

SOFTWARE INTELIGENTE

CNN y modelos detección de patrones en imágenes



David Calderon

Scopus[®]



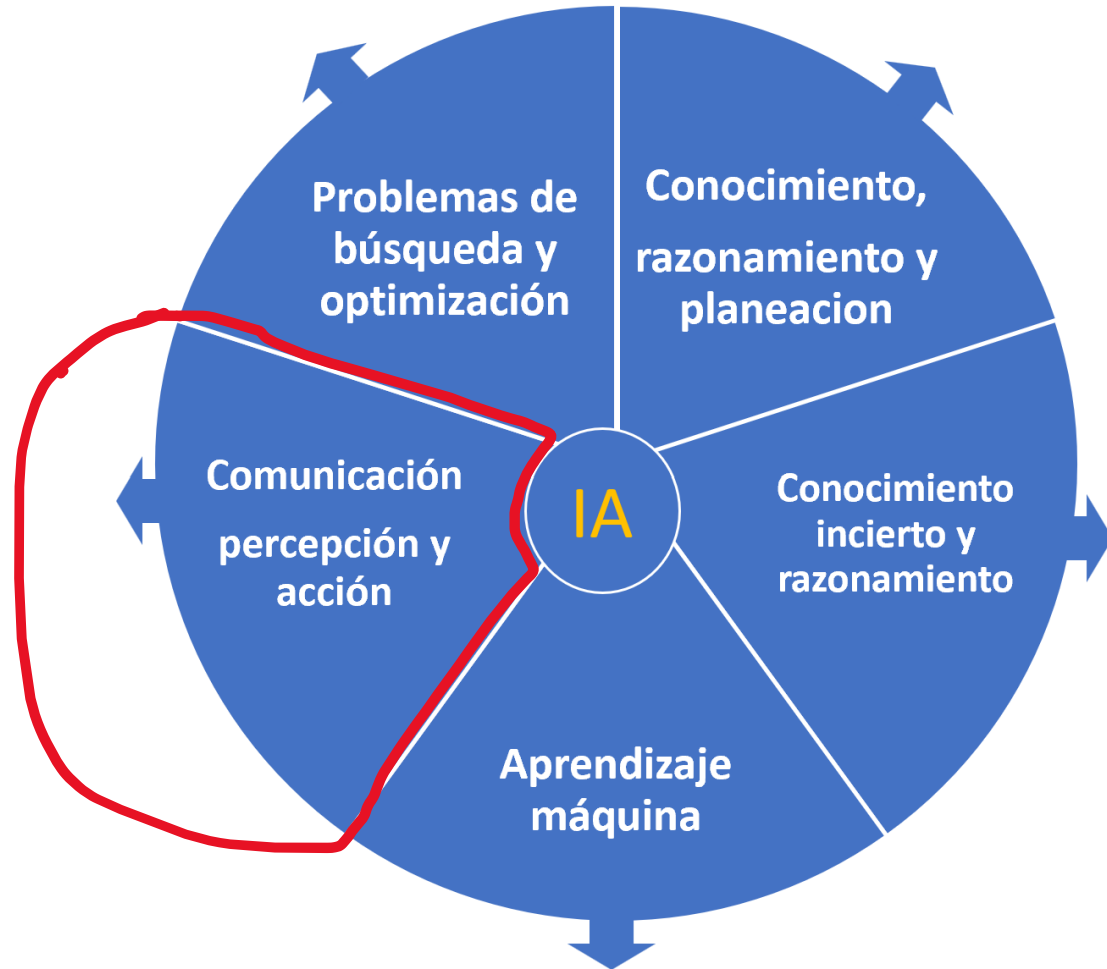
ORCID



ResearchGate



Subcampos de IA



Computer visión

Métodos para adquirir, procesar, analizar y comprender las imágenes del mundo real con el fin de producir información.

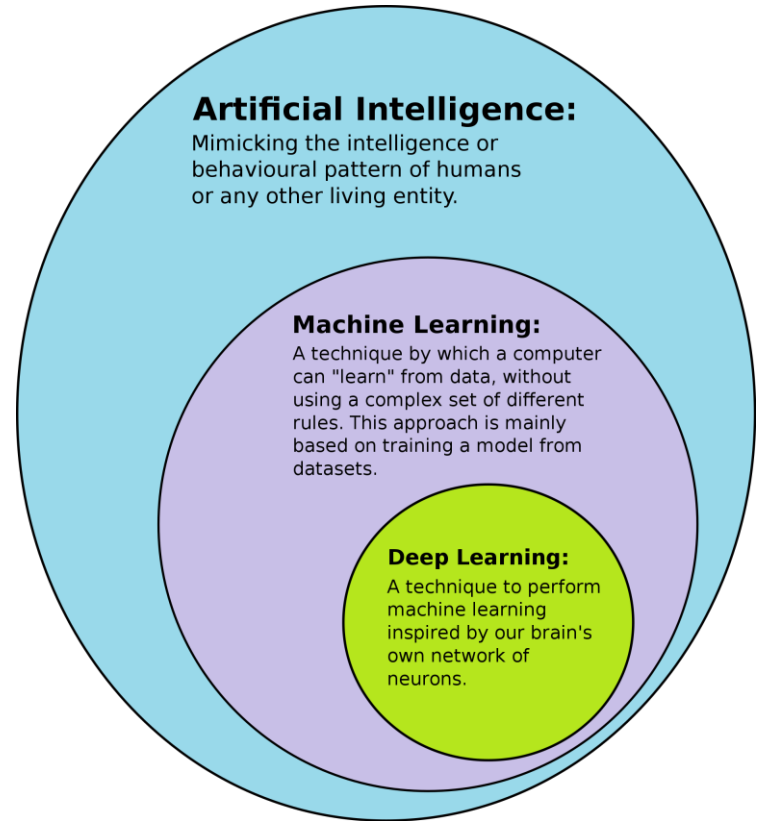


Deep learning

Deep learning es una clase de algoritmos de ML que utiliza múltiples capas para extraer progresivamente características de nivel superior.

En aprendizaje superficial el algoritmo aprende directamente de las características de los datos.

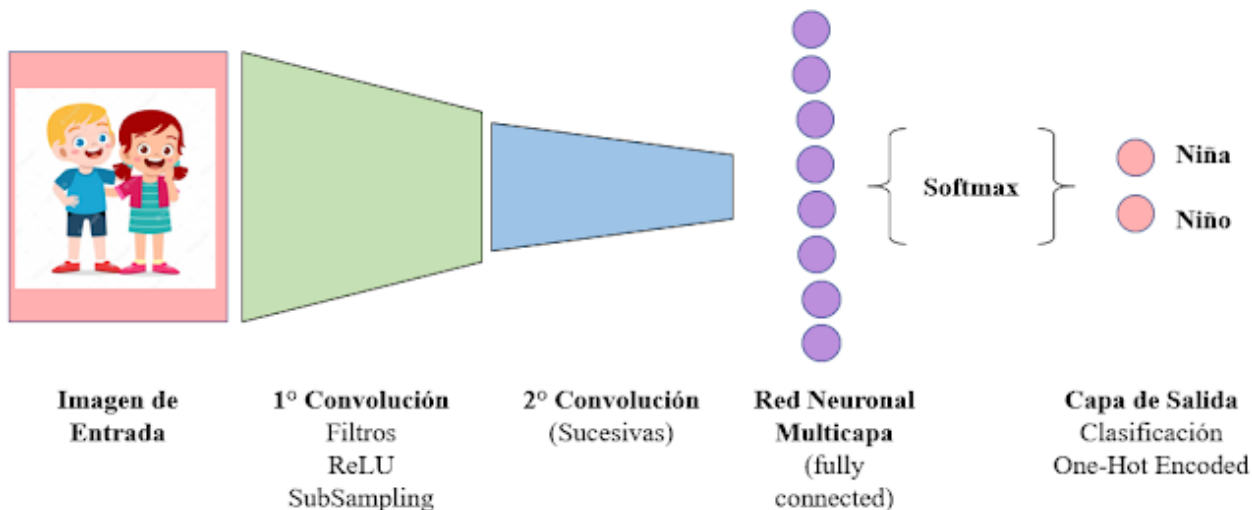
Mientras que en Deep Learning, el algoritmo aprende de características abstractas provenientes de los datos.



