



**ESCUELA UNIVERSITARIA DE POSGRADO**

**DELIMITACIÓN DE ÁREAS AFECTADAS POR INCENDIOS FORESTALES  
MEDIANTE APRENDIZAJE PROFUNDO EN IMÁGENES SATELITALES**

**Línea de investigación:**

**Ingeniería de software, simulación y desarrollo de TICs**

Tesis para optar el grado académico de Maestro en Ingeniería de Sistemas

**Autor**

Quispe Varillas, Javier José

**Asesor**

Zamora Talaverano, Noe Sabino

ORCID: 0000-0002-4368-8955

**Jurado**

Coveñas Lalupu, José

Petrlik Azabache, Ivan Carlo

Peña Carrillo, César Serapio

**Lima - Perú**

**2024**

# DELIMITACIÓN DE ÁREAS AFECTADAS POR INCENDIOS FORESTALES MEDIANTE APRENDIZAJE PROFUNDO EN IMÁGENES SATELITALES

## INFORME DE ORIGINALIDAD

27%

INDICE DE SIMILITUD

25%

FUENTES DE INTERNET

5%

PUBLICACIONES

14%

TRABAJOS DEL ESTUDIANTE

## FUENTES PRIMARIAS

1	<b>torres.ai</b> Fuente de Internet	6%
2	<b>hdl.handle.net</b> Fuente de Internet	2%
3	<b>Submitted to Pontificia Universidad Catolica del Peru</b> Trabajo del estudiante	2%
4	<b>medium.com</b> Fuente de Internet	1%
5	<b>es.scribd.com</b> Fuente de Internet	1%
6	<b>www.ign.es</b> Fuente de Internet	1%
7	<b>repositorio.unfv.edu.pe</b> Fuente de Internet	1%
8	<b>www.slideshare.net</b> Fuente de Internet	1%



Universidad Nacional  
**Federico Villarreal**

**VRIN** | VICERRECTORADO  
DE INVESTIGACIÓN

ESCUELA UNIVERSITARIA DE POSGRADO  
DELIMITACIÓN DE ÁREAS AFECTADAS POR INCENDIOS FORESTALES  
MEDIANTE APRENDIZAJE PROFUNDO EN IMÁGENES SATELITALES

Línea de investigación:

Ingeniería de software, simulación y desarrollo de TICs

Tesis para optar el Grado Académico de Maestro en Ingeniería de Sistemas

Autor:

Quispe Varillas, Javier José

Asesor:

Zamora Talaverano, Noe Sabino

ORCID:0000-0002-4368-8955

Jurado:

Coveñas Lalupu, José

Petrlik Azabache, Ivan Carlo

Peña Carrillo, César Serapio

**Lima - Perú**

**2024**

## ÍNDICE

<b>Resumen</b>	X
<b>Abstract</b>	XI
<b>I Introducción</b>	1
1.1 Planteamiento del Problema . . . . .	2
1.2 Descripción del Problema . . . . .	2
1.3 Formulación del Problema . . . . .	4
1.3.1 Problema General . . . . .	4
1.3.2 Problemas Específicos . . . . .	4
1.4 Antecedentes . . . . .	4
1.5 Justificación de la Investigación . . . . .	9
1.6 Limitaciones de la Investigación . . . . .	10
1.7 Objetivos de la Investigación . . . . .	10
1.7.1 Objetivo General . . . . .	10
1.7.2 Objetivos Específicos . . . . .	10
1.8 Hipótesis . . . . .	10
1.8.1 Hipótesis General . . . . .	10
1.8.2 Hipótesis Específicos . . . . .	11
<b>II Marco Teórico</b>	12
2.1 Monitoreo Satelital - SERFOR . . . . .	12

2.2	Teledetección .....	14
2.2.1	Introducción.....	14
2.2.2	Satélites de Observación Terrestre.....	19
2.3	Machine Learning .....	26
2.4	Deep Learning .....	28
2.5	Redes Neuronales Convolucionales .....	33
2.5.1	Introducción.....	33
2.5.2	Componentes .....	36
2.6	Segmentación Semántica.....	44
2.6.1	Introducción.....	44
2.6.2	Arquitecturas .....	47
2.6.3	Métricas de validación.....	50
<b>III</b>	<b>Método</b> .....	<b>55</b>
3.1	Tipo de Investigación .....	55
3.2	Población y muestra .....	55
3.3	Operacionalización de variables.....	55
3.4	Instrumentos .....	57
3.5	Procedimientos .....	57
3.6	Análisis de datos .....	58

	IV
<b>IV Resultados</b>	60
<b>V Discusión de resultados</b>	90
<b>VI Conclusiones</b>	93
<b>VII Recomendaciones</b>	95
<b>VIII Referencias</b>	97
<b>IX Anexos</b>	104
9.1 Matriz de consistencia .....	104
9.2 Validación y confiabilidad de instrumentos .....	105
9.3 Arquitectura del Modelo UNet propuesto .....	107
9.4 Listado de imágenes satelitales utilizadas .....	108

**ÍNDICE DE TABLAS**

1	Registro Histórico de Áreas Afectadas (Cicatrices) de Incendios Forestales por Departamento .....	14
2	Resolución radiométrica - Niveles Digitales .....	21
3	Bandas Espectrales de Landsat-8 y 9 .....	22
4	Bandas Espectrales de Sentinel-2.....	24
5	Criterios de grado de concordancia basado en índice Kappa .....	54
6	Cicatrices de RIF - 2017 a 2021 .....	60
7	Equivalencia de Bandas Espectrales - S2 y L8.....	61
8	Escenas del sensor Sentinel-2.....	61
9	Tiles para Aprendizaje Supervisado .....	63
10	Datos Grupo 1 - G1: Proporción de tiles con máscaras 15.3 %.....	63
11	Datos Grupo 2 - G2: Proporción de tiles con máscaras 20 %.....	64
12	Datos Grupo 3 - G3: Proporción de tiles con máscaras 25 %.....	64
13	Combinación de Bandas Sentinel-2.....	64
14	Métrica: Precision - Grupo G1 .....	68
15	Métrica: Precision - Grupo G2 .....	68
16	Métrica: Precision - Grupo G3 .....	69
17	Métrica: F1-Score - Grupo G1 .....	70
18	Métrica: F1-Score - Grupo G2 .....	70

19	Métrica: F1-Score - Grupo G3 .....	70
20	Métrica: IoU (índice Jaccard)- Grupo G1.....	71
21	Métrica: IoU (índice Jaccard)- Grupo G2.....	71
22	Métrica: IoU (índice Jaccard) - Grupo G3.....	71
23	Métrica: Recall o Sensibilidad - Grupo G1 .....	72
24	Métrica: Recall o Sensibilidad - Grupo G2 .....	72
25	Métrica: Recall o Sensibilidad - Grupo G3 .....	72
26	Métrica: Coeficiente Kappa - Grupo G1.....	73
27	Métrica: Coeficiente Kappa - Grupo G2.....	73
28	Métrica: Coeficiente Kappa - Grupo G3.....	73
29	Resumen Comparativo de Métricas - Promedio por subgrupos de 3 Bandas (3B) y 6 Bandas (6B) - Datos de Prueba .....	76

## ÍNDICE DE FIGURAS

1	Registro Individual de Cicatriz de Incendio Forestal .....	15
2	Componentes de un sistema de teledetección.....	17
3	Historia de los sistemas de teledetección.....	18
4	Espectro electromagnético.....	19
5	Espectro Electromagnético de Landsat-8 y 9 .....	23
6	Espectro Electromagnético de Sentinel-2, Landsat-7,8 .....	25
7	Procesos del Aprendizaje Automático.....	27
8	Deep Learning, parte de Machine Learning, y a su vez de la Inteligencia Artificial	29
9	Deep Learning - red neuronal de una o más capas ocultas (hidden).....	30
10	Perceptrón .....	31
11	Aprendizaje de una CNN .....	35
12	Esquema de una CNN .....	36
13	Conexión de neurona de capa oculta con capa de entrada.....	38
14	Entrada de 28x28 con ventana de 5x5 que genera una salida de 24x24 neuronas en la capa oculta .....	39
15	Capa CNN compuesta por varios filtros.....	40
16	Capa Pooling construida desde una ventana 2x2 que se aplica a una capa CNN de 24x24.....	41
17	Transformación de pooling mantiene la proporción inicial .....	42

18	Capa CNN y pooling .....	43
19	Imagen y su segmentación semántica.....	45
20	Segmentación de instancia .....	46
21	Red Autoencoder.....	48
22	Red Unet.....	49
23	Matriz de Confusión Binaria .....	51
24	Métrica IoU - Índice de Jaccard .....	52
25	Ejemplo de métrica IoU .....	52
26	Área afectada - Provincia: Cusco / Quispicanchis - Grupo G1 y G2.....	74
27	Área afectada - Provincia: Cusco / Quispicanchis / Chumbivilcas - Grupo G3 .	75
28	Matriz de Confusión - Grupo G1 - D1 y D2.....	78
29	Matriz de Confusión - Grupo G1 - D3 y D4.....	78
30	Matriz de Confusión - Grupo G1 - D5 y D6.....	79
31	Matriz de Confusión - Grupo G2 - D1 y D2.....	79
32	Matriz de Confusión - Grupo G2 - D3 y D4.....	80
33	Matriz de Confusión - Grupo G2 - D5 y D6.....	80
34	Matriz de Confusión - Grupo G3 - D1 y D1-AD (Aumento de Datos) .....	81
35	Matriz de Confusión - Grupo G3 - D2 y D2-AD (Aumento de Datos) .....	81
36	Matriz de Confusión - Grupo G3 - D3 y D3-AD (Aumento de Datos) .....	82

37	Matriz de Confusión - Grupo G3 - D4 y D4-AD (Aumento de Datos) .....	82
38	Matriz de Confusión - Grupo G3 - D5 y D5-AD (Aumento de Datos) .....	83
39	Matriz de Confusión - Grupo G3 - D6 y D6-AD (Aumento de Datos) .....	83
40	Comparación visual de resultados - Grupo G1 - D1 a D3 .....	84
41	Comparación visual de resultados - Grupo G1 - D4 a D6 .....	85
42	Comparación visual de resultados - Grupo G2 - D1 a D3 .....	86
43	Comparación visual de resultados - Grupo G2 - D4 a D6 .....	87
44	Comparación visual de resultados - Grupo G3 - D1 a D3 .....	88
45	Comparación visual de resultados - Grupo G3 - D4 a D6 .....	89

## RESUMEN

El mapeo de las áreas afectadas por incendios forestales (cicatriz) es una tarea importante en nuestro territorio. Esta delimitación de superficies afectadas es crucial para conocer la magnitud real de los incendios con el objetivo de construir políticas adecuadas de prevención, reducción del riesgo, y restauración. El avance en la clasificación de imágenes satelitales mediante algoritmos de aprendizaje profundo basado en redes neuronales convolucionales junto con la gran cantidad de datos de sensores de acceso libre nos permite ámbitos de investigación. Este estudio plantea un modelo de segmentación semántica para delimitar las áreas afectadas producto de los incendios forestales en el departamento de Cusco. Se generó una base de datos de 2898 imágenes (444 con áreas afectadas) del sensor Sentinel-2 de los años 2017 a 2021, lo cual nos brinda una variación anual de los ecosistemas afectados. Se formaron 18 conjuntos de datos donde se distribuyeron un 64 % para entrenamiento, 16 % para validación y 20 % para prueba, tanto para imágenes con o sin áreas afectadas. Los resultados obtenidos para el modelo nos brindan valores promedio: *Accuracy* de 99 %, *Precision* de 92 %, *Recall* de 91 %, *IoU* de 0.84, *F1-Score* de 0.91, y *Kappa* de 0.87 para los datos de prueba. Con lo cual se valida que el modelo propuesto es una herramienta efectiva para la delimitación de áreas afectadas producto de los incendios forestales.

**Palabras clave:** segmentación semántica, aprendizaje profundo, redes neuronales convolucionales, cicatriz, incendios forestales

## ABSTRACT

Mapping affected areas by forests fires (scar) is an important task in our country. This delimitation of affected areas is crucial to know the real magnitude of the fires with the aim of building appropriate prevention, risk reduction, and restoration policies. Progress in the classification of satellite images using algorithms deep learning based on convolutional neural networks along with the large amount of open access sensor data allows us to research areas. This study proposes a semantic segmentation model to delimit affected areas from forest fires in the department of Cusco. A database of 2898 images was generated (444 with affected areas) of the Sentinel-2 sensor from 2017 to 2021, which gives us a variation of affected ecosystems. 18 data sets were formed where a 64 % for training, 16 % for validation and 20 % for testing, both for images with or without affected areas. The results obtained for the model give us average values: *Accuracy* of 99 %, *Precision* of 92 %, *Recall* of 91 %, *IoU* of 0.84, *F1-Score* of 0.91, and *Kappa* of 0.87 for the test data. This validates that the proposed model is a tool effective for the delimitation of affected areas resulting from forest fires.

**Keywords:** semantic segmentation, deep learning, convolutional neural networks, scar, forest fires.

## I. INTRODUCCIÓN

Los incendios forestales son considerados como uno de los peligros priorizados en el Sistema Nacional de Gestión de Riesgos de Desastres (SINAGERD), para lo cual las medidas y acciones implementadas para la reducción de la vulnerabilidad de la población y sus medios de vida ante el riesgo de desastres se integran a través de los procesos de estimación, prevención, reducción del riesgo, preparación, respuesta, rehabilitación y reconstrucción / restauración. (Ley N° 29664; Decreto Supremo N° 048-2011-PCM; Decreto Supremo N° 111-2012-PCM; Resolución Ministerial N° 046-2013-PCM; Decreto Supremo N° 034-2014-PCM; Decreto Supremo N° 115-2022-PCM)

En el Perú, los incendios forestales representan un riesgo para la población y sus medios de vida, así como para los ecosistemas vulnerables y sensibles al fuego, causando en los últimos una afectación y consecuente pérdida de cobertura natural importante. Dentro de este marco, el Servicio Nacional Forestal y de Fauna Silvestre (SERFOR) (Ley N° 29763), como entidad rectora a nivel nacional de la gestión sostenible de la flora y fauna silvestre (artículo 14 inciso d), establece y lidera una serie de acciones en todos los procesos de la Gestión de Riesgos de Desastres (GRD) de los incendios forestales (artículo 24), entre ellos la restauración. Igualmente, en la Octava Disposición complementaria de Ley Marco del Cambio Climático (Ley N° 30754), se señala que el SERFOR y las Autoridades Regionales Forestales y de Fauna Silvestre incorporan la gestión del riesgo de desastres a la gestión del patrimonio forestal y de fauna silvestre, priorizando la implementación de un Plan de Prevención y Control de Incendios Forestales. Finalmente, mencionamos que el SERFOR es responsable del modelo operacional del producto N°1: Actividad 6: generación de información y monitoreo de incendios forestales correspondiente al programa presupuestal 0068. (Decreto de Urgencia N° 024-2010)

En el marco del proceso de restauración de las áreas afectadas por los incendios

forestales, uno de los insumos más importantes es la delimitación de dichas áreas. En este aspecto, las imágenes satelitales proporcionadas en forma gratuita por entidades internacionales nos permiten obtener información sólida y validada para la toma de acciones.

Al respecto, múltiples técnicas para la determinación de áreas afectadas (superficies quemadas) se han desarrollado, entre ellas se encuentra las enmarcadas dentro del “Aprendizaje Automático” (*Machine Learning*).

En los últimos años, ha surgido una rama del *Machine Learning* denominada “Aprendizaje Profundo” (*Deep Learning*), la cual tiene como base el uso de las denominadas “Redes Neuronales Artificiales” (RNA). En base a esto, planteamos una metodología alternativa utilizando las RNA para delimitar las áreas afectadas producto de los incendios forestales en nuestro país. (Knopp et al., 2020; Hoeser and Kuenzer, 2020; Hoeser et al., 2020)

## **1.1 Planteamiento del Problema**

Actualmente el país no cuenta con información aproximada de superficie de área afectada producto de los incendios forestales en un corto tiempo, ya que actualmente la información se determina al culminar la temporada de incendios, lo que genera graves consecuencias para la toma de decisiones de las autoridades como el SERFOR, así como las Autoridades Regionales Forestales y de Fauna Silvestre (ARFFS). Contar con información validada y a escala temporal adecuada, permitirá diseñar medidas apropiadas de restauración a través de lineamientos acorde al tipo de ecosistema y su asociación al uso del fuego, así como implementar las acciones de restauración oportunas.

## **1.2 Descripción del Problema**

A nivel mundial, el cambio climático permitirá que las temporadas de incendios sean cada vez más largas e intensas, producto del aumento de la temperatura y los cambios en las dinámicas climáticas.

En nuestro país se ha visto un aumento en el mayor número e intensidad de incendios, producto de la ocurrencia de sequías cada vez más frecuentes, pero principalmente por la acción del hombre para actividades relacionadas a la agricultura y ganadería. Debido a los factores climáticos, el uso del fuego a través de las quemadas se transforma en incendios difíciles de sofocar en las regiones de la sierra, donde los bosques relictos andinos y los pastizales son los más afectados; en la costa, donde los bosques secos y húmedos son ecosistemas vulnerables al impacto del fuego; y en nuestra región amazónica, donde el impacto del fuego ha generado cambios drásticos en los bosques húmedos tropicales.

En el Perú, es el SERFOR la autoridad nacional que vela por la gestión sostenible de los recursos forestales y de fauna silvestre a nivel nacional, para ello implementó el año 2021, la Unidad Funcional de Monitoreo Satelital (UFMS) (Resolución de Gerencia General N°006-2021-MIDAGRI-SERFOR-GG) dentro de la Dirección General de Información y Ordenamiento Forestal y de Fauna Silvestre (DGIUFFS), que entre sus funciones se encuentra el monitoreo continuo del patrimonio forestal y de fauna silvestre, y la generación de reportes oficiales de afectación. Uno de los drivers de afectación que se monitorean son los incendios forestales, para lo cual diseñaron un reporte que permite identificar las cicatrices (áreas afectadas) de incendios forestales que aproximan la superficie quemada para su distribución a las autoridades regionales y locales. La UFMS es responsable de realizar un análisis técnico y preciso del monitoreo de los incendios forestales, incluyendo la determinación de áreas afectadas, mediante el uso de diversas plataformas de alertas tempranas y el procesamiento y análisis de imágenes satelitales, y su interacción con el patrimonio forestal, catastro forestal y otras áreas categorizadas. La generación de las delimitaciones de las áreas afectadas se realiza a partir de mosaicos anuales de imágenes satelitales.

### 1.3 Formulación del Problema

Tomando en consideración lo visto en el planteamiento y la descripción del problema, podemos mencionar las siguientes preguntas:

#### 1.3.1 *Problema General*

- ¿Cómo puedo obtener en tiempo oportuno al incendio forestal la delimitación del área afectada utilizando técnicas de aprendizaje profundo (RNA) en imágenes satelitales?

#### 1.3.2 *Problemas Específicos*

- A. ¿Cómo se puede obtener un modelo óptimo de las RNA que permita delimitar la superficie afectada?
- B. ¿Cómo se puede obtener un modelo óptimo de las RNA que permita realizar una adecuada estimación del área afectada?
- C. ¿Cómo se puede obtener un modelo óptimo de las RNA que permita determinar las bandas de una imagen satelital para delimitar la superficie afectada?
- D. ¿Cómo se puede obtener un modelo óptimo de las RNA que permita determinar una adecuada estimación del área afectada por incendios forestales en la región Cusco?

### 1.4 Antecedentes

Hoeser y Kuenzer (2020) nos brindan una revisión del estado del arte en el tema del aprendizaje profundo (Deep Learning) relacionado a la evolución y tendencias en las técnicas de segmentación de imágenes y detección de objetos con datos de la observación de la Tierra, pudiendo ser: imágenes satelitales y aéreas. En los últimos años, el aprendizaje profundo (Deep Learning - DL) ha recibido una gran cantidad de atención, en aplicaciones tanto a nivel científico y práctico. Los dos principales factores que son responsables para esta creciente atención son: la accesibilidad de los datos y el incremento en el poder de procesamiento computacional, especialmente con las unidades de procesamiento gráfico (GPU - Graphic

Processing Unit). El incremento de la accesibilidad de datos también puede ser encontrado en el campo de la observación de la Tierra. La disponibilidad de las imágenes multiespectrales y ópticas de alta resolución es particularmente importante.

En un futuro próximo se incrementará rápidamente la adquisición de datos a través de sensores remotos de alta resolución. Sin embargo, los datos ópticos de alta resolución han cimentado la base para transferir los conceptos de DL desde la visión de la computación a las aplicaciones de la observación de la Tierra tales como detectar o segmentar vehículos, caminos y edificios desde imágenes aéreas. (Hoeser y Kuenzer, 2020)

Los modelos de DL son específicos de aprendizaje automático (ML - Machine Learning), conformados por capas apiladas, lo cual permite a estos modelos consecutivamente extraer características más ricas desde el ingreso de los datos. Cuando más capas son apiladas, el modelo se vuelve más profundo, y características más complejas pueden ser aprendidas, de aquí viene el nombre de *aprendizaje profundo* (Deep Learning). En las investigaciones, se han dado terminologías y se incorporan a la literatura de DL relacionado a: reconocimiento de imágenes, segmentación de imágenes y detección de objetos con CNNs (Convolutional Neural Networks).

Hoeser et al. (2020) brindan una revisión de las aplicaciones relacionadas al anterior antecedente. La Parte I nos brinda el fundamento teórico de las CNN, este artículo nos muestra las tendencias para los investigadores que aplican DL en observación de la Tierra. En las diversas aplicaciones se ven temas de combinación de tipos de datos, modelos y marcos de trabajo (frameworks) para satisfactoriamente implementar un flujograma de DL.

Los diversos artículos de investigación sobre el DL aplicado a la observación de la Tierra y las diversas aplicaciones son publicadas en una serie de fuentes. Se observa desde el

2012 con la aparición de las CNNs para procesamiento de imágenes. Las fuentes revisadas son: Remote Sensing, IGARSS papers, Geoscience and Remote Sensing Letters, Transactions on Geoscience and Remote Sensing, Sensors, ISPRS Journal of Photogrammetry and RS, Journal of Selected Topics in Applied EO and RS, Access, Journal of Applied Remote Sensing, International Journal of Remote Sensing, Remote Sensing of Environment, ISPRS International Journal of Geo-Information, Scientific Reports, Science of the Total Environment, Geoscience, e International Journal of Earth Observation and Geoinformation. Mientras que los publicadores visitados son: MDPI, IEEE, SPIE, Taylor and Francis, Elsevier y Nature Publishing. De la revisión de artículos se tomó 429, separados por aplicación: detección de objetos (168) y segmentación de imágenes (261), desde el 2012 al 2019. (Hoeser et al., 2020)

Dentro de este grupo de artículos, son nueve los dominios de aplicación: Transporte, Asentamientos, Cobertura general del suelo y uso del suelo, detección de objetos multiclase, agricultura, vegetación natural, peligros naturales, criósfera y fauna silvestre. Comentamos los dominios afines al estudio a realizar.

La generación de mapas en el dominio: Cobertura general del suelo y uso del suelo, tienen una larga tradición en la observación de la Tierra, y cuenta con un 13 % de las publicaciones, es el tercer grupo más grande de aplicaciones. Las publicaciones se conducen en estudios de pruebas de concepto en una escala local demostrando cómo los modelos DL pueden ser aplicados en escenarios complejos con muchas clases alcanzando altas precisiones espaciales. Hay estudios adicionales en un contexto específico en medios costeros y alpinos, así como los humedales derivando en mapas multiclase.

En el sector de la agricultura hay un 10 % de publicaciones. Típicamente, la clasificación de tipos de cultivo tiene un porcentaje alto con un valor de 29 %. Lo que es particularmente importante para la clasificación de tipos de cultivos es la fenología, por eso

las señales temporales deben ser consideradas como límites espaciales. Las CNNs vieron el tema de espacialidad, sin embargo, la explotación temporal es posible ser vista por las CNNs. Varios estudios usan las CNNs en un contexto multitemporal para investigar la clasificación por tipo de cultivos.

En el dominio de la vegetación natural las publicaciones fueron de 4 %. La vegetación natural semejante a los bosques y a grupos de plantas cercanas a la superficie son estudiadas. Asimismo, se analiza la clasificación de áreas boscosas, la salud de bosques y árboles, donde se observa daños causados por insectos y áreas afectadas por tormentas. A nivel individual, los árboles y su corona son detectados con imágenes ópticas. La vegetación sobre la superficie semejante a arbustos y malas hierbas son vistos para monitorear la distribución de malas hierbas que conducen a un aumento en la severidad del fuego o para generar mapas específicos de especies en peligro.

Knopp et al. (2020) elaboraron una investigación basada en la aplicación de modelos de DL. Los incendios forestales presentan consecuencias a nivel económico, social y ecológico. La información acerca de la extensión de las áreas quemadas es esencial para valorar estas consecuencias y que pueden ser derivadas desde datos de sensores remotos. En los pasados años, varios métodos han sido desarrollados para segmentar áreas quemadas con imágenes satelitales. Sin embargo, estos métodos requieren alto preprocesamiento, mientras que las técnicas de DL - las cuales fueron aplicadas de manera satisfactoria a otras tareas de segmentación- aún no se han explorado por completo.

En este estudio, se combinan desarrollos metodológicos y un sensor en específico de años anteriores y se sugiere una cadena de procesamiento automatizado basado en Deep Learning, para la segmentación de áreas quemadas usando una imagen del sensor Sentinel-2. En forma particular, se crea un nuevo conjunto de datos de entrenamiento y validación, el cual

es usado para entrenar una Red Neuronal Convolutiva (CNN) basada en una arquitectura U-Net. (Knopp et al., 2020)

En Lazo (2019) se propuso el uso de espectroscopia con infrarrojo cercano, y Deep Learning para la detección y clasificación de frutas para la agroindustria. Utilizando los equipos de espectroscopia para generar las firmas espectrales de los arándanos (fruta a detectar y clasificar), posteriormente con estas firmas realizar el entrenamiento de una Red Neuronal utilizando el lenguaje de programación Python y la plataforma Keras con Tensorflow. Los resultados muestran que se pueden clasificar los arándanos con una elevada exactitud. (Lazo, 2019)

En Chicchon (2018) se clasificó las áreas urbanas mediante la teledetección utilizando algoritmos de Deep Learning basados en segmentación semántica sobre la fusión de datos. El crecimiento urbano requiere un mapeo y seguimiento para planificar futuros desarrollos. La teledetección permite obtener información de la superficie de la Tierra. Debido a que esta información puede ser generada desde distintas fuentes, se hace necesario contar con técnicas de fusión multisensorial eficientes y efectivas. El aprendizaje profundo o Deep Learning se ha vuelto muy importante en el enfoque de Visión por Computadora y Teledetección, existiendo diferentes estudios de la aplicación de variantes de Redes Neuronales Convolucionales (CNN) en segmentación semántica. En este estudio se investiga el uso de métodos de fusión de datos basados en algoritmos de aprendizaje profundo para la segmentación semántica en aplicaciones urbanas. (Chicchón, 2018)

Debido a que nuestra geografía es muy diferente, las imágenes utilizadas en los estudios precedentes no son suficientes para poder generalizar los diversos ecosistemas de nuestro territorio, por ello es necesario generar nuestra propia base de datos de imágenes y etiquetas; asimismo, un modelo nuevo que sea entrenado desde cero.

## 1.5 Justificación de la Investigación

Los incendios forestales afectan al patrimonio forestal y al capital natural de la Nación de manera directa, además de la afectación directa a la vida y la salud de la población, así como la infraestructura y medios de vida. Su extinción puede tomar días considerando las condiciones climáticas que se presentan, por lo que el daño puede ser de gran magnitud. Debido a ello, se hace necesario delimitar los límites de las superficies afectadas inmediatamente después de la extinción de los incendios con la intención de diseñar e implementar medidas para la restauración de dichas áreas quemadas.

Una de las metodologías para la delimitación de áreas afectadas por incendios forestales es ir al lugar del siniestro y realizar un levantamiento de campo. Esto puede tomar tiempo y un elevado costo. Sin embargo, la tecnología nos permite contar con imágenes satelitales que en un corto tiempo nos permite observar la realidad de una extensa área de nuestro territorio. Los satélites de diversas constelaciones generan un número cada vez más creciente de imágenes de nuestro territorio y a nivel mundial, conformando de esta manera una gran base de datos de imágenes que a través del uso de diversas herramientas informáticas que procesan dichos datos, nos permiten responder una serie de preguntas.

En este contexto, y existiendo un mayor número de satélites de diversas características que nos permiten obtener información sobre la realidad de nuestro territorio y del mundo; se hace necesario acelerar el procesamiento de dicha cantidad de imágenes. Es en este punto que se presenta la aplicabilidad del procesamiento automatizado por medio de la inteligencia artificial. A través del “Deep Learning” (Aprendizaje Profundo), se nos brinda la alternativa de utilizar los elementos computacionales para el procesamiento de una gran cantidad de imágenes, posibilitando de esta manera delimitar de una forma automatizada en tiempo oportuno las áreas afectadas e identificar los límites del patrimonio forestal afectado (esto debido a la ventaja de

las CNN de encontrar las características representativas sin supervisión humana).

## **1.6 Limitaciones de la Investigación**

- Debido a la gran extensión de nuestro territorio, la investigación se centrará principalmente en el departamento de Cusco, para el período 2017 al 2021.
- El proceso de entrenamiento del modelo de RNA será realizado en el hardware: CPU: Intel i7, GPU: NVIDIA GeForce GTX 1660 de 6 GB RAM, Memoria RAM: 24 GB.

## **1.7 Objetivos de la Investigación**

### **1.7.1 Objetivo General**

Delimitar el área afectada por un incendio forestal en tiempo oportuno mediante el uso de aprendizaje profundo (RNA) en imágenes satelitales.

### **1.7.2 Objetivos Específicos**

- A. Obtener un modelo óptimo de las RNA que permita delimitar la superficie afectada.
- B. Obtener un modelo óptimo de las RNA que permita realizar una adecuada estimación del área afectada.
- C. Obtener un modelo óptimo de las RNA que permita determinar las bandas de una imagen satelital para delimitar la superficie afectada.
- D. Obtener un modelo óptimo de las RNA que permita determinar una adecuada estimación del área afectada por incendios forestales en la región Cusco.

## **1.8 Hipótesis**

### **1.8.1 Hipótesis General**

Un modelo de aprendizaje profundo utilizando segmentación semántica nos permitirá delimitar la superficie afectada denominada cicatriz de área quemada producto de un incendio forestal en tiempo oportuno.

### ***1.8.2 Hipótesis Específicos***

- A. Un modelo óptimo de las RNA permitirá delimitar la superficie afectada.
- B. Un modelo óptimo de las RNA permitirá realizar una adecuada estimación del área afectada.
- C. Un modelo óptimo de las RNA permitirá determinar las bandas de una imagen satelital para delimitar la superficie afectada.
- D. Un modelo óptimo de las RNA permitirá determinar una adecuada estimación del área afectada por incendios forestales en la región Cusco.

## II. MARCO TEÓRICO

### 2.1 Monitoreo Satelital - SERFOR

Como se mencionó en el capítulo anterior, el estudio se realizará con cicatrices de incendios forestales generados por la Unidad Funcional de Monitoreo Satelital (UFMS). Esta unidad es dependiente de la Dirección General de Información y Ordenamiento Forestal y de Fauna Silvestre (DGIOFFS) de SERFOR, conformada el 22 de marzo del 2021 (Resolución de Gerencia General N°006-2021-MIDAGRI-SERFOR-GG). Pero ya desde el 2017 se ha venido generando información relacionada a los incendios forestales. SERFOR como parte integrante de un grupo de entidades que monitorea y responde ante este tipo de desastre, y cumpliendo con sus funciones relacionadas a la temática, elaboró un documento técnico denominado: Plan de Prevención y Reducción de Riesgos de Incendios Forestales, aprobada el 17 de diciembre del 2018, y con una vigencia del 2019-2022. En dicho Plan, se indican las acciones a realizar por parte de SERFOR dentro de sus funciones en el marco de los incendios forestales. (Resolución de Dirección Ejecutiva N°284-2018-MINAGRI-SERFOR-DE)

Dentro de dicho marco se definen los siguientes términos:

**Incendio Forestal:** fuego no deseado de cualquier origen, que no es estructural, que se propaga sin control en los recursos forestales causando daños ecológicos, económicos y sociales. Las coberturas consideradas forestales incluyen ecosistemas naturales como bosques, matorrales, pastizales, bofedales, entre otros; y plantaciones forestales.

