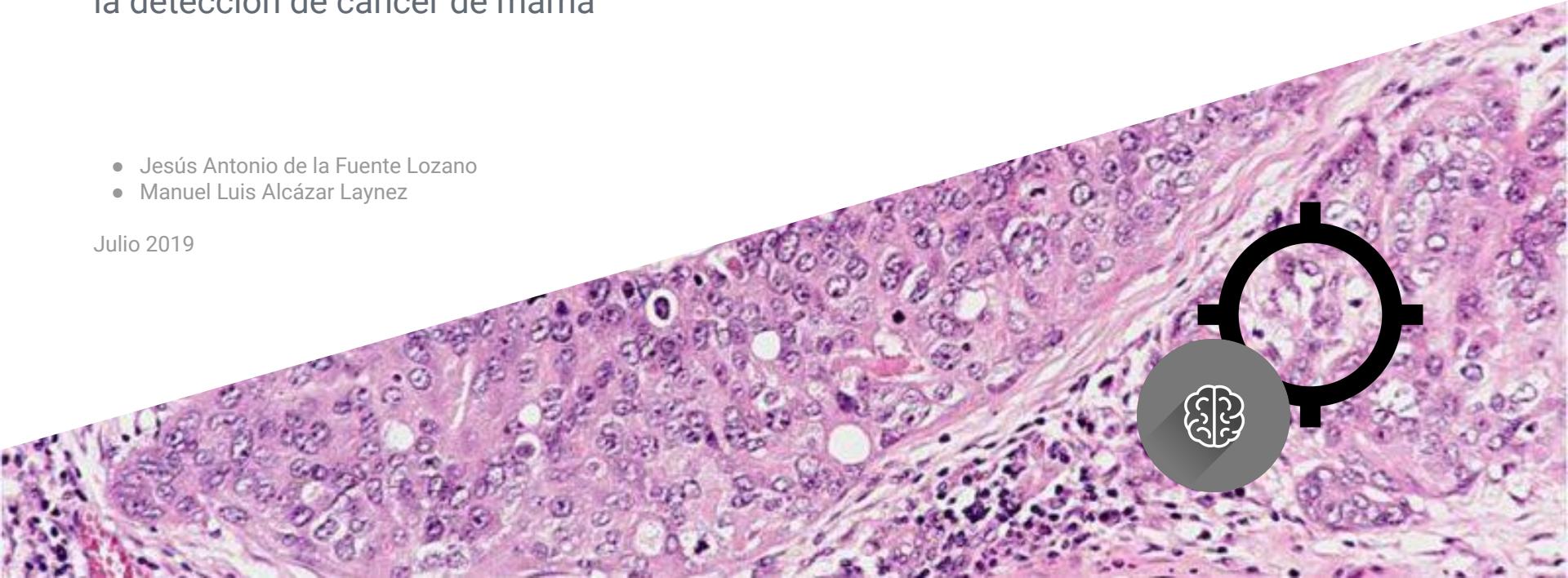


# Cancer.Bot

Clasificación de imágenes histológicas para  
la detección de cáncer de mama

- Jesús Antonio de la Fuente Lozano
- Manuel Luis Alcázar Laynez

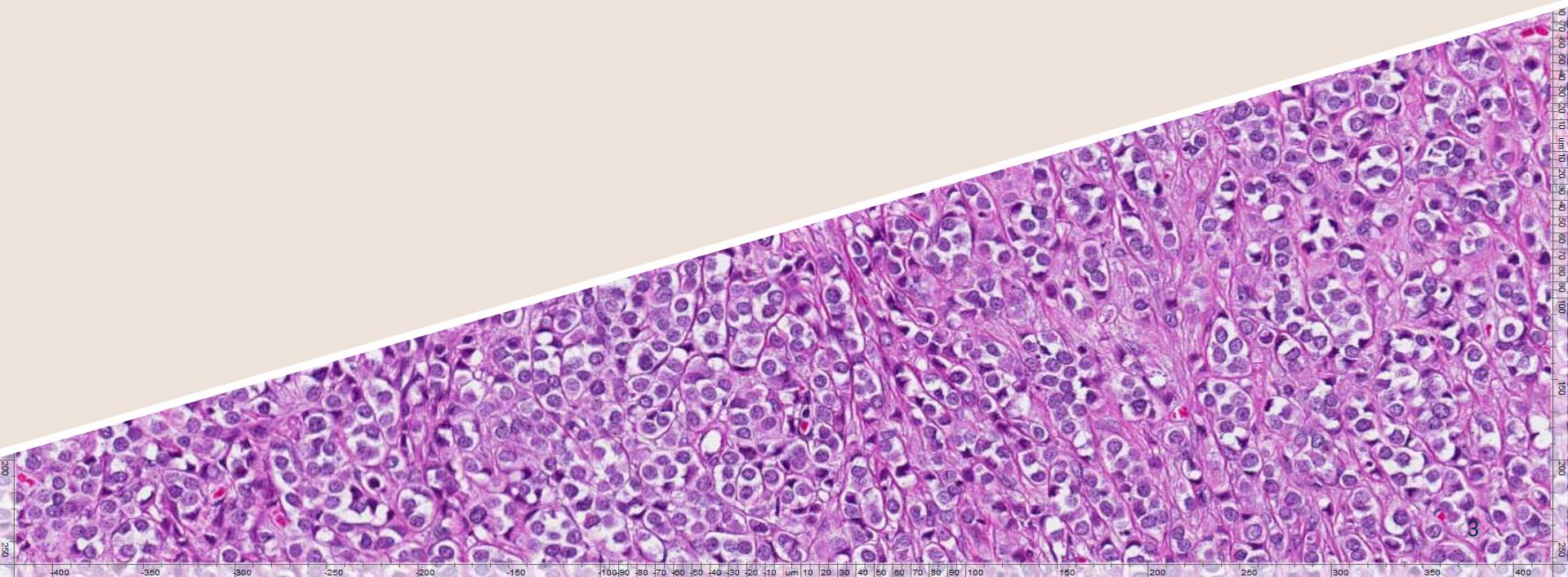
Julio 2019



# Índice

- Cáncer de mama
- Histología
  - Histología mamaria
- Cancer.Bot
  - Datos disponibles
  - Arquitectura
  - Resultados

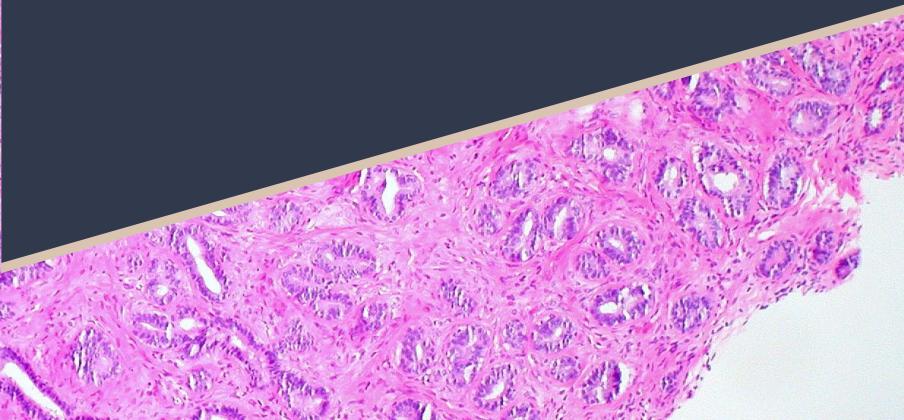
