Mẫu F6

ĐẠI HỌC QUỐC GIA TP. HCM
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

BÁO CÁO TỔNG KẾT
ĐỀ TÀI KHOA HỌC VÀ CÔNG NGHỆ SINH VIÊN NĂM 2022

Tên đề tài tiếng Việt:

Nghiên cứu mô hình dự đoán nguy cơ bệnh tim sử dụng phương pháp Học liên kết với Mã hóa đồng cấu và Blockchain trong ngữ cảnh y tế thông minh.

Tên đề tài tiếng Anh:

A study on Heart Disease prediction models using Federated Learning with Homomorphic Encryption and Blockchain in the smart healthcare context.

Khoa/ Bộ môn: Mạng máy tính và truyền thông

Thời gian thực hiện: 06 tháng

Cán bộ hướng dẫn: Ths.Phan Thế Duy

Tham gia thực hiện

TT	Họ và tên, MSSV	Chịu trách nhiệm	Điện thoại	Email
1.	Bùi Tấn Hải Đăng, 20520173	Chủ nhiệm	0889337917	20520173@gm.uit.edu.vn
2.	Phan Hữu Luân, 20521585	Tham gia	0931328505	20521585@gm.uit.edu.vn
3.	Vương Đinh Thanh Ngân, 20521649	Tham gia	0399832522	20521649@gm.uit.edu.vn

Thành phố Hồ Chí Minh – Tháng ..... /20..

Mẫu F6

ĐẠI HỌC QUỐC GIA TP. HCM
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

Ngày nhận hồ sơ	
Mã số đề tài	
(Do CQ quản lý ghi)	

BÁO CÁO TỔNG KẾT

Tên đề tài tiếng Việt:

Nghiên cứu mô hình dự đoán nguy cơ bệnh tim mạch quỵ sử dụng phương pháp Học liên kết với Mã hóa đồng cấu và Blockchain trong ngữ cảnh y tế thông minh.

Tên đề tài tiếng Anh:

A study on Heart Disease prediction models using Federated Learning with Homomorphic Encryption and Blockchain in the smart healthcare context.

Ngày ... tháng ..... năm ....
Cán bộ hướng dẫn
(Họ tên và chữ ký)

Ngày ... tháng ..... năm ....
Sinh viên chủ nhiệm đề tài
(Họ tên và chữ ký)

Bùi Tấn Hải Đăng

Mẫu F6

THÔNG TIN KẾT QUẢ NGHIÊN CỨU

1. Thông tin chung:

- Tên đề tài: Nghiên cứu mô hình dự đoán nguy cơ mắc bệnh tim mạch sử dụng phương pháp Học cộng tác với Mã hóa đồng cấu và Blockchain trong ngữ cảnh y tế thông minh.

- Chủ nhiệm: Bùi Tấn Hải Đăng - 20520173

- Thành viên tham gia:

- Phan Hữu Luân - 20521585,
- Vương Đinh Thanh Ngân - 20521649.

- Cơ quan chủ trì: Trường Đại học Công nghệ Thông tin.

- Thời gian thực hiện: 06 tháng

2. Mục tiêu:

- Nghiên cứu, thiết kế và xây dựng hệ thống kết hợp phương pháp Học cộng tác (Federated Learning – FL) và Blockchain có khả năng dự đoán các chẩn đoán nguy cơ đột quỵ tốt, đồng thời đảm bảo được quyền riêng tư dữ liệu cũng như có cơ chế khuyến khích chủ sở hữu (người dùng cuối, bệnh viện, các phòng khám, viện nghiên cứu) tham gia đóng góp dữ liệu trong ngữ cảnh y tế thông minh.

- Nghiên cứu, thiết kế, và xây dựng dưới dạng prototype hệ thống các quy tắc tương tác giữa các thành viên để mô hình có thể tự động vận hành một cách chính xác và minh bạch dựa trên blockchain.

- Cải thiện chi phí truyền tải dữ liệu sau khi huấn luyện cục bộ nhờ tích hợp điện toán biên (Mobile Edge Computing - MEC) vào mô hình FL.

- Bảo vệ hệ thống khỏi các cuộc tấn công suy ngược dữ liệu nhờ hệ thống FL kết hợp thuật toán mã hóa đồng cấu (Homomorphic Encryption – HE). Cụ thể chúng tôi sử dụng cấu trúc CKKS, một lược đồ tương thích với số dấu phẩy động, có lợi cho việc đóng gói và thay đổi tỉ lệ bản mã.

3. Tính mới và sáng tạo:

- Mô hình chẩn đoán nguy cơ đột quỵ sử dụng kết hợp FL, HE và Blockchain trong ngữ cảnh y tế thông minh.

- Tích hợp điện toán biên MEC vào mô hình để tối ưu hiệu năng cũng như chi phí liên lạc và chi phí truyền tải dữ liệu.

- Sử dụng Blockchain có cơ chế trả thưởng để các thành viên có thể tham gia hợp tác ngay cả khi họ không tin tưởng lẫn nhau. Nhờ đó có thể khuyến khích các thành viên trong hệ thống tích cực tham gia cộng tác và đóng góp dữ liệu.

- Mã hóa dữ liệu bằng thuật toán HE giúp tăng cường tính riêng tư và toàn vẹn của dữ liệu trong quá trình truyền tải dữ liệu lên MEC Server hoặc Cloud Server.

4. Tóm tắt kết quả nghiên cứu:

4.1 Giới thiệu bài toán

Ngày này, sự phát triển của các thiết bị thông minh đã hỗ trợ con người có thể tự giảm sát sức khoẻ bản thân. Và việc áp dụng các thiết bị công nghệ tiên tiến đối với lĩnh vực y tế cũng đem đến nhiều giải pháp giúp giảm các sức ép lên các bệnh viện, cơ sở y tế, tiết kiệm thời gian và tiền bạc. Các thiết bị chăm sóc khoẻ có thể chẩn đoán được nhiều vấn đề của cơ thể bao gôm nguy cơ đột quỵ. Theo thời gian, dữ liệu sẽ nhiều lên khi số lượng người dùng tăng từ đó chất lương mô hình và tỉ lệ chẩn đoán đúng tăng cao. Tuy nhiên, trên thực tế thì dữ liệu dùng để học tập cho mô hình đều thuộc pham trù thông tin riêng tư của từng người dùng dẫn tới người dùng sẽ ngại chia sẻ các thông tin cá nhân, nhạy cảm cho bên thứ ba. Đồng thời, vấn đề rò rỉ dữ liệu trong quá trình trao đổi, học tập dữ liệu với bên thứ ba cũng là thách thức lớn tới sự phát triển của các thiết bị chăm sóc sức khoẻ thông minh.

Phương pháp học cộng tác có thể giải quyết cho phép đào tạo với dữ liệu phi tập trung được lưu trữ riêng trên nhiều máy. Điều này rất quan trọng vì trong thực tế, dữ liệu có thể quá lớn để gửi đến máy chủ trung tâm hoặc quá nhạy cảm để chia sẻ. Tuy nhiên, FL vẫn còn một số nhược điểm trong việc kiểm soát dữ liệu, có khả năng bị rò rỉ bởi bên thứ ba. Để giải quyết vấn đề đó, trong nghiên cứu này, chúng tôi đề xuất một mô hình dự đoán nguy cơ mắc bệnh tim sử dụng FL tích hợp HE, Blockchain (tên của mô hình). Trong đó, việc áp dụng HE giúp dữ liệu nhạy cảm được bảo vệ tốt hơn, theo đó HE sẽ mã hoá dữ liệu trọng số gửi lên và trả về trong mô hình học cộng tác giúp tăng tính an toàn. Bên cạnh đó, MEC cũng được kết hợp vào mô hình như một giải pháp để tối ưu hóa thời gian giao tiếp giữa các phần tử trong hệ thống, làm giảm tắc nghẽn. Liên quan tới một vấn đề nữa là lưu trữ các bản ghi, truy xuất nguồn gốc, loại bỏ các phần tử độc và làm sao để khuyến khích sự đóng góp của dữ liệu được đào tạo tốt tại cục bộ thì trong mô hình đã tích hợp thêm Blockchain nhằm bảo vệ tính minh bạch.

