

**FACULTAD DE CIENCIAS DE LA SALUD**

Escuela Académico Profesional de Medicina Humana

Tesis

**Precisión diagnóstica del ChatGPT-4 para  
apendicitis aguda en el Servicio de Emergencia del  
Hospital Ramiro Prialé Prialé, Huancayo, 2022**

Diana Evelyn Barrera Chuquiarque  
Victor Ricardo Rojo Camayo

Para optar el Título Profesional de  
Médico Cirujano

Huancayo, 2025

Repositorio Institucional Continental  
Tesis digital



Esta obra está bajo una Licencia "Creative Commons Atribución 4.0 Internacional" .

**INFORME DE CONFORMIDAD DE ORIGINALIDAD DE TRABAJO DE  
INVESTIGACIÓN**

**A** : Decana de la Facultad de Ciencias de la Salud  
**DE** : Luis Tello Dávila  
Asesor de trabajo de investigación  
**ASUNTO** : Remito resultado de evaluación de originalidad de trabajo de investigación  
**FECHA** : 18 de febrero de 2025

Con sumo agrado me dirijo a vuestro despacho para informar que, en mi condición de asesor del trabajo de investigación:

**Título:**

Precisión diagnóstica del ChatGPT-4 para Apendicitis Aguda en el Servicio de Emergencia del Hospital Ramiro Priale Priale, Huancayo 2022

**Autores:**

1. Diana Evelyn Barrera Chuquiarque – EAP. Medicina Humana
2. Víctor Ricardo Rojo Camayo – EAP. Medicina Humana

Se procedió con la carga del documento a la plataforma "Turnitin" y se realizó la verificación completa de las coincidencias resaltadas por el software dando por resultado 5 % de similitud sin encontrarse hallazgos relacionados a plagio. Se utilizaron los siguientes filtros:

- Filtro de exclusión de bibliografía SI  NO
- Filtro de exclusión de grupos de palabras menores  
Nº de palabras excluidas (en caso de elegir "SI"): 40 SI  NO
- Exclusión de fuente por trabajo anterior del mismo estudiante SI  NO

En consecuencia, se determina que el trabajo de investigación constituye un documento original al presentar similitud de otros autores (citas) por debajo del porcentaje establecido por la Universidad Continental.

Recae toda responsabilidad del contenido del trabajo de investigación sobre el autor y asesor, en concordancia a los principios expresados en el Reglamento del Registro Nacional de Trabajos conducentes a Grados y Títulos – RENATI y en la normativa de la Universidad Continental.

Atentamente,

  
Mg Luis Tello Dávila  
Asesor de trabajo de investigación

## Dedicatoria

Dedicamos esta investigación...

A nuestros padres, por ser el pilar fundamental en nuestras vidas, por su amor incondicional, paciencia y apoyo constante en cada paso de este camino académico. Gracias por creer en nosotros y alentarnos a seguir adelante, incluso en los momentos más desafiantes.

A nuestras familias, quienes han sido nuestro motor y fortaleza a lo largo de este proceso. Su apoyo incondicional, sus palabras de aliento y su confianza en nosotros nos han motivado a seguir adelante en cada etapa de esta investigación.

A nosotros mismos, por el esfuerzo, la perseverancia y el compromiso que hemos puesto en este trabajo. Este logro es el resultado de nuestro ímpetu por aprender crecer, y marca el inicio de nuevas metas y desafíos.

## **Agradecimientos**

A nuestra Casa de estudios y a sus docentes, les extendemos nuestra más sincera gratitud por compartir con nosotros su conocimiento, su experiencia y su tiempo. Sus orientaciones no solo han enriquecido este trabajo, sino que también han dejado huellas en nuestra formación profesional y personal.

A nuestro asesor Dr. Luis Tello Dávila, por su guía invaluable, paciencia y compromiso durante el desarrollo de esta investigación. Su experiencia, dedicación y confianza en nuestro trabajo no solo enriquecieron este proyecto, sino que también nos inspiraron a dar lo mejor de nosotros mismos en cada etapa del proceso.

A nuestro docente y amigo el Dr. Carlos Contreras Paucca, queremos dedicar un agradecimiento especial por su acompañamiento, su apoyo desinteresado y su disposición constante para orientarnos en este camino.

A los doctores que nos ayudaron en la validación del instrumento: Dr. Brayan Capcha, Dra. Lizeth Chávez, Dra. Andrea Arbañil, Dr. Max Hernández, Dra. Sandra Barrientos y Dr. José Zarate; les expresamos nuestro más sincero agradecimiento. Su tiempo, conocimiento y rigurosidad fueron fundamentales para garantizar la calidad y pertinencia de esta herramienta, permitiéndonos llevar a cabo un trabajo sólido y confiable.

Al Hospital Ramiro Prialé Prialé y especialmente al Dr. Christian Matamoros Vera y la Dra. Magali Franco Benites quienes colaboraron con esta investigación, gracias por brindarnos los recursos, datos y espacios necesarios para llevar a cabo este proyecto. Su disposición y compromiso han sido clave para alcanzar nuestros objetivos.

A todos, gracias por formar parte de este sueño y por ayudarnos a convertirlo en una realidad.

## Índice de contenidos

Dedicatoria .....	iv
Agradecimientos .....	v
Índice de contenidos .....	vi
Índice de tablas .....	ix
Índice de gráficos .....	x
Abreviaturas .....	xi
Resumen .....	xii
Abstract y keywords .....	xiii
Introducción .....	xiv
<b>CAPÍTULO I .....</b>	<b>1</b>
<b>PLANTEAMIENTO DEL ESTUDIO .....</b>	<b>1</b>
1.1. Delimitación del problema .....	1
1.1.1. Delimitación territorial .....	1
1.1.2. Delimitación temporal .....	1
1.1.3. Delimitación conceptual .....	1
1.2. Planteamiento y formulación del problema .....	1
1.3. Formulación del problema .....	4
1.3.1. Problema general .....	4
1.3.2. Problemas específicos .....	4
1.4. Objetivos .....	4
1.4.1. Objetivo general .....	4
1.4.2. Objetivos específicos .....	4
1.5. Justificación .....	4
1.5.1. Justificación teórica .....	4
1.5.2. Justificación práctica .....	5
1.5.3. Justificación metodológica .....	6
1.5.4. Justificación social .....	6
<b>CAPITULO II .....</b>	<b>8</b>
<b>MARCO TEÓRICO .....</b>	<b>8</b>
2.1. Antecedentes de la investigación .....	8

2.1.1. Antecedentes internacionales .....	8
2.1.2. Antecedentes nacionales .....	11
2.2. Bases teóricas.....	12
2.2.1. Inteligencia artificial.....	12
2.2.2. ChatGPT.....	14
2.2.3. Prompt .....	15
2.2.4. Apendicitis aguda.....	15
2.2.5. Prueba diagnóstica .....	20
2.2.6. Precisión diagnóstica.....	20
2.3. Definición de términos básicos .....	23
CAPÍTULO III .....	25
HIPÓTESIS Y VARIABLES.....	25
3.1. Hipótesis .....	25
3.1.1. Hipótesis general.....	25
3.1.2. Hipótesis específicas .....	25
3.2. Variables de la investigación.....	25
3.3. Operacionalización de la variable.....	26
CAPÍTULO IV.....	27
MATERIALES Y MÉTODOS.....	27
4.1. Métodos, tipo y alcance de la investigación.....	27
4.1.1. Método de la investigación .....	27
4.1.2. Tipo de investigación .....	27
4.1.3. Alcance de la investigación.....	27
4.2. Diseño de la investigación.....	27
4.3. Población y muestra .....	28
4.3.1. Población.....	28
4.3.2. Muestra.....	28
4.4. Técnicas e instrumentos de recolección y análisis de datos .....	29
4.4.1. Técnicas .....	29
4.4.2. Instrumento .....	29
4.4.3. Recopilación y manejo de datos .....	30
4.4.4. Análisis de datos .....	31
4.5. Consideraciones éticas.....	33
CAPÍTULO V.....	34

RESULTADOS Y DISCUSIÓN.....	34
5.1. Resultados.....	34
5.2. Discusión.....	41
Conclusiones.....	51
Recomendaciones.....	52
Referencias bibliográficas .....	54
Anexos .....	64

## Índice de tablas

Tabla N.º 1. Frecuencia según sexo en el tópico de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022.....	34
Tabla N.º 2. Frecuencia según rango de edad en el tópico de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022. ....	35
Tabla N.º 3. Frecuencia de los diagnósticos según CIE-10 en el tópico de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022. ....	36
Tabla N.º 4. Frecuencia de los pacientes con apendicitis aguda total y según sexo en el tópico de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022. ....	37
Tabla N.º 5. Frecuencia de apendicitis aguda según rango de edad de los pacientes en el tópico de cirugía de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022. ....	37
Tabla N.º 6. Contingencia del Chat-GPT 4 para el diagnóstico de apendicitis aguda en el tópico de cirugía de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022. ....	38
Tabla N.º7. Validez y confiabilidad del Chat-GPT 4 para el diagnóstico de apendicitis aguda en el tópico de cirugía de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022. ....	39
Tabla N.º 8. Tabla de frecuencias relativas de los pacientes atendidos en el tópico de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022 ....	84
Tabla N.º 9. Tabla de estadísticos de tendencia central de los pacientes atendidos en el tópico de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022 .....	86
Tabla N.º 10. Tabla de contingencia por sexo de los pacientes atendidos en el tópico de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022 ....	88

## Índice de gráficos

Gráfico N.º 1. Curva ROC de ChatGPT-4 para el diagnóstico de apendicitis aguda en el tópicico de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022.....	41
Gráfico N.º 2. Distribución de edades según anatomopatología y respuesta de ChatGPT-4 para la apendicitis aguda en el tópicico de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022.....	89

## Abreviaturas

°C: Grados Centígrados	MLNN: Red Neuronal Multicapa
dL: decilitro	MM: Milímetros
FID: Fosa Iliaca Derecha	mmHg: Milímetros de Mercurio
FKGL: Flesch Kincaid Grade Level	NLG: Generador de Lenguaje Natural
FN: Falso Negativo	PNN: Red Neural Probabilística
FP: Falso Positivo	RBF: Función de Base Radial
FRE: Flesch Reading Ease	RNA: Redes Neuronales Artificiales
GPT: Generative Pre-trained Transformer	ROC: Receiver Operating Characteristic
Gr: Gramos	RPM: Respiraciones por Minuto
IA: Inteligencia Artificial	SEG: Segundos
IAs: Inteligencias Artificiales	TAC: Tomografía Axial Computarizada
Kg: Kilogramos	U: Unidades
L: Litros	VN: Verdadero Negativo
LPM: Latidos por Minuto	VP: Verdadero Positivo
M: Metros	VPN: Valor Predictivo Negativo
MCC: Coeficiente de Correlación de Matthews	VPP: Valor Predictivo Positivo
mg: Miligramos	VSG: Velocidad de Sedimentación Globular
MIN: Minutos	X: Por
MIR: Médico Interno Residente	XAI: Inteligencia Artificial Explicable
ML: Machine learning - Aprendizaje automático	

## Resumen

La inteligencia artificial (IA) se ha integrado progresivamente en la medicina, con aplicaciones en la práctica clínica, análisis de imágenes y diagnóstico de enfermedades. Sin embargo, la apendicitis aguda, una de las patologías quirúrgicas más frecuentes, aún enfrenta retos diagnósticos, con tasas significativas de diagnósticos erróneos y apendicectomías innecesarias, especialmente en mujeres. Ante este panorama, se exploró el potencial del ChatGPT-4 como una herramienta accesible y comprensible para apoyar en el diagnóstico clínico de la apendicitis aguda, con el fin de reducir errores y optimizar necesidad de la atención quirúrgica. El objetivo de esta investigación fue determinar la precisión diagnóstica del ChatGPT-4 para apendicitis aguda en el servicio de emergencia del Hospital Ramiro Prialé Prialé. Se realizó un estudio cuantitativo, descriptivo y transversal. La población estuvo conformada por pacientes entre 18 a 59 años que fueron atendidos en el tópico de cirugía en el servicio de emergencia en el año 2022. La muestra incluyó 495 historias clínicas, las cuales fueron ingresadas al ChatGPT-4 para determinar la presencia o ausencia de apendicitis aguda. Estos resultados se compararon con la base de datos del servicio de anatomopatología, donde se registran las biopsias consideradas el método diagnóstico de certeza para esta investigación. Según los indicadores de validez, el ChatGPT-4 tuvo una sensibilidad de 89 %, especificidad de 98 %, valor predictivo positivo de 75 % y valor predictivo negativo de 99 %. Respecto al likelihood ratio positivo y negativo sus valores fueron 37 y 0,11 respectivamente; además, el índice de exactitud fue 97 %, el score F1 de 0,815 y el área bajo la curva ROC fue de 0,93. En cuanto a la confiabilidad, el coeficiente de correlación de Matthew fue de 80 %. Los investigadores concluyeron que ChatGPT-4 tiene una adecuada precisión diagnóstica para la apendicitis aguda. Sin embargo, se recomienda realizar más estudios en diferentes hospitales y con un rango de edad más amplio para validar estos hallazgos.

Palabras claves: inteligencia artificial, ChatGPT, apendicitis aguda

## **Abstract y keywords**

Artificial intelligence (AI) has been progressively integrated into medicine, with applications in clinical practice, image analysis and disease diagnosis. However, acute appendicitis, one of the most frequent surgical pathologies, still faces diagnostic challenges, with significant rates of misdiagnoses and unnecessary appendectomies, especially in women. Given this scenario, the potential of ChatGPT-4 was explored as an accessible and understandable tool to support the clinical diagnosis of acute appendicitis, to reduce errors and optimize the need for surgical care. The objective of this research was to determine the diagnostic accuracy of ChatGPT-4 for acute appendicitis in the emergency service of the Ramiro Prialé Prialé Hospital. A quantitative, descriptive and cross-sectional study was carried out. The population consisted of patients between 18 and 59 years of age who were treated in the surgery area in the emergency service in 2022. The sample included 495 medical records, which were entered into ChatGPT-4 to determine the presence or absence of acute appendicitis. These results were compared with the database of the pathology service, where the biopsies considered the diagnostic method of certainty for this research are recorded. According to the validity indicators, ChatGPT-4 had a sensitivity of 89 %, specificity of 98 %, positive predictive value of 75 % and negative predictive value of 99 %. Regarding the positive and negative likelihood ratio, their values were 37 and 0,11 respectively; in addition, the accuracy rate was 97 %, the F1 score was 0,815 and the area under the ROC curve was 0,93. Regarding reliability, Matthew's correlation coefficient was 80 %. The researchers concluded that ChatGPT-4 has adequate diagnostic accuracy for acute appendicitis. However, further studies in different hospitals and with a wider age range are recommended to validate these findings.

Keywords: artificial Intelligence, ChatGPT, acute appendicitis.

## Introducción

La inteligencia artificial (IA) ha demostrado un impacto creciente en la medicina, especialmente en el apoyo al diagnóstico y tratamiento mediante el análisis de imágenes y patrones clínicos. En particular, se ha explorado su aplicación en el diagnóstico de apendicitis aguda, una patología frecuente con riesgo de complicaciones si no se detecta a tiempo, además de otras patologías de gran importancia para la salud pública. ChatGPT-4, con su capacidad de procesamiento de lenguaje natural, podría ser una herramienta accesible para mejorar la precisión diagnóstica, reduciendo errores y facilitando el acceso a la salud en regiones con limitaciones.

En este contexto, los investigadores llevaron a cabo un estudio para evaluar la validez y confiabilidad de ChatGPT-4 en el diagnóstico de apendicitis aguda en el servicio de emergencia del Hospital Ramiro Priale Priale en 2022, con el fin de generar evidencia sobre su aplicabilidad y su potencial para modernizar el sistema de salud en Perú.

Este estudio se estructuró en cinco capítulos. En el Capítulo I, se aborda la delimitación y el planteamiento del problema, analizando la importancia de la inteligencia artificial en la medicina y su posible aplicación en el diagnóstico de apendicitis aguda. Además, se establecen los objetivos de la investigación y se justifica su relevancia en el contexto de la modernización del sistema de salud en el país.

El Capítulo II desarrolla el marco teórico, donde se presentan los antecedentes de investigaciones previas relacionadas con el uso de inteligencia artificial en el diagnóstico médico. Asimismo, se revisan los conceptos fundamentales de la apendicitis aguda y la inteligencia artificial proporcionando la base científica para la investigación.

En el Capítulo III, se establecen las hipótesis del estudio y se identifican las variables principales, definiendo su operacionalización para una correcta medición e interpretación. Se describen las dimensiones e indicadores de cada variable, asegurando una base estructurada para el análisis de datos.

En el Capítulo IV, se describe el diseño metodológico del estudio, detallando el tipo, nivel y enfoque de la investigación. Se especifica el método empleado para evaluar la precisión diagnóstica de ChatGPT-4 en la detección de apendicitis aguda. Además, se presenta la sede del estudio y se define la muestra analizada, explicando los criterios utilizados para su selección. Se detallan las variables del estudio y la manera en que fueron medidas, así como las técnicas e instrumentos empleados para la recolección de datos.

También se expone el procedimiento seguido para la obtención y análisis de la información, incluyendo el tratamiento estadístico aplicado para evaluar los resultados. Finalmente, se abordan las consideraciones éticas, asegurando el cumplimiento de los principios éticos y la protección de los datos de los pacientes.

El Capítulo V presenta los resultados obtenidos y la discusión de estos, analizando la precisión diagnóstica de ChatGPT-4 en la identificación de casos de apendicitis aguda. Se comparan los hallazgos con estudios previos y se discuten las implicancias del uso de esta tecnología en el ámbito clínico.

Finalmente, se exponen las conclusiones del estudio y se proponen recomendaciones para futuras investigaciones y para la posible implementación de este tipo de herramientas en la práctica médica

## **CAPÍTULO I**

### **PLANTEAMIENTO DEL ESTUDIO**

#### **1.1. Delimitación del problema**

##### **1.1.1. Delimitación territorial**

La presente investigación se llevó a cabo en el servicio de Emergencia del Hospital Nacional Ramiro Prialé Prialé, perteneciente a EsSalud y ubicado en el distrito de El Tambo en la ciudad de Huancayo. Se trata de un hospital de categoría III-1 que cuenta con diversas especialidades, como medicina interna, cirugía general y emergencia, además de servicios de apoyo al diagnóstico, incluyendo imagenología, patología y laboratorio.

##### **1.1.2. Delimitación temporal**

Este estudio abarcó la revisión de historias clínicas de pacientes atendidos en el tópico de cirugía del servicio de emergencia, durante el período comprendido entre enero y diciembre de 2022

##### **1.1.3. Delimitación conceptual**

En la investigación en salud, la delimitación conceptual es fundamental para definir con precisión los términos clave y el alcance del estudio (1). Tomando esto en cuenta, definimos que la precisión diagnóstica se refiere a la capacidad de una herramienta o prueba para distinguir correctamente entre la presencia y ausencia de una enfermedad, evaluándose mediante indicadores como sensibilidad, especificidad y valores predictivos positivo y negativo, entre otros.

Este estudio evalúa la precisión diagnóstica de ChatGPT-4 en la detección de apendicitis aguda, una patología caracterizada por la inflamación del apéndice, cuyo diagnóstico temprano es fundamental para prevenir complicaciones graves. Se analizará la capacidad del modelo para distinguir entre pacientes con y sin apendicitis dentro del servicio de Emergencia del Hospital Ramiro Prialé Prialé, en Huancayo, durante el año 2022, comparándolo con los métodos diagnósticos convencionales empleados por los profesionales de la salud.

#### **1.2. Planteamiento y formulación del problema**

La inteligencia artificial (IA) ha ido ganando protagonismo por su acelerado aprendizaje en diversas áreas, siendo la medicina una de las cuales se ha visto una integración adecuada, que es cada vez más importante conforme se desarrolla e investiga en las distintas aplicaciones en cada especialidad medicina. Entre los usos que se ya se

ha implementado para las IAs son ayuda en decisiones clínicas tanto en el diagnóstico como el tratamiento y en el estudio de imágenes donde analiza radiografías, tomografías u otras imágenes y patología, donde es fundamental la búsqueda de patrones que facilitan el reconocimiento de determinadas enfermedades. (2,3)

Se plantea que, en un futuro no muy lejano, las IAs ayudarán a los profesionales sanitarios en el procesamiento de información, automatización de labores repetitivas, en el reconocimiento de equivocaciones en la prescripción de medicamentos y la detección de reacciones adversas medicamentosas; además, los usuarios podrán tener acceso a tratamientos en enfermedades comunes y que no representen riesgo vital, acudiendo al sistema de salud por padecimientos de mayor complejidad o que requieran tratamientos puntuales (4). Sin embargo, es una tecnología que aún está en constante evolución y aprendizaje de una gran cantidad de datos; por ello se requiere más investigación en esta área para alcanzar el potencial previsto, y así pueda existir una colaboración entre este sistema avanzado y los profesionales de la salud para una mejor atención a los pacientes. (5)

El ChatGPT es un tipo de inteligencia artificial de procesamiento de lenguaje natural que puede responder preguntas de todo tipo, incluidas las referidas a la salud, de manera entendible y con un lenguaje parecido al humano (6). No obstante, existen limitaciones en su uso, una de las más importantes es la precisión y confiabilidad de la información que origina, debido a que, aún no se puede determinar exactamente las fuentes bibliográficas que utiliza para llegar a las respuestas planteadas (7). La calidad de las respuestas que puede dar estas nuevas tecnologías, están ligadas a la cantidad y calidad de la información brindada al mismo, en especial en el área de la salud porque tiene potencial de afectar la vida de muchas personas. (8)

La apendicitis aguda es uno de los motivos quirúrgicos más frecuentes en la consulta de emergencia en todo el mundo, se calcula que alrededor del 7 % de personas padecerán esta patología (9). Siendo ligeramente más frecuente en varones que en mujeres (10). Se estima que 10 de cada 10 000 pacientes por año se les realizaron apendicectomías (11), así como, la tasa de mortalidad es relativamente pequeña, calculándose como menor de 1 por 100 000 pacientes (11). Las edades de mayor incidencia están comprendidas entre 20 y 30 años, observándose menor frecuencia en niños muy pequeños y en adultos mayores. (12)

Las primeras apendicetomías se realizaron en el año 1880 y desde ese momento se han mejorado las técnicas de diagnóstico y tratamiento quirúrgico, sin embargo, en el 15 % de los casos se diagnostica erróneamente, generando retraso en el procedimiento

quirúrgico y aumentando las complicaciones de esta patología (11). Por otro lado, existe una cantidad importante de personas que se les realizaron apendicectomías sin tener apendicitis aguda según los hallazgos histopatológicos, oscilando entre 6 a 12 % (13,14). Y en el caso de las mujeres, es más frecuente ya que está relacionado a la presencia de los órganos reproductivos en cavidad abdominal que podría aparentar la apendicitis aguda. (14,15)

En nuestro país, se estima que la tasa de apendicitis es de 9,6 por 10 000 habitantes, encontrándose que la frecuencia fue mayor en departamento de Madre de Dios con 22,5 y la menor fue en Puno con 2,7 por 10 000 habitantes (16). Además, se ha calculado que la mediana del tiempo que pasa hasta que se realice la operación es de 33,5 horas, obteniendo apéndices gangrenados con riesgo elevado de complicaciones. (17)

En la región Junín la tasa fue de 7,1 por 10 000 habitantes encontrándose mayor incidencia en el grupo etario comprendido entre 30 a 59 años con 8,6 por 10 000 habitantes (16). Se encontró que un hospital nacional ubicado en la ciudad de Huancayo, el porcentaje de apendicectomías sin apendicitis aguda fue de 5,2 %, presentándose con más frecuencia en los pacientes de 30 a 59 años, seguido del grupo comprendido entre los 18 a 29 años. De todas las apendicectomías negativas encontradas, la mayoría se debieron a causas gineco obstétricas y en segundo lugar fueron las de origen abdominal. (18)

Teniendo en cuenta la importancia de esta patología a nivel mundial, se ha planteado el uso de la inteligencia artificial para el diagnóstico de la apendicitis a través de redes neuronales y distintos algoritmos de aprendizaje automático, teniendo resultados favorables en distintas investigaciones (19,20); todo esto con un abordaje más ligado a la ingeniería de sistemas, de manejo difícil para un usuario promedio. No obstante, el ChatGPT-4 plantea una mayor facilidad en el manejo, comprensión a nivel usuario con lenguaje natural, interfaz amigable al usuario, accesibilidad que solo necesita conexión a internet y respuestas en tiempo real, nos plantea que este Chabot, podría ser una herramienta útil y accesible para abordar esta patología. Para ello, se debería favorecer su entrenamiento en la rama diagnóstica y en el pronóstico individualizado con el fin de generar mayor precisión y evitar errores que pueden impactar en la vida del paciente. (21)

Teniendo en cuenta que la apendicitis aguda es una enfermedad de alta incidencia a nivel mundial y que el diagnóstico es fundamentalmente clínico, los investigadores de este estudio han considerado importante proponer a la inteligencia artificial como otra ayuda diagnóstica a fin de evitar errores en la prescripción o retrasos en la atención quirúrgica.

### **1.3. Formulación del problema**

#### **1.3.1. Problema general**

- ¿Cuál es la precisión diagnóstica del ChatGPT-4 para apendicitis aguda en el servicio de emergencia del Hospital Ramiro Prialé Prialé en 2022?

