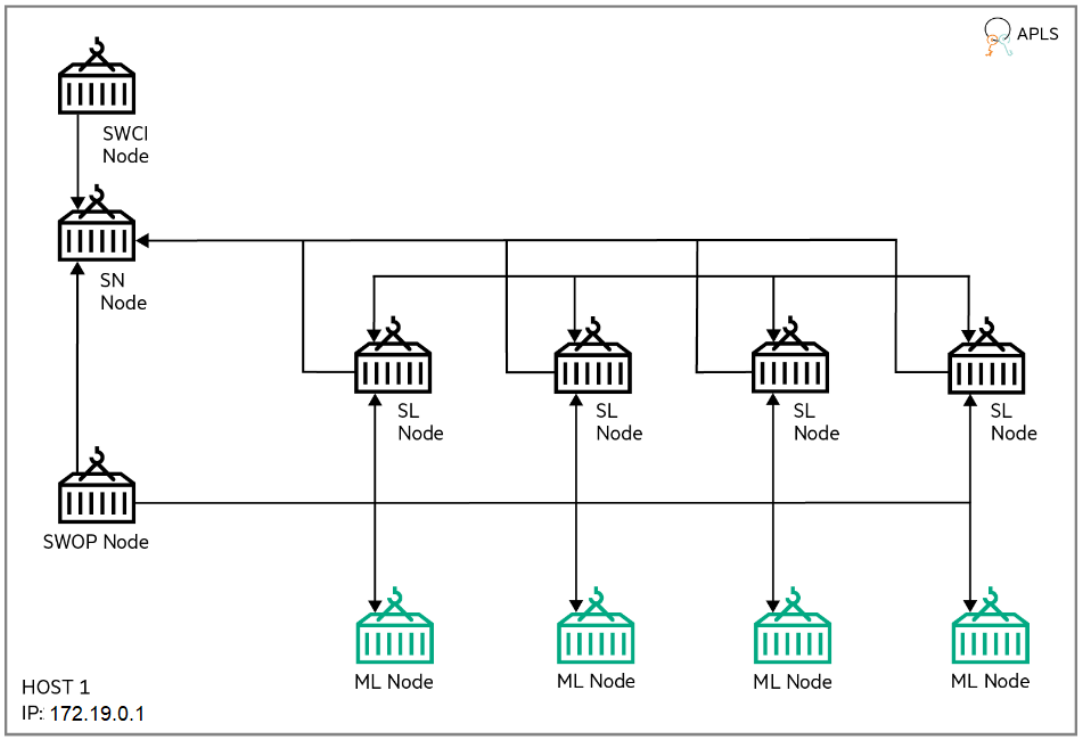
- Tính trung bình có trọng số (FedAvg) của các mô hình cho tất cả các máy khách k tham gia:

$$w_{t+1} \longleftarrow w_t - \eta \sum_{k=1}^{K} \frac{n_k}{n} g_k$$

Bước 3: Trả về mô hình toàn cầu cuối cùng w_t

4.1.6. Huấn luyện dữ liệu trên mô hình Swarm Learning

Mô hình Swarm Learning mà nhóm em sử dụng cho việc thực nghiệm như sau (4.3):



Hình 4.3: *Mô hình Swarm Learning thực nghiệm*

Nhóm em thực hiện chạy với 3 user tương ứng là 3 node, sau quá trình chạy trên khung SL thì sẽ trả về một thư mục trong đó có mô hình huấn luyện tại từng user. Từ đó, nhóm em sự dụng mô hình được tại ra từ quá trình học tập

tập trao đổi trên SL để thực hiện kiểm tra lại với bộ dữ liệu theo từng user để cho ra hiệu suất huấn luyện.

4.1.7. Poisoning trong Swarm Learning và Federated Learning

Có rất nhiều loại tấn công poisoning tới hệ thông FL và SL cũng có thể xảy ra các tấn công poisoning tương tự. Và tấn công Label Flipping Attack thuộc loại tấn công Targeted Attack poisoning là một trong những phương pháp phổ biến. Kẻ xấu sẽ thực hiện lật nhãn dữ liệu tại cái máy địa phương (client) để làm giảm độ tin cậy của mô hình học được.

Để thực nhiệm tấn công lật nhãn dữ liệu, nhóm em đã lật nhãn hai bộ dữ liệu như đã nói ở trên. Sau khi lật nhãn dữ liệu, nhóm em cho học dữ liệu trên khung SL và FL để so sánh kết quả.

