



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



**Instituto Tecnológico de Orizaba División de
Estudio de Posgrado e Investigación
Maestría en Sistemas Computacionales**

PROYECTO DE TESIS

TÍTULO DEL PROYECTO: Aplicación de técnicas de Visión por Computadora y Deep Learning para la realización de un análisis comparativo utilizando como caso de estudio la identificación de instrumental quirúrgico

PRESENTADO POR:

I.S.C. IVONNE JUÁREZ LÓPEZ M09011273

PARA OBTENER EL GRADO DE:

Maestro en Sistemas Computacionales

DIRECTOR DEL PROYECTO DE TESIS:

Dr. José Luis Sánchez Cervantes

CODIRECTOR DEL PROYECTO DE TESIS:

Dr. Luis Rolando Guarneros Nolasco

Orizaba, Veracruz, 07/marzo/2023
Dependencia: División de Estudios de
Posgrado e Investigación
Asunto: Autorización de Impresión
OPCION: I

**C. IVONNE JUÁREZ LÓPEZ
CANDIDATO A GRADO DE MAESTRO EN:
SISTEMAS COMPUTACIONALES
P R E S E N T E.-**

De acuerdo con el Reglamento de Titulación vigente de los Centros de Enseñanza Técnica Superior, dependiente de la Dirección General de Institutos Tecnológicos de la Secretaría de Educación Pública y habiendo cumplido con todas las indicaciones que la Comisión Revisora le hizo respecto a su Trabajo Profesional titulado:

" Aplicación de técnicas de Visión por computadora y Deep Learning para la realización de un análisis comparativo utilizando como caso de estudio, la identificación de instrumental quirúrgico."

comunico a Usted que este Departamento concede su autorización para que proceda a la impresión del mismo.

ATENTAMENTE

Experiencia en Educación Tecnológica
CIENCIA - TÉCNICA - CULTURA

Cuahtemoc Sánchez R.

**DR. CUAUHTÉMOC SÁNCHEZ RAMÍREZ
JEFE DE LA DIVISIÓN DE ESTUDIOS
DE POSGRADO E INVESTIGACIÓN**



OG-13-F06

CARTA DE ORIGINALIDAD

En la ciudad de Orizaba, Veracruz, el día lunes 20 de marzo del año 2023, la que suscribe **Ivonne Juárez López**, alumna del programa de Maestría en Sistemas Computacionales con número de control **M09011273**, manifiesta que es autora del trabajo de tesis titulado **"Aplicación de técnicas de Visión por Computadora y Deep Learning para la realización de un análisis comparativo utilizando como caso de estudio la identificación de instrumental quirúrgico"** y declara que el trabajo es original ya que sus contenidos son producto de mi directa contribución intelectual. Todos los datos y las referencias a materiales ya publicados están debidamente identificados con su respectivo crédito e incluidos en las notas bibliográficas y en las citas que se destacan como tal y, en los casos que así lo requieran, cuento con las debidas autorizaciones de quienes poseen los derechos patrimoniales. Por lo tanto, me hago responsable de cualquier litigio o reclamación relacionada con derechos de propiedad intelectual, exonerando de toda responsabilidad al Tecnológico Nacional de México / Instituto Tecnológico de Orizaba.

ATENTAMENTE

Excelencia en Educación Tecnológica
GENCIA - TÉCNICA - CULTURAL



IVONNE JUÁREZ LÓPEZ
AUTOR DEL TRABAJO DE TESIS





CARTA DE CESIÓN DE DERECHOS

En la ciudad de Orizaba, Veracruz, el día lunes 20 de marzo del año 2023, la que suscribe **Ivonne Juárez López**, alumna del programa de Maestría en Sistemas Computacionales con número de control **M09011273**, **manifiesta que es autora** del trabajo de tesis bajo la dirección del Dr. José Luis Sánchez Cervantes y cede los derechos del trabajo de tesis titulado **“Aplicación de técnicas de Visión por Computadora y Deep Learning para la realización de un análisis comparativo utilizando como caso de estudio la identificación de instrumental quirúrgico”** al Tecnológico Nacional de México / Instituto Tecnológico de Orizaba para su difusión y divulgación, con fines académicos y de investigación.

Queda estrictamente prohibido reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del Tecnológico Nacional de México / Instituto Tecnológico de Orizaba. Este puede obtenerse escribiendo a la siguiente dirección: msc@orizaba.tecnm.mx. Si el permiso se otorga, cualquier usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

ATENTAMENTE

Excelencia en Educación Tecnológica
CIENCIA – TÉCNICA – CULTURA

IVONNE JUÁREZ LÓPEZ
AUTOR DEL TRABAJO DE TESIS



Agradecimientos

En primer lugar agradezco a Dios por brindarme sabiduría y permitirme cerrar este capítulo de mi vida. Agradezco a mi madre, que es la mujer más ejemplar que he conocido, gracias madre sin ti no sería la mujer que soy ahora, sin ti sería nada, gracias por el amor que me das, por cada abrazo, por cada beso, por cada palabra de aliento, por siempre decirme lo orgullosa que estas de mí, por todo el sacrificio que has hecho y por el esfuerzo realizado para que nunca me faltara nada. Sin tu apoyo y sin tus consejos brindados no hubiera podido concluir con éxito la maestría, por todo eso gracias, también agradezco a mis pequeños hijos Dominic y Luz Renata que a pesar de su corta edad me motivan y aplauden cada uno de mis logros. Gracias hijos porque ustedes son lo más valioso que tengo en mi vida y mi mayor motivación para seguir luchando por mis sueños.

De igual manera agradezco a mi hermana Itzel y a mi cuñado Pablo por todo el apoyo brindado, pero principalmente por estar siempre para mí y darme tanto amor. Gracias a mi hermana Karina por las palabras de admiración y aliento brindado. Gracias a todos mis sobrinos por todo el amor que me dan.

También doy gracias a mi director de tesis, el Dr. José Luis Sánchez Cervantes por todos los consejos brindados, por ser un gran guía y por compartir sus conocimientos, pero principalmente porque se comportó como un gran amigo cuando estaba pasando por un momento difícil y doloroso en mi vida, gracias por escucharme, dejarme llorar y tener siempre las palabras correctas para motivarme y así logrará culminar la maestría. Gracias porque hoy soy una Ivonne más segura y todo se lo debo a usted que hizo que me diera cuenta de que puedo lograr todo lo que me propongo. Siempre tendrá mi admiración y respeto. También agradezco a la M.C.E Beatriz Alejandra Olivares Zepahua porque sin saberlo contribuyo con cada una de sus palabras para que me volviera una mujer más segura.

Gracias a mi amiga Yessica y a mi amigo José Antonio por su amistad brindada y por insistir tanto en que ingresaré a la maestría.

Me quedo tan satisfecha con los resultados obtenidos tras la elaboración de este proyecto, el cual fue un gran desafío ya que contaba con nulo conocimiento en el tema de inteligencia artificial. Hoy estoy preparada para los retos que lleguen a mi vida.

Finalmente, doy gracias al Consejo Nacional de Ciencia y Tecnología (CONACYT), por el apoyo económico brindado durante el periodo que duro la maestría (febrero 2021 – febrero 2023).

Tabla de contenido

Índice de Tablas.....	x
Índice de Figuras	xi
Resumen.....	xii
Introducción	1
Capítulo 1. Antecedentes	2
1.1 Marco Teórico	2
1.1.1 Intervención Quirúrgica	2
1.1.2 Repositorio digital de imágenes	5
1.1.3 Inteligencia Artificial.....	6
1.1.4 Visión por Computadora	6
1.1.5 Aprendizaje profundo	16
1.2 Planteamiento del Problema	23
1.3 Objetivo General y Objetivos Específico	24
1.3.1 Objetivo General	24
1.3.2 Objetivos específicos.....	24
1.4 Justificación	25
Capítulo 2. Estado de la práctica	27
2.1 Trabajos relacionados	27
2.2 Análisis Comparativo.....	36
Capítulo 3. Aplicación de la Metodología	43
3.1 Planeación	43
3.2 Diseño.....	45
3.2.1 Arquitectura del sistema	46
3.2.2 Arquitectura de la red neuronal y convolucional.....	48
3.2.3 Análisis de requerimientos.....	49
3.2.4 Nombre de la aplicación.....	59
3.2.5 Diseño de la aplicación	59
3.3 Desarrollo	64
3.3.1 Entrenamiento del modelo	65
3.3.2 Desarrollo de interfaces gráficas.....	71
3.3.3 Creación de la base de datos.....	81

3.3.4	Desarrollo de APIs	82
3.3.5	Documentación	85
3.4	Pruebas.....	85
3.4.1	Inicio de la página web.....	86
3.4.2	Iniciar detección de objetos	86
Capítulo 4.	Resultados.....	91
4.1	Caso de estudio	91
4.2	Evaluación y resultados.....	95
Capítulo 5.	Conclusiones y recomendaciones.....	97
5.1	Conclusiones.....	97
5.2	Recomendaciones	98
Productos académicos	99
6.1	Artículo de congreso	99
6.2	Retribución social	100
6.3	Registro de derecho de autor.....	101
Anexos.....		101
Referencias bibliográficas	102

Índice de Tablas

Tabla 1.1 Comparativa de herramientas más populares para visión por computadora.	9
Tabla 1.2 APIs utilizadas en el aprendizaje profundo.	21
Tabla 2.1 Análisis comparativo de los artículos relacionados.	36
Tabla 3.1 Historia de usuario selección tipo cirugía.	43
Tabla 3.2 Historia de usuario registra participantes cirugía.	43
Tabla 3.3 Historia de usuario identificación de instrumental quirúrgico.	44
Tabla 3.4 Historia de usuario consultar cirugías realizadas.	44
Tabla 3.5 Historia de usuario generar reporte.	45
Tabla 3.6 Requerimientos funcionales.....	49
Tabla 3.7 Requerimientos no funcionales.	50
Tabla 3.8 Narrativa de casos de uso registrar cirugía.	52
Tabla 3.9 Narrativa de curso de excepción valida datos de entrada.	52
Tabla 3.10 Narrativa de casos de uso cargar imagen.	53
Tabla 3.11 Curso alternativo “Identificar instrumental quirúrgico”.	53
Tabla 3.12 Narrativa de casos de uso consultar historial cirugías.	54
Tabla 3.13 Curso alternativo “Generar reporte”.	55
Tabla 3.14 Diccionario de datos tabla tipo cirugía.	56
Tabla 3.15 Diccionario de datos tabla cirugía.	56
Tabla 3.16 Diccionario de datos tabla Tipo de usuario.	57
Tabla 3.17 Diccionario de datos tabla usuario.	57
Tabla 3.18 Diccionario de datos tabla descripciones.	58
Tabla 4.1 Resultados obtenidos de la identificación de objetos.	94
Tabla 4.2 Evaluación de usabilidad de la aplicación.	95

Índice de Figuras

Figura 1.1 <i>Fases del proceso para la interpretación de imágenes de la visión por computadora.</i>	6
Figura 1.3 <i>Fórmula de IOU.</i>	15
Figura 1.4 <i>Algoritmos utilizados en aprendizaje profundo</i>	17
Figura 1.5 <i>Imagen que muestra como visualiza una imagen una computadora.</i>	18
Figura 1.6 <i>Esquema de la detección de objetos mediante Fast R-CNN</i>	19
Figura 1.7 <i>Esquema de la detección de objetos mediante Faster R-CNN</i>	20
Figura 1.8 <i>Esquema de la detección de objetos.</i>	23
Figura 3.2 <i>Arquitectura de la red neuronal convolucional.</i>	49
Figura 3.3 <i>Diagrama de casos de uso.</i>	51
Figura 3.4 <i>Diagrama entidad-relación.</i>	55
Figura 3.5 <i>Diagrama de clases.</i>	59
Figura 3.6 <i>Pantalla de inicio, selección de cirugía.</i>	60
Figura 3.8 <i>Pantalla para la selección de imagen.</i>	62
Figura 3.9 <i>Pantalla que muestra la detección del instrumental quirúrgico.</i>	63
Figura 3.10 <i>Pantalla que despliega todas las cirugías realizadas.</i>	64
Figura 3.11 <i>Instrumental quirúrgico utilizado para una cirugía cesárea.</i>	65
Figura 3.12 <i>Instrumental quirúrgico utilizado para una cirugía parto natural.</i>	66
Figura 3.13 <i>Etiquetado de imágenes desde el programa labelImg.</i>	66
Figura 3.14 <i>Línea de comando para la ejecución del entrenamiento del modelo.</i>	68
Figura 3.15 <i>Métricas obtenidas de los entrenamientos realizados.</i>	69
Figura 3.16 <i>Resultados obtenidos en el antepenúltimo entrenamiento.</i>	70
Figura 3.17 <i>Resultados obtenidos en el penúltimo entrenamiento.</i>	70
Figura 3.20 <i>Prueba que muestra todos los tipos de cirugía.</i>	85
Figura 3.22 <i>Interfaz gráfica de la pantalla principal.</i>	87
Figura 3.23 <i>Interfaz gráfica de editar imagen.</i>	88
Figura 3.25 <i>Interfaz gráfica que muestra el la identificación del instrumental quirúrgico.</i>	89
Figura 3.27 <i>Interfaz gráfica que muestra mensaje de error.</i>	90
Figura 4.1 <i>Instrumental quirúrgico utilizado para cirugía por cesárea.</i>	92
Figura 4.2 <i>Instrumental quirúrgico utilizado pata parto natural.</i>	92
Figura 4.5 <i>Identificación de objetos cirugía por parto natural</i>	94

Resumen

En México existe déficit en la proporción de enfermeras-pacientes de acuerdo a los rangos establecidos por la Organización Mundial de la Salud (OMS) y la Organización Panamericana de la Salud (OPS). Por lo anterior, las comunidades rurales son las más afectadas ya que solo algunas cuentan con servicios de salud y estos únicamente cuentan con un médico pero no con enfermeras, lo que provoca que no se brinden servicios de calidad y en caso de ser necesario realizar algún procedimiento quirúrgico (cirugía parto por cesárea o cirugía parto natural) al médico se le complica realizarlo, ya que no hay quien le proporcione el material quirúrgico para realizar ese tipo de procedimientos.

Por lo anterior se propone desarrollar una aplicación que ayude a la detección de instrumental quirúrgico. Esta aplicación se desarrollará mediante técnicas de visión por computadora y *Deep Learning*. Dicha aplicación obtendrá una imagen en tiempo real a partir de una cámara web, esta será la encargada de transmitir la imagen al módulo de detección de objetos en donde se aplicarán técnicas de visión por computadora para identificar todos los objetos de la imagen; una vez detectados, el módulo de extracción de regiones propuestas será el encargado de detectar todos los objetos que mandó el módulo anterior sin importar la clasificación de estos; una vez detectados, el módulo de detección de agrupamiento será el que analice si existen objetos repetidos, si es así procederá a agruparlos, para posteriormente pasar los objetos al módulo de entrenamiento y detección de características CNN. Este último módulo será el encargado de comparar el objeto recibido mediante una red neuronal y es así como se obtendrá la clasificación del objeto.

El principal beneficio del desarrollo de la aplicación mencionada con anterioridad, es que servirá para capacitar tanto a estudiantes como a parteras, estas ingresarán a una cirugía (cesárea o parto natural) y verán en la pantalla el nombre del material y la descripción de este, esto ayudará a que asistan al médico sin necesidad de tener conocimientos con respecto al instrumental quirúrgico utilizado para la intervención.

Introducción

El aprendizaje profundo y la visión por computadora ganaron gran popularidad en los últimos años, gracias a todas las áreas que abarca (medicina, educación, finanzas, seguridad, entre otros) y los resultados favorables obtenidos.

Por lo tanto, se apuesta por generar proyectos innovadores que ayuden a la sociedad aplicando todas las técnicas que abarcan la visión por computadora y el aprendizaje profundo.

Para contribuir a contrarrestar el déficit de enfermeras instrumentistas que existe en México es idónea la creación de una aplicación que sirva para capacitar tanto a estudiantes como a parteras ya que estas últimas cuentan con conocimientos necesarios para ayudar con el alumbramiento de un bebé pero solo por medio del parto natural, por lo tanto la aplicación ayudará a que detecten de manera rápida el nombre y descripción del instrumental quirúrgico utilizado para la realización de intervenciones quirúrgicas, ya sea por parto natural o por parto por cesárea.

Por lo anteriormente descrito, en este proyecto se propone la creación de una aplicación que sirva para la detección de instrumental quirúrgico, la cual se creará mediante técnicas de visión por computadora y aprendizaje profundo.

Este trabajo se encuentra dividido en cinco capítulos, en el primer capítulo se realiza la presentación del marco teórico en donde se describen los conceptos que son importantes para que el lector conozca, también se da a conocer el planteamiento del problema, el objetivo general, objetivos específicos y la justificación. En el segundo capítulo se presenta el estado de la práctica y el análisis comparativo de estos.

Mientras que en el capítulo número 3 se presenta el desarrollo de la aplicación, este se basa en la aplicación de la metodología seleccionada para el desarrollo del proyecto de tesis, el cual inicia con la planeación y finalizó con las pruebas de la aplicación. Por su parte, en el cuarto capítulo se presentan los resultados obtenidos tras realizar las pruebas de la aplicación de desarrollada con un caso real.

Finalmente, se presentan las conclusiones y referencias del trabajo.

Capítulo 1. Antecedentes

En este capítulo se presentan conceptos claves que son relevantes para el desarrollo de este proyecto de tesis. Cada palabra clave se describe de forma concreta y puntual para que el lector tenga una mejor comprensión del tema de investigación.

1.1 Marco Teórico

A continuación, se describen los conceptos fundamentales relacionados con el tema de investigación.

1.1.1 Intervención Quirúrgica

Se define como cualquier manipulación individual, separada y sistemática, sobre o dentro del cuerpo. Es realizada por un médico u otro profesional de la salud titulado, con o sin instrumentos, se realiza para restaurar partes del cuerpo, extirpar tejidos enfermos o lesionados, extraer cuerpos extraños, asistir en partos o facilitar el diagnóstico. Dichas intervenciones se realizan en quirófanos, salas de parto o salas de procedimientos, además requieren la administración de anestesia local o general [1].

1.1.1.1 Parto por Cesárea

La operación por cesárea es una intervención quirúrgica que consiste en la extracción del feto vía abdominal, el procedimiento se realiza mediante una incisión en el útero. En sus inicios esta operación era indicada solo cuando el parto por vía vaginal era imposible. Actualmente este tipo de operación se realiza incluso si la paciente lo solicita. [2]

En [3], el material quirúrgico que se utiliza en este procedimiento se lista a continuación:

- Mango de bisturí: Instrumento de acero inoxidable para acoplar una hoja de bisturí, usado para realizar incisiones en todo tipo de cirugías de forma segura.
- Tijera de mayo recta: Instrumento quirúrgico de acero inoxidable utilizada para cortar materiales (hilos, gasas, entre otros), seccionar o separar tejidos resistentes, músculos

Capítulo 1. Antecedentes

o hasta partes de algún órgano. Estas tijeras están diseñadas para cortar tejidos corporales que se encuentren cerca de la superficie de la herida.

- Tijera de mayo curva: Instrumento quirúrgico de acero inoxidable, estas tijeras están diseñadas para realizar una penetración más profunda en la herida. Se usa para cortar tejidos gruesos como los que se ubican en los músculos, útero, pecho y pie.
- Tijera *Metzenbaum* Curva: Instrumento quirúrgico de acero inoxidable utilizada para cortar tejidos delicados y finos o aislar vasos o conductos.
- Pinzas *Kelly* Curvas: Instrumento de acero inoxidable utilizada para realizar hemostasia en vasos sanguíneos.
- Pinza de disección con dientes: Instrumento de acero inoxidable, utilizado para sostener el tejido (liso) o ejercer presión en un vaso sanguíneo durante un procedimiento quirúrgico.
- Pinza de disección sin dientes: Instrumento de acero inoxidable, permite sostener tejidos sin necesidad de desgarrar (tejido delicado).
- Pinzas *Allis*: Instrumento quirúrgico con dientes afilados fabricado en acero inoxidable, cuya función es sostener tejido grueso u órganos intraabdominales.
- Pinza de anillo (*Forester*) curva: Se caracteriza por tener forma de anillo en las puntas. Es fabricada en acero inoxidable y se utiliza para sostener un tejido y otros materiales que son peligrosos si son tocados con las manos.
- Pinzas Rochester (Recta y Curva): Instrumento de acero inoxidable, cuya función es traccionar, sujetar y realizar hemostasia en vasos sanguíneos de mayor tamaño.
- Valvas suprapúbicas: Instrumento en forma de pala, se utilizan para separar la pared abdominal hacia arriba, abajo o a los lados.
- Separadores *Farabeuf*: Instrumento utilizado para separar las paredes de una cavidad o labios de una herida.
- Perilla: Instrumento fabricado de goma utilizado para extraer la mucosidad de la nariz o de la boca.

1.1.1.2 Parto Natural

La Organización Mundial de la Salud (OMS) define al parto natural como aquel que “Comienza espontáneamente, bajo riesgo al inicio del parto manteniéndose como tal hasta el

Capítulo 1. Antecedentes

alumbramiento. El niño nace espontáneamente en posición cefálica entre las semanas 37 a 42”.

Aunque la mayoría de las madres desearían tener un parto natural, no todas las circunstancias y situaciones son idóneas para llevarlo a cabo.

El médico encargado de la atención del parto es que el decide si el parto natural es viable o no. Para tomar dicha decisión se basa en 4 requisitos que tiene que cumplir la paciente:

- **Gestación sin factores de riesgo:** Es toda característica o circunstancia que pertenece y asocian a un grupo de personas con la probabilidad de estar especialmente expuesta a desarrollar o padecer un proceso mórbido, los factores de riesgo son universales (edad), sociales, ambientales y biológicos, o asociarse varios de ellos.
- **Trabajo de parto de forma espontánea:** Esta situación se da cuando las contracciones comienzan por si solas, o cuando la bolsa amniótica se rompe y la ruptura de esta estimula las contracciones.
- **Embarazo a término:** Si el embarazo alcanza entre 37 y 42 semanas de gestación se considera un embarazo a término.
- **Presentación fetal:** Se refiere a la posición en la que se encuentra el bebé en el canal de parto. Para que el feto al momento del nacimiento no sufra lesiones y el parto se presente de forma natural, el feto tiene que encontrarse en posición cefálica, cuando el feto está en esta posición se encuentra colocado boca abajo y mirando a la espalda de la madre [4].

Aunque se espera que el bebé logre nacer de manera espontánea los médicos suelen ayudar a la paciente en el momento de su alumbramiento, para eso ocupan algunos instrumentos quirúrgicos [5]:

- **Perilla:** Instrumento fabricado de goma utilizado para extraer la mucosidad de la nariz o de la boca.
- **Porta agujas:** Instrumento de acero inoxidable, cuya función es sujetar la aguja de sutura y así realizar los puntos de sutura en el paciente de forma más fácil, segura y cómoda.

Capítulo 1. Antecedentes

- Pinza de disección con dientes: Instrumento de acero inoxidable, utilizado para sostener el tejido (liso) o ejercer presión en un vaso sanguíneo durante un procedimiento quirúrgico.
- Pinza de disección sin dientes: Instrumento de acero inoxidable, permite sostener tejidos sin necesidad de desgarrar (tejido delicado).
- Pinza de anillo (*Forester*) curva: Se caracteriza por tener forma de anillo en las puntas. Es fabricada en acero inoxidable y se utiliza para sostener un tejido y otros materiales que son peligrosos si son tocados con las manos.
- Pinzas Rochester Curva: Instrumento de acero inoxidable, cuya función es traccionar, sujetar y realizar hemostasia en vasos sanguíneos de mayor tamaño.
- Pinza Campo: Se caracterizan por tener puntas puntiagudas y curvas, su principal función es sujetar el paño de campo a la piel del paciente.
- Tijera de mayo recta: Instrumento quirúrgico de acero inoxidable utilizada para cortar materiales (hilos, gasas, entre otros), seccionar o separar tejidos resistentes, músculos o hasta partes de algún órgano. Estas tijeras están diseñadas para cortar tejidos corporales que se encuentren cerca de la superficie de la herida.
- Tijera de mayo curva: Instrumento quirúrgico de acero inoxidable, estas tijeras están diseñadas para realizar una penetración más profunda en la herida. Se usa para cortar tejidos gruesos como los que se ubican en los músculos, útero, pecho y pie.

1.1.2 Repositorio digital de imágenes

Gracias a los cambios que el internet produjo en las modalidades de intercambio de información, surgen los repositorios digitales como un medio eficaz en línea con una gran diversidad de contenidos (imágenes, video, documentos).

Ernest Abadal, definió un repositorio como una plataforma que “recoge, preserva y difunde la producción académica de una institución y/o de una disciplina científica, permitiendo el acceso a los objetos digitales que contiene y a sus metadatos”[6].

Entre las ventajas que ofrecen los repositorios digitales se tiene que son de acceso libre y gratuito, así como la posibilidad de autoarchivo [7]. Existe una gran cantidad de repositorios

digitales de imágenes enfocados en diferentes áreas, en el área de medicina se tiene a los siguientes repositorios digitales en donde se almacenan imágenes de instrumental quirúrgico.

- Equipo Quirúrgico – Imagen de Stock¹
- Imágenes de instrumental quirúrgico²
- *Depositphotos* – Stock de material quirúrgico³

1.1.3 Inteligencia Artificial

La AI (*Artificial Intelligence*, Inteligencia Artificial) se define como “la habilidad de las computadoras para hacer actividades que normalmente requieren inteligencia humana “. En términos más detallados la AI se define como capacidad de las máquinas para usar algoritmos, aprender de los datos y utilizar lo aprendido en la toma de las decisiones tal y como lo haría un ser humano [8].

1.1.4 Visión por Computadora

Una de las áreas más importantes de la AI es la visión por computadora o también conocida como visión artificial, se define como la ciencia de programar a una computadora para que procese imágenes o videos e incluso los entienda [9]. En la Figura 1.1 se muestran los pasos del proceso que realiza la visión por computadora para lograr la interpretación de una imagen.

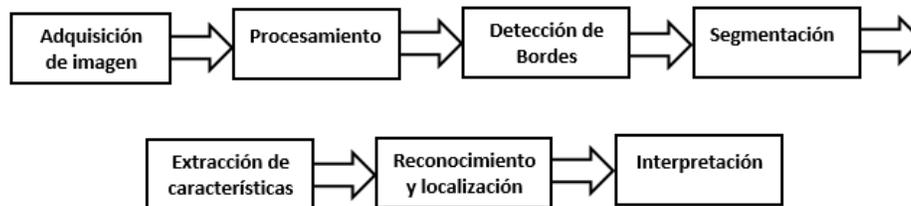


Figura 1.1 Fases del proceso para la interpretación de imágenes de la visión por computadora.

¹ <https://www.istockphoto.com/es/search/2/image?phrase=equipo%20quir%C3%BArgico>

² <https://www.shutterstock.com/es/search/instrumental+quir%C3%BArgico>

³ <https://sp.depositphotos.com/stock-photos/material-quirirgico.html?filter=all>

1.1.4.1 Técnicas de Visión por Computadora

En [10] las técnicas de visión por computadora se clasifican en 6 tipos:

- Clasificación de imágenes: Es la técnica de visión por computadora más conocida, su objetivo es clasificar el contenido de la imagen según su contenido.
- Detección de objetos: Esta técnica se encarga de encontrar todos los objetos que se encuentren en una imagen, una vez detectados los etiqueta y genera en cada uno de ellos cuadros delimitadores.
- Seguimiento de objetos: Es utilizada para el seguimiento de uno o más objetos en movimiento en cualquier escena dada.
- Segmentación semántica: Es la encargada de dividir toda la imagen en grupos de píxeles para que posteriormente sean etiquetados y clasificados los objetos. Sirve para detectar todos los límites de los objetos.
- Segmentación de instancias: Su función es clasificar qué es exactamente la imagen, esto lo logra categorizando todas las distintas clases de instancia.
- Reconstrucción de imágenes: Con la ayuda de conjuntos de datos de fotos actuales esta técnica logra reconstruir imágenes.

