



UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO

TESIS PARA OPTAR POR EL TÍTULO PROFESIONAL DE INGENIERO DE SISTEMAS



“SISTEMA INTELIGENTE BASADO EN DEEP LEARNING PARA EL DIAGNÓSTICO DE CÁNCER DE PRÓSTATA”

TESISTAS:

- SANTAMARIA SANTISTEBAN, Jahir Santos
- SIESQUEN VALDIVIA, Luis Felipe

ASESOR:

MG. ING. VILLEGAS CUBAS, Juan Elías



01

GENERALIDADES



PLANTEAMIENTO DEL PROBLEMA

SITUACIÓN

- El cáncer de próstata es una de las principales causas de mortalidad en hombres a nivel mundial. Sin embargo, la detección temprana de esta enfermedad resulta un desafío debido a la complejidad en la identificación de signos iniciales y a la limitada disponibilidad de tecnologías avanzadas en varias regiones.
- Esta situación genera la necesidad de implementar sistemas de diagnóstico automáticos y accesibles para mejorar la precisión y oportunidad del diagnóstico clínico.

FORMULACIÓN PROBLEMA

- ¿Cómo detectar de forma temprana del cáncer de próstata, permitiendo un diagnóstico más preciso y oportuno en el ámbito de la salud?



PLANTEAMIENTO DEL PROBLEMA

HIPÓTESIS

- La implementación de un sistema inteligente basado en Deep Learning permitirá mejorar la detección temprana del cáncer de próstata, optimizando la precisión y confiabilidad en el análisis de imágenes médicas.

Objetivo general

- Desarrollar un sistema inteligente basado en Deep Learning para predecir el diagnóstico de cáncer de próstata.

OBJETIVOS

Objetivos específicos

- Analizar y preprocesar imágenes de cáncer de próstata para crear un conjunto de datos representativo para el modelo.
- Desarrollar y optimizar un modelo de Deep Learning adecuado para la clasificación de imágenes médicas que permita diferenciar entre muestras con cáncer y sin cáncer.
- Implementar un sistema de predicción que integre el modelo entrenado en una interfaz accesible para profesionales de la salud.
- Evaluar el rendimiento del sistema mediante métricas de precisión, exactitud, sensibilidad y especificidad para validar su eficacia en la detección de cáncer de próstata.



02

FUNDAMENTOS



MARCO TEÓRICO

Sistema Inteligente basado
en Deep Learning para el
Diagnóstico de Cáncer de
Próstata

Guajin y Kup-Sze, 2018

(Antecedente Internacional)

Faez, 2018

(Antecedente Nacional)

Forero Cuellar, 2019

(Antecedente Nacional)

- Desarrollaron un sistema para diagnosticar cáncer de próstata en una población china utilizando biopsias de próstata guiadas por ecografía transrectal (TRUS). Se emplearon varios métodos de aprendizaje automático, entre ellos, Support Vector Machine (SVM) y Red Neuronal Artificial (ANN), destacándose ANN con una precisión del 95.27%
- En su investigación en el hospital Imam Reza de Teherán, empleó técnicas de Deep Learning para mejorar el diagnóstico de cáncer de próstata, logrando una precisión del 86.3% en la clasificación.
- Usó el software "Orange Data Mining" para realizar minería de datos en el diagnóstico de cáncer de próstata mediante imágenes de resonancia magnética, logrando una precisión del 83.1% con regresión logística.



MARCO TEÓRICO

Detección del Cáncer de próstata-Métodos

Lal et al., 2019

(Antecedente Internacional)

- Estudiaron la relación entre características morfológicas del cáncer de próstata y las imágenes de resonancia magnética (MRI). Este enfoque permite una mejor visualización y clasificación de los tumores, proporcionando una herramienta de apoyo en la decisión clínica para el diagnóstico y tratamiento.

Yanes Chacón, Villalobos Campos y Cubas González, 2023

(Antecedente Nacional)

- Mencionan que los factores de riesgo como la edad avanzada, la historia familiar y la raza influyen significativamente en el desarrollo del cáncer de próstata. En el contexto peruano, señalan la importancia de mejorar el acceso a métodos diagnósticos y la educación sobre factores de riesgo para promover la detección temprana.



