

BIG DATA VÀ NHỮNG ỨNG DỤNG TRONG Y TẾ

TS. NGUYỄN MẠNH CƯỜNG
 Chủ nhiệm Bộ môn Điện tử y sinh,
 Học viện Kỹ thuật Quân sự

TÓM TẮT: Cuộc cách mạng công nghiệp lần thứ tư với sự kết hợp giữa các thành tựu trên lĩnh vực công nghệ kỹ thuật số, lĩnh vực vật lý và lĩnh vực công nghệ sinh học đã mở ra nhiều tiềm năng cho sự bứt phá ngoạn mục của lĩnh vực y dược học. Tiếp tục bài viết “Cuộc cách mạng công nghiệp lần thứ tư và những ứng dụng tiềm năng trong y học” đăng trên Tạp chí Y học Quân sự số 329 (tháng 5-6/2018), bài viết này giới thiệu về Big data và các phương thức khai thác, ứng dụng Big data trong một số hoạt động y tế.

Từ khóa: Cách mạng công nghiệp, công nghệ sinh học, dữ liệu lớn.

ABSTRACT: The fourth industry revolution is marked by emerging technologies in a number of fields, including digital technology, physics and biological technologies... It made potential chances for developing medical and pharmacological sciences. This paper continuing the article “Fourth industrial revolution and potential medical applications” published in Military Medicine Journal 329 (May-June 2018). In this paper, we introduce the basic of Big data and its applications in medical field.

Keywords: Industrial revolution, biotechnology, big data.

Chịu trách nhiệm nội dung: TS. Nguyễn Mạnh Cường, Email: nmcuong@mta.edu.vn

Ngày nhận bài: 03/4/2017; chấp nhận đăng: 30/5/2018.

1. TÌM HIỂU VỀ BIG DATA.

Big data (dữ liệu lớn) là thuật ngữ dùng để chỉ một tập hợp dữ liệu rất lớn và rất phức tạp, đến nỗi những công cụ, ứng dụng xử lý dữ liệu truyền thống không thể nào đảm đương được. Thuật ngữ “Big data” không chỉ đề cập tới dữ liệu mà còn chỉ cơ cấu tổ chức dữ liệu, các công cụ và công nghệ liên quan. Tuy nhiên, Big data lại chứa trong mình rất nhiều thông tin quý giá, nếu trích xuất thành công, nó sẽ giúp ích lớn cho rất nhiều lĩnh vực, như kinh doanh, nghiên cứu khoa học, dự báo dịch bệnh, xác định điều kiện giao thông theo thời gian thực... Vì thế, những dữ liệu này phải được thu thập, tổ chức, lưu trữ, tìm kiếm, chia sẻ theo một cách khác so với bình thường [1].

Trước đây, nguồn dữ liệu lớn đã tồn tại dưới nhiều hình thức, do các tổ chức xây dựng phục vụ nhu cầu của riêng mình. Từ những năm 1990, nguồn dữ liệu lớn được thương mại hóa và được thu thập, cung cấp song song với hệ thống quản lý cơ sở dữ liệu. Lợi ích của Big data được nhắc đến thường xuyên hơn, đặc biệt trong lĩnh vực công nghiệp và kinh doanh. Cho đến hết năm 2003, toàn thế giới chỉ có khoảng 5 tỉ gigabyte dữ liệu. Tuy nhiên sau đó, một khối lượng dữ liệu như vậy được tạo ra chỉ trong mỗi 2 ngày vào năm 2011 và chỉ trong mỗi 10 phút vào năm 2013. Sự tiến bộ của công nghệ, sự ra đời của các kênh truyền thông mới như mạng xã hội và các thiết bị công nghệ mới tiên tiến hơn đã đặt ra thách thức cho các nền

