

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA INFORMÁTICA

Proyecto de título

**“Detección automática de células mitóticas en imágenes
histológicas usando redes neuronales convolucionales
profundas”**

Nombre: **Felipe Andrés Acosta Castillo.**

Profesor Guía: **Dr. Héctor Allende Cid.**

Carrera: **Ingeniería Civil Informática**

Octubre 2017.

Índice

Resumen	1
1. Introducción.....	2
2. Objetivos	3
2.1. Objetivo General	3
2.2. Objetivos específicos	3
3. Trasfondo biológico.....	4
3.1. Etapas de la mitosis	4
3.1.1. Profase.....	4
3.1.2. Metafase	5
3.1.3. Anafase	6
3.1.4. Telofase.....	6
3.2. El cáncer.....	7
3.3. Diagnóstico del cáncer	7
3.4. Grado de un tumor.....	8
3.5. Sistema de gradación de Nottingham	8
4. Descripción del problema	9
4.1. Obtención de las imágenes histológicas	10
5. Estudio del estado del arte.....	11
5.1. Normalización de las muestras.....	11
5.1.1. Normalización basada en descomposición de valores singulares [11].	12
5.1.2. Normalización luego de separación basada en regularización dispersa [12]	13
5.2. Segmentación de candidatos.....	14
5.2.1. Basado en operaciones morfológicas.	14
5.2.2. Laplaciano del Gaussiano (LoG) [14].....	14
5.3. Clasificador	15
5.3.1. Red Neuronal Convolutiva [15].....	15
5.3.2. Clasificador por características morfológicas y de canales múltiples [14].....	16
6. Modelo propuesto.....	17
6.1. Selector de candidatos.....	18
7. Creación del set de datos	18
7.1. Aumentación de datos.....	19
8. Redes Neuronales Convolucionales	20

8.1.	Capa Convolutacional	20
8.2.	Rectificador lineal	21
8.3.	Max pooling	22
8.4.	Dropout	22
8.5.	Capa completamente conectada.....	23
9.	Arquitecturas e ideas involucradas.....	23
9.1.	Arquitectura VGG-16 Y VGG-19.....	23
9.2.	Arquitectura Inception	23
9.3.	Arquitectura ResNet	24
9.4.	Arquitectura SqueezeNet.....	25
9.5.	Arquitectura Enet	25
10.	Arquitecturas probadas.....	26
10.1.	SimpleNet	26
10.2.	Mini SqueezeNet.....	30
10.3.	ResNet – 11.....	33
11.	Discusión de los resultados	37
12.	Conclusión.....	39
13.	Referencias.....	41

Lista de figuras

Ilustración 1. Célula en profase junto a un dibujo explicativo a mano	5
Ilustración 2. Célula en metafase junto a un dibujo explicativo a mano	5
Ilustración 3. Célula en anafase junto a un dibujo explicativo a mano	6
Ilustración 4. Célula en telofase junto a un dibujo explicativo a mano.....	6
Ilustración 5. Ejemplo de células mitóticas.....	9
Ilustración 6. Ejemplo de mitosis tripolar	10
Ilustración 7. Normalización de la misma muestra tomada con diferentes escáneres.....	12
Ilustración 8. Normalización basada en regularización dispersa.....	13
Ilustración 9. Imagen binaria donde se muestran los candidatos seleccionados	14
Ilustración 10. Paso a paso de la selección de candidatos basada en LoG	15
Ilustración 11. Arquitectura del modelo propuesto	17
Ilustración 12. Muestra de la salida del selector de candidatos basado en el laplaciano del gaussiano	19
Ilustración 13. Visualización de una red neuronal	20
Ilustración 14. Convolución 2d	21
Ilustración 15. Ejemplo de una operación de max pooling con desplazamiento de 2.....	22
Ilustración 16. Efecto de la regularización mediante dropout en las conexiones.....	22
Ilustración 17. Módulo de la arquitectura ResNet.....	24
Ilustración 18. Representación del módulo fuego	25
Ilustración 19. Métricas de entrenamiento de la red SimpleNet sin aumentación de datos	27
Ilustración 20. Fscore de la red SimpleNet sin aumentación de datos	27
Ilustración 21. Métricas de entrenamiento de la red SimpleNet con aumentación de datos en primer entrenamiento	28
Ilustración 22. Fscore de la red SimpleNet con aumentación de datos en primer entrenamiento	28
Ilustración 23. Fscore del segundo entrenamiento de la red SimpleNet	29
Ilustración 24. Fscore del tercer entrenamiento de la red SimpleNet.....	29
Ilustración 25. Métricas de entrenamiento de la red Mini SqueezeNet sin aumentación de datos ...	31
Ilustración 26. Fscore de la red Mini SqueezeNet sin aumentación de datos	31
Ilustración 27. Métricas del primer entrenamiento de la red Mini SqueezeNet con aumentación de datos	32
Ilustración 28. Fscore del primer entrenamiento de la red Mini SqueezeNet con aumentación de datos	32
Ilustración 29. Fscore de la red mini SqueezeNet durante el segundo entrenamiento con aumentación de datos	33
Ilustración 30. Métricas de entrenamiento de la red ResNet-11 sin aumentación de datos	34
Ilustración 31. Fscore de la red ResNet sin aumentación de datos	35
Ilustración 32. Métricas de entrenamiento de la red ResNet-11 con aumentación de datos	35
Ilustración 33. Fscore de la red ResNet-11 con aumentación de datos	36
Ilustración 34. Fscore de la red ResNet-11 durante el segundo entrenamiento con aumentación de datos	36

Lista de tablas

Tabla 1. Asignación de puntaje para determinar el grado del cáncer de mama	9
Tabla 2. Transformaciones y sus valores máximos aplicados en la aumentación de datos	20
Tabla 3. Arquitectura SimpleNet junto con la cantidad de parámetros.....	26
Tabla 4. Detalle de la arquitectura MiniSqueeze	30
Tabla 5. Detalle de la arquitectura ResNet-11	34
Tabla 6. Comparación del rendimiento de las arquitecturas probadas.....	38
Tabla 7. Comparación del rendimiento de las arquitecturas probadas unidas al selector de candidatos.....	38

Resumen

El recuento de células mitóticas consiste en una etapa fundamental en la gradación de un tumor. Sin embargo el proceso está sujeto a la subjetividad del patólogo que realiza el análisis, por lo que se presenta la necesidad de establecer un marco más objetivo de evaluación. En este proyecto se presenta un método de detección de células mitóticas automático basado en redes convolucionales profundas que logra subsanar en cierto grado el problema anteriormente descrito. Los resultados que logra se acercan bastante a otros logrados como métodos del estado del arte, pero logrando una tiempos de ejecución muchos más cortos.

1. Introducción

En muchos casos, los doctores necesitan hacer una biopsia para diagnosticar el cáncer. El cual es un procedimiento en el cual se analiza una muestra de tejido extraída del paciente. El análisis lo realiza un patólogo usando un microscopio para ver si hay células cancerosas. En los años recientes ha surgido la tendencia de la digitalización de las imágenes histológicas. Escáneres digitales, junto con visualizadores digitales pretenden proveer la experiencia de visualizar una muestra en un monitor de computadora de forma análoga a observarla bajo un microscopio, pero con todos los beneficios adicionales de un formato digital.

La actividad mitótica es uno de los pronosticadores más relevantes para el diagnóstico de tejido canceroso. Se expresa como la cantidad de figuras mitóticas por área de tejido. Los tumores agresivos poseen una alta tasa de proliferación, lo que se refleja en un alto número de células en mitosis presentes en las muestras histológicas. La tarea de contar las células mitóticas la puede realizar casi cualquier patólogo, pero en la práctica esta actividad posee problemas de reproducibilidad que son causados por la subyacente subjetividad del proceso. La detección automática de células en mitosis es vista como una solución al problema de la subjetividad, además de una solución en términos de costo y tiempo en el diagnóstico de cáncer.

Para la detección de células mitóticas se propone la utilización de Redes Neuronales Profundas la cual consiste en una Red Neuronal Convolutiva. Estas redes están inspiradas en la organización de la corteza visual animal y están logrando resultados de vanguardia en tareas difíciles de visión por computador tales como detección facial y reconocimiento óptico de caracteres manuscritos.

En términos de reconocimiento, se busca detectar automáticamente figuras mitóticas en imágenes de una región acotada de una muestra histológica. Puesto que lo importante es detectar el número de células mitóticas, no es relevante clasificarlas por su tamaño o forma. En ese sentido se formula como un problema de detección.

2. Objetivos

Para poder cumplir con la tarea de detección de células mitóticas, se plantean los siguientes objetivos.

2.1. Objetivo General

- Diseñar e Implementar una Red Neuronal Profunda para detectar automáticamente células mitóticas en imágenes histológicas.

2.2. Objetivos específicos

- Investigar sobre el aprendizaje profundo con redes neuronales.
- Habilitar un ambiente de desarrollo usando Theano o TensorFlow junto con Keras.
- Identificar las características de células mitóticas en imágenes histológicas.
- Crear un conjunto de datos para entrenamiento en base a las características identificadas
- Diseñar y definir la mejor arquitectura de una Red Neuronal que replique lo logrado por el estado del arte.
- Implementar la solución en algún lenguaje de alto nivel
- Validar los resultados obtenidos, contrastando con las clasificaciones hechas por oncólogos mediante medidas de desempeño usadas en clasificación automática

3. Trasfondo biológico

La teoría celular señala que la célula es la unidad básica de la vida. Todos los organismos están compuestos de células y todas las células derivan de otras precedentes. En el núcleo de la célula es donde se almacena el material genético o ADN. La reproducción celular se puede producir mediante mitosis o meiosis. La meiosis es el proceso en el cual una célula diploide (que contiene 2 pares de cromosomas) se divide en cuatro haploides (un solo par de cromosomas). Este proceso aporta diversidad genética y es la forma en que se generan los óvulos y espermatozoides. En contraste durante la mitosis se generan dos células con el mismo material genético. Un error durante la clonación de material genético generalmente tiene la consecuencia de que la célula se autodestruya mediante un proceso llamado apoptosis. En caso de que la célula no se autodestruya se produce una mutación genética que puede desarrollarse en cáncer.

El principal resultado de la mitosis es la transferencia del ADN de la célula padre en dos células hijas. Las células hijas deben ser idénticas a la célula padre, por lo que antes de dividirse debe duplicar cada cromosoma. Esto se produce durante la interfase o el periodo que ocurre entre cada mitosis. La interfase es la fase más larga del ciclo celular, ocupando un 90% de éste. Consta principalmente de un periodo de crecimiento celular, seguido de un periodo de duplicamiento de sus estructuras internas y ADN y por último un periodo de preparación para la mitosis.

3.1. Etapas de la mitosis

La mitosis es la etapa del ciclo celular en la cual el ADN replicado se separa en dos núcleos nuevos y consiste en 4 sub etapas [1]:

3.1.1. Profase.

Durante esta etapa se condensa el ADN, que durante la interfase se encontraba descondensado en una forma llamada cromatina, en cromosomas. Se disuelve el núcleo y microtúbulos se conectan con los cromosomas. Durante esta etapa se puede observar distintivamente los cromosomas. En la figura 1.a se observa una imagen histológica de una célula en prometafase, mientras que en la figura 1.b un dibujo a mano en cual se observa con mayor claridad la cromatina nuclear condensada.

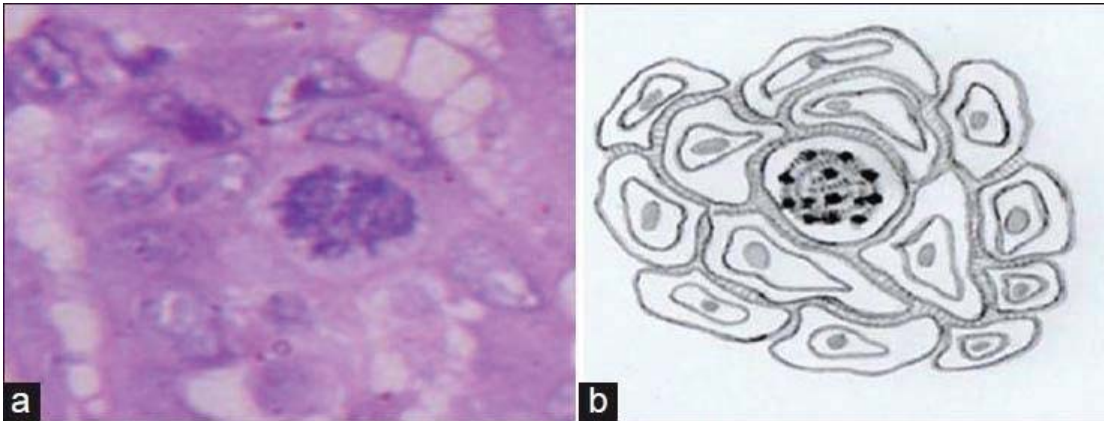


Ilustración 1. Célula en profase junto a un dibujo explicativo a mano

3.1.2. Metafase

Luego que los microtúbulos se hayan anclado a los cromosomas, estos empiezan a jalar a los cromosomas hacia polos opuestos de la célula. Lo anterior resulta en que los cromosomas se alinean en un plano ecuatorial imaginario en la mitad de la célula

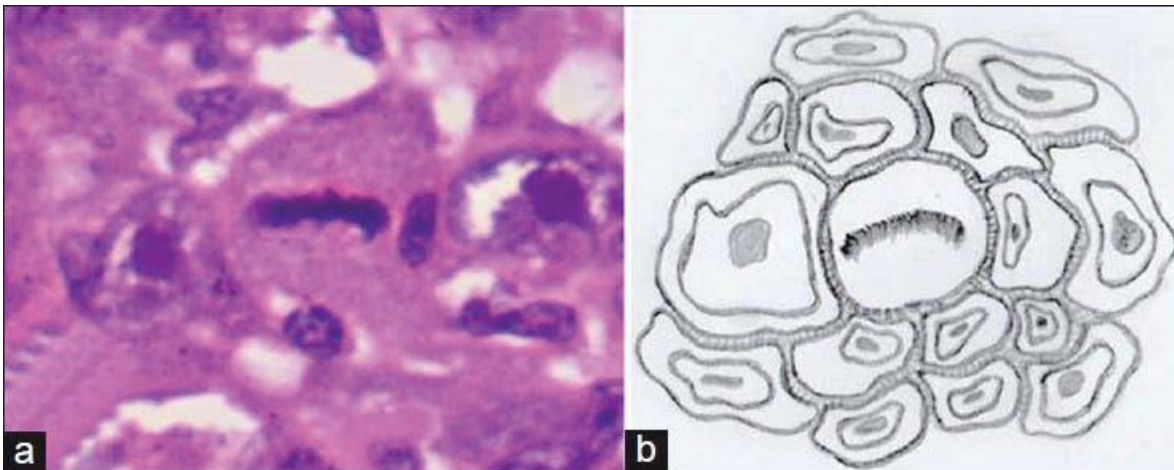


Ilustración 2. Célula en metafase junto a un dibujo explicativo a mano

3.1.3. Anafase

Los cromosomas duplicados se alejan y se comienzan a mover hacia los polos de la célula. Los microtúbulos se empujan entre sí, lo que causa que la célula se alargue. Al final de la anafase, la célula ha conseguido separar dos juegos idénticos de material genético en dos grupos definidos, cada uno alrededor de un centrosoma.

