



(12) **BẢN MÔ TẢ GIẢI PHÁP HỮU ÍCH THUỘC BẰNG ĐỘC QUYỀN
GIẢI PHÁP HỮU ÍCH**

(19) **Cộng hòa xã hội chủ nghĩa Việt Nam (VN)
CỤC SỞ HỮU TRÍ TUỆ**

(11)



2-0004050

(51) **G06N 20/00; G16H 50/00**
2021.01

(13) **Y**

(21) 2-2022-00477

(22) 07/11/2022

(45) 25/04/2025 445

(43) 27/05/2024 434

(71) Trường Đại học Y Hà Nội (VN)

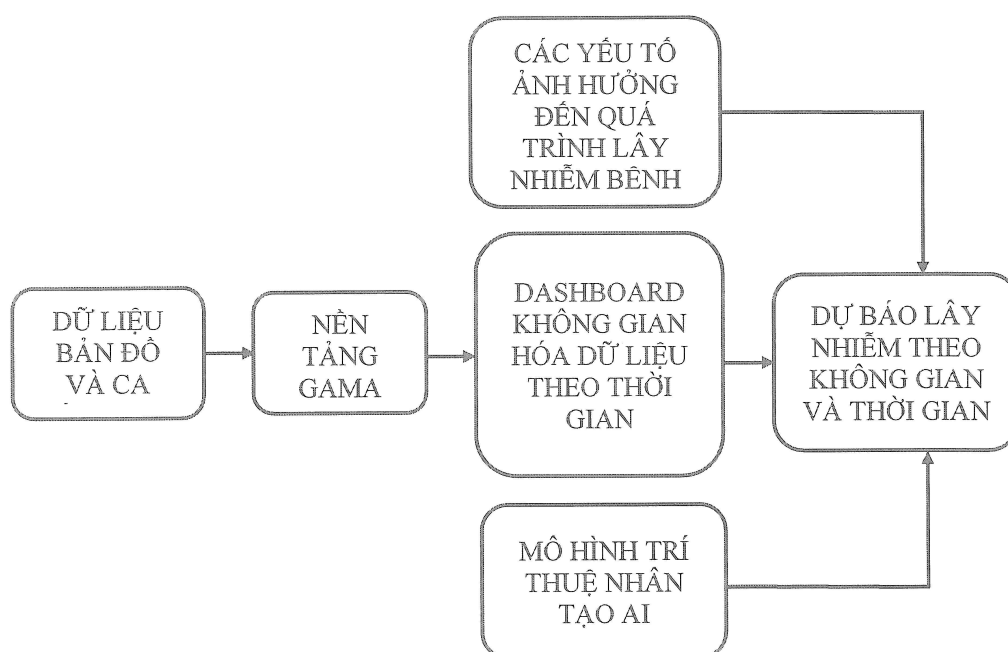
Số 1 Tôn Thất Tùng, phường Trung Tự, quận Đống Đa, Hà Nội

(72) Bùi Mỹ Hạnh (VN); Nguyễn Việt Nhung (VN); Nguyễn Ngọc Doanh (VN).

(54) **PHƯƠNG PHÁP PHÂN TÍCH VÀ DỰ BÁO DỮ LIỆU LỚN BỆNH LAO THEO
KHÔNG GIAN VÀ THỜI GIAN DỰA TRÊN NỀN TẢNG HỖ TRỢ TRỰC QUAN
HÓA**

(21) 2-2022-00477

(57) Giải pháp hữu ích đề cập đến phương pháp phân tích và dự báo dữ liệu lớn bệnh lao theo không gian và thời gian dựa trên nền tảng hỗ trợ trực quan hóa để dự báo xác suất nguy cơ lây nhiễm bệnh lao dựa trên yếu tố cá nhân (tiền sử bệnh tật và các thói quen có hại của bệnh nhân ảnh hưởng trực tiếp đến quá trình phát triển và lây nhiễm của bệnh lao) và các yếu tố môi trường xung quanh (sự tiếp xúc của người bệnh với những người xung quanh thông qua sự di chuyển và tiếp xúc). Trong mô hình mô phỏng, mức độ lây nhiễm sẽ phụ thuộc vào các yếu tố môi trường xung quanh, khả năng lây nhiễm của người bệnh – F0 (hệ số lây nhiễm) và bản thân của những người tiếp xúc – F1 (người có bệnh nền, người già, trẻ em, người có thói quen xấu như hút thuốc lá, uống rượu. Các yếu tố về dân số (mật độ, phân bố, giới tính, độ tuổi,...), các số liệu về kinh tế - xã hội, khí hậu thời tiết, các yếu tố cá nhân (bệnh nền, thói quen xấu,...) được đưa vào mô hình dựa theo mô phỏng AI trên cơ sở học từ bộ dữ liệu đã có. Một số biến sẽ được xây dựng dưới dạng tham số để người sử dụng có thể điều chỉnh theo thời gian thực. Các dữ liệu đầu vào sau khi đi qua mạng LSTM sẽ hội tụ các đặc trưng (attention) và là dữ liệu đầu vào mô hình dịch tễ SEIR.



Hình 1

Lĩnh vực kỹ thuật được đề cập

Giải pháp hữu ích đề cập đến phương pháp phân tích và dự báo dữ liệu lớn bệnh lao theo không gian và thời gian dựa trên nền tảng hỗ trợ trực quan hóa.

Tình trạng kỹ thuật của giải pháp hữu ích

Trí thông minh nhân tạo không gian địa lý (GeoAI) là một ngành khoa học mới nối kết hợp những đổi mới trong khoa học không gian, phương pháp trí tuệ nhân tạo trong học máy, học sâu, khai thác dữ liệu và tính toán hiệu năng cao để trích xuất kiến thức từ dữ liệu lớn về không gian. Khoa học không gian cung cấp các công cụ và công nghệ cho phép hiểu, phân tích và hình dung các hiện tượng trong thế giới thực theo vị trí của chúng. Mục đích của GeoAI là tập trung trong trí thông minh y tế kết hợp vị trí để dùng thông tin diễn biến thực để cải thiện sức khỏe của con người. Các ứng dụng GeoAI phổ biến ở cấp độ quần thể và cá nhân là sử dụng các nguồn dữ liệu lớn mới về không gian như phương tiện xã hội, hồ sơ sức khỏe điện tử, do thám vệ tinh và cảm biến cá nhân để thúc đẩy khoa học về sức khỏe cộng đồng (đặc biệt là trong bối cảnh của 'các thành phố khỏe mạnh thông minh') và y học chính xác một cách tiềm năng, tạo ra những cơ hội mới để trả lời toàn diện hơn các câu hỏi thường được giải quyết trong các lĩnh vực này cũng như các cơ hội duy nhất để trả lời các câu hỏi mới và mới nổi.

Năm 2016 trong Thế vận hội Rio ở Brazil, nền tảng Trí tuệ nhân tạo và Dịch tễ học y tế (AIME) được phát triển ở Malaysia đã sử dụng bộ dữ liệu gồm 272 biến số, bao gồm thông tin như dữ liệu thời tiết, dữ liệu môi trường, dữ liệu di chuyển của con người, dữ liệu so sánh và dữ liệu dịch tễ học để đưa ra dự đoán thời gian thực về dịch Zika và sốt xuất huyết với độ chính xác 84,11%.

Trong dịch tễ học, GeoAI đã được sử dụng để mô tả và phân tích sự phân bố không gian của các bệnh và để nghiên cứu ảnh hưởng của các yếu tố dựa trên vị trí đến kết quả bệnh. Ví dụ, để tạo thuận lợi cho việc tạo ra giả thuyết liên quan đến căn

nguyên của để non, học máy (thuật toán K-means clustering) được sử dụng để xác định mô hình không gian tuổi thai khi sinh cho 145 triệu ca sinh ở hơn 3000 quận ở Hoa Kỳ từ năm 1971 đến 2008 cho tập tin thống kê sức khỏe. Trong một nghiên cứu khác ở Bờ Biển Ngà của Châu Phi, các nhà nghiên cứu nhằm mục đích hiểu rõ hơn về yếu tố quyết định tỷ lệ nhiễm virus suy giảm miễn dịch ở người (HIV) bằng cách sử dụng máy học (hồi quy véc tơ hỗ trợ) để trích xuất dữ liệu di động và dữ liệu kết nối từ dữ liệu điện thoại di động được mã hóa. Những đặc điểm được trích xuất này được phân tích liên quan đến tỷ lệ lưu hành HIV ở cấp bộ, nơi các tác giả nghiên cứu phát hiện ra rằng các yếu tố, như khu vực không gian được người dùng điện thoại và người di cư nói chung, có liên quan đến tỷ lệ hiện mắc HIV.

Sức khỏe môi trường cũng có mối quan hệ chặt chẽ với bệnh học, vì các phơi nhiễm môi trường được đo và/hoặc mô hình hóa có thể được sử dụng để đánh giá phơi nhiễm trong các quần thể nghiên cứu như là một phần của nghiên cứu dịch tễ học môi trường. Ví dụ: một nghiên cứu gần đây đã sử dụng CNN trên hình ảnh của Google Maps Static và Google Place (quan tâm) để trích xuất các yếu tố tự nhiên và đã sửa đổi của môi trường được xây dựng (ví dụ: các tòa nhà, đường băng qua đường, đường phố xanh) để nghiên cứu liên quan đến điều tra dân số-tỷ lệ béo phì ở Hoa Kỳ từ dự án 500 Thành phố của Trung tâm kiểm soát và phòng chống dịch bệnh (CDC). Nghiên cứu này có thể thông báo các can thiệp ở cấp độ vùng để tăng cường hoạt động thể chất và tiếp cận các cửa hàng thực phẩm lành mạnh để giải quyết dịch bệnh béo phì. Một nghiên cứu khác về mặt động lực học đã liên kết các mô hình phơi nhiễm ô nhiễm không khí dựa trên học tập sâu đối với PM2.5 và ozone với Mã ZIP từ dữ liệu lớn của Medicare để kiểm tra mối liên hệ của chúng với nguy cơ tử vong ở Hoa Kỳ. Những ví dụ này minh họa rất nhiều cơ hội cho GeoAI để mô hình hóa các phơi nhiễm môi trường mà sau đó có thể được liên kết với các nguồn dữ liệu khác nhau với thông tin về kết quả bệnh (và các yếu tố gây nhiễu tiềm năng) để tiến hành các nghiên cứu về dịch tễ học.

Hơn nữa, nghiên cứu dịch tễ học trong tương lai có thể khai thác các nguồn dữ liệu lớn về không gian mới nổi để kiểm tra các câu hỏi nghiên cứu về căn nguyên bệnh, có khả năng cung cấp những hiểu biết mới về các yếu tố nguy cơ mới. Ví dụ, cảm biến cá nhân thu thập dữ liệu bằng cách sử dụng các cảm biến được nhúng trong

điện thoại di động cũng như thông qua các thiết bị đeo như Fitbits. Năng lượng không gian là một lĩnh vực tập trung vào việc thu thập dữ liệu độ phân giải không gian cao về năng lượng vị trí và thời gian phù hợp từ GPS, cảm biến gia tốc và GIS để xác định các yếu tố dựa trên không gian có thể liên quan đến việc không hoạt động thể chất và béo phì. GeoAI có thể được sử dụng để xử lý và phân tích các dữ liệu dựa trên vị trí này để xác định loại hoạt động nào tại thời điểm nhất định và phơi nhiễm tại các địa điểm cụ thể cho các loại người khác nhau có liên quan đến kết quả sức khỏe. Các nguồn dữ liệu lớn không gian mới khác bao gồm thông tin từ các dịch vụ chia sẻ đi xe như Uber và Lyft. Có khoảng 5,5 triệu chuyến đi Uber và 1 triệu chuyến đi Lyft được hoàn thành mỗi ngày. Địa điểm đón và trả khách là một khía cạnh quan trọng của việc chia sẻ đi xe vì nó liên quan đến việc có thể được sử dụng để giải quyết các câu hỏi nghiên cứu dịch tễ học liên quan đến chấn thương. Một số nghiên cứu cho thấy các khu vực được đặc trưng bởi việc sử dụng ứng dụng sử dụng xe chung cao hơn có thể liên quan đến tỷ lệ tai nạn liên quan đến giao thông thấp hơn hay việc cung cấp thực phẩm tích hợp dịch vụ sử dụng xe chung (ví dụ: Uber Eats) cũng có thể cung cấp những hiểu biết thú vị về vai trò tiềm năng của nó trong việc cải thiện các hành vi ngồi nhiều và béo phì ở trẻ em.

Trong khoa học xã hội và hành vi, GeoAI đã được sử dụng để giúp xác định các yếu tố quyết định xã hội và hành vi của sức khỏe cũng như tiến hành các can thiệp bằng thông tin định vị. Hồ sơ sức khỏe điện tử (EHRs) là một nguồn dữ liệu lớn có giá trị. EHR cho phép liên kết dữ liệu không gian với các biến địa lý như Mã ZIP khi địa chỉ người bệnh được kiểm tra và cập nhật thường xuyên cho mục đích thanh toán. Ví dụ, phân tích bệnh án điện tử từ Hệ thống Y tế Đại học Duke và Trung tâm Y tế Cộng đồng Lincoln ở Hoa Kỳ đã sử dụng máy học để xác định xem tình trạng kinh tế xã hội cấp độ khu phố có cải thiện dự đoán nguy cơ về kết quả sức khỏe cấp cứu cả ngoại trú và điều trị nội trú. Cục điều tra dân số Hoa Kỳ Dữ liệu khảo sát cộng đồng Mỹ đã được sử dụng để xác định tình trạng kinh tế xã hội cấp khu phố bằng cách tính toán chỉ số tình trạng kinh tế xã hội và chất lượng chăm sóc sức khỏe. Một nghiên cứu khác đã sử dụng học máy để thực hiện can thiệp cho bệnh trầm cảm. Một ứng dụng điện thoại di động đã được phát triển bằng cách sử dụng máy học để dự đoán tâm trạng, cảm xúc, trạng thái nhận thức / động lực của người bệnh, hoạt động, bối

cảnh môi trường và bối cảnh xã hội dựa trên hơn 30 cảm biến điện thoại. Học sâu cũng đã được sử dụng để xác định các yếu tố như thu nhập, sự giàu có, giáo dục dự đoán kết quả sức khỏe bao gồm huyết áp tâm thu, cơ thể chỉ số khối lượng, chu vi vòng eo và độ dài telomere trong nghiên cứu về sức khỏe người cao tuổi.

Học máy (hồi quy véc tơ hỗ trợ) đã được sử dụng để dự báo sốt xuất huyết ở Trung Quốc bằng cách sử dụng dữ liệu về khí hậu, các trường hợp sốt xuất huyết hàng tuần và hàng đợi tìm kiếm của Yahoo. Xác nhận cho thấy dịch bệnh trong 12 tuần trước và đỉnh điểm của đợt bùng phát lớn năm 2014 đã được dự báo chính xác. Một nghiên cứu khác đã phát triển một mô hình học máy có tên là FINDER để phát hiện các bệnh liên quan đến thực phẩm bằng cách sử dụng dữ liệu tìm kiếm và ấn danh trên web của Google, ước tính tỷ lệ những người đã ghé thăm một nhà hàng cụ thể và sau đó đã tìm kiếm các thuật ngữ về ngộ độc thực phẩm (để xác định các nhà hàng đó có khả năng không an toàn).

Một ứng dụng tiềm năng khác của GeoAI cho y học chính xác là thông qua địa kỹ thuật, được sử dụng để mô tả xem xét tầm quan trọng của tiền sử trong chẩn đoán và điều trị bệnh. Các bác sĩ lâm sàng có thể được cung cấp thông tin về phơi nhiễm môi trường của người bệnh, điều này có thể giúp các bác sĩ lâm sàng xác định các yếu tố môi trường có thể ảnh hưởng đến sức khỏe của người bệnh. Các bác sĩ lâm sàng (có quyền truy cập vào lịch sử người bệnh nội trú) cung cấp cho người bệnh thông tin về phơi nhiễm môi trường tiềm ẩn đối với các yếu tố nguy cơ môi trường dựa trên nơi họ sống và làm việc và những phơi nhiễm môi trường này có thể được lấy từ công nghệ GeoAI. Tuy nhiên, các rào cản cho các ứng dụng như vậy bao gồm khả năng chấp nhận lâm sàng, đặc biệt là việc phiên dịch y học chính xác sang chăm sóc lâm sàng và chính sách y tế nói chung đã bị tụt hậu so với nhịp độ khám phá khoa học. Xét nghiệm di truyền có thể dùng để theo dõi và cải thiện sức khỏe cá nhân trong nghiên cứu của GxE phát hiện ra tính nhạy cảm gen đối với các bệnh cụ thể liên quan đến phơi nhiễm môi trường tại địa điểm có thể gây tổn hại đặc biệt đến sức khỏe mỗi cá nhân.