#### **1.3.2. Problemas específicos**

- ¿Cuál es la validez del ChatGPT-4 para apendicitis aguda en el servicio de emergencia del Hospital Ramiro Prialé Prialé en 2022?

- ¿Cuál es la confiabilidad del ChatGPT-4 para apendicitis aguda en el servicio de emergencia del Hospital Ramiro Prialé Prialé en 2022?

### **1.4. Objetivos**

#### **1.4.1. Objetivo general**

- Determinar la precisión diagnóstica del ChatGPT-4 para apendicitis aguda en el servicio de emergencia del Hospital Ramiro Prialé Prialé.

#### **1.4.2. Objetivos específicos**

- Evaluar la validez del ChatGPT-4 para apendicitis aguda en el servicio de emergencia del Hospital Ramiro Prialé Prialé

- Estimar la confiabilidad del ChatGPT-4 para apendicitis aguda en el servicio de emergencia del Hospital Ramiro Prialé Prialé

### **1.5. Justificación**

ChatGPT-4 es un modelo avanzado de inteligencia artificial (IA), este se encuentra diseñado para la generación de texto de manera coherente y precisa, se encuentra basada en técnicas de aprendizaje profundo (Deep learning), por ello, este modelo puede comprender, interpretar y responder preguntas en lenguaje natural, haciéndola útil en gran número de aplicaciones. (8,22)

En el ámbito de la medicina, la IA ha demostrado ser una herramienta revolucionaria, con capacidad de análisis de gran volumen de datos clínicos, identificar patrones complejos y asistir al diagnóstico preciso y oportuno de los pacientes. (8)

#### **1.5.1. Justificación teórica**

La inteligencia artificial se está consolidando como una herramienta esencial en la práctica médica, siendo destacada por su capacidad de mejorar procesos complejos como

el diagnóstico clínico diferencial. Parte de esta nueva tecnología, el aprendizaje profundo, logra con el análisis de grandes volúmenes de datos dar con diagnósticos precisos de diferentes enfermedades. (23)

La presente investigación se fundamenta en explorar la aplicabilidad y precisión de este tipo de tecnología en el diagnóstico de apendicitis aguda, aportando evidencia científica sobre su validez y confiabilidad en entornos hospitalarios, además, nos puede dar una pauta para la implementación en otras patologías frecuentes y si es factible su aplicabilidad en otros entornos como en diferentes hospitales o en el primer nivel de atención.

### **1.5.2. Justificación práctica**

La inteligencia artificial, en este caso ChatGPT-4, está revolucionando la medicina al proporcionar herramientas con la capacidad de analizar extensos volúmenes de datos clínicos de manera rápida, sistematizada y precisa, superando limitaciones humanas en el diagnóstico médico (24,25). En el caso de la apendicitis aguda en el servicio de emergencia del Hospital Ramiro Prialé Prialé, es una de las causas más frecuentes de consulta. Sin embargo, la variabilidad en la presentación clínica y los recursos limitados en ciertas circunstancias pueden llevar a diagnósticos erróneos o retrasados, con sus posteriores consecuencias. Incorporar herramientas como estas pueden reducir significativamente estas limitaciones, apoyando al médico a la toma de decisiones basadas en el diagnóstico ofrecidos luego de un análisis sistemático, no sesgado y basado en datos (26). Esto no solo mejorará la precisión en el diagnóstico, sino optimizará el tiempo y recursos disponibles, impactando positivamente en la salud del paciente.

Además, los serumistas médicos recién graduados que realizan su servicio rural y urbano marginal de salud, desempeñan un papel clave en la atención primaria de salud en localidades alejadas de recursos limitados. En estas zonas escasean los especialistas y tecnologías avanzadas, haciendo que el diagnóstico de condiciones potencialmente mortales o que generen comorbilidades, sean mucho más complicadas (27). La implementación del ChatGPT-4, podría proporcionar un apoyo valioso a profesionales novatos, ofreciéndole una segunda opinión fiable, rápida y basada en evidencia. Esto contribuirá a reducir tasas de error y mejorar la calidad de atención en comunidades remotas.

También, los internos y estudiantes de medicina, en su etapa formativa dentro de hospitales y centros de salud enfrentan desafíos para consolidar sus conocimientos, capacidad diagnóstica y de toma de decisiones en escenarios reales. El uso de ChatGPT-

4 se encuentra estudiada en el ámbito educativo ayudando al estudiante a mejorar sus habilidades y conocimientos, además la capacidad de ChatGPT-4 de funcionar como un tutor ofreciendo información y escenarios simulados hacen que el interno y estudiante de medicina exploren sus habilidades en el diagnóstico diferencial, interpretación de datos clínicos complejos y reforzando su razonamiento médico (28). Al integrar esta tecnología en su formación mejorará sus habilidades clínicas en su futura práctica médica.

### **1.5.3. Justificación metodológica**

El diseño de esta investigación está orientado a evaluar la precisión diagnóstica de ChatGPT-4 en entornos médicos. Al tratar a esta como una herramienta de diagnóstico, se busca determinar su capacidad para generar un diagnóstico acreditado mediante el análisis sistemático de datos clínicos, comparándolo con los resultados definitivos obtenidos de exámenes anatomopatológicos, considerados el estándar de referencia de esta enfermedad.

Entre las innovaciones metodológicas de este estudio esta la incorporación del score F1, un indicador estadístico especialmente útil para evaluar la precisión y el equilibrio entre sensibilidad y especificidad en modelos de inteligencia artificial. Este enfoque permite medir el rendimiento del modelo en términos de concordancia diagnóstica, además, introduce una herramienta avanzada del análisis de pruebas diagnósticas, ampliando las posibilidades metodológicas en investigaciones futuras. (29)

El empleo de una muestra probabilística y representativa, calculada con un nivel de confianza del 98 % y un margen de error del 5 %, garantiza la robustez de los resultados de esta investigación.

La combinación de herramientas tecnológicas avanzadas, métodos innovadores traídos de áreas que antes eran ajenas a la medicina como son las ciencias computacionales y un diseño metodológico sólido refuerza el rigor de esta investigación, contribuyendo al desarrollo de nuevas perspectivas en el uso de inteligencia artificial como apoyo en el diagnóstico clínico. (29)

### **1.5.4. Justificación social**

La implementación de herramientas de IA en salud como es ChatGPT-4 no solo beneficia a la labor del personal médico, sino también a la sociedad en general garantizando un acceso más equitativo a diagnósticos precisos, independientemente de la ubicación geográfica o limitaciones de recursos humanos como técnicos. (30)

En un país como el Perú, que el alcance de la salud todavía se encuentra limitada por el centralismo, y en ciudades como Huancayo, donde las brechas en la atención médica aún persisten, un sistema como este podría democratizar el acceso a diagnósticos oportunos, reduciendo complicaciones y mejorando los desenlaces de enfermedades con potencial mortal como es la apendicitis aguda (30–33). Es de resaltar que el promover el uso de nuevas tecnologías fomentan la modernización del sistema de salud y el fortalecimiento de capacidades locales en salud.

## **CAPITULO II**

### **MARCO TEÓRICO**

#### **2.1. Antecedentes de la investigación**

##### **2.1.1. Antecedentes internacionales**

Reis et al. (34) en el 2024, realizaron una investigación titulada Rendimiento de los chatbots basados en inteligencia artificial (IA) en la evaluación de informes de casos médicos: información cualitativa en escenarios simulados en Alemania, su objetivo fue evaluar tres chatbots populares y emergentes respaldados por IA en el análisis de informes de casos que representan tanto emergencias médicas como casos no urgentes, para este estudio se usaron ChatGPT-3.5, Bing Chat y Pi AI presentándoles 6 casos médicos simulados de casos urgentes (apendicitis aguda, síndrome coronario agudo y tendencia suicida aguda) y no urgentes (infección respiratoria no complicada / bronquitis, lesión cutánea poco clara y hematoma sin complicaciones) donde se evaluó la respuesta de cada chatbot dentro de las categorías de diagnósticos diferenciales, instrucciones adicionales, urgencia, aviso legal médico y empatía, calificándolos en alto, moderado y bajo. Se obtuvo como resultado que todos fueron capaces de diferenciar los casos agudos de los no críticos. Bing Chat obtuvo 21 veces calificaciones altas, Chat-GPT 17 veces y Pi AI 12 veces. Se concluyó que Bing Chat tuvo respuestas más completas y detalladas, por otro lado, Pi AI tuvo un abordaje más conciso y empático.

Ghanem et al. (35) en el 2024 llevaron a cabo una investigación titulada Del Dr. Google al Dr. ChatGPT: evaluación del contenido y la calidad de la información médica generada por inteligencia artificial sobre la apendicitis, tuvo por objetivos evaluar la calidad y legibilidad de la información médica generada por múltiples chatbots de IA populares sobre apendicitis aguda, y evaluar si el público puede utilizar de forma segura el contenido generado por IA como fuente de información médica en línea. Para esto se usó Bard, Claude 2, ChatGPT-3.5 y 4, a quienes se les propuso 4 preguntas de manera secuencial; sus respuestas fueron evaluadas por el instrumento DISCERN modificado para medir la calidad de la información en salud, además se midió legibilidad a través de Flesch Reading Ease (FRE) y Flesch Kincaid Grade Level (FKGL). Los resultados obtenidos fueron de calidad media para los 4 chatbots, siendo el de menor calidad Claude 2, Bard fue la única que entregó fuentes verificables. En cuanto a legibilidad, las 4 indicaron legibilidad difícil, estimado para un estudiante de secundaria y graduado universitario. Por lo tanto, se concluyó que ChatGPT-3.5, ChatGPT-4 y Bard producen contenido de calidad superior, teniendo en cuenta que Bard fue el único con fuentes verificables.

Akbulut et al. (36) en el 2023 realizaron un trabajo titulado Predicción de apendicitis aguda perforada y no perforada mediante inteligencia artificial explicable basada en aprendizaje automático en Turquía, teniendo como objetivo crear un modelo de aprendizaje automático (ML) que pueda predecir la apendicitis aguda perforada y no perforada con alta precisión y demostrar la interpretabilidad clínica del modelo con inteligencia artificial explicable (XAI). Se incluyó 1797 pacientes que se habían sometido a apendicectomía, quienes se dividieron en apendicetomías negativas, apendicitis aguda no perforada y perforada a través del modelo CatBoots, que los distinguió con una precisión de 88,2 % a los que tuvieron apendicitis de los que no, y con 92 % a los de apendicitis perforada y no perforada. Además, se asoció valores altos de bilirrubina, leucocitos, neutrófilos, entre otros aumentaron la predicción de apendicitis aguda perforada, por lo tanto, se concluyó que se pudo predecir apendicitis y apendicitis perforada con alta precisión.

Yingjie et al. (37) en el 2023, investigaron si ¿Puede Chat-GPT sustituir al médico residente de urología en el diagnóstico de enfermedades? en China, para ello estudiaron de manera retrospectiva a pacientes con enfermedades urinarias comunes y raras, siendo en total 306. Esta información se introdujo al ChatGPT y además se evaluó a 3 residentes junior. Se obtuvo como resultados que el ChatGPT tuvo una precisión de 94,1 % que es similar a 2 de los 3 residentes participantes; por ello, se concluyó que el ChatGPT tiene potencial como ayuda diagnóstica en especial para el de enfermedades comunes urológicas, pero no en las enfermedades raras.

Rao et al. (38) en 2023, realizaron una investigación titulada Evaluación de ChatGPT como complemento para la toma de decisiones radiológicas en Estados Unidos, proponiéndose como objetivo evaluar la capacidad de ChatGPT para apoyar la toma de decisiones clínicas en radiología mediante la identificación de servicios de imágenes apropiados para dos presentaciones clínicas importantes: detección de cáncer de mama y dolor de mama. Para lograr esto se comparó el formato abierto (ChatGPT sin lista del Colegio Americano de Radiología) y el formato que incluye los criterios. Se presentó que el ChatGPT tuvo una precisión moderada general, con el modelo abierto obtuvo 1,83 puntos para cáncer de mama y 1,125 para dolor de mama y con el modelo donde se incluyen los criterios del colegio americano de radiología se presentó el 88,9 % para las indicaciones en detección de cáncer de mama y 58,3 % en dolor de mamas. Por lo tanto, se puede concluir que es factible utilizar el ChatGPT para la toma de decisiones radiológicas con el fin de hacer uso responsable de la unidad de radiología.

Ghareeb et al. (39) en el 2022, desarrollaron un trabajo titulado Inteligencia artificial comparada con el sistema de puntuación de Alvarado solo o combinado con criterios

ecográficos en el diagnóstico de apendicitis aguda en Egipto, teniendo como objetivo precisión diagnóstica del modelo derivado de inteligencia artificial (IA) en comparación con la escala de Alvarado sola o combinada con criterios de ecografía, por ello se utilizaron 4 modelos: predictivo basado en IA, escala de Alvarado, criterios ecográficos y, la combinación de puntaje de Alvarado y criterios ecográficos. Participaron un total de 319 pacientes, con incidencia de 10,7 % de apendicitomías negativas; siendo el que obtuvo mayor sensibilidad (100 %) y precisión (97,9 %) el modelo de IA, por otro lado, para la especificidad (100 %) fue mejor la combinación de la Escala de Alvarado y los criterios ecográficos; por lo tanto, los investigadores concluyeron que el modelo IA es un sistema prometedor con predicción muy alta en comparación de los otros modelos.

Carrasco et al. (40) en el 2022, presentaron una investigación titulada ¿Es capaz “ChatGPT” de aprobar el examen MIR de 2022? Implicaciones de la inteligencia artificial en la educación médica en España; teniendo como objetivo evaluar la capacidad de rendimiento de ChatGPT en el examen de acceso a la Formación Sanitaria Especializada en España. Para esto, se consideró las 200 preguntas del examen MIR del 2022 que fueron introducidas en el programa con la pregunta: “¿cuál es la respuesta correcta a la siguiente pregunta del MIR 2022 en España?”. El ChatGPT tuvo el 51,4 % de aciertos de manera general y 54,8 % de aciertos en preguntas que no tenían imágenes, por lo tanto, se pudo concluir que acertó 69 preguntas con lo cual podría haber sido elegible para alguna especialidad médica según los resultados del MIR 2021.

Moreno et al. (41) en el 2017, desarrollaron una investigación denominada Una nueva herramienta de inteligencia artificial para evaluar los síntomas en pacientes que acuden a urgencias: la aplicación Mediktor en España; tenía como objetivo el analizar la concordancia entre los diagnósticos emitidos por la aplicación Mediktor y los de un médico tratante, así como, evaluar la utilidad de esta aplicación en pacientes que buscan atención de emergencia. Para eso, se consideraron 622 pacientes del servicio de urgencia de un hospital universitario, el aplicativo Mediktor se presentó a los pacientes, quienes respondieron directamente las preguntas planteadas por el aplicativo; Mediktor clasificó a todos los pacientes en 10 posibles diagnósticos, siendo el estándar de oro el diagnóstico médico, obteniéndose una sensibilidad de 92 % y especificidad de 91 %. Por lo tanto, se concluyó que este aplicativo es fiable como ayuda diagnóstica para las condiciones más comunes que se trata en un servicio de urgencias y que el sistema es bastante amigable con los usuarios.

Park et al. (20) en el 2015 presentaron una investigación denominada Diagnóstico de apendicitis aguda mediante redes neuronales artificiales en Corea, con el objetivo de

proponer un sistema de diagnóstico de apendicitis utilizando redes neuronales artificiales (RNA). Para ello, se usaron los datos de 801 pacientes para elaborar redes neurales artificiales: función de base radial (RBF), red neuronal multicapa (MLNN) y red neuronal probabilística (PNN); y compararlo con la escala Alvarado. Se tuvo como resultado que la precisión de RBF fue 99,8 %, PNN 99,41 %, MLNN 97,84 % y escala de Alvarado 72,19 %, por otro lado, las curvas ROC fueron RBF 0,998, PNN 0,993, MLNN 0,985 y Alvarado 0,633. Por todo lo mencionado, se concluyó que los modelos de redes neuronales artificiales tuvieron un buen desempeño y fueron mejores que la escala de Alvarado.

Velázquez et al. (42) en el 2010, investigaron acerca de Evaluación prospectiva de la Escala de Alvarado en el diagnóstico de apendicitis aguda en México, su objetivo fue determinar la utilidad y eficacia de la Escala de Alvarado para el diagnóstico de apendicitis aguda, utilizando pruebas clínicas y de laboratorio. Para ello, se incluyeron 113 pacientes de un hospital regional que tenían como diagnóstico apendicitis aguda y que operaron de apendicectomía. Se les aplicó la Escala de Alvarado. Además, se buscó el reporte histopatológico para confirmación diagnóstica. Se encontró que la Escala de Alvarado tiene 99 % de sensibilidad, 72 % de especificidad, VPP de 92 %, VPN de 93 % y exactitud de prueba del 92 %. Por lo tanto, se concluyó que el score de Alvarado puede considerarse como un instrumento útil para la valoración de la apendicitis aguda.

Prabhudesai et al. (43) en el 2008, investigaron acerca de Redes neuronales artificiales: una ayuda útil en el diagnóstico de la apendicitis aguda en Londres, siendo su objetivo evaluar el papel de las redes neuronales artificiales (RNA) en el diagnóstico de apendicitis en pacientes que presentan dolor agudo en la fosa ilíaca derecha (FID) y comparar su desempeño con la evaluación realizada por médicos experimentados y la puntuación de Alvarado; donde se buscaba comparar el diagnóstico médico, la escala de Alvarado y la red neuronal artificial. Para ello, se entrenó una red neuronal artificial, luego se recolectaron información de 60 pacientes con sospecha de apendicitis, obteniéndose que la red neuronal obtuvo 100 % de sensibilidad, 97,2 % especificidad, 96 % de VPP y 100 % VPN, 36 de likelihood ratio positivo y 0 de likelihood ratio negativo, por lo tanto, se concluyó que la red neuronal artificial puede ser una estrategia eficaz para el diagnóstico de apendicitis y así evitar apendicetomías negativas.

### **2.1.2. Antecedentes nacionales**

Quispe et al. (44) en el 2023 realizaron una investigación titulada Explorando la viabilidad de la integración de la inteligencia artificial en la atención médica pediátrica; teniendo como objetivo el evaluar la precisión del modelo ChatGPT en dar pautas a padres con hijos enfermos. Por ello, se realizó una lista con 14 preguntas más frecuentes de

cuidadores de niños y estas fueron ingresadas en el ChatGPT-3 y las respuestas las evaluaron 12 médicos pediatras según la escala de Likert, obteniéndose como resultado que el 90 % de los médicos estuvo de acuerdo o muy de acuerdo con las respuestas brindadas por el ChatGPT-3. Por lo tanto, se pudo concluir que el ChatGPT es una herramienta con mucho potencial para dar orientación y guía en salud pediátrica.

Meléndez et al. (45) en el 2019 investigaron sobre la Sensibilidad y especificidad de la Escala de Alvarado en el diagnóstico de pacientes atendidos por apendicitis aguda en Hospitales del Cusco, cuyo objetivo fue determinar la sensibilidad y especificidad de la Escala de Alvarado para el diagnóstico de pacientes atendidos con apendicitis aguda en Hospitales del Cusco durante el periodo 2011 al 2016. Para ello, se analizaron con la escala de Alvarado 316 historias de pacientes que sufrieron apendicitis aguda contrastándolo con la histopatología; teniendo como resultado que 13,6 % fueron apendicetomías negativas y que la escala de Alvarado presentó 78 % de sensibilidad, 44,2 % de especificidad, 89,8 % de VPP, 13,3 % de VPN y la curva ROC fue de 61,1 %. Por lo que los investigadores concluyeron que la escala de Alvarado es un buen instrumento por su fácil aplicabilidad y accesibilidad para el diagnóstico de apendicitis aguda y así evitar las apendicetomías negativas y otras complicaciones derivadas de esta enfermedad.

Alarcón (46) en el 2012 realizó un trabajo denominado Asociación entre Escala de Alvarado y diagnóstico de apendicitis aguda complicada y no complicada según anatomía patológica en el Centro Médico Naval, que tuvo como objetivo determinar la asociación entre la Escala de Alvarado y diagnóstico de apendicitis aguda complicada y no complicada según anatomía patológica en el Centro Médico Naval. Para esto se revisó historias clínicas de 116 de pacientes que se le realizaron apendicetomías. Teniendo como resultado que en un puntaje mayor igual de 7 en la escala de Alvarado tuvo una sensibilidad de 71,15 % y especificidad de 56,67 % para diagnosticar apendicitis complicada, además, se clasificó correctamente a 63,39 % pacientes y se tuvo un valor bajo la curva de 0,68 en la curva ROC. Por lo tanto, se concluyó que, si existe relación entre la escala de Alvarado, el diagnóstico de apendicitis aguda y la anatomía patológica, así como se observó que a mayor el valor de la escala de Alvarado, aumentó la probabilidad de apendicitis complicada.

## **2.2. Bases teóricas**

### **2.2.1. Inteligencia artificial**

#### **2.2.1.1. Historia**

La inteligencia artificial tiene sus bases en las matemáticas, la estadística y las ciencias computacionales. Los primeros modelos, como el “Discriminante Lineal de Fisher” (1936), que utilizó un método estadístico para la clasificación de datos, llegaron a

sentar las bases a lo que ahora conocemos como aprendizaje automatizado. Este enfoque identifica patrones y relaciones entre variables, un principio fundamental para los algoritmos de inteligencia artificial actual. (47)

En 1936, bajo el contexto de la Segunda Guerra Mundial, Alan Turing actualmente considerado uno de los padres de la inteligencia artificial por el desarrollo de su revolucionaria Máquina de Turing, un modelo teórico con la capacidad de realizar cualquier cálculo matemático a través de un programa. En 1950, también publicó un artículo llamado “Computing Machinery and Intelligence”, donde plantea preguntas filosóficas fundamentales para el desarrollo de la inteligencia artificial como “¿Pueden las máquinas pensar?” en el mismo propone la famosa prueba de Turing como un método para la evaluación de la inteligencia de las máquinas. (24,47)

En 1956, durante la conferencia de verano de Dartmouth College, donde John MacCarthy, Marvin Minski, Nathaniel Richester y Claude Shannon conciertan el término “inteligencia artificial”. En esta conferencia se define a la inteligencia artificial como “La ciencia y la ingeniería para fabricar máquinas inteligentes”. Este es el punto de partida del intento del hombre de emular su propia inteligencia a través de algoritmos y modelos computacionales. (20,44)

Desde entonces hasta la actualidad, la inteligencia artificial ha evolucionado, siendo ahora una rama multidisciplinaria integrando avances de la estadística, informática y neurociencia. Este desarrollo continuo ha permitido su aplicación en una variedad de campos, desde la medicina hasta la exploración espacial. (21,44)

#### **2.2.1.2. Definición**

Se define como la capacidad de una máquina para realizar tareas anteriormente requerían inteligencia humana. Estas tareas incluyen desde el razonamiento, resolución de problemas, aprendizaje, percepción y procesamiento del lenguaje natural. Estos conceptos no solo se restringen en imitar, sino que llegan en algunos casos a superar las capacidades humanas (24,47). La definición también incluye la capacidad de aprender de manera autónoma mediante análisis de datos, lo que permite a las máquinas adaptarse a nuevos desafíos y mejorar su desempeño con el tiempo. Este aprendizaje puede ser supervisado, no supervisado o basado en refuerzo, dependiendo del tipo de tarea y datos disponibles. (47)

#### **2.2.1.3. Funciones y aplicaciones en salud**

La inteligencia artificial encuentra sus aplicaciones prácticas en diversos campos, desde la detección de enfermedades hasta la optimización de recursos en

medicina. En cuanto a procesamiento de datos clínico, apoya al análisis de grandes volúmenes de datos para la detección de patrones de esta manera dando información de decisiones médicas. El diagnóstico y pronóstico médico se ve beneficiado de uso de algoritmos de aprendizaje automático para el diagnóstico de enfermedades a partir de imágenes o datos estructurados, logrando prever resultados. Además, la inteligencia artificial, utilizándose como chatbot, mejora la interacción personalizada entre profesional de salud y los pacientes. (24,47)

La inteligencia artificial está transformando la medicina mediante el análisis de datos médicos para mejorar la precisión diagnóstica y la eficacia de los tratamientos; del mismo modo, facilita la gestión de la salud, ayuda en la prevención de enfermedades y mejora la calidad de vida de los pacientes (4). En el diagnóstico, la IA analiza imágenes radiológicas para detectar enfermedades como la neumonía y el cáncer de mama con mayor precisión que los métodos tradicionales (48,49). También se aplica en el desarrollo de fármacos, acelerando la investigación y el desarrollo de nuevas terapias. La IA se utiliza para desarrollar robots quirúrgicos, capaces de realizar cirugías con mayor precisión y ser menos invasivos (50). Además, la IA puede personalizar los planes de tratamiento para cada paciente, teniendo en cuenta sus características individuales y su respuesta a los medicamentos.