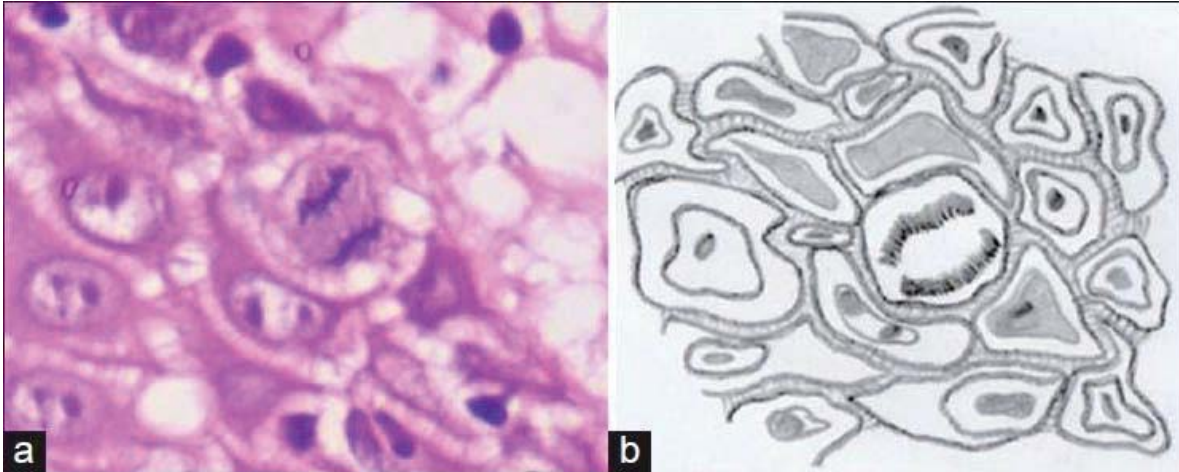


Ilustración 3. Célula en anafase junto a un dibujo explicativo a mano

3.1.4. Telofase

Los microtúbulos continúan alargándose, alargando aún más la célula. Se reforma el núcleo celular, los cromosomas vuelven a descondensarse en cromatina. Simultáneamente está ocurriendo otro proceso llamado citocinesis en el cual la célula se divide como tal, formando dos células idénticas entre sí.

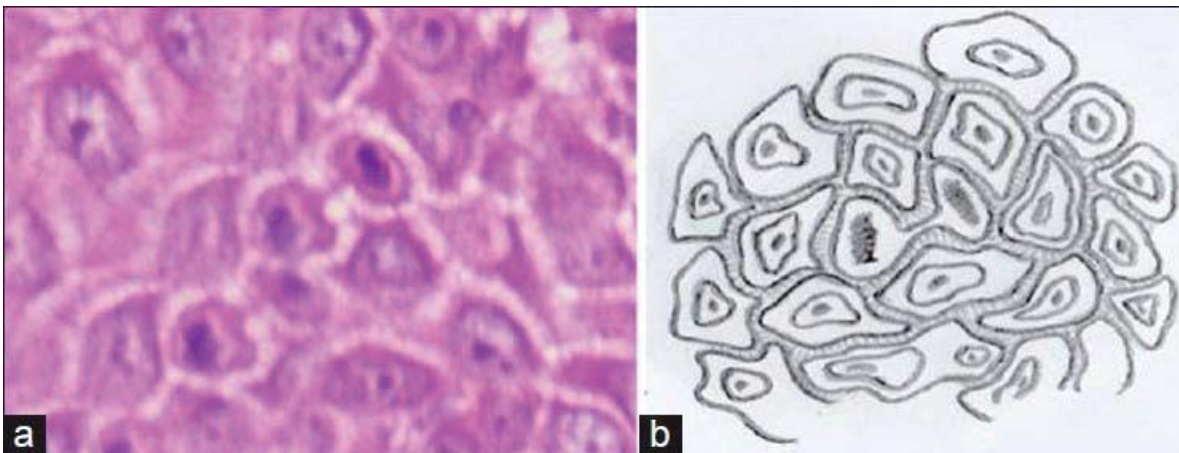


Ilustración 4. Célula en telofase junto a un dibujo explicativo a mano

3.2. El cáncer

Es el nombre que se le da a un conjunto de enfermedades relacionadas. En todos los tipos de cáncer las células comienzan a dividirse mediante mitosis sin detenerse y se diseminan por a los tejidos de su alrededor.

Normalmente, las células humanas crecen y se dividen para formar nuevas células a medida que el cuerpo las necesita. Cuando las células normales envejecen o se dañan, mueren, y células nuevas las remplazan. Sin embargo, en el cáncer, este proceso se descontrola. A medida que las células se hacen más anormales, las células viejas o dañadas sobreviven, y nuevas células anormales se forman cuando no son necesarias. Estas nuevas células se siguen dividiendo sin control y eventualmente forman masas de tejido que se denominan tumores. [2]

Los tumores formados por células cancerígenas se les denominan tumores malignos, y a diferencia de los tumores benignos, estos tienen la capacidad de extenderse a otras áreas del organismo y continuar creciendo en varias zonas. Este proceso se le conoce como metástasis.

3.3. Diagnóstico del cáncer

El primer paso será un análisis a la historia clínica del paciente, incluyendo antecedentes de familiares que pudiesen haber sufrido de cáncer. Posteriormente el médico puede solicitar una serie de pruebas como análisis de laboratorio o de imágenes. [3]

El análisis de sangre, orina u otros fluidos del cuerpo puede indicar concentraciones altas o bajas de ciertos componentes. Estos resultados si bien son útiles no son suficientes para diagnosticar cáncer.

Otro tipo de análisis complementarios son los visuales. En los cuales se generan imágenes de ciertas zonas del cuerpo. Estas imágenes se pueden formar de diferentes formas, tales como por ejemplo: La tomografía computarizada, en la cual una máquina de rayos X toma imágenes detalladas de los órganos. Las ecografías a partir de la emisión de ondas de sonido y sus ecos. La resonancia magnética genera imágenes a partir de un potente imán. La tomografía con emisión de positrones es una técnica nueva que es útil para detectar tumores malignos en etapas tempranas. Consiste en la inyección de un marcador radioactivo que luego es detectado por una máquina que genera una imagen tridimensional indicando donde se acumula éste.

Finalmente para llegar al diagnóstico con certeza se suele realizar una biopsia, una muestra de tejido, con el fin de realizar un análisis histológico por un patólogo en un microscopio.

3.4. Grado de un tumor

Es una forma de clasificar las células de un tumor cuando son observadas bajo un microscopio [4]. El grado de un tumor es una indicación de la rapidez con la que probablemente crecerá y se extenderá el tumor. Si las células del tumor y la organización del tejido del tumor se asemejan a las células y a los tejidos normales, el tumor se llama "bien diferenciado". Estos tumores tienden a crecer y a extenderse a un paso más lento que los tumores que son indiferenciados, los cuales tienen células que se ven anormales y a los que les pueden faltar las estructuras de tejido normal. Con base en estas y en otras diferencias de apariencia al microscopio, los médicos asignan un grado numérico a la mayoría de los cánceres.

El grado de un tumor no es lo mismo que la fase en la que se encuentra. La fase se refiere a lo extendido que está el tumor principal y si las células se han extendido a otras partes del cuerpo. La fase de un tumor se basa en la ubicación del tumor primario, el tamaño, si están afectados los ganglios linfáticos regionales y el número de tumores.

3.5. Sistema de gradación de Nottingham

Es un sistema para asignar el grado de un tumor para el cáncer de mamas. Se le conoce también como la modificación de Elston-Ellis del sistema de Scarff-Bloom-Richardson. Es el sistema que se recomienda utilizar para determinar el grado de un tumor en el caso de cáncer de mamas [5]. Consiste en asignarle una puntuación de 1 a 3 a tres características de la muestra. [6]

- Formación tubular: indica que porcentaje de los conductos de leche materna aún conserva una estructura normal
- Pleomorfismo nuclear: se refiere a la variación en el tamaño y forma de los núcleos de las células del tumor en relación a células normales.
- Conteo mitótico: es cuantas células están en una etapa de división en un área determinada. Una de las características de los tumores cancerígenos es que sus células se dividen sin control. Para juzgar y evaluar el conteo mitótico, este se debe hacer en la periferia del tumor y luego de haber observado 10 zonas diferentes de la muestra

Una vez asignado el puntaje a cada característica, se suman todos estos. Resultando en una puntuación final que se encuentra entre 3 y 9. Una vez realizado esto el grado de un tumor se obtiene a través del siguiente criterio:

- 3 a 5 puntos: grado 1 o tumor bien diferenciado
- 6 a 7 puntos: grado 2 o tumor medianamente diferenciado
- 8 a 9 puntos grado 3 o tumor indiferenciado.

En la tabla 1 se muestra la forma de asignar la puntuación a cada característica.

CARACTERÍSTICA	PUNTUACIÓN
FORMACIÓN TUBULAR	
La mayoría (>75%)	1
Moderadamente (10 – 75%)	2
Poco o nada (< 10%)	3
PLEOMORFISMO NUCLEAR	
Pequeño, núcleo celular uniforme	1
Moderado incremento en tamaño y variabilidad	2
Variación marcada	3
CONTEO MITÓTICO (EN UN DIAMETRO DEL CAMPO VISUAL DE 0.59mm)	
0 - 9	1
10 – 19	2
>20	3

Tabla 1. Asignación de puntaje para determinar el grado del cáncer de mama

4. Descripción del problema

El conteo mitótico es una evaluación importante en el diagnóstico de cáncer, no solo de mamas sino de todos los tipos. La evaluación de la muestra histológica la realizan patólogos de todo el mundo, sin embargo es una tarea tediosa y muy sujeta a la subjetividad del profesional debido a que las formas de las células mitóticas encontradas en las muestras pueden diferir de las encontradas en la literatura, por lo mismo resulta en una tarea no trivial identificarlas.

Debido a la preparación de la muestra, que consiste en teñirla con hematoxilina-eosina (H&E), los núcleos celulares se tiñen de tonalidades azul púrpura, mientras que otras estructuras se tiñen de tonalidades rosa. Si bien como característica general, las células mitóticas se vuelven más oscuras que las demás, esto no es suficiente para identificarlas. Existen otras características, como se apreció durante las etapas de la mitosis, estas células tienden a volverse más alargadas también [7]

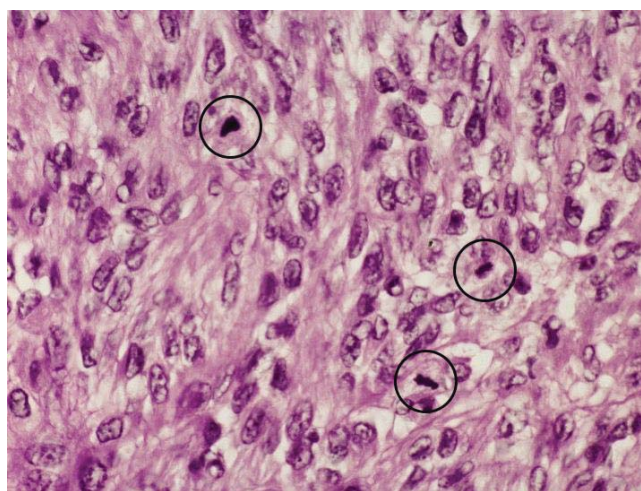


Ilustración 5. Ejemplo de células mitóticas. (Fuente: <https://www.pathologystudent.com/?p=5689>)

Si bien en la imagen 5, pareciera ser fácil la tarea de identificar células mitóticas, la realidad es que no todas las células “más oscuras” son mitosis, existen muchos otros componentes celulares que tienen una apariencia similar, siendo algunos de ellos células apoptóticas (que sufren una muerte programada), con núcleos necróticos, núcleos comprimidos, pequeños defectos en la preparación de la muestra que puede, por ejemplo sobre teñir una célula, entre otros. Hay que tener en cuenta que al tratarse de células que forman parte de un tumor cancerígeno suelen presentar formas deformes en relación a células sanas. Pueden aparecer células con núcleos más grandes e irregulares. Los errores en la duplicación de las células se hacen más frecuente, permitiendo observar por ejemplo mitosis tripolares o multipolares en los que la célula intenta dividirse en tres o más células hijas en vez de dos como es lo habitual

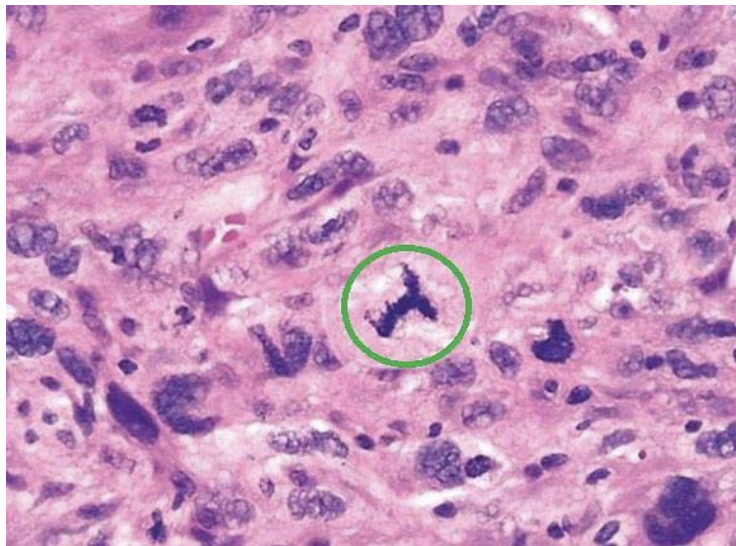


Ilustración 6. Ejemplo de mitosis tripolar

Para apoyar la automatización del conteo de mitosis de manera automatizada es que se organizaron dos competencias. Primero el MITOS Challenge [8] en el 2012 que contaba con solo 5 muestras con 10 campos de gran aumento (HPF, del inglés High Power Field) con 226 mitosis anotadas en total. Al año siguiente se organizó el AMIDA 2013 Grand Challenge on Mitosis Detection [9] que buscaba utilizar un dataset más amplio con 23 muestras y una cantidad de HPF variable por cada muestra, contando 649 mitosis anotadas. Estas muestras son más desafiantes al contener más elementos similares a células mitóticas, por lo que se buscaba avanzar en el estado del arte con clasificadores que diferenciaron de mejor manera los falsos positivos con las células mitóticas.