Vi khuẩn *Mycobacterium tuberculosis* thường gặp trong một hộ gia đình hoặc cộng đồng nhỏ vì thời gian tiếp xúc kéo dài đủ để việc nhiễm trùng xảy ra, tạo ra tiềm năng cho các cụm cục bộ phát triển. Tuy nhiên, các cụm lao không gian địa lý không phải lúc nào cũng do sự lây truyền từ người sang người đang diễn ra mà còn có thể là

do tái kích hoạt nhiễm trùng tiềm ẩn ở một nhóm người có các yếu tố nguy cơ chung. Phân tích không gian và xác định các khu vực có tỷ lệ lao (cụm) cao, tiếp theo là đặc tính của các trình điều khiển động lực học trong các cụm này, đã được thúc đẩy để kiểm soát lao mục tiêu và tăng cường sử dụng các công cụ kiểm soát lao hiện có.

Lao khác với các bệnh truyền nhiễm khác theo một số cách có khả năng ảnh hưởng đến cụm không gian rõ ràng. Ví dụ, độ trễ dài và thời gian không thường xuyên kéo dài của nó cho phép di chuyển dân số đáng kể giữa các trường hợp nối tiếp. Do đó, nhiễm trùng vi khuẩn lao ở một vị trí nhất định có thể tiến triển thành dịch bệnh lao ở một khu vực hoàn toàn khác, do đó, các trường hợp có thể không nhất thiết chỉ ra sự lây truyền mạnh mà có thể phản ánh tập hợp các nhóm quần thể có nguy cơ mắc bệnh cao hơn, chẳng hạn như người di cư. Tương tự như vậy, việc lây nhiễm lao xảy ra từ nơi làm việc và các địa điểm tụ họp khác có thể bị quy kết sai do phơi nhiễm dân cư, vì chỉ một thông tin cư trú của cá nhân được ghi lại chính xác trên các tài liệu giám sát bệnh lao trong nhiều môi trường.

Xác định tính không đồng nhất trong phân bố không gian của các trường hợp mắc lao và mô tả các tác nhân kiểm soát có thể giúp hình thành các hành động mục tiêu tích cực.

Phân tích không gian được tiến hành phổ biến nhất trên dữ liệu cho các loại bệnh lao, sau đó chỉ là lao phổi đơn thuần và lao phổi dương tính. Phân tích không gian của lao kháng thuốc và lao đa kháng thuốc rộng rãi đã được báo cáo trong 15 nghiên cứu và một nghiên cứu tương ứng. Các nghiên cứu trước đó có xu hướng mô tả trực quan, trong khi các nghiên cứu trong 2 thập kỷ qua thường kết hợp phát hiện cụm và dự đoán rủi ro. Các thang đo không gian phổ biến bao gồm đường điều tra dân số, quận, mã bưu chính, khu phố, khu vực, khu vực y tế, đô thị, phường...

Bản chất kỹ thuật của giải pháp hữu ích

Giải pháp hữu ích nhằm đề xuất phương pháp phân tích và dự báo dữ liệu lớn bệnh lao theo không gian và thời gian dựa trên nền tảng hỗ trợ trực quan hóa và dự báo sự lây nhiễm của bệnh theo mô hình trí tuệ nhân tạo AI (Artificial intelligence). Nền tảng được xây dựng dựa trên nền tảng GAMA. GAMA là một nền tảng cung cấp môi trường mô hình hóa và mô phỏng hoàn chỉnh để xây dựng không gian mô phỏng.

Các thành phần của nền tảng bao gồm dữ liệu bản đồ và số ca bệnh, nền tảng GAMA cung cấp môi trường mô phỏng đa tác tử, các yếu tố ảnh hưởng đến quá trình lây nhiễm bệnh lao, mô hình trí tuệ nhân tạo AI mô phỏng các yếu tố tác động đến quá trình lây nhiễm, trong đó tất cả các yếu tố ảnh hưởng đến quá trình lây nhiễm đều phải được đưa vào mô phỏng. Mức độ chính xác của mô phỏng sẽ phụ thuộc hoàn toàn vào chất lượng nguồn dữ liệu, dữ liệu càng chi tiết thì mức độ dự báo chính xác càng cao.

Nền tảng này đáp ứng được mục tiêu cung cấp bộ công cụ hỗ trợ dự báo nguy cơ lây nhiễm lao, phân tích, khai thác, trích xuất số liệu lưu trữ phục vụ cho công tác quản lý điều hành bệnh lao cũng như phục vụ các nghiên cứu khoa học liên quan.

Mô tả vắn tắt các hình vẽ

Hình 1 thể hiện sơ đồ các thành phần của nền tảng theo giải pháp hữu ích.

Hình 2 thể hiện cấu trúc của nền tảng GAMA.

Hình 3 là sơ đồ phân tích AI liên kết từ cấp huyện đến cấp tỉnh.

Hình 4 thể hiện các yếu tố tác động đến chỉ số nguy cơ.

Hình 5 minh họa RNN có các vòng lặp.

Hình 6 là diễn giải một mạng RNN.

Hình 7 thể hiện mô-đun lặp lại trong RNN tiêu chuẩn chứa một lớp đơn.

Hình 8 thể hiện mô-đun lặp lại trong một LSTM chứa bốn lớp tương tác.

Hình 9 thể hiện các ký hiệu trong lớp tương tác.

Hình 10 minh họa tầng cổng vào trong LSTM.

Hình 11 thể hiện cập nhật trạng thái mới trong LSTM.

Hình 12 thể hiện xác định yêu cầu đầu ra trong LSTM.

Hình 13 thể hiện các biến thể được thêm vào các cổng.

Hình 14 là sơ đồ tổng quát các yếu tố ảnh hưởng trong mạng LSTM.

Hình 15 minh họa bản đồ không gian dữ liệu bệnh lao cấp tỉnh.

Hình 16 minh họa bản đồ không gian dữ liệu bệnh lao cấp huyện.

Hình 17 minh họa bản đồ không gian dữ liệu bệnh lao cấp xã.

Hình 18 thể hiện bản đồ dự báo bệnh lao theo không gian và thời gian.

Hình 19 là biểu đồ dự báo nguy cơ lây nhiễm theo tháng.

Hình 20 là Biểu đồ dự báo số ca lây nhiễm theo tháng.

Hình 21 minh họa kết quả mô hình dự báo theo AI và ý kiến chuyên gia.

Hình 22 minh họa bảng chấm điểm các yếu tố tác động.

Mô tả chi tiết giải pháp hữu ích

Tiếp theo, phương pháp phân tích và dự báo dữ liệu lớn bệnh lao theo không gian và thời gian dựa trên nền tảng hỗ trợ trực quan hóa theo giải pháp hữu ích sẽ được mô tả chi tiết có dựa vào các hình vẽ kèm theo.

Hình 1 thể hiện sơ đồ các thành phần của nền tảng theo giải pháp hữu ích.

Dữ liệu bản đồ và số ca bệnh: Dữ liệu sử dụng cho mô hình không gian là sự tổ hợp các lớp dữ liệu cơ bản như lớp ranh giới (tỉnh, huyện, xã, vv...) đường giao thông (đường quốc lộ, tỉnh lộ, huyện lộ,...), sông suối, ao hồ, dân cư, thảm phủ, vv... cấu thành nên bản đồ phân theo các khu vực/cấp tỉnh/huyện/xã bao gồm: Dữ liệu bản đồ hành chính, giao thông, dữ liệu ảnh vệ tinh của 63 tỉnh/thành phố; 705 đơn vị hành chính cấp huyện, gồm: 01 thành phố thuộc thành phố trực thuộc trung ương, 80 thành phố thuộc tỉnh, 50 thị xã, 46 quận và 528 huyện; 10.599 đơn vị hành chính cấp xã, bao gồm 1.723 phường, 612 thị trấn và 8.264 xã, trong đó có 352 xã thuộc các thành phố trực thuộc tỉnh, 313 xã thuộc các thị xã và 7.599 xã thuộc các huyện. Các dữ liệu được tổng hợp thu thập từ Cục đo đạc, bản đồ và thông tin địa lý Việt Nam.

Trong khi đó, dữ liệu bệnh lao do Bệnh viện Phổi Trung ương cung cấp từ tháng 01/2016 đến tháng 7/2020. Các trường thông tin đầu vào (input data) được sử dụng để đưa vào cơ sở dữ liệu bao gồm: ID bệnh nhân; Ngày khám bệnh; Địa chỉ: chi tiết đến xã/phường, quận/huyện, tỉnh/thành phố (Các bệnh nhân có địa chỉ chính xác sẽ được cập nhật vị trí chính xác trên bản đồ); Tuổi; Giới tính; Tình trạng HIV; Cách giải quyết; Đối tượng; Ngày bắt đầu điều trị; Ngày kết thúc điều trị; ID kết quả điều trị; Loại kết quả điều trị: Lao đa kháng thuốc, lao đa kháng thuốc (thô). Các dữ liệu đầu vào được xây dựng ở dạng mở, do vậy có thể bổ sung thêm các trường dữ liệu khi cần thiết.

Nền tảng GAMA: GAMA (GIS & Agent-based Modeling Architecture) là hệ nền mô phỏng giúp cho các chuyên gia, các nhà mô hình hóa hay người làm về lĩnh vực khoa học máy tính có một môi trường thực hiện mô phỏng đa tác tử. GAMA được phát triển tại phòng nghiên cứu MSI (Modélisation et Simulation Informatique de systèmes complexes) và được tài trợ bởi IRD (Institut de Recherche pour le Développement). Mục tiêu của dự án là thiết kế một hệ nền chung cho mô hình hóa và

mô phỏng hướng tác tử đối với những hệ thống phức tạp. Một số đặc điểm của GAMA: Có thể sử dụng được cho nhiều ứng dụng khác nhau; Tích hợp ngôn ngữ mô hình hóa (GAML) và môi trường phát triển: giúp người sử dụng (ngay cả đối với người không phải kỹ sư tin học) có thể xây dựng được mô hình một cách dễ dàng và nhanh chóng; Phát triển theo ngôn ngữ JAVA do vậy dễ dàng mở rộng để đưa các dữ liệu mới vào; Công cụ tích hợp để phân tích mô hình: không gian tham số rộng để nghiên cứu và căn chỉnh mô hình; Hỗ trợ phát triển các mô hình phức tạp; Tích hợp trực tiếp dữ liệu địa lý và hệ thống thông tin địa lý với mô hình dựa trên tác tử; Tích hợp cách tiếp cận phương pháp luận để xây dựng mô hình nhiều cấp độ; Tích hợp nhiều công cụ: công cụ hỗ trợ quá trình quyết định sử dụng nhiều tiêu chí, chức năng phân nhóm, công cụ thống kê...; Dễ mở rộng nhờ cấu trúc mở gắn với công nghệ JAVA: khung OSGI và chú thích trong JAVA. Cấu trúc của nền tảng GAMA được thể hiện trên Hình 2.

Các yếu tố ảnh hưởng đến quá trình lây nhiễm bệnh lao: Các nhóm yếu tố ảnh hưởng đến quá trình lây nhiễm bệnh lao gồm:

(1) Nhân tố lây nhiễm (B): Số lượng và phân bố ca bệnh theo thời gian và không gian;

(2) Nhóm yếu tố cá nhân (NCN): Các loại bệnh nền và thói quen xấu làm tăng nguy cơ lây nhiễm bệnh lao;

(3) Nhóm yếu tố di chuyển (NCD): Là sự di chuyển của dân cư trong khu vực hoặc sự di chuyển giữa các khu vực khác nhau. Quá trình di chuyển, tương tác là quá trình phát tán dịch bệnh từ người bệnh sang những người xung quanh;

(4) Nhóm yếu tố khí hậu thời tiết (NKH): Các yếu tố khí hậu thời tiết bao như nhiệt độ, độ ẩm, bốc hơi, bức xạ, số giờ nắng, chế độ mưa, chế độ gió theo các mùa, vv... tùy theo các thời điểm khác nhau mà có tác động thúc đẩy hoặc hạn chế quá trình lây nhiễm bệnh;

(5) Nhóm yếu tố môi trường phát tán (NMT): Môi trường phát tán được hiểu là các yếu tố cơ sở hạ tầng xung quanh bệnh nhân có tác động trực tiếp đến quá trình lây nhiễm như là các điểm có nguy cơ lây nhiễm cao (bệnh viện, khu công nghiệp, trường học, nhà hàng, khách sạn, khu du lịch, khu vực tập trung đông người), mật độ và phân

bố dân cư (đặc trưng theo các khu vực thành thị, nông thôn, vùng núi), các khu vực nhạy cảm như khu vực biên giới;

(6) Nhóm yếu tố kinh tế, xã hội (NKT): Các yếu tố kinh tế - xã hội sẽ ảnh hưởng đến khả năng chi trả của người bệnh cho quá trình khám và điều trị bệnh. Các yếu tố được xem xét đến trong mô phỏng AI bao gồm là bình quân GDP/người, tỷ lệ hộ nghèo khu vực mô phỏng;

(7) Nhóm yếu tố chính sách (các giải pháp xử lý, can thiệp) (NCS): Đây được xem là nhóm các giải pháp điều hành của các cơ quan nhà nước để ứng phó và xử lý khả năng lây lan của dịch bệnh. Các chính sách can thiệp như khoanh vùng xử lý, hỗ trợ điều trị, vv... sẽ ảnh hưởng trực tiếp đến quá trình lây nhiễm bệnh. Nhóm yếu tố chính sách sẽ tác động đến mức độ lan truyền dịch bệnh thông qua các nhóm yếu tố khác. Ví dụ, chính sách khoanh vùng nguy cơ và thực hiện các biện pháp cách ly, phong tỏa sẽ tác động đến số lượng người di chuyển giữa các khu vực và trong nội tại khu vực đó. Các chính sách về an sinh xã hội sẽ tác động đến các yếu tố chi phí khám chữa bệnh hoặc các đối tượng bệnh nhân dễ bị tổn thương (ví dụ nhóm người nghèo), hoặc các chính sách về an toàn bệnh viện, cơ sở y tế sẽ làm giảm rủi ro và nguy cơ nhiễm bệnh ở các điểm có nguy cơ lây nhiễm cao. Như vậy, các chính sách của cơ quan quản lý sẽ tác động đến các yếu tố lây nhiễm và gián tiếp làm giảm số lượng bệnh nhân trong các khu vực có hiệu lực của chính sách. Vì vậy, trong mô hình mô phỏng sự can thiệp của các chính sách pháp luật sẽ được can thiệp vào mô hình thông qua sự thay đổi của các yếu tố tác động. Các yếu tố tác động này sẽ được xây dựng ở dạng tham số để có thể điều chỉnh trong mô hình.

Đối với bệnh lao, mô hình mô phỏng sự phát tán bệnh lao (SIR) được Kermack và McKendrick đưa ra năm 1927. Gần một thế kỷ sau khi ra đời, mô hình này và các biến thể của nó vẫn là những mô hình được sử dụng rộng rãi nhất để mô phỏng quá trình phát tán và lây lan của dịch bệnh. Trong mô hình SIR, cộng đồng dân số được quan tâm được chia thành ba “khu vực”, tương ứng với ba trạng thái có thể đối với bệnh: S (susceptible) - khỏe mạnh và có thể mắc bệnh, I (infectious) - đang bị bệnh và có thể lây cho người khác, R (recovered) - đã khỏi bệnh và được miễn dịch. Các cá thể có thể chuyển từ trạng thái S sang trạng thái I (bị nhiễm bệnh) hoặc từ trạng thái I sang trạng thái R (khỏi bệnh).

Nguyên nhân gây ra sự lan truyền bệnh là sự tiếp xúc của người bệnh với những người xung quanh thông qua sự di chuyển và tiếp xúc. Mức độ lây nhiễm sẽ phụ thuộc vào các yếu tố môi trường xung quanh, khả năng lây nhiễm của người bệnh – F0 (hệ số lây nhiễm) và bản thân của những người tiếp xúc – F1 (người có bệnh nền, người già, trẻ em, người có thói quen xấu như hút thuốc lá, uống rượu,...). Vì vậy, thành phần chuyển động trong mô hình dự báo là số lượng người di chuyển giữa các vùng. Ở quy mô cấp xã dự báo cấp xã sẽ là số lượng người di chuyển giữa các xã theo thời gian. Các bệnh nhân (F0) sẽ di chuyển và tiếp xúc với những người trong khu vực và những người từ khu vực khác đến theo dạng ngẫu nhiên (random). Sự di chuyển và tiếp xúc của những người trong khu vực và những người từ khu vực khác di chuyển đến cũng được mô phỏng theo dạng ngẫu nhiên.

