



LICENCIATURA EN CIENCIAS DE DATOS
Propuesta de tema de tesis

Desarrollo de una base de datos anotada de imágenes citológicas de Papanicolaou

Este plan se enmarca en un proyecto más general para automatizar el diagnóstico de cáncer de cuello uterino. Para entrenar los modelos de AI, es esencial desarrollar una base de datos de imágenes anotadas que permita identificar las células o regiones patológicas. Esta tarea es realizada por personal especializado, por lo que este plan contempla una interacción con personal médico colaborador en este proyecto. En este plan se propone trabajar en la confección de una base de datos anotada que contemple el acuerdo inter-anotador como parte fundamental de la metadata. En este proyecto proponemos las siguientes tareas para mejorar y escalar el proceso actual: 1. Poner a punto un sistema de organización de las imágenes anotadas con el fin de confeccionar una base de datos de uso público. 2. Levantar la base de datos en un servidor de la FCEN-UBA

Palabras clave: Base de Datos - Anatomía Patológica PAPs - Acuerdo Inter-Anotador - Imágenes Citológicas

Conocimientos deseables

Manejo de Bases de Datos. Desarrollo web

¿Qué podría aprender quien realice esta tesis?

El/La estudiante formará parte de un equipo interdisciplinario que combina expertises de las Cs. de la Computación, Cs. de Datos y Medicina. Trabjará con imágenes de relevancia médica en un problema de aplicación directa a la Salud Pública.

Dirección de la tesis

*Bruno, Luciana
Instituto de Cálculo, FCEN, UBA-CONICET*

Contacto: lucianabrun@gmail.com

Más información en el pdf a continuación.

Título: Desarrollo de una base de datos anotada de imágenes citológicas de Papanicolaou.

Dirección:

[Dra. Luciana Bruno](#). IC-CONICET. Facultad de Cs. Exactas y Naturales. Universidad de Buenos Aires.
[Dra. Viviana Siless](#). y [Emmanuel Iarussi](#). Escuela de Negocios. Universidad Torcuato Di Tella.

Contacto: lucianabrun@gmail.com

Resumen:

Este plan se enmarca en un proyecto más general para automatizar el diagnóstico de cáncer de cuello uterino. El proyecto comienza con la digitalización de las muestras citológicas, y es seguido por el desarrollo de modelos supervisados de IA para asistir en el diagnóstico médico. El proyecto cuenta con financiación de un subsidio otorgado por el [CLIAS](#).

Para entrenar los modelos de AI, es esencial desarrollar una base de datos de imágenes anotadas que permita identificar las células o regiones patológicas. Este es un proceso fundamental que consiste en señalar en la imagen digital las zonas de la imagen que corresponden a las entidades buscadas (ej: patológica de alto, bajo, medio grado. inflamatoria, normal). Esta tarea es realizada por personal especializado, por lo que este plan contempla una interacción con personal médico colaborador en este proyecto. Se cuenta con un software libre Label Studio, que permite al personal técnico detallar en las imágenes las regiones normales y de lesiones (Fig.1). Un aspecto importante para la construcción de este tipo de bases de datos es la fiabilidad del etiquetado humano. Es por esto que es sumamente valioso obtener anotaciones de una misma célula de varios expertos, para así poder también medir: 1- La confianza en el etiquetado, y 2- el error humano vs. el error de IA.

En este plan se propone trabajar en la confección de una base de datos anotada que contemple el acuerdo inter-anotador como parte fundamental de la metadata.

En este proyecto proponemos las siguientes tareas para mejorar y escalar el proceso actual:

- 1) Poner a punto un sistema de organización de las imágenes anotadas con el fin de confeccionar una base de datos de uso público.
- 2) Levantar la base de datos en un servidor de la FCEN-UBA

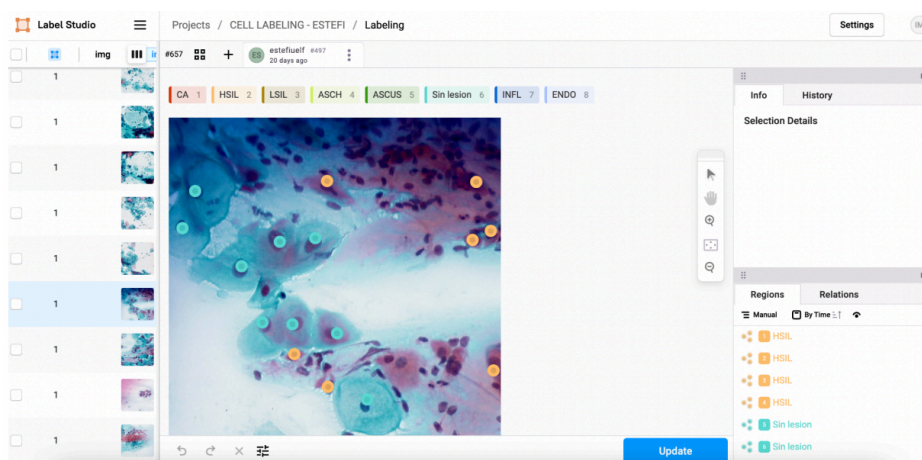


Figura 1. Captura de pantalla del software de anotación donde puede verse una imagen digitalizada de un PAP. Los círculos de colores indican los núcleos de células anotadas. Las células son clasificadas en 8 clases diferentes.