

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

15 / nov / 2022

**1. Nombre del proyecto**

Análisis de imágenes de cáncer cérvico uterino para apoyo en detección de enfermedad por medio de visión evolutiva

**2. Responsables**

Dr. Juan Villegas Cortez  
UAM-Azcapotzalco, oficina: HP-06  
eMail: [juanvc@azc.uam.mx](mailto:juanvc@azc.uam.mx)

Dra. Graciela Román Alonso  
UAM-Iztapalapa, oficina: T-215  
eMail: [gra@xanum.uam.mx](mailto:gra@xanum.uam.mx)

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

- Reconocimiento de patrones: teoría de clasificadores, conformación y estudio de patrones, discriminación de rasgos.
- Técnicas básicas de los algoritmos evolutivos y específicamente de la teoría de la programación genética.
- Redes neuronales artificiales.
- Cómputo paralelo.

**4. Descripción del proyecto**

Resumen

Desarrollar una metodología para el estudio y caracterización de las imágenes de detección de cáncer cérvico uterino, provenientes de las pruebas de papanicolau, a partir del análisis de textura y el desarrollo de rasgos característicos de la propia imagen, por medio de un algoritmo evolutivo, orientado a caracterizar las imágenes que indican la presencia de la enfermedad de las que no son representativas, acorde a imágenes supervisadas por profesionales de la salud. La caracterización obtenida automáticamente sería una propuesta de apoyo alternativo para la detección automática de la enfermedad.

Perfil deseable del alumno:

- Alumno con conocimientos de programación..
- Alumno con interés en tópicos de inteligencia artificial y cómputo paralelo.

Entradas del sistema:

- Patrón en bruto (*raw*) de los datos a analizar.
- Una función de aptitud acorde al patrón.
- Determinación de los parámetros evolutivos: tamaño de la población, número de generaciones máximo, cota máxima/mínima de la función de aptitud, porcentaje de aplicación en operaciones genéticas de cruce y mutación, determinación de elitismo a usar.
- Determinación de un mínimo de procesos evolutivos en paralelo para tener comparaciones de complejidad.

Salidas del sistema:

- Una Memoria Asociativa Evolutiva (MAE) que indica las variables más representativas para realizar la clasificación óptima de los datos estudiados.
- Análisis de complejidad de las MAE obtenidas a partir del número de generaciones necesarias para alcanzar el óptimo.

## 5. Objetivos

- Objetivo general

Implementar la evolución de descriptores de textura de la imagen para el tipo de imágenes digitales obtenidas de la prueba Papanicolau, y poder caracterizar las imágenes a partir de la supervisión de imágenes confirmadas de cáncer y las que no.

- Objetivos específicos
  - ❖ Delimitar el modelo de un algoritmo evolutivo para el tipo de imágenes por interés, justificando el algoritmo.
  - ❖ Análisis del modelo, diseño e implementación paralela del sistema.
  - ❖ Realizar pruebas sobre al menos una base de datos pública reconocida, para poder discriminar las imágenes positivas de las negativas de la enfermedad, con base a la metodología del proceso evolutivo, proporcionando el patrón característico obtenido.
  - ❖ Validar la robustez del modelo con la técnica evolutiva propuesta, con otros resultados reportados en el estado del arte.

## 6. Metodología

Para la realización del presente proyecto la metodología a seguir es la siguiente:

- Obtención y análisis de las imágenes de estudio.
- Análisis del desempeño de la metodología CBIR probada por la literatura para caracterizar las imágenes sobre dos clases: positivas y negativas.
- Estudio y análisis de los algoritmos básicos del reconocimiento de patrones: teoría de clasificadores, conformación y estudio de patrones, discriminación de rasgos.
- Estudio y análisis de las técnicas básicas de los algoritmos evolutivos y específicamente de la teoría de la programación genética.

- Diseño de una arquitectura paralela de una red neuronal artificial de tipo profunda, para ser ejecutada en varios procesadores para la implementación rápida de la metodología.
- Análisis de los patrones acorde a los rasgos de los datos a estudiar.
- Pruebas de la metodología con diversos patrones acorde a precisión versus recuperación para los datos, las bases de datos, reportados en la literatura.
- Generación de resultados.
- Publicación de los resultados y escritura de la tesis

## 7. Calendarización de actividades

Trimestre	Actividad
23-I	Estudio del Estado del Arte, la problemática y la metodología de investigación. Diseño del sistema paralelo a desarrollar. Estos avances serán presentados por escrito y en el seminario organizado para tal fin.
23-P	Implementación y evaluación del desempeño de los algoritmos propuestos. Los avances del proyecto serán presentados por escrito, en formato de artículo. Estos resultados serán presentados en el seminario organizado para tal fin.
23-O	Escritura y entrega de una primera versión de la comunicación idónea de resultados. Los resultados serán presentados en el seminario organizado para tal fin.
24-I	Entrega de la versión final de la idónea comunicación de resultados, revisión por parte del jurado y defensa oral del trabajo de investigación ante un jurado especializado.