4.2. Kết quả thí nghiệm

4.2.1. Khả năng huấn luyện của Swarm Learning và Federated Learning

Về khả năng huấn luyện của Swarm Learning và Federated Learning, nhóm em sử dụng 3 bộ dữ liệu như đã nói ở trên với kết quả được đưa ra như trong hình (4.4, 4.5, 4.6).

		ACCURACY	PRECISION	RECALL	F1_SCORE
Swarm Learning	User 1	0.843	0.854	0.984	0.914
	User 2	0.843	0.854	0.984	0.914
	User 3	0.843	0.854	0.984	0.914
Federated Learning	Client 1	0.699	x	x	x
	Client 2	0.699	x	x	x
	Client 3	0.779	x	x	x
	Global	0.866	1.0	0.866	0.928

Hình 4.4: *Hiệu suất huấn luyện trên trên SL và FL của bộ dữ liệu COVID19 X-ray Chest.*

		ACCURACY	PRECISION	RECALL	F1_SCORE
Swarm Learning	User 1	0.964	0.867	0.935	0.901
	User 2	0.964	0.867	0.935	0.901
	User 3	0.964	0.867	0.935	0.901
Federated Learning	Client 1	0.982	x	x	x
	Client 2	0.980	x	x	x
	Client 3	0.983	x	x	x
	Global	0.982	0.911	0.985	0.947

Hình 4.5: *Hiệu suất huấn luyện trên trên SL và FL của bộ dữ liệu EGG Heartbeat.*

		ACCURACY	PRECISION	RECALL	F1_SCORE
Swarm Learning	User 1	0.958	0.967	0.947	0.957
	User 2	0.958	0.967	0.947	0.957
	User 3	0.958	0.967	0.947	0.957
Federated Learning	Client 1	0.711	x	x	x
	Client 2	0.991	x	x	x
	Client 3	0.991	x	x	x
	Global	0.891	0.812	0.996	0.879

Hình 4.6: *Hiệu suất huấn luyện trên trên SL và FL của bộ dữ liệu Credit Card Fraudulent Detection.*

Nếu xét theo như kết quả được đưa ra thì huấn luyện dựa trên SL với FL có vẻ xấp xỉ nhau và SL có đôi chút nhỉnh hơn. Tuy nhiên, kết quả ở đây đang còn mang tính khách quan, có sự chênh lệch với thực tế bởi SL sẽ có sự mất mát trong qua trình trao đổi trọng số và nhiều vấn đề khác.

4.2.2. Khả năng bị tấn công Poisoning trên Swarm Learning và Federated Learning

Để so sánh khả năng bị tấn công poisoning, nhóm em thực nghiệm với 2 bộ dữ liệu Credit Card Fraudulent Detection và EGG Heartbeat. Trong đó bộ dữ liệu Credit Card Fraudulent Detection sẽ có hai trường hợp lật nhãn với tỉ lệ khác nhau như đã trình bày ở trên. Trước tiên ta sẽ so sánh về hiệu suất của FL và SL khi bị tấn công poisoning trên từng bộ dữ liệu. Hình (4.7, 4.8, 4.9)

		ACCURACY	PRECISION	RECALL	F1_SCORE
Swarm Learning	User 1-10%	0.958	0.851	0.919	0.884
	User 2-20%	0.958	0.851	0.919	0.884
	User 3-30%	0.958	0.851	0.919	0.884
Federated Learning	Client 1-10%	0.982	x	x	x
	Client 2-20%	0.980	x	x	x
	Client 3-30%	0.983	x	x	x
	Global	0.982	0.911	0.985	0.947

Hình 4.7: *Hiệu suất học tập trên SL và FL với bộ dữ liệu ECG Heatbeat khi bị tấn công Poisoning lật nhãn.*

		ACCURACY	PRECISION	RECALL	F1_SCORE
Swarm Learning (TH1)	User 1-10%	0.837	0.827	0.849	0.838
	User 2-20%	0.837	0.827	0.849	0.838
	User 3-30%	0.837	0.827	0.849	0.838
Federated Learning (TH1)	Client 1-10%	0.767	x	x	x
	Client 2-20%	0.767	x	x	x
	Client 3-30%	0.758	x	x	x
	Global	0.732	0.608	0.809	0.694

Hình 4.8: *Hiệu suất học tập trên SL và FL với bộ dữ liệu Credit Card Fraudulent Detectio khi bị tấn công Poisoning lật nhãn TH1.*