En el contexto de la apendicitis aguda, la IA ofrece métodos de diagnóstico y enfoque más precisos, disminuyendo la tasa de intervenciones quirúrgicas innecesarias. Algoritmos de aprendizaje automático, basados en datos demográficos y de laboratorio, alcanzan alta sensibilidad y especificidad en la distinción entre apendicitis aguda y complicada (20,26). Estos modelos, entrenados con grandes flujos de datos, superan en precisión a las escalas predictivas tradicionales, prometiendo mejorar las estadísticas diagnósticas y reducir la morbilidad. (45)

### **2.2.2. ChatGPT**

ChatGPT es un modelo de lenguaje de inteligencia artificial desarrollada por OpenAI. Se basa en el uso de técnicas de procesamiento del lenguaje natural y se somete a un extenso entrenamiento con un gran volumen de texto destinado a producir respuestas a las consultas o ideas planteada por los usuarios. Este modelo pertenece a la familia GPT-3.5, que se considera uno de los modelos de lenguaje más avanzados disponibles hasta la fecha. Debido al aprendizaje automático, este tipo de modelos de procesamiento generador de lenguaje natural (NLG), como es ChatGPT, tienen la capacidad de realizar una amplia gama de funciones en las que se encuentran, su capacidad de condensación de datos,

ofrecer información adicional, traducción de textos, efectuar previsiones, creación de código y la generación de discursos con una alta exactitud. (51)

El uso de ChatGPT debe seguir ciertos parámetros, primero se debe realizar el registro en la aplicación, segundo, una vez se encuentre registrado la herramienta enviará un correo de verificación corroborando que puede ingresar. Luego, deberá aprovechar este modelo de lenguaje para escribir una pregunta con un adecuado prompt, no debe obviar revisar la respuesta de ChatGPT en términos de relevancia y precisión. En caso de ser necesario deberá continuar con el seguimiento de preguntas adicionales para obtener aclaraciones y más información. (52)

### **2.2.3. Prompt**

Los prompts son frases o preguntas empleadas para orientar a un modelo de lenguaje, como es ChatGPT, con la finalidad de obtener de esta una respuesta. La efectividad de los prompts es fundamental para tener una conversación exitosa en ChatGPT, estos deberán encontrarse debidamente formuladas con mucha precisión para dirigir la conversación de manera efectiva, garantizando que se aborden los temas de interés. Por otro lado, los prompts mal formulados pueden llevar a conversaciones sin rumbo y poco productivas. (51)

Como la finalidad de ChatGPT es producir respuestas precisas y de calidad adecuada. Se puede decir que existe una fórmula para crear un rápido e ideal prompt que genere los resultados deseados. (52)

$$[\textit{Contextualización}] + [\textit{Especificación de Información}] + [\textit{Finalidad}] + \dots \\ \dots + [\textit{Formato de contestación}] = [\textit{Respuesta perfecta}]$$

### **2.2.4. Apendicitis aguda**

#### **2.2.4.1. Historia**

En 1735 se realizó la primera apendicectomía, esta fue realizada por Claudius Amyand un cirujano francés quien la realizó en un niño de 11 años con un saco herniario en la ingle perforado por una aguja. Pese a haberse encontrado y descrito casos de apéndices perforadas en necropsias, la recomendación de extirpación y la caracterización de la clínica de esta afección recién fue hecha pública en 1886 por el trabajo de Reginald Heber Fitz miembro perteneciente a la Universidad de Harvard. En la posterioridad se generaron avances importantes en la cirugía de la apendicitis como el de McBurney quien describió una incisión muscular además de la clásica técnica de extirpación en 1894, así como la primera extracción de apéndice a través de laparoscopia

fue hecha por Kurt Semm en el año de 1982, siendo actualmente el estándar quirúrgico para abordar este tipo de patología. El uso de antibióticos de amplio espectro, la radiología intervencionista y mejoras en la atención del paciente críticos lograron ayudar al tratamiento de la apendicitis. (53)

#### **2.2.4.2. Definición**

Es la inflamación del apéndice cecal, se trata de la principal razón detrás del dolor abdominal repentido, es decir agudo o dolor abdominal que requiere intervención quirúrgica. Comúnmente se presenta en población de edades comprendidas desde los 10 a los 30 años, sin embargo, puede surgir a cualquier edad. (54)

Es la inflamación del apéndice cecal, se trata de la principal razón detrás del dolor abdominal repentido, es decir agudo o dolor abdominal que requiere intervención quirúrgica. Comúnmente se presenta en población de edades comprendidas desde los 10 a los 30 años, sin embargo, puede surgir a cualquier edad. (54)

#### **2.2.4.3. Epidemiología**

A nivel mundial se sabe que la incidencia de apendicitis aguda se ha reducido desde 1940. En países con mayor desarrollo esta enfermedad se presenta en tasas de 5,7 a 50 casos por cada 10000 pobladores por año, teniendo mayoría de casos en personas de entre 10 a 30 años. También es importante resaltar que la geografía y sus diferencias hacen que esta enfermedad tenga un riesgo a lo largo de la vida de 9% en Estados Unidos de Norteamérica, 8 % situados en Europa y un 2 % ubicados en África. Asimismo, es importante resaltar la relación que tienen los casos de perforación y las edades jóvenes, contando estos con un porcentaje de casos de entre el 40 % y 57 %, siendo aún mayor en individuos de con edades superiores a los 50 años contando estos con tasas de 55 % al 70 %. (55)

En el Perú, ocupa el segundo lugar en cuanto a motivos que llevan a la hospitalización, con tasas de incidencia al año de alrededor del 3 % siendo una gran amenaza para la salud estatal. En el mismo lapso, la tasa de apendicitis dificultosa fue de 8 de cada 100000 personas, además la tasa incidencia de apendicitis no complicada resultó ser de 34.87 casos de cada 100000 pobladores, en la región San Martín esta tasa resultó siendo de 7,80 casos de cada 10 mil personas. (56)

#### **2.2.4.4. Fisiopatología**

La apendicitis ocurre por la obstrucción interior del apéndice cecal, el cual es susceptible por tener un diámetro mucho más estrecho que su longitud. Cuando su

estrecho proximal es bloqueado, la presión distal se incrementa por el conjunto de moco y gas derivado de las bacterias que contiene en su interior, esta expansión gradual compromete al flujo sanguíneo, lo cual provoca isquemia en el espesor del apéndice por falta de oxígeno, lo que conlleva a la perforación del tejido. La proliferación de bacterias del apéndice es causada por la estasis bacteriana distal a la obstrucción, llevando al incremento de carga bacteriana en casos de perforación. El tiempo que lleva a la perforación de esta víscera es variable, los abscesos en las regiones colindantes del apéndice suelen ser comunes, a veces llegando a perforación libre que resultará en una peritonitis generalizada. (53)

#### **2.2.4.5. Etiología**

Las causas de la obstrucción luminal del apéndice son diversas e incluyen la estasis fecal y los fecalitos, además de neoplasias, restos de fruta o materia vegetal, hiperplasia linfoide, ingesta de bario e incluso la presencia de parásitos como son el áscaris u oxiuros. (53)

#### **2.2.4.6. Clínica**

Los pacientes que padecen de apendicitis aguda experimentan dolor abdominal inespecífico, es decir difuso, comenzando alrededor del ombligo que es consecuencia de la estimulación de vías sensitivas, también denominadas vías aferentes, durante el progreso de la gradual expansión del apéndice. Es frecuente la pérdida de apetito y las náuseas a veces acompañadas con vómito. También se pueden experimentar episodios de constipación o diarrea. Según progrese la enfermedad y el extremo distal del apéndice se inflame esto generará irritación peritoneal lo que hará que el dolor se ubique en el cuadrante inferior derecho, siendo este el indicador más fidedigno de este padecimiento. Es importante resaltar que, pese a que los síntomas son representativos de esta enfermedad, a veces puede manifestarse de manera inusual por la presentación del apéndice u otros. (53)

#### **2.2.4.7. Diagnóstico**

El diagnóstico de la apendicitis aguda se da principalmente por la clínica, sin embargo, se ha utilizado vario métodos para mejorar la certeza diagnóstica y de esa manera reducir las apendicetomías innecesarias. Estos métodos incluyen la cuenta leucocitaria y la proteína C reactiva, aunque uno y otro son marcadores inflamatorios inespecíficos y no se pueden considerar del todo confiables. Otros enfoques diagnósticos comunes son la ultrasonografía, la laparoscopia diagnóstica y la tomografía axial computarizada. La ultrasonografía cuenta con una sensibilidad del 84 % y una especificidad

del 92 % sin embargo, su precisión es operador dependiente. La tomografía axial computarizada, por otro lado, muestra una alta capacidad para detectar apendicitis aguda, con una sensibilidad del 77 %, especificidad del 100 % y un valor predictivo positivo del 96,4 %. La laparoscopia es muy beneficioso en los casos de pacientes femeninas cuyas patologías pélvicas pueden simular la apendicitis aguda, no obstante, su disponibilidad es limitada además de ser costosa en comparación a las técnicas quirúrgicas convencionales. Pese a la creciente utilización de la TAC y la laparoscopia, no se ha observado una disminución significativa en tasa de perforación apendicular ni en apendicectomías innecesarias. Se han propuesto escalas de diagnóstico temprano basados en signos y síntomas, como MANTRELS (por sus siglas en inglés, es también conocido como escala de Alvarado), que asigna puntaje de diferentes características clínicas, en caso de contar con puntajes de 7 puntos a más indica la necesidad cirugía, mientras que puntajes de 5 o 6 sugieren probable apendicitis y requiere una evaluación continua y puntajes de 1 a 4 indica muy poca probabilidad de apendicitis. (53)

Si bien las escalas son una buena opción de diagnóstico este se debe individualizar por existir variaciones en la ubicación del apéndice y diferencias que existan entre cada paciente; en la actualidad la escala de Alvarado es bien considerada por el personal de salud y de fácil implementación. Esta escala fue propuesta en 1986 como una herramienta diagnóstica para la apendicitis aguda, la misma cuenta con ocho apartados que son: migración del dolor, falta de apetito, náuseas y/o vómitos, además de neutrofilia superior al 75 %, signo de rebote positivo, fiebre mayor de 38°C, signos de defensa en el cuadrante inferior derecho y leucocitosis, los puntajes se encuentran distribuidos y catalogados de la siguiente manera: puntajes de 0 a 4 son considerados como con poca probabilidad de que el paciente presente apendicitis, de 5 a 7 como probabilidad media y de 8 a 10 alta probabilidad de que este paciente tenga apendicitis, siendo necesario cirugía inmediata para estos últimos. (45,57)

#### **2.2.4.8. Exámenes auxiliares**

##### **- Laboratoriales**

Se solicita hemograma para evidenciar leucocitosis de entre 10500 y 13500, VSG, glucemia, urea y coagulograma, además de sedimentación de orina que comúnmente se muestra patológico mostrarse patológico en casos de apéndices retrocecales (54). Por lo general la leucocitosis se presenta junto a una “desviación a la izquierda”, que significa un aumento de neutrófilos y a veces glóbulos inmaduros, esto se encuentra en aproximadamente el 90 % de casos de apendicitis aguda, sin embargo, el 10 % de los casos se encontrará en estándares normales, esto no debe ser considerado

suficiente para descartar apendicitis. La prueba de orina suele encontrarse normal, si este contara con esterazas y piuria esto no deberá ser usado para descartar el diagnóstico de apendicitis. Cabe resaltar que en mujeres en edad fértil es obligatorio realizar pruebas de embarazo. Se han investigado varios biomarcadores además del recuento de glóbulos blancos, como la proteína C reactiva, la procalcitonina, la interleucina 6, entre otros, sin embargo, al parecer la proteína C reactiva es la más sensible de los biomarcadores, sin embargo, ninguno ofrece suficiente especificidad diagnóstica para definir apendicitis. (53)

#### - **Otros exámenes auxiliares**

La radiografía de tórax cuenta como indicación la sospecha de perforación de víscera hueca. La radiografía de abdomen puede mostrar asas intestinales delgadas dilatadas a nivel de cuadrante inferior izquierdo. La ecografía abdominal cuenta con sensibilidad del 83 % y especificidad del 93 %, además es una gran ayuda en el diagnóstico diferencial. El electrocardiograma es esencial para la valoración quirúrgica descartando patología cardíaca. (54)

#### **2.2.4.9. Diagnósticos diferenciales**

La apendicitis debe ser considerado como una posibilidad en todo paciente con dolor abdominal sin apendicectomía previa. Se debe resaltar que parte importante para la limitación de posibles diagnósticos son la edad y el sexo del paciente. En niños debería considerarse posibilidades como la adenitis mesentérica, gastroenteritis aguda, invaginación intestinal, persistencia del divertículo Meckel, enfermedad inflamatoria intestinal y en casos de pacientes varones la torsión testicular puede ser una causa. En mujeres en edad fértil los diagnósticos se amplifican pues podrían tratarse de problemas ginecológicos confundidos con apendicitis como quistes ováricos rotos, dolor intermenstrual que coincide con la ovulación, torsión ovárica, embarazo ectópico, endometriosis y enfermedad inflamatoria pélvica. En ancianos pueden ser causas la diverticulitis aguda y a enfermedades malignas. En pacientes neutropénicos las causas pueden ser la enterocolitis neutropénica. La nefrolitiasis e infecciones urinarias pueden ser causa en ambos sexos. (53)

#### **2.2.4.10. Tratamiento**

El tratamiento no quirúrgico con antibióticos es una alternativa propuesta en caso apendicitis simple, esta propuesta hecha en un metanálisis mostró efectividad, sin embargo, el 25 % al 30 % de los pacientes requirieron cirugía en menos de 1 año. Se deben elegir antibióticos que cubran bacterias en el tracto digestivo, tomando en cuenta las resistencias locales a los antibióticos, en este estudio se utilizó combinaciones de

antibióticos intravenosos como meropenem y metronidazol, además de tomar en cuenta la estabilidad del paciente para continuar con el consumo de antibióticos por vía oral. La cirugía tradicionalmente se recomienda lo antes posible para reducir riesgo de perforación, sin embargo, estudios han sugerido que retrasos cortos (de menos de 24 horas) en pacientes estables y no complicados no aumentan el riesgo que el apéndice se perfora. En Estados Unidos la mayoría de las cirugías son realizadas por laparoscopia con una estancia hospitalaria corta y bajas tasas de complicaciones, este abordaje está asociado a menor dolor postoperatorio que la cirugía abierta. En cuanto al manejo del dolor, el uso de analgésicos no afecta la precisión del diagnóstico en el examen físico, está recomendado el uso de analgésicos incluso antes del diagnóstico etiológico del dolor abdominal, para de esta manera garantizar el bienestar del paciente. Se debe realizar una evaluación efectiva de la intensidad del dolor, elegir el analgésico adecuado y vigilar el progreso de disminución de la intensidad del dolor, además de los efectos secundarios de los analgésicos. (58)

#### **2.2.5. Prueba diagnóstica**

La prueba diagnóstica está definida como un procedimiento, que puede ser más o menos complicado, diseñado para la identificación de una condición específica que puede presumirse como enfermedad, que no puede ser percibida directamente a través de los cinco sentidos básicos. (59)

#### **2.2.6. Precisión diagnóstica**

##### **2.2.6.1. Validez de una prueba diagnóstica**

Por la necesidad de la ciencia de hacer de la duda descriptible y cuantificable, y no somera deducción, el ser humano ha ideado formas de reducir la incertidumbre propia de las enfermedades. La validez de una prueba diagnóstica se refiere como la potencial que tiene una prueba de medir o identificar lo que se desea ponderar. En tanto a la validez se puede clasificar en tipos, como son: la validez de contenido se refiere a si los elementos contenidos en una prueba son relevantes y representa lo que se desea medir, la validez de constructo comprende a la medición de un fenómeno externo fijado. La validez de criterio se define como una medida se correlacionará con la prueba Gold Standard, es decir la prueba de predilección para el fenómeno, este valor dado determinará cuál es su capacidad para predecir un fenómeno visible. (59)

##### **- Sensibilidad**

Se puede definir como la habilidad que tiene una prueba para detectar de manera precisa los individuos que se encuentran afectados por una enfermedad. Esto

equivale a la cantidad de individuos que demuestran un resultado positivo a la prueba y que realmente se encuentran afectados por la enfermedad, dividido entre la cantidad total de individuos que verdaderamente tienen la enfermedad. (59)

$$\text{Sensibilidad} = \frac{\text{Verdaderos Positivos}}{\text{Total de Enfermos}}$$

- **Especificidad**

Se define como la facultad que cuenta una prueba para detectar a los individuos sanos, es decir que no se encuentran afectados por la enfermedad, y este se logra mediante la proporción de individuos que cuentan con resultados negativos en la prueba, pero realmente no se encuentran afectados por la enfermedad, dividido entre la cantidad total de individuos que se encuentran libres de la enfermedad, es decir los que se encuentran saludables. (59)

$$\text{Especificidad} = \frac{\text{Verdaderos Negativos}}{\text{Total de Sanos}}$$

- **Valor predictivo positivo**

Esta referido a la probabilidad que tiene una persona de encontrarse enfermo cuando se muestra como una prueba positiva. Es decir, la probabilidad que se encuentra orientada a que un individuo arroje un resultado positivo y se encuentre verdaderamente enfermo. (59)

$$\text{Valor Predictivo Positivo (VPP)} = \frac{\text{Verdaderos Positivos}}{\text{Total de Positivos}}$$

- **Valor predictivo negativo**

Se refiere a la probabilidad de que una persona reciba un resultado negativo en la prueba, sin embargo, este no este afectada por la enfermedad, es decir se encuentre sana. (59)

$$\text{Valor Predictivo Negativo (VPN)} = \frac{\text{Verdaderos Negativos}}{\text{Total de Negativos}}$$

- **Curva ROC**

Representa visualmente la sensibilidad además de la especificidad de varios puntos de corte de una prueba diagnóstica que tenga escala de medición continua. Esta se crea al unir puntos de corte, donde el eje vertical representa la sensibilidad y en eje horizontal representa 1 – especificidad de cada uno de dichos puntos. Estos ejes contienen valores que se encuentran entre 0 a 1, que equivalen al 0 % y al 100 % respectivamente.

Para proporcionar estos puntos de referencia, todas las curvas ROC incluyen una línea que van desde el punto 0.0 hasta el punto 1,1, conocida como la diagonal de referencia, es decir, la línea de no discriminación. (60)

#### - **Likelihood Ratio**

Es un indicador fijo se utiliza para evaluar la precisión de una prueba en ámbito clínico, tiene como ventaja ser independiente de la prevalencia de las enfermedades de una población. Indica como un resultado de una prueba aumenta o disminuye la probabilidad de tener la enfermedad en relación con la probabilidad previa. Es decir, es respuesta a la pregunta ¿cuál es la probabilidad de tener o no la enfermedad? Esta se encuentra compuesta por Likelihood Ratio Positiva y Negativa. (59)

##### • **Likelihood ratio positiva**

Se calcula la probabilidad de tener un resultado positivo en individuos con la enfermedad y se la divide con la probabilidad de tener un resultado positivo en personas sanas. Es decir, es la relación entre la proporción de VP (la sensibilidad) y la proporción de FP (1- especificidad). Esta nos dice la probabilidad de tener la enfermedad si el resultado de la prueba es positivo. (59)

$$\text{Likelihood Ratio Positiva} = \frac{\text{Sensibilidad}}{1 - \text{Especificidad}}$$

##### • **Likelihood ratio negativa**

Es calculado al dividir la probabilidad de obtener resultados negativos cuando se tiene la patología sobre la probabilidad de obtener un resultado negativo cuando no se tiene la patología. Es decir, es la proporción entre FN (1 – sensibilidad) y VN (especificidad), este indicará cuan probable es que un paciente patológico obtenga un resultado negativo en comparación de un paciente sin patología que también obtiene un resultado negativo. (59)

$$\text{Likelihood Ratio Negativa} = \frac{1 - \text{Sensibilidad}}{\text{Especificidad}}$$

#### - **Índice de exactitud**

Es una métrica que estima la relación de predicciones acertadas hechas por un modelo de clasificación, su cálculo se basa en dividir la cantidad de predicciones acertadas con la totalidad de predicciones efectuadas. (59)

$$\text{Índice de Exactitud} = \frac{VP + VN}{VP + VN + FP + FN} \times 100$$

### - **Score F1**

También conocido como medida F, es un métrico empleado en ciencia de datos y algoritmos de inteligencia artificial para medir la eficacia de un modelo o prueba diagnóstica. Este se obtiene mediante la media armónica entre la sensibilidad (habilidad para identificar correctamente los positivos) y el valor predictivo positivo (proporción de positivos identificados que son correctos), combinando ambos aspectos en solo valores representativos. Su rango va de 0 a 1, donde 1 indica un desempeño perfecto. Es especialmente valioso en escenarios con clases desbalanceadas o cuando se busca un balance general en el rendimiento del modelo. (29)

$$Score F1 = 2 \times \frac{VPP \times Sensibilidad}{VPP + Sensibilidad}$$

### 2.2.6.2. **Confiabilidad**

El término “confiabilidad” o “reproducibilidad” hace referencia a la capacidad de error involucrado al llevar a cabo cualquier tipo de medición. (59)

### - **Coficiente de correlación de Matthews**

También conocido como coeficiente de correlación de Pearson resulta ser más compleja pues refleja una correlación entre los valores reales y los predichos por la prueba diagnóstica. La fórmula de este tiene componentes clásicos de las pruebas diagnósticas resultando finalmente en valores entre -1 y 1, siendo -1 un modelo que no está realizando una clasificación adecuada en absoluto, 0 significa que este modelo está generando valores al azar y 1 sugiere que la prueba está generando una clasificación perfecta. (61)

$$MMC = \frac{VP \times VN - FP \times FN}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}}$$

## 2.3. **Definición de términos básicos**

**Anamnesis:** es un proceso de la evaluación clínica que implica la realización de un interrogatorio con la finalidad de identificar de manera personal al paciente, apoya en la comprensión de los problemas de salud que aquejan a este. Además, es necesario para recopilar información sobre el historial médicos de este, establecer los factores ambientales, individuales y ambientales que pueden llegar a ser pertinentes en el tratamiento de las patologías del paciente. (62)

**Examen físico:** se lleva a cabo con el propósito de evaluar de manera comprobable y objetiva el estado físico y mental del individuo en un momento específico. Se emplean los cuatro métodos tradicionales de exploración clínica: la observación, el tacto, la percusión y la auscultación. Aunque a menudo, el personal de salud se guía por la información proporcionada en la historia clínica orientada hacia el órgano o área del cuerpo afectado, es esencial que el examen se realice de manera exhaustiva dirigiéndose de cefálico a podálico en busca de irregularidades. (63)

**Antropometría:** se trata de un marcador objetivo que se utiliza con la finalidad de medir con precisión los cambios de las características físicas y la composición corporal de personas de diversidad de edades e incluso estados nutricionales. (64)

**Diagnóstico:** es un proceso de deducción basado generalmente en la presentación de síntomas y signos denominados cuadro clínico, cuya finalidad yace en la identificación de patologías que perjudican a los aquejados. (65)

**Signos vitales:** son manifestaciones observables en el exterior y reflejan las funciones esenciales del cuerpo, ya sea la respiración, el metabolismo e incluso la circulación. Estos pueden ser evaluados ya sea mediante el examen físico, como también por medio de instrumentación. Las alteraciones de estos signos corresponderían a cambios en el organismo, algunos de carácter fisiológico y otros de características patológicas. (66)

**Exámenes auxiliares:** el diagnóstico clínico puede llegar a ser más preciso al integrar datos proporcionados por exámenes auxiliares. Cuando se solicitan e interpretan adecuadamente, estos resultan ser muy importantes, puesto que ayuda a disminuir la incertidumbre diagnóstica en diversos grados e incluso ayudan a respaldar las suposiciones clínicas, brindando así una mayor confianza tanto al paciente como a los médicos al tomar decisiones médicas. (67)

## **CAPÍTULO III**

### **HIPÓTESIS Y VARIABLES**

#### **3.1. Hipótesis**

##### **3.1.1. Hipótesis general**

La precisión diagnóstica del ChatGPT-4 para el diagnóstico de apendicitis aguda en el servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022 será adecuada.

##### **3.1.2. Hipótesis específicas**

El ChatGPT-4 demostrará ser válido para el diagnóstico de apendicitis Aguda en el Servicio de Emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022, si cumple los siguientes criterios: sensibilidad > 80 %, especificidad > 80 %, valor predictivo positivo (VPP) > 70 %, valor predictivo negativo (VPN) > 90 %, likelihood ratio positivo (LR+) > 10, likelihood negativo (LR-) < 0.2, índice de exactitud > 90 % y Score-F1 > 0.8.

El ChatGPT-4 demostrará ser confiable para el diagnóstico de apendicitis aguda en el servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022, si alcanza un coeficiente de correlación de Matthews (MCC) > 0.5.

#### **3.2. Variables de la investigación**

La presente investigación tiene dos variables, siendo la variable dependiente la precisión diagnóstica y la variable independiente es el ChatGPT-4.