**Foco de calor:** puntos calientes sobre la superficie terrestre, detectados por satélites especializados de la NASA.

**Alerta de incendio forestal:** son los focos de calor agrupados que caen sobre la cobertura vegetal natural, plantaciones y cercanas a las áreas agropecuarias. Estos focos brindan una alerta de la existencia de fuego activo. Sin embargo, estas alertas deben ser confirmadas

por las autoridades locales.

**Cicatriz de incendio forestal:** huella de superficie quemada en consecuencia de la ocurrencia de un incendio sobre cobertura forestal.

**Mosaico Landsat:** compuesto rasterizado de imágenes de los satélites del programa Landsat de la NASA.

En el marco de estudio y en coordinación con los especialistas de la UFMS, los datos de áreas afectadas mediante el uso de mosaicos anuales de imágenes satelitales se han generado desde el 2017 al 2021, con la finalidad de tener un marco de referencia del área quemada a nivel nacional producto de los incendios forestales. En este punto, hay que indicar que el departamento de análisis (Cusco) se encuentra enmarcado dentro de las regiones con mayor afectación en lo relacionado al patrimonio forestal y de fauna silvestre (Resolución de Dirección Ejecutiva N°284-2018-MINAGRI-SERFOR-DE). Las zonas urbanas, fábricas, industrias, y similares no son consideradas. Estas cicatrices son generadas utilizando imágenes satelitales (Landsat-8 y/o Sentinel-2), software de información geográfica (ArcGIS y QGIS), y la plataforma Google Earth Engine.

Observamos en la Tabla 1 un reporte de las cicatrices generadas.

La generación de las cicatrices se realiza de manera anual, pero en algunos casos de acuerdo a su magnitud o solicitud en específico por parte de alguna autoridad gubernamental se genera un reporte individual de un incendio forestal (RIF) y se calculan las áreas afectadas.

De acuerdo al reporte de cicatriz presentado, y lo mencionado anteriormente, nuestro territorio es inmenso y muy vasto, conteniendo una gran variedad de ecosistemas a través de nuestras regiones: Costa, Sierra y Selva. Por ello la decisión de solamente analizar las cicatrices concernientes al departamento de Cusco.

**Tabla 1**

*Registro Histórico de Áreas Afectadas (Cicatrices) de Incendios Forestales por Departamento*

<b>Departamento</b>	<b>2018 (ha)</b>	<b>2019 (ha)</b>	<b>2020 (ha)</b>
Amazonas	7502.49	6390.20	14595.96
Ancash	16687.71	18432.24	31654.61
Apurímac	6185.61	10463.49	40124.21
Arequipa	11460.92	8990.58	8555.05
Ayacucho	7698.74	15512.39	35125.65
Cajamarca	15123.75	11298.98	41396.23
Cusco	27804.10	55744.46	108436.26
Huancavelica	3931.10	5513.81	16956.51
Huánuco	6716.84	19152.58	26349.49
Ica	788.91	321.03	464.06
Junín	6539.31	22627.82	49867.87
La Libertad	15159.63	5392.70	19106.79
Lambayeque	4246.13	223.37	585.35
Lima	3328.12	3992.14	5635.28
Loreto	0.00	550.23	258.26
Madre de Dios	201.64	4566.27	839.80

*Nota.* Tomado de *Reporte de Áreas Afectadas (Cicatrices) de Incendios Forestales 2018-2019-2020 (interno)*, por la UFMS-DGIOFFS, Setiembre 2021, SERFOR.

Las cicatrices son generadas en archivos digitales de tipo vectorial bajo el formato Shapefile, y con geometría de forma poligonal. En la Figura 1 apreciamos una sección de un RIF.

Se utilizan imágenes satelitales pre y post incendio forestal. Adicionalmente de calcular el área de la cicatriz, se calcula su afectación sobre diversos ecosistemas afectados por el siniestro, con la finalidad de acciones a tomar para el proceso de restauración de zonas afectadas.

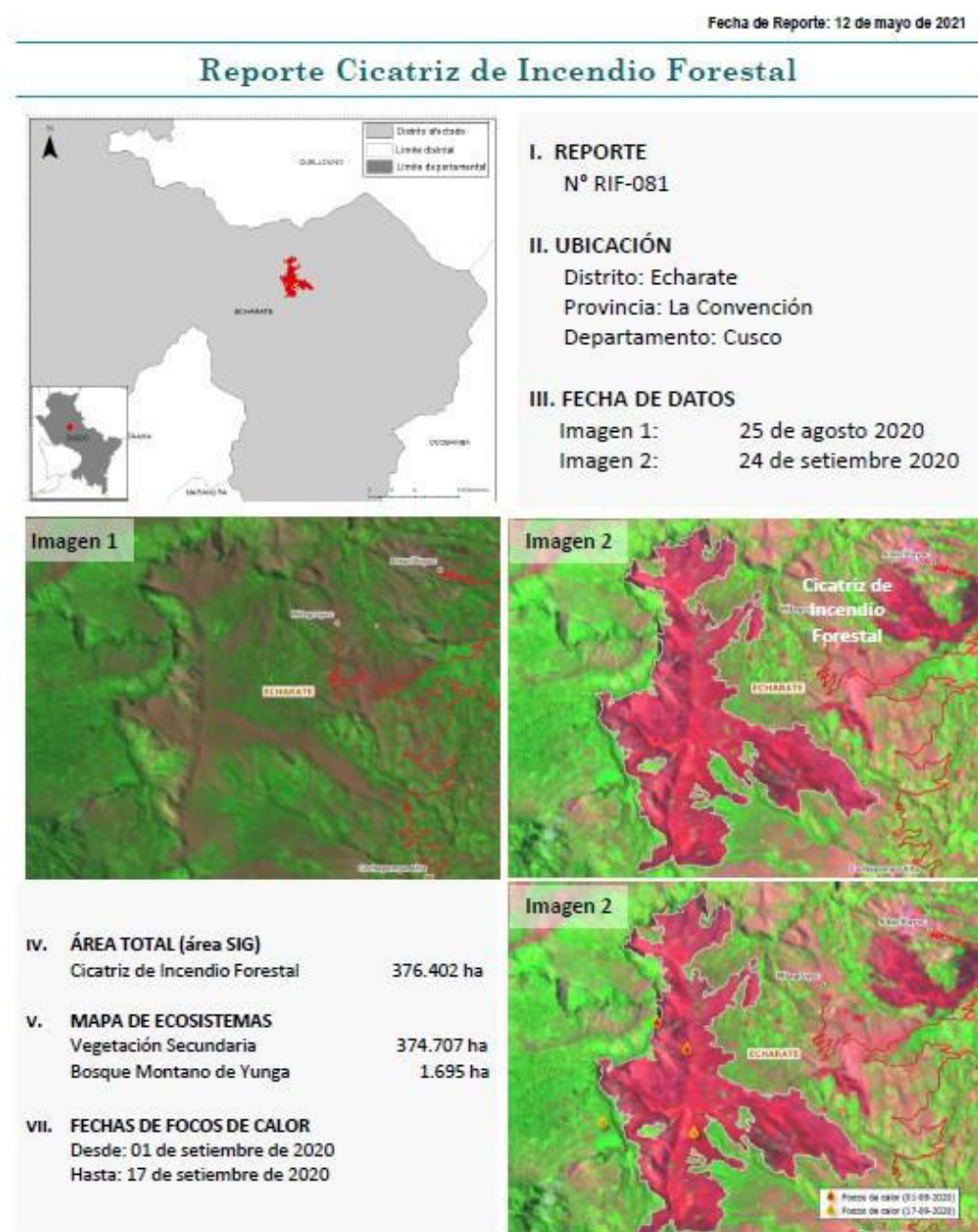
## **2.2 Teledetección**

### **2.2.1 Introducción**

A través de los años, siempre hemos buscado la manera de observar la diversidad de paisajes que nuestro planeta alberga. El levantamiento de datos del terreno siempre se realiza

de manera directa, estando en el lugar con diversos equipos para medir distancias, calcular ángulos, toma de fotos, entre otros. Esta manera de adquirir datos toma un prolongado tiempo y sobre todo es de un alto costo.

**Figura 1**  
*Registro Individual de Cicatriz de Incendio Forestal*



*Nota.* Tomado de *Reporte Cicatriz de Incendio Forestal RIF-081 (interno - Composición de Bandas: 12,8,4 de Sentinel-2)*, por la UFMS-DGIOFFS, 2021, SERFOR.

Tomando en cuenta lo último, estudiar la superficie terrestre se vuelve una necesidad, por ello se constituye el marco de estudio de la **Teledetección**. También llamado Sensado Remoto, proveniente de la traducción del término inglés: Remote Sensing. Se le denomina así a cualquier medio de observación remota de la superficie terrestre. Inicialmente se aplicó a las fotografías aéreas, principal sensor en aquellos tiempos, también se tienen ahora sensores digitales como las cámaras aerotransportadas (aviones) o desde satélites. La teledetección no solo cubre el tema de la adquisición de la imagen, sino su posterior procesamiento e interpretación. (Chuvieco, 2016)

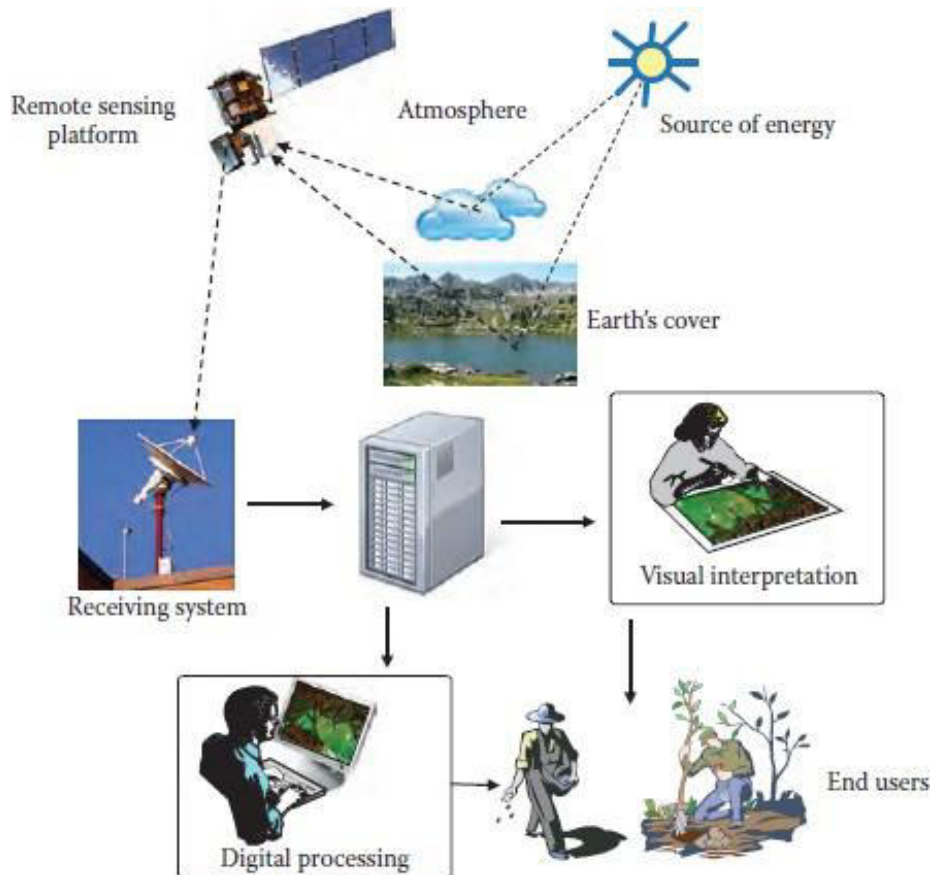
Los elementos de un sistema de teledetección son (Figura 2):

- Fuente de energía (Source of energy)
- Cubierta terrestre (Earth's cover)
- Sistema sensor (Remote sensing platform)
- Sistema de recepción (Receiving system)
- Interpretación (Visual interpretation - Digital processing)
- Usuario final (End users)

En relación a uno de los componentes principales de un sistema de teledetección, se encuentra la fuente de energía. Los sensores captan la radiación electromagnética proveniente de este componente esencial.

De forma natural se tiene al Sol como fuente principal, y a los sensores que utilizan dicha energía se les denomina de tipo pasivo, generando las denominadas imágenes ópticas. De similar manera, se tienen sensores que incorporan su propia fuente de energía, denominados de tipo activo, dando lugar a las imágenes de tipo radar, entre otras.

**Figura 2**  
*Componentes de un sistema de teledetección*



*Nota.* Tomado de *Illustration of the main components associated with remote sensing activities*, por Chuvieco, 2016, *Fundamentals of Satellite Remote Sensing - An Environmental Approach - Second Edition*.

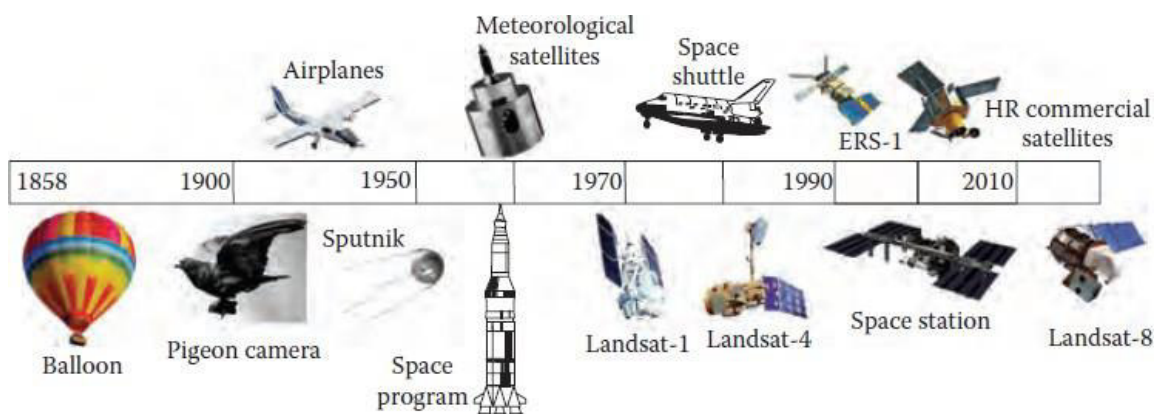
A nivel de los años, la teledetección ha evolucionado de la mano con el avance tecnológico tanto a nivel hardware como software. En la actualidad, ya se poseen sensores de plataforma espacial alrededor del planeta (satélites), aviones aerotransportados, drones o UAV (Sistemas de aeronaves no tripuladas). En la Figura 3 podemos apreciar el desarrollo de estos sistemas.

Tomando conocimiento de lo anterior, y por medio de las ciencias, tenemos el espectro

electromagnético, como una clasificación de toda la energía radiante en función de la longitud de onda o frecuencia. Se suelen establecer una serie de bandas en donde dicha radiación manifiesta un comportamiento similar. En la Figura 4 podemos apreciar el espectro y las bandas mencionadas.

### Figura 3

#### *Historia de los sistemas de teledetección*



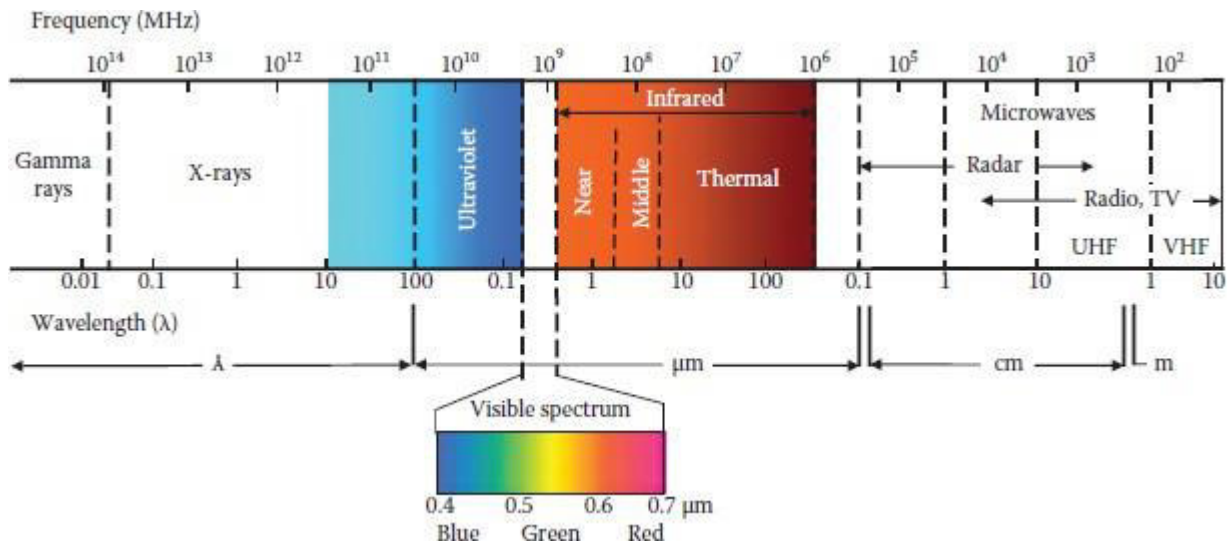
*Nota.* Tomado de *Historical development of remote sensing systems*, por Chuvieco, 2016, *Fundamentals of Satellite Remote Sensing - An Environmental Approach - Second Edition*.

Dentro de estas bandas las más utilizadas para la observación de la tierra son:

- Espectro visible (Visible spectrum)
- Infrarrojo cercano (Infrared Near)
- Infrarrojo medio (Infrared Middle)
- Infrarrojo térmico (Infrared Thermal)
- Microondas (Microwaves)

Para el estudio se utilizarán sensores de plataforma espacial (satélites) de tipo pasivo.

**Figura 4**  
*Espectro electromagnético*



*Nota.* Tomado de *Major spectral bands within the electromagnetic spectrum*, por Chuvieco, 2016, *Fundamentals of Satellite Remote Sensing - An Environmental Approach - Second Edition*.

### 2.2.2 *Satélites de Observación Terrestre*

Como mencionamos en la sección anterior, los sensores se clasifican de dos tipos: pasivos y activos. Para realizar esta investigación observaremos las imágenes ópticas provenientes de los sensores pasivos. Dentro de este grupo se encuentran una gran variedad de sensores: ASTER, MODIS, Landsat, Sentinel, Worldview, entre otros. De manera específica procesaremos imágenes provenientes de los sensores: Landsat-8 y Sentinel-2.

Después de mencionar los dos satélites con lo que se realizará la investigación, es necesario mencionar algunas características principales que nos brindan la calidad del sensor, los cuales son relevantes al momento de seleccionar el apropiado para una aplicación en particular:

- Resolución espacial: es la distancia que corresponde a la unidad mínima de información

incluida en la imagen (píxel). Así, a menor tamaño del píxel mayor será la resolución espacial, lo que implica que el sensor obtendrá más detalle de los objetos.

- Resolución espectral: determina el número y ancho de las bandas espectrales que puede diferenciar el sensor. Entre mayor sea esta resolución se tendrá una mayor información del comportamiento de una misma cobertura en diferentes bandas espectrales.
- Resolución radiométrica: nos indica la sensibilidad del sensor, siendo la capacidad para diferenciar pequeñas variaciones en la radiancia espectral registrada. Es decir, la medida en bits (Dígitos Binarios), la gama de valores de brillo disponible, que corresponde con el alcance máximo de niveles digitales (ND o DN); por ejemplo, una imagen con una resolución de 8 bits tiene 256 niveles de brillo. En una imagen cada píxel, asume una cantidad de valores específico, dependiendo de su resolución radiométrica; viendo el ejemplo anterior, una imagen de 8 bits, puede tomar valores desde 0 a 255 tonalidades de gris. En la Tabla 2 podemos observar detalles de esta resolución.
- Resolución temporal: nos referimos al tiempo por el cual se vuelve a visitar la misma zona de la superficie terrestre. Este valor de tiempo nos permite obtener imágenes de la misma región de la Tierra. Este valor está relacionado con las características propias del sensor, tales como: altura, velocidad, etc.

Las imágenes satelitales también se pueden clasificar en:

- Pancromáticas: son aquellas que disponen de una sola banda espectral, comúnmente gran parte del espectro visible, donde se obtiene una imagen que se representa en una escala de grises (en blanco y negro).
- Multiespectral: son aquellas que disponen de varias bandas, normalmente entre 3 a 20 bandas, no necesariamente contiguas. Mayormente cubren el espectro visible, infrarrojo,

entre otras.

- Hiperespectrales: son aquellas que brindan un mayor número de bandas (>20) y normalmente contiguas.

**Tabla 2**

*Resolución radiométrica - Niveles Digitales*

Número Bits	Niveles de gris	ND - Mínimo	ND - Máximo
8	256	0	255
9	512	0	511
10	1024	0	1023
11	2048	0	2047
12	4096	0	4095
13	8192	0	8191
14	16384	0	16383
15	32768	0	32767
16	65536	0	65535

### **Landsat-8**

El programa Landsat es una serie de satélites de observación de la Tierra liderado por la NASA y el Servicio Geológico de los EEUU. En 1972, se lanzó el Landsat-1, que dio inicio a la obtención de datos adquiridos de forma continua y remota desde el espacio.

A través de los años se lanzaron las versiones: Landsat-2, Landsat-3, Landsat-4, Landsat-5, Landsat-6. En 1999 fue lanzado el Landsat-7 y se mantiene aún operativo, pero presenta problemas técnicos en su sensor principal. Debido a ello, en 2013 se lanzó el Landsat-8 para suplir los problemas de la versión 7, se encuentra operativo y los datos adquiridos nos permiten monitorear la Tierra actualmente. También, es importante mencionar que en 2015 se aprobó el desarrollo del Landsat-9 esperando a lanzarse en el 2023. (Chuvieco, 2016)

El Landsat-9 fue lanzado el 27 de setiembre del 2021, sin presentar contratiempos. El satélite lleva sensores similares al Landsat-8 pero una versión mejorada (OLI-2 y TIRS-2). El satélite se desplazará a la órbita del Landsat-7 (con la finalidad de sustituirlo), el cual

tiene combustible para operar hasta 2021, luego de dicha fecha será puesto fuera de servicio. En noviembre del 2021, se han obtenido las primeras imágenes del Landsat-9. Los datos brindados por el Landsat-9 estarán disponibles al público a principios de 2022. (Fuente: <https://landsat.gsfc.nasa.gov/landsat-9>).

En la Tabla 3 podemos observar las bandas espectrales de Landsat- 8 y 9.

**Tabla 3**

*Bandas Espectrales de Landsat-8 y 9*

<b>Bandas</b>	<b>Longitud de onda (micrometros)</b>	<b>Resolución (m)</b>
Banda 1 - Coastal aerosol	0.43 - 0.45	30
Banda 2 - Blue	0.45 - 0.51	30
Banda 3 - Green	0.53 - 0.59	30
Banda 4 - Red	0.64 - 0.67	30
Banda 5 - Near Infrared (NIR)	0.85 - 0.88	30
Banda 6 - SWIR 1	1.57 - 1.65	30
Banda 7 - SWIR 2	2.11 - 2.29	30
Banda 8 - Panchromatic	0.50 - 0.68	15
Banda 9 - Cirrus	1.36 - 1.38	30
Banda 10 - Thermal Infrared (TIRS) 1	10.6 - 11.19	100
Banda 11 - Thermal Infrared (TIRS) 2	11.5 - 12.51	100

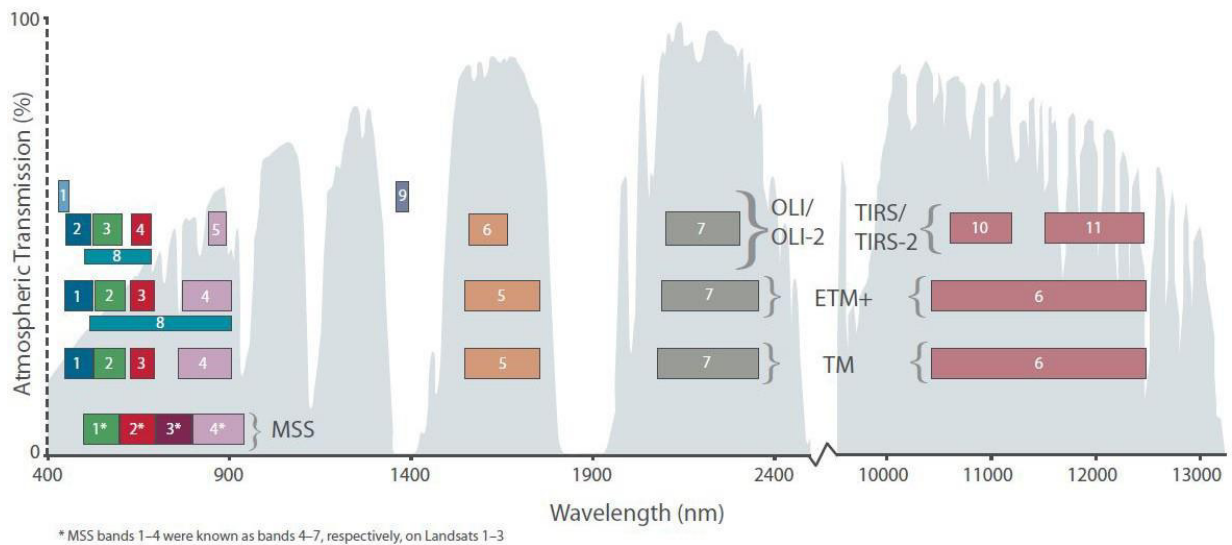
*Nota.* Tomado de *What are the band designations for the Landsat satellites?*, por USGS, 2021, Fuente: <https://www.usgs.gov/faqs/what-are-band-designations-landsat-satellites>.

Como podemos apreciar y se mencionó anteriormente, la misión 9 mantiene los mismos tipos de sensores que la misión 8 pero con tecnología actualizada. Observamos en la Figura 5 que mantienen un similar espectro electromagnético en las bandas. En cuanto a las características de la misión 8, podemos mencionar:

- Resolución espacial: se aprecia en la Tabla 3.
- Resolución espectral: se aprecia en la Figura 5.
- Resolución radiométrica: las bandas se encuentran cuantizadas a 12 bits. (USGS, 2019)
- Resolución temporal: el ciclo de visita de cualquier zona de la Tierra es de 16 días.

(United States Geological Survey [USGS], 2019)

**Figura 5**  
*Espectro Electromagnético de Landsat-8 y 9*



*Nota.* Tomado de *LANDSAT 9 BANDS - The Multispectral Scanner System (MSS)*, por USGS, 2021, Fuente: <https://landsat.gsfc.nasa.gov/satellites/landsat-9/landsat-9-bands>.

Podemos adicionar, que el Landsat-9 se unió a su hermano el Landsat-8, y trabajando en forma conjunta, ambos satélites recopilan imágenes que abarcan todo el planeta cada 8 días, disminuyendo la resolución temporal de las misiones Landsat de 16 a 8 días.

## Sentinel-2

El programa Copernicus bajo la dirección de la Agencia Espacial Europea (ESA) es el encargado de la observación y monitorización de la Tierra, que analiza el planeta y su medio ambiente en beneficio de los ciudadanos europeos. El programa liderado por la Comisión Europea, proporciona datos de manera operacional y servicios de información de forma gratuita sobre numerosas áreas de aplicación, gracias a una gran variedad de tecnologías, que van desde los satélites en el espacio a los sistemas de medición de tierra, mar y aire. (Instituto Geográfico Nacional, 2020)

El programa se apoya en una familia de seis satélites Sentinel. A continuación los describimos en forma sucinta: (Instituto Geográfico Nacional, 2020)

- Sentinel-1: SAR (radar). Zonas con nubes, datos diurnos y nocturnos.
- Sentinel-2: resolución mediana/alta, tipo óptico. Aplicaciones terrestres: urbanas, bosques, agricultura.
- Sentinel-3 y 6: océanos y servicios terrestres a escala global, altimetría oceánica de alta resolución.
- Sentinel-4 y 5: composición atmosférica, con órbita geoestacionaria (S-4) y polar (S-5).

De acuerdo a lo informado, en este estudio se utilizarán las imágenes provenientes de los satélites: Sentinel-2A y Sentinel-2B. Son dos sensores con órbita polar con desfase de 180 grados entre ellos. Debido a estas características el tiempo de revisita de una zona del planeta es de 5 días. (European Space Agency [ESA], 2015).

En la Tabla 4 podemos observar las bandas espectrales del Sentinel-2.

**Tabla 4**  
*Bandas Espectrales de Sentinel-2*

<b>Bandas</b>	<b>Longitud de onda central (um)</b>	<b>Resolución (m)</b>
Banda 1 - coastal aerosol	0.443	60
Banda 2 - azul	0.490	10
Banda 3 - verde	0.560	10
Banda 4 - rojo	0.665	10
Banda 5 - Vegetation Red Edge	0.705	20
Banda 6 - Vegetation Red Edge	0.740	20
Banda 7 - Vegetation Red Edge	0.783	20
Banda 8 - NIR	0.842	10
Banda 8A - Vegetation Red Edge	0.865	20
Banda 9 - Vapor de agua	0.945	60
Banda 10 - SWIR - cirrus	1.375	60
Banda 11 - SWIR	1.610	20
Banda 12 - SWIR	2.190	20

*Nota.* Tomado de *Bandas de los satélites Sentinel-2*, por Instituto Geográfico Nacional,

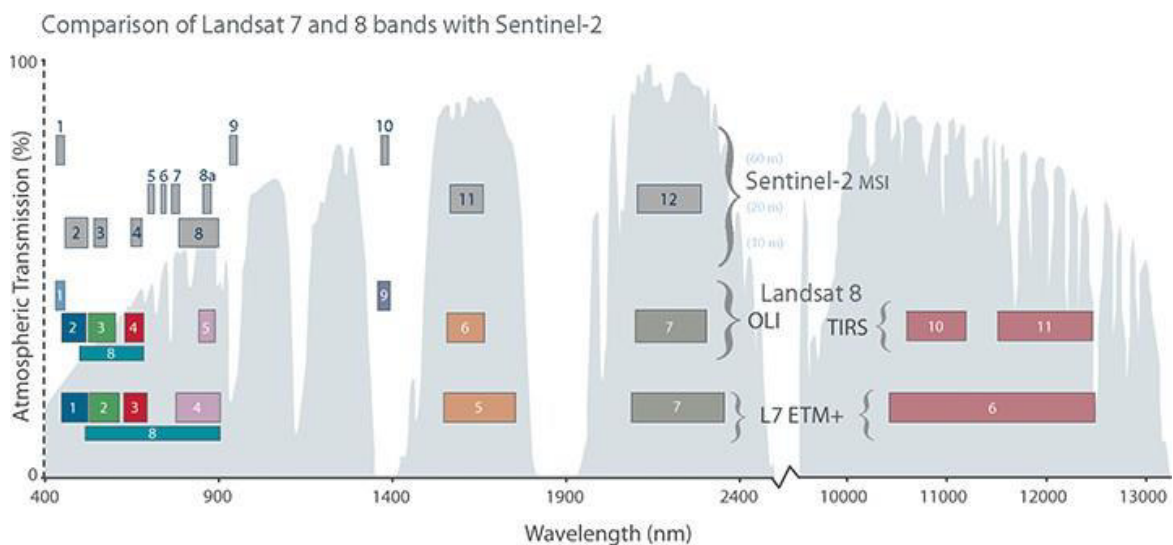
2020, El Programa Copernicus para la monitorización del territorio y los Objetivos del Desarrollo Sostenible.

