THÔNG TIN CHUNG CỦA BÁO CÁO

- Link YouTube video của báo cáo (tối đa 5 phút): https://youtu.be/fBCl7poI-2o
- Link slides (dạng .pdf đặt trên Github):
  *(ví dụ: https://github.com/chauthevi2004/CS519.O21.KHTN/**Slide.pdf**)*
- *Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới*
- *Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in*

• Họ và Tên: Châu Thế Vĩ • MSSV: 22521653 	• Lớp: CS519.O21.KHTN • Tự đánh giá (điểm tổng kết môn): 9.5/10 • Số buổi vắng: 0 • Số câu hỏi QT cá nhân: 9 • Link Github: https://github.com/chauthevi2004/CS519.O21.KHTN

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

ÁP DỤNG TRÍ TUỆ NHÂN TẠO CÓ THỂ GIẢI THÍCH VÀ THÍCH ỨNG MIỀN CHO CHẨN ĐOÁN VÀ ĐIỀU TRỊ TRONG HỆ THỐNG Y TẾ LIÊN KẾT

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

APPLICATION OF EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI) AND DOMAIN-ADAPTATION (DA) FOR DIAGNOSIS AND TREATMENT IN FEDERATED LEARNING-BASED HEALTHCARE SYSTEM

TÓM TẮT *(Tối đa 400 từ)*

Trí tuệ nhân tạo (AI) và học máy (ML) đang cải tiến phân tích hình ảnh y tế, nhưng bảo mật và quyền riêng tư dữ liệu hạn chế việc chia sẻ dữ liệu giữa các địa điểm. Học liên kết (Federated Learning - FL) là giải pháp hứa hẹn, cho phép đào tạo mô hình ML trên dữ liệu phi tập trung mà không cần chia sẻ thông tin nhạy cảm. Nghiên cứu này tập trung cải tiến thích ứng miền (DA) và AI có thể giải thích (XAI) cho chẩn đoán và điều trị y tế trong hệ thống y tế liên kết.

Mục tiêu: (1) cải tiến DA trong y tế, nâng cao độ chính xác và tin cậy của mô hình ML trên dữ liệu mới; (2) phát triển và áp dụng XAI trong y tế, đảm bảo mô hình AI minh bạch và dễ hiểu, tạo niềm tin cho chuyên gia y tế; (3) phát triển ứng dụng AI mới trong nghiên cứu và chăm sóc sức khỏe, hỗ trợ phát hiện sớm bệnh tật, chẩn đoán chính xác và dự đoán tiên lượng.

Nội dung: (1) khảo sát và phân tích các nghiên cứu hiện có về DA và XAI trong FL; (2) thiết kế phương pháp DA cho hệ thống y tế liên kết; (3) phát triển XAI cho học liên kết trong y tế; (4) thử nghiệm các phương pháp XAI và DA trong hệ thống FL, đánh giá hiệu suất trên các bộ dữ liệu khác nhau.

Kết quả: báo cáo về việc áp dụng XAI và DA vào mô hình học máy liên kết trong y tế, đánh giá hiệu quả và độ chính xác của mô hình, cung cấp bằng chứng về khả năng giải thích quyết định của mô hình trong thực tế.

Nghiên cứu này giải quyết các vấn đề trong ứng dụng AI trong y tế, vượt qua thách thức về dữ liệu và quyền riêng tư, nâng cao độ chính xác, tin cậy và minh bạch của mô hình AI trong chẩn đoán và điều trị y tế, đóng góp vào y học cá nhân hóa và nâng cao chất lượng chăm sóc sức khỏe.

GIỚI THIỆU *(Tối đa 1 trang A4)*

Trong những năm gần đây, Trí tuệ nhân tạo (AI) đã và đang thay đổi cách chúng ta phân tích và xử lý dữ liệu trong nhiều lĩnh vực, đặc biệt là y tế. AI mang lại tiềm năng lớn trong việc cải thiện chẩn

đoán và điều trị, nhờ vào khả năng phân tích dữ liệu nhanh chóng, chính xác và hiệu quả. Tuy nhiên, y tế là một lĩnh vực đặc thù với các loại dữ liệu nhạy cảm, các nhiệm vụ phức tạp và mức độ rủi ro cao, do đó yêu cầu một mức độ bảo mật và trách nhiệm giải trình cao.

