

Ficha Técnica de los Algoritmos para predicción de complicación de acuerdo a la Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud (CIE-10)

Proyecto: “Estudio longitudinal para el desarrollo de modelos predictivos de complicaciones crónicas de la diabetes mellitus tipo 2”

Registro: PRONACES 2019-06 número de registro 10410

1. Objetivo general del proyecto	Construir un modelo predictivo con alta sensibilidad y especificidad que identifique el riesgo que tiene un paciente con DM2 para desarrollar complicaciones crónicas.
2. Propósito del registro	Construcción del modelo con algoritmos para predecir las complicaciones crónicas de la diabetes mellitus tipo 2 de acuerdo al CIE-10 (E11.2: Nefropatía; CIE E11.3: Retinopatía; CIE E11.4: Neuropatía; CIE E11.5: Pie Diabético) con su validación interna.
3. Investigador principal a cargo	Dra. Anel Gómez García
4. Recolección / obtención de los datos	Se requirió la extracción de la información contenida en el Sistema de Informática de Medicina Familiar (SIMF)-IMSS del estado de Michoacán, de forma retrospectiva mediante consulta electrónica de todos los pacientes con código de diagnóstico de DM2 según el CIE-10. Se realizó la anonimización de los datos. Una vez anonimizados y etiquetados los datos, se realizaron los algoritmos de extracción de datos clínicos utilizando deep learning y se obtuvo la información de la complicación registrada en el expediente electrónico de acuerdo con el CIE (E11.2: Nefropatía; CIE E11.3: Retinopatía; CIE E11.4: Neuropatía; CIE E11.5: Pie Diabético). Además de otras variables de importancia como la edad del paciente al diagnóstico, sexo, IMC registrado durante 3 años y si tuvo diagnóstico de hipertensión arterial sistémica (HAS)
5. Periodo de recolección / obtención de los datos	Para la construcción del modelo se extrajo la información del periodo del 2010 al 2013.
6. Variables incluidas	520
7. Estrategia de aseguramiento para la protección de datos sensibles /	Anonimización a través de deep learning y Named Entity Recognition (NER) para la eliminación de elementos de información identificable que permite cumplir con las regulaciones en materia de protección de datos personales.

personales	
8. Fecha de última actualización	1 de septiembre del 2021
9. Mencionar el tipo de controles para la validación / verificación de la captura / registro de los datos.	Los datos faltantes se imputaron para todos los diagnósticos, tanto en los conjuntos de datos de entrenamiento como de prueba. Si un diagnóstico apareció temprano en el historial médico, se propagaría para todas las visitas posteriores, incluso si ya no se menciona explícitamente en la lista de diagnósticos. Las variables de diagnóstico se registraron a nivel de capítulo de la CIE-10 (p. ej., neoplasias para todos los códigos C00 a D49). Los capítulos del sistema circulatorio y endocrino se registran en el tercer nivel de la jerarquía ICD-10 (por ejemplo, E10 diabetes mellitus tipo 1). Finalmente, los grupos de diabetes se registraron como variables en el cuarto nivel de la jerarquía ICD-10 (por ejemplo, E11.2, E11.3, E11.4, E11.5). Las medidas antropométricas y los valores de laboratorio se imputaron utilizando el valor medio entre los 10 vecinos más cercanos, controlando por edad, sexo, diabetes e hipertensión. Todos los demás valores faltantes (por ejemplo, en las prescripciones de medicamentos) se consideraron como faltantes al azar (MAR) y, por lo tanto, se asignó un valor de 0, mientras que un valor de 1 si se prescribió.
10. Otras plataformas donde se encuentren disponibles estos recursos	No aplica
11. Otras fuentes de Financiamiento	No aplica
12. Seguimiento de la Cohorte en estudio	No aplica
13. Publicaciones	<p>1.- Cleto Álvarez, César Arcos, Karina Figueroa, Marisol Flores, Luis García, Anel Gómez, Arturo López, Jesús Mercado, Mario Rodríguez y Joaquin Tripp. Estudio longitudinal para el desarrollo de modelos predictivos de complicaciones crónicas de la diabetes mellitus tipo 2. <i>Komputer Sapiens</i> 2022; Vol.III:10</p> <p>2.- Barnard Amozurrutia, A., Bernal Astorga, Y., Cuéllar Hidalgo, R., Escoto Velázquez, C.A. y García-Velázquez, L.M. (2022). "Inteligencia artificial en los archivos. Consideraciones de diseño e implementación". <i>Tábula</i>, n. 25, pp. 41-59 Recibido: 7-9-2022. Aceptado: 26-11-2022. En prensa.</p>

3.- Artículo de divulgación "Cuidado por donde pisas" autores: Luis José Flores Alvarez, Cleto Alvarez Aguilar. Fue enviado a Saber Más, Revista de Divulgación de la UMSNH. Enviado el 10 de Noviembre 2022. En revisión por parte de la revista.

4.-Artículo de divulgación "Perdida de la vista por retinopatía diabética, una visión de la enfermedad" autores: María Cristina Maldonado Pichardo, Anel Gómez García. Fue enviado a Saber Más, Revista de Divulgación de la UMSNH. Enviado el 12 de Enero 2023. En revisión por parte de la revista.

REQUISITOS MÍNIMOS PARA LA ENTREGA DE RECURSOS DIGITALES DE INFORMACIÓN DE LOS PROYECTOS CONACYT

Tipo de recursos de información: bases de datos, imágenes, videos, información geoespacial, etc., contemplando los principios FAIR y procurando tenerlos en los siguientes formatos:

- **Contenedores:** TAR, GZIP, ZIP
- **Bases de datos:** XML, CSV, JSON
- **Geoespacial:** SHP, DBF, GeoTIFF, NetCDF
- **Vídeo:** MPEG, AVI, MXF, MKV
- **Sonido:** WAVE, AIFF, MP3, MXF FLAC
- **Estadísticas:** DTA, POR, SAS, SAV
- **Imágenes:** TIFF, JPEG 2000, PDF, DNG, GIF, BMP, SVG
- **Datos tabulares:** CSV, TXT
- **Texto:** XML, PDF / A, HTML, JSON, TXT, RTF
- **Archivo web:** WARC

FICHA TÉCNICA

1. **Objetivo general del proyecto.** Describir brevemente el objetivo del proyecto que derivó en la generación/uso de estas bases de datos.
2. **Propósito del recurso.** Mencionar los usos que el recurso ha tenido al momento y los usos que se pretende hacer eventualmente (análisis, generación de recomendaciones en salud, usos institucionales, etc.).
3. **Investigador principal a cargo.**
4. **Recolección/obtención de los datos.** Descripción detallada de los métodos y técnicas para la recolección/obtención de los datos.
5. **Periodo de recolección/obtención de datos.** Mencionar el inicio y fin de la recolección/obtención. En caso de ser una colección continua mencionar cuándo se planea cerrarla.
6. **Variables incluidas.** Mencionar el # de variables total colectadas.
7. **Estrategia de aseguramiento para la proyección de datos sensibles/personales.**
8. **Fecha última de actualización.** Mencionar la última fecha de captura/registro de los datos que se envían.
9. Mencionar el tipo de **controles para la validación y verificación** de la captura/registro de los datos.
10. Otras **plataformas** donde se encuentren disponibles estos recursos de información. Incluir URL.
11. Otras fuentes de Financiamiento.
12. Seguimiento de la Cohorte en estudio.
13. Publicaciones.