



Universidad Nacional Autónoma de México
Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas



Predicción de mortalidad hospitalaria para pacientes con SARS-COV-2 basado en aprendizaje de máquinas

Blanca Hilda Vázquez Gómez

Escuela de ciencia de datos

Aprendizaje de máquinas

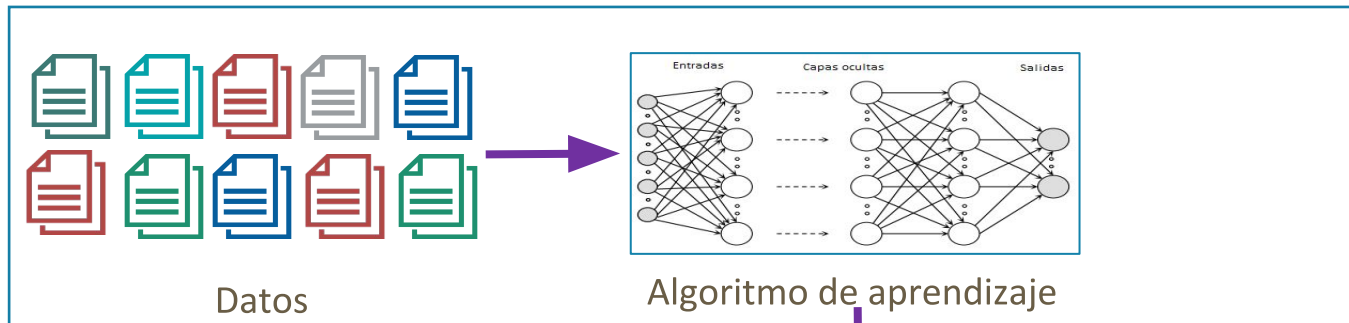
- ❖ Es el estudio de programas que aprenden a partir de ejemplos para estimar comportamientos futuros.
- ❖ El resultado de un programa es la probabilidad de que un evento ocurra.



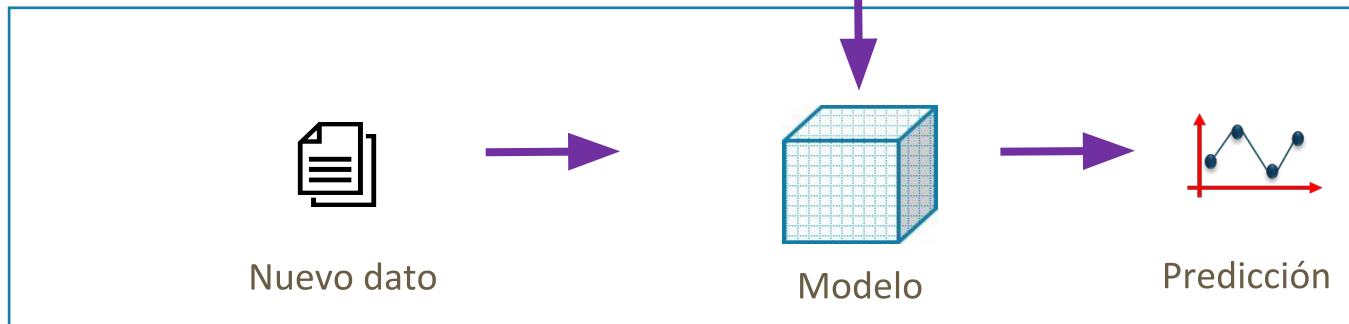
La idea: ¡aprender de los datos!

¿Cómo funciona?

Entrenamiento



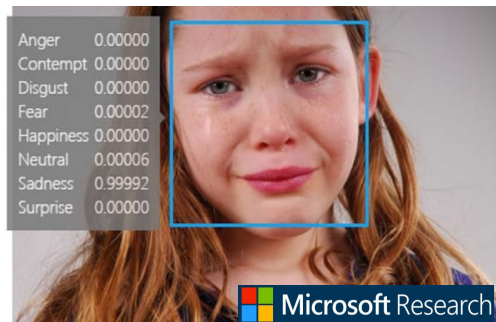
Inferencia



Ejemplos de aplicaciones



Reconocimiento de rostros



Reconocimiento de emociones

Sugerir palabras



Detección de derrame pleural

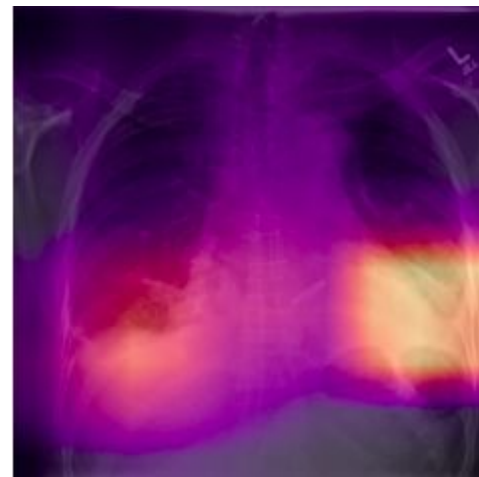


Imagen tomada de CheXpert, 2019.

Registros clínicos electrónicos



Los EHR son colecciones de datos que describen la historia clínica de los pacientes

Aprendizaje de máquinas basado en EHR

Diagnóstico	Pronóstico	Tratamientos
<p>Predicción de una enfermedad:</p> <ul style="list-style-type: none">- Oncología- Patología- Enfermedades neuro-degenerativas- Enfermedades raras- Enfermedades respiratorias	<p>Predecir el riesgo de un evento futuro:</p> <ul style="list-style-type: none">- Falla cardíaca- Mortalidad- Accidente cerebro-vascular- Estimar la supervivencia	<p>Estimar los efectos de un tratamiento:</p> <ul style="list-style-type: none">- Medicina personalizada

