

# SEMINARIO SOBRE LAS TELECOMUNICACIONES Y LA ELECTRÓNICA EN LA MEDICINA



Jueves 5 de diciembre de 2019  
8.00 a 14.00 Hrs.



Academia  
de Ingeniería México



CDA  
Centro de  
Desarrollo Académico



CUDI  
Centro de Desarrollo Académico

Centro de Educación Continua "Eugenio Méndez Docurro" del IPN  
Belisario Domínguez # 22, Colonia Centro, CD de México.



# Ciencia de Datos – Big Data en Medicina de Precisión *Impacto en Cardiología*

Dr Fabian García Nocetti  
IIMAS-UNAM  
CAACFMI-UNAM  
*dfgarcia@unam.mx*



## Que es?

Concepto que hace referencia a la adaptación del tratamiento médico a las características individuales del paciente.



## Que implica?

**Decisiones referentes al tratamiento o a la prevención de enfermedades se tomarán con base en:**

- Características genómicas y moleculares de la enfermedad
- Información sobre la situación clínica
- Hábitos del paciente.
- Expectativas.

- **Aspectos que sustentan esta idea:**
  - Conocimiento de alteraciones genéticas y moleculares
  - Desarrollo de fármacos que actúen a nivel de dichas alteraciones
  - Almacenamiento de datos de pacientes (datos moleculares, genéticos, clínicos, etc.)
  - Nuevas posibilidades de gestión y explotación de grandes volúmenes de datos generados
  - Análisis de Big Data para la creación de nuevo conocimiento



## ¿Qué es?

Conjuntos de datos cuyo tamaño está más allá de la capacidad de las herramientas de software de bases de datos típicas para capturar, almacenar, gestionar y analizar información.



## ¿Cómo se origina?

Por la explosión en la cantidad (velocidad y frecuencia) y diversidad de datos digitales generados en tiempo real como resultado del rol cada vez mayor de la tecnología en las actividades diarias.



## ¿Para qué sirve?

Permite generar información y conocimiento con base en información completa en tiempo real.



## Implicaciones

- Era caracterizada por la abundancia de datos.
- Ha alcanzado todos los sectores en la economía
- Los datos son un nuevo factor de producción y de ventaja competitiva
- Oportunidad: Aprender sobre el comportamiento humano para diversos fines.
- Creación de valor vía innovación, eficiencia y competitividad
- Nuevas formas de competencia y nuevos negocios Almacenamiento y gestión de datos.



## **Analítica: capacidades**

- La analítica de grandes datos se refiere a las herramientas y metodologías para transformar cantidades masivas de datos brutos en “información útil” con propósitos analíticos
- Se originó en las áreas de biología intensiva en cómputo, ingeniería biomédica, medicina y electrónica
- Algoritmos para detectar patrones, tendencias y correlaciones, en varios horizontes temporales, en los datos
- Uso de técnicas avanzadas de visualización: datos que hacen sentido



- Proyectos de desarrollo de software orientado hacia la computación distribuida.
- Buscan resolver parte de los problemas asociados a big data y a la aparición del data science.
- Ofrecen capacidad de almacenamiento y procesamiento local.
  - Permite escalar desde unos pocos servidores hasta miles de máquinas, todas ellas ofreciendo calidad de servicio.
  - Permite el procesamiento distribuido de grandes conjuntos de datos en clusters de computadoras utilizando modelos sencillos de programación.



## Fases de la Analítica de Big Data

### 1. Descubrimiento de grandes datos

- Definir cuáles son los datos de interés.
- Encontrar sus fuentes (históricos o Social Media, entre otros).
- Grabar los datos en el sistema.
- Determinar cómo serán procesados.

### 2. Extracción y limpieza de los grandes volúmenes de datos

- Extraer los datos de la fuente de origen datos.
- Perfillar y limpiar los datos.
- Adecuarlos a las necesidades.
- Aplicar los estándares de calidad de datos.



## Fases de Big Data y sus soluciones

### 3. Estructuración y análisis de big data

- Dotar de estructura lógica a los conjuntos de datos tratados.
- Almacenar los datos en el repositorio elegido (puede ser una base de datos o un sistema)
- Analizar los datos disponibles para hallar relaciones.

### 4. Modelado de datos

- Aplicar algoritmos a los datos.
- Aplicar procesos estadísticos.
- Resolver las peticiones lanzadas mediante el modelado de datos en base a técnicas de minería.

### 5. Interpretación de grandes datos

- Interpretar las distintas soluciones.
- Aportar un resultado final.



- **Crecimiento explosivo cantidad de datos**
  - Generados a grandes velocidades
  - Distintos formatos y diferentes fuentes
    - Redes sociales, dispositivos móviles, sensores, etc.
- **Necesidad de extraer**
  - Patrones, tendencias y/o conocimiento
  - De forma rápida y eficiente



- **Evolución de métodos tradicionales**
  - Rendimiento
  - Escalabilidad
- **Contenido de valor que genera permite**
  - Mejora en toma de decisiones
  - Obtención de ventajas competitivas
  - Diferentes campos de acción



- **Avance tecnológico** ha permitido un **crecimiento explosivo** de datos generados
- **Fuentes:** redes sociales, dispositivos móviles, sensores, máquinas de rayos x, log de aplicativos y, en términos generales, lo que se puede clasificar como Internet de las Cosas"



- Necesidad de extraer **patrones, tendencias** y/o conocimiento para apoyar la toma de decisiones.
- Métodos tradicionales de procesamiento de datos han tenido que evolucionar rápidamente (**escalabilidad y rendimiento**).
- Referencia principalmente a tres términos conocidos como las 3 Vs: **Volumen, Velocidad y Variedad**.