4.2 Các nghiên cứu liên quan

4.2.1 Học cộng tác

Học cộng tác được đề đưa ra bởi Google vào năm 2016 và dần dần thu hút được rất nhiều sự quan tâm trong nghiên cứu và ứng dụng. Học cộng tác đã và đang được sử dụng rộng rãi trong nhiều lĩnh vực: bàn phím Gboard thiết kế bởi Google sử dụng học cộng tác để cải thiện khả năng gợi ý từ mà vẫn bảo vệ được tính riêng tư của người dùng, trong y học các dữ liệu của bệnh nhân rất nhạy cảm nên học cộng tác cũng rất hữu dụng, xử lý ngôn ngữ tự nhiên và hệ thống gợi ý cũng áp dụng học cộng tác.

4.2.2 Giải pháp khuyến khích tham gia đóng góp dữ liệu và giám sát các hành vi bất thường trong học cộng tác

Federated Learning cho phép đào tạo với dữ liệu phi tập trung được lưu trữ riêng trên nhiều máy. Điều này rất quan trọng vì trong thực tế, dữ liệu có thể quá lớn để gửi đến máy chủ trung tâm hoặc quá nhạy cảm để chia sẻ. Tuy nhiên, để FL phát huy hết tiềm năng của mình, có ba vấn đề như sau: *thứ nhất là* làm thế nào để tối đa hóa độ chính xác của việc học dựa trên sự không đồng nhất dữ liệu được đào tạo cục bộ, *thứ hai là* làm sao để khuyến khích sự đóng góp của dữ liệu được đào tạo tốt tại cục bộ và *cuối cùng là* làm sao để giảm thiểu sự phụ thuộc của FL trên bất kỳ máy chủ cục bộ nào đó có thể dễ xảy ra gian lận hay tắc nghẽn làm giảm hiệu suất của mô hình trung tâm. Nhằm góp phần giải quyết vấn đề trên, nhóm chúng tôi đề xuất một mô hình học cộng tác bằng cách tận dụng **Blockchain** để tăng cường tính bảo mật trong quá trình lưu trữ và tạo nên hệ thống khuyến khích tham gia đóng góp dữ liệu từ chủ sở hữu đồng thời giám sát các hành vi bất thường của thành viên trong mạng.

4.2.3 Giải pháp đảm bảo riêng tư trong học cộng tác

Bên cạnh những lợi ích mà phương pháp học cộng tác mang lại, phương pháp đào tạo mô hình học máy phi tập trung này cũng có nhiều điểm yếu về tính bảo mật, an toàn. Không ít nghiên cứu đã chỉ ra những hạn chế của mô hình này trong quá trình triển khai vào thực tế. Những nhược điểm đó được hai tác giả N. Bouacida và P. Mohapatra trình bày cụ thể ở công trình nghiên cứu của họ [1]. Trên thực tế, đã có những nghiên cứu chỉ ra rằng, từ các trọng số của mô hình sau khi huấn luyện vẫn có thể truy xuất ra thông tin riêng tư từ các mô hình đó. Để tăng cường tính bảo mật dữ liệu trong quá trình lưu trữ và truyền tải dữ liệu, Stripelis và các cộng sự, đề cập đến **Mã hóa đồng cấu (Homomorphic Encryption)** kết hợp với học cộng tác (Federated Learning) ở công trình nghiên cứu của mình [2], đây là một hệ thống mà nhóm chúng tôi quan tâm. Cụ thể, nhóm tác giả đã giới thiệu hệ thống với khả năng đảm bảo toàn vẹn dữ liệu riêng tư của người dùng, khi quá trình tính toán được thực hiện trên dữ liệu được **mã hóa**. Ta có thể hiểu đơn giản, bộ trọng số được mã hoá hoàn toàn trong quá trình truyền tải và tính toán.

4.2.4 Giải pháp giảm thiểu chi phí và thời gian truyền liên lạc

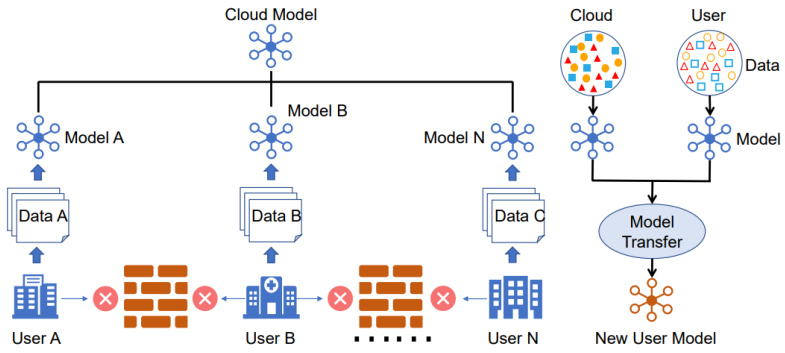
Bên cạnh những tiềm năng mà Homomorphic Encryption mang lại thì một trong những hạn chế của thuật toán này là chi phí truyền tải dữ liệu cao, cụ thể đã được tác giả Kogos và cộng sự đã chỉ ra trong công trình nghiên cứu [3]. Hơn nữa, Rui Wang cũng đã đề cập đến vấn đề về chi phí về thời gian truyền tải dữ liệu khi triển khai FL vào thực tế ở nghiên cứu của mình [4]. Vì vậy tác giả đã đề xuất giải pháp sử dụng **Mobile Edge Computing (MEC)** để cải thiện chi phí truyền tải dữ liệu và mang lại khả năng mạnh mẽ trong việc lưu trữ và xử lý dữ liệu thời gian thực ở mạng biên. MEC có ý nghĩa quan trọng trong việc giảm đáng kể lượng dữ liệu phải di chuyển lên mô hình trung tâm. Điều này sẽ làm giảm chi phí truyền tải, giảm thời gian trễ và nâng cao được chất lượng dịch vụ hệ thống.

Để hiện thực hóa kiến trúc FLchain tích hợp điện toán biên MEC, vào năm 2021, tác giả Dinh C. Nguyen cùng cộng sự trong bài nghiên cứu [5] đã phân tích những cơ hội cũng như thử thách trong quá trình triển khai nhằm đưa ra giải pháp khắc phục khả thi. Với mô hình được tác giả đề xuất, hệ thống có khả năng tận dụng tối đa nguồn tài nguyên bằng cách xây dựng cơ chế trao thưởng cho bên đóng góp dữ liệu cũng như tăng cường giám sát hành vi bất thường của các thành viên trong hệ thống. Hơn nữa, việc tích hợp điện toán biên MEC trong mô hình giúp cải thiện chi phí truyền tải dữ liệu khi vận hành và tăng chất lượng dịch vụ hệ thống.

4.2.5 Hệ thống phát hiện bệnh dựa trên học cộng tác

Sự ra đời của tiêu chuẩn General Data Protection Regulation (GDPR) vào tháng 5 2018, giúp quyền bảo mật dữ liệu của bệnh nhân Châu Âu đã được tăng lên, tuy nhiên lại dẫn đến việc nghiên cứu trong lĩnh vực y tế ngày càng khó khăn. Năm 2021, Pfitzner cùng cộng sự [6] đã nghiên cứu và cho rằng mô hình học cộng tác là một phương pháp học máy đầy hứa hẹn trong tương lai nhờ đặc tính lưu trữ phi tập trung và không vi phạm quyền riêng tư dữ liệu của người dùng.