COVID-19

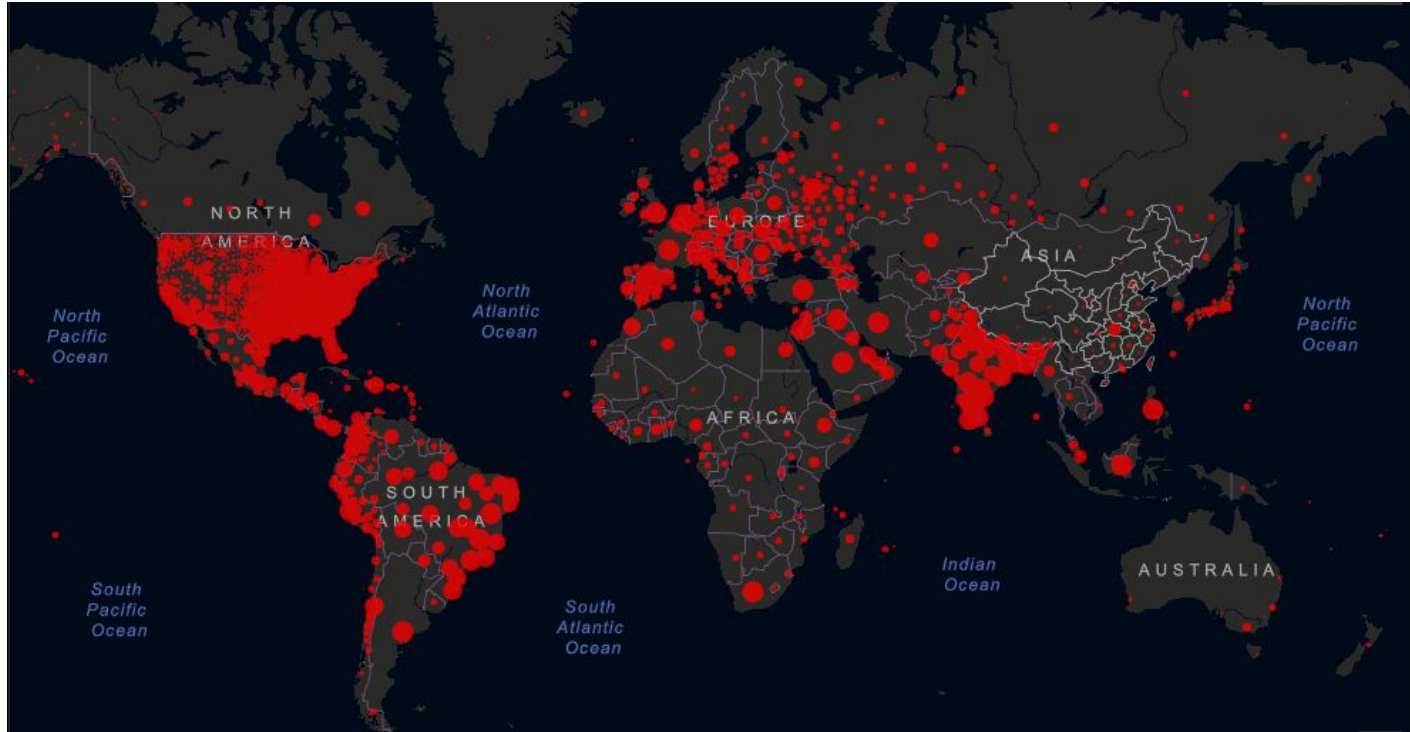


Imagem tomada de *Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)*

Aprendizaje de máquinas ante la pandemia

Objetivo	Modelos empleados	EHR usados en la construcción de los modelos					
		Rayos X	Lab	Demo	Signos vitales	Comorbilidades	Tratamientos
Detección	Máquinas de soporte vectorial [1], árboles aleatorios [2], redes convolucionales [3-4]	x	x	x			
Pronóstico	Regresión logística [5], XGB [6]		x	x	x		
Mortalidad	Algoritmo de XGB [7], análisis multivariable [8]			x	x	x	x

¿Por qué es importante estimar la mortalidad?



- Estratificación del riesgo
- Ajustar estrategias de tratamientos
- Recursos limitados
- Preservar la vida

Objetivos

- Desarrollar un modelo para estimar el riesgo de mortalidad en pacientes con SARS-COV-2 basado en aprendizaje de máquinas.
- Identificar las variables clínicas que incrementan el riesgo de mortalidad en los pacientes.

Base de datos

Para el desarrollo de este estudio, se emplearon los datos abiertos disponibles por la Secretaría de Salud.




The screenshot shows a web interface for the Mexican government's data portal. On the left is a sidebar with the 'SALUD' logo and the text 'SECRETARÍA DE SALUD'. The main content area has a navigation bar with 'Conjunto de datos' (selected), 'Actividad', and 'Grupos'. A red 'Reportar' button is in the top right. The main heading is 'Información referente a casos COVID-19 en México', followed by 'Datos y Recursos'. Below this is a card for 'Bases de datos COVID 19 en México' with a green document icon. The card text includes 'Información del Sistema de Vigilancia Epidemiológica de Enfermedades Respiratoria Viral.' and a note about preliminary data. Two green buttons, 'Descargar' and 'Más información', are on the right side of the card.

SALUD

Conjunto de datos Actividad Grupos Reportar

Información referente a casos COVID-19 en México

Datos y Recursos

 **Bases de datos COVID 19 en México**

Información del Sistema de Vigilancia Epidemiológica de Enfermedades Respiratoria Viral.
Nota: Datos preliminares sujetos a validación por la...

Descargar Más información

Fuente: <https://datos.gob.mx/busca/dataset/informacion-referente-a-casos-covid-19-en-mexico>

Variables disponibles

Demográficos	Comorbilidades	Datos hospitalarios
Sexo, edad, nacionalidad, habla lengua indígena, indígena, migrante, país nacionalidad, país origen, entidad, sector, origen, embarazo.	Diabetes, epoc, asma, obesidad, hipertensión, cardiovascular, renal crónica, tabaquismo, inmunosupresión, otra comorbilidad.	Clasificación final, fecha ingreso, fecha síntomas, fecha def, intubado, tipo paciente, uci, neumonía.