1.1.4.2 Algoritmos utilizados

Para lograr que el procesamiento de la imagen sea más sencillo existen una serie de herramientas, que trabajan con los algoritmos más simples hasta los más complejos. A continuación, se listan los algoritmos más utilizados en visión por computadora.

- SIFT (*Scale Invariant Feature Transform*, Transformación de características invariante de la escala): Se propuso para imágenes en escala de grises, para obtener un conjunto de descriptores de una imagen, es necesario obtener los puntos característicos y posteriormente, para cada punto de interés encontrado se procede a calcular su vector descriptor a partir de la información de los píxeles que lo rodean. Este algoritmo consta de dos partes: (1) Obtención de los puntos característicos (2) Descripción de la región alrededor de cada punto de interés [11].
- OpenCV (*Open Computer Vision*, Visión por Computadora Abierta): Biblioteca de funciones utilizada en visión por computadora y aprendizaje automático. Su

Capítulo 1. Antecedentes

- función principal es la extracción y el proceso de datos relevantes de las imágenes. OpenCV incluye varias herramientas que ayudan a resolver problemas de visión por computadora. Contiene funciones de procesamiento de imágenes de bajo y alto nivel, a su vez contiene varios algoritmos de alto nivel que sirven para detectar rostros, comparar rasgos y seguimiento [12].
- YOLO (*You Only Look Once*, Solo se Mira una Vez): Es un algoritmo que implementa redes neuronales convolucionales (CNN) para detectar y reconocer varios objetos a la vez en tiempo real. Como su nombre lo indica, solo requiere una propagación hacia adelante a través de una red neuronal para detectar los objetos.

Esto significa que el algoritmo solo se ejecuta una vez. La CNN se utiliza para predecir varias probabilidades de clase y cuadros delimitadores simultáneamente [13].

Este algoritmo funciona utilizando las técnicas siguientes:

- Bloques residuales
- Regresión de cuadro delimitador
- Intersección sobre unión (IOU)

1.1.4.3 Interfaces Gráficas de Programación para la aplicación de Técnicas de Visión por Computadora

Existen un gran número de herramientas que se utilizan para la creación de aplicaciones en las diferentes técnicas de visión por computadora, en la tabla 1.1 se presentan las herramientas más populares [14]:

Tabla 1.1 Comparativa de herramientas más populares para visión por computadora.

Nombre	Características	Interfaces	CNN	RNN	DMB/ RBM	Ejecución paralela multinodo	Desarrollador	Licencia
Caffe	Velocidad, estructura modular, esquema de texto plano para el modelado y la optimización, almacenamiento de datos y comunicación mediante blob	C, C++, command line interface, Python, and MATLAB	Sí	Sí		Sí	Berkeley Vision & Learning Center	BSD
Microsoft Cognitive Toolkit CNTK	Manejo de datos densos multidimensionales, ajuste automático de hiperparámetros, normalización por lotes	Python, C++, C#, and command line interface	Sí	Sí		Sí	Microsoft Research	MIT
TensorFlow	Cálculos matemáticos mediante el gráfico de flujo de datos, inicio, clasificación de imágenes, autodiferenciación.	C++, Python, Java, Go	Sí	Sí	Sí	Sí	Google Brain team	Apache 2.0
Theano	Compilación de expresiones matemáticas mediante matrices multidimensionales	Python	Sí	Sí	Sí	Sí	Universidad de Montreal	BSD
Torch	Soporte de matrices N-dimensionales, diferenciación automática de gradientes, soporte para modelos neurales y modelos de energía.	C, C++, Lua	Sí	Sí	Sí	Sí	R. Collobert, K. Kavukcuoglu, C. Faret	BSD
Chainer	Esquema de definición de red, paralelización multi-GPU, cálculo hacia delante/hacia atrás, arquitectura por lotes.	Python	Sí	Sí		Sí	Preferred Networks Inc.	MIT
Keras	Creación rápida de prototipos, módulos modulares y minimalistas, extensibles, esquemas de conexión arbitrarios.	Python	Sí	Sí	Sí	Sí	F. Chollet	MIT

Capítulo 1. Antecedentes

Nombre	Características	Interfaces	CNN	RNN	DMB/ RBM	Ejecución paralela multinodo	Desarrollador	Licencia
Deeplearning4j	Marco de aprendizaje profundo distribuido, arquitectura de microservicios.	Python, Java, Scala	Sí	Sí	Sí	Sí	SkyMind engineering team	Apache 2.0
Apache Singa	Plataforma de formación distribuida escalable	Python, Java, C++	Sí	Sí	Sí	Sí	Apache Incubator	Apache 2.0
MXnet	Mezcla de programación simbólica y programación imperativa, portabilidad, autodiferenciación	Python, R, C++, Julia	Sí	Sí		Sí	Distributed (Deep) Machine Learning Community	Apache 2.0
Neon	El rendimiento más rápido en varios hardware y redes profundas (GoogLeNet, VGG, AlexNet, Generative Adversarial Networks)	Python	Sí	Sí	Sí	Sí	Intel	Apache 2.0

1.1.4.4 Etiquetado de imagen

El etiquetado de datos es el proceso de anotar manualmente el contenido con etiquetas. En visión por computadora la etiqueta identifica a los elementos dentro de la imagen, una vez identificados los elementos etiquetados se utilizan en el aprendizaje supervisado. El conjunto de datos etiquetados se utiliza para enseñar al modelo con el ejemplo.

Existen diferentes tipos de etiquetado de datos, para la detección de objetos por ejemplo, para conocer la clase del objeto y sus ubicaciones, se dibujan cuadros delimitadores alrededor de los objetos destino en las imágenes [15].

Capítulo 1. Antecedentes

Para realizar el etiquetado de imágenes existen un amplio conjunto de herramientas, a continuación, se listan algunas:

1. MakeSense.AI: Herramienta de uso gratuito bajo licencia GPLv3, no requiere instalación ya que se usa desde el navegador web, realizado etiquetado de líneas, puntos y polígonos, los formatos de archivos de salida que admite son YOLO, VOC XML, VGG Json y CSV.
2. LabelImg: Escrita en Python y utiliza como interfaz gráfica Qt. Las anotaciones que se realizan en las imágenes se guardan como archivos XML en formato PASCAL VOC (formato utilizado por ImagiNet), también es compatible con formatos YOLO y CreateML.
3. Anotador de imágenes VGG: Aplicación para etiquetado manual simple e independiente para imagen, audio y vídeo.
4. LabelMe: Herramienta gráfica de etiquetado de imágenes, escrita en Python y como interfaz gráfica utiliza Qt. Las anotaciones de las imágenes las realiza con polígono, rectángulo, círculo, línea y punto; los formatos aceptados son VOC y COCO.
5. Escalable: Herramienta web de código abierto, admite anotaciones de imagen y vídeo. Cuadros delimitadores, segmentación semántica, seguimiento de vídeo, área manejable, señalización de carril y cajas de rebote 3D en nube de puntos son los tipos de anotaciones admitidos.
6. RectLabel: Herramienta para etiquetar imágenes. Cuadros delimitadores, polígono, curva cúbica, línea, punto, entre otras son las técnicas que admite. Permite realizar lectura y escritura en formatos YOLO, Create ML, COCO Json y CSV.

1.1.4.5 Métricas para la identificación de objetos utilizando visión por computadora

Para evaluar qué tan eficaces y precisos son los detectores de objetos, existen diferentes métricas, estas realizan una variedad de cálculos. El aplicar métricas a diferentes algoritmos ayuda a seleccionar aquel que brinde mayor precisión en la detección de objetos. La Figura 1.2 muestra el proceso de análisis realizado con las diferentes métricas (Precisión, Exhaustividad, Área bajo la curva e IOU).

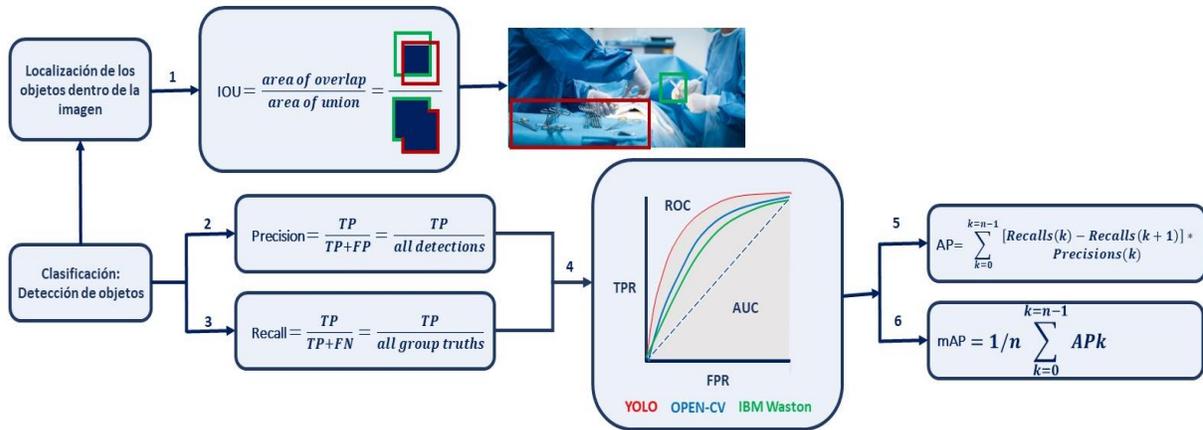


Figura 1.2 Tareas del proceso para analizar la precisión de la detección de objetos.

Proceso de análisis: En el análisis de algoritmos para detección de objetos mediante Visión por Computadora es necesario medir dos tareas distintas: 1) Clasificación, que consiste en determinar si los objetos existen en la imagen y, 2) Localización, es una tarea de regresión que implica determinar la ubicación de los objetos dentro de la imagen.

Clasificación de objetos: Requiere de las métricas Precisión, Exhaustividad, curva AUROC, AP y mAP que se describen brevemente a continuación:

- **Precisión (*Precision*):** La precisión es el número de campos detectados correctamente. Es decir, cuántos objetos se detectaron correctamente con respecto a la etiqueta de clase y el IOU (*Intersection Over Union*, Intersección sobre Unión) en la imagen. Por lo tanto, la precisión es la capacidad de un modelo para identificar sólo los objetos relevantes. Es el porcentaje de predicciones positivas correctas y se representa por:

$$Precisión = \frac{VP}{VP + FP} = \frac{VP}{Todas\ la\ detecciones}$$

- **Exhaustividad (*Recall*):** Se refiere al número de objetos presentes en la imagen que no se predijeron con respecto a la etiqueta de la clase y el IOU. Por lo tanto, la exhaustividad es la capacidad de un modelo para encontrar todos los casos relevantes (todas las cajas de delimitación de la verdad básica o *ground truths*). Es el porcentaje

Capítulo 1. Antecedentes

de verdaderos positivos detectados entre todas las verdades básicas relevantes y se representa por:

$$Exhaustividad = \frac{VP}{VP + FN} = \frac{VP}{Todas\ las\ verdades\ fundamentales}$$

Donde para ambas representaciones:

VP = Verdaderos Positivos

FP = Falsos Positivos

FN = Falsos Negativos

Las descripciones de las métricas de Precisión y Exhaustividad permiten identificar que, a mayor precisión, los objetos predichos por el modelo son correctos con respecto a la clasificación y su localización. Si se obtiene una precisión baja, los objetos de la imagen no se predicen correctamente.

Por otra parte, cuanto mayor es el valor de la Exhaustividad, se tendrá un alto porcentaje de predicción correcta de todos o la mayoría de los objetos de la imagen en lo que respecta a la clasificación y la localización. Por lo que si se obtiene un valor de Exhaustividad bajo, significa que todos o la mayoría de los objetos de la imagen no se predicen correctamente [16].

- **AP:** La precisión media es una forma de resumir la curva de recuperación de precisión en un solo valor que representa el promedio de todas las precisiones. El AP se calcula de acuerdo con la siguiente ecuación. Usando un bucle que pasa por todas las precisiones / exhaustividad, la diferencia entre los retiros actual y siguiente se calcula y luego se multiplica por la precisión actual. En otras palabras, el AP es la suma ponderada de precisiones en cada umbral donde el peso es el aumento en el recuerdo [17].

$$AP = \sum_{k=0}^{k=n-1} \frac{[Recalls(k) - Recalls(k + 1)] *}{Precisions(k)}$$

Donde para:

Capítulo 1. Antecedentes

Recalls (n) = 0

Precisions (n) = 1

n = Valor del umbral

- **mAP:** La precisión media corresponde al promedio de AP y se calcula con la siguiente fórmula.

$$mAP = 1/n \sum_{k=0}^{k=n-1} APk$$

Donde para:

APk = El AP para la clase K

n = Número de clases

- **Curva ROC:** Cuando se requiere comprobar o visualizar el rendimiento de problemas de clasificación multiclase, es necesario utilizar la curva AUC (*Area Under The Curve*, Área Bajo La Curva) ROC (*Receiver Operating Characteristics*, Características de Funcionamiento del Receptor). Es una de las métricas de evaluación más importantes para comprobar el rendimiento de cualquier modelo de clasificación. También se escribe como AUROC (*Area Under the Receiver Operating Characteristics*, Área Bajo las Características de Funcionamiento del Receptor). La curva ROC se traza con el TPR (*True Positive Rate*, Tasa de Verdaderos Positivos) frente al FPR (*False Positive Rate*, Tasa de Falsos Positivos), donde el TPR está en el eje Y y el FPR en el eje X [18].

Localización de objetos dentro de la imagen: Requiere de la métrica Intersección sobre la Unión (IOU), que es una medida basada en el índice de Jaccard que evalúa el solapamiento entre dos cuadros delimitadores, por lo que requiere de un cuadro delimitador del área de verdad sobre el terreno de la imagen. Aplicar el IOU permite saber si una detección es válida (Verdadero Positivo) o no (Falso Positivo). Por lo tanto, la métrica IOU es la relación de la intersección (Solapamiento) del cuadro delimitador de la predicción y el cuadro delimitador

Capítulo 1. Antecedentes

de verdad del terreno de la imagen entre la unión del cuadro delimitador de la predicción y las áreas de verdad del terreno de la imagen. Esto se representa mediante la fórmula siguiente:

$$IOU = \frac{\text{área}(B\rho \cap Bgt)}{\text{área}(B\rho \cup Bgt)}$$

Donde:

$B\rho$ = Cuadro delimitador de la predicción

Bgt = Cuadro delimitador del área de verdad del terreno de la imagen.

La Figura 1.3 muestra el IOU entre un cuadro delimitador de la verdad sobre el terreno (en verde) y un cuadro delimitador detectado (en rojo).

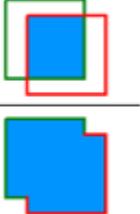
$$IOU = \frac{\text{área de superposición}}{\text{área de unión}} = \frac{\text{área de superposición}}{\text{área de unión}}$$


Figura 1.3 Fórmula de IOU.

Algunos conceptos básicos utilizados por las métricas:

- Verdadero positivo (VP): Una detección correcta. Detección con un umbral de IOU \geq .
- Falso Positivo (FP): Una detección errónea, Detección con IOU < umbral.
- Falso Negativo (FN): Una verdad de fondo no detectada, [si IOU con verdad de fondo >umbral, detección errónea]
- Verdadero negativo (VN): No se aplica. Representaría una detección errónea corregida. En la tarea de detección de objetos hay muchos cuadros delimitadores posibles que no deben ser detectados dentro de una imagen. Por lo tanto, VN sería todas las posibles cajas delimitadoras que no se detectaron correctamente (tantas cajas posibles dentro de una imagen). Por eso no se utiliza por las métricas.
- Umbral: dependiendo de la métrica, suele fijarse en el 50%, el 75% o el 95% [19].

1.1.4.6 Sobreajuste

Se denomina sobreajuste al hecho de hacer un modelo tan ajustado a los datos de entrenamiento que hace que no clasifique correctamente los datos en las imágenes de pruebas. Esto sucede cuando un modelo se entrena demasiado o con datos incorrectos y esto ocasiona que el modelo aprenda patrones que no son generales [20].

1.1.4.7 Subajuste

Este término se refiere a un modelo que no puede modelar los datos de entrenamiento por tal motivo no generaliza a nuevos datos. Esto se presenta cuando el modelo es entrenado con pocos datos, por lo tanto es imposible construir un modelo preciso [21].

1.1.5 Aprendizaje profundo

También conocido como *Deep Learning* es un subconjunto de la IA y del aprendizaje automático, surgió como una poderosa estrategia para aprender representaciones de características directamente a partir de datos. Esta técnica proporcionó grandes mejoras en la detección de objetos ya que utiliza redes neuronales artificiales.

El aprendizaje profundo permite que los modelos computacionales aprendan representaciones complejas, abstractas y sutiles, esta técnica se aplican a una amplia gama de problemas tales como reconocimiento visual, detección de objetos, reconocimiento de voz, procesamiento de lenguaje natural, análisis de imágenes médicas, entre otras [22].

1.1.5.1 Algoritmos Utilizados

Los algoritmos de aprendizaje profundo dependen de una red neuronal, ya que al igual que el cerebro calcula la información utilizando millones de neuronas. Los algoritmos en aprendizaje profundo se dividen en dos tipos: aprendizaje supervisado y aprendizaje no supervisado. En la Figura 1.2 se muestran los algoritmos pertenecientes a cada categoría [23].

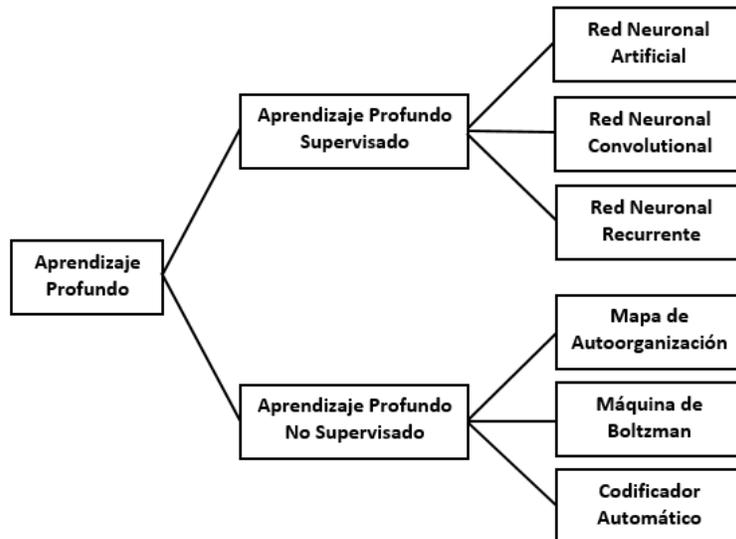


Figura 1.4 Algoritmos utilizados en aprendizaje profundo

1.1.5.2 Redes Neuronales Convolucionales

El algoritmo más utilizado en el aprendizaje profundo es la Red Neuronal Convolucional (CNN), es una clase de red neuronal artificial que es dominante en tareas de visión por computadora. Estas redes están diseñadas para imitar el comportamiento del cerebro humano; el propósito de estas se reduce a 3 aspectos: aprender de la experiencia, obtener información según las entradas que se proveen y generalizar lo ya aprendido en nuevos elementos [24] .

La CNN es una construcción matemática que normalmente se compone de 3 capas: convolución, agrupación y capas completamente conectadas. Las dos primeras capas (convolución y de agrupación) realizan la extracción de características, mientras que la tercera (capa totalmente conectada) convierte las características extraídas en un resultado final. La capa de convolución desempeña un papel fundamental en la CNN, que está compuesta por una pila de operaciones matemáticas, como la convolución, un tipo especializado de operación lineal. En las imágenes digitales, los valores de los píxeles se almacenan en una rejilla bidimensional (2D), es decir, una matriz de números como se muestra en la Figura 1.5, y en cada posición de la imagen se aplica una pequeña rejilla de parámetros llamada *kernel*, un extractor de características optimizable, lo que hace que las CNN sean muy eficientes para el procesamiento de imágenes, ya que una característica de

Capítulo 1. Antecedentes

estas es que aparezcan en cualquier lugar de la imagen. A medida que una capa alimenta su salida a la siguiente, las características extraídas se jerarquizan y se vuelven progresivamente más complejas. El proceso de optimización de parámetros como los *kernels* se denomina entrenamiento, que se lleva a cabo para minimizar la diferencia entre las salidas y las etiquetas de la verdad fundamental mediante un algoritmo de optimización llamado retropropagación y descenso de gradiente, entre otros [25].

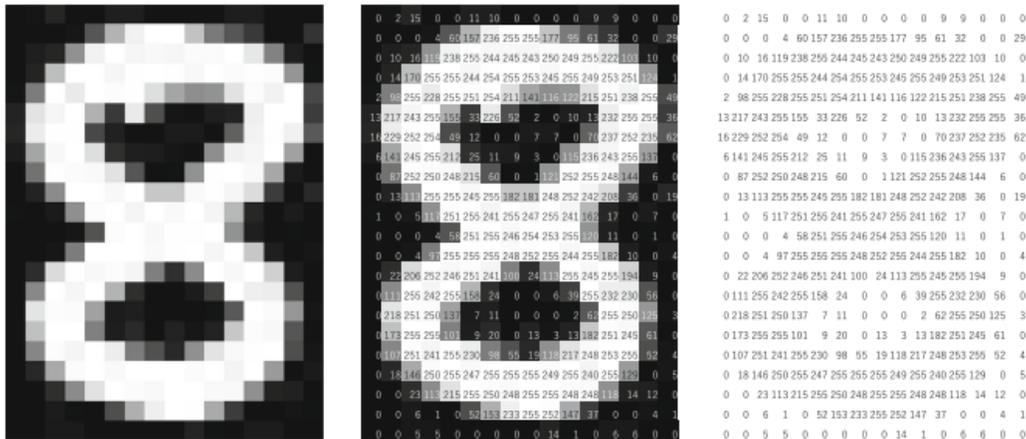


Figura 1.5 Imagen que muestra como visualiza una imagen una computadora.

La matriz de la derecha contiene números entre 0 y 255, cada uno de los cuales corresponde al brillo de los píxeles en la imagen de la izquierda. Ambos se superponen en la imagen del medio.

1.1.5.3 Tipos de CNN

Existen múltiples algoritmos derivados de las Redes Neuronales Convolucionales (CNN), a continuación, se describen brevemente R-CNN, Fast R-CNN y Faster R-CNN:

1. R-CNN: La R-CNN (*Region-Based Convolutional Neuronal Network*, Red Neuronal Convolutacional Basada en Regiones) es un hito importante de la CNN para la detección de objetos. Es la primera red neuronal que propone delimitar regiones para realizar la detección de objetos basándose en la excelente capacidad de extracción y clasificación de características de la CNN. Para obtener las regiones se buscan las propuestas de los objetos potenciales y se delimitan con diferente altura y anchura.

Antes de la CNN (*Convolutional Neuronal Network*), R-CNN propone recortar/deformar (*crop/warp*) para normalizar las imágenes candidatas con tamaño fijo a estándar. Después de la CNN, añade la clasificación SVM (*Support Vector Machines*, Máquinas de Vectores de Soporte) y la regresión de los límites para obtener el resultado exacto de la detección de objetos. Gracias a todo el proceso que realiza la propuesta de región hace que la R-CNN consuma una gran cantidad de datos, tiempo, cálculos y energía [26].

2. Fast R-CNN: Uno de los problemas de la R-CNN es el aumento del tiempo de procesamiento ya que la generación de características de la CNN se procesa para todas propuestas de regiones. La *Fast R-CNN* (*Fast Region-Based Convolutional Neuronal Network*, Red Neuronal Convolutacional Rápida Basada en Regiones) nace para resolver ese problema.

Su esquema se muestra en la Figura 1.6, en primer lugar, el mapa de características de la imagen se genera a partir de toda la imagen de entrada. Posteriormente, las propuestas de regiones de la búsqueda selectiva se proyectan en el mapa de características mediante el *RoI Projection* (Proyección de Región de Interés). Finalmente, se clasifican los vectores de características extraídos de las propuestas de regiones. Este método consigue detectar el objeto más rápidamente reduciendo la generación de características [27].

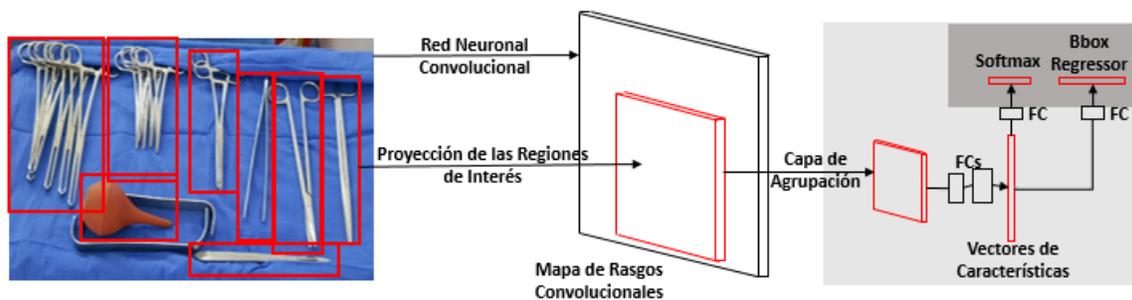


Figura 1.6 Esquema de la detección de objetos mediante Fast R-CNN

3. Faster R-CNN: Aunque la R-CNN rápida redujo el procesamiento de la CNN, seguía existiendo el problema que lleva tiempo el procesamiento de la búsqueda selectiva. Como un método mejorado surgió *Faster R-CNN* (*Faster Region-Based Convolutional Neuronal Network*, Red Neuronal Convolutacional más Rápida Basada

en Regiones), este método utiliza capas de CNN denominadas Red de Propuesta de Regiones (RPN) en lugar de la Búsqueda Selectiva.

El esquema de Faster R-CNN se muestra en la Figura 1.7. Faster R-CNN genera propuestas de regiones a partir de mapas de características generados por RPN. RPN escanea la ventana deslizante en el mapa de características y extrae los objetos candidatos. En este momento, para detectar objetos largos y delgados, cada cuadrícula tiene unas cajas delimitadoras llamadas “cajas de anclaje”. Por último, al igual que la R-CNN rápida, las propuestas de región se proyectan sobre el mapa de características, y el objeto se detecta mediante la clasificación de las propuestas de región. Dado que la R-CNN rápida genera las características de la imagen y las propuestas de región utilizando una sola CNN, existe la ventaja de que se realice un entrenamiento de principio a fin además de una detección más rápida [27].

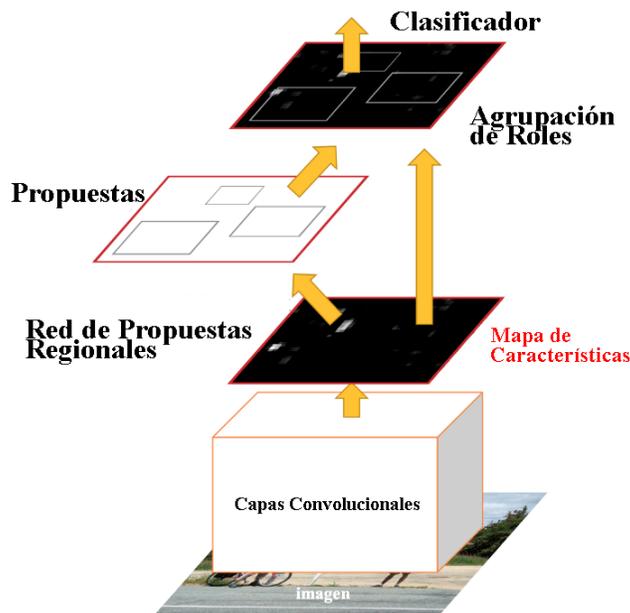


Figura 1.7 Esquema de la detección de objetos mediante Faster R-CNN.

1.1.5.4 Etiquetado de imágenes

Todas las tareas de aprendizaje profundo supervisado dependen de conjuntos de datos etiquetados, lo que significa la necesidad de que los humanos apliquen sus conocimientos para entrenar a la red neuronal sobre lo que está trabajando para identificar. Los objetos

Capítulo 1. Antecedentes

etiquetados se utilizarán por la red neuronal para entrenar un modelo que se utilice para realizar inferencias sobre los datos.