công nghiệp phải tìm những cách thức khác nhau để thu thập và xử lý nguồn dữ liệu. Cho đến nay, Big data được hình thành từ 6 nguồn chủ yếu, là từ hoạt động hành chính (hồ sơ nhân thân, hồ sơ bảo hiểm, hồ sơ ngân hàng, bệnh án điện tử...); từ hoạt động thương mại (giao dịch thẻ tín dụng, giao dịch trên mạng...); từ các thiết bị cảm biến (thiết bị chụp hình ảnh vệ tinh, cảm biến khí hậu...); từ các thiết bị theo dõi (định vị GPS, máy theo dõi sức khỏe...); từ các hành vi (tìm kiếm trực tuyến, các trang trực tuyến truy cập...) và từ các thông tin về ý kiến, quan điểm của các cá nhân, tổ chức trên các phương tiện thông tin xã hội [8]. Do đó, không ngạc nhiên khi 90% dữ liệu của toàn thế giới ngày nay được tạo ra chỉ trong vài năm gần đây. Những dữ liệu cực kỳ hữu ích này gần như chỉ được sử dụng cho rất ít mục đích khác nhau trước khi thuật ngữ “Big data” ra đời. Cùng với vạn vật kết nối internet (IoT) và trí tuệ nhân tạo (AI), Big data ngày nay đã trở thành một thành phần vô cùng quan trọng của lĩnh vực kỹ thuật số. Vai trò của Big data trong cuộc cách mạng công nghiệp 4.0 được đánh giá giống như vai trò của dầu mỏ trong cuộc cách mạng công nghiệp 2.0 ở cuối thế kỷ XIX, đầu thế kỷ XX.

Phương pháp khai thác và quản lý Big data hiện nay được thiết kế phù hợp dựa theo các nguồn hình thành dữ liệu. Mỗi nguồn dữ liệu khác nhau sẽ có phương pháp khai thác và quản lý khác nhau. Tuy nhiên, phần lớn các tổ chức trên thế giới đều đang sử dụng giải pháp tối ưu là Hadoop ecosystem để khai thác và quản lý nguồn dữ liệu lớn.

2. MỘT SỐ ỨNG DỤNG BIG DATA TRONG HOẠT ĐỘNG Y TẾ.

Tuổi thọ trung bình của con người đang gia tăng cùng sự phát triển dân số thế giới. Điều này đặt ra những thách thức mới cho chiến lược phát triển y tế toàn cầu [4]. Những thành tựu của cách mạng công nghiệp 4.0 đã và đang đáp ứng nhu cầu phát triển y tế ở nhiều quốc gia. Nhiều nội dung trong lĩnh vực chăm sóc sức khỏe con người gần đây đã có những thay đổi nhanh chóng và mạnh mẽ. Đây cũng là một trong những lĩnh vực mà ứng dụng Big data hứa hẹn có thể làm thay đổi toàn diện như một cuộc cách mạng. Phân tích những dữ liệu liên quan đến chăm sóc sức khỏe trong Big data có khả năng dự báo bùng phát dịch bệnh lây nhiễm, phòng tránh các bệnh không lây nhiễm, bảo đảm tốt nhất chất lượng chăm sóc người bệnh theo thời gian thực, giảm các chi phí điều trị, nâng cao chất lượng cuộc sống nói chung. Dưới đây là một số ứng dụng của Big data trong các hoạt động y tế.