90% datos son binarios

Análisis exploratorio de los datos

2,183,272

Registros de pacientes

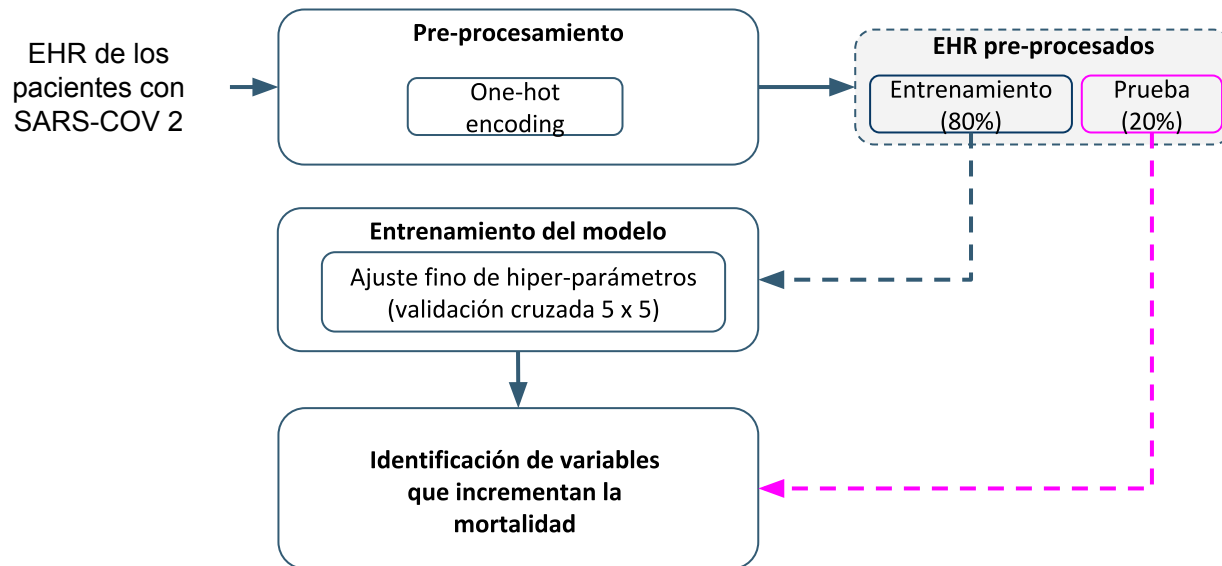
848,382

Pacientes con SARS-COV 2

86,037

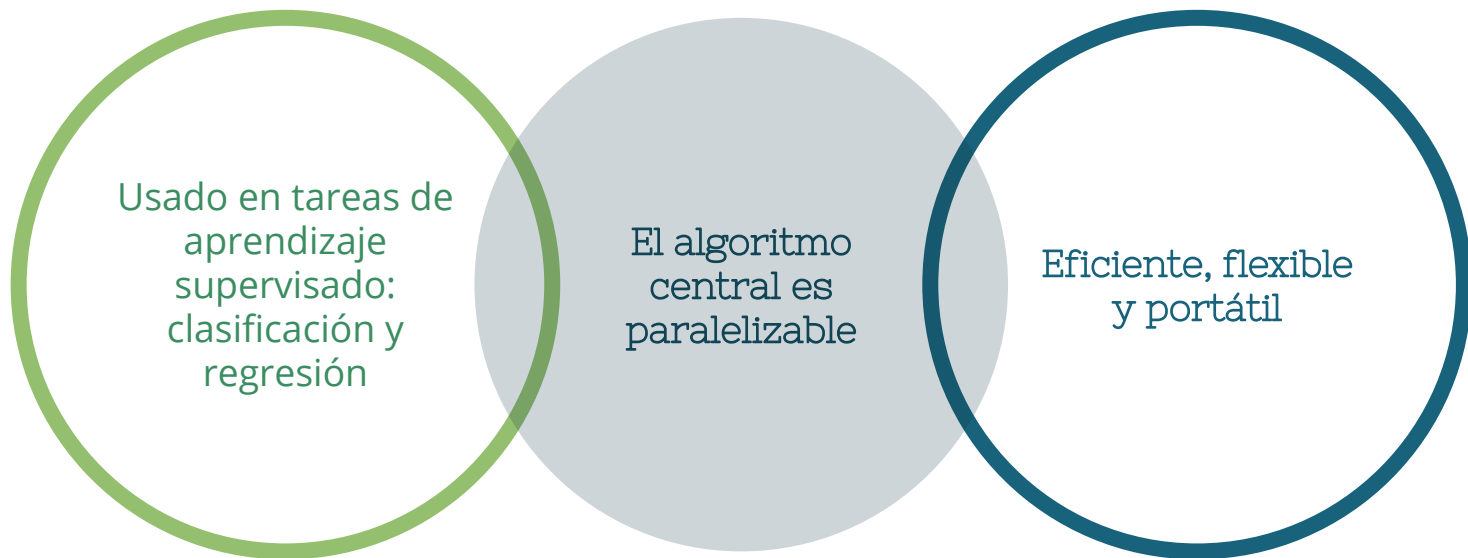
Pacientes fallecidos

Metodología



Árboles de potenciación del gradiente

XGBoost (*eXtreme Gradient Boosting*)



Enfoque de interpretabilidad

Con el objetivo de interpretar e identificar las variables que incrementan la mortalidad, se empleó el enfoque SHAP (*SHapley Additive exPlanations*)[9].

- Es un enfoque basado en la teoría de juegos:
 - * donde el juego es la tarea de predicción para una instancia y los jugadores son los valores de las características.
- Permite obtener una interpretabilidad global, local e individual.
- Ha sido empleado para identificar marcadores de riesgo para enfermedades renales[10] e hipoxemia[11].

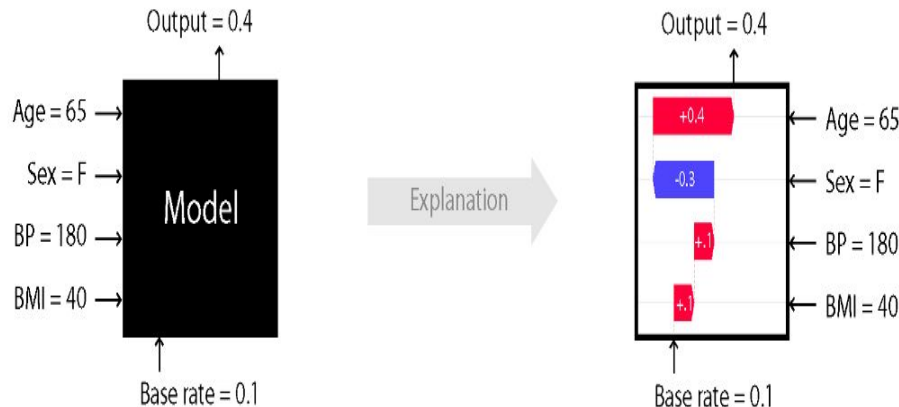
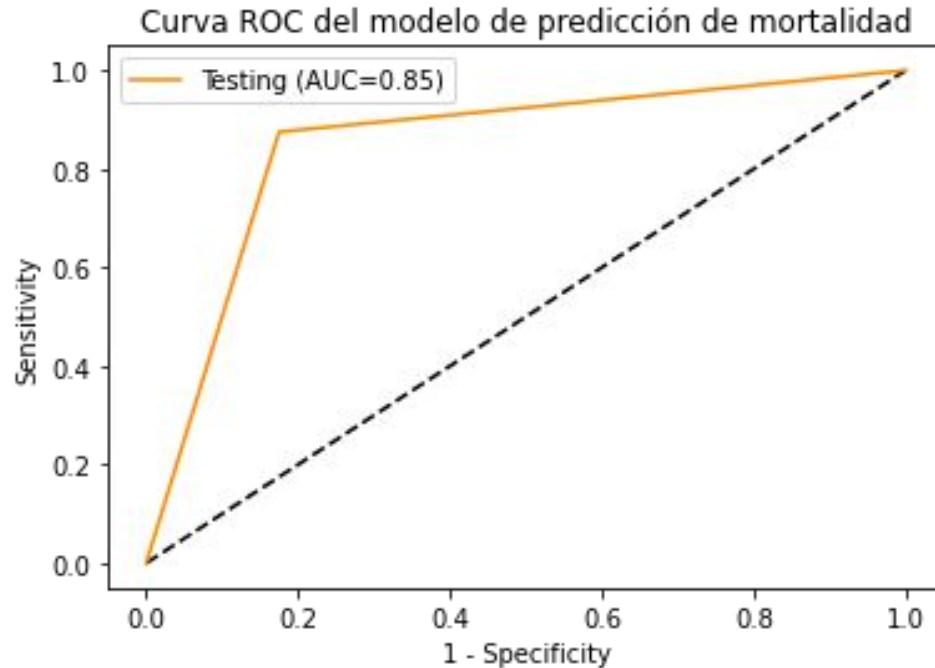


Imagen tomada de <https://github.com/slundberg/shap>

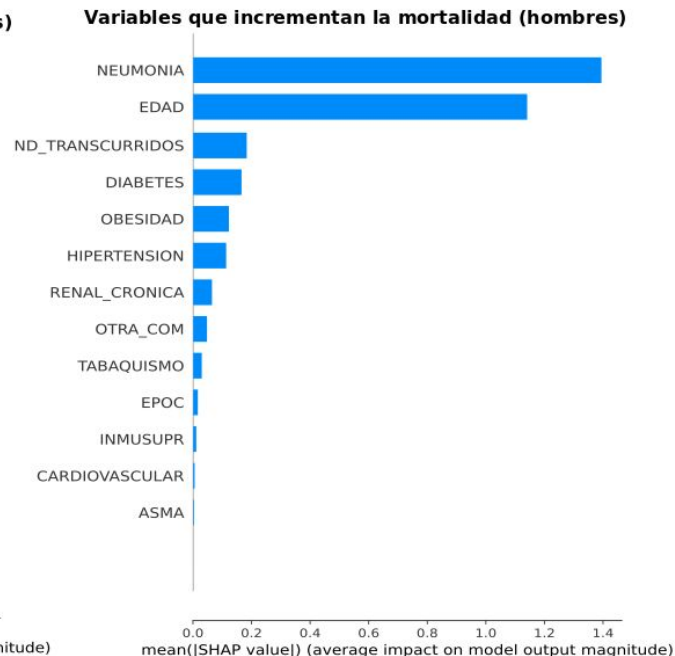
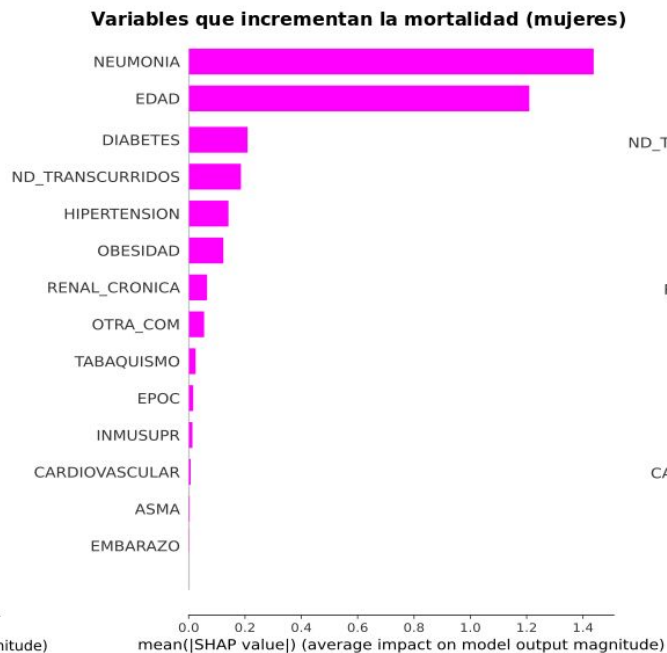
Libreta