4.1. Obtención de las imágenes histológicas

Las muestras histológicas utilizadas en el MITOS dataset fueron anotadas por el departamento de patología del hospital Pitié-Salpêtrière en Paris, Francia. Las muestras se tiñeron con H&E y fueron escaneadas por dos microscopios: Asperio Scanscope XT

(Scanner A) y el Hamamatsu Nanozoomer 2.0-HT (Scanner H). La digitalización de las imágenes se realiza con un zoom de x40 con una resolución espacial de 0.25 $\mu\text{m}/\text{pixel}$. Se seleccionan 10 campos de gran aumento por cada muestra, estos tienen un tamaño de 512 x 512 μm , lo que equivale a un área de 0,262 mm^2 o un diámetro del campo visual de 0.58mm. El MITOS dataset es posible descargarlo desde la dirección http://ludo17.free.fr/mitos_2012/download.html.

El AMIDA13 dataset se digitalizó usando solo el microscopio Asperio Scanscope XT. Las HPF son tratadas de manera diferente, se incluyen todas las zonas que contienen al menos 1 mitosis. El borde de la muestra se marca con un marcador negro y se ignoran todas las HPF que intercepten esta marca. Alrededor del 6% de las figuras mitóticas se encontraban en estas áreas y por lo tanto no fueron incluidas en el dataset [10].

Para determinar las células mitóticas, dos patólogos independientes analizaron las muestras. Las zonas en las que ambos estuvieron de acuerdo fueron inmediatamente anotadas como mitosis, las anotaciones discordantes fueron enviadas a otros dos patólogos que las analizaron para tomar la decisión final. De esta forma todas las anotaciones del dataset han sido aceptadas por al menos dos expertos.

5. Estudio del estado del arte

Como casos de interés sobre la segmentación de mitosis, se ha seleccionado los métodos que han obtenido los mejores resultados de ambas competencias. Los resultados se encuentran en las páginas respectivas de ambas competencias, lamentablemente no todos los participantes de la competencia de MITOS Challenge escribieron artículos sobre sus métodos. Para la competencia AMIDA13 el artículo de evaluación de la competencia [10] ofrece una descripción resumida sobre los métodos participantes.

5.1. Normalización de las muestras

Las muestras histológicas obtenidas después de ser teñidas con H&E pueden presentar una gran variabilidad, las que pueden deberse a una sobrexposición a los tintes, procesos de preparación distintos en los tejidos, el tiempo de exposición al ambiente e incluso al microscopio utilizado. Los métodos tradicionales para imágenes naturales no son aplicables en esta situación (normalización de histogramas, color matching) debido a que se asume que la proporción de los tintes es igual, lo que no necesariamente es correcto, por lo que se pueden introducir errores no deseados. Por otro lado es más apropiado utilizar técnicas de normalización luego de separar el aporte de cada tinte al color final de cada muestra. Los métodos que se presentaran a continuación presentan este enfoque.

5.1.1. Normalización basada en descomposición de valores singulares [11].

Para normalizar los colores de una muestra en primer lugar se transforman los colores a un espacio llamado Densidad Óptica (OD) utilizando la fórmula:

$$OD = -\log_{10}(I) \quad (1)$$

Siendo I un vector RGB cuyos componentes han sido normalizados entre 0 y 1. Posteriormente se calcula la descomposición en valores singulares. Se toman las direcciones de la descomposición correspondientes a los valores singulares (SVD) más grandes y se forma un plano con ellas. Se proyectan los píxeles que se encuentran en el espacio OD en el plano. Se calcula el ángulo de cada punto con respecto a la primera dirección en la descomposición de los valores singulares. Dependiendo del ángulo se determina la concentración de hematoxilina y eosina presentes en el píxel, lo que permite separarlos en dos vectores (o imágenes) distintos. A cada vector, denominados vector del tinte (stain vector), se les aplica una normalización basada en la intensidad del 99% de los píxeles. Una vez normalizado se devuelve al espacio de color RGB.

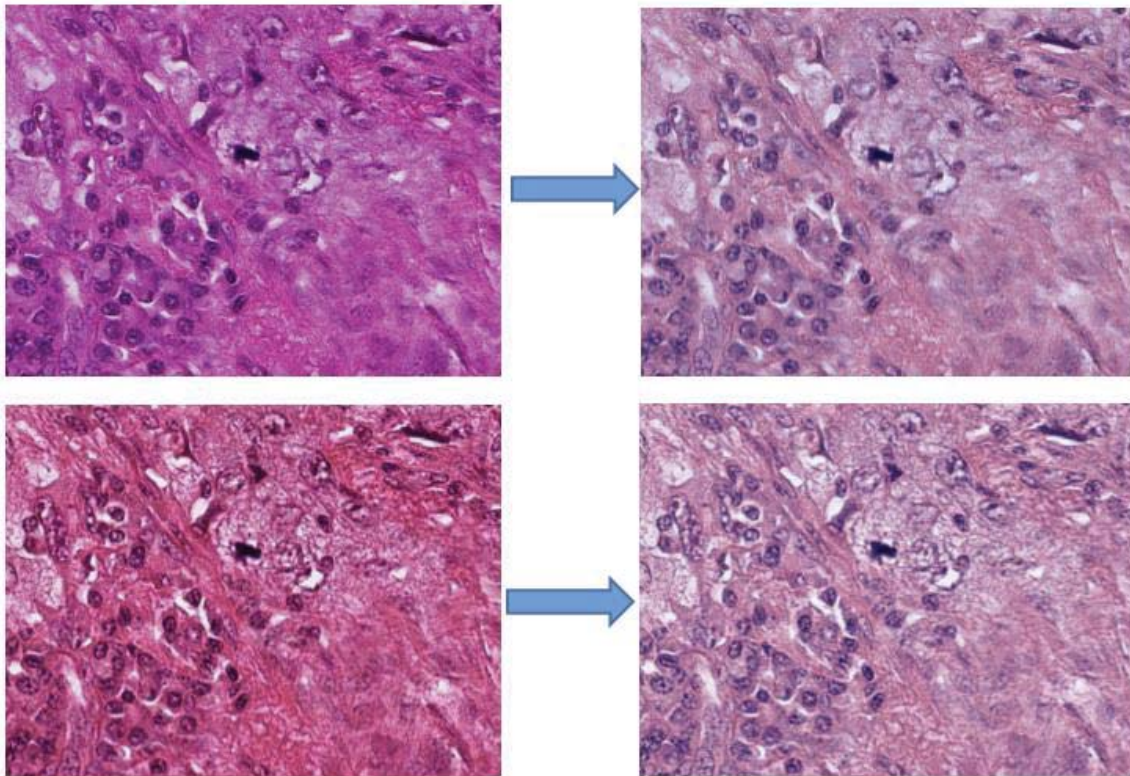


Ilustración 7. Normalización de la misma muestra tomada con diferentes escáneres

5.1.2. Normalización luego de separación basada en regularización dispersa [12]

Está basado en la factorización de matrices no negativas

$$V = WH \quad (2)$$

Donde V es la imagen vectorizada en el espacio de densidad óptica, W es la matriz del tinte y H la saturación. Para encontrar los valores de WH se basa en el siguiente problema de minimización de costo:

$$\min_{W,H} \frac{1}{2} \|V - WH\|_F^2, W, H > 0 \quad (3)$$

A la función objetivo anterior le agrega regularización dispersa L1 en la matriz H, por lo que el problema de optimización final es:

$$\min_{W,H} \frac{1}{2} \|V - WH\|_F^2 + \lambda \sum_{j=1}^r \|H(j)\|_1 \quad (4)$$

La optimización se resuelve dejando un parámetro fijo, mientras se optimiza el otro. Para la implementación el autor utiliza la librería públicamente disponible SPAMS y se utiliza 0,1 como valor por defecto para el parámetro λ .

Una vez separado se busca normalizar la matriz de tinte, pero basado en otra imagen llamada objetivo, esta segunda imagen también se separa en la influencia de los tintes H&E y se determina el rango de color de estos. Se ajusta la saturación de la imagen fuente H_s para que este en el mismo rango que la saturación objetivo H_t . Luego se intercambian las matrices de tinte.

$$V_s = W_t H_s \quad (5)$$

Finalmente se devuelve el resultado al espacio de color RGB.

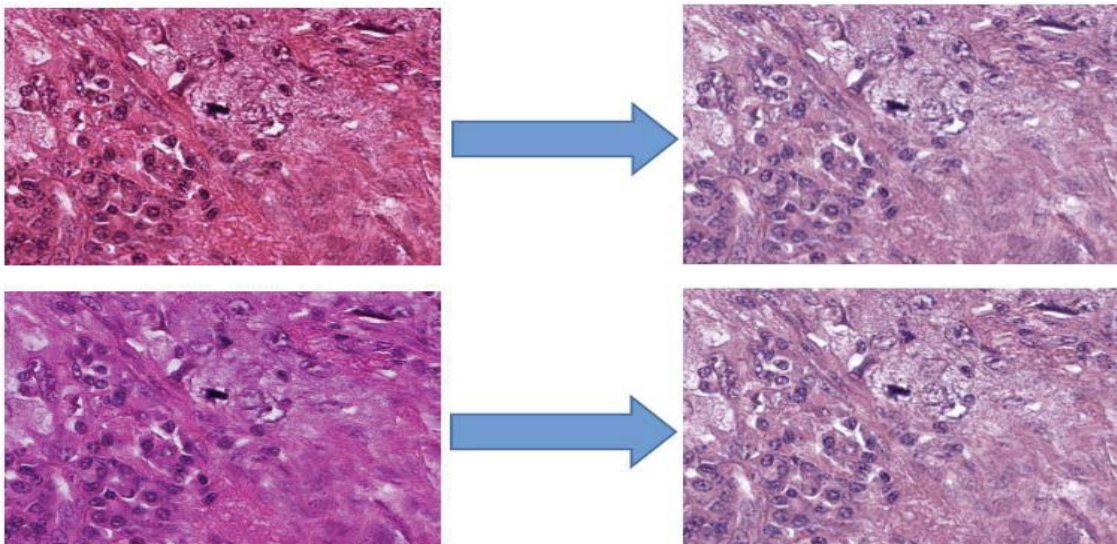


Ilustración 8. Normalización basada en regularización dispersa

5.2. Segmentación de candidatos

5.2.1. Basado en operaciones morfológicas.

La imagen se transforma a un espacio de color donde se muestra la influencia de los tintes de hematoxilina y eosina mediante una desconvolución utilizando el método de Macenko et al [11]. Se descarta el canal de eosina y se segmenta únicamente con el canal de la hematoxilina. Se aplica un filtro de mediana para hacer la imagen más borrosa y así reducir el ruido. Se continúa con una operación morfológica de dilatación seguida de una erosión, con el fin de eliminar pequeños huecos dentro de los núcleos. Se calcula la diferencia con la misma imagen pero con un filtro de mediana con un kernel mucho más grande, los componentes conectados que sobrevivan a dos operaciones de umbral se consideran candidatos [13]. Este método solo falló en identificar dos células mitóticas en la data de entrenamiento en el AMIDA13. Debe ser complementado con algún clasificador para reducir la cantidad de falsos positivos.

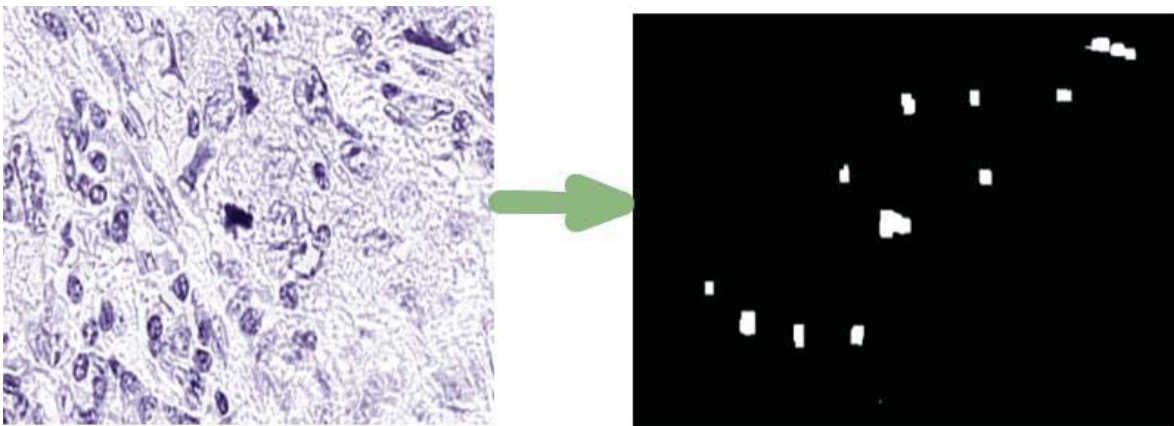


Ilustración 9. Imagen binaria donde se muestran los candidatos seleccionados

5.2.2. Laplaciano del Gaussiano (LoG) [14]

Es un método para detectar regiones (blobs) utilizando una convolución cuyo kernel viene dado por la ecuación:

$$LoG(x, y) = -\frac{1}{\pi\sigma^4} \left[1 - \frac{x^2+y^2}{2\sigma^2} \right] e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (6)$$

El kernel generado tiene un fuerte valor negativo en el centro que se acerca a 0 rápidamente a medida que se aleja del centro.

El LoG se aplica sobre una imagen que se encuentra transformada a proporción de azul (blue ratio, BR) que es una imagen en la que se privilegia el canal azul, por sobre los demás. La ecuación del canal BR está dada por:

$$BR = \frac{100*B}{1+R+G} * \frac{256}{1+B+R+G} \quad (7)$$

Al resultado de LoG se aplica una operación de umbral para binarizar la imagen, seguida opcionalmente de una operación morfológica para cerrar agujeros pequeños. Los candidatos se seleccionan de acuerdo al tamaño que tengan.

