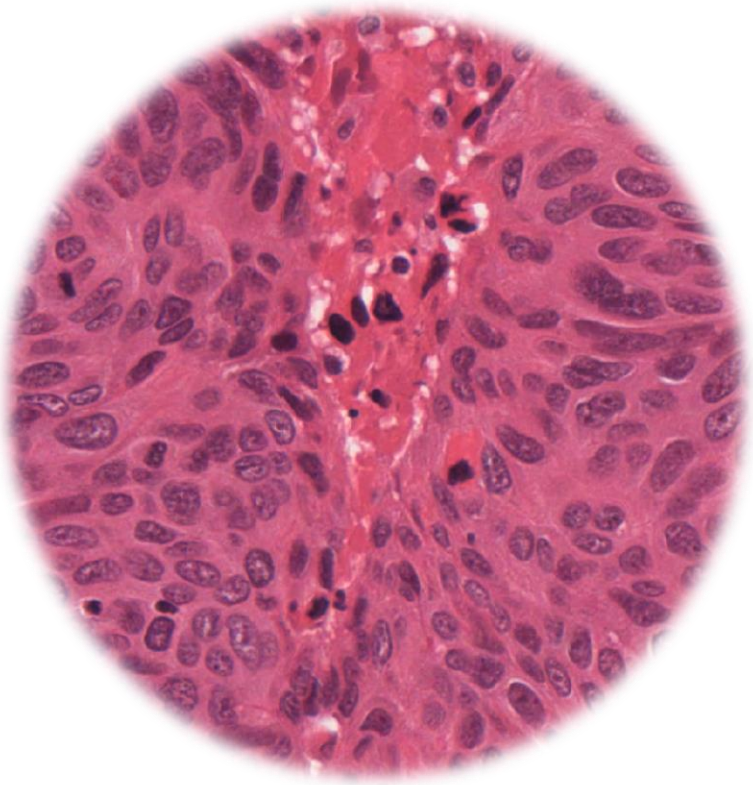
Segmentación y conteo de núcleos celulares en histología patológica

Xabier Jiménez Gómez



Índice

1.Introducción

El método para estudiar la anatomía de un tejido del cual se sospecha una posible patología o enfermedad que se encuentra visible en el sistema celular consiste en hacer cortes histológicos de la zona en cuestión para su estudio .De esta forma aplicando distintos tintes de color y técnicas se resaltan los núcleos o estructuras celulares del tejido afectado consiguiendo de estos un posible análisis biológico.

Para la realización de este proyecto de segmentación y conteo de núcleos celulares en histología de anatomía patológica se han otorgado 30 imágenes en distintas versiones(según el tinte H&E) de la base de datos de MoNuSeg para cumplir una serie de pautas. En los siguientes apartados se explicaran las fases por que paso cada objetivo y se comentarán en varias imágenes.

