

Imagen: Wikimedia Commons

Con adecuada exploración y análisis de datos, los sistemas informáticos pueden ayudar a ver en las imágenes digitales, como esta ecografía de hígado, patrones diagnósticos que escapan al ojo humano. Una de las muchas aplicaciones del big data en salud. Imagen: Wikimedia Commons.

Big data en salud

Sergio A. Urquijo Morales

Establecer un modelo marco eficiente para manejar la enorme cantidad de datos informáticos generados en las entidades de salud resulta clave para alcanzar un sistema más eficiente para administradores, prestadores del servicio, usuarios y pacientes. Los investigadores de la Universidad de Medellín se dedican a generar un mejor modelo de *big data* en salud.

■ Alguna vez le han hecho o ha visto una ecografía? Esa técnica de exploración del organismo mediante sonido no solo sirve para saber si hay un bebé en camino, se usa para muchos diagnósticos que incluyen inflamaciones y tumores. La mayoría de personas, al ver una ecografía, quedan confundidas debido a que “no distinguimos casi nada”. Eso ocurre porque el ojo humano se necesita entrenar para diferenciar tonos y formas en esas imágenes. Y hasta el más experto de los médicos y técnicos llega a distinguir solo una parte de la rica información que contiene una de estas imágenes diagnósticas.

No obstante, existen maneras de captar mucha más información que el ojo humano y el entendimiento de un individuo, estas pueden apoyar a los profesionales al momento de interpretar una imagen diagnóstica, como por ejemplo, un buen algoritmo o programa informático. Si dichas imágenes diagnósticas son digitales o digitalizadas, un adecuado manejo de sus datos y la comparación con los de muchas otras imágenes similares –almacenadas en bases de datos interconectadas– aportará información valiosa al diagnóstico.

Esa es solo una de las aplicaciones en salud de los sistemas de manejo de *big data*, un concepto que está en boga en estos tiempos y que hace referencia a inmensas cantidades de datos generados en las actividades de una organización, sus públicos variados y múltiples procesos. Este es precisamente, el amplio conjunto de temas que la Universidad de Medellín abordó en el proyecto *Framework Big Data Salud*, realizado en alianza con la Escuela de Ingeniería de Antioquia (EIA) y la Universidad Eafit.

Big data: buscar información en el océano de datos

La investigadora Lillyana Giraldo Marín del grupo de investigación Arkadius de la Universidad de Medellín, señala que “al inicio de la Cuarta Revolución Industrial, a eso del 2012, se posicionó el tema de *big data* como una de las grandes tendencias de un área del conocimiento que haría un aporte grandísimo, tanto a los sectores productivos, como a la investigación de los temas que le corresponden”.

Dicha cantidad de datos, por su tamaño y complejidad, no se puede interpretar con técnicas convencionales de estadística, para ello se requiere contar con aplicaciones informáticas no tradicionales de procesamiento de datos, como el aprendizaje de máquinas, la inteligencia artificial y la computación de alto desempeño. “Se necesitan técnicas como la minería de datos, que combina la parte aleatoria con la parte determinística en los procesos de optimización para hallar esa información que a simple vista

no se ve”, comenta Carmen Cecilia Sánchez Zuleta, investigadora del Grupo de Investigación en Modelación y Computación Científica de la Universidad de Medellín.

“Por ejemplo, todos los datos en rayos X de un gran hospital se pueden almacenar de forma equivalente a una cantidad de información de varios terabytes. Sin embargo, si vamos a un nivel más detallado o granular, la información crece de manera exponencial”, comenta el profesor Carlos César Piedrahita Escobar, también investigador del mismo grupo. “Por ejemplo, Hoof (2012) indica que la información genética almacenada en las todas las células de un cuerpo humano es del orden aproximado de 1,5 gigabytes por célula y si consideramos un promedio de 10^{14} células en el cuerpo humano, eso correspondería aproximadamente a 10^{23} bytes de información, es decir, a 100 zettabytes”.

Los datos están por todas partes en una entidad de salud como un hospital, laboratorio clínico, EPS o entidad administradora de salud. Se generan a partir de asuntos administrativos, financieros, de mantenimiento, procesamiento de señales e imágenes de equipos médicos; exámenes de laboratorio y compra y mantenimiento de los equipos e insumos. Se dan desde la gestión del personal hasta la atención al usuario y sus diagnósticos.

Cuando esta información no se maneja adecuadamente se puede presentar un sinnúmero de inconvenientes en los diferentes procesos, por ejemplo, pacientes mal diagnosticados, procedimientos médicos que no son los indicados para un paciente en

Los datos están por todas partes en una entidad de salud, sea esta un hospital, laboratorio clínico, EPS o entidad administradora de salud.

específico, exámenes y procesos duplicados, además de errores en el manejo de presupuestos y facturaciones, y duplicación de gastos en detrimento del presupuesto. Todos esos errores se pueden corregir con un tratamiento de datos que sea óptimo.

