

Estado de la publicación: El preprint no ha sido enviado para publicación

Redes neuronales convolucionales para la efectividad del primer diagnóstico de VPH en Piura 2025

Walter Edgar Abanto Sánchez, Alexis López

<https://doi.org/10.1590/SciELOPreprints.14359>

Enviado en: 2025-12-04

Postado en: 2025-12-18 (versión 1)

(AAAA-MM-DD)

Convolutional Neural Networks for the Effectiveness of the First HPV Diagnosis in Piura 2025

Redes neuronales convolucionales para la efectividad del primer diagnóstico de VPH en Piura 2025

Abanto W.¹, López A.²

¹ <https://orcid.org/0000-0001-7581-025X>

² <https://orcid.org/0000-0001-5259-6086>

Escuela Profesional de Ingeniería de Sistemas, Facultad de Ingeniería y Arquitectura, Universidad César Vallejo, Perú.

RESUMEN

El presente trabajo de investigación desarrolló un aplicativo móvil que integra el modelo MobileNet V2 para detectar lesiones producidas por el Virus del Papiloma Humano (VPH), contribuyendo al ODS 9, enfocado en construir infraestructuras resilientes, promover la industrialización sostenible y fomentar la innovación. El objetivo fue evaluar la efectividad de una red neuronal convolucional (CNN) como herramienta de apoyo en la primera fase diagnóstica. Se empleó un enfoque mixto con diseño experimental, orientado a medir el nivel de efectividad de una CNN preentrenada para el diagnóstico inicial del VPH. El modelo MobileNet V2 fue adaptado mediante transfer learning y optimizado con técnicas de regularización como dropout, fine-tuning, entre otras. La efectividad se determinó mediante métricas de desempeño: precisión (98.63 %), sensibilidad (100 %), especificidad (97.33 %), VPP (97.26 %), VPN (100 %) y F1-Score (98.60 %). El entrenamiento resultó determinante, pues el uso de estrategias de regularización permitió al modelo extraer características relevantes para la identificación de lesiones. Aunque el periodo de prueba fue corto, los resultados fueron prometedores en un entorno clínico simulado. En conclusión, la CNN demostró ser una herramienta válida de apoyo diagnóstico, abriendo paso al uso de la inteligencia artificial en la detección temprana de enfermedades.

Palabras clave

Redes Neuronales Convolucionales, Virus del Papiloma Humano, Aplicativo Móvil, TensorFlow, MobileNet V2.

ABSTRACT

This research involves the development of a mobile application that integrates the MobileNet V2 model to detect lesions caused by the Human Papillomavirus (HPV), contributing to SDG 9, which focuses on building resilient infrastructure, promoting sustainable industrialization, and fostering innovation. The objective was to evaluate the effectiveness of a convolutional neural network (CNN) as a support tool in the initial diagnostic phase. A mixed-method approach with an experimental design was used to measure the effectiveness of a pre-trained CNN for the initial diagnosis of HPV. The MobileNet V2 model was adapted using transfer learning and optimized with regularization techniques such as dropout and fine-tuning. Effectiveness was determined using performance metrics: accuracy (98.63%), sensitivity (100%), specificity (97.33%), PPV (97.26%), NPV (100%), and F1-Score (98.60%). Training was crucial, as the use of regularization strategies allowed the model to extract relevant features for lesion identification. Although the testing period was short, the results were promising in a simulated clinical setting. In conclusion, the CNN proved to be a valid diagnostic support tool, paving the way for the use of artificial intelligence in early disease detection.

Keywords

Convolutional Neural Networks, Human Papillomavirus, Mobile Application, TensorFlow, MobileNet V2.

I. Introducción

La identificación temprana del Virus de Papiloma Humano (VPH), representa un desafío crítico en los sistemas de salud pública, especialmente en regiones donde el sector salud no cuenta con los recursos necesarios para emitir un diagnóstico rápido y preciso. La demora en el diagnóstico puede generar complicaciones graves, no solo para los pacientes, sino, para la población colindante que se expone a un posible contagio por tener múltiples parejas sexuales en su vida, dando como consecuencia un incremento del riesgo de propagación comunitaria. En Piura, la incidencia de esta enfermedad de transmisión sexual detectada tardíamente ha aumentado en los últimos años, evidenciando la falta de campañas de prevención, así como la ausencia de sistemas eficientes para la detección temprana y oportuna. Un grupo de investigadores Colombianos, expusieron que las aplicaciones móviles que integran redes neuronales convolucionales, ayudan en gran parte del diagnóstico de enfermedades bucales, demostrando que estas tecnologías se pueden implementar en distintos ámbitos de la medicina para la detección de diversas patologías [1]. EsSalud destaca los posibles peligros y beneficios de implementar la IA en el sector salud, recalcando que es inevitable pensar en un futuro no tan lejano, donde esta se integre cada vez más a un entorno cambiante, propio de este sector; por ello, es importante que los profesionales de la salud se familiaricen con el correcto uso de estas tecnologías y los agentes reguladores del Perú estén bien informados para enfrentar de la mejor manera los retos que se puedan presentar, resaltando los potenciales beneficios que estas aportan en el diagnóstico de enfermedades [2].

Estudios realizados en diferentes contextos, demuestran que las redes neuronales convolucionales han mostrado resultados prometedores, sobre todo las que fueron entrenadas con múltiples imágenes, ya que demuestran resultados aceptables dentro de las lesiones escogidas, facilitando la identificación de la enfermedad que padece el paciente [1]. Asimismo, se evidenció que el uso de estas mismas en la detección de retinopatía diabética, pueden presentar bajo porcentaje de error, pero ante estos problemas se destacó que no cualquier modelo de red neuronal es capaz de manejar la información como se espera; por ende, un modelo no preentrenado, resultó ser más efectivo para este tipo de estudios [3], aplicando el aprendizaje profundo (Deep Learning), demostraron la efectividad que puede llegar a tener una CNN en la identificación de imágenes, proporcionando resultados alentadores ante problemas no lineales y resaltando la obtención de diagnósticos precisos [4]; yendo un poco más allá, se combinaron el deep learning con el machine learning para detectar a tiempo el COVID-19 y, mediante esa información, predecir qué tan rápido podría progresar la enfermedad en los pacientes, aumentando la tasa de efectividad en los diagnósticos, para de ese modo tratar a tiempo y prevenir daños en los pulmones [5]. Ahora bien, el uso de IA en su variante de aprendizaje profundo, da resultados muy prometedores en la ayuda de

prevención e identificación de patologías antes de generar daños en la salud de los pacientes, que muchas veces son mortales.

Ampliando la información en esta línea, se evidenció que la detección de cáncer cervical en sus diferentes tipos, alcanzaron hasta un 97,35% de precisión, con una leve pérdida de información en el modelo VGG16; por otra parte el modelo VGG19, obtuvo 97,10% de precisión, con una pérdida creciente de datos; son detalles que se pueden mejorar con la realización de una buena segmentación de imágenes y un correcto balance en las clases, que implica el aumento de datos proporcionados para el entrenamiento de las redes neuronales aplicadas [6].

En este contexto, la implementación de redes neuronales convolucionales en el sector salud, ha reflejado cierta eficacia en las distintas ramas donde fueron aplicadas. Un estudio hecho hace algunos años atrás, indica el alto grado de efectividad que tienen estas tecnologías, ayudando a reducir el tiempo del proceso diagnóstico, que conlleva a una reducción considerable para la autorización de inicio al tratamiento correspondiente [1]. Así como estos estudios, existen gran variedad que respaldan el uso e implementación de estas tecnologías para el correcto diagnóstico de VPH en la ciudad de Piura.

La incidencia del VPH está influenciada por la práctica sexual sin precauciones, ya sea por el no uso de preservativos o falta de vacunas para prevenirlo [7]. Desde otro punto de vista, múltiples estudios demuestran una fuerte tendencia al uso de redes neuronales en la detección de distintas patologías, debido a los buenos resultados evidenciados en investigaciones que las utilizan para evaluar el aprendizaje y desempeño por las que fueron entrenadas, para esos estudios se han utilizado desemejantes modelos, donde destacan las CNN, ANN, SVM, LSTM y varios más, siendo entrenados para identificar las lesiones producidas por el contagio de alguna enfermedad [8].

Existen más de 200 variantes del VPH, de las cuales, algunas si son consideradas de alto riesgo. En Piura existe una prevalencia muy amplia, siendo las variantes más frecuentes en mujeres de 31 a 35 años el tipo VPH (-16, -56) y el VPH-P2(VPH-56, -59, -66) [9], por otro lado, en hombres de 18 a 70 años se identifican variantes como la de tipo VPH (-16, -18, -26, -31, -33, -35, -39, -45, -51, -52, -53, -56, -58, -59, -66, -68, -69, -73 y -82). Se conoce que este virus causa cáncer en ambos sexos, pero más comúnmente en mujeres, ya que su útero es el principal afectado [10].

La problemática de este proyecto, surge de la necesidad de implementar nuevas formas tecnológicas para la eficiente identificación del VPH a través de las lesiones cutáneas y mucosas que pueda presentar el paciente. La ausencia de herramientas avanzadas que integren redes neuronales convolucionales para el análisis y automatización del diagnóstico de esta enfermedad, limita la capacidad

de los médicos al identificar las lesiones producidas a raíz del contagio. En contexto del primer diagnóstico visual, los médicos dependen de su propio criterio, lo que fomenta posibles diagnósticos erróneos y tardíos que pueden desencadenar tratamientos equívocos. Las redes neuronales convolucionales juegan un papel muy importante para solucionar mediante el análisis de imágenes, posibles errores en la percepción humana, que se demuestra en la primera fase de diagnóstico. Sin embargo, estos incluyen la necesidad de grandes volúmenes de datos previamente etiquetados, junto con la variabilidad en la calidad de imágenes médicas y por consecuencia, el correcto entrenamiento de la red neuronal.

Para este estudio, se creó un aplicativo móvil innovador que integra un modelo de CNN para mejorar la detección oportuna y eficaz del VPH en Piura. Esta herramienta combina inteligencia artificial con accesibilidad móvil, permitiendo analizar imágenes clínicas, para brindar diagnósticos preliminares con la mayor precisión posible. En un escenario donde la tecnología puede desempeñar un rol clave en el diagnóstico médico, este estudio se presenta como una solución prometedora para optimizar la atención. A través del entrenamiento con imágenes referenciales de las lesiones cutáneas y mucosas de los pacientes diagnosticados con VPH, se buscó no solo acelerar el proceso de diagnóstico, sino también reducir el margen de error en la identificación de dicha patología, para de esta manera poder contribuir con el sistema de salud y promuevan la inversión en estas tecnologías.

Referente al problema general, este fue planteado de la siguiente forma, ¿Qué nivel de efectividad tuvo el diagnóstico de VPH con el uso de un modelo de redes neuronales convolucionales en Piura?. Además de los problemas específicos planteados: (a) ¿Cuál fue la tasa de aciertos y desaciertos de la red neuronal convolucional para detectar VPH?. (b) ¿Qué tan efectivo resultó el modelo de CNN entrenado para identificar el VPH?. (c) ¿Cómo afectó el tamaño y calidad del conjunto de datos en el desempeño de la red neuronal convolucional para el diagnóstico de VPH?.

La justificación principal propuso: El respaldo teórico de esta investigación se fundamentó en diversos estudios que demuestran la eficacia del uso de redes neuronales convolucionales como herramienta tecnológica en la salud pública. Por ejemplo, Mohamad y Suriani, verificaron que la precisión total de su aplicación móvil alcanzó el 91%. La proporción de identificación fue del 100%, con un periodo de inferencia de 317 ms. El modelo MobileNet V2 es el más apropiado para dispositivos móviles, gracias a su gran eficacia, tamaño reducido y rápido entrenamiento, demostrando el potencial tecnológico en la usabilidad de sus recursos. Además, esta evidencia destacó la relevancia de implementar un sistema que utilice redes neuronales convolucionales [11].

Respecto a la justificación social, estuvo sustentada en la necesidad de optimizar

los diagnósticos médicos en la zona de Piura, donde la prevalencia del VPH ha experimentado un incremento en años recientes. Al haberse puesto en marcha una aplicación móvil con integración de un modelo convolucional para el diagnóstico automatizado de VPH, mediante la comparación de imágenes referentes a lesiones en la piel y mucosas, se buscó acelerar significativamente el proceso de diagnóstico. El propósito de la investigación fue no solo evaluar la efectividad de la CNN al detectar VPH, sino también optimizar la precisión en la primera etapa de diagnóstico brindado por el personal médico, al facilitar una identificación inmediata ante la posible presencia de lesiones causadas por esta patología. Esta innovación tecnológica permitió la detección y puesta en marcha del tratamiento, evitando futuras condiciones precancerosas, añadiendo veracidad a los resultados clínicos de los pacientes. Asimismo, se presentó como una herramienta útil para beneficiar a las personas que carecen de recursos económicos, validando el acceso sin discriminación al servicio de salud que promueve el estado peruano.

La justificación tecnológica estuvo amparada en la incorporación de métodos avanzados, tales como la técnica propuesta en este estudio para el beneficio de los pacientes y sector salud. Este estudio propuso el entrenamiento de un modelo de CNN para procesar volúmenes de datos clínicos (imágenes) y realizar el análisis de los patrones causados por VPH. Esta implementación actuó como una tecnología líder en el enfoque de identificación temprana a irregularidades causadas por este virus. Además, continúa profundizando e impulsando la investigación de futuros ingenieros que deseen innovar con nuevas herramientas para contribuir en este sector, ofreciendo un producto que puede adaptarse a la evolución de nuevas formas de usar la IA.

En última instancia, la justificación metodológica se basó en el enfoque estricto que se implementó para realizar el entrenamiento de la CNN. Se utilizaron métodos cuantitativos para evaluar la efectividad del modelo de red neuronal convolucional escogido, sin infringir las normas éticas y confidenciales que por ley mantienen los hospitales y pacientes ante este tipo de investigaciones. Fue esencial recopilar un porcentaje considerable de imágenes clínicas referente a la enfermedad, para garantizar un correcto entrenamiento del modelo, asegurando la fiabilidad y solidez de los resultados. Dicho enfoque metodológico no sólo afirmó la validez del estudio, sino que también estableció un esquema replicable en futuras investigaciones orientadas a la identificación de enfermedades mediante el uso de esta tecnología.

Este estudio tuvo como objetivo general: Evaluar el nivel de efectividad de las redes neuronales convolucionales en el primer diagnóstico de VPH en Piura, 2025. Acorde a ello, se establecieron los siguientes objetivos específicos: (a) Medir la sensibilidad de la CNN al identificar correctamente las lesiones del VPH. (b) Elegir un modelo adecuado de CNN para procesar imágenes reales de la patología. (c) Evaluar la calidad de la CNN considerando las dimensiones de validez diagnóstica

y aceptación por el médico.

Una vez definidos los objetivos, resulta pertinente presentar los antecedentes que respaldaron el desarrollo de esta investigación, según el artículo en Bangladesh, cuyo objetivo fue desarrollar un modelo eficiente de red neuronal convolucional propia, con el fin de clasificar tipos de cuello uterino y detectar anomalías celulares. La metodología fue de naturaleza experimental que duró 6 meses. Entre los resultados que destacan, es la utilización del conjunto de datos SiPaKMeD para entrenar desde cero una arquitectura de red neuronal convolucional personalizada, comparándola con modelos populares como ResNet50, VGG16 y MobileNetV2, logrando una precisión superior al 98%, tanto en la clasificación de anomalías, como en tipos de cuello uterino, destacando además por su baja complejidad computacional. Dicha conclusión, la perfila como una herramienta viable para hospitales con recursos reducidos, gracias a su balance entre precisión y eficiencia [12].

Por otro lado, investigadores estadounidenses y etíopes, analizaron la sinergia entre arquitecturas SWIN Transformer y CNN, para identificar cáncer cervical en etapas iniciales mediante imágenes colposcópicas. Adoptando la metodología experimental, por el cual llegaron a tener como resultado un rendimiento consistente en diversos conjuntos de datos, con un área bajo la curva (AUC) de 80 a 82% y precisión entre 75 a 78%. Estos resultados reflejan una capacidad excepcional para clasificar tejidos normales, precancerosos y cancerosos. Donde concluyeron que la fusión de estas arquitecturas optimiza la precisión diagnóstica, ofreciendo una herramienta valiosa para áreas con limitado acceso a especialistas [13].