En cuanto a las características de la misión Sentinel-2, podemos mencionar:

- Resolución espacial: se aprecia en la Tabla 4.
- Resolución espectral: se aprecia en la Figura 6.
- Resolución radiométrica: las bandas se encuentran cuantizadas a 12 bits. (ESA, 2015)
- Resolución temporal: el ciclo de repetición de visita de cualquier zona de la Tierra es de 5 días. (ESA, 2015)

En la Figura 6 se observa el espectro electromagnético de la misión Sentinel-2.

**Figura 6**  
*Espectro Electromagnético de Sentinel-2, Landsat-7,8*



*Nota.* Tomado de *Comparison of Landsat 7 and 8 bands with Sentinel-2*, por USGS, 2021, Fuente: <https://www.usgs.gov/media/images/comparison-landsat-7-and-8-bands-sentinel-2>.

Como se puede apreciar, las bandas espectrales de los satélites Landsat-7, Landsat-8

son en la mayoría de rangos similares de la misión Sentinel-2. Lo cual permite un análisis complementario entre dichos sensores. Asimismo, el Landsat-9 al ser hermano del Landsat-8, sus imágenes podrán utilizarse en forma conjunta con Sentinel-2. Las bandas del Landsat-8 en donde pueden utilizarse en forma conjunta son: B2, B3, B4, B5 (B8 o B8A en S-2), B6 (B11 en S-2), B7 (B12 en S-2), B9 (B10 en S-2). De este trabajo conjunto entre ambas plataformas (Landsat-8 y Sentinel-2) se obtendrá una visión más completa de la Tierra, y de acuerdo a la órbita de los mismos se tendrá observaciones de cualquier lugar del planeta cada dos días de manera aproximada.

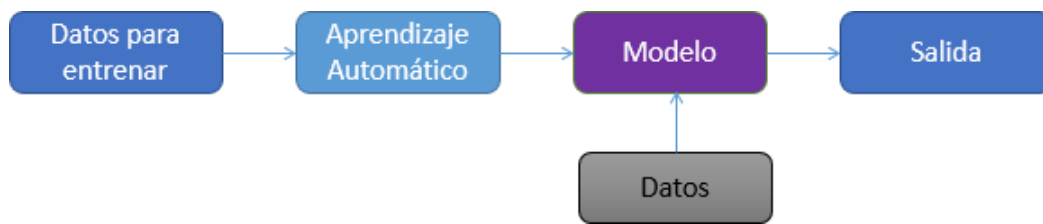
### 2.3 Machine Learning

Una primera definición de *Aprendizaje Automático* (*Machine Learning - ML*, en el idioma inglés) fue introducida en 1959 por Arthur Samuel, uno de los pioneros de la inteligencia artificial. Esta definición indica que el aprendizaje automático es una rama de la inteligencia artificial cuyo fin es dotar a una máquina, a través de algoritmos, la capacidad de entrenar y aprender en base a datos, sin ser explícitamente programada, imitando la capacidad que tienen las personas de aprender mediante ejemplos sin recurrir a fórmulas ni reglas entre las variables y posibilitando al término del entrenamiento un modelo que permite la *generalización*; es decir, la obtención de resultados en nuevas situaciones no conocidas durante el aprendizaje. Este comportamiento es muy importante porque permite resolver situaciones en donde no existe o es muy difícil encontrar una fórmula que facilite una respuesta exacta a partir del conocimiento de ciertas variables. De esta manera el modelo podrá, por ejemplo, reconocer y clasificar nuevas imágenes si se ha “entrenado” en un conjunto de casos o ejemplos para los cuales se han presentado determinadas características y la imagen correspondiente. Surge, de esta manera, el aprendizaje automático como una *ciencia de los algoritmos* que brinda a las máquinas la capacidad de aprender para resolver problemas que a “mano” son imposibles de

ejecutar. (Véliz, 2018)

De acuerdo a la Figura 7, el aprendizaje automático comprende una serie de técnicas con la finalidad de obtener un modelo a partir del aprendizaje en un conjunto de datos, sin que medie la intervención de las personas. (Véliz, 2018)

**Figura 7**  
*Procesos del Aprendizaje Automático*



*Nota.* Tomado de *Aprendizaje automático*, por Veliz, 2018, *Aprendizaje Automático - Análisis para la minería de datos y Big Data*.

Los problemas a resolver son mayormente de dos tipos:

- Regresión: calcular un valor numérico.
- Clasificación: predecir o seleccionar una categoría como salida.

Desde que la tarea principal es seleccionar un algoritmo de aprendizaje y entrenar con un conjunto de datos, las dos ideas que pueden salir equivocadas son: mal algoritmo o data errónea. Podemos mencionar algunos ejemplos de data errada:

- Insuficiente cantidad de datos de entrenamiento
- Datos de entrenamiento no representativa
- Datos de pobre calidad
- Datos irrelevantes
- Sobreajuste de los datos de entrenamiento: el modelo se desempeña bien en los datos de

entrenamiento, pero no generaliza correctamente.

- Subajuste de los datos de entrenamiento: lo opuesto al sobreajuste, el modelo es muy simple para aprender desde los datos.

El único camino para conocer que un modelo generaliza bien en nuevos casos es actualmente probarlo en casos no entrenados. Un camino para realizar ello es desplegar el modelo en producción y monitorear cómo se desempeña. Esto trabaja bien, pero si el modelo genera resultados erróneos, los usuarios se quejarán, siendo no la mejor idea.

La mejor opción es dividir los datos dentro de dos conjuntos: *entrenamiento* y *prueba*. Como estos nombres implican, se entrenará el modelo en el conjunto de entrenamiento, y se probará usando el conjunto de prueba. La tasa de error en nuevos casos se denomina error de generalización, y por evaluar tu modelo en el conjunto de prueba, se obtendrá una estimación de este error. Este valor nos indica que tan bien tu modelo se desempeñará en instancias que nunca ha visto. (Géron, 2019)

## 2.4 Deep Learning

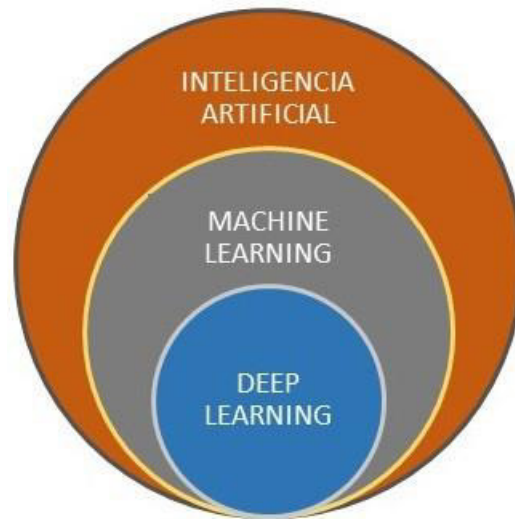
Se está considerando la Inteligencia Artificial como la nueva revolución industrial, muchos lo denominan la industria 4.0. Lo que mueve ello es el *Deep Learning*. Los sistemas informáticos actuales ya traducen textos en cualquier idioma, responden automáticamente correos electrónicos o crean noticias falsas que abundan a nivel mundial. Todo esto es producido por la inteligencia artificial que se denomina *Deep Learning* - DL (traducido como *Aprendizaje Profundo*). El término Deep Learning agrupa una serie de técnicas de aprendizaje automático (ML) que se basan en modelos de redes neuronales y cuya aplicación se observa en multitud de usos a nivel mundial. (Torres, 2020).

En la Figura 8 podemos apreciar la relación en forma gráfica de: Inteligencia Artificial,

Machine Learning y Deep Learning.

**Figura 8**

*Deep Learning, parte de Machine Learning, y a su vez de la Inteligencia Artificial*



El auge del Deep Learning, se inicia aproximadamente en el 2012 de la mano de las mejoras computacionales con la aparición de los CPUs. A esto se suma la aparición de los GPU (Graphic Processing Unit) creados inicialmente para los videojuegos. El aprendizaje profundo requiere gran capacidad de procesamiento la cual comenzó a facilitarse por el sistema de procesamiento en paralelo de los GPU. Resultando estos más eficientes que los CPU para determinados algoritmos.

Sumado a esto se suma una era de gran generación de datos, denominada *Big Data*, debido a que todos estos son registrados y se encuentran al alcance de los investigadores para el entrenamiento de redes neuronales. Con el pasar de los años y de los diversos avances tecnológicos, así como su aplicación en diversos sectores de la industria y del quehacer humano, se puede encontrar gran cantidad de modelos preentrenados y que pueden reutilizarse.

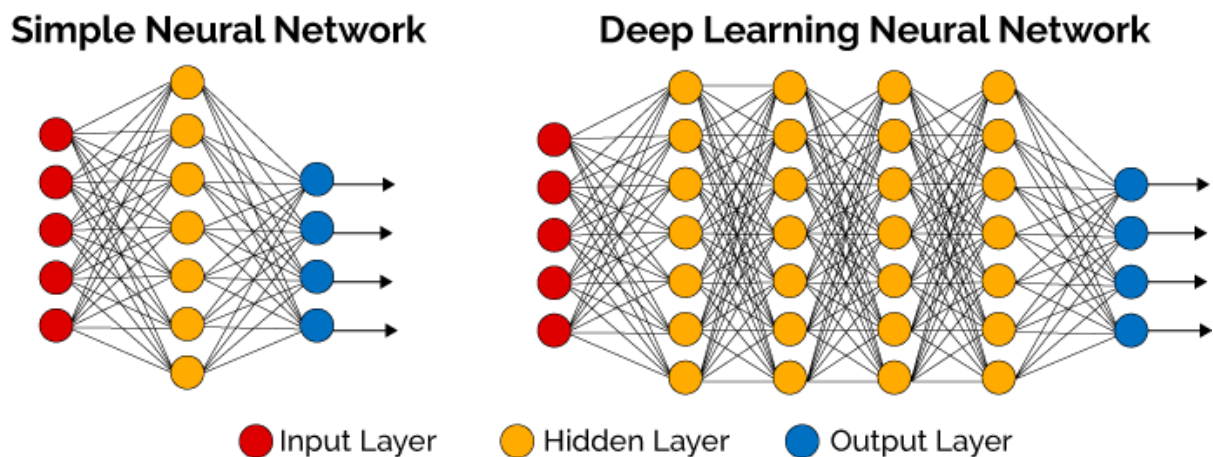
Siendo las redes neuronales uno de los algoritmos de Machine Learning, son utilizados para el mismo fin, pero presentan un rendimiento mayor que otros algoritmos. Una de sus principales diferencias es su mayor capacidad de abstracción. Los algoritmos de Deep Learning

tienen la capacidad de realizar una abstracción semejante a propia cuenta, sin necesidad de un diseño previo. Ello nos indica que el Deep Learning no solo es capaz de aprender, sino también encontrar un significado entre los datos.

En la Figura 9 podemos apreciar ejemplos de Deep Learning: redes neuronales de una o más capas.

### Figura 9

*Deep Learning - red neuronal de una o más capas ocultas (hidden)*



*Nota.* Fuente: <https://www.ackstorm.com/deep-learning-clasificacion-imagenes/>.

El número de investigaciones en Deep Learning crece cada vez, esto debido a su capacidad de adaptación. De similar manera se aplica a tareas de regresión como de clasificación. Se utilizan en sectores como: transporte, medicina, agricultura, aeroespacial, defensa, automatización industrial, electrónica y en procesamiento de imágenes para reconocimiento de objetos o patrones. En la actualidad, hay dos grandes áreas de investigación: Procesamiento del Lenguaje Natural (NLP - *Natural Language Processing*) y Visión por Computador (CV - *Computer Vision*).

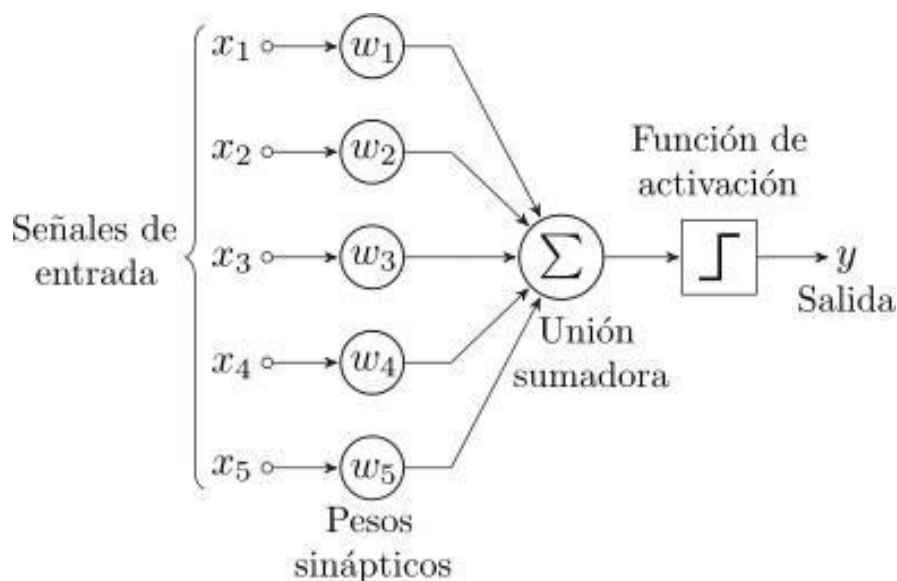
### Redes Neuronales Artificiales

Como ya mencionamos, es un caso especial de *Machine Learning*, que en cierta manera

son un intento de imitar la actividad en capas de neuronas en la neocorteza, que es la parte del cerebro humano donde ocurre el pensamiento. Estas redes neuronales aprenden estructuras jerárquicas y niveles de representación y abstracción, para comprender los patrones de datos que provienen de varios tipos de fuentes, como imágenes, videos, sonido o texto. (Torres, 2020)

El perceptrón es la unidad básica de una red neuronal. Varios de ellos forman una red neuronal. Cada unidad está conformada por varias entradas y una única salida. La salida como se aprecia en la Figura 10, es una sumatoria de las entradas ponderadas adicionando un sesgo o bias de la neurona (también ponderada) a la cual se le aplica una función de activación. La variación de los pesos para obtener la salida deseada es el resultado del proceso de aprendizaje.

**Figura 10**  
*Perceptrón*



*Nota.* Fuente: <https://es.wikipedia.org/wiki/Perceptron>.

Una de sus características principales es la arquitectura. A mayor complejidad en la arquitectura de la red, el coste computacional aumenta y el nivel de abstracción crece. En este punto, se puede considerar que no existe una arquitectura mejor que otra, todo dependerá del tipo de problema que se requiere solucionar. La arquitectura se encuentra definida por:

el número de capas, número de neuronas por capa, y sus conexiones. En una red totalmente conectada, todas las neuronas de la capa predecesora tienen una conexión para cada neurona de la capa siguiente.

Capa de entrada: es una capa única de la red, y por donde se ingresan las características de los datos (una neurona por cada característica).

Capa oculta: no se tiene un número definido de capas y neuronas. Son aquellas que procesan los datos y brindan una salida esperada. En Deep Learning, existe una gran cantidad de capas ocultas como ya se mencionó anteriormente.

Capa de salida: es aquella que brinda la salida esperada, donde se tiene una neurona por cada clase que tenga la salida.

Si volvemos a observar la Figura 9, en la imagen del lado izquierdo apreciamos una red neuronal con tres capas. Una capa de entrada con 5 neuronas, una capa oculta con 6 neuronas, y una capa de salida con 4 neuronas. Asimismo, se encuentra totalmente conectada. Mientras, que en el lado derecho podemos observar una red neuronal pero más profunda: 4 capas ocultas, donde cada una tiene 6 neuronas, lo cual es solo por el ejemplo, ya que no necesariamente la cantidad de neuronas tienen que ser iguales.

Las principales clases de redes neuronales son: (Torres, 2020)

- Perceptrón multicapa - (MLP - Multi-layer Perceptron): similar a la imagen del lado derecho de la Figura 9.
- Redes Neuronales Convolucionales - (CNN - Convolutional Neural Networks): es una red con varios tipos de capas especiales, como lo veremos en la sección siguiente. En la actualidad es muy utilizada en la industria y de manera especial en: Visión por Computador.

- **Redes Neuronales Concurrentes (RNN - Recurrent Neural Networks):** este tipo de red contiene un estado interno (memoria) que se crea con los datos vistos por la red. Debido a ello son buenas candidatas para utilizarse con datos secuenciales, como texto o series de tiempo.

En la actualidad, con el aumento de la disponibilidad de datos y del poder computacional debido a la utilización de los GPU, se pueden procesar grandes redes neuronales con docenas de capas y millones de parámetros (pesos), obteniendo resultados cada vez más precisos. De esta manera, logrando algoritmos óptimos en tareas como: clasificación de imágenes, detección de objetos y segmentación semántica.

El aprendizaje profundo o Deep Learning está siendo utilizado con mayor frecuencia en el tema de Teledetección o Sensado Remoto, principalmente por la disponibilidad de grandes cantidades de datos, como imágenes de satélite provistas por las misiones Landsat y Sentinel, asimismo, por el buen rendimiento de algoritmos basados en CNN obteniendo resultados óptimos en clasificación de cultivos, segmentación semántica, detección y reconocimiento de objetos, entre otras. (Firigato, 2021)

## **2.5 Redes Neuronales Convolucionales**

### **2.5.1 Introducción**

Como se mencionó en la sección anterior, las CNN son usadas intensivamente en tareas de Visión por Computador. En esta sección explicaremos con un mayor detalle estas redes con el fin de comprender sus características.

Una red neuronal convolucional (CNN) es un caso concreto de redes neuronales que fueron usadas a finales de los 90 pero en los últimos años se han popularizado enormemente al conseguir resultados muy buenos en el reconocimiento de imagen, impactando de manera profunda en el área de visión por computador. Estas son muy similares a las redes neuronales

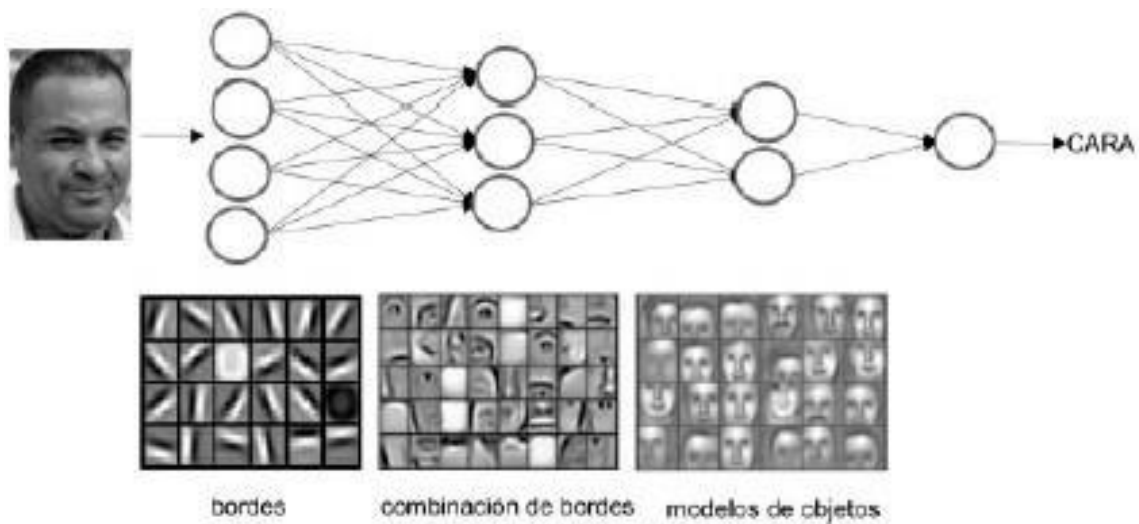
artificiales ya que están formadas por neuronas que tienen parámetros en forma de pesos y sesgos que se pueden aprender. Pero un rasgo diferencial de las CNN es que hacen la suposición explícita de que las entradas son imágenes, lo que nos permite codificar ciertas propiedades en la arquitectura para reconocer elementos concretos en las imágenes. (Torres, 2020)

Para hacernos una idea intuitiva sobre cómo funcionan estas redes neuronales, pensemos en cómo nosotros reconocemos las cosas. Por ejemplo, si vemos una cara, la reconocemos porque tiene orejas, ojos, nariz, cabello, etc. Entonces, para decidir si algo es una cara, lo hacemos como si tuviéramos unas casillas mentales de verificación de las características que vamos marcando. Algunas veces una cara puede no tener una oreja por estar cubierta por el cabello, pero igualmente la clasificamos como cara porque vemos los ojos, la nariz y la boca. En este caso, podemos ver a una red neuronal convolucional como un clasificador, que predice una probabilidad de que la imagen de entrada sea una cara o no. Pero, en realidad, antes de poder clasificar debemos saber cómo es una oreja o una nariz para saber si están en una imagen; es decir, previamente debemos identificar líneas, bordes, texturas o formas que sean similares a las que contienen las orejas o narices que hemos visto antes. Esto es lo que las capas de una red neuronal convolucional tienen encomendado hacer. Pero identificar estos elementos no es suficiente para poder decir que una imagen es una cara. Además, debemos poder identificar cómo las partes de una cara se encuentran colocadas entre sí, tamaños relativos, etc. De lo contrario, la cara no se parecería a lo que estamos acostumbrados. Es decir, una cara está formada por una boca, una nariz y dos ojos, pero la disposición entre ellos es importante, ya que si la misma no es la adecuada —por ejemplo, si la nariz se encuentra por debajo de la boca— no podemos considerar la imagen como una cara, a pesar de contener todos los elementos que conforman una cara. (Torres, 2020)

Podemos ver una idea de lo comentado en la Figura 11. Lo que se quiere expresar

es que cada capa de una CNN va aprendiendo diferentes niveles de abstracción. Una primera capa CNN va aprendiendo elementos básicos como aristas, y una segunda aprende patrones compuestos de elementos básicos aprendidos en la capa anterior. Y de esta manera, sucesivamente, en cada capa hasta ir aprendiendo patrones más complejos.

**Figura 11**  
*Aprendizaje de una CNN*



*Nota.* Tomado de *Aprendizaje de una CNN*, por Torres, 2020, Python Deep Learning - Introducción práctica con Keras y TensorFlow 2.

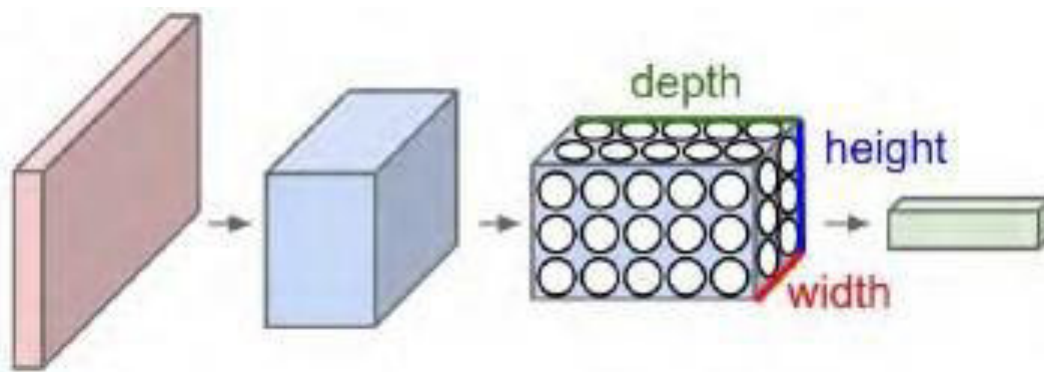
La CNN nos ayuda a procesar datos que presentan una tipología similar a una cuadrícula. Por ejemplo, en datos de series de tiempo, puede darse como una cuadrícula 1-D tomando muestras a intervalos de tiempo regulares, y en datos de imágenes, se puede considerar como una cuadrícula 2-D de píxeles. También, se le puede denominar como ConvNet para referirse a ellas.

Las CNN tienen un mejor desempeño sobre imágenes del mundo real en comparación con las clásicas Redes Neuronales Artificiales (RNA) como los MLP, y esto se debe a que ellas como mencionamos en párrafos anteriores diferencian patrones de datos multidimensionales. Las CNN organizan sus neuronas en tres dimensiones: ancho, altura, y profundidad. Cada capa

transforma su volumen de entrada 3D en un volumen de neuronas de salida 3D utilizando funciones de activación. Como vemos en la Figura 12, la capa de entrada roja contiene la imagen, por dicha razón su ancho y alto son las primeras dos dimensiones de la imagen, y la profundidad sería la tercera, debido a los canales en rojo, verde y azul de la imagen. (Chicchón, 2018)

**Figura 12**

*Esquema de una CNN*



*Nota.* Tomado de *Esquema de una red neuronal artificial y modelo de una neurona*, por Chicchón, 2018, Fusión de Datos para Segmentación Semántica en Aplicaciones Urbanas de Teledetección Aérea usando Algoritmos de Aprendizaje Profundo.

### 2.5.2 Componentes

Ahora que hemos revisado cómo clasifican las imágenes las CNN, veremos un ejemplo de reconocimiento de dígitos del conjunto de datos MNIST, e introduciremos las dos capas que definen las CNN, que se pueden delinear como grupo de neuronas especializadas en dos operaciones: convolución y pooling. (Torres, 2020)

#### **Operación de Convolución**

La diferencia fundamental entre una capa densamente o totalmente conectada y una capa especializada en la operación de convolución, que llamaremos capa convolucional, es

que la capa densa aprende patrones globales en su espacio global de entrada, mientras que la capa convolucional aprende patrones locales dentro de la imagen en pequeñas ventanas de dos dimensiones. De manera intuitiva, podríamos decir que el propósito principal de una capa convolucional es detectar características o rasgos visuales en las imágenes, como aristas, líneas, gotas de color, etc. Esta es una propiedad muy interesante porque una vez aprendida una característica en un punto concreto de la imagen, la puede reconocer después en cualquier parte de la misma. En cambio, una red neuronal densamente conectada tiene que aprender el patrón nuevamente si este aparece en una nueva localización de la imagen.

Otra característica importante es que las capas convolucionales pueden aprender jerarquías espaciales de patrones, preservando así relaciones espaciales, como hemos visto en el ejemplo intuitivo de la Figura 12. Esto permite que las redes neuronales convolucionales aprendan eficientemente conceptos visuales cada vez más complejos y abstractos.

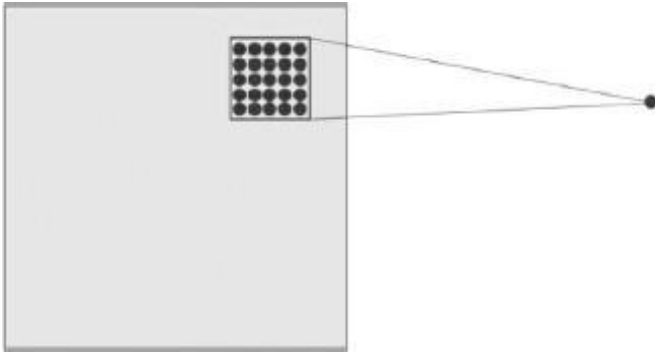
En general, las capas convolucionales operan sobre tensores 3D, llamados mapas de características (*feature maps* en inglés), con dos ejes espaciales de altura y anchura (*height* y *width*), además de un eje de canal (*channels*) también llamado profundidad (*depth*). Para una imagen de color RGB, la dimensión del eje *depth* es 3, pues la imagen tiene tres canales: rojo, verde y azul (*red*, *green* y *blue*). Para una imagen en blanco y negro, como es el caso de los dígitos MNIST, la dimensión del eje *depth* es 1 (nivel de gris). (Torres, 2020)

En el caso de los dígitos MNIST, como entrada en nuestra red neuronal podemos pensar en un espacio de neuronas de dos dimensiones  $28 \times 28$ , que transformaremos en un tensor 3D (*height* = 28, *width* = 28, *depth* = 1), aunque la tercera dimensión en este caso sea de tamaño 1. Una primera capa de neuronas ocultas conectadas a las neuronas de la capa de entrada que hemos comentado realizará las operaciones convolucionales que acabamos de describir. Pero, como hemos avanzado, no se conectan todas las neuronas de entrada con todas las neuronas

de este primer nivel de neuronas ocultas (como en el caso de las redes neuronales densamente conectadas); solo se hace por pequeñas zonas localizadas del espacio de las neuronas de entrada que almacenan los píxeles de la imagen. Lo podemos observar en la Figura 13.

### Figura 13

*Conexión de neurona de capa oculta con capa de entrada*



*Nota.* Tomado de *Conexión de neurona de capa oculta con capa de entrada*, por Torres, 2020, Python Deep Learning - Introducción práctica con Keras y TensorFlow 2.

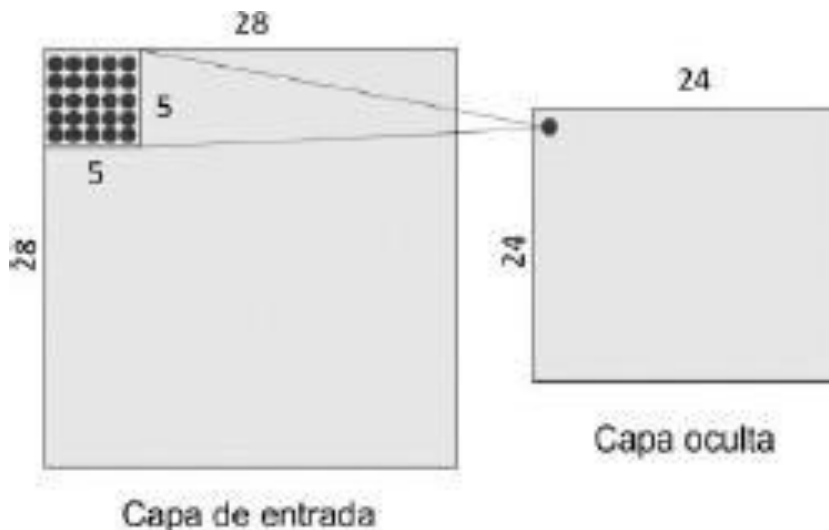
Como podemos apreciar en el ejemplo, cada neurona de la capa oculta será conectada a una pequeña región de  $5 \times 5$  neuronas (es decir, 25 neuronas) de la capa de entrada (de  $28 \times 28$ ). Intuitivamente, se puede pensar en una ventana del tamaño de  $5 \times 5$  que va recorriendo toda la capa de  $28 \times 28$  que contiene la imagen. Esta ventana va deslizándose a lo largo de toda la capa de neuronas. Por cada posición de la ventana hay una neurona en la capa oculta que procesa esta información. La ventana empieza en la esquina superior izquierda de la imagen, y esto le da la información necesaria a la primera neurona de la capa oculta; ver Figura 14.

A continuación, la ventana se desliza una posición hacia la derecha para «conectar» las  $5 \times 5$  neuronas de la capa de entrada incluidas en esta ventana con la segunda neurona de la capa oculta. Y así, sucesivamente, va recorriendo todo el espacio de la capa de entrada, de izquierda a derecha y de arriba abajo. Analizando un poco el ejemplo concreto que hemos propuesto, observemos que si tenemos una entrada de  $28 \times 28$  píxeles y una ventana de  $5 \times 5$ , nos define un

espacio de  $24 \times 24$  neuronas en la primera capa oculta, debido a que la ventana solo se puede desplazar 23 neuronas hacia la derecha y 23 hacia abajo antes de chocar con el lado derecho (o inferior) de la imagen de entrada.

#### Figura 14

Entrada de  $28 \times 28$  con ventana de  $5 \times 5$  que genera una salida de  $24 \times 24$  neuronas en la capa oculta



*Nota.* Tomado de *Entrada de  $28 \times 28$  con ventana de  $5 \times 5$  que genera una salida de  $24 \times 24$  neuronas en la capa oculta*, por Torres, 2020, Python Deep Learning - Introducción práctica con Keras y TensorFlow 2.

