**Towards multi-modal causability with Graph Neural Networks enabling information fusion for explainable AI**

**Abstract**

* Explainable AI (xAI) in medical using GNN with multi-modal input

**Introduction**

* Current medical AI is successful in various task such as classification of skin cancer via CNN. They achieved human expert performance
* AI trong lĩnh vực y tế đã thành công trong nhiều task như phân loại ung thư da bằng CNN đạt độ chính xác ngang bằng với các chuyên gia hay …
* Các phương pháp hoạt động tốt nhất hiện nay đều thiếu khả năng diễn giải. Tính không rõ ràng này được gọi là [black-box](http://black-box.in) in model free stocschastic approaches.
* Vì vậy, các vấn đề này nằm trong lĩnh vực explainable AI như thể hiện tính minh bạch và khả năng giải thích các phương pháp tiếp cận, nguyên nhân thúc đẩy bởi vì liên quan đến các vấn đề pháp lý. Tuy nhiên, trong y tế, còn có các vấn đề như việc tổng hợp nhiều dữ liệu không đồng nhất (multi-modal, đa phương thức). Điển hình là nghiên cứu ung thư hoặc phóng xạ trong đó xử lí nhiều loại đầu vào giúp góp phần quyết định đầu ra của mô hình
* Đồng thời trong y tế, cần phải có nguyên nhân hay nguồn gốc làm cho kết quả cải thiện độ chính xác, gọi là tính nhân quả (causality) và một lời giải thích minh bạch cho những nguyên nhân đó .
* Các quyết định bởi mô hình chỉ có thể là bước trung gian do tầm quan trọng của tính hợp lệ trong y học.
* Trong trường hợp này, điều bắt buộc là các quyết định của AI phải có thể truy nguyên nguồn gốc, các phương thức liên quan giúp các chuyên gia trong lĩnh vực y tế có quyền hiểu, xác nhận hoặc bác bỏ chúng.
* Ở đây, việc sử dụng đồ thị hoặc biểu diễn đồ thị có thể có lợi trong việc diễn giải → mô tả một số các tiếp cận ở phần 3
* Đề xuất ra 3 thách thức trong việc thiết lập cấu trúc học máy đa phương thức và có thể giải thích được trong lĩnh vực y tế. Mỗi giai đoạn trong quy trình có thể thực hiện riêng biệt, nhưng việc tổng hợp lại các cách tiếp cận mang lại lợi ích tối đa.

**Challenge 1: Constructing a multi-modal (MM) embedding space**

* Học trên dữ liệu hợp nhất từ các nguồn khác nhau hoặc các phương thức khác nhau có thể đem lại kết quả tốt hơn bằng việc học trên 1 cấu trúc dữ liệu.
* Thử thách ở đây là thu hẹp khoảng cách về mặt ngữ nghĩa trong kết hợp dữ liệu nhưng các phương pháp hiện nay là tổng hợp lại các đặc trưng không chú trọng khoảng cách hay ngữ nghĩa của chúng.
* Để nắm được những đặc trưng như tính quan hệ, thứ bậc, loại suy. Các hướng giải quyết có khả năng là Joint embedding và Graph Representation Learning - unsupervised way. GRL có thể học bằng supervised nhưng nó sẽ làm cho đầu vào giảm tính tổng quát trừ khi sử dụng các thuật toán tăng dữ liệu (data augmentation). Hơn nữa, các embedding layer trong NN sẽ không mang các khái niệm về sự hiểu biết của con người, thiếu tính nhân quả.

**Challenge 2: Distributed GRL**

* Vì lý do hạn chế bộ nhớ chứa dữ liệu hoặc hiệu suất xử lý, các vấn đề pháp lý trong việc trao đổi dữ liệu, điều này đặc biệt đúng trong lĩnh vực y tế, nơi trao đổi dữ liệu là không được phép và bộ dữ liệu rất lớn.
* Thử thách chính ở phần này là to keep the rep- resentation space aligned across location boundaries — akin to the feature alignment problem in the MM setting (lười dịch)
* Distributed Learning còn có vấn đề không đảm độ trễ giữa các kết nối → Thông tin tổng hợp trong các đồ thị có còn giữ được hiệu suất và độ chính xác hay không ?
* Quy trình được đề xuất: Cải thiện kỹ thuật neighborhood aggregation, lựa chọn và thử nghiệm (mean, max, avg), hay tự động lựa chọn như LSTM, được sử dụng thành công trong GraphSage.
* Tối ưu hoá các kĩ thuật tổng hợp góp phần tối đa hoá hiệu suất và độ chính xác, do đó tránh được các quy tắc logic cứng nhắc áp đặt trong phần tổng hợp và lựa chọn đặc trưng, trong khi có thể thu được các chiến lược hay hướng tiếp cận khác.

**Challenge 3: Explainable AI / Interpretable ML**

* Three prominent explainability methods of convolutional neural networks include: Contrastive Gradient-based (CG) saliency maps, Class Activation Mapping (CAM), and Excitation Backpropagation (EB) as well as their variants: Gradient-weighted CAM (Grad-CAM) and contrastive EB (c-EB) are extended to GCNNs.
* Mô hình GNN có thể được xem là mô hình black-box, thiếu tính minh bạch. Vì vậy số lượng nghiên cứu giúp diễn giải mô hình ngày càng tăng. Trong đó, nghiên cứu được đề xuất của bài báo này là GNN-LRP, được hình thành từ 2 phương pháp là kỹ thuật lan truyền ngược với max-pooling và phân rã Taylor. Ý tưởng tổng thể là xác định đầu vào liên kết với một activation nào đó.
* Một đống công thức + tại sao, lười quá.
* Phương pháp GNN-LRP giải thích dự đoán của mô hình bằng cách attribute các cạnh liên kết từ các node đầu vào đến các node đầu ra.
* LRP được áp dụng thành công trong Mạng kết hợp đồ thị (GCNN) được áp dụng trong phân tích tình cảm văn bản bằng cách phân loại đồ thị, hóa học lượng tử và xử lý hình ảnh

**Towards a integrated vision**