La anotación de imágenes, o el etiquetado, es vital para las tareas de aprendizaje profundo. Se necesita una gran cantidad de datos etiquetados para entrenar un buen modelo de aprendizaje profundo. Cuando se dispone de los datos de entrenamiento adecuados, los sistemas de aprendizaje profundo son muy precisos en la extracción de características, el reconocimiento de patrones y la resolución de problemas complejos [28].

1.1.5.5 APIs Utilizadas

Las API (*Application Programming Interfaces*, Interfaces de Programación de Aplicaciones) son un conjunto de herramientas y modelos utilizadas para crear software y modelos. Existen varios tipos (local, web y programa) estas ayudan a los desarrolladores de aprendizaje automático a comunicarse entre sí y compartir conocimientos entre varias plataformas.

En la tabla 1.2 se muestran una comparativa de las APIs utilizadas en el aprendizaje profundo [14].

Tabla 1.2 APIs utilizadas en el aprendizaje profundo.

Nombre	Servicio	Características	Accede a
Clarifai	Servicio de reconocimiento de imágenes y vídeo	Etiquetado de imágenes y vídeos, personalización del modelo, búsqueda de imágenes basada en la similitud visual, compatibilidad con varios idiomas, procesamiento escalable de imágenes y vídeos, modelo personalizado (modelo preentrenado) para categorías específicas (como bodas, viajes, comida, reconocimiento de rostros, color, modelo no seguro para el trabajo)	Biblioteca cliente para acceder a la API, HTTPS
Google Cloud Vision API	Servicio de análisis de imagen y API	Detección de etiquetas, detección de contenido explícito, detección facial, detección de puntos de referencia, detección	API REST integrada

Capítulo 1. Antecedentes

		de logotipos, análisis de sentimientos de imágenes, soporte multilingüe	
IBM Watson Vision Recognition Service	Servicio de reconocimiento visual para comprender el contenido de las imágenes	Imagen: Descripción y taxonomía de clases, detección de rostros, soporte multilingüe, identificación de coincidencias de imágenes con puntuación de confianza.	HTTP
Amazon Rekognition	Reconocimiento y análisis de imágenes basado en el aprendizaje profundo	Detección de objetos y escenas, análisis facial, comparación de caras, reconocimiento facial, integración con (AWS) Amazon Web Services, soporte multilingüe.	Interfaz de línea de comandos, HTTP
CloudSight	API de comprensión de imágenes	La descripción de la imagen se envía como respuesta cuando la imagen se envía a través de la API REST	API REST (biblioteca REST disponible en Ruby, Objective-C / Swift, Go, Python)

1.1.5.6 Proceso básico para la detección de imágenes utilizado en el aprendizaje profundo

El proceso para la detección de objetos se describe a continuación [29]:

1. Imagen de entrada: El modelo recibe la imagen para proceder a realizar la detección de objetos.
2. Extracción de regiones: Se generan propuestas de regiones independientes de la categoría.
3. Delimitación de áreas (*Wrapped*): Se extrae un vector de características de longitud fija de cada propuesta de región, posteriormente se obtiene un conjunto de SVM (*Support vector machine*, Máquinas de Vectores de Soporte) lineales para clasificar

los objetos de una imagen. Una vez que se identifican los objetos con precisión se ponen cajas delimitadoras en los objetos detectados.

4. Aplicación del CNN: Se utiliza una CNN para extraer un vector de características de 4096 dimensiones de cada propuesta de región. Como la capa totalmente conectada necesita vectores de entrada de longitud fija, las características de las propuestas de regiones requieren tener las mismas dimensiones.
5. Clasificación: Una vez aplicadas las CNNs se procede a la clasificación de objetos detectados en la imagen, es decir qué tipo de objeto encontró (perro, carro, flor).

En la Figura 1.8 se muestra el proceso para la detección de objetos en aprendizaje profundo.

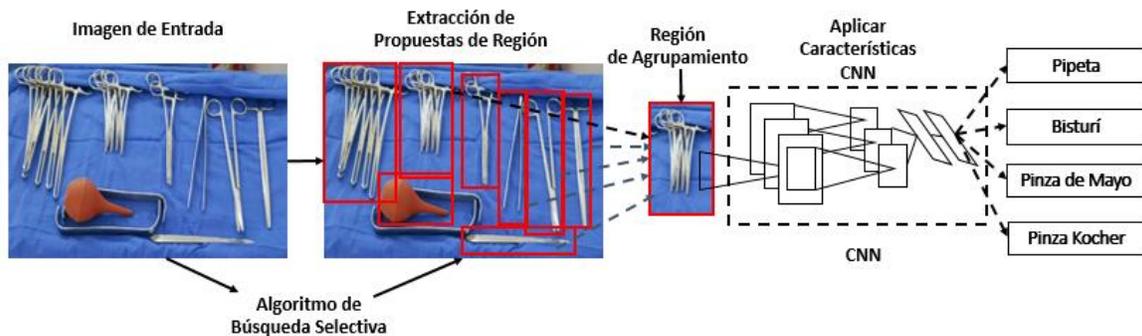


Figura 1.8 Esquema de la detección de objetos.

1.2 Planteamiento del Problema

En México existe déficit en la proporción de enfermeras-pacientes de acuerdo a los rangos que establece la Organización Mundial de la Salud (OMS) y la Organización Panamericana de la Salud (OPS). Existen aproximadamente 315 mil profesionales de enfermería en el sistema de salud lo que corresponde a 2.5 enfermeras por cada mil pacientes, por lo tanto aún faltan 115 mil profesionales de la salud (enfermeros) para alcanzar el índice de cobertura de países como Perú, Ecuador, Chile y Brasil donde cuentan con 6 a 7 enfermeros por cada mil pacientes [30].

Capítulo 1. Antecedentes

De acuerdo a lo anteriormente descrito las zonas rurales en México son las más afectadas ya que solo algunas cuentan con médicos, pero estos no tienen quien los asista (enfermera) en los diferentes procedimientos quirúrgicos (parto por cesárea y parto natural).

Por lo tanto, es importante implementar una solución que ayude a la capacitación de parteras o personal de la salud para que ingresen a una cirugía (parto por cesárea y/o parto natural) y reconozca de manera fácil y rápida el instrumental quirúrgico, de esta manera fungiría el papel de asistente médico.

Por lo anterior, se propone el desarrollo de una aplicación que realice la detección de material quirúrgico mediante técnicas de visión por computadora y *Deep Learning* (Aprendizaje Profundo).

1.3 Objetivo General y Objetivos Específico

1.3.1 Objetivo General

Aplicar técnicas de Visión por Computadora y *Deep Learning* para la realización de un análisis comparativo que permita el desarrollo de una aplicación Web para la identificación de instrumental quirúrgico.

1.3.2 Objetivos específicos

- Analizar trabajos relacionados para identificar diferencias entre iniciativas existentes y el tema que se aborda, para establecer la principal aportación del proyecto de tesis.
- Identificar las técnicas de Visión por Computadora que se implementarán para llevar a cabo el análisis comparativo para la identificación de objetos.
- Delimitar el instrumental quirúrgico que será parte del caso de estudio, sus usos y al menos un procedimiento quirúrgico en el que se utiliza.
- Diseñar e implementar los experimentos necesarios para el análisis comparativo de técnicas de Visión por Computadora que permita determinar la técnica más apropiada para la identificación de instrumental quirúrgico.

- Definir la arquitectura y el modelo de entrenamiento basado en *Deep Learning* con base alguna de las formas siguientes: a) Entrenamiento desde cero; b) Transferencia de aprendizaje y, c) Extracción de características.
- Recopilar un conjunto de imágenes etiquetadas para implementar el modelo de entrenamiento para el reconocimiento de instrumental quirúrgico.
- Implementar la aplicación Web basada en el resultado del análisis de técnicas de Visión por computadora y el modelo de entrenamiento del *Deep Learning* para la detección de instrumental quirúrgico.
- Realizar las pruebas de concepto para validar la aplicación Web desarrollada a través de al menos un caso de estudio que incluya la detección de instrumental quirúrgico en un entorno controlado.

1.4 Justificación

La falta de personal del sector salud provoca que muchas zonas rurales no cuenten con personal de asistencia médica, principalmente enfermeras, lo que conlleva a que los médicos requieran la asistencia de personal con conocimientos basados en costumbres culturales que si bien son útiles, carecen de los fundamentos médicos, como es el caso de las parteras quienes junto con los médicos se enfrentan a situaciones de partos naturales y cesáreas en las que comúnmente los médicos requieren soporte respecto al uso de instrumental quirúrgico y las parteras carecen del conocimiento respecto a este tipo de instrumentos generando complicaciones para dicho procedimiento. Por otro lado, los avances de la visión por computadora y aprendizaje profundo son cada vez más potentes, gracias a todas las áreas que abarca (medicina, educación, finanzas, seguridad, entre otros) y los resultados favorables que se obtuvieron en los últimos años.

Dicho lo anterior, el presente proyecto es de gran importancia para la educación y para el área de salud, ya que servirá para capacitar tanto a estudiantes y/o parteras, donde el objetivo es desarrollar una aplicación que sea capaz de detectar instrumental quirúrgico para que una persona conozca el nombre, la descripción, el uso y la ubicación de dicho instrumental de manera que facilite la asistencia a los médicos cirujanos, el involucrar a parteras en este tipo de procedimientos quirúrgicos ayuda a que se conserve y respete la riqueza cultural del

Capítulo 1. Antecedentes

personal de parto de los pueblos indígenas, a su vez estas adquirirán conocimientos respecto al instrumental quirúrgico utilizado en los partos por cesaría y partos naturales. El proceso anteriormente descrito se realizará mediante técnicas de visión por computadora y *Deep Learning*, donde se diseñará una interfaz que muestre todas las características del material quirúrgico detectado utilizado en los partos por cesárea y natural.

Capítulo 2. Estado de la práctica

El aprendizaje profundo es un tema que cada vez adquiere mayor relevancia, ya que usa las redes neuronales para mejorar la detección de voz, de rostros, de textos y de vídeo, por mencionar algunas de sus aplicaciones. En el presente capítulo se presenta un resumen detallado de 15 artículos analizados, además se muestra una comparativa de las características principales que se abordaron en cada uno de los trabajos analizados.

2.1 Trabajos relacionados

En [31] elaboraron un análisis de métodos de *Deep Learning* para la clasificación de imágenes y la detección de objetos. Realizaron una comparación de algunas herramientas de software existentes para problemas de *Deep Learning*. Los modelos analizados por los autores fueron: *autoencoders* (AE), RBM (*Restricted Boltzmann Machines*, Máquinas de Boltzmann Restringidas) y CNN (*Convolutional Neural Networks*, Redes Neuronales Convolucionales). Por otro lado, realizaron la comparación de 17 herramientas que incluyen bibliotecas y paquetes, extensiones de lenguajes de programación y lenguajes autónomos. La funcionalidad proporcionada varía mucho, lo que plantea el problema de elegir una herramienta adecuada. Un gran lote de herramientas admite una amplia serie de modelos que incluyen FCNNs (*Fully Connected Neural Networks*, Redes Neuronales Totalmente Conectadas), CNNs, AEs y RBMs. Deep Learn Toolbox, Theano, Pylearn2, Deepnet y DeepMap se encuentran entre ellas. Torch, Caffe, nnForge, CXXNET, Cuda Convent y Cuda CNN representan el grupo de herramientas de aprendizaje profundo destinadas al entrenamiento de alto rendimiento de CNN. El resto de las herramientas implementan principalmente variedades de redes neuronales profundas utilizando otras bibliotecas. Por ejemplo, Crino y R CNN se basan en Theano y Caffe respectivamente.

De manera similar al autor anterior, en [14] mencionaron que la tecnología de aprendizaje profundo (*Deep Learning*) se convirtió en una palabra de moda hoy en día a causa de los resultados de vanguardia obtenidos en el dominio de la clasificación de imágenes, la detección de objetos y el procesamiento del lenguaje natural. Por lo tanto evaluaron los marcos de aprendizaje profundo basados en CNN (*Convolutional Neural Network*, Red

Neuronal Convolutacional) para la detección de objetos, los cuales fueron: Caffe, Microsoft cognitivo, TensorFlow, Theano, Torch, Chainer, Keras, DeepLearning, Apache Signa, MXnet, Neon, Clarifai, Google Cloud API de visión, Microsoft Cognitive Service, IBM Watson Vision Recongition Service, Amazon Rekognition y CloudSight. Los conjuntos de datos proporcionan una forma de entrenar y verificar los algoritmos de visión por computadora y, por lo tanto, desempeñan un papel crucial para impulsar la investigación. Por lo tanto, los autores mencionaron 6 conjuntos de datos que son de gran ayuda para entrenar y verificar los algoritmos, los cuales son: Microsoft COCO, ImageNet, CUB-200-2011, Caltech 256, ILSVRC y PASCAL VOC Challenge.

En los últimos años con el rápido desarrollo del aprendizaje profundo, varias áreas de investigación lograron buenos resultados y junto con la mejora continua de las redes neuronales de convolución, la visión por computadora llegó a un nuevo pico. Por lo tanto en [32] analizaron algunos algoritmos relacionados con el aprendizaje profundo para la detección de objetos, posteriormente eligieron el algoritmo más eficiente. Para alimentar los datos investigaron sobre los siguientes conjuntos de datos ImageNet, Pascal Voc y Coco. Llegaron a la conclusión que el conjunto de datos ImageNet es uno más utilizados en el campo de aprendizaje profundo ya que cuenta con más de 14 millones de imágenes que cubren más de 20,000 categorías. Por otra parte los algoritmos analizados fueron R-CNN (*Regions with Convolutional Neural Network*), SPP-Net (*Spatial Pyramid Pooling*), Fast R-CNN (*Fast Regions with Convolutional Neural Network*) y Faster R-CNN (*Faster Regions with Convolutional Neural Network*). Los autores eligieron el algoritmo Faster R-CNN ya que obtuvo una mejor precisión en comparación con el resto de los algoritmos analizados. El algoritmo R-CNN obtuvo una precisión promedio del 0.66, mientras que SPP-NET obtuvo 0.631 mAP (*mean Average Precision*, Precisión Media promedio), por su parte Fast R-CNN obtuvo 0.669 mAP y Faster R-CNN obtuvo una precisión promedio de 0.732. Los autores concluyeron que los datos experimentales muestran que la tecnología del aprendizaje profundo es una herramienta eficaz para transmitir las características creadas por el hombre que se basa en el impulso de la experiencia de aprendizaje.

En [33] la aplicación del aprendizaje profundo demostró un rendimiento impresionante en varias áreas de aplicación, particularmente en la clasificación de imágenes, segmentación y detección de objetos. Por lo tanto, los autores analizaron y presentaron el funcionamiento de las arquitecturas CNN (*Convolutional Neural Network*, Redes Neuronales Convolucionales), a su vez presentaron una descripción detallada de sus modelos. Los modelos presentados fueron: LeNet, AlexNet, ZFNet, GoogleNet, VGGNet, ResNet, ResNeXt, SEnet, DenseNet, Xception y PNAS/ENAS. Se plantearon algunas de las aplicaciones de CNN para la detección de objetos como son: detección de animales salvajes, detección de seres humanos, detección de brazo pequeño (pistola) y la detección de objetos diversos. Es importante mencionar que los métodos de detección de objetos basados en el aprendizaje profundo lograron un progreso asombroso en los últimos años, por lo que será un área de investigación activa.

Wu et al. [34] mencionaron que al integrar características de múltiples escalas, los métodos basados en el aprendizaje profundo son capaces de detectar los objetos destacados desde una vista global y aproximada. Aunque se logró obtener un buen desempeño, todavía sufren de dos grandes inconvenientes: el primero, la totalidad de los objetos sobresalientes difícilmente se resaltan uniformemente por sus complejas estructuras internas; el segundo se refiere a que las predicciones alrededor de los contornos de los objetos son inexactas por la pérdida de información causada por las operaciones de convolución escalonada y agrupación. Para abordar esos problemas los autores propusieron un algoritmo para la detección de objetos sobresalientes, la detección de contornos en primer plano y la detección de bordes, para ello emplearon una estrategia de supervisión entrelazada. Además, los autores desarrollaron un módulo de aprendizaje mutuo (MLM) que sirve como componente básico del algoritmo propuesto. Los autores compararon su método con otros 15 encargados de la detección de objetos salientes en 7 conjuntos de datos desafiantes y 6 métodos de detección de bordes en BSD500 (*Berkeley Segmentation Dataset 500*, Conjunto de Datos de Segmentación de Berkeley 500), los resultados mostraron que estos mecanismos resultan en mapas de prominencia más precisos sobre una variedad de imágenes, a su vez el modelo desarrollado por los autores es capaz de detectar los bordes satisfactoriamente y a una mayor velocidad.

Mientras que en [35] con el rápido desarrollo de las técnicas de aprendizaje profundo, las DCNNs (*Deep Convolutional Neuronal Network*, red neuronal convolucional profunda) se volvieron más importantes para la detección de objetos. Los métodos de detección de objetos basados en aprendizaje profundo aprenden de características de imagen de bajo y alto nivel. Por lo tanto, los autores realizaron una revisión que se centra en los algoritmos de detección de objetos basados en DCNNs. Presentaron a detalle la estructura de la de las redes troncales (*Backbone*) las cuales se dividen en CBNs (*Complex Backbone Neuronal*, Red Troncal Compleja) y LBNs (*Complex backbone network*, Redes Troncales Ligeras). El primer grupo está conformado por redes como AlexNet, ZFNet, GoogLeNet, VGGNet, ResNet y DetNet; mientras que el segundo grupo está conformado por SqueezeNet, Xception, MobileNet, MobileNetV2, ShuffleNet, ShuffleNetV2 y PeleeNet. Las arquitecturas principales de detección de objetos se basan en DCNNs y se dividen en dos categorías, arquitectura de detección de objetos de dos etapas y arquitectura de detección de objetos de una etapa. Cada arquitectura cuenta con su grupo de algoritmos, RCNN, SPPNet, Fast RCNN, Faster RCNN, RFCN, Mask RCNN y FPN forman parte de la primera arquitectura anteriormente mencionada, mientras que OverFeat, YOLO, YOLO9000/V2, YOLOV3, SSD, DSSD y FSSD son parte de la segunda arquitectura. El entrenamiento DCNNs es inseparable de varios conjuntos de datos de imágenes. Los conjuntos de datos juegan un papel indeleble en el desarrollo DCNNs, ya sea que se trate del reconocimiento de imágenes, detección de objetos o segmentación de instancias, los conjuntos de datos presentados por los autores fueron: MINST, CIFAR-10, CIFAR-100, ImageNet, Open Images, MS COCO, PASCAL VOC, SUN, Tiny Images y Places. La detección de objetos tiene una amplia gama de aplicaciones en escenarios del mundo real, como lo son detección de rostros, detección de textos, detección peatonal, detección de logo, detección de vídeo, detección de vehículos y detección de imágenes médicas.

Por su parte en [29] analizaron los métodos de los modelos de detección de objetos típicos existentes y describieron los conjuntos de datos para alimentar de información los modelos descritos. Los detectores de objetos de imagen específicos de dominio preexistentes

generalmente se dividen en dos categorías, uno es detector de dos etapas, el más representativo, Faster R-CNN (*Faster Region-Convolutional Neural Network*). El otro es un detector de una etapa, como YOLO (*You-Only-Look-Once*) y SSD (*Single Shot Detector*). Los detectores de dos etapas tienen una alta precisión de localización y reconocimiento de objetos, mientras que los detectores de una etapa logran una alta velocidad de inferencia. Los autores describieron 10 conjuntos de datos (Pascal Voc, MS COCO, BENCHMARK, VisDrone2018, OPEN IMAGES V5, Caltech, KITTI, CityPersons, TDC, Euro City Persons) los últimos cinco son conjuntos de datos basados en la detección de peatones. Las áreas de aplicación típicas para la detección de objetos son campo de seguridad (detección de rostros, detección de peatones, detección de anomalías), campo militar (detección de objetos por teledetección), campo de transporte (reconocimiento de matrícula, reconocimiento de señales de tráfico), campo médico (sistemas de diagnóstico asistido por computadora, detección de glaucoma, detección automatizada de sinapsis y análisis de lesiones cutáneas) y campos de vida (detección de eventos, detección de patrones y generación de título de imágenes). Por otra parte, la detección de objetos tiene una amplia gama de escenarios de aplicación. La investigación de este dominio contiene una gran variedad de ramas, a continuación, se listan las que se analizaron: detección de objetos débilmente supervisada (WSOD), detección de objetos salientes, detección destacada, detección de bordes, detección de texto, detección de objetos en vídeo, detección de objetos 3D en nubes de puntos, reconocimiento visual fino, detección de postura 2D y 3D.

En [36] el auge del aprendizaje profundo en los últimos años surgió gracias en gran medida a los avances que permitió el campo de visión por computadora. Se revisaron 3 principales categorías de aprendizaje profundo para la visión por computadora, CNN (*Convolutional Neural Network*/ red neuronal convolucional profunda), SdAs (*Stacked Denoising Autoencoder*) y familia Boltzmann que incluye DBNs (*Deep Belief Networks*, Red de Creencia Profunda) y DBMs (*Deep Boltzmann Machines*), estas técnicas se emplearon para lograr tasas de rendimiento significativas en una variedad de tareas de comprensión visual como la detección de objetos, el reconocimiento de rostros, el reconocimiento de acciones y actividades, la estimación de poses humanas, la recuperación de imágenes y la segmentación semántica. Sin embargo, cada categoría tiene ventajas y desventajas distintas. Las CNN son

invariantes a las transformaciones, es un gran activo para ciertas aplicaciones de visión por computadora, a su vez las CNN tienen la capacidad única de aprendizaje de funciones. Por otro lado, dependen en gran medida de la existencia de datos etiquetados, a diferencia de los DBN, DBM y SdAs que funcionan sin supervisión. De los modelos analizados, tanto los CNN como los DBN son computacionalmente exigentes cuando se trata de entrenamiento, mientras que los SdAs se entrenan en tiempo real bajo ciertas circunstancias.

En [37] los ecologistas de poblaciones usan cámaras trampa para monitorear el tamaño de la población animal y administrar los ecosistemas de todo el mundo, ya que responden al movimiento, que generalmente se corresponde con la entrada de un animal en el marco. Los análisis de datos de cámaras trampa implican cuantificar manualmente la especie y el número de individuos en miles de imágenes. Automatizar este proceso tiene ventajas obvias, que incluyen una reducción del trabajo humano, una estimación imparcial de análisis y la disponibilidad de identificación de especies sin experiencia en el dominio. Por lo tanto, los autores utilizaron aprendizaje profundo basado en la detección de objetos para localizar, cuantificar e identificar especies animales dentro de las imágenes de la cámara trampa. Compararon dos métodos *Faster Region-Convolutional Neural Network* (Faster R-CNN) y *You-Only-Look-Once v2.0* (YOLO v2.0), para alimentar de información a los modelos ocuparon como conjuntos de datos *Reconyx CameraTramp* (RCT), que contiene 946 imágenes etiquetadas con 20 clasificaciones de especies y coordenadas de cuadro delimitador, así como el conjunto de datos *Gold Standard Snapshot Serengeti* (GSSS) que contiene 4096 imágenes etiquetadas de 48 clasificaciones de especies. Se obtuvieron los siguientes resultados: Faster R-CNN arrojó una precisión de 93.0% y 76.7%, y valores IOU (*Intersection Over Union*) de 0.804 y 0.722 en el conjunto de datos RCT y el conjunto de datos GSSS, mientras que YOLO arrojó una precisión de 73.0% y 40.3% y IOU de 0.570 y 0.221, respectivamente. Con los resultados obtenidos se concluye que el modelo Faster R-CNN tuvo mucha mayor precisión que YOLO.

En [38] la detección de objetos consiste en detectar un objeto específico a partir de una imagen de líneas y formas múltiples y complejas, se utiliza en la detección de rostros, seguimiento de objetos, recuperación de imágenes, entre algunos otros. Por lo tanto, se

propuso el uso de redes neuronales profundas, especialmente redes neuronales de convolución para realizar la detección de objetos. AlexNet, VGG Net, GoogLeNet y *Regional-Convolutional Neural Networks* (R-CNN) son los algoritmos que se ocuparon para realizar el experimento. Con el conjunto de datos PASCAL VOC se alimentaron de información a los algoritmos, el conjunto de datos contiene 20 etiquetas con un mínimo de 1000 imágenes. Los resultados obtenidos demostraron que R-CNN da un mejor resultado, mientras que Alex Net es el algoritmo menos eficaz ya que arrojó un 20% de error en su precisión.

En [39] la detección de objetos es un problema de reconocimiento visual fundamental en la visión por computadora y ha sido ampliamente estudiado en las últimas décadas. La detección de objetos tiene como objetivo encontrar objetos de ciertas clases con localización precisa en una imagen dada y asignar a cada instancia de objeto una etiqueta de clase correspondiente. Por lo tanto, los autores presentaron un estudio de los avances recientes en la detección de objetos con aprendizaje profundo, los puntos que abordaron fueron componentes de detección, estrategias de aprendizaje y aplicaciones. Se presentaron diferentes componentes de detección de objetos, el primero trata de un esquema sobre la elección de configuración de detección de objetos Bounding Box y Pixel Mask, el segundo es un paradigma para la detección de objetos y se divide en detectores de dos etapas y detectores de una etapa. Bajo estos esquemas los detectores utilizan una variedad de arquitecturas troncales de aprendizaje profundo (VGG16, RestNext, DenseNet, MobileNet, ResNeXt, DetNet, Hourglass Net), generadores de propuestas (métodos tradicionales de visión por computadora, métodos basados en puntos clave, métodos basados en anclajes) y aprendizaje de representación de características (características de múltiples escalas, codificación de características de región, razonamiento contextual). Por otro lado presentaron un análisis de la etapa de entrenamiento y la etapa de prueba que forman parte de la estrategia de aprendizaje. La detección de objetos es una tarea fundamental de la visión por computadora, hay muchas aplicaciones del mundo real basadas en esta tarea como son detección de rostros y detección de peatones.

Capítulo 2. Estado de la práctica

En [40] la detección de objetos en imágenes aéreas se ve desafiada por la variación de los colores de los objetos, las relaciones de aspecto, los fondos desordenados y en particular, las orientaciones indeterminadas. Por ello se propuso utilizar las características de D-CNN (*Deep Convolutional Neural Network*/ red neuronal convolucional profunda) de capas combinadas para realizar una detección de objetos aéreos robustos. Se evaluó el enfoque propuesto en un conjunto de datos de vehículos y en un conjunto de datos de aviones recompilados a partir de imágenes aéreas de Google Earth. Se obtuvieron los siguientes resultados para la característica de una sola capa, FC7 supera a POOL5 y FC6 en el conjunto de datos del vehículo, con una precisión de 0.861 cuando la recuperación se establece en 0.8. Mientras que, para el conjunto de datos de avión, POOL5 es el mejor entre los tres, con 0,891 tasa de precisión cuando la recuperación se establece en 0.9. Por otro lado, las funciones combinadas mejoran significativamente el rendimiento, especialmente para las funciones relacionadas con la capa POOL5. En la detección de vehículos, la combinación POOL5 + FC6 logró un 0.945 tasa de precisión con un 0.8 tasa de recuperación. En la detección de planos, las combinaciones POOL5 + FC6, POOL5 + FC7 y POOL5 + FC6 + FC7 muestran resultados similares, mientras que POOL5 + FC7 es un poco mejor.

En [41] durante años se creyó que modelar las relaciones entre objetos ayudaría al reconocimiento de los mismos, no hay evidencia de que la idea esté funcionando en el área de aprendizaje profundo. Por lo anteriormente descrito se propuso un módulo llamado módulo de relación de objetos, que procesa un conjunto de objetos simultáneamente a través de la interacción entre sus características de apariencia y de geometría, permitiendo así modelar sus relaciones. El módulo está basado en R-CNN (*Regions with Convolutional Neural Network*) que incluye la mayoría de los detectores de objetos modernos. El modelo propuesto está basado en cuatro pasos, el primero genera características de imagen completas, el segundo genera características regionales, el tercero realiza reconocimientos de instancias y el último realiza la eliminación de duplicados. Una vez entrenados los modelos se alimentaron de información por medio de COCO con 80 categorías de objetos. Los experimentos realizados demostraron que el módulo de relación de objetos mejora el reconocimiento de instancias.