- CD incluye **técnicas, tecnologías, métodos y paradigmas** que apoyan la solución de problemas relacionados con datos de una **forma diferente y más adecuada** que los métodos tradicionales.
- CD permite nuevas y mejores formas de procesar información.
- Ventajas sobre los enfoques tradicionales (**velocidad, costos de implementación, escalabilidad, flexibilidad y elasticidad**)



- Enfoques orientados a la **computación distribuida** y el **procesamiento paralelo masivo**.
- Convergentes también con la **computación en la nube** y nuevas formas de almacenar los datos mediante **modelos no relacionales**.
- Existen arquitecturas de referencia, **patrones de diseño** y tendencia de S/W y H/W para facilitar el uso de BD asociado con CD.



- Pero los datos almacenados y gestionados no representan por si solos una ventaja.
- Lo verdaderamente importante: **VALOR** que se puede generar a partir de estos datos.





- Desarrollo de medicamentos inductivos
- Los genes juegan un papel importante en la progresión de la enfermedad y son objetivos terapéuticos
- Debe identificar los pacientes que reúnen las condiciones necesarias y suficientes entre el objetivo a lograr y la fisiopatología de la enfermedad cardiovascular.
- Posible alcanzar este objetivo mediante BDA (variables epidemiológicas, registros electrónicos y BD de genómica).

# Impacto en Cardiología



Academia  
de Ingeniería México

- Diversos estudios sobre tendencias de enfermedades cardiovasculares
- Necesidad de estudios epidemiológicos que permitan el reconocimiento preciso de los factores de riesgo, su distribución y efectos sinérgicos
- Lograr una modificación a corto y mediano plazo de los mismos
- Prevención efectiva de la cardiopatía isquémica e insuficiencia cardíaca a nivel primario, secundario y terciario.

# Impacto en Cardiología



Academia  
de Ingeniería México

- De la misma manera plantean que la Analítica de Big Data (BDA) y la Minería de Datos como vías de obtenerlo.
- Señales para llamar la atención a los cardiólogos, sobre la necesidad de BDA y el resto de los componentes de esta nueva visión de la medicina
- Existen realidades tecnológicas irreversibles que es imprescindible conocer



## ANÁLISIS DE BIG DATA EN EPIDEMIOLOGÍA CARDIOVASCULAR

- ❑ La computación de alto rendimiento (HPC) que usa el procesamiento paralelo para ejecutar aplicaciones avanzadas de manera rápida, eficiente y confiable.
- ❑ La proporción de centros de supercomputo que emplean co-procesadores y aceleradores se ha duplicado en los últimos 2 años.
- ❑ Los recursos de HPC en la nube son cada vez más accesibles lo que constituye para los consumidores un servicio.

# Impacto en Cardiología



Academia  
de Ingeniería México

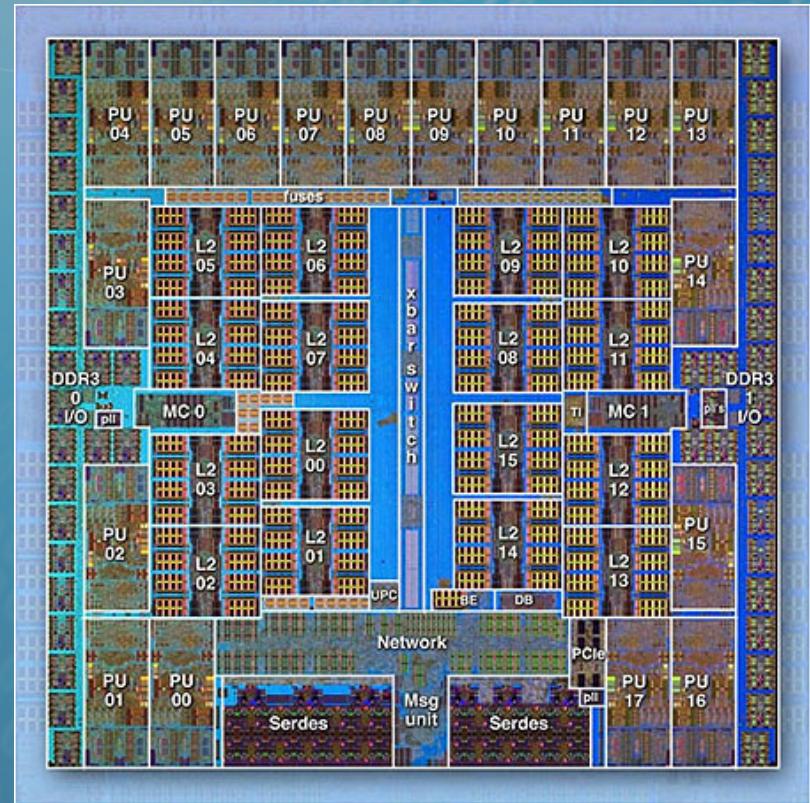
## ANÁLISIS DE BIG DATA EN EPIDEMIOLOGÍA CARDIOVASCULAR

- ❑ Estamos inmersos en una nueva plataforma tecnológica que está constituida por la informática móvil, servicios en la nube, Big Data, analítica y las redes sociales.
- ❑ El gasto mundial en software y hardware relacionado con el Big Data y analítica crecerá y las grandes cadenas de suministro de datos como servicio crecerán en importancia como plataforma en la nube.
- ❑ La adopción de la infraestructura de nube como servicio (IaaS) crecerá ostensiblemente en un 30%-40% en este año.

