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

Tipo de recursos de información: bases de imágenes, software:

 • **Contenedores:** TAR, GZIP, ZIP

# FICHA TÉCNICA

1. **Objetivo general del proyecto**. Diseñar y crear una plataforma para coordinar esfuerzos de intercambio seguro de datos de salud y la creación de servicios seguros de analítica de datos sin realizar cambios mayores en los sistemas de expediente Clínico Electrónico (SECE) existentes permitiendo su adecuación al cumplimiento de normas oficiales y considerando los estándares que rigen su implementación. La plataforma se basará en servicios configurables para la gestión, aseguramiento, control de acceso, intercambio y preservación de grandes volúmenes de datos en salud (Big Data) enfocados en avanzar hacia la creación de un SECE único nacional y coadyuvar en la edificación de un repositorio nacional de ejecución segura de servicios de análisis de datos médicos.
2. **Propósito del recurso**. Diseñar e implementar un sistema de diagnóstico asistido por computadora para la detección, caracterización y clasificación de cáncer primario de huesos largos usando imágenes de tomografía computada para su integración a un repositorio nacional de servicios de análisis de datos de salud. Análisis, usos institucionales.
3. **Investigador principal a cargo.** Miguel Contreras Murillo, José Luis González Compeán. Marco Antonio Núñez Gaona.
4. **Recolección/obtención de los datos.**

Creación del repositorio de imágenes médicas implementando una aplicación de consulta y recuperación en lenguaje estructurado de imágenes de tomografía con filtrado de información (fecha de estudio, modalidad y descripción de estudio). Extracción de la información contenida en los metadatos y anonimización, excluyendo la información sensible de cada imagen del banco de datos, eliminando los metadatos asociados a cada imagen en formato DICOM referente a pacientes, médicos e institución entre otros. Generación de los planos anatómicos sagital y coronal a partir del plano axial de cada serie de imágenes. Se convierten las imágenes del formato DICOM a PNG. Etiquetado de imágenes manual; para el entrenamiento del modelo se requiere información previamente analizada por el médico especialista donde cada región de interés es definida por medio de un etiquetado manual, esto es enmarcar cara región de interés y definir el tipo de objeto del que se trata. Posteriormente, es necesario entrenar el modelo de detección de imágenes; durante esta etapa, las redes neuronales convoluciones aprenden a identificar diferentes patrones presentes en las zonas de interés en cada una de las imágenes etiquetadas dadas; se encuentran las zonas con mayor presencia de estas características por medio de diferentes técnicas (dependiendo del diseño de la red que se utilice) y se obtiene un modelo entrenado capaz de realizar nuevas detecciones.

1. **Periodo de recolección/obtención de datos**.

La fuente de información está formada por imágenes de tomografía computada de huesos largos almacenados en el PACS-INR desde el año 2014 hasta el año 2020 y que en la descripción de estudio contenga las etiquetas de Fémur.

1. **Variables incluidas**. Se incluyen contendores con el software que permite el entrenamiento de nuevos modelos, y contenedores con modelos entrenados para la detección automática de objetos con características tumorales en pulmones y huesos largo (fémur). Los resultados obtenidos de cada detección son: Nombre de la imagen correspondiente, *Coordenadas* y *clase* de cada región de interés en formato XML-PascalVoc (xmin, ymin, xmax, ymax del BoundingBox), y en formato XML-CVAT (x,y de cada punto del polígono).
2. **Estrategia de aseguramiento para la proyección de datos sensibles/personales**.

Se elimina la información contenida en los metadatos de la imagen DICOM, como datos sensibles del paciente, médicos e institución (anonimización de datos) antes de la fase del etiquetado. Y las imágenes y etiquetas resultantes se proporcionan en formato PNG que sólo contiene la imagen sin metadatos, y xml que sólo contienen las coordenadas, tipo de cada tumor, y orden en el que se deben visualizar.

1. **Fecha última de actualización**. 2023
2. Mencionar el tipo de **controles para la validación y** **verificación** de la captura/registro de los datos. *No aplica.*
3. Otras **plataformas** donde se encuentren disponibles estos recursos de información. Incluir URL.

[https://1drv.ms/u/s!AtMgnjYpElvzg5diSBI09xv-Kyh91A?e=KyZIwv](https://1drv.ms/u/s%21AtMgnjYpElvzg5diSBI09xv-Kyh91A?e=KyZIwv)

1. Otras fuentes de Financiamiento.

*No aplica.*

1. Seguimiento de la Cohorte en estudio.

*No aplica.*

1. Publicaciones.

Miguel Contreras-Murillo, José Luis González- Compeán, “BigData en Salud: Tecnologías Emergentes y Aplicaciones", Montiel & Soriano Editores, S.A. de C.V. ISBN: 978-607-8857-25-8. Pag. 122-138.

1. Otros. Cualquier información extra que sea pertinente precisar o agregar.

Este software está relacionado con la Base de imágenes de cáncer de hueso largo, la cual ha sido agregada como uno de los productos digitales a la CRIP.