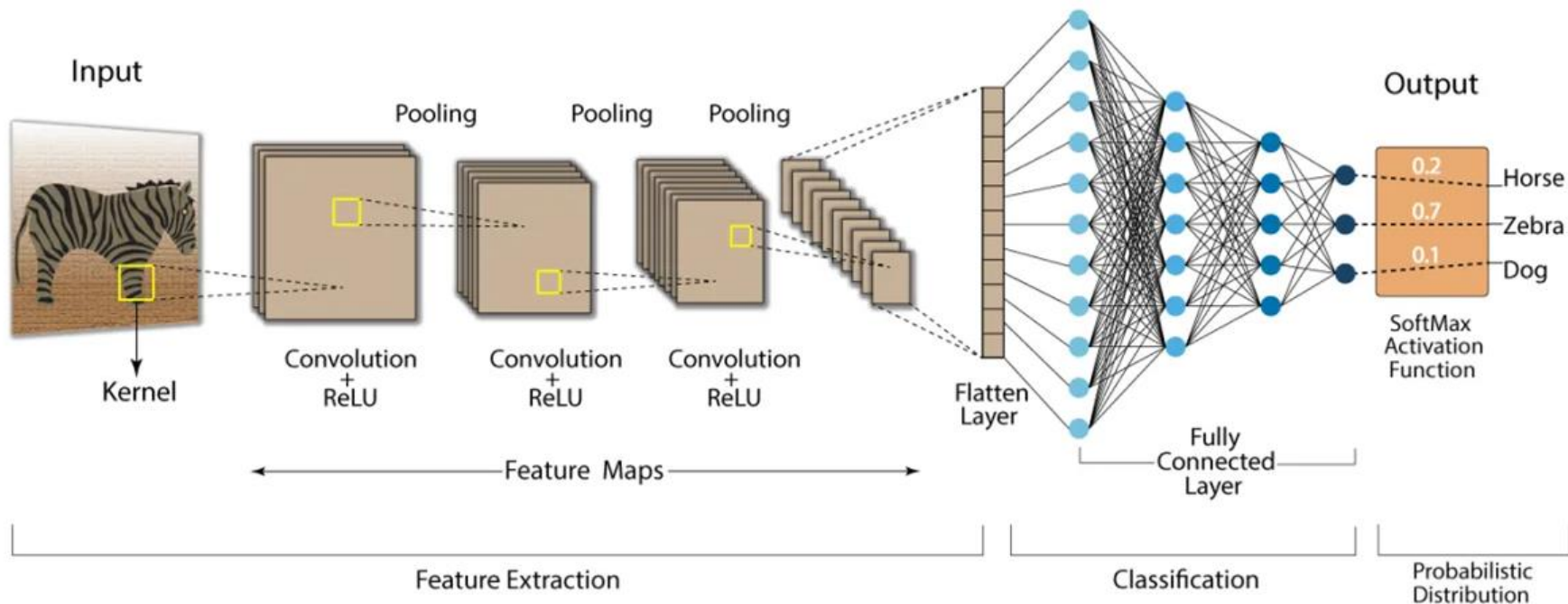
Convolutional neural network

Es un tipo de red neuronal artificial que trabaja similar a las neuronas en la corteza visual primaria de un cerebro biológico.

Arquitectura de Redes Neuronales Convolucionales – CNN



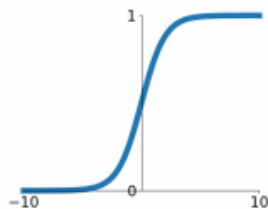
Arquitecturas Convolucionales para clasificación



Funciones de Activación

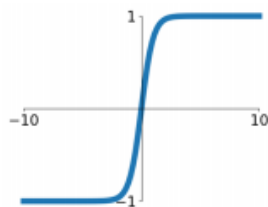
Sigmoid

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$



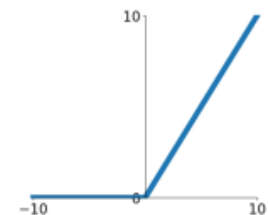
tanh

$$\tanh(x) = \frac{e^{2x} - 1}{e^{2x} + 1}$$



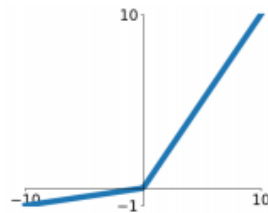
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.2x, x)$$

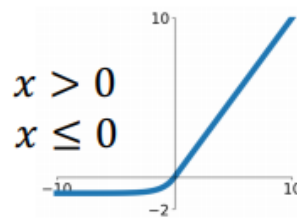


Softplus

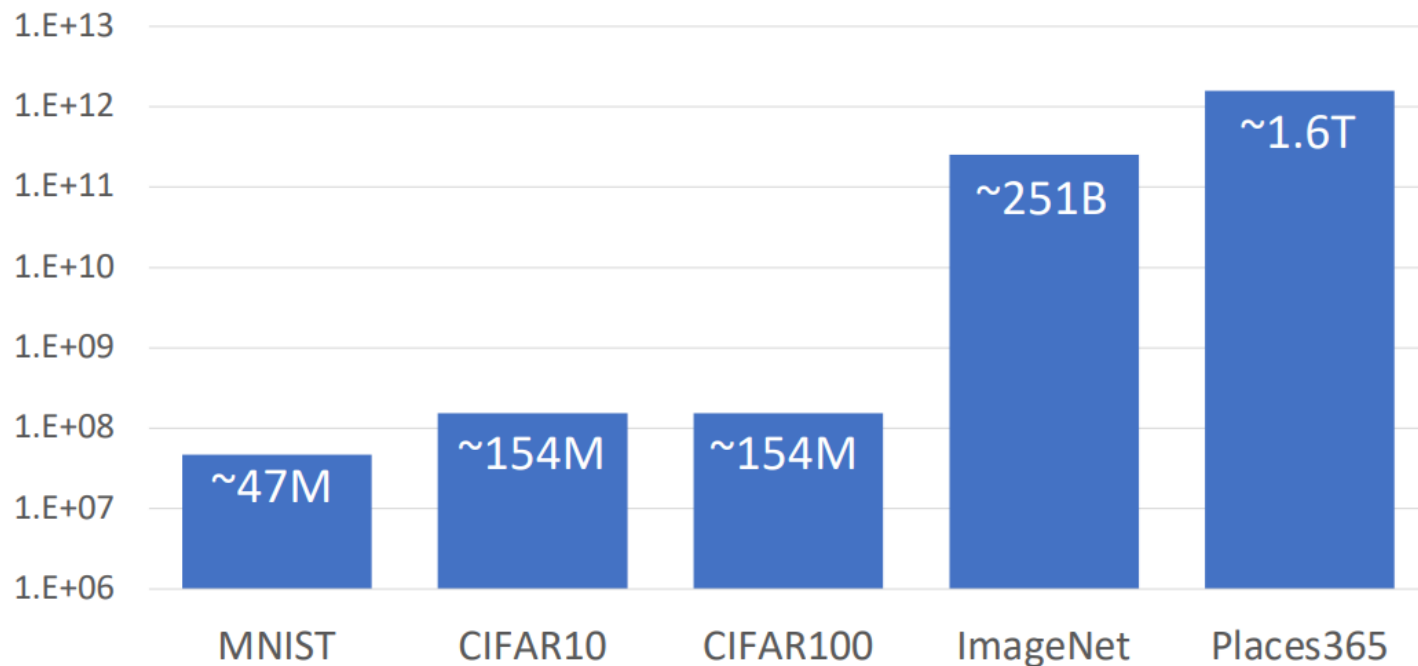
$$\log(1 + \exp(x))$$

ELU

$$f(x) = \begin{cases} x, & x > 0 \\ \alpha(\exp(x) - 1), & x \leq 0 \end{cases}$$