Investigadores chinos, diseñaron un sistema de inteligencia artificial basado en CNN para optimizar el cribado de cáncer cervical en regiones con recursos limitados. Mediante el análisis de imágenes citológicas recolectadas en múltiples centros hospitalarios, entrenaron un modelo Faster R-CNN con un extenso dataset de más de 13,000 imágenes, validando su desempeño con una cohorte masiva de 400,000 muestras. Los resultados revelaron una sensibilidad excepcional (99.4%), aunque con una especificidad moderada de 34.8% y un AUC de 0.67. Si bien el modelo demostró alta eficacia en la detección, el estudio concluyó que era necesario mejorar la especificidad para minimizar diagnósticos con falsos positivos y garantizar su viabilidad en la práctica clínica real [14].

También en China, investigadores formalizaron como objetivo, realizar un análisis comparativo a modelos de aprendizaje profundo pre-entrenados, los elegidos fueron Xception, ResNet-50 y VGG-19 para la diferenciación automatizada entre adenocarcinoma cervical y tejido benigno a partir de imágenes histológicas. La investigación tuvo como resultados que la arquitecta Xception, superó significativamente a los demás modelos, alcanzando una precisión diagnóstica del

94.5% y un área bajo la curva (AUC) de 0.966, demostrando una capacidad de clasificación comparable al del criterio de patólogos expertos. En conclusión el estudio demostró el potencial de las CNN como herramienta de apoyo en el diagnóstico histopatológico, ofreciendo precisión y eficiencia en la identificación de neoplasias cervicales [15]. En este mismo país, desarrollaron un sistema basado en redes neuronales convolucionales para determinar el estado de infección por VPH en casos de cáncer orofaríngeo, utilizando imágenes de tomografía computarizada. La investigación implementó una novedosa metodología de aprendizaje por transferencia, donde los modelos se preentrenaron con datos de videos deportivos antes de su adaptación al análisis de imágenes médicas tridimensionales. Los resultados mostraron un rendimiento consistente, alcanzando valores del 75% en sensibilidad y 72% en especificidad. Este desempeño se mantuvo a pesar de la variabilidad inherente en los datos provenientes de diferentes centros hospitalarios, lo que sugiere la robustez del enfoque propuesto. La investigación concluyó que los algoritmos de aprendizaje profundo pueden extraer patrones relevantes de las imágenes médicas para predecir el estado viral, incluso cuando se trabaja con conjuntos de datos heterogéneos [16].

Estudiantes peruanos de la Universidad Nacional Mayor de San Marcos, elaboraron un estudio, cuyo objetivo principal fue crear un sistema automatizado para detectar cáncer de cuello uterino en hospitales públicos, a través del procesamiento inteligente de imágenes asociadas a citologías líquidas, obtenidas de pruebas de Papanicolaou, enfocándose especialmente en entornos con pocos recursos médicos. Para alcanzar este objetivo, implementaron un método que combinó el algoritmo de segmentación k-means con redes neuronales convolucionales, usando en particular la arquitectura ResNet50. El modelo se entrenó con un banco de 963 imágenes médicas, categorizadas en cuatro grupos diagnósticos: sin indicios de malignidad intraepitelial, lesión intraepitelial escamosa leve, lesión intraepitelial escamosa grave y carcinoma escamoso invasivo. Entre los resultados más relevantes, ResNet50 alcanzó un 91% de exactitud en la identificación de las imágenes, mostrando un desempeño consistente para diferenciar las etapas del cáncer cervical. En conclusión, confirmaron la propuesta como opción viable y eficaz para facilitar el diagnóstico oportuno en centros de salud públicos del Perú, favoreciendo la identificación temprana de anomalías relacionadas con el VPH [17].

Investigadores en Alemania desarrollaron recientemente un innovador sistema de diagnóstico asistido por inteligencia artificial para la esclerosis múltiple, que combina el poder predictivo de las redes neuronales convolucionales, con técnicas avanzadas de explicabilidad algorítmica. Haciendo uso de imágenes referentes a resonancias magnéticas del proyecto ADNI sobre Alzheimer, aplicaron el aprendizaje por transferencia, para posteriormente distinguir entre pacientes con esclerosis múltiple, con resultados de precisión del 87.04% y un AUC de 96.08%,

demostrando la eficacia que poseen las CNN como herramienta diagnóstica. En conclusión, afirmaron que es posible desarrollar herramientas de IA que potencialmente superen el rendimiento diagnóstico de los especialistas humanos, facilitando así la integración de estas mismas en los flujos de trabajo y permitiendo una mejor colaboración entre sistemas inteligentes y profesionales médicos [18].

Finalmente, investigadores marroquíes propusieron una metodología avanzada basada en CNN, para mejorar la detección de patologías foliares. El objetivo principal fue desarrollar un sistema integral que combinara alta precisión con interpretabilidad del modelo, incorporando técnicas innovadoras de preprocesamiento y visualización en tiempo real. La metodología implementó un robusto flujo de actividades que incluyeron: (1) técnicas avanzadas de aumento de datos para optimizar el conjunto de entrenamiento, (2) una arquitectura CNN profunda con activaciones ReLU para extraer características complejas, (3) un novedoso módulo de visualización 2D que permite monitorear el aprendizaje de los filtros y, (4) una interfaz interactiva para ajustar parámetros dinámicamente. Los resultados en el dataset PlantVillage demostraron un rendimiento excepcional, superando el 99% de precisión, mientras que el módulo de visualización proporcionó insights valiosos sobre las características aprendidas por la red. La conclusión destaca que el enfoque integral no solo logra maximizar la precisión diagnóstica, sino que también resuelve el problema de interpretabilidad típico en modelos de deep learning, facilitando su adopción en aplicaciones agrícolas reales [19].

El virus de papiloma humano contiene ADN que al ingresar en el huésped, genera lesiones que mayormente aparecen en las mucosas orales y genitales, rara vez, algunos de sus síntomas son picazón, sangrado o dolor, mientras que el tiempo de aparición de lesiones posteriores al contagio, va a depender del estado inmunológico del huésped [20]. En gran parte de casos, el sistema inmune logra eliminar el virus, sin embargo, cuando no logra combatirlo, se desencadenan las lesiones típicas del VPH [21], la incidencia clínica abarca desde erupciones, hasta casos de cáncer en ambos sexos. La sintomatología son las propias verrugas generalmente en zonas íntimas, pero algunas personas no las presentan hasta muchos años después, donde, si el virus permanece por largo tiempo en el cuerpo genera cambios a nivel celular que derivan en cáncer [22].

Según LEY N° 26482, “establece la protección de la salud como un interés público y la responsabilidad del estado en su regulación, vigilancia y promoción”, en pocas palabras, permite a los ciudadanos la oportunidad de recibir atención de calidad para que incentiven la prevención y control de las enfermedades, asimismo el artículo 2° menciona, toda persona tiene derecho a exigir que los bienes destinados a la atención de la salud correspondan a las características y atributos indicados en su presentación y, a todas aquellas que acrediten su validación [23]. Accesibilidad a la salud es la capacidad de toda persona para ser atendidos sin

discriminación sociocultural o geográfica, de tal manera que reciban dicho servicio de manera óptima y sin restricciones, esto engloba todos los servicios de salud que promueve el estado [24].

IBM define las CNN como un tipo de modelo dentro del machine learning y constituyen la base de los algoritmos de deep learning. Estas redes están formadas por capas de nodos interconectados, que incluyen una capa de entrada, una o más capas ocultas y una capa de salida. Cada nodo posee un peso y un umbral asociados que determinan su activación, si la salida supera el umbral, el nodo transmite información a la siguiente capa, caso contrario, no envía datos. Estas procesan información tridimensional, lo que las hace especialmente eficaces para tareas como clasificación de imágenes y reconocimiento de objetos. Este enfoque resulta fundamental en aplicaciones avanzadas de inteligencia artificial, donde la jerarquía de características aprendidas permite una interpretación automatizada de datos visuales [25]. En las primeras etapas, las capas identifican características básicas como bordes, texturas y cambios de luminosidad. A medida que la información avanza hacia capas más profundas, los filtros convolucionales capturan patrones cada vez más complejos y específicos, como formas geométricas o componentes distintivos de los objetos. El proceso se realiza mediante operaciones de convolución, donde filtros de diferentes tamaños recorren la imagen generando mapas de activación. Estos mapas, que resaltan las características detectadas, sirven como entrada para las siguientes capas. De esta manera, la red construye progresivamente una representación más abstracta y discriminativa de los datos, optimizando su capacidad para tareas como clasificación y detección de objetos [26].

Entre las herramientas empleadas en el desarrollo de CNN destaca Keras, una biblioteca de código abierto que facilita la creación, entrenamiento y evaluación de modelos de aprendizaje profundo. Su diseño modular y su interfaz de alto nivel permiten estructurar las redes de manera más intuitiva [27]; asimismo, complementos como el optimizador Adam contribuye al ajuste eficiente de los parámetros mediante la actualización adaptativa de los pesos, mejorando el proceso de aprendizaje y convergencia [28], mientras que la capa GlobalAveragePooling2D cumple con la función de reducir las dimensiones espaciales de los mapas de características generados por las capas convolucionales, promediando cada mapa con un solo valor [29].

TensorFlow es una plataforma desarrollada por Google, pensada principalmente para el desarrollo y ejecución de modelos de aprendizaje profundo. Proporciona un entorno flexible que permite implementar cálculos mediante grafos de flujo de datos, lo que facilita optimizar operaciones matemáticas que mantienen cierto grado de complejidad [30]. Incluido en este entorno se puede encontrar el modelo MobileNet V2, desarrollada principalmente para ser usada en dispositivos con limitaciones en sus recursos, como los dispositivos móviles y sistemas embebidos.

Esta red neuronal emplea un diseño basado en bloques inversos residuales y convoluciones separables en profundidad, lo que se traduce a una significativa reducción en la cantidad de parámetros y operaciones sin comprometer la precisión [31].

Matplotlib está orientada a la generación de gráficos y representaciones visuales de datos, usada en entornos científicos para analizar el comportamiento y desempeño de los modelos [32]. Numpy constituye una herramienta esencial para la realización de cálculos numéricos, ofrece estructuras de datos basadas en arreglos multidimensionales y operaciones matemáticas avanzadas, que son fundamentales en el procesamiento de información dentro de las redes neuronales [33]. Complementariamente, Seaborn es una biblioteca de visualización estadística que permite la representación de las relaciones entre variables y patrones presentes en los datos [34].

Dropout es una técnica que se define como un método de regularización que congela aleatoriamente cierta cantidad de neuronas durante el entrenamiento, lo que reduce ampliamente el riesgo de posibles sobreajustes, mejorando la capacidad de generalización [35]. Los pesos (Weights) representan parámetros que son ajustables en la red neuronal que determinan la fuerza entre las conexiones neuronales, siendo su actualización un factor clave en el proceso de aprendizaje [36].

Desde una postura innovadora la integración de arquitecturas como ResNet50 y VGG19 ha demostrado mejoras notables en la clasificación de imágenes cervicales, logrando una precisión del 94.53%. Las técnicas de aprendizaje por transferencia permiten utilizar modelos previamente entrenados con grandes volúmenes de datos y adaptarlos a contextos específicos donde los datos disponibles son más limitados [37].

Desde una postura educativa, se afirmó que el trabajo debe ser evaluado con cautela, debido a limitaciones metodológicas clave. La muestra reducida y la falta de validación externa comprometen la robustez estadística y la generalización de los resultados. Además, la simplificación del diagnóstico a una clasificación binaria, ignora la gradación de enfermedades, lo cual es crítico en la práctica clínica. Aunque la alta sensibilidad y especificidad reportadas son alentadoras, estos indicadores podrían estar sobreestimados, si es que el modelo no fue probado en condiciones reales con variables no controladas como la iluminación, calidad de imagen, etc. Para que la herramienta tenga un impacto real, se requiere: (1) ampliar el dataset con imágenes representativas de diversas poblaciones, (2) incorporar etapas de validación clínica independiente y, (3) evaluar el rendimiento frente a métodos diagnósticos convencionales. En conclusión la investigación es valiosa como prueba de concepto, pero debe superar desafíos metodológicos y clínicos antes de considerarse una herramienta confiable en entornos médicos [38].

Desde una postura tecnológica, se han consolidado como una herramienta clave en diversos campos, especialmente en el sector médico, donde su implementación en sistemas de diagnóstico asistido por IA, ha demostrado una alta efectividad. Estas arquitecturas han sobresalido en la detección y clasificación de anomalías, como el cáncer de piel, gracias a su capacidad para identificar patrones específicos en imágenes médicas. Sin embargo, el proceso enfrenta desafíos técnicos, como la heterogeneidad en las características de las imágenes y la complejidad de los distintos tipos de cáncer. En este contexto, se examinan diversos modelos de CNN entrenados con datasets de acceso abierto, donde mediante el uso de técnicas han optimizado la precisión y eficiencia en el diagnóstico automatizado, reforzando su valor como solución innovadora en la medicina moderna [39].

Desde una postura médica, han transformado el análisis de imágenes médicas, al permitir la identificación automática de características jerárquicas a partir de grandes y complejos conjuntos de datos (imágenes). La evolución de diversas arquitecturas aplicadas al análisis de imágenes médicas, resaltan su impacto y eficacia en múltiples áreas de la medicina como oncología, neurología, cardiología, neumología, oftalmología, dermatología y ortopedia. Además, se abordan los desafíos específicos que presenta la imagenología médica y se describen las tendencias actuales y futuras líneas de investigación. Esta revisión tiene como objetivo ser una herramienta útil para investigadores y profesionales de la salud pública, contribuyendo a mejorar la precisión diagnóstica y la atención al paciente en diversos contextos clínicos [40].

II. Metodología

Este trabajo se delimitó a una investigación de tipo aplicada, ya que estuvo orientada al uso de una red neuronal convolucional basada en un modelo de deep learning, para la identificación de lesiones referentes al VPH en Piura. El propósito fue aplicar conocimientos científicos de manera práctica, dando como resultado una herramienta que permitió identificar la enfermedad mediante la comparación de imágenes referentes. De este modo se pretendió añadir veracidad en la primera fase de diagnóstico, facilitando así la adopción de medidas preventivas y minimizar el riesgo de que llegue a etapas cancerígenas. Maldonado, destaca que la investigación aplicada está vinculada al ámbito experimental, donde se definen metodologías y se elaboran prototipos susceptibles de validación. Esto conlleva un abordaje técnico-práctico orientado a resolver desafíos concretos mediante enfoques sistemáticos [41].

El enfoque considerado para esta investigación fue de tipo mixto, ya que se integraron métodos cuantitativos y cualitativos, con el objetivo de analizar de forma integral la efectividad del modelo MobileNet V2, en el proceso diagnóstico del Virus del Papiloma Humano. En primera instancia, se emplearon técnicas de tipo cuantitativas como la precisión y sensibilidad, que permitieron medir la efectividad del modelo frente a la detección de la patología, a partir de imágenes clínicas. Asimismo, se aplicó el índice Kappa de Cohen para evaluar la concordancia intermétodo, es decir, el grado de acuerdo entre el diagnóstico realizado por el médico especialista y el generado por la CNN, con resultados interpretados cuantitativamente. Complementariamente, se empleó una lista de cotejo que fue el método de recolección referente a las percepciones de los médicos sobre la aceptación del aplicativo, considerando los indicadores grado de utilidad percibida y nivel de confianza. Estas opiniones fueron analizadas mediante el índice de Fleiss Kappa, permitiendo evaluar el nivel de acuerdo entre los evaluadores, lo que seguidamente fue interpretado desde un enfoque cualitativo. Los métodos mixtos en una investigación se centran entre los enfoques cualitativos y cuantitativos, para fomentar la complejidad de los fenómenos estudiados. Estos procedimientos están respaldados por una serie de teorías que fundamentan su uso y proporcionan un marco conceptual concreto [42].