En [42] el uso de EHR (*Electronic Health Records* / Historias Clínicas Electrónicas) se cuadruplicó del 11,8% al 39,6% entre los médicos de oficina de los EE.UU. de 2007 a 2012. Las imágenes médicas son una parte integral de la EHR de un paciente y actualmente son analizadas por radiólogos, quienes están limitados por la velocidad, la fatiga y experiencia. Se necesitan años y un gran costo financiero para capacitar a un radiólogo calificado, y algunos sistemas de atención de salud subcontratan los informes de radiología a países de menor costo. Un diagnóstico tardío o erróneo causa daño al paciente. Por lo tanto, es ideal que el análisis de imágenes médicas se lleve a cabo mediante un algoritmo de aprendizaje automático, preciso y eficiente. Las CNN son el algoritmo de aprendizaje automático más popular en reconocimiento de imágenes y tareas de aprendizaje, gracias a su característica única de preservar las relaciones de la imagen local, al mismo tiempo que realiza la reducción de dimensiones. Se realizaron múltiples CNN para la detección de enfermedades mediante imágenes médicas, entre ellas están detección de nódulos pulmonares en radiografías de tórax, los resultados arrojaron que el método clasificó los nódulos con un 86% de precisión y también encontraron que era robusto frente a diferentes niveles de entrada de ruido, se aplicaron capas convolucionales llamadas CheXNet para la detección de 14 enfermedades diferentes observadas en la radiografía del tórax; la clasificación en la neumonía en particular logró una puntuación de área bajo curva (AUC) de 0,7632 con el análisis de características operativas del receptor (ROC). La retinopatía diabética (RD) también se diagnostica utilizando CNN, los resultados obtuvieron un 75% de precisión en la clasificación.

En [43], mencionaron que la *Convolutional Neural Networks* (CNN) logró un gran éxito en muchas tareas de visión por computadora. Parcialmente inspirado por la neurociencia, la CNN comparte muchas propiedades con el sistema visual del cerebro. Una de las diferencias que se presenta es que CNN es típicamente una arquitectura de retroalimentación, mientras que en el sistema visual las conexiones recurrentes son abundantes. Inspirados en este hecho se propuso una CNN recurrente (R-CNN) para el reconocimiento de objetos mediante la incorporación de conexiones en cada capa convolucional. El R-CNN (*Regional-Convolutional Neural Networks*) se implementó usando *cuda-convnet2* desarrollado por Alex Krizhevsky. Se realizaron experimentos en dos GPU con paralelismo de datos. Los modelos se evaluaron con cuatro conjuntos de datos CIFAR-10, CIFAR-100, MNIST y

SVHN. Se descubrió que algunas iteraciones del procesamiento dinámico de RCNN produjeron excelentes resultados. Los resultados experimentales demostraron la ventaja de usar R-CNN sobre CNN para el reconocimiento de objetos.

2.2 Análisis Comparativo

En la tabla 2.1 se presenta una comparativa de los trabajos analizados anteriormente, con el fin de identificar las diferencias, similitudes y resultados obtenidos.

Tabla 2.1 Análisis comparativo de los artículos relacionados.

Autor	Enfoque	Problemática	Tecnologías utilizadas	Objetos que analizaron	Resultados	Estatus
Druzhkov and Kustikova [31]	Análisis de métodos y herramientas de <i>Deep Learning</i> para la clasificación de imágenes y detección de objetos.	Existe una gran cantidad de herramientas y métodos para <i>Deep Learning</i> , por tal motivo es importante conocer sus características y usos para aplicarlas correctamente.	Modelos <ul style="list-style-type: none"> • Autoencoders • Restricted Boltzmann Machines) Herramientas <ul style="list-style-type: none"> • Torch • Caffe • nnForge • CXXNET • Cuda Convent • Cuda CNN 	-	La funcionalidad que muestra cada una de las herramientas varía mucho, lo cual plantea el problema de elegir una herramienta adecuada.	Finalizado
Pathak et al.[14]	Análisis de marcos de aprendizaje profundo basados en CNN.	Los sistemas de reconocimiento visual que abarcan la clasificación, localización y detección de imágenes alcanzaron un gran impulso en la investigación.	APIs: Caffe, Microsoft cognitivo, TensorFlow, Theano, Torch, Chainer, Keras, Deeplearning, Apache Signa, MXnet, Neon, Clarifai, Google Cloud API de visión, Microsoft Cognitive Service, IBM Watson Vision Recognition Service, Amazon	General	Se analizaron diferentes herramientas basadas en CNN para la detección de objetos, la elección de cual utilizar depende de las necesidades que se tengan.	Finalizado

Autor	Enfoque	Problemática	Tecnologías utilizadas	Objetos que analizaron	Resultados	Estatus
			Rekognition y CloudSight Conjunto de datos: Microsoft COCO, ImageNet, CUB-200-2011, Caltech 256, ILSVRC y PASCAL VOC Challenge.			
Zhou et al.[32]	Análisis de algoritmos relacionados con el aprendizaje profundo para la detección de objetos.	En los últimos años, con el rápido desarrollo del aprendizaje profundo, varias áreas de investigación lograron buenos resultados.	Conjunto de datos: <ul style="list-style-type: none"> • ImageNet • PascalVoc • Coco Algoritmos: <ul style="list-style-type: none"> • R-CNN • SPP-Net • Fast R-CNN • Faster R-CNN 	General	El algoritmo que obtuvo mejor precisión es Faster R-CNN y el conjunto de datos que se eligió es ImageNet ya que cuenta con más de 14 millones de imágenes.	Finalizado
Dhillon and Verma [33]	Análisis, funcionamiento de las arquitecturas CNN y descripción de sus modelos.	Gracias al éxito y rápido desarrollo del aprendizaje profundo, varios campos lograron resultados considerables.	Modelos: <ul style="list-style-type: none"> • LeNet • AlexNet • ZFNet • GoogleNet • VGGNet • ResNet • ResNeXt • SEnet • DenseNet • Xception • PNAS/ENAS 	General	Se revisaron los métodos de detección de objetos basados en el aprendizaje profundo.	Finalizado
Wu et al.[34]	Desarrollo de un módulo de aprendizaje mutuo.	Los métodos basados en aprendizaje profundo todavía sufren de dos inconvenientes, el primero la	Arquitectura <ul style="list-style-type: none"> • R-FCN 	General	El modelo que se desarrollo es capaz de detectar bordes a una mayor velocidad.	Finalizado

Capítulo 2. Estado de la práctica

Autor	Enfoque	Problemática	Tecnologías utilizadas	Objetos que analizaron	Resultados	Estatus
		<p>totalidad de los objetos sobresalientes difícilmente se resaltan uniformemente por sus complejas estructuras internas; el segundo se refiere a que las predicciones alrededor de los contornos de los objetos son inexactas</p>				
Xiao et al.[35]	<p>Revisión de algoritmos para la detección de objetos basados en D-CNNs.</p>	<p>Con el rápido desarrollo de técnicas de aprendizaje profundo, las D-CNNs se volvieron más importantes para detección de objetos.</p>	<p>Algoritmos</p> <ul style="list-style-type: none"> • R-CNN • SPP-Net • Fast R-CNN • Faster R-CNN • RFCN • Mask RCNN • FPN • OverFeat • YOLO • SSD • DSSD • FSSD 	-	<p>Se realizó una revisión completa y detallada de los métodos de detección de objetos basados en el aprendizaje profundo.</p>	Finalizado
Jiao et al.[29]	<p>Análisis de los métodos de los modelos para la detección de objetos y análisis de conjuntos de datos.</p>	<p>Con el rápido desarrollo de algoritmos de aprendizaje profundo para tareas de detección, el rendimiento de los detectores de objetos se mejoró enormemente.</p>	<p>Algoritmos</p> <ul style="list-style-type: none"> • Faster R-CNN • YOLO • SSD <p>Conjunto de datos</p> <ul style="list-style-type: none"> • Pascal Voc • MS COCO • BENCHMARK • VisDrone2018 • OPEN IMAGES V5 • Caltech 	General	<p>Se revisaron algoritmos basados en la detección de objetos y se analizaron diferentes conjuntos de datos.</p>	Finalizado

Capítulo 2. Estado de la práctica

Autor	Enfoque	Problemática	Tecnologías utilizadas	Objetos que analizaron	Resultados	Estatus
			<ul style="list-style-type: none"> • KITTI • CityPersons • TDC • Euro City Persons 			
Voulodimos et al. [36]	Revisión de 3 categorías para el aprendizaje profundo.	En los últimos años, se demostró que los métodos de aprendizaje profundo superan a las técnicas anteriores de aprendizaje automático de última generación en varios campos, la visión por computadora es uno de los casos más destacados.	Modelos <ul style="list-style-type: none"> • CNN • DBN • DBM • SDAs 	-	Cada categoría tiene sus ventajas y desventajas distintas. Las CNN dependen en gran medida de la existencia de datos etiquetados, a diferencia de los DBN, DBM y SdAs que funcionan sin supervisión.	Finalizado
Schneider et al. [37]	Aplicación de aprendizaje profundo basado en la detección de objetos para localizar, cuantificar e identificar especies de animales dentro de las imágenes captadas por la cámara trampa.	El análisis de datos de cámaras trampa implica cuantificar manualmente la especie y el número de individuos en miles de imágenes.	Métodos <ul style="list-style-type: none"> • Faster R-CNN • YOLO2 Conjunto de datos <ul style="list-style-type: none"> • Reconyx CameraTramp • Gold Standard Snapshot Serengeti 	Animales	El modelo Faster R-CNN en conjunto con Reconyx Camera Tramp tuvo mucha mayor precisión que YOLO.	Finalizado

Capítulo 2. Estado de la práctica

Autor	Enfoque	Problemática	Tecnologías utilizadas	Objetos que analizaron	Resultados	Estatus
Shah [38]	Análisis de redes neuronales profundas.	Detección de objetos usando redes neuronales profundas	Algoritmos <ul style="list-style-type: none"> • AlexNet • VGG Net • GoogLe Net • R-CNN Conjunto de datos <ul style="list-style-type: none"> • PASCAL VOC 	-	El algoritmo R-CNN dio un mejor resultado, mientras que AlexNet es el algoritmo con menor precisión ya que presentó un 20% de error.	Finalizado
Wu, Sahoo and Hoy [39]	Estudio de los avances recientes en la detección de objetos con aprendizaje profundo.	La detección de objetos es un problema de reconocimiento visual fundamental en la visión por computadora.	Modelos <ul style="list-style-type: none"> • VGG16 • RestNext • DenseNet • MobileNet • ResNeXt • DetNet • Hourglass Net 	-	Se presentó un análisis detallado de modelos para la detección de objetos.	Finalizado
Zhu et al. [40]	Aplicación de D-CNN de capas para la detección de objetos aéreos.	La detección de objetos en imágenes aéreas se ve desafiada por la variación de los colores de los objetos, las relaciones de aspecto, los fondos desordenados y en particular, las orientaciones indeterminadas.	Conjunto de datos <ul style="list-style-type: none"> • Google Earth Algoritmo <ul style="list-style-type: none"> • D-CNN 	Aviones Vehículos	La capa FC7 superó a POOL5 y FC6 en el conjunto de datos de vehículo con una precisión del 0,861. Para el conjunto de datos de avión POOL5 es el mejor con 0,891 tasa de precisión.	Finalizado
Hu et al. [41]	Módulo encargado de procesar un conjunto de objetos simultáneamente a través de la interacción	Durante años se creyó que modelar las relaciones entre objetos ayudaría al reconocimiento de los mismos, no hay	Algoritmo <ul style="list-style-type: none"> • R-CNN Conjunto de datos <ul style="list-style-type: none"> • COCO 	-	El módulo de relación de objetos desarrollado mejora el reconocimiento de instancias.	Finalizado

Capítulo 2. Estado de la práctica

Autor	Enfoque	Problemática	Tecnologías utilizadas	Objetos que analizaron	Resultados	Estatus
	entre sus características de apariencia y geometría.	evidencia de que la idea esté funcionando en el área de aprendizaje profundo.				
Wiestler and Menze [42]	Análisis de imágenes médicas mediante un algoritmo de aprendizaje automático.	Las imágenes médicas son una parte integral de las historias clínicas electrónicas de un paciente y actualmente son analizadas por radiólogos, quienes están limitados por la velocidad, la fatiga y experiencia. Un diagnóstico tardío o erróneo causa daño al paciente.	Algoritmo <ul style="list-style-type: none"> • CNN 	Nódulos pulmonares, clasificación de neumonía y clasificación de retinopatía diabética.	El método clasificó nódulos con un 86% de precisión, la clasificación de la neumonía logró un puntuación de área bajo la curva de 0,7632, mientras que la retinopatía diabética obtuvo un 75% de precisión en la clasificación.	Finalizado
Liang and Hu [43]	Modelo basado en R-CNN para la detección de objetos mediante la incorporación de conexiones en cada capa convolucional.	La CNN comparte muchas propiedades con el sistema visual del cerebro. Una de las diferencias que se presenta es que CNN es típicamente una arquitectura de retroalimentación, mientras que en el sistema visual las conexiones	Algoritmo <ul style="list-style-type: none"> • R-CNN Herramienta <ul style="list-style-type: none"> • Cuda-convnet2 Conjunto de datos <ul style="list-style-type: none"> • CIFAR-10 • CIFAR-100 • MNIST • SVHN 	-	Los resultados experimentales demostraron la ventaja de usar R-CNN sobre CNN para el reconocimiento de objetos.	Finalizado

Capítulo 2. Estado de la práctica

Autor	Enfoque	Problemática	Tecnologías utilizadas	Objetos que analizaron	Resultados	Estatus
		recurrentes son abundantes.				

De acuerdo al análisis de la tabla anterior se observa que los algoritmos que presentan mejor rendimiento son YOLO y *Faster* CNN, por lo tanto en este anteproyecto de tesis se considera realizar un análisis de tecnologías (APIs y algoritmos) con el propósito de identificar cual es el algoritmo que presenta una detección de objetos más eficiente y corroborar si con los resultados obtenidos en el análisis de tecnologías se obtiene la misma conclusión que los autores de los artículos que se presentaron en la tabla comparativa.

Capítulo 3. Aplicación de la Metodología

El objetivo de este capítulo es presentar las etapas del desarrollo de la aplicación que consiste en la detección de instrumental quirúrgico utilizando técnicas de visión por computadora y aprendizaje profundo, los cuales están basados en la metodología XP (*eXtreme Programming*, Programación Extrema).

La metodología anteriormente mencionada consta de 4 fases las cuales son: planeación, diseño, desarrollo y pruebas.

3.1 Planeación

Esta fase consiste en la identificación de las historias de usuario, las cuales ayudan a obtener los requerimientos del sistema y la estimación de tiempos de desarrollo. A continuación, se presentan las historias de usuario.

La Tabla 3.1 muestra la historia de usuario para la selección del tipo de cirugía.

Tabla 3.1 Historia de usuario selección tipo cirugía.

Número: 1	
Nombre Historia: Selección tipo cirugía	
Prioridad: Media (Alta, Media, Baja)	Riesgo en Desarrollo: Bajo (Alta, Media, Baja)
Puntos Estimados: 1	Iteración Asignada: 1
Descripción: El usuario podrá seleccionar el tipo de cirugía a realizar.	
Observaciones: Tipo de cirugías definidas Cesárea y Parto Natural.	

La Tabla 3.2 muestra la historia de usuario para el registro de los participantes de la cirugía.

Tabla 3.2 Historia de usuario registra participantes cirugía.

Número: 2
Nombre Historia: Registra participantes cirugía

Capítulo 3. Aplicación de la Metodología

Prioridad: Media (Alta, Media, Baja)	Riesgo en Desarrollo: Bajo (Alta, Media, Baja)
Puntos Estimados: 1	Iteración Asignada: 1
Descripción: El usuario ingresara todos los campos que se visualizan en el formulario, posteriormente se procederá a guardar la información capturada en la base de datos.	
Observaciones: El número de cédula profesional no se ingresara solo si el tipo de usuario corresponde a un asistente.	

La Tabla 3.3 muestra la historia de usuario para la visualización de identificación de instrumental quirúrgico.

Tabla 3.3 Historia de usuario identificación de instrumental quirúrgico.

Número: 3	
Nombre Historia: Identificación de instrumental quirúrgico	
Prioridad: Alta (Alta, Media, Baja)	Riesgo en Desarrollo: Alta (Alta, Media, Baja)
Puntos Estimados: 1	Iteración Asignada: 1
Descripción: El usuario visualizara la detección de instrumental quirúrgico, si se observa un cuadro delimitador sobre el instrumental quirúrgico significa que la identificación se está realizando correctamente.	
Observaciones: La detección se instrumental se realiza a partir de la imagen captada desde una cámara.	

La Tabla 3.4 muestra la historia de usuario consultar cirugías realizadas.

Tabla 3.4 Historia de usuario consultar cirugías realizadas.

Número: 4	
Nombre Historia: Consultar cirugías realizadas	
Prioridad: Media (Alta, Media, Baja)	Riesgo en Desarrollo: Media (Alta, Media, Baja)
Puntos Estimados: 1	Iteración Asignada: 1

Descripción: El usuario visualizara todas las cirugías realizadas.
Observaciones:

La Tabla 3.5 muestra la historia de usuario generar reporte.

Tabla 3.5 Historia de usuario generar reporte.

Número: 5	
Nombre Historia: Generar reporte	
Prioridad: Media (Alta, Media, Baja)	Riesgo en Desarrollo: Media (Alta, Media, Baja)
Puntos Estimados: 1	Iteración Asignada: 1
Descripción: El usuario descargará un reporte de las cirugías realizadas en formato PDF.	
Observaciones: El reporte se generara de acuerdo a los filtros que se hayan aplicado al momento de consultar las cirugías realizadas.	

La detección de objetos se realiza a partir de una imagen, la cual se obtiene mediante una cámara que se sitúa en algún lugar de la sala de operaciones, una vez obtenida se procederá a realizar la identificación del instrumental quirúrgico. Este procedimiento se llevará a cabo mediante el entrenamiento de un modelo, el cual se entrenará con un set de datos descargado de diferentes repositorios de imágenes.

El usuario podrá visualizar la detección de instrumental quirúrgico mediante una aplicación web, pero antes tiene que registrar datos de perfiles de usuario y paciente.

3.2 Diseño

A continuación, se presenta el diseño de la aplicación, la cual se basa en los requerimientos identificados en las historias de usuario.

3.2.1 Arquitectura del sistema

Para lograr el resultado deseado es importante analizar las funciones que se esperan realice el sistema, una vez identificadas se procede a generar la arquitectura que brindará una solución al problema presentado en este proyecto de tesis. La Figura 3.1 muestra la arquitectura de la aplicación, en ella se observan los módulos que son encargados de realizar diferentes actividades.

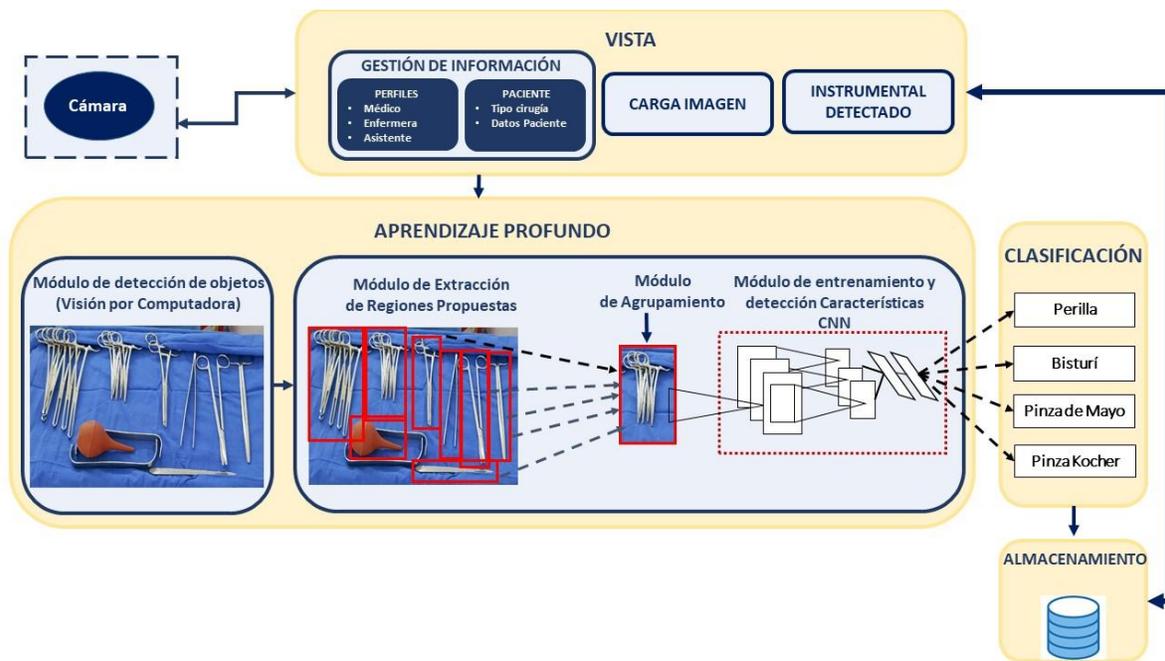


Figura 3.1 Arquitectura de la aplicación.

3.2.1.1 Capas de la arquitectura

A continuación, se describen las capas que conforman la arquitectura.

- **Vista:** En esta capa es la encargada de presentar la interfaz gráfica al usuario final, en dicha interfaz se presentan formularios para que se ingrese la información solicitada y se muestra el resultado de la detección del instrumental quirúrgico.
- **Aprendizaje Profundo:** Es la capa más importante de la aplicación, en ella se realizan técnicas de visión por computadora y aprendizaje profundo para la identificación de instrumental quirúrgico.
- **Clasificación:** Esta capa se obtiene la clasificación de los objetos identificados.

- **Almacenamiento:** Capa compuesta por la BD en la que se almacena toda la información de la cirugía a realizar, así como la ruta de la imagen cargada por el usuario y la lista del instrumental quirúrgico identificado.

3.2.1.2 Módulos que conforman el sistema

A continuación, se presenta la descripción de los módulos que conforman la arquitectura del sistema.

- **Gestión de información:** Este módulo es el encargado de gestionar la información de los perfiles de usuario (médico, enfermera, asistente) y los datos del paciente.
- **Carga imagen:** Módulo cuya función es permitir al usuario cargar una imagen a la cual se le aplicará la identificación de objetos.
- **Instrumental detectado:** Módulo encargado de presentar la imagen y lista con los nombres del instrumental quirúrgico identificado. A su vez se encarga de extraer las descripciones de la base de datos que corresponden al instrumental quirúrgico identificado, con el fin de que el usuario final conozca el uso de cada uno.
- **Detección de objetos:** En este módulo se aplican técnicas de visión por computadora (detección de objetos), como primer paso se recibe una imagen, posteriormente se realiza el procesamiento de esta y se identifican todos los objetos que se encuentren, una vez identificados se procede a generar en cada uno de ellos cuadros delimitadores y etiquetado de imagen.
- **Extracción de regiones propuestas:** Este módulo es el encargado de extraer todas las regiones de la imagen recibida sin importar la categoría. Estas regiones las enmarcara con cuadros delimitadores.
- **Módulo de agrupamiento:** Esta etapa es la encargada de analizar todos los objetos, en caso de que existiera más de un objeto del mismo tipo procederá a agruparlos.
- **Módulo de entrenamiento y detección de características con Redes Neuronales Convolucionales:** En este módulo se realiza el entrenamiento de la información, dicho entrenamiento se lleva a cabo mediante Redes Neuronales Convolucionales (CNN, *Convolutional Neural Network*). Este módulo recibe objeto por objeto, estos se comparan con todas las imágenes que tiene la red para encontrar a qué clasificación pertenece.

- **Módulo de visualización y descripción:** En esta etapa se muestra la clasificación de los objetos encontrados en la imagen, dicha información se presentará mediante una página web.

3.2.1.3 Flujo de trabajo

1. El usuario ingresa la información solicitada para el perfil usuario y el paciente, esta se almacena en la base de datos.
2. Se realiza la obtención de la imagen de instrumental quirúrgico, la cual se obtiene mediante una cámara que se encontrará situada en alguna parte de la sala de operaciones.
3. La imagen obtenida se procesa para posteriormente identificar todos los objetos que se encuentren en ella.
4. Se extraen todas las regiones propuestas de la imagen sin importar la categoría a la que pertenece.
5. A partir de las regiones propuestas, se realiza la agrupación de todos los objetos que pertenecen al mismo tipo.
6. Cada objeto identificado es evaluado mediante redes neuronales convolucionales, como resultado se obtiene la categoría a la que pertenece.
7. Las categorías de los objetos identificadas se visualizan en pantalla mediante una aplicación web, en cada objeto identificado se observan cuadros delimitadores con su respectivo nombre de clase.

3.2.2 Arquitectura de la red neuronal y convolucional

Como se observó en la sección anterior, una parte de la arquitectura del sistema está compuesto por una red neuronal convolucional. Por lo tanto, es importante presentar la arquitectura de dicha red; la Figura 3.2 presenta la arquitectura de la red en formato AlexNet.

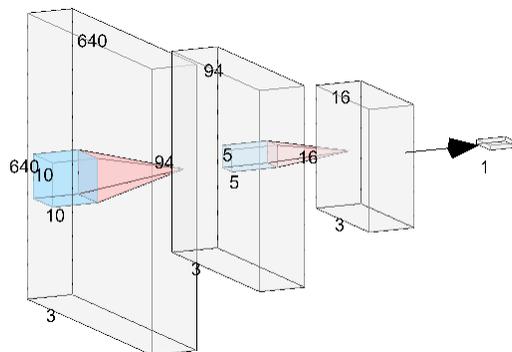


Figura 3.2 *Arquitectura de la red neuronal convolucional.*

La arquitectura de la red neuronal convolucional está formada por 3 capas. A continuación, se describe el uso de cada una.

- La primera, es una capa convolucional, recibe la imagen para posteriormente aplicar en ella un filtro (*kernel*) que devolverá un mapa con las características de la imagen original. Esta capa está conformada por el tamaño de la imagen recibida, por lo tanto, su dimensión es de 640 x 640 X 3, donde 640 corresponde al ancho y alto de la imagen en pixeles, debido a que la imagen que recibe es a color el número de canales (rojo, verde y azul) es 3. Por lo tanto, la primera capa está compuesta por 1, 228, 800 convoluciones.
- La segunda, corresponde a una campaña de reducción. Su principal función es reducir la cantidad de parámetros, esto sucede gracias a que únicamente se queda con las características más comunes.
- La tercera, es una capa clasificadora totalmente conectada, la cual da el resultado final de la red. En esta última capa, el número de neuronas corresponde al número de clases que se va a predecir.

3.2.3 Análisis de requerimientos

A continuación, se presentan los requerimientos identificados para el sistema. Iniciando por los requerimientos funcionales (Tabla 3.6) y no funcionales (Tabla 3.7).

Tabla 3.6 Requerimientos funcionales.

ID	Nombre	Descripción	Observaciones
RF1	Seleccionar tipo de cirugía	El sistema mostrará la lista de las cirugías disponibles y	El usuario podrá seleccionar el tipo de

Capítulo 3. Aplicación de la Metodología

		permitirá la selección del tipo de cirugía.	cirugía a practicar (cesárea o parto natural).
RF2	Registrar participantes.	El sistema mostrara un formulario con los datos requeridos y permitirá almacenar la información en la base de datos.	El usuario ingresará los datos solicitados por el sistema.
RF3	Identificación de instrumental quirúrgico.	El sistema mostrará el instrumental quirúrgico identificado en la mesa de mayo.	El sistema almacenará en base de datos el material identificado.
RF4	Mostrar historial de cirugías	El sistema mostrará en forma de lista todas las cirugías realizadas.	No se necesita autenticación para realizar esta función.
RF5	Generar reporte Historial	El sistema permitirá descargar el historial de las cirugías seleccionada.	El sistema generará un documento PDF de la cirugía seleccionada.