https://github.com/blancavazquez/escuelaCD_covid

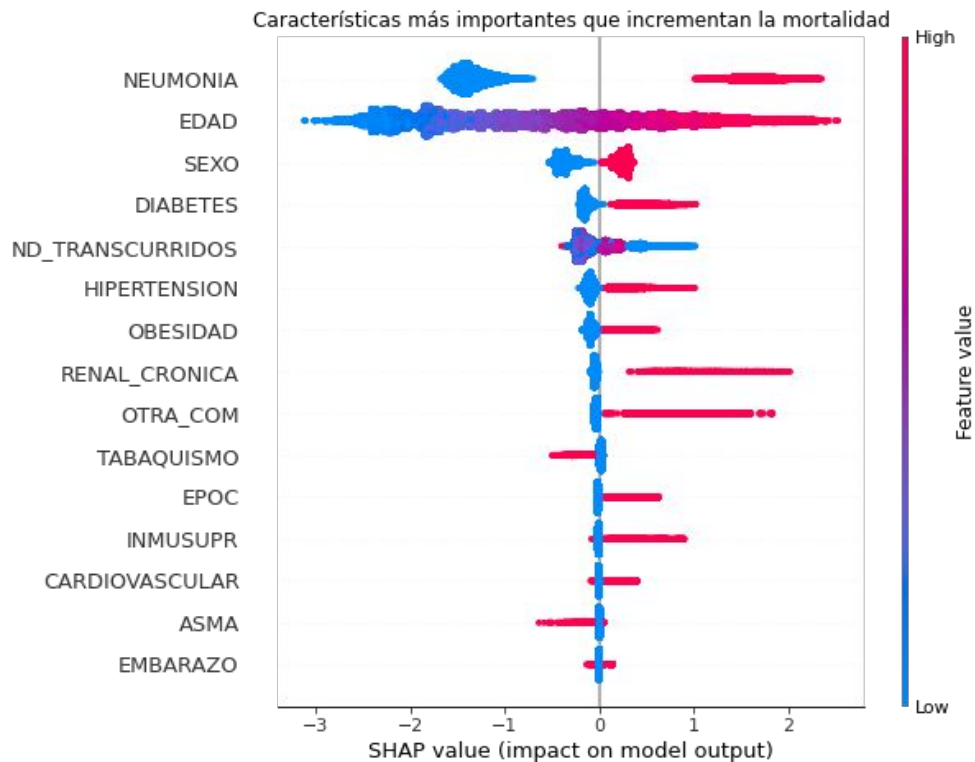
Rendimiento del modelo de predicción de mortalidad



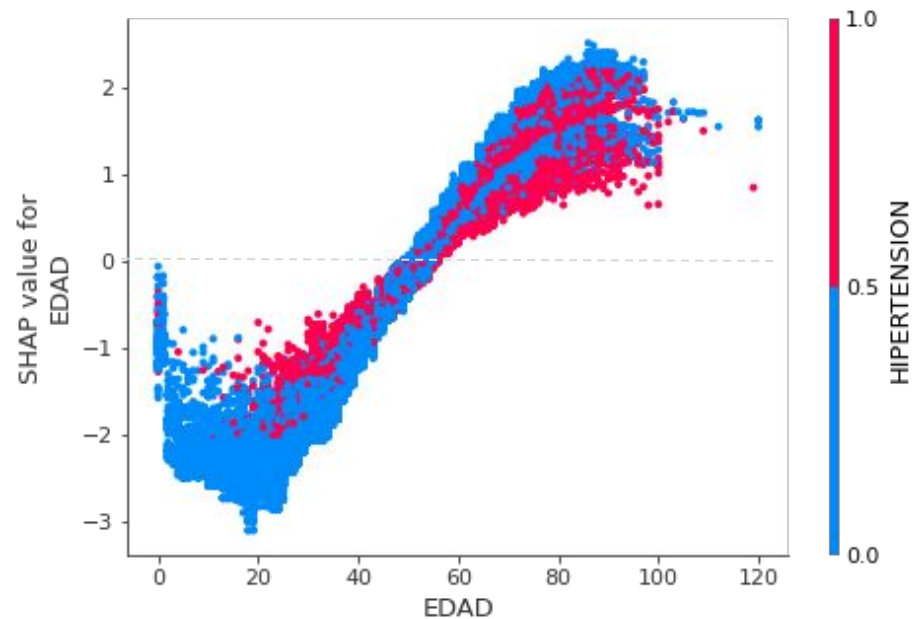
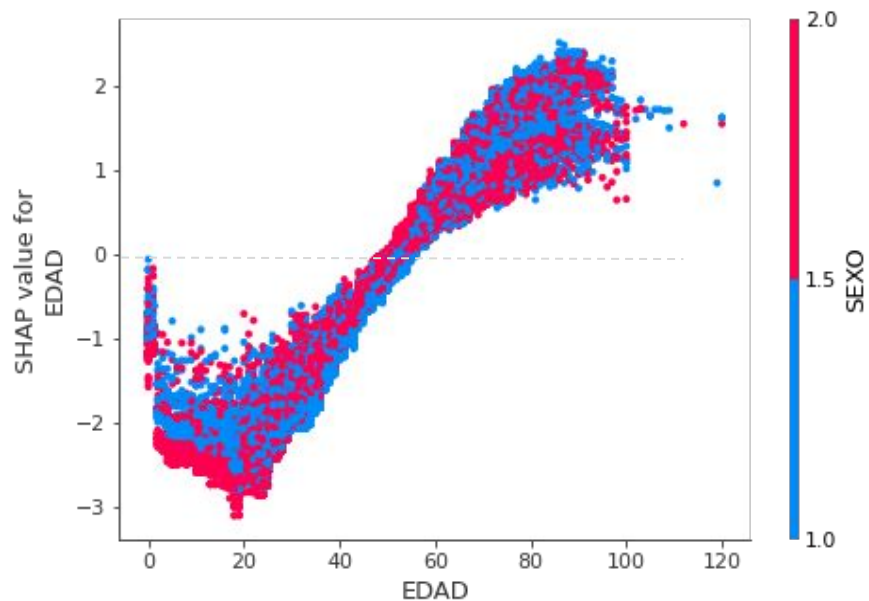
Variables que incrementan la probabilidad de mortalidad en pacientes con SARS-COV 2



Variables que incrementan la probabilidad de mortalidad en pacientes con SARS-COV 2



Interacciones entre variables clínicas



Hallazgos / limitantes

Técnico:

- Naturaleza de los datos
 - Sí / No
- Datos faltantes
- Número limitado de variables clínicas
- Para la recolección, se empleó un cuestionario predefinido

Clínico:

- Desconocimiento del padecimiento de comorbilidades

Áreas de oportunidad

- Promover iniciativas para compartir datos clínicos (preservando la privacidad y seguridad de los pacientes).
- Capturar datos que incluyan:
 - Resultados de laboratorio
 - Signos vitales
 - Tratamientos
 - Síntomas
- Capturar la temporalidad de los datos / eventos clínicos

Áreas de oportunidad

Se ha identificado que la calidad y el tipo de datos, pueden contribuir a:

- Mejorar el rendimiento de los modelos
 - Diagnóstico
 - Readmisión hospitalaria
 - Calcular el tamaño de la estancia en UCI
 - Mortalidad
 - Generar de planes de tratamiento
- Identificar mejores marcadores de riesgo (patrones ocultos)



Identificación de marcadores de riesgo de mortalidad hospitalaria por sexo y grupo de edad en pacientes con STEMI y NSTEMI basado en aprendizaje de máquinas

Blanca Vázquez, Gibran Fuentes, Fabian Nocetti,
Gabriela Borrayo, Juan Prohías

Infarto agudo al miocardio

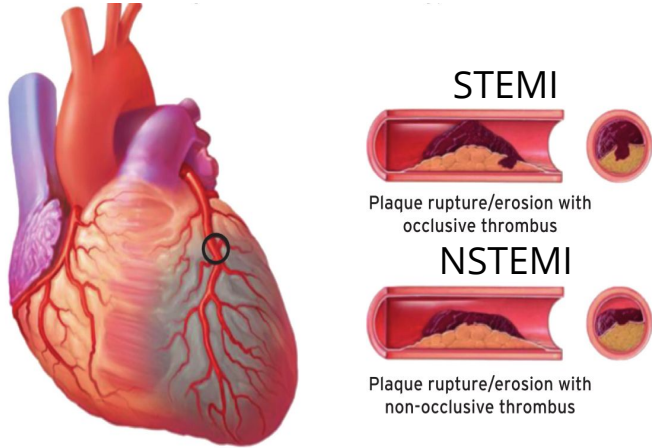


Imagen tomada de: Fourth Universal Definition of Myocardial Infarction (2018).

- Son la primera causa de muerte en el mundo.
- De acuerdo al INEGI, en el 2016 se reportaron 175,078 muertes (1 cada 3 minutos), 70% por Infarto Agudo al Miocardio.
- Los tipos más comunes de infarto:
 - Infarto agudo con elevación del segmento ST (STEMI)
 - Infarto agudo sin elevación del segmento ST (NSTEMI)

Problema

En el pasado, las enfermedades cardiovasculares fueron percibidas como un problema de salud solo para hombres.



Las mujeres fueron sub-representadas en los ensayos clínicos

Recibieron la misma atención clínica

Es necesario distinguir marcadores por sexo y edad

Ambos sexos, tienen diferentes características biológicas y fisiológicas

Múltiples investigaciones han abordado la distinción de marcadores por género usando enfoques estadísticos: readmisión, complicaciones, mortalidad ...



Otros enfoques, han usado aprendizaje de máquinas para predecir eventos clínicos: regresión logística, árboles aleatorios, análisis multivariado ...

Cada tipo de infarto difiere en la ubicación y la extensión del daño al miocardio

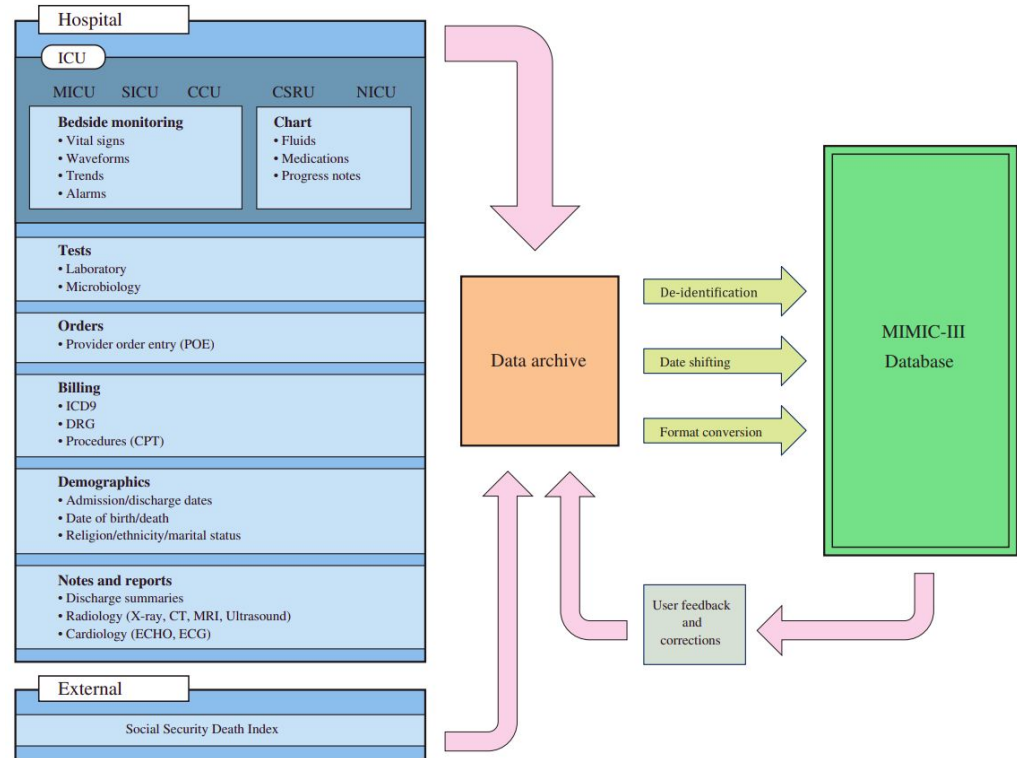
No es clara la distinción de marcadores de mortalidad por sexo y edad en pacientes con STEMI y NSTEMI usando aprendizaje de máquinas

Objetivos

- Explorar modelos de aprendizaje de máquinas para predecir mortalidad hospitalaria durante las primeras 24 horas de admisión a ICU después de un STEMI o NSTEMI.
- Distinguir los marcadores de riesgo de mortalidad hospitalaria por sexo y grupo de edad para cada sub-población.

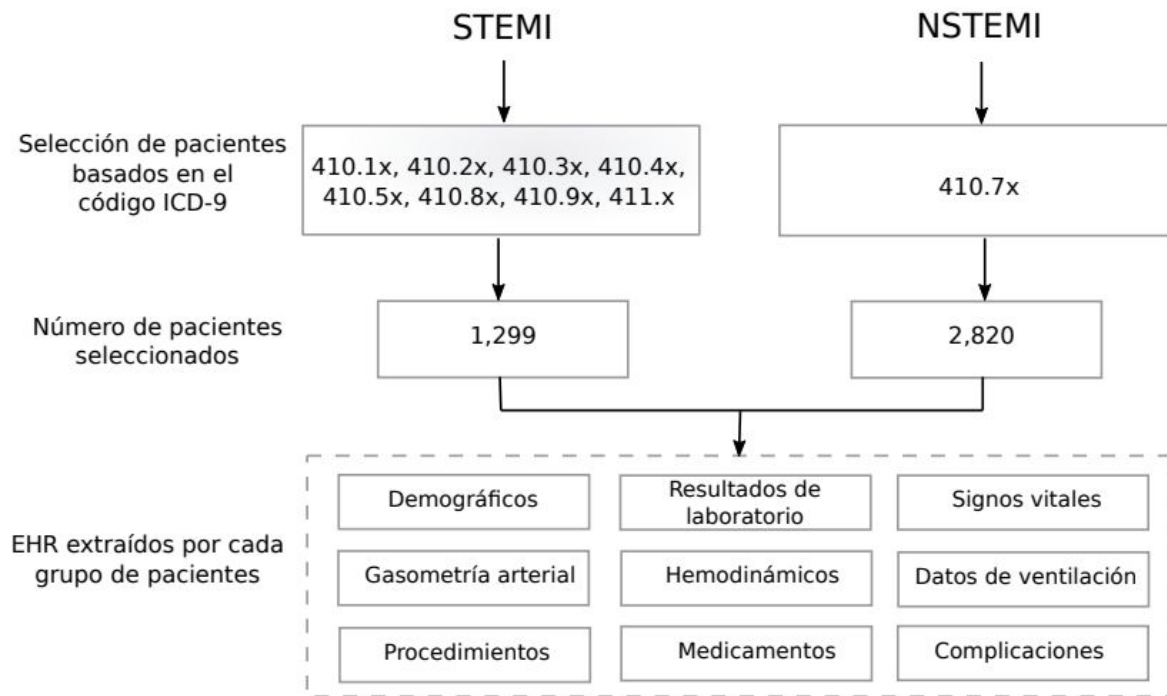
Base de datos

Se empleó la base de datos MIMIC-III la cual describe pacientes admitidos a la UCI ingresados al *Beth Israel Deaconess Medical Center* entre el 2001 y el 2012.