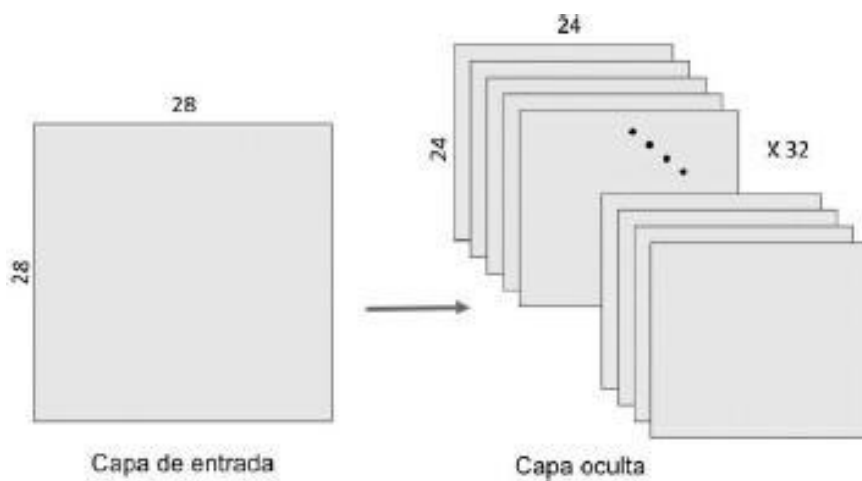
Para «conectar» cada neurona de la capa oculta con las 25 neuronas que le corresponden de la capa de entrada usaremos un valor de sesgo  $b$  y una matriz de pesos  $W$  de tamaño  $5 \times 5$ , que llamaremos filtro (o *kernel* y *filter* en inglés). El valor de cada punto de la capa oculta corresponde al producto escalar entre el filtro y el conjunto de 25 neuronas ( $5 \times 5$ ) de la capa de entrada. Ahora bien, lo particular y muy importante de las redes convolucionales es que se usa el mismo filtro (la misma matriz  $W$  de pesos y el mismo sesgo  $b$ ) para todas las neuronas de la capa oculta: en nuestro caso para las  $24 \times 24$  neuronas (576 neuronas en total) de la primera capa oculta. Se puede comprobar que esta forma de compartir reduce drásticamente el número

de parámetros que tendría una red neuronal si no la hiciéramos: pasa de 14400 parámetros que tendrían que ser ajustados ( $5 \times 5 \times 24 \times 24$ ) a  $25(5 \times 5) = 625$  parámetros más los sesgos  $b$ .

Finalmente, la convolución es el proceso matemático sobre una matriz de entrada por otra que llamamos filtro. Pero un filtro definido por una matriz  $W$  y un sesgo  $b$  solo permiten detectar una característica concreta en una imagen. Por tanto, para poder realizar el reconocimiento de imágenes se propone usar varios filtros a la vez, uno para cada característica que queramos detectar. Por eso una capa convolucional completa en una red neuronal convolucional incluye varios filtros. Una manera habitual de representar visualmente esta capa convolucional es la que se muestra en la Figura 15, donde se visualiza que la capa convolucional está compuesta por varios filtros.

**Figura 15**

*Capa CNN compuesta por varios filtros*



*Nota.* Tomado de *Capa CNN compuesta por varios filtros*, por Torres, 2020, Python Deep Learning - Introducción práctica con Keras y TensorFlow 2.

En nuestro ejemplo proponemos 32 filtros, donde cada filtro se define con una matriz  $W$  de pesos compartida de  $5 \times 5$  y un sesgo  $b$ . En este ejemplo, la primera capa convolucional recibe un tensor de entrada de tamaño  $(28, 28, 1)$  y genera una salida de tamaño  $(24, 24, 32)$ ,

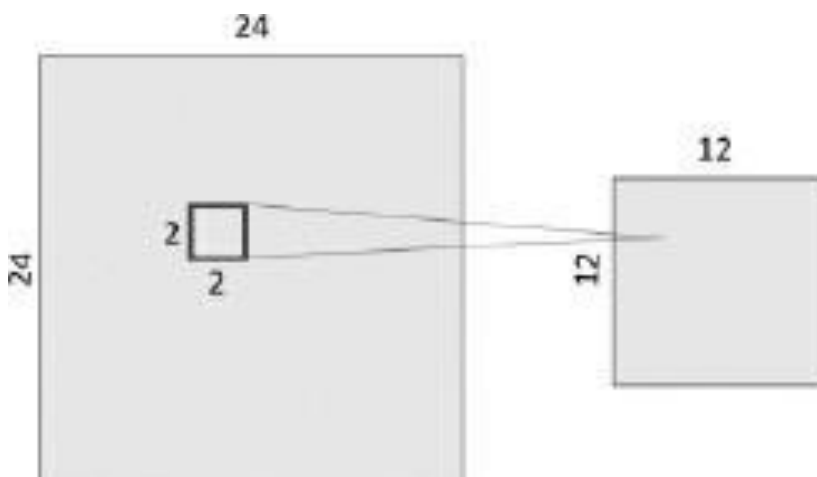
un tensor 3D que contiene las 32 salidas de  $24 \times 24$  píxeles resultado de computar los 32 filtros sobre la entrada.

### Operación de Pooling

Adicionalmente de las capas convolucionales que acabamos de describir, las redes neuronales convolucionales acompañan a la capa de convolución con unas capas de pooling —que podríamos traducir por agrupación—, que suelen ser aplicadas inmediatamente después de las capas convolucionales. Una primera aproximación para entender para qué sirven estas capas es considerar que las capas de pooling hacen una simplificación de la información recogida por la capa convolucional y crean una versión condensada de la información contenida en esta capa. En nuestro ejemplo de dígitos MNIST, vamos a escoger una ventana de  $2 \times 2$  sobre la capa convolucional y sintetizar la información en la capa de pooling; ver Figura 16.

#### Figura 16

*Capa Pooling construida desde una ventana  $2 \times 2$  que se aplica a una capa CNN de  $24 \times 24$*



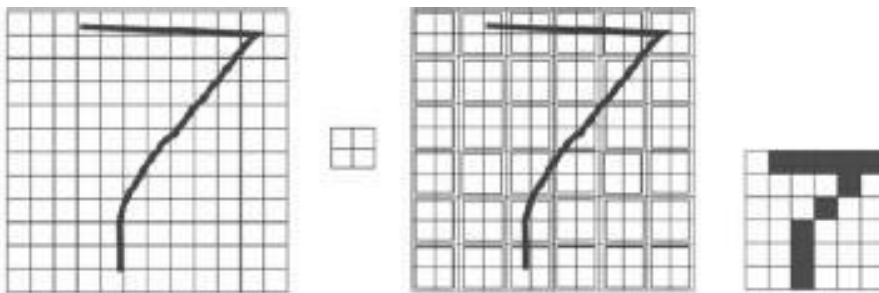
*Nota.* Tomado de *Capa Pooling construida desde una ventana  $2 \times 2$  que se aplica a una capa CNN de  $24 \times 24$* , por Torres, 2020, Python Deep Learning - Introducción práctica con Keras y TensorFlow 2.

Hay varias maneras de condensar la información, pero una habitual y que usaremos en

nuestro ejemplo es la conocida como *max-pooling*. Como valor, se queda con el valor máximo de los datos en la ventana de entrada de  $2 \times 2$  que, en nuestro caso, ha sido «dividido» en  $12 \times 12$  ventanas la capa de pooling. En este caso, se divide por 4 el tamaño de la salida de la capa de pooling en relación a la capa convolucional donde se aplica el pooling, y queda con tamaño de  $12 \times 12$ . También se puede utilizar *average-pooling* en lugar de *max-pooling*, donde cada grupo de puntos de entrada se transforma en el valor promedio del grupo de puntos, en vez de su valor máximo. Pero, en general, el *maxpooling* tiende a funcionar muy bien. Es interesante remarcar que con la transformación de *pooling* mantenemos la relación espacial. Para verlo visualmente, cogemos el siguiente ejemplo de una matriz de  $12 \times 12$  donde tenemos representado un 7 (imaginemos que los píxeles por los que pasamos por encima contienen un 1 y el resto 0; no lo hemos añadido al dibujo para simplificar su visualización). Esto se representa visualmente en la Figura 17.

**Figura 17**

*Transformación de pooling mantiene la proporción inicial*



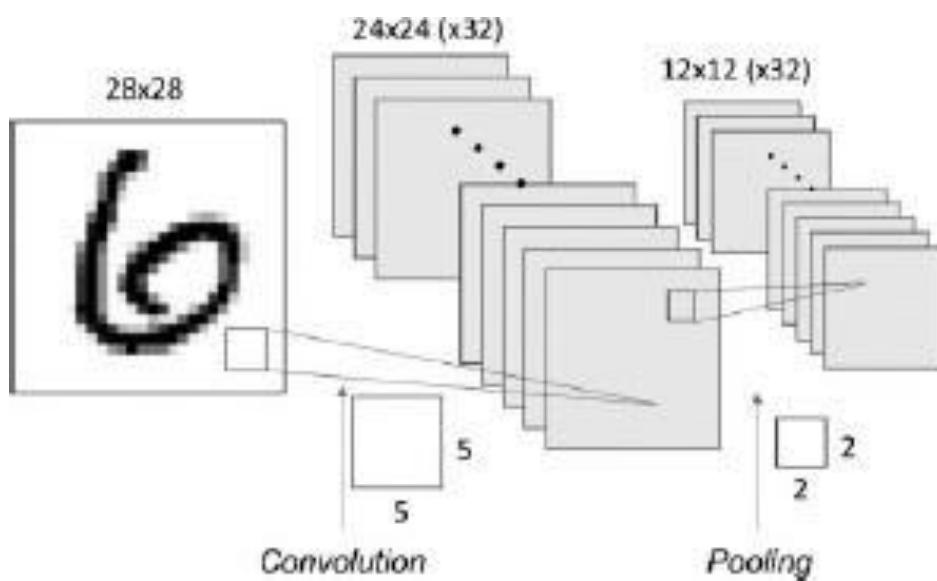
*Nota.* Tomado de *Transformación de pooling mantiene la proporción inicial*, por Torres, 2020, Python Deep Learning - Introducción práctica con Keras y TensorFlow 2.

Si aplicamos una operación de *max-pooling* con una ventana de  $2 \times 2$  (lo representamos en la matriz central que divide el espacio en un mosaico con regiones del tamaño de la ventana), obtenemos una matriz de  $6 \times 6$  donde se mantiene una representación que nos recuerda sin ninguna duda al número 7 (lo podemos ver en la imagen de la derecha, donde hemos marcado

en blanco los ceros y en negro los puntos con valor 1).

Como hemos mencionado anteriormente, la capa convolucional alberga más de un filtro y, por tanto, como aplicamos el *max-pooling* a cada uno de estos filtros separadamente, la capa de pooling contendrá tantos filtros de pooling como filtros convolucionales había, tal como se representa en la Figura 18.

**Figura 18**  
*Capa CNN y pooling*



*Nota.* Tomado de *Capa CNN y pooling*, por Torres, 2020, Python Deep Learning - Introducción práctica con Keras y TensorFlow 2.

Dado que teníamos un espacio de  $24 \times 24$  neuronas en cada filtro convolucional, después de hacer el *pooling* tenemos  $12 \times 12$  neuronas, que corresponden a las  $12 \times 12$  regiones de tamaño  $2 \times 2$  que aparecen al dividir el espacio de neuronas del espacio del filtro de la capa convolucional y este por los 32 filtros convolucionales.

## 2.6 Segmentación Semántica

### 2.6.1 Introducción

La segmentación semántica es un tipo de segmentación de imagen. Este tipo se entiende como la clasificación a nivel de píxel de una imagen, o de un grupo reducido de píxeles, según una cierta categoría, pudiendo pertenecer a un objeto o a una parte de este. Por ejemplo, tronco y hojas de un árbol son categorías diferentes del mismo objeto. (Pajares et al., 2021)

La resolución de problemas por segmentación es una extensión del problema de detección de objetos. La segmentación es muy útil en aplicaciones para el entendimiento de imágenes médicas y de satélite. Es la tarea de predecir píxeles por etiquetas. En la Figura 19 vemos un ejemplo de una imagen y su correspondiente segmentación semántica. (Shanmugamani, 2018)

Como se observa en la Figura 19, apreciamos una imagen de entrada que es predicha con etiquetas para cada píxel. Las etiquetas o clases son: el cielo (sky), mar (sea), persona (person), montaña (mountain), y puente (bridge). Antes que asignar una etiqueta a toda la imagen, estas son asignadas a cada píxel. La segmentación semántica etiqueta los píxeles de forma independiente. Se aprecia que no todas las personas se distinguen. Todas las personas de la imagen son etiquetadas de la misma manera.

Para diferenciar lo último vemos en la Figura 20, el mismo ejemplo anterior donde cada instancia de una misma etiqueta se distingue (persona).

La tarea de segmentar cada instancia con etiquetado a nivel de píxel es llamada: segmentación de instancia. Este tipo de segmentación también se puede considerar como una extensión de la detección de objetos con etiquetas a nivel de píxel. Las aplicaciones de ambos tipos de segmentación: semántica e instancia, son muchas. En este estudio veremos la

aplicación de la segmentación semántica para la comprensión de las imágenes satelitales.

### Figura 19

*Imagen y su segmentación semántica*



*Nota.* Tomado de *Imagen y su segmentación semántica*, por Shanmugamani, 2018, Deep Learning for Computer Vision.

### Comprender la Tierra a partir de imágenes satelitales

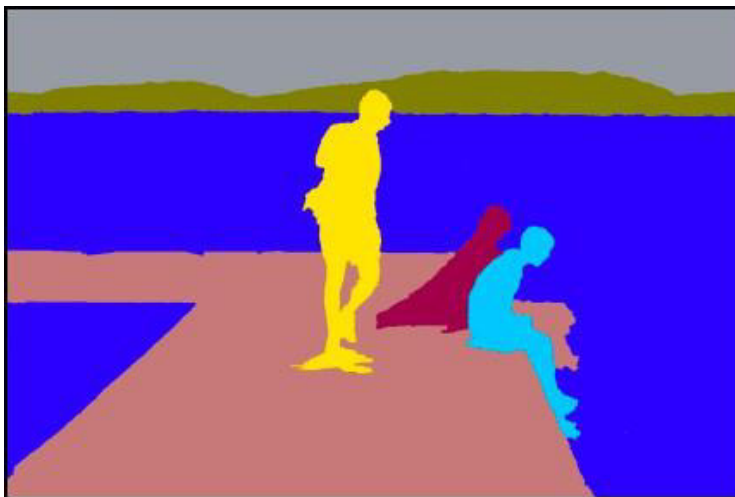
Las imágenes satelitales como se mencionó anteriormente han aumentado en los últimos años. Se tiene satélites tanto de acceso libre y sin costo, así como de tipo comercial. Las imágenes capturadas por satélites nos brindan una vista de alta resolución de la superficie terrestre. Analizando las mismas podemos entender varios aspectos de la Tierra como:

- Medición del nivel de infraestructura construida relacionado al crecimiento económico

- Medición de tanques de combustible
- Planificación y organización del tráfico
- Cálculo de la deforestación y sus efectos
- Ayudando a la preservación de la vida silvestre contabilizando los animales y rastreando sus movimientos
- Descubriendo sitios arqueológicos
- Mapeo de regiones dañadas debido a desastres naturales

### Figura 20

#### *Segmentación de instancia*



*Nota.* Tomado de *Segmentación de instancia*, por Shanmugamani, 2018, Deep Learning for Computer Vision.

Hay más aplicaciones posibles, para la mayoría de ellas la solución inicia con la segmentación de estas imágenes. Este estudio trata sobre la segmentación semántica de imágenes satelitales para el mapeo o delimitación de áreas o superficies quemadas por incendios forestales.

### 2.6.2 Arquitecturas

Un enfoque de ventana deslizante puede ser aplicado a nivel del píxel para segmentación. Este enfoque de ventana deslizante toma una imagen y divide la misma en pequeñas secciones. Cada sección de la imagen es clasificada por una etiqueta. Este enfoque es ineficiente y costoso debido a que no reusa las características compartidas entre secciones superpuestas. A continuación, describiremos dos modelos de segmentación semántica en imágenes mediante las CNN que aplicaremos en este estudio.

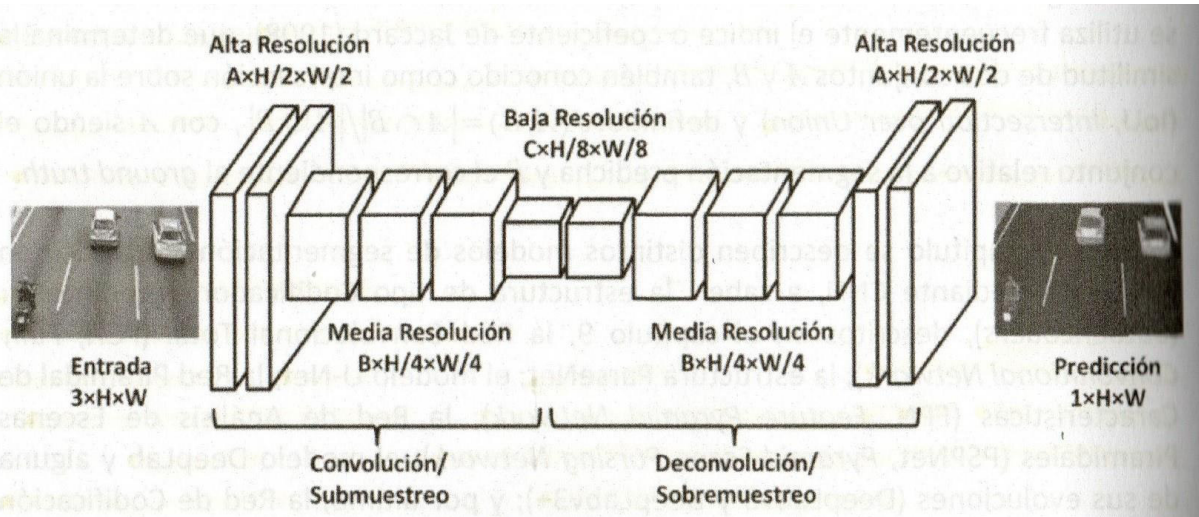
#### **Codificador - Decodificador (Autoencoder)**

Un esquema general de segmentación semántica es una estructura codificador - decodificador (Encoder-Decoder), denominado *Autoencoder*. El planteamiento adaptado a la segmentación semántica, consiste en construir una secuencia de capas convolucionales que realizan un submuestreo (contracción) y sobremuestreo (expansión) (*downsampling-upsampling*) y en este orden, de forma de que en primer lugar se reducen las resoluciones espaciales progresivamente,  $(H \times W)$ ,  $(H/2 \times W/2)$ ,  $(H/4 \times W/4)$  hasta un determinado nivel  $(H/8 \times W/8)$  mediante operaciones de convolución y pooling, para luego incrementar las resoluciones en el orden inverso por medio de operaciones de deconvolución y unpooling. En la parte de submuestreo se capturan principalmente las características discriminativas a través de los correspondientes mapas de características para luego llegar a la predicción semántica a resolución completa. En cada uno de esos niveles aparecen mapas de características con un número determinado de filtros (A, B, C). En definitiva, se trata de un planteamiento multiescala que codifica la información contextual a través de una serie de filtros u operaciones de pooling, mientras que las capas posteriores capturan características más finas tales como detalles y bordes existentes en los objetos. En la Figura 21 observamos una estructura típica de un autoencoder, con una imagen de entrada y salida de la misma

resolución. Además, en este caso, ambas imágenes son diferentes y se necesita un conjunto de datos etiquetados (ya que durante el entrenamiento la salida de la red se compara con la imagen segmentada verdadera disponible). (Pajares et al., 2021)

**Figura 21**

*Red Autoencoder*



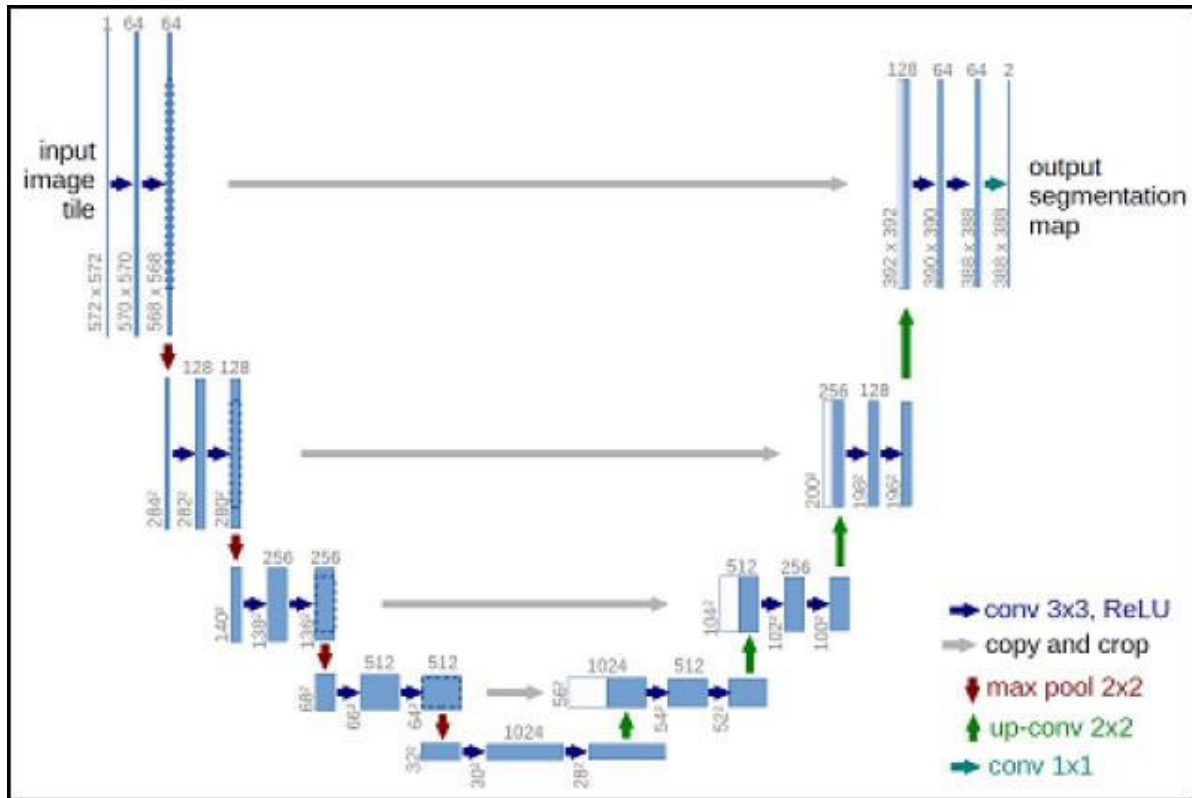
*Nota.* Tomado de *Esquema general Codificador-Decodificador*, por Pajares, 2021, Aprendizaje Profundo.

## UNET

Este modelo fue propuesto en 2015, para imágenes microscópicas biológicas, es en realidad una extensión del encoder-decoder (Autoencoder). La red se compone de las dos típicas partes de submuestreo y sobremuestreo, tal y como lo vemos en la Figura 22. Una modificación importante en relación con las arquitecturas anteriores se da en la parte de sobremuestreo o expansión donde aparecen una serie de operaciones y mapas de características intermedios, que permiten la propagación de información contextual de la imagen hasta las capas de mayor resolución. Como consecuencia, la ruta expansiva es más o menos simétrica a la ruta de contracción y produce una arquitectura de U. La red carece de capas completamente conectadas y solo usa la parte válida de cada convulsión, es decir, solo la parte del mapa de

segmentación que contiene los píxeles para los cuales el contexto completo está disponible en la imagen de entrada. (Pajares et al., 2021)

**Figura 22**  
Red Unet



*Nota.* Tomado de *Architecture of the Unet model*, por Shanmugamani, 2018, Deep Learning for Computer Vision.

De forma más concreta, la ruta de contracción sigue la arquitectura típica de una red convolucional. Consiste en la aplicación repetida de dos convoluciones de dimensiones 3x3, sin padding, cada una seguida de una ReLU (función de activación) y una operación de pooling (máximo) de dimensión 2x2 con desplazamiento (stride) de paso 2 para disminuir la resolución. En cada paso de reducción se duplica el número de mapas de características. Cada paso en la ruta expansiva consiste en un sobremuestreo del mapa de características seguido de una deconvolución de tipo 2x2 que reduce a la mitad el número de canales de características,

una concatenación con el mapa de características recortado que se corresponde con el mapa de la ruta de contracción (son los recuadros que aparecen en blanco en la Figura 22), y dos convoluciones de tamaño 3x3, cada una seguida de una ReLU. El recorte es necesario debido a la pérdida de píxeles de tamaño 1x1 para asignar cada vector de características de 64 componentes al número deseado de clases. En total, la red tiene 32 capas convolucionales. (Pajares et al., 2021)

### 2.6.3 Métricas de validación

En particular, se definen distintas métricas para determinar el desempeño de los modelos de segmentación. Si consideramos un clasificador binario donde cada categoría tiene asignada una etiqueta ( $x$ ), Verdadera (V - True T) o Falsa (F - False F), y realiza una asignación sobre la categoría ( $y$ ), también V o F. Con ellos se caracterizan los siguientes valores:

- VP (Verdadero Positivo - True Positive - TP) con  $x = V$  e  $y = V$
- VN (Verdadero Negativo - True Negative - TN) con  $x = F$  e  $y = F$
- FP (Falso Positivo - False Positive - FP) con  $x = F$  e  $y = V$
- FN (Falso Negativo - False Negative - FN) con  $x = V$  e  $y = F$

A partir de aquí se pueden construir las denominadas matrices de confusión, incluyendo los sistemas multiclase. En la Figura 23 podemos apreciar una matriz de confusión binaria.

Además, podemos definir las siguientes medidas de desempeño.

- Exactitud - Accuracy =  $(VP + VN) / (VP + FP + FN + VN)$
- Precisión - Precision (P) =  $(VP) / (VP + FP)$
- Sensibilidad - Recall (R) =  $(VP) / (VP + FN)$
- Especificidad - Specificity (S) =  $(VN) / (VN + FP)$

- F-score -  $F1 = (2PR)/(P + R)$

**Figura 23**

*Matriz de Confusión Binaria*

		Clase Predicha	
		Positivo -1	Negativo - 0
Clase Real	Positivo - 1	Verdadero Positivos (VP - TP)	Falso Negativos (FN)
	Negativo - 0	Falsos Positivos (FP)	Verdaderos Negativos (VN - TN)

Exactitud: es la fracción de predicciones correctas sobre el total de observaciones, siendo apropiada cuando existe equilibrio entre observaciones verdaderas y falsas.

Precisión: es la fracción de predicciones verdaderas realizadas.

Sensibilidad: es la razón de predicciones verdaderas detectadas.

F-score F1: es un promedio ponderado de P y R.

Debemos tener mucho cuidado con las métricas mencionadas. En otras tareas se utiliza la Exactitud en forma mayoritaria. Esta métrica mide el número de píxeles acertados sobre el número de píxeles total. No obstante, no es la mejor métrica para la segmentación, ya que es muy habitual que el número de píxeles pertenecientes a cada clase esté muy desbalanceado y como sabemos, las medidas de exactitud en estos escenarios son poco informativas. Debido a lo mencionado se utiliza otra métrica que pueda medir la segmentación de una manera más adecuada:

**IoU o Índice de Jaccard**

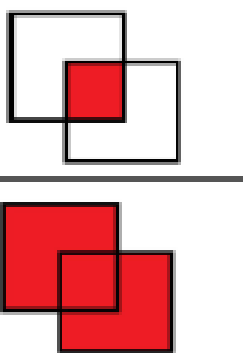
Es la tasa entre el área de la intersección de dos cajas y la unión de las dos cajas. Si dos cajas son exactamente coincidentes este valor será 1, mientras que si no se solapan el valor es 0.

$$\text{IoU} = (\text{Área de la intersección}) / (\text{Área de la unión})$$

En la Figura 24 se observa este concepto.

**Figura 24**

*Métrica IoU - Índice de Jaccard*

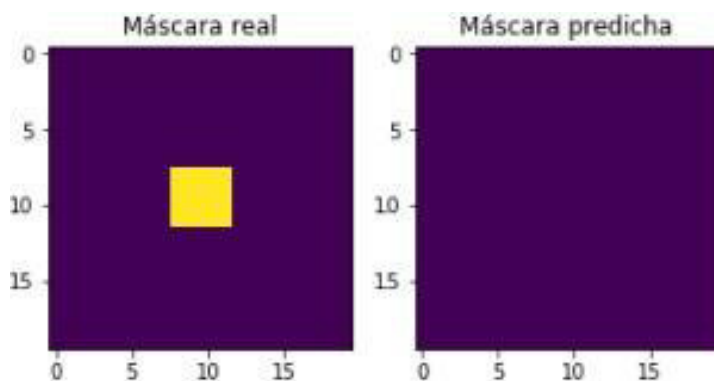
$$\text{IoU} = \frac{\text{Área de intersección de dos cajas blancas}}{\text{Área de intersección de dos cajas rojas}}$$


La IoU se calcula para cada clase presente en el problema y posteriormente se promedian dichos valores.

Veamos un ejemplo para apreciar su utilidad, ver Figura 25.

**Figura 25**

*Ejemplo de métrica IoU*



Imaginemos que tenemos un modelo de segmentación de dos clases en imágenes de

20x20 píxeles que genera la máscara indicada en la Figura 24. Como vemos, predice toda la imagen de una sola clase (clase1). Mientras que la imagen original contiene un área cuadrada de 4 píxeles de la clase2.

La exactitud sería calculada como:

$$\text{Exactitud} = (\text{Píxeles bien clasificados}) / (\text{Píxeles totales})$$

$$\text{Exactitud} = (384) / (400) = 0.96$$

El modelo tendría una Exactitud del 96%

Calculemos la IoU. Se calcula el índice para cada clase y luego promediamos.

$$\text{Intersección clase 1} = 384, \text{ Unión clase 1} = 400$$

$$\text{IoU (clase1)} = (384) / (400) = 0.96$$

$$\text{Intersección clase 2} = 0, \text{ Unión clase 2} = 16$$

$$\text{IoU (clase2)} = (0) / (16) = 0$$

La IoU total sería la media:

$$\text{IoU} = (0.96 + 0) / 2 = 0.48$$

Se observa que la IoU es una métrica más resistente al desbalanceo de clases.

Para un clasificador binario (por ejemplo, píxeles quemados y píxeles no quemados) podemos también expresar de la forma:

$$\text{IoU} = (\text{VP}) / (\text{VP} + \text{FN} + \text{FP})$$

Debido a las propiedades de los datos de observación de la Tierra, las tareas de reconocimiento de imágenes, segmentación de imágenes, y detección de objetos pueden ser consideradas como las más retadoras. Para estos problemas, los modelos de Deep Learning que

fueron afinados para conjuntos de datos de Visión por Computador se ajustan con precisión a los conjuntos de datos de observación de la Tierra. Estos datos de observación de la Tierra son a menudo pequeños y es una práctica común refinar modelos que ya han aprendido a extraer características jerárquicamente de datos de imágenes de Visión por Computador más grandes. Este refinamiento de modelos, originalmente entrenados con otros conjuntos de datos, se conoce como: Transferencia de aprendizaje. Sin embargo, aún con este refinamiento, el conjunto de datos de observación de la Tierra tiene que ser creado. En segmentación de imágenes, los modelos Autoencoder dominan su utilización, debido a su diseño complejo pero modular la cual puede ser adaptado a las propiedades de los datos de observación de la Tierra. (Hoeser y Kuenzer, 2020)

Una métrica que también se utiliza bastante en clasificaciones de imágenes por teledetección es el índice Kappa. Con base en los valores de concordancia - como vemos en la Tabla 5, los valores totales de las métricas mencionadas anteriormente serán evaluadas para conseguir un rendimiento satisfactorio en la clasificación supervisada por píxeles. (Firigato, 2021)

**Tabla 5**

*Criterios de grado de concordancia basado en índice Kappa*

<b>Valor de Kappa</b>	<b>Concordancia</b>
$K \leq 0$	Pésimo
$0.0 < K \leq 0.2$	Malo
$0.2 < K \leq 0.4$	Razonable
$0.4 < K \leq 0.6$	Bien
$0.6 < K \leq 0.8$	Muy bien
$0.8 < K \leq 1.0$	Excelente

*Nota.* Fuente: (Congalton y Green, 2019)

### III. MÉTODO

#### 3.1 Tipo de Investigación

La investigación a implementar será de tipo: cuantitativa, aplicada, descriptiva, y de tipo no experimental; que nos permitirá evaluar el rendimiento del algoritmo propuesto. Pasamos a describir los puntos indicados. (Hernández et al., 2014)

Es de tipo cuantitativa ya que es necesario describir, analizar o experimentar con las variables de estudio; basado en aportes científicos de la referencia se pretende brindar un aporte científico para la delimitación de las áreas afectadas o quemadas.