Dự báo nhằm cải thiện dịch vụ y tế tại bệnh viện. Bất kỳ nhà quản lý nào cũng đều phải tính toán để bố trí tối ưu số lượng nhân viên vào từng thời điểm nhất định. Nếu đưa vào quá nhiều nhân công, các chi phí lao động không cần thiết sẽ tăng lên. Nhưng nếu bố trí quá ít, chất lượng dịch vụ sẽ kém đi. Trong lĩnh vực chăm sóc sức khỏe, nguồn nhân công thiếu có thể dẫn đến hậu quả nghiêm trọng, thậm chí gây tử vong cho người bệnh. Big data đang giúp giải quyết hiệu quả cách thức quản lý nguồn nhân lực. Tại Paris, một dự án phân tích dữ liệu đã được thử nghiệm cho 4 bệnh viện. Dự án sử dụng dữ liệu từ nhiều nguồn khác nhau để đưa ra dự báo hàng ngày, hàng giờ về số lượng người bệnh có thể sẽ nhập vào mỗi bệnh viện. Một trong những dữ liệu này là hồ sơ nhập viện của bệnh viện trong 10 năm. Sử dụng kỹ thuật “phân tích chuỗi thời gian”, các chuyên gia phân tích dữ liệu tìm ra mối tương quan giữa tỉ lệ nhập viện với các yếu tố. Sau đó, họ sử dụng máy học để tìm ra các thuật toán chính xác nhất dự đoán xu hướng nhập nhập viện trong tương lai. Kết quả nghiên cứu là một giao diện dựa trên trình duyệt web được thiết kế dành cho nhân viên quản trị bệnh viện để dự báo tỉ lệ người bệnh tới khám và nhập viện trong 15 ngày tiếp theo. Nhờ đó, bệnh viện sẽ bổ sung nhân viên khi số lượng lớn người bệnh được dự báo, giúp giảm thời gian chờ đợi cho người bệnh, nâng cao chất lượng chăm sóc. Các bệnh viện tại Hoa Kỳ trước đây cũng không có sự chia sẻ thông tin với nhau. Điều này dẫn đến những khó khăn trong việc phân tích đúng các dữ liệu trên phạm vi rộng. Giờ đây, nguồn thông tin liên tục được cập nhật, chia sẻ, giúp việc phân tích dữ liệu dễ dàng và chính xác hơn. Nhờ vào nguồn thông tin và các

kết quả phân tích, người bệnh không chỉ tiết kiệm chi phí, thời gian phải trực tiếp đến bệnh viện khám chữa bệnh mà còn giúp các bệnh viện giảm số lượng bệnh nhân, góp phần nâng cao chất lượng dịch vụ, giảm đáng kể các chi phí khám bệnh, theo dõi và điều trị. Big data cũng góp phần giúp các bác sĩ chẩn đoán bệnh dựa trên bằng chứng là bệnh án nhiều năm của người bệnh, các dữ liệu lâm sàng lớn, tránh việc chẩn đoán bệnh chỉ dựa trên kinh nghiệm, gây sai lệch kết quả.

Cảnh báo theo thời gian thực. Cảnh báo theo thời gian thực đi liền với các thiết bị y tế cá nhân có thể mang theo người. Các thiết bị đeo trên người giúp giám sát các chỉ số sinh học của cơ thể người bệnh, như huyết áp, nhịp tim, thân nhiệt... và trong tương lai không xa là các thông số huyết học, sinh hóa... Tất cả các dữ liệu này đều được ghi lại theo thời gian thực với một lượng thông tin khổng lồ. Những dữ liệu từ xa này kết hợp với bệnh án điện tử sẽ nhanh chóng cảnh báo cho bác sĩ và người nhà người bệnh trong những trường hợp có bất thường hoặc trường hợp khẩn cấp. Một ví dụ là, nếu huyết áp của người bệnh tăng đáng báo động, hệ thống sẽ gửi cảnh báo ngay lập tức đến bác sĩ. Thông tin này sẽ được bác sĩ đáp ứng phù hợp như liên hệ, tư vấn điều trị cho người bệnh hoặc có biện pháp can thiệp, cấp cứu nhanh nhất có thể. Với các hệ thống cảnh báo thời gian thực như thế, thực tế đã cứu sống hàng triệu con người trên toàn thế giới những năm gần đây [3]. Ngay tại Madison, Winsconsin, Công ty Asthmapolis đã sử dụng ồng hất gắn định vị GPS, cho phép theo dõi để xác định mức độ, xu hướng hen suyễn của từng cá nhân người bệnh. Từ dữ liệu đó, đã có những phân tích tình hình hen suyễn trong một khu vực dân cư lớn, kết hợp với các dữ liệu y tế khác, công ty này đã phát triển các phác đồ điều trị hữu hiệu hơn cho bệnh hen suyễn [9].