### 3.3. Operacionalización de la variable

VARIABLE	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIONES	INDICADOR	TIPO DE DIMENSIÓN
Variable Independiente: <b>Modelo ChatGPT-4</b>	Modelo de lenguaje basado en inteligencia artificial creado por OpenAI (51).	Ficha de recolección de datos basada en la historia clínica digital del Hospital Nacional Ramiro Prialé Prialé	Prompt	Instructivo o comandos previos para mejorar la precisión y el rendimiento del diagnóstico	Nominal
			Diagnóstico preoperatorio del Modelo ChatGPT - 4	Tiene apendicitis aguda No tiene apendicitis aguda	Nominal
			Diagnóstico anatomopatológico	Resultado de biopsia	Nominal
Variable Dependiente: <b>Precisión diagnóstica</b>	Grado de concordancia de informaciones recopiladas de la prueba diagnóstica estudiada y la proporcionada por la prueba de referencia, esta se manifiesta como nivel de sensibilidad más especificidad, entre otras (68).	Estadísticas de análisis de datos recolectados con las fichas de recolección de datos evaluadas a través de métodos estadísticos.	Validez	Sensibilidad	Continuo
				Especificidad	Continuo
				Valor predictivo positivo	Continuo
				Valor predictivo negativo	Continuo
				Curva de ROC	Continuo
				Likelihood ratio	Continuo
				Índice de exactitud	Continuo
			Score F1	Continuo	
Confiabilidad	Coefficiente de Correlación de Matthews	Continuo			

## **CAPÍTULO IV**

### **MATERIALES Y MÉTODOS**

#### **4.1. Métodos, tipo y alcance de la investigación**

##### **4.1.1. Método de la investigación**

En esta investigación se consideró:

El método general se realizó siguiendo el procedimiento del método científico a través del cual, los elementos básicos tales como definiciones, variables entre otros son indispensables para poder desarrollarlo y así lograr los objetivos o realizar las comprobaciones de las hipótesis planteadas. (69)

En tanto, el método específico que se consideró fue el inductivo, debido a que, en esta investigación, se tomó en cuenta los resultados particulares buscando obtener relaciones entre ellos y así poder entregar un resultado general del trabajo.(69)

##### **4.1.2. Tipo de investigación**

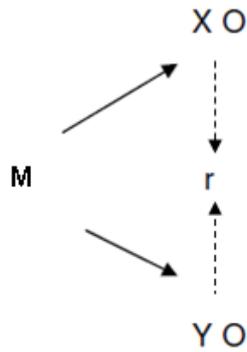
Supo (70) nos refiere que esta investigación fue observacional ya que solo se describió la realidad sin ningún tipo de intervención; así como, retrospectivo y transversal, porque los datos fueron tomados de la historia clínica y solo una vez en la investigación y, además, fue descriptiva, debido a que se recopilaron datos de variables, se describieron y los datos fueron operados estadísticamente.

##### **4.1.3. Alcance de la investigación**

El alcance que utilizamos fue descriptivo; porque nos limitamos a recolectar y medir data particular o global de manera independiente o conjunta de nuestras variables en sí, con ello mostrar nuevos ángulos para abordar una enfermedad. (1)

#### **4.2. Diseño de la investigación**

Este estudio fue no experimental cuantitativo, porque no se manipularon las variables de una manera intencional y solo se observó y analizó a través de las fuentes de información tal como están. Se utilizó el método tipo descriptivo donde usamos la frecuencia de eventos o actos de las variables; esto ayudó a categorizar la población. Además, fue de corte transversal o transeccional debido a que solo se tomó la recolección de la muestra en un único momento. (1)



### 4.3. Población y muestra

#### 4.3.1. Población

La población contemplada fue de 5619 historias clínicas de los pacientes entre 18 a 59 años que fueron atendidos en el tópic de cirugía en el servicio de emergencia del hospital Nacional Ramiro Prialé Prialé en el año 2022.

#### 4.3.2. Muestra

A diferencia del nivel de confianza del 95 % comúnmente empleado en estudios similares, en esta investigación se aseguró un nivel de confianza del 98 %, con un margen de error del 5 %. Se estimó el tamaño de la muestra utilizando el software estadístico STATA 18, que indicó que se requerían 495 historias clínicas para la realización del análisis.

La muestra para esta investigación se calculó con la fórmula de tamaño de muestra para población finita dado por STATA 18: Ecuación estadística para proporciones poblaciones

$$n = \frac{z^2(p \times q)}{e^2 + \left(\frac{z^2(p \times q)}{N}\right)}$$

Donde:

$n$  = tamaño de la muestra

$z$  = Nivel de confianza deseado

$p$  = Proporción de la población con la característica deseada (éxito)

$q$  = Proporción de la población sin la característica deseada (fracaso)

$e$  = Nivel de error dispuesto a cometer

$N$  = Tamaño de la población

Además, la muestra para este trabajo fue probabilística, debido a que, las historias clínicas se eligieron al azar por un proceso de aleatorización, permitiendo realizar un análisis representativo y robusto.

Asimismo, se tomaron en cuenta los siguientes criterios:

#### **Criterios de inclusión**

- Pacientes atendidos en el tópico cirugía del servicio de Emergencia en el Hospital Ramiro Nacional Prialé Prialé en el año 2022.
- Pacientes adultos en el rango de edad comprendido de 18 a 59 años, de ambos sexos.

#### **Criterios de exclusión**

- Pacientes gestantes.
- Pacientes con historias clínicas incompletas.
- Pacientes que no se presentaron a la atención.

### **4.4. Técnicas e instrumentos de recolección y análisis de datos**

#### **4.4.1. Técnicas**

Para el mejor desarrollo de la investigación, se empleó la técnica denominada Análisis documental donde se analizó el contenido de estos documentos a fin de investigar algún fenómeno en particular (71). Teniendo en cuenta lo antes mencionado, se consideró esta técnica como la más adecuada, debido a que se recolectó información completa de las historias clínicas de los pacientes atendidos en el tópico de cirugía del servicio de emergencia del Hospital Nacional Ramiro Prialé Prialé en el año 2022.

#### **4.4.2. Instrumento**

Se usó como instrumento la Ficha de Recolección de Datos (Anexo B) que ha sido elaborado por los investigadores según la información brindada en la historia clínica sin modificaciones, únicamente obviando datos como nombres, DNI, datos de filiación o que puedan intervenir en la confidencialidad de los pacientes.

La ficha consta de 7 dimensiones:

- Datos generales: Codificación asignada a cada participante para asegurar la confidencialidad del paciente, edad en años, sexo, fecha y hora de atención.
- Anamnesis: donde se consignó textualmente lo registrado por el médico tratante. Esta incluyó elementos como el tiempo de enfermedad, los signos y síntomas de la enfermedad actual, antecedentes médicos y quirúrgicos, así como alergias

medicamentosas. La ausencia de alguno de estos elementos no invalidó la historia clínica, siempre que se detallen los signos y síntomas de la enfermedad actual.

- Funciones vitales: se colocaron los valores correspondientes a la presión arterial en mmHg, frecuencia cardíaca en latidos por minuto, frecuencia respiratoria en respiraciones por minuto, temperatura en grados centígrados, peso en kilogramos y talla en metros.

- Examen físico: se registró textualmente y sin modificaciones lo escrito por el médico tratante, incluyendo abreviaturas u otra nomenclatura que se haya escrito en la historia clínica.

- Diagnóstico: se consideró el código según el CIE 10, de acuerdo la cantidad de diagnósticos que hayan escrito en la historia clínica.

- Exámenes auxiliares: donde se incluyeron exámenes de laboratorio como el hemograma, perfil de coagulación, bioquímicos y examen de orina. Además, se encuentra un apartado para imágenes en el que se podrán indicar los informes de radiología, ecografía, tomografía y otros exámenes relacionados.

- Otros: aquí se colocó información relevante o exámenes auxiliares que no se han registrado en los ítems anteriores.

El presente instrumento ha sido sometido a validez por jueces expertos, la cual se calculó con el alfa de Cronbach (Anexo C) considerando la participaron 5 médicos cirujanos generales (Anexo D) con amplia experiencia que trabajan actualmente en distintos hospitales y clínicas del departamento de Junín; quienes dieron su apreciación y recomendaciones al instrumento presentado, teniendo como resultado 0.96, categorizado con una excelente consistencia interna. Por lo tanto, se consideró este instrumento como válido.

#### **4.4.3. Recopilación y manejo de datos**

En este estudio, los datos recopilados de las historias clínicas fueron organizados y estructurados utilizando Google Forms, donde se colocó toda la información siguiendo el formato propuesto en la ficha de recolección de datos, permitiendo así, la sistematización de la información, evitando errores si el procesamiento fuera manual.

Una vez teniendo esa base de datos (denominada BD1 que se encuentra en el Anexo F) con información de cada historia clínica, se procedió a generar texto tipo caso clínico consignando la información del paciente y los demás datos, todo esto con el fin que sea de mejor entendimiento y comprensión para el ChatGPT-4, añadiéndole las indicaciones

precisas (prompt) (Anexo E) para obtener el diagnóstico de apendicitis aguda o la negación de este.

En esta investigación se utilizó el ChatGPT-4 en su versión ChatGPT-4o mini (con un valor de suscripción de 20 dólares mensuales). El texto generado se introdujo al cuadro de entrada y se presionó el botón de envío. El chatGPT-4 se tomó unos segundos y arrojó alguna de las dos respuestas: la afirmativa (“El paciente tiene apendicitis aguda”) o la negativa (“El paciente no tiene apendicitis aguda”). Estas respuestas fueron recopiladas en una nueva base de datos (BD2 que se encuentra en el Anexo G) donde se colocó el código de cada historia clínica, la respuesta afirmativa o negativa y el resultado de la histopatología. Este proceso se repitió hasta culminar con la totalidad de la muestra.

El Hospital Nacional Ramiro Prialé Prialé facilitó a los investigadores la relación de anatomía patológica de los pacientes que fueron operados de apendicectomías, donde se obtuvo el resultado con la confirmación de apendicitis aguda o la negación de esta, con la inscripción de Apéndice Cecal sin alteración significativas; siendo, en última instancia, el diagnóstico definitivo de Apendicitis aguda. Esta información fue colocada en BD2, siendo el diagnóstico de confirmación para esta investigación.

#### **4.4.4. Análisis de datos**

El análisis de datos en esta investigación se llevó a cabo mediante métodos descriptivos e inferenciales, con el objetivo de evaluar la precisión diagnóstica de ChatGPT-4 como herramienta para el diagnóstico de apendicitis aguda. Estos métodos se justifican por los objetivos del estudio y la naturaleza de las variables analizadas.

##### **4.4.4.1. Análisis descriptivo**

El análisis descriptivo se utilizó para resumir y caracterizar las variables principales de la investigación, como la distribución de edad y sexo. Este enfoque permitió representar nuestra muestra mediante el cálculo de medidas de tendencia central (media, mediana) y dispersión (desviación estándar, rango) para las variables, además las variables cualitativas se describieron mediante frecuencia absoluta y relativa (porcentajes). (72)

##### **4.4.4.2. Análisis inferencial**

El análisis inferencial se aplicó para evaluar la capacidad diagnóstica de ChatGPT-4 mediante comparaciones entre sus resultados y el estándar de referencia (anatomía patológica). Este tipo de análisis permitió generalizar los hallazgos de la muestra a la población total, garantizando la validez estadística de los resultados obtenidos. (73)

### **Métricas diagnósticas**

Las métricas específicas que se calcularon para evaluar la precisión diagnóstica de ChatGPT-4 incluyen:

- **Sensibilidad:** proporción de casos positivos correctamente identificados por ChatGPT-4 como apendicitis aguda.
- **Especificidad:** proporción de casos negativos correctamente identificados como no apendicitis aguda.
- **Valor predictivo positivo (VPP):** probabilidad de que los pacientes clasificados como positivos por ChatGPT-4 realmente tengan apendicitis aguda.
- **Valor predictivo negativo (VPN):** probabilidad de que los pacientes clasificados como negativos por ChatGPT-4 realmente no tengan apendicitis aguda.
- **Curva ROC:** evaluó la discriminación global del modelo al diferenciar entre casos positivos y negativos, representando el balance entre sensibilidad y especificidad.
- **Likelihood ratio:** evaluó la probabilidad de obtener resultados positivos o negativos en función del estado real del paciente.
- **Índice de exactitud:** proporción total de diagnósticos correctos realizados por ChatGPT-4.
- **Score F1:** evaluó el balance entre precisión y sensibilidad.
- **Coficiente de correlación de Matthews (MCC):** evaluó la calidad de las respuestas dadas por el ChatGPT-4 correlacionando las clasificaciones del modelo y los resultados reales.

### **Herramientas estadísticas**

El procesamiento y análisis de los datos se realizó utilizando STATA 18 para el análisis estadístico avanzado, incluyendo el cálculo de indicadores de precisión diagnóstica (sensibilidad, especificidad, VPP, VPN, likelihood ratio entre otros) además de los análisis inferenciales con intervalos de confianza al 95 %. Microsoft Excel fue empleado para la organización inicial de los datos, la realización de cálculos básicos, la creación de gráficos y tablas descriptivas para facilitar la visualización de los resultados. La combinación de análisis descriptivo e inferencial, junto con el uso de estas métricas avanzadas, asegura un enfoque sólido y riguroso en la evaluación de ChatGPT-4 como prueba diagnóstica.

#### **4.5. Consideraciones éticas**

Se garantizó la confidencialidad y privacidad de los datos obtenidos en las historias clínicas, las cuales fueron utilizadas única y exclusivamente para los fines estipulados en esta investigación. Para asegurar el anonimato de los pacientes, se implementó un proceso de codificación y aleatorización de las historias clínicas, de manera que cada registro fue renombrado e introducido por ChatGPT-4 sin incluir datos filiales, como nombres, apellidos, DNI u otra información que permitiera la identificación del paciente. Cabe destacar que nuestro instrumento fue diseñado en Google Forms sin apartados de datos de filiación, garantizando así el total anonimato de los pacientes involucrados.

Dado que se trató de una investigación retrospectiva basada en historias clínicas, no fue necesaria la obtención de consentimiento informado, situación que fue aprobada por el Comité de Ética de la Universidad Continental (Anexo H) y el Comité de Ética del Hospital Ramiro Prialé Prialé (Anexo I). Además, se siguió en todo momento el principio de beneficencia, asegurando que el potencial conocimiento generado pudiera beneficiar a futuros pacientes con apendicitis aguda, y se garantizó la transparencia en la presentación y análisis de los resultados, evitando cualquier sesgo o distorsión en la interpretación de los hallazgos. No existe conflicto de interés entre los autores.

## CAPÍTULO V

### RESULTADOS Y DISCUSIÓN

#### 5.1. Resultados

El análisis de los resultados reveló una muestra diversa de 495 de historias clínicas de pacientes atendidos en el tópico de cirugía del servicio de emergencia del Hospital Nacional Ramiro Prialé Prialé en el año 2022. En la tabla N.º 1, la distribución por género mostró una mayoría de pacientes femeninas, alcanzando el 56,97 % (282 pacientes), mientras el 43,03 % restante (213 pacientes) correspondió a pacientes masculinos.

**Tabla N.º 1. Frecuencia según sexo en el tópico de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022.**

Sexo	N	%
Femenino	282	56,97
Masculino	213	43,03
Total	495	100,00

Al evaluar la edad de los pacientes, se identificó una dispersión en el rango de edades, con una edad mínima de 18 años y una máxima de 59 años. La mediana de edad fue de 40 años, mientras que la media se situó en 40.93 años, indicando una distribución central alrededor de estos valores, así como, la moda, que representa el valor más frecuente, fue de 35 años. La desviación estándar de 10.21 años y el coeficiente de variación del 24,95 % señalan una dispersión moderada alrededor de la media, indicando cierta variabilidad en las edades de la muestra, lo que nos indica que la media es representativa. Además, en la tabla N.º 2 se aprecia la cantidad de participantes según rangos de edad, donde el 18,4 % (91 pacientes) tienen las edades de 33 a 37 años, seguido de los de 38 a 42 años con el 15,6 % (77 participantes). El rango con menor cantidad fue de 18 a 22 años con un 1,8 % (9 pacientes). Otros rangos significativos son 43 a 47 años (13,9 %) y 53 a 57 años (13,7 %). Los rangos de 28 a 32 años y 48 a 52 años muestran porcentajes menores, el 11,9 % y un 10,7 %, respectivamente. Por otro lado, los grupos etarios menos representados son 23 a 27 años (9,3 %), 58 a 59 años (4,6 %) y 18 a 22 años (1,8 %).

**Tabla N.º 2. Frecuencia según rango de edad en el tópic de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022.**

<b>Rango de edad</b>	<b>N</b>	<b>%</b>
<b>18 - 22</b>	9	1,82
<b>23 - 27</b>	46	9,29
<b>28 - 32</b>	59	11,92
<b>33 - 37</b>	91	18,38
<b>38 - 42</b>	77	15,56
<b>43 - 47</b>	69	13,94
<b>48 - 52</b>	53	10,71
<b>53 - 57</b>	68	13,74
<b>58 - 59</b>	23	4,65
<b>Total</b>	495	100,00

Los diagnósticos que se observaron con más frecuencia en esta muestra fueron los relacionados con el abdomen, tal y como se aprecian en la tabla N.º 3. Se consideraron en total 558 diagnósticos según CIE-10, incluyendo 164 diferentes. El diagnóstico más común fue “Otros dolores abdominales y los no especificados”, que representa el 10,75 % (60 casos) del total, reflejando la relevancia de los cuadros inespecíficos en el contexto de emergencias abdominales que requieren una evaluación clínica y paraclínica exhaustiva. En segundo lugar, abdomen agudo, con un 4,66 % (26 casos), resalta como una condición de alta prioridad en la atención quirúrgica. En cuanto a las condiciones traumáticas, la contusión del tórax se reportó en un 4,30 % (24 casos), mientras que los traumatismos superficiales múltiples no especificados y los traumatismos superficiales de la cabeza ocuparon un 3,94 % (22 casos) y un 3,76 % (21 casos), respectivamente, subrayando la prevalencia de lesiones traumáticas en emergencias, probablemente relacionadas con accidentes de tránsito o caídas. Por su parte, las patologías específicas del sistema biliar y abdominal, como el cálculo de la vesícula biliar con colecistitis y la colecistitis aguda, representaron un 3,76 % (21 casos) y un 3,41 % (19 casos), lo que evidencia la carga de estas afecciones en la población atendida.

**Tabla N.º 3. Frecuencia de los diagnósticos según CIE-10 en el tópico de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022.**

	<b>CIE - 10</b>	<b>N</b>	<b>%</b>
<b>R10.4</b>	Otros dolores abdominales y los no especificados	60	10,75
<b>R10.0</b>	Abdomen agudo	26	4,66
<b>S20.2</b>	Contusión del tórax	24	4,30
<b>T00.9</b>	Traumatismos superficiales múltiples, no especificadas	22	3,94
<b>S00.9</b>	Traumatismo superficial de la cabeza, parte no especificada	21	3,76
<b>R10.1</b>	Dolor abdominal localizado en parte superior	21	3,76
<b>K80.1</b>	Cálculo de la vesícula biliar con otra colecistitis	21	3,76
<b>K81.0</b>	Colecistitis aguda	19	3,41
<b>R10.3</b>	Dolor localizado en otras partes inferiores del abdomen	14	2,51
<b>M54.5</b>	Lumbago no especificado	13	2,33
	Otros diagnósticos	317	56,81
	<b>Total</b>	<b>558</b>	<b>100,00</b>

En relación con los dolores localizados, el dolor abdominal en la parte superior fue identificado en el 3,76 % (21 casos), mientras que el dolor localizado en otras partes inferiores del abdomen fue menos prevalente, con un 2,51 % (14 casos). Finalmente, el lumbago no especificado se reportó en el 2,33 % (13 casos), probablemente asociado a patologías musculoesqueléticas comunes que requieren un enfoque clínico diferenciado. Los datos resaltan la alta proporción de emergencias relacionadas con patologías abdominales inespecíficas y abdomen agudo, que requieren un manejo quirúrgico oportuno, además de la significancia de las lesiones traumáticas y las patologías biliares.

**Tabla N.º 4. Frecuencia de los pacientes con apendicitis aguda total y según sexo en el tópico de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022.**

<b>Sexo</b>	<b>AA</b>	<b>%</b>
<b>Masculino</b>	17	3.43
<b>Femenino</b>	20	4.04
<b>Total</b>	37	7.47

AA: apendicitis aguda

**Tabla N.º 5. Frecuencia de apendicitis aguda según rango de edad de los pacientes en el tópico de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022.**

<b>Rango de edad</b>	<b>N</b>	<b>%</b>
<b>18 - 22</b>	1	2,70
<b>23 - 27</b>	1	2,70
<b>28 - 32</b>	7	18,92
<b>33 - 37</b>	3	8,11
<b>38 - 42</b>	6	16,22
<b>43 - 47</b>	5	13,51
<b>48 - 52</b>	8	21,62
<b>53 - 57</b>	6	16,22
<b>58 - 59</b>	0	0,00
<b>Total</b>	37	100,0

En la muestra que se consideró para esta investigación, los pacientes que fueron diagnosticados con apendicitis aguda, según anatomía patológica, fueron 37 pacientes que corresponden al 7,47 % de la muestra. De estos, 20 fueron de sexo femenino y 17 de sexo masculino, según la tabla N.º 4. De acuerdo con los datos representados, hubo una ligera predominancia de casos en mujeres, representando aproximadamente el 54 % del total de la muestra, frente al 46 % de los hombres.

En la tabla N.º 5, se observa que el grupo etario más afectado por la apendicitis aguda corresponde a los pacientes de 48 a 52 años, representando el 21,62 % (8 pacientes), seguido por el grupo de 28 a 32 años con el 18,92 % (7 pacientes). Posteriormente, los rangos de 53 a 57, 43 a 47 y 33 a 37 años tienen frecuencias de 16,22 % (6 pacientes), 13,51 % (5 pacientes) y 8,11 % (3 pacientes) respectivamente. Los grupos de 23 a 27 años y 18 a 22 años muestran una menor proporción con 2,70 % (1 paciente) cada uno, mientras que el rango de 58 a 59 años no registró casos. Estos datos sugieren que la apendicitis aguda, en esta muestra específica, tuvo una mayor prevalencia en pacientes adultos de mediana edad, particularmente en aquellos entre los 48 y 52 años. Es relevante destacar la ausencia de casos en edades avanzadas (mayores de 58 años) y la baja incidencia en adultos jóvenes (18 a 27 años).

**Tabla N.º 6. Contingencia del Chat-GPT 4 para el diagnóstico de apendicitis aguda en el tópico de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022.**

Resultados		Respuestas de ChatGPT-4		Total
		Sin AA	Con AA	
<b>Anatomo patología</b>	Sin AA	447	11	458
	Con AA	4	33	37
<b>Total</b>		451	44	495

AA: apendicitis aguda

En la tabla N.º 6 se presentan los resultados de la comparación entre las respuestas generadas por chatGPT-4 y los hallazgos anatomopatológicos, los cuales constituyen el estándar de referencia para el diagnóstico definitivo de apendicitis aguda para esta investigación. El análisis de los datos revela que, de un total de 495 pacientes evaluados, ChatGPT-4 identificó correctamente 447 casos sin apendicitis, lo que representa una proporción significativa de clasificaciones acertadas dentro de este grupo. Del mismo

modo, en 33 pacientes se confirmó la presencia de apendicitis aguda, coincidiendo con los hallazgos histopatológicos. No obstante, se observaron 11 falsos positivos, es decir, casos en los que chatGPT-4 predijo la presencia de apendicitis, pero la evaluación anatomopatológica no lo confirmó. Estos resultados sugieren que el modelo podría sobreestimar la presencia de apendicitis en ciertos escenarios clínicos.

Por otro lado, se registraron 4 falsos negativos, lo que indica que el modelo no detectó la enfermedad en pacientes que sí fueron diagnosticados con apendicitis en la evaluación histopatológica. Este hallazgo resalta la importancia de considerar siempre la evaluación clínica complementaria y otros métodos de diagnóstico en la toma de decisiones médicas. Los valores de la tabla N.º 2 permitieron a los investigadores calcular los indicadores clave de diagnóstico de desempeño de ChatGPT-4 como herramienta de apoyo en el diagnóstico de apendicitis aguda, tales como la sensibilidad, especificidad, valores predictivos positivos y negativos, entre otros.

**Tabla N.º7. Validez y confiabilidad del Chat-GPT 4 para el diagnóstico de apendicitis aguda en tópico de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022.**

Precisión Dx	Indicadores	Valor
<b>Validez</b>	Sensibilidad	89,19 %
	Especificidad	97,60 %
	VPP	75 %
	VPN	99,11 %
	Likelihood ratio positivo	37,14
	Likelihood ratio negativo	0,11
	Índice de exactitud	96,97 %
	Score - F1	0,815
	Curva ROC	93,4 %
<b>Confiabilidad</b>	Coeficiente de correlación de Matthews	0.802

VPP: valor predictivo positivo, VPN: valor predictivo negativo

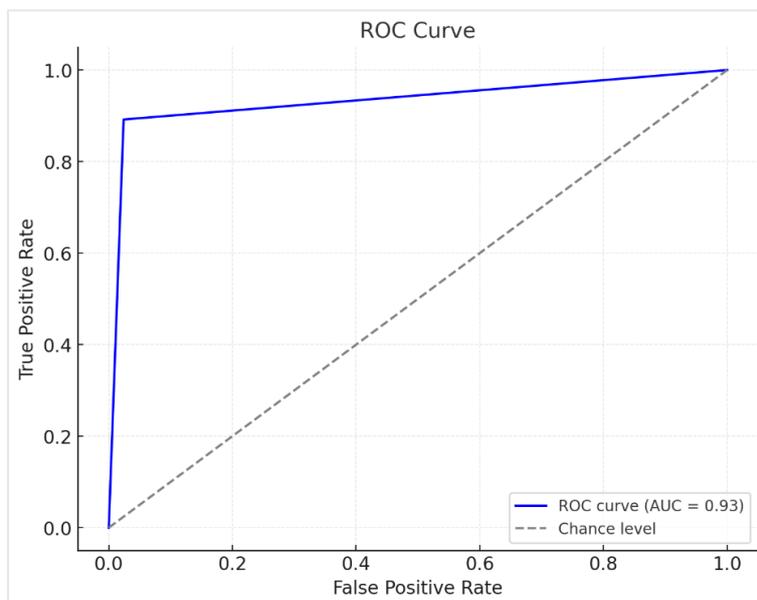
En la tabla N.º 7, los resultados revelaron que, en tanto a la validez, se observa una sensibilidad del 89,19 % lo que representa la capacidad de ChatGPT-4 para identificar correctamente los pacientes con apendicitis aguda, es decir los verdaderos positivos; dado el valor de esta métrica podemos decir que este modelo detecta un porcentaje significativo de los casos reales. En cuanto a la especificidad de ChatGPT-4 resultó ser un 97,60 %, lo cual nos ayuda a evaluar su capacidad de distinguir correctamente a los pacientes sin apendicitis, es decir, los verdaderos negativos. Este valor nos refleja que este modelo es altamente confiable al descartar apendicitis en pacientes que realmente no la tienen.