Datasets de clasificación: número de pixels de entrenamiento



Datasets de clasificación: MNIST



10 classes: Digits 0 to 9

28x28 grayscale images

50k training images

10k test images

Datasets de clasificación: CIFAR-10

airplane



automobile



bird



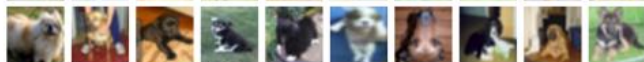
cat



deer



dog



frog



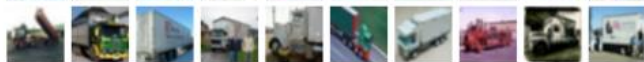
horse



ship



truck



10 classes

50k training images (5k per class)

10k testing images (1k per class)

32x32 RGB images

We will use this dataset for
homework assignments

Datasets de clasificación: CIFAR-100



100 classes

50k training images (500 per class)

10k testing images (100 per class)

32x32 RGB images

20 superclasses with 5 classes each:

Aquatic mammals: beaver, dolphin, otter, seal, whale

Trees: Maple, oak, palm, pine, willow

Datasets de clasificación: ImageNet

1000 classes

~1.3M training images (~1.3K per class)

50K validation images (50 per class)

100K test images (100 per class)

Performance metric: **Top 5 accuracy**

Algorithm predicts 5 labels for each image; one of them needs to be right



Deng et al, "ImageNet: A Large-Scale Hierarchical Image Database", CVPR 2009

Russakovsky et al, "ImageNet Large Scale Visual Recognition Challenge", IJCV 2015

Datasets de clasificación: ImageNet

1000 classes

~1.3M training images (~1.3K per class)

50K validation images (50 per class)

100K test images (100 per class)
test labels are secret!



flamingo

cock

ruffed grouse

quail

partridge

...



Egyptian cat

Persian cat

Siamese cat

tabby

lynx

...



dalmatian

keeshond

miniature schnauzer

standard schnauzer

giant schnauzer

Images have variable size, but often
resized to **256x256** for training

There is also a 22k category version of
ImageNet, but less commonly used

Deng et al, "ImageNet: A Large-Scale Hierarchical Image Database", CVPR 2009

Russakovsky et al, "ImageNet Large Scale Visual Recognition Challenge", IJCV 2015

Datasets de clasificación: MIT Places



365 classes of different scene types

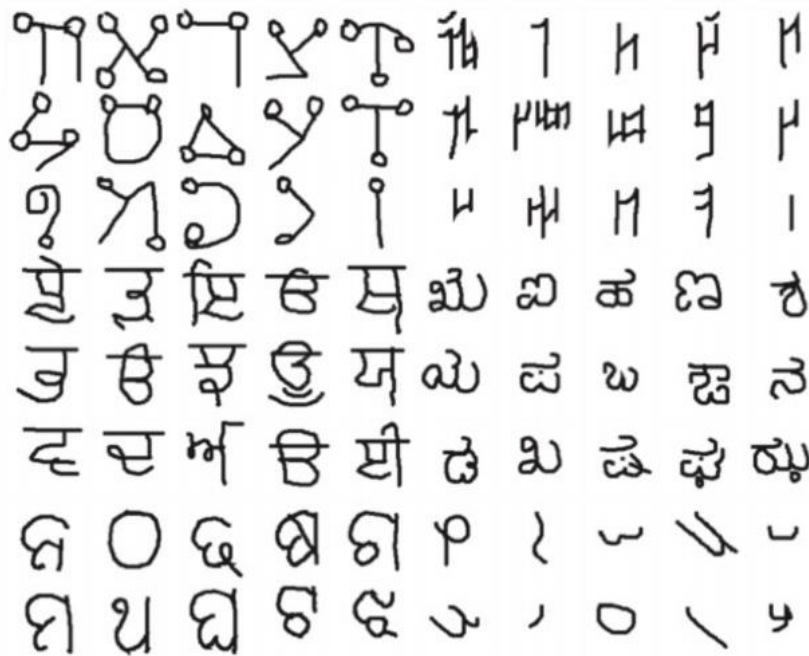
~8M training images

18.25K val images (50 per class)

328.5K test images (900 per class)

Images have variable size, often
resize to **256x256** for training

Datasets de clasificación: Omniglot



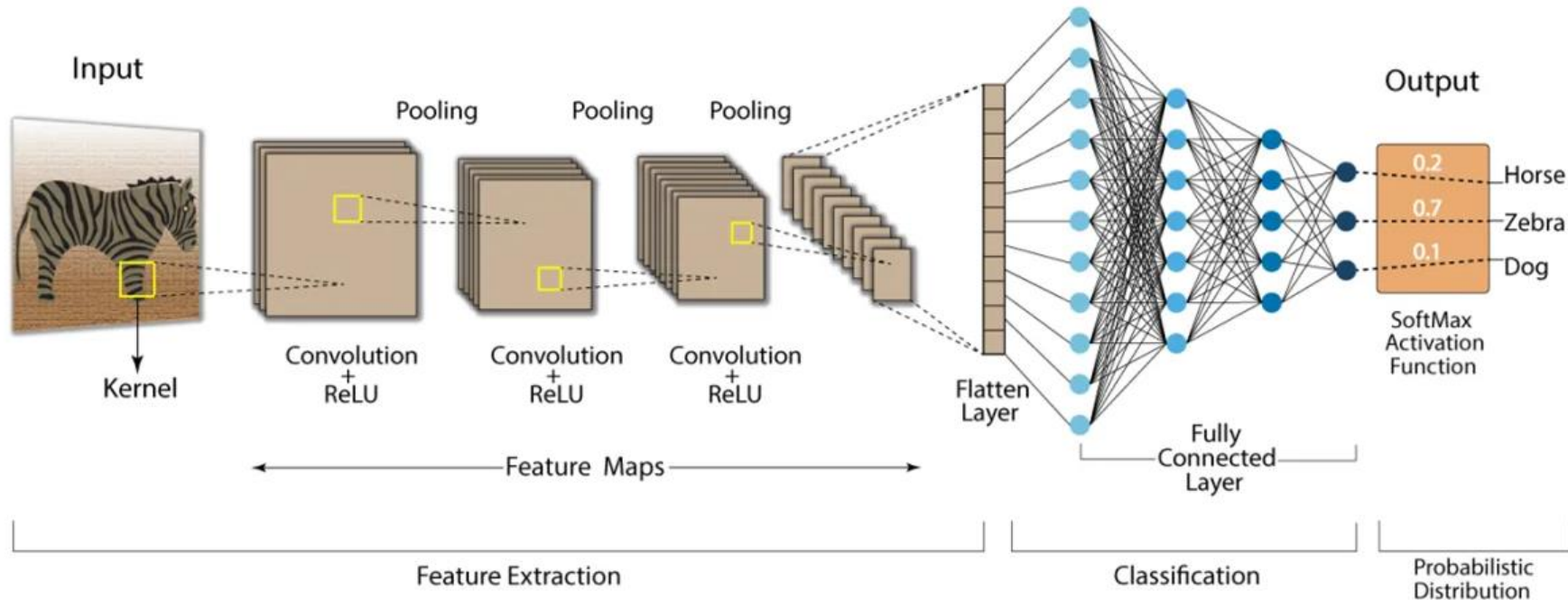
1623 categories: characters
from 50 different alphabets

20 images per category

Meant to test **few shot learning**

Lake et al, "Human-level concept learning through probabilistic program induction", Science, 2015

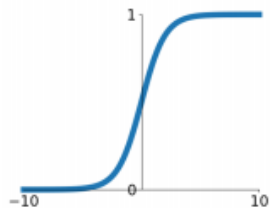
Arquitecturas Convolucionales para clasificación



Funciones de Activación

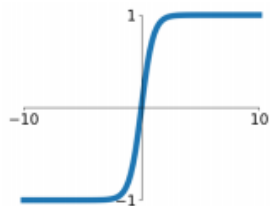
Sigmoid

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$



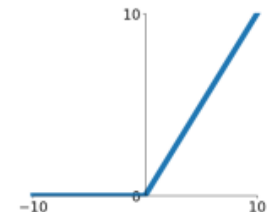
tanh

$$\tanh(x) = \frac{e^{2x} - 1}{e^{2x} + 1}$$



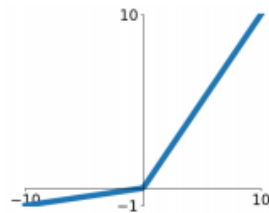
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.2x, x)$$

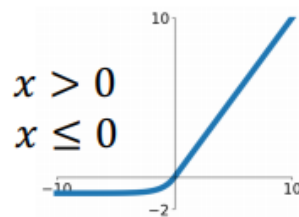


Softplus

$$\log(1 + \exp(x))$$

ELU

$$f(x) = \begin{cases} x, & x > 0 \\ \alpha(\exp(x) - 1), & x \leq 0 \end{cases}$$



Pixels



Una imagen...

| | | | | | |
|--|-----|-----|-----|-----|--|
| | | | | | |
| | | 0.6 | 0.6 | | |
| | 0.6 | | | 0.6 | |
| | 0.6 | 0.6 | 0.6 | 0.6 | |
| | 0.6 | | | 0.6 | |
| | | | | | |

...es una matriz de pixeles.