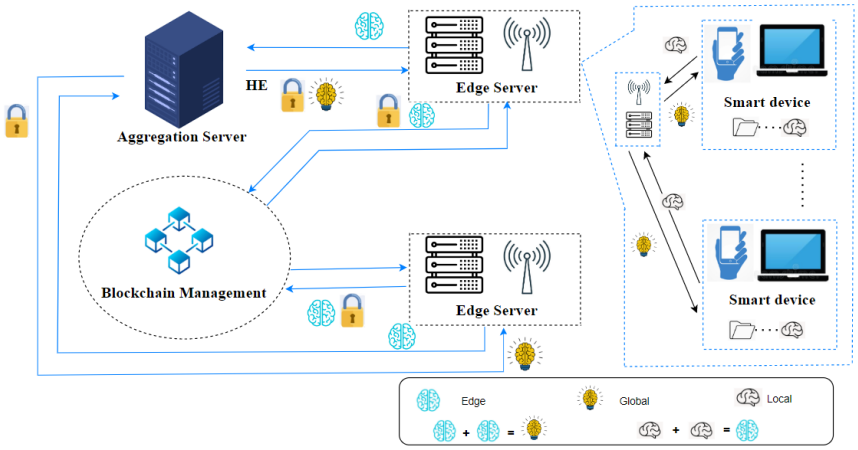
Nhận thấy tiềm năng của phương pháp học cộng tác, các nhà khoa học đã nghiên cứu ứng dụng Federated Learning vào nhiều khía cạnh trong đời sống như xe tự hành, các thiết bị IoT và đặc biệt là trong ngữ cảnh chăm sóc sức khỏe thông minh. Trong công trình nghiên cứu của Yiqiang Chen và cộng sự [7], nhóm tác giả đã phân tích về tiềm năng cũng như thách thức trong tương lai của FedHealth ở bối cảnh thu nhập dữ liệu qua thiết bị IoT, cụ thể là thiết bị đeo tay. Tác giả cũng trình bày khái quát một bộ khung cơ bản của FedHealth trong ngữ cảnh y tế thông minh, trong đó 'User' không chỉ đơn thuần là các người dùng mà còn là các cơ quan như bệnh viện, phòng khám tư nhân hay thậm chí là các thiết bị IOT tham gia vào quá trình đào tạo.



Hình 1: Bộ khung cơ bản của FedHeath trong ngữ cảnh y tế thông minh

4.3 Phương pháp nghiên cứu

4.3.1. Mô hình học cộng tác tích hợp Blockchain và HE có hỗ trợ điện toán biên



Hình 2: Mô hình học cộng tác tích hợp Blockchain và HE có hỗ trợ điện toán biên

Hình 2 mô tả cấu trúc của mô hình học cộng tác tích hợp Blockchain và HE có hỗ trợ điện toán biên trong bối cảnh chăm sóc sức khỏe thông minh, huấn luyện cộng tác trên bộ dữ liệu được tạo ra thông qua việc sử dụng các thiết bị theo dõi sức khỏe thông minh. Mô hình thông qua Mobile Edge Computing (MEC) đã được thêm vào FL để giảm sức ép lên máy chủ và một hệ thống blockchain ghi lại toàn bộ hoạt động của hệ thống. Dữ liệu trong quá trình học tập luôn được đảm bảo độ an toàn cao thông qua HE. Mô hình sẽ hỗ trợ phát hiện nguy cơ đột quỵ ở người dùng. Chi tiết về mô hình được trình bày trong các phần tiếp theo bên dưới.

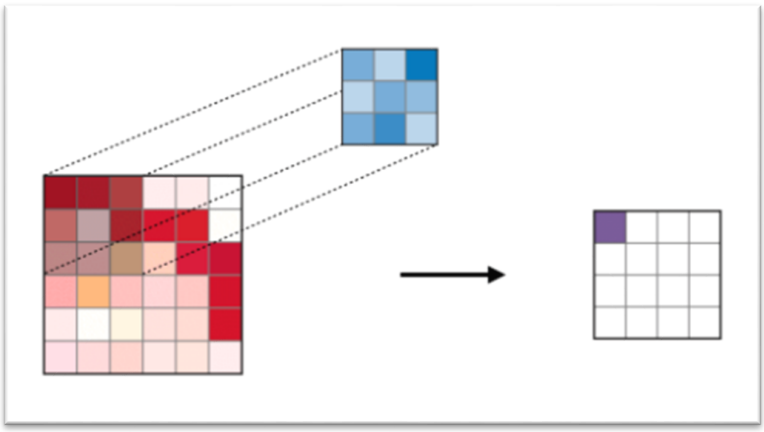
4.3.2. Mô hình học sâu được đề xuất cho nghiên cứu dự đoán bệnh tim:

Convolutional neural network – CNN:

- Mạng nơ-ron tích chập (CNN) là một trong những mô hình Deep Learning tiên tiến. Nó giúp cho chúng ta xây dựng được những hệ thống thông minh với độ chính xác cao như hiện nay. CNN được sử dụng nhiều trong các bài toán nhận dạng các object trong ảnh.

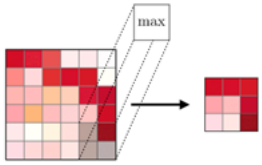
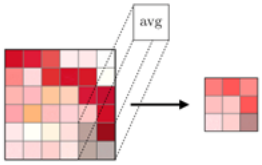
- Các kiểu tầng trong kiến trúc của một mạng CNN:

+ **Tầng tích chập (CONV)** sử dụng các bộ lọc để thực hiện phép tích chập khi đưa chúng đi qua đầu vào I theo các chiều của nó. Các siêu tham số của các bộ lọc này bao gồm kích thước bộ lọc F và độ trượt (stride) S. Kết quả đầu ra O được gọi là feature map hay activation map.



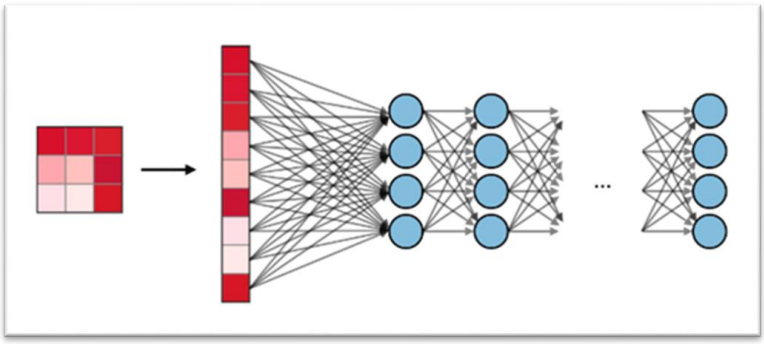
Hình 3: Mô phỏng tầng tích chập với kích thước bộ lọc F và độ trượt S

+ Pooling (POOL) là một phép downsampling, thường được sử dụng sau tầng tích chập, giúp tăng tính bất biến không gian. Cụ thể, max pooling và average pooling là những dạng pooling đặc biệt, mà tương ứng là trong đó giá trị lớn nhất và giá trị trung bình được lấy ra.

Kiểu	Max pooling	Average pooling
Chức năng	Từng phép pooling chọn giá trị lớn nhất trong khu vực mà nó đang được áp dụng	Từng phép pooling tính trung bình các giá trị trong khu vực mà nó đang được áp dụng
Minh họa		
Nhận xét	• Bảo toàn các đặc trưng đã phát hiện • Được sử dụng thường xuyên	• Giảm kích thước feature map • Được sử dụng trong mạng LeNet

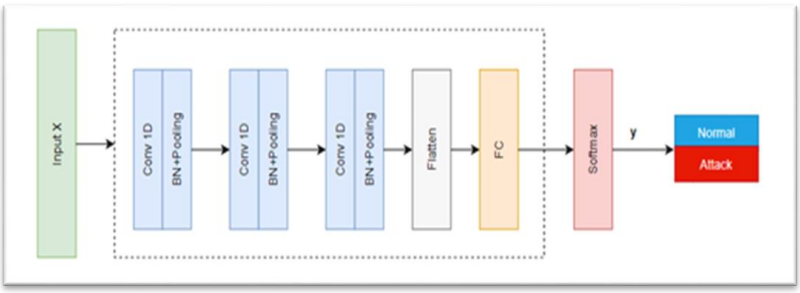
Hình 4: Max Pooling và Average Pooling trong CNN

+ Fully Connected (FC) nhận đầu vào là các dữ liệu đã được làm phẳng, mà mỗi đầu vào đó được kết nối đến tất cả neuron. Trong mô hình mạng CNNs, các tầng kết nối đầy đủ thường được tìm thấy ở cuối mạng và được dùng để tối ưu hóa mục tiêu của mạng ví dụ như độ chính xác của lớp.