Tabla 3.7 Requerimientos no funcionales.

ID	Nombre	Descripción	Observaciones
RNF1	Diseño Resposivo	La aplicación web tiene que poseer un diseño responsivo, esto para garantizar la adecuada visualización en cualquier dispositivo.	Ninguna
RNF2	Visualización en navegadores	La aplicación web debe funcionar correctamente para los navegadores google Chrome y Microsoft Edge.	Ninguna
RNF3	Tiempo identificación de instrumental quirúrgico.	El sistema identificará el instrumental quirúrgico en	Ninguna

ID	Nombre	Descripción	Observaciones
		un tiempo no mayor a 5 segundos.	

Una vez analizados los requerimientos funcionales, se procede a la obtención del diagrama de casos de uso, a continuación, se muestran los casos de uso identificados para el actor usuario. La Figura 3.3 muestra las actividades que realizará dicho actor.

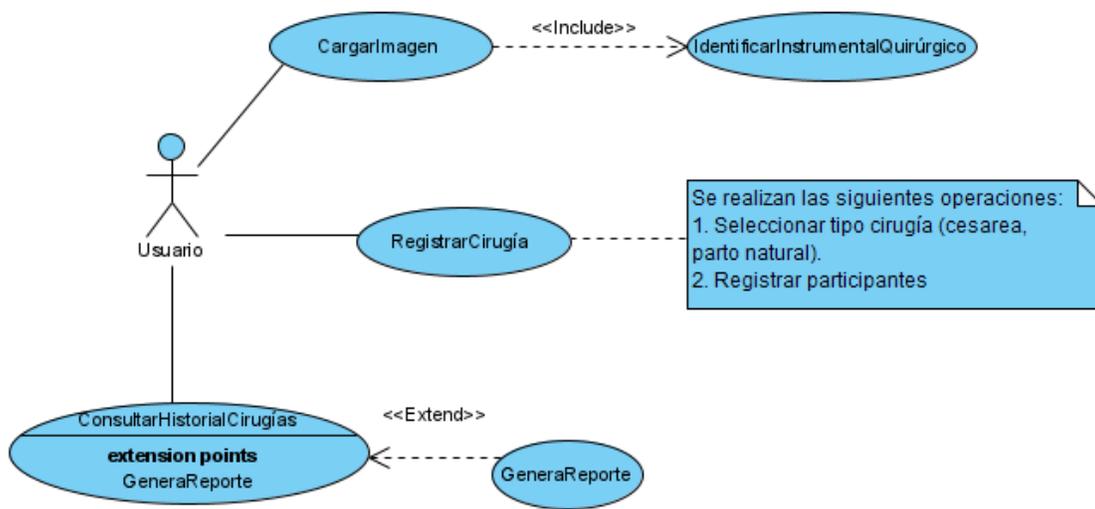


Figura 3.3 Diagrama de casos de uso.

A continuación, se detallan los casos de uso mostrados en la figura anterior. La Tabla 3.8 muestra la narrativa para el caso de uso seleccionar tipo cirugía.

Caso de Uso: Registrar cirugía.

Actor principal: Usuario

Precondiciones: No tiene

Postcondiciones:

- El sistema almacenara la información ingresada por el actor.

Flujo básico

Tabla 3.8 Narrativa de casos de uso registrar cirugía.

Actor	Sistema
1. Selecciona tipo cirugía y presiona el botón aceptar.	
	2. Muestra un formulario con los siguientes campos: nombre, apellido paterno, apellido materno, cédula, tipo usuario y nombre del paciente.
3. Ingresa todos los campos solicitados y presiona el botón guardar.	
	<p>4. Valida que todos los campos obligatorios sean ingresados de lo contrario ver curso de excepción “Valida datos de entrada” Tabla 3.9.</p> <p>5. Se genera una nueva clave (última clave más uno) para la cirugía a realizar, se almacena la clave generada, el nombre del paciente y el tipo de cirugía.</p> <p>6. Se genera una nueva clave (última más uno) por cada participante agregado y se almacenan los datos de nombre, apellido paterno, apellido materno, tipo de usuario, número de cedula profesional, fecha y hora de registro.</p> <p>7. Fin de caso de uso</p>

Curso de Excepción “Valida datos de entrada”

Tabla 3.9 Narrativa de curso de excepción valida datos de entrada.

Actor	Sistema
	1. Valida que el nombre, apellido paterno, apellido materno, nombre del

Actor	Sistema
	<p>paciente hayan sido ingresados de lo contrario muestra el mensaje “Nombre del campo + es obligatorio”.</p> <p>2. Valida si el tipo de usuario es enfermera o médico y el campo cédula profesional es vacío muestra el mensaje “Cédula profesional obligatoria”.</p>

La Tabla 3.10 muestra la narrativa para el caso de uso cargar imagen.

Caso de Uso: Cargar imagen.

Actor principal: Usuario

Precondiciones: Se tiene que registrar previamente una cirugía.

Postcondiciones:

- El sistema realiza la identificación de instrumental quirúrgico y almacena en la base de datos la lista del instrumental quirúrgico identificado.

Flujo básico

Tabla 3.10 Narrativa de casos de uso cargar imagen.

Actor	Sistema
1. Arrastra o carga una imagen.	
	<p>2. Realiza la detección de instrumental quirúrgico. Ver curso alternativo “Identificar instrumental quirúrgico”.</p> <p>3. Fin de caso de uso.</p>

Curso alternativo “Identificar instrumental quirúrgico”

Tabla 3.11 Curso alternativo “Identificar instrumental quirúrgico”.

Actor	Sistema
	1. Muestra la imagen con el instrumental quirúrgico identificado, la cual está

Capítulo 3. Aplicación de la Metodología

Actor	Sistema
	rodeado de cuadros delimitadores y se visualiza el nombre de cada uno. 2. Almacena la lista de instrumental quirúrgico identificado en la tabla cirugía.
3. Presiona salir.	
	4. Regresa a flujo básico de “Seleccionar tipo de cirugía”.

La Tabla 3.12 muestra la narrativa para el caso de uso consultar historial cirugías.

Caso de Uso: Consultar historial cirugías.

Actor principal: Usuario

Precondiciones: Tienen que existir cirugías realizadas en la base de datos.

Postcondiciones:

- Se visualizan las cirugías realizadas

Flujo básico

Tabla 3.12 Narrativa de casos de uso consultar historial cirugías.

Actor	Sistema
1. Selecciona historial	
	2. Muestra en forma de lista todas las cirugías que se han realizado.
3. Si desea, selecciona generar reporte. Ver curso alternativo “Generar reporte” Tabla 3.13	
	3. Fin caso de uso.

Curso alternativo “Generar reporte”

Tabla 3.13 Curso alterno “Generar reporte”.

Actor	Sistema
	1. Genera un documento PDF con todas las cirugías realizadas. El reporte incluye clave de la cirugía, datos de los participantes, tipo de cirugía, datos del paciente y la lista del instrumental quirúrgico utilizado.

A continuación, en la Figura 3.4 se presenta el diagrama entidad-relación. Las tablas identificadas para el almacenamiento de la información son 5.

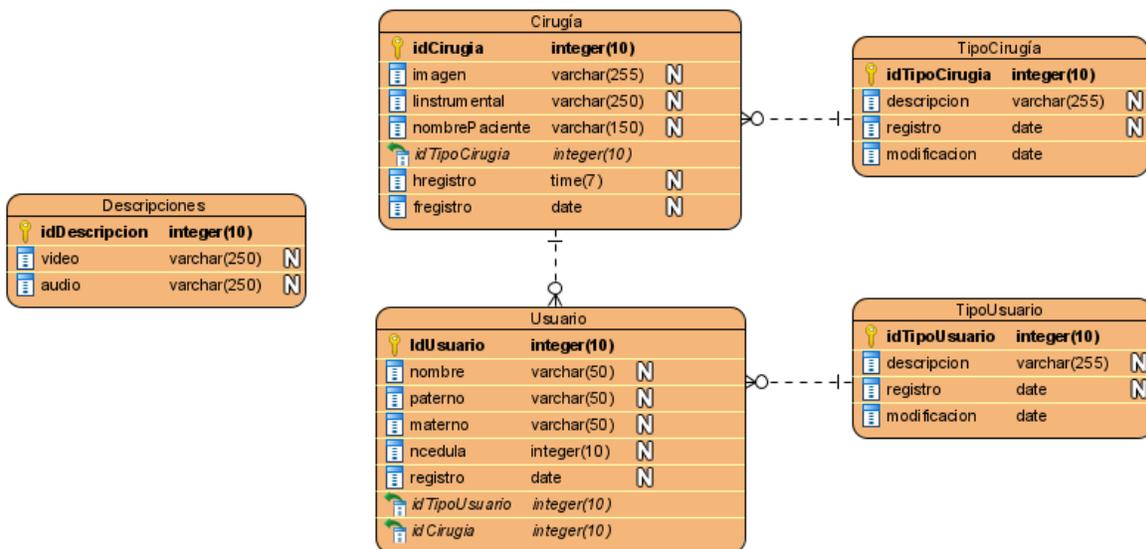


Figura 3.4 Diagrama entidad-relación.

Una vez que se tiene bien definido el diagrama entidad relación se procede a realizar el diccionario de datos, con el fin de describir los valores que contendrán cada uno de los campos que componen las diferentes tablas. La Tabla 3.14 muestra el diccionario de datos para la entidad Tipo Cirugía.

Tabla Tipo Cirugía

En esta tabla se almacenan los tipos de cirugías disponibles.

Tabla 3.14 Diccionario de datos tabla tipo cirugía.

Nombre del campo	Tipo	Tamaño	Nulo	Llave primaria	Descripción
idTipoCirugia	Integer	10	Not null	Sí	Identificador del tipo de cirugía, se incrementa de uno en uno.
Descripción	Varchar	255	Not null	No	Nombre de la cirugía.
Registro	Date		Not null	No	Fecha de alta.
Modificación	Date		null	No	Fecha de modificación.

La Tabla 3.15 muestra el diccionario de datos para la entidad Cirugía.

Tabla Cirugía

En esta tabla se almacenan los datos de las cirugías realizadas.

Tabla 3.15 Diccionario de datos tabla cirugía.

Nombre del campo	Tipo	Tamaño	Nulo	Llave primaria	Descripción
idCirugia	Integer	10	Not null	Sí	Identificador de la cirugía, se incrementa de uno en uno.
Imagen	Varchar	255	Not null	No	Ruta en donde se almacena la imagen a la que se le aplico la identificación de instrumental quirúrgico.
Listinstrumental	Varchar	255	Not null	No	Lista del instrumental quirúrgico identificado durante la cirugía.
nombrePaciente	Varchar	150	Not null	No	Nombre completo del paciente al que se le realiza la cirugía.
idTipoCirugia	Integer	10	Not null	No	Identificador que corresponde al tipo de cirugía a realizar.
Fregistro	Date		Not null	No	Fecha de alta.

Capítulo 3. Aplicación de la Metodología

Nombre del campo	Tipo	Tamaño	Nulo	Llave primaria	Descripción
Hregistro	Time		Not null	No	Hora de alta.

La Tabla 3.16 muestra el diccionario de datos para la entidad Tipo de Usuario.

Tabla Tipo de Usuario

En esta tabla se almacenan los tipos de usuarios que participarán en una cirugía.

Tabla 3.16 Diccionario de datos tabla Tipo de usuario.

Nombre del campo	Tipo	Tamaño	Nulo	Llave primaria	Descripción
idTipoUsuario	Integer	10	Not null	Sí	Identificador del tipo de usuario se incrementa de uno en uno.
Descripción	Varchar	255	Not null	No	Tipo de usuario que participara en una cirugía (médico, enfermera, asistente).
Registro	Date		Not null	No	Fecha de alta.
Modificación	Date		Null	No	Fecha de modificación.

La Tabla 3.17 muestra el diccionario de datos para la entidad Usuario.

Tabla Usuario

En esta tabla se almacenan los usuarios que participan en la cirugía.

Tabla 3.17 Diccionario de datos tabla usuario.

Nombre del campo	Tipo	Tamaño	Nulo	Llave primaria	Descripción
idUsuario	integer	10	Not null	Sí	Identificador de usuario se incrementa de uno en uno.
nombre	varchar	50	Not null	No	Nombre del participante.
paterno	varchar	50	Not null	No	Apellido paterno del participante,
materno	varchar	50	Not null	No	Apellido materno del participante.

Capítulo 3. Aplicación de la Metodología

Nombre del campo	Tipo	Tamaño	Nulo	Llave primaria	Descripción
ncedula	integer	10	Null	No	Número de la cédula profesional del participante.
registro	Date		Not null	No	Fecha de alta.
idTipoUsuario	integer	10	Not null	No	Identificador del tipo de usuario participante.
idCirugia	integer	10	Not null	No	Identificador de la cirugía que realizan los participantes.

La Tabla 3.18 muestra el diccionario de datos para la entidad Descripciones.

Tabla Descripciones

En esta tabla se almacenan las rutas en donde están guardados los audios y videos con la descripción del instrumental quirúrgico.

Tabla 3.18 Diccionario de datos tabla descripciones.

Nombre del campo	Tipo	Tamaño	Nulo	Llave primaria	Descripción
idDescripcion	integer	10	Not null	Sí	Identificador de la descripción se incrementa de uno en uno.
video	varchar	250	Not null	No	Corresponde a la ruta en la que está almacenado el video con la descripción del instrumental quirúrgico.
audio	varchar	250	Not null	No	Corresponde a la ruta en la que está almacenado el audio con la descripción del instrumental quirúrgico.

A continuación, en la Figura 3.5 se muestra la estructura del diagrama de clases el cual esta conformado por 6 clases.

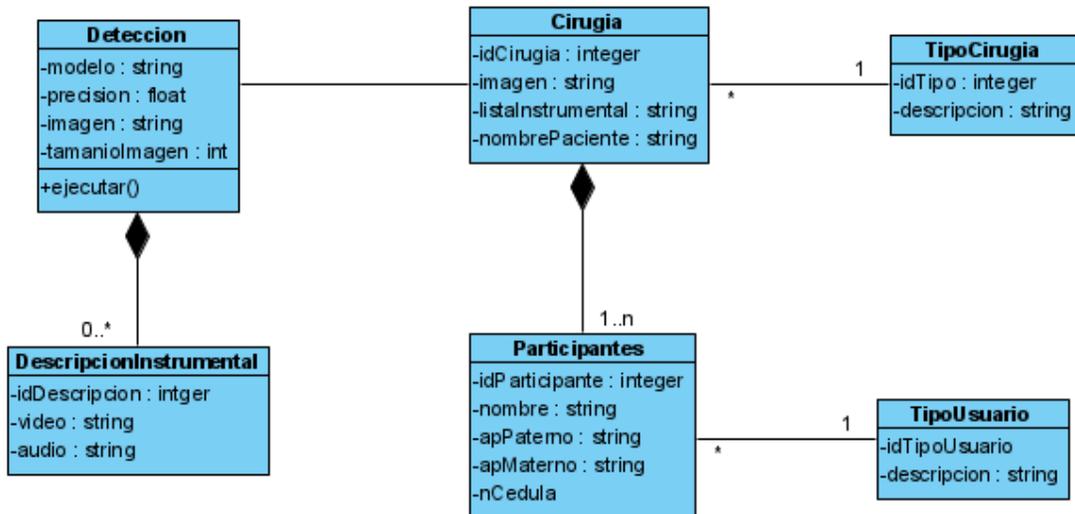


Figura 3.5 Diagrama de clases.

3.2.4 Nombre de la aplicación

Para darle identidad al sistema se le asignó el nombre de SurIAs, el cual nace al abreviar “Surgical Instrument Assistant”, por otra parte se realiza énfasis en las siglas IA ya que el proyecto consiste en aplicar técnicas de algunas de las ramas de la inteligencia artificial para la detección de instrumental quirúrgico.

3.2.5 Diseño de la aplicación

Una vez que se tienen claras las reglas del negocio es importante realizar el bosquejo de la aplicación web, ya que este sirve para presentar al cliente el posible funcionamiento que tendrá el sistema, una vez aceptado por el cliente se procede a iniciar con la programación. A continuación, se presentan los *wireframes* de las interfaces gráficas del sistema SurIAs.

De acuerdo a los requisitos identificados es necesario que el usuario (participante) seleccione el tipo de cirugía, la Figura 3.6 muestra la pantalla de selección de tipo de cirugía.

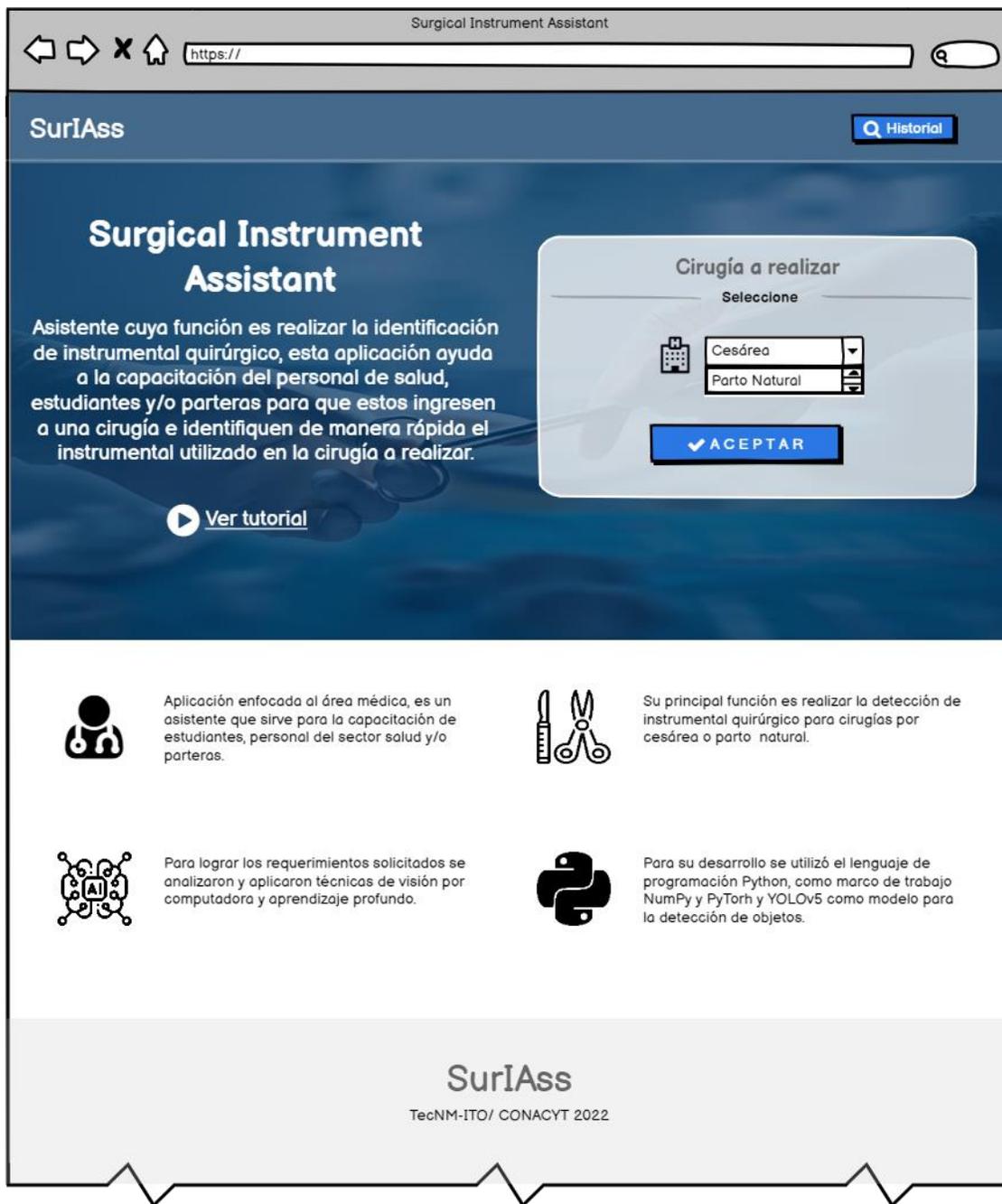


Figura 3.6 Pantalla de inicio, selección de cirugía.

El usuario selecciona el tipo de cirugía a realizar y presiona el botón aceptar. Posteriormente el sistema mostrará un formulario (Figura 3.7) para que se ingresen los datos de las personas que participaran en la cirugía y el nombre del paciente.

Surgical Instrument Assistant

[← Regresar](#)

Surgical Instrument Assistant

Por favor en el siguiente formulario ingrese todos los datos solicitados, para agregar un nuevo participante para la realización de la cirugía presione el botón de +. Una vez ingresada toda la información presione el botón guardar para continuar con la detección de instrumental quirúrgico. Para regresar a la página anterior presione el botón regresar.

Lista de participantes

Nombre	Cédula	Usuario	
Iván Rene Peralta Rodriguez	8870168	Médico	✖
Marisol Huerta Martínez	98326460	Asistente	✖

Ingrese los siguientes datos

Participantes

Nombre:

Paterno:

Materno:

Cédula:

Usuario: +

Enfermera

Asistente

Pacientes

Nombre:

Paterno:

Materno:

GUARDAR

SurIAss
TecNM-ITO/ CONACYT 2022

Figura 3.7 Formulario, ingreso de los participantes en la cirugía y nombre del paciente.

Una vez que el usuario presione guardar se muestra la siguiente interfaz (Figura 3.8) en donde se carga la imagen del instrumental quirúrgico a identificar.

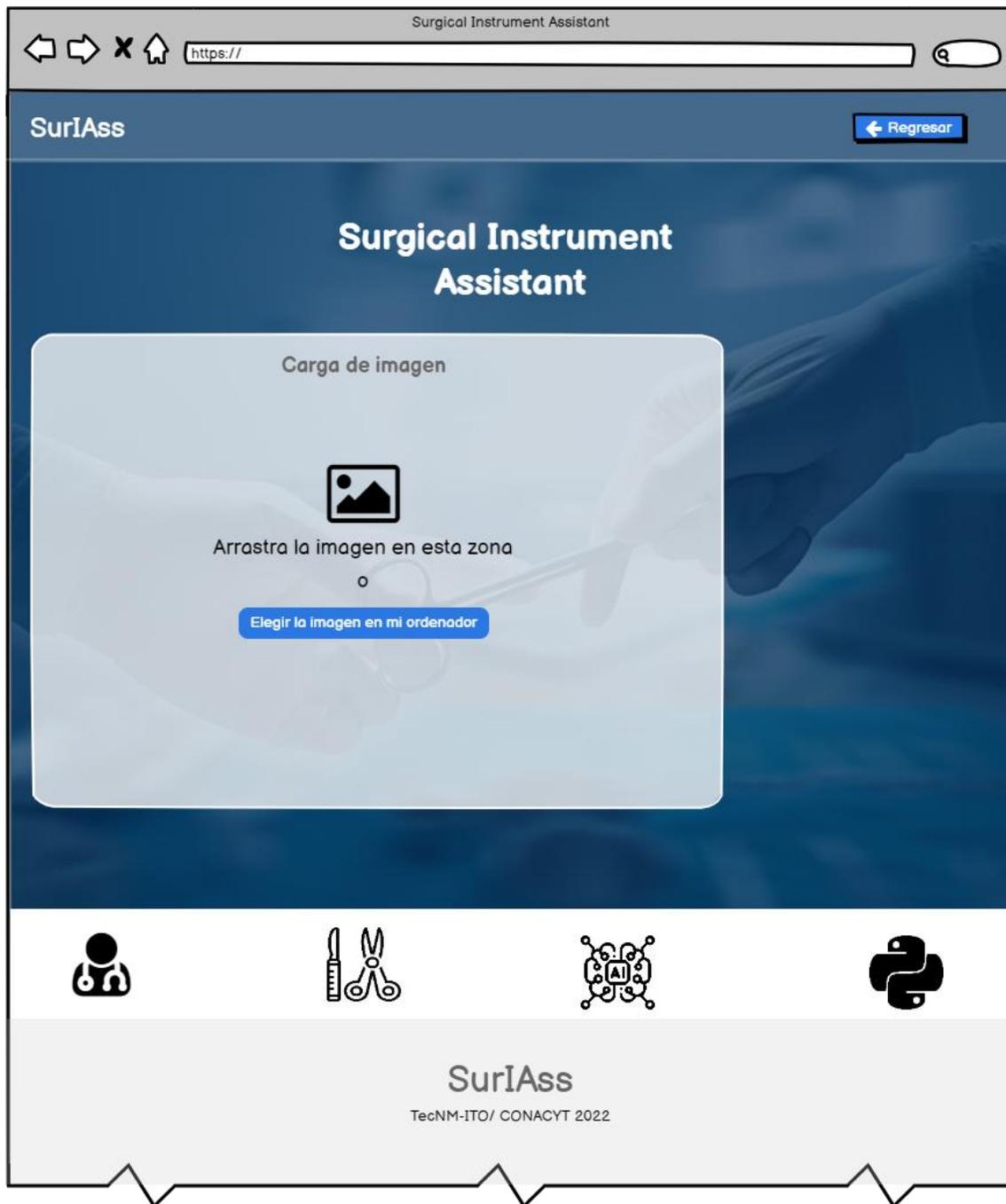


Figura 3.8 *Pantalla para la selección de imagen.*

Una vez cargada la imagen se inicia con la detección de instrumental quirúrgico, el resultado se presenta en la Figura 3.9.



Figura 3.9 Pantalla que muestra la detección del instrumental quirúrgico.

Si el usuario requiere consultar el historial de todas las cirugías realizadas, desde la página principal (inicio) selecciona historial, posteriormente se observa la lista de todas las cirugías realizadas como lo muestra la Figura 3.10.

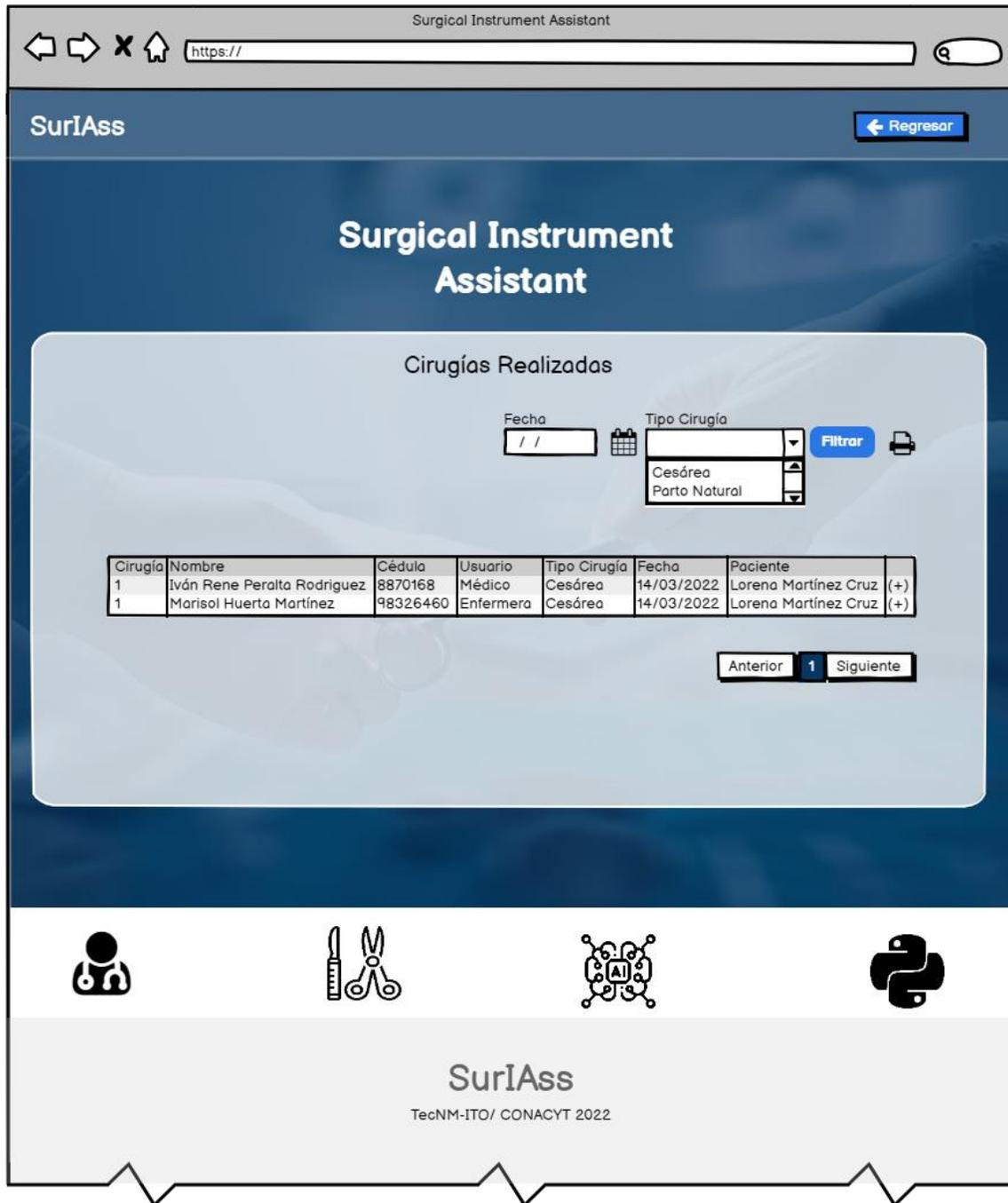


Figura 3.10 Pantalla que despliega todas las cirugías realizadas.