Una brújula para navegar en los datos

El sector salud anualmente genera grandes cantidades de datos que se convierten en mares de información, este es uno de los motivos por los cuales investigadores de los grupos de Modelación y Computación Científica, y Arkadius, de la Universidad de Medellín, junto a pares de la Universidad Eafit y la Universidad EIA, desarrollaron un proyecto de investigación en aras de identificar el diseño más apropiado de un *framework* o marco para el *big data* con modelos, tecnologías, herramientas, guías y servicios para el desarrollo y la competitividad del sector salud en Colombia. Una especie de mapa y brújula para el arduo trabajo.

Buscar en el vasto océano de datos la información relevante o necesaria para llegar a la solución adecuada frente a una situación problemática es como buscar oro en las entrañas de una montaña. Por eso, resulta acertado que a esa labor se le llame minería de datos, como lo ejemplifica la profesora Sánchez Zuleta, “es escarbar, extraer de una gran nube de información datos que a simple vista no se ven, conectarlos y sacarlos a la luz, para luego comprender de qué hablan esos datos que se develaron”.

Para tratar adecuadamente dicho volumen de datos se requiere contar con aplicaciones informáticas no tradicionales de procesamiento de datos, como el aprendizaje de máquinas, la inteligencia artificial y la computación de alto desempeño.

El primer reto que debió resolver el equipo de investigadores fue encontrar las instituciones aliadas dispuestas a permitir el acceso a sus nubes de datos y contar sus principales problemáticas referentes a la gestión de la información en sus entidades. No fue una tarea fácil, las entidades de salud manejan datos muy sensibles y que, en muchas ocasiones, si salen a la luz, podrían comprometer la privacidad de sus pacientes o la gestión de los administradores y el personal de salud. Por ello, algunos de los resultados tienen carácter confidencial.

Después del proceso de caracterización, una de las postuladas fue la IPS universitaria de la Universidad de Antioquia. Inicialmente, los investigadores identificaron que, por la

complejidad que implica trabajar con datos masivos, era necesario partir de un modelo de madurez de *big data*. Este permite, mediante la evaluación de varios criterios, diagnosticar la capacidad de la organización para enfrentar este tipo de análisis.

El nivel de madurez se refiere al estado de conocimiento que tiene una organización sobre los datos que genera, metodologías para gestionar los datos, tecnologías de información que utiliza para soportarlos, conectividad y disponibilidad de fuentes. Este análisis permitió validar la capacidad organizacional de la IPS universitaria para la implementación del estudio.

Un trabajo de muchas áreas

Luego del análisis de madurez se definieron los grupos de profesionales necesarios para desarrollar el proyecto con la IPS, fueron expertos del área administrativa puesto que era indispensable conocer la organización, entender su misión y visión, estrategias, objetivos e intenciones. “Este es un trabajo articulado e interdisciplinario, no lo hacemos solo nosotros o el experto en matemáticas, estadística o sistemas. Si no generamos un equipo de trabajo articulado la probabilidad de éxito se reduce significativamente”, señala Liliana Giraldo Marín.

El grupo de trabajo que la IPS requería, entre otros, de expertos en urgencias, epidemiólogos, medicina general, debido al proceso organizacional que se quería intervenir. Por su parte, los investigadores del proyecto fueron expertos en *big data*, analítica