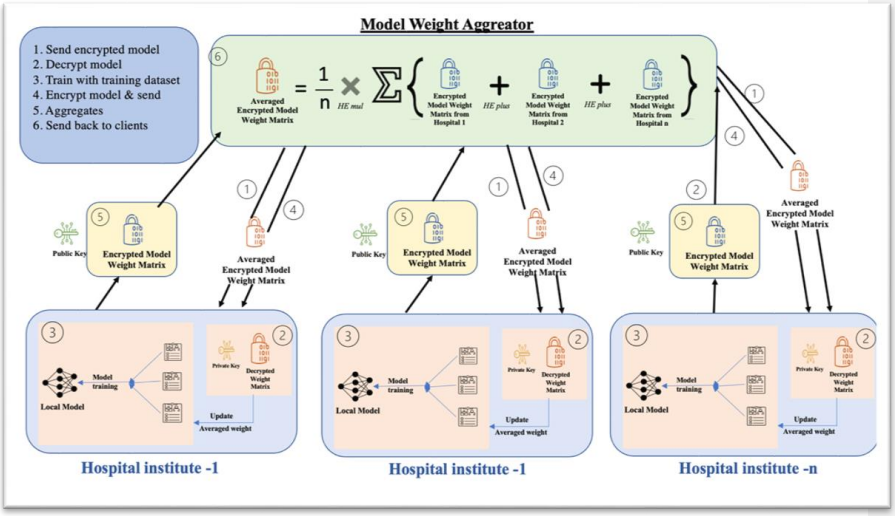
Hình 5: Mô phỏng cụ thể các Neural Network trong Fully Connected

- Dựa vào kiến thức trên, nhóm chúng tôi đề xuất mô hình mạng tích chập để phục vụ nghiên cứu dự đoán bệnh tim mạch mang lại hiệu quả cao như bên dưới:



Hình 6: Mô hình CNN được đề xuất

4.3.3. Mô hình học cộng tác kết hợp mã hoá đồng cấu:



Hình 7 : Mô hình học cộng tác kết hợp mã hoá đồng cấu

- Mã hoá đồng cấu sẽ được kết hợp trong mô hình học cộng tác tổng quát ở phía trên. Mỗi bên tham gia sẽ mã hoá tham số mô hình (bộ trọng số) sau khi huấn luyện và gửi chúng lên máy chủ tập trung để tổng hợp. Luồng hoạt động của mô hình, ta có thấy cơ bản bao gồm 4 giai đoạn :

1. Máy chủ tổng hợp khởi tạo mô hình cục bộ : Đây là giai đoạn đầu của quá trình, một máy chủ tin cậy bên ngoài sẽ khởi tạo bộ khoá (khoá riêng tư, khoá công khai), phân phối đến toàn bộ các bên tham gia, trong đó máy chủ tổng hợp chỉ nhận được khoá công khai. Điểm khác với mô hình FL thông thường, ở mô hình học công tác kết hợp mã hoá đồng cấu, máy chủ tổng hợp sẽ **mã hoá** bộ trọng số **W0** ban đầu với PK thành **eW0** sau đó gửi Wo1 đến từng bên tham gia.

2. Tự huấn luyện mô hình cục bộ : Khi nhận được bộ trọng số Wo, mỗi bên tham gia sẽ dùng khoá riêng tư để giải mã bộ trọng số Wo, mỗi bên sẽ tiếp tục tự huấn luyện mô hình học máy dựa trên dữ liệu cục bộ. Khi đã huấn luyện xong, từng bên tham gia tiến hành mã hoá bộ trọng số mới : W1, W2, W3,…,Wn thành eW1, eW2, eW3,…, eWn. Sau đó các trọng số đã được mã hoá này sẽ được gởi đến máy chủ tổng hợp.

3. Tổng hợp bộ trọng số các mô hình : Sau khi nhận được tất cả các bộ trọng số đã mã hoá từ tất cả bên tham gia, máy chủ tiến hành tổng hợp bằng FedAvg, ở đây việc tính toán sẽ thực hiện hoàn toàn trên bản mã (nhờ vào thuộc tính tính toán cộng, trừ của mã hoá đồng cấu). Kết quả nhận được là một bản mã đã được tổng hợp và tiếp tục gửi nó trở lại cho tất cả các bên tham gia.

4. Cập nhật lại mô hình cục bộ và tiếp tục huấn luyện : Từng bên tham gia tiến hành giải mã bằng khoá riêng tư thu được các trọng số chung và dùng nó để cập nhật mô hình cục bộ, tiếp tục huấn luyện ở vòng tiếp theo.

4.3.4. Giám sát các hoạt động huấn luyện mô hình học cộng tác sử dụng Permisioned Blockchain

- Trong đề tài này chúng tôi tận dụng tính minh bạch và bất biến của Blockchain để triển khai song song với FL nhằm hỗ trợ ghi lại mọi hành vi của các thành viên trong hệ thống. Tuy nhiên, Blockchain được phân thành nhiều loại khác nhau dựa trên tính chất và yêu cầu của tổ chức. Sau khi hệ thống được triển khai, một loại Blockchain cụ thể phù hợp sẽ được chọn tùy thuộc vào các yêu cầu, bao gồm kiểm soát truy cập, tốc độ, quy mô, v.v.

- Kiến trúc hệ thống mà chúng tôi đề xuất nhằm tự động hóa hệ thống cho việc phục vụ kết nối giữa những người có dữ liệu và những người muốn tạo mô hình ML từ dữ liệu đó. Ngoài ra, một cơ chế trả thưởng để khuyến khích các tổ chức tham gia đóng góp dữ liệu cũng sẽ được triển khai một cách chính xác và công bằng. Để làm được điều đó, chúng tôi phát triển một hợp đồng thông minh với Transaction's Type và Transaction's Attributes được miêu tả như ở bảng 1.

- Tất cả các hoạt động xuyên suốt quá trình học cộng tác của chủ sở hữu dữ liệu và server đều được lưu lại trên Blockchain, do đó có thể đánh giá được những đóng góp của các thành viên đó và có thể phát hiện nhanh chóng những hành vi bất thường. Khi có cuộc tấn công xảy ra, dựa vào tính minh bạch của Blockchain ta có thể truy vết và xóa bỏ kẻ tấn công ra khỏi hệ thống dễ dàng hơn.

Bảng 4.1: Cấu trúc của Blockchain Transaction trong việc giám sát hệ thống

Transaction's Type	**Transaction's Attributes**
Register Task	TaskID, Key, Timestamp
Submit Task	TaskID, Key, Model Hash, Round
Get Model	Model_ID, Round
UpdateModel	Model_ID, newTimestamp, newModelHashing, newTask, newQuality, newModel_Link, NewNote
DeleteModel	Model_ID, Round
CreateModel	Model_ID, Timestamp, modelHashing, task, quality, modelLink, Note
Claim Reward	TaskID, Key

4.3.5. Mô hình MEC sử dụng kết hợp phương pháp Học công tác.

- Mobile Edge Computing là một mạng lưới các trung tâm lưu trữ dữ liệu cục bộ trước khi chúng được gửi tới trung tâm dữ liệu chính. Trong mô hình học cộng tác, Edge Server đóng vai trò như một trung tâm dữ liệu thứ cấp có khả năng tổng hợp mô hình từ các bên tham gia và gửi mô hình đó lên trung tâm dữ liệu chính.