Tổ hợp của các xã sẽ là mô hình dự báo cấp huyện và tổ hợp của mô hình dự báo cấp huyện sẽ là mô hình dự báo cấp tỉnh. Các yếu tố về dân số (mật độ, phân bố, giới tính, độ tuổi,...), các số liệu về kinh tế - xã hội, khí hậu thời tiết, các yếu tố cá nhân (bệnh nền, thói quen xấu,...) được đưa vào mô hình dựa theo mô phỏng AI trên cơ sở học từ bộ dữ liệu đã có. Một số biến sẽ được xây dựng dưới dạng tham số để người sử dụng có thể điều chỉnh theo thời gian thực. Sơ đồ liên kết AI theo phạm vi được minh họa trên Hình 3.

Hình 4 minh họa các nhóm yếu tố tác động đến chỉ số nguy cơ lây nhiễm.

Nhóm yếu tố chính sách sẽ tác động đến mức độ lan truyền dịch bệnh thông qua các nhóm yếu tố khác. Ví dụ, chính sách khoanh vùng nguy cơ và thực hiện các biện pháp cách ly, phong tỏa sẽ tác động đến số lượng người di chuyển giữa các khu vực và trong nội tại khu vực đó. Các chính sách về an sinh xã hội sẽ tác động đến các yếu tố chi phí khám chữa bệnh hoặc các đối tượng bệnh nhân dễ bị tổn thương (ví dụ nhóm người nghèo), hoặc các chính sách về an toàn bệnh viện, cơ sở y tế sẽ làm giảm rủi ro và nguy cơ nhiễm bệnh ở các điểm có nguy cơ lây nhiễm cao. Như vậy, các chính sách của cơ quan quản lý sẽ tác động đến các yếu tố lây nhiễm và gián tiếp làm giảm số lượng bệnh nhân trong các khu vực có hiệu lực của chính sách. Vì vậy, trong mô hình mô phỏng sự can thiệp của các chính sách pháp luật sẽ được can thiệp vào mô hình thông qua sự thay đổi của các yếu tố tác động. Các yếu tố tác động này sẽ được xây dựng ở dạng tham số để có thể điều chỉnh trong mô hình.

Trong mô hình mô phỏng dự báo quá trình lây nhiễm bệnh theo mô hình AI, tất cả các yếu tố ảnh hưởng đến quá trình lây nhiễm đều phải được đưa vào mô phỏng. Mức độ chính xác của mô phỏng sẽ phụ thuộc hoàn toàn vào chất lượng nguồn dữ liệu, dữ liệu càng chi tiết thì mức độ dự báo chính xác càng cao.

Mô hình trí tuệ nhân tạo AI mô phỏng các yếu tố tác động đến quá trình lây nhiễm: Việc mô phỏng dự báo các yếu tố tác động đến nguy cơ lây nhiễm bệnh lao sẽ được mô phỏng bởi mô hình mạng bộ nhớ dài-ngắn (Long Short Term Memory networks), thường được gọi là LSTM - là một dạng đặc biệt của RNN, nó có khả năng học được các phụ thuộc xa. LSTM được giới thiệu bởi Hochreiter & Schmidhuber (1997), và sau đó đã được cải tiến và phổ biến bởi rất nhiều người trong ngành. Chúng hoạt động cực kì hiệu quả trên nhiều bài toán khác nhau nên dần đã trở nên phổ biến như hiện nay. LSTM là một mạng cải tiến của RNN nhằm giải quyết vấn đề nhớ các bước dài của RNN.

Mạng hồi quy RNN

Con người không bắt đầu suy nghĩ của họ từ đầu tại tất cả các thời điểm. Cũng như bạn đang đọc báo cáo này, bạn hiểu mỗi chữ ở đây dựa vào từ bạn đã hiểu các chữ trước đó chứ không phải là đọc tới đâu ném hết đi tới đó, rồi lại bắt đầu suy nghĩ lại từ đầu tới chữ bạn đang đọc. Tức là tư duy đã có một bộ nhớ để lưu lại những gì diễn ra trước đó. Tuy nhiên các mô hình mạng nơ-ron truyền thống thì không thể làm được việc đó, đó có thể coi là một khuyết điểm chính của mạng nơ-ron truyền thống. Ví dụ, bạn muốn phân loại các bối cảnh xảy ra ở tất cả các thời điểm trong một bộ phim, thì đúng là không rõ làm thế nào để có thể hiểu được một tình huống trong phim mà lại phụ thuộc vào các tình huống trước đó nếu sử dụng các mạng nơ-ron truyền thống. Mạng nơ-ron hồi quy (Recurrent Neural Network) sinh ra để giải quyết vấn đề đó. Mạng này chứa các vòng lặp bên trong cho phép thông tin có thể lưu lại được.

Hình 5 minh họa một đoạn của mạng nơ-ron hồi quy A với đầu vào là x_t và đầu ra là h_t . Một vòng lặp cho phép thông tin có thể được truyền từ bước này qua bước này qua bước khác của mạng nơ-ron. Các vòng lặp này khiến cho mạng nơ-ron hồi quy trông có vẻ khó hiểu. Tuy nhiên, nếu bạn để ý một chút thì nó không khác mấy so với các mạng nơ-ron thuần. Một mạng nơ-ron hồi quy có thể được coi là nhiều bản sao

chép của cùng một mạng, trong đó mỗi đầu ra của mạng này là đầu vào của một mạng sao chép khác.

Chuỗi lặp lại các mạng này chính là phân giải của mạng nơ-ron hồi quy, các vòng lặp khiến chúng tạo thành một chuỗi danh sách các mạng sao chép nhau. Các nút của mạng vẫn nhận đầu vào và có đầu ra như mạng nơ-ron thuần như được thể hiện trên Hình 6.

Một điểm nổi bật của RNN chính là ý tưởng kết nối các thông tin phía trước để dự đoán cho hiện tại. Việc này tương tự như ta sử dụng các cảnh trước của bộ phim để hiểu được cảnh hiện thời. Nếu mà RNN có thể làm được việc đó thì chúng sẽ cực kỳ hữu dụng, tuy nhiên liệu chúng có thể làm được không? Câu trả lời là còn tùy trường hợp cụ thể. Đôi lúc ta chỉ cần xem lại thông tin vừa có thôi là đủ để biết được tình huống hiện tại. Ví dụ, ta có câu: “các đám mây trên bầu trời” thì ta chỉ cần đọc tới “các đám mây trên bầu” là đủ biết được chữ tiếp theo là “trời” rồi. Trong tình huống này, khoảng cách tới thông tin có được cần để dự đoán là nhỏ, nên RNN hoàn toàn có thể học được.

Nhưng trong nhiều tình huống ta buộc phải sử dụng nhiều ngữ cảnh hơn để suy luận. Ví dụ, dự đoán chữ cuối cùng trong đoạn: “I grew up in France... I speak fluent French”. Rõ ràng là các thông tin gần (“I speak fluent”) chỉ có phép ta biết được đằng sau nó sẽ là tên của một ngôn ngữ nào đó, còn không thể nào biết được đó là tiếng gì. Muốn biết là tiếng gì, thì ta cần phải có thêm ngữ cảnh “I grew up in France” nữa mới có thể suy luận được. Rõ ràng là khoảng cách thông tin lúc này có thể đã khá xa rồi. Thật không may là với khoảng cách càng lớn dần thì RNN bắt đầu không thể nhớ và học được nữa.

Về mặt lý thuyết, rõ ràng là RNN có khả năng xử lý các phụ thuộc xa (long-term dependencies). Chúng ta có thể xem xét và cài đặt các tham số sao cho khéo là có thể giải quyết được vấn đề này. Tuy nhiên, đáng tiếc trong thực tế RNN có vẻ không thể học được các tham số đó. Vấn đề này đã được khám phá khá sâu bởi Hochreiter (1991) và Bengio, et al. (1994), trong các bài báo của mình, họ đã tìm được nhưng lý do căn bản để giải thích tại sao RNN không thể học được. Vì vậy, sự ra đời của LSTM đã giúp giải quyết vấn đề này.

Trong vài năm gần đây, việc ứng dụng RNN đã đưa ra được nhiều kết quả không thể tin nổi trong nhiều lĩnh vực: nhận dạng giọng nói, mô hình hóa ngôn ngữ, dịch máy, mô tả ảnh,... Danh sách vẫn còn đang được mở rộng tiếp. Đằng sau sự thành công này chính là sự đóng góp của LSTM. LSTM là một dạng đặc biệt của mạng nơ-ron hồi quy, với nhiều bài toán thì nó tốt hơn mạng hồi quy thuần. Hầu hết các kết quả thú vị thu được từ mạng RNN là được sử dụng với LSTM.

Mạng LSTM

Mạng bộ nhớ dài-ngắn (Long Short Term Memory networks), thường được gọi là LSTM - là một dạng đặc biệt của RNN, nó có khả năng học được các phụ thuộc xa. LSTM được giới thiệu bởi Hochreiter & Schmidhuber (1997), và sau đó đã được cải tiến và phổ biến bởi rất nhiều người trong ngành. Chúng hoạt động cực kì hiệu quả trên nhiều bài toán khác nhau nên dần đã trở nên phổ biến như hiện nay. LSTM được thiết kế để tránh được vấn đề phụ thuộc xa (long-term dependency). Việc nhớ thông tin trong suốt thời gian dài là đặc tính mặc định của chúng, chứ ta không cần phải huấn luyện nó để có thể nhớ được. Tức là ngay nội tại của nó đã có thể ghi nhớ được mà không cần bất kì can thiệp nào.

Mọi mạng hồi quy đều có dạng là một chuỗi các mô-đun lặp đi lặp lại của mạng nơ-ron. Với mạng RNN chuẩn, các mô-đun này có cấu trúc rất đơn giản, thường là một tầng *tanh* như được minh họa trên Hình 7.

LSTM cũng có kiến trúc dạng chuỗi như vậy, nhưng các mô-đun trong nó có cấu trúc khác với mạng RNN chuẩn. Thay vì chỉ có một tầng mạng nơ-ron, chúng có tới 4 tầng tương tác với nhau một cách rất đặc biệt.

Hình 8 minh họa mô-đun lặp lại trong một LSTM chứa bốn lớp tương tác. Trong sơ đồ này, mỗi một đường mang một véc-tơ từ đầu ra của một nút tới đầu vào của một nút khác. Các hình trong màu hồng biểu diễn các phép toán như phép cộng véc-tơ chẳng hạn, còn các ô màu vàng được sử dụng để học trong các tầng mạng nơ-ron. Các đường hợp nhau kí hiệu việc kết hợp, còn các đường rẽ nhánh ám chỉ nội dung của nó được sao chép và chuyển tới các nơi khác nhau. Các ký hiệu trong lớp tương tác được thể hiện trên Hình 9.

Chìa khóa của LSTM là trạng thái tế bào (cell state) - chính đường chạy thông ngang phía trên của sơ đồ hình vẽ. Trạng thái tế bào là một dạng giống như băng

truyền. Nó chạy xuyên suốt tất cả các mắt xích (các nút mạng) và chỉ tương tác tuyến tính đôi chút. Vì vậy mà các thông tin có thể dễ dàng truyền đi thông suốt mà không sợ bị thay đổi. LSTM có khả năng bỏ đi hoặc thêm vào các thông tin cần thiết cho trạng thái tế bào, chúng được điều chỉnh cẩn thận bởi các nhóm được gọi là cổng (gate). Các cổng là nơi sàng lọc thông tin đi qua nó, chúng được kết hợp bởi một tầng mạng sigmoid và một phép nhân.

Tầng sigmoid sẽ cho đầu ra là một số trong khoảng $[0,1]$ mô tả có bao nhiêu thông tin có thể được thông qua. Khi đầu ra là 0 thì có nghĩa là không cho thông tin nào qua cả, còn khi là 1 thì có nghĩa là cho tất cả các thông tin đi qua nó. Một LSTM gồm có 3 cổng như vậy để duy trì và điều hành trạng thái của tế bào.

Bước đầu tiên của LSTM là quyết định xem thông tin nào cần bỏ đi từ trạng thái tế bào. Quyết định này được đưa ra bởi tầng sigmoid - gọi là “tầng cổng quên” (forget gate layer). Nó sẽ lấy đầu vào là h_{t-1} và x_t rồi đưa ra kết quả là một số trong khoảng $[0,1]$ cho mỗi số trong trạng thái tế bào C_{t-1} . Đầu ra là 1 thể hiện rằng nó giữ toàn bộ thông tin lại, còn 0 chỉ rằng toàn bộ thông tin sẽ bị bỏ đi.

Quay trở lại với ví dụ mô hình ngôn ngữ dự đoán từ tiếp theo dựa trên tất cả các từ trước đó, với những bài toán như vậy, thì trạng thái tế bào có thể sẽ mang thông tin về giới tính của một nhân vật nào đó giúp ta sử dụng được đại từ nhân xưng chuẩn xác. Tuy nhiên, khi đề cập tới một người khác thì ta sẽ không muốn nhớ tới giới tính của nhân vật nữa, vì nó không còn tác dụng gì với chủ thể mới này.

Bước tiếp theo là quyết định xem thông tin mới nào ta sẽ lưu vào trạng thái tế bào. Việc này gồm 2 phần. Đầu tiên là sử dụng một tầng sigmoid được gọi là “tầng cổng vào” (input gate layer) để quyết định giá trị nào ta sẽ cập nhật. Tiếp theo là một tầng *tanh* tạo ra một véc-tơ cho giá trị mới C_t nhằm thêm vào cho trạng thái. Trong bước tiếp theo, ta sẽ kết hợp 2 giá trị đó lại để tạo ra một cập nhật cho trạng thái. Chẳng hạn với ví dụ mô hình ngôn ngữ của ta, ta sẽ muốn thêm giới tính của nhân vật mới này vào trạng thái tế bào và thay thế giới tính của nhân vật trước đó.

Giờ là lúc cập nhật trạng thái tế bào cũ C_{t-1} thành trạng thái mới C_t . Ở các bước trước đó đã quyết định những việc cần làm, nên giờ ta chỉ cần thực hiện là xong. Ta sẽ nhân trạng thái cũ với f_t để bỏ đi những thông tin ta quyết định quên lúc trước. Sau đó cộng thêm $i_t * C_t$. Trạng thái mới thu được này phụ thuộc vào việc ta quyết định

cập nhập mỗi giá trị trạng thái ra sao. Với bài toán mô hình ngôn ngữ, chính là việc ta bỏ đi thông tin về giới tính của nhân vật cũ, và thêm thông tin về giới tính của nhân vật mới như ta đã quyết định ở các bước trước đó.

Cuối cùng, ta cần quyết định xem ta muốn đầu ra là gì. Giá trị đầu ra sẽ dựa vào trạng thái tế bào, nhưng sẽ được tiếp tục sàng lọc. Đầu tiên, ta chạy một tầng sigmoid để quyết định phần nào của trạng thái tế bào ta muốn xuất ra. Sau đó, ta đưa nó trạng thái tế bào qua một hàm *tanh* để co giá trị nó về khoảng $[-1,1]$, và nhân nó với đầu ra của công sigmoid để được giá trị đầu ra ta mong muốn. Với ví dụ về mô hình ngôn ngữ, chỉ cần xem chủ thể mà ta có thể đưa ra thông tin về một trạng từ đi sau đó. Ví dụ, nếu đầu ra của chủ thể là số ít hoặc số nhiều thì ta có thể biết được dạng của trạng từ đi theo sau nó phải như thế nào.

Quá trình trên được minh họa trên các hình vẽ từ Hình 10 đến Hình 12.

Các biến thể của bộ nhớ dài hạn:

Những thứ ta vừa mô tả ở trên là một LSTM khá bình thường. Nhưng không phải tất cả các LTSM đều giống như vậy. Thực tế, các bài báo về LTSM đều sử dụng một phiên bản hơi khác so với mô hình LTSM chuẩn. Sự khác nhau không lớn, nhưng chúng giúp giải quyết phần nào đó trong cấu trúc của LTSM. Một dạng LTSM phổ biến được giới thiệu bởi Gers & Schmidhuber (2000) được thêm các đường kết nối “peephole connections”, làm cho các tầng công nhận được giá trị đầu vào là trạng thái tế bào.