Phân tích hình ảnh y tế đã được thúc đẩy bởi các phương pháp thị giác máy tính và học máy [1, 2, 3, 4] hiện đại, điển hình như học sâu (deep learning) [5]. Trong lĩnh vực y tế, một trong những thách thức lớn là vấn đề kích thước mẫu nhỏ [6, 7], làm giảm sức mạnh thống kê và độ tin cậy của các mô hình học máy. Để giải quyết vấn đề này, việc sử dụng dữ liệu đa miền từ nhiều địa điểm thu thập khác nhau là cần thiết. Tuy nhiên, các quy định bảo vệ quyền riêng tư nghiêm ngặt như HIPAA [8] và GDPR [9] đã hạn chế việc chia sẻ dữ liệu giữa các trung tâm y tế.

Để vượt qua các thách thức này, Học tập liên kết (Federated Learning - FL) [10, 11, 12] đã nổi lên như một giải pháp tiềm năng, cho phép đào tạo mô hình cộng tác sử dụng dữ liệu từ nhiều nguồn khác nhau mà không cần chia sẻ thông tin nhạy cảm. FL giúp xây dựng một mô hình toàn cầu từ các mô hình cục bộ tại từng địa điểm, sau đó mô hình này được gửi đến các địa điểm để tinh chỉnh và triển khai. Đây là một giải pháp hứa hẹn để giải quyết vấn đề quy mô mẫu nhỏ và bảo vệ quyền riêng tư trong y tế.

Trong lĩnh vực y tế, các quyết định dựa trên AI có thể ảnh hưởng trực tiếp đến sức khỏe bệnh nhân. Do đó, Trí tuệ nhân tạo giải thích (Explainable AI - XAI) là một phương pháp được đề xuất nhằm tăng cường tính minh bạch và khả năng hiểu biết của các mô hình AI. Trong lĩnh vực y tế, các quyết định dựa trên AI có thể ảnh hưởng lớn đến sức khỏe bệnh nhân. Do đó, XAI giúp giải thích và làm rõ các quyết định của mô hình, từ đó nâng cao độ tin cậy và khả năng chấp nhận AI trong y tế [13, 14]. XAI có thể bổ trợ cho cách mà con người quan sát và hiểu dữ liệu, giúp các chuyên gia y tế dễ dàng tiếp cận và tin cậy hơn vào các quyết định của mô hình AI [15, 16].

Đầu vào (Input):

- Dữ liệu y tế từ nhiều nguồn: Bao gồm hình ảnh y tế, hồ sơ bệnh án và các dữ liệu lâm sàng từ nhiều thiết bị và trung tâm y tế khác nhau.
- Các mô hình AI hiện có: Sử dụng các mô hình học máy hiện đại như học sâu (deep learning) và học liên kết (federated learning).

Đầu ra (Output): Mô hình AI cải tiến: Các mô hình AI có khả năng thích ứng với sự thay đổi của miền dữ liệu và đưa ra các quyết định minh bạch, dễ hiểu.

Đề tài nghiên cứu này tập trung vào việc đánh giá và cải tiến các phương pháp thích ứng miền (Domain Adaptation - DA) và XAI trong y tế. Mục tiêu chính là cải thiện độ chính xác và tin cậy của các mô hình học máy khi áp dụng trên dữ liệu mới hoặc chưa được nhìn thấy trước đó. Đề tài cũng khám phá và phát triển các ứng dụng AI trong chăm sóc sức khỏe và nghiên cứu y tế, bao gồm phát hiện sớm bệnh tật, chẩn đoán chính xác và dự đoán tiên lượng, từ đó đóng góp vào việc nâng cao chất lượng chăm sóc sức khỏe và kết quả điều trị cho bệnh nhân. [17, 18]

Với sự tiến bộ không ngừng của AI và sự phát triển của các kỹ thuật FL và XAI, đề tài này hứa hẹn mang lại những đóng góp quan trọng trong lĩnh vực y tế, giúp nâng cao hiệu quả và độ tin cậy của

các mô hình AI, đồng thời bảo vệ quyền riêng tư của bệnh nhân.

