



# **Universidad Nacional Mayor de San Marcos**

**Universidad del Perú. Decana de América**

**Facultad de Ingeniería Geológica, Minera, Metalúrgica y Geográfica**

**Escuela Profesional de Ingeniería Geográfica**

## **Recuperación de imágenes Landsat MSS (1972-1999) en el Perú mediante inteligencia artificial hacia una armonización efectiva del monitoreo global y a largo plazo**

### **TESIS**

Para optar el Título Profesional de Ingeniero Geógrafo

### **AUTOR**

Julio Cesar CONTRERAS HUERTA

### **ASESOR**

Dr. Francisco Alejandro ALCÁNTARA BOZA

Lima - Perú

2024

*Dedico esta tesis a mi madre, mi hermana  
y mis amigos, cuyo apoyo incondicional  
y aliento constante hicieron posible  
la realización de mi carrera  
profesional. Gracias por  
siempre estar ahí.*

## **AGRADECIMIENTOS**

Deseo comenzar esta tesis expresando mi más sincera y profunda gratitud hacia todas las personas e instituciones que me brindaron su apoyo y contribuyeron a la realización de esta investigación.

Mi agradecimiento se extiende a ProCiencia, la Universidad Cayetano Heredia, la Universidad Nacional Mayor de San Marcos y la Universidad de Valencia. El financiamiento proporcionado a través del Proyecto: “AndesDataCube: Desarrollo de algoritmos de inteligencia artificial para la cuantificación de activos naturales en los Andes del Perú a partir del procesamiento de imágenes de satélite”, con contrato N°PE501083135-2023-PROCIENCIA, ha sido esencial para el éxito de esta investigación.

Al Dr. Alejandro Alcántara Boza, asesor de Tesis, por sus sugerencias y constante apoyo en el presente trabajo de investigación. Asimismo, debo agradecer al Dr. Cesar Aybar, quien estuvo conmigo desde el inicio de esta tesis, por siempre estar dispuesto a ayudar y compartir sus conocimientos de manera incondicional.

Quisiera expresar mi gratitud a mis colegas Roy Yali, Fernando Prudencio, Jair Flores y José Del Castillo por su apoyo emocional y motivación a lo largo de esta jornada académica. Su presencia en los momentos más retadores de este proceso ha sido de un valor incalculable para mí.

Y por último, quiero dar las gracias a mi madre Celia y mi hermana Luz María. Sin su amor, paciencia y comprensión, enfrentar los desafíos de este proyecto de tesis no habría sido posible. Su apoyo incondicional ha sido el pilar de mi fortaleza y fuente constante de ánimo.

## ÍNDICE

<b>CAPÍTULO I</b>	<b>1</b>
<b>1. INTRODUCCIÓN</b>	<b>2</b>
1.1 Introducción . . . . .	2
1.2 Planteamiento del problema . . . . .	3
1.2.1 Descripción de la problemática . . . . .	3
1.2.2 Formulación del problema . . . . .	4
1.2.2.1 Pregunta general . . . . .	4
1.2.2.2 Preguntas específicas . . . . .	4
1.3 Objetivos de la investigación . . . . .	4
1.3.1 Objetivo general . . . . .	4
1.3.2 Objetivos específicos . . . . .	4
1.4 Importancia y alcance de la investigación . . . . .	5
<b>CAPÍTULO II</b>	<b>7</b>
<b>2. MARCO TEÓRICO</b>	<b>8</b>
2.1 Antecedentes de la investigación . . . . .	8
2.2 Bases teóricas . . . . .	10
2.2.1 Armonización de imágenes satelitales . . . . .	10
2.2.1.1 Landsat: satélites de observación terrestre . . . . .	10
2.2.1.2 Landsat MSS . . . . .	13
2.2.1.3 Landsat TM . . . . .	20
2.2.1.4 Armonización de imágenes en Teledetección . . . . .	22
2.2.1.5 Armonización geométrica de datos de Teledetección . . . . .	23

2.2.1.6	Armonización espectral de datos de Teledetección . . . . .	24
2.2.1.7	Normalización y ajuste de datos con aplicación de parámetros BRDF . . . . .	25
2.2.2	Inteligencia artificial . . . . .	26
2.2.2.1	Deep learning en el procesamiento de imágenes de Teledetección . . . . .	26
2.2.2.2	Redes neuronales convolucionales (CNN) . . . . .	26
2.2.2.3	Tipos de capas en una red neuronal convolucional . . . . .	34
2.3	Definición de términos . . . . .	39
2.3.1	Armonización de datos . . . . .	39
2.3.2	BRDF (Bidirectional Reflectance Distribution Function) . . . . .	39
2.3.3	CNN (Redes Neuronales Convolucionales) . . . . .	39
2.3.4	Co-registración . . . . .	40
2.3.5	Cubo de datos . . . . .	40
2.3.6	Deep learning . . . . .	40
2.3.7	Entornos conda . . . . .	40
2.3.8	ESRCNN (Extended Super-Resolution Convolutional Neural Network)	41
2.3.9	GAN (Generative Adversarial Network) . . . . .	41
2.3.10	Generación de banda virtual . . . . .	41
2.3.11	Histogram matching . . . . .	41
2.3.12	Landsat . . . . .	41
2.3.13	MSS (Multi-Spectral Scanner) . . . . .	42
2.3.14	NDVI . . . . .	42
2.3.15	Segmentación de imágenes . . . . .	42
2.3.16	Super-resolución . . . . .	42

2.3.17 TM (Thematic Mapper) . . . . .	42
2.3.18 Transformadas de Fourier . . . . .	42
<b>CAPÍTULO III</b>	<b>44</b>
<b>3. HIPÓTESIS Y VARIABLES</b>	<b>45</b>
3.1 Las hipótesis . . . . .	45
3.1.1 Hipótesis general . . . . .	45
3.1.2 Hipótesis específicas . . . . .	45
3.2 Las variables . . . . .	45
3.2.1 Variable independiente . . . . .	45
3.2.2 Variable dependiente . . . . .	45
3.3 Operacionalización de variables . . . . .	46
<b>CAPÍTULO IV</b>	<b>47</b>
<b>4. METODOLOGÍA</b>	<b>48</b>
4.1 Área de estudio (unidad de análisis) . . . . .	48
4.2 Diseño de investigación . . . . .	50
4.2.1 Tipo . . . . .	50
4.2.2 Nivel . . . . .	50
4.2.3 Enfoque . . . . .	51
4.2.4 Diseño . . . . .	51
4.3 Población y muestra . . . . .	51
4.3.1 Población . . . . .	51
4.3.2 Muestra . . . . .	51
4.4 Procedimiento, técnicas e instrumentos de recolección de información . . . . .	52

4.4.1	Obtención de datos de entrada . . . . .	52
4.4.1.1	Selección por localización . . . . .	52
4.4.1.2	Selección por tiempo y calidad . . . . .	54
4.4.2	Corrección geométrica de las imágenes MSS . . . . .	55
4.4.2.1	Lectura y preprocesamiento de imágenes . . . . .	55
4.4.2.2	Extracción y coincidencia de características . . . . .	55
4.4.2.3	Eliminación de errores calidad y la usabilidad de las MSS . .	56
4.4.2.4	Determinación del desplazamiento espacial . . . . .	57
4.4.2.5	Refinamiento subpíxel y convolución en 2D . . . . .	58
4.4.2.6	Filtro final de los pares de imágenes . . . . .	59
4.4.3	Preparación y organización de datos para deep learning . . . . .	60
4.4.3.1	Conversión de formato de datos . . . . .	60
4.4.3.2	Estandarización y homogeneización . . . . .	60
4.4.3.3	Integración con Hugging Face . . . . .	60
4.4.3.4	Segmentación y distribución de datos . . . . .	61
4.4.3.5	Procesamiento y alineación de imágenes . . . . .	61
4.4.3.6	Normalización y preparación de datos . . . . .	61
4.4.3.7	Optimización y manejo eficiente de datos . . . . .	61
4.4.4	Desarrollo del modelo MSS2TM . . . . .	63
4.4.4.1	Configuración del entorno de trabajo . . . . .	63
4.4.4.2	Elección de arquitecturas para el generador y discriminador en GANs . . . . .	63
4.4.4.3	Definición de parámetros y funciones de pérdida . . . . .	64
4.4.4.4	Proceso de entrenamiento . . . . .	65
4.4.4.5	Ajuste y optimización . . . . .	66
4.4.4.6	Configuración de las funciones de pérdida . . . . .	66
4.4.5	Generación de bandas virtuales en imágenes MSS . . . . .	68
4.4.5.1	Adquisición y procesamiento inicial de datos . . . . .	68

4.4.5.2	Entrenamiento del perceptrón multicapa (MLP) . . . . .	68
4.4.5.3	Proceso de entrenamiento y ajuste fino . . . . .	69
4.5	Análisis estadístico . . . . .	72
4.5.1	Validación cuantitativa . . . . .	72
4.5.1.1	Error Absoluto Medio (MAE) . . . . .	72
4.5.1.2	Raíz del Error Cuadrático Medio (RMSE) . . . . .	72
4.5.1.3	Coeficiente de Correlación de Pearson (R) . . . . .	72
4.5.2	Evaluación de métricas de rendimiento . . . . .	72
4.5.2.1	CLIP Loss . . . . .	72
4.5.2.2	L1 Loss . . . . .	72
4.5.2.3	LPIPS (Learned Perceptual Image Patch Similarity) Loss . .	73
<b>CAPÍTULO V</b>		<b>74</b>
<b>5. RESULTADOS</b>		<b>75</b>
5.1	Evaluación de la armonización de imágenes MSS y TM . . . . .	75
5.1.1	Corrección geométrica y alineación espacial . . . . .	75
5.1.2	Calidad de la armonización espectral y espacial . . . . .	76
5.1.3	Generación de bandas virtuales . . . . .	77
5.1.4	Evaluación de la armonización de imágenes . . . . .	79
5.1.4.1	Índices espectrales . . . . .	79
5.2	Análisis estadístico de los resultados . . . . .	84
5.2.1	Contexto y metodología de Validación . . . . .	84
5.2.2	Evaluación cuantitativa de la superresolución . . . . .	85
5.2.2.1	Interpretación de métricas: . . . . .	86
5.2.3	Ánálisis detallado por bandas espectrales . . . . .	86
5.2.4	Distribución de métricas y variabilidad del modelo . . . . .	87
5.2.5	Ánálisis de las bandas SWIR y térmica . . . . .	88

<b>CAPÍTULO VI</b>	<b>90</b>
<b>6. DISCUSIÓN</b>	<b>91</b>
6.1 Evaluación comparativa con estudios previos . . . . .	91
6.2 Variabilidad geográfica y rendimiento del modelo . . . . .	91
6.3 Implicaciones prácticas y limitaciones . . . . .	91
<b>7. CONCLUSIONES Y RECOMENDACIONES</b>	<b>93</b>
7.1 Conclusiones . . . . .	93
7.2 Recomendaciones . . . . .	93
<b>REFERENCIAS BIBLIOGRÁFICAS</b>	<b>95</b>
<b>ANEXOS</b>	<b>99</b>
Operacionalización de variables . . . . .	99
Matriz de consistencia . . . . .	100

# **CAPÍTULO I**

## **1. INTRODUCCIÓN**

### **1.1 Introducción**

La exploración de nuestro planeta desde el espacio ha revolucionado el entendimiento del medio ambiente y ha sido fundamental en la planificación y gestión de recursos naturales. Dentro de este marco, el programa Landsat ha desempeñado un rol vital, suministrando una secuencia continua de datos sobre las transformaciones terrestres desde la década de 1970. No obstante, la evolución tecnológica entre las distintas generaciones de sensores Landsat ha presentado obstáculos para la comparación homogénea de datos a través del tiempo.

La presente tesis, titulada Recuperación de imágenes Landsat MSS (1972-1999) mediante inteligencia artificial hacia una armonización efectiva del monitoreo global y a largo plazo”, enfrenta este desafío mediante la aplicación de avanzadas técnicas de inteligencia artificial para la estandarización de datos obtenidos por el sensor Multi-Spectral Scanner (MSS) de Landsat, abarcando desde 1972 hasta 1999. Este enfoque no solo mejora la capacidad de analizar cambios en la superficie terrestre a largo plazo, sino que también enriquece el valor histórico de los registros de Landsat, fundamentales para comprender las dinámicas de cambio en uso del suelo, cobertura vegetal y fenómenos geográficos a escala global.

En el capítulo uno, se introduce el contexto y la importancia de la observación terrestre a través de Landsat, subrayando las dificultades técnicas surgidas de las diferencias entre generaciones de sensores y cómo estas impactan en la comparabilidad de los datos. El segundo capítulo se centra en el marco teórico, delineando los fundamentos de la teledetección y la inteligencia artificial como herramientas para la armonización de datos satelitales. El capítulo tres detalla la metodología empleada, describiendo el diseño y la implementación de modelos de aprendizaje profundo destinados a ajustar las discrepancias entre los sensores MSS y TM, permitiendo así una integración efectiva de los datos históricos con registros más recientes.

El cuarto capítulo expone los resultados obtenidos, evidenciando la efectividad de los modelos de aprendizaje profundo en la armonización de imágenes MSS con las de TM, y destaca cómo este enfoque mejora significativamente la utilidad de los datos de Landsat para análisis a

largo plazo. Por último, el capítulo cinco reflexiona sobre las implicaciones de estos hallazgos, tanto para la comunidad científica como para la práctica en campos relacionados con la geografía y la ecología, resaltando el aporte de la investigación al avance en la ingeniería geográfica y subrayando el valor incalculable de los registros Landsat como crónica de la historia ambiental del planeta.

## 1.2 Planteamiento del problema

### 1.2.1 Descripción de la problemática

La serie de datos Landsat ha sido una herramienta esencial para la investigación de la superficie terrestre debido a su larga serie temporal. Sin embargo, la armonización de imágenes satelitales de diferentes sensores Landsat, como MSS y TM, representa un desafío en la comunidad de teledetección. Las discrepancias se deben, en gran medida, a las diferencias en las bandas espectrales, resoluciones espaciales y la calidad inherente de las imágenes.

Aunque se proporcionan productos de Reflectancia de la Superficie Terrestre (LSR) para Landsat TM, ETM+ y OLI, las primeras imágenes MSS de Landsat 1-5 carecen de dicho producto debido a desafíos como la calidad de imagen inferior y diferencias en las bandas (Zhao et al., 2022). Las imágenes MSS, adquiridas desde 1972, carecen de ciertas bandas espectrales que están presentes en las imágenes TM, complicando aún más su integración multiespectral.

Los métodos basados en regresiones han sido propuestos para mapear las bandas entre sensores Landsat, facilitando la integración de datos de múltiples plataformas satelitales (Roy, Kovalskyy et al., 2016). Sin embargo, estos enfoques a menudo requieren múltiples pasos de preprocesamiento y pueden no ser óptimos para todas las aplicaciones.

Por su parte, algoritmos de deep learning, como U-Net (Ronneberger et al., 2015), se han presentado como herramientas poderosas para el procesamiento y análisis de imágenes satelitales. Una red neuronal convolucional de super-resolución extendida (ESRCNN) fue desarrollada para fusionar datos de Landsat-8 y Sentinel-2, demostrando su eficacia en la producción de conjuntos de datos coherentes (Shao et al., 2019). Sin embargo, la aplicación de estas técnicas para la armonización de imágenes satelitales aún está en sus etapas iniciales.