2.Objetivos

1. Segmentar cada uno de los núcleos celulares en la imagen.

2. Contar el número de núcleos en la imagen.

3. Medir el área en px2 de cada núcleo, y el área media.

4. Proporcionar una evaluación cuantitativa de la segmentación, conteo y medición en 1, 2 y 3.

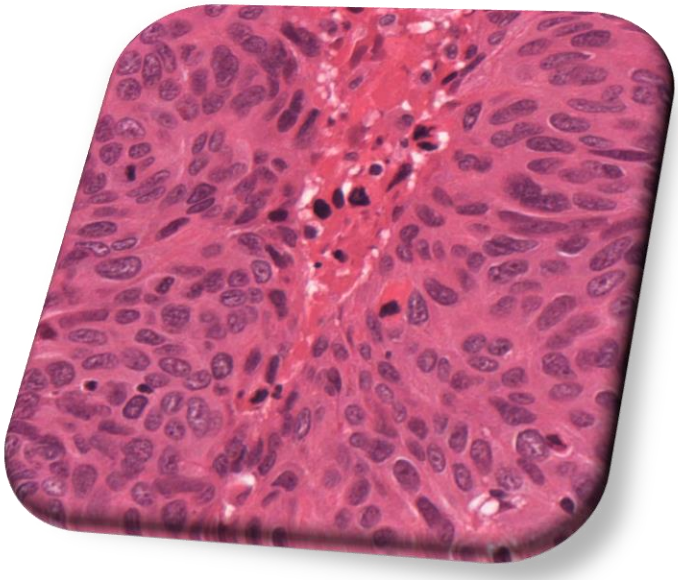
3.Realización

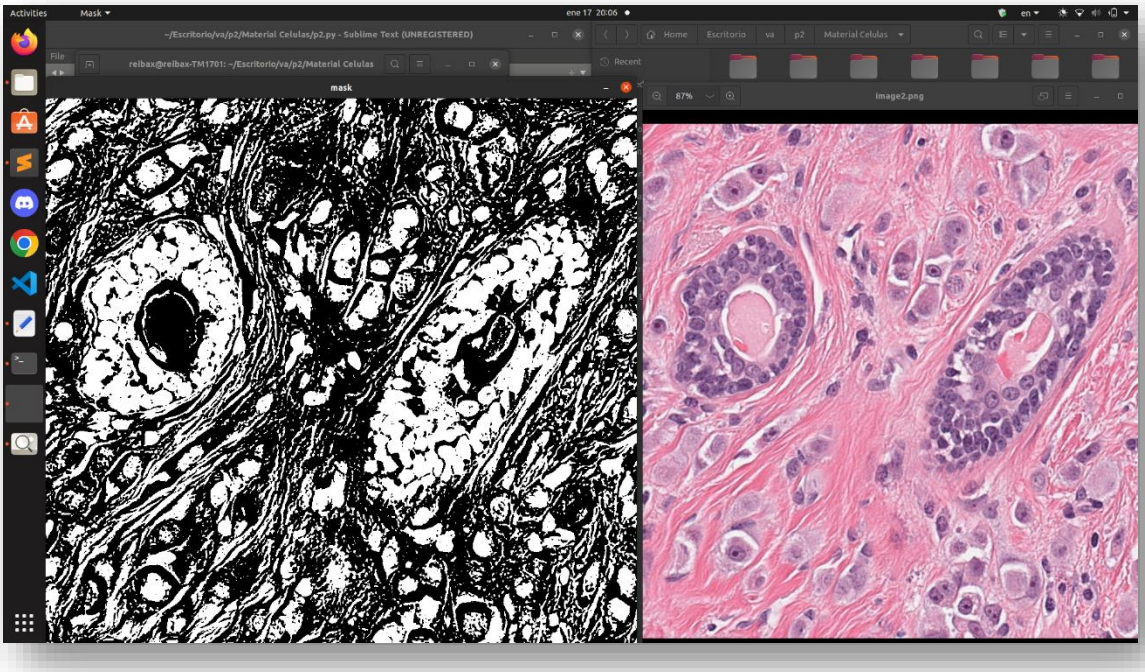
Para el conteo de núcleos celulares se realizaron diversos métodos en las imágenes siendo el que mejor resultados da el siguiente paso a paso:

Se carga la imagen y se convierte a escala de grises primariamente,posteriormente se crea una mascara binaria a la cual se le aplicaron operaciones morfológicas para limpiarla. Concretamente se creó un kernel 3X3 de forma elipse (después de diferentes pruebas y ver ejemplos en la web se optó por esta forma) y se realizó opening sobre la misma.

Después aplicamos la función connectedComponentWithStats a la que dándole un shape conseguimos su conteo. En esta medición no se tuvieron en cuenta los resaltes de contorno y se obtuvieron mejores resultados, sin embargo en el conteo final al resaltarlos salen cantidades mas altas.

Dando un resultado de 556 el que bajo medición visual parece el más cercano a la realidad.