## 8. Infraestructura necesaria y disponible

Para llevar a cabo esta propuesta se hará uso de los siguientes recursos como entorno de desarrollo.

### Hardware

- Computadora tipo WorkStation serie Dell Precision 2, 4, 8 y 12 núcleos, con 4, 8 y 16 GB de RAM, con procesador Intel XEON 64 bit.
- Los recursos del Laboratorio de Sistemas Paralelos y Distribuidos T-169 UAM Iztapalapa

### Software

- Sistema operativo Linux / 64 bits.
- Matlab con toolboxes: Digital Signal Analysis, Statistical Analysis, GPLab, y GPTips V2.

- **Compilador ANSI C/C++ , MPI, OpenMP y/o Cuda**

## **Datos**

- Bases de datos de cáncer del tipo público.

## **9. Lugar de realización**

El proyecto se podrá realizar en la modalidad mixta (presencial y/o a distancia), en instalaciones de UAM-Izt, UAM-Azc y con sesiones remotas via Zoom.

## **10. Entregables**

Considerar los distintos entregables que están establecidos para las UEA de Proyecto de Investigación. Por ejemplo: Memoria in extenso en congreso nacional e ICR.

Sistema paralelo para el estudio y caracterización de las imágenes de detección de cáncer cérvico uterino, provenientes de las pruebas de papanicolau.

Documento de la Idónea Comunicación de Resultados.

## **11. Referencias bibliográficas básicas**

- J. Villegas-Cortez, J. Sossa, C. Aviles-Cruz, and G. Olague, “Evolutionary associative memories through genetic programming,” *Revista mexicana de física*, vol. 57, pp. 110 – 116, 04 2011.
- J. Villegas-Cortez, G. Olague, H. Sossa, and C. Avilés, *Evolutionary Associative Memories through Genetic Programming*, pp. 171–188. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.
- F. Chávez, F. Fernández, D. Lanza, C. Benavides, J. Villegas, L. Trujillo, G. Olague, and G. Román, “Deploying massive runs of evolutionary algorithms with ecj and hadoop: Reducing interest points required for face recognition,” *The International Journal of High Performance Computing Applications*, p. 1094342016678302, 2017/01/18 2016.
- Erick Esteven Montelongo González. «Modelos de aprendizaje automático para el apoyo en la clasificación de tipos de cáncer a partir de datos estructurados y no estructurados de expedientes clínicos». Maestría en Ciencias en Computación. División de Ciencias Básicas e Ingeniería. Universidad Autónoma Metropolitana – Unidad Azcapotzalco. 14 de julio de 2020.
- Erick E. Montelongo-González, Alejandro Reyes-Ortíz y Alejandro González-Beltrán Beatriz A. «Machine Learning Models for Cancer Type Classification with

Unstructured Data». En: *Computación y Sistemas (CyS) 24.2* (2020). ISSN impreso 1405-5546, ISSN electrónico 2007-9737, págs. 403-411.

- A. E. W. Johnson *et al.*, “MIMIC-III, a freely accessible critical care database,” *Scientific Data*, vol. 3, no. 1, p. 160035, May 2016, doi: [10.1038/sdata.2016.35](https://doi.org/10.1038/sdata.2016.35). ● [1]
- T. Xu *et al.*, “A New Image Data Set and Benchmark for Cervical Dysplasia Classification Evaluation,” in *Machine Learning in Medical Imaging*, Cham, 2015, pp. 26–35.
- Hu L, Schiffman M, et al. An observational study of deep learning and automated evaluation of cervical images for cancer screening. *Journal of the National Cancer Institute*. Jan. 10, 2019. DOI: 10.1093/jnci/djy225.
- Karla Mandujano López, Juan Villegas Cortez, Graciela Román Alonso, Arturo Zúñiga López, Edgar Rangel López, César Benavides Alvarez. Identificación y localización de células gliales en imágenes de tejido cerebral. *Research in Computing Science (Latindex, DBPL)* . 2021. Pp. 297-309. Vol. 149, No. 8. ISSN 18704069. ● [1]
- E. Hussain, L. B. Mahanta, H. Borah, and C. R. Das, “Liquid based-cytology Pap smear dataset for automated multi-class diagnosis of pre-cancerous and cervical cancer lesions,” *Data in Brief*, vol. 30, p. 105589, Jun. 2020, doi: [10.1016/j.dib.2020.105589](https://doi.org/10.1016/j.dib.2020.105589).
- William W, Ware A, Basaza-Ejiri AH, Obungoloch J. A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images. *Comput Methods Programs Biomed*. 2018;164:15-22. doi:10.1016/j.cmpb.2018.05.034 ● [1]
- M. E. Plissiti, P. Dimitrakopoulos, G. Sfikas, C. Nikou, O. Krikoni, and A. Charchanti, “Sipakmed: A New Dataset for Feature and Image Based Classification of Normal and Pathological Cervical Cells in Pap Smear Images,” in *2018 25th IEEE International Conference on Image Processing (ICIP)*, Oct. 2018, pp. 3144–3148, doi: [10.1109/ICIP.2018.8451588](https://doi.org/10.1109/ICIP.2018.8451588).