

**Propuesta de Proyecto de Investigación  
Maestría en Ciencias y Tecnologías de la Información**

**21 de Noviembre de 2018**

**1. Nombre del proyecto:**

Algoritmo inteligente para la gestión de documentos médicos

**2. Responsable(s):**

Dr. Ricardo Marcelín Jiménez (calu@xanum.uam.mx)  
Departamento de Ingeniería Eléctrica, UAM-Iztapalapa.

Dr. Pedro Lara Velázquez (plara@xanum.uam.mx)  
Departamento de Ingeniería Eléctrica, UAM-Iztapalapa.

**3. Área(s) de conocimiento relacionada(s) con el proyecto**

Big Data / Optimización e Inteligencia Artificial

**4. Descripción del proyecto**

Las instituciones de salud producen un número masivo de imágenes digitales como resultado de los estudios clínicos que practican. Sólo por mencionar un ejemplo, un hospital puede alcanzar miles de procedimientos radiológicos en un año. Lo anterior implica la necesidad de preservar decenas de terabytes de estudios sobre períodos de tiempo que pueden extenderse más allá de un lustro e, incluso, alcanzar algunas décadas. Los archivos digitales correspondientes pueden pesar entre 3.5 y 1500 MB, y su conservación requiere cumplir con exigentes normas relacionadas con su integridad, disponibilidad y confidencialidad. Aun con la aparición de nuevas tecnologías, la imagenología médica seguirá siendo una herramienta clave para el soporte de los servicios de salud.

Un PACS (Picture Archiving and Communications System) es un sistema integrado para la administración de imágenes médicas digitales. A grandes rasgos, se compone de los subsistemas para adquisición de imágenes, para su almacenamiento y para despliegue, todos conectados por una red de datos de alto desempeño. Por otra parte, la cantidad de información que debe preservarse obliga a pensar el diseño de la infraestructura de almacenamiento y dividir estas capacidades en 2 grandes categorías: primario y secundario. El almacenamiento primario es el lugar donde los estudios se colocan para la consulta del médico especialista, se requiere de dispositivos de bajas latencias de acceso. Por su parte, el almacenamiento secundario es el soporte de la preservación de largo plazo, se requiere en este caso de un conjunto de dispositivos capaz de acomodar un volumen masivo de estudios. Este diseño permite acomodar en el primario sólo aquellos estudios que son requeridos en el día a día del hospital, mientras que el almacenamiento secundario contiene la totalidad de los estudios.

Por otra parte, este diseño tiene algunas limitaciones, una de ellas es que si un archivo no se encuentra en el primario y se requiere en el instante, puede tomar algunos segundos su recuperación desde el secundario, lo que implica obtener información de una base de datos compactada. Esta operación es susceptible a fallos que pueden llegar a limitar la disponibilidad de los estudios y depende fuertemente de la calidad de la red de conexión. Nos gustaría adelantarnos a las solicitudes de los médicos, y poner en el almacenamiento primario aquellos archivos que, estimamos, requerirán en sus próximas consultas. Existen varias razones para automatizar este procedimiento, incluyendo aspectos como la confidencialidad o el número de los archivos que deben recuperarse.

Un sistema clasificador es la abstracción informática de la habilidad humana de reconocer clases o categorías de varios objetos, ya sean caras, voces, letras, seres vivos, etc. Para la tarea que implica el

reconocimiento de patrones, se busca clasificar ciertos objetos en una de las k categorías preestablecidas. Resultando la clase a la que pertenece el objeto o en algunos casos no poder determinar la clase a la que se llega.

El objetivo de este trabajo es generar un sistema que, mediante reconocimiento de patrones y/o análisis de series de tiempo, pueda predecir con buena calidad, los archivos que deben recuperarse y emplazarse en el almacenamiento primario de un PACS, con una anticipación de 24 horas.

## 5. Objetivos

Objetivo general

- Generar un modelo de almacenamiento primario para bases de datos médicas que considere periodicidad y frecuencia de uso.

Objetivos particulares

- Obtener la base de datos de un hospital y acoplarla a las necesidades requeridas. Dicha base de datos es posible obtenerla en la actualidad pero por confidencialidad no se puede decir a que hospital pertenece.
- Familiarización del estudiante en algoritmos clasificadores, series de tiempo y manejo de bases de datos.
- Generar un modelo de almacenamiento primario con base en un algoritmo clasificador.
- Probar el modelo en una instancia particular de hospitales.
- Reportar los resultados de la investigación en la idónea comunicación de resultados.

## 6. Metodología

- Estudio y análisis de la base de datos, sistemas clasificadores y otras herramientas que se requieran.
- Proponer e implementar un modelos clasificador que gestione la base de datos.
- Escribir la idónea comunicacion de resultados (ICR) a lo largo de todo el proyecto.
- Entregar de la version final de la idonea comunicación de resultados.
- Presentar y aprobar el examen de grado.

## 7. Calendarización de actividades

**Trimestre 1:** Estudio y análisis de la base de datos, sistemas clasificadores y otras herramientas que se requieran. Proponer un modelo clasificador que gestione la base de datos

**Trimestre 2:** Implementar un modelos clasificador que gestione la base de datos. Escritura de la idónea comunicación de resultados.

**Trimestre 3:** Mejoras al modelo propuesto. Entrega de la version final de la idónea comunicación de resultados.

**Trimestre 4:** Revisión de los sinodales de la idónea comunicación de resultados. Presentación del examen de grado.

## 8. Infraestructura necesaria y disponible

Una computadora con sistema operativo Linux o Windows con Python.

## **9. Lugar de realización**

Cubículo de estudiantes o Laboratorio T326B

## **10. Entregables**

Idónea comunicación de resultados en tiempo y forma.

## **11. Referencias básicas**

Piankyh, O: Digital Imaging and Communications in Medicine (DICOM) Cap 11. DICOM Media and Security, Springer 2nd Edition, 2012.

Huang, HK. PACS and Imaging Informatics. Basic Principles and Applications; New Jersey: Wiley Blackwell 2nd Edition, 2010.

Gutiérrez-Martínez, J., Núñez-Gaona, M. A., Aguirre-Meneses, H., & Delgado-Esquerro, R. E. (2012). Software and hardware architecture for a high-availability PACS. *Journal of Digital Imaging*, 25(4), 471–479. <https://doi.org/10.1007/s10278-012-9494-2>

Núñez-Gaona M.A., Marcelín-Jiménez R., González-Compeán J.L., Aguirre-Meneses H., & Gutiérrez-Martínez J., A Dependable Massive Storage System for Medical Imaging, *Journal of Digital Imaging*, May 2018 (JCR impact factor 1.407) /doi.org/10.1007/s10278-018-0091-x ISSN: 0897-1889 (Print) 1618-727X (Online)

**12. En caso de que ya fue revisada y aprobada en convocatorias anteriores favor de indicar los cambios correspondientes e indicar que será sometida para la Generación 2018.**