Một biến thể khác là nối 2 cổng loại trừ và đầu vào với nhau. Thay vì phân tách các quyết định thông tin loại trừ và thông tin mới thêm vào, ta sẽ quyết định chúng cùng với nhau luôn. Ta chỉ bỏ đi thông tin khi mà ta thay thế nó bằng thông tin mới đưa vào. Ta chỉ đưa thông tin mới vào khi ta bỏ thông tin cũ nào đó đi.

Một biến thể khác khá thú vị của LSTM là Gated Recurrent Unit, hay GRU được giới thiệu bởi Cho, et al. (2014). Nó kết hợp các cổng loại trừ và đầu vào thành một cổng “cổng cập nhập” (update gate). Nó cũng hợp trạng thái tế bào và trạng thái ẩn với nhau tạo ra một thay đổi khác. Kết quả là mô hình của ta sẽ đơn giản hơn mô hình LSTM chuẩn và ngày càng trở nên phổ biến.

Trên đây chỉ là một vài biến thể được chú ý nhiều nhất, thực tế có rất nhiều các biến thể khác nhau của LSTM như Depth Gated RNNs của Yao, et al. (2015). Cũng

có những biến thể mà chiến lược xử lý phụ thuộc xa hoàn toàn khác như Clockwork RNNs của Koutnik, et al. (2014). Ngoài ra, Jozefowicz, et al. (2015) thậm chí còn thử hàng chục nghìn kiến trúc RNN khác nhau và tìm ra một vài mô hình hoạt động tốt hơn cả LSTM ở một số bài toán. LSTM là một bước lớn trong việc sử dụng RNN. Ý tưởng của nó giúp cho tất cả các bước của RNN có thể truy vấn được thông tin từ một tập thông tin lớn hơn.

Sử dụng LSTM trong dự báo các yếu tố ảnh hưởng

Các nhóm yếu tố ảnh hưởng đến quá trình lây nhiễm bệnh lao như yếu tố cá nhân, yếu tố di chuyển, yếu tố môi trường - kinh tế xã hội,... sẽ là các thông số đầu vào của mô hình dự báo. Mỗi loại dữ liệu sẽ được gắn vào một lớp tương ứng trong mô hình LSTM và được hiển thị tổng thể trong mô hình được thể hiện trên Hình 14.

Các số liệu của các yếu tố ảnh hưởng đến quá trình lây nhiễm của bệnh lao được đưa vào mô hình với hệ số của các yếu tố được tiến hành như sau:

- Bước 1: Gán cho chúng các giá trị bất kỳ;
- Bước 2: Lập hàm lost;
- Bước 3: Qua mỗi bước chạy với dữ liệu theo thời gian => các hệ số được điều chỉnh theo hướng tối ưu hàm lost;
- Bước 4: Quá trình chạy hết các dữ liệu thì hệ số sẽ hội tụ về một giá trị => đó chính là giá trị tối ưu.

Các dữ liệu đầu vào sau khi đi qua mạng LSTM sẽ hội tụ các đặc trưng (attention) và là dữ liệu đầu vào mô hình dịch tễ SEIR.

Mô phỏng dự báo số ca bệnh

Số ca bệnh lao theo thời gian và không gian sẽ được mô phỏng dự đoán theo mô hình mạng LSTM. Đối với việc dự báo chuỗi dữ liệu bệnh lao, LSTM sẽ học kinh nghiệm từ chuỗi số liệu bệnh lao từ năm 2016 đến năm 2020. Chuỗi dữ liệu hiện trạng được xây dựng ở dạng mở, theo thời gian chuỗi dữ liệu thực tế sẽ được cập nhật vào mô hình để làm tăng tính chính xác của mô hình dự báo. Trong mô hình này, dữ liệu đầu vào ($X_1, X_2, \dots, X_n$) là số lượng các ca bệnh lao trong giai đoạn từ năm 2016 - 2020, vì các số liệu về ca bệnh có vai trò tương đương nên các trọng số ($W_1, W_2, \dots, W_n$) trong trường hợp này có giá trị như nhau bằng 1.

Mô phỏng dự báo các yếu tố cá nhân

Yếu tố cá nhân là các yếu tố bệnh nền và thói quen có hại làm cho hệ thống miễn dịch yếu dẫn đến nguy cơ dễ bị mắc bệnh hoặc khó khăn hơn trong quá trình điều trị. Một hệ thống miễn dịch khỏe mạnh thường chiến đấu thành công với vi khuẩn lao, nhưng cơ thể sẽ không thể phòng thủ hiệu quả nếu sức đề kháng thấp. Một số bệnh và các loại thuốc có thể làm suy yếu hệ thống miễn dịch, bao gồm:

- HIV/AIDS;
- Bệnh tiểu đường;
- Bệnh thận nặng;
- Mắc một số bệnh ung thư;
- Điều trị ung thư, như hóa trị;
- Sử dụng thuốc ức chế miễn dịch như corticoid, hóa chất điều trị ung thư;
- Sử dụng một số loại thuốc dùng để điều trị viêm khớp dạng thấp, bệnh vẩy nến và bệnh Crohn;
- Suy dinh dưỡng;
- Trẻ em và người cao tuổi;
- Sử dụng các loại chất kích thích: Lạm dụng ma túy, rượu bia sẽ khiến cho hệ thống miễn dịch dần bị suy yếu và làm cho cơ thể dễ bị nhiễm bệnh lao;
- Sử dụng thuốc lá: Sử dụng thuốc lá làm tăng đáng kể nguy cơ mắc bệnh lao và tử vong cho người bệnh.

Việc đưa các yếu tố cá nhân này vào mô hình mô phỏng được thực hiện trong mạng LSTM dựa trên chuỗi số liệu đã có. Chuỗi dữ liệu hiện trạng được xây dựng ở dạng mở, theo thời gian chuỗi dữ liệu thực tế sẽ được cập nhật vào mô hình để làm tăng tính chính xác của mô hình dự báo. Các yếu tố sẽ được mô phỏng độc lập trong mô hình dự báo.

Mô hình dự báo chính xác nhất nếu tất cả các yếu tố trên đều được mô phỏng trong mô hình. Do sự hạn chế về mặt thời gian, kinh phí và nguồn dữ liệu yếu tố cá nhân nên trong giai đoạn này mô hình dự báo chỉ đưa vào yếu tố dữ liệu HIV/AIDS, trong khi các yếu tố dữ liệu khác sẽ để ở dạng mở để có thể tiếp tục mở rộng và nâng cấp mô hình mô phỏng trong tương lai.

Mô phỏng dự báo các yếu tố di chuyển

Yếu tố di chuyển là sự di chuyển của dân cư trong khu vực hoặc sự di chuyển giữa các khu vực khác nhau. Quá trình di chuyển, tương tác là quá trình phát tán dịch bệnh từ người bệnh sang những người xung quanh.

Dữ liệu về các phương tiện vận chuyển hành khách được thu thập từ Tổng cục Đường bộ Việt Nam và dữ liệu trích xuất từ một số camera lắp đặt tại một số vị trí nút giao (các camera được tích hợp công nghệ AI để đếm, thống kê số lượng và các loại phương tiện đi qua các nút giao thông). Các dữ liệu điển hình này sẽ được phân tích và tính toán cho các khu vực điển hình để làm cơ sở ước đoán số lượng người di chuyển giữa các khu vực. Như vậy, tổng dân số trong mô hình là thông số đầu vào được nhập theo số dân của từng vùng với phạm vi chi tiết đến cấp huyện. Sự di chuyển của cư dân trong khu vực sẽ được mô phỏng theo dạng ngẫu nhiên, trong khi sự di chuyển của dân cư giữa các khu vực (trong đề tài này là đến phạm vi cấp huyện) được tính là tổng số lượt người di chuyển được ước đoán theo tháng và được dự báo theo mô hình mạng LSTM.

Mô phỏng dự báo các yếu tố khí hậu thời tiết

Yếu tố môi trường được hiểu là các yếu tố cơ sở hạ tầng xung quanh bệnh nhân có tác động trực tiếp đến quá trình lây nhiễm. Các yếu tố trong nhóm này được chia thành 2 loại: (1) nhóm điểm nguy cơ được hiểu là các điểm có nguy cơ lây nhiễm cao như bệnh viện, khu công nghiệp, trường học, nhà hàng, khách sạn, khu du lịch, khu vực tập trung đông người và (2) nhóm nguy cơ theo phân bố đều được thể hiện qua mật độ dân số của khu vực theo đặc trưng các vùng như khu vực thành thị, nông thôn, vùng núi, các khu vực nhạy cảm như khu vực biên giới.

(1) Nhóm điểm nguy cơ: Các điểm nguy cơ này sẽ được đánh giá theo thang điểm từ 1 đến 5 với điểm 1 là nguy cơ rất thấp, điểm 2 là nguy cơ thấp, điểm 3 là nguy cơ trung bình, điểm 4 là nguy cơ cao và điểm 5 là nguy cơ rất cao. Các điểm nguy cơ này sẽ được xác định trên bản đồ không gian.

(2) Nhóm nguy cơ theo phân bố đều: Mật độ dân số sẽ được nhập vào mô hình cho từng khu vực mô phỏng (đến quy mô cấp huyện). Căn cứ theo Quyết định phân cấp đô thị, các khu vực sẽ được chấm điểm nguy cơ theo thang điểm từ 1 đến 5 tương tự như nhóm 1.

Trong mô hình dự báo, yếu tố điểm nguy cơ sẽ thay đổi không đáng kể vì cơ sở hạ tầng thông thường ít có biến động lớn. Trong trường hợp các điểm nguy cơ này thay đổi (ví dụ xây mới bệnh viện) thì việc bổ sung/thay đổi cũng khá đơn giản bằng cách thay đổi trực tiếp các chức năng của điểm nguy cơ trên bản đồ mô phỏng. Về dự báo mức độ thay đổi của mật độ dân số, thông số này sẽ được xác định thông qua kế hoạch phát triển của khu vực.

Mô phỏng dự báo yếu tố kinh tế xã hội

Các yếu tố kinh tế - xã hội sẽ ảnh hưởng đến khả năng chi trả của người bệnh cho quá trình khám và điều trị bệnh. Các yếu tố được xem xét đến trong mô phỏng AI bao gồm là bình quân GDP/người, tỷ lệ hộ nghèo khu vực mô phỏng.

Trong mô hình dự báo, bình quân GDP/người và tỷ lệ hộ nghèo có thể xác định thông 2 cách: (1) Dựa vào các báo cáo định hướng phát triển của từng khu vực, (2) Dựa vào mô hình mạng LSTM: Căn cứ vào dữ liệu hiện trạng của từng khu vực, mô hình mạng LSTM sẽ dự báo xu hướng biến đổi của các thông số này trong tương lai.

Yếu tố chính sách

Đây được xem là nhóm các giải pháp điều hành của các cơ quan nhà nước để ứng phó và xử lý khả năng lây lan của dịch bệnh. Các chính sách can thiệp như khoanh vùng xử lý, hỗ trợ điều trị, vv... sẽ ảnh hưởng trực tiếp đến quá trình lây nhiễm bệnh. Khi đó mô hình sẽ là nơi thử nghiệm các giải pháp để dự báo và đánh giá mức độ hiệu quả của các phương án can thiệp. Cụ thể, mô hình sẽ giúp dự báo các khu vực có nguy cơ lây nhiễm cao để các cơ quan quản lý có thể khoanh vùng các khu vực nguy cơ theo các cấp độ và thực hiện các giải pháp hạn chế sự lây nhiễm. Hiệu quả của các giải pháp sẽ được thể hiện ngay trên mô hình dự báo.

Sự can thiệp của các yếu tố chính sách vào mô hình mô phỏng sẽ được can thiệp thông qua các yếu tố tác động dưới dạng tham số. Ví dụ, giải pháp khoanh vùng, hạn chế đi lại vào các khu vực có nguy cơ cao sẽ được điều chỉnh bằng cách giảm số lượng người di chuyển giữa các khu vực.

Phần trên đã trình bày phương pháp sử dụng mô hình AI để mô phỏng và dự báo sự lan truyền của bệnh lao. Nội dung này sẽ sử dụng phương pháp ý kiến đánh giá của chuyên gia để dự báo sự lan truyền của dịch bệnh.

Tương tự như phương pháp sử dụng mô hình AI, phương pháp chuyên gia cũng phân tích các yếu tố tác động đến sự lan truyền của dịch bệnh bao gồm các nhóm yếu tố: Yếu tố cá nhân, yếu tố di chuyển, yếu tố khí hậu thời tiết, yếu tố môi trường, yếu tố kinh tế xã hội. Tuy nhiên, từng nhóm yếu tố này sẽ được các chuyên gia đánh giá chấm điểm thông qua các chỉ tiêu (mỗi chỉ tiêu bao gồm điểm và trọng số). Điểm (S) và trọng số (W) của từng chỉ tiêu sẽ được đánh giá dựa trên kinh nghiệm của các chuyên gia.

Các tiêu chí được đánh giá theo thang điểm từ 1 đến 5 tương ứng với các mức độ ảnh hưởng từ 1 – Rất thấp, 2 – Thấp, 3 – Trung bình, 4 – Cao, 5 – Rất cao. Mỗi tiêu chí tùy theo mức độ quan trọng sẽ được chấm điểm trọng số từ 1 đến 5 tương ứng với các mức 1 – Rất ít quan trọng, 2 – Ít quan trọng, 3 – Khá quan trọng, 4 – Quan trọng, 5 – Rất quan trọng. Tổng điểm của từng tiêu chí sẽ bằng điểm (S) nhân với trọng số (W). Tổng điểm của từng tiêu chí sẽ nằm trong phạm vi từ 1 đến 25 điểm. Ví dụ, điểm tiêu chí sẽ thấp nhất (1 điểm) nếu S và W của tiêu chí đó đồng thời bằng 1, hoặc điểm tiêu chí sẽ là cao nhất (25 điểm) nếu S và W đồng thời bằng 5. Để dễ quy đổi và đánh giá tổng điểm của từng tiêu chí sẽ được quy đổi về thang điểm 5 bằng cách lấy tổng điểm chia 5, khi đó công thức tính điểm cho từng tiêu chí (I) sẽ được tính theo công thức:

$$I = \frac{S \times W}{5}$$

Với cách tính này điểm đánh giá của từng tiêu chí sẽ nằm dao động trong khoảng từ 0 đến 5. Tương ứng, thang điểm đánh giá mức độ tác động của tiêu chí (I) được quy đổi tương đương như sau:

- $0 < I \leq 1$: Tiêu chí có mức độ tác động rất thấp
- $1 < I \leq 2$: Tiêu chí có mức độ tác động thấp
- $2 < I \leq 3$: Tiêu chí có mức độ tác động trung bình
- $3 < I \leq 4$: Tiêu chí có mức độ tác động cao
- $4 < I \leq 5$: Tiêu chí có mức độ tác động rất cao

Điểm tổng hợp của từng nhóm (IG) sẽ được tính theo công thức trung bình cộng:

$$I_G = \sum \frac{I_1 + I_2 + \dots + I_n}{n}$$

Thang điểm của nhóm yếu tố (I_G) cũng có giá trị dao động từ 0 đến 5. Khi đó, mức độ tác động của các nhóm sẽ được đánh giá tương tự như sau:

- $0 < I_G \leq 1$: Nhóm yếu tố có mức độ tác động rất thấp
- $1 < I_G \leq 2$: Nhóm yếu tố có mức độ tác động thấp
- $2 < I_G \leq 3$: Nhóm yếu tố có mức độ tác động trung bình
- $3 < I_G \leq 4$: Nhóm yếu tố có mức độ tác động cao
- $4 < I_G \leq 5$: Nhóm yếu tố có mức độ tác động rất cao

Các nhóm yếu tố cũng được đánh giá trọng số (WG) theo mức độ quan trọng từ 1 đến 5, tương ứng với các mức 1 – Rất ít quan trọng, 2 – Ít quan trọng, 3 – Khá quan trọng, 4 – Quan trọng, 5 – Rất quan trọng. Khi đó, tổng điểm của khu vực đánh giá (Z) sẽ được tính theo công thức sau:

$$Z = \sum \frac{I_{G1} \times WG_1 + I_{G2} \times WG_2 + \dots + I_{Gn} \times WG_n}{n}$$

Trong đó:

- $I_{G1}, I_{G2}, \dots, I_{Gn}$: điểm của các nhóm yếu tố tác động
- $WG_1, WG_2, \dots, WG_n$: Trọng số của các nhóm yếu tố tác động

Mô hình không gian sẽ hiện thị phổ màu mức độ nguy cơ căn cứ vào điểm đánh giá (Z) của khu vực.