- MEC được sử dụng trong mô hình học cộng tác như một giải pháp giúp hệ thống trở nên phân tán hơn, đồng thời làm giảm áp lực và tính toán tổng hợp lên trung tâm dữ liệu chính. Những lợi ích của một Edge Server trong hệ thống có thể kể đến như là tối ưu hóa các hệ thống truyền dẫn, giảm tình trạng gián

đoạn hay làm chậm trễ quá trình gửi/nhận dữ liệu giữa chủ sở hữu dữ liệu và trung tâm dữ liệu chính.

4.4 Thực nghiệm và đánh giá

4.4.1. Môi trường thực nghiệm

A. Cấu hình

- Các kịch bản đánh giá hiệu quả của mô hình Học sâu cộng tác nhiều tầng kết hợp Mã hóa đồng cấu được thực nghiệm trên Google Colab, với cấu hình phần cứng thiết bị là Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz, RAM 12.7GB và ổ cứng 78.2GB.

Đã chú thích [PTD1]: Bổ sung

- Môi trường thực nghiệm Blockchain:

- Cấu hình phần cứng thiết bị: Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz chạy hệ điều hành Ubuntu 18.04 với RAM 8GB,
- Hyperledger Fabric phiên bản v2.4.7 được chạy trên Docker sử dụng Hệ điều hành Ubuntu 18.04 RAM 32GB, ổ cứng 60GB
- Chaincode được viết trên Node.js phiên bản mới nhất – 19.7.0 (bao gồm npm 9.5.0).

- Đánh giá hiệu năng của hệ thống Blockchain:

- Hyperledger Caliper phiên bản v.0.5.0 chạy trên Node.js v19.7.0

B. Xây dựng hệ thống theo mô hình

- Mô hình học sâu cộng tác nhiều tầng được xây dựng dựa trên thư viện TensorFlow, trong đó máy chủ tổng hợp (Aggregator), máy chủ biên (Edge Server) và các bên tham gia (Participant) được giả lập để trao đổi mô hình với nhau.

- Quá trình làm việc chi tiết của mô hình học sâu cộng tác nhiều tầng như sau:

1. Các máy chủ biên sẽ ghi danh với máy chủ tổng hợp, mỗi máy chủ biên sẽ được cung cấp số thứ tự để nhận dạng và số lượng vòng cần phải huấn luyện.
2. Các bên tham gia sẽ ghi danh với máy chủ biên quản lý nó, mỗi bên tham gia sẽ được cung cấp số thứ tự để nhận dạng và số lượng vòng cần phải huấn luyện. Máy chủ biên sẽ chọn ngẫu nhiên một bên tham gia làm máy lãnh đạo.
3. Bên tham gia được chọn làm máy lãnh đạo sẽ gửi trọng số khởi đầu đến cho máy chủ biên quản lý nó và các bên tham gia còn lại sẽ nhận trọng số từ máy chủ biên.
4. Các bên tham gia huấn luyện mô hình với dữ liệu cục bộ. Sau đó, các bên tham gia gửi mô hình cục bộ vừa được huấn luyện và số lượng dữ liệu của mình đến máy chủ biên quản lí.
5. Máy chủ biên nhận được mô hình cục bộ và kiểm tra số lượng mô hình cục bộ nhận được, nếu số lượng mô hình nhận được đã đủ (bằng với số lượng các bên tham gia đã ghi danh ở bước 2), nó sẽ đọc từng tập tin lên và thực hiện tổng hợp trung bình (FedAvg) theo công thức như sau:

Đã chú thích [PTD2]: Viêts cong thức, ko chụp

$$W_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} W_{t+1}^{k}$$

6. Máy chủ biên sau đó gửi mô hình vừa được tổng hợp trung bình đến máy chủ tổng hợp.
7. Máy chủ tổng hợp nhận được mô hình từ các máy chủ biên và kiểm tra số lượng các mô hình nhận được, nếu số lượng mô hình nhận được đã đủ (bằng với số lượng các máy chủ biên đã ghi danh ở bước 1), nó sẽ đọc từng tập tin lên và thực hiện tổng hợp trung bình (với công thức tương tự bước 5). Kết quả của bước này sẽ cho ra mô hình toàn cục.
8. Các bên tham gia lấy mô hình toàn cục đã được tổng hợp trung bình từ máy chủ tổng hợp và cập nhật mô hình cho mình.

TenSEAL - thư viện mã hóa đồng cấu:

TenSEAL là một thư viện để thực hiện các hoạt động mã hóa đồng cấu trên tensors, được xây dựng dựa trên Microsoft SEAL. Nó cung cấp tính dễ sử dụng thông qua API Python, đồng thời duy trì hiệu quả bằng cách triển khai hầu hết các hoạt động của nó bằng C++. Ở đồ án của nhóm, mong muốn sử dụng mã hoá đồng cấu CKKS, hỗ trợ tính toán số thực, phù hợp với việc tính toán trên dữ liệu học máy. Chúng tôi sử dụng khóa có poly_modulus_degree = 8096 bởi vì sử dụng phương pháp batching, phương pháp này thực hiện nối nhiều số để mã hóa cùng lúc nên plaintext sẽ bị giới hạn bởi độ dài của khóa, đồng thời nó giúp giảm thiểu chi phí, thời gian tính toán vì số lượng dữ liệu của các mô hình tương đối lớn.

4.4.2. Kịch bản thực nghiệm

- Chúng tôi thử nghiệm qua các kịch bản để đánh giá về hiệu năng, khả năng bảo đảm quyền riêng tư.

1. **FL:** Huấn luyện mô hình học cộng tác lần lượt giữa các Round và số lượng Client tham gia khác nhau, dữ liệu được chia điều cho các bên tham gia.

2. **FL-HE**: Huấn luyện các mô hình với phương pháp học cộng tác kết hợp mã hóa đồng cấu, dữ liệu được chia điều cho các bên tham gia.

3. Blockchain – FL: Đo hiệu năng của hệ thống Blockchain kết hợp mô hình học cộng tác dựa trên các thông số như Độ trễ lớn nhất, độ trễ thấp nhất và Throughput.

4. **FL-MEC**: Huấn luyện các mô hình với phương pháp học cộng tác kết hợp điện toán đám mây, dữ liệu được chia điều cho các bên tham gia.

5. **FL-MEC-HE:** Huấn luyện các mô hình với phương pháp học cộng tác kết hợp điện toán đám mây và mã hóa đồng cấu, dữ liệu được chia điều cho các bên tham gia.

4.4.3. Thu thập và xử lý dữ liệu

- ECG Heartbeat Categorization Dataset được sử dụng để đánh giá hiệu năng của mô hình FL trong ngữ cảnh dự đoán bệnh tim mạch. Bộ dataset này cũng được dùng cho việc đánh giá độ hiệu quả của mô hình phát hiện rối loạn nhịp tim trong nghiên cứu của Moody[8]. Nó là tập hợp những thông số liên quan đến tim mạch được lấy từ hai bộ dataset nổi tiếng là: MIT-BIH

- Arrhythmia Dataset và PTB Diagnostic ECG Database. Số lượng mẫu dữ liệu trong hai bộ dữ liệu này đủ lớn để phục vụ cho mục đích huấn luyện mô hình học sâu. Các bộ dữ liệu được thu thập thông qua phép đo điện tâm đồ (ECG) của nhịp tim ở những người bình thường và những người bị ảnh hưởng bởi rối loạn nhịp tim.

- Các mẫu dữ liệu trong bộ dataset được chia thành 5 nhãn: Normal, Atrial, Premature, Premature ventricular contraction, Fusion of ventricular and normal, Fusion of paced and normal.

- Tuy nhiên trong quá trình huấn luyện mô hình học sâu, chúng tôi đã thay đổi nhãn thành Normal và Abnormal để xây dựng mô hình phân loại nhị phân với hiệu năng tốt hơn.