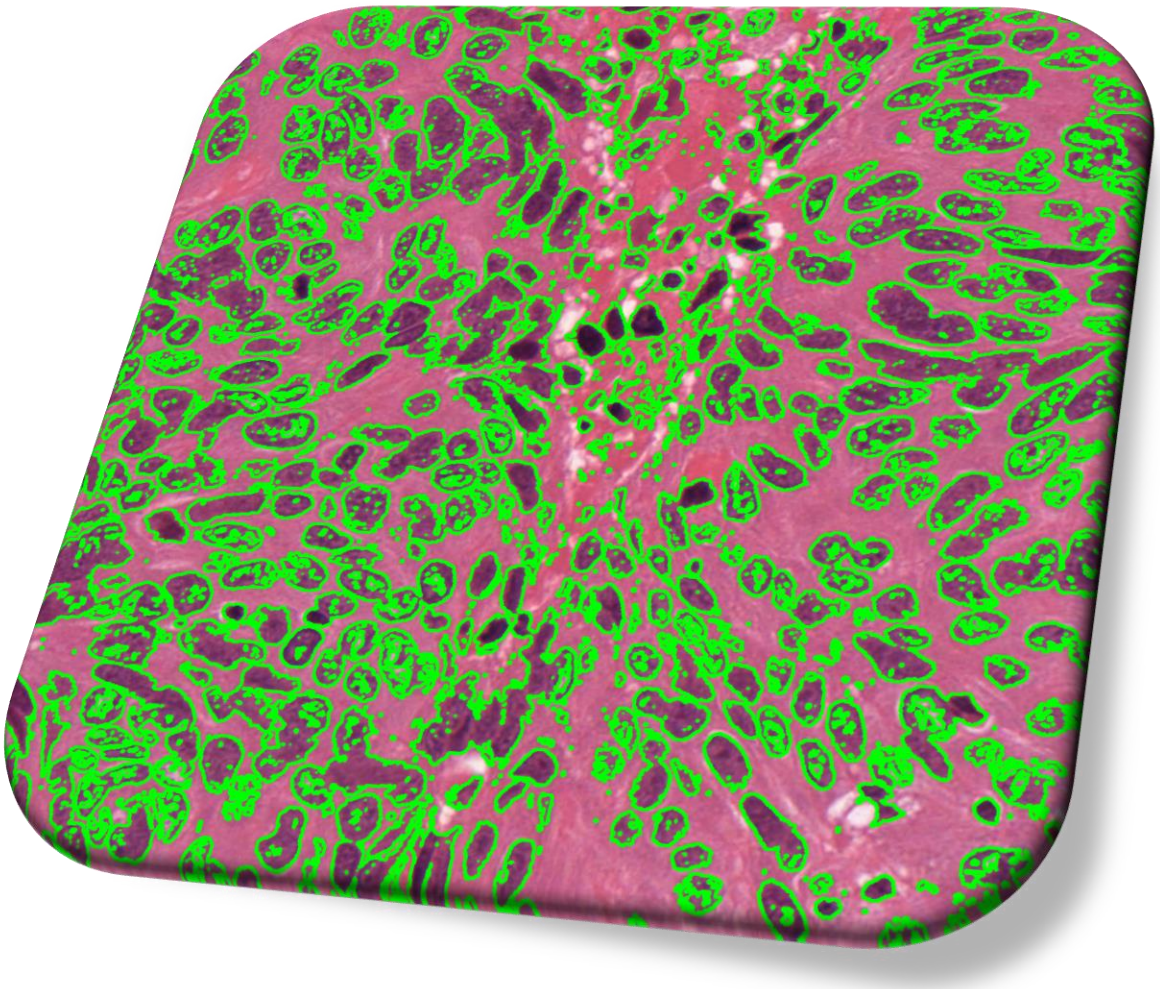




En estas imágenes de arriba se realizó el mismo proceso pero los resultados fueron peores por la claridad de la imagen dando cantidades más altas que antes pese a que esperábamos menores inicialmente.

Primariamente poner en escala de grises la imagen facilitó su uso, posteriormente aplicar Gaussian blur reduce el ruido y la suaviza para obtener mejores resultados, seguidamente Otsu ,se hacen los contornos y se dibujan.

Más en detalle en este código se usa el método de Otsu para la segmentación ,técnica que automáticamente calcula el valor threshold para separar el fondo de los objetos. Se aplicó Gaussian blur para reducir el ruido de la imagen en escala de grises devolviendo de esta el método Otsu dos valores siendo el segundo(según he visto en documentación) el que usaremos. Para marcar los contornos se utilizo una función de OpenCV llamada findContours al resultado de lo anterior. Posteriormente se dibujaron dando como resultado que la función anterior daba soluciones satisfactorias en cuanto a verdaderos positivos(coge todos los núcleos prácticamente) pero suma algunos falsos positivos ,como son el borde de toda la figura y pequeños puntos que podrían ser quitados poniendo más adelante un mínimo de tamaño al área. Finalmente en un bucle se va mostrando el área de cada contorno creado y se muestra una media de los mismos.