# Impacto en Cardiología

A la fecha virtualmente todas las computadoras “stand-alone” son paralelas desde una perspectiva de hardware:

- ◆ Multiple functional units (L1 cache, L2 cache, branch, prefetch, decode, floating-point, graphics processing (GPU), integer, etc.)
- ◆ Multiple execution units/cores
- ◆ Multiple hardware threads



IBM BG/Q Compute Chip with 18 cores (PU) and 16 L2 Cache units (L2)

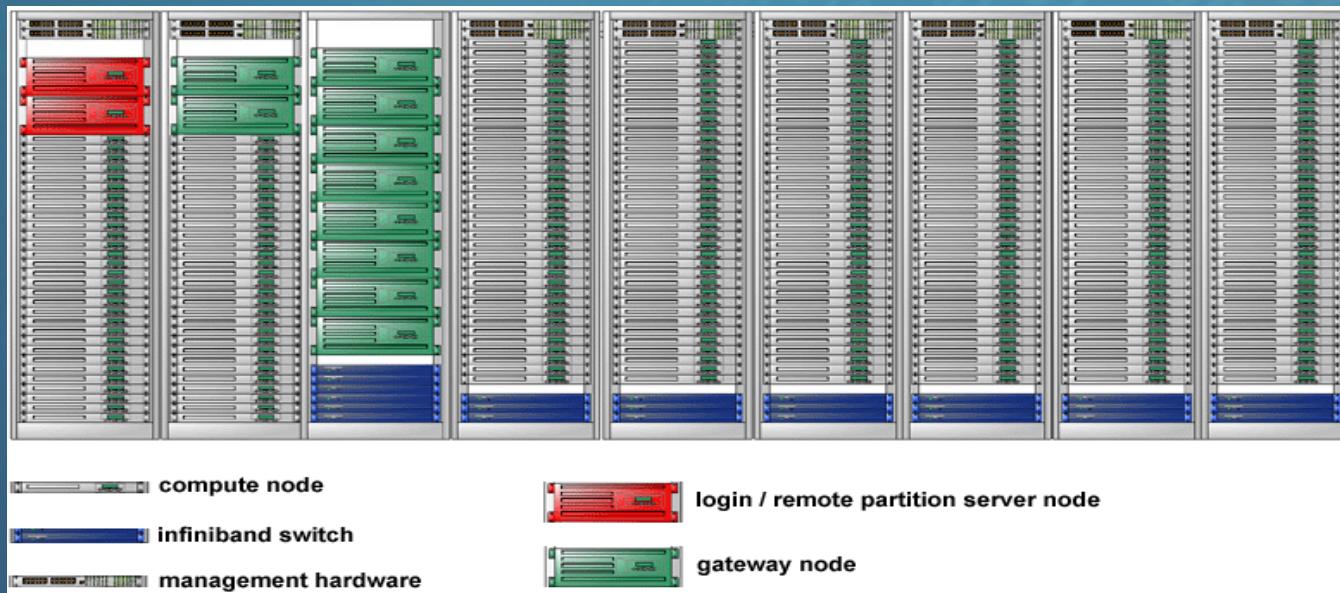
# Impacto en Cardiología



Academia  
de Ingeniería México

La figura muestra un cluster de computadoras paralelas:

- ◆ Cada nodo de cálculo es un equipo paralelo multiprocesador
- ◆ Múltiples nodos de cálculo se conectan con una red Infiniband
- ◆ Los nodos de propósito especial, también multiprocesador, se utilizan para otros fines



# Impacto en Cardiología



Academia  
de Ingeniería México

- HPC integra un conjunto de herramientas y técnicas para la distribución y ejecución de trabajos a lo largo de un cluster de computadoras que acceden a un sistema de archivos compartidos ubicados en un sistema de almacenamiento en red.
- La ejecución de trabajos de forma paralela y distribuida puede resolverse utilizando especificaciones como la interfaz de paso de mensajes “*Message Passing Interface*”
- MPI define cómo los procesos se comunican entre sí mediante el envío y recepción de mensajes.

# Impacto en Cardiología



Academia  
de Ingeniería México

- El **BDA** en la epidemiología cardiovascular permite el estudio en topografías y grupos demográficos diversos
- Puede modelarse según los propósitos de **examen de prevalencias**, o seleccionando subpoblaciones en áreas específicas.
- Estudiar la salud de la población en áreas pequeñas permite el diseño de **políticas de salud locales**
- Planificación de los recursos

# Impacto en Cardiología



Academia  
de Ingeniería México

- La Ciencia de Datos y HPC permite detectar nuevos patrones dentro de amplios datasets
- Facilitar la decisión médica (correlación del Score Syntax con las diferentes modalidades de imagen en diferentes subgrupos de pacientes)
- Diagnóstico no invasivo más simple de las cardiopatías congénitas.