Es de tipo aplicada ya que basados en un marco teórico, lo aplicamos al contexto de nuestro territorio en el problema de la delimitación de las áreas quemadas en forma concreta. Debido a ello esta investigación es una aplicación inmediata.

Es de tipo no experimental ya que las recolecciones de datos se toman de dos momentos de tiempos diferentes, uno previo y otros post-incendio forestal. Asimismo, los polígonos conformantes de las áreas quemadas utilizan estos datos para ser generados. La investigación propone un algoritmo utilizando solo un dato post-incendio forestal con la finalidad de la delimitación de los polígonos en mención (cicatrices).

#### 3.2 Población y muestra

La población se encuentra determinada por todos los departamentos de nuestro país debido a que todos presentan ecosistemas asociados al patrimonio forestal y de fauna silvestre.

El tamaño de la muestra se encuentra determinada por las cicatrices del departamento de Cusco del período 2017 al 2021. Dicha muestra es de un área afectada de aproximadamente 12000 hectáreas, y cubierta por alrededor de 72 imágenes satelitales.

#### 3.3 Operacionalización de variables

##### Variable Independiente

Un modelo generado mediante técnicas de aprendizaje profundo utilizando segmentación semántica

#### *Dimensiones*

- Precisión

#### *Indicadores*

- Métrica Precisión

### **Variable Dependiente**

Delimitación de la superficie afectada (cicatriz) realizada por un modelo de aprendizaje profundo

#### *Dimensiones*

- Calidad general
- Certeza sobre área predicha
- Sensibilidad
- Certeza sobre área afectada

#### *Indicadores*

- Métrica F-Score o Coeficiente Dice
- Métrica IoU o índice Jaccard
- Métrica Recall o Sensibilidad
- Coeficiente Kappa

### 3.4 Instrumentos

- Lenguaje de programación Python: herramienta tecnológica que permite desarrollar aplicaciones multiplataforma de código abierto
- Librerías Tensorflow, Keras, y otras: bibliotecas de software de código abierto para Python. A través de los mismos se implementará el modelo de Deep Learning, así como su entrenamiento y validación
- Imágenes Satelitales brindadas por una misión de observación terrestre implementada por la Agencia Espacial Europea (ESA - European Space Agency) dentro del programa Copérnico. Estas imágenes son obtenidas en forma gratuita. Se utilizarán aquellas proporcionadas por los satélites: Sentinel-2A y Sentinel-2B.
- Imágenes Satelitales brindadas por una misión de observación terrestre implementada por la Administración Nacional de Aeronáutica y el Espacio (NASA - National Aeronautics and Space Administration) y el Servicio Geológico de los Estados Unidos (USGS - United States Geological Survey) dentro del proyecto Landsat. Estas imágenes son obtenidas en forma gratuita. Se utilizarán aquellas proporcionadas por el satélite: Landsat-8.
- Software de Procesamiento Digital de Imágenes para el tratamiento de imágenes satelitales para la generación de las cicatrices
- Software de Sistema de Información Geográfica para la generación de los archivos digitales de las cicatrices

### 3.5 Procedimientos

En base al marco teórico mencionado, se procede con los siguientes pasos:

- Descarga de las imágenes satelitales con fecha más cercana a la extinción del incendio

forestal desde la plataforma web para los satélites mencionados

- Se realiza una etapa preprocesamiento para su adecuación al modelo
- Se generan tres conjuntos de datos: entrenamiento, validación y prueba
- Se entrena el modelo con los conjuntos de: entrenamiento y validación
- Se evalúa el modelo con el conjunto de prueba
- Se calculan las métricas del modelo, y de ser necesario se vuelve a la etapa de entrenamiento para optimizar el modelo

### **3.6 Análisis de datos**

Los datos a ser utilizados consisten de dos grupos, en donde debido a sus características su tratamiento es un desafío:

- Imágenes satelitales
- Cicatrices de áreas afectadas o quemadas

#### **Imágenes satelitales**

Las imágenes son procesadas por software de Procesamiento Digital de Imágenes o de Sistemas de Información Geográfica. En este caso, se utiliza el software QGIS, el cual nos permite apilar las bandas, luego remuestrear a una resolución de 10m, y finalmente extraer una zona rectangular en donde se encuentre la cicatriz del incendio forestal.

Luego de obtenida la imagen recortada, se divide la misma en pequeños recortes de tamaño de 64, 128 o 256 filas y columnas. Dando como resultado imágenes de tamaño cuadrado. Estas son ya ingresadas al modelo para su entrenamiento. Recordar que el entrenamiento es de tipo aprendizaje supervisado por lo que se requieren las salidas verdaderas (etiquetas) de las cicatrices.

### **Cicatrices de áreas afectadas o quemadas**

Los archivos digitales de formato vectorial de las cicatrices proporcionados por la UFMS de la DGIOFFS de SERFOR, son convertidos a formato raster con la finalidad de que también ingresen al modelo. Dicha conversión se realiza tomando en consideración el tamaño de su correspondiente imagen satelital recortada que contiene a la cicatriz. Esto con la finalidad de que ambos archivos raster tengan las mismas dimensiones de filas por columnas. Este archivo raster obtenido también se le conoce con el nombre de “máscara de la cicatriz”.

Posteriormente, los archivos raster de las cicatrices se dividen de igual manera que el paso anterior en pequeñas imágenes de tamaño de 64, 128 o 256 filas y columnas. De esta manera se tienen pequeñas imágenes tanto satelital como de la cicatriz (máscara) con las mismas dimensiones, similar ubicación geográfica, y de esta manera a partir de estas son divididas en conjuntos de datos de entrenamiento, validación y prueba.

### **Modelo de Deep Learning**

Mediante el uso de técnicas de segmentación semántica se elabora un modelo de RNA el cual es entrenado con los conjuntos de datos de entrenamiento y validación. De acuerdo a los valores de las métricas obtenidas, se optimiza el modelo mediante la configuración de sus hiperparámetros. Finalmente, cuando se evalúa el modelo optimizado con el conjunto de datos de prueba nos devuelve las máscaras predichas por el modelo que deberán ser muy similares a las etiquetas verdaderas.

## IV. RESULTADOS

### 4.1 Procesamiento de conjunto de datos

Las imágenes satelitales a utilizarse para el desarrollo de la tesis fueron descargadas desde las plataformas siguientes:

- Landsat-8, brindada por la USGS - United States Geological Survey. (USGS, 2023)
- Sentinel-2, brindada por la ESA - Europe Spatial Agency. (ESA, 2023)

Antes de ver el tema de las imágenes a utilizar, es bueno tener en cuenta los datos de las cicatrices provenientes de los RIF. Para ello en la Tabla 6, podemos apreciar un cuadro resumen de la muestra a utilizar.

**Tabla 6**  
*Cicatrices de RIF - 2017 a 2021*

<b>Año</b>	<b>Nro. RIF</b>	<b>Nro. Cicatriz</b>	<b>Área (ha)</b>
2017	4	5	364.84
2018	4	5	1540.14
2019	1	1	123.39
2020	13	26	5637.128
2021	82	89	20675.49
<b>TOTALES</b>	<b>104</b>	<b>126</b>	<b>28340.988</b>

*Nota.* Es bueno mencionar que a la fecha de la presentación de la tesis, la UFMS aún se encuentra procesando los RIF de los años: 2022 - 2023.

Podemos afirmar que se tiene 126 formas geométricas poligonales de los incendios forestales descritos en los RIF, y que el área quemada es de 28340.988 hectáreas.

Debido a que se desea aplicar el modelo tanto a imágenes de los sensores: Sentinel-2 (S2) y Landsat-8 (L8), observamos que la resolución espacial es mayor en bandas espectrales de sensor S2 que de sensor L8. Debido a ello descargaremos escenas completas que cubran la superficie poligonal de las cicatrices de los RIF. Para poder utilizar el mismo modelo, es importante que las bandas espectrales a utilizar por ambos sensores presenten las mismas

características en el espectro electromagnético. En la Tabla 7 podemos apreciar la equivalencia de bandas lo cual nos permite aplicar el modelo a ambos sensores sobre las bandas indicadas.

**Tabla 7**

*Equivalencia de Bandas Espectrales - S2 y L8*

<b>Espectro Electromagnético</b>	<b>Sentinel-2</b>	<b>Landsat-8</b>
Blue - Azul	B2	B2
Green - Verde	B3	B3
Red - Rojo	B4	B4
Near Infrared - NIR	B8, B8A	B5
SWIR 1	B11	B6
SWIR 2	B12	B7

Ahora vemos la descarga de imágenes satelitales para la generación de las etiquetas según las cicatrices de los RIF. Luego de haber visto el tema espectral de ambos sensores, evaluamos el tema de la resolución espacial. El sensor S2 presenta valores de 10 y 20 m. en las bandas a utilizar, mientras que el sensor L9 presenta 30 m. en las bandas equivalentes. Debido a ello se eligió descargar escenas completas del sensor Sentinel-2 con fecha posterior del incendio forestal según se indica en los RIF, desde la plataforma web: (ESA, 2023). Las imágenes del sensor L8 solo fueron visualizadas desde la plataforma web: (USGS, 2023) en fechas cercanas. En la Tabla 8, podemos apreciar el número de escenas descargadas del sensor Sentinel-2. Dichas escenas son ortorectificadas, y en el tope de la atmósfera (TOA).

**Tabla 8**

*Escenas del sensor Sentinel-2*

<b>Año</b>	<b>Escenas</b>
2017	4
2018	3
2019	1
2020	25
2021	46

*Nota.* En el Anexo 9.4 se describe el detalle de las escenas utilizadas del Sentinel-2.

#### **4.1.1 Preprocesamiento de datos**

Las imágenes descargadas vienen separadas por bandas espectrales. Debido a ello, usando el software QGIS procedemos a apilar las bandas. Debido a la diferencia de resoluciones espaciales entre los sensores y entre las mismas bandas, evaluamos dichos valores. Se decidió remuestrear todas las bandas a 10m., lo cual también es aplicable a las imágenes del sensor Landsat-8. Finalmente, se tiene una imagen apilada y remuestreada.

Debido al remuestreo cada píxel tendrá una resolución espacial de 10m., y para la conformación de las imágenes a utilizar se eligió recortes de 128 x 128 píxeles como mínimo que nos permita con los recursos informáticos poder entrenar, validar y probar el modelo. Tomando como referencia las poligonales se genera un archivo vectorial rectangular que las contenga y que cuya extensión tanto a nivel horizontal como vertical sean múltiplos de 1280 m. De esta manera obtenemos las imágenes recortadas finales a utilizar.

Para la generación de las imágenes etiquetas para el modelo procedemos a convertir los archivos vectoriales de las cicatrices a formato raster tomando como referencia el archivo vectorial rectangular mencionado en párrafo anterior y una resolución de 10m. por píxel. De esta forma, se tiene la misma superficie, pero con píxeles marcados como área quemada. Finalmente, obtenemos imágenes etiquetadas a utilizar en el modelo.

#### **4.1.2 Procesamiento de datos**

El proceso de un sistema de Deep Learning requiere varios conjuntos de datos: entrenamiento, validación y prueba, este último no lo conoce el modelo. Las imágenes tanto recortadas y etiquetadas de 128 x 128 píxeles, se les denomina: tiles. Tanto los tiles de imágenes y etiquetas (también llamadas máscaras) deben tener el mismo nombre para un aprendizaje supervisado. Luego de la generación de tiles de la totalidad de las cicatrices mencionadas se obtuvo: 2898 tiles, de los cuales, no solo se tiene tiles con etiquetas de área quemada

sino también sin etiquetas. Los tiles sin etiquetas representan los diversos ecosistemas del departamento de Cusco. En la siguiente Tabla 9 podemos apreciar los tiles totales y aquellos con o sin etiquetas (máscaras).

**Tabla 9**

*Tiles para Aprendizaje Supervisado*

<b>Tiles</b>	<b>Nro</b>	<b>Porcentaje %</b>
Sin etiqueta - ecosistemas	2454	84.7
Con etiqueta - máscaras	444	15.3
<b>TOTAL</b>	<b>2898</b>	<b>100.0</b>

De acuerdo a la cantidad de tiles con etiquetas, se aprecia un conjunto de datos desbalanceado. Lo recomendable es tener 50 % con etiquetas para que los resultados sean los más equitativos. En la Tabla 10 siguiente podemos apreciar la conformación del grupo de datos así como la cantidad de tiles asignados, recuerden que cada grupo desde desdoblarse en: entrenamiento (training), validación (validation) y prueba (test).

**Tabla 10**

*Datos Grupo 1 - G1: Proporción de tiles con máscaras 15.3 %*

<b>Datos</b>	<b>Porcentaje - %</b>	<b>TIT</b>	<b>TSE</b>	<b>TCE</b>
Entrenamiento	64	1854	1570	284
Validación	16	464	393	71
Prueba	20	580	491	89
<b>TOTALES</b>	<b>100</b>	<b>2898</b>	<b>2454</b>	<b>444</b>

*Nota.* La columna TIT nos indica el número de tiles totales. La columna TSE nos menciona el número de tiles sin etiqueta, que son aquellos libres de área quemada - ecosistemas. La columna TCE nos indica el número de tiles con etiqueta - máscaras (aquellas que contienen píxeles de área quemada pero no necesariamente todo el tile).  $TIT = TSE + TCE$ . Todo conjunto de tiles con etiqueta mantiene la proporción del 15 % sobre los totales.

En la temática de segmentación semántica casi siempre tendremos datos desbalanceados. Debido a ello vamos a generar otros dos grupos de datos desde el inicial.

Para ambos grupos se reducirán los tiles sin etiqueta con la finalidad de subir el porcentaje de tiles con etiqueta.

En la Tabla 11 el porcentaje sube de 15.3 % a alrededor del 20 %.

**Tabla 11**

*Datos Grupo 2 - G2: Proporción de tiles con máscaras 20 %*

Datos	Porcentaje - %	TIT	TSE	TCE
Entrenamiento	64	1421	1137	284
Validación	16	355	284	71
Prueba	20	444	355	89
<b>TOTALES</b>	<b>100</b>	<b>2220</b>	<b>1776</b>	<b>444</b>

Mientras que en la Tabla 12 sube aproximadamente a un 25 %.

**Tabla 12**

*Datos Grupo 3 - G3: Proporción de tiles con máscaras 25 %*

Datos	Porcentaje - %	TIT	TSE	TCE
Entrenamiento	64	1137	853	284
Validación	16	284	213	71
Prueba	20	355	266	89
<b>TOTALES</b>	<b>100</b>	<b>1776</b>	<b>1332</b>	<b>444</b>

Cada grupo de datos tiene 6 combinaciones de bandas - D1 a D6. Podemos apreciar el orden de apilamiento por cada banda en la Tabla 13.

**Tabla 13**

*Combinación de Bandas Sentinel-2*

Datos	Nro. Bandas	Bandas					
D1	3	B12	B8	B4	-	-	-
D2	3	B12	B8A	B4	-	-	-
D3	3	B12	B11	B8A	-	-	-
D4	3	B12	B11	B8	-	-	-
D5	6	B12	B11	B8	B4	B3	B2
D6	6	B12	B11	B8A	B4	B3	B2

Finalmente, tomando en cuenta los grupos de datos (G1 a G3) y las combinaciones de bandas por cada grupo (D1 a D6); se tendrá 18 conjuntos de datos en total con lo que se probará el modelo a implementar.

Como se conoce las redes neuronales requieren de una gran cantidad de datos para poder aprender las características de los datos a entrenar. Pero debido a que la información proveniente de los RIF es limitada, y la generación de informes similares implica un alto costo de tiempo, hay una técnica útil para este caso denominada: Aumento de Datos (Data Augmentation).

Utilizaremos dicha técnica en los datos del Grupo 3 con la finalidad de evaluar sus resultados con respecto a los demás grupos.

## **4.2 Arquitectura de red neuronal convolucional**

### **4.2.1 Modelo**

Para el desarrollo de la tesis se eligió la arquitectura encoder-decoder, utilizando el modelo UNet. Tomando como base la configuración básica desde (Ronneberger et al., 2015), procedemos a personalizarlo para este estudio.

Recordamos que los datos de entrada están conformados por 128x128 píxeles por 3 y 6 canales de profundidad. El número de canales en este caso bandas espectrales no modifica el modelo en sí, simplemente toma más tiempo en el entrenamiento al tener el doble de capas de datos (cada canal o banda es una capa cuadrada de 128x128).

En el Anexo 9.3 se muestra la arquitectura UNet personalizada y que se describe a continuación.

### **Codificador - Encoder**

Se inicia con la aplicación de dos capas convolucionales de 32 filtros con dimensiones 3x3, padding de tipo 'same', stride de valor = 1, seguida de una función de activación 'elu', una operación de pooling (máximo) de dimensión 2x2, y una de Batch Normalization. Ahora, aplicamos dos capas convolucionales de 64 filtros de similares propiedades, una operación de

pooling (máximo) de dimensión 2x2, una de Dropout de un 50 % y una de Batch Normalization. Luego, aplicamos repetidamente capas convolucionales similares a la última de 128, 256, y 512 filtros. En cada paso se produce un submuestreo. Finalmente aplicamos un par de capas convolucionales de 1024 filtros de mismas propiedades, pero ahora seguido solo de una operación Dropout de 50 %.

### **Decodificador - Decoder**

En esta parte consiste en realizar un sobremuestreo de los datos. Utilizamos la operación de Convolución Transpuesta de 512 filtros con dimensiones 2x2, strides 2x2, padding de tipo 'same', y luego se concatena con el mapa de características que corresponde. A esta concatenación aplicamos una operación de Batch Normalization. Luego, aplicamos con dos capas de convoluciones similares a la sección del Encoder, pero de 512 filtros, y una operación de Dropout de 50 %. Este resultado subimos el muestreo con una operación Convolución Transpuesta de 256 filtros y similares propiedades que el anterior, y concatenamos el mapa que corresponde, y luego una operación de Batch Normalization. Luego, aplicamos con dos capas de convoluciones similares a la sección del Encoder, pero de 256 filtros, y una operación de Dropout de 50 %. Subimos el muestreo nuevamente con una operación Convolución Transpuesta de 128 filtros y similares propiedades que anterior, y concatenamos el mapa que corresponde, y luego una operación de Batch Normalization. Luego, aplicamos con dos capas de convoluciones similares a la sección del Encoder, pero de 128 filtros, y una operación de Dropout de 50 %. Subimos el muestreo nuevamente con una operación Convolución Transpuesta de 64 filtros y similares propiedades que anterior, y concatenamos el mapa que corresponde, y luego una operación de Batch Normalization. Luego, aplicamos con dos capas de convoluciones similares a la sección del Encoder, pero de 64 filtros, y una operación de Dropout de 50%. Subimos el muestreo nuevamente con una operación Convolución

Transpuesta de 32 filtros y similares propiedades que el anterior, y concatenamos el mapa que corresponde, y luego una operación de Batch Normalization. Luego, aplicamos con dos capas de convoluciones similares a la sección del Encoder, pero de 32 filtros. Finalmente, aplicamos una última capa convolucional de 1 filtro de dimensiones 1x1 con función de activación de tipo 'sigmoid' ya que la salida es de tipo binaria.

#### **4.2.2 Configuración de entrenamiento**

Luego de la descripción del modelo, debemos definir los hiperparámetros del modelo para la fase de entrenamiento, validación y prueba. Como toda red neuronal no hay una regla exacta que nos indique el número de filtros ni las dimensiones de una capa convolucional, la definición realizada anteriormente se puede considerar también como hiperparámetros a nivel del modelo. Ahora pasamos a definir los relacionados al entrenamiento del modelo. Hay tres principales asociados a nuestra capacidad de recursos informáticos: tamaño de lote o conocido como: 'batch size', pasos por época y número de épocas. En este caso, elegimos un 'batch size'=4, los pasos por época son la división del tamaño de datos de entrenamiento por el 'batch size', y las épocas a iterar son de 100. No se utilizará 'transferencia de aprendizaje' para el cálculo de los pesos de las neuronas o filtros en el modelo, será un entrenamiento desde cero como se conoce.

También elegimos el método de optimización Adam con una tasa de aprendizaje de '1e-5', y un retardo de '1e-6'. Como función de pérdida, muy importante de acuerdo a la temática, elegimos el 'Jaccard Loss' y como métrica asociada 'Coeficiente Jaccard', ambos relacionados a la métrica IoU.

Para la fase de prueba sobre datos a los cuales el modelo no ha tenido acceso asumimos un valor de probabilidad mayor a 0.5 como de valor binario 1, significando: pixel quemado.

### 4.3 Contrastación de Hipótesis

#### *Hipótesis General*

**Ho:** un modelo de aprendizaje profundo utilizando segmentación semántica no permitirá delimitar la superficie afectada denominada cicatriz de área quemada producto de un incendio forestal en tiempo oportuno.

**Ha:** un modelo de aprendizaje profundo utilizando segmentación semántica permitirá delimitar la superficie afectada denominada cicatriz de área quemada producto de un incendio forestal en tiempo oportuno.

**Tabla 14**

*Métrica: Precision - Grupo G1*

Conjunto de Datos	Métricas	D1	D2	D3	D4	D5	D6
Entrenamiento	Exactitud	0.9945	0.9924	0.9948	0.9944	0.9949	0.9946
Entrenamiento	IoU	0.9401	0.9196	0.9404	0.935	0.9462	0.9419
Validación	Exactitud	0.9959	0.9952	0.9957	0.9959	0.996	0.9963
Validación	IoU	0.9558	0.9498	0.9487	0.9564	0.9568	0.9612
Prueba	Exactitud	0.9937	0.9929	0.9941	0.9929	0.9938	0.9943
Prueba	Precision	0.9091	0.9085	0.8944	0.8948	0.9333	0.9555

**Tabla 15**

*Métrica: Precision - Grupo G2*

Conjunto de Datos	Métricas	D1	D2	D3	D4	D5	D6
Entrenamiento	Exactitud	0.9916	0.9909	0.9934	0.9912	0.9934	0.994
Entrenamiento	IoU	0.922	0.9157	0.9332	0.9126	0.9329	0.9427
Validación	Exactitud	0.994	0.9929	0.9951	0.9944	0.9947	0.9947
Validación	IoU	0.9418	0.9261	0.9548	0.9507	0.9467	0.9523
Prueba	Exactitud	0.9913	0.9915	0.991	0.9918	0.9906	0.9933
Prueba	Precision	0.9113	0.8983	0.8938	0.9232	0.9044	0.9429

**Tabla 16***Métrica: Precision - Grupo G3*

<b>Conjunto de Datos</b>	<b>Métricas</b>	<b>D1</b>	<b>D1-AD</b>	<b>D2</b>	<b>D2-AD</b>	<b>D3</b>	<b>D3-AD</b>
Entrenamiento	Exactitud	0.9889	0.9862	0.9842	0.9764	0.9855	0.9857
Entrenamiento	IoU	0.9098	0.8809	0.8703	0.8082	0.8884	0.8682
Validación	Exactitud	0.9921	0.9913	0.9894	0.984	0.9905	0.9922
Validación	IoU	0.9266	0.9277	0.9029	0.8632	0.9209	0.9225
Prueba	Exactitud	0.9892	0.9875	0.9855	0.9821	0.9872	0.9868
Prueba	Precision	0.9307	0.9485	0.9164	0.9678	0.9513	0.8637
<b>Conjunto de Datos</b>	<b>Métricas</b>	<b>D4</b>	<b>D4-AD</b>	<b>D5</b>	<b>D5-AD</b>	<b>D6</b>	<b>D6-AD</b>
Entrenamiento	Exactitud	0.9915	0.9859	0.9914	0.9886	0.9916	0.9903
Entrenamiento	IoU	0.9203	0.8659	0.9281	0.8973	0.9291	0.9175
Validación	Exactitud	0.9939	0.9934	0.9934	0.9932	0.9936	0.9937
Validación	IoU	0.9484	0.9336	0.9437	0.9321	0.943	0.946
Prueba	Exactitud	0.9898	0.9886	0.9915	0.989	0.9909	0.9916
Prueba	Precision	0.9113	0.8907	0.9258	0.886	0.9047	0.9314

*Nota.* Mencionamos que en etapa de entrenamiento el modelo tuvo acceso tanto al conjunto de datos de entrenamiento y validación. Además, se adicionó el cálculo de la métrica 'Exactitud' solo para observar el desempeño del modelo, pero no para temas de rendimiento. Para ello se utilizó la métrica: 'Precision' (por tener clases desbalanceadas), y 'Coeficiente Jaccard' o 'IoU' como se mencionó para segmentación semántica.

**Interpretación:** según los resultados obtenidos para verificar la hipótesis general, podemos observar en las Tablas 14, 15 y 16 que los valores de la métrica 'Precision' tiene valores desde 0.89 hasta 0.95, equivalentes a 89 % hasta 95 %; en el caso del uso de aumento de datos en el Grupo G3, se observa que en la mayoría de combinaciones de bandas se aprecia un aumento de 1 a 2 puntos porcentuales. Finalmente, con los valores obtenidos en el conjunto de prueba a pesar de los datos limitados en número podemos afirmar que la hipótesis alterna

se cumple: un modelo de aprendizaje profundo utilizando segmentación semántica permitirá delimitar el área afectada denominada cicatriz de área quemada producto de un incendio forestal en tiempo oportuno.

#### 4.3.1 Hipótesis Específica A

**Ho:** un modelo óptimo de las RNA no permitirá delimitar la superficie afectada.

**Ha:** un modelo óptimo de las RNA permitirá delimitar la superficie afectada.

**Tabla 17**

*Métrica: F1-Score - Grupo G1*

Conjunto de Datos	Métricas	D1	D2	D3	D4	D5	D6
Prueba	F1-Score	0.9174	0.905	0.9237	0.9071	0.9174	0.9237

**Tabla 18**

*Métrica: F1-Score - Grupo G2*

Conjunto de Datos	Métricas	D1	D2	D3	D4	D5	D6
Prueba	F1-Score	0.9119	0.9154	0.9105	0.916	0.9042	0.9317

**Tabla 19**

*Métrica: F1-Score - Grupo G3*

Conjunto de Datos	Métricas	D1	D1-AD	D2	D2-AD	D3	D3-AD
Prueba	F1-Score	0.9104	0.8926	0.8776	0.8347	0.8897	0.8973
Conjunto de Datos	Métricas	D4	D4-AD	D5	D5-AD	D6	D6-AD
Prueba	F1-Score	0.918	0.9096	0.9312	0.9136	0.9282	0.9317

**Interpretación:** según los resultados obtenidos para verificar la hipótesis en las Tablas 17, 18 y 19, la métrica F1-Score tiene valores desde 0.87 hasta 0.93, equivalentes a 87 % hasta 93 %; en el caso del uso de aumento de datos en el Grupo G3, se observa también variabilidad como en métricas anteriores. Finalmente, podemos afirmar que la hipótesis alterna se cumple: un modelo óptimo de las RNA permitirá delimitar la superficie afectada.

### 4.3.2 Hipótesis Específica B

**Ho:** un modelo óptimo de las RNA no permitirá realizar una adecuada estimación del área afectada.

**Ha:** un modelo óptimo de las RNA permitirá realizar una adecuada estimación del área afectada.

**Tabla 20**

*Métrica: IoU (índice Jaccard)- Grupo G1*

Conjunto de Datos	Métricas	D1	D2	D3	D4	D5	D6
Prueba	IoU	0.8474	0.8265	0.8583	0.8299	0.8474	0.8583

**Tabla 21**

*Métrica: IoU (índice Jaccard)- Grupo G2*

Conjunto de Datos	Métricas	D1	D2	D3	D4	D5	D6
Prueba	IoU	0.838	0.8439	0.8358	0.845	0.8251	0.8721

**Tabla 22**

*Métrica: IoU (índice Jaccard) - Grupo G3*

Conjunto de Datos	Métricas	D1	D1-AD	D2	D2-AD	D3	D3-AD
Prueba	IoU	0.8355	0.8061	0.7818	0.7164	0.8013	0.8137
Conjunto de Datos	Métricas	D4	D4-AD	D5	D5-AD	D6	D6-AD
Prueba	IoU	0.8485	0.8342	0.8713	0.841	0.866	0.8721

**Interpretación:** según los resultados obtenidos para verificar la hipótesis en las Tablas 20, 21 y 22, la métrica IoU tiene valores desde 0.78 hasta 0.87, equivalentes a 78 % hasta 87 %; en el caso del uso de aumento de datos en el Grupo G3, se observa que hay una mayor rango de variabilidad dependiendo de las combinaciones de bandas. Finalmente, podemos afirmar que la hipótesis alterna se cumple: un modelo óptimo de las RNA permitirá realizar una adecuada estimación del área afectada.

### 4.3.3 Hipótesis Específica C

**Ho:** Un modelo óptimo de las RNA no permitirá determinar las bandas de una imagen satelital para delimitar la superficie afectada.

**Ha:** Un modelo óptimo de las RNA permitirá determinar las bandas de una imagen satelital para delimitar la superficie afectada.

**Tabla 23**

*Métrica: Recall o Sensibilidad - Grupo G1*

Conjunto de Datos	Métricas	D1	D2	D3	D4	D5	D6
Prueba	Recall	0.9258	0.9015	0.9551	0.9197	0.9021	0.894

**Tabla 24**

*Métrica: Recall o Sensibilidad - Grupo G2*

Conjunto de Datos	Métricas	D1	D2	D3	D4	D5	D6
Prueba	Recall	0.9124	0.933	0.9279	0.9088	0.9039	0.9207

**Tabla 25**

*Métrica: Recall o Sensibilidad - Grupo G3*

Conjunto de Datos	Métricas	D1	D1-AD	D2	D2-AD	D3	D3-AD
Prueba	Recall	0.891	0.843	0.8419	0.7339	0.8355	0.9336
Conjunto de Datos	Métricas	D4	D4-AD	D5	D5-AD	D6	D6-AD
Prueba	Recall	0.9248	0.9293	0.9367	0.943	0.9529	0.932

**Interpretación:** según los resultados obtenidos para verificar la hipótesis en las Tablas 23, 24 y 25, la métrica Recall tiene valores desde 0.83 hasta 0.95, equivalentes a 83 % hasta 95 %; en el caso del uso de aumento de datos en el Grupo G3, se observa que hay una mayor rango de variabilidad dependiendo de las combinaciones de bandas. Finalmente, podemos afirmar que la hipótesis alterna se cumple: un modelo óptimo de las RNA permitirá determinar las bandas de una imagen satelital para delimitar la superficie afectada.

#### 4.3.4 Hipótesis Específica D

**Ho:** Un modelo óptimo de las RNA no permitirá determinar una adecuada estimación del área afectada por incendios forestales en la región Cusco.

**Ha:** Un modelo óptimo de las RNA permitirá determinar una adecuada estimación del área afectada por incendios forestales en la región Cusco.