El valor de los pixeles va de 0 a 255 pero se normaliza para la red neuronal de 0 a 1

<https://www.juanbarrios.com/redes-neurales-convolucionales/>

Pixels y neuronas

La red toma como entrada los píxeles de una imagen. Si tenemos una imagen con apenas 28×28 píxeles de alto y ancho, esto equivale a utilizar 784 neuronas. Y eso es si sólo tenemos 1 color (escala de grises). Si tuviéramos una imagen a color, necesitaríamos 3 canales RGB (red, green, blue) y entonces usaríamos $28 \times 28 \times 3 = 2352$ neuronas. Estas neuronas constituyen nuestra capa de entrada.

<https://www.juanbarrios.com/redes-neurales-convolucionales/>

Pre-procesamiento



| | | | | | | |
|--|-----|-----|-----|-----|--|--|
| | | | | | | |
| | | 0.2 | 0.2 | | | |
| | 0.2 | | | 0.2 | | |
| | 0.2 | 0.2 | 0.2 | 0.2 | | |
| | 0.2 | | | 0.2 | | |
| | | | | | | |

| | | | | | | |
|--|-----|-----|-----|-----|--|--|
| | | | | | | |
| | | 0.4 | 0.4 | | | |
| | 0.4 | | | 0.4 | | |
| | 0.4 | 0.4 | 0.4 | 0.4 | | |
| | 0.4 | | | 0.4 | | |
| | | | | | | |

| | | | | | | |
|--|-----|-----|-----|-----|--|--|
| | | | | | | |
| | | 0.2 | 0.2 | | | |
| | 0.2 | | | 0.2 | | |
| | 0.2 | 0.2 | 0.2 | 0.2 | | |
| | 0.2 | | | 0.2 | | |
| | | | | | | |

Si la imagen es a color, estará compuesta de tres canales: rojo, verde, azul.

Antes de alimentar la red, se convierte los valores entre 0 y 1, se divide todos entre 255 .

Convoluciones

| | | | | | |
|--|-----|-----|-----|-----|--|
| | | | | | |
| | | 0.6 | 0.6 | | |
| | 0.6 | | | 0.6 | |
| | 0.6 | 0.6 | 0.6 | 0.6 | |
| | 0.6 | | | 0.6 | |
| | | | | | |

Imagen de
entrada

| | | |
|---|---|----|
| 1 | 0 | -1 |
| 2 | 0 | -2 |
| 1 | 0 | -1 |

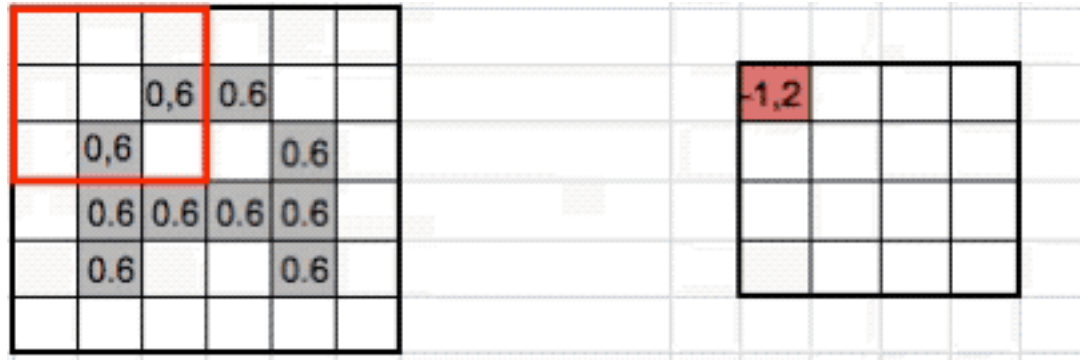
kernel

Toma **grupos de pixeles cercanos** de la imagen de entrada operando (producto escalar) contra una pequeña matriz que se llama kernel. Ese kernel supongamos que tiene un tamaño de de 3×3 pixels y con ese tamaño logra «visualizar» todas las neuronas de entrada (de izquierda-derecha, de arriba-abajo) y así logra generar una nueva matriz de salida, que en definitiva será nuestra nueva capa de neuronas ocultas.

NOTA: si la imagen fuera a color, el kernel realmente sería de $3 \times 3 \times 3$ es decir: un filtro con 3 kernels de 3×3 ; luego esos 3 filtros se suman (y se le suma una unidad bias) y conformarán 1 salida (cómo si fuera 1 solo canal).

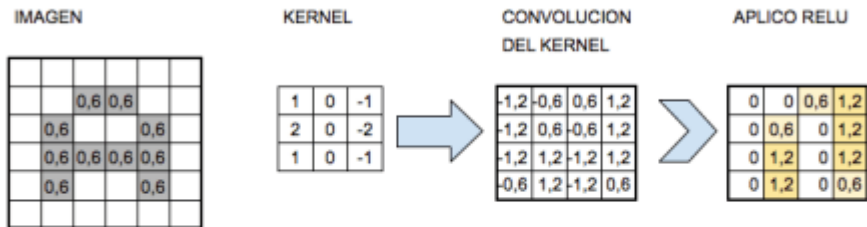
Filtro: conjunto de kernels

Se aplican muchos kernel (les llama filtros). Si en primer convolución tiene 32 filtros, se obtiene 32 matrices de salida (feature mapping), cada una de $28 \times 28 \times 1$ dan un total del 25.088 neuronas para nuestra PRIMER CAPA OCULTA (perteneciente a una imagen de 28 pixeles).



Filtro: conjunto de kernels

A medida que vamos desplazando el kernel y vamos obteniendo una «nueva imagen» filtrada por el kernel. En esta primer convolución y siguiendo con el ejemplo anterior, es como si obtuviéramos 32 «imágenes filtradas nuevas». Estas imágenes nuevas lo que están «dibujando» son ciertas características de la imagen original. Esto ayudará en el futuro a poder distinguir un objeto de otro (por ej. gato ó un perro).



FINALMENTE
OBTENGO UN MAPA
DE DETECCIÓN DE
CARACTERÍSTICAS

Función de activación

La función de activación más utilizada para este tipo de redes neuronales es la llamada ReLu por Rectifier Linear Unit y consiste en una función $f(x)=\max(0,x)$.

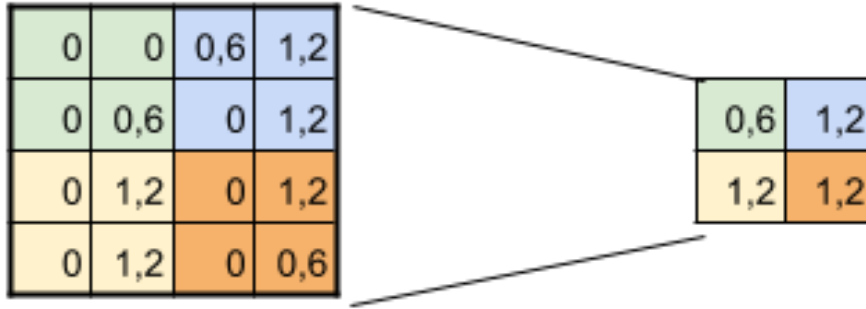
Muestreo (subsampling)

Ahora viene un paso en el que tomamos una muestra de las neuronas mas representativas antes de hacer una nueva convolución. En el ejemplo se tiene una imagen blanco y negro de 28x28pixels tenemos una primer capa de entrada de 784 neuronas y luego de la primer convolución obtenemos una capa oculta de 25.088 neuronas -que realmente son nuestros 32 mapas de características de 28x28 $((28 \times 28) \times 32)$.