03

METODOLOGÍA



OPERACIONALIZACIÓN DE VARIABLES

Variable	Dimensión	Indicadores	
Rendimiento	Eficiencia del sistema inteligente basado en Deep Learning.	Exactitud: Mide el rendimiento del sistema en la detección de cáncer de próstata, reflejando cuántas veces clasifica correctamente tanto los casos con cáncer como los sin cáncer.	$Exactitud = \frac{VP + VN}{VP + FP + FN + VN}$
		Precisión: Evalúa la confiabilidad del sistema al identificar correctamente los casos de cáncer de próstata positivos.	$Precisión = \frac{VP}{VP + FP}$
		Sensibilidad: Indica la capacidad del sistema para detectar todos los casos verdaderos de cáncer de próstata, asegurando que los pacientes con cáncer no pasen desapercibidos y reciban la atención necesaria.	$Sensibilidad = \frac{VP}{VP + FN}$
		Especificidad: Mide la habilidad del sistema para identificar correctamente los casos sin cáncer de próstata, reduciendo el número de falsos positivos brindando confiabilidad a los pacientes.	$Especificidad = \frac{VN}{VN + FP}$

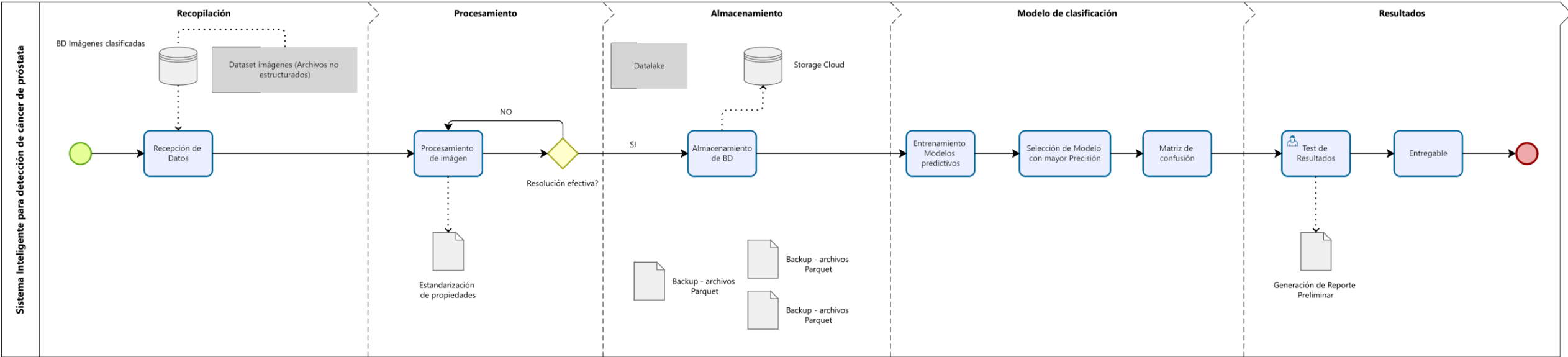


04

PROPUESTA



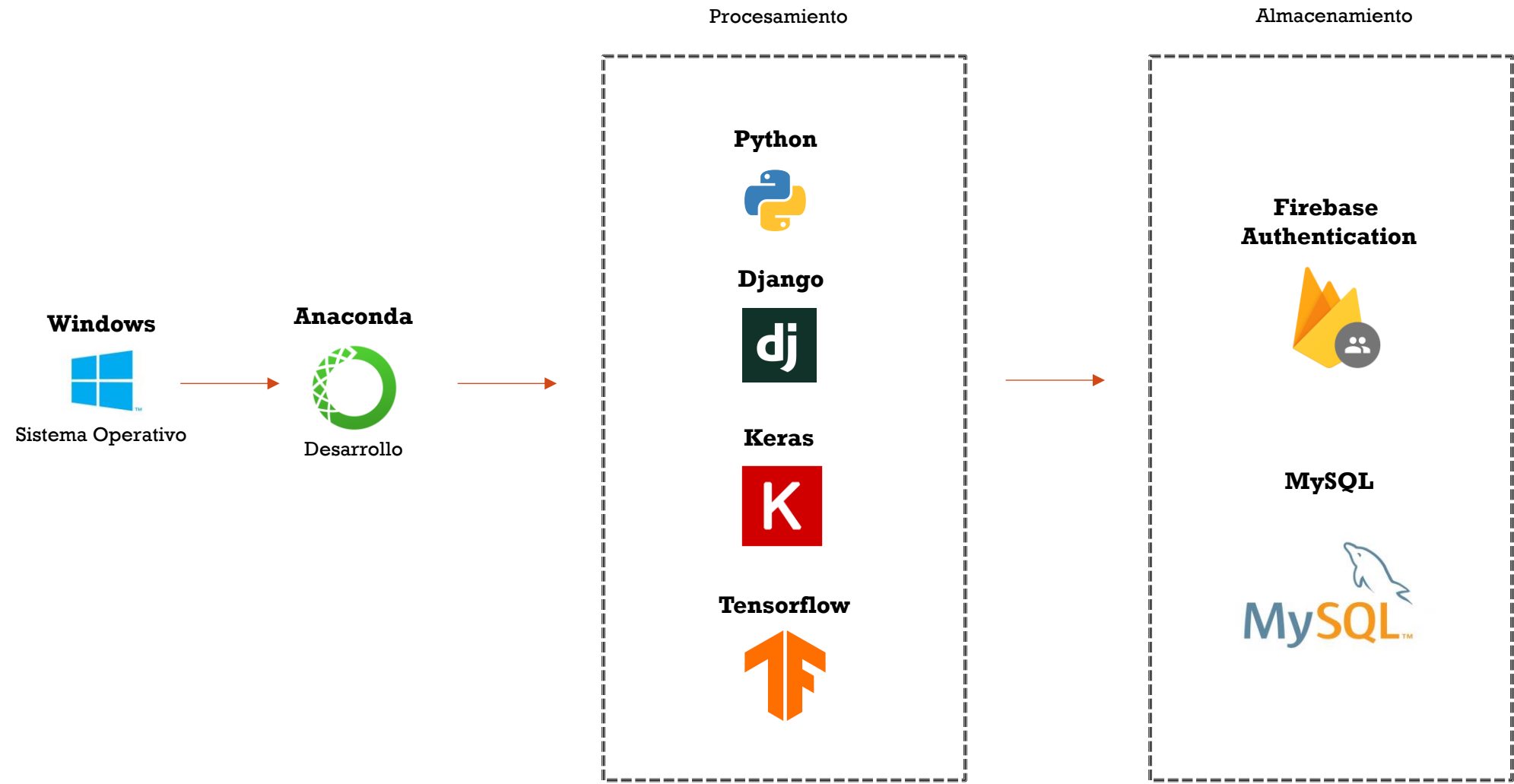
DESARROLLO DEL PROYECTO



Nota: Flujograma elaborado por los autores en base a la propuesta del modelo de predicción



HERRAMIENTAS PARA EL DESARROLLO DEL PROYECTO



Nota: Flujo elaborado por los autores de los recursos utilizados en el sistema inteligente



05

RESULTADOS



ANÁLISIS DE RESULTADOS

DATASET SICAP V2

Dataset	Entrenamiento(train)	Validación(val)	Total	Porcentaje
ConCancer	7,367	2,043	9,410	68%
SinCancer	3,590	827	4,417	32%

Nota: Tabla elaborada por los autores especificando la cantidad de imágenes utilizadas para el modelo