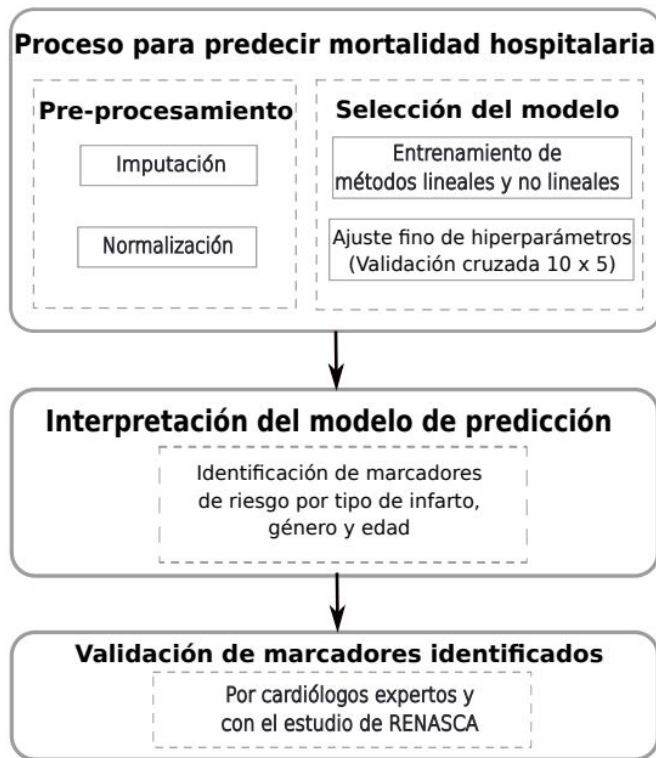


Vista general de la base de datos MIMIC-III
Imagen tomada de Johnson, 2016 [12]

Selección de la población de pacientes



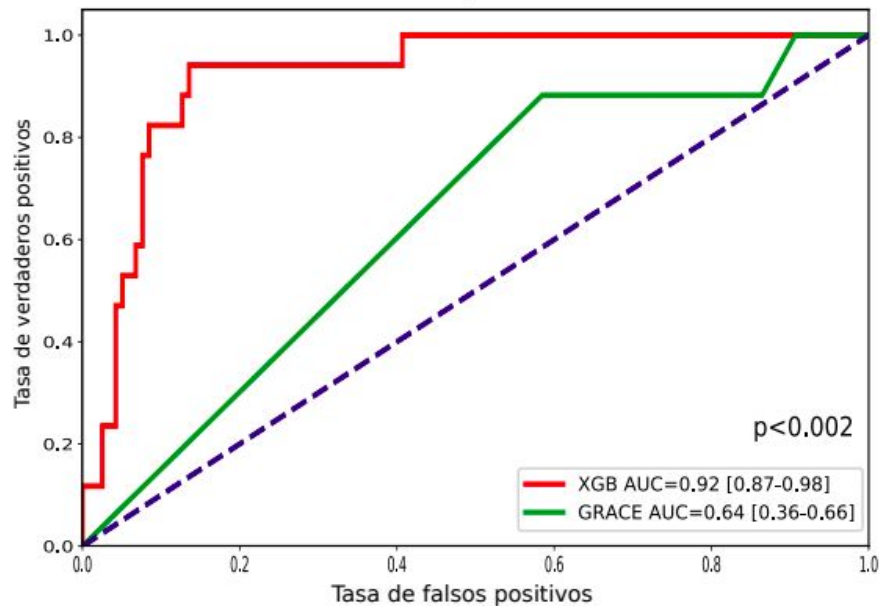
Proceso para predecir mortalidad hospitalaria e identificación de marcadores para pacientes ingresados a una UCI con STEMI y NSTEMI



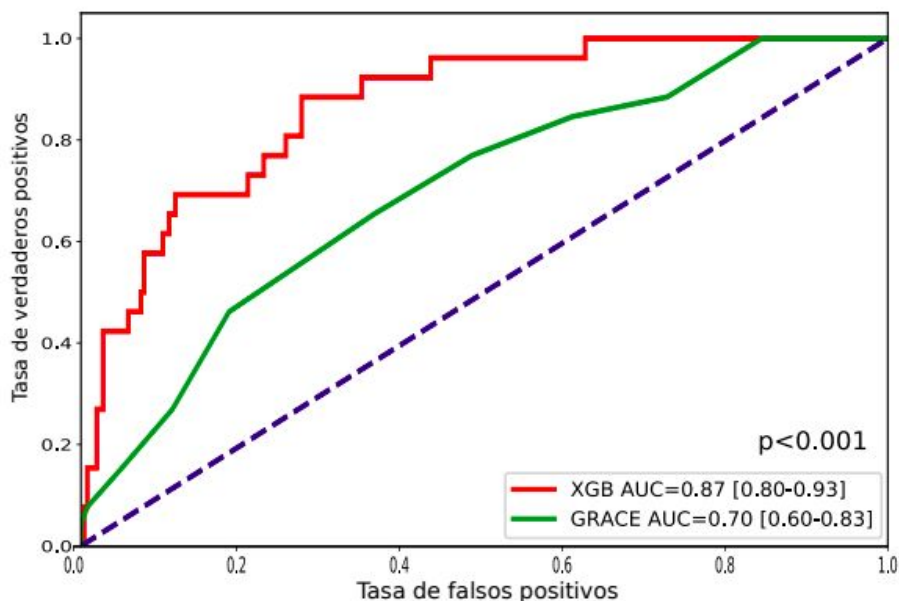
Resultados

Predicción de mortalidad hospitalaria para pacientes ingresados a una UCI después de un STEMI y NSTEMI