# Cáncer de mama



# Cáncer de mama

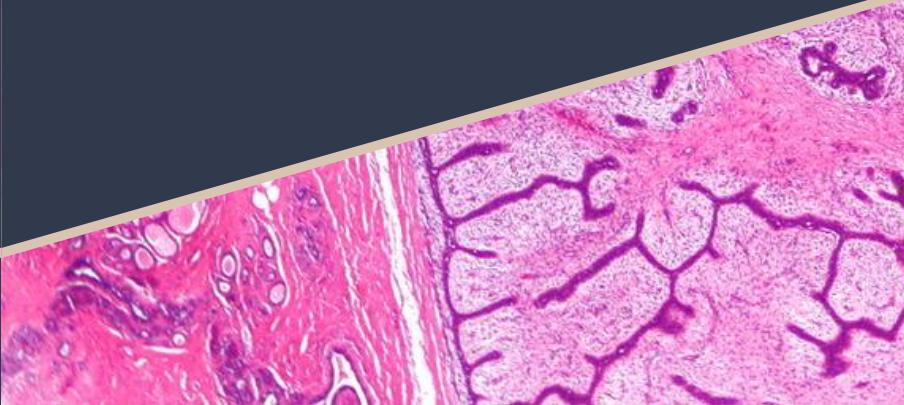
- Adenocarcinoma: tumor originado en las células y estructuras de glándulas
  - En nuestro caso, **glándulas mamarias**
- Es el tipo de tumor más frecuente en mujeres
- CDI: Carcinoma Ductal Invasivo
  - Se originan en las células que revisten los conductos galactóforos
  - Es el tipo más frecuente, representando el 80% de los cánceres infiltrantes de mama.
    - Nos centraremos en la detección de éste