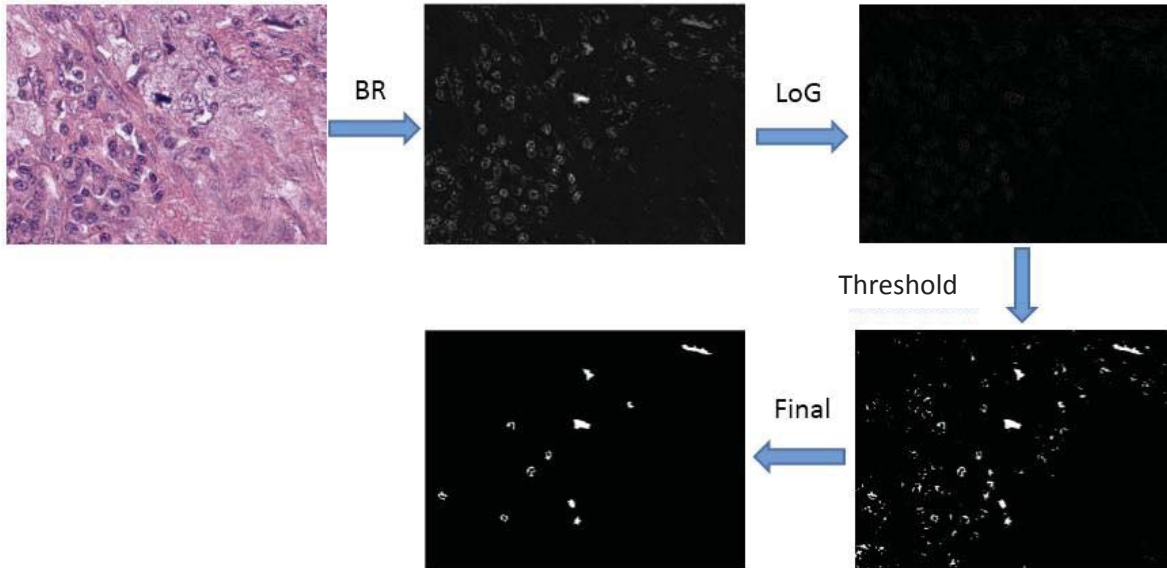


Ilustración 10. Paso a paso de la selección de candidatos basada en LoG

5.3. Clasificador

5.3.1. Red Neuronal Convolutiva [15]

El método ganador tanto de la MITOS Challenge, como de AMIDA13 estuvo basado en redes neuronales convolucionales. La entrada a la red es una ventana deslizante de tamaño 63 x 63, en la que ingresan los píxeles RGB sin ningún pre procesamiento. La salida de la red es la probabilidad de que el centro de la ventana se encuentre en un radio de 20 píxeles del centroide de una célula mitótica. Al ir deslizando la ventana en la imagen de prueba se puede ir mapeando las salidas de la red neuronal en una matriz, que puede ser observada como una imagen en tonalidades de blanco y negro. Idealmente se espera que solo los 20 píxeles alrededor del centroide obtengan una probabilidad mayor a 0, pero en la práctica la salida es bastante ruidosa. Para suavizar la salida se le aplica un filtro de desenfoque a través de una convolución. Como paso final se buscan los máximos locales en el mapeo de probabilidad y ellos se agregan al grupo de mitosis detectadas como el centroide de la célula.

Este método no solo no cuenta con ningún pre procesamiento, sino que tampoco ningún selector de candidatos. Para lo anterior trata de representar lo mejor posible células no mitóticas que son muy parecidas a las reales, en desmedro de zonas menos desafiante que son evidentemente pertenecientes a la clase “no mitosis”. Al final de la construcción del dataset de entrenamiento se obtienen los siguientes porcentajes: alrededor de 10% de células mitóticas, 40% de muestras aleatorias de pixeles no mitóticos y 50% de células muy similares a las mitóticas. Debido a que la cantidad de muestras es bajo en relación a la variabilidad del problema, se aumenta de forma artificial el set de entrenamiento, para ello se aplica un filtro de espejo, además de rotar las muestras buscando abordar el problema de la invariabilidad de la rotación.

5.3.2. Clasificador por características morfológicas y de canales múltiples [14]

La imagen se transforma a distintos espacios de color, como HSV, Lab, Luv, BR. Se realiza un análisis de histograma a cada canal de los espacios de color seleccionado y discrimina aquellos que mejor capturan la información de las células mitóticas. Para clasificar un candidato se extraen una serie de características morfológicas con respecto a su forma, además de una serie de características estadísticas de los canales seleccionados. En total se recolectan 143 características, pero debido a su gran dimensión se realiza una evaluación de consistencia en subgrupos con el fin de reducir la dimensionalidad de las características. Finalmente se determinó que 20 era el subgrupo mínimo con características consistentes. Por ultimo las características seleccionadas son usadas para entrenar un árbol de decisión de profundidad 10. Este método logró una precisión de 69%, un recall de 74% y un f-score de 71% en la detección en el MITOS dataset.

6. Modelo propuesto

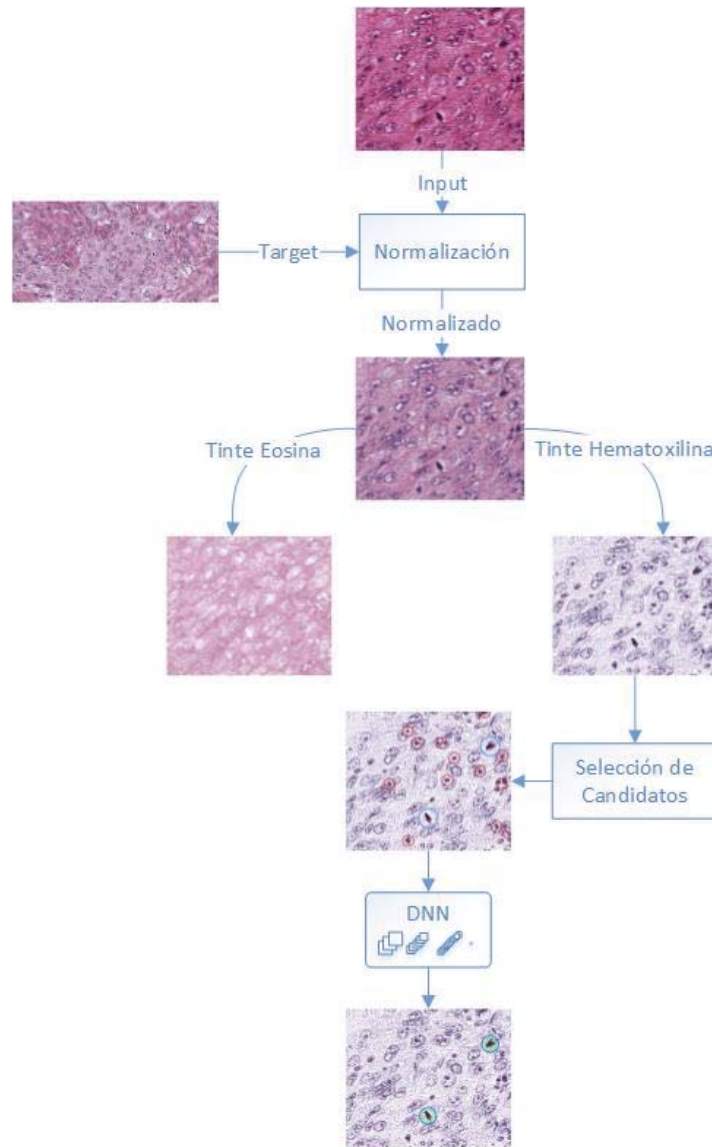


Ilustración 11. Arquitectura del modelo propuesto

Al igual que lo fue presentado en el estudio del arte, el modelo propuesto consta de tres partes. En primer lugar la normalización mediante el método propuesto por Vahadane [12] utilizando como target la muestra de color que se propuso en la competencia ICPR 2012. La normalización actúa separando la influencia de los tintes Hematoxilina y eosina, pero puesto que solo la Hematoxilina tiñe al núcleo celular, se procede usándola en las siguientes etapas. La selección de candidatos se discute en la sección 6.1 y su salida es una cantidad de muestras que pueden ser o no células mitóticas. La idea del selector de candidatos es que sea un método simple y rápido que pueda discriminar de forma efectiva y eficiente posibles células mitóticas. Finalmente se encuentra una red neuronal

convolucional profunda que se encarga de discriminar si efectivamente una célula seleccionada por el selector de candidatos es mitótica o no.

6.1. Selector de candidatos

El selector de candidatos es el encargado de seleccionar posibles ubicaciones donde encontrar células mitóticas. Debido a su simplicidad es muy rápido, de forma de evitar que el clasificador deba tomar la imagen completa como entrada, o utilice ventanas deslizantes que serían métodos mucho más lentos.

En primer lugar la imagen del canal de la hematoxilina se transforma a Blue Ratio, con lo que el fondo queda de un color oscuro, mientras que las zonas más azules son más claras. Para eliminar pequeños agujeros se recurre a la operación morfológica apertura utilizando un kernel de 5x5 compuesto únicamente de 1. Para binarizar la imagen, se procede al uso de un umbral. Los valores por debajo del umbral se les asigna 0, mientras que por encima 255. Por lo tanto la elección del umbral es muy importante, para evitar las variaciones de las imágenes es que se propone el uso de un umbral adaptativo a esta. O sea, el umbral se adapta a la tonalidad de la imagen, lo cual se logra en base a la media y la desviación estándar de los valores de la imagen. Siendo el umbral final de binarización la media + desviación estándar.

Una vez binarizada la imagen, se vuelve a aplicar la operación morfológica apertura, pero esta vez con un kernel de 7x7. Las células con un área menor a 70 y mayor a 1700 se filtran. Finalmente los candidatos que están a menos de 35 píxeles se fusionan como uno solo en el punto medio de sus centros.

Este método logra contener la mayoría de las células mitóticas dentro de los candidatos. En el set de pruebas falló en detectar 7 células. Estas 7 células fueron falsos negativos en el método de Ciresan [15] por lo que no se considera como un fallo importante.

7. Creación del set de datos

El set de datos usado en la competencia de detección de mitosis ICPR2012 consta de la muestra de 50 zonas de alto aumento tomadas de 5 pacientes. Se encontraba dividida en 35 zonas para entrenamiento y 15 para testing. Cada zona tiene asociado un archivo .csv en el cual aparece anotadas las células mitóticas presentes. El set de datos consta de 226 células mitóticas anotadas en las zonas para entrenamiento y 101 en las de testing, formando un total de 327 anotaciones. Es por lo tanto tarea del investigador utilizar algún método con el cual definir la clase no Mitosis.

Para la creación de la clase no-mitosis en primera instancia se utilizó un método presentado en la sección 5.2.2 para seleccionar candidatos. El método basado en el Laplaciano del Gaussiano aplicado sobre las imágenes en bruto, o sea sin normalizar. El

problema encontrado con este acercamiento es que es demasiado sensible a al valor de umbral establecido para realizar la bancarización candidatos (separar los candidatos del fondo). Esto se suma, además, a que las muestras histológicas en la vida real presentan variaciones que se deben a diversas causas, tales como: variaciones en la preparación de la muestra, cantidad de tinte, el microscopio usado, etc. Es por eso que existieron algunas imágenes donde se seleccionaba una gran cantidad de candidatos, habiendo otras donde selecciona demasiado pocos, habiendo incluso evitado seleccionar células mitóticas. Tal como se muestra en la figura 11, donde las marcas blancas corresponden a los candidatos detectados, mientras que los cuadros rojos son zonas donde estaban anotadas células mitóticas.

Es por eso, que se utilizan las células del selector de candidatos como muestras de la clase no-mitosis. Lo anterior resulta en una cantidad de 5290 muestras para la clase no mitosis para entrenamiento, y 2287 para pruebas.

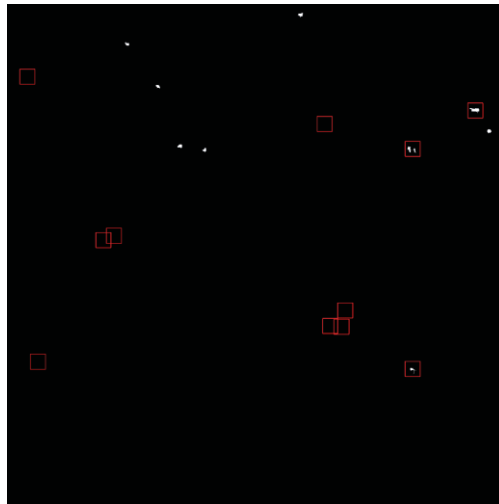


Ilustración 12. Muestra de la salida del selector de candidatos basado en el laplaciano del gaussiano

7.1. Aumentación de datos

El problema de detección de células mitóticas es rotacionalmente invariante, además para compensar la escasez del dataset de entrenamiento, se recurre a una aumentación de datos. Durante la fase de entrenamiento se le aplica a cada imagen una serie de transformaciones de manera aleatoria en un rango de valores. De esta manera, en cada iteración el clasificador recibirá de entrada una imagen que es diferente a las que ha sido entrenado, de forma de evitar que se sobreajuste en exceso. El rango de valores varía entre 0 y los especificados en la tabla 2. Además, se aplica el efecto espejo tanto vertical como horizontal. Los bordes se rellenan con la imagen girada, o sea con los pixeles del otro extremo de la misma.

TRANSFORMACIÓN	VALOR MÁXIMO
Rotación	44 grados
Desplazamiento vertical/horizontal	20 píxeles
Cizallado (shear)	0.21 radianes
Zoom	0.21%

Tabla 2. Transformaciones y sus valores máximos aplicados en la aumentación de datos

8. Redes Neuronales Convolucionales

Puesto que las redes neuronales convolucionales son las mejores en el análisis de imágenes, es que se propone su uso para la detección de células mitóticas. Las redes neuronales convolucionales son muy similares a las demás redes, la diferencia radica en que utilizan la convolución como función de propagación. Funcionan mejor cuando la entrada contiene información espacial como imágenes. La convolución actúa como una especie de filtro que va detectando ciertas características específicas de una imagen. Las redes convolucionales suelen utilizar ciertos tipos especiales de capas, a continuación se muestran los más comunes.

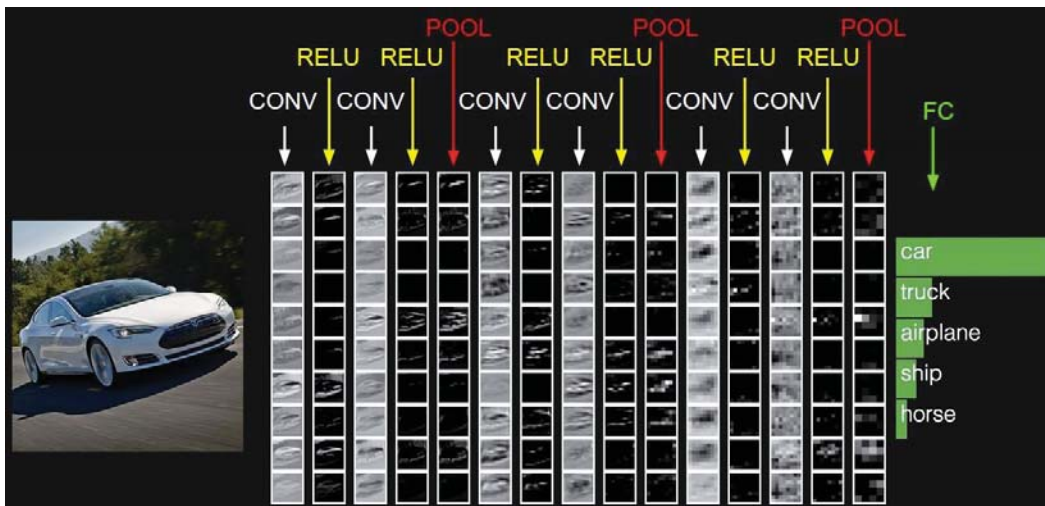


Ilustración 13. Visualización de una red neuronal. (Fuente: <http://cs231n.github.io/convolutional-networks/>)

8.1. Capa Convolutiva

Es una capa que consiste en una serie de operaciones de convolución sobre su entrada. La Convolución funciona aplicando un filtro, formalmente llamado kernel. Este kernel actúa como una ventana deslizante sobre el impulso de entrada, lo cual permite ahorrar parámetros para entrenar una red. Por ejemplo si una imagen posee un tamaño de $32 \times 32 \times 3$ (los 3 canales de colores rgb) entonces la primera capa de una red mlp tendría $32 \times 32 \times 3 = 3072$ neuronas, que además perderían información espacial de la imagen, puesto

que sería mucho más difícil reconocer las figuras si la imagen está aplanada. Sin embargo con una convolución los parámetros se reducen al tamaño del kernel, ej si se usa un kernel de 3x3x3 con 32 neuronas se tiene 864 pesos que optimizar.