RESOLUCIÓN DE LA IMAGEN

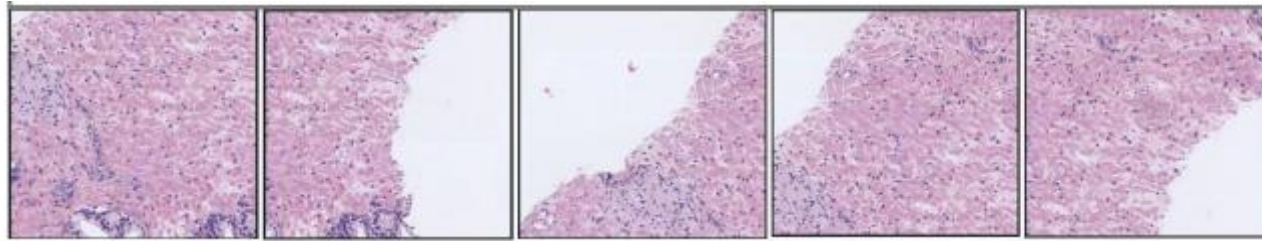
Imagen	
Id. de imagen	
Dimensiones	512 x 512
Ancho	512 pixeles
Alto	512 pixeles
Resolución horizontal	96 ppp
Resolución vertical	96 ppp
Profundidad en bits	24
Compresión	
Unidad de resolución	
Representación del color	
Bits comprimidos/pixel	

Nota: Resolución del tipo de imagen utilizada que contiene el dataset SICAP V2

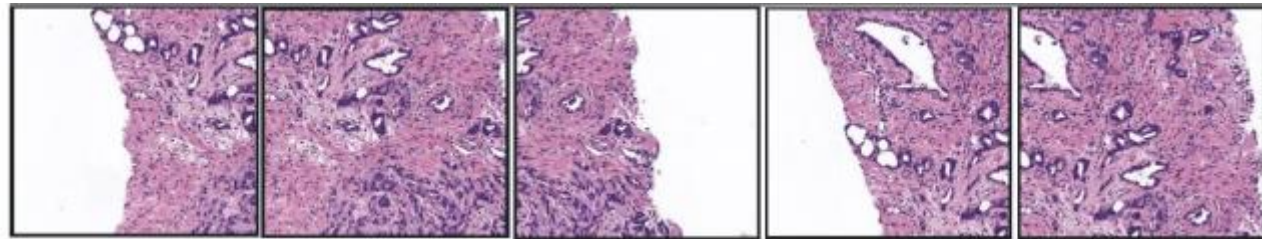


IMÁGENES CON PATRONES CRIBIFORMES

a) Imágenes con patrones cribiformes, no contiene signos de cáncer de próstata



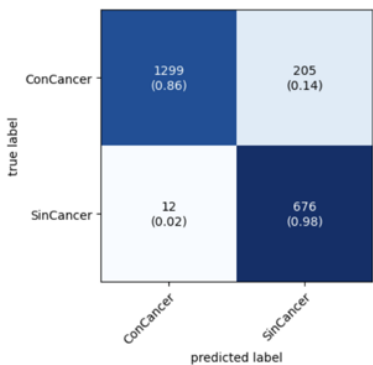
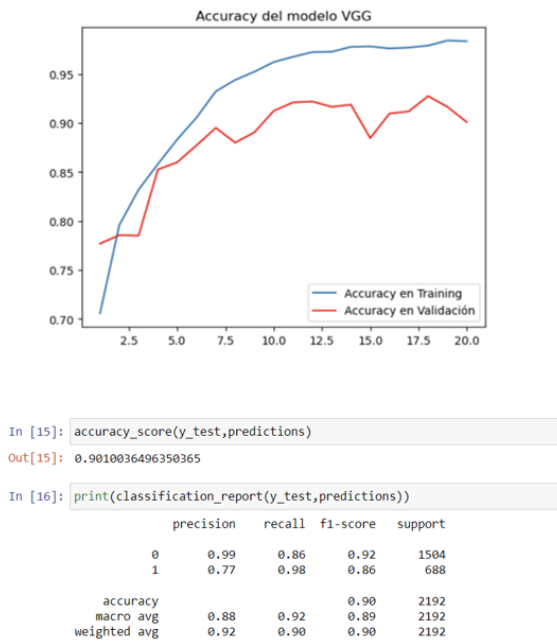
b) Imágenes con patrones cribiformes, contiene signos de cáncer de próstata