# Histología



- **Técnica histológica:** Conjunto de operaciones a que se somete un tejido biológico a fin de que sea posible su estudio por medio del microscopio.
- Las células, por sí mismas, no poseen coloración. Por tanto, para poder observar la morfología tisular deben "teñirse"
- Técnica muy empleada en laboratorios y hospitales. Permite hacer diagnósticos de distintas enfermedades, como para saber si un tumor es maligno o benigno, etc.

# Histología mamaria



- Células deformes, alargadas
- Nuestro objetivo: Crear **modelo predictivo** que aprenda a **distinguir y localizar** zonas con alta probabilidad de **carcinomas**.

# Cancer.Bot

- Datos disponibles
- Arquitectura
- Resultados

# Datos disponibles

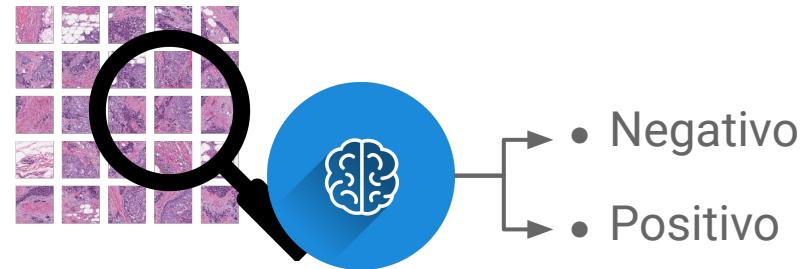
## Dataset:

- Breast Histopathology Images - Paul Mooney (Kaggle)
- **162** imágenes repartidas en pequeñas muestras de 50x50 px.
  - Total: **277.524** muestras
  - Positivas: **78.786** muestras
  - Negativas: **198.738** muestras
- Dataset claramente desbalanceado



# Datos disponibles

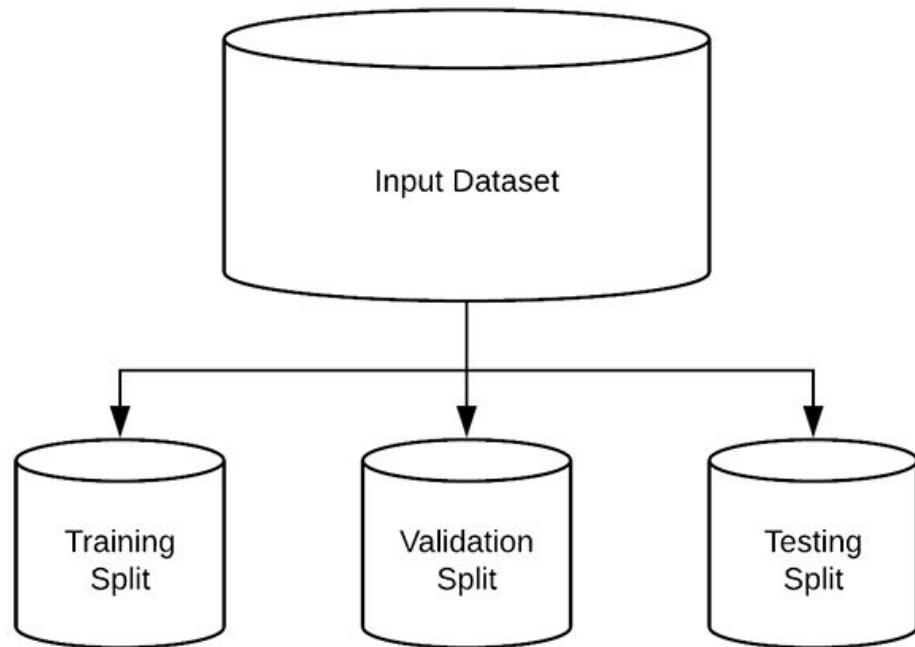
- Técnica
  - **Modelo clasificador** de Visión Artificial a través de dataset desbalanceado
  - **Keras + OpenCV + Python**
  - **PylImageSearch** - Adrian Rosebrock
    - Número masivo de tutoriales, referencias y guías sobre modelos predictivos y visión artificial
      - ¡Visita más que recomendada!
  - **Atom** - A hackable text editor for the 21st Century