Dentro de este contexto, surge la necesidad de explorar cómo la inteligencia artificial,

especialmente el aprendizaje profundo, puede desempeñar un papel crucial en la armonización efectiva de las imágenes Landsat MSS, haciendo viable su uso en monitoreos globales de largo plazo. Esta investigación busca abordar precisamente este desafío, proporcionando soluciones innovadoras para una armonización efectiva de las imágenes MSS y TM.

### ***1.2.2 Formulación del problema***

#### **1.2.2.1 Pregunta general.**

- ¿Cómo armonizar las imágenes Landsat MSS para su uso en el monitoreo global y a largo plazo, utilizando inteligencia artificial?

#### **1.2.2.2 Preguntas específicas.**

- ¿Cómo integrar el aprendizaje profundo y procesamiento de imágenes en la corrección geométrica de las imágenes Landsat MSS y alinearlas con las TM a nivel de pixel y sub-pixel?
- ¿De qué manera puede el modelo de aprendizaje profundo MSS2TM alinear espectral y espacialmente las imágenes Landsat MSS y TM?
- ¿Mediante qué técnica de aprendizaje profundo se puede generar bandas faltantes en imágenes MSS que existen en las TM?

## **1.3 Objetivos de la investigación**

### ***1.3.1 Objetivo general***

- Armonizar las imágenes Landsat MSS utilizando inteligencia artificial para su uso en el monitoreo global y a largo plazo.

### ***1.3.2 Objetivos específicos***

- Integrar técnicas de aprendizaje profundo y procesamiento de imágenes para la corrección geométrica de las imágenes Landsat MSS, buscando una alineación precisa con las imágenes TM a niveles de pixel y sub-pixel.

- Desarrollar el modelo de aprendizaje profundo MSS2TM para alinear espectral y espacialmente las imágenes Landsat MSS y TM.
- Implementar una técnica de aprendizaje profundo para generar bandas ausentes en imágenes MSS que existen en las TM.

#### **1.4 Importancia y alcance de la investigación**

El programa Landsat, a lo largo de sus décadas de existencia, ha proporcionado datos esenciales que han informado y guiado la investigación sobre la superficie terrestre. La armonización de imágenes satelitales, especialmente las de Landsat MSS (1972- 1999), es esencial para garantizar la continuidad y coherencia de las series temporales de datos satelitales. Estas imágenes, aunque valiosas por su extenso registro temporal, presentan desafíos inherentes debido a las diferencias en las bandas espectrales y resoluciones espaciales en comparación con sensores más modernos. La falta de coherencia espectral y temporal puede llevar a interpretaciones erróneas, especialmente cuando se intenta combinar o comparar datos de diferentes series de Landsat para análisis multitemporales.

La inteligencia artificial, particularmente a través del aprendizaje profundo, se presenta como una solución potente para estos desafíos. Con técnicas avanzadas, es posible abordar las diferencias espectrales y geométricas, logrando una integración más exacta y, consecuentemente, análisis multitemporales más fidedignos. La eficacia de estas técnicas es resaltada por la eficiencia de arquitecturas como U-Net y modelos de armonización como ESRCNN. Sin embargo, aún queda por determinar la capacidad integral de estas y otras técnicas al enfrentar el desafío de alinear imágenes de diferentes sensores satelitales.

La importancia de esta investigación radica en su potencial para llenar un vacío existente en la teledetección. Al optimizar la calidad y compatibilidad de las imágenes Landsat MSS, esta investigación facilitará la utilización de datos que abarcan desde 1972 a 1999 combinando imágenes de sensores Landsat MSS y TM, enriqueciendo el alcance temporal del monitoreo global. Esto es esencial para entender y rastrear cambios a largo plazo en la superficie terrestre, lo que tiene implicaciones en áreas tan diversas como el cambio climático, la conservación ambiental y el desarrollo urbano.

Además, la metodología de armonización desarrollada en este estudio puede ser adaptada y utilizada para otros sensores y tecnologías emergentes en el campo de la teledetección. Esto incluye sensores con resoluciones espaciales más finas, que son cada vez más comunes en la observación terrestre. Al establecer un marco sólido para la armonización de datos, esta investigación permitirá mejorar la calidad y coherencia de los datos satelitales, promoviendo así estudios más complejos y detallados en diversas aplicaciones prácticas. La planificación urbana, la gestión de recursos naturales, la mitigación de desastres ambientales y muchos otros campos podrán beneficiarse directamente de los datos armonizados y las técnicas desarrolladas en este estudio. De esta manera, se sentarán las bases para futuras investigaciones y aplicaciones prácticas en teledetección y análisis ambiental, permitiendo que estas metodologías se apliquen también a sensores de mayor resolución espacial, facilitando el análisis de imágenes con mayor detalle y precisión.

## **CAPÍTULO II**

## **2. MARCO TEÓRICO**

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

La armonización de imágenes de diferentes sensores es esencial para el monitoreo continuo y preciso de la superficie terrestre. En este contexto, un estudio en el 2019 abordó este desafío mediante la implementación de un marco de fusión basado en deep learning para mejorar la resolución espacial de las imágenes Landsat-8 utilizando imágenes Sentinel-2 como auxiliares. Los resultados cuantitativos del estudio mostraron que este enfoque, denominado ESRCNN, superó al algoritmo de referencia, el area-to-point regression kriging (ATPRK), en términos de precisión. En particular, para las imágenes del 15 de junio de 2017, el RMSE para la banda 1 fue de 0.0243, el coeficiente de correlación (CC) alcanzó valores de hasta 0.9860 y el Universal Image Quality Index (UIQI) registró valores de hasta 0.9710. Estos indicadores reflejan la alta eficacia del método propuesto en la preservación de la distribución de reflectancia de la imagen original y en la mejora de la resolución espacial. Estos hallazgos subrayan el potencial del deep learning en la armonización y recuperación de imágenes satelitales, lo que es relevante para investigaciones que buscan mejorar la calidad y consistencia de las series temporales de imágenes (Shao et al., 2019).

Posteriormente, se llevó a cabo un estudio para entender mejor la relación entre las longitudes de onda reflectivas de Landsat-7 ETM+ y Landsat-8 OLI, y también para analizar la consistencia del Índice de Vegetación de Diferencia Normalizada (NDVI). Al analizar las trayectorias coincidentes, se examinaron 59 millones de observaciones de sensores de 30 m, obtenidas de 6317 imágenes. Los resultados mostraron que, en promedio, la reflectancia TOA de OLI supera a la de ETM+ en todas las bandas. Asimismo, en áreas con vegetación y terrenos, el NDVI de OLI tiende a ser superior al de ETM+. Para garantizar una transición fluida entre los datos de ambos, se introdujeron métodos estadísticos de transformación que mostraron un ajuste con coeficientes de determinación  $r^2$  de más de 0.7 para reflectancia y más de 0.9 para NDVI, con significancias estadísticas menores a 0.0001 (Roy, Kovalskyy et al., 2016).

Adicionalmente, en un estudio se aplicó el modelo 'ArithFusion' basado en deep learning

para fusionar imágenes de teledetección temporales multimodales. Esta investigación se centró en combinar imágenes de diferentes resoluciones y períodos de revisita, un aspecto crucial para la armonización efectiva de imágenes satelitales. El modelo propuesto, UMSEh, demostró ser eficaz, logrando un RMSE de 2.06 (0.57) en experimentos con escala 4x, lo que indica una alta precisión en la fusión de imágenes. Este estudio resalta el potencial del deep learning, como la arquitectura U-Net, en la armonización de imágenes satelitales, proporcionando una base sólida para futuras investigaciones en este campo (Hoque et al., 2022).

Asimismo, una investigación llevada a cabo por Saunier et al. (2022) exploró la armonización y fusión de datos de sensores satelitales Landsat 8, Landsat 9 y Sentinel-2 utilizando técnicas avanzadas de inteligencia artificial. Mediante el software S2L, desarrollado en Python 3.7 e incorporando librerías como Numpy, Pandas y Scikit-Learn, se generaron conjuntos de datos armonizados (L2H) y fusionados (L2F). Estos conjuntos, tras meticulosos ajustes espectrales, se presentaron en mosaicos de 110 km × 110 km utilizando el Sistema de Referencia UTM-MGRS. Se aplicaron correcciones geométricas, apoyadas en herramientas como OpenCV, y se implementaron correcciones atmosféricas con la ayuda de SMAC y Sen2Cor 3.0. Uno de los avances más significativos fue la corrección BRDF, que logró reducir coeficientes hasta un 0.01°. Además, se puso de manifiesto la variabilidad en respuestas espectrales entre diferentes instrumentos, llevando a la introducción del Factor de Ajuste de Banda Espectral (SBAF). La innovación de este estudio sirve como pilar para futuros esfuerzos enfocados en la integración y optimización de imágenes de diferentes sensores, estableciendo un precedente clave para trabajos que buscan la armonización en el monitoreo satelital.

Recientemente, un avance importante se destacó en el campo de la super-resolución para teledetección, donde las redes generativas adversarias (GANs) han mostrado un potencial significativo en la producción de imágenes de super-resolución de alta fidelidad. La introducción de CLIPSCORE, una nueva métrica, ha mejorado notablemente la evaluación de las imágenes super-resueltas alineándose estrechamente con el juicio humano. Utilizando CLIPSCORE, las GANs demostraron ser superiores en comparación con métodos basados en pérdida L2 y mostraron una precisión semántica más alta que los modelos de difusión modernos. Este descubrimiento, apoyado en el uso del conjunto de datos S2-NAIP que abarca 113,000 km<sup>2</sup> con

imágenes de Sentinel-2 y NAIP, facilita una investigación más profunda en este campo. La capacidad de las GANs para acelerar el entrenamiento hasta 18 veces, al tiempo que mejora la calidad de los resultados, subraya un avance crucial hacia la armonización eficaz y el monitoreo global, marcando un hito significativo en la aplicación de técnicas de inteligencia artificial en la teledetección (Wolters et al., 2023)

## 2.2 Bases teóricas

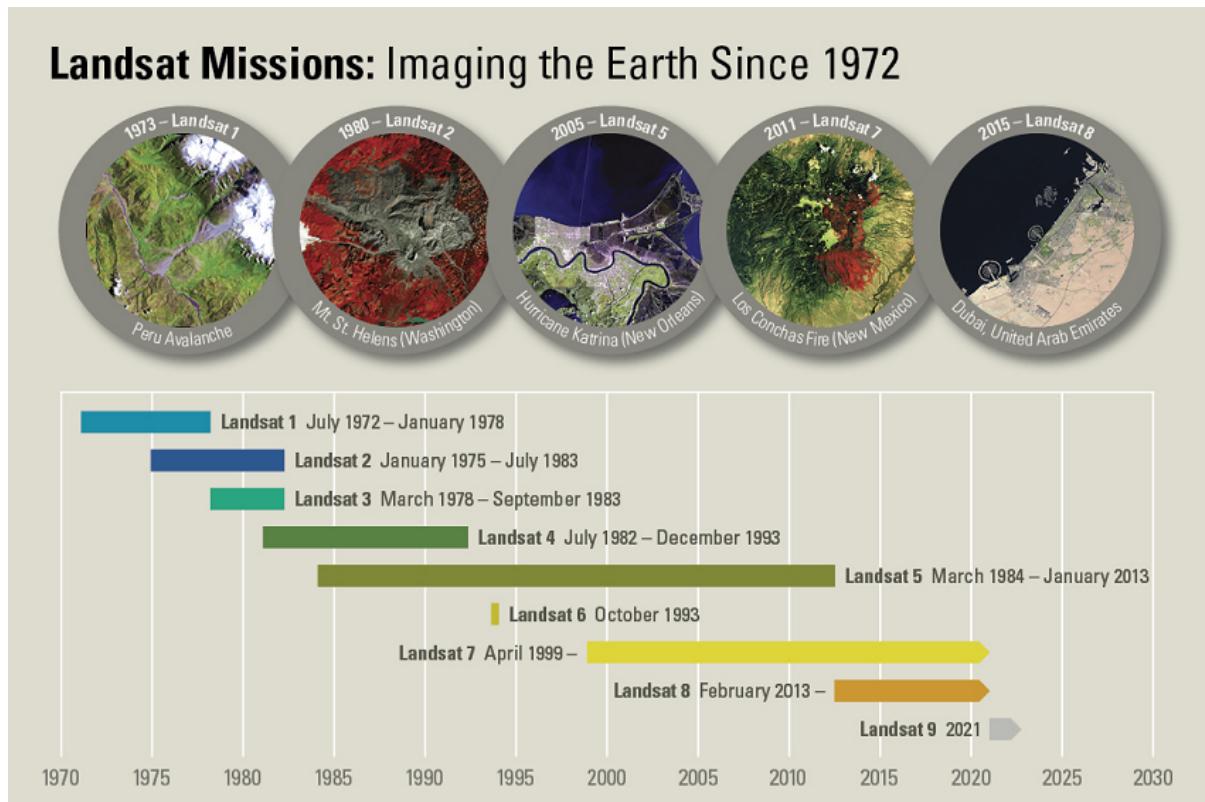
### 2.2.1 Armonización de imágenes satelitales

**2.2.1.1 Landsat: satélites de observación terrestre.** Desde 1972, los satélites Landsat han adquirido de manera continua imágenes basadas en el espacio de la superficie terrestre de la Tierra. Estos datos son esenciales para investigaciones relacionadas con el uso y cambio del suelo, con aplicaciones en áreas como la silvicultura, agricultura, geología, planificación regional y educación.

**Historia y misión de Landsat.** Landsat es una colaboración entre el U.S. Geological Survey (USGS) y la National Aeronautics and Space Administration (NASA). Mientras NASA se encarga del desarrollo de los instrumentos y satélites, así como de su lanzamiento y validación, el USGS toma control sobre la operación de los satélites, la recepción de datos, archivo, generación de productos y distribución (USGS, 2023a).

**Figura 1.**

*Línea de tiempo e historia de las misiones Landsat, iniciadas en 1972.*



*Nota.* La línea de tiempo destaca que, desde su inicio en 1972, las misiones Landsat han brindado una secuencia constante de datos que revelan las transformaciones de la Tierra, apoyando así diversas investigaciones y evidenciando su desarrollo en la vigilancia ambiental global. Tomado de USGS (2023a).

La idea de crear un satélite de observación civil surgió en la década de 1960, lográndose el 23 de julio de 1972 con el lanzamiento de ERTS-1, luego renombrado Landsat 1. A este le siguieron varias versiones, siendo Landsat 5 el que estableció un récord por ser el satélite de observación terrestre en funcionamiento más longevo.

**Adquisiciones satelitales.** Tanto Landsat 8 como Landsat 9 orbitan la Tierra a una altitud de 705 km, completando una órbita en 99 minutos y pasando por cualquier punto de la Tierra cada 16 días. Entre los dos satélites, se agregan más de 1,500 escenas al archivo USGS diariamente. Landsats 4, 5 y 7 siguieron la misma órbita, mientras que Landsats 1, 2 y 3 tenían

una altitud de 920 km (USGS, 2023a).

**Sensores y designaciones de bandas.** Los sensores han variado a lo largo de las versiones de Landsat. Desde el Multispectral Scanner (MSS) en los primeros Landsats hasta el Operational Land Imager (OLI) y el Thermal Infrared Sensor (TIRS) en los Landsats más recientes.