MỤC TIÊU

(Viết trong vòng 3 mục tiêu, lưu ý về tính khả thi và có thể đánh giá được)

1. Nghiên cứu và xây dựng ứng dụng AI trong chăm sóc sức khỏe và nghiên cứu y tế: Mục tiêu thúc đẩy sự phát triển của các ứng dụng AI mới trong nghiên cứu y tế và chăm sóc sức khỏe. Điều này bao gồm việc tìm ra cách thức mới để AI có thể hỗ trợ trong việc phát hiện sớm bệnh tật, chẩn đoán chính xác, và dự đoán tiên lượng, từ đó đóng góp vào việc cải thiện chất lượng chăm sóc sức khỏe và kết quả điều trị cho bệnh nhân.

2. Nghiên cứu và áp dụng thích ứng miền (Domain Adaptation) để cải thiện dữ liệu trong y tế: Mục tiêu chính là đánh giá và cải thiện cách thức thích ứng miền trong lĩnh vực y tế. Điều này bao gồm việc giải quyết sự thay đổi giữa các bộ dữ liệu từ các nguồn khác nhau (chẳng hạn như từ các thiết bị khác nhau hoặc các môi trường lâm sàng khác nhau), để cải thiện độ chính xác và độ tin cậy của các mô hình học máy khi chúng được áp dụng trên dữ liệu mới hoặc chưa được nhìn thấy trước đó.

3. Nghiên cứu và áp dụng AI giải thích (Explainable AI/XAI) trong y tế: Xác định cách thức dự đoán của các mô hình AI có thể được làm cho minh bạch hơn và dễ hiểu hơn trong bối cảnh y tế, nơi mà quyết định dựa trên các mô hình này có thể có hậu quả lớn đối với sức khỏe của bệnh nhân. Việc nâng cao khả năng giải thích của AI không chỉ giúp các bác sĩ và nhân viên y tế hiểu và tin tưởng vào quyết định mà còn đảm bảo rằng các quyết định này có thể được giám sát, đánh giá một cách hiệu quả và có độ tin cậy.

NỘI DUNG VÀ PHƯƠNG PHÁP

(Viết nội dung và phương pháp thực hiện để đạt được các mục tiêu đã nêu)

Nội dung 1: Tìm hiểu các kiến thức nền tảng và các công trình liên quan hiện có của chuyển đổi miền (Domain Shift) và AI có thể giải thích (XAI) cho học liên kết (FL).

Mục tiêu:

- Khảo sát và phân tích các nghiên cứu hiện có về domain shift và XAI trong FL.
- Đánh giá tác động của domain shift đối với hiệu suất của các mô hình học liên kết.
- Tìm hiểu các phương pháp XAI có thể giúp giải thích và cải thiện mô hình trong môi trường FL.

Phương pháp thực hiện:

- Thực hiện tổng quan các bài báo, nghiên cứu đã công bố liên quan đến domain shift và XAI trong FL, từ các hội nghị và tạp chí có uy tín.
- Phân tích các phương pháp hiện có, nhằm mục đích xác định lỗ hổng kiến thức và những thách thức cụ thể trong việc áp dụng các kỹ thuật này trong FL.
- Thử nghiệm một số kỹ thuật XAI để đánh giá và giải thích hiệu suất của mô hình dưới ảnh hưởng của domain shift trong môi trường FL.

Kết quả dự kiến:

- Hiểu được cơ bản về các nghiên cứu hiện có, bao gồm đánh giá các kỹ thuật, công cụ và phương pháp đã được sử dụng để giải quyết domain shift và áp dụng XAI trong FL.
- Xác định các thách thức chính và đề xuất các giải pháp tiềm năng cho việc cải thiện hiệu

suất và độ minh bạch của mô hình trong học liên kết.
- Hiểu được hiệu quả của các phương pháp XAI trong việc giải thích và giảm thiểu tác động của domain shift.

Nội dung 2: Thiết kế phương pháp dùng thích ứng miền (Domain Adaptation) cho hệ thống y tế dựa trên học liên kết.

Mục tiêu:
- Hiểu rõ được cách thức áp dụng của thích ứng miền
- Thiết kế mô hình học máy liên kết áp dụng thích ứng miền để giải quyết vấn đề chuyển đổi miền.