Además, el valor predictivo positivo (VPP) es del 75 %, lo que sugiere una probabilidad adecuada de que un paciente con un resultado positivo en la prueba diagnóstica con ChatGPT-4 presente apendicitis aguda. Respecto al Valor Predictivo Negativo (VPN), se obtuvo el valor de 99,11 %, lo que nos muestra que ChatGPT-4 tiene un excelente desempeño para descartar apendicitis aguda en pacientes a quienes este modelo resulte negativo para apendicitis aguda.

También, el índice de probabilidad positivo (Likelihood ratio positivo) resultó ser de 37.1, lo que respalda la utilidad de ChatGPT-4 para confirmar diagnósticos, mientras que el índice de probabilidad negativo (Likelihood ratio negativo) llegó a ser de 0.11 demostrando que un resultado negativo dado por ChatGPT-4 disminuye significativamente la probabilidad que el diagnóstico del paciente sea apendicitis aguda.

El índice de exactitud fue del 96,97 %, reflejando la proporción de diagnósticos correctos de apendicitis aguda dadas por ChatGPT-4 respecto al total de casos evaluados, Asimismo, el Score-F1, que integra sensibilidad y precisión, alcanzó un valor de 0.815, evidenciando un equilibrio adecuado entre la identificación de casos positivos de apendicitis aguda y la reducción de falsos positivos.

**Gráfico N.º 1. Curva ROC de ChatGPT-4 para el diagnóstico de apendicitis aguda en el tópico de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022.**



La curva Roc presentó un área del 93,4 %, como muestra el Gráfico N°1, lo que confirma la capacidad general de ChatGPT-4 para distinguir entre pacientes con y sin apendicitis aguda. Por último, el coeficiente de correlación de Matthews (MCC) alcanzó un valor de 0.802, lo que indica una alta consistencia entre las predicciones de ChatGPT-4 y los diagnósticos reales.

En conjunto, estos resultados demuestran que ChatGPT-4 tuvo una excelente validez y confiabilidad para el diagnóstico de apendicitis aguda en esta muestra. Su alta especificidad y VPN lo convierten en una herramienta útil para descartar casos, mientras que la sensibilidad y el Score-F1 respaldan su capacidad para identificar pacientes con apendicitis aguda.

## **5.2. Discusión**

El presente estudio tuvo como objetivo evaluar el uso de una herramienta actual y novedosa, ChatGPT-4, para el diagnóstico de apendicitis aguda en un entorno hospitalario. Para ello, se consideró a una muestra representativa de pacientes atendidos en el tópico de cirugía del Servicio de Emergencia. La muestra fue mayoritariamente de pacientes femeninas con un promedio de edad de 41 años que se atendieron por diagnósticos diversos, siendo los más frecuentes los relacionados con las patologías abdominales. La apendicitis aguda estuvo presente en el 7,49 % de los casos, porcentaje similar al estimado a nivel mundial para esta patología (9). De este grupo, tuvo ligera predominancia en el sexo

femenino y en los adultos de mediana edad, especialmente, los que se encontraron en el rango de 48 a 52 años, similar a la encontrada por Tejada en la región Junín. (16)

Para determinar la precisión en el diagnóstico de la Apendicitis aguda, se propuso el ChatGPT-4 como herramienta diagnóstica, para ello, se le proporcionó la información consignada en las historias clínicas. El diagnóstico de apendicitis aguda o la negación de esta fue contrastada con el resultado de las anatomía-patológica. Según esto, el ChatGPT-4 tuvo un alto porcentaje de aciertos, lo que nos indicaría que, para esta población, puede ser útil en determinar Apendicitis Aguda. Además, con los resultados obtenidos, se calculó los indicadores de validez y confiabilidad para establecer su idoneidad en el diagnóstico.

Los hallazgos de este estudio nos refieren una sensibilidad del 89,19 % lo cual supera a la reportada en un estudio hecho por Florence, donde se evaluó la sensibilidad de estudios de imagen en el diagnóstico de apendicitis aguda como son la tomografía computarizada con un resultado del 88 % y el ultrasonido con el 69,5 % (13). Además, resultó ser superior a la Escala de Alvarado en un estudio realizado en Cusco por Meléndez, en el que nos refiere que este presenta una sensibilidad del 78 %, todo esto siendo comparado con los hallazgos histopatológicos (45). Si bien otros sistemas diseñados especialmente para diagnóstico de apendicitis aguda como los modelos de la investigación de Shahmoradi (Multilayer Perceptron y Support Vector Machine) obtuvieron una sensibilidad del 91,7 % con Support Vector Machine (26), queda por resaltar que estos últimos a comparación de ChatGPT-4 son sistemas diseñados especialmente para este, mientras que ChatGPT-4 resulta ser un modelo multipropósito.

En este estudio, la especificidad obtenida fue del 97,60 %, superando significativamente lo reportado por Velásquez, quien evaluó la utilidad de la Escala de Alvarado en un hospital de segundo nivel, obteniendo una especificidad del 62 % (42). Asimismo, en un estudio realizado por Saverio, se reportó que la tomografía computarizada en casos de apendicitis aguda complicada presenta una especificidad del 93 % (55); estos hallazgos sugieren que ChatGPT-4 presenta una mayor especificidad en el diagnóstico de apendicitis aguda en comparación con la Escala de Alvarado y la tomografía computarizada. Además, a diferencia de la tomografía computarizada, ChatGPT-4 representa una alternativa de menor costo, lo que podría convertirlo en una herramienta diagnóstica accesible y de alta especificidad.

Por otra parte, el valor predictivo positivo (VPP) obtenido en esta investigación fue del 75 %, lo que indica que cuando ChatGPT-4 predice un caso positivo de apendicitis aguda, hay un 75 % de probabilidad de que el diagnóstico sea correcto. Si bien este porcentaje indica una capacidad de diagnóstico moderado, su desempeño es inferior en

comparación con otras herramientas y estudios previamente reportados. Un estudio realizado por Meléndez encontró un VPP del 89.8 % de la Escala de Alvarado (45), lo que indica un rendimiento intermedio; sin embargo, es superior a ChatGPT-4 evaluado en la presente investigación. Estas diferencias pueden atribuirse a factores metodológicos. En el estudio de Meléndez, la población analizada incluyó pacientes con edades entre 11 y 81 años, mientras que en nuestra investigación se limitó el rango de edad a 18-59 años; lo que generaría un sesgo al existir mayor incidencia en pacientes pediátricos. Además, el estudio de Meléndez se centró exclusivamente en pacientes con apendicitis, mientras que este estudio evaluó a todos los pacientes independientemente de su diagnóstico, lo que permitió reducir sesgos de repetición por parte de ChatGPT-4.

Asimismo, al comparar estos resultados con los métodos de imagen más utilizadas reportada por el estudio de Florence, se observa que la tomografía computarizada tiene un VPP del 97 % (13), lo que la posiciona como una herramienta muy confiable para confirmar el diagnóstico. Por otro lado, el ultrasonido presenta un VPP de 94 % (13), lo que también demuestra un alto potencial diagnóstico, aunque ligeramente menor que la tomografía. Ambos métodos superan el desempeño de ChatGPT-4 para la confirmación del diagnóstico de apendicitis aguda. No obstante, su aplicación sería relevante en contextos donde el acceso a métodos de diagnóstico avanzado, como son tomografía computarizada y el ultrasonido, tengan accesibilidad limitada, tampoco se descarta la mejora de este parámetro si se usa ChatGPT-4 en combinación de estos mismos métodos de imágenes o incluso la capacitación en *prompt engineering* del personal de la salud que utilice este modelo de inteligencia artificial.

De la misma manera, el valor predictivo negativo (VPN) de ChatGPT-4 en el diagnóstico de apendicitis aguda alcanzó un 99,11 %, lo que indica una capacidad muy elevada para descartar apendicitis aguda cuando el modelo de IA emite un resultado negativo. Este parámetro es de suma importancia en un contexto clínico, ya que un alto VPN reduce la probabilidad de falsos negativos, asegurando que los pacientes sin apendicitis aguda no sean erróneamente excluidos del diagnóstico y tratamiento oportuno. Al comparar estos resultados con investigaciones previas, se observa que la Escala de Alvarado presenta una variabilidad significativa en su VPN. En el estudio de Meléndez, realizado en Cusco, se reportó un VPN de solo el 24 % (45), lo que sugiere una elevada tasa de falsos negativos y una baja capacidad para descartar la enfermedad con confianza. En contraste, Velázquez obtuvo un VPN del 93 % (42) para la misma escala, mostrando una mejor precisión en la identificación de casos negativos, aunque aún por debajo del desempeño observado con ChatGPT-4. Estos resultados resaltan la ventaja de los

modelos de inteligencia artificial en el ámbito diagnóstico, ya que, a diferencia de la Escala de Alvarado, ChatGPT-4 no se limita a una serie de criterios clínicos predefinidos, sino que analiza múltiples variables mediante procesamiento de lenguaje natural. El alto valor del VPN en este estudio sugiere que ChatGPT-4 podría ser una herramienta de apoyo confiable en el servicio de emergencia, contribuyendo a reducir errores en la evaluación de pacientes con sospecha de apendicitis aguda.

Los valores que obtuvo el ChatGPT-4 para la likelihood ratio positivo (LR+) fue de 37.14 y negativa (LR-) fue 0.11, los cuales reflejan un desempeño diagnóstico sobresaliente, debido a que el elevado LR+ demuestra una alta capacidad para confirmar apendicitis aguda en pacientes con resultados positivos, mientras que el valor bajo LR- indica que la herramienta es confiable para descartar la enfermedad en casos negativos (59). En comparación con la investigación realizada por Velázquez en México, donde se evaluó la escala de Alvarado obteniendo un LR+ de 2.6 y de LR- de 0.02 (42). Esto indica que, aunque la Escala de Alvarado tiene un desempeño excepcional para descartar apendicitis aguda (LR- extremadamente bajo), su capacidad para confirmar la enfermedad es limitada (LR+ bajo). Por otro lado, el ChatGPT-4 no solo supera ampliamente a la Escala de Alvarado en su capacidad para confirmar apendicitis (LR+ > 10 se considera excelente), sino que también mantiene una sólida capacidad para descartar la enfermedad.

Un LR- cercano a 0.02, como el reportado por la Escala de Alvarado, sugiere que un resultado negativo reduce drásticamente la probabilidad de apendicitis, prácticamente descartándola. En cambio, un LR- de 0.11, aunque no tan bajo, sigue siendo clínicamente relevante, ya que reduce la probabilidad a niveles bajos. Este desempeño es particularmente útil en entornos con recursos limitados o donde se necesita tomar decisiones rápidas, ya que el ChatGPT-4 proporciona un balance adecuado entre sensibilidad y especificidad. Además, es importante señalar que, aunque la Escala de Alvarado tiene una ligera ventaja en términos de LR-, este valor por sí solo no es suficiente para garantizar un diagnóstico confiable en escenarios donde la confirmación también es crucial. El ChatGPT-4, con su combinación de un LR+ excelente (37.14) y un LR- clínicamente sólido (0.11), ofrece una solución más equilibrada para escenarios donde tanto la exclusión como la confirmación diagnóstica son igualmente importantes.

En esta investigación se obtuvo un índice de exactitud del 96,97 %, lo que demuestra que ChatGPT-4 presenta un alto grado de confiabilidad en la clasificación y predicción de los casos evaluados. Este resultado es comparable con los hallazgos de Park, quien analizó tres modelos distintos y obtuvo índices de exactitud del 99,80 %, 99,41 % y 97,84 % respectivamente (20). Sin embargo, ChatGPT-4 supera ampliamente la

precisión de la Escala de Alvarado, la cual, según el mismo estudio, presentó una exactitud de solo el 72,19 % (20). Esta diferencia resalta la superioridad de ChatGPT-4 sobre los métodos tradicionales, sugiriendo que su capacidad para integrar múltiples variables y patrones clínicos lo convierte en una herramienta con potencial de utilidad en el diagnóstico de apendicitis aguda.

El Score F1 obtenido por el ChatGPT-4 en esta investigación fue de 0.815, lo cual refleja un equilibrio adecuado entre precisión y sensibilidad, siendo esencial en escenarios donde el diagnóstico oportuno y preciso es crítico, como en el caso de la apendicitis aguda. Este resultado indica que la herramienta evaluada tiene un desempeño confiable para discriminar entre pacientes con y sin apendicitis aguda, siendo capaz de identificar un alto porcentaje de casos positivos mientras minimiza los errores de clasificación (29). Este valor respalda la viabilidad del ChatGPT-4 como un apoyo confiable en la toma de decisiones clínicas, mejorando tanto la eficiencia diagnóstica como los resultados para los pacientes.

El valor obtenido en el área bajo la curva ROC (AUC) fue de 0.93, el cual es considerado un valor excelente, ya que esto nos indica que tiene una alta capacidad para distinguir entre los casos positivos de los negativos (60), por lo tanto, el ChatGPT-4 tiene un 93 % de probabilidad de asignar una mayor puntuación a un caso positivo que a un caso negativo. Esto evidencia una elevada efectividad en la discriminación entre pacientes con y sin apendicitis aguda. En las investigaciones previas realizadas en nuestro país, donde se evaluaron la Escala de Alvarado, se obtuvieron valores de AUC de 0.61 y 0.68 (45,46), lo cual reflejan una capacidad de discriminación moderado, siendo valores con desempeño inferior en comparación a los alcanzados en esta investigación.

En la investigación que realizó Park en Corea, donde se construyó 3 tipos de redes neuronales, donde el AUC fue de 0.998 en el modelo que tuvo mayor valor (20), estos resultados difieren ligeramente al de nuestro estudio debido a que fueron modelos especialmente diseñados con el fin de diagnosticar exclusivamente apendicitis aguda, teniendo como datos de entrada los signos y síntomas específicos para apendicitis aguda tales como el dolor abdominal, migración del dolor, náuseas, recuentos de leucocitos, entre otros; solo incluyéndose en la investigación, los pacientes con dolor abdominal. A diferencia de esto, nuestra investigación, consideró toda la información consignada en la historia clínica, dándole mucha más información que procesar, teniendo que, el propio ChatGPT-4, discriminar los datos más importantes de otros y concluir con estos, la presencia de apendicitis aguda; además, incluimos historias clínicas de pacientes con otras patologías y no solamente las referidas a dolor abdominal, aumentando la complejidad para el procesamiento de los datos.

Respecto a la confiabilidad del ChatGPT-4, se halló con el índice de correlación de Matthews, obteniéndose un valor de 0.8, considerándose un valor bastante alto, lo que sugiere que el ChatGPT-4 como herramienta diagnóstica tuvo un buen desempeño en esta población. Según Hicks, el coeficiente de correlación de Matthews es una herramienta útil para evaluar modelos de inteligencia artificial (machine learning), donde el valor de 1 nos indicaría una predicción perfecta, el 0 una predicción aleatoria y el -1 una correlación inversa (61). Por ello, podemos considerar que existe una fuerte correlación entre las predicciones realizadas por el ChatGPT-4 y los resultados reales hallados con la anatomía patológica.

Por lo que consideramos que ChatGPT-4 tiene un alto desempeño en el diagnóstico de apendicitis aguda, con una sensibilidad del 89,19 % y especificidad del 97,60 %, superando herramientas convencionales como la Escala de Alvarado, la ecografía y la tomografía computarizada en algunos parámetros. Además, su elevado VPN (99,11 %) lo posiciona como una alternativa útil para descartar casos negativos. Aunque su VPP (75 %) es inferior a otras técnicas, su accesibilidad y menor costo lo convierten en una herramienta complementaria viable en entornos con recursos limitados. Su índice de exactitud del 96,97 % y su área bajo la curva ROC (AUC) de 0.93 evidencian una elevada capacidad de discriminación diagnóstica. Además, el Score F1 obtenido (0.815) refleja un equilibrio adecuado entre precisión y sensibilidad, lo que refuerza su utilidad en el diagnóstico de apendicitis aguda. Finalmente, el coeficiente de correlación de Matthews (0.8) sugiere una fuerte correlación entre las predicciones realizadas por ChatGPT-4 y los hallazgos histopatológicos, lo que respalda su confiabilidad como herramienta diagnóstica. Su accesibilidad y menor costo podrían convertirlo en un complemento útil en entornos con recursos limitados, especialmente si se optimiza su uso mediante estrategia como el *prompt engineering*.

La relevancia clínica de esta investigación radica en su aplicación en el diagnóstico de apendicitis aguda, una de las principales causas de cirugía de emergencia en el mundo. La detección oportuna de esta patología es crucial para reducir complicaciones como la perforación y peritonitis, pero también para evitar apendicectomías innecesarias (18). La incorporación de inteligencia artificial en este proceso ofrece una alternativa prometedora para mejorar la precisión diagnóstica, especialmente en contextos donde los recursos diagnósticos avanzados, como la tomografía computarizada, no están fácilmente disponibles. La alta sensibilidad y especificidad obtenida por ChatGPT-4 en este estudio sugieren que su uso podría ser parte de la evaluación clínica de los médicos, optimizando la toma de decisiones y mejorando los desenlaces para los pacientes.

Además de su impacto inmediato en la práctica clínica, esta investigación tiene una gran proyección en el ámbito de la digitalización en salud. La accesibilidad de herramientas basadas en inteligencia artificial permite su implementación en centros de salud con recursos limitados, donde los especialistas pueden no estar siempre disponibles, como en hospitales periféricos o rurales (44). En estos centros, la falta de acceso a tecnología avanzada y personal especializado puede generar retrasos en el diagnóstico y tratamiento de enfermedades críticas, lo que impacta directamente en la calidad de la atención. La implementación de la inteligencia artificial en estos entornos permite que el personal médico, incluso con menos experiencia o recursos, pueda obtener diagnósticos más certeros y rápidos, basados en grandes volúmenes de datos y algoritmos avanzados. Esto no solo mejora la eficiencia del sistema de salud, sino que también ayuda a reducir las desigualdades en la atención médica entre hospitales de alta complejidad y centros de salud periféricos, contribuyendo a un sistema de salud más equitativo y accesible.

También, la capacidad de ChatGPT-4 para minimizar la tasa de apendicectomías negativas representa una ventaja significativa, tanto para los pacientes como para los sistemas de salud. Las apendicectomías innecesarias no solo exponen a los pacientes a riesgos quirúrgicos y complicaciones posoperatorias, sino que también generan una carga económica considerable para los hospitales. Al utilizar la inteligencia artificial para identificar de manera más precisa los casos de apendicitis aguda y evitar intervenciones quirúrgicas innecesarias, se reduce el riesgo de complicaciones como infecciones, hemorragias y problemas derivados de la anestesia, lo que contribuye a una recuperación más rápida y menos costosa para los pacientes. Desde un punto de vista económico, esta reducción en las apendicectomías negativas también disminuye los costos hospitalarios y los tiempos de hospitalización, lo que libera recursos que pueden ser utilizados para tratar a otros pacientes o mejorar la infraestructura sanitaria. A largo plazo, esto podría generar un ahorro significativo para los sistemas de salud, mejorando la eficiencia operativa y reduciendo la sobrecarga que enfrentan muchos hospitales, particularmente aquellos con recursos limitados.

En tanto al uso de ChatGPT-4 en la toma de decisiones médicas, especialmente en contextos críticos como la emergencia médica, plantea una serie de implicaciones éticas y sociales que deben ser cuidadosamente consideradas. Una de las preocupaciones más importantes es la dependencia del juicio clínico humano en situaciones de alta presión, donde la rapidez y precisión en las decisiones pueden salvar vida. Aunque herramientas como Chat-GPT-4 pueden proporcionar diagnósticos rápidos y sugerencias basadas en grandes volúmenes de datos, existe el riesgo de que los profesionales de salud puedan,

en ocasiones, confiar excesivamente en la IA y sustituir su juicio experto por la información generada por el modelo, lo que podría llevar a errores si no se combina adecuadamente con la evaluación clínica directa. Además, el uso de IA en la emergencia médica implica que las decisiones que afectan directamente a la salud de los pacientes sean tomadas parcialmente por algoritmos, lo que plantea la cuestión de la responsabilidad en caso de errores médicos.

Finalmente, este estudio sienta las bases para futuras investigaciones que amplíen el alcance de la inteligencia artificial en el diagnóstico médico. La posibilidad de aplicar modelos similares en otras patologías comunes, como colecistitis aguda o diverticulitis, podría revolucionar el enfoque diagnóstico en distintas áreas de la medicina. Además, investigaciones adicionales que incluyan una mayor diversidad de poblaciones y escenarios clínicos permitirán mejorar la generalización de estos hallazgos, consolidando el papel de la inteligencia artificial como un apoyo confiable en la toma de decisiones médicas.

Si bien la presente investigación ha demostrado que ChatGPT-4 posee una alta precisión diagnóstica en la detección de apendicitis aguda, es fundamental reconocer las limitaciones metodológicas y operativas que podrían influir en su aplicabilidad en la práctica clínica. En primer lugar, el diseño retrospectivo del estudio introduce restricciones inherentes a este tipo de investigación, ya que se basa en la evaluación de historias clínicas previamente registradas. Esto limita la capacidad de ChatGPT-4 para interactuar con escenarios clínicos dinámicos, en los que las decisiones médicas deben ser tomadas en tiempo real y en función de una evaluación clínica continua y contextualizada. La naturaleza retrospectiva impide también controlar factores que pueden haber influido en el diagnóstico, como cambios en la práctica médica a lo largo del tiempo o la variabilidad en la documentación médica de los casos. Además, dado que se trata de un estudio retrospectivo, la información disponible es estática y no refleja la evolución en tiempo real de la apendicitis aguda ni la intervención médica a lo largo del proceso de atención al paciente.

Otro aspecto relevante es que la precisión diagnóstica de ChatGPT-4 depende en gran medida de la calidad de los datos ingresados en las historias clínicas. Cualquier error o inconsistencia en la documentación médica, o la falta de estandarización en la forma en que se registran los datos, puede influir negativamente en los resultados generados por la inteligencia artificial. Aunque el modelo puede identificar patrones y relaciones complejas entre los datos, su efectividad está directamente vinculada a la calidad de la información disponible para su análisis. En este sentido, el uso de ChatGPT-4 como herramienta

médica de apoyo, y no como sustituto del juicio clínico, debe ser considerado un principio fundamental. La inteligencia artificial puede ser una poderosa aliada en el diagnóstico y la toma de decisiones, pero no puede reemplazar la evaluación clínica integral que realiza un profesional de salud experimentado. La decisión final sobre el diagnóstico y el tratamiento debe seguir siendo responsabilidad del personal médico, quien debe integrar los resultados proporcionados por el modelo con su propio juicio clínico y las condiciones específicas de cada paciente.

Aunado a esto, otra limitación crítica de este estudio es la ausencia de investigaciones prospectivas y multicéntricas que validen el rendimiento de ChatGPT-4 en entornos hospitalarios y poblaciones diversas. La realización de estudios prospectivos permitiría evaluar cómo se comporta el modelo en escenarios clínicos en tiempo real, lo que ofrecería información más precisa sobre su capacidad para manejar situaciones dinámicas en el día a día de los hospitales. A través de un estudio prospectivo, se podría analizar el desempeño de ChatGPT-4 a lo largo del curso de la enfermedad en pacientes con apendicitis aguda, permitiendo observar cómo la inteligencia artificial maneja variables cambiantes y cómo contribuye a la toma de decisiones médicas en tiempo real. Esta información sería esencial para evaluar la efectividad del modelo en un entorno clínico real, donde la situación del paciente puede evolucionar rápidamente.

Además, los estudios multicéntricos proporcionarán una visión más amplia sobre la reproducibilidad de los resultados en diferentes hospitales y regiones geográficas, cada uno con características clínicas, recursos y realidades epidemiológicas diversas. La variabilidad en la infraestructura, el personal médico, los protocolos y los recursos disponibles entre centros de salud puede afectar la capacidad del modelo para ser implementado de manera consistente. Así, un estudio multicéntrico contribuiría a evaluar cómo el rendimiento de ChatGPT-4 varía en función del contexto local y, de ser necesario, a realizar ajustes en el modelo para garantizar su eficacia en una gama más amplia de entornos. Asimismo, la inclusión de hospitales con diferentes niveles de atención permitiría evaluar si el modelo es igualmente efectivo en centros de alta complejidad en comparación con hospitales de menor capacidad.