Bệnh án điện tử (EHRs - Electronic health records). Đây là ứng dụng phổ biến nhất của Big data trong chăm sóc sức khỏe. Mỗi người bệnh có hồ sơ kĩ thuật số của riêng mình, bao gồm nhân khẩu học, lịch sử y tế, dị ứng, kết quả xét nghiệm... Hồ sơ được chia sẻ thông qua hệ thống thông tin an toàn và có sẵn cho các nhà cung cấp dịch vụ y tế, cả khu vực công và tư. Mỗi bản ghi bao gồm một tệp tin có thể sửa đổi, bổ sung. Bệnh án điện tử không chỉ đơn thuần là quản lí dữ liệu mà nó còn có thể kích hoạt các cảnh báo, nhắc nhở khi người bệnh nên có xét nghiệm mới hoặc theo dõi đơn thuốc để xem người bệnh có tuân thủ hướng dẫn của bác sĩ hay không. Bệnh án điện tử cũng giúp các bác sĩ đưa ra các chẩn đoán, điều trị trong một thời gian ngắn, dựa trên cơ sở dữ liệu đã được lưu trữ. Điều này đặc biệt hữu ích trong trường hợp

người bệnh có tiền sử bệnh lý phức tạp, có nhiều bệnh. Tại Hoa Kỳ, 94% các bệnh viện đã áp dụng bệnh án điện tử. Riêng chuỗi bệnh viện Kaiser Permanente tại đây đang dẫn đầu với việc triển khai đầy đủ hệ thống có tên Health connect để chia sẻ dữ liệu trong tất cả các cơ sở của họ để giúp sử dụng bệnh án điện tử dễ dàng, hiệu quả hơn. Báo cáo của McKinsey về tình trạng ứng dụng Big data trong chăm sóc sức khỏe cho biết: “Hệ thống tích hợp đã cải thiện kết quả trong theo dõi, điều trị bệnh tim mạch và tiết kiệm được hàng tỉ đô la từ việc cắt giảm chi phí thăm khám và xét nghiệm” [9].

Như trên đã đề cập, phân tích dự báo là một trong những xu hướng đang rất được quan tâm những năm gần đây. Các ứng dụng tiềm năng của phân tích dự báo sẽ còn vượt xa hơn nữa trong tương lai. Optum Labs, một hợp tác nghiên cứu của Hoa Kỳ, đã thu thập bệnh án điện tử của hơn 30 triệu người bệnh để tạo ra một cơ sở dữ liệu cho các công cụ phân tích, nhằm cải thiện việc chẩn đoán và điều trị bệnh. Mục tiêu của dự án là giúp bác sĩ đưa ra quyết định chính xác hơn, phác đồ điều trị hiệu quả hơn dựa trên sự hỗ trợ của dữ liệu lớn trong vòng vài giây. Chỉ trong vài phút, máy tính Watson của hãng IBM có thể đưa ra khuyến nghị về các phương pháp điều trị người bệnh ung thư bằng cách so sánh lịch sử bệnh lý, quy trình điều trị, kết quả chụp cắt lớp, dữ liệu di truyền cũng như sử dụng các dữ liệu về kiến thức y học cập nhật trên toàn cầu. Ứng dụng Google flu trend là một trong những ứng dụng y tế thành công của Google. Ứng dụng này dựa trên từ khóa tìm kiếm ở một khu vực, sau đó bộ máy của Google sẽ phân tích và đối chiếu kết quả tìm kiếm, đưa ra dự báo về xu hướng dịch cúm tại khu vực đó. Qua đó, cho biết tình hình cúm tại khu vực để đưa ra các giải pháp phòng tránh. Những kết quả mà Google flu trend đã đưa ra hoàn toàn phù hợp với báo cáo của Tổ chức Y tế thế giới WHO [9].

Dữ liệu lớn giúp quá trình điều trị ung thư.