El diseño fue de tipo experimental, ya que se evaluó la efectividad de la red neuronal en la detección oportuna de VPH. Este estudio se enfocó en evaluar la influencia de un algoritmo de aprendizaje profundo, ante la detección de esta patología, analizando cómo la variable independiente (redes neuronales convolucionales), influyó sobre la variable dependiente (diagnóstico de VPH). A raíz de imágenes patológicas referentes a la enfermedad, la investigación evaluó cómo el modelo MobileNet V2 puede mejorar la primera fase del diagnóstico en la región de Piura 2025. El diseño de la investigación experimental se basa en la

manipulación deliberada de una o más variables independientes que tuvieron influencia en una o más variables dependientes [43].

El alcance fue de tipo explicativa, porque la CNN permitió detectar VPH mediante las lesiones cutáneas y mucosas, se estableció el analizar en qué medida el uso de esta tecnología puede contribuir a un diagnóstico rápido y preciso. El estudio se enfocó en imágenes de VPH asociadas a pacientes en Piura, evaluando la efectividad del diagnóstico que ésta pueda emitir, adaptando la arquitectura MobileNet V2 a datos locales. La investigación explicó el porqué este modelo pudo ser efectivo, teniendo en cuenta factores de calidad de imagen y el tamaño de la base de datos. Además se discutió de manera secundaria los desafíos prácticos de implementación en hospitales.

En el presente estudio se estableció como variable independiente “redes neuronales convolucionales”, ya que las CNN forman parte del subconjunto dentro de machine learning y se sitúan en el corazón de los algoritmos de deep learning, se distinguen de otras por el rendimiento superior al momento de clasificar y comparar objetos, ya que cuentan con entradas de señal del tipo imagen, voz o audio [12]. El modelo se evaluó con el uso de métricas cuantitativas adaptadas a este contexto, las cuales dieron resultados precisos y comparables. Estas métricas proporcionaron valores del tipo numérico dentro de un rango predefinido. Entre estas métricas se calcularon la sensibilidad, que determinó la proporción de evaluaciones positivas reales respecto a las estimaciones positivas y, la precisión, que indicó el número de aciertos entre las predicciones del modelo [12]. Aquí se contempla la dimensión efectividad del modelo que comprende dos métricas: (1) Precisión (Accuracy): Se determinó mediante TP (Verdaderos Positivos), TN (Verdaderos Negativos), FP (Falsos Positivos), FN (Falsos Negativos), esto dio como resultado la capacidad general de la CNN en determinar si es o no VPH, en pocas palabras, qué tan frecuentemente la CNN clasificó correctamente los casos positivos. (2) Sensibilidad: contempla TP (Verdaderos Positivos), FN (Falsos Negativos), dando como resultado la capacidad de la CNN para identificar correctamente casos verdaderamente positivos de VPH.

La escala de medición fue la razón, dado que, mediante los parámetros de precisión y sensibilidad, se facilitó la interpretación objetiva referente a la efectividad del modelo en la identificación del VPH. El análisis se centró en determinar si la red neuronal convolucional detecta correctamente los verdaderos casos positivos, siendo de 0% a 100%, el nivel de sensibilidad. Asimismo, la precisión permitió evaluar qué proporción de las predicciones realizadas por el modelo resultaron ser correctas, también de un rango de 0% a 100%.

La variable dependiente “diagnóstico de VPH”, estuvo referida a la capacidad de identificar a las personas que padecen una enfermedad de las que no la padecen [38]. Se evaluó con métricas cuantitativas y cualitativas, contó con criterios

previamente establecidos que permitieron recopilar datos para este estudio. La información obtenida fue registrada para su posterior análisis estadístico, facilitando su interpretación al medir la validez en el proceso de diagnóstico, además de medir el grado de aceptación del médico hacia el aplicativo móvil. Las dimensiones evaluadas fueron: (1) Validez diagnóstica, mediante el indicador: Número de casos resueltos correctamente por un experto, tomando en cuenta el total de diagnósticos correctamente clasificados como positivos por un doctor con el método tradicional. (2) Tiempo promedio de obtención de resultados de VPH: Se estudió el tiempo que le toma a un doctor emitir un diagnóstico con el método tradicional. (3) Concordancia intermétodo: Hizo referencia al grado de concordancia entre los diagnósticos brindados por la CNN y el emitido por un médico utilizando el método tradicional. Como segunda dimensión está la aceptación por el médico, que engloba los indicadores: (4) Grado de utilidad percibida: Se investigó si el aplicativo móvil resulta ser útil para el médico en la primera fase del diagnóstico. (5) Nivel de confianza: demostró si los diagnósticos dados por la CNN generan que el médico los tome como verídicos.

La concordancia intermétodo obtenida a partir de la comparación entre el método tradicional y los diagnósticos brindados por la CNN, evidenciaron la existencia de una relación consistente que respalda la consideración del aplicativo móvil como una herramienta confiable. Para su interpretación, se utilizó una escala de -1 a 1, donde < 0.00 representa concordancia totalmente pobre, 0.00 - 0.20 concordancia igual al azar, 0.21 - 0.40 concordancia baja pero superior al azar, 0.41 - 0.60 concordancia aceptable, 0.61 - 0.80 alto grado de concordancia y 0.81 - 1.00 concordancia casi perfecta o perfecta [44].

La población total estuvo conformada por 500 pacientes. Con base en esta cifra, se procedió a calcular el tamaño de la muestra utilizando la fórmula para poblaciones finitas, lo que permitió garantizar un nivel adecuado de representatividad, considerando un nivel de confianza correspondiente al 95% y un margen de error aceptable, con el propósito de asegurar la validez y precisión de los resultados obtenidos en este estudio. Por lo tanto, el tamaño muestral mínimo se estableció en 218 pacientes, de los cuales fueron divididos en 2 grupos de 73 y uno de 72. En dichos grupos se incluyeron pacientes que presentaban diferentes tipos de lesiones como aquellos que no padecían, con el fin de realizar una evaluación comparativa más precisa y confiable. Para el registro de información clínica se emplearon imágenes obtenidas de 3 fuentes, algunas proporcionadas por los médicos tratantes, otras fueron tomadas directamente mediante el uso del aplicativo y unas cuantas obtenidas de fuentes confiables, lo que dio garantía a una documentación homogénea y estandarizada. De esta manera, se completó satisfactoriamente el periodo de prueba, validando su utilidad en el contexto clínico.

No obstante, se utilizaron un total de 1294 imágenes médicas, organizadas en 3 etiquetas principales:

- Etiqueta 1: **VPH**, contando con 386 imágenes referentes a la patología. Estas imágenes contuvieron lesiones ubicadas en zonas íntimas, mucosa bucal y en áreas del cuerpo como manos y pies.
- Etiqueta 2: **SIN VPH**, integrada por 197 imágenes, de las cuales fueron cicatrices, piel normal, manchas, pecas, lunares, acné, heridas causadas por cortes, moretones, vitíligo, zonas íntimas y mucosas sanas.
- Etiqueta 3: **OTRAS ETS**, dividida en 2 subetiquetas conformadas por VHS (Virus del herpes simple) con 381 imágenes y Sífilis con 330 imágenes, ambas con lesiones que se manifiestan en diferentes zonas del cuerpo.

Definir las etiquetas fue un aspecto clave para mejorar el rendimiento y generalización del modelo en la fase de entrenamiento. El término “población” es el total de personas, grupos, organizaciones, conductas u otras entidades que un estudio pretende analizar o sobre las que se desean obtener conclusiones generales [45].

Para la delimitación adecuada de la población de estudio, se aplicaron criterios de inclusión y exclusión. Dentro de los criterios de inclusión se consideraron los casos registrados durante los últimos tres años, comprendidos entre 2022 y 2025. Asimismo, se incluyeron pacientes con edades desde los 18 a 60 años, rango considerado de mayor incidencia y relevancia clínica para esta investigación. Pacientes cuyas lesiones fueron compatibles con el VPH, a tal punto de presentar crecimientos de color piel o grises. Pacientes que padecían de afecciones cutáneas sin diagnosticar, pero que el médico tratante al visualizar las lesiones estuvo seguro de que se trataba de alguna enfermedad compatible con las consideradas en este estudio. Imágenes de repositorios públicos que el médico experto validó como reales y compatibles con las enfermedades contempladas. Por otro lado, se excluyeron aquellos casos cuyo tiempo de infección no había dado las lesiones típicas de VPH, debido a que no presentaron manifestaciones visibles lo suficientemente desarrolladas para un análisis adecuado mediante el modelo de red neuronal convolucional. También se excluyeron pacientes cuyo proceso de diagnóstico no estuvo confirmado por laboratorio, siempre y cuando la enfermedad requería dicho proceso para su confirmación. Sumado a esto, imágenes clínicas de baja calidad. Pacientes que no autorizaron el uso del aplicativo móvil en su proceso de diagnóstico. Los criterios de inclusión contemplan características predefinidas, que son utilizadas en su gran mayoría para identificar a los sujetos que se incluyen en un estudio de investigación, mientras que los de exclusión, conforman criterios de selección o elegibilidad que se usan para excluir a la población objetivo en una investigación [46].

El tipo de muestreo que se aplicó para el presente estudio fue por conveniencia, lo que implicó que se seleccionarán aquellas imágenes médicas que estén fácilmente

disponibles y accesibles. Las imágenes que se utilizaron para el estudio no incluyeron datos personales del paciente y fueron recortadas lo máximo posible para enfocarse únicamente en la lesión presente. Este enfoque facilitó el acceso a los datos necesarios para el entrenamiento y validación de la red neuronal convolucional.

La unidad de análisis estuvo conformada por imágenes clínicas de pacientes con lesiones compatibles con VPH. Dichas imágenes fueron obtenidas de un repositorio público y con la colaboración de un especialista en el área, fueron validadas para asegurar su correcta clasificación en las etiquetas correspondientes. Adicionalmente, se contó con la contribución de médicos, quienes proporcionaron imágenes de pacientes destinadas exclusivamente a fines académicos, garantizando en todo momento la completa privacidad y anonimato. Todo este conjunto de datos fue empleado para entrenar la red neuronal convolucional, con el propósito de evaluar su efectividad en la primera etapa de diagnóstico referente a dicha patología. La unidad de análisis representa al sujeto, entidad o fenómeno que constituye el centro de estudio en una investigación. Esta puede abarcar desde individuos o comunidades que incluyen hasta instituciones, naciones, artefactos tecnológicos y otros elementos comparables [47].

Las técnicas de recolección de datos que se emplearon fueron la observación, revisión cruzada y matriz de confusión, técnica de evaluación referente al desempeño del modelo. Mediante la observación, se recopiló información relevante sobre la efectividad de la red neuronal convolucional en un entorno clínico simulado. Asimismo, se empleó una lista de cotejo con una escala de valoración con dos categorías, diseñada para recopilar la percepción de los médicos evaluadores con respecto al aplicativo que integra redes neuronales convolucionales.

Para medir la concordancia intermétodo entre el proceso de diagnóstico tradicional vs el usando CNN, se aplicó el coeficiente Kappa de Cohen, que permitió evaluar el acuerdo entre ambos métodos. Del mismo modo, para analizar la concordancia de opiniones médicas hacia el aplicativo, se empleó el coeficiente Kappa de Fleiss para determinar el nivel de acuerdo entre múltiples evaluadores al clasificar en índices de categorías nominales.

Tabla 1. Variable 1: Diagnóstico de VPH

Indicadores	Técnica	Instrumento
Número de casos resueltos correctamente por un experto	Observación estructurada	Ficha de observación
Tiempo promedio de obtención de resultados de VPH		
Concordancia intermétodo (Índice Kappa de Cohen)	Revisión cruzada	Matriz de confusión de 2x2
Grado de utilidad percibida	Observación sistemática	Lista de cotejo
Nivel de confianza		

Tabla 2. Variable 2: Redes neuronales convolucionales

Indicadores	Técnica	Instrumentos
Precisión	matriz de confusión	Fórmula de accuracy
Sensibilidad		Fórmula de sensibilidad

Como método de análisis de datos en la fase de término referente a la recopilación de los mismos mediante los instrumentos, se evaluó la efectividad del diagnóstico que tuvo lugar en la primera fase, buscando el cumplimiento de los objetivos establecidos. Para la evaluación estadística, se utilizaron herramientas de software como lo es IBM SPSS Statistics, que permitió analizar la concordancia intermétodo y la coincidencia de opiniones entre los especialistas, considerando las dimensiones de validez diagnóstica, aceptación por el médico y efectividad del modelo.

El desarrollo del estudio abordó fundamentos éticos que garantizan la validez y confidencialidad de los datos utilizados para el entrenamiento. Estos datos, obtenidos con la debida autorización de las autoridades correspondientes, fueron

empleados para evaluar la efectividad de la primera fase diagnóstica. Se aseguró la total confidencialidad de las imágenes médicas utilizadas, respetando en todo momento la privacidad de los pacientes y cumpliendo con los protocolos éticos y normativas vigentes respecto a investigaciones en salud. De este modo, se garantizó que el estudio no solo fuera técnicamente riguroso, sino también éticamente responsable. Además se utilizaron herramientas de búsqueda para fundamentar la conceptualización de ciertos temas necesarios para el estudio. Sin embargo, la prioridad fue la originalidad de la información recopilada, para ello se consultó en fuentes como Scopus, SciELO, ScienceDirect, Mendeley, etc, tomando en cuenta los derechos de autor, citando adecuadamente todas las fuentes en formato IEEE. Además se utilizó turnitin para detectar la originalidad del proyecto y evitar similitud o plagio.

III. Resultados

En esta sección se presentan los resultados obtenidos a partir de la implementación del aplicativo móvil desarrollado, el cual integra el modelo MobileNet V2 para la detección del virus del papiloma humano. Los hallazgos evidencian el potencial de esta tecnología para optimizar los procesos diagnósticos, aportando información relevante que respalda su propósito, el cual es servir como una herramienta para el personal médico.

Objetivo específico 1: Medir la sensibilidad de la CNN al identificar correctamente las lesiones del VPH.

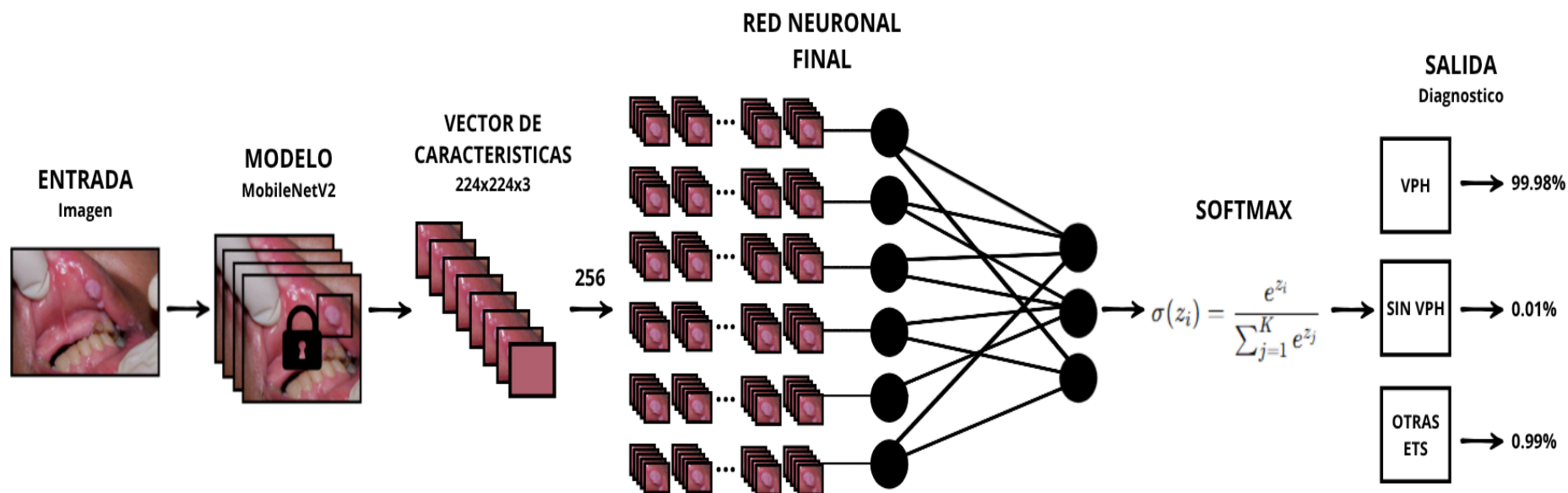
Los resultados mostraron que la CNN alcanzó un buen porcentaje de aciertos, siendo de un 100%, confirmando su capacidad de identificar en su totalidad los casos positivos de VPH, generando la inexistencia de falsos negativos. La sensibilidad al ser de suma relevancia en este objetivo, evidencia el buen desempeño del modelo al momento de clasificar la patología; de manera complementaria, la especificidad obtenida alcanza el 97.33%, confirmando así que el modelo es confiable en la identificación de casos negativos. Estos resultados reflejan un rendimiento sobresaliente en la detección de casos positivos y consistente en la exclusión de negativos.