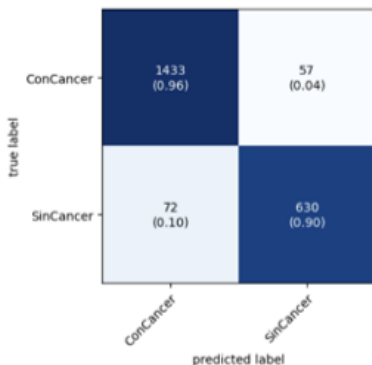
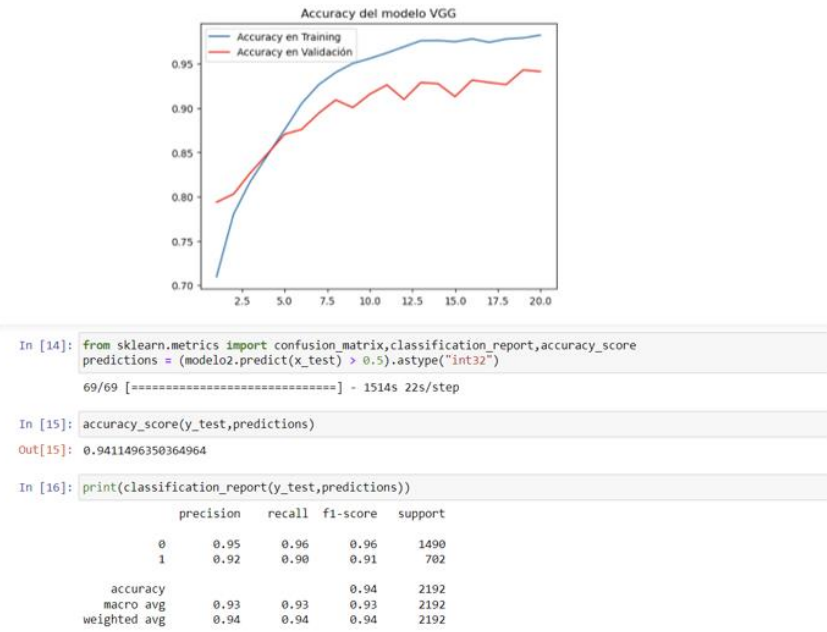
Nota: Obtenido y modificado de Rodríguez S. (2020)



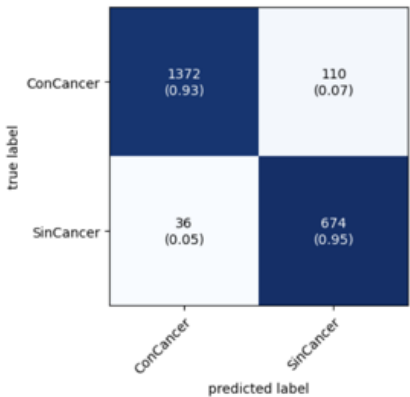
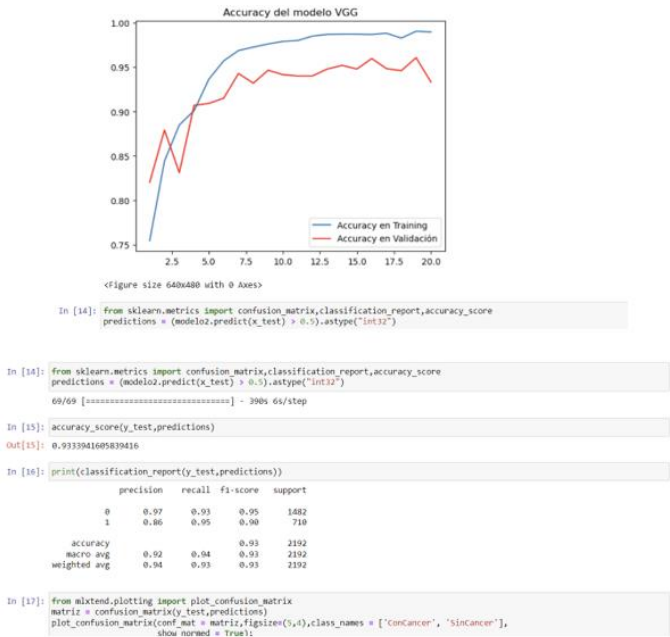
VGG16