**Tabla 26**

*Métrica: Coeficiente Kappa - Grupo G1*

Conjunto de Datos	Métricas	D1	D2	D3	D4	D5	D6
Prueba	Kappa	0.9141	0.9013	0.9206	0.9034	0.9142	0.9208

**Tabla 27**

*Métrica: Coeficiente Kappa - Grupo G2*

Conjunto de Datos	Métricas	D1	D2	D3	D4	D5	D6
Prueba	Kappa	0.9073	0.9109	0.9058	0.9117	0.8992	0.9282

**Tabla 28**

*Métrica: Coeficiente Kappa - Grupo G3*

Conjunto de Datos	Métricas	D1	D1-AD	D2	D2-AD	D3	D3-AD
Prueba	Kappa	0.8355	0.8061	0.7818	0.7164	0.8013	0.8137
Conjunto de Datos	Métricas	D4	D4-AD	D5	D5-AD	D6	D6-AD
Prueba	Kappa	0.8485	0.8342	0.8713	0.841	0.866	0.8721

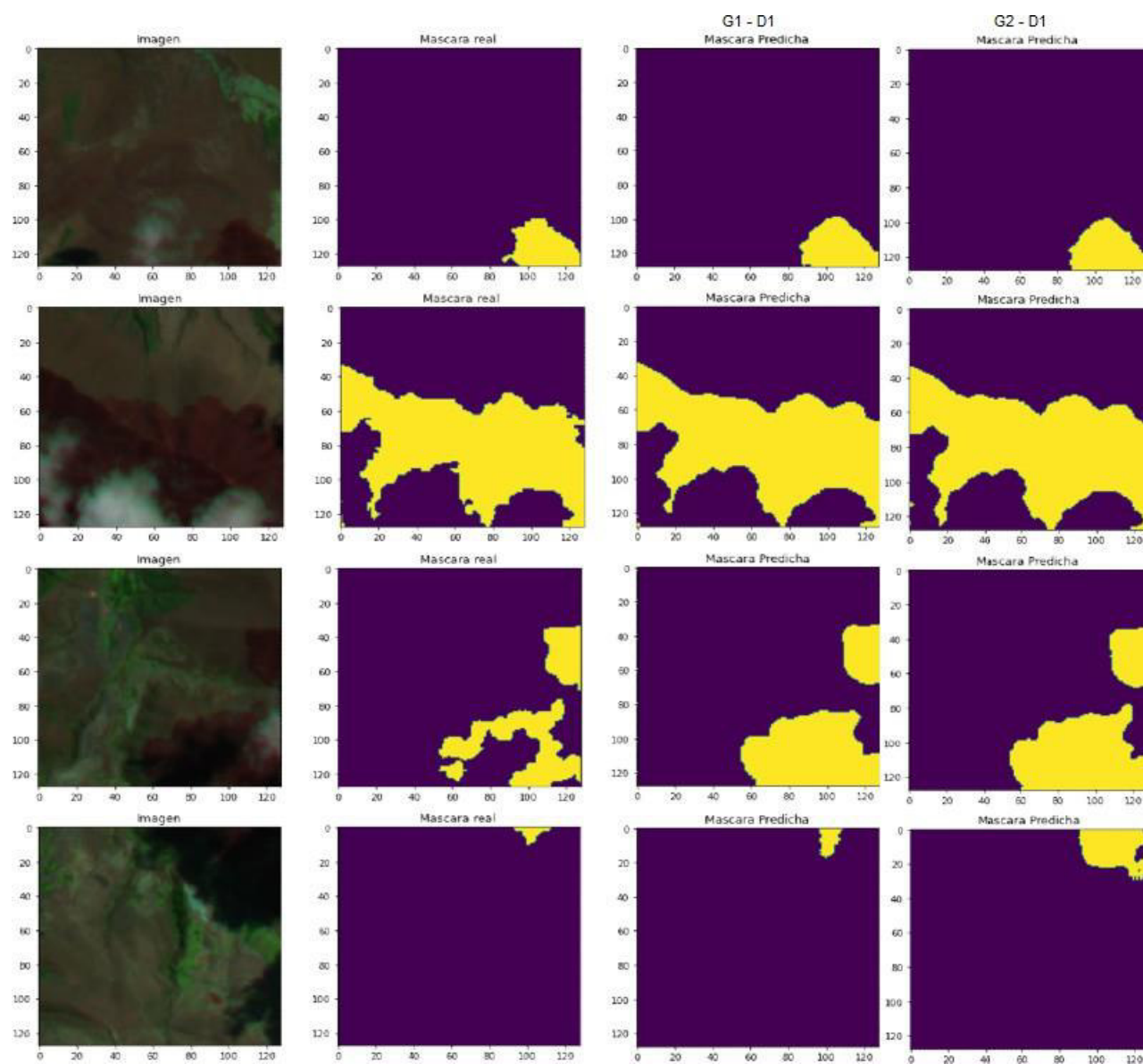
**Interpretación:** según los resultados obtenidos para verificar la hipótesis en las Tablas 26, 27 y 28, la métrica Kappa tiene valores desde 0.78 hasta 0.92, equivalentes a 78 % hasta 92 %; en el caso del uso de aumento de datos en el Grupo G3, se observa que hay una mayor rango de variabilidad dependiendo de las combinaciones de bandas. Finalmente, podemos afirmar que la hipótesis alterna se cumple: un modelo óptimo de las RNA permitirá determinar una adecuada estimación del área afectada por incendios forestales en la región Cusco.

A continuación observamos gráficos de estimación de áreas afectadas en la región del Cusco, provincias de: Cusco, Quispicanchis y Chumbivilcas, correspondientes al conjunto de prueba de los años 2020 y 2021.

Las figuras 26 y 27 nos muestran áreas afectadas por incendios forestales en la región del Cusco tomando en cuenta el subgrupo D1 de cada grupo de datos de prueba.

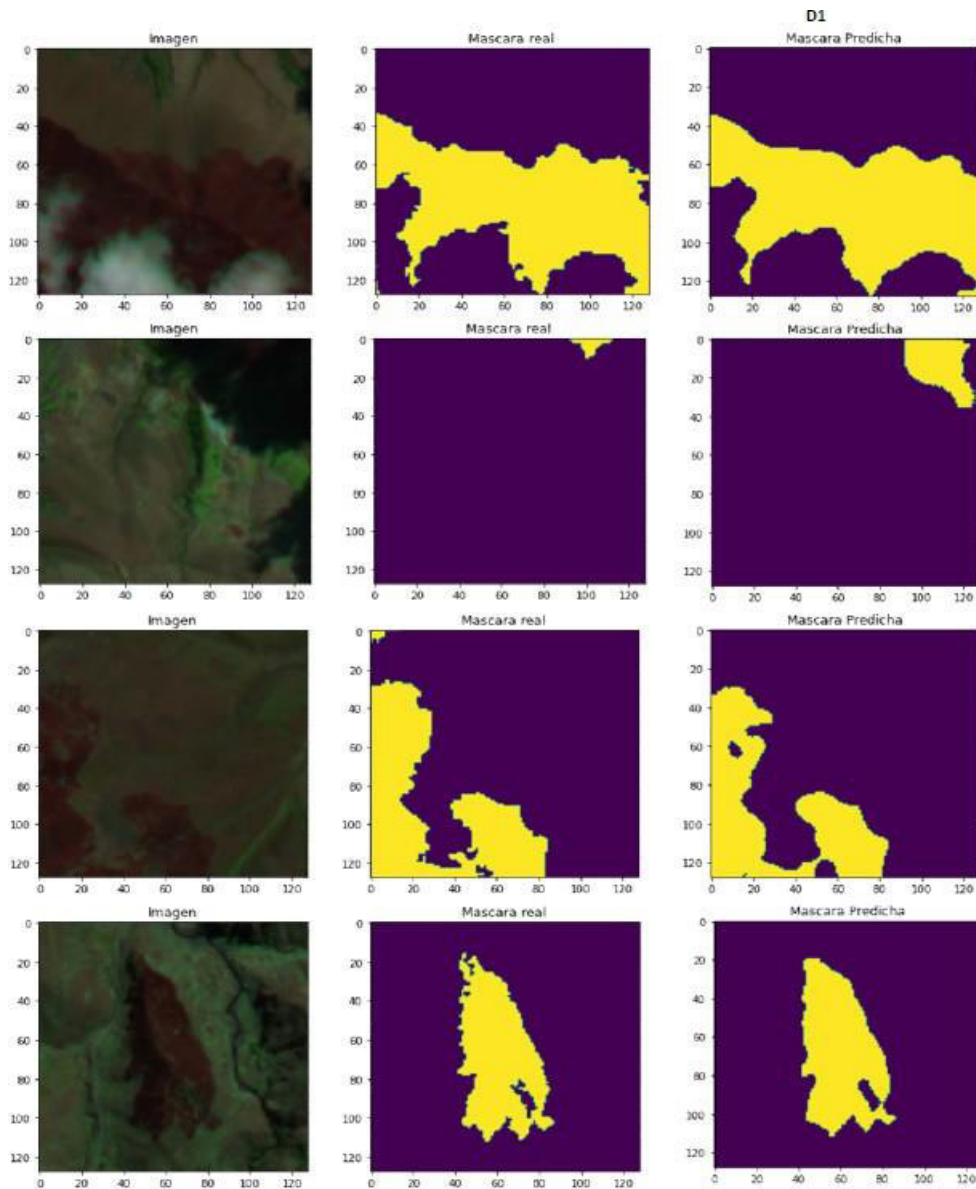
### Figura 26

Área afectada - Provincia: Cusco / Quispicanchis - Grupo G1 y G2



**Figura 27**

Área afectada - Provincia: Cusco / Quispicanchis / Chumbivilcas - Grupo G3



#### 4.4 Análisis e Interpretación

Como mencionamos en el marco teórico, se tiene una serie de métricas a utilizar en el desarrollo del modelo. En la fase de entrenamiento y validación se utilizó la métrica de Coeficiente Jaccard (IoU), solo viendo los valores de dicha métrica en el conjunto de datos de validación podemos afirmar que son muy buenos. Pero a pesar de ello, decidimos calcular las métricas asociadas según la Matriz de Consistencia con el conjunto de datos de prueba

(estos datos no los conoce el modelo) ello brinda una verificación más exhaustiva y cercana a la realidad. Dentro de ellas se encuentra el coeficiente Kappa y según la Tabla 5 los rangos con los cuales podemos comparar en forma cualitativa a los valores obtenidos.

En las Tablas 26, 27 y 28, podemos apreciar el cálculo del coeficiente Kappa para grupo de datos de prueba: G1, G2, y G3; además por cada combinación de bandas. El mayor porcentaje de valores obtenidos de todas las métricas se encuentran en niveles de concordancia 'Excelente' y el resto en 'Muy Bueno'.

Los grupos de datos G1, G2 y G3 fueron desdoblados en 6 subgrupos por agrupamiento de bandas según la Tabla 13. Hay 4 subgrupos de 3 bandas y 2 de 6 bandas. A mayor número de bandas hay mayor información espectral, mayor número de canales, mayor tiempo de entrenamiento, y por ello un mayor valor de las métricas, pero de acuerdo al tema de investigación no siempre es así. Para ello elaboramos la Tabla 29 en donde observamos los promedios de los subgrupos según su número de bandas (canales).

**Tabla 29**

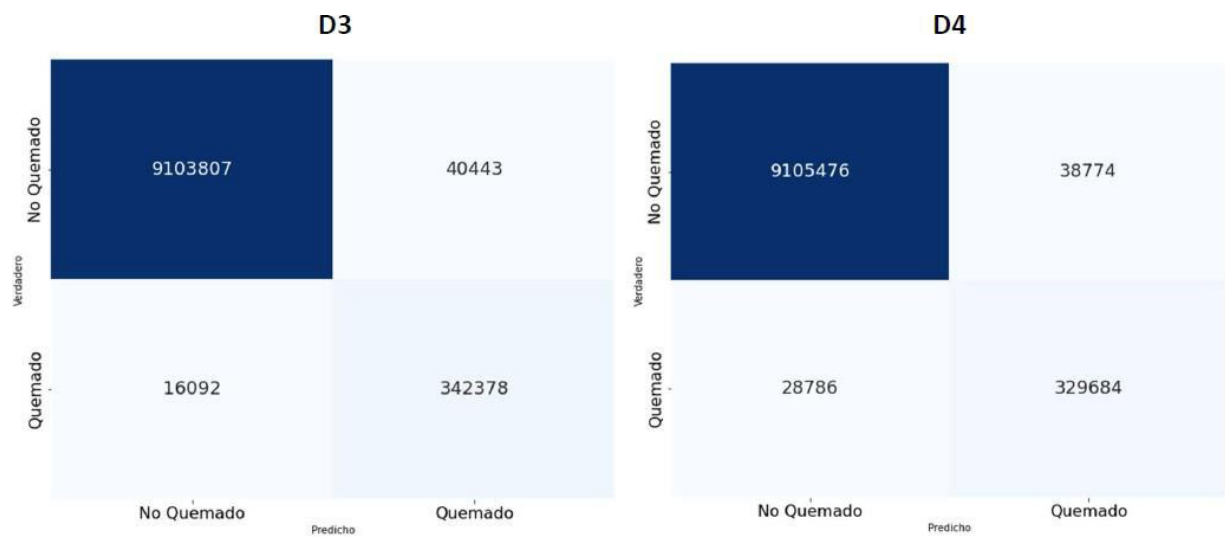
*Resumen Comparativo de Métricas - Promedio por subgrupos de 3 Bandas (3B) y 6 Bandas (6B) - Datos de Prueba*

Métricas	G1-3B	G1-6B	G2-3B	G2-6B	G3-3B	G3-3B-AD	G3-6B	G3-6B-AD
Exactitud	0.9934	0.9941	0.9914	0.9920	0.9879	0.9863	0.9912	0.9903
Precision	0.9017	0.9444	0.9067	0.9237	0.9274	0.9177	0.9153	0.9087
Recall	0.9255	0.8981	0.9205	0.9123	0.8733	0.8600	0.9448	0.9375
IoU	0.8405	0.8529	0.8407	0.8486	0.8168	0.7926	0.8687	0.8566
F1-Score	0.9133	0.9206	0.9135	0.9180	0.8989	0.8836	0.9297	0.9227
Kappa	0.9099	0.9175	0.9089	0.9206	0.8168	0.7926	0.8687	0.8566

Luego del entrenamiento del modelo donde se calculan los pesos del modelo de red neuronal implementado, y realizar la verificación sobre el conjunto de datos de pruebas, aparte de calcular las métricas ya mencionadas y con concordancia 'Muy Bueno' y 'Excelente',

elaboramos las Matrices de Confusión que nos brindan desde otro punto de vista una comparativa a nivel de los píxeles predichos: quemados (1) y no quemados (0). De las Figuras 28 hasta 39, apreciamos la cantidad de píxeles quemados predichos y verdaderos, así como los píxeles no quemados predichos y verdaderos. Entre ambos valores se encuentra la mayor parte de píxeles evaluados por el modelo. Esto se encuentra en concordancia con los valores de las métricas obtenidas ya que se encuentran relacionadas con los elementos de la matriz de confusión.

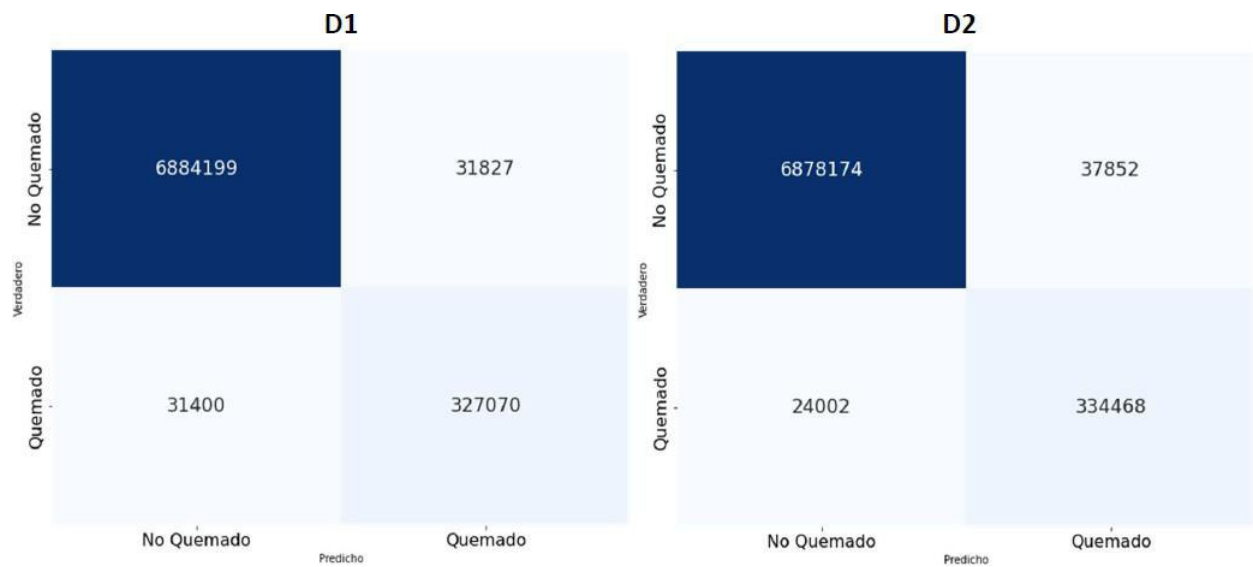
De las Figuras 40 hasta 45, podemos observar la comparación visual de los resultados en algunas imágenes sobre el conjunto de datos de prueba. Para efectos de usar una misma imagen visual se combinó las bandas: B12-B8-B4, pero el modelo utilizó las bandas de acuerdo a la Tabla 13.

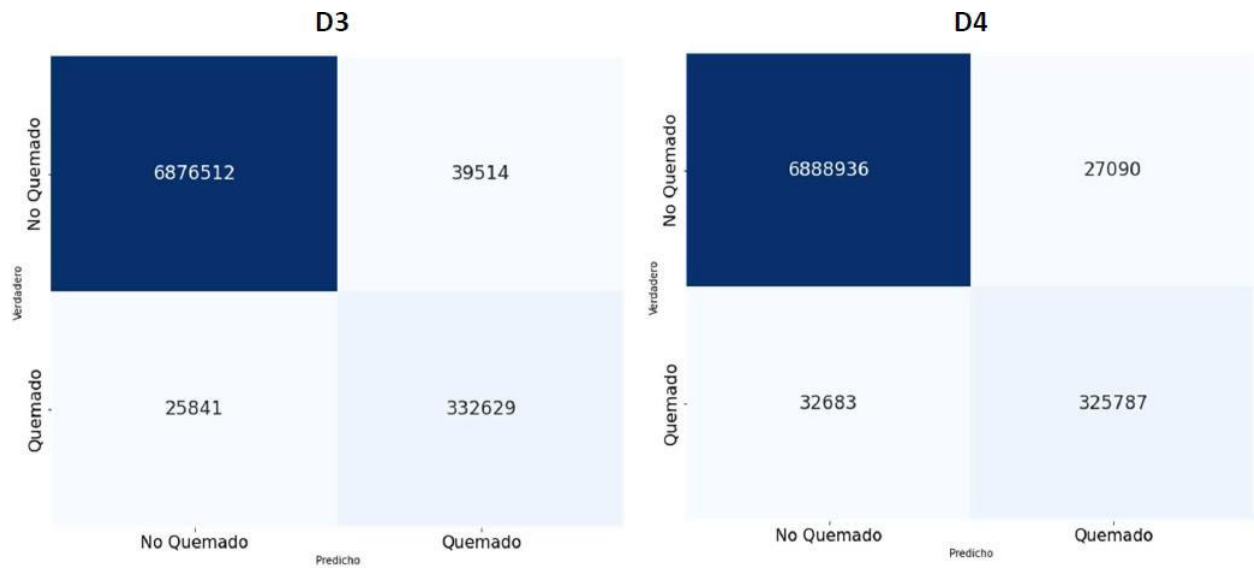
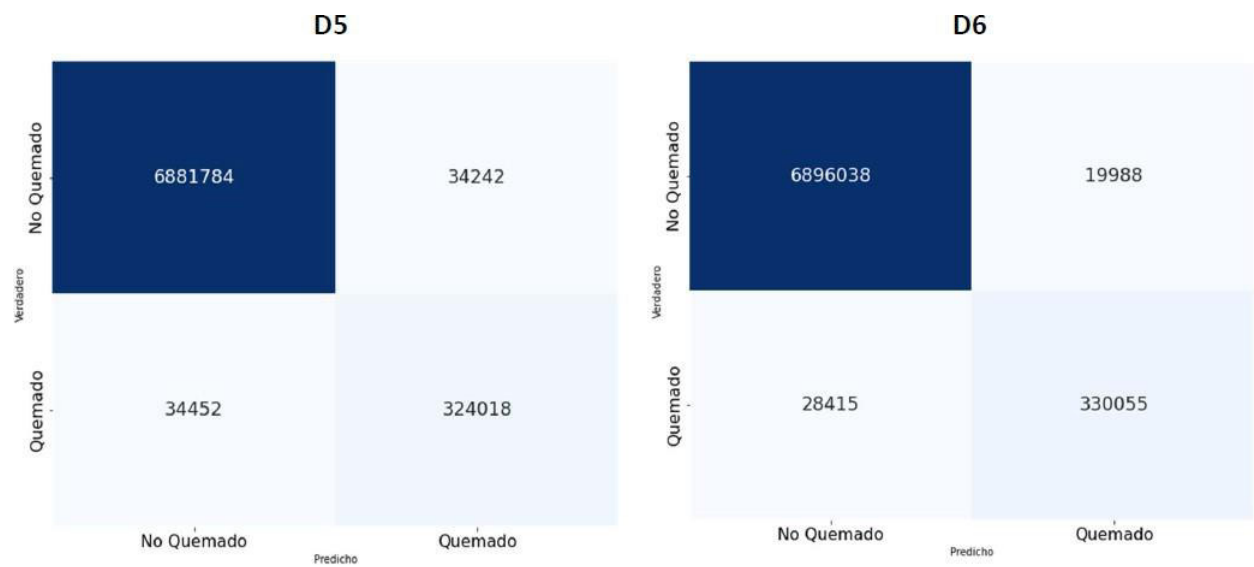
**Figura 28***Matriz de Confusión - Grupo G1 - D1 y D2***Figura 29***Matriz de Confusión - Grupo G1 - D3 y D4*

**Figura 30**  
*Matriz de Confusión - Grupo G1 - D5 y D6*



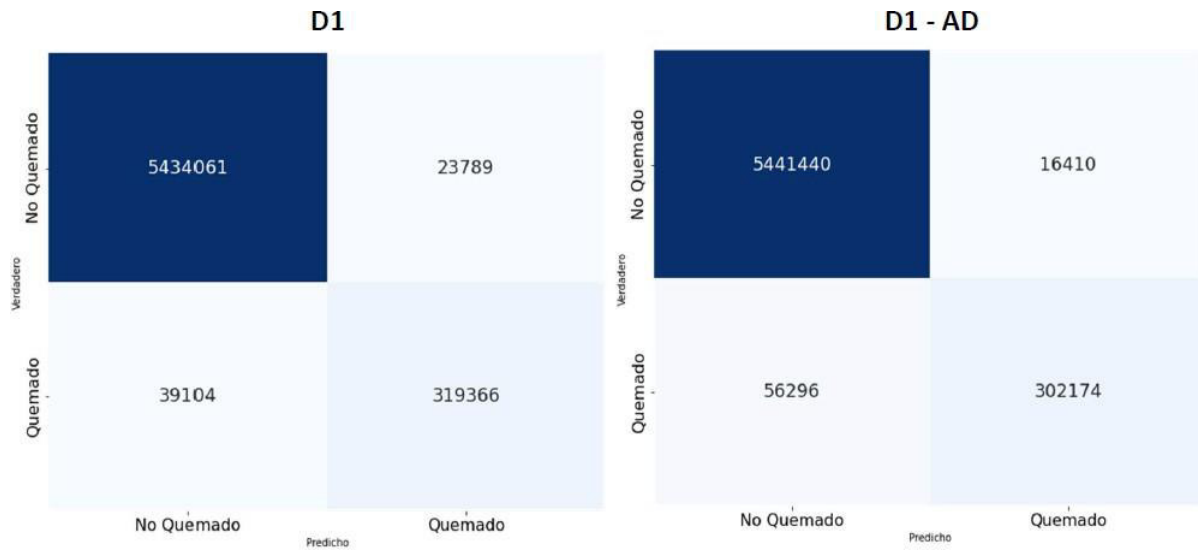
**Figura 31**  
*Matriz de Confusión - Grupo G2 - D1 y D2*



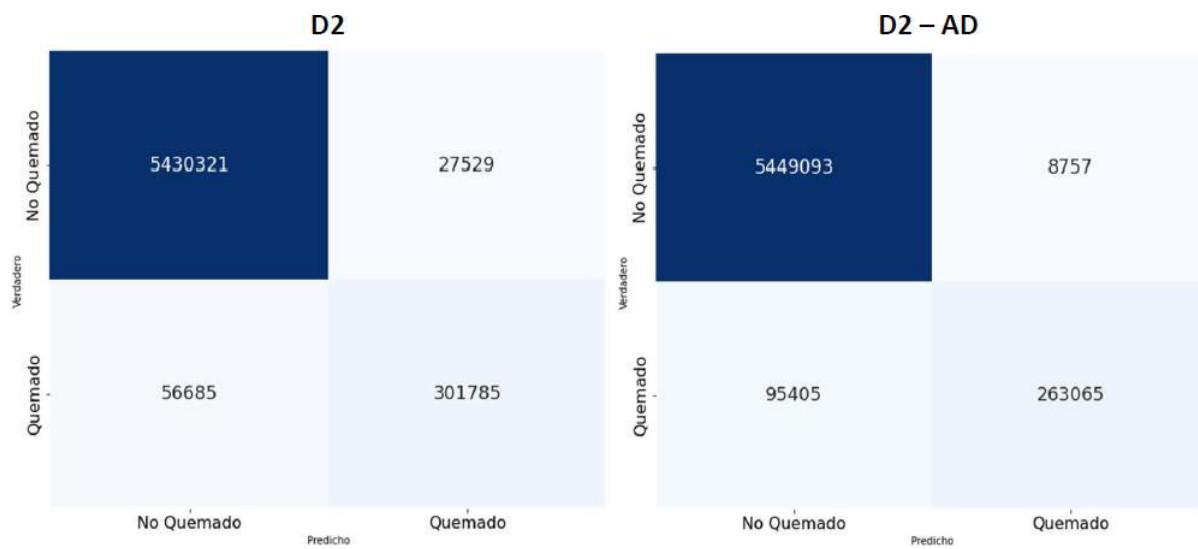
**Figura 32***Matriz de Confusión - Grupo G2 - D3 y D4***Figura 33***Matriz de Confusión - Grupo G2 - D5 y D6*