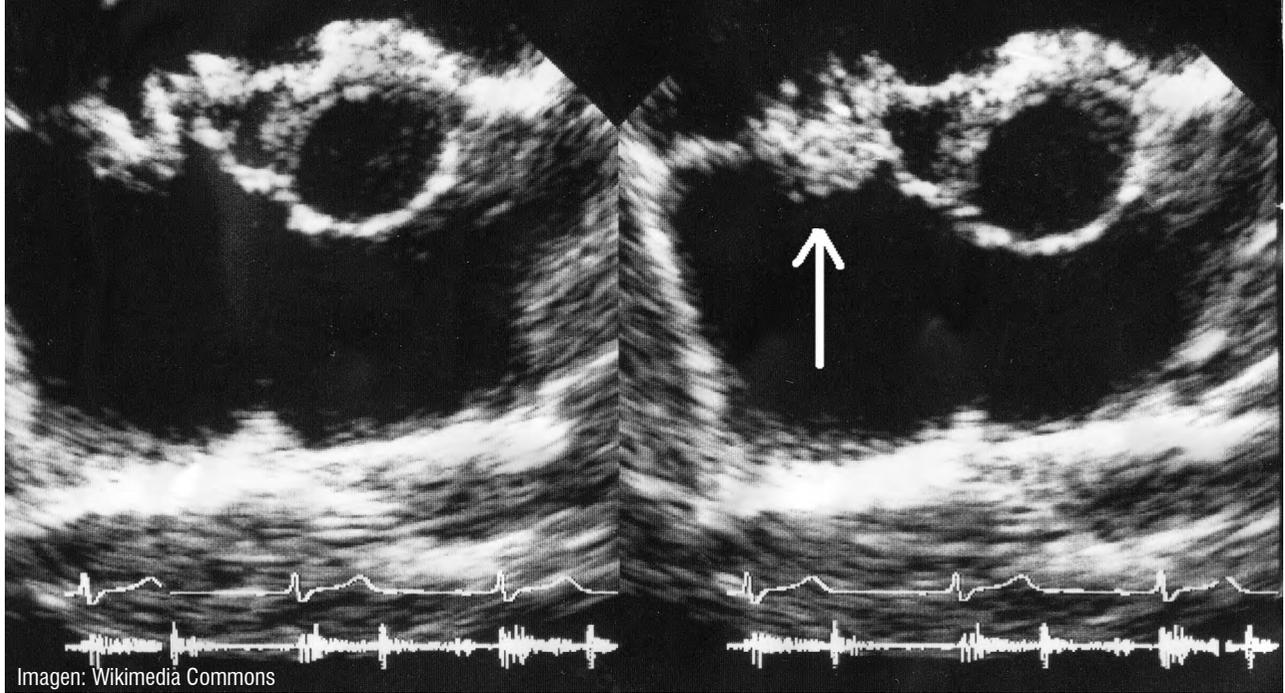


Imagen: Wikimedia Commons

Ultrasonido de una endocarditis, un desafío diagnóstico para el especialista y para los sistemas de big data. Imagen: Wikimedia Commons.

de datos, *machine learning*, estadística, gestión de la información y matemáticas. Estos debían gestionar las fuentes de información, saber dónde están ubicadas, en qué formatos se encuentran, su capacidad de procesamiento y los accesos para poder llegar a estos datos.

Los equipos de trabajo utilizaron metodologías, en gran medida, propias, con la intención de dar pasos lentos pero acertados. De esta manera, si se presentaba algún inconveniente en el proceso de análisis de la data podrían hacer ingeniería inversa para identificar, corregir la situación problemática y continuar con el estudio. Una vez determinados y gestionados todos los datos para el análisis, los investigadores iniciaron un proceso de interpretación, ajuste y toma de decisiones según el objetivo y las estrategias planteadas inicialmente. Los componentes del proceso mencionado se agrupan en lo que el equipo de investigadores denominó *framework* o marco.

El futuro de los sistemas de salud es muy prometedor gracias al

potencial que yace bajo el océano de datos, pero hay que encontrar el modelo más adecuado para compilar, almacenar, gestionar, minar e interpretar los datos. “En los grandes sectores de la economía y la ciencia casi siempre hay un sistema de detección automático que le ayuda a los expertos a tomar decisiones frente a grandes volúmenes de datos, que, de otra forma sería imposible de analizar, como ocurre en áreas como geofísica, sector financiero, física de altas energías, astrofísica y meteorología, entre otros”; explica el investigador Carlos César Piedrahita Escobar.

Con datos organizados y un modelo eficiente para acceder a ellos y comprenderlos, no solo se pueden realizar diagnósticos más precisos con imágenes que a nuestros ojos son casi manchas, también se mejoran los procesos de afiliación y traslado, se exploran las historias clínicas de manera más eficiente –uno de los dolores de cabeza del sistema– y hasta se facilita compartir información entre instituciones.

Todo esto se encuentra mediado por aquello que, hasta el momento, las máquinas no hacen tan bien como los humanos. “Los análisis basados en *big data*, si bien se apoyan en sistemas tecnológicos, no reemplazan al experto como muchas personas piensan, por el contrario, lo apoyan en la toma de decisiones más acertadas”, concluye el profesor Piedrahita Escobar. La tecnología siempre ha sido una gran aliada, pero en la Cuarta Revolución Industrial quizás comencemos a apreciarla como colega de trabajo. ○

Proyecto de investigación	<i>Framework big data para apoyar el desarrollo del sector salud en Colombia</i>
Investigadores	Carmen Cecilia Sánchez Zuleta Carlos César Piedrahita Escobar Lillyana Giraldo Marín Andrés David Blandón Restrepo Christian David Orrego Suaza Manuel Alejandro Ramírez Claudia María Tabares Betancur Marta Silvia Tabares Betancur
Entidades participantes	Universidad de Medellín Universidad EIA Universidad Eafit
Estado del proyecto	Terminado
Palabras claves	Big data, sector salud, <i>machine learning</i> , minería de datos