# Impacto en Cardiología



Academia  
de Ingeniería México

- La **medicina basada en la evidencia** que hemos conocido hasta ahora, debe ser modificada a una nueva dimensión
- **Convergencia** de los avances tecnológicos
- Datos móviles y **las metodologías** para transformar cantidades masivas de datos sin sentido aparente, en datos sobre datos con propósitos analíticos.
- Esto se une a un **procesamiento en la nube** cada vez más eficaz y a la aparición de nuevos sensores y dispositivos.

# Impacto en Cardiología



Academia  
de Ingeniería México

- En **Cardiología**, el empoderamiento del paciente a través de **dispositivos móviles y aplicaciones** será un campo de desarrollo decisivo
- Fundamentalmente en **la insuficiencia cardiaca para su manejo extrahospitalario** y la disminución de las costosas rehospitalizaciones,
- También en **la fibrilación auricular** y los aspectos terapéuticos específicos.

# Impacto en Cardiología



Academia  
de Ingeniería México

- Quedan problemas para fortalecer el papel del BDA pendientes de resolver
- Incentivos para compartir datos y la preservación de la privacidad y establecimiento de los límites al anonimato del conjunto de datos.

# Impacto en Cardiología



# **ANALISIS DE BIG DATA EN EPIDEMIOLOGIA CARDIOVASCULAR**

# Análisis de Big Data en epidemiología cardiovascular

**DISCA-IIMAS-UNAM**  
**DCC-IIMAS-UNAM**

# FABIAN GARCIA NOCETTI GIBRAN FUENTES

FAC. DE MEDICINA  
IMSS

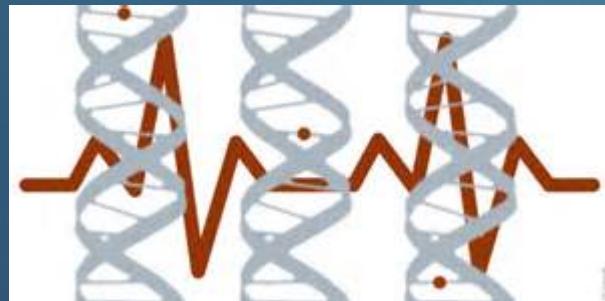
GABRIELA BORRAYO

CARDIOCENTRO  
CLINICO QUIRURGICO

JUAN PROHIAS

POSGRADO CIC-UNAM

BLANCA VAZQUEZ





- Las **enfermedades cardiovasculares** son la **primera causa** de muerte en el mundo. (17 millones de personas)
- Análisis de los **registros electrónicos clínicos (EHR)**
- Enfoque novedoso ofrece nuevas **oportunidades para la detección** de enfermedades cardiovasculares.
- **Retos:** la naturaleza de los datos y la temporalidad entre eventos.



- EHR se caracterizan por su alta dimensionalidad, complejidad, heterogeneidad, ruido y ausencia de datos.
- EHR registran la dinamicidad de los eventos clínicos en un espacio de tiempo
- Construir un modelo predictivo para identificar pacientes en riesgo de infarto agudo al miocardio a partir del análisis de EHR utilizando un enfoque de aprendizaje profundo no supervisado



- El modelo predictivo tiene características principales:
  - Procesar toda la información contenida en los EHR (respetando la naturaleza de los datos y considerando temporalidad)
  - Construir una red neuronal profunda (eficientes en el tratamiento de temporalidad y generación de soluciones escalares)
- Contribuciones:
  - identificación de pacientes en riesgo
  - Identificación de fenotipos clínicos
  - Jerarquización de factores de riesgo.



- Elementos a considerar:
  - Transformación Digital Hospitalaria
  - Sistemas de historia clínica electrónica interoperables
  - Crecimiento moderado (firme) de EHR en la nube
  - Incremento uso de aplicaciones móviles en salud
  - Déficit de infraestructura seguirá presionando los modelos de asociación público privada APP
  - Interés en la tecnologías BDA, IA, IoT
  - Signos de la inmadurez del mercado
  - Problemas de seguridad
  - Impacto de inversiones en el sector



Academia  
de Ingeniería México

**Posgrado en Ciencia e Ingeniería de la Computación  
Instituto de Investigaciones en Matemáticas  
Aplicadas y en Sistemas  
Universidad Nacional Autónoma de México**

**Modelo predictivo para la detección  
de enfermedades cardiovasculares  
basado en EHR**

**M.C Blanca Hilda Vázquez Gómez  
Dr. Gibran Fuentes Pineda  
Dr. Fabián García Nocetti**

# Motivación

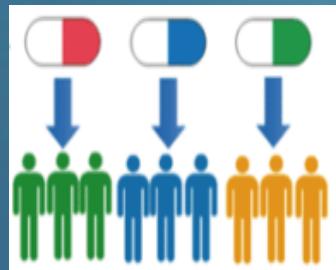


Academia  
de Ingeniería México

- Las enfermedades cardiovasculares son la primera causa de muerte en el mundo.
- Las enfermedades tiene un impacto negativo en los sistemas del sector salud.