Đã chú thích [PTD3]: Abnormal chứ

- Trong đề tài này, chúng tôi sử dụng bộ dữ liệu Arrhythmia để huấn luyện mô hình, chi tiết của bộ dữ liệu này:

  - Số lượng mẫu dữ liệu: 109446
  - Số lượng nhãn: 5
  - Nguồn dữ liệu: Physionet's MIT-BIH Arrhythmia Dataset
  - Các nhãn dán nhãn theo số thứ tự như sau: ['N': 0, 'S': 1, 'V': 2, 'F': 3, 'Q': 4]

4.4.4 Tiêu chí đánh giá

- Để đánh giá hiệu năng của mô hình, chúng tôi sử dụng các giá trị accuracy, precision, recall và cuối cùng là F1 – score. Các chỉ số này được tính từ các thuộc tính trong Confusion Matrix, bao gồm có dương tính thật (True Positive – TP), dương tính giả (False Positive – FP), âm tính thật (True Negative – TN) và âm tính giả (False Negative – FN). Trong ngữ cảnh an ninh mạng, giả sử positive là mẫu tấn công và negative là mẫu bình thường thì các thuộc tính này sẽ có ý nghĩa như sau:

• **TP**: là số lượng mẫu tấn công được phân loại đúng.

• **FP:** là số lượng mẫu dữ liệu bị phân loại là tấn công nhưng thực chất là các mẫu bình thường.

• **TN:** là số lượng mẫu bình thường được phân loại đúng.

• **FN:** là số lượng mẫu dữ liệu bị phân loại là bình thường nhưng thực chất là các mẫu tấn công.

Bảng 4.2: Định nghĩa toán học của các tiêu chí đánh giá mô hình

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - score = 2\,x\,\frac{Precision * Recall}{Precision + Recall}$$

4.4.5 Kết quả thực nghiệm

4.4.5.1. Hiệu năng của Blockchain trong ngữ cảnh Học cộng tác.

- Hyperledger Fabric là một Blockchain Framework với mã nguồn mở do Linux Foundation cung cấp. Đặc trưng của Fabric là một Permisioned Blockchain, nghĩa là tất cả các thành viên tham gia đều được chứng thực trước khi vào mạng lưới này. Tận dụng điều đó, chúng tôi sử dụng Fabric như một nền tảng Blockchain của kiến trúc hệ thống đã đề xuất.

- Để đánh giá hiệu năng của công nghệ này, chúng tôi sử dụng Hyperledger Caliper - một công cụ có khả năng đánh giá toàn diện các nền tảng Private/Public Blockchain như Hyperledger Fabric, Hyperledger Sawtooth và Ethereum dựa trên các thông số như Succesful Request, Latency, Throughput và Resource Consumption của các Transaction.

- Một trong những nhiệm vụ chính của hợp đồng thông minh khi tương tác với hệ thống là Get Model nên phương thức này sẽ được dùng để đánh giá. Để có cái nhìn tổng quát hơn, chúng tôi tăng số lượng các transaction lên 1000, 5000, 10000, 50000, 100000 để thu được bảng số liệu như Bảng 2.

Bảng 4.3: Hiệu năng của giao dịch Get Model trong hợp đồng thông minh

Successful request	Min latency(s)	Max latency (s)	Avg latency (s)	Throughput (TPS)
1000	0.001	0.04	0.00	17.18
5000	0.003	0.17	0.01	42.34
10000	0.003	0.25	0.01	56.10
50000	0.012	0.42	0.05	209.94
100000	0.018	0.64	0.07	335.52

- **Kết luận:** Kết quả nhận được cho thấy hệ thống kiến trúc Blockchain đề xuất hoàn toàn đáp ứng được nhu cầu giám sát dữ liệu. Thông số Successfull Request là 100% cho thấy hệ thống Blockchain xử lí và lưu trữ dữ liệu rất hiệu quả ngay cả khi số lượng transaction lên tới 100000. Bên cạnh đó, ta thấy độ trễ hệ thống càng tăng khi số lượng transaction tăng lên và lượng Throughput cũng vậy.

4.4.5.2. Hiệu năng của mô hình Học cộng tác

- Huấn luyện với mô hình CNN được đề xuất ở mục 4.3.2, chúng tôi sẽ thực hiện đánh giá trên toàn bộ tập dữ liệu đã tiền xử lý và cân bằng, mỗi lần huấn luyện qua 5 epoch, *batch_size* bằng 64, sử dụng hàm mất mát Categorical CrossEntropy và optimizer Adam với learning rate bằng 0.01. Chúng tôi thực hiện huấn luyện với số lượng các bên tham gia lần lượt là 2, 4, 6 và 8; với số vòng huấn luyện là 20. Kết quả của các kịch bản thực nghiệm như sau:

Bảng 4.4: Hiệu năng của mô hình Học cộng tác truyền thống (FL)

Round	Client	Accuracy	Recall	Precision	F1-Score
10	**2**	0.9688	0.8468	0.9682	0.9033
	4	0.9793	0.8967	0.9814	0.9371
	6	0.9815	0.9187	0.9728	0.9449
	8	0.9837	0.9306	0.9739	0.9518
20	**2**	0.9784	0.9189	0.9543	0.9363
	4	0.9810	0.9295	0.9590	0.9440
	6	**0.9829**	**0.9780**	**0.9752**	**0.9696**
	8	0.9792	0.9010	0.9764	0.9764

- Kết luận: Kết quả thực nghiệm cho thấy mô hình học cộng tác có độ chính xác khá tốt, khi càng nhiều client tham gia vào quá trình đóng góp dữ liệu thì mô hình sẽ ổn định hơn.

4.4.5.3. Hiệu năng của mô hình Học cộng tác kết hợp Mã hóa đồng cấu (FL + HE):

- Việc bộ trọng số được mã hóa với CKKS có sử dụng kỹ thuật batch để đo lường kích thước của mô hình sau khi mã hóa và thời gian mã hóa. Ở **bảng 4.5.1** khi không áp dụng batch, dữ liệu khi tính toán tăng lên cấp số nhân dẫn đến mô hình không thể tính toán được. Việc sử dụng batch mang lại nhiều lợi ích về thời gian tính toán cũng như kích thước bản mã.

- Kết quả **bảng 4.5.2**, ta có thể thấy mô hình sau khi mã hoá có kích thước rất lớn so với ban đầu. Tuy nhiên, khi áp dụng vào mô hình Học cộng tác, hiệu năng không có nhiều sự chênh lệch.

Bảng 4.5.1 : Kích thước dữ liệu sau khi mã hoá chưa kết hợp batch

Model	Parameters	Plaintext size	Ciphertext size	Time to encrypt
CNN	21058	1.7 MB	Out of memory	

Bảng 4.5.2 : Kích thước dữ liệu sau khi mã hoá kết hợp batch

Model	Parameters	Plaintext size	Ciphertext size	Time to encrypt
CNN	21058	1.7 MB	689.6 MB	6.5s

Bảng 4.6:Hiệu năng của mô hình Học cộng tác kết hợp mã hoá đồng cấu(FL-HE)

Round	Client	Accuracy	Recall	Precision	F1-Score
10	**2**	0.9737	0.8789	0.9654	0.9201
	4	0.9766	0.8953	0. 9665	0.9296
	6	0.9771	0.8847	0.9806	0.9302
	8	0.9787	0.9041	0.9707	0.9362
20	**2**	0.9773	0.8858	0.9806	0.9308
	4	0.9804	0.9112	0.9734	0.9413
	6	**0.9821**	**0.9261**	**0.9687**	**0.9469**
	8	0.9805	0.9091	0.9766	0.9417

- Kết luận: Kết quả đo được cho thấy với số lượng Client giống nhau, khi số lượng Round được tăng lên thì mô hình FL-HE cho kết quả cao, các thông số Accuracy, Recall, Precision, F1-Score có xu hướng tăng theo.

- Việc đánh đổi về kích thước bản mã, thời gian tăng lên nhiều lần, nhưng kết quả cho thấy độ chính xác của mô hình khá tốt kèm theo việc bảo mật an toàn của mô hình được cải thiện đáng kể.