**Figura 34**

*Matriz de Confusión - Grupo G3 - D1 y D1-AD (Aumento de Datos)*

**Figura 35**

*Matriz de Confusión - Grupo G3 - D2 y D2-AD (Aumento de Datos)*

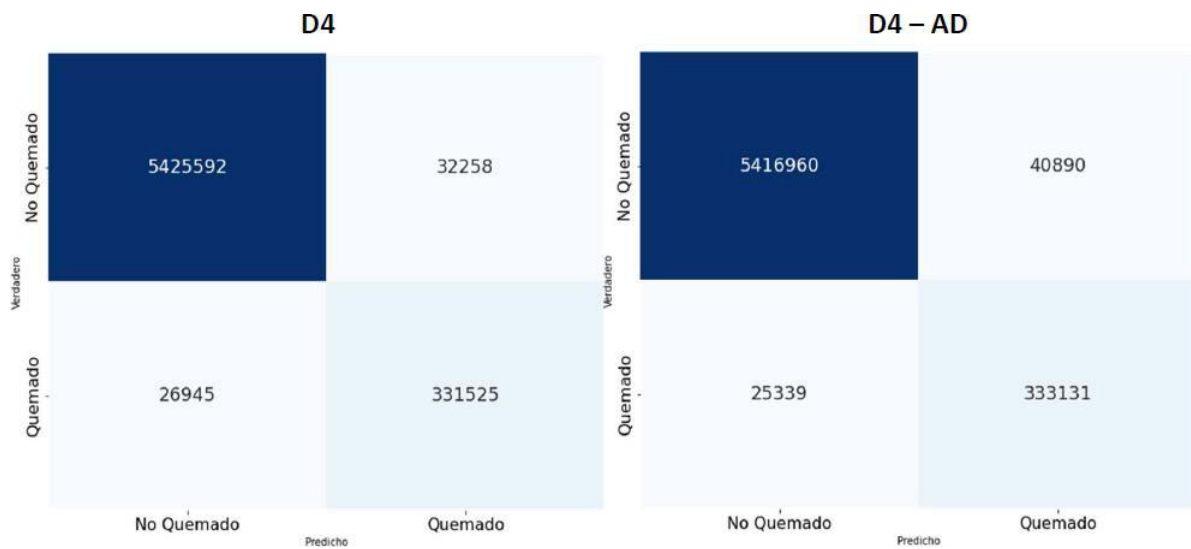


**Figura 36**

*Matriz de Confusión - Grupo G3 - D3 y D3-AD (Aumento de Datos)*

**Figura 37**

*Matriz de Confusión - Grupo G3 - D4 y D4-AD (Aumento de Datos)*

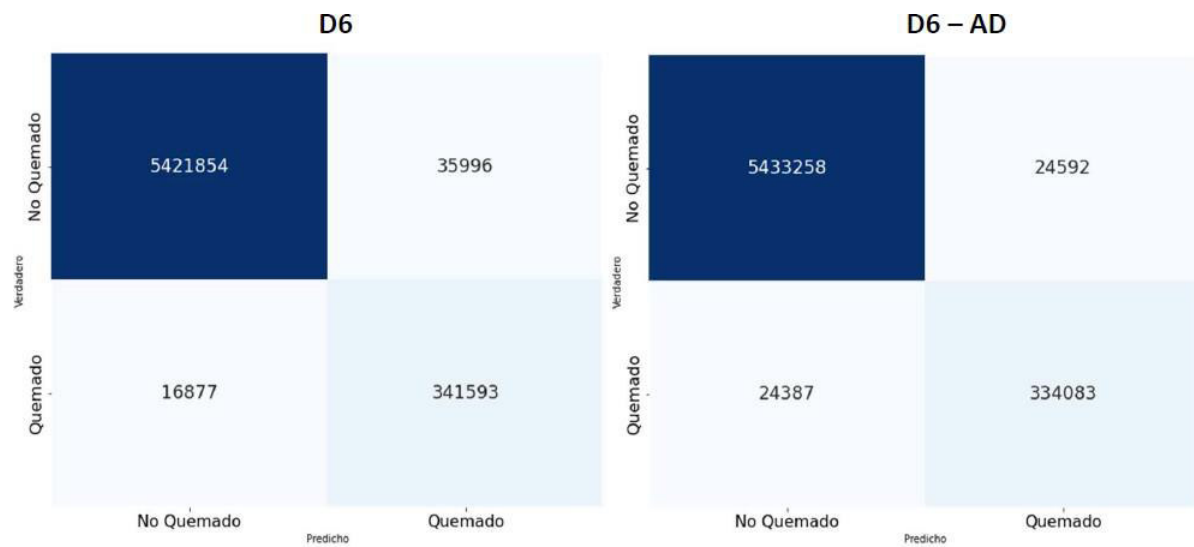


**Figura 38**

*Matriz de Confusión - Grupo G3 - D5 y D5-AD (Aumento de Datos)*

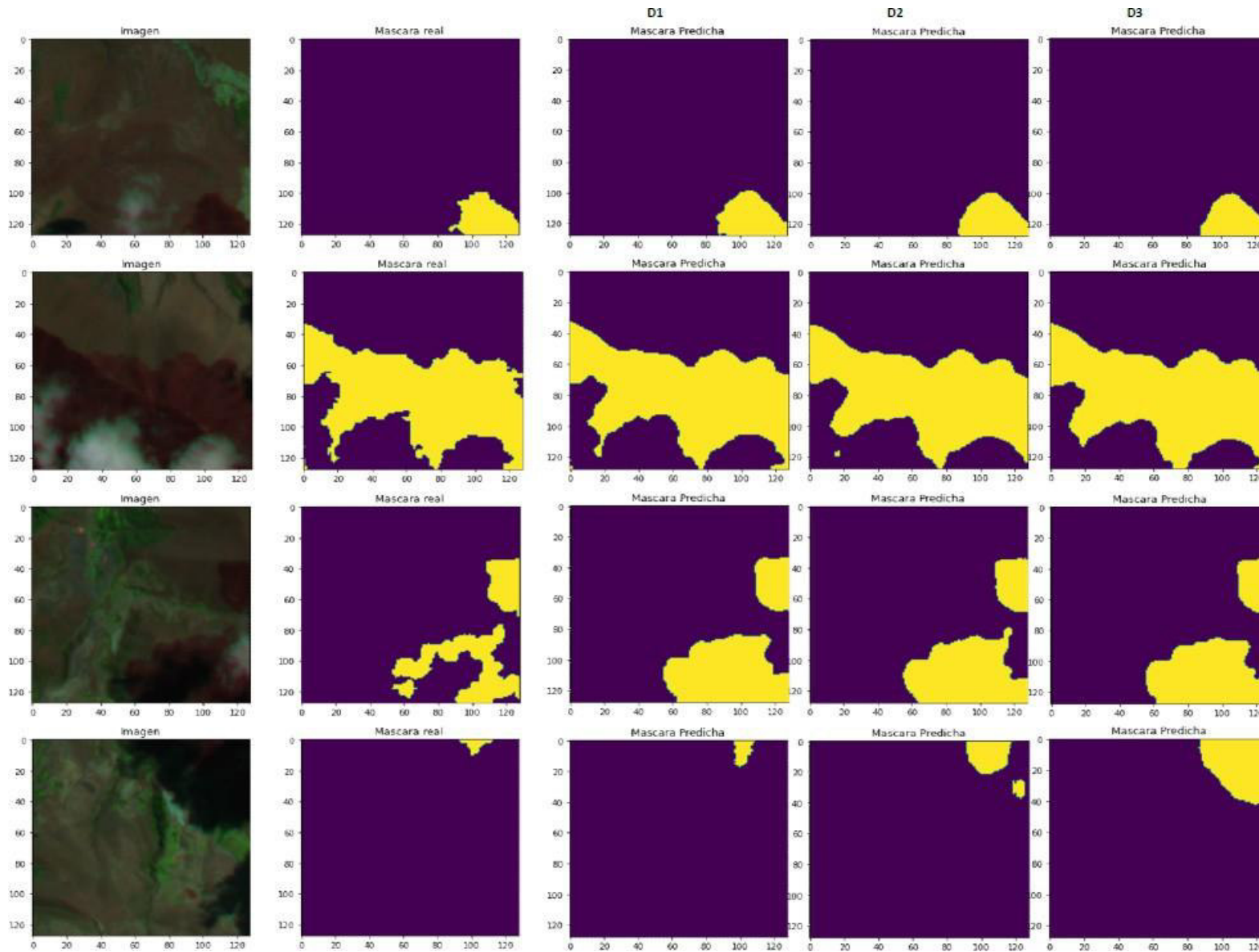
**Figura 39**

*Matriz de Confusión - Grupo G3 - D6 y D6-AD (Aumento de Datos)*



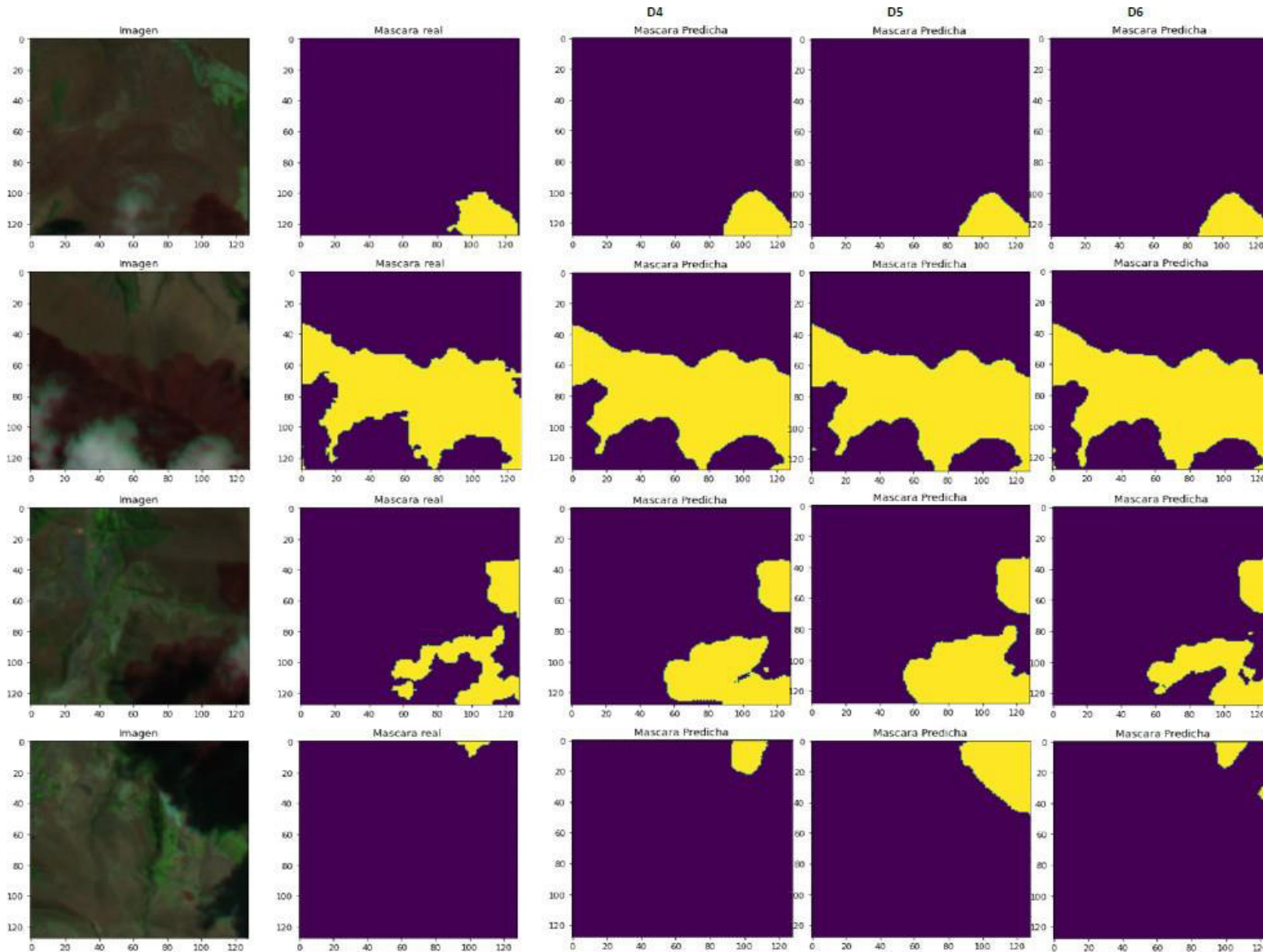
**Figura 40**

*Comparación visual de resultados - Grupo G1 - D1 a D3*



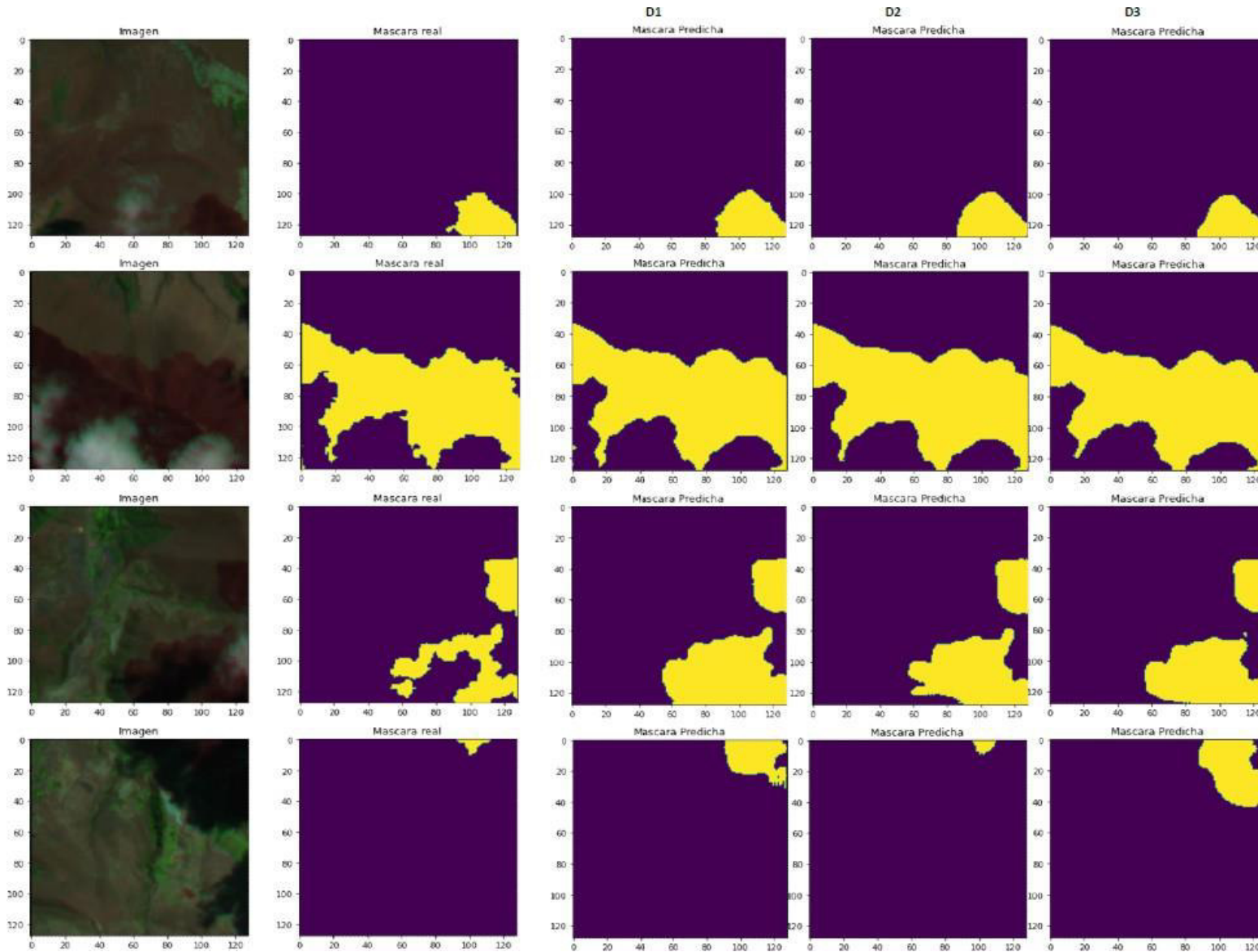
**Figura 41**

*Comparación visual de resultados - Grupo G1 - D4 a D6*



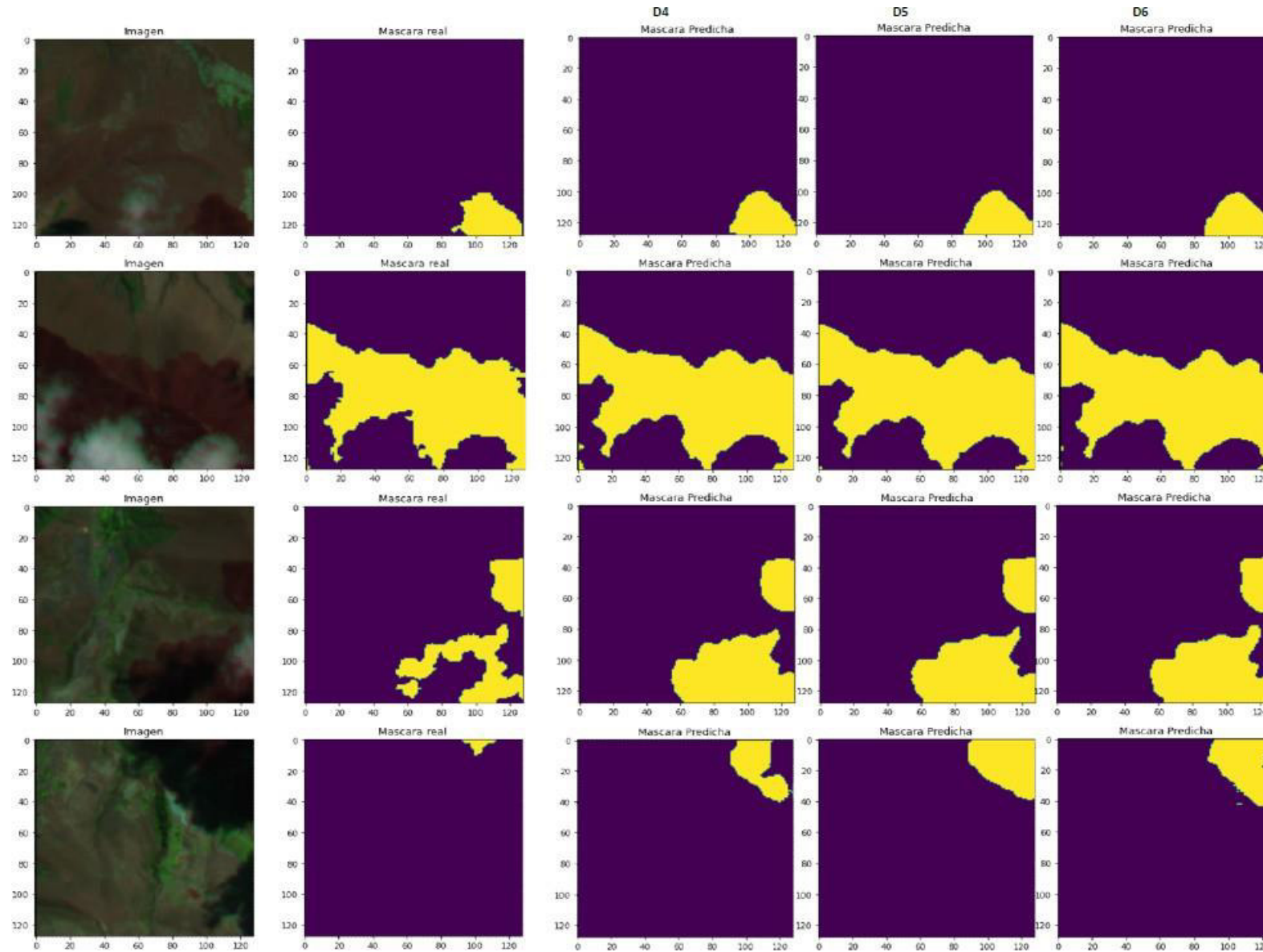
**Figura 42**

*Comparación visual de resultados - Grupo G2 - D1 a D3*



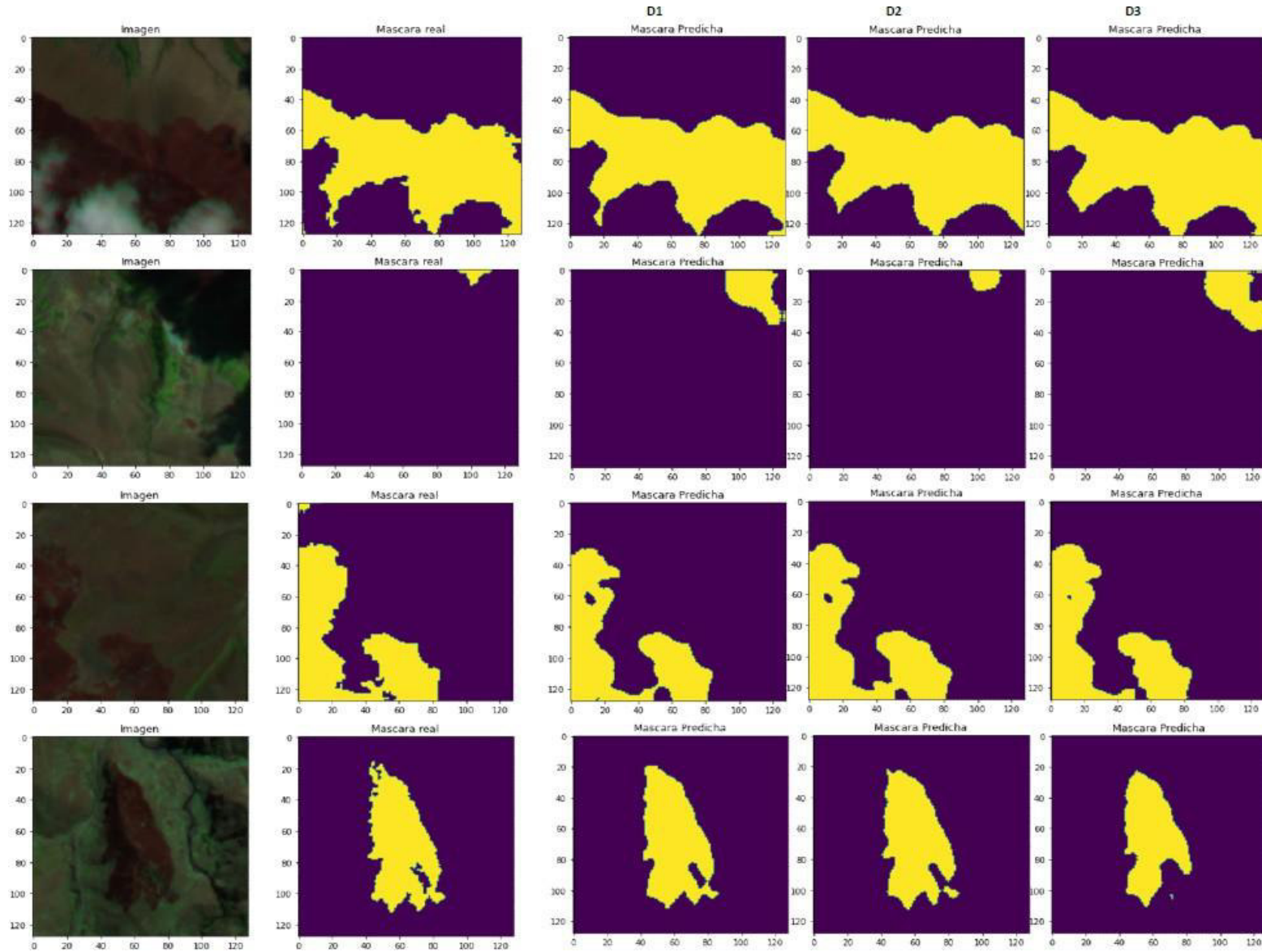
**Figura 43**

*Comparación visual de resultados - Grupo G2 - D4 a D6*



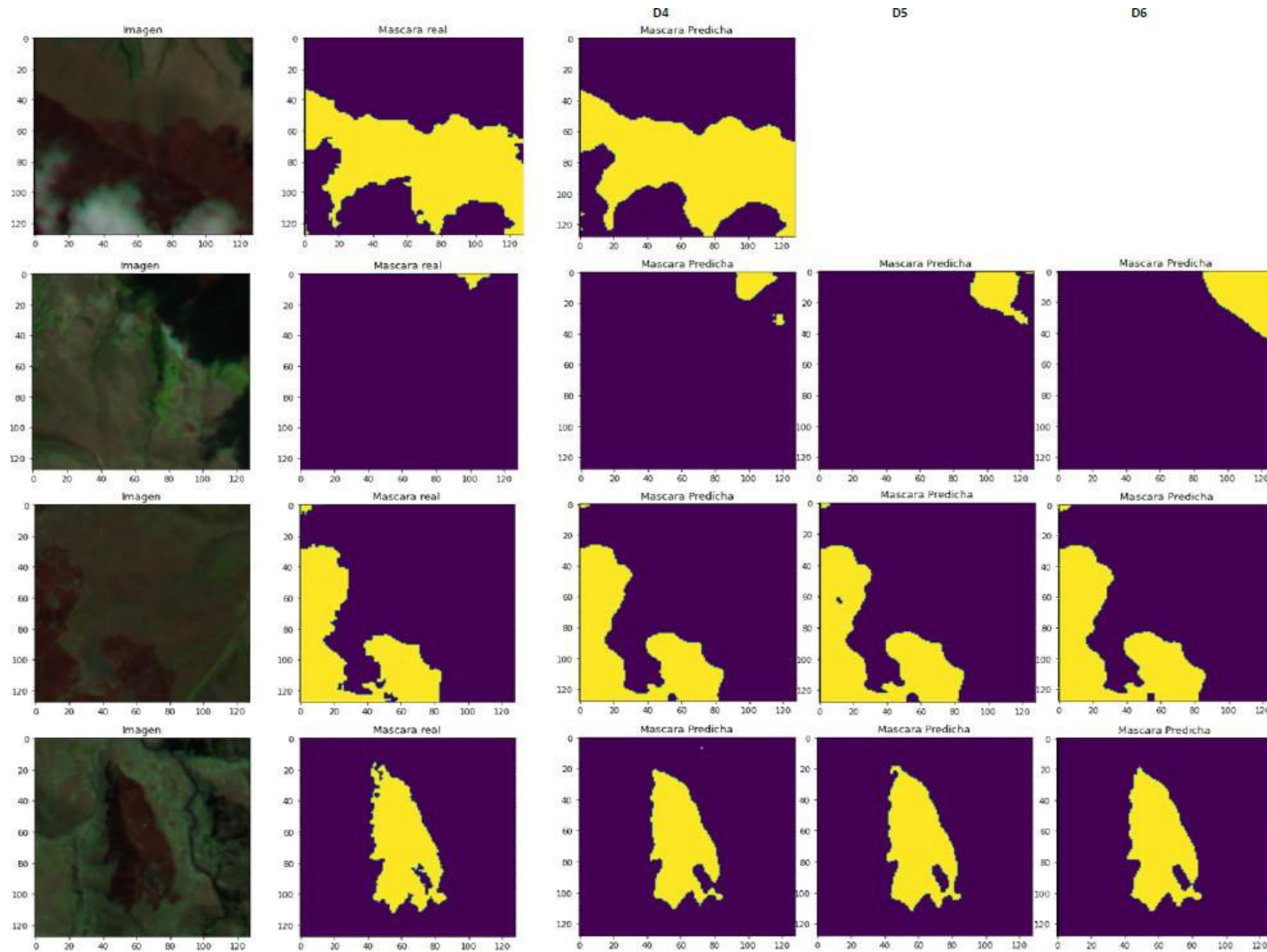
**Figura 44**

*Comparación visual de resultados - Grupo G3 - D1 a D3*



**Figura 45**

*Comparación visual de resultados - Grupo G3 - D4 a D6*



## V. DISCUSIÓN DE RESULTADOS

Los resultados obtenidos han demostrado la posibilidad del uso de técnicas de segmentación semántica con redes neuronales convolucionales mediante la arquitectura Encoder-Decoder usando el modelo UNet para la delimitación de la superficie quemada (cicatriz) de un incendio forestal. El modelo resultante fue entrenado con varios grupos de conjunto de datos, en todos se tiene datos desbalanceados y a su vez cada grupo tiene subgrupos por combinación de bandas para ver cuál de ellos impacta de mayor manera en la comprobación de las hipótesis. Según podemos observar en la Tabla 29, se tiene las métricas promedio: Exactitud de 99 %, Precision de 92 %, Recall de 91 %, IoU de 0.84, F1-Score de 0.91, y Kappa de 0.87; todos en el conjunto de datos de prueba (no conocidos por el modelo). Dentro del Grupo G1, observamos que las métricas aumentan de valor en subgrupos de 6 bandas en lugar de 3 bandas, eso es comprensible ya que hay mayor información espectral; a excepción del 'Recall'. De similar manera se presenta en el Grupo G2. En el Grupo G3, observamos una variación, todas aumentan a excepción de 'Precision'. Todos estos promedios obtenidos aún el de la métrica 'IoU' con el valor más bajo son considerados 'Muy Bueno' como se manifestó anteriormente.

De las tablas de resultados podemos apreciar lo siguiente. En el grupo G1, la combinación D1 (B12-B8-B4) obtuvo mayor valor de 'Precision' entre los subgrupos de 3 bandas, mientras que la combinación D6 (B12-B11-B8A-B4-B3-B2) obtuvo mayor valor de dicha métrica entre los subgrupos de 6 bandas. Esta última combinación también tiene el mayor valor de todo el grupo. De similar forma podemos apreciar las métricas restantes, y en el resto de los grupos. Cuando el desbalance de los datos se reduce, se aprecia un aumento en los valores de las métricas.

De las figuras de la Matrices de Confusión, podemos afirmar que siguen la misma

tendencia de las métricas. Los valores de Falsos Positivos (píxeles estimados como quemados pero que no lo son) son los menores en todas ellas, esto se aprecia en la métrica 'Precision'. En segundo lugar, se encuentran los Falsos Negativos (píxeles estimados no quemados pero lo son) asociado con la métrica 'Recall'. A pesar de ello, los valores de ambas métricas son entre 'Muy Bueno' y 'Excelente'. Asimismo, la métrica 'IoU' va de acorde con estos valores, y la tendencia a reducir estos dos elementos de las Matrices genera un aumento en la 'IoU'.

De las figuras de comparación visual, a nivel de grupos se observa una alta semejanza en las máscaras predichas en forma general. A nivel del grupo G1, del subgrupo de 3 bandas las combinaciones D1 y D4 (B12-B11-B8) son más cercanas a la máscara real, mientras del subgrupo de 6 bandas es la combinación D6. En el grupo G2, a nivel de 3 bandas los subgrupos D1 y D4 siguen sobresaliendo, mientras que a nivel de 6 bandas ambos subgrupos brindan máscaras predichas similares. En el grupo G3, el subgrupo D1 sobresale sobre las demás, mientras que el subgrupo D5 a nivel de 6 bandas. Debido a la reducción del desbalance se aprecia que las máscaras predichas van acercándose a sus reales. De las combinaciones que resaltan podemos afirmar que las bandas: B12, B11, B8 (B8A) y B4 son las que brindan una mayor ayuda para la predicción de los píxeles quemados (y reales). Hay que considerar que los datos se encuentran en ecosistemas diferentes por lo que pueden ser afectadas por nubosidad, pendientes, cascadas, etc., todo esto afecta de alguna manera la estimación. Debido a ello se realizó sobre un solo departamento, ya que los restantes tienen ecosistemas similares o diferentes.

En Knopp et al. (2020), también se estudia la demarcación de áreas quemadas utilizando el modelo UNet pero más semejante a la estructura original. Los datos utilizados son más balanceados y cubren una mayor área quemada lo cual genera mayor cantidad de imágenes de entrenamiento en diferentes zonas geográficas de Europa. Las imágenes son procedentes

del sensor Sentinel-2 de manera exclusiva. Verifica en varias combinaciones de bandas proporcionados por el sensor: B1 al B12, todas fueron remuestreadas a 10m. La combinación de bandas: B2, B3, B4, B8, B11, B12, brindó los mejores valores de métricas: 'Precision', 'Recall', y 'Kappa'. Es bueno mencionar que la configuración de entrenamiento del modelo es diferente. Presenta valores de: 'Kappa'=0.94, 'Precision'=95 %, y 'Recall'=95 %.

En Hu et al. (2021), se estudia la aplicación de diversos modelos: UNet, HRNet, Fast-SCNN, KNN, Random Forest, y otros a nivel de umbrales. Menciona que el modelo sea uno que se pueda aplicar a sensores tanto Sentinel-2 como Landsat-8. La configuración de los modelos es diferente, y en cuanto a la UNet es diferente al utilizado en el desarrollo de esta tesis. Menciona que las bandas del Sentinel-2: SWIR (B12 y B11), NIR (B8 y B8A), y color rojo (B5-B7) son más sensitivas al cambio de reflectancia espectral causada por el fuego. Se produce una disminución en las bandas NIR y un incremento en bandas SWIR dependiendo del ecosistema. Para posibilitar la transferencia a datos de Landsat-8, elige la combinación: B8A, B11 y B12. Estas tres bandas se equiparán con las bandas B7, B6 y B5 de imágenes Landsat-8. Obtiene valores de 'mIoU'=0.77 a 0.82 y 'Kappa'=0.75 a 0.8. (Hu et al., 2021)

En Seydi et al. (2022), de manera similar utiliza imágenes postincendio de Sentinel-2. El modelo se inspira en el modelo UNet pero es modificado por los autores en la sección de extracción de características. La data utilizada es balanceada y cubre también una mayor área quemada. Se utilizan las bandas: B2 al B8A, B11 y B12 (10 bandas). Se obtienen valores de 'Exactitud'=96.5 %, 'F1-Score'=0.91, e 'IoU'=0.83, en promedio entre todas sus zonas de prueba. (Seydi et al., 2022)

Viendo los resultados y comparando con los valores de estudios citados podemos afirmar que se respalda las métricas obtenidas y que por lo tanto se confirma que las hipótesis son validadas.

## VI. CONCLUSIONES

- Mediante la presente investigación, se puede concluir que un modelo de aprendizaje profundo utilizando segmentación semántica puede delimitar la superficie afectada denominada cicatriz de área quemada producto de un incendio forestal en tiempo oportuno. Según el ítem 4.3 Hipótesis General podemos ver los resultados obtenidos a través de la métrica 'Precision'. Dicha métrica nos presenta valores entre 0.89 a 0.95, siendo considerados 'Muy Bueno' o 'Excelente'. De esta manera se propone un modelo que pueda apoyar a la UFMS que aplicando en la resolución temporal de los sensores Sentinel-2 y Landsat-8, determinar los píxeles quemados en una escena completa en un tiempo oportuno a su extinción, o quizás en plena propagación.
- De igual manera, se puede concluir que un modelo de RNA aplicando segmentación semántica puede delimitar la superficie afectada en un incendio forestal. Según el ítem 4.3.1 podemos apreciar los resultados obtenidos a través de la métrica 'F1-Score' con valores desde 0.87 a 0.93 consiguiendo una grado de concordancia 'Excelente'.
- De la misma forma, se concluye que a través de los resultados del ítem 4.3.2 que un modelo de RNA mediante un proceso de segmentación semántica puede estimar el área quemada de un incendio forestal con valores de métrica 'IoU' entre 0.78 a 0.87, siendo considerados 'Muy Bueno'. Los resultados obtenidos validan que los modelos de *Deep Learning* ya mencionados usando las CNN pueden ser aplicados no solo para delimitar la cicatriz de un incendio forestal, sino también una adecuada estimación del área afectada. Siendo una alternativa más para que la UFMS pueda calcular las áreas quemadas.
- A la vez, se concluye que a través de los resultados del ítem 4.3.3 que un modelo de RNA mediante un proceso de segmentación semántica puede determinar las bandas de una imagen satelital para delimitar la superficie afectada con valores de métrica 'Recall' entre

0.83 a 0.95, siendo considerados 'Muy Bueno' o 'Excelente'. Los resultados obtenidos validan que los modelos de *Deep Learning* ya mencionados usando las CNN pueden ser aplicados para determinar las bandas de una imagen satelital para delimitar la superficie afectada.

- Finalmente, se concluye que a través de los resultados del ítem 4.3.4 que un modelo de RNA mediante un proceso de segmentación semántica puede determinar una adecuada estimación del área afectada por incendios forestales en la región Cusco con valores de métrica 'Kappa' entre 0.78 a 0.92, siendo considerados 'Muy Bueno' o 'Excelente'. Los resultados obtenidos validan que los modelos de *Deep Learning* ya mencionados usando las CNN pueden ser aplicados para determinar una adecuada estimación del área afectada por incendios forestales en la región Cusco en un porcentaje basado en el valor de la métrica obtenida. Siendo una alternativa más para que la UFMS pueda calcular las áreas quemadas.