Cụ thể từng tiêu chí trong các nhóm yếu tố tác động được đánh giá như sau:

Yếu tố cá nhân

Các yếu tố cá nhân về bệnh nền và thói quen có hại làm suy giảm hệ thống miễn dịch dẫn đến tăng nguy cơ nhiễm bệnh lao được đưa vào bảng đánh giá gồm:

- HIV/AIDS;
- Bệnh tiểu đường;
- Bệnh thận nặng;
- Mắc một số bệnh ung thư;

- Điều trị ung thư, như hóa trị;
- Sử dụng thuốc ức chế miễn dịch như corticoid, hóa chất điều trị ung thư;
- Sử dụng một số loại thuốc dùng để điều trị viêm khớp dạng thấp, bệnh vẩy nến và bệnh Crohn;
- Suy dinh dưỡng;
- Trẻ em và người cao tuổi;
- Sử dụng các loại chất kích thích: Lạm dụng ma túy, rượu bia sẽ khiến cho hệ thống miễn dịch dần bị suy yếu và làm cho cơ thể dễ bị nhiễm bệnh lao;
- Sử dụng thuốc lá: Sử dụng thuốc lá làm tăng đáng kể nguy cơ mắc bệnh lao và tử vong cho người bệnh.

Các chuyên gia căn cứ vào kinh nghiệm để đánh giá điểm (theo mức độ ảnh hưởng từ thấp đến cao) và trọng số (theo mức độ quan trọng) cho từng loại bệnh nền và thói quen ở trên vào ô tương ứng trong mục ý kiến chuyên gia. Mô hình sẽ tự động tính điểm tổng hợp cho nhóm yếu tố này.

Yếu tố di chuyển

Yếu tố di chuyển gồm dự di chuyển của dân cư trong khu vực và sự di chuyển giữa các khu vực khác nhau. Sự di chuyển của cư dân trong khu vực sẽ được mô phỏng theo dạng ngẫu nhiên, trong khi sự di chuyển của dân cư giữa các khu vực được tính là tổng số lượt người di chuyển được ước đoán theo tháng. Căn cứ vào kết quả phân tích dữ liệu mẫu của một số khu vực điển hình cho thấy, tổng lượng người di chuyển giữa các khu vực có thể giao động từ vài nghìn đến vài trăm nghìn/tháng. Do vậy, trong bảng đánh giá của chuyên gia, số lượt di chuyển được chia thành các mức như sau:

- Dưới 10.000 lượt
- Từ 10.000 đến 20.000 lượt
- Từ 20.000 đến 30.000 lượt
- Từ 30.000 đến 50.000 lượt
- Từ 50.000 đến 70.000 lượt
- Từ 70.000 đến 100.000 lượt
- Trên 100.000 lượt

Các chuyên gia căn cứ vào số lượt người di chuyển giữa các khu vực, dân số, mật độ dân số cũng như các yếu tố về vị trí địa lý (khu vực biên giới, trung tâm hành chính, khu vực du lịch, vv....) để cho điểm và trọng số tương ứng của các khu vực.

Yếu tố khí hậu thời tiết

Việc phân tích từng dữ liệu khí hậu thời tiết như mưa, nhiệt độ, độ ẩm, bốc hơi, chế độ gió, nắng, vv... sẽ rất phức tạp vì các yếu tố thời tiết tác động đồng thời và tương hỗ đến khả năng lây nhiễm bệnh. Vì vậy, mô hình xây dựng lựa chọn đánh giá theo mùa trong năm.

Nếu các chuyên gia không có thông tin hoặc kiến thức sâu về vấn đề này có thể không lựa chọn, khi đó chế độ AI để mô hình mặc định sử dụng dữ liệu từ mô hình mô phỏng AI để tính toán dữ liệu này.

Yếu tố môi trường

Các yếu tố trong nhóm này được chia thành 2 loại: (1) nhóm vị trí có nguy cơ được hiểu là các điểm có nguy cơ lây nhiễm cao như bệnh viện, khu công nghiệp, trường học, nhà hàng, khách sạn, khu du lịch, khu vực tập trung đông người và (2) nhóm nguy cơ theo phân bố đều được thể hiện qua mật độ dân số của khu vực theo đặc trưng các vùng như khu vực thành thị, nông thôn, vùng núi, các khu vực nhạy cảm như khu vực biên giới. Vì vậy, chuyên gia đánh giá theo 2 nhóm như sau:

Nhóm 1: Vị trí có nguy cơ

- Bệnh viện lao phổi
- Bệnh viện khác
- Khu công nghiệp, trường học
- Nhà hàng
- Chung cư, các địa điểm du lịch, công cộng khác

Nhóm 2: Mật độ dân số

- Dưới 500 người/km²
- Từ 500 đến 1000 người/km²
- Từ 1000 từ 2000 người/km²
- Từ 2000 đến 3000 người/km²

- Từ 3000 đến 5000 người/km²
- Từ 5000 đến 10.000 người/km²
- Từ 10.000 đến 20.000 người/km²
- Từ 20.000 đến 30.000 người/km²
- Trên 30.000 người/km²

Các chuyên gia sẽ đánh giá điểm và trọng số căn cứ vào vị trí và mật độ dân số của khu vực đánh giá. Phần mềm sẽ tự động tính điểm cho nhóm yếu tố này.

Nếu các chuyên gia không có thông tin hoặc kiến thức sâu về vấn đề này có thể không lựa chọn, khi đó chế độ AI để mô hình mặc định sử dụng dữ liệu từ mô hình mô phỏng AI để tính toán dữ liệu này.

Yếu tố kinh tế xã hội

Các yếu tố kinh tế - xã hội được xem xét bao gồm là bình quân GDP/người, tỷ lệ hộ nghèo khu vực mô phỏng. Căn cứ vào số liệu bình quân GDP/người và tỷ lệ hộ nghèo của Việt Nam giai đoạn từ năm 2010 - 2020, các số liệu này có thể chia thành các mức sau:

(1) GDP bình quân khu vực

- Dưới 500 USD/năm
- Từ 500 đến 1000 USD/năm
- Từ 1000 đến 2000 USD/năm
- Từ 2000 đến 3000 USD/năm
- Trên 3000 USD/năm

(2) Tỷ lệ hộ nghèo khu vực

- Dưới 1%
- Từ 1 đến 2%
- Từ 2 đến 3%
- Từ 3 đến 4%
- Từ 4 đến 5%
- Từ 5 đến 6%
- Từ 6 đến 7%

- Từ 7 đến 8%
- Trên 8%

Các chuyên gia sẽ đánh giá điểm và trọng số căn cứ vào số liệu bình quân GDP/người và tỷ lệ hộ nghèo của khu vực đánh giá. Phần mềm sẽ tự động tính điểm cho nhóm yếu tố này.

Nếu các chuyên gia không có thông tin hoặc kiến thức sâu về vấn đề này có thể không lựa chọn, khi đó chế độ AI để mô hình mặc định sử dụng dữ liệu từ mô hình mô phỏng AI để tính toán dữ liệu này.

Chi tiết bảng chấm điểm từng tiêu chí được thể hiện trong bảng dưới đây.

Bảng 1. Chấm điểm các yếu tố ảnh hưởng

TT	Yếu tố ảnh hưởng đến lây nhiễm bệnh lao	Điểm đánh giá					Trọng số					Điểm tiêu chí (I) - Mức độ tác động				
		Rất thấp	Thấp	Trung bình	Cao	Rất cao	Rất ít quan trọng	Ít quan trọng	Khá quan trọng	Quan trọng	Rất quan trọng	Rất thấp	Thấp	Trung bình	Cao	Rất cao
		1	2	3	4	5	1	2	3	4	5	0<I≤1	1<I≤2	2<I≤3	3<I≤4	4<I≤5
I	Yếu tố cá nhân người bệnh															
1	HIV/AIDS															
2	Bệnh tiểu đường															
3	Bệnh thận nặng															
4	Mắc một số bệnh ung thư															

TT	Yếu tố ảnh hưởng đến lây nhiễm bệnh lao	Điểm đánh giá					Trọng số					Điểm tiêu chí (I) - Mức độ tác động				
		Rất thấp	Thấp	Trung bình	Cao	Rất cao	Rất ít quan trọng	Ít quan trọng	Khá quan trọng	Quan trọng	Rất quan trọng	Rất thấp	Thấp	Trung bình	Cao	Rất cao
		1	2	3	4	5	1	2	3	4	5	0<I≤1	1<I≤2	2<I≤3	3<I≤4	4<I≤5
5	Điều trị ung thư, như hóa trị															
6	Sử dụng thuốc ức chế miễn dịch như corticoid, hóa chất điều trị ung thư															
7	Sử dụng một số loại thuốc dùng để điều trị viêm khớp dạng thấp, bệnh vẩy nến và bệnh															

TT	Yếu tố ảnh hưởng đến lây nhiễm bệnh lao	Điểm đánh giá					Trọng số					Điểm tiêu chí (I) - Mức độ tác động				
		Rất thấp	Thấp	Trung bình	Cao	Rất cao	Rất ít quan trọng	Ít quan trọng	Khá quan trọng	Quan trọng	Rất quan trọng	Rất thấp	Thấp	Trung bình	Cao	Rất cao
		1	2	3	4	5	1	2	3	4	5	0<I≤1	1<I≤2	2<I≤3	3<I≤4	4<I≤5
	Crohn															
8	Suy dinh dưỡng															
9	Trẻ em và người cao tuổi															
10	Nghiện rượu															
11	Nghiện thuốc lá															
12	Sử dụng các loại chất kích thích: Ma túy,...															

TT	Yếu tố ảnh hưởng đến lây nhiễm bệnh lao	Điểm đánh giá					Trọng số					Điểm tiêu chí (I) - Mức độ tác động				
		Rất thấp	Thấp	Trung bình	Cao	Rất cao	Rất ít quan trọng	Ít quan trọng	Khá quan trọng	Quan trọng	Rất quan trọng	Rất thấp	Thấp	Trung bình	Cao	Rất cao
		1	2	3	4	5	1	2	3	4	5	0<I≤1	1<I≤2	2<I≤3	3<I≤4	4<I≤5
II	Yếu tố di chuyển															
1	Dưới 10.000 lượt															
2	10.000 - 20.000 lượt															
3	20.000 - 30.000 lượt															
4	30.000 - 50.000 lượt															
5	50.000 - 70.000 lượt															
6	70.000 - 100.000 lượt															

TT	Điểm đánh giá	Trọng số					Điểm tiêu chí (I) - Mức độ tác động				
		Rất thấp	Thấp	Trung bình	Cao	Rất cao	Rất thấp	Thấp	Trung bình	Cao	Rất cao
		1	2	3	4	5	0<I≤1	1<I≤2	2<I≤3	3<I≤4	4<I≤5
Yếu tố ảnh hưởng đến lây nhiễm bệnh lao											
7	Trên 100.000 lượt										
III	Yếu tố khí hậu thời tiết										
1	Tháng 2,3,4										
2	Tháng 5,6,7										
3	Tháng 8,9,10										
4	Tháng 11,12,1										
IV	Yếu tố môi trường										

TT	Yếu tố ảnh hưởng đến lây nhiễm bệnh lao	Điểm đánh giá					Trọng số					Điểm tiêu chí (I) - Mức độ tác động				
		Rất thấp	Thấp	Trung bình	Cao	Rất cao	Rất ít quan trọng	Ít quan trọng	Khá quan trọng	Quan trọng	Rất quan trọng	Rất thấp	Thấp	Trung bình	Cao	Rất cao
		1	2	3	4	5	1	2	3	4	5	0<I≤1	1<I≤2	2<I≤3	3<I≤4	4<I≤5
4.1	Vị trí có nguy cơ cao															
1	Bệnh viện lao phổi															
2	Bệnh viện khác															
3	Khu công nghiệp, trường học															
4	Nhà hàng															
5	Chung cư, các địa điểm du lịch, công cộng khác															

TT	Yếu tố ảnh hưởng đến lây nhiễm bệnh lao	Điểm đánh giá					Trọng số					Điểm tiêu chí (I) - Mức độ tác động				
		Rất thấp	Thấp	Trung bình	Cao	Rất cao	Rất ít quan trọng	Ít quan trọng	Khá quan trọng	Quan trọng	Rất quan trọng	Rất thấp	Thấp	Trung bình	Cao	Rất cao
		1	2	3	4	5	1	2	3	4	5	0<I≤1	1<I≤2	2<I≤3	3<I≤4	4<I≤5
4.2	Mật độ dân số khu vực															
1	Dưới 500 người/km2															
2	500-1000 người/km2															
3	1000-2000 người/km2															
4	2000-3000 người/km2															
5	3000-5000 người/km2															
6	5000-10.000 người/km2															

TT	Yếu tố ảnh hưởng đến lây nhiễm bệnh lao	Điểm đánh giá					Trọng số					Điểm tiêu chí (I) - Mức độ tác động				
		Rất thấp	Thấp	Trung bình	Cao	Rất cao	Rất ít quan trọng	Ít quan trọng	Khá quan trọng	Quan trọng	Rất quan trọng	Rất thấp	Thấp	Trung bình	Cao	Rất cao
		1	2	3	4	5	1	2	3	4	5	0<I≤1	1<I≤2	2<I≤3	3<I≤4	4<I≤5
7	10.000-20.000 người/km ²															
8	20.000-30.000 người/km ²															
9	Trên 30.000 người/km ²															
V	Yếu tố kinh tế xã hội															
5.1	GDP bình quân khu vực															
1	Dưới 500 USD/năm															

TT	Yếu tố ảnh hưởng đến lây nhiễm bệnh lao	Điểm đánh giá					Trọng số					Điểm tiêu chí (I) - Mức độ tác động				
		Rất thấp	Thấp	Trung bình	Cao	Rất cao	Rất ít quan trọng	Ít quan trọng	Khá quan trọng	Quan trọng	Rất quan trọng	Rất thấp	Thấp	Trung bình	Cao	Rất cao
		1	2	3	4	5	1	2	3	4	5	0<I≤1	1<I≤2	2<I≤3	3<I≤4	4<I≤5
2	500-1000USD															
3	1000-2000USD															
4	2000-3000USD															
5	Trên 3000USD															
5.2	Tỷ lệ hộ nghèo khu vực															
1	Dưới 1%															
2	1-2%															

TT	Yếu tố ảnh hưởng đến lây nhiễm bệnh lao	Điểm đánh giá					Trọng số					Điểm tiêu chí (I) - Mức độ tác động				
		Rất thấp	Thấp	Trung bình	Cao	Rất cao	Rất ít quan trọng	Ít quan trọng	Khá quan trọng	Quan trọng	Rất quan trọng	Rất thấp	Thấp	Trung bình	Cao	Rất cao
		1	2	3	4	5	1	2	3	4	5	0<I≤1	1<I≤2	2<I≤3	3<I≤4	4<I≤5
3	2-3%															
4	3-4%															
5	4-5%															
6	5-6%															
7	6-7%															
8	7-8%															
9	Trên 8%															

Bản đồ không gian dữ liệu bệnh lao được thể hiện trên các hình vẽ từ Hình 15 đến Hình 17.

Phần bên trái màn hình với giao diện dạng bản đồ thể hiện số ca nhiễm theo phổ màu. Trên giao diện này, người sử dụng có thể chọn chế độ quan sát toàn màn hình (ấn vào biểu tượng ) , chế độ quan sát cấp độ tỉnh bằng cách ấn vào biểu tượng chữ T, chế độ quan sát cấp độ huyện bằng cách ấn vào biểu tượng chữ H, chế độ quan sát cấp độ xã bằng cách ấn vào biểu tượng chữ X. Ngoài ra, người sử dụng có thể nhấn vào nút Home để chở lại màn hình chính và các thao tác phóng to hay thu nhỏ màn hình quan sát khi nhấn vào biểu tượng + và -.

Màn hình bên phải là các biểu đồ phân tích số ca nhiễm bệnh như là: Tỷ lệ ca nhiễm trên số dân, số ca nhiễm theo khu vực, số ca và tỷ lệ theo thời gian, phân bố lao đa kháng thuốc, phân bố lao đa kháng thuốc thô, phân bố bệnh lao theo giới tính độ tuổi, phân bố bệnh lao theo kết quả điều trị, phân bố tình trạng HIV của Việt Nam. Tất cả các phân tích này đều được hiển thị theo các cấp độ: cả nước, tỉnh/thành phố, huyện, xã.