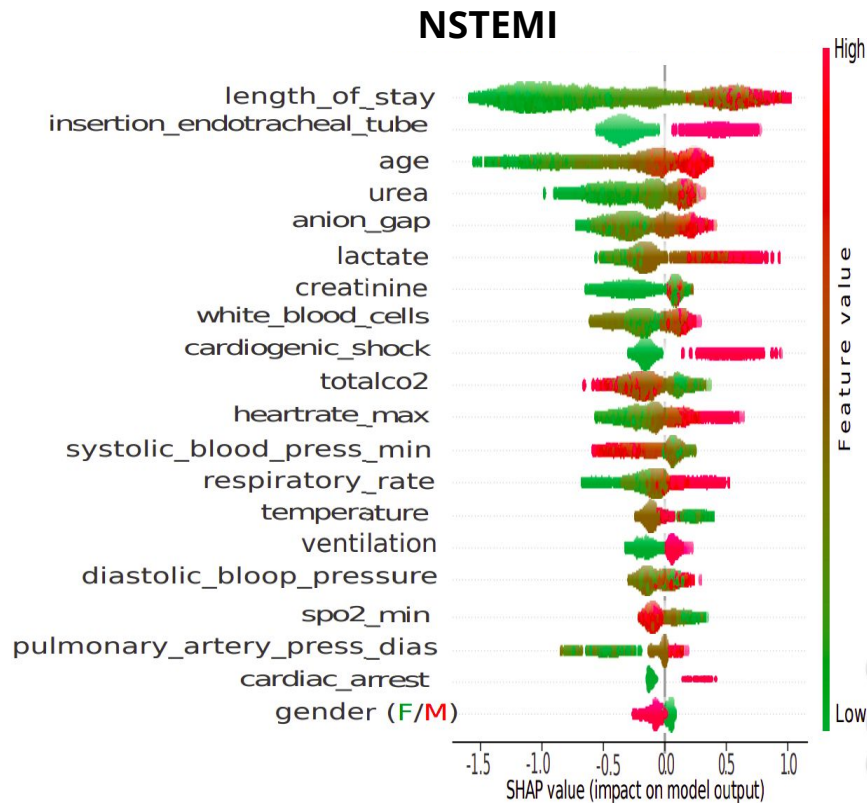
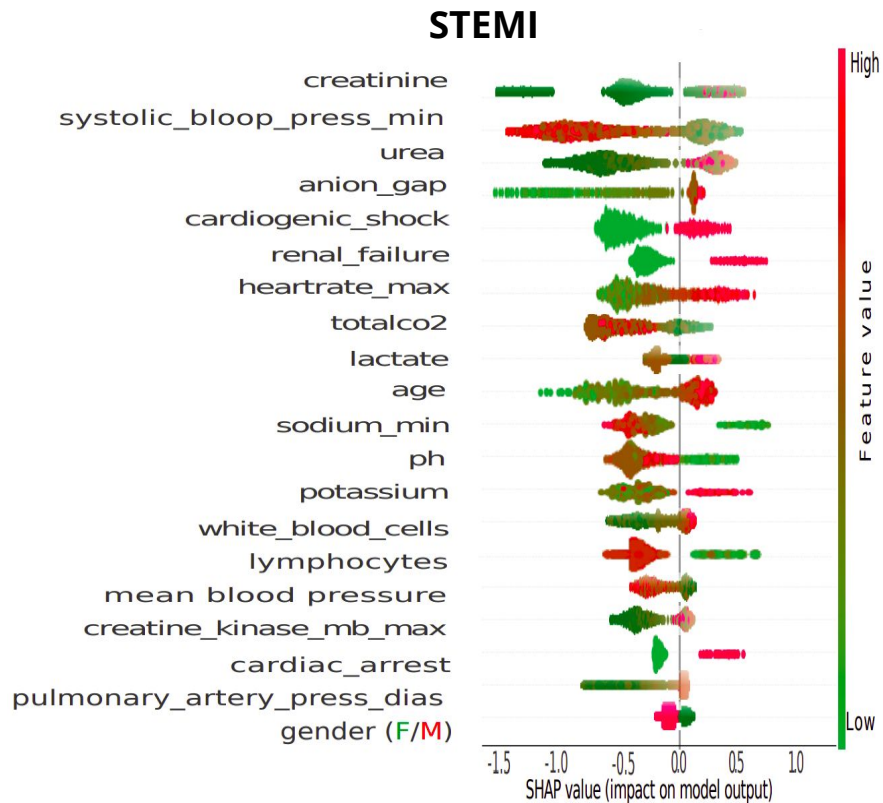
Curva ROC para STEMI



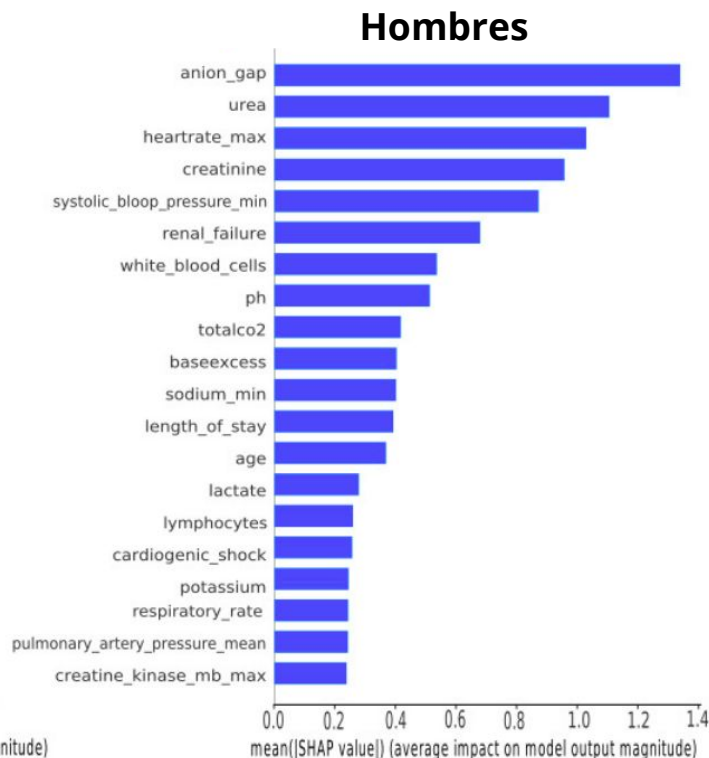
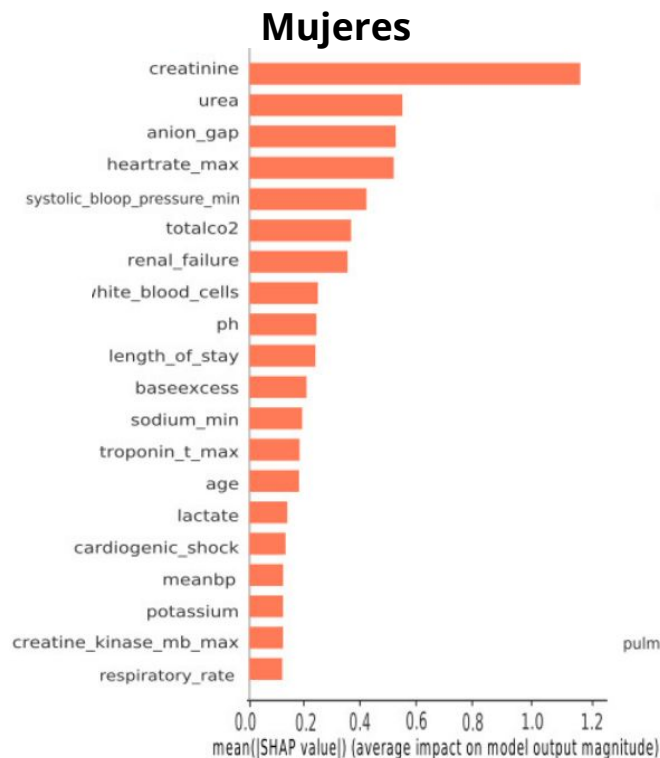
Curva ROC para NSTEMI



Marcadores de riesgo que incrementan la probabilidad de mortalidad hospitalaria

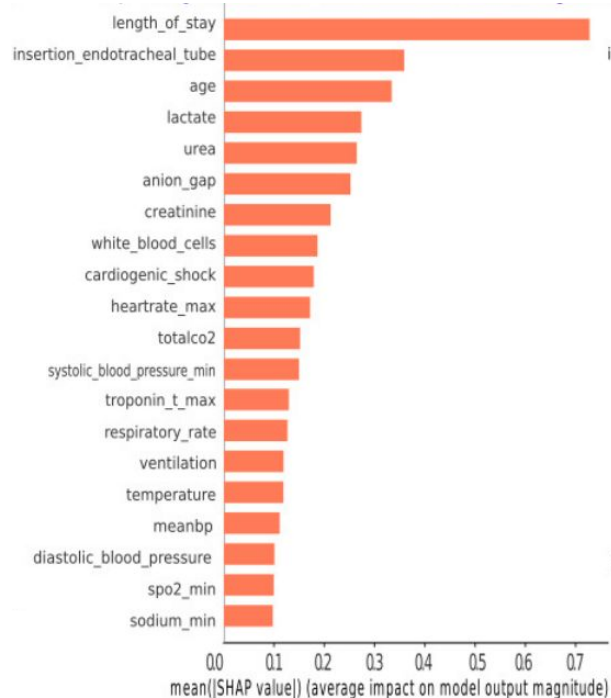


Marcadores de riesgo que incrementan la probabilidad de mortalidad hospitalaria para pacientes con STEMI