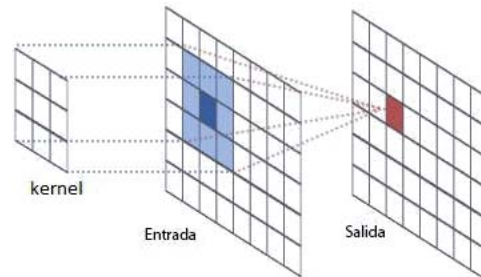


Ilustración 14. Convolución 2d

La convolución tiene la desventaja que reduce el tamaño de la imagen, un kernel de 3x3 reduce en 1 píxel, esto se puede obviar si lo que se desea es ir reduciendo gradualmente la imagen de los bordes, o se puede aplicar una estrategia llamada rellenar con ceros, para mantener el tamaño de la imagen. El relleno con ceros consiste en incrementar artificialmente el tamaño de la imagen original con ceros de forma que al aplicar una convolución la imagen no cambie de tamaño con respecto a la original. Para calcular la convolución muchas implementaciones recurren a la eficiencia que muestran muchas librerías blas para implementar una convolución como multiplicación de matrices. En primer lugar se cada campo receptivo posible se estira en base al tamaño del kernel, en el ejemplo de una imagen de entrada de 32x32x3 y un kernel de 3x3 se estiraría cada campo en un vector de largo 27, además se ejecutarían $32 \times 32 = 1024$ convoluciones (asumiendo que se halla usado relleno con ceros). Lo que resulta en una matriz de 27×1024 , notar que como los desplazamientos del kernel se superponen algunos miembros quedarán repetidos.

Los kernel presente en las neuronas también sufren una transformación similar, por ejemplo si la capa posee 32 neuronas, entonces se obtiene una matriz de la forma 32×27 . En este punto ya es posible realizar la multiplicación de matrices implementada en las librerías blas. Esta forma de realizar la convolución posee la desventaja de consumir demasiada memoria, puesto que algunos valores se almacenan dos veces y además la multiplicación de matrices es una operación costosa, sin embargo se obvia esta desventaja en pos de la velocidad en tiempo de ejecución de las ya mencionadas librerías blas.

8.2. Rectificador lineal

Es la función de activación más común en las redes neuronales convolucionales[16]. Consiste en la aplicación de la formula $\text{relu}(x) = \max(0, x)$ que básicamente impide que valores negativos se propaguen por la red.

8.3. Max pooling

Las redes profundas ocupan muchos parámetros, sobre todo para el procesamiento de imágenes, por eso es común insertar en ciertos lugares una capa que reduzca la representación de la señal de entrada. Esa capa es el max pooling, la cual reduce el tamaño de la imagen haciendo uso de la operación max sobre una ventana. El tamaño más común de la ventana es de 2x2 con un desplazamiento de 2, lo cual reduce el tamaño de la imagen a la mitad, requiriendo un 75% menos de parámetros y propaga solo las mejores características.

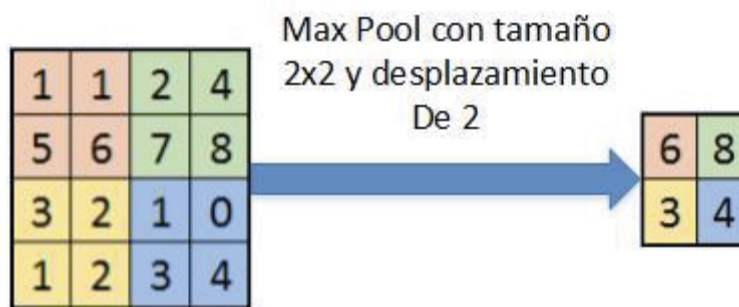


Ilustración 15. Ejemplo de una operación de max pooling con desplazamiento de 2

8.4. Dropout

Es una capa enfocada en la regularización de la red para reducir el sobreajuste, generalmente se aplica antes de la última capa. Durante la fase de entrenamiento se toma una probabilidad p y en base a ella se eliminan p por ciento de las conexiones. Solo se actualizan los pesos de los nodos que no fueron desconectados. La idea es que la red no vea que dos muestras sean totalmente iguales, a pesar de serlo, y de esta forma generalice de mejor manera su poder predictivo.

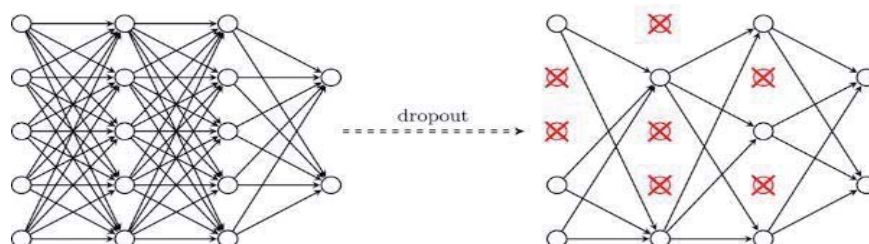


Ilustración 16. Efecto de la regularización mediante dropout en las conexiones. (Fuente: <https://cambridgespark.com/content/tutorials/convolutional-neural-networks-with-keras/index.html>)

8.5. Capa completamente conectada

Las capas completamente conectadas son las capas tradicionales de una red neuronal y se utilizan en una red convolucional al final de ésta para la toma de decisiones. La última capa convolucional se “aplana” para permitir una capa tradicional, a la cual se le aplica la regularización mediante dropout, y se conecta a la capa de salida con tantas neuronas como clases se busquen clasificar.

9. Arquitecturas e ideas involucradas

Gracias a la aplicación de distintas competencias de clasificación o segmentación del contenido de una imagen, tales como la ImageNet Large Scale Recognition Challenge [17] o el Microsoft Coco [18] es que la investigación de redes neuronales ha estado avanzado rápidamente, probando nuevas ideas y comparándolas en estas competencias.

9.1. Arquitectura VGG-16 Y VGG-19

La arquitectura VGG [19] destaca por su simplicidad de diseño y su profundidad, de 16 y 19 capas para las arquitecturas VGG-16 y 19 respectivamente. Utiliza solamente convoluciones de 3x3 que hasta la fecha se habían querido evitar, puesto que se pensaba que las convoluciones más grandes capturaban mejor las características adyacentes. Sin embargo el diseño de apilar convoluciones pequeñas, puede emular a una convolución grande, idea que será emulada por arquitecturas más recientes y además mantiene un número más pequeño de parámetros. Esta red no solo posee una profundidad mayor a la de otras arquitecturas a la fecha, sino que también posee mucha anchura, 512 nodos en las convoluciones más profundas y dos capas completamente conectadas de 4096 nodos hacen que esta red posea un número enorme de parámetros. Es más, al guardar los pesos en disco esta red ocupa 528 mb, la versión de 19 capas ocupa 548 mb.

9.2. Arquitectura Inception

La arquitectura inception, creada por Google, Busca generar una refactorización de las convoluciones tratando de reducir lo más posible el número de parámetros, de forma de poder crear arquitecturas más profundas ahorrando memoria. La red está formada por una arquitectura más pequeña que se repite varias veces, esta idea será usada también por otras arquitecturas que aparecieron luego. Se Retoma la idea de que varias convoluciones más pequeñas pueden emular a una más grande y se le suma la idea de que una convolución de 1x1 antes de una más grande, reduce la cantidad de características de entrada, lo que se le bautiza como cuello de botella [20]. En la versión 3 de la red inception se reautorizaron todas los kernels de más de 3, las arquitecturas anteriores poseían algunos de 5x5 u algunos

de 7x1. Todos ellos se eliminaron en favor de convoluciones más pequeñas, inclusive se juega con kernels asimétricos argumentando que ahorran parámetros, tomando el ejemplo presentado en el paper, “una convolución de 3x1 seguida de una de 1x3 es equivalente a dos capas de 3x3” [16].

Otra idea que se retoma es la presentada por Springenberg et al [21] bajo la cual propone desechar completamente las capas de max pooling para reducir el tamaño de las representaciones, y en su lugar reemplazarlas por convoluciones con un desplazamiento mayor a 1 (un desplazamiento de 2 también reduce el ancho y alto a la mitad)

Todas estas factorizaciones de los kernels son lo que les ha permitido crear una arquitectura mucho más profunda que vgg-19 y aun así posee menos parámetros. La arquitectura original de inception pesa 51 Mb, la versión 3 pesa 93 Mb.

9.3. Arquitectura ResNet

Viene dado por acortar Residual Network y su idea es simple. Saltarse capas y enviarla hacia capas más adelante y agregarlas al recorrido normal [22]. Los autores argumentan que este residuo de capas anteriores es más fácil de entrenar que utilizar las capas apiladas tradicionales. En el paper se muestra que solo hacen falta cuatro iteraciones para obtener buenos resultados en el dataset de cifar 10. Otro punto que los autores argumentan a favor es que incrementar la cantidad de capas de la red siempre aumenta la capacidad de generalizar, aunque no se algo proporcional (aumentar el doble de capas no reduce a la mitad el error). Sin embargo es posible que se necesite más evidencia empírica puesto que durante la construcción por parte del equipo de Google de la inception v4 y la Google res net, pusieron en duda esa afirmación [23].

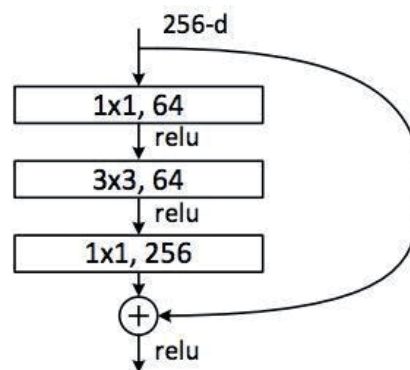


Ilustración 17. Módulo de la arquitectura ResNet

9.4. Arquitectura SqueezeNet

Es una arquitectura que busca ser profunda, pero con muchísimos menos parámetros y aun así sacrificar la precisión lo menos posible. En ese sentido toma todas las ideas mencionadas y las implementa para lograr la máxima disminución de parámetros posibles. En términos de precisión, logra resultados ligeramente a la red AlexNet en el set de datos de ImageNet 2012 [24]. La arquitectura busca cumplir con tres estrategias: tratar de utilizar lo más posibles filtros de 1×1 , puesto que tienen 9 veces menos parámetros que uno de 3×3 . Los filtros que se conectan a uno de 3×3 deben tener una cantidad reducida de parámetros y finalmente tratar de reducir la resolución tarde en la red, para no perder capacidad de generalizar. El módulo con el que se pretende lograr las estrategias anteriormente mencionadas se le bautizó como el módulo fuego el cual consiste en un filtro de 1×1 para reducir los parámetros, denominado squeeze filter, seguido de convoluciones para expandir, una de 3×3 concatenada a otra de 1×1 .

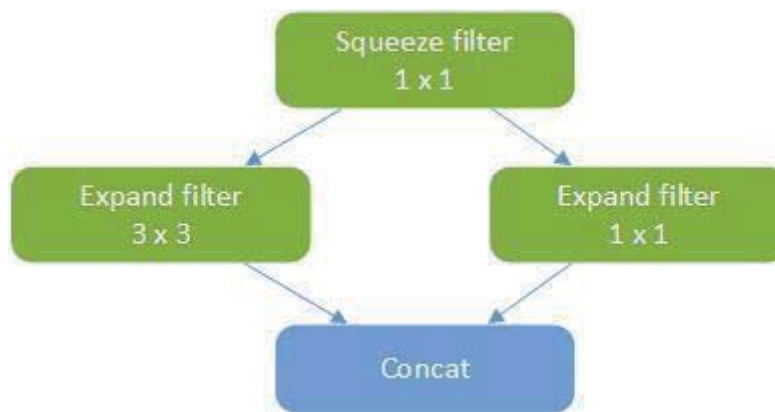


Ilustración 18. Representación del módulo fuego

La arquitectura consiste en utilizar 9 módulos fuego y una capa de global average pool como salida. El autor utiliza además métodos para compresión de arquitecturas que han sido utilizadas en la ya mencionada arquitectura AlexNet, consiguiendo así un modelo que ocupa 0.5 Mb en disco duro.

9.5. Arquitectura Enet

Es otra arquitectura que busca ser ligera para poder ser desplegada en entornos de poca capacidad de cómputo [25]. Esta red busca reducir el tamaño de la imagen lo más pronto posible, para ello concatena una capa de max pooling con un filtro de 3×3 . El módulo de esta red es bautizado como cuello de botella y hace un uso extensivo de filtros de 1×1 seguido de una capa de batch normalization. La red consiste en 14 módulos cuello de botella. El modelo pesa 0.7 mb utilizando precisión de 16 bit para el punto flotante.

10. Arquitecturas probadas

Para determinar el clasificador, es que se probaron tres arquitecturas diferentes. Estas estuvieron inspiradas en parte en las presentadas en la sección 9.

10.1. SimpleNet

Es una arquitectura de tres capas convolucionales con relativamente pocos parámetros y un diseño muy similar a una red feedforward.

Tipo	Filtro	Salida	Parámetros
Conv	32 x (3 x 3)	80 x 80 x 32	$896 = (3 \times 3 \times 3 \times 32) + 32$
Max pool	2 x 2, despl = 2	40 x 40 x 32	
Conv	32 x (3 x 3)	40 x 40 x 32	$9.248 = (3 \times 3 \times 32 \times 32) + 32$
Max pool	2 x 2, despl = 2	20 x 20 x 32	
Conv	64 x 3 x 3	20 x 20 x 64	$18.946 = (3 \times 3 \times 32 \times 64) + 64$
Max pool	2 x 2, despl = 2	9 x 9 x 64	
FC		64	$331.840 = (9 \times 9 \times 64 \times 64) + 64$
Total			360.545

Tabla 3. Arquitectura SimpleNet junto con la cantidad de parámetros

En una primera instancia se entrenó a la red con una aumentación de datos muy básicos, solamente en la clase mitosis. La cual consistía únicamente en rotar la imagen en 90, 180 y 270 grados, sumado a un filtro espejo. Para lidiar con las clases desbalanceadas se utiliza sensitive learning [26] aplicando un mayor error al equivocarse en la clase mitosis.

En esta primera instancia se observa que la red aprende bastante rápido, pero luego de alrededor de la iteración 60, el error es casi 0. Además de clasificar correctamente todas las muestras de entrenamiento. En el gráfico para que la curva error fscore equivale a 1 – fscore, esto se grafica de esta manera para que la curva siguiera a la curva de loss.