### Tabla 1.

*Comparación y visualización de bandas y longitudes de onda de sensores Landsat mediante Spectral Viewer del U.S. Geological Survey.*

Designaciones de banda	L8-9 OLI/TIRS		L7 ETM+		L4-5 TM		L4-5 MSS*		L1-3 MSS*	
	B	Longitud	B	Longitud	B	Longitud	B	Longitud	B	Longitud
Costera/Aerosol	1	0.43–0.45	–	–	–	–	–	–	–	–
Azul	2	0.45–0.51	1	0.45–0.52	1	0.45–0.52	–	–	–	–
Verde	3	0.53–0.59	2	0.52–0.60	2	0.52–0.60	1	0.5–0.6	4	0.5–0.6
Pancromática**	8	0.50–0.68	8	0.52–0.90	–	–	–	–	–	–
Rojo	4	0.64–0.67	3	0.63–0.69	3	0.63–0.69	2	0.6–0.7	5	0.6–0.7
Infrarrojo cercano	5	0.85–0.88	4	0.77–0.90	4	0.76–0.90	3	0.7–0.8	6	0.7–0.8
Infrarrojo cercano	–	–	–	–	–	–	4	0.8–1.1	7	0.8–1.1
Cirrus	9	1.36–1.38	–	–	–	–	–	–	–	–
Infrarrojo corto-1	6	1.57–1.65	5	1.55–1.75	5	1.55–1.75	–	–	–	–
Infrarrojo corto-2	7	2.11–2.29	7	2.09–2.35	7	2.08–2.35	–	–	–	–
Térmico	10 T1	10.60–11.19	6 T2	10.40–12.50	6 T2	10.40–12.50	–	–	–	–
Térmico	11 T1	11.50–12.51	–	–	–	–	–	–	–	–

*Nota.* Hay observaciones que se debe tener en cuenta. Adapatado de USGS (2023b).

\* Adquirido a 80 metros, remuestreado a 60 metros.

\*\* 15 metros (pancromático).

T1 = Térmico (adquirido a 100 metros, remuestreado a 30 metros).

T2 = Térmico (adquirido a 120 metros, remuestreado a 30 metros).

**Tabla 2.***Comparación de sensores Landsat*

Banda	L8–9 OLI/TIRS	L7 ETM+	L4–5 TM	L4–5 MSS	L1–3 MSS	Uso
Costero/Aerosol	Banda 1	–	–	–	–	Observaciones costeras y detección de aerosoles.
Azul	Banda 2	Banda 1	Banda 1	–	–	Mapeo batimétrico y discriminación suelo/vegetación.
Verde	Banda 3	Banda 2	Banda 2	Banda 1	Banda 4	Evaluación de la vegetación.
Rojo	Banda 4	Banda 3	Banda 3	Banda 2	Banda 5	Identificación de vegetación y suelos.
Infrarrojo Cercano	Banda 5	Banda 4	Banda 4	Banda 3	Banda 6	Análisis y detección de vegetación.
	–	–	–	Banda 4	Banda 7	
Infrarrojo Onda Corta-1	Banda 6	Banda 5	Banda 5	–	–	Análisis de humedad y detección de incendios.
Infrarrojo Onda Corta-2	Banda 7	Banda 7	Banda 7	–	–	Detección de incendios y análisis de humedad.
Panchromático	Banda 8	Banda 8	–	–	–	Mejora de resolución en imágenes multiespectrales.
Cirrus	Banda 9	–	–	–	–	Detección de nubes cirrus.
Térmico	Banda 10	Banda 6	Banda 6	–	–	Mapeo de temperatura y estimación de humedad.

*Nota.* Se observa una tendencia en el uso de diferentes bandas para monitorear vegetación, analizar humedad, clima y mejorar la resolución de imágenes. Adaptado de USGS (2023b).

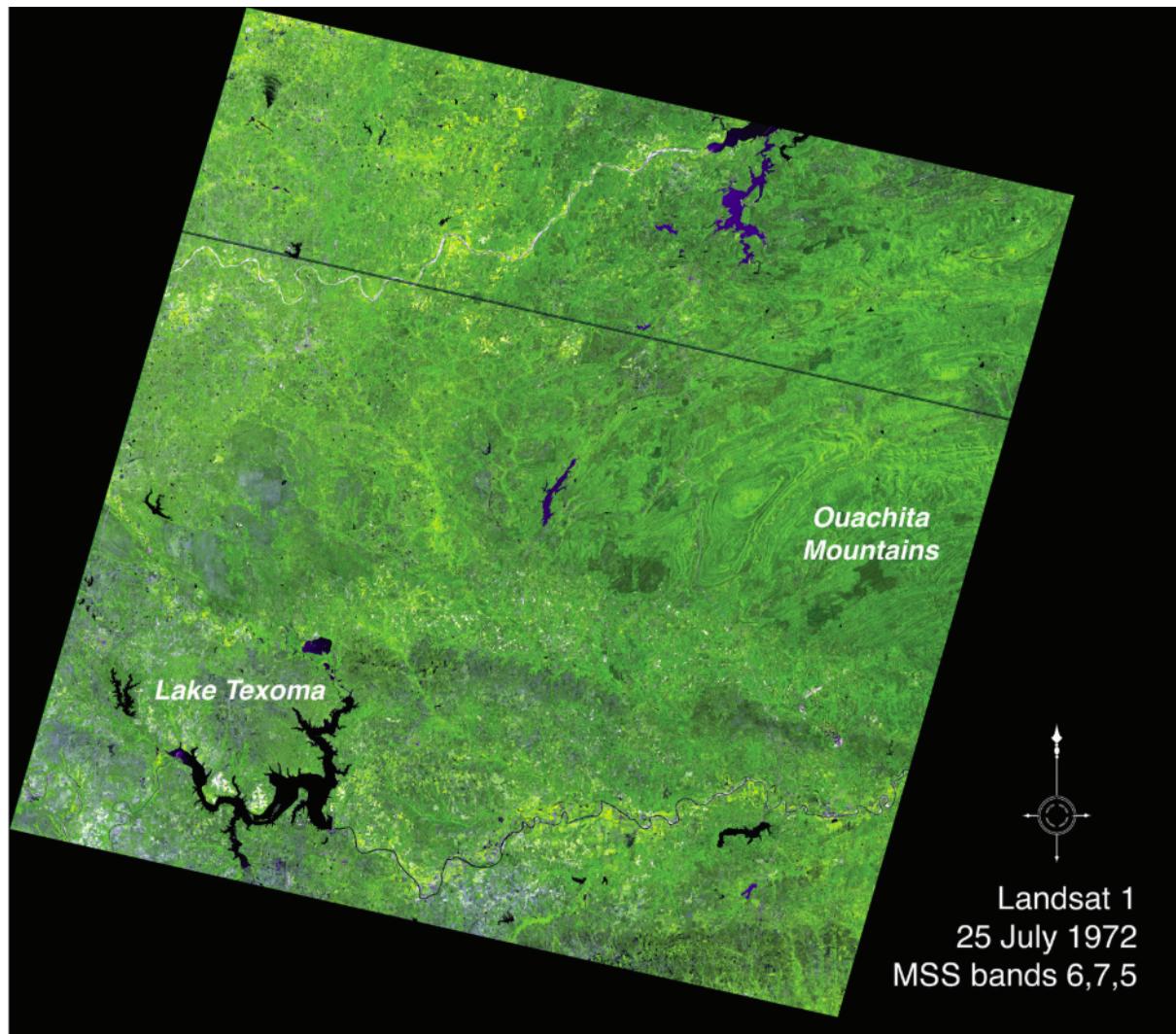
**Aplicaciones de los datos Landsat.** Los datos de Landsat respaldan una amplia gama de aplicaciones, desde investigación del cambio global hasta seguimiento de derrames de petróleo y monitoreo de contaminación por residuos mineros.

**2.2.1.2 Landsat MSS.** El Multispectral Scanner (MSS) fue el primer sensor de Landsat, utilizado en las misiones Landsat 1 a 5. Este sensor capturaba imágenes en cuatro bandas espectrales, con una resolución espacial de 79 metros para las bandas 1 a 4 y de 185 metros para la banda 5. Aunque el MSS fue reemplazado por el Thematic Mapper (TM) en Landsat 4, 5 y 7, sus datos siguen siendo valiosos para estudios de largo plazo y análisis de series temporales (Council, 2013).

**Landsat 1 (ERTS-1) y su MSS.** Originalmente denominado Earth Resources Technology Satellite (ERTS-1), fue lanzado el 23 de julio de 1972. El satélite llevaba a bordo el MSS, un sistema de escaneo multiespectral que capturaba imágenes en cuatro bandas espectrales (verde, rojo, y dos infrarrojas cercanas) con una resolución espacial de 80 metros . Esta tecnología permitía detectar y analizar cambios en el uso del suelo y la vegetación con mayor detalle que los sistemas anteriores. Durante su operación, Landsat 1 experimentó problemas técnicos, como la falla de una de sus grabadoras de cinta de video en agosto de 1972, seguida por la segunda en 1974, lo que limitó la capacidad de grabar datos de manera continua . A pesar de estos desafíos, el MSS proporcionó imágenes claras y precisas que demostraron ser esenciales para una variedad de aplicaciones científicas y de gestión de recursos.

**Figura 2.**

Primera imagen terrestre sin nubes adquirida por el Landsat 1 MSS, 25 de julio de 1972.

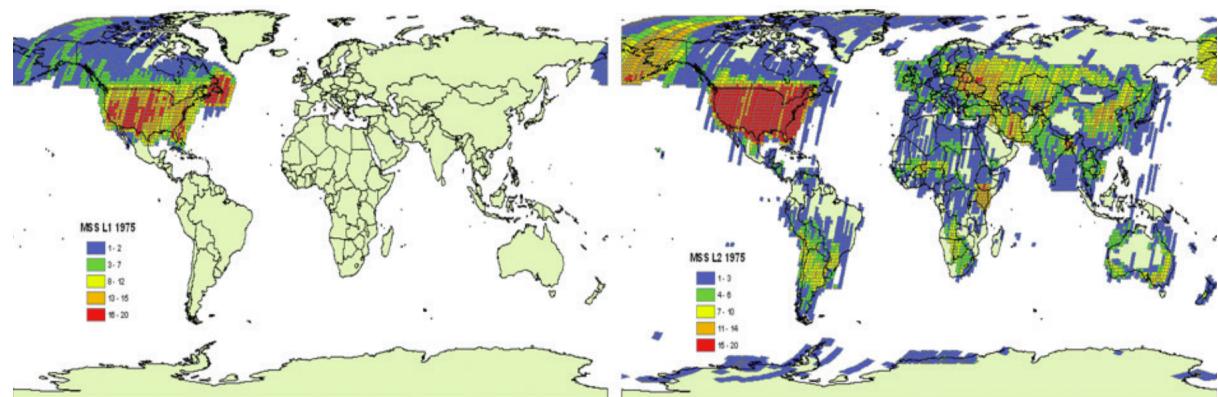


*Nota.* Imagen sin nubes del sistema MSS de Landsat 1 mostrando las Montañas Ouachita en el sureste de Oklahoma. Tomado de Council (2013).

**Landsat 2.** Lanzado el 22 de enero de 1975, fue diseñado de manera similar a Landsat 1 y continuó la misión de su predecesor con mejoras en la recopilación y manejo de datos . Landsat 2 también estaba equipado con el MSS, que siguió siendo el instrumento principal para la recolección de datos. A lo largo de su misión, Landsat 2 recopiló una vasta cantidad de datos MSS, aunque enfrentó problemas similares con sus grabadoras de cinta, lo que eventualmente afectó su capacidad operativa.

**Figura 3.**

*Fondos de escenas de USGS EROS MSS archivados en 1975 por Landsat 1 y Landsat 2.*

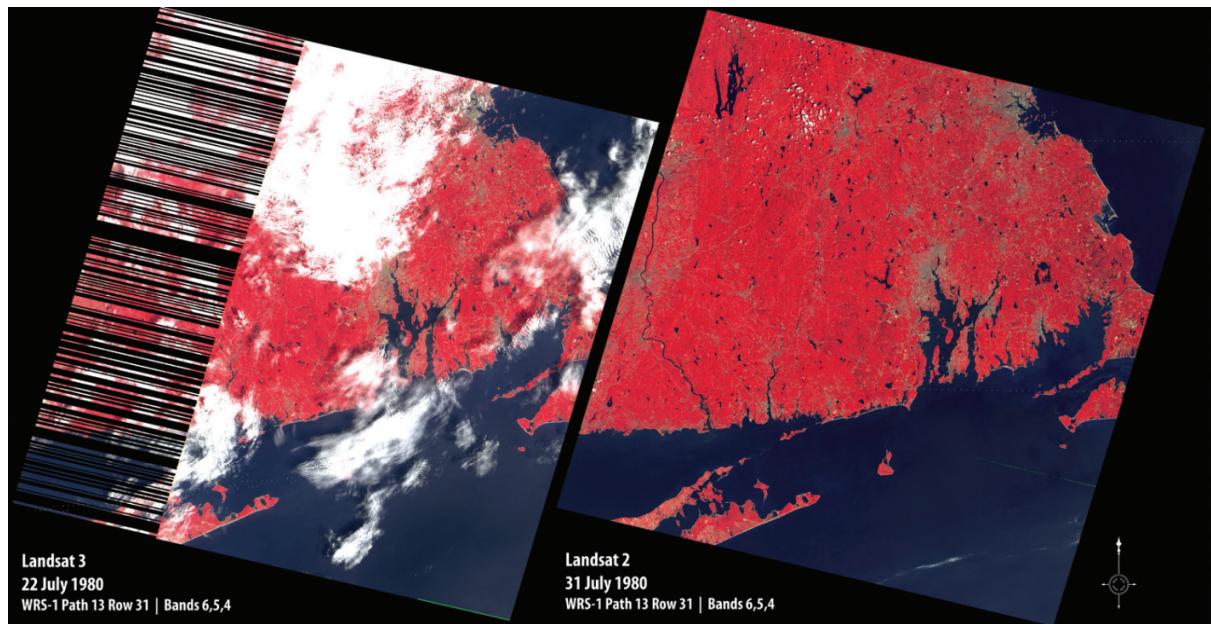


*Nota.* Comparación de las escenas MSS archivadas de Landsat 1 y 2 en 1975, mostrando la cantidad y distribución de datos recopilados. Tomado de Council (2013).

**Landsat 3.** Lanzado el 5 de marzo de 1978, introdujo algunas mejoras en el MSS, incluyendo una banda infrarroja térmica para la recolección de datos nocturnos . Sin embargo, el satélite enfrentó varios problemas técnicos que afectaron su rendimiento. La calidad de los componentes del MSS era inferior a la de los satélites anteriores, lo que resultó en fallas recurrentes y una disminución en la calidad de los datos . A pesar de estos problemas, Landsat 3 contribuyó con valiosos datos MSS hasta su retiro en 1983.