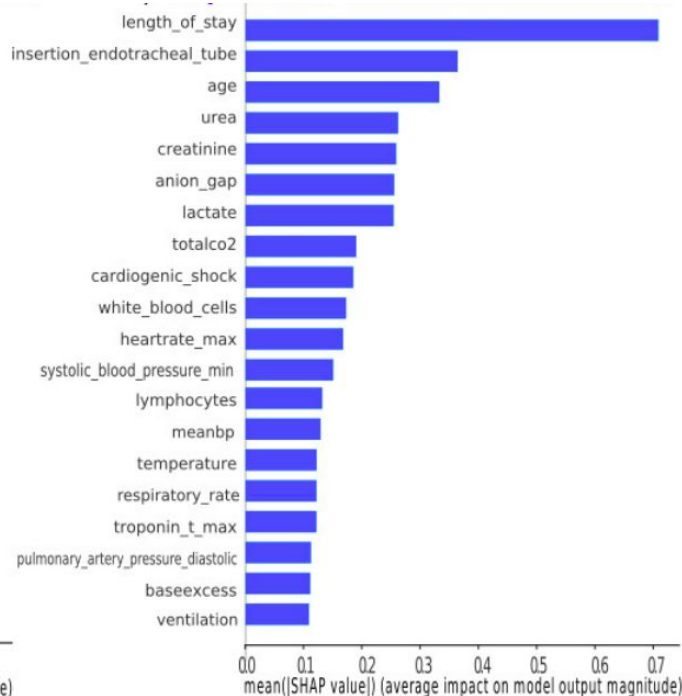


Marcadores de riesgo que incrementan la probabilidad de mortalidad hospitalaria para pacientes con NSTEMI

Mujeres



Hombres



Resumen de marcadores de riesgo identificados

STEMI - MUJERES

- Insuficiencia renal aguda
- Troponina > 5 ng/L
- Sodio < 135 mEq/L
- PH > 7.6
- Edad > 75 años.

STEMI - HOMBRES

- Insuficiencia renal crónica
- Presión arterial pulmonar alta
- Sodio < 135 mEq/L
- PH > 7.32
- Edad > 70

NSTEMI - MUJERES

- Estadías prolongadas
- Creatinina (<20 umol/L)
- CO2 total < 25 mEq/L
- Tasa de respiration > 25
- Edad > 60 años

NSTEMI - HOMBRES

- Estadías prolongadas
- CO2 total < 23 mEq/L
- Troponin > 2 ng/L
- Daño en la aurícula izquierda
- Edad > 70 años.

Hallazgos

- Evidenciamos que los modelos de predicción basados en XGB, logran una estimación del riesgo más precisa, en comparación con el sistema de puntaje más utilizado.
- Nuestros resultados mostraron que el aprendizaje de máquinas puede contribuir en la identificación de marcadores específicos para sub-poblaciones de pacientes con enfermedades cardiovasculares.
- Consideramos que la distinción de marcadores entre mujeres y hombres puede ayudar a definir estrategias de tratamiento más adecuados, reduciendo así los riesgos y mejorando los resultados clínicos.

Conclusiones y retos

- La construcción de modelos pueden contribuir a generar herramientas auxiliares en el área clínica.
- Es necesario promover iniciativas de apertura de datos (preservando la privacidad de los pacientes).
- La integración de equipos multidisciplinario pueden contribuir a refinar los resultados y a mejorar los resultados clínicos
- La calidad y los tipos de datos pueden contribuir a construir mejores modelos de predicción para mejorar y acelerar el diagnóstico, el pronóstico y el tratamiento de COVID-19.

Gracias

Blanca Hilda Vázquez Gómez
Posgrado en Ciencia e Ingeniería de la Computación
IIMAS / UNAM

Referencias

- [1] Sun L, Liu G, Song F, Shi N, Liu F, Li S, Li P, Zhang W, Jiang X, Zhang Y, Sun L, Chen X, Shi Y. Combination of four clinical indicators predicts the severe/critical symptom of patients infected COVID-19.. J Clin Virol 2020:104431. doi:10.1016/j.jcv.2020.104431.
- [2] Wu, J., Zhang, P., Zhang, L., Meng, W., Li, J., Tong, C., Li, Y., Cai, J., Yang, Z., JZhu, J., Zhao, M., Huang, H., Xie, X. and Li, S. (2020). Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results. medRxiv Preprint doi: <https://www.medrxiv.org/content/10.1101/2020.04.02.20051136v1>
- [3] Ardakani AA, Kanafi AR, Acharya UR, Khadem N, Mohammadi A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: results of 10 convolutional neural networks. Comput Biol Med 2020;121:103795.2020 <https://doi.org/10.1016/j.compbiomed.2020.103795>.
- [4]] Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra Acharya U. Automated detection of COVID-19 cases using deep neural networks with Xray images. Comput Biol Med 2020:103792. doi:10.1016/j.compbiomed.2020.103792.
- [5] Khanday, A.M.U.D., Rabani, S.T., Khan, Q.R. et al. Machine learning based approaches for detecting COVID-19 using clinical text data. Int. j. inf. tecnol. 12, 731–739 (2020). <https://doi.org/10.1007/s41870-020-00495-9>
- [6] Ribeiro MHDM, da Silva RG, Mariani VC, Coelho LDS. Short-term forecasting COVID-19 cumulative confirmed cases: perspectives for Brazil.. Chaos, Solitons Fractals 2020:109853. doi:10.1016/j.chaos.2020.109853.

Referencias

- [7] Arjun S Yadaw, Yan-chak Li, Sonali Bose, Ravi Iyengar, Supinda Bunyavanich, Gaurav Pandey, Clinical features of COVID-19 mortality: development and validation of a clinical prediction model, The Lancet Digital Health, Volume 2, Issue 10, 2020, Pages e516-e525, ISSN 2589-7500, [https://doi.org/10.1016/S2589-7500\(20\)30217-X](https://doi.org/10.1016/S2589-7500(20)30217-X).
- [8] Yufeng Shang, Tao Liu, Yongchang Wei, Jingfeng Li, Liang Shao, Minghui Liu, Yongxi Zhang, Zhigang Zhao, Haibo Xu, Zhiyong Peng, Xinghuan Wang, Fuling Zhou, Scoring systems for predicting mortality for severe patients with COVID-19, EClinicalMedicine, Volume 24, 2020, 100426, ISSN 2589-5370, <https://doi.org/10.1016/j.eclinm.2020.100426>.
- [9] Lundberg, S.M., Lee, S.-I.: A Unified Approach to Interpreting Model Predictions. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems 30, pp. 4765–4774. Curran Associates, Inc., (2017).
- [10] Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I.: From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence 2 (2020). doi:10.1038/s42256-019-0138-9. Accessed 2020-02-13.
- [11] Lundberg, S.M., Nair, B., Vavilala, M.S., Horibe, M., Eisses, M.J., Adams, T., Liston, D.E., Low, D.K.-W., Newman, S.-F., Kim, J., Lee, S.-I.: Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nature Biomedical Engineering 2 (2018). doi:10.1038/s41551-018-0304-0.
- [12] A. E. W. Johnson, T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R. G. Mark, “MIMIC-III, a freely accessible critical care database,” Scientific Data, vol. 3, p. 160035, May 2016.