4.4.5.4. Mô hình học cộng tác nhiều tầng (FL+MEC) trong ngữ cảnh không sử dụng mã hóa đồng cấu

- Ở mô hình này, chúng tôi sử dụng 2 Edge Server ở biên mạng có nhiệm vụ làm trung gian trao đổi bộ weights giữa Local và Server.

- Trong mục tiêu sử dụng MEC trong mô hình học cộng tác, chúng tôi mong muốn giảm áp lực xử lí, truyền tải dữ liệu lên Server chính, do đó chúng tôi đã tạo một bảng thống kê để nghiệm thu.

- Hơn nữa, việc sử dụng các Edge Server ở biên cũng có một vai trò khác đó là Backup trong trường hợp một trong các Server chính bị hư hỏng.

- Để so sánh chi phí truyền thông giữa chiến lược FL truyền thống và FL kết hợp MEC, giả sử mô hình FL được huấn luyện với tổng số Round là 20 (R=20). Kích thước của mô hình CNN đã được huấn luyện của chúng tôi là 1.7MB, trong trường hợp 10 Client (K=10) tham gia vào quá trình đóng góp dữ liệu, tổng số dữ liệu trao đổi giữa Server và Client được tính là 20 * 10 * 2 * 1.7

= 680 MB, trong đó tính *2 là chi phí truyền thông đi và về giữa local và Server trong 1 round.

- Bên cạnh đó, nếu áp dụng MEC với 2 Edge Server, tổng số dữ liệu mà Server chính nhận được là 20 * 2 * 2 * 1.7 = 136 MB, bởi vì mỗi Edge Server đã xử lí lần lượt 5 client và sau khi xử lí xong sẽ gửi mô hình tổng hợp cuối cùng của mình cho Server. Do đó Server thật sự chỉ nhận được 2 mô hình từ 2 Edge Server.

- Với số Round = 20, kích thước của mô hình 1.7 MB, kết quả thống kê sau khi tăng số Client lên 100, 1000, đồng thời tăng số lượng Edge Server lên 5, 10 được thể hiện ở **bảng 4.7.**

Bảng 4.7 Thống kê tổng dữ liệu Server phải xử lí trong 20 Round.

Number of Client	10	100	1000
FL	680 MB	6800 MB	68000 MB
FL + 10 Edge Server	680 MB	680 MB	680 MB
FL+ 5 Edge Server	340 MB	340 MB	340 MB
FL+ 2 Edge Server	136 MB	136 MB	136MB

➔ **Nhận xét:**

- Ta có thể thấy khi tăng số lượng Client tham gia vào quá trình đóng góp dữ liệu thì chi phí truyền thông ở FL truyền thống tăng lên trong khi với FL + MEC, chi phí truyền thông không bị ảnh hưởng.

- Càng nhiều Edge Server được dựng lên thì càng nhiều tài nguyên hệ thống bị tiêu tốn, nhưng đồng thời áp lực tính toán, truyền tải dữ liệu ở các Edge Server và Server đều được giảm tải đáng kể.

- Chi phí truyền thông ở Server chính được giảm xuống ít nhất khi số lượng Edge Server ít, do đó chúng ta cần có những chiến lược phân phối dữ liệu sao cho phù hợp để hệ thống các Server hoạt động kết hợp một cách trơn tru, hiệu quả.

- Mặt khác, để đánh giá hiệu năng về độ chính xác của mô hình học sâu nhiều tầng, chúng tôi cũng thực hiện huấn luyện mô hình này với số lượng Round khác nhau để thu được kết quả như bảng.

Bảng 4.8: Hiệu năng của mô hình Học cộng tác kết hợp MEC

Round	Client	Accuracy	Recall	Precision	F1-Score
10	**2**	0.9727	0.8580	0.9815	0.9156
	4	0.9812	0.9197	0.9896	0.9440
	6	0.9834	0.9279	0.9744	0.9506
	8	0.9823	0.9229	0.9744	0.9506

20	**2**	0.9813	0.9096	0.9806	0.9438
	4	0.9837	0.9364	0.9680	0.9519
	6	**0.9847**	**0.9372**	**0.9733**	**0.9549**
	8	0.9842	0.9306	0.9769	0.9532

- Kết luận: Nghiệm thu cho biết rằng khi sử dụng mô hình học cộng tác sâu nhiều tầng thì độ chính xác thu được vẫn cao và xấp xỉ với mô hình học cộng tác truyền thống (chỉ một server tổng hợp). Bên cạnh đó, Edge Server có thể giảm tải áp lực tính toán lên trung tâm Server và cải thiện tốc độ truyền, hạn chế mất mát dữ liệu trong quá trình trao đổi các Local với Server.

Đã chú thích [PTD4]: Chỗ này có bảng kết quả nào thể hiện lượng dữ liệu trao đổi giữa client và server giảm ko? V thì mới rõ, còn bảng bên trên chỉ nói về độ chính xác, k nói về áp lực dữ liệu phải xử lí đã giảm

Đã chú thích [BTHĐ5R4]: Em đã thêm hihi

4.4.5.5. Mô hình học cộng tác nhiều tầng (FL+MEC) trong ngữ cảnh sử dụng mã hóa đồng cấu

- Với mô hình đề xuất trên, Mã hóa đồng cấu đóng vai trò tăng tính bảo mật dữ liệu của các chủ thể đóng góp dữ liệu và hạn chế các cuộc tấn công Man in the Middle nhằm trích xuất và truy ngược bộ weights.

Bảng 4.9: Hiệu năng của mô hình Học cộng tác kết hợp MEC, mã hoá đồng cấu

Round	Client	Accuracy	Recall	Precision	F1-Score
10	**2**	0.9748	0.9009	0.9503	0.9249
	4	0.9757	0.8834	0.9734	0.9262
	6	0.9787	0.9030	0.9712	0.9358
	8	0.9777	0.9038	0.9603	0.9312
20	**2**	0.9793	0.9054	0.9727	0.9378
	4	0.9740	0.9348	0.9163	0.9255
	6	**0.9803**	**0.9067**	**0.9774**	**0.9407**
	8	0.9779	0.9010	0.9521	0.9385

Kích thước Model 1.7 MB

- Kết luận:

- Đối với các kết quả huấn luyện dựa trên mô hình CNN, có thể thấy rằng sự chênh lệch trong độ chính xác giữa việc sử dụng phương pháp học học cộng tác truyền thống (chỉ có một máy chủ tổng hợp) và phương pháp học cộng tác kết hợp MEC là không quá lớn và vẫn có thể chấp nhận được vì các kết quả huấn luyện vẫn rất tốt, thậm chí là gần bằng nhau.

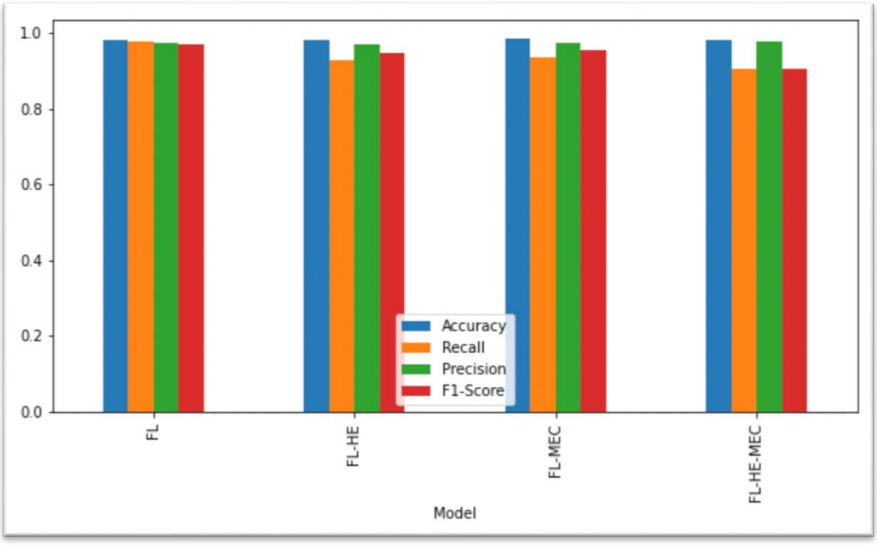
- Mặt khác, khi sử dụng tích hợp với mã hóa đồng cấu, có thể thấy rằng độ chính xác của mô hình vẫn được đảm bảo bên cạnh việc đảm bảo quyền riêng tư dữ liệu của người dùng.