# Datos disponibles

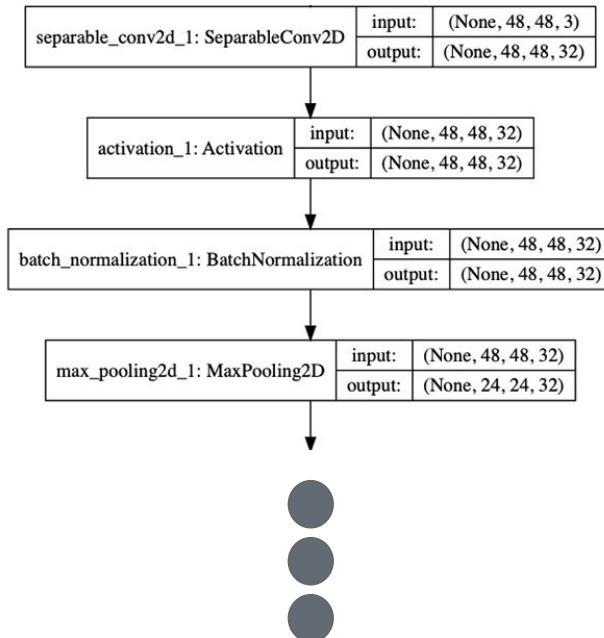
## División de los datos:

- Training split: **80%**
- Validation split: **10%**
- Testing split: **10%**



# Arquitectura

## CancerNet



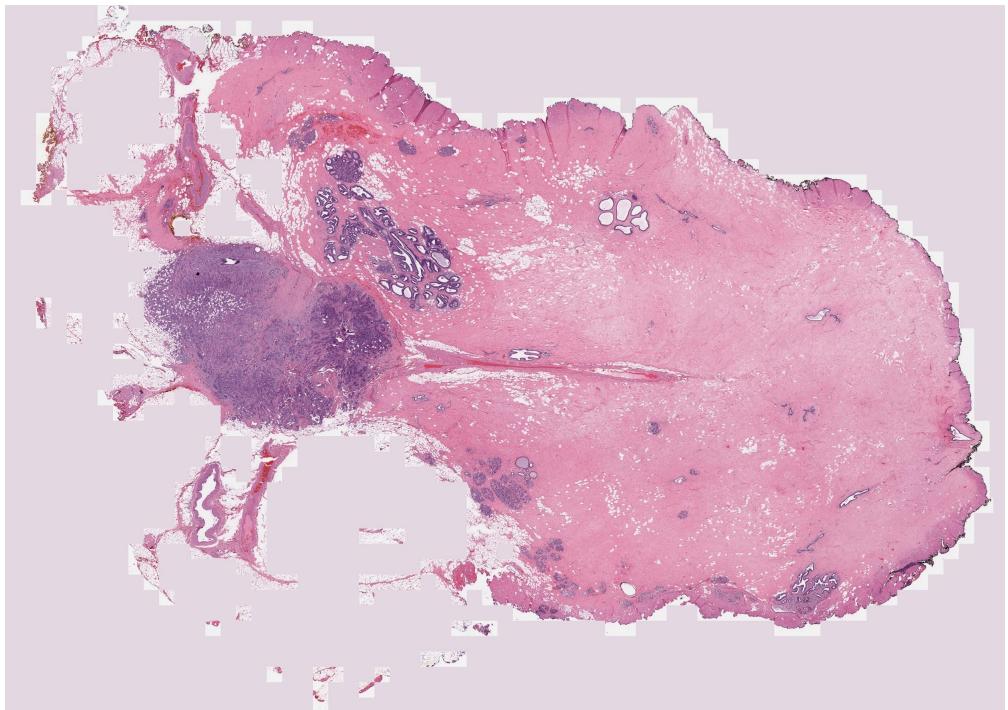
### CancerNet:

- Creado por **Adrian Rosebrock** (pyImageSearch)
- 29 Capas ocultas
  - Similares a VGGNet pero con capas **convolucionales separables en profundidad** (depthwise separable convolution)
  - Más eficiente en memoria

# Resultados

# Resultados

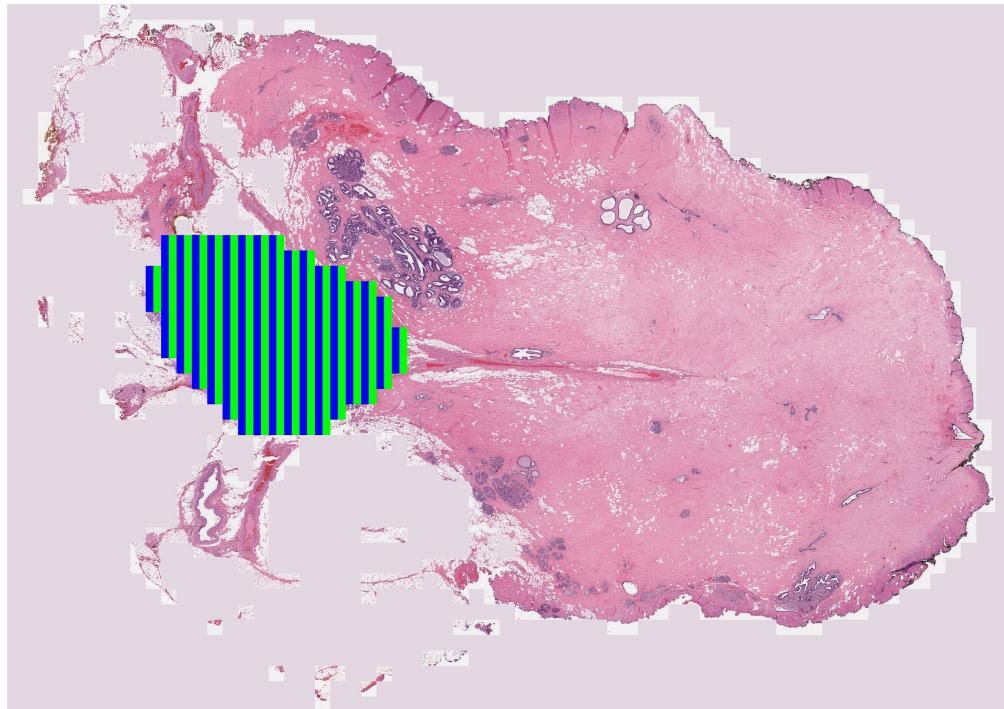
Imagen original



# Resultados

Imagen original con  