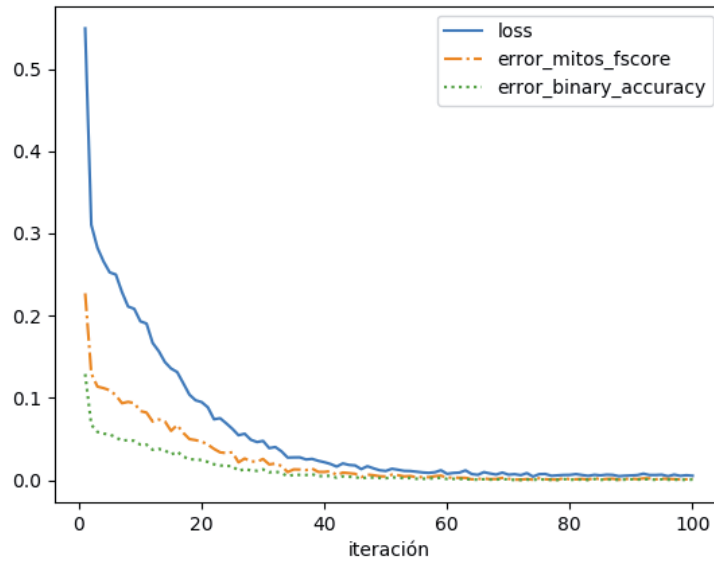


Ilustración 19. Métricas de entrenamiento de la red SimpleNet sin aumentación de datos

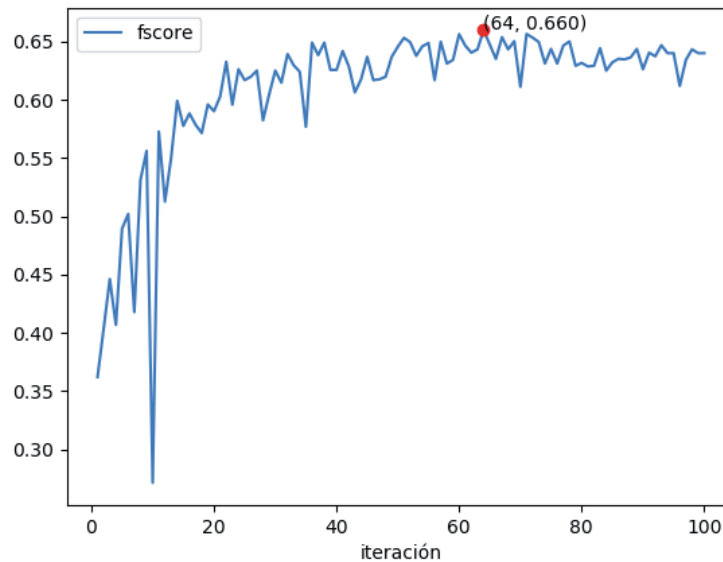


Ilustración 20. Fscore de la red SimpleNet sin aumentación de datos

Como se puede apreciar el fscore alcanzado rodea el 0.63 y se mantiene constante, por lo que entrenar la red más allá de la iteración 60 no tiene mucho sentido.

Puesto que los resultados sin utilizar aumentación de datos agresiva no fueron muy buenos, es que se aplicó. La aumentación se realiza en tiempo de ejecución, por lo que genera una sobrecarga al entrenamiento. Además, se genera una muestra distinta en cada iteración. Con el uso de aumentación de datos se logra que la red no se sobreajuste tanto como se observó con anterioridad, logrando mucho mejor resultado.

Debido a la aumentación de datos, la red se entrena de manera más lenta. Puesto que, como se mencionó, en cada iteración observa una muestra distinta. Además, una vez completado el entrenamiento se utiliza la distribución de pesos que mejor rendimiento tuvo, como inicialización para un nuevo entrenamiento. Pero esta vez, con una menor cantidad de iteraciones. Esto se debe a que al ir reduciendo el learning rate de la gradiente descendiente, las últimas iteraciones solo realizan pequeños retoques a los pesos. En contraste realizar un

nuevo entrenamiento con una buena inicialización permite explorar en mayor medida el espacio, sin el riesgo de reducir la precisión, puesto que se va guardando el modelo con mejor rendimiento.

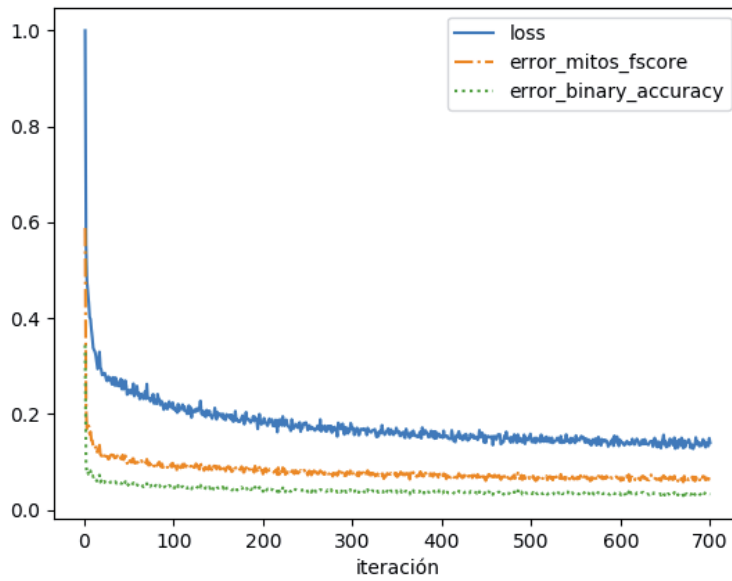


Ilustración 21. Métricas de entrenamiento de la red SimpleNet con aumentación de datos en primer entrenamiento

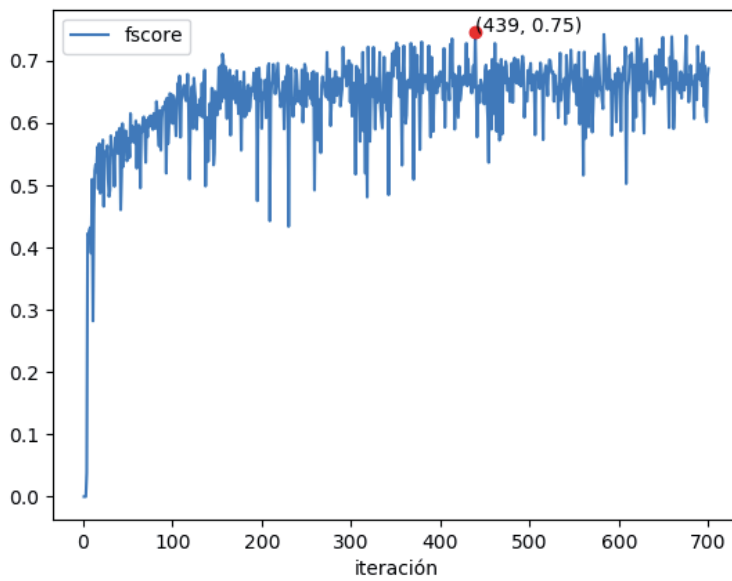


Ilustración 22. Fscore de la red SimpleNet con aumentación de datos en primer entrenamiento

Tal como se aprecia en la ilustración 23, al entrenar nuevamente la red cuya inicialización es la misma red previamente entrenada, se logra un fscore de 0.788 (en el gráfico aparece 0.79 que se debe a que el método `string.format()` de python redondeo el valor). En la ilustración 24 se presenta el fscore versus iteración de un tercer entrenamiento, tomando como inicialización a la red previa. En la cual se consigue un fscore de 0.79 que es ligeramente superior al método del estado del arte.

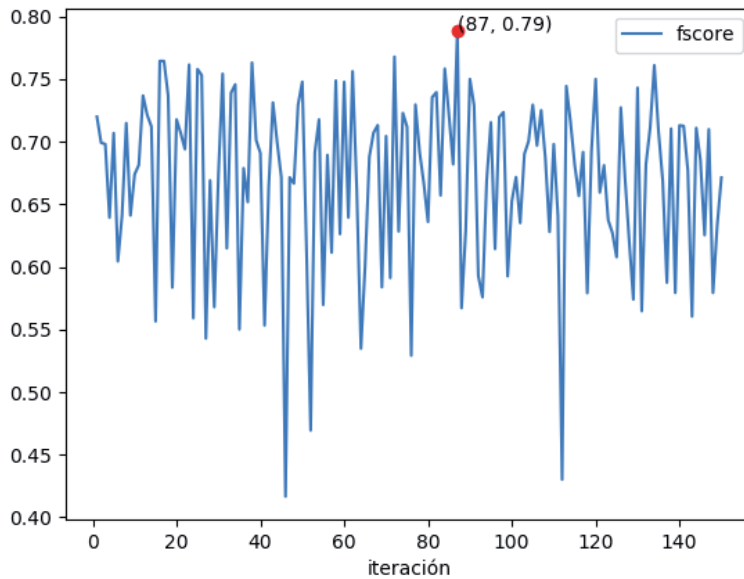


Ilustración 23. Fscore del segundo entrenamiento de la red SimpleNet

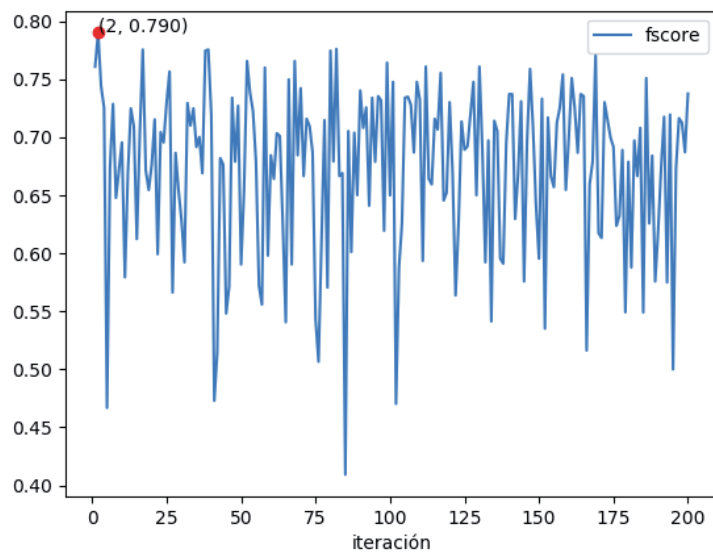


Ilustración 24. Fscore del tercer entrenamiento de la red SimpleNet

10.2. Mini SqueezeNet

Es una red inspirada en la red SqueezeNet, pero con menor cantidad de módulos fuego, por lo tanto de menor profundidad. Esta decisión se toma principalmente a limitaciones de memoria. Para la creación del módulo fuego se utiliza una función de conveniencia con dos parámetros: Sn y En. El parámetro Sn controla la cantidad de neuronas del filtro Squeeze 1x1 que se encarga de reducir la representación del mapa de activación. Por otro lado, el parámetros En controla la cantidad de neuronas que poseen los filtros expand de 3x3 y 1x1. Es requisito evidente que el parámetro Sn sea menor a En para la aplicación correcta de la estrategia de diseño de la arquitectura.

Tipo	Filtro	Parámetro Sn	Parámetro En	Salida	Cantidad de Parámetros
Conv	3x3 x 64			80 x 80 x 64	1.792
Max pool	3x3 displ=2			39 x 39 x 64	
Módulo fuego		64	128	39 x 39 x 256	86.336
Max pool	3x3 displ=2			19 x 19 x 256	
Módulo fuego		64	128	19 x 19 x 256	98.624
Max pool	3x3 displ=2			9 x 9 x 256	
Conv	1x1 x 32			9 x 9 x 32	8.224
FC				64	165.952
FC				1	65
Total					360.993

Tabla 4. Detalle de la arquitectura MiniSqueeze

Debido a lo que se determinó con la arquitectura SimpleNet, un entrenamiento demasiado extenso sin una aumentación de datos agresiva no aumentaba la precisión más allá de las iteraciones 50 – 60. Por lo que esta red se entrenó en primera instancia con 60 iteraciones.

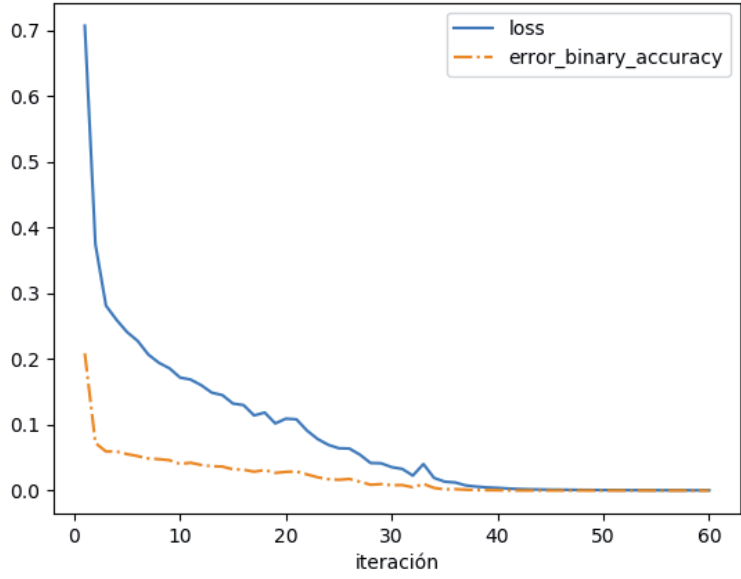


Ilustración 25. Métricas de entrenamiento de la red Mini SqueezeNet sin aumentación de datos

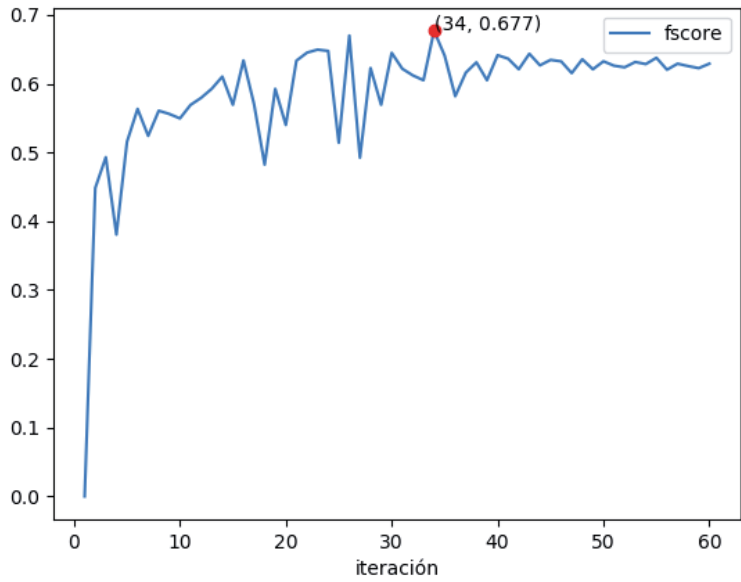


Ilustración 26. Fscore de la red Mini SqueezeNet sin aumentación de datos

De igual modo que sucedió con SimpleNet esta red no obtuvo buenos resultados sin usar aumentación de datos agresiva. Por lo que con el fin de compararla de manera más justa, ésta fue aplicada en un entrenamiento con una mayor cantidad de iteraciones. De igual manera se tomó el mejor modelo de este entrenamiento como inicialización para un nuevo entrenamiento, sin embargo esto último no conllevó a una mejora en el rendimiento. Es más, durante el segundo entrenamiento, se pierde un poco el rendimiento.