Si hiciéramos una nueva convolución a partir de esta capa, el número de neuronas de la próxima capa requeriría un poder computacional importante. Por ello y para reducir el tamaño de la próxima capa de neuronas se hace un muestreo preservando las características más importantes que detectó cada filtro. Hay diversos tipos de muestreo (subsampling) , El más usado es: Max-Pooling

<https://www.juanbarrios.com/redes-neurales-convolucionales/>

Muestreo (subsampling)



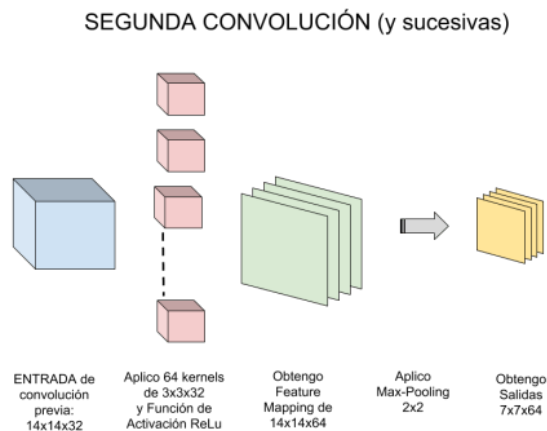
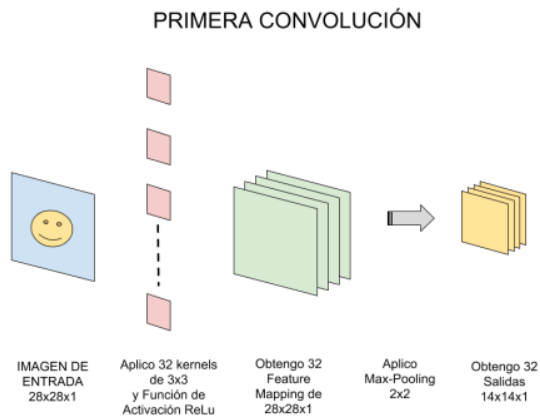
SUBSAMPLING:

Aplico Max-Pooling de 2x2
y reduzco mi salida a la mitad

Muestreo (subsampling)

El paso siguiente sería utilizar la técnica de «Max-pooling» con un tamaño de 2×2 . Esto quiere decir que recorreremos cada una de las 32 imágenes de características obtenidas anteriormente de 28×28 px de izquierda-derecha, arriba-abajo PERO en vez de tomar de a 1 pixel, tomaremos de « 2×2 » (2 de alto por 2 de ancho = 4 píxeles) e iremos preservando el valor «más alto» de entre esos 4 píxeles (por eso lo de «Max»). En este caso, usando 2×2 , la imagen resultante es reducida «a la mitad» y quedará de 14×14 píxeles. Luego de este proceso de submuestreo nos quedarán 32 imágenes de 14×14 , pasando de 25.088 neuronas a 6272, las cuales son bastantes menos y que -en teoría- deberían seguir almacenando la información más importante para detectar características deseadas.

Convoluciones

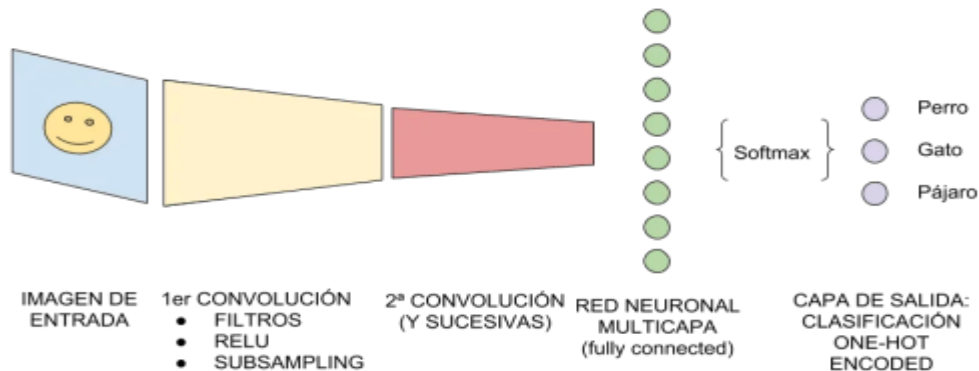


<https://www.juanbarrios.com/redes-neurales-convolucionales/>

Conectar con una red neuronal tradicional

Finalmente la última capa oculta a la que hicimos subsampling (tridimensional) 3x3x128 (alto, ancho, mapas) la cual se aplanan y se deja ser tridimensional, y pasa a ser una capa de neuronas tradicionales, se conecta a una nueva capa oculta de neuronas tipo feedforward.

ARQUITECTURA DE UNA CNN



Una vista general de modelos pre entrenados

- <https://medium.com/@sohaib.zafar522/choosing-the-right-pre-trained-model-a-guide-to-vggnet-resnet-googlenet-alexnet-and-inception-db7a8c918510>
- <https://www.kaggle.com/code/deepakat002/vgg-resnet-inception-xception-comparision>

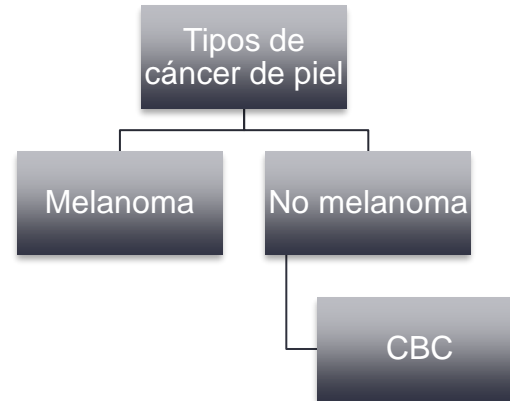
Detección de carcinoma basocelular utilizando red neuronal convolucional y support vector machine

Carcinoma basocelular

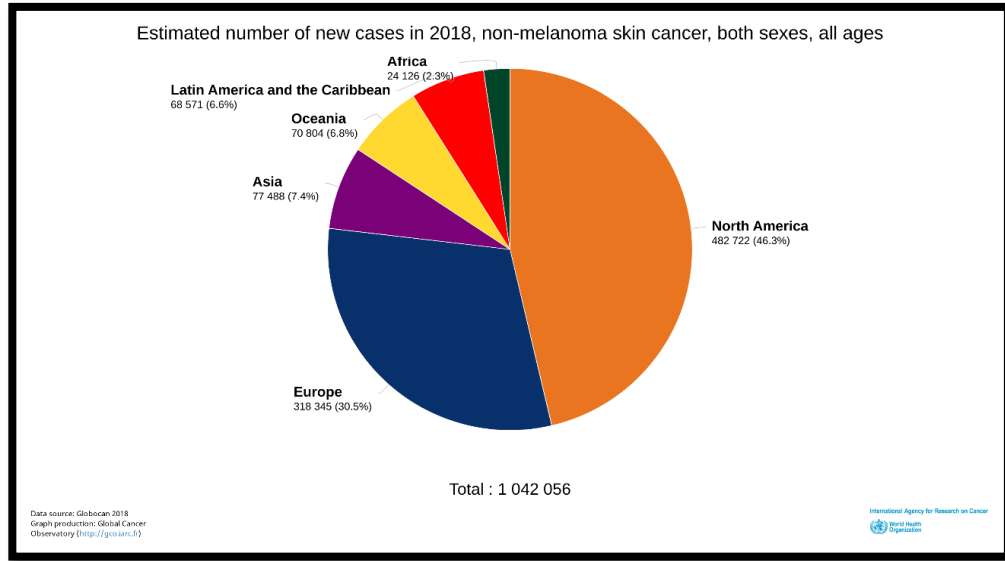
El carcinoma basocelular (CBC) es un tumor epitelial maligno que se caracteriza por un crecimiento lento, pero que tiene capacidad destructora local y compromete extensas áreas de tejido, cartílago e incluso huesos. (Darias Domínguez & Garrido Celis, 2018).