3.3 Desarrollo

En esta sección se presenta la tercera fase de la metodología XP (eXtreme Programming, Programación Extrema) la cual consiste en realizar todo el proceso de programación con el que se le dará funcionalidad a la aplicación web solicitada.

3.3.1 Entrenamiento del modelo

Para realizar la identificación del instrumental quirúrgico utilizado en la cirugía cesárea o parto natural, es necesario aplicar técnicas de visión por computadora y aprendizaje profundo. En los puntos siguientes se presentan las etapas (adquisición de imágenes, etiquetado de imágenes, implementación del modelo y evaluación del modelo) que se llevaron a cabo para realizar el entrenamiento del modelo.

3.3.1.1 Adquisición de imágenes

Es el punto más importante y en el que se tiene que tener gran cuidado, ya que las imágenes obtenidas tienen que ser de calidad para que el modelo logre obtener detalladamente las características que presentan los objetos a identificar.

A continuación, se muestran algunas de las imágenes obtenidas para la identificación de instrumental quirúrgico, estas imágenes se descargaron de diferentes repositorios, videos y otras fueron capturadas.

La Figura 3.11, representa el instrumental quirúrgico utilizado para el parto por cirugía cesárea.

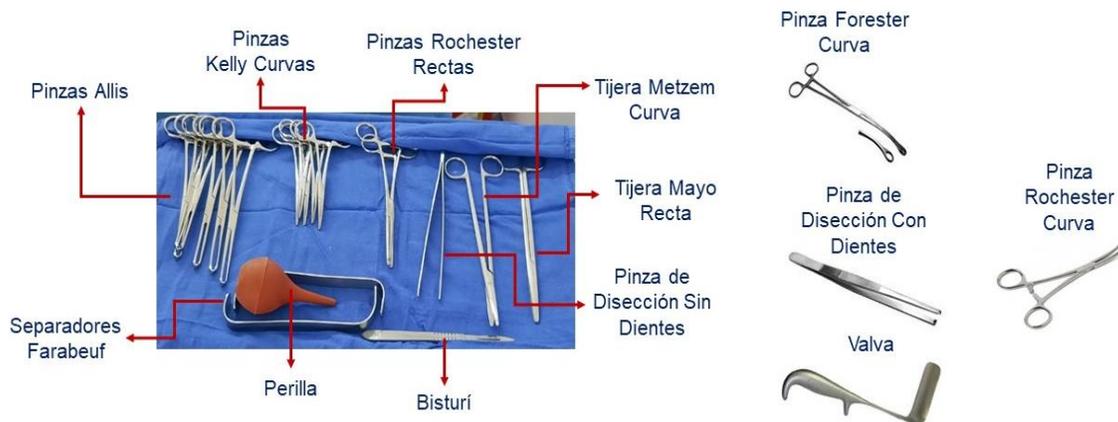


Figura 3.11 Instrumental quirúrgico utilizado para una cirugía cesárea.

Mientras que el instrumental utilizado para el parto natural se muestra en la figura 3.12.

Capítulo 3. Aplicación de la Metodología



Figura 3.12 Instrumental quirúrgico utilizado para una cirugía parto natural.

3.3.1.2 Etiquetado de imágenes

Una vez que se ha obtenido el conjunto de datos se procede a realizar el etiquetado de imágenes, el cual se realizó con ayuda de la aplicación labelImg como lo muestra la Figura 3.13.

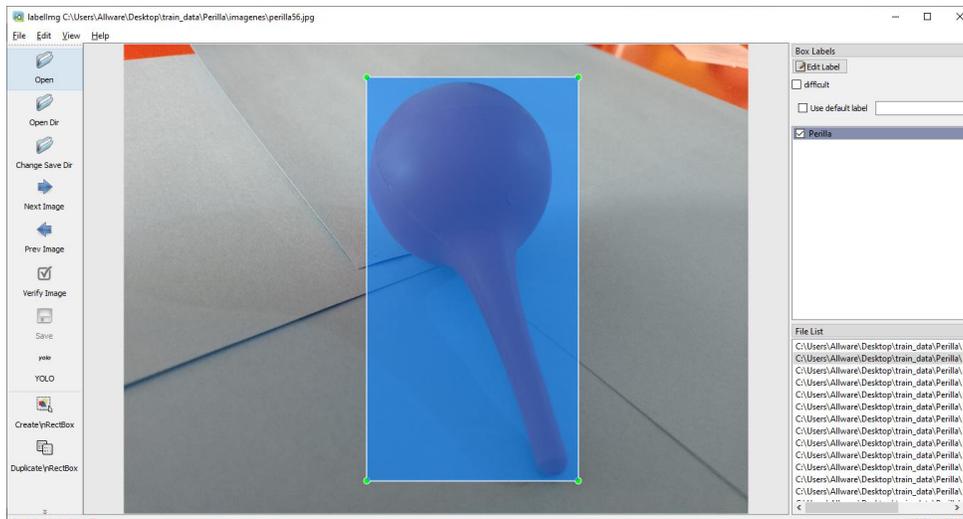


Figura 3.13 Etiquetado de imágenes desde el programa labelImg.

Como resultado del etiquetado de imágenes se obtuvo un archivo de texto plano por cada imagen etiquetada. La estructura obtenida es la siguiente:

```
1 0.493333 0.506667 0.648889 0.657778
```

Capítulo 3. Aplicación de la Metodología

El primer número indica la clase a la que pertenece, posteriormente se encuentran las coordenadas las cuales se generaron a partir de los cuadros delimitadores con los que se enmarcaron las secciones de interés.

3.3.1.3 Implementación del modelo

Después de etiquetar el set de datos se procede a realizar el entrenamiento del modelo con el algoritmo YOLO en su versión 5. Para lograrlo, se descarga el código del repositorio gitHub <https://github.com/ultralytics/yolov5>,

Una vez descargado, es necesario generar un archivo .yaml cuya función es indicar al algoritmo la ubicación las imágenes de entrenamiento y validación, además contiene el número y el nombre de las clases (nombre del instrumental quirúrgico a identificar). A continuación, se muestra el código que contiene el archivo generado.

Listado 3.1 *Código que contiene el archivo .yaml*

```
1 path: ../train_data # dataset root dir
2 train: ../train_data/images/train/ # train images (relative to
  'path') 128 images
3 val: ../train_data/images/val/ # val images (relative to 'path')
  128 images
4 test: # test images (optional)
5
6 # Classes
7 nc: 16 # number of classes
8 names: ['Bisturí', 'Perilla', 'P. Campo', 'P. Dis CD', 'P. Dis SD',
  'P. For C', 'P. Roch C', 'P. Roch R', 'T. Mayo R', 'P. Kelly
  C', 'P. Allis', 'Porta Aguja', 'Sep Fara', 'T. Mayo C', 'T.
  Metzem C', 12 'Valva']
```

Una vez que se tiene listo el archivo. yaml y el set de datos, se procede a realizar el entrenamiento del modelo.

La figura 3.14 muestra la línea de comando para ejecutar el entrenamiento, en ella se observa el nombre del programa encargado de realizar el entrenamiento (train.py) y los parámetros necesarios para realizar el entrenamiento los cuales se describen a continuación:

- **Img:** Tamaño de la imagen en pixeles.
- **Batch:** Corresponde al número de imágenes que va a procesar un lote (tamaño de lote).
- **Epoch:** Corresponde al número de veces que se entrenara el algoritmo de aprendizaje automático, es decir las veces por las que tiene que pasar por el conjunto de datos.
- **Data:** Nombre del archivo. yaml el cual contiene el nombre de las clases y las rutas de las imágenes de entrenamiento y validación.
- **Weights:** Nombre del archivo de los pesos pree entrenados.

```
!python train.py --project yolo-wandb-all --img 640 --batch 32 --epochs 100 --data customdata.yaml --weights yolov5s.pt --cache
```

Epoch	gpu_mem	box	obj	cls	labels	img_size					
0/99	6.29G	0.07409	0.03339	0.07268	92	640: 100%	94/94	[00:38<00:00,	2.42it/s]		
	Class	Images	Labels	P	R	mAP@.5	mAP@.5:.95:	100%	10/10	[00:04<00:00,	2.48it/s]
	all	634	760	0.192	0.296	0.114	0.0568				
Epoch	gpu_mem	box	obj	cls	labels	img_size					
1/99	8.63G	0.05247	0.02493	0.06283	108	640: 100%	94/94	[00:37<00:00,	2.54it/s]		
	Class	Images	Labels	P	R	mAP@.5	mAP@.5:.95:	100%	10/10	[00:03<00:00,	2.63it/s]
	all	634	760	0.399	0.341	0.174	0.0851				
Epoch	gpu_mem	box	obj	cls	labels	img_size					
2/99	8.63G	0.04805	0.02288	0.0576	107	640: 100%	94/94	[00:36<00:00,	2.54it/s]		
	Class	Images	Labels	P	R	mAP@.5	mAP@.5:.95:	100%	10/10	[00:03<00:00,	2.65it/s]
	all	634	760	0.452	0.311	0.257	0.155				

Figura 3.14 Línea de comando para la ejecución del entrenamiento del modelo.

3.3.1.4 Evaluación del modelo

Una vez que se entrenado el modelo, es necesario realizar la evaluación del mismo. Esto con el fin de validar qué tan preciso es al momento de realizar la detección del instrumental quirúrgico.

Para cada modelo entrenado se calcularon las métricas de precisión, *recall* y mAP, además se analizó el comportamiento de cada clase, esto con el fin de determinar en qué instrumentos quirúrgicos estaba siendo más preciso al momento de realizar la detección. A continuación, en la Figura 3.15 se muestran los valores obtenidos en las diferentes métricas para cada entrenamiento realizado.

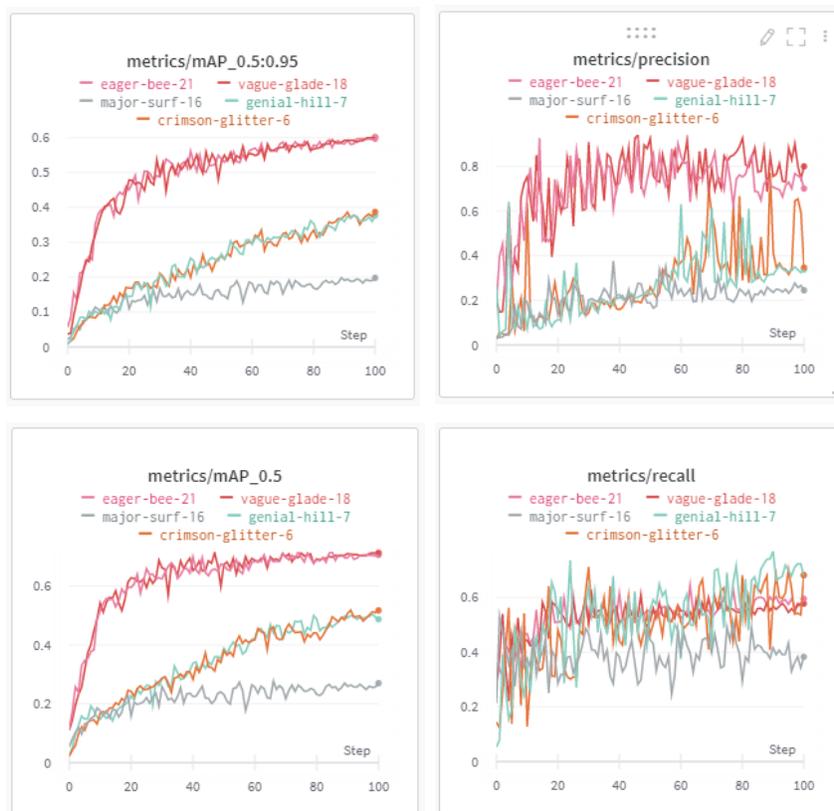


Figura 3.15 Métricas obtenidas de los entrenamientos realizados.

Como se observa en la Figura 3.15 para la métrica mAP, los últimos dos entrenamientos tienen un comportamiento similar, por lo tanto, es necesario analizar el comportamiento obtenido de cada una de las clases declaradas para nuestro modelo. A continuación, se presentan las métricas obtenidas por clase para los últimos tres entrenamientos.

La Figura 3.16 muestra las métricas obtenidas para las 16 clases definidas en el antepenúltimo entrenamiento realizado. Dicho entrenamiento se realizó con 1710 imágenes para entrenamiento y 658 imágenes para validación. La imagen muestra que se obtuvo un mAP general de 0.265 y hay clases que no llegaron al 0.1 de mAP. Al realizar la prueba de identificación de instrumental quirúrgico únicamente el modelo detecta dos de 16 clases declaradas, por lo tanto, en este modelo se presenta subajuste, para solucionarlo es necesario ingresar más imágenes al set de datos.

Capítulo 3. Aplicación de la Metodología

Entrenamiento: 1710		Validación: 658		
Class	P	R	mAP@.5	mAP@.5:.95
all	0.258	0.358	0.265	0.195
Bisturí	0.125	0.746	0.317	0.216
Perilla	0.855	0.981	0.984	0.871
P. Campo	0.37	0.309	0.236	0.196
P. Dis CD	0.384	0.517	0.464	0.327
P. Dis SD	0.352	0.577	0.393	0.269
P. For C	0.194	0.0465	0.133	0.086
P. Roch C	0.0834	0.0233	0.0935	0.0689
P. Roch R	0.0448	0.0213	0.0697	0.0552
T. Mayo R	0.281	0.25	0.209	0.136
P. Kelly C	0.207	0.279	0.251	0.2
P. Allis	0.196	0.298	0.182	0.141
Porta Aguja	0.129	0.35	0.126	0.102
Sep Fara	0.411	0.48	0.434	0.223
T. Mayo C	0.13	0.727	0.144	0.098
T. Metzerm C	0.37	0.125	0.183	0.116
Valva	0	0	0.0285	0.016

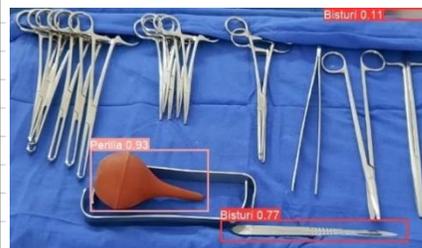


Figura 3.16 Resultados obtenidos en el antepenúltimo entrenamiento.

La Figura 3.17 muestra las métricas obtenidas para el penúltimo entrenamiento realizado. El modelo se entrenó con 2364 imágenes y se validó con 638 imágenes, se obtuvo un mAP de 0.713 comparado con el entrenamiento anterior se observa que el mAP aumento considerablemente. Al realizar la prueba del modelo se identifica y clásica el instrumental quirúrgico en un 100% correctamente.

Entrenamiento: 2364		Validación: 638		
Class	P	R	mAP@.5	mAP@.5:.95
all	0.8	0.578	0.713	0.602
Bisturí	0.663	0.492	0.545	0.367
Perilla	0.963	0.957	0.99	0.886
P. Campo	0.884	1	0.995	0.902
P. Dis CD	0.626	0.156	0.481	0.341
P. Dis SD	0.299	0.0701	0.371	0.253
P. For C	0.789	1	0.993	0.903
P. Roch C	0.442	0.976	0.875	0.812
P. Roch R	0.858	0.0244	0.378	0.307
T. Mayo R	0.809	0.926	0.959	0.87
P. Kelly C	1	0.0369	0.25	0.214
P. Allis	1	0	0.229	0.193
Porta Aguja	0.816	1	0.995	0.864
Sep Fara	1	0.0596	0.463	0.252
T. Mayo C	1	0.546	0.905	0.78
T. Metzerm C	0.746	1	0.99	0.911
Valva	0.902	1	0.995	0.783

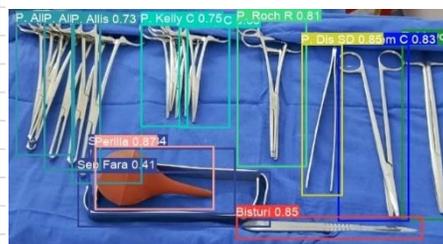


Figura 3.17 Resultados obtenidos en el penúltimo entrenamiento.

Capítulo 3. Aplicación de la Metodología

A continuación, en la Figura 3.18 se muestran las métricas obtenidas en el último entrenamiento ejecutado, se obtuvo un mAP menor al entrenamiento anterior, el modelo se validó con 638 imágenes y se entrenó con 3011 imágenes. Al realizar la prueba del modelo se identifica al 100% del instrumental, pero solo el 68% se clasificó correctamente.

Entrenamiento: 3011		Validación: 638		
Class	P	R	mAP@.5	mAP@.5:.95
all	0.701	0.595	0.706	0.597
Bisturí	0.505	0.542	0.531	0.361
Perilla	0.974	0.926	0.986	0.869
P. Campo	0.749	1	0.994	0.93
P. Dis CD	0.606	0.0381	0.491	0.357
P. Dis SD	0.636	0.356	0.52	0.334
P. For C	0.661	1	0.977	0.923
P. Roch C	0.479	0.93	0.757	0.704
P. Roch R	0.988	0.0488	0.386	0.319
T. Mayo R	0.737	0.926	0.958	0.866
P. Kelly C	0.642	0.0714	0.204	0.171
P. Allis	0	0	0.2	0.172
Porta Aguja	0.805	1	0.994	0.874
Sep Fara	0.778	0.14	0.422	0.213
T. Mayo C	1	0.546	0.894	0.771
T. Metzerm C	0.72	1	0.995	0.907
Valva	0.93	1	0.991	0.784

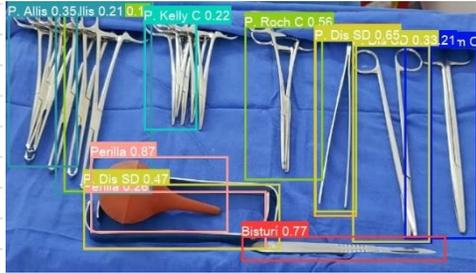


Figura 3.18 Resultados obtenidos en el último entrenamiento.

3.3.2 Desarrollo de interfaces gráficas

En esta fase se realizó el desarrollo para generar la interfaz gráfica de usuario (GUI); para el diseño del *frontend* se utilizaron las siguientes herramientas: django, JavaScript, HTML, CSS, entre otras.

La siguiente lista de código (Listado 3.2) muestra las bibliotecas utilizadas para realizar la representación gráfica de la interfaz (línea 2), para realizar la carga de las plantillas creadas (línea 3) y para realizar el consumo de los servicios backend (línea 1).

Listado 3.2 Fragmento de código de Bibliotecas utilizadas.

```
1 from django.http import HttpResponse
2 from django.shortcuts import render
3 from django.template.loader import get_template
```

La siguiente lista de código muestra las vistas generadas, las cuales nos servirán para darle funcionalidad a la interfaz gráfica.

Listado 3.3 *Fragmento de código de vistas declaradas.*

```
1     def index(request):
2         plantilla = get_template('index.html')
3         return HttpResponse(plantilla.render())
4
5     def registrar(request):
6         plantilla = get_template('registrar.html')
7         return HttpResponse(plantilla.render())
8
9     def cargarImagen(request):
10        plantilla = get_template('cargarImagen.html')
11        return HttpResponse(plantilla.render())
12
13    def deteccion(request):
14        plantilla = get_template('deteccion.html')
15        return HttpResponse(plantilla.render())
16
17    def reporte(request):
18        plantilla = get_template('reporte.html')
19        return HttpResponse(plantilla.render())
```

Ya que el encabezado y pie de página se tiene que mostrar en todas las vistas, se generó un solo archivo para que este pueda ser invocado desde el resto de las vistas. La siguiente lista (Listado 3.4) muestra el código HTML generado.

Listado 3.4 *Fragmento de código en HTML del encabezado y pie de página.*

```
1     <body>
2         <!-- ===== Header ===== -->
3         <header id="header" class="fixed-top ">
4             <div class="container d-flex align-items-center">
5
6                 <h1 class="logo me-auto"><a href="index.html">SurIAss</a></h1>
7                 <!-- Uncomment below if you prefer to use an image logo -->
8                 <!-- <a href="index.html" class="logo me-auto"></a-->
```

```

9         <nav id="navbar" class="navbar">
10             <ul>
11                 <li><a class="getstarted scrollto" href="#about">Historial</a></li>
12             </ul>
13             <i class="bi bi-list mobile-nav-toggle"></i>
14         </nav><!-- .navbar -->
15
16     </div>
17 </header><!-- End Header -->
18
19 {% block content %}{% endblock %}
20 <main id="main">
21
22     <!-- ===== Info del sistema ===== -->
23     <section id="team" class="team section-bg">
24         <div class="container" data-aos="fade-up">
25             <div class="row">
26
27                 <div class="col-lg-6">
28                     <div class="member d-flex align-items-start" data-aos="zoom-in"
29                         data-aos-delay="100">
30                         <div class="pic"><img src='{% static
31                             "assets/img/icon/medico.png" %}'
32                             class="img-fluid" alt=""></div>
33                         <div class="member-info">
34                             <p>Aplicación enfocada al área médica, es un asistente que
35                                 sirve para la capacitación de estudiantes, personal del
36                                 sector salud y/o parteras.</p>
37                         </div>
38                     </div>
39
40                     <div class="col-lg-6 mt-4 mt-lg-0">
41                         <div class="member d-flex align-items-start" data-aos="zoom-in"
42                             data-aos-delay="100">
43                             <div class="pic"><img src='{% static
44                                 "assets/img/icon/instrumental.jpg" %}' class="img-fluid"
45                                 alt=""></div>
46                             <div class="member-info">
47                                 <p>Su principal función es realizar la detección de
48                                     instrumental quirúrgico para cirugías por cesárea o
49                                     parto natural.</p>
50                             </div>
51                         </div>
52                     </div>
53                 </div>
54             </div>
55         </div>
56     </section>
57 </main>
58 </div>
59 </div>
60 </div>
61 </div>
62 </div>
63 </div>
64 </div>
65 </div>
66 </div>
67 </div>
68 </div>
69 </div>
70 </div>
71 </div>
72 </div>
73 </div>

```

```

44     <div class="col-lg-6 mt-4">
45         <div class="member d-flex align-items-start" data-aos="zoom-in"
46             data-aos-delay="100">
47             <div class="pic"><img src='{% static "assets/img/icon/ia.png"
48                 %}' class="img-fluid" alt=""></div>
49             <div class="member-info">
50                 <p>Para lograr los requerimientos solicitados se analizaron y
51                 aplicaron técnicas de visión por computadora y
52                 aprendizaje profundo.</p>
53             </div>
54         </div>
55     </div>
56
57     <div class="col-lg-6 mt-4">
58         <div class="member d-flex align-items-start" data-aos="zoom-in"
59             data-aos-delay="100">
60             <div class="pic"><img src='{% static
61                 "assets/img/icon/python.png" %}' class="img-fluid" alt=""></div>
62             <div class="member-info">
63                 <p>Para su desarrollo se utilizó el lenguaje de programación
64                 Python, como marco de trabajo NumPy y PyTorch y YOLOv5
65                 como modelo para la detección de objetos.</p>
66             </div>
67         </div>
68     </div>
69
70 </div>
71
72 </div>
73 </section><!-- End Team Section -->
74
75 </main><!-- End #main -->
76
77 <!-- ===== Footer ===== -->
78 <footer id="footer">
79     <div class="container footer-bottom clearfix">
80         <div style="text-align:center">
81             <h1 style="font-size: 30px">SurIAss</h1>
82             <h2 style="font-size: 18px">TecNM-ITO/ CONACYT 2022</h2>
83         </div>
84     </div>
85 </footer><!-- End Footer -->

```

Para la interfaz gráfica de página de inicio se generaron las siguientes líneas de código en HTML.

Listado 3.5 Fragmento de código para la página index.

```
1      {% extends "base.html" %}
2      {% load static %}
3      {% block content %}
4          <!-- ===== Sección Tipo Cirugía ===== -->
5          <section id="hero" class="d-flex align-items-center ">
6              <div class="container">
7                  <div class="row">
8                      <div class="col-lg-6 d-flex flex-column justify-content
9                          center pt-4 pt-lg-0 order-2 order-lg-1 "
10                         data-aos="fade-up" data-aos-delay="200">
11                          <h1>Surgical Instrument Assistant</h1>
12                          <h2>Asistente cuya función es realizar la identificación
13                          de instrumental quirúrgico, esta aplicación ayuda a la
14                          capacitación del personal de salud, estudiantes y/o parteras
15                          para que estos ingresen a una cirugía e identifiquen y
16                          reconozcan de manera rápida el instrumental utilizado en la
17                          cirugía a realizar.</h2>
18                          <div class="d-flex justify-content-center justify-
19                              content-lg-start">
20                              <a href="{% url 'registrar' %}" class="btn-get-started
21                                  scrollto">Iniciar</a>
22                              <a href="https://www.youtube.com/watch?v=jDDapla0z7Q"
23                                  class="glightbox btn-watch-video"><i class="bi bi-play-
24                                      circle"></i><span>Ver tutorial</span></a>
25                          </div>
26                      </div>
27                      <div class="col-lg-6 order-1 order-lg-2 hero-img"
28                          data-aos="zoom-in" data-aos-delay="200">
29                          
31                      </div>
32                  </div>
33              </div>
34          </section><!-- Fin Tipo Cirugía -->
35      {% endblock %}
```

Para realizar la identificación de instrumental quirúrgico el usuario tiene que ingresar todos los datos solicitados y una vez ingresados presiona el botón iniciar (línea 64). El listado siguiente muestra parte del fragmento de código HTML de la interfaz registrar.

Listado 3.6 Fragmento de código para la página registrar.

```
1    {% extends "base.html" %}
2    {% load static %}
3    {% block content %}
4    <section id="form">
5        ...
36        <div class="col-md-3" id="div-usuario">
37            </div>
38        <div class="col-md-3" id="div-cirugía">
39            <span class="select-btn">
40                <i class="zmdi zmdi-chevron-down"></i>
41            </span>
42        </div>
43        <div class="col-md-5">
44            <div class="form-heading">
45                <label for="room" class="special-label-1">Nombre completo
46                    paciente</label>
47                <input class="controls" type="text" name="nombrePaciente"
48                    id="nombrePaciente" placeholder="Ingrese nombre">
49            </div>
50        </div>
51        <div class="col-md-1" style="text-align:center">
52            <br>
53            <button type="button" class="btn btn-primary" onclick="validar
54                ()"></button>
55        </div>
56    </div>
57    <div class="row" id="div-tabla" style="display: none">
58        <div class="col-md-12">
59            <p>Lista de Participantes</p>
60            <table class="table" id="usuarios" style="color: white;">
61            </table>
62        </div>
63    </div>
64    <div class="row" id="div-siguiente">
65        <div class="col-md-4"></div>
66        <div class="col-md-4">
67            <input class="botons" type="button" id="iniciar" value=
68                "Iniciar" onclick="iniciar()">
```

Capítulo 3. Aplicación de la Metodología

Al presionar el botón iniciar se invoca una función en JS, cuyo objetivo es invocar a las APIs generadas para realizar las consultas y altas a la base de datos, además de invocar a la API encargada de realizar la identificación de instrumental quirúrgico. El listado siguiente muestra el código de la función iniciar.