Esto en base a los resultados obtenidos en la segunda fase, donde se puso a prueba el aplicativo móvil en un entorno clínico simulado. Esto, a su vez, refleja que el modelo escogido y el entrenamiento realizado en la primera fase, cumple con los criterios de sensibilidad y especificidad.

Objetivo específico 2: Elegir un modelo adecuado de CNN para procesar imágenes reales de la patología.

En esta investigación fue seleccionado como base el modelo MobileNet V2, preentrenado con un conjunto de datos correspondientes a ImageNet, ya que está comprobada su eficiencia en tareas de clasificación, especialmente en escenarios donde es requerido un balance entre rendimiento y consumo, lo que es ideal para que en combinación con el Framework Flutter, pudiera crearse un aplicativo que integre esta arquitectura, para de ese modo correr en prácticamente cualquier dispositivo android con versión mayor o igual a 10. El modelo fue adaptado con el uso de la estrategia transfer learning, incorporando una arquitectura personalizada en la parte superior que fue diseñada específicamente para la clasificación de imágenes asociadas a las patologías escogidas en este estudio.

Figura 1. Flujo del modelo CNN correspondiente a la clasificación de imágenes



La arquitectura final está compuesta por dos secciones diferenciadas, el modelo base cuyas capas convolucionales profundas fueron congeladas para de ese modo conservar el conocimiento previamente adquirido en ImageNet, posteriormente mediante fine-tuning las últimas 30 capas sean descongeladas para adaptar las características de bajo nivel y la parte personalizada que compone dos capas densas ocultas y una capa de salida, puedan usarse en el proceso de análisis de imágenes.

La primera capa densa que está oculta cuenta con 256 neuronas y emplea la función de activación ReLU, permitiendo transformar la salida del bloque GlobalAveragePooling2D en una representación de menor dimensionalidad, con el objetivo de reducir los riesgos correspondientes al sobreajuste, también se implementó una capa Dropout (0.3), que desactiva aleatoriamente el 30% de las conexiones neuronales durante el entrenamiento, para asegurar una mayor capacidad de generalización, adicionalmente esta capa incorpora regularización L2 ($\lambda=0.01$), cuya función es penalizar los pesos elevados en la función de pérdida. Posteriormente una segunda capa que mediante Dropout con una tasa de 0.5 (50% de desactivación), proporciona una regularización más agresiva para que la segunda capa densa de 128 neuronas, también con activación ReLU, aporte una refinación a lo aprendido, para reducir la dimensionalidad de manera progresiva, estableciendo una base sólida para posterior clasificación de imágenes.

La capa de salida incorpora 3 neuronas con activación mediante softmax, permitiendo generar una correcta distribución probabilística ante las 3 etiquetas definidas en este estudio, las cuales son VPH, Sin VPH y Otras ETS.

Este diseño permitió consolidar un modelo final con dos capas ocultas densas y funciones de activación no lineal que estuvieron complementadas con la activación softmax en su salida, asegurando que su capacidad de aprendizaje correspondientes a patrones complejos como lo es la interpretación probabilística de los resultados esperados, vaya de la mano con lo que se buscó evidenciar. La inclusión de la capa Dropout se tradujo en una reducción bastante significativa respecto a la varianza, lo que es particularmente relevante considerando el tamaño limitado del conjunto de datos.

En desempeño MobileNet V2, resultó ser adecuado para el objetivo planteado, dado que cuenta con una arquitectura ligera que facilita un correcto entrenamiento, sin sacrificar capacidad de representación; el añadir un conjunto de datos personalizados permitió ajustar el modelo para analizar imágenes médicas asociadas a este estudio, generalizando un equilibrio entre precisión y capacidad de generalización.

En conclusión la arquitectura seleccionada es realmente una buena alternativa

para el momento de procesar y clasificar imágenes reales de la patología, aprovechando el conocimiento previamente aprendido por MobileNet V2 y adaptarlo a un contexto clínico, confirmando la eficiencia del aprendizaje por transferencia como una herramienta en el ámbito biomédico, donde los datasets suelen ser limitados y la necesidad de modelos robustos es necesariamente prioritaria. Asimismo al ser ligero y optimizado, se resalta su compatibilidad con aplicativos móviles, permitiendo la integración de este modelo para ser convertido en una herramienta portátil, esta característica facilita ampliamente su integración en entornos clínicos con tecnologías que son variables, potenciando su accesibilidad, permitiendo que profesionales de la salud dispongan de un apoyo en los procesos diagnósticos que realizan, resaltando la aplicabilidad del modelo en entornos clínicos reales.

Objetivo específico 3: Evaluar la calidad de la CNN considerando las dimensiones de validez diagnóstica y aceptación por el médico.

Para entender cómo es que un modelo de CNN resulta ser bueno en un aspecto crucial de la detección de enfermedades en el ámbito clínico, hay que remontarse a comprender el porqué de la necesidad de implementar nuevas tecnologías, para ello se observó el tiempo que toma diagnosticar a una persona que padece de esta patología, obteniendo estos resultados:

Tabla 3. Tiempo de resolución con el método tradicional

CASOS	TIEMPO DE RESOLUCIÓN	TIEMPO EN MINUTOS
Caso 1 (Mujer, 32 años)	20 min	20 min
Caso 2 (Hombre, 27 años)	42 min	42 min
Caso 3 (Mujer, 45 años)	3 h 40 min	220 min
Caso 4 (Mujer, 51 años)	1 día 16 horas	2,400 min
Caso 5 (Hombre, 40 años)	45 min	45 min
Caso 6 (Hombre, 34 años)	16 días	23,040 min
Caso 7 (Hombre, 56 años)	3 semanas	30,240 min
Caso 8 (Mujer, 29 años)	2 horas 20 min	140 min
Caso 9 (Mujer, 38 años)	1 hora 13 min	73 min
Caso 10 (Hombre, 49 años)	3 semanas	30,240 min
Caso 11 (Mujer, 44 años)	1 semana	10,080 min
Caso 12 (Hombre, 36 años)	1h 20 min	80 min
Caso 13 (Mujer, 32 años)	9 días	12,960 min

Las personas para poder recibir un tratamiento adecuado, deben esperar los resultados de sus pruebas, las cuales muchas veces pueden tomar semanas, dependiendo de la zona donde se realicen; es cierto que la falta de herramientas médicas que optimicen el proceso de diagnóstico es muy necesario para evitar tiempos de espera largos y por consecuencia iniciar con el tratamiento correspondiente. Por ello se calculó el promedio general del tiempo que toma recibir un diagnóstico. Para llegar al resultado final se debe hacer la conversión del promedio a días, horas y minutos, resultando en datos que se aplicaron para determinar el tiempo promedio, con el fin de contar con información más apegada a la realidad.

Entonces el promedio general fue de 5 días, 20 horas y 29 min, cada caso atendido por un médico utilizando el método tradicional demora casi 6 días en resolverse, lo que traducido en términos prácticos simboliza que el proceso es algo lento a comparación de un tiempo óptimo de diagnóstico que podría brindar una CNN. Bajo esta premisa, se procedió con el entrenamiento de la red neuronal, donde lo primordial fue que pueda lograr captar con precisión las características propias de las lesiones causadas por el VPH.

El entrenamiento del modelo mantuvo parámetros que fueron fundamentales para que todo se haga de la mejor manera, las imágenes que fueron alojadas en las correspondientes etiquetas fueron redimensionadas a 224x224 píxeles, utilizando un batch size de 32 imágenes. Asimismo, fueron definidas 3 clases de salida que corresponden a las etiquetas relevantes del diagnóstico, la función de pérdida escogida fue la de Categorical Crossentropy, que es una de las más apropiadas para clasificación multiclase.

Para la preparación de los datos, se empleó la librería ImageDataGenerator de Keras, que ayudó a facilitar no solo la normalización de los valores de píxeles entre 0 y 1, sino también la aplicación de técnicas de aumentación de datos, dichas transformaciones incluyen rotaciones, desplazamientos, cizallamientos, zooms y volteos horizontales con el propósito de incrementar la diversidad de los ejemplos de entrenamiento, reduciendo el sobreajuste. TensorFlow y Keras fueron los pilares para la construcción, entrenamiento y despliegue del modelo, Numpy fue utilizado para la manipulación de datos y operaciones matemáticas, mientras Matplotlib y Seaborn jugaron un papel muy importante en la visualización de métricas como lo son la precisión, pérdida y el recall.

Se hizo uso de la estrategia de entrenamiento en dos fases:

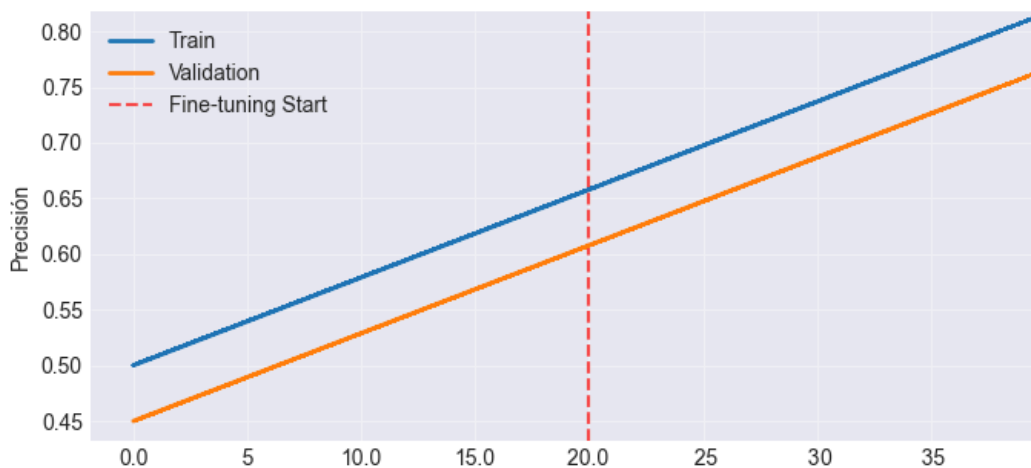
- Fase 1: 20 épocas con el modelo base congelado y un learning rate de 0.001.
- Fase 2: Épocas restantes más Fine-tuning descongelando las últimas 30

capas con learning rate reducido a 1×10^{-5} .

Complementariamente con técnicas de regularización que implementaron Class Weights balanceados para compensar el desbalance existente en el dataset, además de tres capas Dropout de 0.3, 0.5 y 0.4 para apagar aleatoriamente un porcentaje de neuronas durante el entrenamiento, con el fin de que el modelo no dependa de neuronas específicas y de ese modo mejore la generalización, con la Regularización L2 ($\lambda=0.01$) en las capas densas; para finalizar se configuraron callbacks de EarlyStopping (patience=10), para que si no hay una mejora significativa se detenga y restaure los pesos de la mejor época, complementando con ReduceLROnPlateau para reducir la tasa de aprendizaje si la métrica deja de mejorar y ModelCheckpoint para observar la retrollamada cuando guarda los pesos optimizando el entrenamiento.

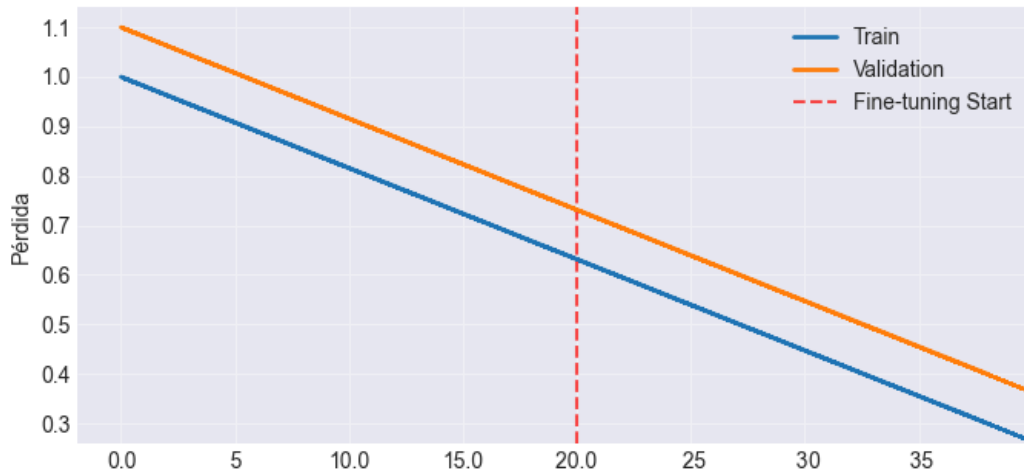
Todo esto dio como resultado un modelo que fue guardado de 2 formas, el tradicional HDF5 (.h5) de Keras y el TensorFlow Lite (.tflite), los cuales están optimizados para la implementación en aplicativos móviles.

Figura 2. Precisión del modelo en el entrenamiento



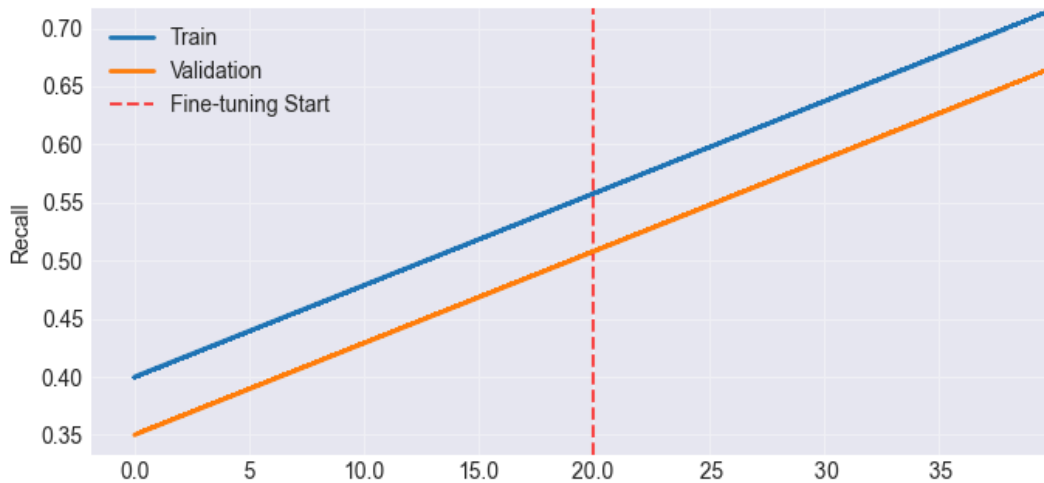
La figura representa un entrenamiento exitoso con un progreso constante y sin evidencia de sobreajuste, durante las primeras 20 épocas la precisión de entrenamiento más la validación, aumenta de manera paralela desde un aproximado de 47% hasta el 62% con una brecha mínima entre ambas. Al iniciar el fine-tuning en la época 20 el modelo continúa mejorando, la distancia reducida entre ambas líneas confirman que las técnicas de regularización fueron efectivas para prevenir sobreajustes, esto permitió que el modelo generalice adecuadamente datos no vistos durante el entrenamiento.