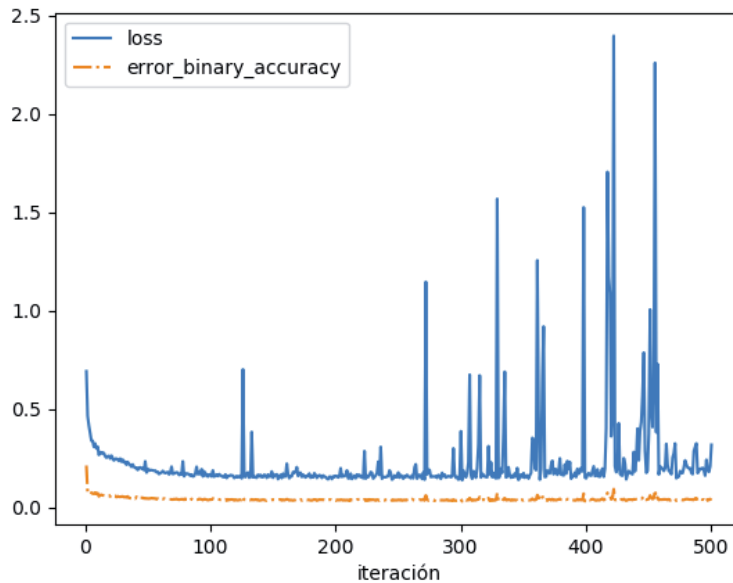


Ilustración 27. Métricas del primer entrenamiento de la red Mini SqueezeNet con aumentación de datos

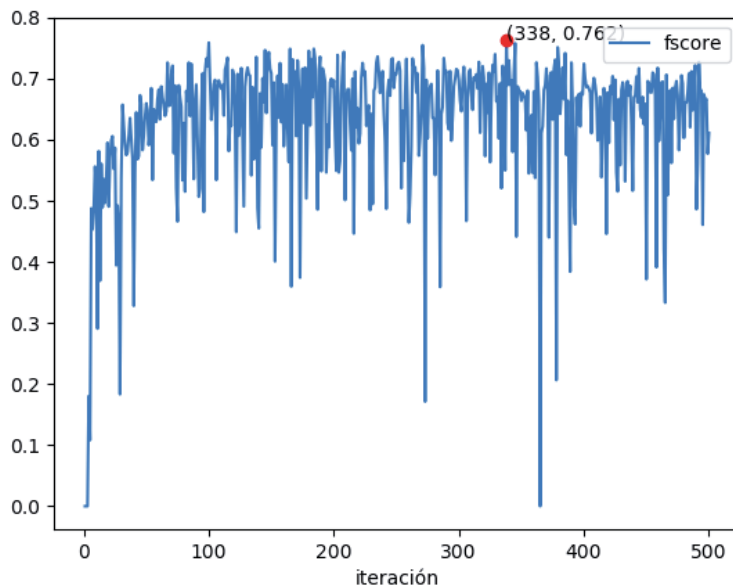


Ilustración 28. Fscore del primer entrenamiento de la red Mini SqueezeNet con aumentación de datos

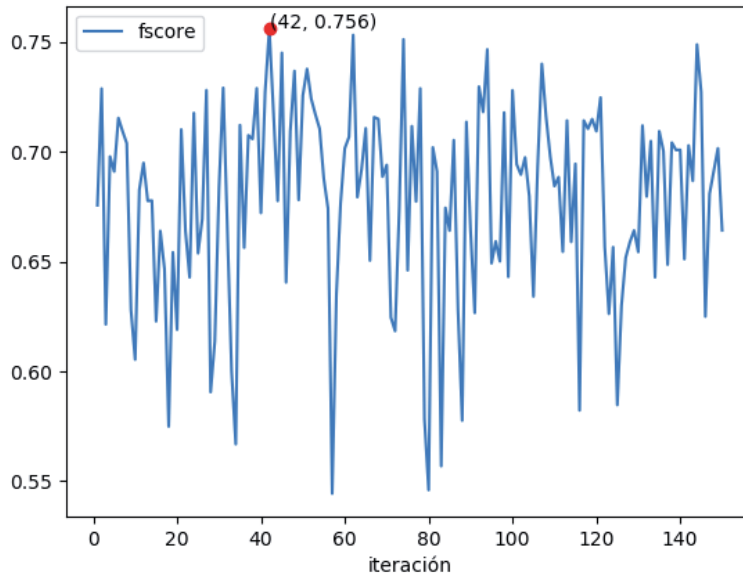


Ilustración 29. Fscore de la red mini SqueezeNet durante el segundo entrenamiento con aumentación de datos

Es destacable que, como a diferencia de la red SimpleNet, esta arquitectura no logra mejorar su rendimiento más allá del primer entrenamiento. Por lo tanto no se procedió a la ejecución de un tercer entrenamiento. Cabe recalcar que, gracias a que se va guardando constantemente el mejor modelo, no existe un riesgo de perder rendimiento.

10.3. ResNet – 11

Es una red inspirada en la idea de la ResNet, pero menos profunda. La arquitectura ResNet original tenía 50 capas. Las cuales estaban contenidas en 16 módulos Highway mas una capa de convolución al comienzo, y la completamente conectada al final. Debido a que esa arquitectura era demasiado costosa en término de memoria para los recursos disponibles, se utilizó una versión más pequeña de 11 capas, con solo dos módulos Highway.

Nuevamente se utiliza una función de conveniencia para construir el módulo característico de la red. Cuyos parámetros son similares a los utilizados en la red mini SqueezeNet. S_n es el parámetro que controla la cantidad de neuronas utilizadas para reducir la representación del mapa de activación, mientras que E_n controla la cantidad de neuronas sobre la cual expandir el mapa de activación. Debido a que en esta arquitectura no se usa una concatenación de las capas, sino una adición se deben cumplir con requisitos más estrictos en el uso de los parámetros. El parámetro S_n debe ser menor a E_n y E_n debe ser el mismo que la capa anterior, vale decir que el tamaño del mapa de activación de salida del módulo debe tener la misma profundidad que el módulo anterior.

Tipo	Filtro	Parámetro Sn	Parámetro En	Salida	Cantidad de parámetros
Conv	3x3 x 64			78 x 78 x 64	1.792
Max pool	3x3 despl=2			38 x 38 x 64	
Highway		32	64	38 x 38 x 64	46.208
Max pool	3x3 despl=2			18 x 18 x 64	
Highway		32	64	18 x 18 x 64	46.208
Max pool	3x3 despl=2			8 x 8 x 64	
Conv	1x1 x 32			8 x 8 x 32	2080
FC				128	262.272
FC				512	66.048
FC				1	513
Total					425.121

Tabla 5. Detalle de la arquitectura ResNet-11

Para la comparación con redes anteriores, se entrenó esta arquitectura sin aumentación de datos sobre un número pequeño de iteraciones. Luego se probó con una aumentación más agresiva sobre una gran cantidad de iteraciones. Esta red fue la que obtuvo mejor rendimiento sin aumentación de datos de las tres. Llegando a un fscore de 0.71 que es el resultado que obtuvo el segundo lugar en la competencia ICPR2012.

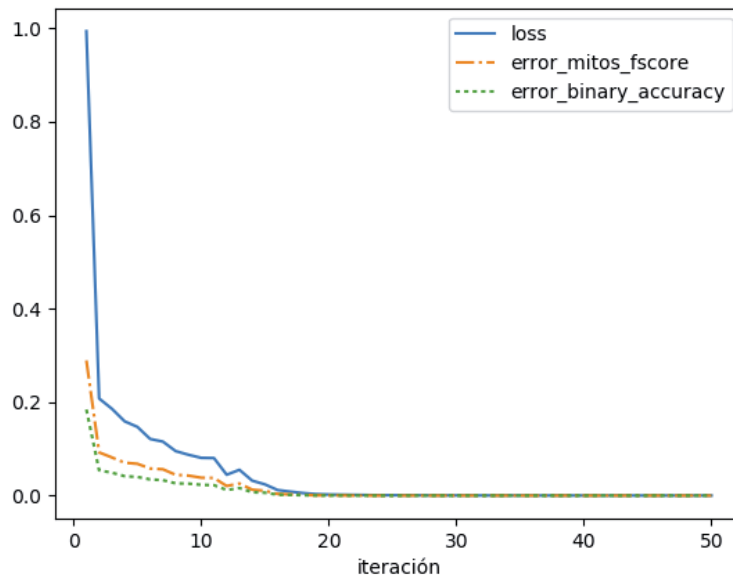


Ilustración 30. Métricas de entrenamiento de la red ResNet-11 sin aumentación de datos

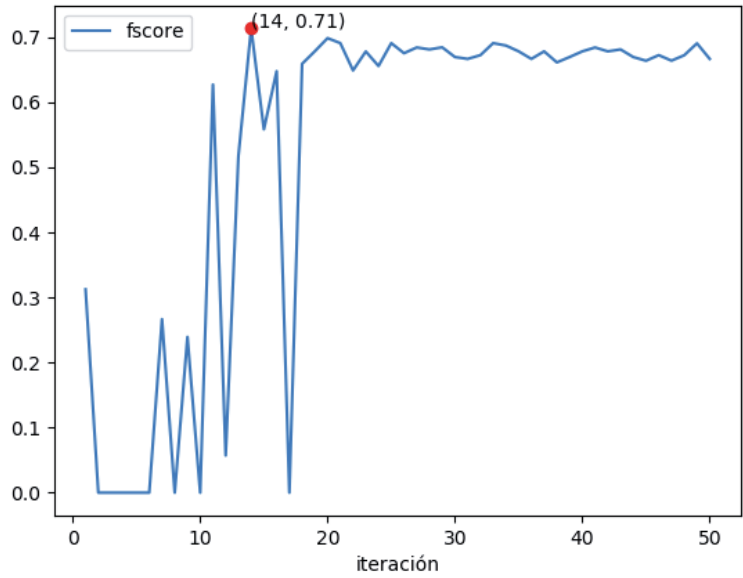


Ilustración 31. Fscore de la red ResNet sin aumentación de datos

Al igual que en arquitecturas anteriores, al aplicar la aumentación de datos se vio una mejora substancial en el rendimiento de la red. Además, nuevamente se entrenó a la red más de una vez, logrando una ligera mejora en la segunda ocasión. Durante el primer entrenamiento se logra alcanzar un fscore de 0.75, mientras que para la segunda se logra uno de 0.77, que es ligeramente inferior al logrado por la red SimpleNet y el método del estado del arte.

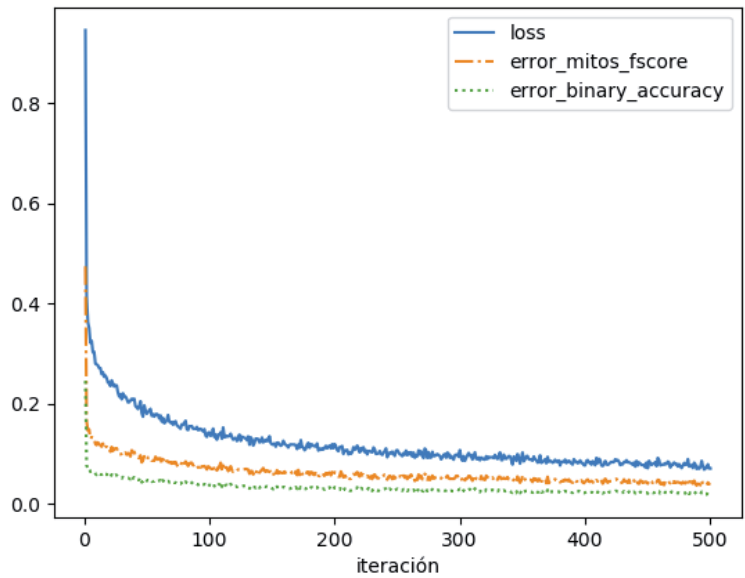


Ilustración 32. Métricas de entrenamiento de la red ResNet-11 con aumentación de datos

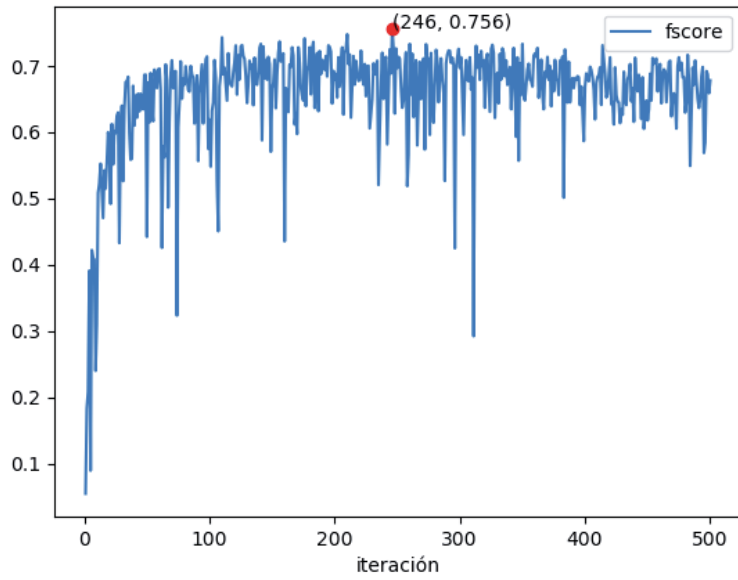


Ilustración 33. Fscore de la red ResNet-11 con aumentación de datos

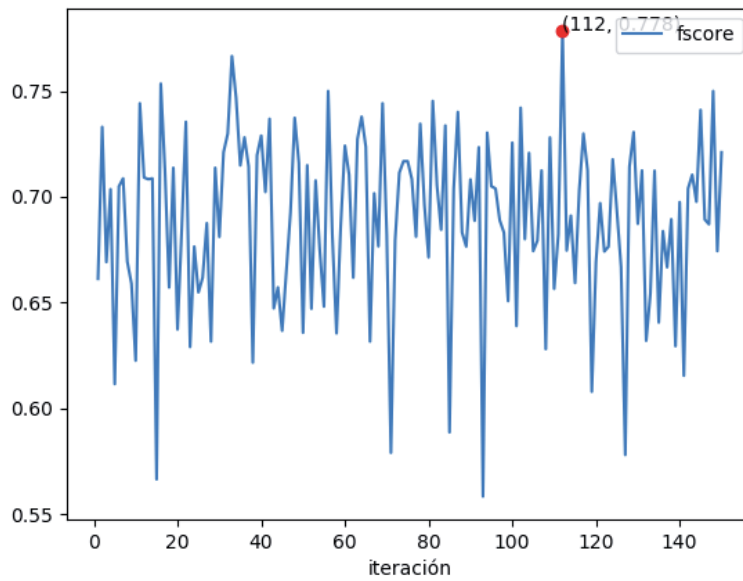


Ilustración 34. Fscore de la red ResNet-11 durante el segundo entrenamiento con aumentación de datos

11. Discusión de los resultados

La competencia ICPR determina que una célula es mitótica si esta se encuentra a una distancia de 8 μm . El escáner Asperio tiene una resolución de 0.2456 μm por pixel, lo que significa que cualquier detección a una distancia de 32 pixeles de una anotación será contada como un verdadero positivo. Sin embargo, en la etapa de selector de candidatos existen algunas células que no son mitóticas a menos de 32 pixeles de una anotación. Para evitar confusiones es que se hizo más estricto la decisión sobre una célula mitótica. Se redujo a 20 pixeles el criterio para contar a una célula como verdadero positivo.