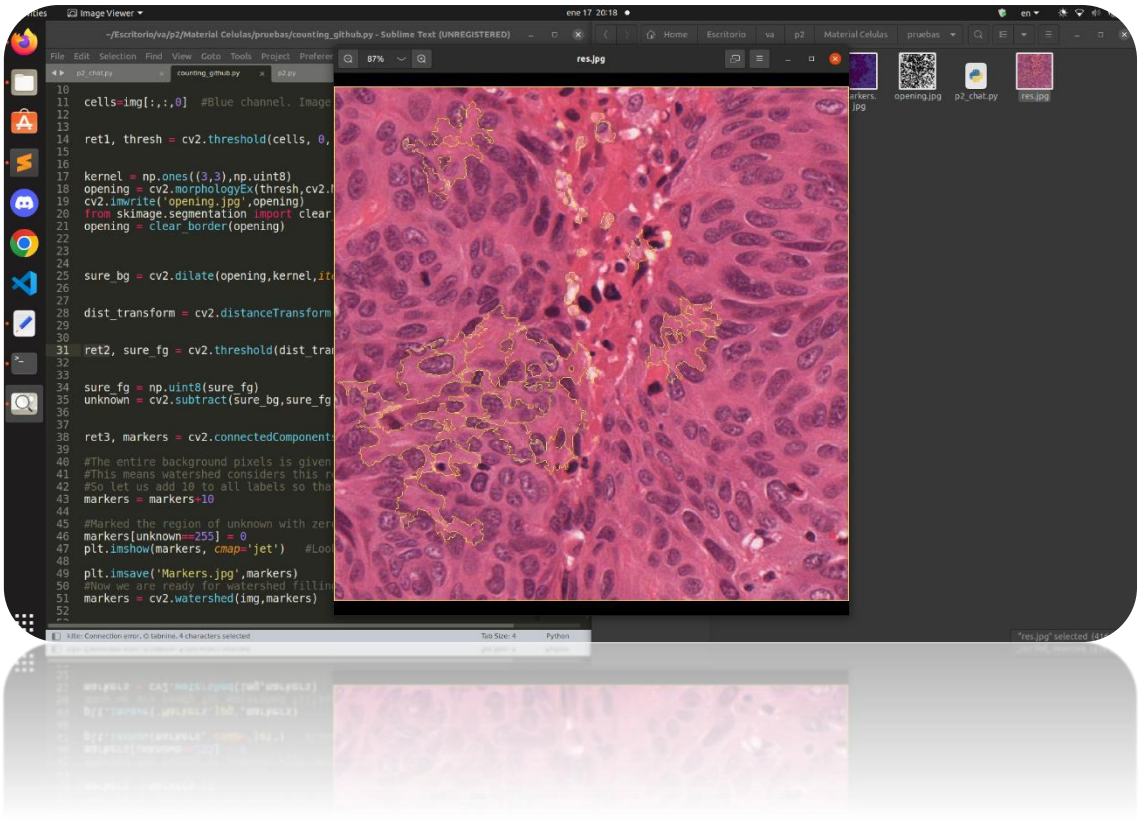


4.Otras pruebas

Se probó a procesar la imagen y oscurecerla con un filtro gaussiano con kernel 5x5 en escala de grises. Posteriormente se uso threasholding para separa el fondo de los objetos a contar.

Para la segmentación utilizamos un método complejo que conseguía segmentar en zonas pero no las deseadas. Mediante el uso de opening y posteriormente dilatación. Volvimos a utilizar la función de threshold pero marcaba cosas que no queríamos

Evidencia:



Marcaba estructuras celulares no deseadas en prácticamente todos los casos.

Para la obtención de tamaños en pixeles de las células se intentó sacar por pantalla los resultados al lado de cada núcleo lo que dió lugar a un solapamiento en la imagen que no dejaba visible nada, en su lugar se han sacado por consola los tamaños separados y la media total. El problema de esto es que no ha sido posible relacionar a cada célula o contorno su tamaño por el momento.