zona marcada por  
especialista

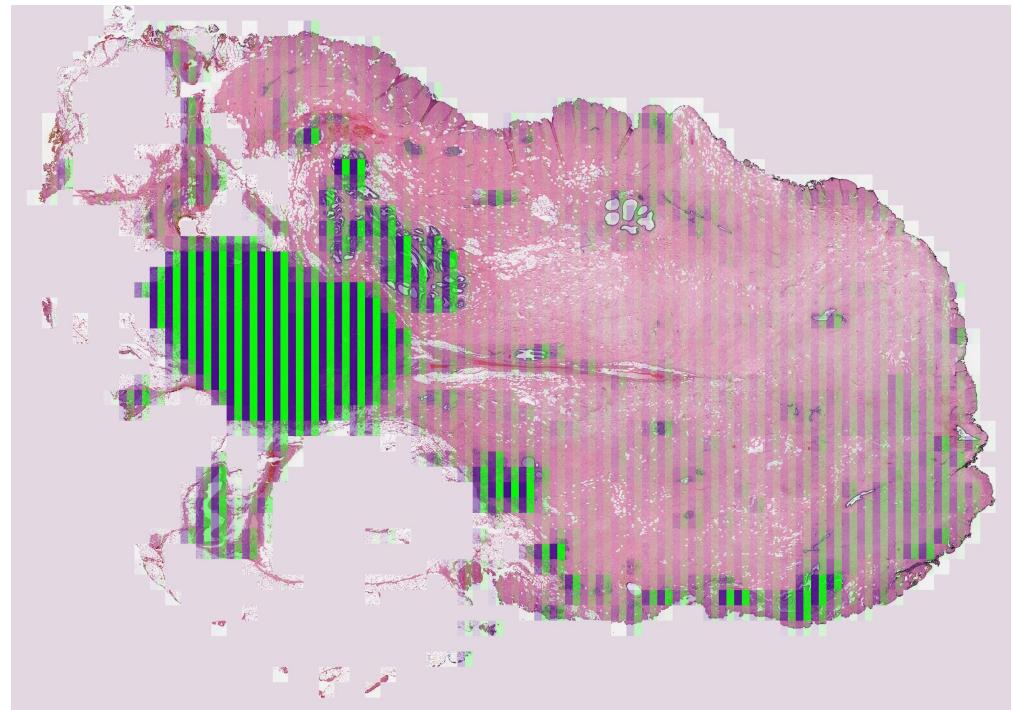
- Zonas “lilas” pueden ser:
  - Capilares o tejido vascularizado
  - Carcinoma
- Reto: Distinguir las



# Resultados

Zonas de interés para la red CancerNet

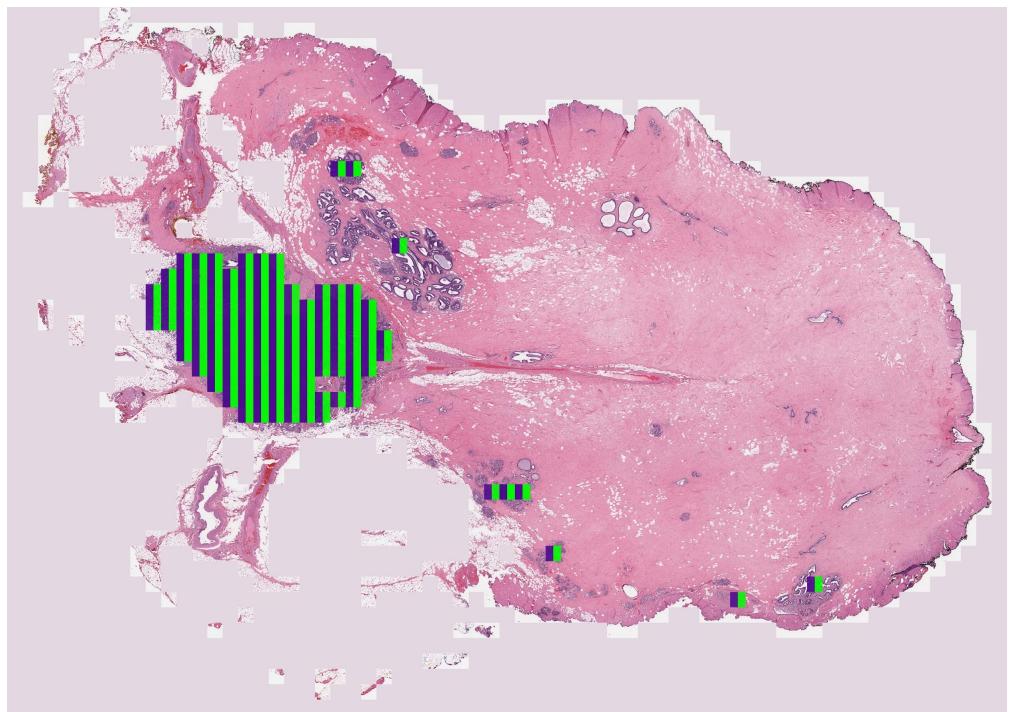
- Zonas “lilas” más marcadas que tejido sano