Imagen de carcinoma basocelular



Problema



- Hubo un aproximado de 833,644 de nuevos casos de carcinoma basocelular a nivel mundial
- Hubo un aproximado de 41,569 de nuevos casos de carcinoma basocelular en Latinoamerica
- Perú: 2,036 nuevos casos.

Descripción del problema

El médico tiene dificultades en lectura de imágenes dermatoscópicas puede ser menos preciso al realizar el diagnóstico.

La precisión en el análisis de imágenes dermatoscópicas depende de la experiencia del médico.

En la medicina se busca alcanzar un buen diagnóstico.

Motivación

Implementar un modelo híbrido de inteligencia artificial el cual permita clasificar eficazmente el carcinoma basocelular, puesto que actualmente hay trabajos (Ahmed et al., 2019; Haghighi et al., 2020; Kurek, 2018; Liu et al., 2018) que obtuvieron mejores resultados de clasificación utilizando modelos híbridos frente a los modelos clásicos de inteligencia artificial.

- Ahmed, T., Wahid, Md. F., & Hasan, Md. J. (2019). Combining Deep Convolutional Neural Network with Support Vector Machine to Classify Microscopic Bacteria Images. *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 1-5. <https://doi.org/10.1109/ECACE.2019.8679397>
- Haghighi, S. N., Danyali, H., Helfroush, M. S., & Karami, M. H. (2020). A Deep Convolutional Neural Network for Melanoma Recognition in Dermoscopy Images. *2020 10th International Conference on Computer and Knowledge Engineering (ICCKE)*, 453-456. <https://doi.org/10.1109/ICCKE50421.2020.9303684>
- Kurek, J. (2018). Hybrid approach towards the assessment of a drill condition using deep learning and the Support Vector Machine. *2018 22nd International Computer Science and Engineering Conference (ICSEC)*, 1-5. <https://doi.org/10.1109/ICSEC.2018.8712622>
- Liu, T., Ye, X., & Sun, B. (2018). Combining Convolutional Neural Network and Support Vector Machine for Gait-based Gender Recognition. *2018 Chinese Automation Congress (CAC)*, 3477-3481. <https://doi.org/10.1109/CAC.2018.8623118>

Formulación del problema

Problema general

- ¿Se puede diseñar un sistema el cual diagnostique eficazmente el carcinoma basocelular utilizando inteligencia artificial?

Problemas específicos

- ¿Qué tipo de arquitectura de red neuronal convolucional es mejor para clasificar el carcinoma basocelular?
- ¿Es posible construir el sistema que diagnostica el carcinoma basocelular?
- ¿Cómo podemos validar el sistema que diagnostica el carcinoma basocelular?

Importancia y justificación



Objetivos

Objetivo general

- Diseñar el sistema automático de detección de carcinoma basocelular basado en el modelo híbrido de red neuronal convolucional y Support Vector Machine.

Objetivos específicos

- Diseñar el modelo de detección de CBC utilizando redes neuronales convolucionales y máquina de vectores de soporte.
- Construir un sistema que diagnostica el carcinoma basocelular.
- Validar el modelo que diagnostica el carcinoma basocelular.

Estado del arte

| Categorías | Cantidad de artículos | Artículo |
|---|-----------------------|--|
| Arquitecturas para clasificar el cáncer de piel | 2 | Basal cell carcinoma detection in full field OCT images using convolutional neural networks. (Mandache et al.) |
| Métodos para clasificar el cáncer de piel | 10 | Deep Neural Network for Melanoma Classification in Dermoscopic Images. (Jiahao et al.) |
| Modelos para clasificar el cáncer de piel | 7 | Melanoma Detection Using Convolutional Neural Network. (Zhang et al.) |
| Framework para clasificar el cáncer de piel | 1 | Melanoma Recognition in Dermoscopy Images via Aggregated Deep Convolutional Features. (Yu et al.) |

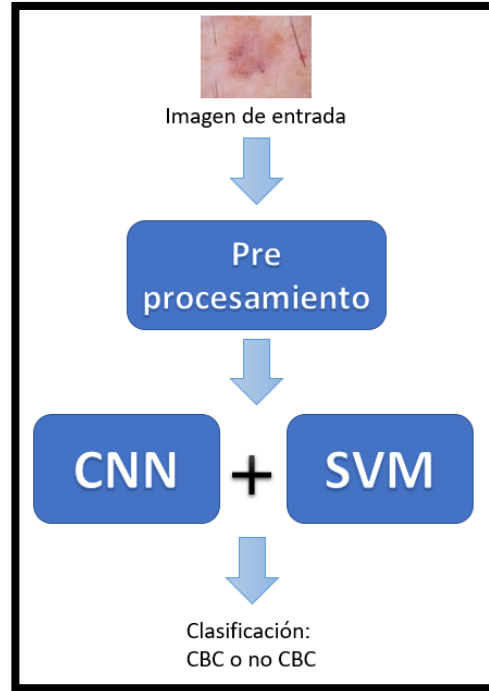
Estado del arte: Investigaciones que diagnosticaron

| Categoría | Título | Autor | Técnicas | Resultados |
|--------------|---|---|----------------------------|-------------------|
| Método | Deep residual neural networks for automated Basal Cell Carcinoma detection | E. V. Putten, A. Kambod and M. Kambod | ResNet | Exactitud: 93% |
| Método | Automatic detection of basal cell carcinoma using vascular-extracted features from dermoscopy images | P. Kharazmi, H. Lui, Z. J. Wang and T. K. Lee | Random Forest | Precisión: 82.7% |
| Arquitectura | Basal cell carcinoma detection in full field OCT images using convolutional neural networks | D. Mandache, E. Dalimier, J. R. Durkin, C. Boceara, J. -. Olivo-Marin and V. Meas-Yedid | Red neuronal convolucional | Precisión: 95.35% |
| Método | Automatic detection of translucency using a deep learning method from patches of clinical basal cell carcinoma images | H. Huang, P. Kharazmi, D. I. McLean, H. Lui, Z. J. Wang and T. K. Lee | Red neuronal convolucional | Exactitud: 93% |

Metodología e implementación: Dataset

| Dataset | License | Total images | Pathologic verification (%) | akiec | bcc | bkl | df | mel | nv | vasc |
|------------------------|---------------------------------|--------------|-----------------------------|-------|-----|------|-----|------|-------|------|
| PH2 | Research&Education ^a | 200 | 20.5% | - | - | - | - | 40 | 160 | - |
| Atlas | No license | 1024 | unknown | 5 | 42 | 70 | 20 | 275 | 582 | 30 |
| ISIC 2017 ^b | CC-0 | 13786 | 26.3% | 2 | 33 | 575 | 7 | 1019 | 11861 | 15 |
| Rosendahl | CC BY-NC 4.0 | 2259 | 100% | 295 | 296 | 490 | 30 | 342 | 803 | 3 |
| ViDIR Legacy | CC BY-NC 4.0 | 439 | 100% | 0 | 5 | 10 | 4 | 67 | 350 | 3 |
| ViDIR Current | CC BY-NC 4.0 | 3363 | 77.1% | 32 | 211 | 475 | 51 | 680 | 1832 | 82 |
| ViDIR MoleMax | CC BY-NC 4.0 | 3954 | 1.2% | 0 | 2 | 124 | 30 | 24 | 3720 | 54 |
| HAM10000 | CC BY-NC 4.0 | 10015 | 53.3% | 327 | 514 | 1099 | 115 | 1113 | 6705 | 142 |

Modelo para clasificar el CBC en alto nivel



Metodología e implementación: Pre procesamiento

- Redimensionamiento
- Aumento de datos (rotaciones, zoom y volteos)
- Normalización.