La métrica principal que se utilizó para evaluar los modelos fue el fscore sobre la clase mitosis. El fscore está definido como:

$$\textit{Precisión} = \frac{TP}{TP+FP} \quad (8)$$

$$\textit{Sensibilidad} = \frac{TP}{TP+FN} \quad (9)$$

$$\textit{Fscore} = \frac{2*\textit{Precisión}*\textit{Sensibilidad}}{\textit{Precisión}+\textit{Sensibilidad}} \quad (10)$$

Donde TP (True Positives) son los verdaderos positivos, FP (False Positives) son los falsos positivos y FN (False Negatives) son los falsos negativos.

Durante la sección anterior se observó que se presentaron gráficas de las distintas arquitecturas que mostraban el fscore versus las iteraciones. Esto se logra probando el modelo al final de cada iteración y guardando su desempeño. Sin embargo, este desempeño es en base únicamente al clasificador. Para entender un poco más lo anterior es necesario repasar que el dataset de entrenamiento fue construido sobre la base del selector de candidatos. Los candidatos a más de 20 pixeles de una anotación son etiquetados como no-mitosis, mientras que los candidatos que se encuentran dentro de ese rango son ignorados. Las células mitóticas se extraen directamente en base a las anotaciones, esto es así tanto para el dataset de entrenamiento como para el de pruebas. Es por lo mismo, que una vez entrenado el modelo se procede a evaluarlo tomando únicamente la salida del selector de candidatos como entrada.

En las tablas 6 y 7 se comparan los modelos que se entrenaron. Se comparan en base a los verdaderos positivos (TP), falsos positivos (FP), falsos negativos (FN), verdaderos negativos (TN) y el fscore. El paréntesis indica la iteración del entrenamiento de la red, o sea si fue la primera, segunda o tercera red entrenada consecutivamente. Es necesario resaltar que el valor del fscore fue truncado en contraposición de los gráficos donde fue redondeado (debido al método `string.format()` de python). Otro dato a resaltar es que la arquitectura ResNet-11 sin aumentación de datos es la que resulto luego de las 50 iteraciones, puesto que en ese momento no se respaldó el mejor modelo.

En la tabla 6 se comparan los modelos únicamente en base a su capacidad de clasificación, mientras que en la tabla 7 se comparan en base a su clasificación tomando

como entrada únicamente la salida del selector de candidatos. Es por eso, que la tabla 7 es la más acertada para evaluar cuál es el modelo con mejor rendimiento.

Arquitectura	TP	FP	FN	TN	Fscore
SimpleNet sin aumentación de datos	63	27	38	2260	0.65
SimpleNet (1) con aumentación de datos	82	38	19	2249	0.74
SimpleNet(2) con aumentación de datos	80	22	21	2265	0.78
SimpleNet(3) con aumentación de datos	81	23	20	2264	0.79
SqueezeNet sin aumentación de datos	67	33	34	2254	0.66
SqueezeNet (1) con aumentación de datos	77	24	24	2263	0.76
SqueezeNet (2) con aumentación de datos	79	32	22	2255	0.74
ResNet-11 sin aumentación de datos	83	65	18	2222	0.66
ResNet-11(1) con aumentación de datos	85	42	16	2245	0.74
ResNet-11(2) con aumentación de datos	86	35	15	2252	0.77

Tabla 6. Comparación del rendimiento de las arquitecturas probadas

Arquitectura	TP	FP	FN	TN	Fscore
SimpleNet sin aumentación de datos	55	27	46	2259	0.62
SimpleNet (1) con aumentación de datos	84	47	17	2239	0.72
SimpleNet(2) con aumentación de datos	74	21	27	2265	0.75
SimpleNet(3) con aumentación de datos	72	22	29	2264	0.73
SqueezeNet sin aumentación de datos	54	30	47	2256	0.58
SqueezeNet (1) con aumentación de datos	72	24	29	2262	0.73
SqueezeNet (2) con aumentación de datos	74	32	27	2254	0.71
ResNet-11(1) con aumentación de datos	79	41	22	2245	0.71
ResNet-11(2) con aumentación de datos	82	35	19	2251	0.75

Tabla 7. Comparación del rendimiento de las arquitecturas probadas unidas al selector de candidatos

Tal como se aprecia en la tabla 7, el rendimiento decrece cuando el problema deja de ser clasificar células mitóticas, sino detectarlas. Esto se ve más marcado en los clasificadores con una aumentación de datos muy básicos (que para fines comparativos se les denominó sin aumentación de datos). Lo cual se podría explicar a que se acostumbraron a recibir células centradas, mientras que el selector de candidatos suele entregar imágenes ligeramente distinta, o a veces muy fuera del centro. Es el caso para células mitóticas en la fase de anafase o telofase, donde se encuentran dos manchas teñidas independientemente. Idealmente para que no hubiese una reducción en el rendimiento, se necesitaría que las células no detectadas por el selector de candidatos fueran catalogadas como no mitóticas por el clasificador, lo cual no siempre es el caso. Sumado además a que debería entregar como centro la misma posición que se encuentra anotada por el experto, lo cual es sumamente difícil.

12. Conclusión

En el presente informe se he mostrado la necesidad de encontrar modelos para detectar células mitóticas en cuanto es un paso fundamental para realizar la gradación de un tumor. Examen que está sujeto a la subjetividad del experto, lo que puede generar un diagnóstico erróneo. Razón por la cual se está comenzando a realizar investigación en el área de asistir en el diagnóstico sobre la gradación de un tumor.

Para solucionar la problemática se presentó un modelo basado en tres partes. En primer lugar una normalización de las imágenes para reducir las diferencias en la preparación de las muestras y el microscopio utilizado. Seguido de una selección de candidatos, cuya salida alimenta a un modelo más complejo que discriminará cuales células son realmente mitóticas. Y finalmente una red neuronal convolucional que determina la salida final del modelo.

Durante la preparación de la muestra, se utilizan dos tintes que tiñen zonas distintas del tejido. Por un lado la eosina tiñe el citoplasma de color rosa, mientras que la hematoxilina tiñe el núcleo celular de color purpura. Debido a que el citoplasma no entrega información relevante para determinar si una célula es mitótica o no, se utiliza solo la información visual entregada por la hematoxilina. Lo anterior permite que modelos relativamente poco profundos en comparación a los usados en competencias como ImageNet tengan un buen rendimiento en este problema.

Dentro de los modelos comparados, se determinó que los que obtuvieron mejor rendimiento fueron las redes SimpleNet en su segundo entrenamiento y ResNet-11 en su segundo entrenamiento, ambas con un fscore de 0.75. Sin embargo es posible encontrar diferencias en la sensibilidad, o sea en la cantidad de células mitóticas que lograron detectar. Por un lado SimpleNet obtuvo una sensibilidad de 0.73, mientras que ResNet-11 0.81. Lo cual indica que en el contexto de un problema médico, ResNet logró un mejor desempeño, puesto que es más peligroso entregar un diagnóstico errado diciendo que el tumor no es peligroso, que errar diciendo que lo es.

El método del estado del arte reportó un fscore de 0.78, el cual obtuvo mediante una precisión de 0.88 y una sensibilidad de 0.7. Si bien los modelos presentados en este informe no superan al método del estado del arte en fscore. La gran mayoría si lo superan en la sensibilidad. Sumado además que el tiempo necesario para ejecutar todo el modelo propuesto es mucho menor al que reportó Ceressan. El cual, señaló que tomó alrededor de 8 minutos por imagen. Mientras que tomando solo el tiempo del selector de candidatos y el clasificador, el modelo propuesto necesita menos de 10 segundos en todo el dataset de pruebas. De hecho la etapa más costosa en término de tiempo es la normalización, que toma alrededor de 6 segundos por imagen.

Es necesario resaltar también que la aumentación de datos fue mucho menor a la que utilizó Ciressan, la cual llegó a generar 66.000 muestras de la clase mitosis, que consistió en aproximadamente el 6% de todo el dataset de entrenamiento. Motivo por el cual se necesitó un día para entrenar la red. En contraste, la red ResNet-11, que fue la que

más parámetros contiene, tomó poco más de dos horas en entrenar las 500 iteraciones. Con una aumentación de datos menos agresiva, pero en tiempo de ejecución.

Dada la complejidad del problema, principalmente debido a la poca cantidad de datos disponibles, se considera que se logró construir un modelo que permite ser utilizado en aplicaciones reales en cuanto a su tiempo de ejecución. El peso completo de los candidatos extraídos en las 15 imágenes del dataset de prueba, ocupa alrededor de 25Mb, lo que eventualmente permitiría el uso de este tipo de modelos sobre una plataforma web, dependiendo de la calidad de la conexión y la cantidad de muestras que se requieran analizar al mismo tiempo.

En un futuro es necesario evaluar los modelos con una mayor cantidad de datos, puesto que 327 muestras de células mitóticas no son suficientes para determinar el universo completo del problema. En ese sentido, y dado lo limitado del dataset, es probable que no sea posible lograr un rendimiento alto sobre el mismo. En el aprendizaje de maquina los datos son lo más importante, y si los datos no logran ser una muestra representativa del universo, es poco realista pensar que permitan obtener una conclusión no sesgada.

13. Referencias

- [1] Bavlle RM. Enigmatic Morpho Insight: Mitosis At A Glance. *J Oral Maxillofac Pathol* 2014;18, Suppl S1:2-5
- [2] What Is Cancer?. (2016). National Cancer Institute. Retrieved 16 October 2016, from <https://www.cancer.gov/about-cancer/understanding/what-is-cancer>
- [3] Diagnosis. (2016). National Cancer Institute. Retrieved 16 October 2016, from <https://www.cancer.gov/about-cancer/diagnosis-staging/diagnosis>
- [4] *Tumour grading - Canadian Cancer Society*. (2016). *www.cancer.ca*. Retrieved 16 October 2016, from <http://www.cancer.ca/en/cancer-information/diagnosis-and-treatment/staging-and-grading/tumour-grading/?region=on>
- [5] *NCCN Clinical Practice Guidelines in Oncology, Breast Cancer Version 2.2011*. (2011). *National Comprehensive Cancer Network*. Retrieved 16 October 2016, from http://www.lecba-rakoviny.cz/dokumenty/NCCN_Guidelines_breast_2011.pdf
- [6] Elston, C.W. and Ellis I.O. (1991) Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*, **19**, 403-10.
- [7] *How do you identify mitoses?*. (2012). *Pathologystudent.com*. Retrieved 18 October 2016, from <http://www.pathologystudent.com/?p=5689>
- [8] *Template with article element*. (2012). *Ludo17.free.fr*. Retrieved 18 October 2016, from http://ludo17.free.fr/mitos_2012/
- [9] [MICCAI 2013 Grand Challenge on Mitosis Detection](#), organised by M. Veta, M.A. Viergever, J.P.W. Pluim, N. Stathonikos, P. J. van Diest of University Medical Center Utrecht
- [10] Veta, M., van Diest, P., Willems, S., Wang, H., Madabhushi, A., & Cruz-Roa, A. et al. (2015). Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Medical Image Analysis*, *20*(1), 237-248. <http://dx.doi.org/10.1016/j.media.2014.11.010>
- [11] Macenko, M., Niethammer, M., Marron, J., Borland, D., Woosley, J., & Xiaojun Guan, et al. (2009). A method for normalizing histology slides for quantitative analysis. *2009 IEEE International Symposium On Biomedical Imaging: From Nano To Macro*. <http://dx.doi.org/10.1109/isbi.2009.5193250>
- [12] Vahadane, A., Peng, T., Sethi, A., Albarqouni, S., Wang, L., & Baust, M. et al. (2016). Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images. *IEEE Transactions On Medical Imaging*, *35*(8), 1962-1971. <http://dx.doi.org/10.1109/tmi.2016.2529665>

- [13] Walter, T., Held, M., Neumann, B., Hériché, J., Conrad, C., Pepperkok, R., & Ellenberg, J. (2010). Automatic identification and clustering of chromosome phenotypes in a genome wide RNAi screen by time-lapse imaging. *Journal Of Structural Biology*, 170(1), 1-9. <http://dx.doi.org/10.1016/j.jsb.2009.10.004>
- [14] Irshad, H. (2013). Automated mitosis detection in histopathology using morphological and multi-channel statistics features. *Journal Of Pathology Informatics*, 4(1), 10. <http://dx.doi.org/10.4103/2153-3539.112695>
- [15] Cireşan, D., Giusti, A., Gambardella, L., & Schmidhuber, J. (2013). Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks. *Medical Image Computing And Computer-Assisted Intervention – MICCAI 2013*, 411-418. http://dx.doi.org/10.1007/978-3-642-40763-5_51
- [16] L. Yann, B. Yoshua and H. Geoffrey, *Deep learning*, 1st ed. Nature, 2015, pp. 436-444.
- [17] Olga Russakovsky*, Jia Deng*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *IJCV*, 2015
- [18] "COCO – Common Objects in Context", *Mscoco.org*, 2017. [Online]. Available: <http://mscoco.org/>. [Accessed: 28- Apr- 2017].
- [19] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", arXiv preprint arXiv:1409.1556, 2014 – arxiv.org.
- [20] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going deeper with convolutions", *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [21] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision", *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [22] J. Springenberg, A. Dosovitskiy, T. Brox and M. Riedmiller, "Striving for Simplicity: The All Convolutional Net", *arXiv preprint arXiv:1412.6806.*, 2014.
- [23] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition", *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [24] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <0.5mb model size," arXiv:1602.07360, 2016
- [26] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "Enet: A deep neural network architecture for real-time semantic segmentation," arXiv preprint arXiv:1606.02147, 2016.
- [27] N. Thai-Nghe, Z. Gantner and L. Schmidt-Thieme, "Cost-sensitive learning methods for imbalanced data", *The 2010 International Joint Conference on Neural Networks (IJCNN)*, 2010.