Một ứng dụng dữ liệu lớn trong chăm sóc sức khỏe nổi bật là chương trình ung thư Moonshot. Trước khi kết thúc nhiệm kỳ thứ hai, tổng thống Hoa Kỳ Obama đã đưa ra chương trình này với mục tiêu đạt được các tiến bộ trong chữa bệnh ung thư. Các nhà nghiên cứu y khoa có thể sử dụng một lượng lớn dữ liệu về phác đồ điều trị và tỉ lệ phục hồi của người bệnh ung thư để tìm xu hướng và phương pháp điều trị có tỉ lệ thành công cao nhất hiện nay trên thế giới. Họ có thể kiểm tra các mẫu ung thư trong ngân hàng mẫu biobanks (có sự liên kết với hồ sơ điều trị của người bệnh). Sử dụng dữ liệu, các nhà nghiên cứu có thể phát hiện một số đột biến và protein ung thư tương tác với các phương pháp điều trị khác nhau và tìm ra hướng điều trị

cho kết quả tốt hơn [4]. Cũng nhờ phân tích dữ liệu đó, Desipramine, một chất chống trầm cảm có khả năng giúp chữa khỏi một số loại ung thư phổi nhất định đã được phát hiện [9].

Tuy có nhiều tiện ích, song việc ứng dụng Big data còn gặp rất nhiều trở ngại, đặc biệt là khó khăn do hệ thống dữ liệu không tương thích. Đây là thách thức kĩ thuật lớn nhất vì việc đưa ra giải pháp làm cho các hệ thống dữ liệu có thể giao tiếp với nhau là điều hết sức khó khăn. Vấn đề bảo mật của người bệnh cũng là một trở ngại vì nó liên quan đến các thông tin nhạy cảm, thông tin cá nhân và phụ thuộc vào các quy định pháp luật của từng nước hay quyền sở hữu dữ liệu của các tổ chức. Cũng có thể do phải dành rất nhiều thời gian, công sức, chi phí để phát triển bộ dữ liệu ung thư của riêng mình, nên các tổ chức không mong muốn chia sẻ nó.

Ứng dụng của Big data trong lĩnh vực y tế và nhiều lĩnh vực khác ở nước ta hiện vẫn đang trong quá trình nghiên cứu, triển khai từng bước. Nhà nước ta đã và đang có chủ trương xây dựng bộ cơ sở dữ liệu lớn, như để phát triển, xây dựng thành phố thông minh, chính quyền điện tử giai đoạn 2020-2025. Năm 2017, nhiều hội thảo về Big data, chính quyền điện tử, thành phố thông minh... đã được tổ chức. Trong đó, nhiều giải pháp tạo lập cơ sở dữ liệu, ứng dụng công nghệ thông tin được đánh giá tốt, như giao dịch ngân hàng điện tử, thương mại điện tử, phân tích hành vi khách hàng, số hóa tài liệu... Hi vọng rằng, các ứng dụng hữu ích của Big data sẽ sớm được quan tâm nghiên cứu, triển khai, ứng dụng trong các hoạt động y tế nước nhà, mang lại nhiều lợi ích thiết thực trong công tác bảo vệ, chăm sóc, nâng cao sức khỏe bộ đội và nhân dân trong tương lai không xa.

TÀI LIỆU THAM KHẢO:

1. Big data. https://en.wikipedia.org/wiki/Big_data
2. Vũ Công Lập (2018), “Y học và kỹ thuật y tế thời đại 4.0”, *Kỷ yếu hội nghị khoa học lần thứ 4*, Hội thiết bị y tế Việt Nam, 25/1/2018.
3. Three ways big data is transforming the medical field.
4. 9 Examples of Big Data Analytics in Health care that can save people.
6. Industry 4.0 in the medical technology and pharmaceutical industry sectors.
7. The Impact of Industry 4.0 on the Pharma Industry, Eli Pelleg, Ilanan Weissberg Doron.
8. Nguyễn Công Hoan, Tổng quan về dữ liệu lớn (Big data).
9. <https://www.datapine.com/blog/big-data-examples-in-healthcare/> □