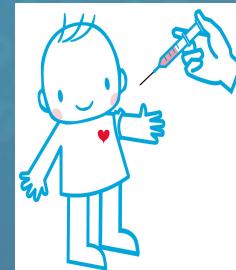
# Retos y oportunidades

Medicina de precisión  
(correctivo)



<http://amitray.com/artificial-intelligence-precision-medicine/>

Detección  
temprana de  
enfermedades  
(preventivo)



<http://judypaez.blogspot.com>

# Capitalizar la información



Registros Electrónicos Clínicos

- Fueron construidos con propósitos clínicos y de facturación.
- Le dan seguimiento al ciclo de vida del paciente
- Ofrecen nuevas oportunidades para la detección temprana de enfermedades

# Datos comunes en los EHR



Academia  
de Ingeniería México

- ✓ Alergias
- ✓ Medicamentos
- ✓ Diagnóstico
- ✓ Procedimientos
- ✓ Admisiones / readmisiones
- ✓ Examen físico
- ✓ Historia clínica
- ✓ Reportes de rayos X
- ✓ Signos vitales
- ✓ Laboratorio
- ✓ Notas clínicas
- ✓ Fluidos

# Retos en los EHR

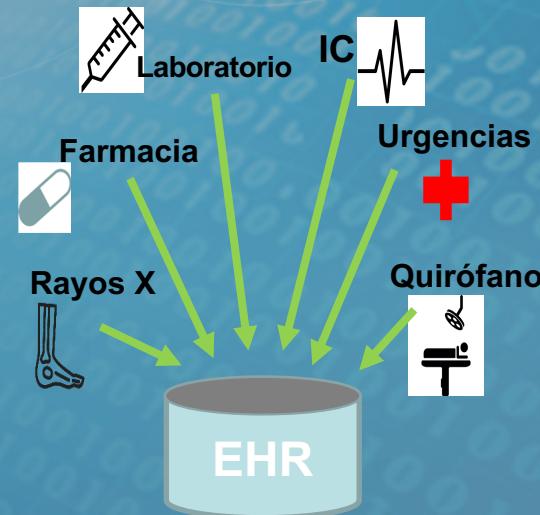


Mediciones periódicas		Datos del paciente		Diagnóstico		Datos de admisión	
pao2							
fio2							
pco2	Edad						
SVR	Género		CódigoICD9			Lugar de admisión	
SVRI	Etnia		Diagnóstico			Fecha de admisión	
PVR	Peso		Especialidad				
PVRI	Altura		Prioridad				
ph	Religión		Diagnóstico				
wbc	Estado C		previo				
Fecha	IMC		Tratamiento				

Alta dimensionalidad



✓ Ruido



✓ Complejidad / heterogeneidad

age	admissiondx	hospitaladmitsource	oncology
52	Emergency Department	Emergency Department	Yes
68	Floor	Floor	No
71	Emergency Department		
77	Arrest, respiratory	Floor	No
25	Asthma		No
82	Sepsis, pulmonary	Floor	Yes