Listado 3.7 Fragmento de código de la función iniciar.

```
1  function iniciar() {
2      if (info.length == 0) {
3          Swal.fire({
4              icon: 'error',
5              title: 'Oops...',
6              text: 'Por favor cargue participantes de cirugía',
7          });
8      } else {
9          if ($('#nombrePaciente').val() == "") {
10             Swal.fire({
11                 icon: 'error',
12                 title: 'Oops...',
13                 text: 'Por favor ingrese nombre del paciente',
14             });
15         } else {
16             generarAlta();
17         }
18     }
19 }
```

Para darle un estilo genérico a toda la aplicación, se generó código en CSS. A continuación, se muestran algunas de las líneas de código tipeadas para dar el estilo deseado.

Listado 3.8 Fragmento de código CSS.

```
1  # General
2  -----*/
3  body {
4      font-family: "Open Sans", sans-serif;
5      color: #444444;
6  }
7
8  a {
9      color: #47b2e4;
10     text-decoration: none;
11 }
```

```
12
13  a:hover {
14    color: #73c5eb;
15    text-decoration: none;
16  }
17
18  h1, h2, h3, h4, h5, h6 {
19    font-family: "Jost", sans-serif;
20  }
21  /*-----
22  # Preloader
23  -----*/
24  #preloader {
25    position: fixed;
26    top: 0;
27    left: 0;
28    right: 0;
29    bottom: 0;
30    z-index: 9999;
31    overflow: hidden;
32    background: #37517e;
33  }
34
35  #preloader:before {
36    content: "";
37    position: fixed;
38    top: calc(50% - 30px);
39    left: calc(50% - 30px);
40    border: 6px solid #37517e;
41    border-top-color: #fff;
42    border-bottom-color: #fff;
43    border-radius: 50%;
44    width: 60px;
45    height: 60px;
46    -webkit-animation: animate-preloader 1s linear infinite;
47    animation: animate-preloader 1s linear infinite;
48  }
49
50  @-webkit-keyframes animate-preloader {
51    0% {
52      transform: rotate(0deg);
53    }
54    100% {
55      transform: rotate(360deg);
56    }
57  }
```

Capítulo 3. Aplicación de la Metodología

Para realizar las validaciones de los campos obligatorios y para cargar imágenes se ocuparon bibliotecas y código en JavaScript, a continuación, se muestra con mayor detalle una parte del código tipeado (Listado 3.7).

Listado 3.9 Fragmento de código JS.

```
1     function validar(){
2
3         if($('#nombre').val() == ""){
4             Swal.fire({
5                 icon: 'error',
6                 title: 'Oops...',
7                 text: 'Por favor ingrese nombre',
8             });
9         }else {
10            if($('#paterno').val() == ""){
11                Swal.fire({
12                    icon: 'error',
13                    title: 'Oops...',
14                    text: 'Por favor ingrese apellido paterno',
15                });
16            }else{
17                if($('#materno').val() == ""){
18                    Swal.fire({
19                        icon: 'error',
20                        title: 'Oops...',
21                        text: 'Por favor ingrese apellido materno',
22                    });
23                }else{
24                    var user = document.getElementById("usuario").value;
25                    if(user === "Médico" || user === "Enfermera"){
26                        if($('#cedula').val() == null){
27                            Swal.fire({
28                                icon: 'error',
29                                title: 'Oops...',
30                                text: 'Por favor ingrese cédula profesional',
31                            });
32                        }
33                    }else{
34                        generarTabla();
35                    }
36                }
37            }
38        }
39    }
```

```

36         else{
37             generarTabla();
38         }
39     }
40
41     }
42 }
43 }
44
45 let info = []
46
47 function generarTabla(){
48     let nombre = document.getElementById("nombre").value + " " +
49         document.getElementById("paterno").value + " "
50         + document.getElementById("materno").value
51     let cedula = document.getElementById("cedula").value
52     let usuario = document.getElementById("usuario").value
53
54     info.push([nombre, cedula, usuario]);
55
56     let tabla = "<thead><tr><th>Nombre</th><th>Número de Cédula</th><th>Tipo
57         Usuario</th></tr></thead>"
58     for(let i = 0; i<info.length; i++){
59
60         tabla += '<tr><td>'+ info[i][0]+'</td><td>'+ info[i][1]+'</td><td>'+
61             info[i][2]+'</td></tr>'
62     }
63
64     document.getElementById("usuarios").innerHTML = tabla;
65
66     Swal.fire({
67         /*position: 'top-end',*/
68         icon: 'success',
69         title: 'Participante cargado exitosamente',
70         showConfirmButton: false,
71         timer: 1500
72     });
73
74     // Limpiamos los parametros
75     document.getElementById("nombre").value = "";
76     document.getElementById("paterno").value = "";
77     document.getElementById("materno").value = "";
78     document.getElementById("cedula").value = "";
79 }

```

3.3.3 Creación de la base de datos

En esta fase se generó la base de datos desde PostgreSQL, la cual servirá para almacenar toda la información necesaria de las cirugías a realizar dentro del sistema SurIAs.

A continuación, se muestra parte del fragmento de código (Listado 3.8) del script utilizado para la creación de la base de datos, el orden se presenta de la siguiente forma:

1. Creación de la base de datos surias
2. Creación de las tablas
3. Inserción de datos a las tablas

Listado 3.10 *Script para generar la BD.*

```
1      -- DROP DATABASE surias;
2      CREATE DATABASE surias;
3      CREATE TABLE cirugia (
4          idcirugia serial,
5          imagen character varying(255) NOT NULL,
6          listinstrumental character varying(255) NOT NULL,
7          nombrepaciente character varying(150) NOT NULL,
8          idtipocirugia integer NOT NULL,
9          hregistro time(6) without time zone,
10         fregistro date,
11         CONSTRAINT cirugia_pkey PRIMARY KEY (idcirugia),
12         CONSTRAINT tipocirugiafk FOREIGN KEY (idtipocirugia)
13             REFERENCES tipocirugia (idtipocirugia) MATCH SIMPLE
14             ON UPDATE CASCADE ON DELETE RESTRICT);
15         . . .
49     -- INSERCCIÓN DE DATOS EN LA TABLA TIPO CIRUGÍA
50     INSERT INTO TIPOCIRUGIA (DESCRIPCION, REGISTRO) VALUES ('Parto Natural',
51         '2022/07/01');
51     INSERT INTO TIPOCIRUGIA (DESCRIPCION, REGISTRO) VALUES ('Cesárea',
52         '2022/07/01');
```

3.3.4 Desarrollo de APIs

Para que la aplicación sea escalable se generaron APIs, esto con el fin de que otros usuarios consuman el modelo entrenado. Estos servicios se generaron con ayuda del marco de trabajo flask. A continuación, se presentan fragmentos de código de los servicios generados.

Listado 3.11 Código de la API para consultar todas las cirugías.

```
1 @app.route('/typeSurgeryAll, methods=['GET'])
2 def geTypeSurgery():
3     try:
4         cadena = restServiceBD.select()
5         response = jsonify({"TipoCirugía" : cadena, "Mensaje": "Datos
6             obtenidos correctamente"})
7         response.headers.add('Access-Control-Allow-Origin', '*')
8         return response
9     except:
10        return jsonify({'Mensaje': 'Error'})
```

Listado 3.12 Código de la API para consultar todos los roles.

```
1 @app.route('/rolesAll', methods=['GET'])
2 def getRoles():
3     try:
4         cadena = restServiceBD.selectRoles()
5         response = jsonify({"Roles" : cadena, "Mensaje": "Datos obtenidos
6             correctamente"})
7         response.headers.add('Access-Control-Allow-Origin', '*')
8         return response
9     except:
10        return jsonify({'Mensaje': 'Error'})
```

Listado 3.13 Código de la API para dar de alta una nueva cirugía.

```
1 @app.route('/insertSurgery', methods=['POST'])
2 @cross_origin(origins=['http://127.0.0.1:8000'])
3 def postInsertSurgery():
4     try :
5         fregistro = request.json['fregistro']
6         hregistro = request.json['hregistro']
7         idtipocirugia = request.json['idtipocirugia']
8         imagen = request.json['imagen']
9         listInstrumental = request.json['listInstrumental']
10        nombrePaciente = request.json['nombrePaciente']
11        result = restServiceBD.insertSurgery(idtipocirugia, fregistro, imagen,
12                                             listInstrumental, nombrePaciente, hregistro)
13        response = jsonify({'id':result, 'message': 'Alta Generada
14                          Exitosamente'})
15        return response
16    except:
17        return jsonify({'message': 'Error al conectar con la BD'})
```

Listado 3.14 Código de la API para dar de alta participantes de la cirugía.

```
1 @app.route('/insertUsers', methods=['POST'])
2 @cross_origin(origins=['http://127.0.0.1:8000'])
3 def postInsertUsers():
4     try :
5         nombre = request.json['nombre']
6         paterno = request.json['paterno']
7         materno = request.json['materno']
8         ncedula = request.json['ncedula']
9         idtipousuario = request.json['idtipousuario']
10        idcirugia = request.json['idcirugia']
11        fregistro = request.json['fregistro']
12
13        cadena = restServiceBD.insertUser(nombre, paterno, materno, ncedula,
14                                         idtipousuario, idcirugia, fregistro)
15        response = jsonify({'message': cadena})
16        return response
17    except:
18        return jsonify({'message': 'Error al conectar con la BD'})
```

Listado 3.15 Código de la API para realizar la detección de objetos.

```
1 # Get Data Routes
2 @app.route('/detect/', methods=['GET'])
3 def getDetect():
4     try:
5         rutaLectura = request.args.get('ruta1')
6         rutaAlmacenamiento = request.args.get('ruta2')
7         cadena = detect.run(source = rutaLectura)
8         ruta = cadena[1]
9         ruta = ruta[21:len(ruta) - 4]
10        rutaDescarga = descargaImg.guardarImg(ruta1 = ruta, ruta2 =
11            rutaAlmacenamiento, nombre = cadena[3])
12
13        response = jsonify({'ObjIdent': cadena[2], 'Ruta': rutaAlmacenamiento,
14            'NombreImg': cadena[3], 'Mensaje': 'Identificación de objetos exitosa'})
15        response.headers.add('Access-Control-Allow-Origin', '*')
16        return response
17    except:
18        return jsonify({'Mensaje': 'Error'})
```

3.3.4.1 Prueba de las APIs

Al concluir la programación de cada API, se realizó el consumo de cada una con ayuda de la aplicación Postman. A continuación, se agregan las pruebas realizadas a algunas de las API.

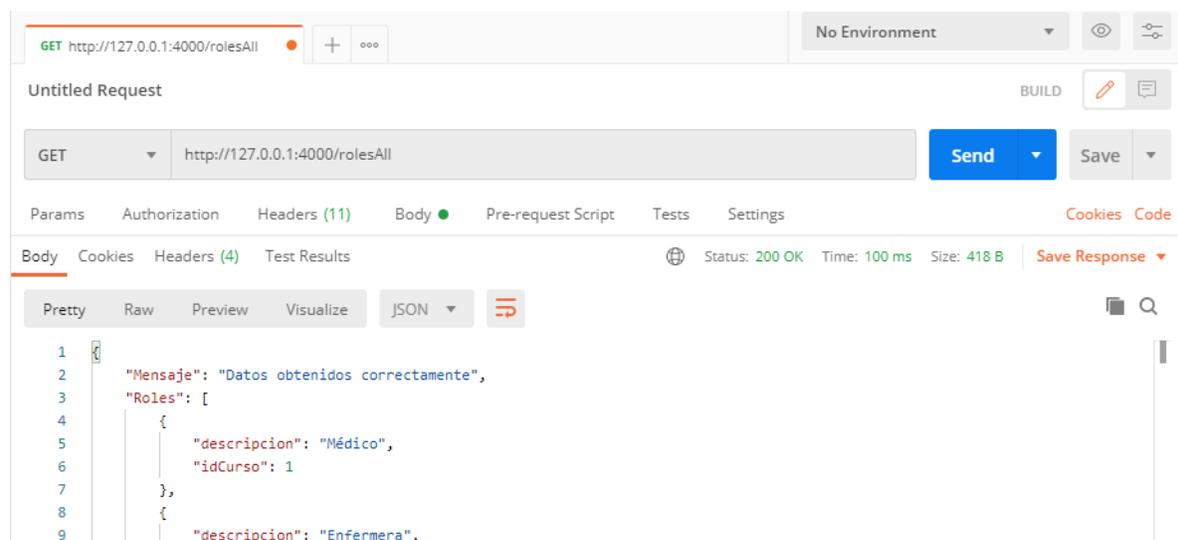


Figura 3.19 Prueba que muestra todos los tipos de usuario.

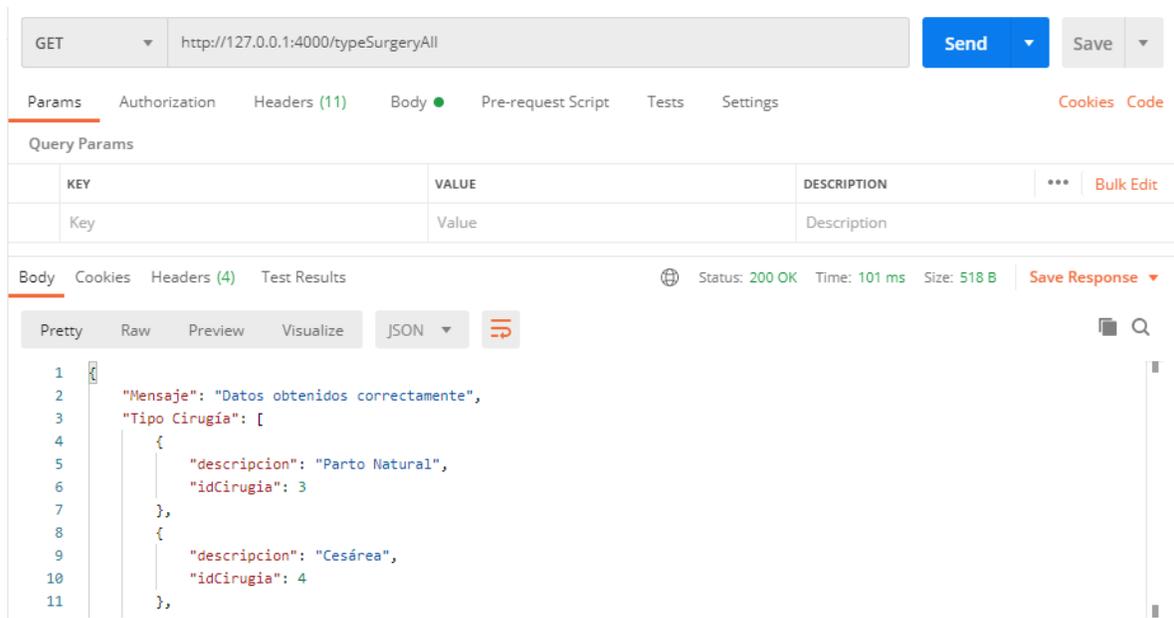


Figura 3.20 Prueba que muestra todos los tipos de cirugía.

3.3.5 Documentación

Para que cualquier otra persona pueda hacer uso de la aplicación SurIAs, se realizaron dos manuales. El primero es un manual que describe la funcionalidad y el contrato de las APIs generadas, el segundo es un manual de usuario de la aplicación SurIAs. Los cuales se anexan en formato PDF.

3.4 Pruebas

Al concluir el desarrollo, se procedió a realizar la unión de todos los componentes con el fin de comprobar que la aplicación funciona correctamente y que da solución al problema planteado. A continuación, se describe el comportamiento obtenido por cada interfaz generada.

3.4.1 Inicio de la página web

Para darle mayor información al usuario acerca de la aplicación SurIAs se generó un interfaz sencilla y elegante, la cual proporciona la información necesaria sobre la aplicación. La figura 3.21 muestra la pantalla de inicio de la aplicación web.

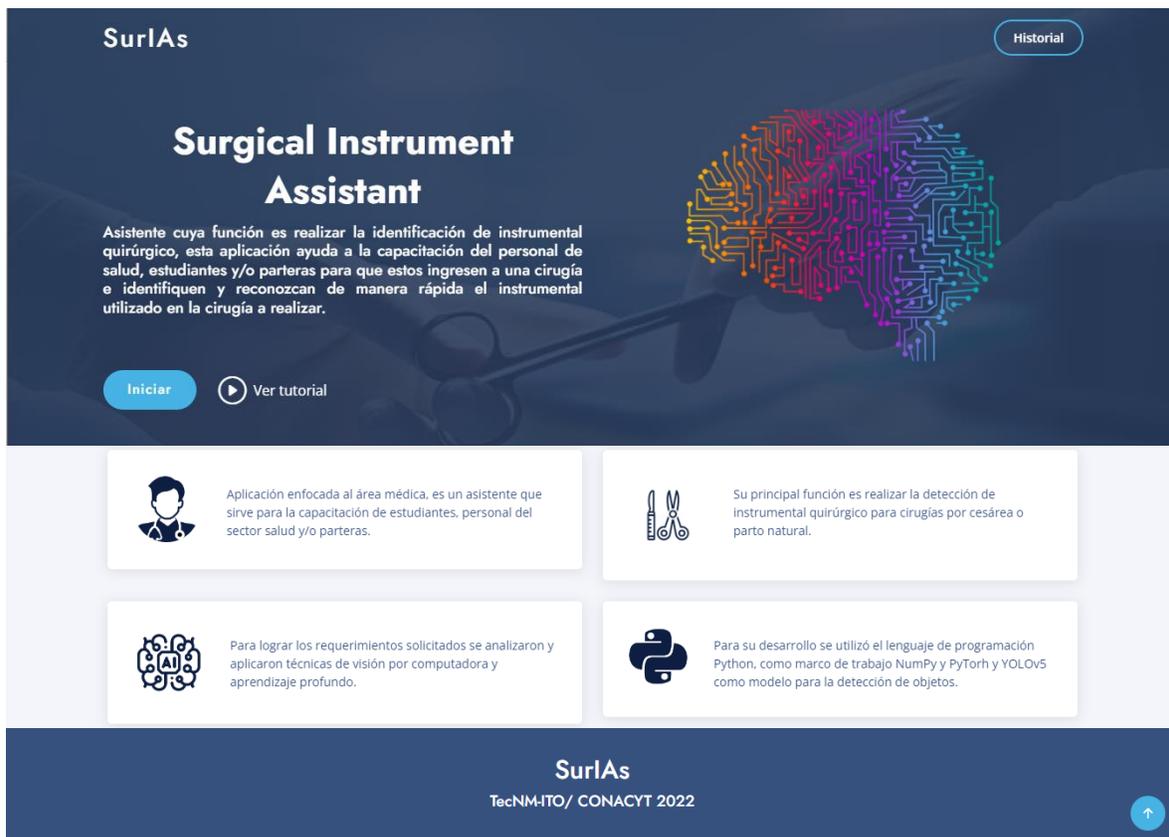


Figura 3.21 Interfaz gráfica de la pantalla principal.

3.4.2 Iniciar detección de objetos

Una vez que el usuario presiona el botón iniciar se mostrará la interfaz gráfica para realizar la identificación de objetos (Figura 3.22)

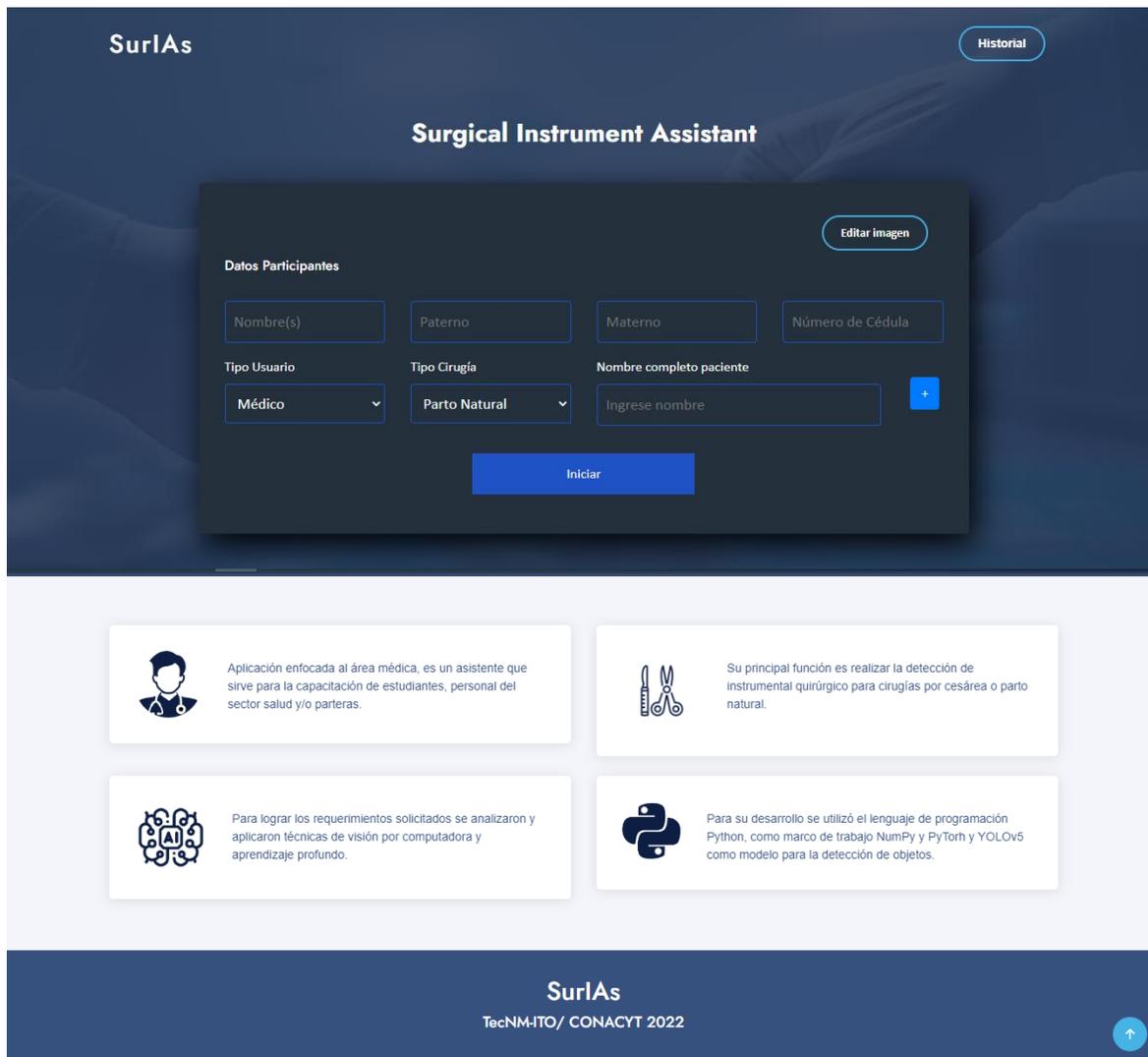


Figura 3.22 Interfaz gráfica de la pantalla principal.

Si el usuario desea editar la imagen a la cual se le aplicará la identificación de instrumental quirúrgico, será necesario que presione el botón editar imagen. La figura 3.23 muestra la página que servirá para realizar la edición de la imagen.

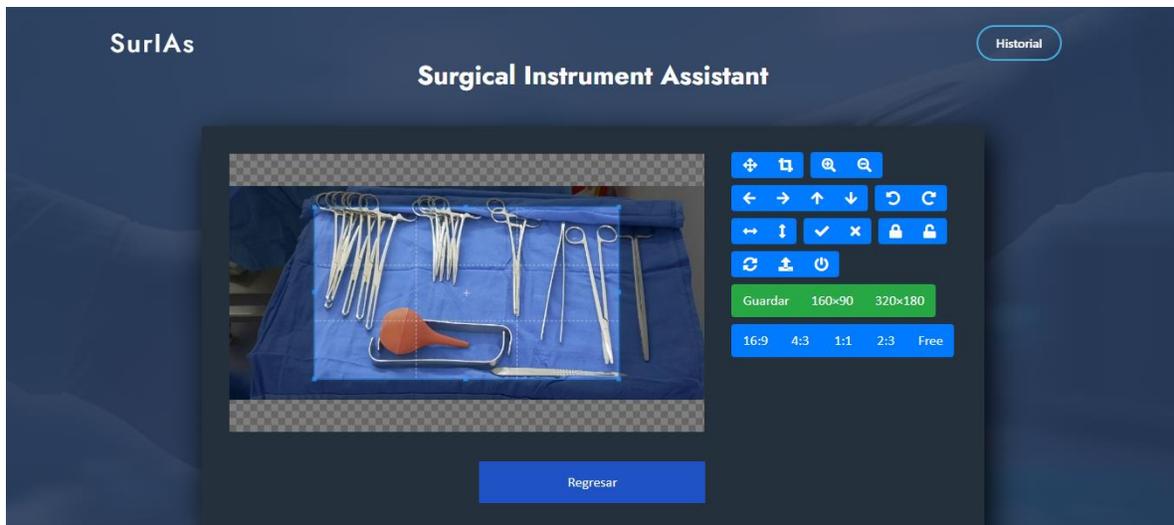


Figura 3.23 *Interfaz gráfica de editar imagen.*

Una vez finalizada la edición, el usuario tiene que regresar a la interfaz anterior presionando el botón regresar.

Antes de iniciar con la identificación de instrumental quirúrgico, será necesario que el usuario ingrese algunos datos, esto con el fin de tener un historial de cirugías realizadas. Los datos solicitados son: nombre, apellido paterno, apellido materno, número de cédula (solo si el participante es médico o enfermera), tipo usuario, tipo de cirugía y nombre del paciente (Figura 3.24). Una vez capturados todos los datos se tiene que presionar el botón de más (+), este proceso se realizará por cada persona que participe en la cirugía (médico, enfermera, asistente).

SurlAs Surgical Instrument Assistant Historial

Datos Participantes Editar imagen

Nombre(s) Paterno Materno Número de Cédula

Tipo Usuario Tipo Cirugía Nombre completo paciente

Enfermera Cesárea Esperanza López Cruz +

Lista de Participantes

Nombre	Número de Cédula	Tipo Usuario
Ivonne Juárez López	9562236	Enfermera

Iniciar

Figura 3.24 Interfaz gráfica que muestra el formulario de registro para los participantes de la cirugía.

Una vez que se han ingresado todos los datos de los participantes, se procede a presionar el botón iniciar y se mostrará el resultado de la identificación de instrumental quirúrgico (figura 3.25) en la cual se visualiza la imagen con el instrumental quirúrgico identificado, así como la lista con los nombres del instrumental identificado.

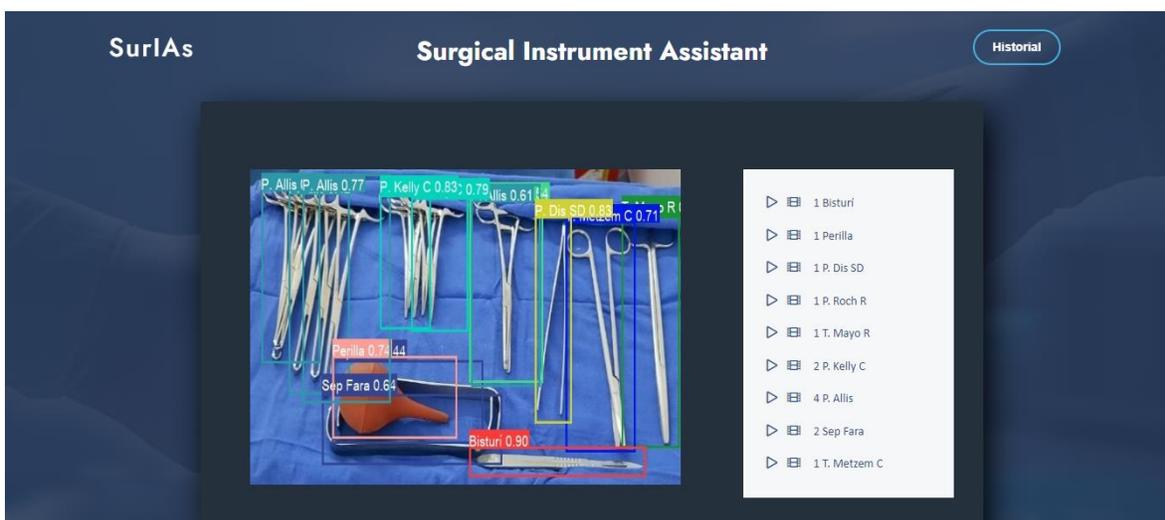


Figura 3.25 Interfaz gráfica que muestra el la identificación del instrumental quirúrgico.