Figura 3. Pérdida del modelo en el entrenamiento

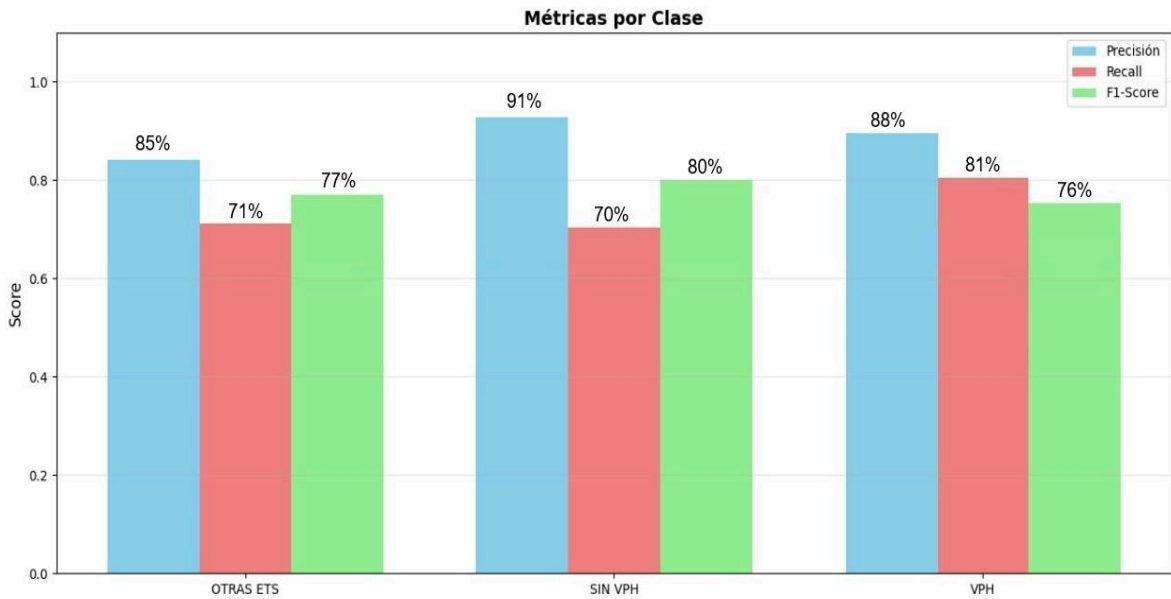


Esta figura confirma el correcto aprendizaje del modelo, tanto el entrenamiento y validación descienden consistentemente desde valores iniciales altos hasta 0.3 y 0.4 aproximadamente al final de cada entrenamiento, demostrando que el modelo optimizó correctamente la función de pérdida sin memorizar durante el entrenamiento.

Figura 4. Recall del modelo en el entrenamiento



El Recall muestra la capacidad de detectar casos positivos, mejorando progresivamente, lo que evidencia un crecimiento en la sensibilidad al identificar los casos positivos sin comprometer la generalización.

Figura 5. Métricas por etiquetas antes de probar en entorno clínico simulado

El desempeño por clase revela un modelo muy balanceado con ligeras variaciones, la etiqueta SIN VPH alcanza la mayor precisión y mejor F1-Score, mientras que VPH muestra el recall más alto y OTRAS ETS mantiene sus métricas balanceadas. Este comportamiento es el fiel reflejo de la estrategia empleada con el modelo para priorizar la detección de lesiones relacionadas al VPH a costa de generar algunos falsos negativos en el proceso.

Para llegar al modelo final se entrenó varias veces MobileNet V2, empezando solo con el modelo base sin complementos, a medida que se fue profundizando con las técnicas de regularización, se fueron obteniendo resultados cada vez más lineales, lo que dejó como producto la última versión del modelo que se usó en la segunda fase de validación.

De esta manera, los resultados obtenidos permiten concluir con esta etapa, evidenciando un desempeño diferencial, para dar pase a la fase de prueba en un entorno clínico simulado, donde tras la aplicación y el análisis de casos clínicos utilizando el aplicativo móvil, se obtuvo la siguiente concordancia:

Tabla 4. Concordancia intermétodo CNN vs Método tradicional

Medidas simétricas					
		Valor	Error estándar asintóticos	T aproximada	Significación aproximada
Medidas de acuerdo	Kappa	,973	,019	11,756	,000
N de casos válidos		146			

El análisis de concordancia mediante el índice Kappa de Cohen, reveló un acuerdo casi perfecto entre las predicciones del modelo y los resultados respaldados por laboratorio; de 146 muestras válidas, el modelo clasificó correctamente 144, con únicamente 2 discordancias que se clasificaron como falsos positivos. El valor de Kappa es superior a 0.9, lo que indica un nivel de concordancia excelente según los criterios de Landis y Koch, validando estadísticamente la confiabilidad del uso del aplicativo móvil.

Tabla 5. Kappa general

Overall Kappa						
	Kappa	Asymptotic Standard Error	Z	P value	Lower 95% Asymptotic CI Bound	Upper 95% Asymptotic CI Bound
Overall	,814	,082	9,965	,000	,654	,974

Tabla 6. Kappa para categorías individuales

Kappas for Individual Categories					
Rating Category	Conditional Probability	Kappa	Asymptotic Standard Error	Z	P Value
0	,857	,814	,082	9,965	,000
1	,957	,814	,082	9,965	,000

Conforme al coeficiente Fleiss Kappa se logró un alto nivel de concordancia entre los médicos que participaron en la investigación, referente a sus opiniones sobre el

uso del aplicativo móvil como una herramienta de diagnóstico. Según la interpretación estándar de Landis y Koch, el valor 0.814 refleja una concordancia casi perfecta. Los resultados de la lista de cotejo evidencian que las percepciones sobre el grado de utilidad percibida y el nivel de confianza tienen consistencia, aunque existen algunas diferencias menores en ciertos criterios como lo es la integración del aplicativo en el flujo del trabajo o una posible sustitución del criterio médico.

Bien, el aplicativo que incluye el modelo MobileNet V2 presenta una calidad sobresaliente, lo que está evidenciado por su alta concordancia con el diagnóstico tradicional, además de su aceptación positiva por parte de los médicos, aunque con ciertas discrepancias. Estos resultados demuestran que la CNN posee validez en sus diagnósticos y es bien valorada como una herramienta de apoyo.

Objetivo General: Evaluar el nivel de efectividad de las redes neuronales convolucionales en el primer diagnóstico de VPH en Piura, 2025.

Tabla 7. Matriz de confusión CNN vs Método tradicional

	POSITIVO	NEGATIVO	TOTAL REAL
POSITIVO REAL	TP= 71	FN= 0	71
NEGATIVO REAL	FP= 2	TN= 73	75
TOTAL PREDICHO	73	73	146

Durante la fase de prueba, el modelo MobileNet V2 logró identificar correctamente 71 positivos reales de VPH, lo que resulta bastante alentador, ya que clasificó de manera correcta los casos que se introdujeron para su análisis; por otro lado, los casos negativos analizados fueron clasificados correctamente en 73 ocasiones, pero dentro de estos análisis, hubieron 2 ocasiones que resultaron en falsos positivos, estos errores se originaron debido a la presencia de una etiqueta que alojaba 2 subetiquetas con imágenes referenciales a herpes simple y sífilis, resultando en una mala clasificación de las mismas.

En este estudio se buscó determinar la efectividad de esta tecnología, para llegar a esa conclusión se aplicaron las siguientes fórmulas:

Precisión (Accuracy):

$$Accuracy = \frac{71+73}{71+73+2+0} \times 100 = \frac{144}{146} \times 100 = 98.63\%$$

Sensibilidad:

$$\text{Sensibilidad} = \frac{71}{71+0} \times 100 = \frac{71}{71} \times 100 = 100\%$$

El modelo logró llegar a una precisión del 98.63%, reflejando que la gran mayoría de casos fueron clasificados adecuadamente, con 2 errores de por medio en todo ese conjunto. Además, alcanzó una sensibilidad de 100%, no dejó pasar ningún caso positivo en el proceso sin detectar. Siendo más directos, el modelo casi siempre acierta, pero lo fundamental es que no deja escapar ningún caso positivo ante la presencia de una lesión originada por VPH o de aquellas incluidas en la etiqueta de OTRAS ETS.

Por otro lado, para complementar los dos indicadores previamente mencionados, se calculó también:

Especificidad:

$$\text{Especificidad} = \frac{73}{73+2} \times 100 = \frac{73}{75} \times 100 = 97.33\%$$

Valor Predictivo Positivo:

$$\text{VPP} = \frac{71}{71+2} \times 100 = \frac{71}{73} \times 100 = 97.26\%$$

Valor Predictivo Negativo:

$$\text{VPN} = \frac{73}{73+0} \times 100 = \frac{73}{73} \times 100 = 100\%$$

F1 Score (media armónica entre VPP y Sensibilidad):

$$F1 = \frac{2 \cdot (0.9726 \cdot 1.0)}{0.9726+1.0} = \frac{1.9452}{1.9726} = 0.986 \text{ (98.6\%)}$$

La especificidad evidencia que el modelo también es confiable, con un 97.33% al momento de reconocer casos negativos, aunque en 2 ocasiones clasificó erróneamente como positivos a VPH. Asimismo, el VPP con un 97.26% indica que al predecir un caso positivo, casi siempre es correcto en contraste con el VPN que refleja un 100% cuando no es ninguna de las enfermedades consideradas en el estudio. Finalmente, el F1 Score resultó con un valor de 98.6% que evidencia el equilibrio entre sensibilidad y precisión, reafirmando el alto desempeño del modelo, convirtiéndolo en una herramienta sólida para apoyar en la detección de VPH.

Expuestos estos resultados, MobileNet V2 demostró un nivel de efectividad muy alto, validando su utilidad como una herramienta de diagnóstico en Piura durante el año 2025, con un nivel de confiabilidad suficiente para ser considerada en futuros estudios.

IV. Discusión

Este estudio formuló como primer objetivo específico el medir la sensibilidad de la CNN al identificar correctamente las lesiones del VPH, que estuvo basado en los aportes de González y Quintero [1], quienes midieron la sensibilidad del modelo MobileNet V2 ante la identificación de lesiones orales, lo que ante distintos niveles de desempeño, indicó una alta sensibilidad en los casos positivos de ciertas enfermedades. Considerando aquello, en este estudio, una vez analizada la información, se encontró que efectivamente este modelo es capaz de identificar la patología con un amplio acierto cuando se le dieron a analizar imágenes correspondientes al VPH, esto gracias a la correcta segmentación en el proceso de entrenamiento.

Por otra parte, Mohamad y Suriani [11], quienes también midieron la sensibilidad en la detección de enfermedades de la piel mediante un aplicativo móvil, concuerdan con lo expuesto anteriormente. Los autores señalan que el modelo presenta un amplio nivel de detección en los casos positivos, aspecto que resultó importante para que confirmen la utilidad del sistema en la eficiente identificación de las enfermedades propuestas en el análisis; no obstante, evidencian leves deficiencias en algunos casos, lo que indica la necesidad de mejorar los procesos de entrenamiento para que logre diferenciar de mejor manera las lesiones con características visualmente similares.

En este sentido, lo evidenciado en el presente estudio, al considerar las recomendaciones planteadas por investigaciones precedentes, confirman que el modelo puede mejorar notablemente aplicando métodos que potencien la fase de entrenamiento, de este modo se tiene como resultado un modelo robusto, capaz de generar diagnósticos que son comparables de manera positiva con los emitidos por un especialista en el área.

Finalmente, la sensibilidad se mantuvo como un indicador consistente en los resultados obtenidos, debido a que realmente el modelo se desempeñó de manera óptima durante su periodo de prueba. Los complementos utilizados resultaron ser eficientes para hacer que MobileNet V2 lograra alcanzar un adecuado entrenamiento por transferencia y por consecuencia, pudiera considerarse como un aporte relevante al uso de modelos de Deep Learning en el sector salud, especialmente en la detección de la enfermedad abordada en el presente estudio.

El segundo objetivo específico es elegir un modelo adecuado de CNN para procesar imágenes reales de la patología, fundamentado en el estudio que realizaron Sharma *et al.* [37], donde analizaron diversos modelos de aprendizaje profundo en la identificación de cáncer cervical, dichos autores justificaron la elección de un modelo por resaltar en su bajo consumo de recursos computacionales, sin sacrificar la exactitud que pueda lograr en su fase de

entrenamiento. A partir de esa investigación y, al constatar que MobileNet V2 fue considerado como un modelo apropiado para investigaciones que buscan automatizar los procesos diagnósticos, se procedió a investigar y analizar su arquitectura, la cual posteriormente fue adaptada a los datos locales disponibles.

Debido a su arquitectura liviana y modular facilitó el proceso de entrenamiento, lo que concuerda significativamente con el estudio previo. El contar con una base de datos desbalanceada no es impedimento para considerar el modelo, por ello fue muy importante el anterior aporte, ya que evidenciaron que se pueden obtener resultados competitivos en tareas de clasificación referentes al cáncer cervical. Con base en toda esa información, fue complementado el entrenamiento mediante la incorporación de arquitecturas y técnicas de aprendizaje por transferencia, que fortalecen inmensamente la extracción de características relevantes de las imágenes locales. Los resultados obtenidos concuerdan con aquel estudio, ya que MobileNet V2 es una muy buena opción para integrar en un aplicativo móvil sin perder rapidez y precisión cuando se le da a analizar imágenes consideradas en la experimentación realizada.

A la postre, los hallazgos obtenidos confirman que el modelo es viable para implementar en entornos clínicos simulados, donde los recursos de procesamiento son limitados y las condiciones de pruebas son controladas, bajo el respaldo de un especialista que valida la precisión del aplicativo sin los riesgos que pueda presentar un contexto real.

Como tercer objetivo específico se consideró evaluar la calidad de la CNN considerando las dimensiones de validez diagnóstica y aceptación por el médico, que parte de la base obtenida del estudio realizado por García Cortegano [38], quien menciona que los entornos clínicos no pueden limitarse a métricas técnicas de desempeño, sino que deben integrar dimensiones clínicas y humanas para respaldar su utilidad en la práctica médica. Este enfoque concuerda con lo expuesto en la presente investigación, al permitir que el aplicativo móvil sea utilizado por médicos, de modo que puedan familiarizarse con su funcionamiento y emitir valoraciones sobre la precisión que este pueda tener, lo que facilita en cierta parte realizar un análisis integral sobre la calidad de la CNN.

Por otro lado, García Cortegano [38], aborda de forma explícita y ética la aceptación por el médico, enfatizando que el sistema no sustituye al juicio clínico, sino que actúa como una herramienta de apoyo en la fase preliminar de un diagnóstico, cuyos resultados deben estar corroborados. Esto va de la mano con el dataset utilizado en el proceso de entrenamiento, ya que el modelo para emitir resultados concordantes con los de un experto, sus etiquetas deben contener imágenes validadas por un médico especialista y todas sean referentes a lo que se desea diagnosticar, lo que da como resultado un modelo basado en autoridad diagnóstica del profesional de la salud. En base a esto, se contempló la

participación directa de médicos especialistas que con sus años de experiencia en la identificación de VPH, fueron garantes en la robustez de las etiquetas para evitar malas calificaciones en las mismas, lo que de este modo trasciende la propuesta brindada por este autor al operacionalizar la aceptación por el médico. Dicho enfoque permitió analizar la calidad de la CNN desde una doble perspectiva, la concordancia intermétodo, que permitió determinar el grado de coincidencia entre el diagnóstico brindado por el aplicativo y el que emite un médico experto, además del cómo estos diagnósticos son aceptados por el especialista, dándoles la libertad de si los consideran en la primera fase diagnóstica o no. Así, la investigación no se limita al modelo de la CNN, sino que evalúa su integración en un contexto bastante sensible como lo es un entorno clínico, en coherencia con los principios éticos que rigen en las instituciones de salud.

Sintetizando, la calidad de la CNN no se mide solo por la capacidad algorítmica que esta pueda alojar, sino que debe mantenerse alineada con la lógica clínica, incorporando el juicio y la experiencia médica como elementos esenciales de validación. En este sentido, la presente investigación deja en evidencia que la integración del criterio profesional en el diseño y evaluación de herramientas diagnósticas que integran modelos de redes neuronales en un diagnóstico asistido, no solo fortalece su validez, sino que garantiza su aceptación ética y funcional en los entornos clínicos, donde se necesita la mayor precisión y sensibilidad posible ante la identificación de lesiones referentes a la patología.