**Figura 4.**

*Inicio de los problemas de inicio de línea Landsat 3 MSS.*

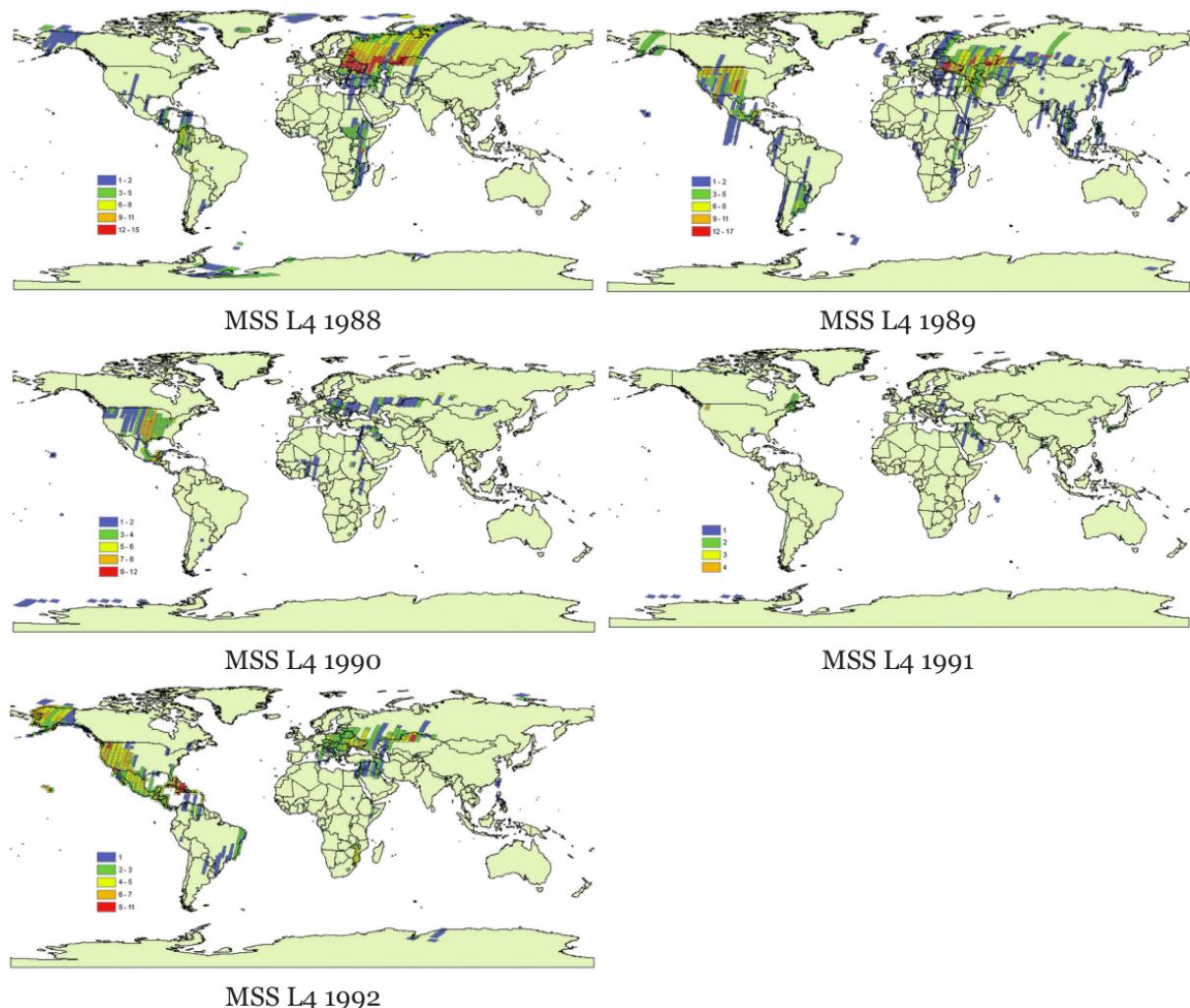


*Nota.* Imagen que muestra los problemas de inicio de línea del MSS de Landsat 3 en comparación con una imagen MSS de Landsat 2. Tomado de Council (2013).

**Landsat 4.** Lanzado el 16 de julio de 1982, marcó una evolución significativa con la introducción del Thematic Mapper (TM), aunque el MSS continuó operando como un sensor secundario . Landsat 4 enfrentó desafíos técnicos con su sistema de transmisión de datos, lo que limitó inicialmente la cantidad de datos MSS que pudo recolectar. No obstante, cuando la capacidad de transmisión se stabilizó, Landsat 4 proporcionó una cobertura casi completa a través de TDRS (Tracking and Data Relay Satellite) hasta 1987.

**Figura 5.**

*Mapas de cobertura anual de Landsat 4 MSS (1982-1992).*

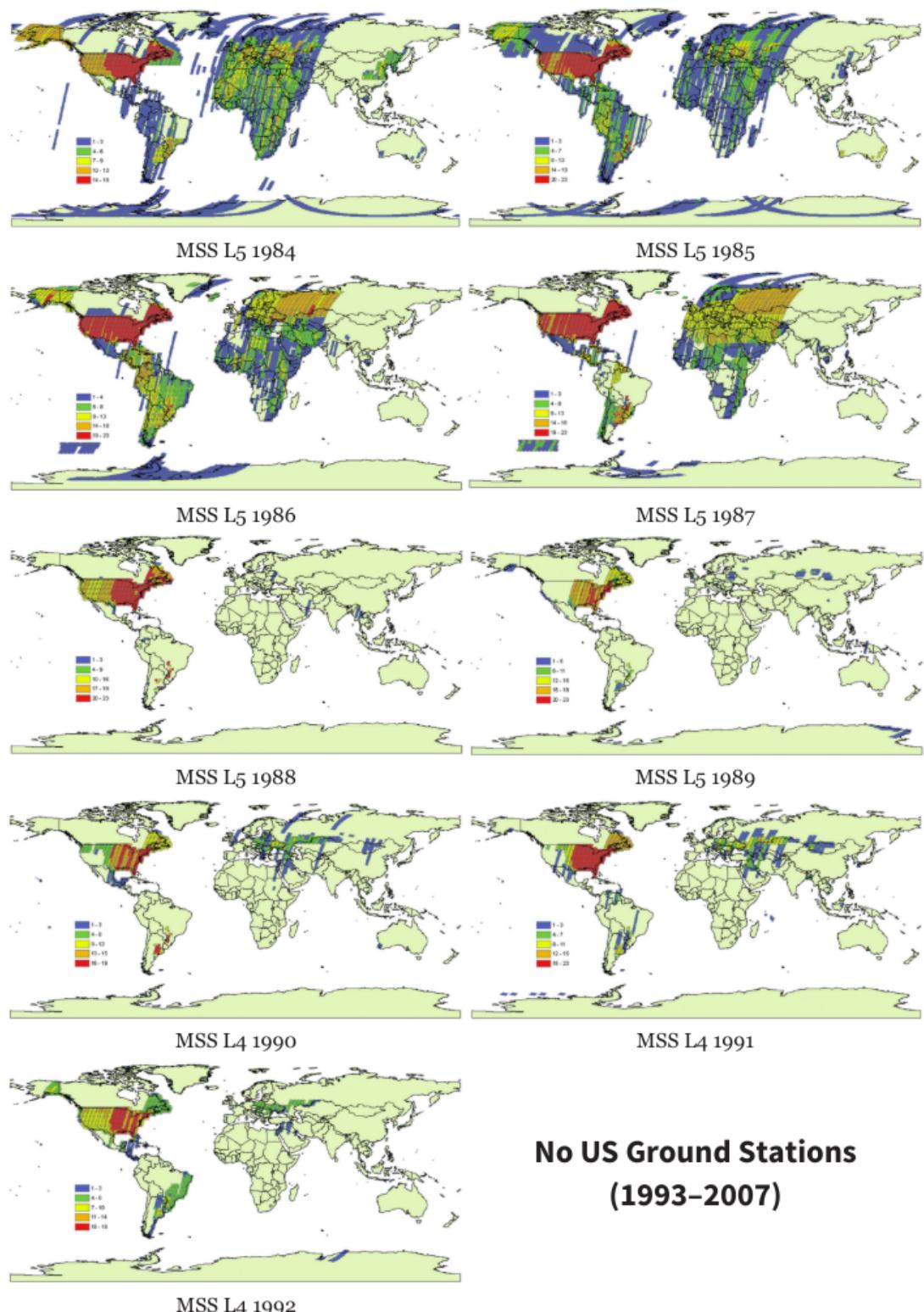


*Nota.* Mapas de cobertura anual del MSS de Landsat 4, mostrando la cobertura geográfica durante su operación. Tomado de Council (2013).

**Landsat 5.** Lanzado el 1 de marzo de 1984, es conocido por su longevidad, operando mucho más allá de su vida útil diseñada . Al igual que Landsat 4, Landsat 5 estaba equipado tanto con el TM como con el MSS. Aunque el MSS dejó de ser el sensor principal, continuó recolectando datos valiosos hasta mediados de la década de 1990. Landsat 5 es reconocido por el Guinness World Records como el satélite de observación de la Tierra más duradero en la historia.

**Figura 6.**

*Mapas de cobertura anual de Landsat 5 MSS (1984-2007).*



*Nota.* Mapas de cobertura anual del MSS de Landsat 5, mostrando la cobertura geográfica durante su operación. Tomado de Council (2013).

**Fechas del Landsat MSS.** Las fechas operativas de los sensores MSS de Landsat pueden parecer confusas debido a las múltiples fases de operación y reactivación. El sensor MSS del Landsat cubrió el período desde 1972 hasta 1999, reflejando el uso operativo en los satélites Landsat 1 hasta Landsat 5. En el caso del Landsat 5 MSS, estuvo operativo hasta 2012 porque, aunque dejó de ser la carga principal a finales de los años 90, se reactivó temporalmente en 2012 para llenar brechas de datos antes de la desactivación final del satélite. La reactivación en 2012 se debió a la suspensión de la recopilación de datos TM a finales de 2011, lo que llevó a la reanudación rutinaria de la adquisición MSS hasta diciembre de 2012.

**2.2.1.3 Landsat TM.** El Landsat Thematic Mapper (TM) es un sensor clave en la historia de la observación de la Tierra, destacándose por sus mejoras significativas en resolución espacial y espectral en comparación con sus predecesores. Diseñado para captar imágenes multiespectrales, el TM está equipado con bandas espectrales que cubren desde el azul hasta el infrarrojo térmico, permitiendo una amplia gama de aplicaciones en monitoreo ambiental, gestión de recursos naturales y estudios científicos (Council, 2013).

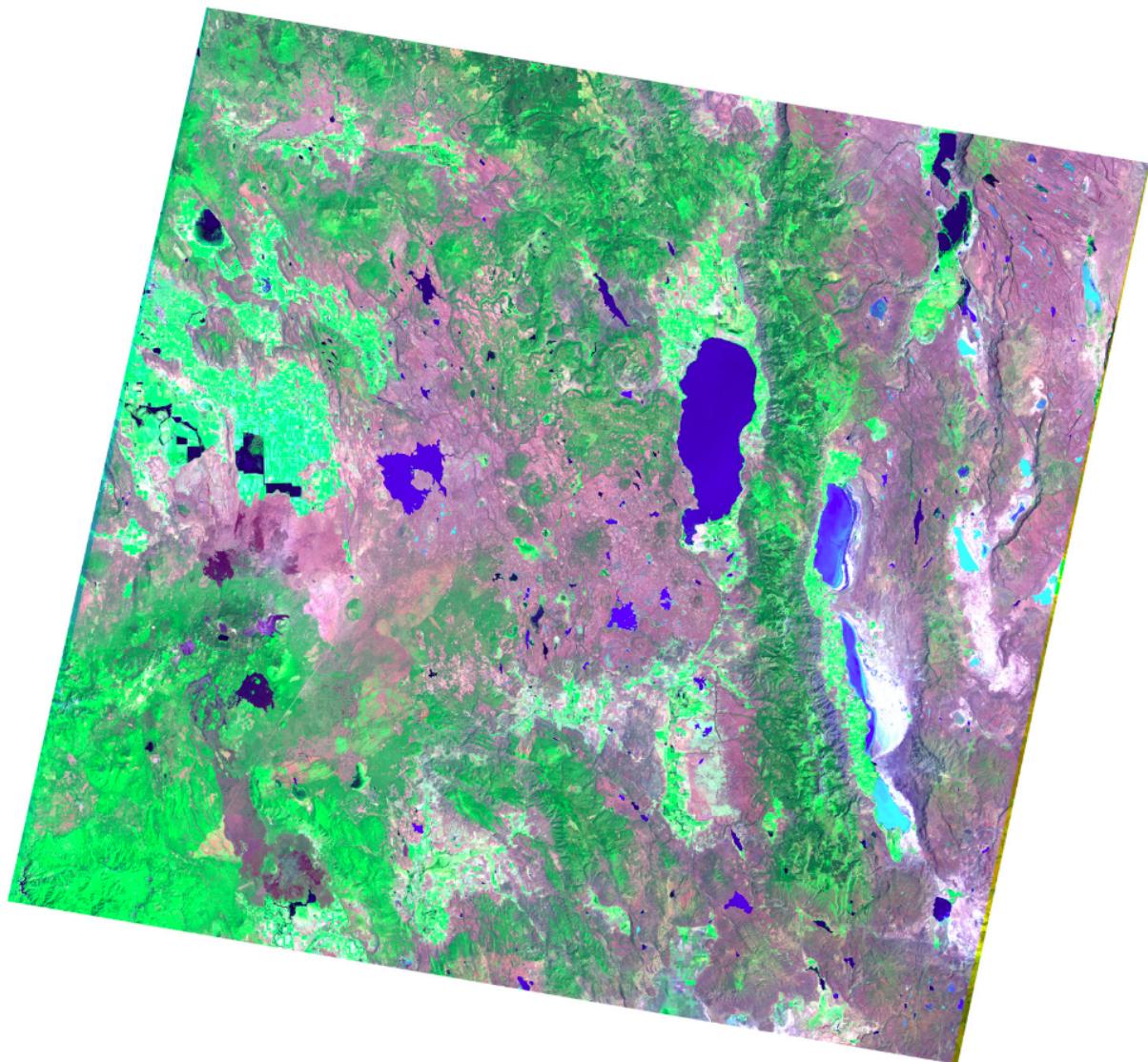
**Landsat 4.** Fue lanzado el 16 de julio de 1982, representando un avance significativo en la teledetección debido a la inclusión del sensor Thematic Mapper (TM). A pesar de sus innovaciones, Landsat 4 enfrentó varios desafíos operacionales desde el inicio, incluyendo fallos en los transmisores de banda X, lo que limitó la capacidad de transmisión de datos del TM hasta que se estableció una conexión a través del sistema de satélites de retransmisión de datos (TDRSS).

- **Problemas de comunicación:** Poco después del lanzamiento, Landsat 4 experimentó fallos en sus transmisores de banda X, dificultando la transmisión de imágenes TM. El lanzamiento del satélite TDRS-E permitió eventualmente la transmisión de datos del TM, aunque con cobertura limitada.

- **Fallos en paneles solares:** En 1983, Landsat 4 perdió el 50 % de su capacidad de generación de energía debido a fallos en los paneles solares. Este problema limitó las operaciones del satélite, afectando su capacidad de recopilación de datos.

**Figura 7.**

*Imagen de prueba del Thematic Mapper descargada vía TDRS-E el 12 de agosto de 1983.*



*Nota.* Se muestra un área alrededor de la frontera California/Oregón, incluyendo Goose Lake, Clear Lake National Refuge, Lava Beds National Monument y Modoc National Forest (TM Bandas 7, 4, 2). Tomado de Council (2013).

**Landsat 5.** Fue lanzado el 1 de marzo de 1984, y se destacó por su longevidad operativa, superando significativamente su vida útil de diseño inicial de tres años. Este satélite continuó recopilando datos hasta 2013, estableciendo un récord como la misión satelital de observación de la Tierra más duradera.