- Ngoài ra, việc tích hợp HE vào mô hình học cộng tác nhiều tầng (FL+MEC) cũng mang lại độ chính xác cao, tốc độ nhanh hơn và đảm bảo

an toàn dữ liệu hơn so với mô hình chỉ có FL-MEC. Tuy nhiên, khi kết hợp Mã hóa đồng cấu vào các mô hình học sâu cộng tác thì sẽ tốn nhiều thời gian hơn trong lúc huấn luyện và sử dụng nhiều tài nguyên hơn so với mô hình thông thường. Điều này có thể dẫn đến hiện tượng "Bùng nổ dữ liệu" làm mô hình bị hỏng.

4.4.6. Đánh giá tổng quan các mô hình

- Các kết quả trên cho thấy rằng thực nghiệm trên số Round = 20 và số Client = 6 cho kết quả tốt hơn các trường hợp khác. Chúng tôi tiến hành đánh giá kết quả của các mô hình theo **hình 8.**



Hình 8: Kết quả so sánh các kịch bản đề xuất

- Dựa trên các kết quả có được, ta có thể thấy rằng mô hình áp dụng mã hóa đồng cấu (FL-HE) là một giải pháp đảm bảo riêng tư, an toàn dữ liệu tốt trong mô hình học sâu cộng tác, đồng thời cũng đảm bảo độ chính xác của mô hình không quá chênh lệch so với mô hình học máy truyền thống. Tuy nhiên, thời gian mã hoá và kích thước mô hình sau khi mã hoá tăng lên khá lớn nên ta phải đánh đổi hiệu năng của máy tính.

- Trong khi đó mô hình học sâu cộng tác nhiều tầng (FL+MEC) có xu hướng bị giảm nhẹ các thông số Recall và F1-Score so với mô hình học sâu cộng tác truyền thống. Tuy nhiên, lợi ích giảm tải chi phí truyền thông mà MEC mang lại cũng là điều mà ta có thể cân nhắc sử dụng.

- Như vậy, đối với các mô hình có độ phức tạp nhỏ, có thể áp dụng cả HE và MEC để đạt được một mô hình có độ chính xác ổn mà vẫn đạt được mức độ bảo mật riêng tư cao. Còn đối với các mô hình có độ phức tạp lớn, nên cân nhắc khi sử dụng HE vì có thể dẫn tới hiện tượng bùng nổ dữ liệu.

Đã chú thích [PHL6]: @Bùi Tấn Hải Đăng

- Qua biểu đồ tổng quan của 4 mô hình, có thể thấy với mô hình học sâu cộng tác truyền thống vẫn đạt độ chính xác cao nhất nhưng vẫn tiềm ẩn những mối lo ngại về bảo mật. Trong khi đó, thông số của mô hình học sâu cộng tác nhiều tầng kết hợp HE có thông số thấp nhất nhưng lại đảm bảo cải thiện quyền riêng tư và giảm chi phí truyền thông dữ liệu giữa các bên.

5. Tên sản phẩm: Mô hình học sâu cộng tác tích hợp Blockchain và Mã hóa đồng cấu có hỗ trợ điện toán biên trong ngữ cảnh y tế thông minh TrustFedHealth.

6. Hiệu quả, phương thức chuyển giao kết quả nghiên cứu và khả năng áp dụng

Việc tích hợp Federated Learning với Blockchain cũng như với các thành phần liên quan như điện toán biên Mobile, Homormophic Encryption trong ngữ cảnh y tế thông minh đã giúp cho hệ thống đề xuất của chúng tôi chẩn đoán nguy cơ đột quỵ với chi phí triển khai và duy trì thấp, đồng thời cải thiện được những vấn đề về quyền riêng tư so với mô hình học máy truyền thống.

Trong nghiên cứu này, chúng tôi đã đạt được một số kết quả như sau:

- Xây dựng mô hình học cộng tác nhiều tầng kết hợp Blockchain và mã hóa đồng cấu trong ngữ cảnh y tế thông minh.
- Áp dụng và đánh giá hai mô hình CNN dựa trên tập dữ liệu ECG Heartbeat Categorization bằng phương pháp học sâu cộng tác.
- Đánh giá kết quả giữa phương pháp học cộng tác chỉ sử dụng một máy chủ tổng hợp và phương pháp học cộng tác nhiều tầng.

Tóm lại, đề tài này đã chứng minh được tính hiệu quả của mô hình học sâu cộng tác nhiều tầng mà vẫn giữ được quyền riêng tư dữ liệu của các bên tham gia cũng như khuyến khích được các bên tham gia nhiều hơn.

Bên cạnh những kết quả đạt được, chúng tôi vẫn còn một số hạn chế nhất định như là vẫn chưa triển khai được trên nhiều mô hình học máy khác nhau, nhiều tập dữ liệu khác nhau.

7. Các hướng mở rộng nghiên cứu trong tương lai của đề tài

- Nghiên cứu kết hợp nhiều mô hình học máy với nhau để đưa ra kết quả huấn luyện cao.
- Triển khai huấn luyện trên nhiều tập dữ liệu khác nhau để nâng cao khả năng dự đoán của mô hình.
- Kết hợp với các thuật toán tổng hợp mô hình khác để tối ưu hóa mô hình.
- Triển khai lưu trữ phân tán trên hệ thống IPFS của Blockchain.

8. Hình ảnh, sơ đồ minh họa chính

Hình ảnh và sơ đồ chi tiết có tại mục số 4.

TÀI LIỆU THAM KHẢO

[1]	Bouacida, N., & Mohapatra, P. (2021). Vulnerabilities in federated learning. *IEEE Access*, *9*, 63229-63249.
[2]	Stripelis, D., Saleem, H., Ghai, T., Dhinagar, N., Gupta, U., Anastasiou, C., ... & Ambite, J. L. (2021, December). Secure neuroimaging analysis using federated learning with homomorphic encryption. In 17th International Symposium on Medical Information Processing and Analysis (Vol. 12088, pp. 351-359). SPIE.

[3]	Kogos, K. G., Filippova, K. S., & Epishkina, A. V. (2017, February). Fully homomorphic encryption schemes: The state of the art. In 2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus) (pp. 463-466). IEEE.
[4]	Wang, R., Li, H., & Liu, E. (2021). Blockchain-based federated learning in mobile edge networks with application in internet of vehicles. *arXiv preprint arXiv:2103.01116*.
[5]	Nguyen, Dinh C., Ming Ding, Quoc-Viet Pham, Pubudu N. Pathirana, Long Bao Le, Aruna Seneviratne, Jun Li, Dusit Niyato, and H. Vincent Poor. "Federated learning meets blockchain in edge computing: Opportunities and challenges." *IEEE Internet of Things Journal* (2021).[[]
[6]	Pfitzner, B., Steckhan, N., & Arnrich, B. (2021). Federated learning in a medical context: A systematic literature review. *ACM Transactions on Internet Technology (TOIT)*, *21*(2), 1-31.
[7]	Chen, Y., Qin, X., Wang, J., Yu, C., & Gao, W. (2020). Fedhealth: A federated transfer learning framework for wearable healthcare. *IEEE Intelligent Systems*, *35*(4), 83-93
[8]	G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," IEEE Engineering in Medicine and Biology Magazine, vol. 20, no. 3, pp. 45–50, 2001.

Cơ quan Chủ trì
(ký, họ và tên, đóng dấu)

Chủ nhiệm đề tài
(ký, họ và tên)

Bùi Tấn Hải Đăng

Mẫu F6