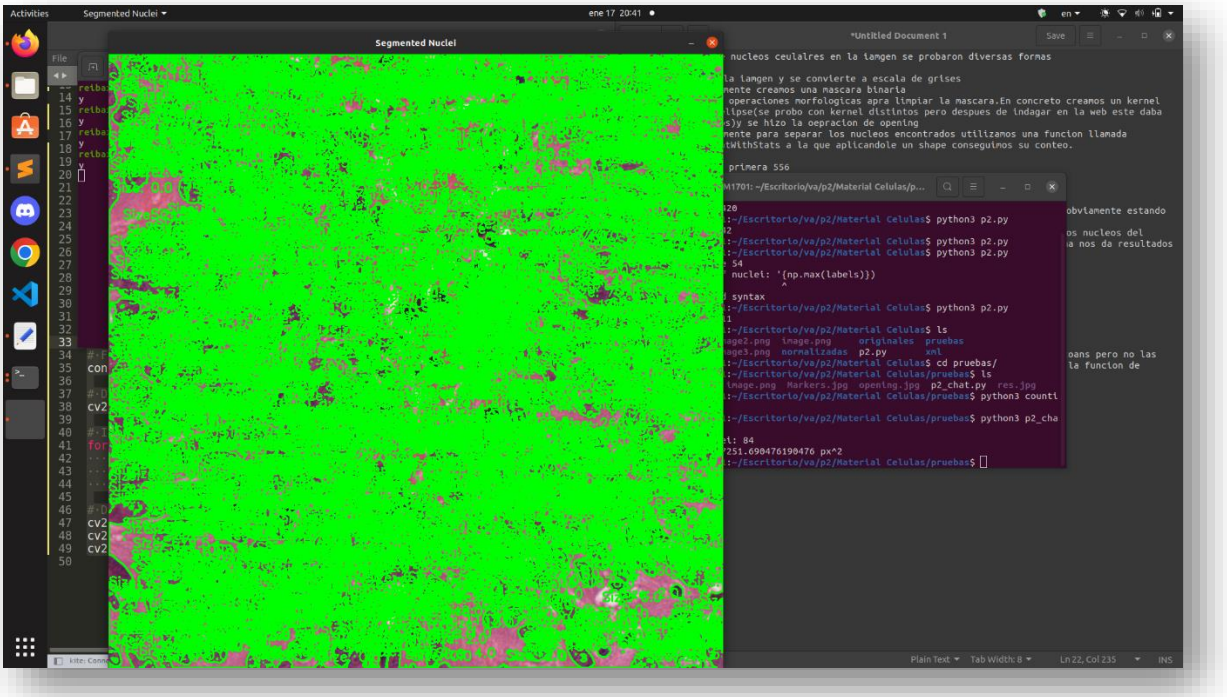
graficamente -> for cnt in contours:

size = cv2.contourArea(cnt)

(x, y, w, h) = cv2.boundingRect(cnt)

cv2.putText(img, f'Size: {size}', (x, y-10),

cv2.FONT_HERSHEY_SIMPLEX, 0.7,(0, 255, 0),2)



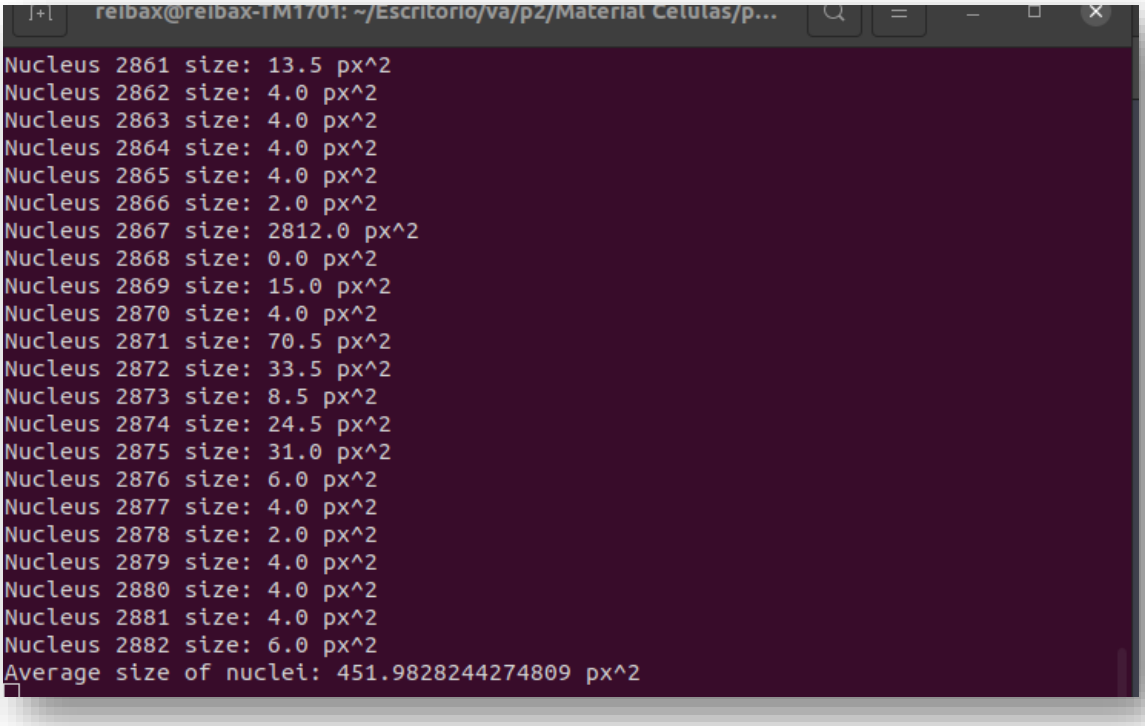
NOTA: ContourArea() calcula el tamaño de cada contorno que corresponde al resultado de las segmentaciones de los núcleos celulares.