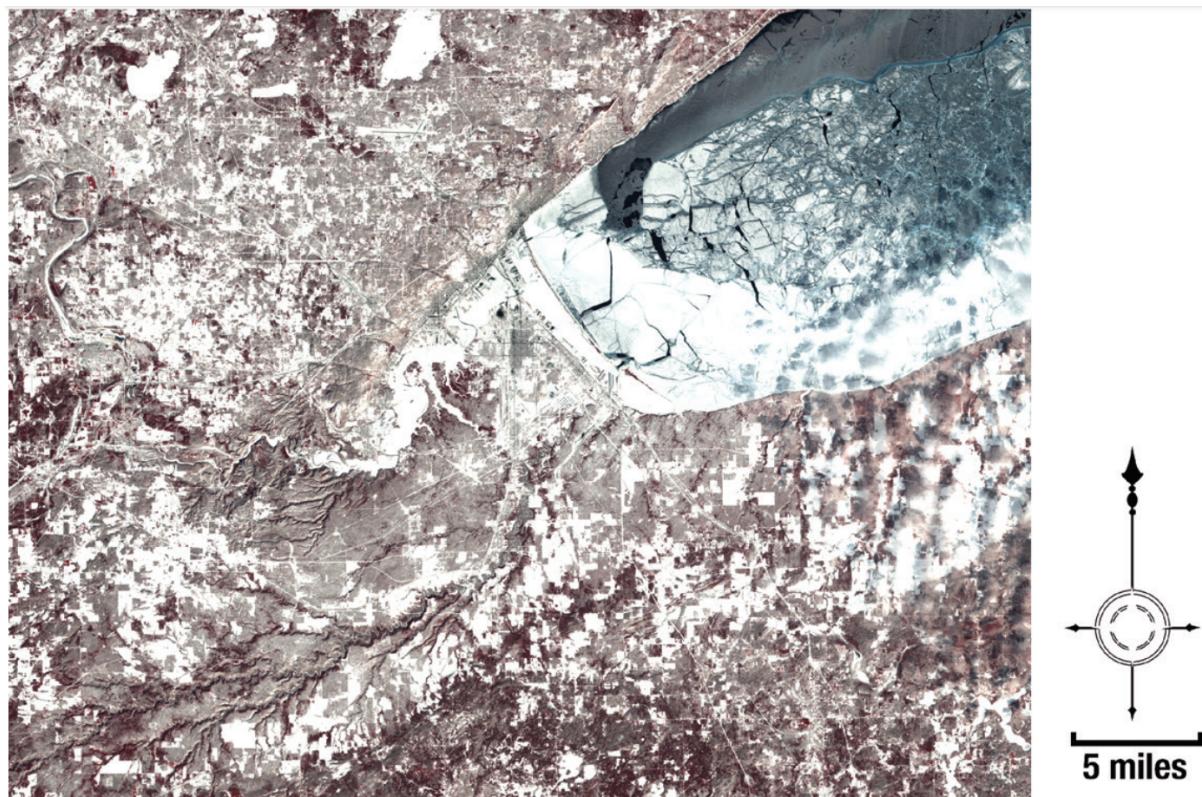
- **Desempeño del TM:** A lo largo de su misión, Landsat 5 proporcionó datos de alta

calidad gracias al sensor TM, a pesar de enfrentar problemas de comunicación y fallos técnicos. La redundancia en los sistemas críticos y una mayor capacidad de combustible contribuyeron a su longevidad.

- **Innovaciones y aplicaciones:** Las mejoras en resolución espectral y radiométrica del TM permitieron aplicaciones avanzadas en clasificación de tierras y monitoreo ambiental. Estudios realizados con los datos del TM demostraron mejoras significativas en la precisión de clasificación, aunque la resolución espacial mejorada del TM no siempre resultó en mejoras proporcionales en todas las aplicaciones.

**Figura 8.**

*Sub-sección de una de las primeras imágenes registradas por el Thematic Mapper de Landsat 5.*



*Nota.* Se muestra Duluth, Minnesota y el rompimiento del hielo en el Lago Superior, adquirida el 6 de marzo de 1984. Tomado de Council (2013).

**2.2.1.4 Armonización de imágenes en Teledetección.** La armonización se refiere al ajuste radiométrico y geométrico de imágenes de múltiples sensores para garantizar la confia-

bilidad en teledetección (Ichikawa et al., 2022). A medida que aumenta la disponibilidad de datos de teledetección, es esencial normalizar imágenes con propiedades radiométricas variables para mosaicos coherentes (Langheinrich et al., 2017). Además, implica alinear datos de diferentes fuentes para interpretaciones uniformes, utilizando modelos específicos como Fiware (Vosnakis et al., 2022). Sin embargo, al estudiar la temperatura de la superficie terrestre, la armonización enfrenta desafíos debido a sesgos entre diferentes satélites (Adeniran et al., 2022).

**2.2.1.5 Armonización geométrica de datos de Teledetección.** Los datos de teledetección a menudo presentan desafíos inherentes. Uno de los principales es la falta de alineación perfecta entre imágenes, lo que da lugar a lo que se conoce como desajustes espaciales o misregistrations (Storey et al., 2016; Wolfe et al., 2002; Yan et al., 2018). Estos desajustes pueden oscilar desde sub-píxeles hasta unos pocos píxeles.

**Causas de desalineaciones.** Estos desajustes pueden ser atribuidos a diversas causas, desde la densidad desigual de puntos de control en tierra, el uso de conjuntos de datos de referencia desiguales, hasta problemas en el software de preprocesamiento de los proveedores o movimientos no modelados del sensor (Storey et al., 2016). Estos desajustes, si no se corrigen, pueden afectar gravemente productos derivados, como clasificaciones de imágenes y detección de cambios.

**Proceso de armonización.** La armonización geométrica tiene como objetivo abordar y corregir estas heterogeneidades espaciales. Se compone de dos fases principales: (1) la corrección de misregistrations, y (2) el remuestreo y reproyección de los datos (Roy, 2000). Sin embargo, corregir estos desajustes no es una tarea sencilla. Las imágenes pueden haber sido adquiridas en diferentes ángulos, fechas, y horarios, lo que complica aún más el proceso de armonización.

**Resampling y Pan-sharpening.** A pesar de los desafíos, se han desarrollado técnicas para adaptar diferentes cuadrículas de píxeles mediante remuestreo espacial o pan-sharpening (Ghassemian, 2016; Javan et al., 2021; Li et al., 2017). Estas técnicas están implementadas en varios paquetes de software de código abierto.

**Técnicas de co-registración.** Las técnicas de co-registración son esenciales para alinear correctamente las imágenes. Estas técnicas se pueden clasificar en dos categorías principales: basadas en características y basadas en intensidad (Zitova & Flusser, 2003). Mientras que las primeras se centran en características notables en las imágenes, como edificios o carreteras, las segundas se basan en patrones de valores de grises. Ambas técnicas ofrecen sus propios beneficios y desafíos.

**Necesidad de un método universal.** Dada la diversidad y complejidad de los datos de teledetección, es esencial contar con un método de co-registración que sea universalmente aplicable (Chen et al., 2003). Este método debería ser capaz de manejar desajustes técnicos entre múltiples sensores y desafíos presentados por paisajes y atmósferas cambiantes.

**2.2.1.6 Armonización espectral de datos de Teledetección.** Los sensores de teledetección óptica se caracterizan por presentar diversos bandas espectrales, abarcando no solo el espectro electromagnético visible (VIS, 400 – 700 nm), sino también el infrarrojo cercano (NIR, 750 – 1000 nm) y el infrarrojo de onda corta (SWIR, 1000 – 2500 nm). Estos sensores facilitan la diferenciación de materiales superficiales y permiten identificar ciertas propiedades del cobertura terrestre o composiciones atmosféricas.

**Discrepancias espectrales entre sensores.** Aunque satélites de Observación de la Tierra (EO) como Landsat (NASA) o Sentinel-2 (ESA) suelen diseñarse con características espectrales similares para garantizar la comparabilidad a largo plazo, aún existen discrepancias en el número de bandas y las respuestas espectrales de cada sensor (Drusch et al., 2012).

**Objetivo de la armonización espectral.** La armonización espectral busca transformar la información espectral de un sensor al dominio espectral de otro, con el fin de estandarizar el número de bandas y reducir inconsistencias entre imágenes de múltiples sensores, facilitando así análisis y flujos de trabajo posteriores.

**Técnicas existentes.** Existen técnicas basadas en relaciones estadísticas de bandas equivalentes de múltiples sensores para lograr una transformación espectral (Chastain et al., 2019; Claverie et al., 2018; Flood, 2014; Roy, Kovalskyy et al., 2016). Sin embargo, estas técnicas

están limitadas a bandas con equivalentes directos entre sensores, lo que deja un potencial no aprovechado en regiones espectrales como el borde rojo (Filella & Penuelas, 1994).

**Necesidad de avances metodológicos.** Es fundamental avanzar en las técnicas existentes de armonización espectral para estimar con precisión la información espectral faltante y reducir los errores de armonización que varían espacialmente. Además, es esencial evaluar el rendimiento de diferentes técnicas de armonización y cuantificar los errores de estimación según los materiales superficiales y las longitudes de onda de las bandas estimadas.

#### **2.2.1.7 Normalización y ajuste de datos con aplicación de parámetros BRDF.**

**Fundamentos de la función de distribución bidireccional de reflectancia (BRDF).** En estudios anteriores, la BRDF se estableció como una función esencial que caracteriza cómo la luz es reflejada por las superficies terrestres en varias direcciones. Se observó que las superficies no son Lambertianas, lo que significa que la reflectancia varía con la geometría solar y de observación. Esto es crucial para corregir y comparar datos de reflectancia obtenidos en diferentes momentos y bajo diversas condiciones de iluminación (Roy, Zhang et al., 2016).

**Aplicación de parámetros BRDF en la normalización de datos.** La práctica de normalizar la reflectancia a través de parámetros BRDF se mostró eficaz para estandarizar las observaciones de Landsat a una vista de nadir. Esto permite mitigar los efectos direccionales y obtener una reflectancia consistente y comparable en el tiempo. Se utilizó un enfoque de c-factor, multiplicando la reflectancia observada por el cociente de la reflectancia modelada, lo que resultó ser poco sensible al tipo de cobertura terrestre y por lo tanto aplicable a todo el registro de Landsat (Roy, Zhang et al., 2016).

**Integración de datos TM y ETM+ en modelos BRDF.** La integración de datos de los sensores TM y ETM+ en modelos BRDF se abordó para proporcionar una reflectancia ajustada al nadir más consistente. Se descubrió que los parámetros de BRDF de MODIS podían ser aplicados a estos datos, lo que sugiere que las formas de BRDF de diversas superficies terrestres son lo suficientemente similares como para permitir esta aplicación directa, facilitando la generación de NBAR para una amplia gama de aplicaciones de monitoreo terrestre.

**Tabla 3.**

*Parámetros BRDF MODIS globales fijos para todas las bandas Landsat.*

Banda Landsat	n	fiso	fgeo	fvol
1 (azul)	15,551,077,545	0.0774	0.0079	0.0372
2 (verde)	16,362,112,402	0.1306	0.0178	0.0580
3 (rojo)	16,095,103,393	0.1690	0.0227	0.0574
4 (NIR)	16,260,280,058	0.3093	0.0330	0.1535
5 (1.6 $\mu$ m)	16,176,131,413	0.3430	0.0453	0.1154
7 (2.1 $\mu$ m)	16,149,440,059	0.2658	0.0387	0.0639

*Nota.* La tabla muestra los parámetros BRDF fijos de MODIS para cada banda de Landsat, utilizados en normalización y comparación con datos satelitales; estos parámetros son constantes y globales durante 12 meses en todas las bandas, reflejando la cantidad 'n' de píxeles de alta calidad y sin nieve a 500 m en los espectrales MODIS BRDF, donde 'n' varía según la banda, lo que corresponde a las variaciones en la calidad de parámetros en el producto MODIS BRDF/Albedo (MCD43A2) (Roy, Zhang et al., 2016).

## 2.2.2 *Inteligencia artificial*

**2.2.2.1 Deep learning en el procesamiento de imágenes de Teledetección.** El deep learning, especialmente en el ámbito de la teledetección, ha emergido como una técnica central y de vanguardia para diversas aplicaciones de visión por computadora. Los investigadores están en constante búsqueda de mejorar el rendimiento de los métodos de deep learning mediante el desarrollo de nuevos diseños arquitectónicos de redes y/o la implementación de técnicas innovadoras, como los mecanismos de atención (Ghaffarian et al., 2021). Estos avances han demostrado ser esenciales para mejorar la precisión y eficiencia en el procesamiento de imágenes de teledetección (Wang et al., 2022).

**2.2.2.2 Redes neuronales convolucionales (CNN).** Las Redes Neuronales Convolucionales, o CNN, son una subcategoría de redes neuronales diseñadas específicamente para manejar datos con una estructura topológica definida. Estas redes son esenciales para el análisis de datos estructurados en series temporales, que se pueden visualizar como una línea continua con intervalos de tiempo uniformes, o imágenes, que se representan como una matriz bidimensional

de píxeles. Lo que distingue a las CNN de otras redes neuronales es su uso de la convolución, una operación matemática lineal específica. En lugar de depender únicamente de las multiplicaciones matriciales, las CNN incorporan la operación de convolución en al menos una de sus capas (Goodfellow et al., 2016).

**Convolución y su aplicación en aprendizaje profundo.** La convolución es una operación matemática esencial en el procesamiento de señales, especialmente en el análisis de imágenes dentro del aprendizaje profundo. Es particularmente relevante en las Redes Neuronales Convolucionales (CNNs) (Goodfellow et al., 2016). En términos generales, la convolución evalúa cómo se superpone una función con otra cuando una de ellas se desplaza sobre la otra.

La operación matemática de la convolución entre dos funciones,  $x(a)$  y  $w(a)$ , se representa como:

$$s(t) = \int x(a)w(t-a)da \quad (2.1)$$

Donde  $x(a)$  es conocida como la función de entrada y  $w(a)$  es el kernel o filtro. El resultado,  $s(t)$ , es a menudo denominado mapa de características o "feature map". En aprendizaje automático, tanto la función de entrada como el kernel suelen ser matrices multidimensionales o tensores.

Para visualizarlo en el contexto de imágenes, consideremos una imagen bidimensional  $I$  y un kernel bidimensional  $K$ :

$$S(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n) \quad (2.2)$$

Donde  $I(m, n)$  representa el valor del píxel en la posición  $(m, n)$  de la imagen y  $K(i - m, j - n)$  es el valor correspondiente del kernel. Dada la propiedad conmutativa de la convolución, esta ecuación también puede expresarse como:

$$S(i, j) = \sum_m \sum_n I(i + m, j + n)K(m, n) \quad (2.3)$$

Es importante señalar que, en el ámbito del aprendizaje profundo, a menudo se utiliza

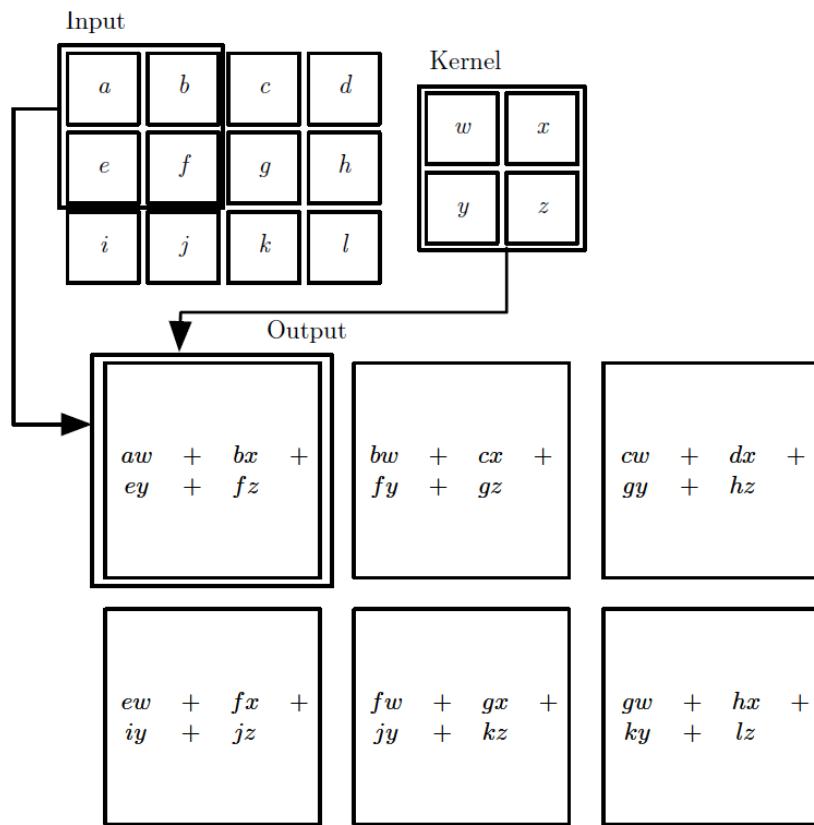
una operación similar llamada *correlación cruzada*:

$$S(i, j) = \sum_m \sum_n I(i + m, j + n)K(m, n) \quad (2.4)$$

A diferencia de la convolución, en la correlación cruzada no se invierte el kernel.