Các thanh trạng thái: Người dùng có thể lựa chọn quan sát số ca bệnh theo thời gian bằng cách kéo thời gian trên thanh thời gian. Lựa chọn quan sát theo tỉnh/thành, quận/huyện, phường/xã và năm bằng cách lựa chọn theo danh sách đã được tích hợp sẵn trong hệ thống.

Kết quả mô hình AI dự báo nguy cơ lây nhiễm bệnh lao sẽ được hiển thị tại một tab độc lập với tab hiện trạng. Tab này sẽ gồm có bản đồ dự báo theo thời gian (bằng cách kéo chọn khoảng thời gian phù hợp trên thanh thời gian) khi đó nguy cơ lây nhiễm sẽ được hiển thị đến cấp huyện (màn hình bên trái). Trong khi đó màn hình bên phải sẽ hiển thị biểu đồ dự báo chỉ số nguy cơ theo tháng và dự báo số ca nhiễm theo tháng.

Giao diện ở mục này có cấu trúc giống như mục hiện trạng, nhưng các dữ liệu hiển thị là dữ liệu dự báo. Người dùng lựa chọn thời gian trên thanh thời gian để xem dự báo mức độ lây nhiễm của bệnh lao theo thời gian, khi đó tất cả các dữ liệu trên bản đồ và các biểu đồ, bảng biểu thống kê sẽ được tự động cập nhật theo thời gian lựa chọn trên thanh thời gian. Bản đồ không gian dữ liệu bệnh lao (màn hình bên trái): Hiển thị không gian dữ liệu bệnh lao theo các cấp độ tỉnh, huyện, xã.

Màn hình bên phải sẽ hiển thị chỉ số nguy cơ và số ca bệnh theo tháng. Đường màu đỏ là số liệu hiện trạng được nhập vào mô hình, trong khi đó đường màu xanh là kết quả dự báo (kết quả hiển thị của mô hình dự báo theo AI) như được minh họa trên Hình 19 và Hình 20.

Phân tích sâu

Mục này sẽ so sánh kết quả mô hình dự báo theo hai cách: Theo mô phỏng AI và theo ý kiến chuyên gia. Kết quả hiển thị sẽ so sánh dự báo theo hai cách này trên cùng một biểu đồ.

1. Thanh trạng thái: Không thay đổi so với tab hiện trạng bệnh lao
2. Bản đồ không gian dữ liệu bệnh lao (màn hình bên trái): Hiển thị không gian dữ liệu bệnh lao theo các cấp độ tỉnh, huyện, xã.
3. Dữ liệu thống kê (màn hình bên phải)

Hiển thị biểu đồ kết quả mô hình dự báo theo AI (màu xanh nước biển) và ý kiến chuyên gia (màu xanh lá cây) như được minh họa trên Hình 21.

Đánh giá chuyên gia

Như được minh họa trên Hình 22, mục này hiển thị bảng thông số các yếu tố tác động đến quá trình lây nhiễm của bệnh lao. Các chuyên gia sẽ cho điểm cho từng yếu tố tác động (yếu tố nào không được lựa chọn phần mềm sẽ tự động tính theo mô hình dự báo AI). Dựa trên điểm số chuyên gia đánh giá (điểm và trọng số), phần mềm sẽ tự động tính toán, tổng hợp và đưa ra kết quả dự báo hiển thị trên bản đồ ở màn hình bên trái.

1. Thanh trạng thái: Không thay đổi so với tab hiện trạng bệnh lao
2. Bản đồ không gian dữ liệu bệnh lao (màn hình bên trái): Hiển thị không gian dữ liệu bệnh lao theo các cấp độ tỉnh, huyện, xã.
3. Dữ liệu thống kê (màn hình bên phải)

Từ viết tắt

WebGIS

GeoAI Trí thông minh nhân tạo không gian

B Nhân tố lây nhiễm

N _{CN}	Nhóm yếu tố cá nhân
N _{KH}	Nhóm yếu tố khí hậu thời tiết
N _{MT}	Nhóm yếu tố môi trường phát tán
N _{KT}	Nhóm yếu tố kinh tế, xã hội
N _{CS}	Nhóm yếu tố chính sách (các giải pháp xử lý, can thiệp)

YÊU CẦU BẢO HỘ

1. Phương pháp phân tích và dự báo dữ liệu lớn bệnh lao theo không gian và thời gian dựa trên nền tảng hỗ trợ trực quan hóa bao gồm các bước:

thu thập dữ liệu bản đồ và số ca bệnh là sự tổ hợp các lớp dữ liệu cơ bản như lớp ranh giới (tỉnh, huyện, xã) đường giao thông (đường quốc lộ, tỉnh lộ, huyện lộ), sông suối, ao hồ, dân cư, thảm phủ cấu thành nên bản đồ phân theo các khu vực/cấp tỉnh/huyện/xã gồm dữ liệu bản đồ hành chính, giao thông, dữ liệu ảnh vệ tinh của 63 tỉnh/thành phố; 705 đơn vị hành chính cấp huyện, gồm: 01 thành phố thuộc thành phố trực thuộc trung ương, 80 thành phố thuộc tỉnh, 50 thị xã, 46 quận và 528 huyện; 10.599 đơn vị hành chính cấp xã, bao gồm 1.723 phường, 612 thị trấn và 8.264 xã, trong đó có 352 xã thuộc các thành phố trực thuộc tỉnh, 313 xã thuộc các thị xã và 7.599 xã thuộc các huyện; dữ liệu bệnh lao do Bệnh viện Phổi Trung ương cung cấp từ tháng 01/2016 đến tháng 7/2020, các trường thông tin đầu vào (input data) được sử dụng để đưa vào cơ sở dữ liệu bao gồm: ID bệnh nhân; Ngày khám bệnh; Địa chỉ: chi tiết đến xã/phường, quận/huyện, tỉnh/thành phố (Các bệnh nhân có địa chỉ chính xác sẽ được cập nhật vị trí chính xác trên bản đồ); Tuổi; Giới tính; Tình trạng HIV; Cách giải quyết; Đối tượng; Ngày bắt đầu điều trị; Ngày kết thúc điều trị; ID kết quả điều trị; Loại kết quả điều trị: Lao đa kháng thuốc, lao đa kháng thuốc (thô); các dữ liệu đầu vào được xây dựng ở dạng mở, do vậy có thể bổ sung thêm các trường dữ liệu khi cần thiết;

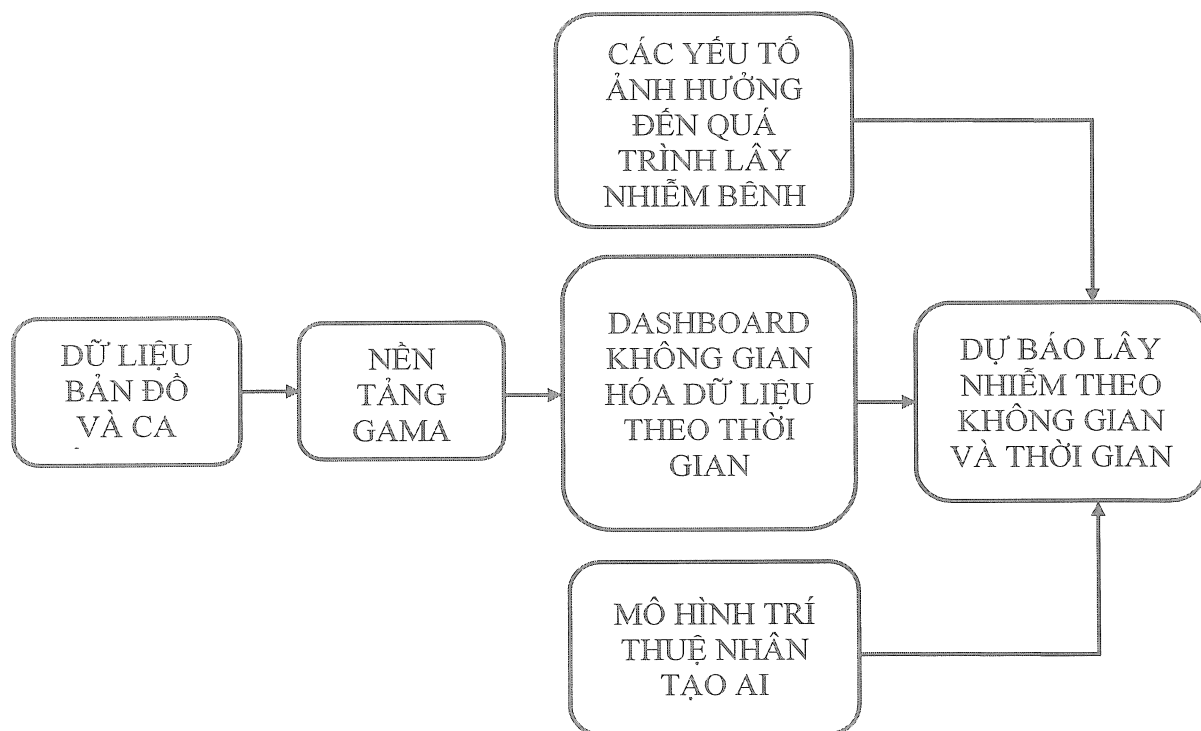
sử dụng nền tảng GAMA: GAMA (GIS & Agent-based Modeling Architecture) thực hiện mô phỏng đa tác tử;

sử dụng mô hình trí tuệ nhân tạo AI mô phỏng các yếu tố tác động đến quá trình lây nhiễm sử dụng mô hình mạng bộ nhớ dài-ngắn (Long Short Term Memory networks) để mô phỏng dự báo các yếu tố tác động đến nguy cơ lây nhiễm bệnh lao, trong đó các nhóm yếu tố ảnh hưởng cơ bản gồm: nhóm yếu tố lây nhiễm (B), nhóm yếu tố cá nhân (NCN), nhóm yếu tố di chuyển (NCD), nhóm yếu tố khí hậu thời tiết (NKH), nhóm yếu tố môi trường phát tán (NMT), nhóm yếu tố kinh tế, xã hội (NKT), nhóm yếu tố chính sách (các giải pháp xử lý, can thiệp) (NCS); các nhóm yếu tố ảnh hưởng đến quá trình lây nhiễm là các thông số đầu vào của mô hình dự báo, mỗi loại dữ liệu sẽ được gắn vào một lớp tương ứng trong mô hình LSTM, các số liệu của các

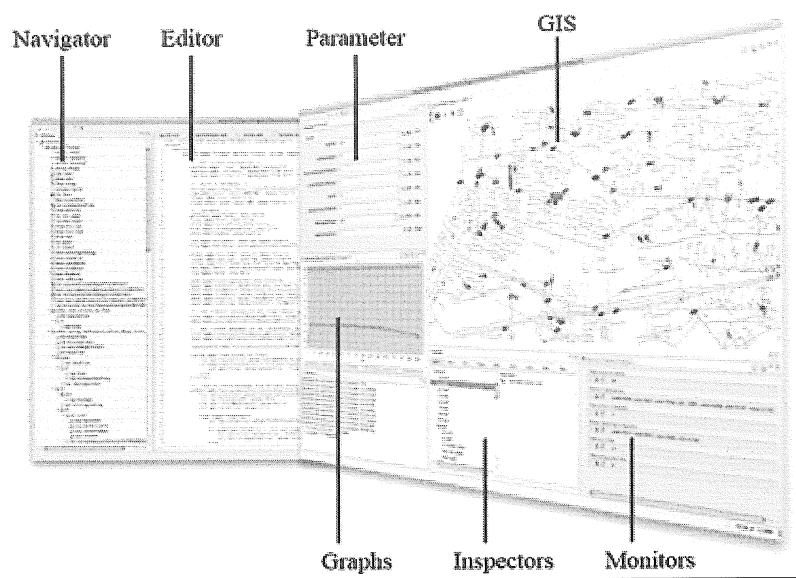
yếu tố ảnh hưởng đến quá trình lây nhiễm của bệnh lao được đưa vào mô hình với hệ số của các yếu tố được tiến hành như sau:

- Bước 1: gán cho chúng các giá trị bất kỳ;
- Bước 2: lập hàm lost;
- Bước 3: qua mỗi bước chạy với dữ liệu theo thời gian, các hệ số được điều chỉnh theo hướng tối ưu hàm lost;
- Bước 4: quá trình chạy hết các dữ liệu thì hệ số sẽ hội tụ về một giá trị, đó chính là giá trị tối ưu;

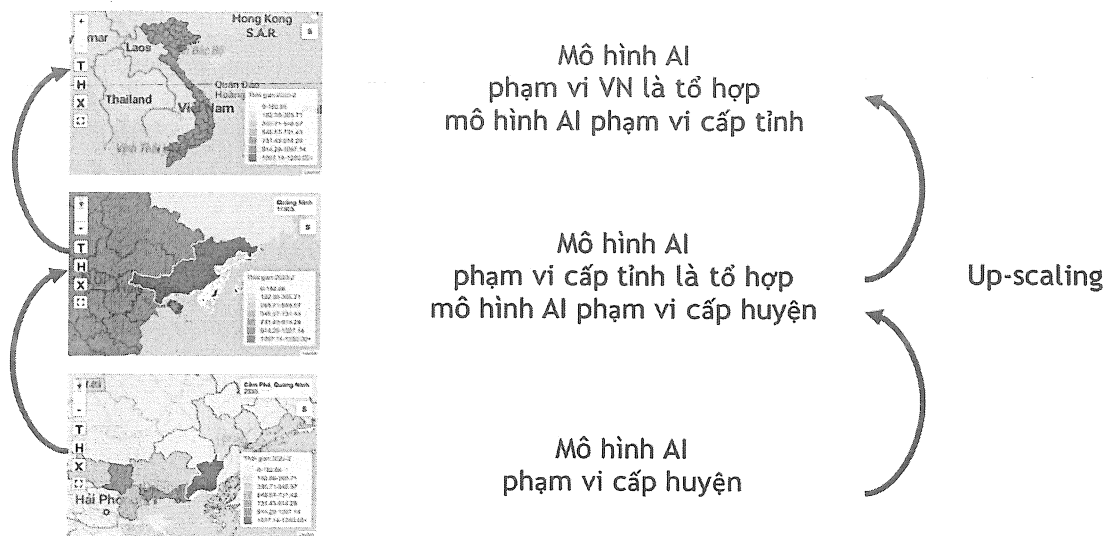
các dữ liệu đầu vào sau khi đi qua mạng LSTM sẽ hội tụ các đặc trưng (attention) và là dữ liệu đầu vào mô hình dịch tễ SEIR.