La generalización de los resultados a otros entornos hospitalarios o geográficos podría verse influenciada por factores tales como la variabilidad en los métodos diagnósticos, las características demográficas de los pacientes y la infraestructura hospitalaria. Por ejemplo, en áreas rurales o en establecimiento de salud de primer nivel, el acceso a herramientas de diagnóstico avanzadas, como la tomografía computarizada o la ecografía, puede ser limitado, lo que afectaría la calidad y disponibilidad de la

información para la inteligencia artificial. Además, las características epidemiológicas de las poblaciones atendidas en diferentes regiones también pueden influir en la precisión diagnóstica del modelo, ya que algunos grupos pueden presentar características clínicas atípicas que dificulten su diagnóstico.

Es decir, las limitaciones de este estudio, tales como su diseño retrospectivo y la falta de validación prospectiva y multicéntrica, podrían restringir la generalización de los resultados a otros entornos hospitalarios o geográficos. La realización de investigaciones prospectivas y multicéntricas sería crucial para validar el rendimiento de ChatGPT-4 en diversos contextos clínicos, asegurar su aplicabilidad en la práctica médica diaria y explorar su capacidad para mejorar la precisión diagnóstica en una amplia variedad de situaciones. Estas futuras investigaciones permitirían optimizar el uso de la inteligencia artificial en la atención médica, mejorando la calidad del diagnóstico y el tratamiento de enfermedades como la apendicitis aguda.

Los resultados obtenidos en esta investigación sugieren que ChatGPT-4 posee una alta precisión diagnóstica para la apendicitis aguda, como una sensibilidad del 89 % y una especificidad del 98 %. Estas cifras lo posicionan como una herramienta prometedora en el apoyo diagnóstico, con potencial para reducir la cantidad de apendicectomías innecesarias gracias a su alto valor predictivo negativo (99 %).

Además, su rapidez en la evaluación de casos clínicos podría optimizar la toma de decisiones en emergencias, facilitando el manejo eficiente de pacientes y reduciendo tiempos de espera. En combinación con herramientas diagnósticas tradicionales como la escala de Alvarado, ecografía y la tomografía computarizada, podría mejorar la precisión diagnóstica, funcionando como un complemento para el médico tratante.

Su aplicación en hospitales con recursos limitados es otra ventaja relevante, ya que permitiría democratizar el acceso a diagnósticos oportunos en zonas con escasez de especialistas. Sin embargo, existen limitaciones, como la menor precisión en pacientes con síntomas atípicos o en mujeres con patologías ginecológicas similares a la apendicitis, lo que indica la necesidad de validación en diversos grupos poblacionales.

También resalta su potencial en la educación médica, donde podría ser utilizado para entrenar a estudiantes e internos en el razonamiento clínico. No obstante, su implementación en la práctica clínica requiere marcos regulatorios adecuados y supervisión médica, garantizando su seguridad y precisión.

## **Conclusiones**

Los hallazgos de esta investigación permiten determinar que ChatGPT-4 posee una alta precisión diagnóstica para la apendicitis aguda en el servicio de emergencia del Hospital Ramiro Prialé Prialé en 2022. Los resultados muestran una capacidad significativa para diferenciar entre pacientes con y sin esta patología, evidenciando su utilidad como herramienta de apoyo en la toma de decisiones médicas.

En cuanto a su validez, se observa que el modelo de inteligencia artificial presenta una alta sensibilidad, especificidad, VPP, índice de exactitud, LR+, score F1 y área bajo la curva ROC; lo que respalda su potencial en la evaluación diagnóstica de la apendicitis aguda. Además, su desempeño en la clasificación de casos fue adecuado en relación con el estándar de referencia, reflejando una concordancia importante con los diagnósticos confirmados.

Respecto a la confiabilidad, los análisis realizados indican que ChatGPT-4 mantiene un desempeño estable en la evaluación de casos clínicos, con una correlación significativa con los resultados histopatológicos. Su capacidad para generar diagnósticos consistentes sugiere que puede ser una herramienta útil en entornos hospitalarios, complementando los métodos diagnósticos convencionales.

## Recomendaciones

Se recomienda al personal del Hospital Ramiro Prialé Prialé mejorar el uso y la accesibilidad del sistema de gestión de la documentación del Servicio de Anatomía Patológica, con el fin de optimizar su acceso y legibilidad para los investigadores. Aunque el hospital cuenta con un sistema para digitalizar y gestionar estas anatomopatologías, el problema radica en que no se utiliza adecuadamente, lo que dificulta el acceso a los datos y afecta la eficiencia en el desarrollo de futuras investigaciones. Además, se aconseja realizar auditorías integrales de las historias clínicas, especialmente las del Servicio de Emergencia, con énfasis en el registro completo y preciso de la información esencial, como funciones vitales, exámenes físicos, anamnesis, entre otros apartados, ya que la falta de datos adecuados puede comprometer tanto la calidad de la atención médica como la validez de los estudios diagnósticos. Asimismo, es fundamental que el personal de salud reciba formación especializada en el uso adecuado de herramientas basadas en inteligencia artificial, enfocándose en la correcta interpretación de los resultados generados por ChatGPT-4 y en la aplicación de nuevas metodologías como el Prompt Engineering. Esta capacitación permitirá una integración más eficiente de estas tecnologías en el proceso de toma de decisiones, identificando sus limitaciones y maximizando sus beneficios dentro del entorno hospitalario.

Además, se sugiere que la Universidad Continental incorpore dentro de sus líneas de investigación en las carreras de ciencias de la salud, temas vinculados a nuevas tecnologías, especialmente en el ámbito de la inteligencia artificial aplicada a la salud, ya que esta área emergente tiene un potencial de revolucionar tanto el diagnóstico como el tratamiento médico. La integración de la inteligencia artificial puede mejorar la precisión y rapidez en la toma de decisiones clínicas, facilitar el análisis de grandes volúmenes de datos y optimizar la eficiencia en la gestión hospitalaria, lo que se traduce en una atención al paciente más oportuna y personalizada. Fomentar la investigación en este campo posicionará a la universidad como pionera en la innovación tecnológica dentro del sector salud en el Perú.

También, se recomienda la realización de estudios prospectivos, ya que estos permiten evaluar de forma continua y en tiempo real la aplicabilidad de ChatGPT-4 en el diagnóstico de apendicitis aguda, superando las limitaciones inherentes a los estudios retrospectivos en cuanto a control de variables y actualización de datos. Para ello, se sugiere llevar a cabo investigaciones en hospitales de distintos niveles de atención y en diversas regiones geográficas, abarcando además una amplia gama de poblaciones que incluyan pacientes pediátricos, obstétricos y geriátricos. Asimismo, se recomienda realizar

estudios comparativos entre ChatGPT-4 y métodos diagnósticos tradicionales, como la Escala de Alvarado y técnicas de imagen (ecografía y tomografía computarizada), con el fin de determinar si la inteligencia artificial puede complementar o incluso mejorar la precisión diagnóstica actual. Adicionalmente, es fundamental evaluar el impacto del uso de esta herramienta en la reducción de apendicectomías innecesarias, los tiempos de hospitalización y los costos asociados a diagnósticos erróneos, además de analizar posibles sesgos relacionados con características clínicas de los pacientes, para garantizar una aplicación equitativa, eficaz en la práctica médica.

## Referencias bibliográficas

1. Hernández Sampieri R, Fernández Collado C, Baptista Lucio P. Metodología de la investigación. 6° ed. Metodología de la investigación. España: McGraw Hill España; 2014. 91 p.
2. IBM. ¿Qué es la Inteligencia Artificial en Medicina? [Internet]. [citado el 16 de enero de 2025]. Disponible en: <https://www.ibm.com/mx-es/topics/artificial-intelligence-medicine>
3. Rozillio Mercado E, López Anza D, Ortega Ortuño GL, Lee Lee SHJ, Minian Okon J, Gutiérrez Gurza RA, et al. Inteligencia Artificial en Medicina, usos Actuales y Futuras Perspectivas. Ciencia Latina: Revista Multidisciplinar, ISSN-e 2707-2215, ISSN 2707-2207, Vol 7, No 6, 2023 [Internet]. 2023 [citado el 16 de enero de 2025];7(6):14. Disponible en: [https://doi.org/10.37811/cl\\_rcm.v7i6.9167](https://doi.org/10.37811/cl_rcm.v7i6.9167)
4. Lanzagorta-Ortega D, Carrillo-Pérez DL, Carrillo-Esper R. Inteligencia artificial en medicina: presente y futuro. Gac Med Mex [Internet]. 2022 [citado el 16 de enero de 2025];158(Supl.1):55–9. Disponible en: <https://doi.org/10.24875/gmm.m22000688>
5. Universidad Europea. IA en medicina: revolución en la salud digital [Internet]. [citado el 16 de enero de 2025]. Disponible en: <https://universidadeuropea.com/blog/inteligencia-artificial-medicina/>
6. Romero-Pérez S, Romero-Pérez E. Inteligencia artificial y medicina. Rev Cir Gal [Internet]. 2022 [citado el 25 de septiembre de 2023];6(7):1–2. Disponible en: <https://www.revistas.usb.edu.co/index.php/IngUSBmed/article/view/5098/4134>
7. Murad Gutiérrez V. Inteligencia artificial a solo un clic: ¿cuál es su implicación en la escritura médica? Rev Colom Radiol [Internet]. 2022 [citado el 25 de septiembre de 2023];33(2). Disponible en: <https://rcr.acronline.org/index.php/rcr/article/download/235/295>
8. Arias V, Salazar J, Garicano C, Contreras J, Chacón G, Chacín-González M, et al. Una introducción a las aplicaciones de la inteligencia artificial en Medicina: Aspectos históricos. Revista Latinoamericana de Hipertensión

- [Internet]. 2019 [citado el 25 de septiembre de 2023];14(5):590–600. Disponible en: <https://www.redalyc.org/articulo.oa?id=170262877013>
9. Rodríguez González HM, Portillo Yáñez IE, Soto Fajard RC, Martínez Hernández JE, Morales Chávez NA. Prevalencia de apendicitis aguda en un centro de segundo nivel de atención. *Cir Gen [Internet]*. 2014 [citado el 28 de enero de 2025];3(2):87–90. Disponible en: <https://www.elsevier.es/es-revista-cirujano-general-218-pdf-X1405009914551822>
  10. Brunicardi C, Andersen D, Billiar T, Dunn D, Hunter J, Matthews J, et al. Schwartz: Principios de Cirugía. 9° ed. Mexico: McGraw-Hill Interamericana Editores; 2012. 1863 p.
  11. Archundia Garcia A. Cirugía 2 . 1° ed. Mexico DF: McGraw-Hill Interamericana Editores; 2014. 331 p.
  12. Fallas González J. Revision bibliografica: Apendicitis Aguda. *Medicina Legal de Costa Rica [Internet]*. 2012 [citado el 21 de septiembre de 2023];29(1):83–90. Disponible en: <https://www.scielo.sa.cr/pdf/mlcr/v29n1/art10.pdf>
  13. Florence M, Flum DR, Jurkovich GJ, Lin P, Steele SR, Symons RG, et al. Negative appendectomy and imaging accuracy in the washington state surgical care and outcomes assessment program. *Ann Surg [Internet]*. 2008 [citado el 21 de septiembre de 2023];248(4):557–62. Disponible en: [10.1097/SLA.0b013e318187aeca](https://doi.org/10.1097/SLA.0b013e318187aeca)
  14. Seetahal SA, Bolorunduro OB, Sookdeo TC, Oyetunji TA, Greene WR, Frederick W, et al. Negative appendectomy: A 10-year review of a nationally representative sample. *Am J Surg [Internet]*. 2011 [citado el 20 de septiembre de 2023];201(4):433–7. Disponible en: [10.1016/j.amjsurg.2010.10.009](https://doi.org/10.1016/j.amjsurg.2010.10.009)
  15. Ávila MJ, García-Acero M. Apendicitis aguda: revisión de la presentación histopatológica en Boyacá, Colombia. *Revista Colombiana de Cirugía [Internet]*. 2015 [citado el 23 de septiembre de 2023];30(2):125–30. Disponible en: <https://www.redalyc.org/articulo.oa?id=355540058005>
  16. Tejada-Llacsá PJ, Melgarejo-García GC. Incidencia de apendicitis aguda y su relación con factores ambientales, Perú, 2013. *Anales de la Facultad de Medicina [Internet]*. 2015 [citado el 23 de septiembre de 2023];76(3):253–6. Disponible en: [dx.doi.org/10.15381/anales.v76i3.11234](https://doi.org/10.15381/anales.v76i3.11234)

17. Mansilla Sandoval MG, Perez Palma Diaz CR. Tiempo de espera quirúrgico en apendicitis aguda y complicaciones postoperatorias en un hospital nivel 3 de Lima-Perú, entre los años 2019- 2021 [Internet]. [Tesis de grado]: Lima: Universidad Peruana de Ciencias Aplicadas; 2022 [citado el 23 de septiembre de 2023]. Disponible en: [https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/663425/Mansilla\\_SM%20.pdf?sequence=11&isAllowed=y](https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/663425/Mansilla_SM%20.pdf?sequence=11&isAllowed=y)
18. Ravelo Bendezu LD. Apendicectomías negativas en pacientes de un hospital regional de Huancayo en el periodo 2015 - 2016 [Internet]. [Tesis de grado]: Huancayo: Universidad Peruana Los Andes; 2017 [citado el 23 de septiembre de 2023]. Disponible en: [https://repositorio.upla.edu.pe/bitstream/handle/20.500.12848/189/Luz\\_Ravelo\\_Tesis\\_Titulo\\_2017.pdf?sequence=1&isAllowed=y](https://repositorio.upla.edu.pe/bitstream/handle/20.500.12848/189/Luz_Ravelo_Tesis_Titulo_2017.pdf?sequence=1&isAllowed=y)
19. Roshanaei G, Salimi R, Mahjub H, Faradmal J, Yamini A, Tarokhian A. Accurate diagnosis of acute appendicitis in the emergency department: an artificial intelligence-based approach. *Intern Emerg Med* [Internet]. 2024 [citado el 7 de enero de 2025];19(8):2347–57. Disponible en: <https://link.springer.com/article/10.1007/s11739-024-03738-w>
20. Park SY, Kim SM. Acute appendicitis diagnosis using artificial neural networks. En: *Technology and Health Care* [Internet]. IOS Press BV; 2015 [citado el 7 de enero de 2025]. p. S559–65. Disponible en: 10.3233/THC-150994
21. Avila-Tomás JF, Mayer-Pujadas MA, Quesada-Varela VJ. La inteligencia artificial y sus aplicaciones en medicina I: introducción antecedentes a la IA y robótica. *Aten Primaria* [Internet]. 2020 [citado el 23 de septiembre de 2023];52(10):778–84. Disponible en: <https://doi.org/10.1016/j.aprim.2020.04.013>
22. Lugo-Reyes SO, Maldonado-Colín G, Murata C. Inteligencia artificial para ayudar al diagnóstico clínico en medicina. *Rev Alerg Mex* [Internet]. 2014 [citado el 11 de enero de 2025];61(2):110–20. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/24915623/>
23. Busnatu S, Niculescu AG, Bolocan A, Petrescu GED, Paduraru DN, Nastasa I, et al. Clinical Applications of Artificial Intelligence: An Updated Overview. *J*

- Clin Med [Internet]. 2022 [citado el 11 de enero de 2025];11(8):2265. Disponible en: <https://doi.org/10.3390/jcm11082265>
24. Basáez E, Mora J. Salud e inteligencia artificial: ¿cómo hemos evolucionado? Revista Médica Clínica Las Condes [Internet]. el 1 de noviembre de 2022 [citado el 16 de enero de 2025];33(6):556–61. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0716864022001262>
  25. Ruiz RB, Velásquez JD. Inteligencia artificial al servicio de la salud del futuro. Revista Médica Clínica Las Condes [Internet]. 2023 [citado el 16 de enero de 2025];34(1):84–91. Disponible en: <https://www.elsevier.es/es-revista-revista-medica-clinica-las-condes-202-articulo-inteligencia-artificial-al-servicio-salud-S0716864023000032>
  26. Shahmoradi L, Safdari R, Mirhosseini MM, Rezayi S, Javaherzadeh M. Development and evaluation of a clinical decision support system for early diagnosis of acute appendicitis. Sci Rep [Internet]. 2023 [citado el 11 de enero de 2025];13:19703. Disponible en: <https://doi.org/10.1038/s41598-023-46721-9>
  27. El impacto crítico de la escasez de especialistas médicos en Perú - Infobae [Internet]. [citado el 11 de enero de 2025]. Disponible en: <https://www.infobae.com/peru/2024/12/08/el-impacto-critico-de-la-escasez-de-especialistas-medicos-en-peru/>
  28. Hernández-Borroto CE, Medrano-Plana Y. La integración de la inteligencia artificial en la educación médica y su impacto en la práctica clínica. FEM: Revista de la Fundación Educación Médica [Internet]. 2024 [citado el 11 de enero de 2025];27(2):59–61. Disponible en: <https://scielo.isciii.es/pdf/fem/v27n2/2014-9832-fem-27-2-59.pdf>
  29. Molina Arias M. Un intruso de otro mundo: F1-score. Revista electrónica de AnestesiaR [Internet]. 2024 [citado el 7 de enero de 2025];16(4):6. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=9532114>
  30. Dirección General de Seguimiento y Evaluación. Reporte Regional de Indicadores Sociales de Junín [Internet]. Lima; 2024 [citado el 11 de enero de 2025]. Disponible en: <https://sdv.midis.gob.pe/redinforma/Upload/Mancomunidad/Junin.pdf>

31. Aguirre Martens G. Salud en el Perú: De la cobertura en papel a la cobertura real [Internet]. 2023 [citado el 11 de enero de 2025]. Disponible en: <https://blogs.worldbank.org/es/latinamerica/cobertura-salud-peru>
32. Chávarry F. SuSalud: acceso a servicios y falta de información son los principales problemas que pacientes reportan -RPP Noticias [Internet]. 2024 [citado el 11 de enero de 2025]. Disponible en: <https://rpp.pe/peru/actualidad/susalud-acceso-a-servicios-y-falta-de-informacion-son-los-principales-problemas-que-pacientes-reportan-en-hospitales-noticia-1581968>
33. Diario Correo. Mujer con peritonitis muere por falta de atención [Internet]. 2013 [citado el 11 de enero de 2025]. Disponible en: <https://diariocorreo.pe/peru/mujer-con-peritonitis-muere-por-falta-de-ate-73909/>
34. Reis F, Lenz C. Performance of Artificial Intelligence (AI)-Powered Chatbots in the Assessment of Medical Case Reports: Qualitative Insights From Simulated Scenarios. *Cureus* [Internet]. 2024 [citado el 7 de enero de 2025];16(2):e53899. Disponible en: [10.7759/cureus.53899](https://doi.org/10.7759/cureus.53899)
35. Ghanem YK, Rouhi AD, Al-Houssan A, Saleh Z, Moccia MC, Joshi H, et al. Dr. Google to Dr. ChatGPT: assessing the content and quality of artificial intelligence-generated medical information on appendicitis. *Surg Endosc* [Internet]. 2024 [citado el 7 de enero de 2025];38(5):2887–93. Disponible en: <https://doi.org/10.1007/s00464-024-10739-5>
36. Akbulut S, Yagin FH, Cicek IB, Koc C, Colak C, Yilmaz S. Prediction of Perforated and Nonperforated Acute Appendicitis Using Machine Learning-Based Explainable Artificial Intelligence. *Diagnostics* [Internet]. 2023 [citado el 7 de enero de 2025];13(6):1173. Disponible en: <https://doi.org/10.3390/diagnostics13061173>
37. Xv Y, Peng C, Wei Z, Liao F, Xiao M. Can Chat-GPT a substitute for urological resident physician in diagnosing diseases?: a preliminary conclusion from an exploratory investigation. *World J Urol* [Internet]. 2023 [citado el 24 de septiembre de 2023];41(9):2569–71. Disponible en: <https://doi.org/10.1007/s00345-023-04539-0>

38. Rao A, Kim J, Kamineni M, Pang M, Lie W, Dreyer KJ, et al. Evaluating GPT as an Adjunct for Radiologic Decision Making: GPT-4 Versus GPT-3.5 in a Breast Imaging Pilot. *Journal of the American College of Radiology* [Internet]. 2023 [citado el 22 de septiembre de 2023];20(10):990–7. Disponible en: [10.1016/j.jacr.2023.05.003](https://doi.org/10.1016/j.jacr.2023.05.003)
39. Ghareeb WM, Emile SH, Elshobaky A. Artificial Intelligence Compared to Alvarado Scoring System Alone or Combined with Ultrasound Criteria in the Diagnosis of Acute Appendicitis. *Journal of Gastrointestinal Surgery* [Internet]. 2022 [citado el 7 de enero de 2025];26(3):655–8. Disponible en: <https://doi.org/10.1007/s11605-021-05147-2>
40. Carrasco JP, García E, Sánchez DA, Porter E, De la Puente L, Navarro J, et al. ¿Es capaz “ChatGPT” de aprobar el examen MIR de 2022? Implicaciones de la inteligencia artificial en la educación médica en España. *Revista Española de Educación Médica* [Internet]. 2023 [citado el 24 de septiembre de 2023];4(1):55–69. Disponible en: <https://doi.org/10.6018/edumed.556511>
41. Moreno Barriga E, Pueyo Ferrer I, Sánchez Sánchez M, Martín Baranera M, Masip Utset J. Experiencia de Mediktor®: un nuevo evaluador de síntomas basado en inteligencia artificial para pacientes atendidos en el servicio de urgencias. *Emergencias* [Internet]. 2017 [citado el 26 de septiembre de 2023];29:391–6. Disponible en: [https://revistaemergencias.org/wp-content/uploads/2023/08/Emergencias-2017\\_29\\_6\\_391-396.pdf](https://revistaemergencias.org/wp-content/uploads/2023/08/Emergencias-2017_29_6_391-396.pdf)
42. Velázquez Mendoza D, Godínez Rodríguez C, Vázquez Guerrero MA. Evaluación prospectiva de la Escala de Alvarado en el diagnóstico de apendicitis aguda. *Cirujano General* [Internet]. 2010 [citado el 26 de septiembre de 2023];32(1):17–23. Disponible en: <https://www.medigraphic.com/pdfs/cirgen/cg-2010/cg101c.pdf>
43. Prabhudesai SG, Gould S, Rekhraj S, Tekkis PP, Glazer G, Ziprin P. Artificial neural networks: Useful aid in diagnosing acute appendicitis. *World J Surg* [Internet]. 2008 [citado el 7 de enero de 2025];32(2):305–9. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/18043966/>
44. Quispe-Juli CU, Kirschbaum JP, Caceres-Alban J, Ugas-Charcape CF. Explorando la viabilidad de la integración de la inteligencia artificial en la

- atención médica pediátrica: Un estudio preliminar con ChatGPT. Investigación e Innovación Clínica y Quirúrgica Pediátrica [Internet]. 2023 [citado el 29 de septiembre de 2023];1(1):14–23. Disponible en: <https://investigacionpediatria.insnsb.gob.pe/index.php/iicqp/article/view/5>
45. Meléndez Flores JE, Cosio Dueñas H, Sarmiento Herrera WS. Sensibilidad y especificidad de la Escala de Alvarado en el diagnóstico de pacientes atendidos por apendicitis aguda en Hospitales del Cusco. Horizonte Médico (Lima) [Internet]. 2019 [citado el 29 de septiembre de 2023];19(1):13–8. Disponible en: <https://doi.org/10.24265/horizmed.2019.v19n1.03>
  46. Alarcón Thompson N. Asociación entre Escala de Alvarado y diagnóstico de apendicitis aguda complicada y no complicada según anatomía patológica en el Centro Médico Naval. Revista Horizonte Médico [Internet]. 2012 [citado el 7 de enero de 2025];12(2):14–20. Disponible en: <https://doi.org/10.24265/horizmed.2012.v12n2.02>
  47. Caiafa CF, Lew SE. ¿Qué es la Inteligencia Artificial? Boletín Radio@stronómico [Internet]. 2020 [citado el 16 de enero de 2025];69(1):1–7. Disponible en: <https://ri.conicet.gov.ar/handle/11336/110093>
  48. Escalante González M. Aplicación de la inteligencia artificial para la detección del cáncer de mama. Revista Medica Sinergia [Internet]. 2023 [citado el 6 de febrero de 2025];8(12):e1113. Disponible en: <https://doi.org/10.31434/rms.v8i12.1113>
  49. Bermejo-Peláez D, San José Estépar R, Fernández-Velilla M, Palacios Miras C, Gallardo Madueño G, Benegas M, et al. Deep learning-based lesion subtyping and prediction of clinical outcomes in COVID-19 pneumonia using chest CT. Sci Rep [Internet]. 2022 [citado el 6 de febrero de 2025];12(9387):1–11. Disponible en: <https://doi.org/10.1038/s41598-022-13298-8>
  50. Knudsen JE, Ghaffar U, Ma R, Hung AJ. Clinical applications of artificial intelligence in robotic surgery. J Robot Surg [Internet]. 2024 [citado el 6 de febrero de 2025];18(1):102. Disponible en: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10907451/>
  51. Morales-Chan MA. Explorando el potencial de Chat GPT: Una clasificación de Prompts efectivos para la enseñanza [Internet]. Guatemala: Universidad