✓ Ausencia de datos 44

# Retos en los EHR

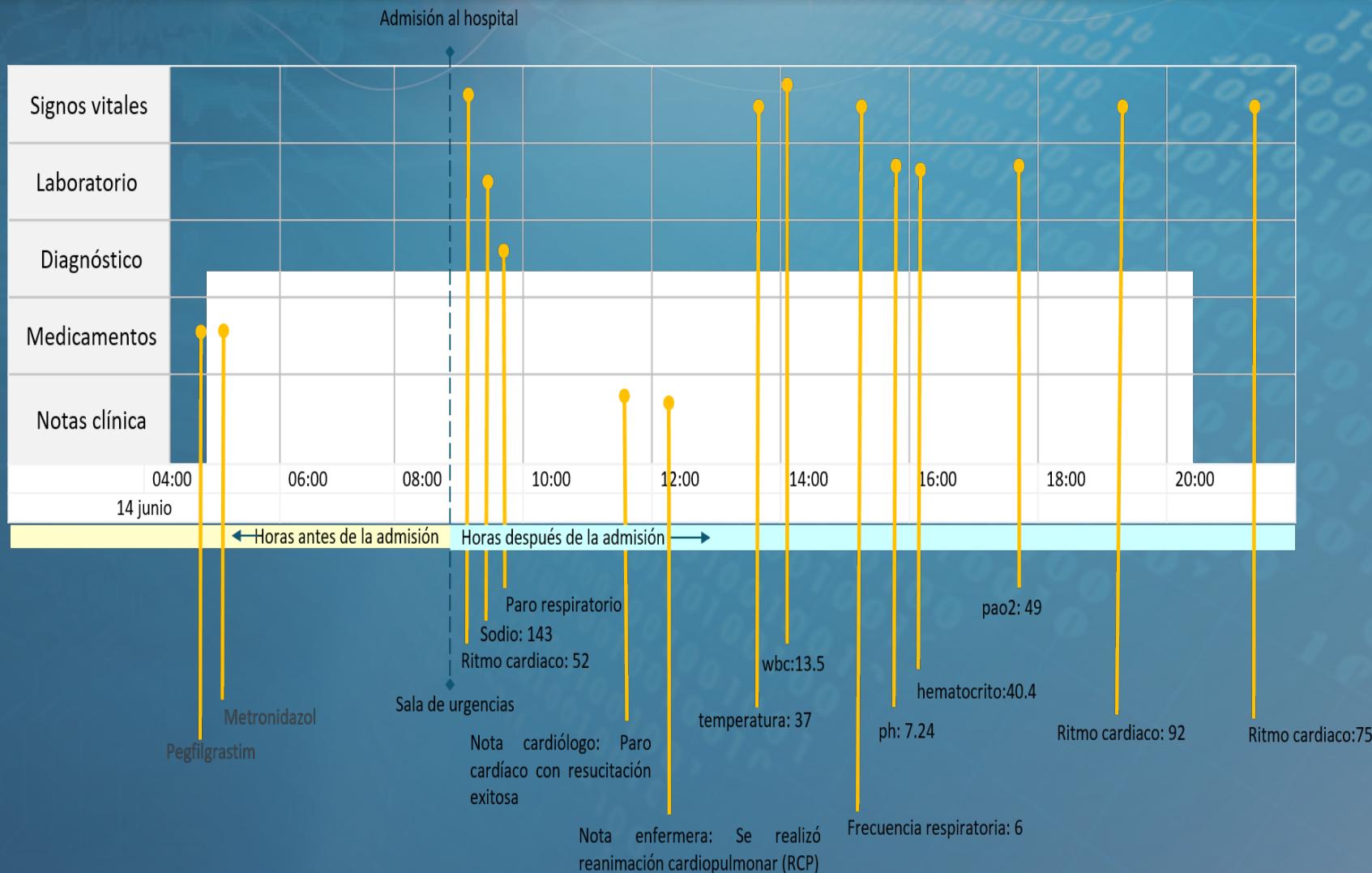


Academia  
de Ingeniería México

- **Diagnósticos confusos**
- **Referencia a medicamentos, alergias**
- **Redundancias en diagnósticos**
- **Inconsistencias**
- **Abreviaciones y tecnicismos (notas clínicas)**

**y más....**

# Secuencia de eventos clínicos



# Modelos predictivos



Academia  
de Ingeniería México

- ✓ Un modelo predictivo consiste en analizar datos históricos y actuales para predecir comportamientos futuros.
- ✓ El resultado de un modelo es la probabilidad de que un evento ocurra.



<https://www.exact.com/nl/software/over-exact/contact/locatie-zwolle/205-business-box/business/4171-digitale-transformatie-en-hr-een-blik-in-de-glazenbol/>