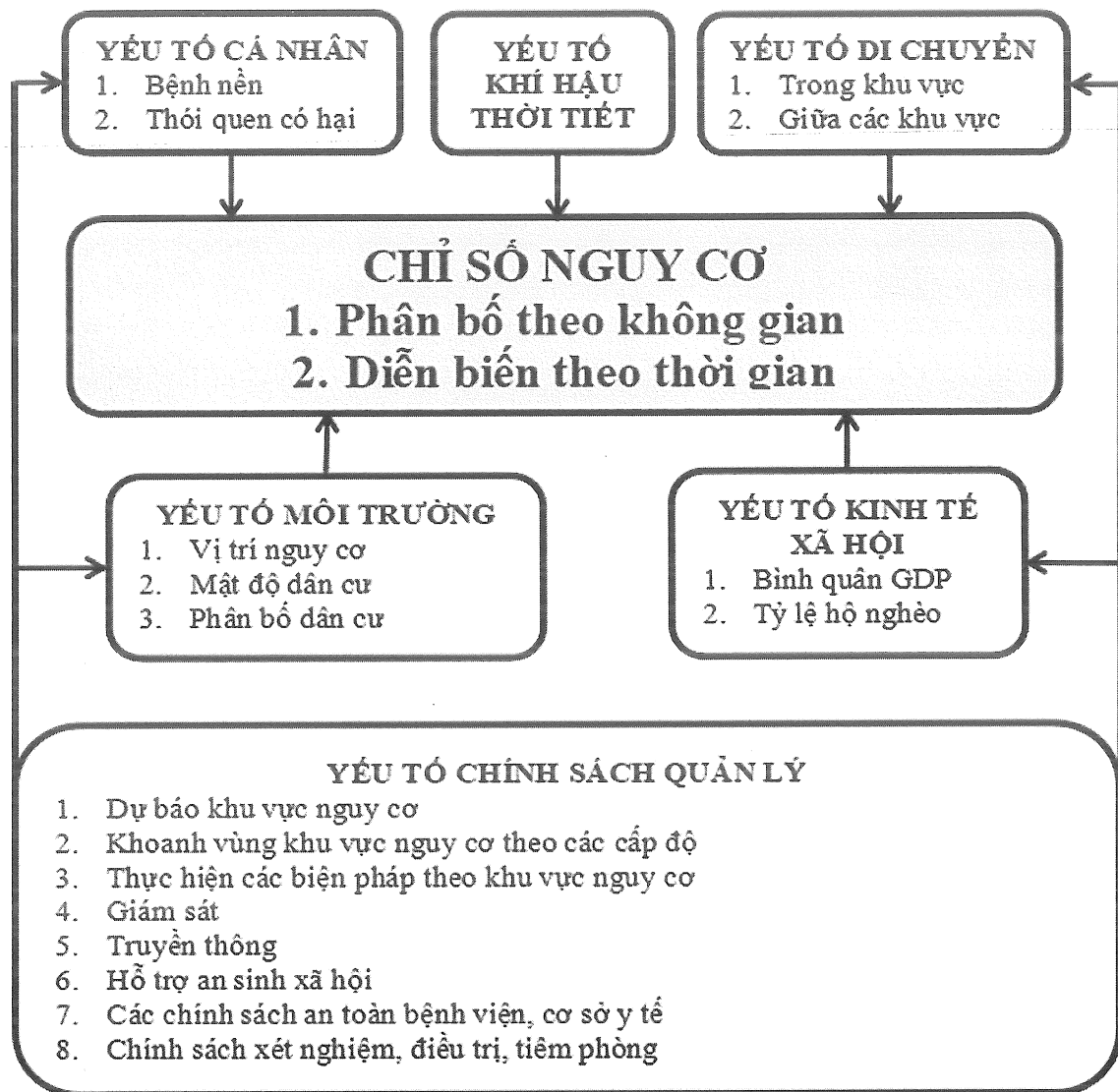
Hình 1



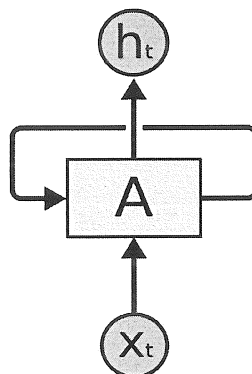
Hình 2



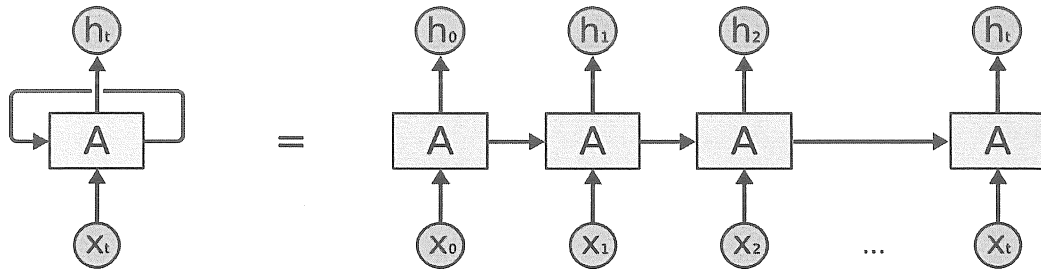
Hình 3



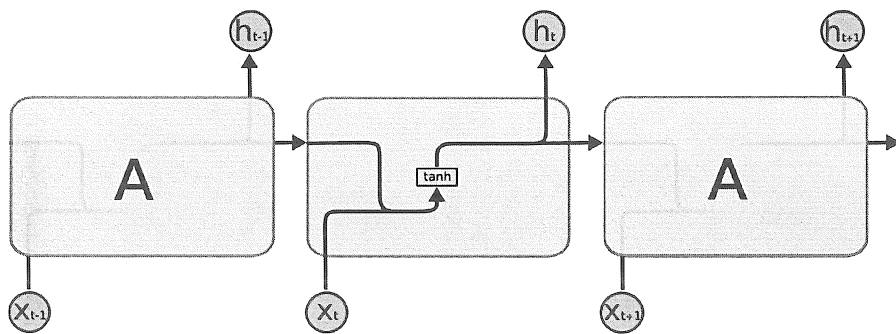
Hình 4



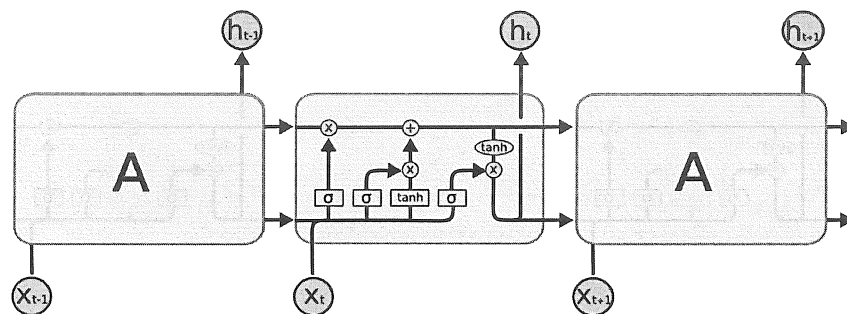
Hình 5



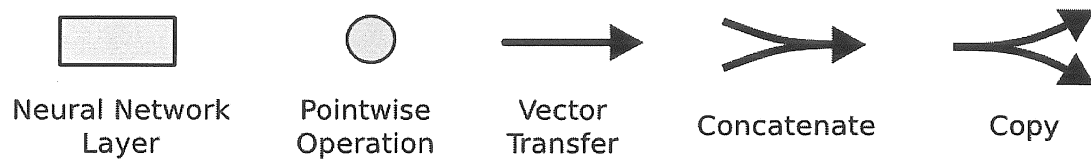
Hình 6



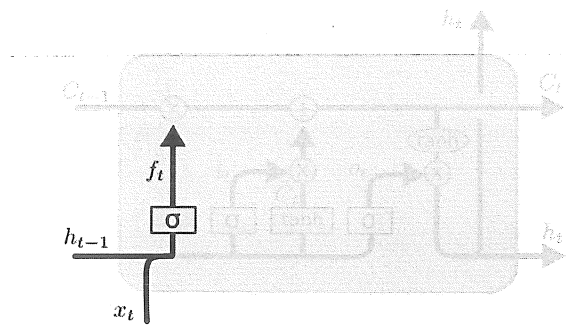
Hình 7



Hình 8

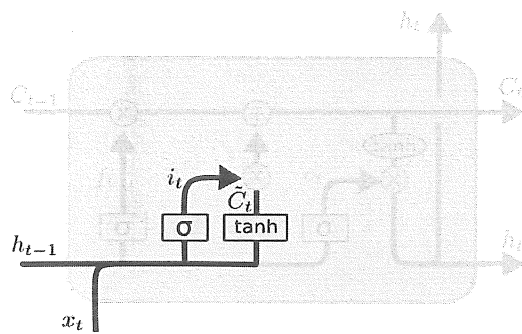


Hình 9



$$f_t = \sigma (W_f \cdot [h_{t-1}, x_t] + b_f)$$

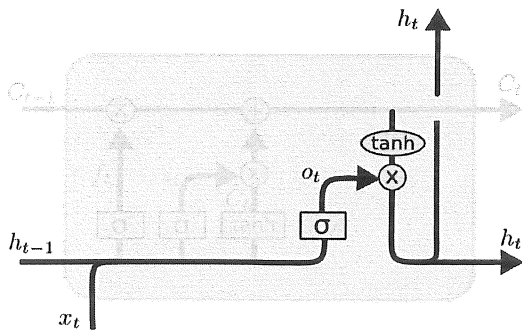
Hình 10



$$i_t = \sigma (W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

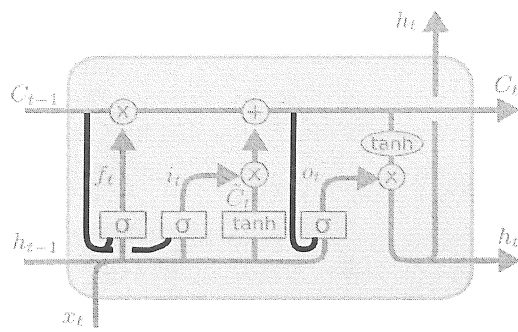
Hình 11



$$o_t = \sigma (W_o [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh (C_t)$$

Hình 12

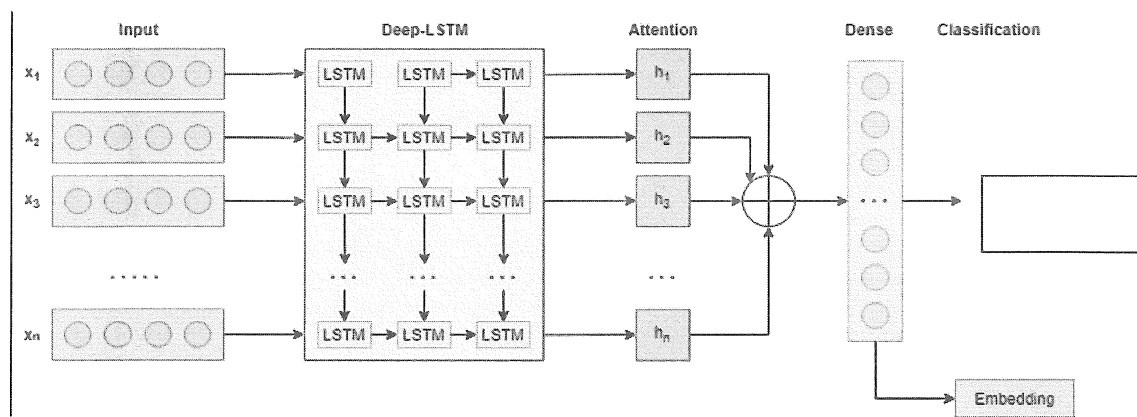


$$f_t = \sigma(W_f \cdot [C_{t-1}, h_{t-1}, x_t] + b_f)$$

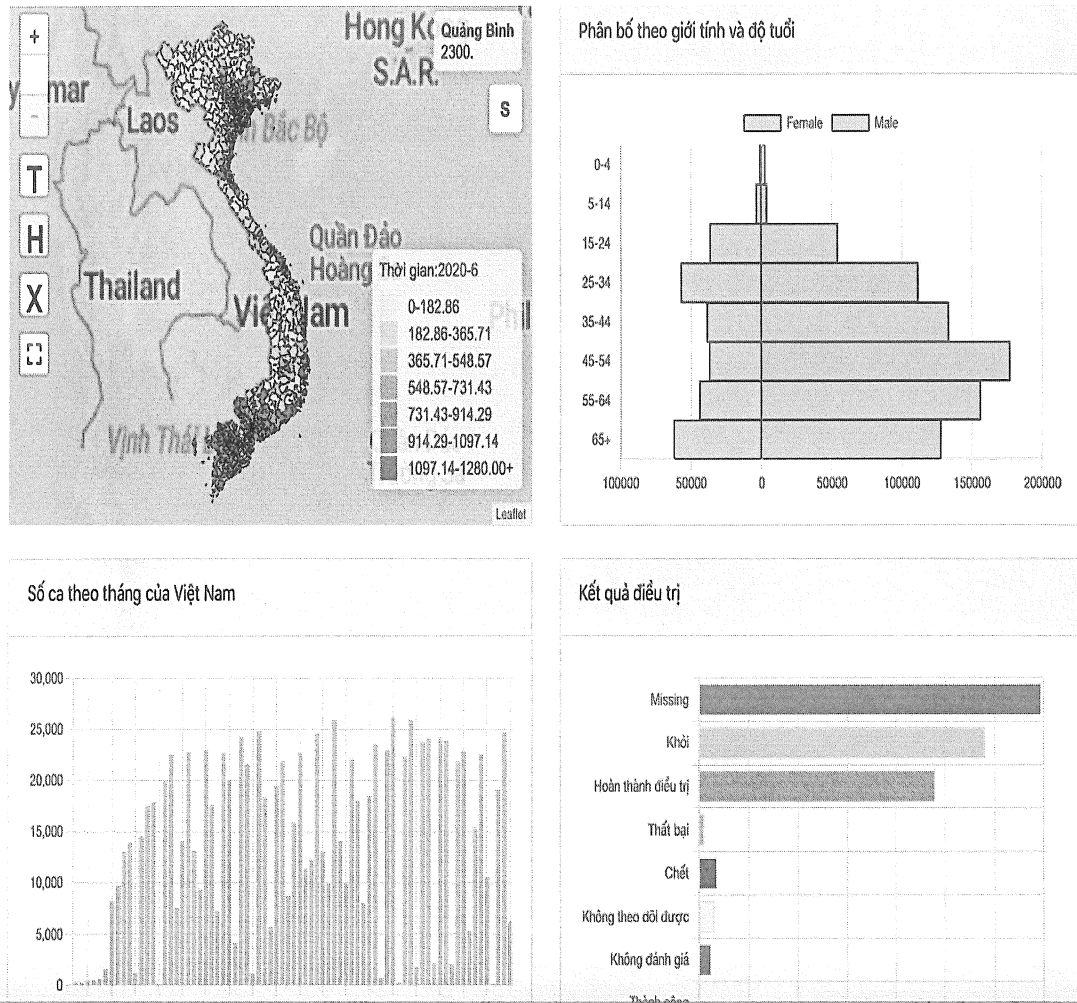
$$i_t = \sigma(W_i \cdot [C_{t-1}, h_{t-1}, x_t] + b_i)$$