- Galileo; 2023 [citado el 1 de octubre de 2023]. Disponible en: <https://biblioteca.galileo.edu/xmlui/handle/123456789/1348>
52. López Galisteo AJ, Rodríguez Calzada L, Montes Diez R. Guía de uso de ChatGPT para potenciar el aprendizaje activo e interactivo en el aula universitaria [Internet]. Madrid: Universidad Rey Juan Carlos; 2023 [citado el 6 de octubre de 2023]. Disponible en: <https://hdl.handle.net/10115/22149>
  53. Townsend CM, Beauchamp D, Evers M, Mattox K. Sabiston: Tratado de cirugía. 21° ed. Barcelona: Elsevier; 2022. 2100 p.
  54. Nogales JA. Apendicitis aguda. Criterios de atención médica. Rev Asoc Med Argent [Internet]. 2017 [citado el 27 de septiembre de 2023];130(4):20–4. Disponible en: [https://ama-med.org.ar/uploads\\_archivos/1362/Rev-4-2017-Pag-20-Nogales.pdf](https://ama-med.org.ar/uploads_archivos/1362/Rev-4-2017-Pag-20-Nogales.pdf)
  55. Di Saverio S, Podda M, De Simone B, Ceresoli M, Augustin G, Gori A, et al. Diagnosis and treatment of acute appendicitis: 2020 update of the WSES Jerusalem guidelines. World Journal of Emergency Surgery [Internet]. 2020 [citado el 5 de octubre de 2023];15(27):1–42. Disponible en: <https://doi.org/10.1186/s13017-020-00306-3>
  56. Coronel Vasquez J, Rodriguez Gomez JH. Sensibilidad y especificidad de biomarcadores inflamatorios como predictores de apendicitis aguda en pacientes hospitalizados. Hospital II-2 Tarapoto. Enero – diciembre 2020 [Internet]. Repositorio - UNSM. [Tesis de grado]: Tarapoto: Universidad Nacional de San Martín; 2021 [citado el 28 de septiembre de 2023]. Disponible en: <http://repositorio.unsm.edu.pe/handle/11458/4167>
  57. Instituto de evaluación de tecnologías en salud e investigación. Guía de práctica clínica para el manejo de la apendicitis aguda en versión corta [Internet]. EsSalud. Lima; 2022 [citado el 6 de octubre de 2023]. Disponible en: [https://ietsi.essalud.gob.pe/wp-content/uploads/2022/12/GPC-manejo-de-Apendicitis-aguda\\_Version-corta.pdf](https://ietsi.essalud.gob.pe/wp-content/uploads/2022/12/GPC-manejo-de-Apendicitis-aguda_Version-corta.pdf)
  58. Garro Urbina V, Rojas Vázquez S, Thuel Gutiérrez M. Diagnóstico, evaluación y tratamiento de la apendicitis aguda en el servicio de emergencias. Revista Medica Sinergia [Internet]. 2019 [citado el 29 de septiembre de 2023];4(12):e316–e316. Disponible en: <https://doi.org/10.31434/rms.v4i12.316>

59. Donis JH. Evaluación de la validez y confiabilidad de una prueba diagnóstica. *Avances en Biomedicina* [Internet]. 2012 [citado el 29 de septiembre de 2023];1(2):73–81. Disponible en: <https://www.redalyc.org/articulo.oa?id=331328015005>
60. Cerda J, Cifuentes L. Uso de curvas ROC en investigación clínica. Aspectos teórico-prácticos. *Revista chilena de infectología* [Internet]. 2012 [citado el 30 de septiembre de 2023];29(2):138–41. Disponible en: <http://dx.doi.org/10.4067/S0716-10182012000200003>
61. Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, et al. On evaluation metrics for medical applications of artificial intelligence. *Sci Rep* [Internet]. 2022 [citado el 30 de septiembre de 2023];12(5979):1–9. Disponible en: <https://doi.org/10.1038/s41598-022-09954-8>
62. Rodríguez García PL, Rodríguez Pupo L. Principios técnicos para realizar la anamnesis en el paciente adulto. *Revista Cubana de Medicina General Integral* [Internet]. 1999 [citado el 30 de septiembre de 2023];15(4):409–14. Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S0864-21251999000400011&lng=es](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21251999000400011&lng=es)
63. Rodríguez García PL, Rodríguez Pupo L. Técnicas para el examen físico en medicina interna. *Medicentro Electrónica* [Internet]. 2003 [citado el 30 de septiembre de 2023];7(1):1–13. Disponible en: <https://medicentro.sld.cu/index.php/medicentro/article/view/1025/1034>
64. Montesinos-Correa H. Crecimiento y antropometría: aplicación clínica. *Acta Pediátr Mex* [Internet]. 2014 [citado el 30 de septiembre de 2023];35(2):159–65. Disponible en: <https://www.scielo.org.mx/pdf/apm/v35n2/v35n2a10.pdf>
65. Capurro D, Rada G. El proceso diagnóstico. *Rev Méd Chile* [Internet]. 2007 [citado el 30 de septiembre de 2023];135:534–8. Disponible en: <https://www.scielo.cl/pdf/rmc/v135n4/art18.pdf>
66. Cobo D, Daza P. Signos vitales en pediatría. *Revista Gastrohup* [Internet]. 2011 [citado el 30 de septiembre de 2023];13(1):s58–70. Disponible en: <https://hdl.handle.net/10893/5810>
67. Corona Martínez LA. Los exámenes complementarios en la práctica médica asistencial. Algunas consideraciones útiles para el médico en formación.

- MediSur [Internet]. 2010 [citado el 30 de septiembre de 2023];8(5):98–100. Disponible en: <https://www.redalyc.org/articulo.oa?id=180020098019>
68. Altman DG, Bossuyt PM. Estudios de precisión diagnóstica (STARD) y pronóstica (REMARK). *Med Clin (Barc)* [Internet]. 2005 [citado el 4 de octubre de 2023];125(Suppl 1):49–55. Disponible en: <https://www.elsevier.es/es-revista-medicina-clinica-2-pdf-13083734>
69. Gomez Bastar S. Metodología de la investigación. 1° ed. Mexico DF: Red tercer milenio; 2012.
70. Supo J. Seminarios de investigación científica. 2° ed. Editorial Bioestadístico EIRL; 2014. 340 p.
71. Dulzaides Iglesias ME, Molina Gómez AM. Análisis documental y de información: dos componentes de un mismo proceso. *ACIMED* [Internet]. 2004 [citado el 5 de octubre de 2023];12(2):1–1. Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1024-94352004000200011&lng=es](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352004000200011&lng=es)
72. Castro EMM. Bioestadística aplicada en investigación clínica: conceptos básicos. *Revista Médica Clínica Las Condes* [Internet]. el 1 de enero de 2019 [citado el 14 de enero de 2025];30(1):50–65. Disponible en: <https://www.elsevier.es/es-revista-revista-medica-clinica-las-condes-202-articulo-bioestadistica-aplicada-investigacion-clinica-conceptos-S0716864019300045>
73. Flores-Ruiz E, Miranda-Novales MG, Villasís-Keever MÁ. El protocolo de investigación VI: cómo elegir la prueba estadística adecuada. *Estadística inferencial. Rev Alerg Mex* [Internet]. el 16 de octubre de 2017 [citado el 14 de enero de 2025];64(3):364–70. Disponible en: <http://revistaalergia.mx/ojs/index.php/ram/article/view/304>

## **Anexos**

## Índice de anexos

Anexo A: Matriz de consistencia

Anexo B: Ficha de recolección de datos

Anexo C: Tabla de validación de instrumento por jueces expertos – Alfa de Cronbach

Anexo D: Rúbrica de evaluación por expertos

Anexo E: Prompt para ChatGPT-4

Anexo F: Base de datos 1 (BD1)

Anexo G: Base de datos 2 (BD2)

Anexo H: Aprobación del Comité de Ética Universidad Continental

Anexo I: Aprobación del Comité de Ética del Hospital Ramiro Prialé Prialé

Anexo J: Tabla N° 8

Anexo K: Tabla N° 9

Anexo L: Tabla N° 10

Anexo M: Gráfico N° 2

## Anexo A

### Matriz de consistencia

Problema	Objetivos	Marco Teórico	Hipótesis	Variables	Metodología
<p><b>Problema general</b> ¿Cuál es la precisión diagnóstica del ChatGPT-4 para Apendicitis Aguda en el servicio de emergencia del Hospital Ramiro Prialé Prialé?</p> <p><b>Problemas específicos</b> ¿Cuál es la validez del ChatGPT-4 para Apendicitis Aguda en el servicio de emergencia del Hospital Ramiro Prialé Prialé? ¿Cuál es la confiabilidad del ChatGPT-4 para</p>	<p><b>Objetivo general</b> Determinar la precisión diagnóstica del ChatGPT-4 para Apendicitis Aguda en el servicio de emergencia del Hospital Ramiro Prialé Prialé.</p> <p><b>Objetivos específicos</b> Evaluar la validez del ChatGPT-4 para Apendicitis Aguda en el servicio de emergencia del Hospital Ramiro Prialé Prialé Estimar la confiabilidad del ChatGPT-4 para Apendicitis Aguda en el servicio de</p>	<p><b>Base teórica</b> <b>Inteligencia Artificial</b></p> <ul style="list-style-type: none"> <li>• Historia</li> <li>• Definición</li> <li>• Funciones y aplicaciones en la salud</li> <li>• ChatGPT-4</li> <li>• Prompt</li> </ul> <p><b>Apendicitis aguda</b></p> <ul style="list-style-type: none"> <li>• Historia</li> <li>• Definición</li> <li>• Epidemiología</li> <li>• Fisiopatología</li> <li>• Etiología</li> <li>• Clínica</li> <li>• Diagnóstico</li> <li>• Exámenes auxiliares</li> </ul>	<p><b>Hipótesis general</b> La precisión diagnóstica del ChatGPT-4 para el diagnóstico de Apendicitis Aguda será adecuada</p> <p><b>Hipótesis específicas</b> El ChatGPT-4 demostrara ser válido si cumple con los siguientes criterios: sensibilidad &gt; 80%, especificidad &gt; 80% VPP &gt; 70%, VPN &gt; 90%, LR+ &gt; 5, LR- &gt; 0.2, índice de exactitud &gt; 90% y Score F1 &gt; 0.8</p>	<p>La presente investigación tiene 2 variables, siendo la variable dependiente la Precisión y la variable independiente es el ChatGPT-4.</p>	<p><b>1. Método, tipo, nivel y diseño</b> <b>Método general:</b> Método científico <b>Método específico:</b> Inductivo. <b>Tipo:</b> Observacional, retrospectivo y transversal <b>Nivel:</b> Descriptivo <b>Diseño:</b> No experimental, descriptivo, corte transversal</p> <p><b>2. Población y muestra</b> La población por investigar serán 5619 pacientes que se atendieron en el tópico de cirugía en el servicio de emergencia en el Hospital Nacional Ramiro Prialé Prialé en el año 2022. La muestra para este trabajo será probabilística, con 495 pacientes con un nivel de confianza de 98% y en margen de error del 5%.</p>

<p>Apendicitis Aguda en el servicio de emergencia del Hospital Ramiro Prialé Prialé?</p>	<p>emergencia del Hospital Ramiro Prialé Prialé</p>	<ul style="list-style-type: none"> <li>• Diagnósticos diferenciales</li> <li>• Tratamiento</li> </ul> <p><b>Prueba diagnóstica</b></p> <p><b>Precisión diagnóstica</b></p> <p>Validez de una prueba diagnóstica</p> <ul style="list-style-type: none"> <li>• Sensibilidad</li> <li>• Especificidad</li> <li>• Valor Predictivo Positivo</li> <li>• Valor Predictivo Negativo</li> <li>• Curva de ROC</li> <li>• Likelihood Ratio</li> <li>• Índice de Exactitud</li> <li>• Score F1</li> </ul> <p><b>Confiabilidad</b></p> <ul style="list-style-type: none"> <li>• Coeficiente de correlación de Matthews</li> </ul>	<p>El ChatGPT-4 demostrará ser confiable si alcanza un coeficiente de correlación de Matthews &gt; 0.5</p>	<p><b>3. Técnicas e instrumentos de recolección de datos</b></p> <p>Ficha de recolección de datos</p> <p><b>4. Recopilación y manejo de datos</b></p> <p>El estudio organiza datos clínicos en Google Forms, generando casos en la BD1 para que ChatGPT-4 diagnostique apendicitis aguda. Las respuestas, afirmativas o negativas, junto con resultados histopatológicos, se registran en BD2. El Hospital Ramiro Prialé proporcionó datos anatómicos que confirman diagnósticos definitivos, integrados a BD2 como referencia para validar la investigación.</p> <p><b>5. Análisis de los datos</b></p> <p>Se analizarán estadísticamente las respuestas de ChatGPT-4 junto a las anatomo patologías. Se usarán los programas STATA 18 y Microsoft Excel</p>
--	---	---	--	---

## Anexo B

### Ficha de recolección de datos

Código		Fecha de atención		Hora		
	Edad		Sexo			
<b>ANAMNESIS</b>						
<b>FUNCIONES VITALES</b>						
	PA		T°		FC	
	FR		Peso		Talla	
<b>EXAMEN FÍSICO</b>						
<b>DIAGNÓSTICO</b>						
	Diagnóstico N°1					
	Diagnóstico N°2					
	Diagnóstico N°3					
	Diagnóstico N°4					
<b>EXÁMENES AUXILIARES</b>						
Hemograma	Leucocitos		Eritrocito		Hb	
	VCM		HCM		CHCM	
	Plaquetas		Mielocito		Metamielocitos	
	Abastionados		Segmentados		Eosinófilos	
	Baófilos		Linfocitos		Monocitos	
Hematológico	Grupo		Factor			
	TPT		INR		PPT	
	T Coagulación		T Sangría			
Bioquímico	Creatinina		Glucosa		Urea	
	Amilasa		BT		BI	
	BD		TGO		TGP	
	Gamma GT		FA			
Examen de Orina	Color		Aspecto		pH	
	Densidad		Glucosa		C. Cetónico	
	Ac Ascorbico		Bilirrubina		Urobilinógeno	
	Nitritos		Thevenon		Proteínas	
	Leucocitos		Hematíes		Cel epiteliales	
	Bacterias		Cilindros		Cristales	
Imágenes						
Otros						

### Anexo C

Tabla de validación de instrumento por jueces expertos – Alfa de cronbach

Encuestados	ITEMS					SUMA
	I1	I2	I3	I4	I5	
E1	5	5	5	5	5	25
E2	5	5	5	5	5	25
E3	4	4	4	4	4	20
E4	5	5	5	5	5	25
E5	5	5	4	5	5	24
Varianza	0.16	0.16	0.24	0.16	0.16	
Sumatoria de Varianzas	0.88					
Varianza de la suma de los elementos	3.76					

Coeficiente de confiabilidad del instrumento	<b>0.95744681</b>
Número de ítems del instrumento	5
Sumatoria de la varianza de los ítems	0.88
Varianza total del instrumento	3.76

E1	BRAYAN KEVIN CAPCHA LOPEZ
E2	LIZETH CHAVEZ QUISPE
E3	ANDREA ESTEFANIA ALBAÑIL GARCIA
E4	MAX FREY HERNANDEZ ZEVALLOS
E5	SANDRA ENEDINA BARRIENTOS VERA

## Anexo D

### Rúbrica de evaluación por expertos

#### INFORMACIÓN DEL ESPECIALISTA

Nombres y Apellidos	BRAYAN KEVIN CAPCHA LOPEZ
Profesión y Grado Académico	CIRUJANO GENERAL.
Especialidad	CIRUJANO GENERAL.
Institución y años de experiencia	04 años
Cargo que desempeña actualmente	CIRUJANO GENERAL.

Puntaje del Instrumento Revisado: 25

#### Opinión de aplicabilidad:

APLICABLE (X)      APLICABLE LUEGO DE REVISIÓN ( )      NO APLICABLE ( )



Firma: \_\_\_\_\_

Nombres y apellidos: Brayan Kevin Capcha Lopez

DNI: 72908186

CMP: 084759

**INFORMACIÓN DEL ESPECIALISTA**

Nombres y Apellidos	Lizeth Chávez Quirope
Profesión y Grado Académico	Cirujana general
Especialidad	Cirugía general
Institución y años de experiencia	IREN centro (15)
Cargo que desempeña actualmente	CIRUJANA GENERAL.

Puntaje del Instrumento Revisado: 25

**Opinión de aplicabilidad:**

APLICABLE (  )

APLICABLE LUEGO DE REVISIÓN ( )

NO APLICABLE ( )

  
 Dra. Norka Chávez Quirope  
 CIRUGIA GENERAL Y LAPAROSCOPIA  
 CMP 75575

Firma: \_\_\_\_\_

Nombres y apellidos: Lizeth Chávez Quirope

DNI: 71594376

CMP: 75575

**INFORMACIÓN DEL ESPECIALISTA**

Nombres y Apellidos	Max Frey Hernández Zevallos
Profesión y Grado Académico	Médico especialista en Cirugía General
Especialidad	Cirugía General
Institución y años de experiencia	Hospital Daniel A. Carrion Huancayo 21 años de especialista
Cargo que desempeña actualmente	Médico asistente

Puntaje del Instrumento Revisado: 25

**Opinión de aplicabilidad:**

APLICABLE     APLICABLE LUEGO DE REVISIÓN ( )    NO APLICABLE ( )

Firma:

  
HOSPITAL REGIONAL ORIENTE CLÍNICO QUIRÚRGICO  
"DANIEL ALCHIBES CARRIÓN" - HUANCAYO  
**MAX FREY HERNÁNDEZ ZEVALLOS**  
CIRUGÍA GENERAL  
C.M.P. 33318 F.N.E. 14607

Nombres y apellidos:

Max Frey Hernández Zevallos

DNI:

29638172

CMP:

33318

**INFORMACIÓN DEL ESPECIALISTA**

Nombres y Apellidos	Sandra Enedine Barrientos Vera
Profesión y Grado Académico	Médico Cirujano
Especialidad	Cirujía General
Institución y años de experiencia	H.D.C.A Daniel A. Carrion - 8a
Cargo que desempeña actualmente	Cirujana general.

Puntaje del Instrumento Revisado: 24

**Opinión de aplicabilidad:**

APLICABLE       APLICABLE LUEGO DE REVISIÓN ( )      NO APLICABLE ( )

Firma:   
 Nombres y apellidos: Sandra E. Barrientos Vera  
 DNI: 4412812  
 CMP: 57564

GOBIERNO REGIONAL JUNÍN  
 HOSPITAL REGIONAL DOCENTE CLÍNICO QUIRÚRGICO  
 "DANIEL ALCIDES CARRIÓN" - HUANCAYO

SANDRA ENEDINA BARRIENTOS VERA  
 CIRUJIA GENERAL  
 C.M.P. 07006 R.N.E. 31886

## Anexo E

### Prompt para Chatgpt-4

ACTÚA COMO UN CIRUJANO GENERAL DE LARGA EXPERIENCIA, ACTUALMENTE TE ENCUENTRAS EN UN ESTUDIO EN EL QUE DEBERÁS DETERMINAR A TRAVÉS DE HISTORIAS CLÍNICAS DE LOS PACIENTES SI ESTOS PRESENTAN, O NO PRESENTAN APENDICITIS AGUDA. DENTRO DE ESTA HISTORIA CLÍNICA EXISTEN APARTADOS CON LAS SIGLAS “NRD” LO CUAL SIGNIFICA QUE NO SE REALIZÓ DICHO EXAMEN O NO SE TIENE ESA INFORMACIÓN. LA HISTORIA CLÍNICA N° «Código» ES LA SIGUIENTE:

PACIENTE «Sexo» DE «Edad» AÑOS REFIERE QUE «ANAMNESIS», CUENTA CON LOS SIGUIENTES FUNCIONES VITALES, PRESIÓN ARTERIAL «Presión\_Arterial» MMHG, TEMPERATURA «Temperatura» °C, FRECUENCIA CARDIACA «Frecuencia\_Cardiaca» LPM, FRECUENCIA RESPIRATORIA «Frecuencia\_Respiratoria» RPM, PESO «Peso» KG, TALLA «Talla» M, AL EXAMEN FÍSICO: «EXAMEN\_FÍSICO» ADEMÁS CUENTA CON HEMOGRAMA: LEUCOCITOS «Hg\_Leucocitos»  $10^3/\text{MM}^3$ , ERITROCITOS «Hg\_Eritrocitos»  $10^6/\text{MM}^3$ , HEMOGLOBINA «Hg\_Hemoglobina» G/DL, HEMATOCRITO «Hg\_Hematocrito» %, VCM «Hg\_VCM» FL, HCM «Hg\_HCM» PG, CHCM «Hg\_CHCM» G/DL, PLAQUETAS «Hg\_Plaquetas»  $10^3/\text{MM}^3$ , MIELOCITOS «Hg\_Mielocitos»%, METAMIELOCITOS «Hg\_Metamielocitos»%, ABASTONADOS «Hg\_Abastonados»%, SEGMENTADOS «Hg\_Segmentados»%, EOSINÓFILOS «Hg\_Eosinófilos»%, BASÓFILOS «Hg\_Basófilos»%, LINFOCITOS «Hg\_Linfocitos»%, MONOCITOS «Hg\_Monocitos»%, GRUPO Y FACTOR «Ht\_G\_y\_F». BIOQUÍMICA SANGUÍNEA: CREATININA «Bq\_Creatinina» MG/DL, GLUCOSA «Bq\_Glucosa» MG/DL, UREA «Bq\_Urea» MG/DL, AMILASA «Bq\_Amilasa» U/L, BILIRRUBINA TOTAL «Bq\_Bilirrubina\_Total» MG/DL, BILIRRUBINA INDIRECTA «Bq\_Bilirrubina\_Indirecta» MG/DL, BILIRRUBINA DIRECTA «Bq\_Bilirrubina\_Directa» MG/DL, TGO «Bq\_TGO» U/L, TGP «Bq\_TGP» U/L, GAMMA GT «Bq\_Gamma\_GT» U/L, FOSFATASA ALCALINA «Bq\_Fosfatasa\_Alcalina» U/L. GRUPO Y FACTOR «Ht\_G\_y\_F». PERFIL DE COAGULACIÓN: TIEMPO DE PROTROMBINA «Ht\_Tiempo\_de\_protrombina» SEG, INR «Ht\_INR», TIEMPO DE TROMBOPLASTINA PARCIAL «Ht\_Tiempo\_de\_Tromboplastina\_parcial\_P» SEG, TIEMPO DE SANGRÍA «Ht\_Tiempo\_de\_Sangría» MIN, TIEMPO DE COAGULACIÓN «Ht\_Tiempo\_de\_Coagulación» MIN. EXAMEN DE ORINA: COLOR «EO\_Color», ASPECTO «EO\_Aspecto», PH «EO\_pH», DENSIDAD «EO\_Densidad», GLUCOSA «EO\_Glucosa» MG/DL, CUERPOS CETÓNICOS, «EO\_Cuerpo\_Cetónico» MG/DL, ÁCIDO ASCÓRBICO «EO\_Ácido\_Ascórbico» MG/DL, BILIRRUBINA «EO\_Bilirrubina» MG/DL, UROBILINÓGENO «EO\_Urobilinógeno» MG/DL, NITRITOS «EO\_Nitritos», THEVENON «EO\_Thevenon» ERY/UL, PROTEÍNAS «EO\_Proteinas» MG/DL, LEUCOCITOS «EO\_Leucocitos» X CAMPO, HEMATÍES «EO\_Hematíes» X CAMPO, CÉLULAS EPITELIALES «EO\_Células\_epiteliales» X CAMPO, BACTERIAS «EO\_Bacterias», CILINDROS «EO\_Cilindros», CRISTALES «EO\_Cristales», ADEMÁS CUENTA CON IMÁGENES «Tipo\_de\_estudio», «Otros\_Introducir\_en\_texto\_plano».

AHORA TENIENDO ESTA INFORMACIÓN ÚNICAMENTE DEBERÁS RESPONDER TEXTUALMENTE “EL PACIENTE TIENE APENDICITIS AGUDA” O “EL PACIENTE NO TIENE APENDICITIS AGUDA”.