VGG19



ResNet50



Entrenamiento con 20 épocas

Modelos	Precisión (Accuracy)	Exactitud	Sensibilidad	Especificidad
VGG16	90.10 %	77%	98%	76.7%
VGG19	94.11%	92%	90%	89.7%
ResNet50	93.33%	82%	95%	94.9%

Nota: Tabla comparativa de los modelos utilizados para la predicción con 20 épocas



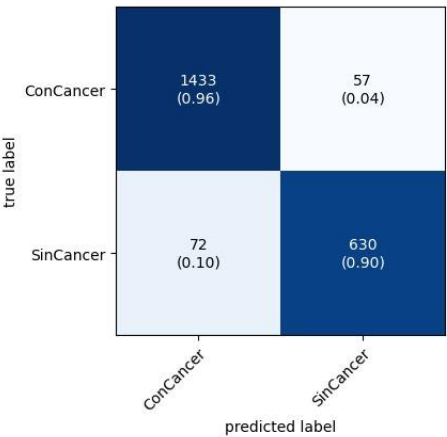
MODELO SELECCIONADO: VGG19

Después de entrenar el modelo con el algoritmo VGG19, se observa que la precisión de los datos de entrenamiento se mantiene constante en las últimas épocas, alcanzando un resultado del 98.22%. En cuanto a la precisión de los datos de validación, también se mantiene estable y finaliza con un resultado del 94.11%.

La función de pérdida para los datos de entrenamiento disminuye de manera lineal hasta alcanzar un 0.18%, lo cual es un resultado positivo. Sin embargo, la función de pérdida para los datos de validación muestra resultados constantes, por lo que se decidió finalizar la ejecución.

Precisión del modelo

Se empleó un modelo para evaluar su precisión utilizando un conjunto de imágenes. A través de la biblioteca Numpy, se realizó la clasificación de las imágenes y se generó la matriz de confusión, que permitió analizar el desempeño del modelo al mostrar las proporciones de clasificaciones correctas e incorrectas.



Clase 'ConCancer':

•**Verdaderos Positivos (TP): 1433 (0.96)**

El modelo identificó correctamente al 96% de los casos que realmente tienen cáncer, lo que demuestra una alta sensibilidad o capacidad para detectar la enfermedad.

•**Falsos Negativos (FN): 72 (0.10)**

El 10% de las predicciones dentro de esta clase fueron clasificadas incorrectamente como negativas, lo que implica que algunos casos con cáncer no fueron detectados. Aunque este porcentaje es bajo, reducir los falsos negativos es crítico en aplicaciones médicas para minimizar riesgos en pacientes.

Clase 'SinCancer':

•**Falsos Positivos (FP): 57 (0.04)**

El modelo clasificó erróneamente al 4% de los casos como positivos cuando en realidad no tenían cáncer. Este valor es aceptablemente bajo y refleja un buen nivel de especificidad, reduciendo diagnósticos falsos.

•**Verdaderos Negativos (TN): 630 (0.90)**

El modelo identificó correctamente al 90% de los casos que no tenían cáncer, lo que indica un buen desempeño en evitar alarmas innecesarias.



Análisis del modelo

El modelo muestra un desempeño sólido, especialmente en la detección de la clase "ConCancer", donde la proporción de verdaderos positivos es muy alta. Sin embargo, aún hay un margen de mejora para reducir los falsos negativos, dado el impacto crítico de no detectar casos positivos en un contexto de salud. Los falsos positivos, aunque menos críticos, también pueden ser minimizados para reducir posibles tratamientos innecesarios.

En general, el balance entre sensibilidad y especificidad parece razonable, pero ajustar el umbral de decisión o entrenar con más datos equilibrados podría optimizar aún más los resultados.



SISTEMA WEB

Login

Loguearse

Registrarse

Usuario

luisfelpiesquen22@gmail.com

Contraseña

Ingresar

CP

CORE

Inicio

Análisis de imagen

Historial del análisis

Pacientes

Ingrese una imagen para la predicción!

Título:

Análisis #1

Descripción:

....