cmd -> for idx, cnt in enumerate(contours):

size = cv2.contourArea(cnt)

print(f'Nucleus {idx+1} size: {size}' px'2)

por defecto ya es en pixeles



Para hacer la media simplemente con una línea de código la variable total que dentro del bucle sume cada tamaño serviría para posteriormente dividirla entre el número total de núcleos.

5.Ventajas y desventajas de alternativas

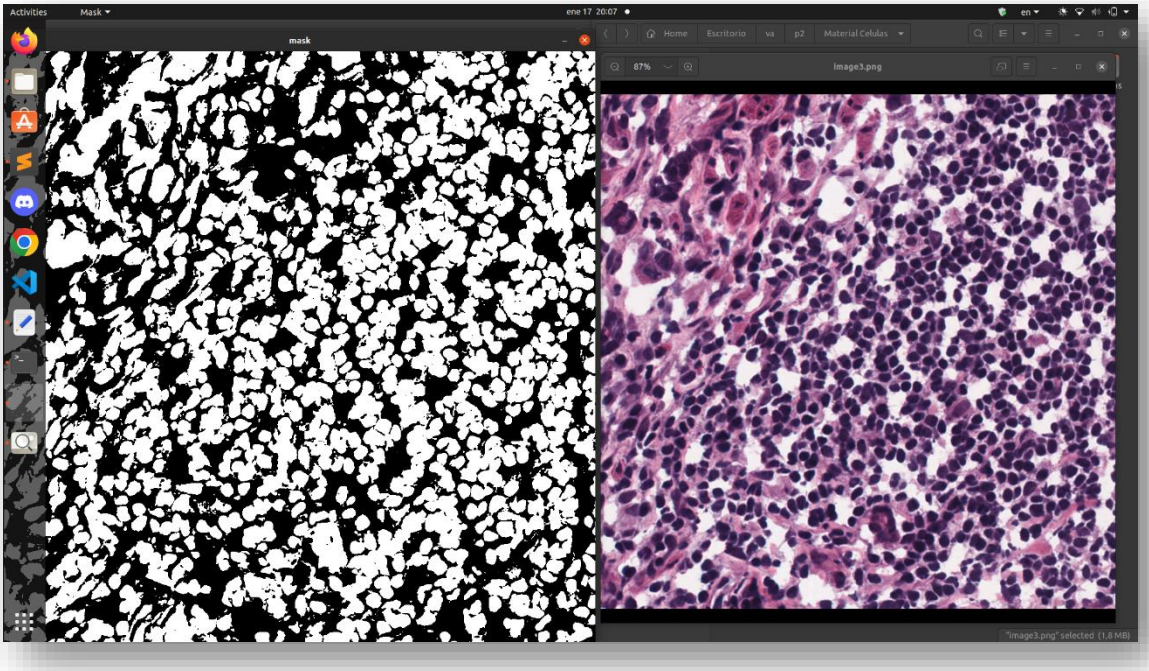
Las ventajas de estas técnicas utilizadas es que permiten hacen la segmentación con precisión y haciendo rápido su estudio y análisis en seguimientos científicos o médicos con resultados obviamente en línea con su precocidad (mejorables).