**Figura 9.**

*Una convolución 2-D sin invertir el núcleo*



*Nota.* En la convolución 2-D un núcleo de 2x2 se desplaza sobre una matriz 4x4, multiplicando y sumando correspondencias sin invertir el núcleo, formando así la matriz de salida, un proceso clave en procesamiento de imágenes y redes neuronales convolucionales. Tomado de USGS (2023a).

**Padding.** En el contexto de las Redes Neuronales Convolucionales (CNN), el *padding* es un mecanismo crucial que permite mantener la dimensionalidad de una imagen de entrada durante el proceso de convolución. La idea principal detrás del *padding* es añadir ciertos valores alrededor de una matriz de entrada, y estos valores adicionales son comúnmente representados

por el símbolo  $p$  (Goodfellow et al., 2016).

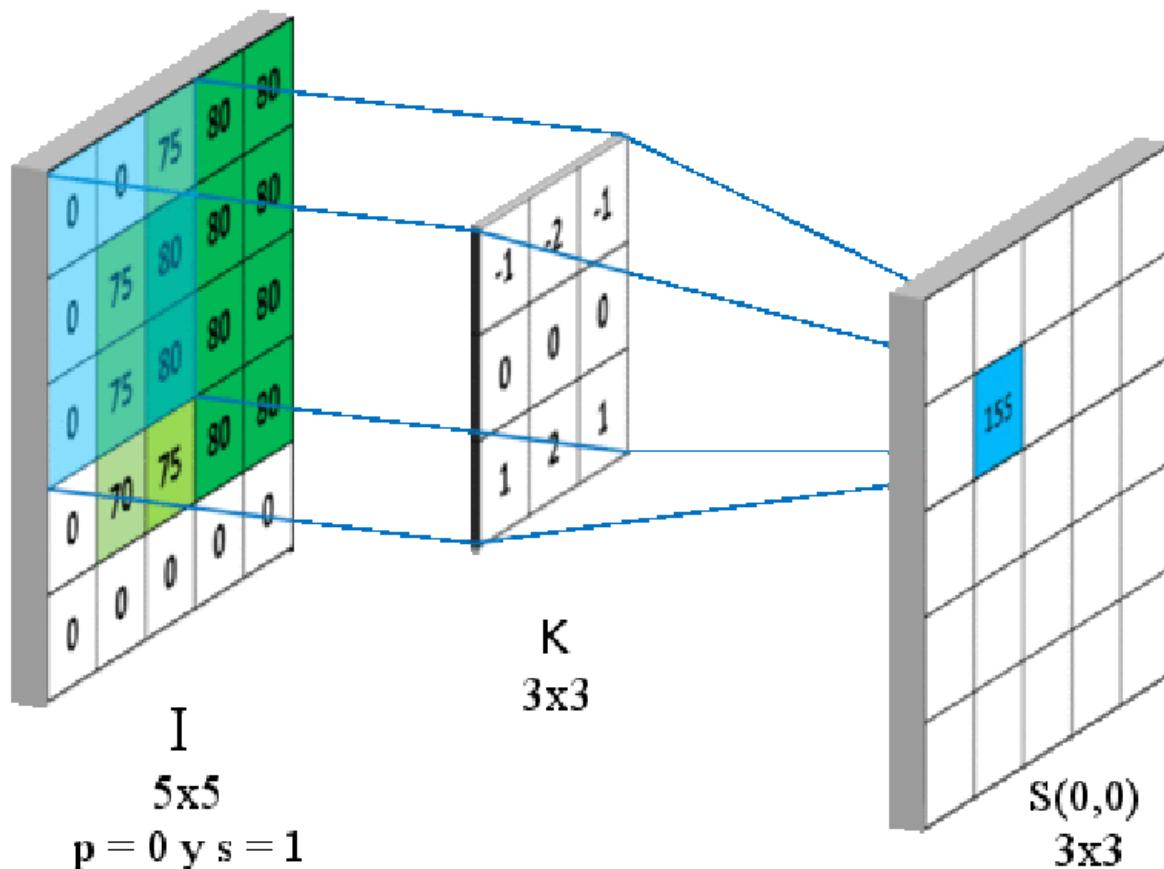
Considere una imagen de entrada representada por una matriz  $I$  de dimensiones  $n \times n$ . Si aplicamos un kernel (o filtro) de tamaño  $f \times f$  a esta imagen, la matriz resultante (sin aplicar *padding*) tendrá un tamaño de  $n - f + 1 \times n - f + 1$ . Esto se debe a que, durante la operación de convolución, el kernel se desliza sobre la imagen de entrada, y, dependiendo del tamaño del kernel, algunas partes de la imagen no son cubiertas.

En la Figura 10, se ilustra un proceso de convolución donde la matriz de entrada  $I$ , de dimensiones  $5 \times 5$ , se convoluciona con un kernel  $K$  de dimensiones  $3 \times 3$ . Se utiliza un *padding*  $p$  de valor 0 y un *stride*  $s$  de valor 1, que es el valor comúnmente asumido. A través de esta operación, obtenemos una matriz de salida  $S$  con dimensiones  $3 \times 3$ , calculadas mediante la fórmula  $5 - 3 + 1 \times 5 - 3 + 1$ . Específicamente, para la posición  $S(0, 0)$  de la matriz de salida, el valor es 155, determinado por la operación de convolución especificada en la figura. Notamos que, debido a que no hay *padding* (es decir,  $p = 0$ ), la matriz de salida es dos unidades menor en dimensiones que la matriz original.

De manera similar, en la Figura 11, se presenta un cálculo para la posición  $S(0, 1)$  con un resultado de 85.

**Figura 10.**

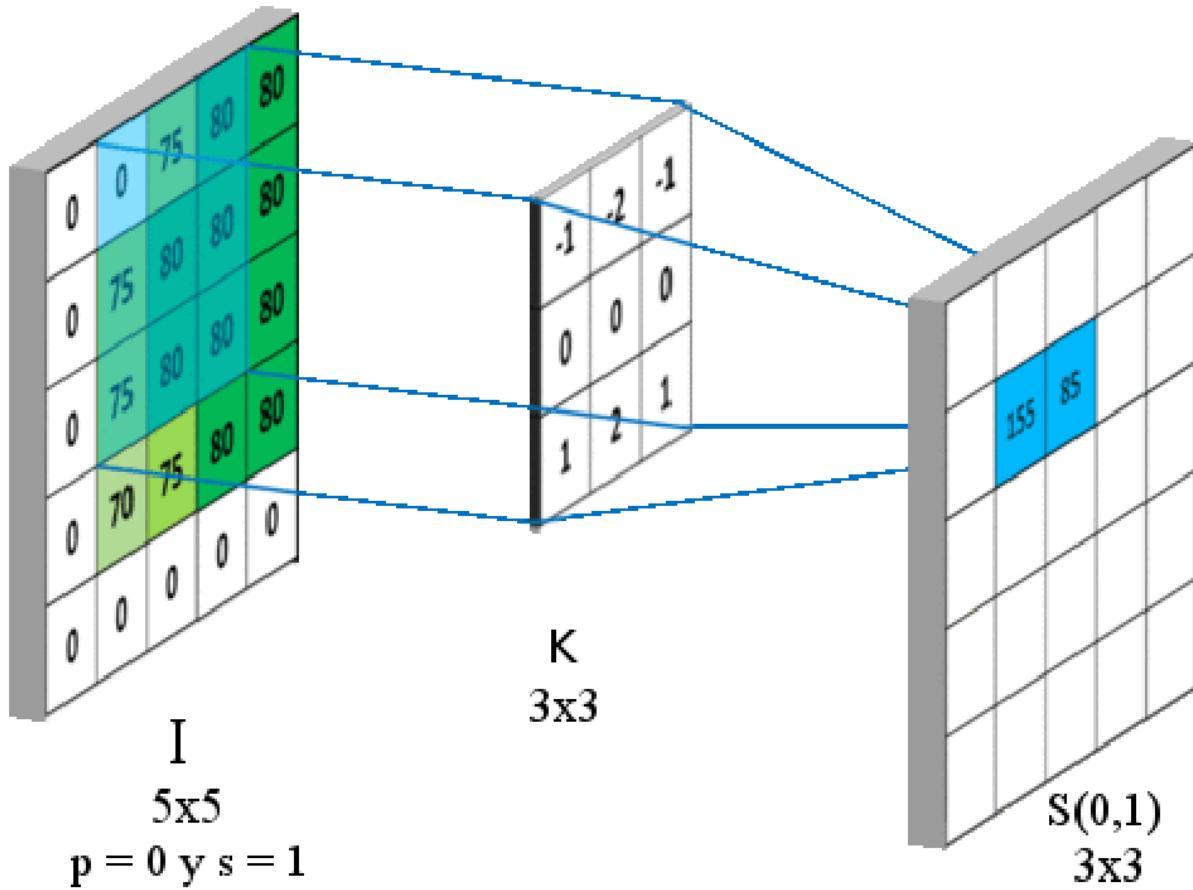
Cálculo de la convolución de una matriz de entrada  $5 \times 5$  con un kernel  $3 \times 3$  con  $p = 0$  y  $s = 1$



*Nota.* La operación genera una matriz de salida de  $3 \times 3$  donde, por ejemplo,  $S(0,0)$  tiene un valor de 155 después de la convolución; la matriz resultante es dos unidades menor que la original por no usar padding. Adaptado de Goodfellow et al. (2016).

**Figura 11.**

Cálculo de la convolución para la posición  $S(0, 1)$  en la matriz de salida.



*Nota.* La convolución involucra una matriz  $5 \times 5$  y un kernel  $3 \times 3$ , resultando en  $S(0, 1) = 85$ , con líneas azules que representan la transformación. Adaptado de Goodfellow et al. (2016).

En la mayoría de los casos,  $p = 0$ , pero si  $p > 0$ , la dimensión de la matriz de salida se calcula como:

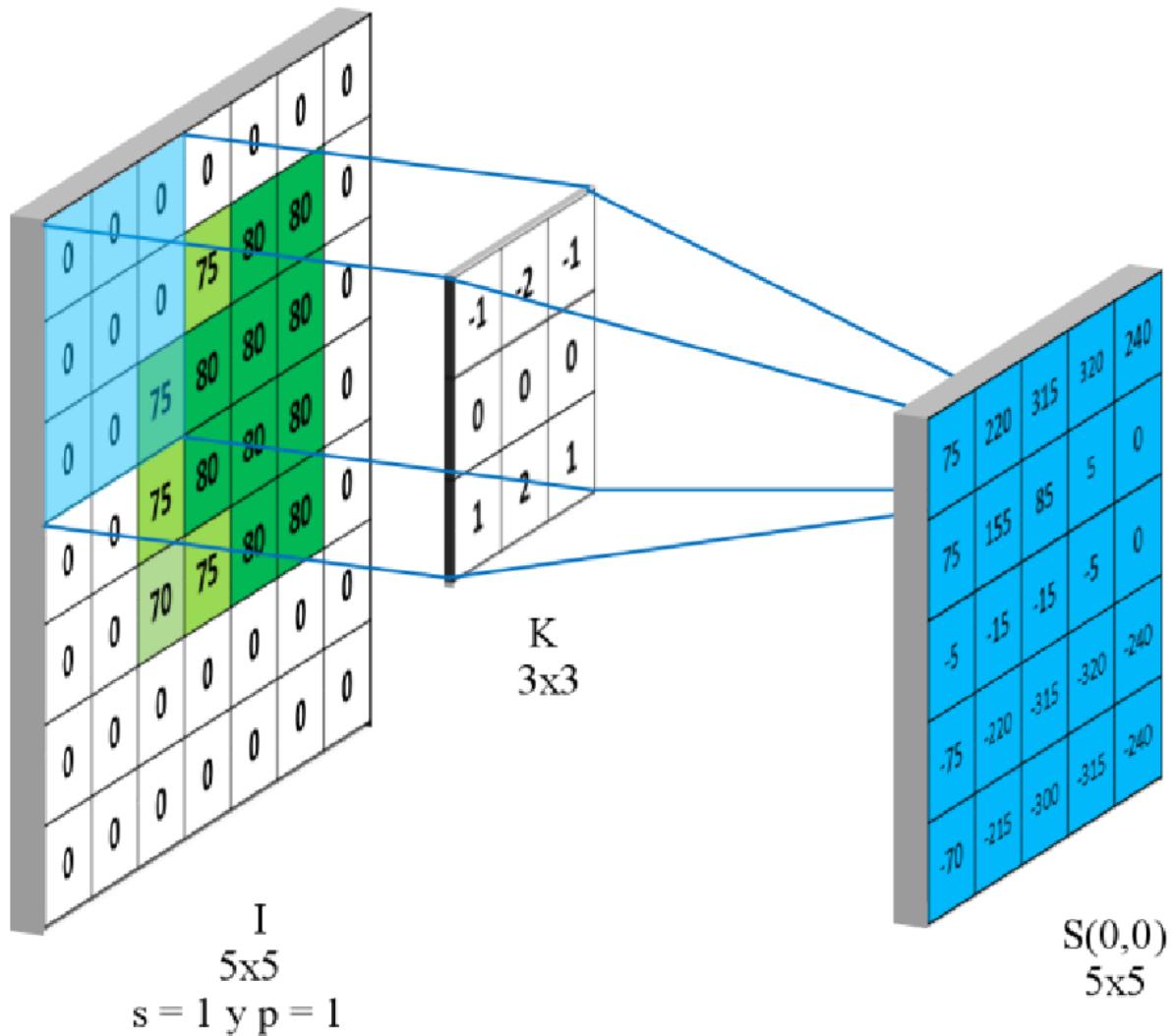
$$\text{Dimensión de Salida} = n + 2 \cdot p - f + 1 \times n + 2 \cdot p - f + 1 \quad (2.5)$$

donde  $n$  es la dimensión de la matriz de entrada y  $f$  es la dimensión del filtro.

Por ejemplo, en la Figura 12, considerando un padding  $p = 1$ , una matriz de entrada  $I$  de dimensiones  $5 \times 5$ , y un filtro de dimensiones  $3 \times 3$ , la matriz de salida resultante tendrá dimensiones  $5 \times 5$ , manteniendo así la dimensión espacial de la matriz de entrada.

**Figura 12.**

Cálculo de la convolución de una matriz de entrada  $5 \times 5$  con un kernel  $3 \times 3$  con  $p = 1$  y  $s = 1$



Nota. Padding de  $p = 1$  mantiene dimensiones de la imagen original, esencial para que redes neuronales aprendan patrones locales eficientemente. Adaptado de Goodfellow et al. (2016).