Para cerrar este capítulo se expone el objetivo general: Evaluar el nivel de efectividad de las redes neuronales convolucionales en el primer diagnóstico de VPH en Piura, 2025. Anteriormente los investigadores Lang *et al.* [16], sustentaron la efectividad que pueden presentar las redes neuronales convolucionales, gracias a la creciente evidencia científica que las pone como una herramienta diagnóstica, aplicaron un modelo C3D tridimensional preentrenado con el conjunto de datos Sports-1M para predecir el estado del VPH en pacientes con cáncer orofaríngeo, esto en base a imágenes de tomografías computarizadas, resultando con una sensibilidad de 75%, valor que refleja una validez diagnóstica comparable con métodos histológicos convencionales, pero con la ventaja de que no es invasivo, rápido y con un costo considerablemente bajo. Además, evidenciaron que los modelos de aprendizaje profundo con transferencia de conocimiento previo, son capaces de mejorar la generalización y robustez diagnóstica, incluso con pequeñas bases de datos.

Durante la fase de prueba computacional del presente estudio, a diferencia del modelo tridimensional C3D empleado por los autores mencionados, se utilizó la arquitectura MobileNet V2, una CNN bidimensional optimizada para presentar una buena eficiencia computacional. Gracias a los métodos utilizados se mejoró el proceso de entrenamiento, alcanzando en la etiqueta VPH una sensibilidad del 81%, valor que supera al reportado en el anterior estudio. Esta diferencia puede

atribuirse al uso de una arquitectura más ligera, al refinamiento de los procesos de aprendizaje y a la capacidad del modelo para conservar un equilibrio entre precisión y generalización con un conjunto de datos reducido.

Li *et al.* [15] reforzaron la superioridad de las CNN frente a los métodos tradicionales de diagnóstico, los cuales dependen muchas veces de la experiencia médica en la fase visual. Según los autores, este método depende de la extracción manual o visual de características, lo que incrementa la posibilidad de sesgos y errores humanos, mientras que las CNN aprenden de manera automática patrones morfológicos complejos. En coherencia con ello se procedió a la segunda fase de prueba, donde se puso en marcha la usabilidad del aplicativo móvil como una herramienta de apoyo. Lo obtenido en ese tiempo resultó ser alentador, resaltando lo expuesto anteriormente por los investigadores, la CNN puede constituir una herramienta eficaz para complementar la labor médica, esto a raíz de que la concordancia observada entre el método tradicional y el que se propuso en esta investigación, reafirma que el uso de estos modelos no busca reemplazar la evaluación clínica, sino el fortalecer la toma de decisiones en la primera fase diagnóstica, ofreciendo un soporte cuantitativo basado en métricas de desempeño, como la precisión, sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo y F1-Score, que dejan evidenciado el potencial de MobileNet V2 en la identificación de lesiones causadas por VPH en un contexto clínico simulado.

V. Conclusiones

MobileNet V2 constituye una buena solución tecnológica para apoyar el diagnóstico que puede emitir un médico. Los resultados obtenidos evidencian que la CNN posee un alto potencial para ser considerada en el ámbito clínico, siempre y cuando sea utilizada en conjunto con un aplicativo móvil.

La implementación de técnicas para la optimización y regularización, permitió alcanzar un desempeño sobresaliente sin comprometer la eficiencia computacional, lo que favoreció su correcto despliegue hacia el aplicativo móvil. A pesar de los distintos dispositivos usados para la instalación y uso del mismo, resultó notable comprobar cómo una app puede convertirse en una herramienta capaz de generar resultados diagnósticos comparables a los obtenidos por un profesional médico, en un tiempo significativamente menor, lo que consolida en cierta parte la aceptación de la herramienta pensada en este estudio.

Sintetizando, la investigación confirma que la aplicación de inteligencia artificial, a través de un modelo que es relativamente liviano como el que se utilizó, representa un significativo avance en la digitalización y automatización de procesos diagnósticos. Con los resultados presentados se sientan las bases para futuras investigaciones que se quieran orientar a este sector, ya sea para clasificar o identificar anomalías que puedan presentar los pacientes, consolidando así el potencial de estas tecnologías como apoyo complementario en la práctica médica.

En conclusión, el estudio deja en evidencia que la integración de MobileNet V2 con un aplicativo móvil constituye una herramienta válida para procesos de clasificación e identificación, es una alternativa viable, eficiente y adaptable al contexto clínico, aportando evidencia empírica sobre la existente posibilidad de combinar los criterios médicos y la precisión algorítmica de las CNN. La investigación reafirma que la opinión médica sigue siendo esencial, ya que una herramienta tecnológica no sustituye el juicio de un experto, sino que lo complementa, ofreciendo una segunda opinión de manera rápida. El enfoque ético y colaborativo es un precedente importante en la incorporación de modelos de CNN en la salud y abre la posibilidad de seguir buscando soluciones más accesibles y humanizadas, orientadas a mejorar la calidad de un diagnóstico.

VI. Recomendaciones

Se recomienda seguir poniendo en práctica el uso de MobileNet V2 para comprobar que se puede adaptar a múltiples entornos, además de siempre buscar la mejor resolución en imágenes para un correcto aprendizaje del modelo, en primera fase es crucial utilizar épocas mayores a 15, ya que si se busca alcanzar la mayor precisión y sensibilidad debe ir de la mano con arreglos pensados para buscar la mayor convergencia entre métricas de entrenamiento y validación, además de aumentar la cantidad de imágenes en las respectivas etiquetas, lo que permitirá que la red neuronal no se estanque en el proceso de aprendizaje, de este modo se permite aprovechar las técnicas de optimización y visualización que van de la mano con el monitoreo.

El hacer un buen seguimiento facilitará la identificación de los sobreajustes o posibles subentrenamientos que puedan afectar a la estabilidad del modelo. Con respecto al tamaño de la data, un incremento de las mismas ayudaría en gran parte al momento de extraer características relevantes, lo que mejoraría la precisión, sensibilidad y generalización. Los detalles de las zonas, ya sea sana o con lesiones, es bastante importante en procesos de aprendizaje por transferencia, por ende recortar solamente el área donde están estas mismas, resulta crucial para refinar la detección de patrones y evitar interpretaciones erróneas.

Con respecto al tiempo de prueba, es pertinente ampliarlo para evaluar el comportamiento del modelo en escenarios clínicos que presentan diversidad de detalles, porque aún falta mucho por descubrir. Ampliar la fase experimental no sólo consolidará la robustez del modelo, sino que también contribuirá al poder descubrir nuevas oportunidades de optimización y validación en este contexto. Sumado a esto, es esencial que las investigaciones futuras, se realicen usando el 100% de imágenes reales pertenecientes a pacientes del mismo sector o población de referencia. Esta recomendación permitirá adaptar de una manera más precisa el modelo a particularidades clínicas, demográficas y tecnológicas del entorno donde se quiera aplicar. La incorporación de etiquetas que almacenen imágenes locales contribuirá a minimizar posibles sesgos derivados del uso de repositorios públicos y a mejorar la capacidad de generalización del modelo en contextos reales.

Una línea que podría ser relevante para futuras investigaciones es el profundizar en la interpretabilidad que puede tener un modelo con técnicas como Grad-CAM, la cual permite visualizar las áreas de atención del algoritmo durante el proceso de clasificación. Esto con el fin de comprender mejor las decisiones diagnósticas que realiza la CNN, lo que puede facilitar la interpretación en los posibles errores que pueda cometer y pulir aquellos detalles. La interpretabilidad se puede constituir como un componente esencial para garantizar la transparencia y trazabilidad en el uso de esta tecnología.

Asimismo, se recomienda fortalecer las etapas de validación clínica, involucrando activamente a los médicos expertos, con el fin de evaluar de una manera más completa el desempeño de los modelos que consideren en futuras investigaciones. Además, se sugiere impulsar nuevas líneas de investigación que estén orientadas a explorar la dimensión ética del uso de la IA en el sector salud, esto en base para garantizar la transparencia, responsabilidad y confiabilidad de las decisiones asistidas por esta tecnología.

Finalmente, se recomienda ajustar periódicamente los hiperparámetros y explorar nuevas estrategias referentes a la regularización o arquitecturas que puedan complementar al modelo, esto con el fin de mantenerlo vigente ante la evolución de los frameworks y librerías de Deep Learning.

CONFLICTO DE INTERÉS

Los autores declaran no tener conflicto de interés

CONTRIBUCIÓN DE LA AUTORÍA

Abanto Sánchez, Walter Edgar: Conceptualización, Metodología, Software, Validación, Investigación, Redacción - borrador original, Redacción - revisión y edición.

López Salinas, Alexis Josué: Conceptualización, Software, Validación, Investigación.

Todos los autores han leído y aceptado la versión publicada del manuscrito

DECLARACIÓN DE APROBACIÓN DEL COMITÉ DE ÉTICA

Esta investigación fue revisada y aprobada por el Comité de Ética en Investigación de la Escuela de Ingeniería de Sistemas de la Universidad César Vallejo. Informe N.º 00091-2025/CEI-EIS, emitido el 11 de junio de 2025. El proyecto titulado "*Redes neuronales convolucionales para la efectividad del primer diagnóstico de VPH en Piura 2025*" obtuvo un dictamen favorable para su ejecución.

La revisión fue realizada por:

- Cieza Mostacero Segundo Edwin, Presidente (ORCID: 0000-0002-3520-4383)
- Vega Gavidia Edward Alberto, Miembro 04 (ORCID: 0000-0001-6716-581X)
- Tavera Ramos Anthony Paul, Miembro 08 (ORCID: 0000-0002-4159-930X)

DISPONIBILIDAD DE DATOS

Los datos generados y analizados durante esta investigación están disponibles previa solicitud razonable al autor correspondiente. Por razones éticas y de privacidad, los datos no pueden ser publicados de manera abierta.

Referencias

- [1] J.-D. González, y J. Quintero-Rojas, «Use of Convolutional Neural Networks in Smartphones for the Identification of Oral Diseases Using a Small Dataset», *Rev. Fac. Ing.*, vol. 30, n.º 55, mar. 2021, doi: 10.19053/01211129.v30.n55.2021.11846.
- [2] «EsSalud realizó simposio sobre oportunidades y retos de la inteligencia artificial en salud». Accedido: 13 de abril de 2025. [En línea]. Disponible en: <https://www.gob.pe/institucion/essalud/noticias/990134-essalud-realizo-simposio-sobre-oportunidades-y-retos-de-la-inteligencia-artificial-en-salud>
- [3] D. Patiño-Pérez *et al.*, «REDES NEURONALES CONVOLUCIONALES PARA DETECCIÓN DE RETINOPATÍA DIABETICA», *Ingenius Rev. Cienc. Technol.*, n.º 33, pp. 91-101, jun. 2025, doi: 10.17163/ings.n33.2025.08.
- [4] C. E. Belman-López, «Detection of COVID-19 and Other Pneumonia Cases using Convolutional Neural Networks and X-ray Images», *Ing. E Investig.*, vol. 42, n.º 1, abr. 2022, doi: 10.15446/ing.investig.v42n1.90289.
- [5] H. Catal Reis, «COVID-19 Diagnosis with Deep Learning», *Ing. E Investig.*, vol. 42, n.º 1, abr. 2022, doi: 10.15446/ing.investig.v42n1.88825.
- [6] M. A. Bravo-Ortíz *et al.*, «Clasificación de cáncer cervical usando redes neuronales convolucionales, transferencia de aprendizaje y aumento de datos», *Rev. EIA*, vol. 18, n.º 35, pp. 100-111, jun. 2021, doi: 10.24050/reia.v18i35.1462.
- [7] «Lesiones en los genitales externos: Enfoque diagnóstico y conducta que debe seguir el médico de la familia – DOAJ». Accedido: 18 de mayo de 2025. [En línea]. Disponible en: <https://doaj.org/article/2a397d80c7e242149b5a74112bd07693>
- [8] M. Llamas-Velasco y E. Ovejero-Merino, «Inteligencia artificial en el diagnóstico dermatopatológico», *Piel*, vol. 39, n.º 8, pp. 512-517, oct. 2024, doi: 10.1016/j.piel.2024.01.002.
- [9] Araujo JM, Ramon-Cotrino M, Quispe-Martinez R *et al.*, «Prevalence of HPV in a Peruvian Healthcare Network: A Descriptive Cross-Sectional Study», *Cancer Control*, feb. 2025, Accedido: 13 de abril de 2025. [En línea]. Disponible en: <https://journals.sagepub.com/doi/full/10.1177/10732748251318386>

- [10] C. Tapia *et al.*, «Distribución de genotipos de virus papiloma humano de alto riesgo en mujeres y hombres atendidos en una red asistencial privada en la Región Metropolitana, Chile», *Rev. Chil. Infectol.*, vol. 41, n.º 1, pp. 20-26, feb. 2024, doi: 10.4067/s0716-10182024000100020.
- [11] N. F. Mohamad y N. S. Suriani, «Skin Disease Classification using Convolutional Neural Network via Android Smartphone Application», *Evol. Electr. Electron. Eng.*, vol. 3, n.º 1, Art. n.º 1, jun. 2022. [En línea]. Disponible en: <https://publisher.uthm.edu.my/periodicals/index.php/eeee/article/view/6776>
- [12] R. S. Akash, R. Islam, S. S. I. Badhon, y K. T. Hossain, «CerviXpert: A multi-structural convolutional neural network for predicting cervix type and cervical cell abnormalities», *Digit. Health*, vol. 10, p. 20552076241295440, sep. 2024, doi: 10.1177/20552076241295440.
- [13] F. A. Mohammed, K. K. Tune, J. A. Mohammed, T. A. Wassu, y S. Muhie, «Early Cervical Cancer Diagnosis with SWIN-Transformer and Convolutional Neural Networks», *Diagnostics*, vol. 14, n.º 20, Art. n.º 20, ene. 2024, doi: 10.3390/diagnostics14202286.
- [14] X. Tan *et al.*, «Automatic model for cervical cancer screening based on convolutional neural network: a retrospective, multicohort, multicenter study», *Cancer Cell Int.*, vol. 21, n.º 1, Art. n.º 1, ene. 2021, doi: 10.1186/s12935-020-01742-6.
- [15] Y.-X. Li, F. Chen, J.-J. Shi, Y.-L. Huang, y M. Wang, «Convolutional Neural Networks for Classifying Cervical Cancer Types Using Histological Images», *J. Digit. Imaging*, vol. 36, n.º 2, Art. n.º 2, abr. 2023, doi: 10.1007/s10278-022-00722-8.
- [16] D. M. Lang, J. C. Peeken, S. E. Combs, J. J. Wilkens, y S. Bartzsch, «Deep Learning Based HPV Status Prediction for Oropharyngeal Cancer Patients», *Cancers*, vol. 13, n.º 4, p. 786, feb. 2021, doi: 10.3390/cancers13040786.
- [17] A. J. Ccopa Mamani y E. Y. Diaz Quiroz, «Sistema para el diagnóstico de cáncer de cuello uterino utilizando redes neuronales convolucionales con imágenes de citología en base líquida de Papanicolaou», abr. 2025, Accedido: 24 de abril de 2025. [En línea]. Disponible en: <https://cybertesis.unmsm.edu.pe/item/34b6eb6f-c082-4d41-8db0-17c7a8c2e439>
- [18] F. Eitel *et al.*, «Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional MRI using layer-wise relevance

- propagation», 18 de abril de 2019, *arXiv*: arXiv:1904.08771. doi: 10.48550/arXiv.1904.08771.
- [19] M. H. Aabidi, A. E. Makrani, B. Jabir, y I. Zaimi, «A Model Proposal for Enhancing Leaf Disease Detection Using Convolutional Neural Networks (CNN): Case Study», *Int. J. Online Biomed. Eng. IJOE*, vol. 19, n.º 12, Art. n.º 12, ago. 2023, doi: 10.3991/ijoe.v19i12.40329.
- [20] «¿Qué es y qué consecuencias trae el Virus del Papiloma Humano? - OPS/OMS | Organización Panamericana de la Salud». Accedido: 25 de abril de 2025. [En línea]. Disponible en: <https://www.paho.org/es/campanas/chile-tu-vida-importa-hazte-pap/que-es-que-e-consecuencias-trae-virus-papiloma-humano>
- [21] «Infección por VPH: La vacuna puede prevenir los tipos más comunes-Infección por VPH - Síntomas y causas», Mayo Clinic. Accedido: 25 de abril de 2025. [En línea]. Disponible en: <https://www.mayoclinic.org/es/diseases-conditions/hpv-infection/symptoms-causes/syc-20351596>
- [22] «Virus del papiloma humano». Accedido: 25 de abril de 2025. [En línea]. Disponible en: <https://medlineplus.gov/spanish/hpv.html>
- [23] «Ley N.º 26842». Accedido: 25 de abril de 2025. [En línea]. Disponible en: <https://www.gob.pe/institucion/minsa/normas-legales/256661-26842>
- [24] «Salud Universal - OPS/OMS | Organización Panamericana de la Salud». Accedido: 3 de mayo de 2025. [En línea]. Disponible en: <https://www.paho.org/es/temas/salud-universal>
- [25] «¿Qué son las redes neuronales convolucionales? | IBM». Accedido: 25 de abril de 2025. [En línea]. Disponible en: <https://www.ibm.com/es-es/think/topics/convolutional-neural-networks>
- [26] «¿Qué son las redes neuronales convolucionales?» Accedido: 25 de abril de 2025. [En línea]. Disponible en: <https://es.mathworks.com/discovery/convolutional-neural-network.html>
- [27] K. Team, «Keras documentation: Getting started with Keras». Accedido: 7 de octubre de 2025. [En línea]. Disponible en: https://keras.io/getting_started/
- [28] D. P. Kingma y J. Ba, «Adam: A Method for Stochastic Optimization», 30 de enero de 2017, *arXiv*: arXiv:1412.6980. doi: 10.48550/arXiv.1412.6980.