## VII. RECOMENDACIONES

- Se ha utilizado una cantidad limitada de cicatrices generadas por los reportes RIF. Para una mejora en los resultados se recomienda generar más RIF de ser posible de los años utilizados con la finalidad de afinar el entrenamiento. Ya teniendo un modelo entrenado se puede reutilizar para mejorar sus métricas. Mientras se van generando más cicatrices, se va entrenando mejor el modelo, por lo cual año a año se tendrá mejores resultados en la determinación de cicatrices.
- Debido a la geografía variable de nuestro país, sobre todo de los diversos ecosistemas presentes en nuestras regiones naturales, se recomienda generar un modelo personalizado para una o varias regiones políticas, esto dependiendo de un análisis previo de los especialistas de la UFMS. Tomando como base el modelo elaborado y realizando pruebas en diferentes regiones podemos lograr su generalización de ser posible o determinar su personalización.
- El modelo utilizado basado en arquitectura Encoder-Decoder y red UNet para el entrenamiento de todos los conjuntos de datos no fue modificado, ni tampoco sus hiperparámetros. Se recomienda realizar evaluaciones modificando la configuración de entrenamiento como la función de pérdida, función de optimización, número de épocas, y tamaño de lotes; luego podría ser el número de ramas de la Unet así como el número de filtros por capa convolucional. Todas estas modificaciones afectan los resultados. Finalmente, en este punto se debe solicitar el apoyo de los especialistas de la UFMS para determinar la combinación de bandas a aplicar estos cambios.
- A pesar de haber obtenido un modelo validado, las técnicas de Deep Learning en el campo de la segmentación semántica poseen otras redes y arquitecturas más aún con el avance de la tecnología. Debido a ello se recomienda el uso de otras técnicas como:

Attention-UNet (Oktay et al., 2018), ResUNet (Zhang et al., 2018), DeepLabV3+ (Chen et al., 2018), HRNet (Wang et al., 2021), Transformer ((Gao et al., 2021), (Yuan et al., 2022), (Gonçalves et al., 2023), (Jamali et al., 2023)).

- Se recomienda la utilización de otros sensores con igual o mejor resolución espacial, como por ejemplo: la constelación Planet de 3m. (Gonçalves et al., 2023), imágenes Radar SAR ((Belenguer-Plomer et al., 2021), (Zhang, 2021), (Hu, 2022)) lo cual debería mejorar los valores obtenidos. Sin embargo, no todos son productos gratuitos, por lo cual no son útiles y eficientes, pudiendo ser una desventaja.

## VIII. REFERENCIAS

- Belenguer-Plomer, M. A., Tanase, M. A., Chuvieco, E., y Bovolo, F. (2021). CNN-based burned area mapping using radar and optical data. *Remote Sensing of Environment*, 260(112468). <https://doi.org/10.1016/j.rse.2021.112468>.
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., y Adam, H. (8 al 14 de setiembre del 2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation[conferencia]. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Alemania. [https://link.springer.com/chapter/10.1007/978-3-030-01234-2\\_49](https://link.springer.com/chapter/10.1007/978-3-030-01234-2_49).
- Chicchón, M. A. (2018). *Fusión de Datos para Segmentación Semántica en Aplicaciones Urbanas de Teledetección Aérea usando Algoritmos de Aprendizaje Profundo*. [Tesis de maestría, Pontificia Universidad Católica del Perú]. Repositorio Institucional PUCP. <https://tesis.pucp.edu.pe/repositorio/handle/20.500.12404/14459>.
- Chuvieco, E. (2016). *Fundamentals of Satellite Remote Sensing - An Environmental Approach*. (2da ed.). CRC Press - Taylor & Francis Group.
- Congalton, R. G. and Green, K. (2019). *Assessing the Accuracy of Remotely Sensed Data - Principles and Practices*. (3ra ed.). CRC Press - Taylor & Francis Group.
- Decreto de Urgencia N° 024-2010. Dictan medidas económicas y financieras para la creación del programa presupuestal estratégico de reducción de la vulnerabilidad y atención de emergencias por desastres en el marco del presupuesto por resultados. (31 de marzo del 2010). <https://www.gob.pe/institucion/mef/normas-legales/226514-024-2010>.

Decreto Supremo N° 034-2014-PCM. Decreto Supremo que aprueba el Plan Nacional de Gestión del Riesgo de Desastres (PLANAGERD 2014-2021). (13 de mayo del 2014). <https://sinia.minam.gob.pe/normas/decreto-supremo-que-aprueba-plan-nacional-gestion-riesgo-desastres>.

Decreto Supremo N° 048-2011-PCM. Decreto Supremo que aprueba el Reglamento de la Ley N° 29664, que crea el Sistema Nacional de Gestión del Riesgo de Desastres (SINAGERD). (26 de mayo del 2011). <https://www.gob.pe/institucion/indeci/informes-publicaciones/1093199-decreto-supremo-n-048-2011-pcm-y-su-modificatoria>.

Decreto Supremo N° 111-2012-PCM. Decreto Supremo que incorpora la Política Nacional de Gestión del Riesgo de Desastres como Política Nacional de obligatorio cumplimiento para las entidades del Gobierno Nacional. (2 de noviembre del 2012). <https://www.gob.pe/institucion/pcm/normas-legales/466435-111-2012-pcm>.

Decreto Supremo N° 115-2022-PCM. Decreto Supremo que aprueba Plan Nacional de Gestión del Riesgo de Desastres (PLANAGERD 2022-2030). (12 de setiembre del 2022). <https://www.gob.pe/institucion/pcm/normas-legales/3465952-115-2022-pcm>.

ESA (2015). *Sentinel-2 User Handbook*. European Space Agency (Issue 1 Rev 2). [https://sentinel.esa.int/documents/247904/685211/Sentinel-2\\_User\\_Handbook](https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook).

ESA (2023). Copernicus Open Access Hub - Plataforma de Descarga de imágenes Sentinel: S1, S2 y S3. European Space Agency. <https://scihub.copernicus.eu/dhus/#/home>.

- Firigato, J. O. (2021). *FORESTRYVIEW APP: MAPEAMENTO DA SILVICULTURA COM USO DE SENSORES REMOTOS, DEEP LEARNING E COMPUTAÇÃO EM NUVEM*. [Tesis de maestría, Universidade Federal de Mato Grosso Do Sul]. Repositorio Institucional UFMS. <https://repositorio.ufms.br/handle/123456789/3856>.
- Gao, L., Liu, H., Yang, M., Chen, L., Wan, Y., Xiao, Z., y Qian, Y. (2021). STransFuse: Fusing Swin Transformer and Convolutional Neural Network for Remote Sensing Image Semantic Segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14,10990-11003. <https://doi.org/10.1109/JSTARS.2021.3119654>.
- Gonçalves, D. N., Marcato, J., Carrilho, A. C., Acosta, P. R., Ramos, A. P. M., Gomes, F. D. G., Osco, L. P., da Rosa Oliveira, M., Martins, J. A. C., Damasceno, G. A., de Araújo, M. S., Li, J., Roque, F., de Faria Peres, L., Gonçalves, W. N., y Libonati, R. (2023). Transformers for mapping burned areas in Brazilian Pantanal and Amazon with PlanetScope imagery. *International Journal of Applied Earth Observation and Geoinformation*, 116(103151). <https://doi.org/10.1016/j.jag.2022.103151>.
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow*. (2da ed.). O'Reilly Media.
- Hernández, R., Baptista, P., y Fernández, C. (2014). *Metodología de la investigación*. (6ta ed.). McGraw-Hill.
- Hoeser, T., Bachofer, F., y Kuenzer, C. (2020). Object Detection and Image Segmentation

- with Deep Learning on Earth Observation Data: A Review-Part II: Applications. *Remote Sensing*, 12(18),3053. <https://doi.org/10.3390/rs12183053>.
- Hoeser, T. y Kuenzer, C. (2020). Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part I: Evolution and Recent Trends. *Remote Sensing*, 12(10),1667. <https://doi.org/10.3390/rs12101667>.
- Hu, X. (2022). *Large-Scale Multi-Source Satellite Data for Wildfire Detection and Assessment Using Deep Learning*. [Tesis de doctorado, KTH Royal Institute of Technology]. Repositorio Institucional KTH. <http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1658529&dswid=-4448>.
- Hu, X., Ban, Y., y Nascetti, A. (2021). Uni-Temporal Multispectral Imagery for Burned Area Mapping with Deep Learning. *Remote Sensing*, 13(8),1509. <https://doi.org/10.3390/rs13081509>.
- Instituto Geográfico Nacional (2020). *El Programa Copernicus para la monitorización del territorio y los Objetivos del Desarrollo Sostenible*. Centro Nacional de Información Geográfica (Ministerio de Transportes, Movilidad y Agenda Urbana). <https://www.ign.es/web/libros-digitales/programa-copernicus-monitorizacion-territorio>.
- Jamali, A., Roy, S. K., Li, J., y Ghamisi, P. (2023). TransU-Net++: Rethinking attention gated TransU-Net for deforestation mapping. *International Journal of Applied Earth Observation and Geoinformation*, 120(103332). <https://doi.org/10.1016/j.jag.2023.103332>.
- Knopp, L., Wieland, M., Rättich, M., y Martinis, S. (2020). A deep learning approach

for burned area segmentation with sentinel-2 data. *Remote Sensing*, 12(15),2422.  
<https://doi.org/10.3390/rs12152422>.

Lazo, W. A. (2019). *Espectroscopía con Infrarrojo y Técnicas de Machine Learning y Deep Learning para la Detección y Clasificación de Frutas para la Agroindustria. Caso: Arándanos - Empresa TalSA - 2018*. [Tesis de maestría, Universidad Privada Antenor Orrego]. Repositorio Institucional UPAO. <https://repositorio.upao.edu.pe/handle/20.500.12759/5106>.

Ley N° 29664. Ley que crea el Sistema Nacional de Gestión del Riesgo de Desastres (SINAGERD). (19 de febrero del 2011). <https://www.leyes.congreso.gob.pe/Documentos/Leyes/29664.pdf>.

Ley N° 29763. Ley Forestal y de Fauna Silvestre, creación del Servicio Nacional Forestal y de Fauna Silvestre (SERFOR). (22 de julio del 2011). <https://www.leyes.congreso.gob.pe/documentos/leyes/29763.pdf>.

Ley N° 30754. Ley Marco sobre Cambio Climático. (18 de abril del 2018). <https://busquedas.elperuano.pe/dispositivo/NL/1638161-1>.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., y Rueckert, D. (4 al 6 de julio del 2018). Attention U-Net: Learning Where to Look for the Pancreas[conferencia]. In *MIDL 2018 - Medical Imaging with Deep Learning Conference*, Amsterdam,Holanda. <https://openreview.net/pdf?id=Skft7cijM>.

Pajares, G., Herrera, P., y Besada, E. (2021). *Aprendizaje Profundo*. RC Libros.

Resolución de Dirección Ejecutiva N°284-2018-MINAGRI-SERFOR-DE. Plan de Prevención y Reducción de Riesgos de Incendios Forestales. (17 de diciembre del 2018). Servicio Nacional Forestal y de Fauna Silvestre - Ministerio de Agricultura y Riego.

Resolución de Gerencia General N°006-2021-MIDAGRI-SERFOR-GG. Conformar la Unidad Funcional de Monitoreo Satelital (UFMS). (22 de marzo del 2021). Servicio Nacional Forestal y de Fauna Silvestre - Ministerio de Desarrollo Agrario y Riego.

Resolución Ministerial N° 046-2013-PCM. Aprueban Directiva "Lineamientos que definen el Marco de Responsabilidades en Gestión del Riesgo de Desastres, de las entidades del estado en los tres niveles de gobierno" y su anexo. (16 de febrero del 2013). <https://www.gob.pe/institucion/indeci/informes-publicaciones/1093220-resolucion-ministerial-n-046-2013-pcm>.

Ronneberger, O., Fischer, P., y Brox, T. (5 al 9 de octubre del 2015). U-Net: Convolutional Networks for Biomedical Image Segmentation[conferencia]. In *MICCAI 2015 - Medical Image Computing-Assisted Intervention*, Munich,Alemania. [https://link.springer.com/chapter/10.1007/978-3-319-24574-4\\_28](https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28).

Seydi, S. T., Hasanlou, M., y Chanussot, J. (2022). Burnt-Net: Wildfire burned area mapping with single post-fire Sentinel-2 data and deep learning morphological neural network. *Ecological Indicators*, 140(180999). <https://doi.org/10.1016/j.ecolind.2022.108999>.

Shanmugamani, R. (2018). *Deep Learning for Computer Vision*. Packt Publishing.

Torres, J. (2020). *Python Deep Learning - Introducción práctica con Keras y TensorFlow 2*. Editorial Marcombo.

- USGS (2019). *Landsat 8 (L8) Data Users Handbook*. U.S. Geological Survey (Versión 5.0).  
<https://www.usgs.gov/media/files/landsat-8-data-users-handbook>.
- USGS (2023). EarthExplorer - Plataforma de Descarga de imágenes Landsat: L8, entre otros sensores. U.S. Geological Survey. <https://earthexplorer.usgs.gov/>.
- Véliz, C. (2018). *Aprendizaje Automático - Análisis para la minería de datos y big data*. Pontificia Universidad Católica del Perú.
- Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., y Xiao, B. (2021). Deep High-Resolution Representation Learning for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10),3349-3364. <https://doi.org/10.1109/TPAMI.2020.2983686>.
- Yuan, J., Wang, L., y Cheng, S. (2022). STransUNet: A Siamese TransUNet-Based Remote Sensing Image Change Detection Network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15,9241-9253. <https://doi.org/10.1109/JSTARS.2022.3217038>.
- Zhang, P. (2021). *Deep Learning for Wildfire Progression Monitoring Using SAR and Optical Satellite Image Time Series*. [Tesis de doctorado, KTH Royal Institute of Technology]. Repositorio Institucional KTH. <http://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A1557429&dswid=-801>.
- Zhang, Z., Liu, Q., y Wang, Y. (2018). Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5),749-753. <https://doi.org/10.1109/LGRS.2018.2802944>.

## IX. ANEXOS

### 9.1 Matriz de consistencia

PROBLEMA	OBJETIVOS	HIPOTESIS	VARIABLES	DIMENSIONES	INDICADORES	INSTRUMENTOS
<b>Problema General</b> ¿Cómo puedo obtener en tiempo oportuno al incendio forestal la delimitación del área afectada utilizando técnicas de aprendizaje profundo (RNA) en imágenes satelitales?	<b>Objetivo General</b> Delimitar el área afectada por un incendio forestal en tiempo oportuno mediante el uso de técnicas de aprendizaje profundo (RNA) en imágenes satelitales.	<b>Hipótesis General</b> Un modelo de aprendizaje profundo utilizando segmentación semántica nos permitirá delimitar la superficie afectada denominada cicatriz de área quemada producto de un incendio forestal en tiempo oportuno.	<b>Independiente</b> Un modelo generado mediante técnicas de aprendizaje profundo utilizando segmentación semántica	Precisión	Métrica: Precisión - estimación de áreas afectadas	- Imágenes Satelitales de la ESA (Sentinel-2) y de la NASA-USGS (Landsat-8) - Software de Procesamiento Digital de Imágenes - Software de Sistema de Información Geográfica - Lenguaje de programación Python - Librerías Tensorflow, Keras, y otras
<b>Problemas Específicos</b> a) ¿Cómo se puede obtener un modelo óptimo de las RNA que permita delimitar la superficie afectada?	<b>Objetivos Específicos</b> a) Obtener un modelo óptimo de las RNA que permita delimitar la superficie afectada	<b>Hipótesis Específicas</b> a) Un modelo óptimo de las RNA permitirá delimitar la superficie afectada	<b>Dependiente</b> Delimitación de la superficie afectada (cicatriz) realizada por un modelo de aprendizaje profundo	Calidad general	Métrica: F-Score o Coeficiente Dice - calidad general del modelo para delimitar la superficie	
b) ¿Cómo se puede obtener un modelo óptimo de las RNA que permita realizar una adecuada estimación del área afectada?	b) Obtener un modelo óptimo de las RNA que permita realizar una adecuada estimación del área afectada	b) Un modelo óptimo de las RNA permitirá realizar una adecuada estimación del área afectada		Certeza sobre área afectada predicha	Métrica: IoU o índice Jaccard - certeza sobre área afectada estimada versus área afectada real	
c) ¿Cómo se puede obtener un modelo óptimo de las RNA que permita determinar las bandas de una imagen satelital para delimitar la superficie afectada?	c) Obtener un modelo óptimo de las RNA que permita determinar las bandas de una imagen satelital para delimitar la superficie afectada	c) Un modelo óptimo de las RNA permitirá determinar las bandas de una imagen satelital para delimitar la superficie afectada		Sensibilidad	Métrica: Recall o Sensibilidad - determinación de bandas de una imagen satelital	
d) ¿Cómo se puede obtener un modelo óptimo de las RNA que permita determinar una adecuada estimación del área afectada por incendios forestales en la región Cusco?	d) Obtener un modelo óptimo de las RNA que permita determinar una adecuada estimación del área afectada por incendios forestales en la región Cusco	d) Un modelo óptimo de las RNA permitirá determinar una adecuada estimación del área afectada por incendios forestales en la región Cusco		Kappa	Métrica: Coeficiente Kappa - estimación de área afectada de región Cusco	

## 9.2 Validación y confiabilidad de instrumentos

La validación y confiabilidad de los instrumentos viene dada por la gran cantidad de estudios tanto a nivel de centros de investigación como de entidades gubernamentales a nivel mundial.

El lenguaje de programación Python es una herramienta de código libre utilizado a nivel mundial para diversas aplicaciones, pero en los últimos años ha subido en el nivel de preferencia en el ámbito de la Inteligencia Artificial, más específicamente en una de sus ramas: Machine Learning. Y a su vez, en una rama del Machine Learning: Deep Learning.

Como se menciona en la sección de antecedentes, se detalla los diversos estudios donde se utilizan las librerías de Tensorflow, Keras y otras en diversos campos de aplicación: Transporte, Coberturas de uso de la Tierra, Asentamientos Humanos, Peligros Naturales, Vegetación Natural, Agricultura, y otros.

De igual manera, en dichos estudios, la Teledetección utiliza datos remotos de observación terrestre, siendo uno de ellos las imágenes satelitales. Los satélites mencionados Sentinel-2 y Landsat-8 son utilizados como fuente de datos de tipo óptico para diversas aplicaciones, más aún debido a su naturaleza gratuita es usada a nivel mundial.

Todos los profesionales inmersos en los estudios de investigación en el campo de la Teledetección utilizan software de Procesamiento Digital de Imágenes, así como de Sistemas de Información Geográfica. Los diversos estudios dan como resultados datos de vectorial o raster dependiendo de la naturaleza de la investigación, que luego son complementados con otros datos similares para que a través de estos programas informáticos se generen los mapas que luego son mostrados al público en general en forma digital o impresa.

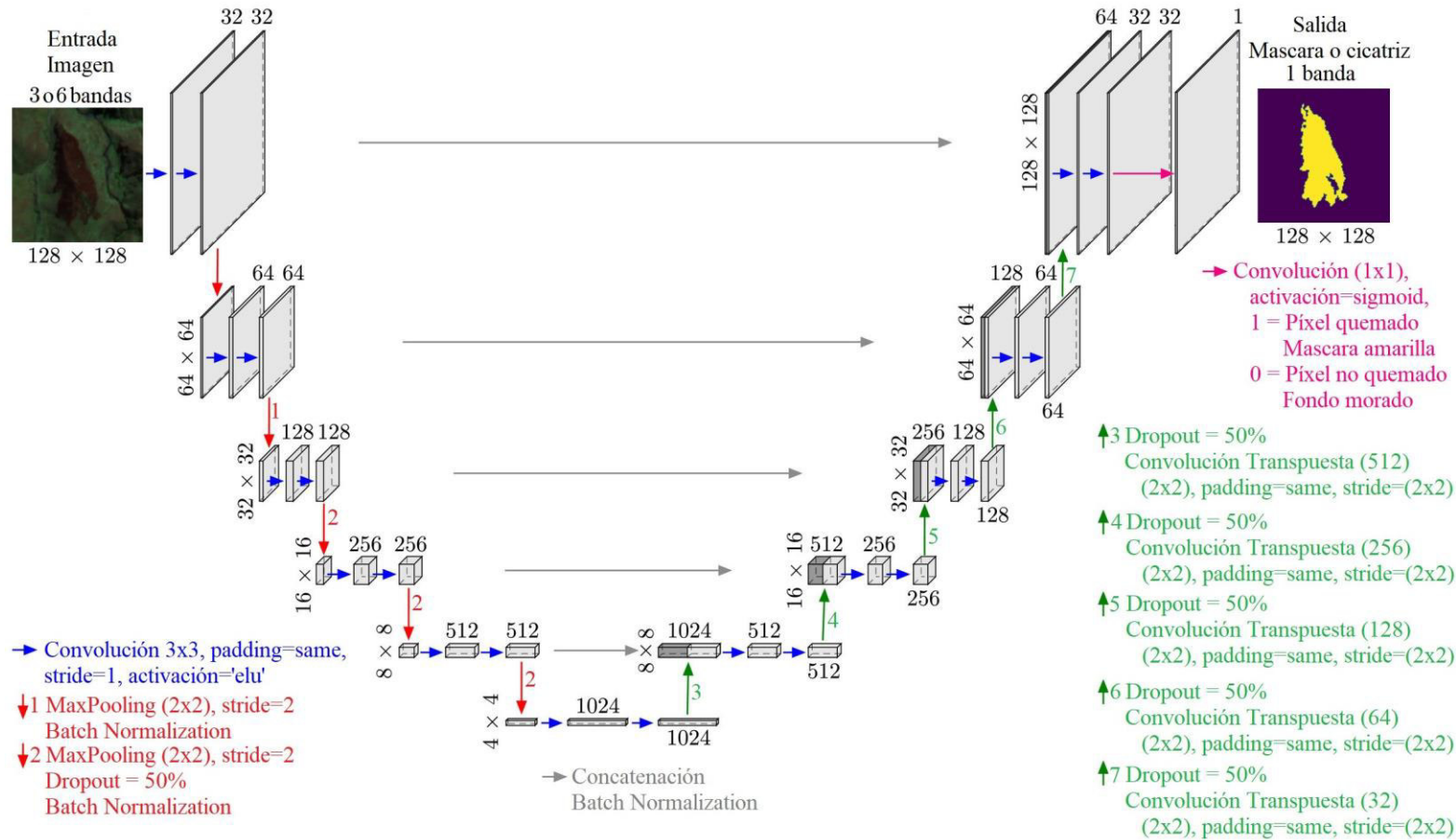
Para reforzar la confiabilidad de los instrumentos podemos mencionar que (Knopp et al.,

2020) obtuvo los siguientes rangos de valores de métricas en su investigación:

- Métrica Recall o Sensibilidad: 0.97 a 0.98, o 97 a 98 %
- Métrica Precisión: 0.77 a 0.96, o 77 a 96 %
- Métrica F1-Score: 0.86 a 0.93, o 86 a 93 %

Siendo estos valores muy buenos, y teniendo el valor de 1 como máximo en estas métricas. Similar rango de valores presenta la métrica IoU.

### 9.3 Arquitectura del Modelo UNet propuesto



#### 9.4 Listado de imágenes satelitales utilizadas

##### Año 2017:

- S2A\_MSIL1C\_20171005T150721\_N0205\_R082\_T18LYL\_20171005T151102
- S2A\_MSIL1C\_20171002T150621\_N0205\_R039\_T18LZK\_20171002T150618
- S2A\_MSIL1C\_20171121T145721\_N0206\_R039\_T18LZL\_20171121T164001
- S2A\_MSIL1C\_20170902T145731\_N0205\_R039\_T19LBE\_20170902T145910

##### Año 2018:

- S2A\_MSIL1C\_20180907T145721\_N0206\_R039\_T18LZK\_20180907T194320
- S2B\_MSIL1C\_20181214T150719\_N0207\_R082\_T18LYL\_20181214T195124
- S2A\_MSIL1C\_20180927T145721\_N0206\_R039\_T18LZK\_20180927T184107

##### Año 2019:

- S2B\_MSIL1C\_20190729T145739\_N0208\_R039\_T18LYL\_20190729T181225

##### Año 2020:

- S2A\_MSIL1C\_20200529T145731\_N0209\_R039\_T19LBE\_20200529T181559
- S2A\_MSIL1C\_20200529T145731\_N0209\_R039\_T18LZK\_20200529T181559
- S2A\_MSIL1C\_20200807T145731\_N0209\_R039\_T19LBE\_20200807T183352
- S2A\_MSIL1C\_20200807T145731\_N0209\_R039\_T18LZK\_20200807T183352
- S2A\_MSIL1C\_20200810T150731\_N0209\_R082\_T18LYL\_20200810T183915
- S2B\_MSIL1C\_20200825T150729\_N0209\_R082\_T18LYL\_20200825T184343
- S2B\_MSIL1C\_20200726T150719\_N0209\_R082\_T18LYL\_20200726T184122
- S2B\_MSIL1C\_20200924T150719\_N0209\_R082\_T18LYM\_20200924T184416
- S2A\_MSIL1C\_20201115T145731\_N0209\_R039\_T18LZL\_20201115T182617

- S2A\_MSIL1C\_20201006T145731\_N0209\_R039\_T18LZL\_20201006T183129
- S2B\_MSIL1C\_20200703T145729\_N0209\_R039\_T18LZL\_20200703T181136
- S2A\_MSIL1C\_20201115T145731\_N0209\_R039\_T18LZL\_20201115T182617
- S2A\_MSIL1C\_20201125T145731\_N0209\_R039\_T18LZL\_20201125T182935
- S2A\_MSIL1C\_20200728T145731\_N0209\_R039\_T18LZL\_20200728T182912
- S2A\_MSIL1C\_20201115T145731\_N0209\_R039\_T18LYL\_20201115T182617
- S2A\_MSIL1C\_20200807T145731\_N0209\_R039\_T18LYL\_20200807T183352
- S2A\_MSIL1C\_20201115T145731\_N0209\_R039\_T18LYL\_20201115T182617
- S2A\_MSIL1C\_20201115T145731\_N0209\_R039\_T18LZL\_20201115T182617
- S2B\_MSIL1C\_20201213T150719\_N0209\_R082\_T18LXL\_20201213T183731
- S2A\_MSIL1C\_20201108T150721\_N0209\_R082\_T18LYL\_20201108T183706
- S2A\_MSIL1C\_20201009T150731\_N0209\_R082\_T18LYL\_20201009T184252
- S2A\_MSIL1C\_20200830T150731\_N0209\_R082\_T18LYL\_20200830T184442
- S2B\_MSIL1C\_20200825T150729\_N0209\_R082\_T18LYL\_20200825T184343
- S2B\_MSIL1C\_20200924T150719\_N0209\_R082\_T18LYL\_20200924T184416
- S2B\_MSIL1C\_20200924T150719\_N0209\_R082\_T18LYM\_20200924T184416

**Año 2021:**

- S2B\_MSIL1C\_20210817T145729\_N0301\_R039\_T18LZK\_20210817T182021
- S2A\_MSIL1C\_20210904T150721\_N0301\_R082\_T18LYL\_20210904T183612
- S2A\_MSIL1C\_20210825T150721\_N0301\_R082\_T18LYL\_20210825T183706
- S2B\_MSIL1C\_20210926T145729\_N0301\_R039\_T18LZK\_20210926T182713

- S2B\_MSIL1C\_20210926T145729\_N0301\_R039\_T18LZL\_20210926T182713
- S2B\_MSIL1C\_20210926T145729\_N0301\_R039\_T18LYL\_20210926T182713
- S2A\_MSIL1C\_20211021T145731\_N0301\_R039\_T18LZK\_20211021T182821
- S2A\_MSIL1C\_20211004T150731\_N0301\_R082\_T18LYL\_20211004T184005
- S2A\_MSIL1C\_20210805T150721\_N0301\_R082\_T18LYL\_20210805T183535
- S2A\_MSIL1C\_20210723T145731\_N0301\_R039\_T18LZK\_20210723T182224
- S2A\_MSIL1C\_20210713T145731\_N0301\_R039\_T18LZK\_20210713T182332
- S2B\_MSIL1C\_20210708T145729\_N0301\_R039\_T18LZK\_20210708T182133
- S2A\_MSIL1C\_20210713T145731\_N0301\_R039\_T18LZL\_20210713T182332
- S2B\_MSIL1C\_20210628T145729\_N0300\_R039\_T18LZJ\_20210628T182224
- S2A\_MSIL1C\_20210703T145731\_N0301\_R039\_T18LZK\_20210706T140541
- S2B\_MSIL1C\_20210807T145729\_N0301\_R039\_T18LZK\_20210807T185658
- S2A\_MSIL1C\_20210706T150721\_N0301\_R082\_T18LXL\_20210706T183517
- S2B\_MSIL1C\_20210701T150719\_N0301\_R082\_T18LXL\_20210701T183442
- S2A\_MSIL1C\_20211021T145731\_N0301\_R039\_T18LYL\_20211021T182821
- S2A\_MSIL1C\_20210901T145731\_N0301\_R039\_T18LZK\_20210901T182339
- S2B\_MSIL1C\_20211026T145729\_N0301\_R039\_T18LZK\_20211026T182618
- S2B\_MSIL1C\_20210810T150719\_N0301\_R082\_T18LYL\_20210810T184620
- S2A\_MSIL1C\_20210812T145731\_N0301\_R039\_T18LZK\_20210812T182223
- S2A\_MSIL1C\_20211113T150721\_N0301\_R082\_T18LYL\_20211113T183645
- S2A\_MSIL1C\_20210812T145731\_N0301\_R039\_T18LZL\_20210812T182223

- S2A\_MSIL1C\_20211031T145731\_N0301\_R039\_T18LZK\_20211031T182400
- S2B\_MSIL1C\_20210916T145719\_N0301\_R039\_T18LZK\_20210916T201122
- S2B\_MSIL1C\_20210820T150719\_N0301\_R082\_T18LYM\_20210820T183848
- S2A\_MSIL1C\_20211001T145731\_N0301\_R039\_T18LZJ\_20211001T183529
- S2B\_MSIL1C\_20211225T145719\_N0301\_R039\_T18LZL\_20211225T164524
- S2B\_MSIL1C\_20210817T145729\_N0301\_R039\_T19LBE\_20210817T182021
- S2A\_MSIL1C\_20210924T150721\_N0301\_R082\_T18LXL\_20210924T183853
- S2B\_MSIL1C\_20210817T145729\_N0301\_R039\_T19LBD\_20210817T182021
- S2B\_MSIL1C\_20210817T145729\_N0301\_R039\_T18LZL\_20210817T182021
- S2A\_MSIL1C\_20210822T145731\_N0301\_R039\_T18LZL\_20210822T182354
- S2B\_MSIL1C\_20210807T145729\_N0301\_R039\_T18LZJ\_20210807T185658
- S2B\_MSIL1C\_20210817T145729\_N0301\_R039\_T19LBF\_20210817T182021
- S2A\_MSIL1C\_20211001T145731\_N0301\_R039\_T19LBD\_20211001T183529
- S2A\_MSIL1C\_20210901T145731\_N0301\_R039\_T18LYL\_20210901T182339
- S2B\_MSIL1C\_20210919T150719\_N0301\_R082\_T18LYM\_20210919T183513
- S2B\_MSIL1C\_20210807T145729\_N0301\_R039\_T18LZL\_20210807T185658
- S2A\_MSIL1C\_20210822T145731\_N0301\_R039\_T18LZK\_20210822T182354
- S2B\_MSIL1C\_20211026T145729\_N0301\_R039\_T18LZL\_20211026T182618
- S2B\_MSIL1C\_20211026T145729\_N0301\_R039\_T18LZJ\_20211026T182618
- S2B\_MSIL1C\_20211105T145729\_N0301\_R039\_T18LZK\_20211105T182422
- S2A\_MSIL1C\_20211021T145731\_N0301\_R039\_T18LZL\_20211021T182821