# Resultados

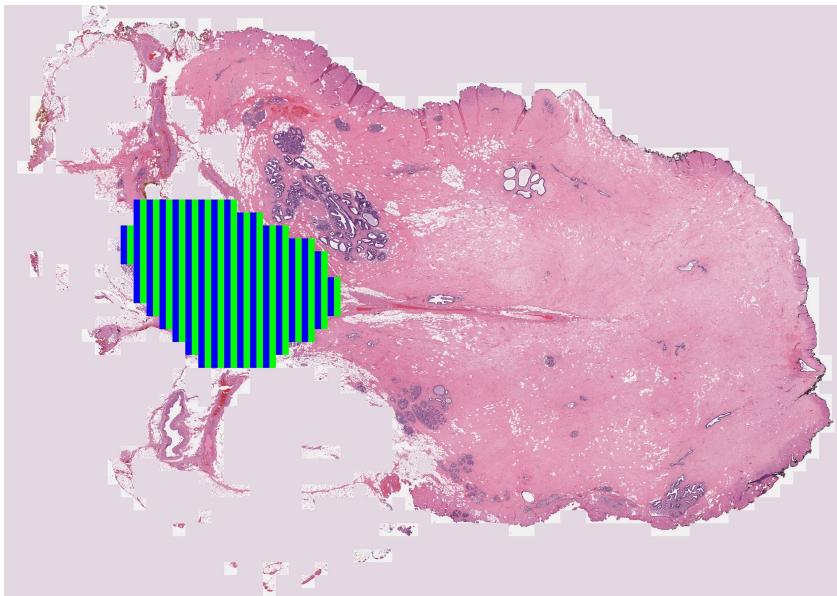
## Zonas con alta probabilidad

- Se marcan las zonas para la cual la predicción ha superado el umbral de 85%

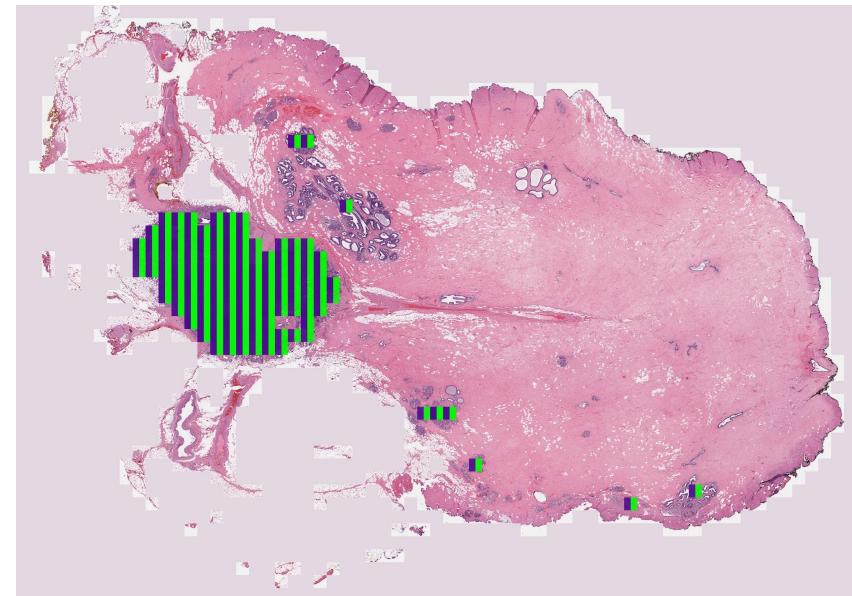


# Resultados

Localización real del carcinoma



Localización tras predicción (generación nº25)