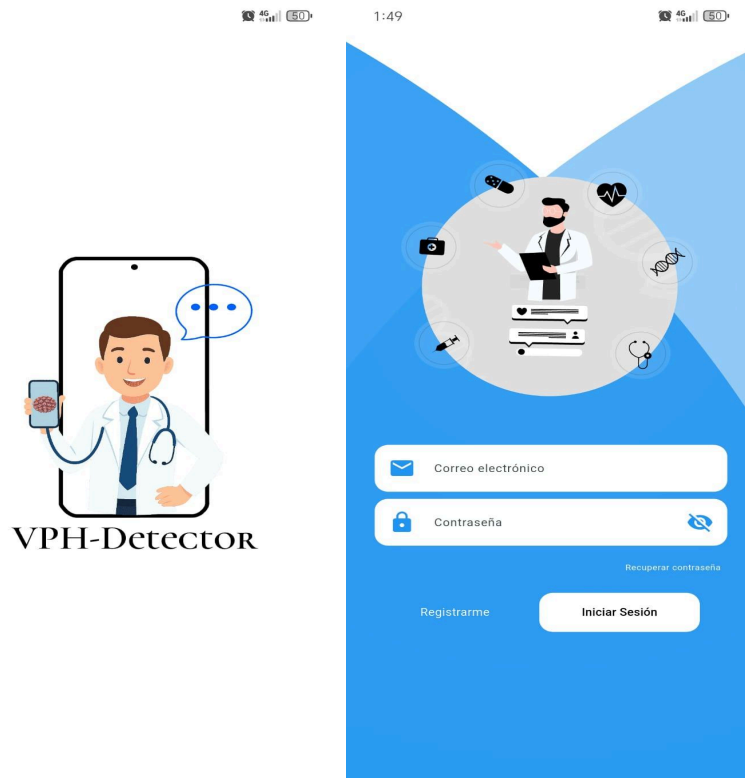
- [29] «tf.keras.layers.GlobalAveragePooling2D | TensorFlow v2.16.1», TensorFlow. Accedido: 7 de octubre de 2025. [En línea]. Disponible en: https://www.tensorflow.org/api_docs/python/tf/keras/layers/GlobalAveragePooling2D
- [30] «Citando TensorFlow», TensorFlow. Accedido: 7 de octubre de 2025. [En línea]. Disponible en: <https://www.tensorflow.org/about/bib?hl=es>
- [31] K. Team, «Keras documentation: MobileNet, MobileNetV2, and MobileNetV3». Accedido: 7 de octubre de 2025. [En línea]. Disponible en: <https://keras.io/api/applications/mobilenet/>
- [32] «Guía de inicio rápido — Documentación de Matplotlib 3.10.6». Accedido: 7 de octubre de 2025. [En línea]. Disponible en: https://matplotlib.org/stable/users/explain/quick_start.html
- [33] «Documentación de NumPy — Manual de NumPy v2.3». Accedido: 7 de octubre de 2025. [En línea]. Disponible en: <https://numpy.org/doc/stable/>
- [34] M. Waskom, «seaborn: statistical data visualization», *J. Open Source Softw.*, vol. 6, n.º 60, p. 3021, abr. 2021, doi: 10.21105/joss.03021.
- [35] «21. Dropout Neural Networks in Python | Machine Learning». Accedido: 7 de octubre de 2025. [En línea]. Disponible en: <https://python-course.eu/machine-learning/dropout-neural-networks-in-python.php>
- [36] A. Zhou, A. Yao, Y. Guo, L. Xu, y Y. Chen, «Incremental Network Quantization: Towards Lossless CNNs with Low-Precision Weights», 25 de agosto de 2017, *arXiv*: arXiv:1702.03044. doi: 10.48550/arXiv.1702.03044.
- [37] A. K. Sharma, A. Nandal, A. Dhaka, A. Alhudhaif, K. Polat, y A. Sharma, «Diagnosis of cervical cancer using CNN deep learning model with transfer learning approaches», *Biomed. Signal Process. Control*, vol. 105, p. 107639, jul. 2025, doi: 10.1016/j.bspc.2025.107639.
- [38] C. A. García Cortegano, «Sistema inteligente móvil basado en redes neuronales convolucionales para el diagnóstico previo de enfermedades renales», *Univ. Nac. Federico Villarreal*, 2020, Accedido: 30 de mayo de 2025. [En línea]. Disponible en: <https://repositorio.unfv.edu.pe/handle/20.500.13084/4481>

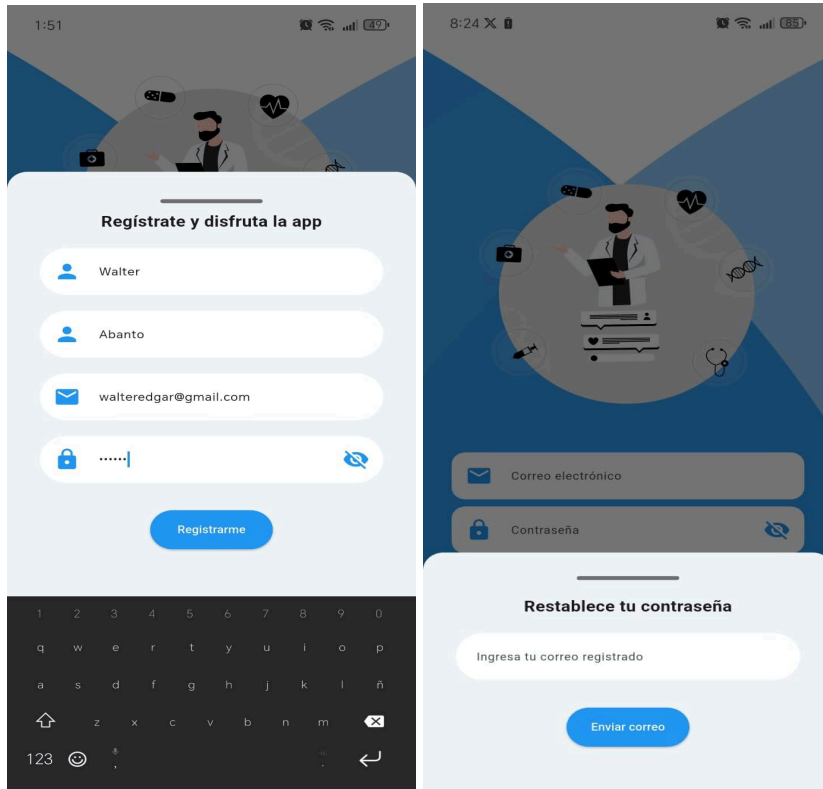
- [39] D. Raval y J. N. Undavia, «A Comprehensive assessment of Convolutional Neural Networks for skin and oral cancer detection using medical images», *Healthc. Anal.*, vol. 3, p. 100199, nov. 2023, doi: 10.1016/j.health.2023.100199.
- [40] I. D. Mienye, T. G. Swart, G. Obaido, M. Jordan, y P. Ilono, «Deep Convolutional Neural Networks in Medical Image Analysis: A Review», *Information*, vol. 16, n.º 3, Art. n.º 3, mar. 2025, doi: 10.3390/info16030195.
- [41] J. J. Castro Maldonado, L. K. Gómez Macho, E. Camargo Casallas, J. J. Castro Maldonado, L. K. Gómez Macho, y E. Camargo Casallas, «La investigación aplicada y el desarrollo experimental en el fortalecimiento de las competencias de la sociedad del siglo XXI», *Tecnura*, vol. 27, n.º 75, pp. 140-174, mar. 2023, doi: 10.14483/22487638.19171.
- [42] T. C. Luza, O. J. Córdova, J. L. A. Gonzáles, F. A. F. Limo, y C. A. B. Flores, *Métodos mixtos de investigación para principiantes*. Instituto Universitario de Innovación Ciencia y Tecnología Inudi Perú, 2023. doi: 10.35622/inudi.b.106.
- [43] «(PDF) Método de investigación científica: Diseño de proyectos y elaboración de protocolos en las Ciencias Sociales», ResearchGate. Accedido: 4 de mayo de 2025. [En línea]. Disponible en: https://www.researchgate.net/publication/377225864_Metodo_de_investigacion_cientifica_Disenio_de_proyectos_y_elaboracion_de_protocolos_en_las_Ciencias_Sociales
- [44] G. Rau y Y.-S. Shih, «Evaluation of Cohen's kappa and other measures of inter-rater agreement for genre analysis and other nominal data», *J. Engl. Acad. Purp.*, vol. 53, p. 101026, sep. 2021, doi: 10.1016/j.jeap.2021.101026.
- [45] V. Sheppard, «7.2 Population versus Samples», abr. 2020, Accedido: 10 de mayo de 2025. [En línea]. Disponible en: <https://pressbooks.bccampus.ca/jibcresearchmethods/chapter/7-2-population-versus-samples/>
- [46] E. Velasco, «Inclusion Criteria», en *Encyclopedia of Research Design*, SAGE Publications, Inc., 2010, pp. 589-591. doi: 10.4135/9781412961288.
- [47] «¿Qué es: Unidad de Análisis?», LEARN STATISTICS EASILY. Accedido: 10 de mayo de 2025. [En línea]. Disponible en: <https://es.statisticseasily.com/glosario/%C2%BFQu%C3%A9-es-la-unidad-de-an%C3%A1lisis-en-la-investigaci%C3%B3n%3F/>

Anexos

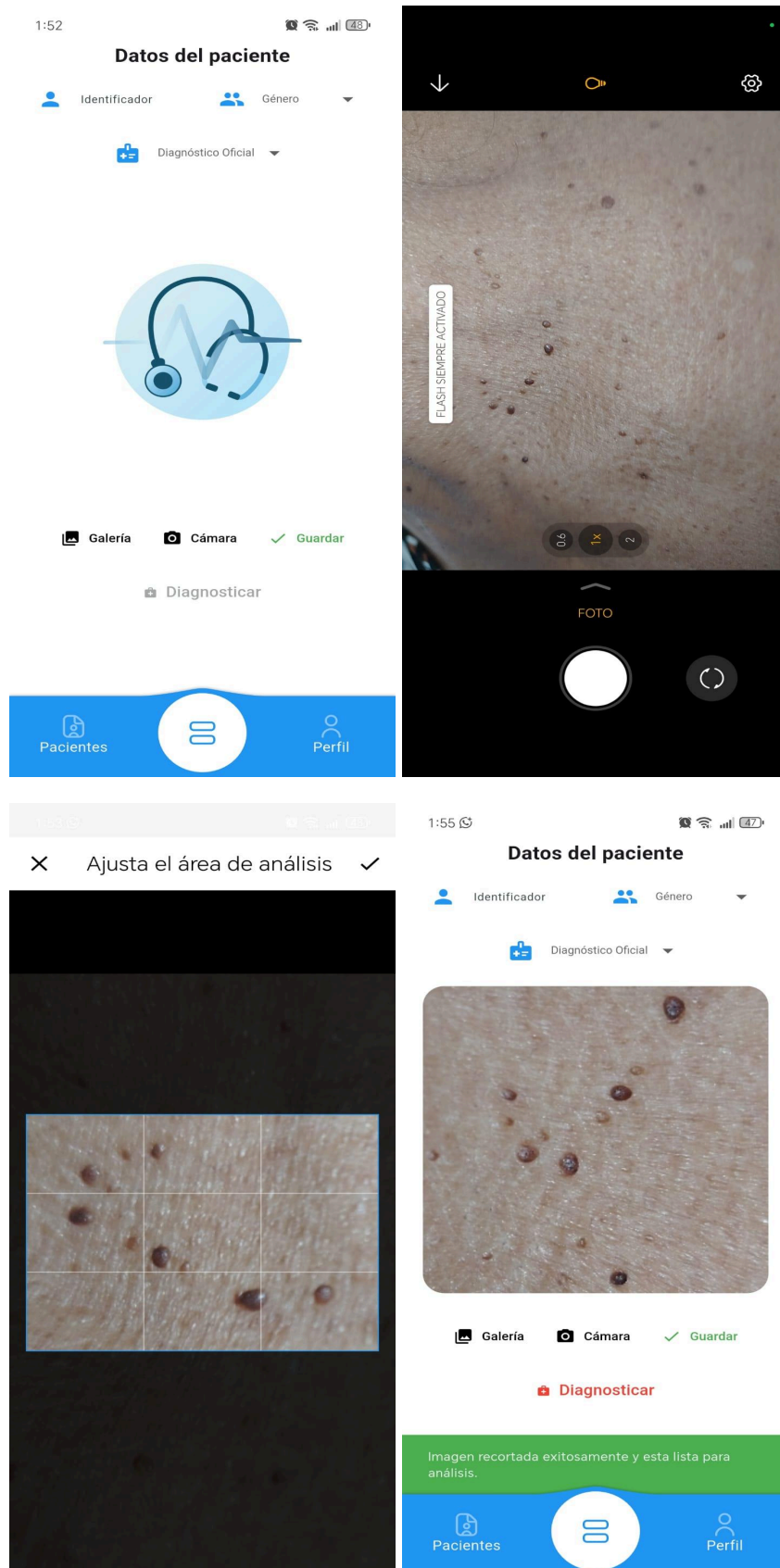
Anexo 1. Aplicativo Móvil

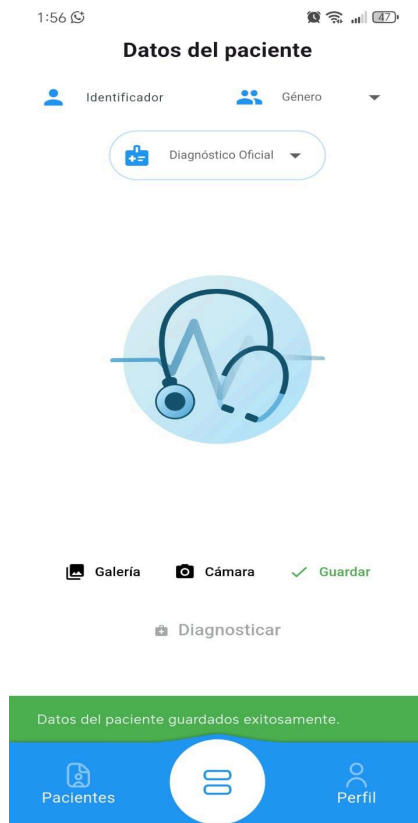
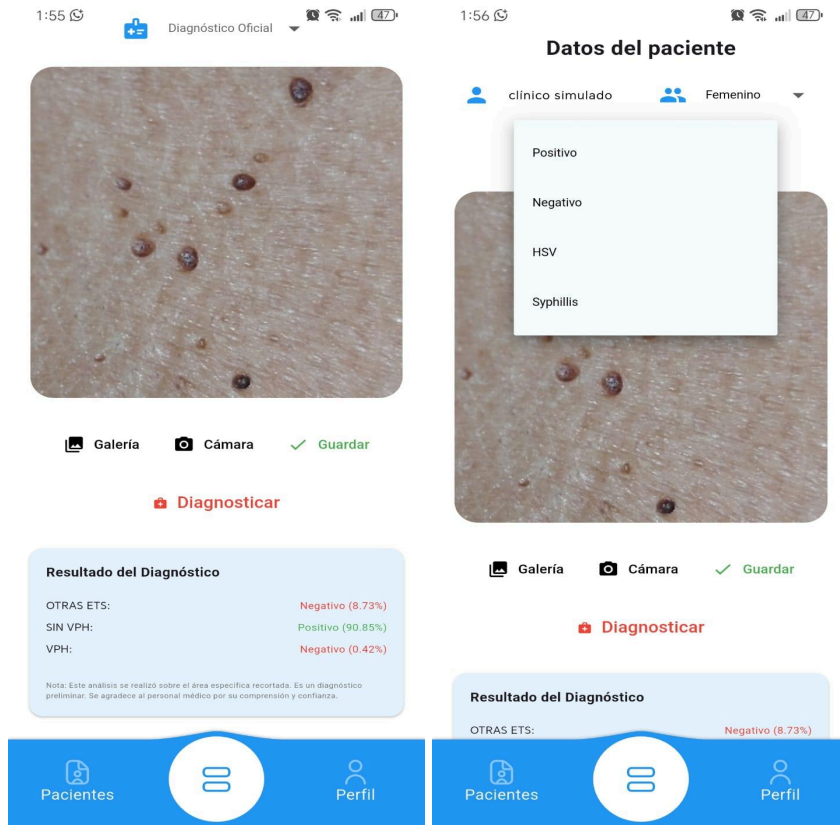
1.1. Splash, login, recuperar contraseña, registro y perfil