Capítulo 3. Aplicación de la Metodología

Si el usuario desea mayor información del uso del instrumental quirúrgico identificado, solo bastara con que presione el icono de play e inmediatamente escuchara la descripción del instrumental. Finalmente, con el objetivo de que SurIAs sea una aplicación inclusiva se agregaron videos en lenguajes de señas los cuáles muestran la descripción de cada instrumento identificado. Para reproducirlo bastara con presionar el icono de video. La figura 3.26 muestra el resultado de la reproducción del video.

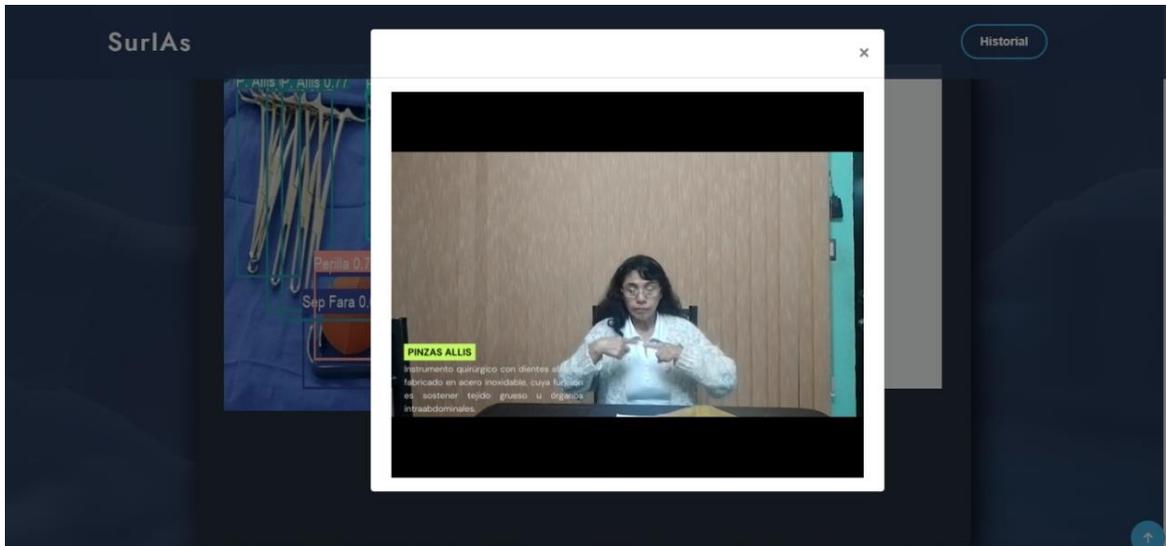


Figura 3.26 Descripción en lenguaje de señas del instrumental quirúrgico.

En caso de que no se haya ingresado algún campo el sistema no procede con la identificación de objetos y manda un mensaje de error (Figura 3.27).

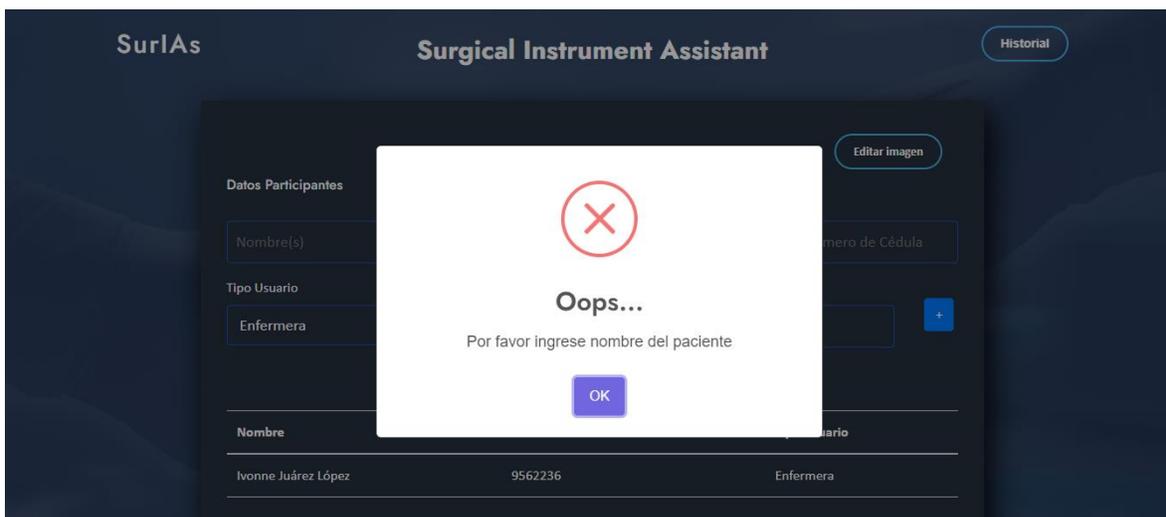


Figura 3.27 Interfaz gráfica que muestra mensaje de error.

Capítulo 4. Resultados

El objetivo de este capítulo es presentar los resultados obtenidos tras la elaboración del presente proyecto.

Se generó una aplicación web con interfaces elegantes y sencillas de entender, lo que permite que el usuario navegue fácilmente en ella. La cual ayuda al personal de salud a capacitar a estudiantes y/o parteras, ya que ayuda al reconocimiento de instrumental quirúrgico utilizado para partos naturales o cirugías por cesárea.

A continuación, se presenta el caso de estudio con el que se realizó la prueba de la funcionalidad de la solución.

4.1 Caso de estudio

Para generar las pruebas de la aplicación se contó con la colaboración de la enfermera Marisol Huerta Martínez con número de cédula 10445382 que actualmente labora como enfermera instrumentista en el Hospital General de Zona 33 IMSS de Tierra Blanca, Ver.

La enfermera comenta que el tipo de instrumental quirúrgico utilizado para las cirugías de parto por cesárea y parto natural varía de acuerdo a la técnica de cada ginecólogo, por lo tanto proporcionó la lista del instrumental quirúrgico más común utilizado en los procedimientos anteriormente mencionados. A continuación, se presenta la lista del instrumental quirúrgico proporcionado.

- Pinzas *Allis*
- Pinzas *Kelly* Curvas
- Pinzas *Rochester*
- Separadores *Farabeuf*
- Bisturí
- Perilla
- Tijeras *Metzemaum*
- Tijera de Mayo
- Pinza de Disección Con/Sin Dientes

Capítulo 4. Resultados

- Valva
- Pinza *Forester*
- Porta Agujas
- Pinza Campo

A su vez la enfermera explicó la técnica utilizada para cada procedimiento quirúrgico, el uso y funcionalidad de cada instrumental quirúrgico. Finalmente mostró cómo se acomoda el instrumental en una mesa de mayo.

Una vez que se tuvo conocimiento necesario acerca de los procedimientos quirúrgicos para el caso de estudio de este proyecto, la enfermera capturó imágenes de instrumental quirúrgico utilizado para cirugía por cesárea (Figura 4.1) y parto natural (Figura 4.2).



Figura 4.1 Instrumental quirúrgico utilizado para cirugía por cesárea.



Figura 4.2 Instrumental quirúrgico utilizado para parto natural.

Capítulo 4. Resultados

Una vez obtenidas las imágenes se le explicó a la enfermera el funcionamiento de la aplicación y procedió a realizar la prueba de la misma. Los resultados obtenidos se muestran a continuación (Figura 4.3, Figura 4.4 y Figura 4.5).



Figura 4.3 Identificación de objetos cirugía por cesárea.

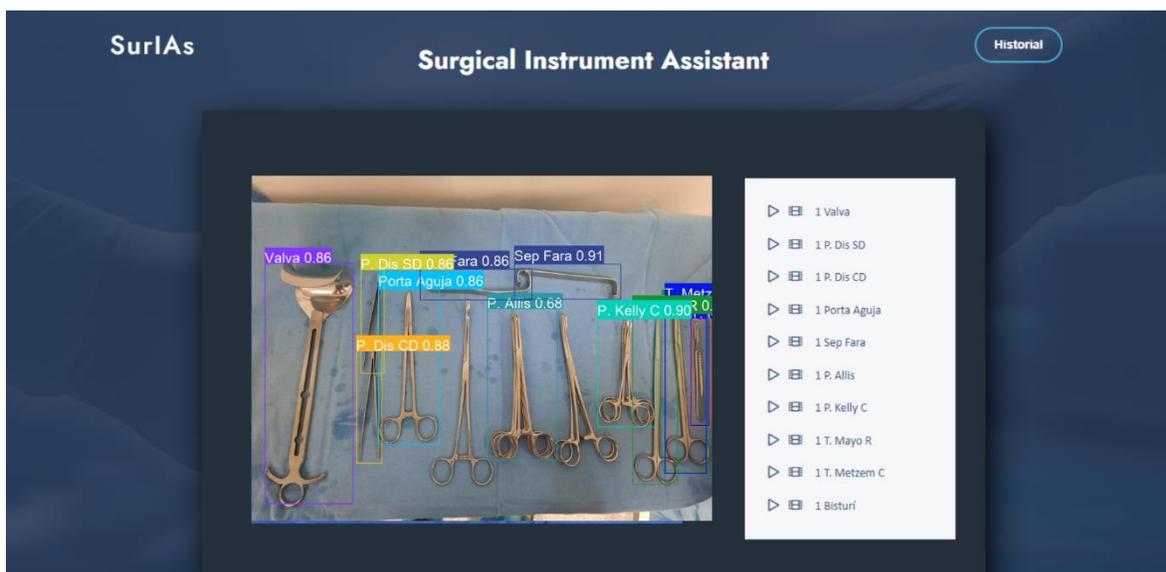


Figura 4.4 Identificación de objetos cirugía por cesárea.



Figura 4.5 Identificación de objetos cirugía por parto natural.

Una vez que la enfermera realizó la identificación de objetos, proporcionó por cada imagen la lista del instrumental quirúrgico que aparecía en ella. A continuación se presenta la lista del instrumental existente en la imagen y la lista que la aplicación identificó.

Tabla 4.1 Resultados obtenidos de la identificación de objetos.

#Figura	Lista de Instrumental real	Lista de instrumental identificado.	%
4.3	<ul style="list-style-type: none"> ● Porta Agujas ● Pinza Forester Curva ● Separadores Farabeuf ● Pinza Allis ● Pinza Rochester Recta ● Pinza Kelly Curva ● Tijera Metzembraum Curva ● Tijera de Mayo Recta. ● Bisturí ● Valva 	<ul style="list-style-type: none"> ● Porta Agujas ● Pinza Forester Curva ● Separadores Farabeuf ● Pinza Allis ● Pinza Rochester Recta ● Pinza Kelly Curva ● Tijera Metzembraum Curva ● Tijera de Mayo Recta. ● Bisturí 	90%
4.4	<ul style="list-style-type: none"> ● Valva ● Pinza de Disección con Dientes ● Pinza de Disección sin Dientes ● Porta Aguja ● Separadores Farabeuf ● Pinza Allis 	<ul style="list-style-type: none"> ● Valva ● Pinza de Disección con Dientes ● Pinza de Disección sin Dientes ● Porta Aguja ● Separadores Farabeuf ● Pinza Allis ● Pinza Kelly Curva ● Tijera de Mayo Recta 	90.9 %

#Figura	Lista de Instrumental real	Lista de instrumental identificado.	%
	<ul style="list-style-type: none"> • Pinzas Rochester Curva • Pinza Kelly Curva • Tijera de Mayo Recta • Tijera Metzembraum Curva • Bisturí 	<ul style="list-style-type: none"> • Tijera Metzembraum Curva • Bisturí 	
4.5	<ul style="list-style-type: none"> • Pinza Campo • Pinzas Kelly • Pinza de Disección con Dientes • Tijera de Mayo Recta • Porta Aguja • Tijera Metzembraum Curva • Pinza Rochester Curva 	<ul style="list-style-type: none"> • Pinza Campo • Pinzas Kelly • Pinza de Disección con Dientes • Tijera de Mayo Recta • Porta Aguja • Tijera Metzembraum Curva • Pinza Rochester Curva 	100%

De acuerdo a los resultados obtenidos, la enfermera considera que la identificación del instrumental quirúrgico es buena.

4.2 Evaluación y resultados

Para obtener una evaluación concisa de la aplicación desarrollada, se le aplicó a la enfermera un pequeño cuestionario, a continuación se presentan las respuestas proporcionadas,

Tabla 4.2 Evaluación de usabilidad de la aplicación.

Pregunta	Respuesta
1. ¿Qué le pareció la aplicación?	Entendible y fácil de usar. No está saturada de información.
2. La navegación en la aplicación ¿es fácil?	Sí
3. ¿Qué le parecieron los colores?	Adecuados y no lastiman la vista.
4. ¿Piensa que cualquier persona puede usar la aplicación?	Sí
5. ¿Piensa que la aplicación puede servir para capacitar a estudiantes?	Sí

Capítulo 4. Resultados

Pregunta	Respuesta
6. ¿Qué le pareció el resultado obtenido tras realizar la identificación de objetos?	Bueno, ya que identifiqué la mayoría del instrumental existente en las imágenes.
7. ¿Piensa que la aplicación puede ser de gran apoyo para los médicos de las zonas rurales?	Sí
8. ¿Qué le parece la sección de descripción de instrumental quirúrgico?	Es buena ya que las personas que no tienen conocimiento sobre el instrumental quirúrgico podrían saber más detalle del uso de estos.
9. ¿Tiene alguna duda?	No
10. ¿Tiene alguna sugerencia?	No

De acuerdo a la encuesta realizada y a los resultados obtenidos tras la realización del caso de estudio se concluye que SurIAs es de gran apoyo para personas que carecen de poco o nulo conocimiento sobre instrumental quirúrgico, ya que una vez cargada la imagen a la cual se le aplicará la identificación de instrumental quirúrgico la aplicación realiza la identificación de objetos en cuestión de segundos, además muestra la descripción de cada instrumental identificado con el fin de que el usuario conozca más sobre el uso de cada uno.

Por otro lado SurIAs es intuitiva y no se necesita realizar mayor capacitación en los usuarios.

Capítulo 5. Conclusiones y recomendaciones

El objetivo de este capítulo es presentar las conclusiones y recomendaciones sobre la solución generada para resolver la problemática planteada.

5.1 Conclusiones

Con el proceso del desarrollo de la tesis se lograron cumplir el objetivo general y los objetivos específicos, los cuales consistían en aplicar técnicas de Visión por Computadora y *Deep Learning* para la realización de un análisis comparativo que permitiera el desarrollo de una aplicación Web para la identificación de instrumental quirúrgico.

Para lograr con éxito el objetivo específico se realizó un análisis extenso de trabajos relacionados con la identificación de objetos, posteriormente se identificaron las técnicas de visión por computadora que se implementarían para llevar a cabo el análisis de la identificación de objetos.

Ya que existen diferentes tipos de cirugías y cada una tiene instrumental quirúrgico específico para realizar dicho procedimiento, se procedió a delimitar el instrumental quirúrgico. A su vez se analizaron las tecnologías que se utilizarían para el entrenamiento del modelo y el desarrollo de la aplicación, una vez identificadas se realizó un análisis comparativo de diferentes algoritmos tales como Faster RCNN, RetinaNet, YoloV4 y YoloV5, donde el algoritmo que dio mejor resultado fue YOLO v5 ya que obtuvo un mAP de 0.757.

Una vez elegido el algoritmo más eficiente, se realizaron n cantidad de entrenamientos con diferentes conjuntos de datos de instrumental quirúrgico utilizado para cirugía y parto natural.

Finalmente se desarrolló una aplicación web la cual hace uso del modelo entrenado mediante una API. Dicha aplicación se bautizó con el nombre de SurIAs por sus siglas en inglés *Surgical Instrument Assistant*. Cuya función es servir como asistente virtual, esta ayudará a que cualquier persona con poco o nulo conocimiento reconozca de manera fácil y rápida el instrumental quirúrgico utilizado durante una cirugía por cesárea o parto natural. Por lo tanto SurIAs beneficiará tanto al sector salud como al sector educativo, ya que por medio de la

aplicación se les podrá mostrar a los estudiantes el instrumental quirúrgico utilizado para las cirugías anteriormente mencionadas, además de que estos podrán escuchar las descripciones de cada instrumental quirúrgico las veces que ellos quieran.

5.2 Recomendaciones

A continuación, se presenta una lista de recomendaciones para futuros trabajos o para continuar con el desarrollo del mismo.

1. Entrenar el modelo para que pueda reconocer instrumental quirúrgico utilizado para otras cirugías.
2. Agregar la opción para descargar el historial de cirugías.
3. Para que el modelo realice una correcta identificación de instrumental quirúrgico se recomienda que la imagen sea recortada, de forma que solo se visualice el área de interés.

Productos académicos

El objetivo de esta sección es proporcionar los productos académicos que se han realizado.

6.1 Artículo de congreso

SurIAss: An architecture for the detection of cesarean section surgical instruments using convolutional neural networks.

Ivonne Juárez López¹, José Luis Sánchez Cervantes², Luis Rolando Guarneros Nolasco¹, Giner Alor Hernández¹, Beatriz Alejandra Olivares Zepahua¹



6.2 Retribución social



6.3 Registro de derecho de autor

Anexos

Cuestionario de la iteración

1. ¿Qué le pareció la aplicación?
2. La navegación en la aplicación ¿es fácil?
3. ¿Qué le parecieron los colores?
4. ¿Piensa que cualquier persona puede usar la aplicación?
5. ¿Piensa que la aplicación puede servir para capacitar a estudiantes?
6. ¿Qué le pareció el resultado obtenido tras realizar la identificación de objetos?
7. ¿Piensa que la aplicación puede ser de gran apoyo para los médicos de las zonas rurales?
8. ¿Qué le parece la sección de descripción de instrumental quirúrgico?
9. ¿Tiene alguna duda?
10. ¿Tiene alguna sugerencia?

Referencias bibliográficas

- [1] D. Mauricio *et al.*, “Procedimientos de enfermería en el área quirúrgica edimec 2015.”
- [2] S. Carlos Schnapp, S. Eduardo Sepúlveda, and S. Jorge Andrés Robert, “Operación cesárea,” *Rev. Médica Clínica Las Condes*, vol. 25, no. 6, pp. 987–992, Nov. 2014, doi: 10.1016/S0716-8640(14)70648-0.
- [3] León Puy, “Procedimiento quirúrgico enfermero en la cesarea urgente,” *Univ. Zaragoza*, 2012.
- [4] M. Viciania Sánchez Tutora and D. Carolina González Hernando, “Universidad de Valladolid Facultad de Enfermería GRADO EN ENFERMERÍA EL PARTO NATURAL.”
- [5] “¿Qué instrumentos se utilizan en el parto? | CuidatePlus.” <https://cuidateplus.marca.com/reproduccion/parto/2017/01/08/-instrumentos-utilizan-parto-137604.html> (accessed Aug. 11, 2021).
- [6] E. Abadal Barcelona, “Acceso abierto a la ciencia,” 2012.
- [7] L. Salvo, “La construcción de un repositorio digital de imágenes fotográficas desde un abordaje comunitario y participativo: la experiencia de la Fototeca Digital de Ciencias Humanas,” *Masquedós*, no. 1, pp. 25–35, 2016.
- [8] “INTELIGENCIA ARTIFICIAL 101 COSAS QUE DEBES SABER HOY SOBRE NUESTRO FUTURO INTELIGENCIA ARTIFICIAL,” 2018, Accessed: Aug. 06, 2021. [Online]. Available: www.planetadelibros.com.
- [9] U. Enfoque, “Internet de las Cosas y Visión Artificial, Funcionamiento y Aplicaciones: Revisión de Literatura (Internet of Things and Artificial Vision, Performance and Applications: Literature Review),” no. 1, pp. 244–256, 2017, Accessed: Aug. 06, 2021. [Online]. Available: <http://ingenieria.ute.edu.ec/enfoqueute/>.
- [10] “Las 6 mejores técnicas y algoritmos de visión por computadora que cambian la percepción del mundo - Skywell Software.” <https://skywell.software/blog/top-6-computer-vision-techniques-and-algorithms/> (accessed Aug. 06, 2021).
- [11] J. A. Cancelas, R. C. González, I. Álvarez, and J. M. Enguita, *Procesamiento*

Morfológico, Visión 3D: Estereoscopia, Álgebra lineal básica para visión por computador, Geometría Proyectiva para Visión 3D, vol. 1. 2016.

- [12] M. Naveenkumar, “OpenCV for Computer Vision Applications,” no. April, 2016.
- [13] “Introducción al algoritmo YOLO para la detección de objetos | Programa de Educación en Ingeniería (EngEd) | Sección.” <https://www.section.io/engineering-education/introduction-to-yolo-algorithm-for-object-detection/> (accessed Aug. 10, 2021).
- [14] A. R. Pathak, M. Pandey, and S. Rautaray, “Application of Deep Learning for Object Detection,” *Procedia Comput. Sci.*, vol. 132, pp. 1706–1717, Jan. 2018, doi: 10.1016/J.PROCS.2018.05.144.
- [15] “Data Labeling of Images for Supervised Learning - Landing AI.” <https://landing.ai/data-labeling-of-images-for-supervised-learning/> (accessed Aug. 11, 2021).
- [16] “Cómo calcular la precisión, el recuerdo y la medida F para una clasificación desequilibrada.” <https://machinelearningmastery.com/precision-recall-and-f-measure-for-imbalanced-classification/> (accessed Nov. 05, 2021).
- [17] “Precisión media media (p.mAP) explicada | Paperspace Blog.” <https://blog.paperspace.com/mean-average-precision/> (accessed May 17, 2022).
- [18] A. E. Bradley, “The use of the area under the {ROC} curve in the evaluation of machine learning algorithms,” *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, 1997.
- [19] “Intersección sobre unión (IoU) para la detección de objetos - PyImageSearch.” <https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/> (accessed Nov. 04, 2021).
- [20] “¿Qué es el sobreajuste u overfitting y por qué debemos evitarlo?” <https://machinelearningparatodos.com/que-es-el-sobreajuste-u-overfitting-y-por-que-debemos-evitarlo/> (accessed May 17, 2022).
- [21] “Introducción a Sobreajuste y Subajuste para Machine Learning - Aprende IA.” <https://aprendeia.com/sobreajuste-y-subajuste-en-machine-learning/> (accessed May 17, 2022).
- [22] L. Liu *et al.*, “Deep Learning for Generic Object Detection : A Survey,” *Int. J.*

- Comput. Vis.*, vol. 128, no. 2, pp. 261–318, 2020, doi: 10.1007/s11263-019-01247-4.
- [23] “Supervised Deep Learning Algorithms : Types and Applications.”
<https://www.analyticsvidhya.com/blog/2021/05/introduction-to-supervised-deep-learning-algorithms/> (accessed Aug. 10, 2021).
- [24] “Redes neuronales. Programa de Visión... | by Bootcamp AI | Medium.”
<https://bootcampai.medium.com/redes-neuronales-13349dd1a5bb> (accessed Aug. 10, 2021).
- [25] R. Yamashita, M. Nishio, R. Kinh, G. Do, and K. Togashi, “Convolutional neural networks : an overview and application in radiology,” pp. 611–629, 2018.
- [26] Y. Zhang *et al.*, “Understanding of Object Detection Based on CNN Family and YOLO Related content Real-time vehicle detection and tracking in video based on faster R-CNN A Real-Time Method to Estimate Speed of Object Based on Object Detection and Optical Flow Calculation Temporal Classification Error Compensation of Convolutional Neural Network for Traffic Sign Recognition This content was downloaded from IP address Understanding of Object Detection Based on CNN Family and YOLO,” *J. Phys. Conf. Ser.*, vol. 1004, p. 12029, 2018, doi: 10.1088/1742-6596/1004/1/012029.
- [27] H. Yanagisawa, T. Yamashita, and H. Watanabe, “A Study on Object Detection Method from Manga Images using CNN.”
- [28] “Label objects for deep learning—ArcGIS Pro | Documentación.”
<https://pro.arcgis.com/es/pro-app/latest/help/analysis/image-analyst/label-objects-for-deep-learning.htm> (accessed Aug. 11, 2021).
- [29] L. Jiao, F. A. N. Zhang, F. Liu, and S. Member, “A Survey of Deep Learning-Based Object Detection,” *IEEE Access*, vol. 7, pp. 128837–128868, 2019, doi: 10.1109/ACCESS.2019.2939201.
- [30] “México tiene un déficit de 115 mil enfermeras.”
<https://www.reporteindigo.com/reporte/mexico-tiene-un-deficit-de-115-mil-enfermeras/> (accessed Aug. 18, 2021).
- [31] P. N. Druzhkov and V. D. Kustikova, “A survey of deep learning methods and software tools for image classification and object detection,” *Pattern Recognit. Image Anal.*, vol. 26, no. 1, pp. 9–15, Jan. 2016, doi: 10.1134/S1054661816010065.

- [32] X. Zhou, W. Gong, W. Fu, and F. Du, "Application of Deep Learning in Object Detection," pp. 631–634, 2017.
- [33] A. Dhillon and G. K. Verma, "Convolutional neural network : a review of models , methodologies and applications to object detection," *Prog. Artif. Intell.*, vol. 9, no. 2, pp. 85–112, 2020, doi: 10.1007/s13748-019-00203-0.
- [34] R. Wu, M. Feng, W. Guan, D. Wang, H. Lu, and E. Ding, "A Mutual Learning Method for Salient Object Detection with intertwined," pp. 8150–8159.
- [35] Y. Xiao, Z. Tian, D. X. Lan, J. Yu, Y. Zhang, and S. Liu, *A review of object detection based on deep learning*. Multimedia Tools and Applications, 2020.
- [36] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep Learning for Computer Vision : A Brief Review," vol. 2018, 2018.
- [37] S. Schneider, G. W. Taylor, and S. C. Kremer, "Deep Learning Object Detection Methods for Ecological Camera Trap Data," *2018 15th Conf. Comput. Robot Vis.*, pp. 321–328, 2018, doi: 10.1109/CRV.2018.00052.
- [38] M. Shah, "Object Detection Using Deep Neural Networks," pp. 787–790, 2017.
- [39] X. Wu, D. Sahoo, and S. C. H. Hoi, "Neurocomputing Recent advances in deep learning for object detection," *Neurocomputing*, vol. 396, pp. 39–64, 2020, doi: 10.1016/j.neucom.2020.01.085.
- [40] H. Zhu, X. Chen, W. Dai, K. Fu, Q. Ye, and J. Jiao, "ORIENTATION ROBUST OBJECT DETECTION IN AERIAL IMAGES USING DEEP CONVOLUTIONAL NEURAL NETWORK School of Electronic , Electrical and Communication Engineering University of Chinese Academy of Sciences , Beijing , China Institute of Electronics , Chinese Aca," pp. 3735–3739, 2015.
- [41] H. Hu, J. Gu, Z. Zhang, and J. Dai, "Relation Networks for Object Detection," pp. 3588–3597.
- [42] B. Wiestler and B. Menze, "Neuro-Oncology Advances Deep learning for medical image analysis : a brief introduction," vol. 2, pp. 35–41, 2020, doi: 10.1093/noajnl/vdaa092.
- [43] M. Liang and X. Hu, "Recurrent Convolutional Neural Network for Object Recognition," no. Figure 1.