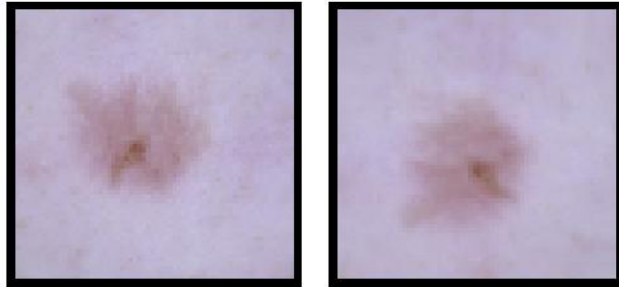
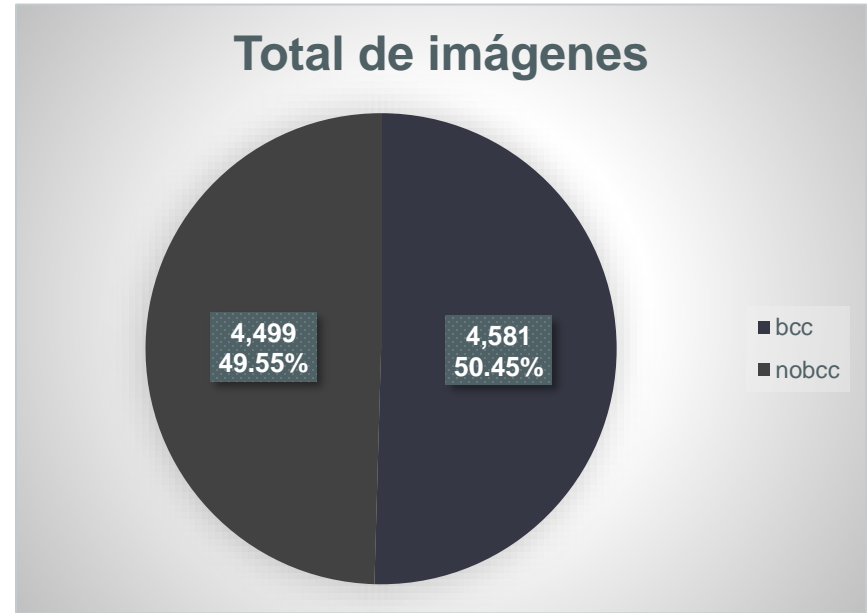
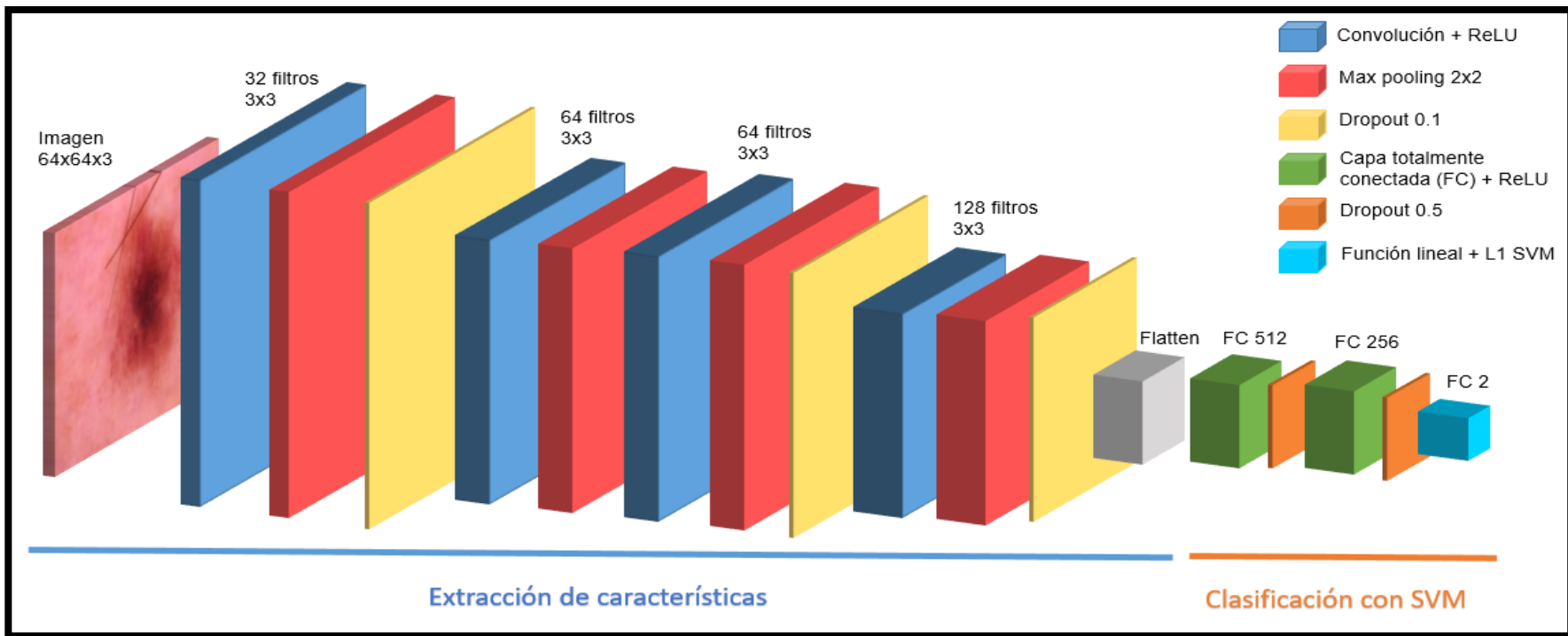


Imagen dermatoscópica (64x64) con rotación aleatoria.