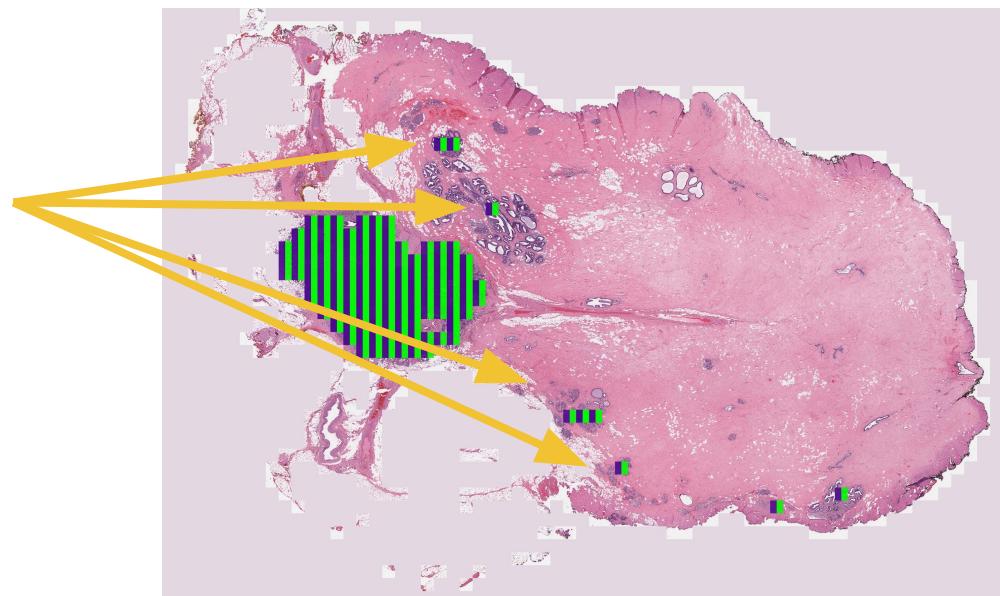


# Resultados

Zonas con falsos positivos pueden ayudar a reducir error humano

Doble-checking del falso positivo por el especialista → Podría descubrir zonas positivas erróneamente descartadas

Localización tras predicción (generación nº25)



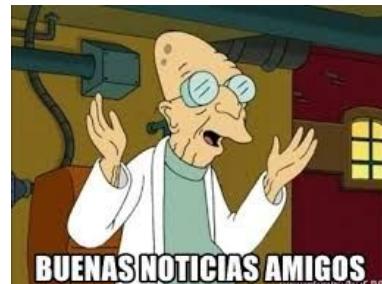
# Resultados

```
[INFO] evaluating network...
      precision    recall    f1-score   support
          0         0.90      0.90      0.90     39847
          1         0.75      0.75      0.75     15658

accuracy                           0.86     55505
macro avg       0.83      0.82      0.83     55505
weighted avg    0.86      0.86      0.86     55505

[[ 36028  3819]
 [ 3980 11678]]
acc: 0.8595
sensitivity: 0.9042
specificity: 0.7458
```

	Predicted: YES	Predicted: NO	
Actual: YES	TP 36.028	FN 3.819	39.847
Actual: NO	FP 3.980	TN 11.678	15.658
	40.008	15497	55505



- Aunque los datos no sean balanceados, FN y FP aproximadamente iguales y tenemos alta sensibilidad

# Posibles trabajos futuros

- Balancear datos
- Entrenar más generaciones
- **Tratar de disminuir los falsos negativos**
  - Es el caso más peligroso
  - Aumenta el número de falsos positivos

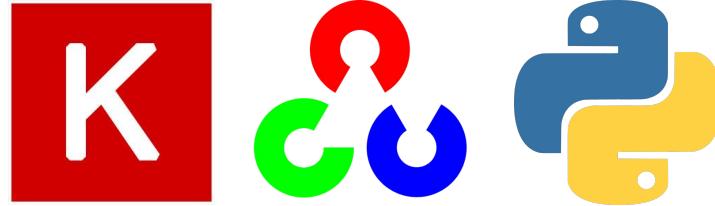


¡Muchas gracias!

# Bibliografía



- [Cáncer de Mama: Todo lo que Necesitas Saber | AECC](#)
- [Invasive ductal carcinoma, NOS - Pathology Outlines](#)
- [Técnica histológica - Wikipedia, la enciclopedia libre](#)
- [Breast Histopathology Images](#)
- [PylImageSearch - Be awesome at OpenCV, Python, deep learning, and computer vision](#)
- [Breast cancer classification with Keras and Deep Learning](#)
- [Keras – Save and Load Your Deep Learning Models](#)
- [OpenCV-Python Tutorials](#)



# Contacto

- Jesús Antonio de la Fuente Lozano
  - Twitter: @coronelyisus
  - Correo: jesusfuentes242@gmail.com
- Manuel Luis Alcázar Laynez
  - Twitter: @mrmanlu
  - Github: MrManlu
  - LinkedIn:  
<https://www.linkedin.com/in/mrmanlu/>
  - Correo: manlu.tomcat@gmail.com