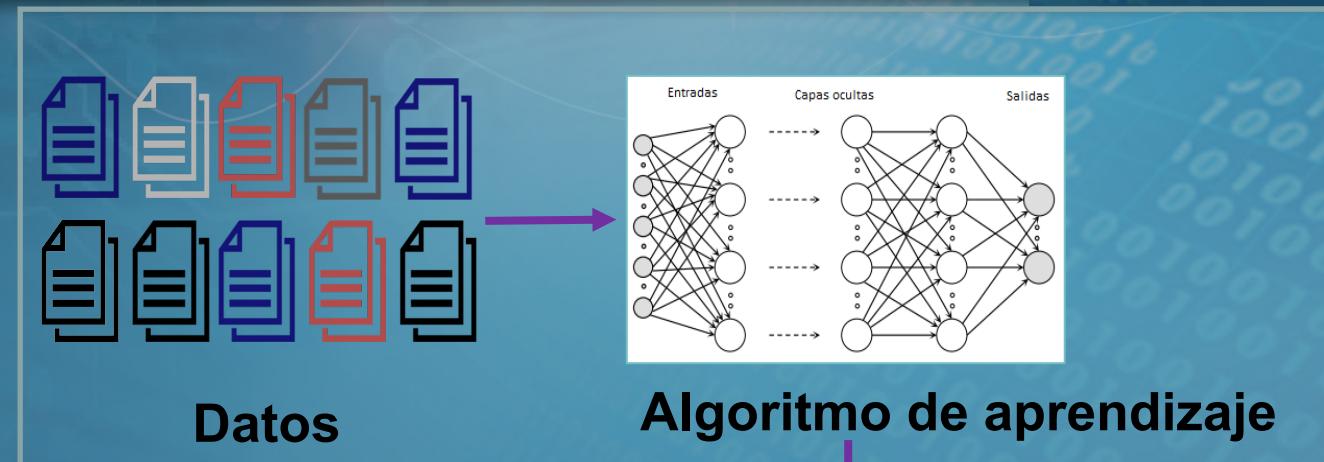
**Objetivo: aprender de los datos**

# ¿Cómo funcionan?

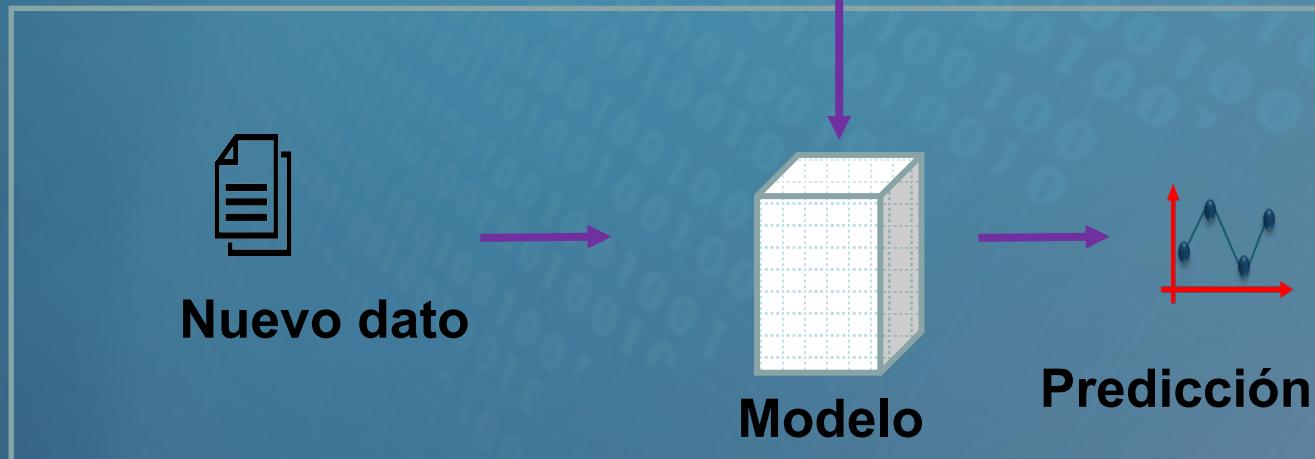


Academia  
de Ingeniería México

Entrenamiento



Inferencia



## Avances en técnicas de aprendizaje profundo:

- ✓ Procesamiento de grandes volúmenes de datos heterogéneos
- ✓ Tratamiento de relaciones temporales
- ✓ Generación de soluciones escalables
- ✓ Aprendizaje a partir datos complejos

# IA en el área de salud?



<https://southwesthealthcollaborative.org/workgroups/emergency-department-utilization/>

- **Recursos limitados en áreas clínicas**
- **Decisiones críticas**
- **Cada minuto cuenta**

# Estado del arte sobre EHR



Academia  
de Ingeniería México

Predicción de  
enfermedades

Predicción de eventos

Predicción de  
tratamientos

# Estado del arte sobre EHR



Academia  
de Ingeniería México

## Predicción de enfermedades

- Predicción de diagnósticos (ICD)
- Identificación de fenotipos

## Predicción de eventos

- Icahn School of Medicine at Mount Sinai, NY, USA.
  - Georgia Institute of Technology
  - KTH Royal Institute Of Technology
  - University of Southern California, USA
  - Google
  - IBM
  - Loyola University Chicago, USA
  - University of California, USA
  - University College London, UK
- New York University, USA

## Predicción de tratamientos

# Estado del arte sobre EHR



Academia  
de Ingeniería México

## Predicción de enfermedades

- Detección de infecciones
- Predicción de mortalidad
- Predicción de intervención clínica
- Predicción de la siguiente visita al hospital
- Predicción de sobrevivencia
- Predicción de readmisión hospitalaria

## Predicción de eventos

- School of Data Science, Shanghai, China
- Intel
- Department of Computer Science, CA, USA
- Google
- MIT
- Georgia Institute of Technology
- Columbia University, NY, USA
- Google

## Predicción de tratamientos

# Estado del arte sobre EHR



Academia  
de Ingeniería México

## Predicción de enfermedades

- Identificación de tratamientos individualizados
- Predicción de medicamentos

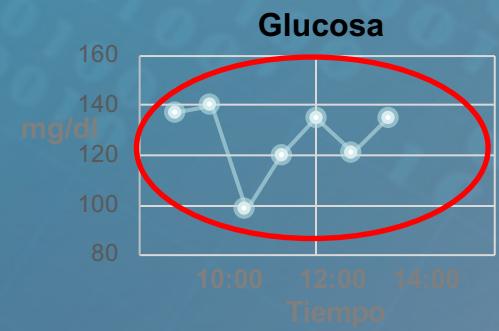
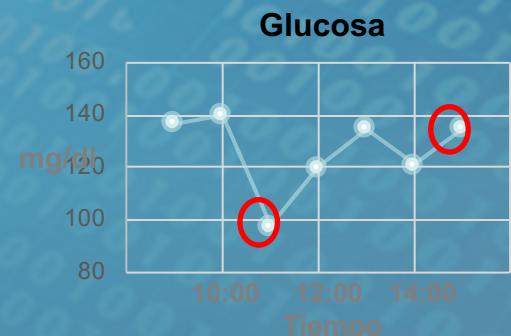
## Predicción de eventos

- Department of Biostatistics, Columbia University
- Georgia Institute of Technology

## Predicción de tratamientos

# Análisis del estado del arte

- Desarrollo de modelos enfocados más en tareas de preprocesamiento que en el análisis de datos.
- Enfoques que excluyen la temporalidad entre eventos
- Análisis de datos sin considerar diversidad poblacional
- No interpretables
- Soluciones no escalables



# Oportunidades



La mayoría de las enfermedades cardiovasculares son de progresión lenta, heterogéneas y con procesos crónicos, donde la patogénesis puede comenzar décadas antes de cualquier manifestación definitiva de la enfermedad.

Por tal motivo, consideramos importante la predicción de enfermedades cardiovasculares con anticipación, con el propósito de que el paciente tenga un **tratamiento oportuno y evitar daños irreversibles**, incluso evitar la mortalidad del paciente.