Phương pháp thực hiện:
- Tham khảo, tổng hợp các bài báo, tài liệu liên quan.
- Hiện thực mô hình học máy liên kết có áp dụng thích ứng miền.

Kết quả dự kiến:

Thiết kế được mô hình học máy liên kết có áp dụng thích ứng miền để giải quyết vấn đề chuyển đổi miền một cách hiệu quả.

Nội dung 3: Thiết kế phương pháp dùng AI có thể giải thích (XAI) cho học liên kết trong y tế thông minh.

Mục tiêu:
- Hiểu rõ tính minh bạch và khả năng giải thích của các mô hình AI, đặc biệt là trong lĩnh vực y tế, để giải quyết vấn đề tin cậy và hiểu biết giữa các nhà khoa học dữ liệu và các chuyên gia y tế.
- Phát triển các mô hình AI có thể giải thích được quyết định của chúng, giúp nâng cao độ chính xác và tin cậy trong các quyết định lâm sàng.
- Giảm bớt sự phụ thuộc vào các mô hình "hộp đen", qua đó cải thiện sự chấp nhận của công nghệ AI trong y tế.

Phương pháp nghiên cứu:
- Tham khảo, tổng hợp các bài báo, tài liệu liên quan.
- Tìm hiểu các tập dữ liệu như: ADNI, CheXpert, BRATS,...
- Thực nghiệm các mô hình học máy liên kết có thể giải thích.

Kết quả dự kiến:
- Nắm được phương pháp hoạt động của các mô hình học máy.
- Đánh giá được độ hiệu quả giữa các mô hình khi áp dụng các tập dữ liệu khác nhau.
- Cải thiện sự hiểu biết và tin cậy của các mô hình AI trong ngành y tế, giúp chúng trở thành công cụ hỗ trợ đắc lực cho việc chẩn đoán và điều trị.

Nội dung 4: Hiện thực hệ thống và thực nghiệm cho AI có thể giải thích (XAI) và thích ứng miền (Domain Adaptation) cho hệ thống y tế liên kết.

Mục tiêu:
- Xây dựng một mô hình AI có khả năng hiểu và thích ứng với sự thay đổi của miền dữ liệu mà không cần can thiệp thủ công quá nhiều.
- Cung cấp khả năng minh bạch và giải thích các quyết định của AI, giúp người dùng hiểu được cách thức và lý do mà hệ thống đưa ra các quyết định nhất định.
- Kiểm tra hiệu suất của mô hình trong các tình huống thực tế để đảm bảo nó có thể hoạt động hiệu quả trong một loạt các điều kiện ứng dụng.

Phương pháp nghiên cứu:

- Thực hiện các thử nghiệm để đánh giá hiệu năng và độ chính xác của mô hình thông qua các tập dữ liệu khác nhau như ADNI, CheXpert, BRATS,...
- Các metrics dự kiến để đánh giá: ACC, F1-Score, SPE, AUC,...
- Các framework dự kiến: FedAvg, HFL, MLP,...

Kết quả dự kiến:

- Mô hình AI được phát triển có khả năng làm việc hiệu quả trên nhiều miền dữ liệu khác nhau.
- Báo cáo chi tiết về hiệu suất của mô hình trên các bộ dữ liệu thử nghiệm, bao gồm cả đánh giá về độ chính xác và độ tin cậy.
- Cung cấp bằng chứng thực nghiệm cho thấy khả năng của mô hình trong việc giải thích các quyết định của nó trong các tình huống thực tế.

KẾT QUẢ MONG ĐỢI

(Viết kết quả phù hợp với mục tiêu đặt ra, trên cơ sở nội dung nghiên cứu ở trên)

- Tài liệu báo cáo về việc áp dụng AI giải thích và thích ứng miền vào mô hình học máy liên kết ứng dụng trong y tế.
- Báo cáo đánh giá độ hiệu quả, chính xác của mô hình khi áp dụng trên các tập dữ liệu khác nhau.