		ACCURACY	PRECISION	RECALL	F1_SCORE
Swarm Learning (TH2)	User 1-0%	0.688	0.715	0.685	0.700
	User 2-100%	0.688	0.715	0.685	0.700
	User 3-30%	0.688	0.715	0.685	0.700
Federated Learning (TH2)	Client 1-0%	0.938	x	x	x
	Client 2-100%	0.507	x	x	x
	Client 3-30%	0.627	x	x	x
	Global	0.638	0.672	0.630	0.650

Hình 4.9: *Hiệu suất học tập trên SL và FL với bộ dữ liệu Credit Card Fraudulent Detectio khi bị tấn công Poisoning lật nhãn TH2.*

Theo đó ta có thể thấy là quá trình tấn công poisoning bằng cách lật nhãn trên SL và FL cùng một bộ dữ liệu và tỉ lệ lật tại từng client (user) là như nhau nhưng SL vẫn cho ra kết quả huấn luyện ổn định hơn FL.

Và để có thể hình dung rõ hơn về khả năng bị tấn công poisoning thì nhóm em thực hiện so sánh giữa SL trước và sau khi lật nhãn, FL trước và sau khi lật nhãn của từng bộ dữ liệu. (Hình (4.10, 4.11, 4.12, 4.13)

- Đối với bộ dữ liệu EGG Heartbeat:

	ACCURACY	PRECISION	RECALL	F1_SCORE
Federated Learning	0.982	0.911	0.985	0.947
Federated Learning(flip)	0.971	0.847	0.988	0.912

Hình 4.10: *Bảng so sánh kết quả huấn luấn dữ liệu trên FL trước và sau khi lật nhãn của bộ dữ liệu ECG Heatbeat.*

		ACCURACY	PRECISION	RECALL	F1_SCORE
Swarm Learning	User 1	0.964	0.867	0.935	0.901
	User 2	0.964	0.867	0.935	0.901
	User 3	0.964	0.867	0.935	0.901
Swarm Learning (flip)	User 1-10%	0.958	0.851	0.919	0.884
	User 2-30%	0.958	0.851	0.919	0.884
	User 3-30%	0.958	0.851	0.919	0.884

Hình 4.11: *Bảng so sánh hiệu suất huấn luyện trên SL giữa trước và sau khi lật nhãn của bộ dữ liệu EGG Heartbeat.*

- Đối với bộ dữ liệu Credit Card Fraudulent Detection:

	ACCURACY	PRECISION	RECALL	F1_SCORE
Federated Learning	0.891	0.812	0.996	0.879
Federated Learning (TH1)	0.732	0.608	0.809	0.694
Federated Learning (TH2)	0.638	0.672	0.630	0.650

Hình 4.12: *So sánh hiệu suất huấn luyện trên FL giữa trước và sau khi lật nhãn của bộ dữ liệu Credit Card Fraudulent Detectio.*

		ACCURACY	PRECISION	RECALL	F1_SCORE
Swarm Learning	User 1	0.958	0.967	0.947	0.957
	User 2	0.958	0.967	0.947	0.957
	User 3	0.958	0.967	0.947	0.957
Swarm Learning (TH1)	User 1-10%	0.837	0.827	0.849	0.838
	User 2-20%	0.837	0.827	0.849	0.838
	User 3-30%	0.837	0.827	0.849	0.838
Swarm Learning (TH2)	User 1-0%	0.688	0.715	0.685	0.700
	User 2-100%	0.688	0.715	0.685	0.700
	User 3-30%	0.688	0.715	0.685	0.700

Hình 4.13: *So sánh hiệu suất huấn luyện trên SL giữa trước và sau khi lật nhãn của bộ dữ liệu Credit Card Fraudulent Detectio.*

Từ những kết quả bên trên nhóm em nhận thấy rằng: Hiệu suất học tập tại các User của SL là như nhau, tương tự như FL thì SL vẫn bị tấn công poisoning. Tuy nhiên, tỉ lệ thiệt hại khi tấn công Flip Label trên SL thấp hơn FL. Theo đó, một số lý do mà nhóm chúng em đưa ra khi khả năng chống chọi lại Poisoning của SL tốt hơn FL là:

- Trong Swarm Learning, các mô hình được lựa chọn ngẫu nhiên dựa trên hiệu suất và đóng góp của chúng cho tổng thể hợp. Các mô hình liên tục cung cấp dự đoán sai hoặc có hành vi bất thường có thể được phát hiện và loại bỏ khỏi hợp tác (có thể dựa theo cơ chế đồng thuận loại bỏ tác động của phân từ độc hại)

- Cơ chế P2P (Peer-to-Peer) của Swarm learning, các bên tham gia giao tiếp với nhau và trao đổi cập nhật mô hình để cùng nhau cải thiện mô hình cá nhân và tổng thể. Quá trình học tập và xác thực đồng nghiệp này cho phép thành viên so sánh mô hình của mình với những người khác và phát hiện sự không nhất quán hoặc hành vi độc hại từ đó cuộc tấn công lật nhãn dữ liệu của cơ chế poisoning có thể được phát hiện và giảm thiểu hiệu quả hơn

so với federated learning - các thành viên thường không tương tác trực tiếp với nhau.

CHƯƠNG 5. KẾT LUẬN

Ở chương này, nhóm em đưa ra những kết luận về đồ án chuyên ngành, những hạn chế, và đồng thời đưa ra hướng cải thiện và phát triển.

5.1. Kết luận

5.1.1. Kết quả đạt được

Với những vấn đề về SL đã được đề cập ở phần trước, nhóm đã thực hiện tìm hiểu, đưa ra phân tích, khả năng của SL so với FL khi đứng trước cuộc tấn công poisoning. Sau đây là tóm tắt kết quả đã đạt được:

- Triển khai SL
- Xử lý dữ liệu cho đầu vào SL và FL, lật nhãn dữ liệu
- Thực hiện huấn luyện dữ liệu trên SL và FL để so sánh hiệu suất huấn luyện.
- Thử nghiệm tấn công poisoning lật nhãn trên SL và FL, đưa ra đánh giá

5.1.2. Định hướng phát triển

Bên cạnh những thành quả đặt được thì vẫn còn tồn tại những vấn đề chưa thể hoàn thành vì sự hạn chế về thời gian, tài nguyên thực nghiệm.

- Dữ liệu chưa thật sự được xử lý tốt
- Môi trường SL chạy còn gặp nhiều lỗi: chưa đưa về được giá trị accuracy mà chỉ trả về mô hình học tập từng user

Trong thời gian tiếp theo nhóm em có thể sẽ xây dựng lại và phát triển thêm các tính năng, khắc phục các vấn đề còn tồn đọng:

- Nhóm em có thể sẽ tạo một khung SL mới dựa trên khung SL hiện tại
- Đầu tư hơn vào dữ liệu huấn luyện
- Thử nghiệm thêm một số dạng tấn công poisoning

5.1.3. Kết luận

Sau quá trình nghiên cứu và triển khai hệ thống, nhóm em học hỏi được nhiều điều về điểm mạnh điểm yếu của Swarm Learning, Federated Learning và biết thêm về các cuộc tấn công poisoning. Thông qua đó tìm cách thực nghiệm, so sánh với mong muốn hiểu được khả năng của chúng. Federated Learning hay Blockchain có thể đã quá phổ biến với chúng ta với nhiều ứng dụng liên quan tới các lĩnh vực khác nhau. Tuy nhiên, Swarm Learning lại là một công nghệ có thể xem là mới mẻ, dường như là sự kết hợp giữa FL và BC để đem tới những tính năng ưu việt, bảo mật tốt hơn. Tổng kết lại, với công nghệ mới như Swarm Learning nhóm em muốn mang đến một cái nhìn rõ hơn về khả năng của SL. Qua đó, nhóm em cũng mong rằng SL có thể phát triển và được ứng dụng rộng rãi hơn bởi những ưu điểm mà nó mang tới.

TÀI LIỆU THAM KHẢO

Tiếng Anh:

[1] Jialiang Han, Yun Ma, and Yudong Han (2022), "Demystifying swarm learning: A new paradigm of blockchain-based decentralized federated learning", *arXiv preprint arXiv:2201.05286*.

[2] Brendan McMahan et al. (2017), "Communication-efficient learning of deep networks from decentralized data", pp. 1273–1282.

[3] Oliver Lester Saldanha et al. (2022), "Swarm learning for decentralized artificial intelligence in cancer histopathology", *Nature Medicine*, 28 (6), pp. 1232–1239.

[4] Stefanie Warnat-Herresthal et al. (2021), "Swarm learning for decentralized and confidential clinical machine learning", *Nature*, 594 (7862), pp. 265–270.