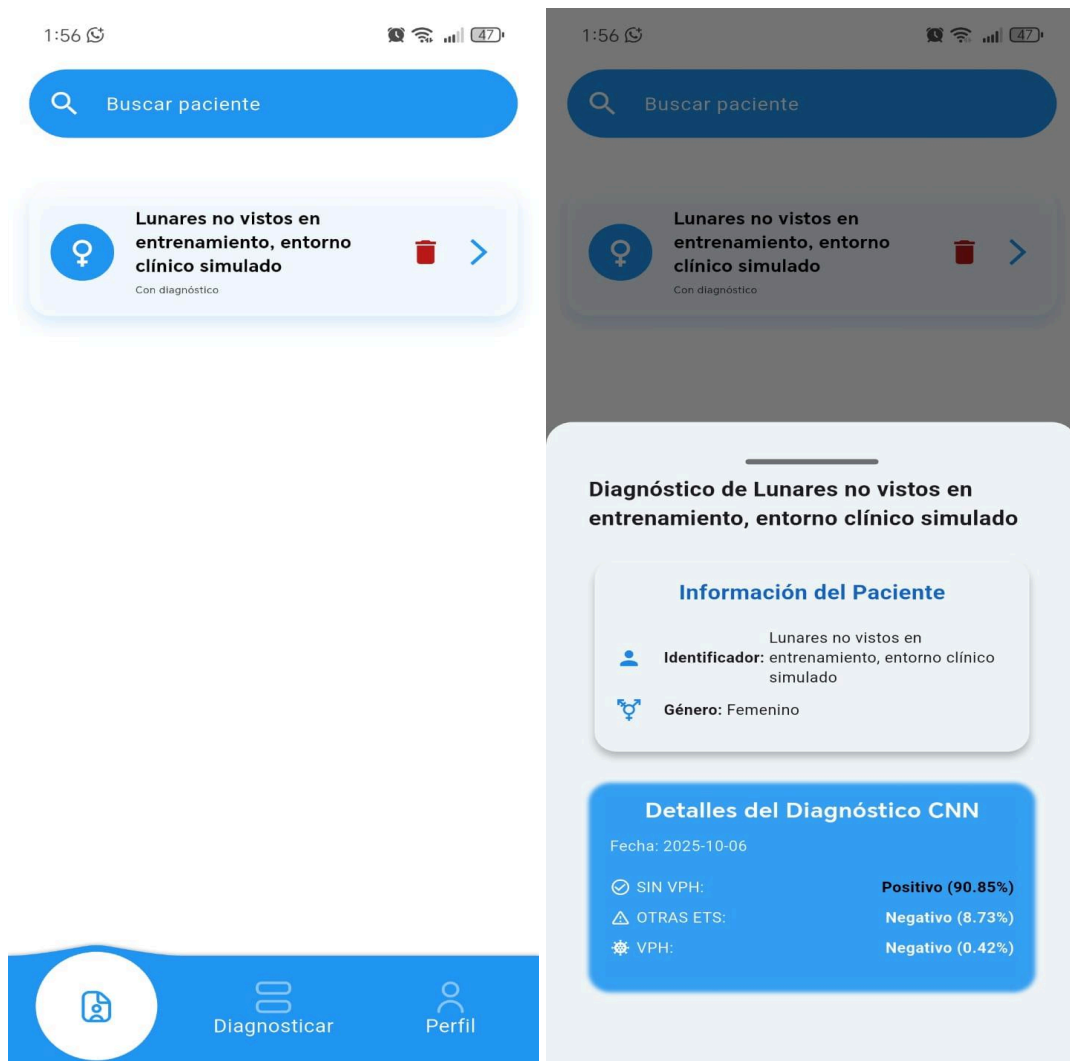


1.2. Proceso de toma, recorte y diagnóstico

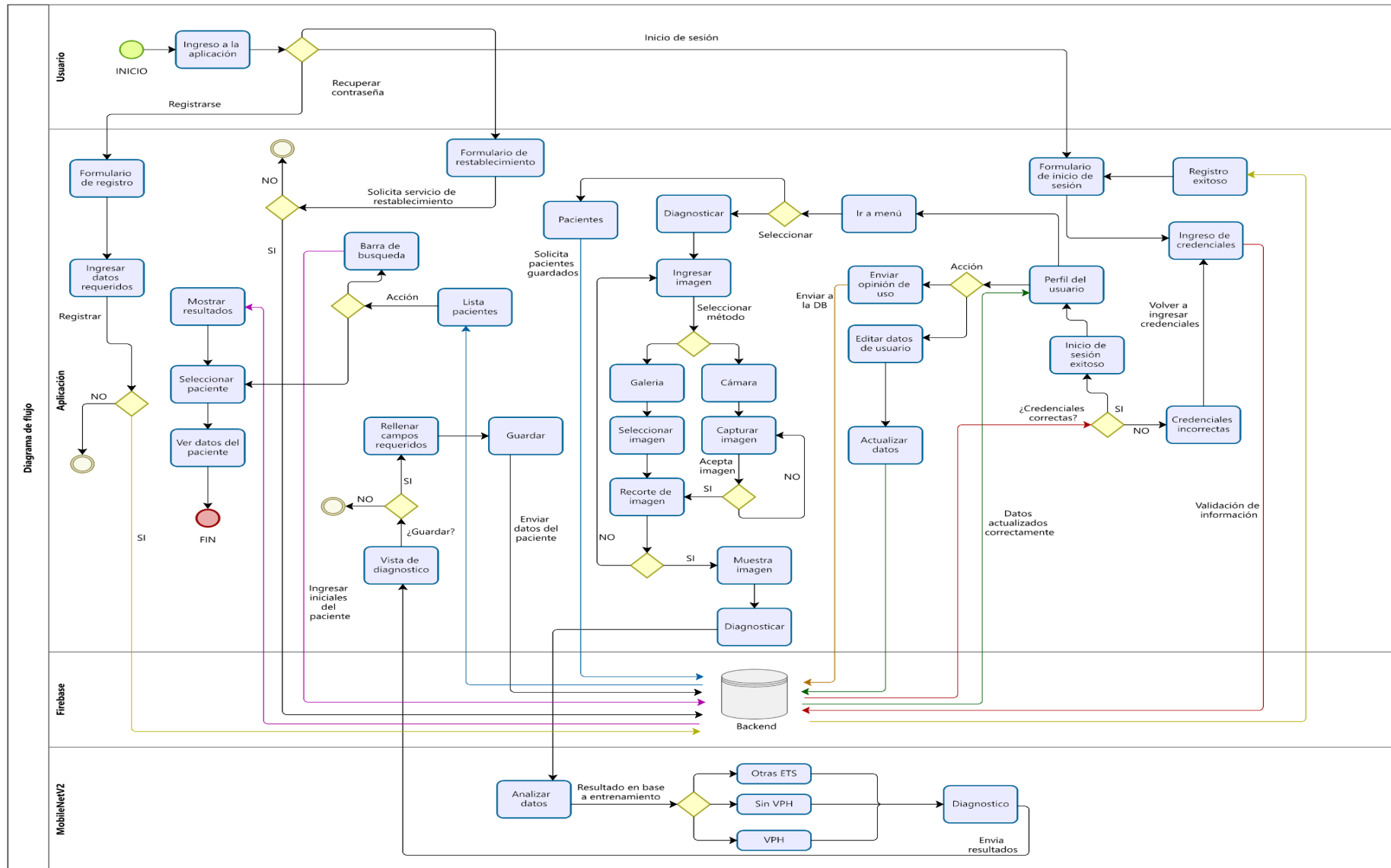




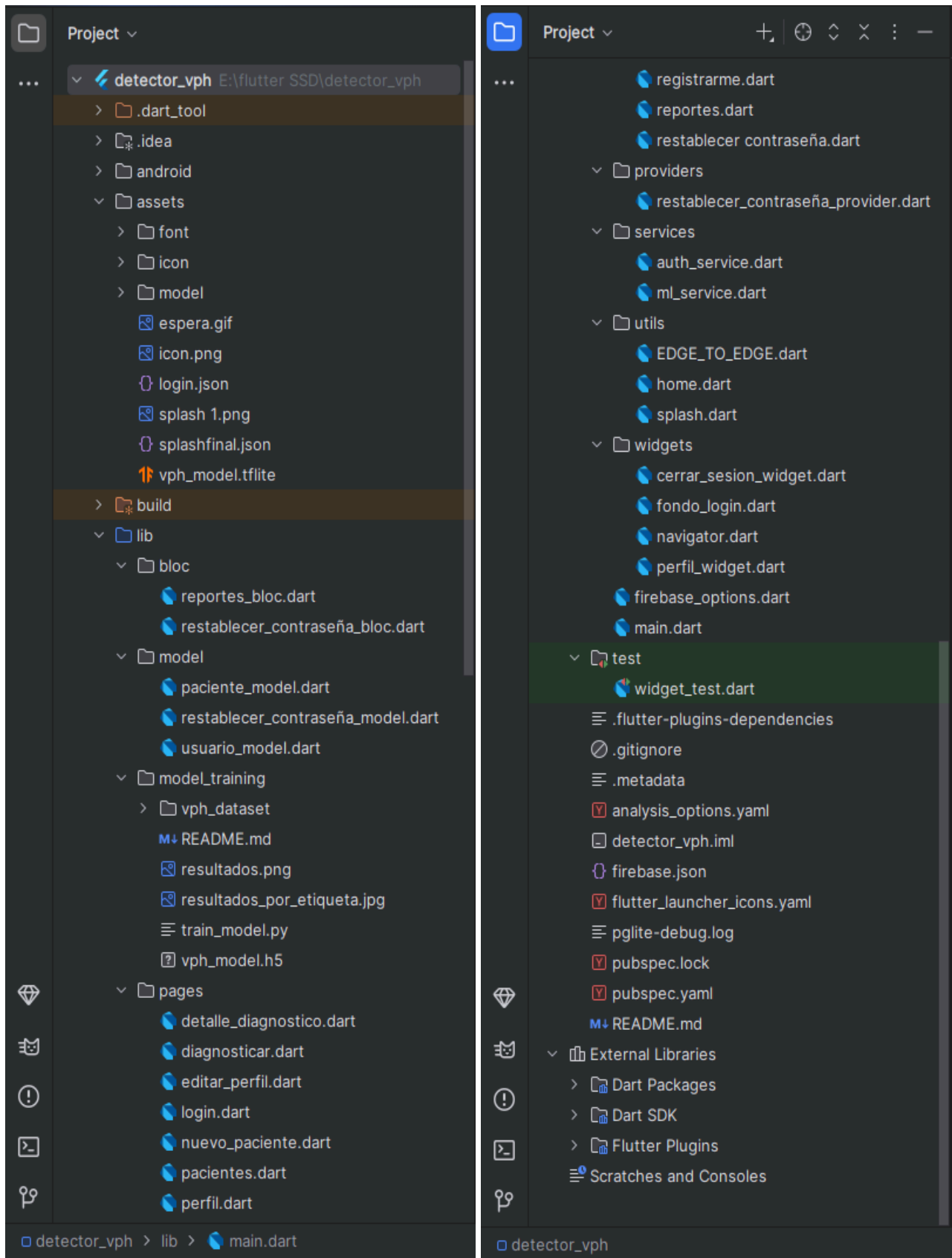
1.3. Clase pacientes y detalles



Anexo 2. Diagrama de flujo del sistema



Anexo 3. Arquitectura usada (Bloc)



Anexo 4. Librerías usadas en entrenamiento (fase computacional)

```
import tensorflow as tf
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from tensorflow.keras.applications import MobileNetV2
from tensorflow.keras.layers import Dense, GlobalAveragePooling2D, Dropout
from tensorflow.keras.models import Model
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import EarlyStopping, ReduceLROnPlateau, ModelCheckpoint
import os
import matplotlib.pyplot as plt
import numpy as np
from sklearn.metrics import classification_report, confusion_matrix
import seaborn as sns

# Directorio del dataset
dataset_dir = 'model_training/vph_dataset/'

# Parámetros del modelo
IMG_HEIGHT = 224
IMG_WIDTH = 224
BATCH_SIZE = 32
EPOCHS = 40
NUM_CLASSES = 3

# Data augmentation SOLO para entrenamiento
train_datagen = ImageDataGenerator(
    rescale=1./255,
    rotation_range=15, # para evitar distorsión excesiva
    width_shift_range=0.15,
    height_shift_range=0.15,
    shear_range=0.15,
    zoom_range=0.15,
    horizontal_flip=True,
    fill_mode='nearest'
)

# Solo rescaling para validación sin augmentation
validation_datagen = ImageDataGenerator(rescale=1./255, validation_split=0.2)
```

Este preprint fue presentado bajo las siguientes condiciones:

- Los autores declaran que se obtuvieron los términos necesarios del consentimiento libre e informado de los participantes o pacientes en la investigación y se describen en el manuscrito, cuando corresponde.
- Los autores declaran que la preparación del manuscrito siguió las normas éticas de comunicación científica.
- Los autores declaran que son conscientes de que son los únicos responsables del contenido del preprint y que el depósito en SciELO Preprints no significa ningún compromiso por parte de SciELO, excepto su preservación y difusión.
- Los autores declaran que los datos, las aplicaciones y otros contenidos subyacentes al manuscrito están referenciados.
- El manuscrito depositado está en formato PDF.
- Los autores declaran que la investigación que dio origen al manuscrito siguió buenas prácticas éticas y que las aprobaciones necesarias de los comités de ética de investigación, cuando corresponda, se describen en el manuscrito.
- Los autores declaran que una vez que un manuscrito es postado en el servidor SciELO Preprints, sólo puede ser retirado mediante solicitud a la Secretaría Editorial deSciELO Preprints, que publicará un aviso de retracción en su lugar.
- Los autores aceptan que el manuscrito aprobado esté disponible bajo licencia [Creative Commons CC-BY](#).
- El autor que presenta el manuscrito declara que las contribuciones de todos los autores y la declaración de conflicto de intereses se incluyen explícitamente y en secciones específicas del manuscrito.
- Los autores declaran que el manuscrito no fue depositado y/o previamente puesto a disposición en otro servidor de preprints o publicado en una revista.
- Si el manuscrito está siendo evaluado o siendo preparando para su publicación pero aún no ha sido publicado por una revista, los autores declaran que han recibido autorización de la revista para hacer este depósito.
- El autor que envía el manuscrito declara que todos los autores del mismo están de acuerdo con el envío a SciELO Preprints.