TÀI LIỆU THAM KHẢO *(Định dạng DBLP)*

[1]. Ana Barragán-Montero, Usman Javaid, Gonzalo Valdés, Dan Nguyen, Paul Desbordes, Benoit Macq, Simon Willems, Laura Vandewinckele, Mats Holmström, Fredrik Löfman, et al.: Artificial intelligence and machine learning for medical imaging: A technology review. Phys. Medica 83 (2021), pp. 242–256

[2]. Veronika Cheplygina, Marleen de Bruijne, Josien P. W. Pluim: Not-so-supervised: A survey of semisupervised, multi-instance, and transfer learning in medical image analysis. Med. Image Anal. 54 (2019), pp. 280–296

[3]. Hongzhi Guan, Meng Liu: Domain adaptation for medical image analysis: A survey. IEEE Trans. Biomed. Eng. 69.3 (2022), pp. 1173–1185

[4]. Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A. W. M. Van Der Laak, Bram Van Ginneken, Clara I. Sánchez: A survey on deep learning in medical image analysis. Med. Image Anal. 42 (2017), pp. 60–88

[5]. Yann LeCun, Yoshua Bengio, Geoffrey Hinton: Deep learning. Nature 521.7553 (2015), pp. 436–444

[6]. Sarunas J Raudys, Anil K Jain: Small sample size effects in statistical pattern recognition: Recommendations for practitioners. IEEE Trans. Pattern Anal. Mach. Intell. 13.3 (1991), pp. 252–264

[7]. Antanas Vabalas, Emma Gowen, Ellen Poliakoff, Alex J Casson: Machine learning algorithm validation with a limited sample size. PLOS ONE 14.11 (2019), pp. 1–20

[8]. US Department of Health and Human Services: HIPAA, 2020, https://www.hhs.gov/hipaa/index.html

[9]. General Data Protection Regulation: GDPR, 2019, https://gdpr-info.eu/

[10]. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, Blaise Agüera y Arcas: Communication efficient learning of deep networks from decentralized data. Artificial Intelligence

and Statistics. PMLR. 2017, pp. 1273–1282
[11]. Keith Bonawitz, Hubert Eichner, Wolfgang Grieskamp, Dzmitry Huba, Alex Ingerman, Vladimir Ivanov, Chloé Kiddon, Jakub Konečný, Stefano Mazzocchi, Brendan McMahan, et al.: Towards federated learning at scale: System design. Proceedings of Machine Learning and Systems. Vol. 1. 2019, pp. 374–388
[12]. Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Keith Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al.: Advances and open problems in federated learning. Found. Trends Mach. Learn. 14.1-2 (2021), pp. 1–210
[13]. Amina Adadi, Mohammed Berrada: Explainable AI for healthcare: From black box to interpretable models. Embedded Systems and Artificial Intelligence. Ed. by Vikrant Bhateja, Suresh Chandra Satapathy, Hassan Satori. Springer, 2020, pp. 327–337
[14]. Guang Yang, Qianhua Ye, Jun Xia: Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A minireview, two showcases and beyond. Inf. Fusion 77 (2022), pp. 29–52
[15]. Andreas Holzinger: The next frontier: AI we can really trust. Proc. Joint European Conf. Machine Learning and Knowledge Discovery in Databases, pp. 427–440
[16]. Andreas Holzinger, Matthias Dehmer, Frank Emmert-Streib, Rita Cucchiara, Isabelle Augenstein, Javier Del Ser, Wojciech Samek, Igor Jurisica, Natalia Díaz-Rodríguez: Information fusion as an integrative cross-cutting enabler to achieve robust, explainable, and trustworthy medical artificial intelligence. Inf. Fusion 79 (2022), pp. 263–278
[17]. Hao Guan, Pew-Thian Yap, Andrea Bozoki, Mingxia Liu: Federated Learning for Medical Image Analysis: A Survey. arXiv:2306.05980 [cs.CV], 2023, https://arxiv.org/abs/2306.05980
[18]. Ahmad Chaddad, Qizong Lu, Jiali Li, Yousef Katib, Reem Kateb, Camel Tanougast, Ahmed Bouridane, Ahmed Abdulkadir: Explainable, Domain-Adaptive, and Federated Artificial Intelligence in Medicine. IEEE/CAA Journal of Automatica Sinica 10(4): 859–876 (2023)