# Oportunidades



Academia  
de Ingeniería México

**Predecir la probabilidad de que una persona desarrolle una determinada enfermedad, dentro de un cierto período de tiempo, dado su estado clínico actual.**



## Preguntas abiertas a responder:

- ✓ Detección temprana de enfermedades cardiovasculares isquémicas
- ✓ Detección oportuna de eventos / episodios crónicos
  - Detección temprana de inflamación severa (sepsis)
  - Predicción de episodios hipotensores agudos
  - Detección de ataques agudos (cerebrovascular, hipertensión, isquémico)
- ✓ Probabilidad de sobrevivencia después de episodio / evento crónico

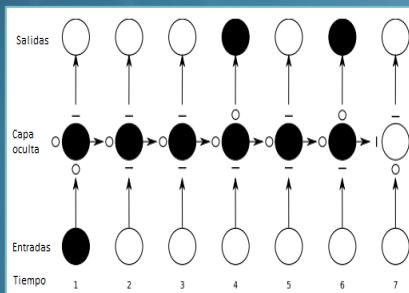
# Metodología propuesta



Academia  
de Ingeniería México



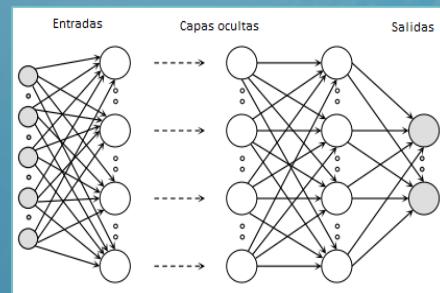
## Fase: Conversión de datos



Notas clínicas  
Diagnósticos  
Tratamientos  
Laboratorio  
Admisiones  
Reportes rayos X

Todos los datos de los pacientes se convertirán a eventos colocados en orden temporal (línea de tiempo)

## Fase: Predicción



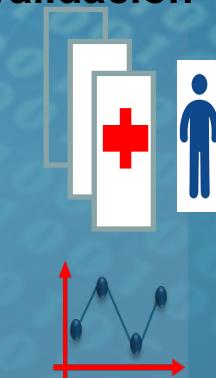
Se usará la historia completa de cada paciente, ordenada temporalmente, para realizar la predicción

## Fase: Definición de patrones

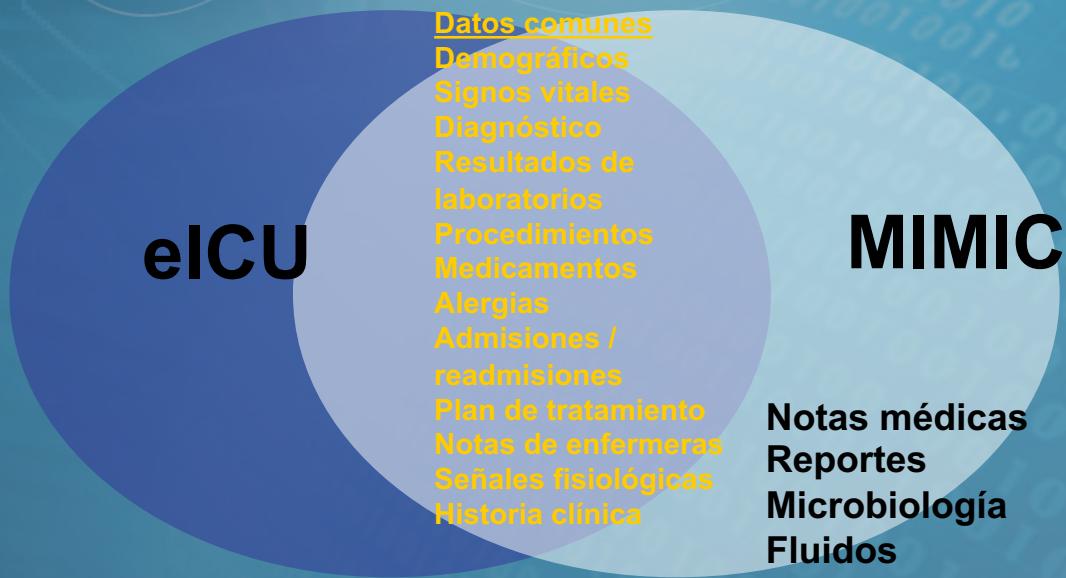
Hombr e	Mujer
Smokin g	Edad
Coleste ro	Sobrep eso
COPD	Fumad or

Identificación de patrones clínicos

## Fase: Validación



# Bases de datos



Total de pacientes: 139,368

Total de admisiones: 200,859

Periodo de recolección: 2011-2015

Fuente de datos: 212 hospitales

Tamaño: 18.4 GB

Total de pacientes: 38,597

Total de admisiones: 49,785

Periodo de recolección: 2001-2012

Fuente de datos: 1 hospital

Tamaño: 31 GB

# Experiencia del grupo



Academia  
de Ingeniería México

- Desarrollo de arquitecturas y algoritmos para cómputo de alto rendimiento
- Procesamiento de señales, imágenes y control en tiempo real
- Aprendizaje automático y visión por computadora
- Procesamiento y análisis de grandes cúmulos de información (Big data)
- Desarrollo de sistemas Doppler para valoración de flujo en cirugía de revascularización coronaria.