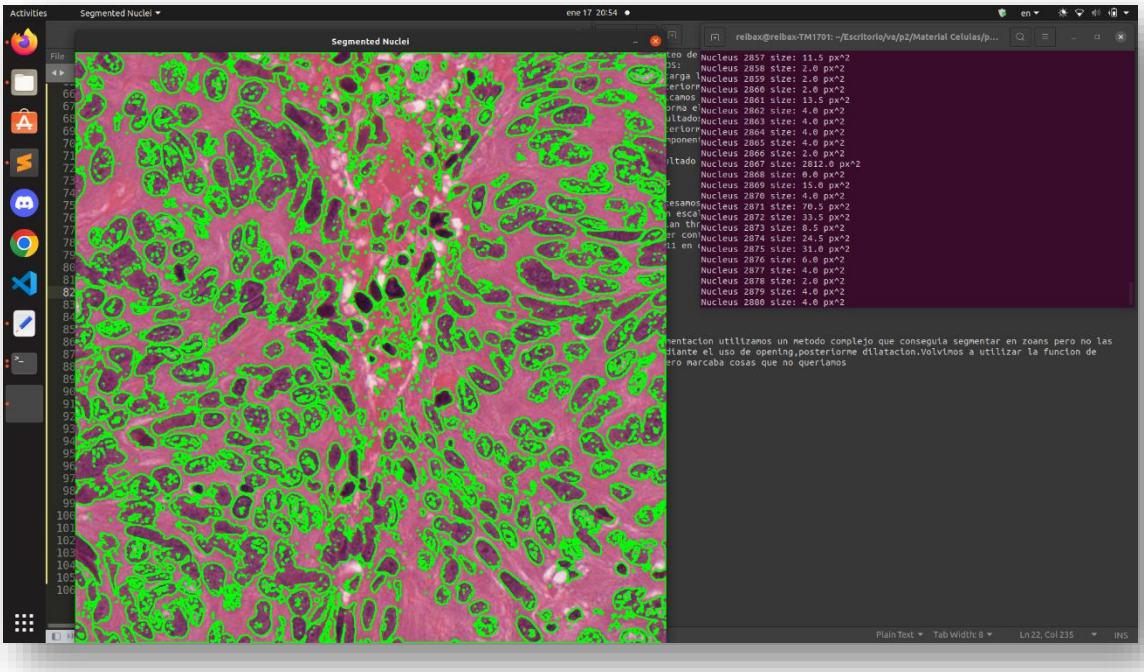
Alternativas serian el uso de algoritmo Watershed, kmeans agrupando pixeles similares (núcleos) según el color o machine learning entrenando redes neuronales lo cual sería lo más óptimo.

Preferiblemente el método usado parece el más simple y común para este tipo de problemas, sin embargo para problemas específicos podría ser más óptimo alguno de los citados.

6.Conclusión y resultados cuantitativos

En cuanto a la medición de resultados es complicado obtener un porcentaje de precisión para el código utilizado ya que no se sabe con precisión el total de núcleos celulares de la imagen. Con el código utilizado se obtiene prácticamente un resultado de 1 en cuanto a verdaderos positivos dando siempre el resultado si existen células patológicas siendo optimo para el campo medico(no pueden escaparse patologías) pero dando resultados de mas en la mayoría de ocasiones. Además se presenta el problema de según la imagen ,si cuenta con un fondo mas claro o cercano al color de las células patológicas el resultado pasa a ser peor ya que la función de detección cuenta con mas problemas. Por otro lado la idea de poner un tamaño mínimo de contornos podría quitarnos muchos falsos positivos pero no se ha llevado a cabo pues no sabría decir cual seria un mínimo pertinente, es decir, no se tiene actualmente certeza de cual sería el tamaño mínimo para tener en cuenta o no un núcleo celular. Se adjuntan imágenes que presentan contornos realizados.





En cuanto a las mediciones parecen optimas tanto por separado como la media global y el conteo es directamente dependiente y equivalente del resultado de segmentación previo.

A grandes rasgos podemos indicar una precisión alta para los positivos , pero la tasa de error es considerable.