$$o_t = \sigma(W_o \cdot [C_t, h_{t-1}, x_t] + b_o)$$

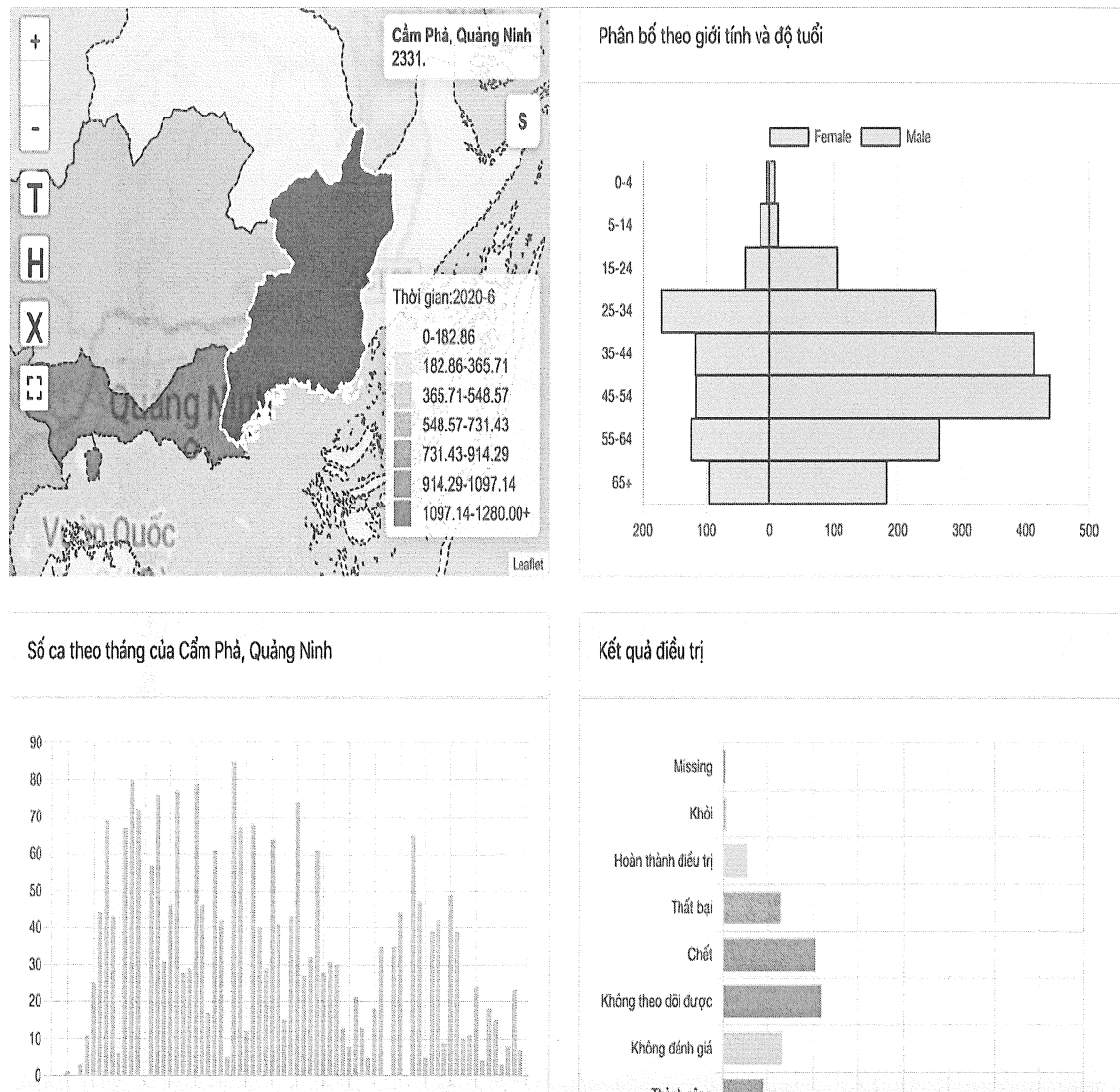
Hình 13



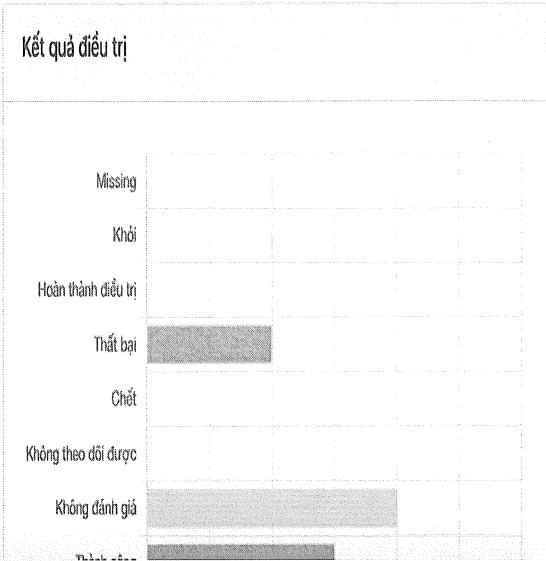
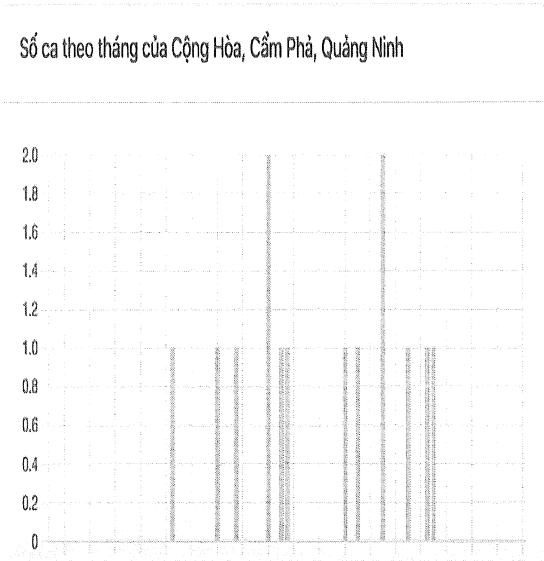
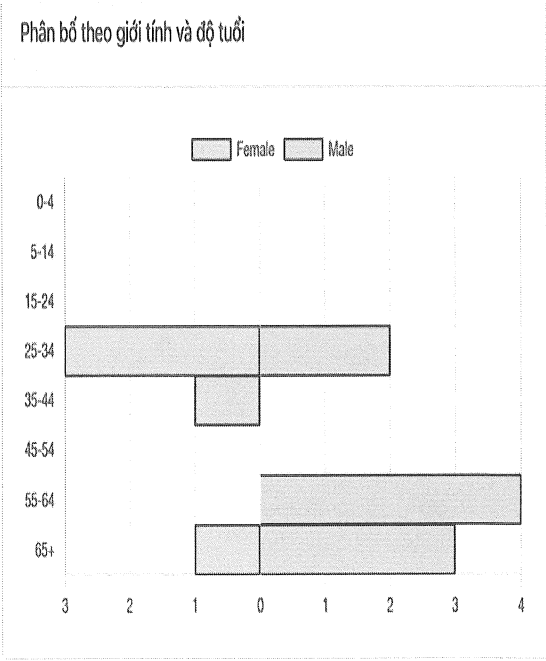
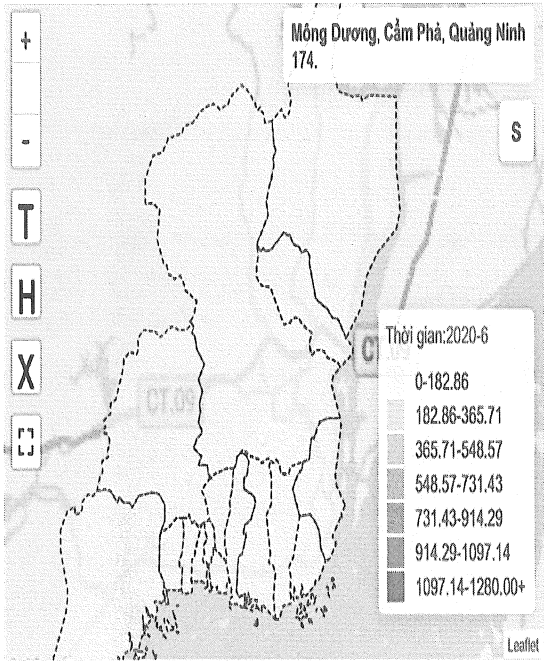
Hình 14



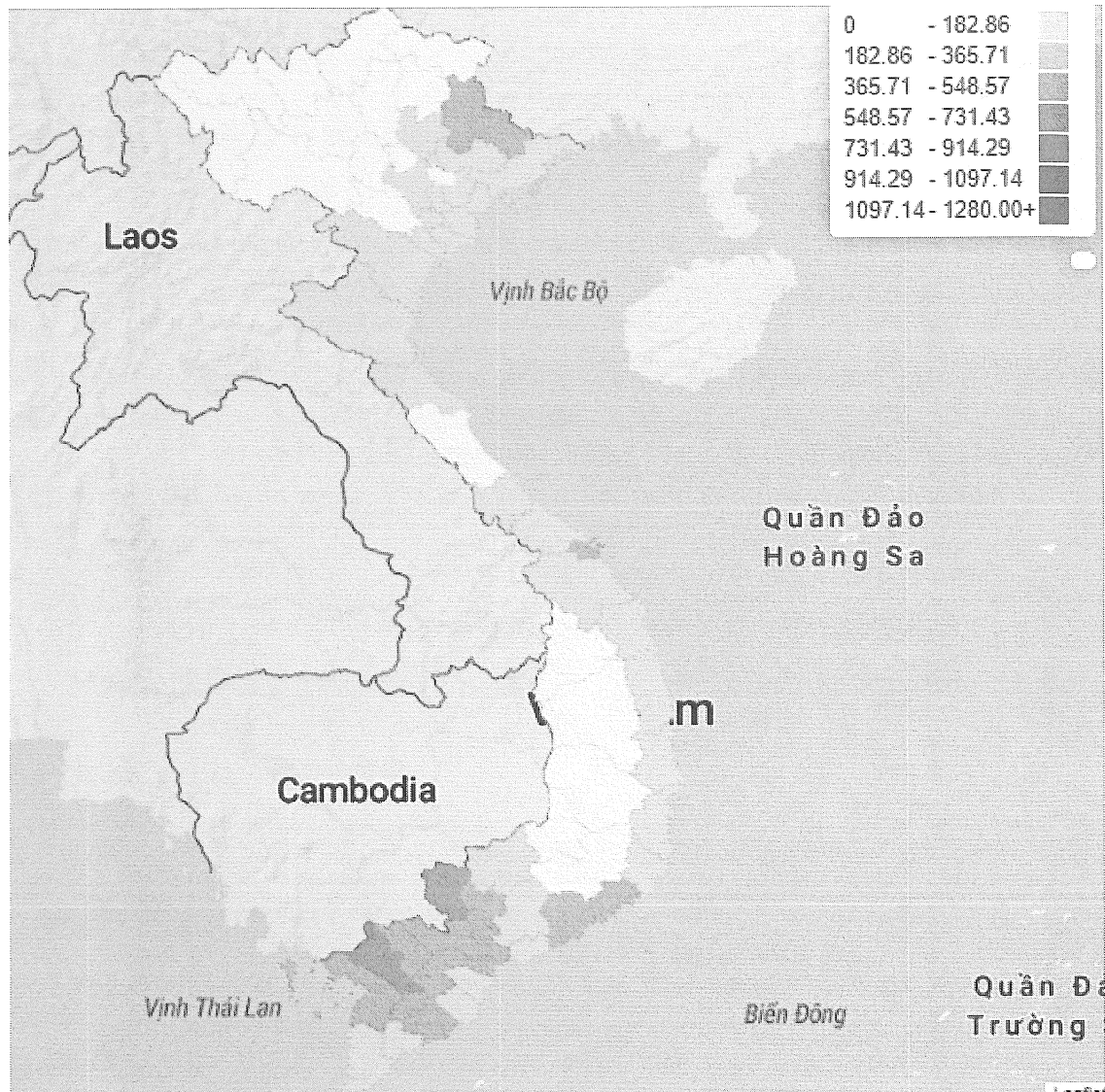
Hình 15



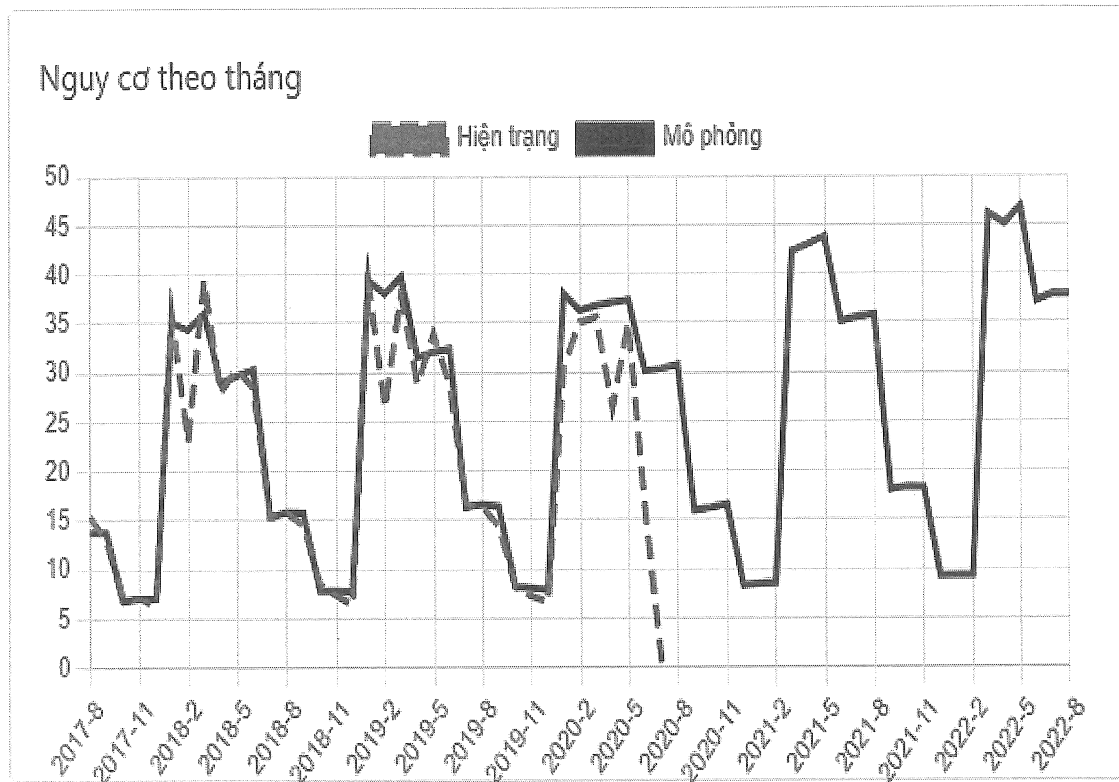
Hình 16



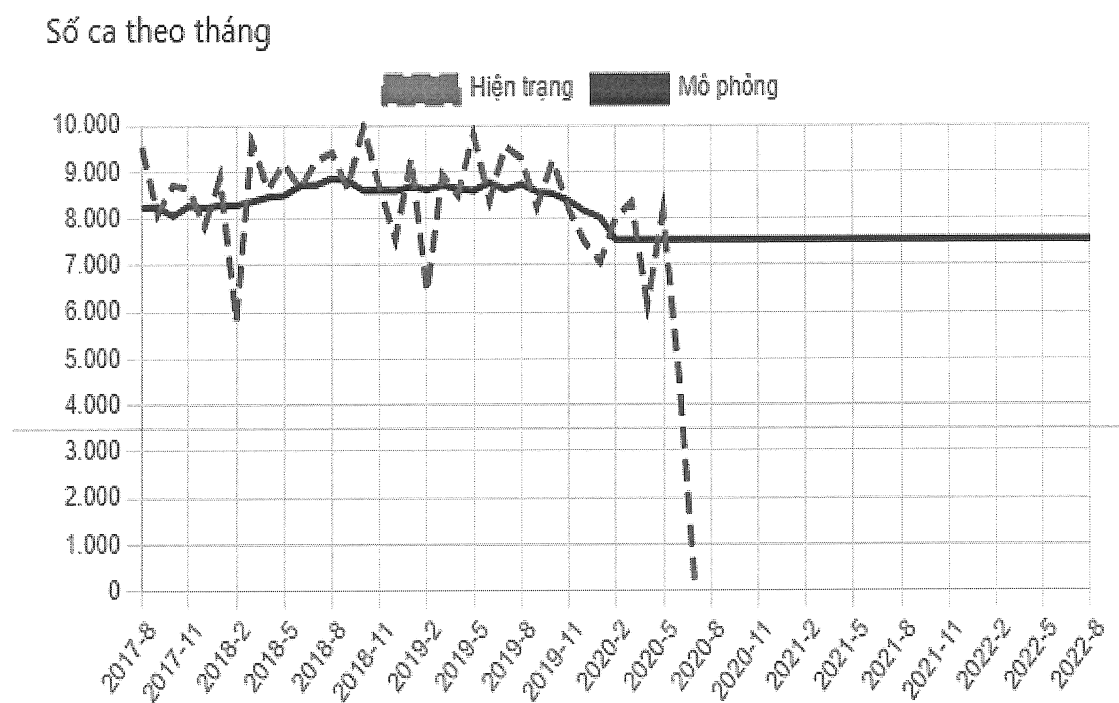
Hình 17



Hình 18

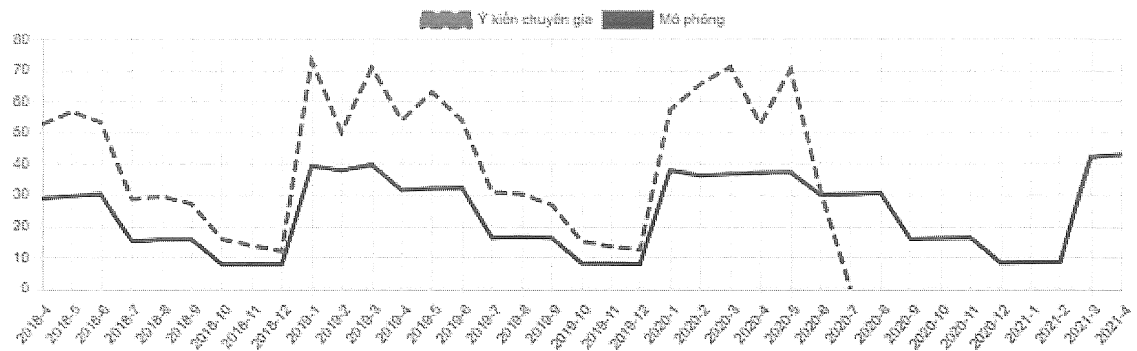


Hình 19



Hình 20

Nguy cơ theo tháng



Hình 21

Yếu tố ảnh hưởng đến lây nhiễm bệnh lao

phamtatthang@wru.vn (chưa chia sẻ) Chuyển đổi tài khoản

Yếu tố cá nhân người bệnh

Điểm 1 = Rất thấp 2 = Thấp 3 = Trung bình 4 = Cao 5 = Rất cao Trọng số

	1	2	3	4	5
HIV/AIDS	☐	☐	☐	☐	☐
Trọng số	☐	☐	☐	☐	☐
Bệnh tiểu đường	☐	☐	☐	☐	☐
Trọng số	☐	☐	☐	☐	☐

Hình 22