Imagen:

Seleccionar archivo

1680024173_Block_Region_8_2_4_xini_9934_yini_3735.jpg

Dni:

72314652

Nombre:

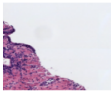
LUIS FELIPE

Apellido Materno:

VALDIVIA

Apellido Paterno:

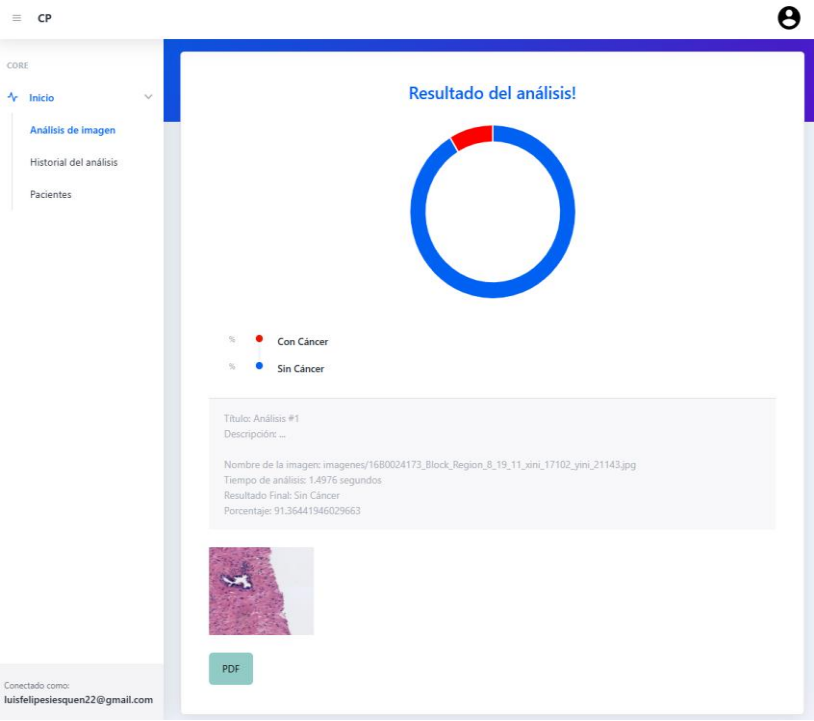
SIESQUEN



Analizar

Conectado como:

luisfelpiesquen22@gmail.com



06

CONCLUSIONES



CONCLUSIONES

01

El desarrollo de un sistema basado en Deep Learning para predecir el diagnóstico de cáncer de próstata tiene el potencial de mejorar la precisión y la rapidez en los diagnósticos. Sin embargo, requiere un conjunto de datos de alta calidad, un preprocesamiento adecuado y una evaluación rigurosa para garantizar su efectividad en entornos clínicos reales.

02

El preprocesamiento de imágenes de cáncer de próstata mejora la calidad y diversidad del conjunto de datos, asegurando un modelo preciso. La normalización, segmentación y aumento de datos son esenciales para entrenar modelos efectivos. Estos pasos son clave para obtener diagnósticos confiables y precisos

03

La integración de Deep Learning en el diagnóstico médico optimiza la detección temprana de cáncer, mejorando la precisión y reduciendo el margen de error en la clasificación de imágenes.

04

La implementación de este sistema inteligente en entornos hospitalarios podría transformar los procesos de diagnóstico, mejorando la atención al paciente y optimizando los recursos médicos.

05

El modelo VGG19 alcanzó una precisión del 94.11%, exactitud de 92%, sensibilidad del 90% y especificidad del 89.7% en el diagnóstico del cáncer de próstata, superando a VGG16 y ResNet50, demostrando ser una herramienta eficiente y precisa para el apoyo clínico.

07

RECOMENDACIONES



RECOMENDACIONES

01

Se recomienda integrar datos clínicos adicionales (edad, antecedentes familiares, marcadores biológicos) junto con las imágenes médicas para mejorar la precisión y personalización del diagnóstico.

02

Es esencial establecer protocolos estandarizados para la captura de imágenes médicas, garantizando consistencia y calidad, lo que aumentará la fiabilidad y reproducibilidad de los resultados.

03

Futuras investigaciones deben explorar nuevas arquitecturas de redes neuronales o enfoques híbridos, buscando mejorar aún más la precisión diagnóstica alcanzada con el modelo VGG19.