Hg. Hemoglobina	Hg. Hemocrito	Hg. HCM	Hg. HCM	Hg. DCM	Hg. Plaquetas	Hg. Leucocitos	Hg. Neutrofilos	Hg. Linfoцитos	Hg. Monocitos	Hg. Eosinofilos	Hg. Basofilos	Hg. Linfocitos	Hg. Monocitos	IGMANTIGLOBULINA	nt. G.F.	H. Tiempo de protrombina	H. INR	H. Tiempo de Tromboplastina parcial (TPT)	H. Tiempo de Coagulacion	H. Tiempo de Sangria	BDQBRACOS	Bq. Creatinina	Bq. Ureica	Bq. Urea	Bq. Amilasa
16.9	47.9	95.1	32	35.3	379	0	0	0	76.6	0.1	0.3	20.7	1.9 SI	0*					5	1.3 SI	0.92	125	30.9	100	
16.6	47.9	86.9	30.1	34.7	270	0	0	0	77.9	0.2	0.3	13.7	7.7 SI	0*	12	1		7	2 SI	0.83	97	23.6	53		
18.1	53.6	92.4	32.9	35.6	167	0	0	1	89.7	0	0.2	4.3	5.4 SI	0*		13.5	1.15		6	1.3 SI	0.91	110	27.4		
18.3	55.9	89.9	31.6	34.9	378	0	0	1	88	0.2	0.4	7.4	4.3 SI	0*		14.9	1.1	29.6	7	2 SI	1	102	25.2		
17.9	48.9	90.4	32.4	35.9	183	0	0	0	21.1	2	0.4	69.5	4.8 SI	0*		15.5	1.33	39.7	6	2 SI	1.06	96			
15	43.9	91.9	31.8	34.6	307	0	0	0	59	3.3	0.3	36.5	6.7 SI	0*		14.6			7	2 SI	0.91	109	28.7		
10.4	31.2	89.9	30.5	34	108	0	0	0	26	0.6	0.6	58.9	12.8 SI	0*		19.7	1.15	31.2	4.3	2 SI	0.5	104			
15.3	41.3	90.8	31.4	34.3	269	0	0	0	24	1	0.2	15.3	6.1 SI	0*					7	2 SI	0.98	87	20.3	48	
17.4	51.3	87.8	29.8	33.9	184	0	0	0	33.3	2	0.5	37.5	7.3 SI	0*				8.30'	2.00'		0.97	83			
15	43.2	87.1	30.2	34.7	276	0	0	0	95.4	0.1	0.1	1.2	2.9 No											46	
16.7	47.9	89.7	31.3	34.9	259	0	0	0	47.1	1.4	0.6	44.1	6.6 SI	0*		14.8	1.25	7.00'	1.40'		0.92	145	33		
15.4	44.9	88.6	30.4	34.3	269	0	0	0	71.1	0.9	0.1	21.6	6.1 SI	0*					7	2 SI	0.98	75	22.3	47	
15.9	46.7	90.7	30.9	34	305	0	0	0	85.9	0.1	0.2	7.3	6 SI	0*		14.1					0.84	108			
16	44.1	88.8	29.2	34	317	0	0	0	89.7	0.6	0.2	25.5	5.7 SI	0*					6	3 SI	0.84	90	18.2	90	
14.6	43.5	87.7	29.4	33.6	242	0	0	0	69	1.5	0	21.3	8.3 SI			15.6	1.33	39.7	8	2.3 SI	0.91	88	18	71	
16.4	46.1	92.1	31.4	34.1	308	0	0	0	83.7	0	0.1	16.6	5 SI	0*		10.2			4.3	2.3 SI	0.76	117	32.9	75	



## Anexo G

### BASE DE DATOS 2 (BD2)

Código	Respuesta ChatGPT	Anatomopatología
5	1	0
15	0	0
16	0	0
40	0	0
48	0	0
106	0	0
107	0	0
109	1	1
110	0	0
121	0	0
131	0	0
152	0	0
161	0	0
164	0	0
178	0	0
180	1	0
200	0	0
202	0	0
212	0	0
224	0	0
226	0	0
241	0	0
242	0	0
248	0	0
262	0	0
271	0	0
282	0	0
285	0	0
288	1	0
294	0	0
327	0	0
355	0	0
357	1	0
370	0	0
372	0	0
378	0	0
384	0	0
389	0	0
404	0	0
411	0	0
427	0	0
429	0	0
445	0	0
450	0	0
469	0	0
480	1	1
482	0	0
484	0	0
491	0	0
500	0	0
512	0	0
514	0	0
517	0	0
521	0	0
523	0	0
558	0	0
599	0	0
606	0	0
615	0	0
635	0	0
640	0	0
646	1	1
660	0	0
698	0	0
699	1	1
706	0	0
728	0	0
743	0	0
748	0	0
756	1	1
763	0	0
770	0	0
771	0	0
799	0	0
824	0	0
825	0	0
832	0	0
833	0	0
854	0	0
870	0	0
886	1	1
902	0	0
903	0	0
908	0	0
910	0	0
941	0	0
944	0	0
984	0	0
1002	0	0
1019	0	0
1020	0	0
1028	0	0
1037	0	0
1041	0	0
1047	0	0
1076	0	0
1088	0	0

1090	0	0
1099	0	0
1104	1	1
1106	1	1
1114	0	0
1121	0	0
1126	1	1
1145	0	0
1149	0	0
1158	0	0
1161	0	0
1180	0	0
1203	0	0
1214	0	0
1215	0	0
1238	0	0
1255	0	0
1270	0	0
1276	0	0
1285	0	0
1303	0	0
1308	0	0
1337	0	0
1351	0	0
1352	0	0
1369	0	0
1384	0	0
1402	0	0
1408	0	0
1412	0	0
1413	0	0
1431	0	0
1454	0	0
1455	0	0
1457	0	0
1471	0	0
1478	0	0
1479	0	0
1489	0	0
1495	0	0
1498	0	0
1504	1	1
1551	0	0
1561	0	0
1568	0	0
1573	0	0
1582	0	0
1583	0	0
1589	0	0
1593	0	0
1595	0	0
1599	1	0
1605	0	0
1609	0	0
1612	0	0
1617	0	0
1618	0	0
1629	0	0
1656	0	0
1664	0	0
1665	0	0
1679	0	0
1695	0	0
1699	0	0
1706	0	0
1708	0	0
1716	0	0
1729	0	0
1734	0	0
1747	0	0
1752	0	1
1757	0	0
1760	0	0
1768	0	0
1772	0	0
1779	0	0
1781	0	0
1782	0	0
1785	0	0
1788	0	0
1808	0	0
1820	0	0
1825	0	0
1845	0	0
1863	0	0
1882	0	0
1919	0	0
1927	0	0
1935	0	0
1961	0	0
1962	0	0
1966	0	0
1967	0	0
1973	0	0
1980	0	0
1982	0	0
1988	0	0
1999	0	0

2017	0	0
2025	0	0
2057	0	0
2063	0	0
2068	0	0
2085	0	0
2086	0	0
2087	1	1
2103	0	0
2104	0	0
2153	1	1
2161	0	0
2163	0	0
2164	0	0
2169	0	0
2174	0	0
2175	0	0
2179	0	0
2184	0	0
2191	0	0
2201	0	0
2206	0	0
2273	0	0
2275	0	0
2289	0	0
2322	0	0
2326	0	0
2332	0	0
2333	0	0
2344	0	0
2349	0	0
2368	0	0
2373	0	0
2375	0	0
2377	0	0
2395	0	0
2419	0	0
2424	0	0
2435	0	0
2450	0	0
2470	0	0
2480	0	0
2485	0	0
2488	0	0
2491	1	1
2493	0	0
2503	0	0
2504	1	0
2524	1	1
2527	0	0
2537	0	0
2543	0	0
2567	0	0
2604	0	0
2614	1	1
2617	0	0
2620	0	0
2633	0	0
2675	0	0
2682	0	0
2703	0	0
2719	0	0
2739	0	0
2747	1	1
2784	0	0
2807	0	0
2810	0	0
2814	0	0
2821	1	0
2829	0	0
2857	0	0
2860	0	0
2869	0	0
2873	0	0
2878	0	0
2891	0	0
2910	0	0
2916	0	0
2934	0	0
2938	0	0
2942	0	0
2945	0	0
2958	0	0
2961	0	0
2964	0	0
2971	1	1
2978	0	0
3024	1	0
3025	0	0
3037	0	0
3039	0	0
3069	0	0
3086	0	0
3098	0	0
3105	0	0
3106	0	0
3116	0	0
3167	0	0

3172	0	0
3177	0	0
3188	0	0
3189	0	0
3192	0	0
3208	0	0
3211	0	0
3214	0	0
3220	0	0
3237	0	0
3268	0	0
3272	0	0
3290	0	0
3301	0	0
3305	0	0
3324	0	0
3328	1	1
3335	0	0
3340	0	1
3342	0	0
3350	0	0
3357	0	0
3397	0	0
3424	0	0
3433	0	0
3440	0	0
3465	0	0
3484	0	0
3492	0	0
3567	1	1
3590	0	0
3597	0	0
3604	0	0
3614	0	0
3624	0	0
3641	0	0
3645	0	0
3648	0	0
3653	0	0
3669	0	0
3694	0	0
3696	0	0
3705	0	0
3709	0	0
3710	0	0
3738	0	0
3742	0	0
3756	0	1
3780	0	0
3790	0	0
3794	0	0
3798	0	0
3801	0	0
3812	0	0
3815	0	0
3816	0	0
3817	0	0
3826	1	1
3859	0	0
3872	0	0
3885	0	0
3906	0	0
3921	0	0
3934	0	0
3977	0	0
3983	0	0
3994	0	0
3995	0	0
4009	1	1
4012	1	0
4024	0	1
4027	0	0
4038	0	0
4080	0	0
4087	0	0
4089	0	0
4105	0	0
4107	0	0
4109	0	0
4114	0	0
4146	0	0
4149	0	0
4206	0	0
4211	1	1
4221	0	0
4230	0	0
4235	0	0
4272	0	0
4333	0	0
4346	0	0
4358	1	1
4364	0	0
4371	0	0
4381	1	1
4394	0	0
4395	0	0
4406	0	0
4418	0	0

4443	0	0
4451	0	0
4453	0	0
4462	0	0
4463	1	1
4508	0	0
4524	1	0
4534	1	1
4544	0	0
4551	1	1
4574	0	0
4578	0	0
4583	0	0
4604	0	0
4621	0	0
4626	0	0
4628	0	0
4630	0	0
4631	0	0
4638	0	0
4639	0	0
4654	0	0
4676	0	0
4688	1	1
4715	0	0
4718	0	0
4729	0	0
4731	0	0
4734	0	0
4739	0	0
4744	0	0
4746	0	0
4751	0	0
4774	0	0
4812	0	0
4813	0	0
4843	0	0
4848	0	0
4911	0	0
4916	0	0
4917	0	0
4932	0	0
4937	0	0
4939	0	0
4941	0	0
4951	1	1
4952	0	0
4961	0	0
4982	0	0
4996	0	0
4999	0	0
5002	0	0
5028	0	0
5029	1	1
5035	1	1
5037	0	0
5045	0	0
5055	0	0
5063	1	1
5100	0	0
5133	0	0

5185	0	0
5187	0	0
5189	0	0
5191	0	0
5202	0	0
5216	0	0
5224	0	0
5231	0	0
5232	0	0
5250	0	0
5265	0	0
5268	0	0
5273	0	0
5320	0	0
5324	0	0
5330	0	0
5344	1	1
5346	0	0
5349	0	0
5358	0	0
5408	0	0
5415	1	0
5444	0	0
5453	0	0
5465	0	0
5473	0	0
5476	0	0
5480	0	0
5504	0	0
5528	0	0
5543	0	0
5544	0	0
5556	0	0
5560	0	0
5562	0	0
5568	0	0
5574	0	0
5583	0	0
5597	0	0
5600	0	0
5602	0	0
5609	0	0
5617	0	0

## Anexo H

### Aprobación del comité de ética de la Universidad Continental



"Año de la unidad, la paz y el desarrollo"

Huancayo, 29 de noviembre del 2023

#### OFICIO N°0748-2023-CIEI-UC

Investigadores:

**BARRERA CHUQUIARQUE DIANA EVELYN  
ROJO CAMAYO VÍCTOR RICARDO**

#### Presente-

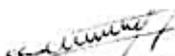
Tengo el agrado de dirigirme a ustedes para saludarles cordialmente y a la vez manifestarles que el estudio de investigación titulado: **PRECISIÓN DIAGNÓSTICA DEL CHATGPT-4 PARA APENDICITIS AGUDA EN EL SERVICIO DE EMERGENCIA DEL HOSPITAL RAMIRO PRIALÉ PRIALÉ, HUANCAYO 2022.**

Ha sido **APROBADO** por el Comité Institucional de Ética en Investigación, bajo las siguientes precisiones:

- El Comité puede en cualquier momento de la ejecución del estudio solicitar información y confirmar el cumplimiento de las normas éticas.
- El Comité puede solicitar el informe final para revisión final.

Aprovechamos la oportunidad para renovar los sentimientos de nuestra consideración y estima personal.

Atentamente


Walter Calderón Gerstein  
Presidente del Comité de Ética  
Universidad Continental

C.c. Archivo.

**Arequipa**  
Av. Los Inca S/N,  
José Luis Bustamante y Rivero  
(054) 412 000

Calle Alfonso Ugarte 607, Yanahuara  
(054) 412 000

**Huancayo**  
Av. San Carlos 1880  
(064) 481430

**Cusco**  
Urb. Manuel Prado - Lote B, N°7 Av. Colcauyo  
(084) 480 070

Sector Angostura KM. 10,  
carretera San Jerónimo - Sayta  
(084) 480 070

**Umas**  
Av. Alredo Mendola 520, Los Olivos  
(01) 213 2780

h. Junín 355, Misaflores  
(01) 213 2780

## Anexo I

### Aprobación del comité de ética del Hospital Ramiro Prialé Prialé



COD. CIEIRAJ 48-24

"Decenio de la igualdad de oportunidades para mujeres y hombres"  
"Año del Bicentenario, de la consolidación de nuestra Independencia, y  
de la conmemoración de las heroicas batallas de Junín y Ayacucho"

#### CONSTANCIA DE APROBACIÓN DE PROYECTO DE INVESTIGACIÓN

##### CONSTANCIA N° 48-CIEI-GRAJ-ESSALUD-2024

Huancayo, 17 de junio de 2024

Investigador (a) Principal  
**Victor Ricardo Rojo Camayo**  
Presente. –

**Título del Protocolo** : "Precisión diagnóstica del Chatgpt-4 para apendicitis aguda en el Servicio de Emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022"  
**Versión y Fecha del Protocolo:** Versión 03, 10 de abril del 2024  
**Tipo de Estudio** : Observacional

De nuestra consideración:

El Comité Institucional de ética en Investigación ha revisado la solicitud de revisión de protocolo de investigación expresada en su carta del 15 de diciembre del 2023.

Para la aprobación se ha considerado el cumplimiento de las consideraciones éticas para la investigación en salud con seres humanos señaladas en la Resolución Ministerial N°233-2020. En virtud a ello ha **aprobado** el siguiente documento:

- Protocolo "Precisión diagnóstica del Chatgpt-4 para apendicitis aguda en el Servicio de Emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022" Versión 03, 10 de abril del 2024.

El periodo de vigencia de la presente aprobación será de (03) meses; desde el 17 de junio al 16 de setiembre de 2024, debiendo solicitar la renovación con 30 días de anticipación al Comité Institucional de Ética en la Investigación de la Red Asistencial Junín.

Asimismo, mencionar que cualquier enmienda en los objetivos secundarios, metodología y aspectos éticos debe ser solicitada a este CIEI.

Sírvase hacernos llegar los **informes de avance del estudio en forma digital semestralmente** al correo electrónico [ciei.redjunin@gmail.com](mailto:ciei.redjunin@gmail.com) a partir la presente aprobación y el informe final una vez concluido el estudio.

Dr. FRANCK QUISPE PARI  
PRESIDENTE  
COMITÉ INSTITUCIONAL DE ÉTICA EN LA INVESTIGACIÓN  
RED ASISTENCIAL JUNÍN

C.c. Unidad de Capacitación, Investigación y Docencia  
JFQP/ascp  
NIT : 1302-2023-11843

[www.essalud.gob.pe](http://www.essalud.gob.pe)

Av. Independencia 296  
El Tambo Huancayo  
Junín Perú  
T.: (064) 248366





PERÚ

Ministerio  
de Trabajo  
y Promoción del Empleo

Seguro Social de Salud  
EsSalud



"Decenio de la Igualdad de Oportunidades para mujeres y hombres"  
"Año del Bicentenario, de la consolidación de nuestra Independencia, y de la  
conmemoración de las heroicas batallas de Junín y Ayacucho"

**CARTA N° 69-UCID-GRAJ-ESSALUD-2024**

Huancayo, 17 de junio de 2024

Señor (a)  
**Victor Ricardo Rojo Camayo**  
Investigador(a) Principal  
**Presente.** –

**ASUNTO :** Autorización para la ejecución de proyecto de Investigación

**Referencia :** a. Carta N° 65-GRAJ-ESSALUD-2024  
b. Constancia N° 48-CIEI-GRAJ-ESSALUD-2024

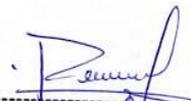
Es grato dirigirme a usted para saludarlo cordialmente y en mención al documento de la referencia b) el Comité Institucional de Ética en la Investigación de la Red Asistencial Junín, aprueba el Proyecto de Investigación, en el cual se ha considerado el cumplimiento de pautas éticas en investigación, incluyendo el balance beneficio/riesgo, confidencialidad de los datos y otros.

En ese sentido, la Unidad de Capacitación, Investigación y Docencia de la Red Asistencial Junín **AUTORIZA LA EJECUCIÓN** de la investigación titulada "*Precisión diagnóstica del Chatgpt-4 para apendicitis aguda en el Servicio de Emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022*".

Es preciso señalar, que el periodo de vigencia de la presente autorización será de **03 meses**; desde el **17 de junio al 16 de setiembre de 2024** debiendo solicitar la renovación de ejecución al Comité de Ética en Investigación, si transcurrido el tiempo de autorización señalado líneas arriba no se culminó con la investigación.

Sin otro particular, me despido de usted.

Atentamente.

  
LIC. NATALY DEL ROCIO FONSECA SEGURA  
JEFE (a) UNIDAD DE CAPACITACIÓN  
INVESTIGACIÓN Y DOCENCIA  
RED ASISTENCIAL JUNÍN  


NRFS/ascp  
NIT°: 1302-2023-11843  
Folios: ( )

[www.essalud.gob.pe](http://www.essalud.gob.pe)

Av. Independencia 296  
El Tambo Huancayo  
Junín Perú  
T.: (064) 248366



## Anexo J

**Tabla N.º 8. Tabla de frecuencias relativas de los pacientes atendidos en el tópico de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022.**

		<b>N</b>	<b>%</b>
<b>Hora del día</b>	00h - 05h	43	8,7 %
	06h - 11h	171	34,5 %
	12h - 17h	169	34,1 %
	18h - 23h	112	22,6 %
	Total	495	100,0 %
<b>Sexo</b>	Femenino	282	57,0 %
	Masculino	213	43,0 %
	Total	495	100,0 %
<b>Con funciones vitales descritas</b>	Sin	479	96,8 %
	Con	16	3,2 %
	Total	495	100,0 %
<b>Cantidad de diagnósticos</b>	1	439	88,7 %
	2	49	9,9 %
	3	4	0,8 %
	4	3	0,6 %
	Total	495	100,0 %
<b>Capítulo de diagnóstico</b>	A	3	0,6 %
	B	3	0,6 %
	C	4	0,8 %
	D	1	0,2 %
	E	3	0,6 %
	F	0	0,0 %
	G	0	0,0 %
	H	16	3,2 %
	I	14	2,8 %
	J	3	0,6 %
	K	79	16,0 %
	L	10	2,0 %
	M	24	4,8 %
	N	6	1,2 %
	Q	1	0,2 %
	R	135	27,3 %
	S	114	23,0 %
	T	56	11,3 %
	W	12	2,4 %
	X	1	0,2 %
Z	10	2,0 %	
Total	495	100,0 %	

<b>Con hemograma</b>	Sin	353	71,3 %
	Con	142	28,7 %
	Total	495	100,0 %
<b>Con perfil de coagulación</b>	Sin	366	73,9 %
	Con	129	26,1 %
	Total	495	100,0 %
<b>Con mediciones bioquímicas</b>	Sin	355	71,7 %
	Con	140	28,3 %
	Total	495	100,0 %
<b>Con exámen de orina</b>	Sin	388	78,4 %
	Con	107	21,6 %
	Total	495	100,0 %
<b>Con imágenes</b>	Sin	412	83,2 %
	Con	83	16,8 %
	Total	495	100,0 %
<b>Con otras mediciones</b>	Sin	391	79,0 %
	Con	104	21,0 %
	Total	495	100,0 %

## Anexo K

**Tabla N.º 9. Tabla de estadísticos de tendencia central de los pacientes atendidos en el tópico de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022.**

		Anatomopatología				Respuesta ChatGPT-4			
		Sin apendicitis		Con apendicitis		Sin apendicitis		Con apendicitis	
		N	%	N	%	N	%	N	%
<b>Hora del día</b>	00h - 05h	37	86,0 %	6	14,0 %	38	88,4 %	5	11,6 %
	06h - 11h	160	93,6 %	11	6,4 %	159	93,0 %	12	7,0 %
	12h - 17h	159	94,1 %	10	5,9 %	156	92,3 %	13	7,7 %
	18h - 23h	102	91,1 %	10	8,9 %	98	87,5 %	14	12,5 %
	Total	458	92,5 %	37	7,5 %	451	91,1 %	44	8,9 %
<b>Sexo</b>	Femenino	262	92,9 %	20	7,1 %	259	91,8 %	23	8,2 %
	Masculino	196	92,0 %	17	8,0 %	192	90,1 %	21	9,9 %
	Total	458	92,5 %	37	7,5 %	451	91,1 %	44	8,9 %
<b>Con funciones vitales descritas</b>	Sin	443	92,5 %	36	7,5 %	437	91,2 %	42	8,8 %
	Con	15	93,8 %	1	6,3 %	14	87,5 %	2	12,5 %
	Total	458	92,5 %	37	7,5 %	451	91,1 %	44	8,9 %
<b>Cantidad de diagnósticos</b>	1	407	92,7 %	32	7,3 %	401	91,3 %	38	8,7 %
	2	44	89,8 %	5	10,2 %	43	87,8 %	6	12,2 %
	3	4	100,0 %	0	0,0 %	4	100,0 %	0	0,0 %
	4	3	100,0 %	0	0,0 %	3	100,0 %	0	0,0 %
	Total	458	92,5 %	37	7,5 %	451	91,1 %	44	8,9 %
<b>Capítulo de diagnóstico</b>	A	3	100,0 %	0	0,0 %	3	100,0 %	0	0,0 %
	B	3	100,0 %	0	0,0 %	3	100,0 %	0	0,0 %
	C	4	100,0 %	0	0,0 %	4	100,0 %	0	0,0 %
	D	1	100,0 %	0	0,0 %	1	100,0 %	0	0,0 %
	E	3	100,0 %	0	0,0 %	3	100,0 %	0	0,0 %
	F	0	0,0 %	0	0,0 %	0	0,0 %	0	0,0 %
	G	0	0,0 %	0	0,0 %	0	0,0 %	0	0,0 %
	H	16	100,0 %	0	0,0 %	16	100,0 %	0	0,0 %
	I	14	100,0 %	0	0,0 %	14	100,0 %	0	0,0 %
	J	3	100,0 %	0	0,0 %	3	100,0 %	0	0,0 %
	K	70	88,6 %	9	11,4 %	67	84,8 %	12	15,2 %
	L	10	100,0 %	0	0,0 %	10	10,0 %	0	0,0 %
	M	24	100,0 %	0	0,0 %	24	100,0 %	0	0,0 %
	N	6	100,0 %	0	0,0 %	6	100,0 %	0	0,0 %
	Q	1	100,0 %	0	0,0 %	1	100,0 %	0	0,0 %
	R	108	80,0 %	27	20,0 %	103	76,3 %	32	23,7 %
	S	114	100,0 %	0	0,0 %	114	100,0 %	0	0,0 %
	T	56	100,0 %	0	0,0 %	56	100,0 %	0	0,0 %
	W	12	100,0 %	0	0,0 %	12	100,0 %	0	0,0 %
X	1	100,0 %	0	0,0 %	1	100,0 %	0	0,0 %	

<b>Con hemograma</b>	Z	9	90,0 %	1	10,0 %	10	100,0 %	0	0,0 %
	Total	458	92,5 %	37	7,5 %	451	91,1 %	44	8,9 %
	Sin	347	98,3 %	6	1,7 %	345	97,7 %	8	2,3 %
	Con	111	78,2 %	31	21,8 %	106	74,6 %	36	25,4 %
<b>Con perfil de coagulación</b>	Total	458	92,5 %	37	7,5 %	451	91,1 %	44	8,9 %
	Sin	361	98,6 %	5	1,4 %	360	98,4 %	6	1,6 %
	Con	97	75,2 %	32	24,8 %	91	70,5 %	38	29,5 %
<b>Con mediciones bioquímicas</b>	Total	458	92,5 %	37	7,5 %	451	91,1 %	44	8,9 %
	Sin	349	98,3 %	6	1,7 %	348	98,0 %	7	2,0 %
	Con	109	77,9 %	31	22,1 %	103	73,6 %	37	26,4 %
<b>Con examen de orina</b>	Total	458	92,5 %	37	7,5 %	451	91,1 %	44	8,9 %
	Sin	376	96,9 %	12	3,1 %	373	96,1 %	15	3,9 %
	Con	82	76,6 %	25	23,4 %	78	72,9 %	29	27,1 %
<b>Con imágenes</b>	Total	458	92,5 %	37	7,5 %	451	91,1 %	44	8,9 %
	Sin	379	92,0 %	33	8,0 %	372	90,3 %	40	9,7 %
	Con	79	95,2 %	4	4,8 %	79	95,2 %	4	4,8 %
<b>Con otras mediciones</b>	Total	458	92,5 %	37	7,5 %	451	91,1 %	44	8,9 %
	Sin	382	97,7 %	9	2,3 %	378	96,7 %	13	3,3 %
	Con	76	73,1 %	28	26,9 %	73	70,2 %	31	29,8 %

Anexo L

Tabla N.º 10. Tabla de contingencia por sexo de los pacientes atendidos en el tópico de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022

		Respuesta ChatGPT-4		
		Sin apendicitis	Con apendicitis	
<b>Sexo</b>	Femenino	Sin apendicitis	256	6
		Con apendicitis	3	17
	Masculino	Sin apendicitis	191	5
		Con apendicitis	1	16

## Anexo M

**Gráfico N.º 2. Distribución de edades según anatomopatología y respuesta de ChatGPT-4 para la apendicitis aguda en el tópico de cirugía del servicio de emergencia del Hospital Ramiro Prialé Prialé, Huancayo 2022**