Las convoluciones pueden clasificarse en:

- **Convoluciones válidas:** No utilizan padding, resultando en una matriz de salida de dimensiones menores.
- **Convoluciones iguales:** Utilizan padding para asegurar que la matriz de salida tenga las mismas dimensiones que la matriz de entrada.

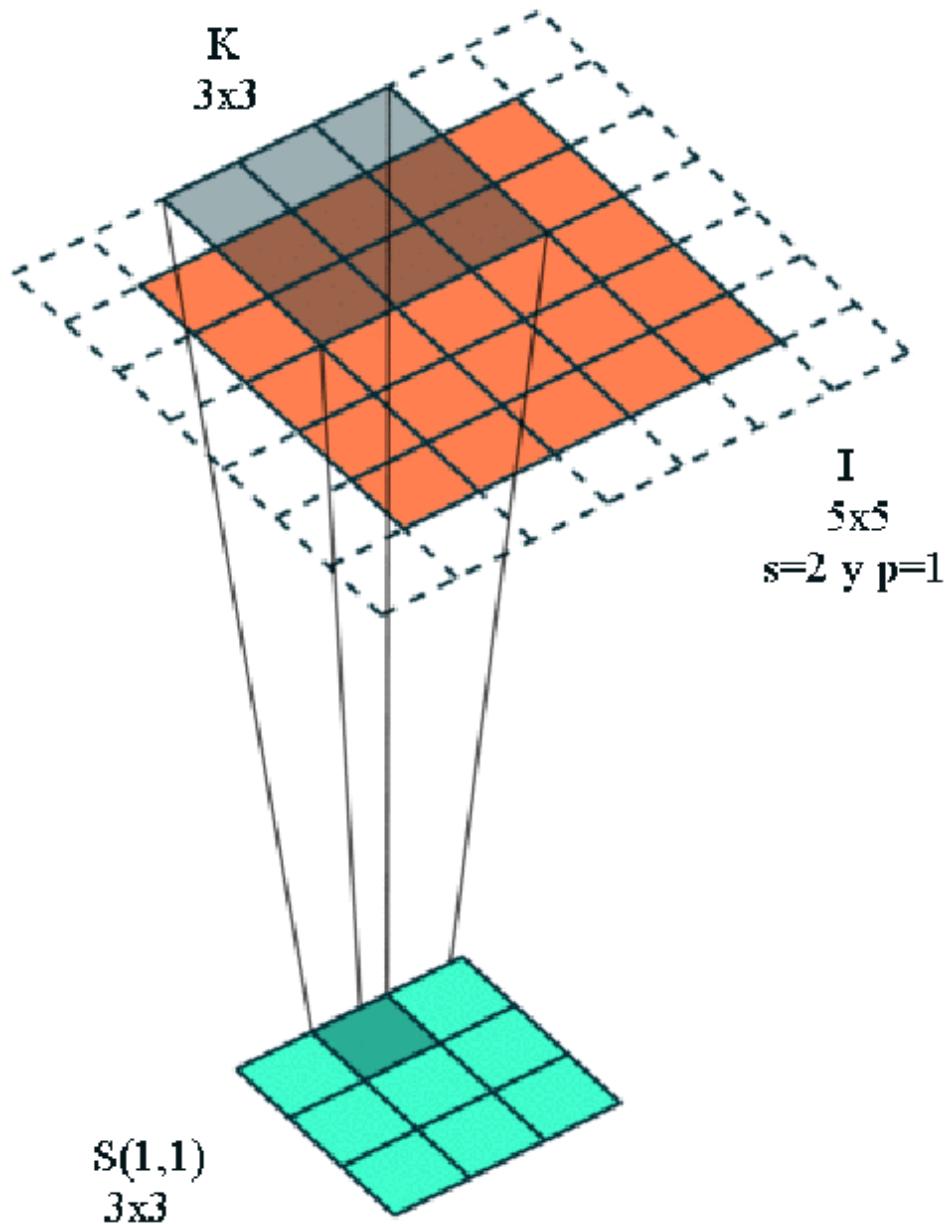
**Stride.** Anteriormente ha sido denotado como  $s$ , representa la cantidad de píxeles que el filtro se moverá durante la operación de convolución. Actualizando la fórmula de la dimensión de salida para incluir el stride, tenemos:

$$\text{Dimensión de Salida} = \left\lfloor \frac{n + 2 \cdot p - f}{s} + 1 \right\rfloor \times \left\lfloor \frac{n + 2 \cdot p - f}{s} + 1 \right\rfloor \quad (2.6)$$

Por ejemplo, en la Figura 13, con un stride  $s = 2$ , un padding  $p = 1$ , una matriz de entrada  $I$  de dimensiones  $5 \times 5$ , y un filtro de dimensiones  $3 \times 3$ , la matriz de salida resultante tendrá dimensiones  $3 \times 3$ .

**Figura 13.**

Convolución 2D con Padding de 1 y Stride de 2 en la Posición S(1, 1).



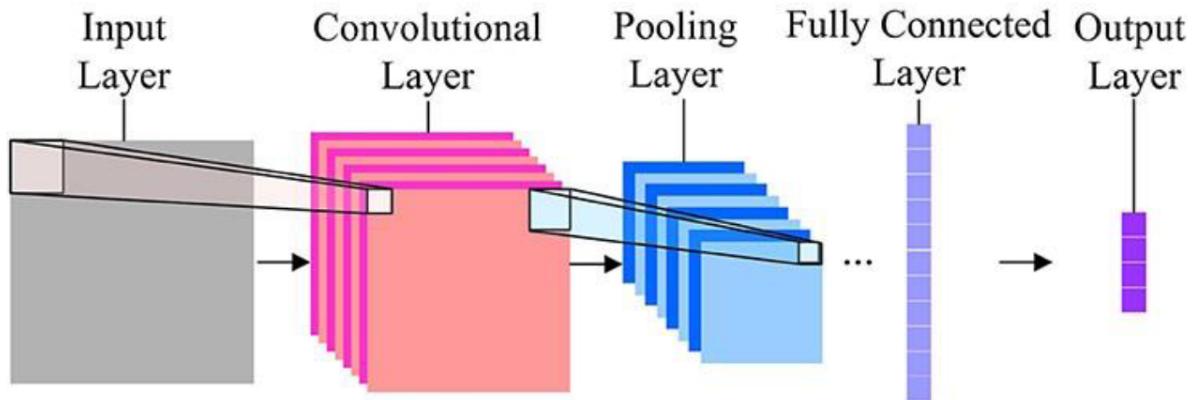
*Nota.* Convolución con stride de 2 y filtro de  $3 \times 3$ , desplazando dos píxeles a la vez y padding que preserva información en bordes. Adaptado de Goodfellow et al. (2016).

**2.2.2.3 Tipos de capas en una red neuronal convolucional.** Las Redes Neuronales Convolucionales emplean una serie de capas específicas que incluyen: Capa de Entrada, Capa Convolutacional, Capa de Agrupación, Capa Totalmente Conectada y Capa de Salida. Cada capa

tiene funciones específicas y contribuye al rendimiento general de la red.

**Figura 14.**

*Representación esquemática de una Red Neuronal Convolucional.*

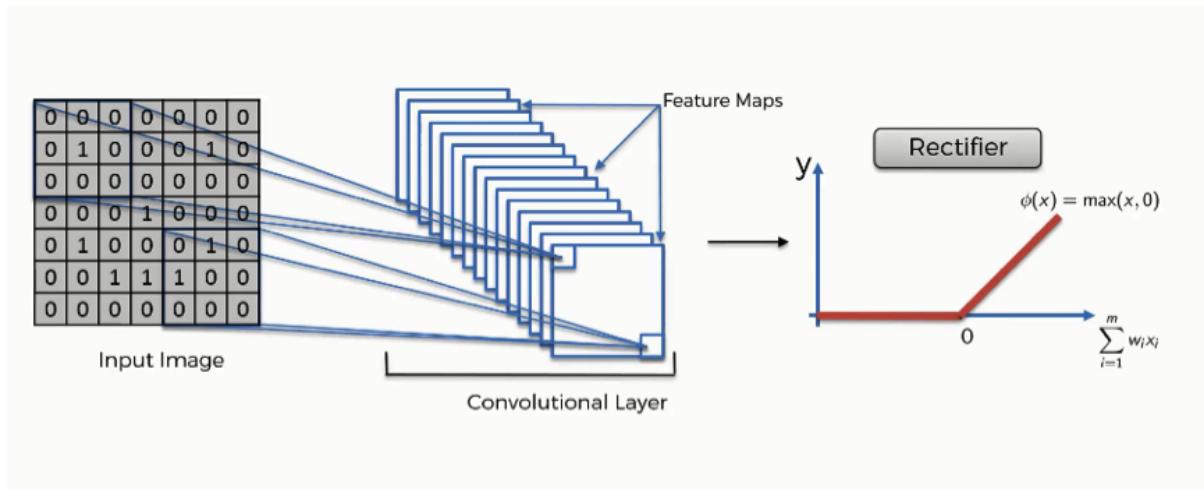


*Nota.* La estructura CNN: capas convolucionales detectan patrones y reducen tamaño; una capa conectada integra rasgos para la clasificación final, crucial en procesamiento visual. Adaptado de Goodfellow et al. (2016).

**Capa de Convolución.** Esta capa realiza operaciones de convolución en la entrada utilizando varios *kernels*. Posteriormente, se agrega un sesgo al resultado y se pasa por una función de activación no lineal como ReLU. Esta función introduce no linealidad tras realizar operaciones lineales en las capas convolucionales. Aunque en el pasado se usaron funciones como tanh y sigmoid, ReLU ha demostrado ser más eficiente en términos de tiempo de entrenamiento y precisión.

**Figura 15.**

Función de activación ReLU.



*Nota.* La función ReLU introduce no linealidad, ayudando a la red a capturar patrones complejos y optimizando el entrenamiento y precisión, evidenciando su ventaja sobre funciones previas como tanh o sigmoid. Tomado de Kalita (2022).

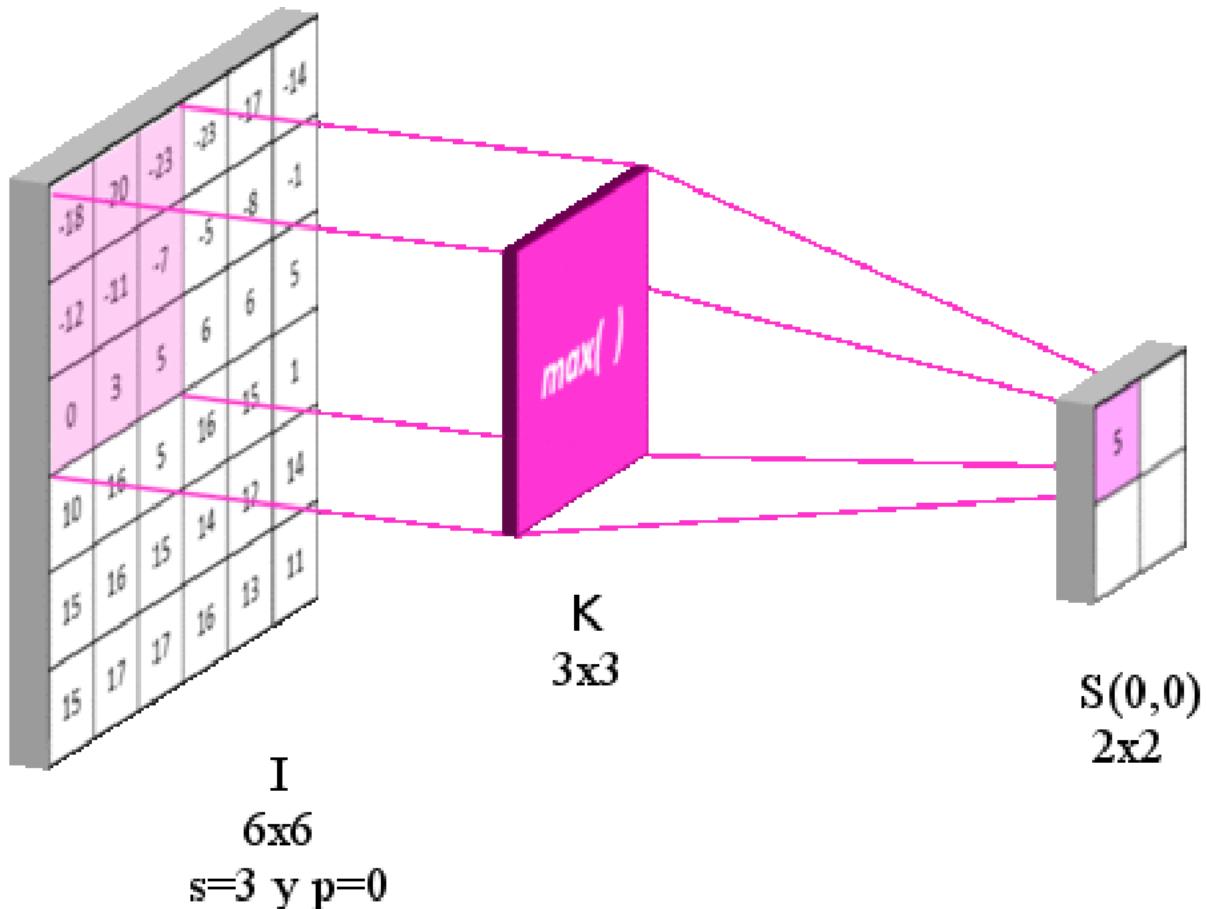
**Capa de agrupación.** Destinada a reducir las dimensiones de la entrada, la Capa de Agrupación mejora la eficiencia y robustez de la red. Una técnica popular aquí es el *Max Pooling*, que selecciona el valor máximo de un conjunto específico, en la Figura 16 se muestra un ejemplo donde una matriz I de tamaño 6 x 6, con un s igual a 3 y p igual a 0 . Esto es especialmente útil porque no introduce parámetros adicionales y es computacionalmente eficiente.

Diversas técnicas de agrupación incluyen:

- Agrupación Máxima:  $s_j = \max_{i \in R_j} a_i$
- Agrupación Promedio:  $s_j = \frac{1}{|R_j|} \sum_{i \in R_j} a_i$
- Agrupación Estocástica:  $p_j = \frac{a_i}{\sum_{k \in R_j} a_k}$

**Figura 16.**

Max-Pooling con padding de 0 y stride de 3 en posición  $S(0, 0)$ .



*Nota.* Max-Pooling con kernel de  $3 \times 3$  y stride de 3 sobre una matriz  $6 \times 6$  extrae máximos locales, resultando en una matriz  $2 \times 2$ , compactando información y resaltando características prominentes sin pérdida significativa. Adaptado de Goodfellow et al. (2016).

**Capa totalmente conectada.** Después de la agrupación, la entrada se pasa a una capa totalmente conectada que la conecta con la capa de salida, estableciendo conexiones entre todas las unidades (Goodfellow et al., 2016).

**Hiperparámetros.** Los hiperparámetros, a diferencia de los parámetros, son valores pre-establecidos que determinan cómo se comporta un algoritmo durante el entrenamiento. Estos se definen antes del proceso de aprendizaje y no se adaptan automáticamente durante el mismo (Goodfellow et al., 2016).