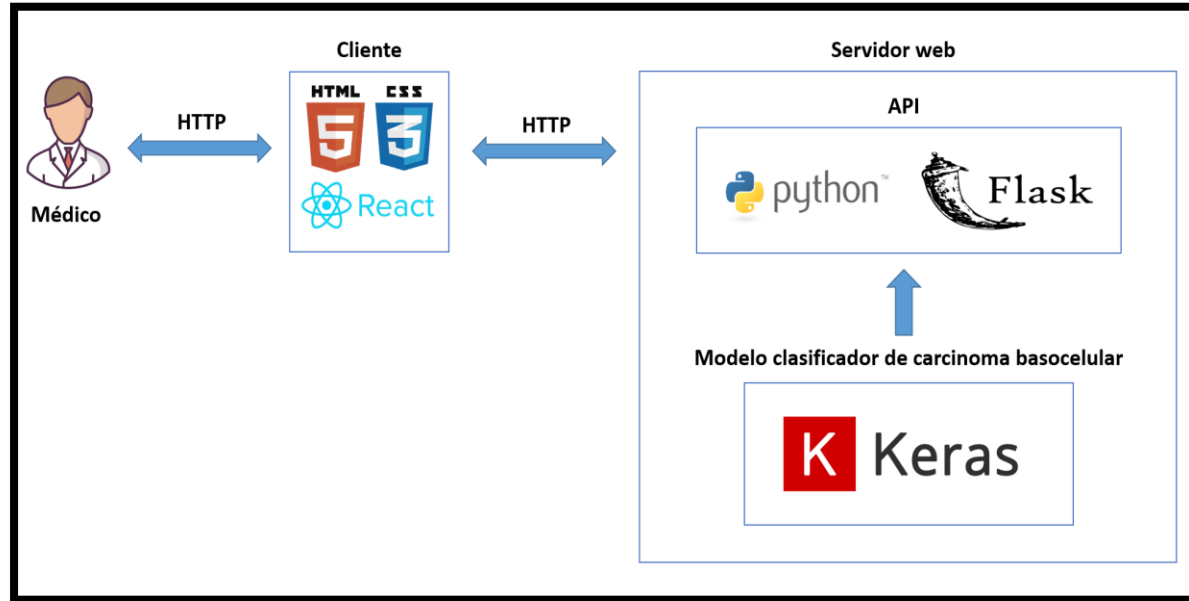


Total: 9,080 imágenes

Modelo híbrido CNN+SVM para clasificar el CBC

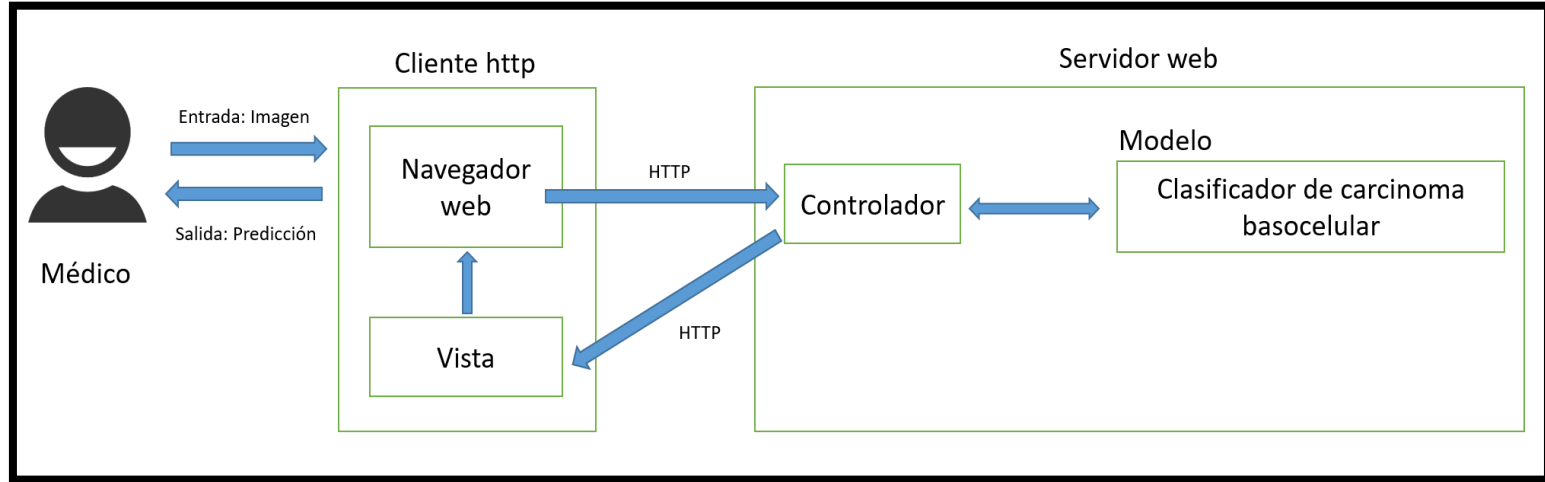


Metodología e implementación: Diseño



Arquitectura en alto nivel. Elaboración propia.

Metodología e implementación: Diseño



Representación de interacción de componentes. Elaboración propia.

Metodología e implementación: Diseño

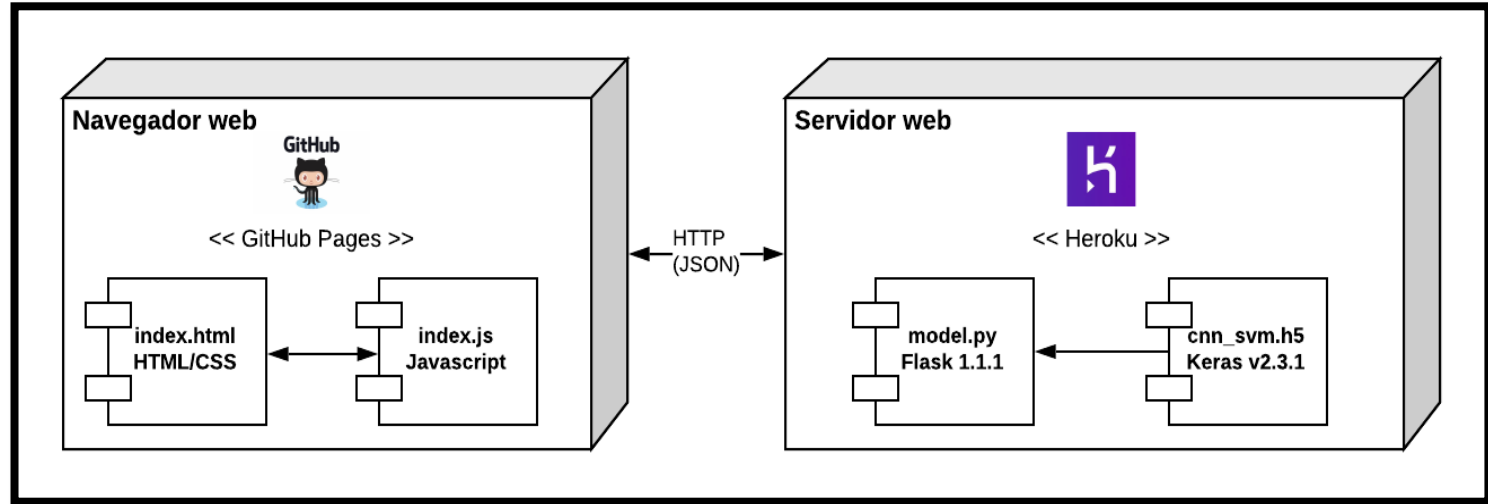


Diagrama de despliegue. Elaboración propia.

Metodología e implementación: Sistema

Predicción del carcinoma basocelular

Cargar la imagen en formato .png o .jpg

Examinar... ISIC_0028542.jpg

Predicir

Imagen



Resultados

Es carcinoma basocelular: 76 %

No es carcinoma basocelular: 24 %

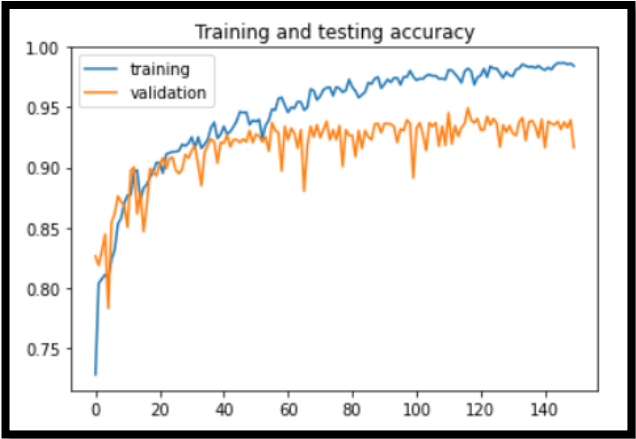
Resultados

- Métricas: Exactitud, precisión, recall y f1-score
- Inicialmente se contó con 5,008 imágenes dermatoscópicas.

| Proceso | Número de imágenes |
|---------------|--------------------|
| Entrenamiento | 5,107 |
| Evaluación | 1,703 |
| Prueba | 2,270 |
| Total | 9,080 |

| Parámetro | Valor |
|---------------------|--------|
| Épocas | 150 |
| Batch size | 80 |
| Optimizador | Adam |
| Función de pérdidas | L1-SVM |

Resultados: CNN + SVM



Evaluación de la exactitud para el modelo híbrido



Evaluación de la pérdida para el modelo híbrido

| Clasificación del CBC (2270) | | Predicción | |
|------------------------------|--------|------------|--------|
| | | CBC | No CBC |
| Observación | CBC | 1016 | 133 |
| | No CBC | 59 | 1062 |

| | Precisión | Recall | F1-score | Exactitud |
|----------|-----------|---------|----------|-----------|
| Promedio | 94.51 % | 88.42 % | 91.36 % | 91.54 % |

| | |
|---------------------|---------------------|
| True positive (TP) | False negative (FN) |
| False positive (FP) | True negative (TN) |

Discusión

| Autor | Técnica | Dataset | Exactitud | Precisión | Recall | F1-score |
|---------------------------|---------------|--|-----------|---------------|----------------|----------|
| D. Mandache et. al | CNN | Dataset propio (TCO) | - | 95.35% | 95.2% | - |
| H. Huang et. al | CNN | Dataset propio (Imágenes clínicas) | 93% | - | 77% | - |
| E. V. Putten et. al | ResNet | Skin Lesion Analysis Towards Melanoma Detection ISBI 2016 | 93% | - | 97% | - |
| P. Kharazmi et. al | Random Forest | Atlas of dermoscopy, la universidad de Missouri y Vancouver Skin Care Centre | - | 82.7% | 82.4% | - |
| Esta investigación | CNN + SVM | HAM10000 | 91.54% | 94.51% | 88.42 % | 91.36% |

Conclusiones

Se propuso un sistema de detección del carcinoma basocelular implementando un modelo híbrido de redes neuronales convolucional y Support Vector Machine y utilizando el dataset HAM10000, el cual clasifica eficazmente el carcinoma basocelular.

Trabajos futuros

- Se podría probar diferentes técnicas de segmentación de imagen, las cuales permitan aumentar la exactitud de la clasificación.
- Se puede adaptar este modelo híbrido CNN+SVM para clasificar otros tipos de cáncer de piel.

Preguntas?