Al diseñar arquitecturas de Redes Neuronales Convolucionales (CNN), se consideran diversos hiperparámetros, entre los cuales se encuentran:

- **Tamaño del kernel:** Definido por la dimensión  $f \times f$ , se utiliza en la capa convolucional o de agrupación.
- **Padding:** Indica cuántos píxeles se añaden alrededor de la imagen, normalmente llenados con cero, para ajustar las dimensiones durante las convoluciones.
- **Stride:** Define cuántos píxeles se desplaza el kernel durante la convolución.
- **Número de kernels:** La cantidad de filtros en una capa convolucional.
- **Número de capas:** El total de capas en una CNN.
- **Función de activación:** Función no lineal aplicada tras la convolución, comúnmente ReLU, Sigmoid o Tanh.

También se deben considerar hiperparámetros específicos al definir el conjunto de entrenamiento:

- **Muestras de entrenamiento:** Total de ejemplos utilizados en el entrenamiento.
- **Tamaño del lote (batch size):** Número de ejemplos procesados en una sola iteración de entrenamiento.
- **Iteraciones:** Total de lotes necesarios para completar una época.
- **Épocas:** Número de veces que se procesa el conjunto completo de entrenamiento.
- **Tasa de aprendizaje ( $\varepsilon$ ):** Define la magnitud de ajuste de los pesos en función del gradiente de pérdida.

El número de épocas se calcula según la relación:

$$\text{epochs} = \frac{\text{batch\_size} \times \text{iterations}}{\text{training\_samples}} \quad (2.7)$$

Existen otros hiperparámetros que se emplean como métodos de regularización:

- **Regularización L2:** Penaliza la magnitud de los pesos para evitar sobreajuste.
- **Regularización L1:** Similar a L2, pero puede hacer que algunos pesos sean exactamente cero.
- **Dropout:** Durante el entrenamiento, desactiva aleatoriamente algunas neuronas para prevenir la dependencia excesiva en cualquier neurona individual.
- **Augmentation:** Amplía el conjunto de entrenamiento al introducir variaciones menores, como rotaciones y escalas, para mejorar la generalización del modelo.

## 2.3 Definición de términos

### 2.3.1 Armonización de datos

Se refiere al proceso de ajustar conjuntos de datos para que sean consistentes entre sí en términos de escala, resolución, calibración y geometría. En la teledetección, la armonización es clave para comparar y analizar imágenes tomadas por diferentes sensores o tecnologías a lo largo del tiempo, como las de las distintas generaciones del sensor Landsat MSS.

### 2.3.2 BRDF (*Bidirectional Reflectance Distribution Function*)

Función que describe cómo la reflectancia de una superficie varía con la geometría de la iluminación y la observación. Es esencial para la corrección y normalización de imágenes satelitales, permitiendo comparaciones más precisas entre diferentes fechas y condiciones de iluminación.

### 2.3.3 CNN (*Redes Neuronales Convolucionales*)

Tipo de red neuronal diseñada para procesar datos con una estructura de grilla, como imágenes, utilizando convoluciones para extraer características relevantes. Son ampliamente utilizadas en la teledetección para tareas como la segmentación de imágenes y la clasificación de objetos.

#### **2.3.4 Co-registración**

Proceso de alinear geométricamente varias imágenes de diferentes sensores para que coincidan espacialmente. Es esencial para el análisis multitemporal y la fusión de imágenes, permitiendo comparaciones precisas y la integración de datos de diversas fuentes.

#### **2.3.5 Cubo de datos**

En el contexto de la teledetección y la ciencia de la Tierra, un cubo de datos se refiere a una colección multidimensional de datos que se ha estructurado de tal manera que permite un análisis eficiente y flexible. Los cubos de datos generalmente organizan la información en tres dimensiones: espacial (latitud y longitud), temporal (tiempo) y espectral (bandas de captura del sensor), lo que facilita diversas operaciones analíticas como la comparación temporal y el análisis de cambios.

#### **2.3.6 Deep learning**

Es una rama del aprendizaje automático basada en algoritmos que modelan abstracciones de alto nivel en datos usando arquitecturas compuestas de múltiples transformaciones no lineales. En el contexto de la teledetección, se utiliza para interpretar y procesar grandes volúmenes de datos de imágenes satelitales, permitiendo la identificación de patrones y la realización de tareas como la clasificación de la cobertura terrestre y la detección de cambios.

#### **2.3.7 Entornos conda**

Conda es un sistema de gestión de paquetes y entornos que permite a los usuarios instalar y gestionar bibliotecas y dependencias en múltiples lenguajes de programación, como Python y R. Los entornos Conda permiten aislar diferentes proyectos y sus dependencias, evitando conflictos entre bibliotecas y facilitando la reproducibilidad de los experimentos. Son particularmente útiles en la teledetección y el procesamiento de imágenes, donde diferentes proyectos pueden requerir versiones específicas de librerías como TensorFlow, PyTorch, GDAL, y otras herramientas de análisis y visualización de datos.

### **2.3.8 *ESRCNN (Extended Super-Resolution Convolutional Neural Network)***

Modelo de red neuronal utilizado para mejorar la resolución espacial de imágenes satelitales mediante la fusión de datos de diferentes fuentes. Este modelo ha demostrado ser eficaz en la producción de imágenes coherentes y de alta calidad a partir de datos de menor resolución.

### **2.3.9 *GAN (Generative Adversarial Network)***

Tipo de red neuronal que consiste en dos modelos en competencia, un generador y un discriminador, utilizada para generar datos sintéticos que parecen reales. Es particularmente útil en la generación de imágenes de alta resolución y en la simulación de datos faltantes.

### **2.3.10 *Generación de banda virtual***

En la teledetección, la generación de bandas virtuales implica la creación de nuevas bandas de datos que no se capturan directamente por el sensor, generalmente a través de algoritmos que simulan estas bandas basándose en las bandas existentes y las relaciones conocidas entre ellas. Esto puede ayudar a mejorar la interpretación y análisis de las imágenes, especialmente cuando se combinan datos de diferentes sensores que tienen distintas capacidades espectrales.

### **2.3.11 *Histogram matching***

Es una técnica de procesamiento de imágenes que se utiliza para ajustar la distribución de brillo de una imagen para que coincida con la distribución de otra imagen. Esto se utiliza en la armonización de imágenes de diferentes sensores para que las características similares tengan una apariencia similar en términos de intensidad y contraste.

### **2.3.12 *Landsat***

Serie de satélites de observación terrestre gestionados por la NASA y el USGS, utilizados para monitorear y estudiar la superficie de la Tierra desde 1972. Proporcionan datos esenciales para investigaciones en áreas como la agricultura, la silvicultura, y el monitoreo ambiental.

### **2.3.13 MSS (*Multi-Spectral Scanner*)**

Primer sensor utilizado en los satélites Landsat 1 a 5, que captura imágenes en varias bandas espectrales con una resolución espacial específica. Este sensor ha sido crucial para el estudio de cambios en la superficie terrestre a lo largo de varias décadas.

### **2.3.14 NDVI**

Índice utilizado para evaluar la presencia y condición de la vegetación en la superficie terrestre, calculado a partir de las bandas espectrales del rojo y el infrarrojo cercano. Es una herramienta crucial para el monitoreo de la salud de los ecosistemas y la gestión de recursos agrícolas.

### **2.3.15 Segmentación de imágenes**

Técnica de procesamiento de imágenes que divide una imagen en regiones o segmentos distintos para facilitar su análisis y procesamiento posterior. Es fundamental en la teledetección para la identificación y clasificación de diferentes tipos de cobertura terrestre.

### **2.3.16 Super-resolución**

Técnica para aumentar la resolución de imágenes utilizando algoritmos avanzados, como redes neuronales, para generar detalles adicionales a partir de datos de baja resolución. Esto mejora la calidad y la utilidad de las imágenes satelitales para análisis detallados.

### **2.3.17 TM (*Thematic Mapper*)**

Sensor avanzado utilizado en los satélites Landsat 4 y 5 que ofrece mejoras en la resolución espacial y espectral en comparación con el MSS. Es capaz de capturar imágenes en siete bandas espectrales, permitiendo análisis más detallados y precisos.

### **2.3.18 Transformadas de Fourier**

En el procesamiento de imágenes, la transformada de Fourier es una herramienta matemática que descompone una imagen en sus frecuencias espaciales. Esto es útil para analizar patrones periódicos y filtrar ruido o para operaciones de corrección geométrica y espectral en

el dominio de la frecuencia.

## **CAPÍTULO III**

### **3. HIPÓTESIS Y VARIABLES**

#### **3.1 Las hipótesis**

##### ***3.1.1 Hipótesis general***

- El uso de inteligencia artificial para armonizar las imágenes Landsat MSS permitirá que adquieran propiedades de las imágenes TM, haciendo viable su uso en monitoreos globales de largo plazo

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

- La integración de aprendizaje profundo y procesamiento de imágenes mejorará la corrección geométrica de las imágenes Landsat MSS, facilitando su armonización con las imágenes TM.
- El modelo MSS2TM, basado en aprendizaje profundo, logrará una alineación precisa tanto espectral como espacial entre las imágenes Landsat MSS y TM.
- La técnica de aprendizaje profundo propuesta permitirá completar las bandas ausentes en las imágenes Landsat MSS, logrando una similitud significativa con las bandas presentes en las TM.

#### **3.2 Las variables**

##### ***3.2.1 Variable independiente***

Inteligencia artificial.

##### ***3.2.2 Variable dependiente***

Armonización de imágenes satelitales.

### 3.3 Operacionalización de variables

**Tabla 4.**

*Operacionalización de variables*

Variables	Dimensiones	Indicadores	Unidad de medida	Instrumento
Inteligencia artificial (independiente)	Corrección geométrica	Error RMS después del ajuste geométrico	Píxeles (px)	Python (LightGlue)
	Alineación espectral	Coefficiente de correlación entre las imágenes MSS y TM armonizadas espacialmente	Coefficiente de correlación (r)	Python (Pytorch)
	Generación de bandas faltantes	Número de bandas generadas para completar MSS comparable con TM	Número de bandas (nb)	Python (Pytorch)
Armonización de imágenes satelitales (dependiente)	Precisión de alineación	Precisión de la superposición de píxeles en imágenes armonizadas	Metros (m)	Python (GDAL, Rasterio)
	Similitud espectral	Índice de similitud espectral entre imágenes MSS y TM	Sin unidades	Python (PyTorch)
	Resolución espacial	Resolución espacial de las imágenes armonizadas	Metros por píxel (m/px)	Python (Rasterio)
Integridad de datos temporales		Cobertura temporal completa en el cubo de datos armonizado	Porcentaje (%)	Python (xarray)

*Nota.* Esta tabla muestra las variables operacionalizadas, destacando cómo la inteligencia artificial contribuye a la armonización de imágenes satelitales con modelos de aprendizaje profundo reflejados en las dimensiones y métricas, utilizando Python como instrumento clave de implementación.

## **CAPÍTULO IV**

## **4. METODOLOGÍA**

### **4.1 Área de estudio (unidad de análisis)**

La investigación inicialmente se enfocó en la armonización y superresolución de imágenes Landsat MSS y TM de diversas regiones a nivel mundial, excluyendo la Antártida por su homogeneidad espectral, principalmente dominada por extensas áreas de hielo y nieve, lo que dificulta la diferenciación de coberturas y complica la armonización satelital (Kokhanovsky et al., 2019). De forma similar, se descartaron las regiones oceánicas debido a su uniformidad espectral, lo que limita la variabilidad en la reflectancia (Estrella et al., 2021). En consecuencia, el estudio se centró en áreas continentales no homogéneas en términos de reflectancia.

**Figura 17.**

*Mapa del área de estudio.*



*Nota.* Se muestra un mapa satelital global, resaltando las zonas terrestres verdes, omitiendo océanos y la Antártida.

En la siguiente fase del estudio, se seleccionó estratégicamente Perú como área de la aplicación del autocompletado de bandas o canales que le faltan a las imágenes MSS para parecerse espectralmente a las TM. Esta elección se justifica por la variada topografía y diversidad de ecosistemas del país, que van desde la costa del Pacífico hasta las alturas de los Andes, ofreciendo un escenario desafiante y representativo para la superresolución satelital. Se identificaron múltiples puntos de muestreo distribuidos a lo largo del país, en los cuales se extrajeron segmentos de imágenes MSS para su transformación a la resolución de imágenes TM.

**Figura 18.**

*Mapa del área para la validación de los modelos.*



*Nota.* Se muestra un mapa de Perú, teniendo en cuenta toda su área, por su diversidad ecosistémica.

## **4.2 Diseño de investigación**

En el marco del diseño de investigación de la tesis que aborda la armonización de imágenes Landsat MSS mediante inteligencia artificial, se adoptan criterios metodológicos que se alinean con los planteamientos de (Sampieri, 2018). Esta estructura metodológica se desglosa en las siguientes categorías:

### **4.2.1 Tipo**

La investigación se clasifica como aplicada, ya que tiene el objetivo de resolver problemas específicos relacionados con la armonización de imágenes satelitales, aplicando teorías y conocimientos de inteligencia artificial. Este tipo de investigación busca generar soluciones prácticas que puedan ser implementadas en el ámbito del monitoreo global y a largo plazo.

### **4.2.2 Nivel**

Corresponde al nivel descriptivo, ya que se centra en describir las características y el comportamiento de un fenómeno o realidad específica, en este caso, la capacidad de la inteli-

gencia artificial para mejorar la armonización de las imágenes Landsat MSS. A través de este nivel, se busca detallar las particularidades, diferencias o modalidades de un fenómeno sin ejercer control sobre las variables de estudio.

#### **4.2.3 Enfoque**

El enfoque de la investigación es cuantitativo, dado que se recopilan y analizan datos numéricos para evaluar la eficacia de los métodos de inteligencia artificial en la armonización de imágenes. Este enfoque permite una medición objetiva y estadística de los resultados, facilitando la comparación y el análisis de la efectividad de las técnicas empleadas.

#### **4.2.4 Diseño**

El diseño de esta investigación es no experimental, transeccional o transversal descriptivo. Esto significa que se observan fenómenos tal como se dan en su contexto natural para luego describirlos, sin manipular las variables de estudio. En este caso, se analiza la efectividad de la inteligencia artificial en la armonización de imágenes Landsat MSS, evaluando los datos existentes en un único momento, o en varios momentos bajo el mismo criterio, sin intervenir o modificar las condiciones bajo las cuales se obtuvieron dichas imágenes.

### **4.3 Población y muestra**

#### **4.3.1 Población**

La población de este estudio incluye imágenes capturadas por los sensores MSS y TM de los satélites Landsat 4 y 5, enfocándose en regiones continentales globales, excluyendo la Antártida y zonas oceánicas, entre 1982 y 1999. Estas imágenes, accesibles a través de las bases de datos de Landsat, ofrecen un registro histórico integral para la extracción de datos necesarios en esta investigación, permitiendo analizar las características y cambios en las regiones seleccionadas durante el periodo especificado.

#### **4.3.2 Muestra**

El enfoque de muestreo implementado es probabilístico, lo cual implica la selección aleatoria de unidades de muestreo para asegurar la independencia estadística y representativi-

dad de las imágenes MSS y TM de Landsat en la muestra. Estas imágenes son elegidas basándose en criterios específicos, incluyendo un intervalo de tiempo de captura menor a 10 minutos entre ellas y una cobertura de nubosidad inferior al 15 %, además de cumplir con requisitos de concordancia espacial. Siguiendo los lineamientos propuestos por Roberto et al. (2014) sobre recolección de datos, se ha establecido un tamaño de muestra de 5431 pares de imágenes. Esta cifra refleja adecuadamente la totalidad de la población estudiada, garantizando un análisis robusto y representativo de los datos.

## **4.4 Procedimiento, técnicas e instrumentos de recolección de información**

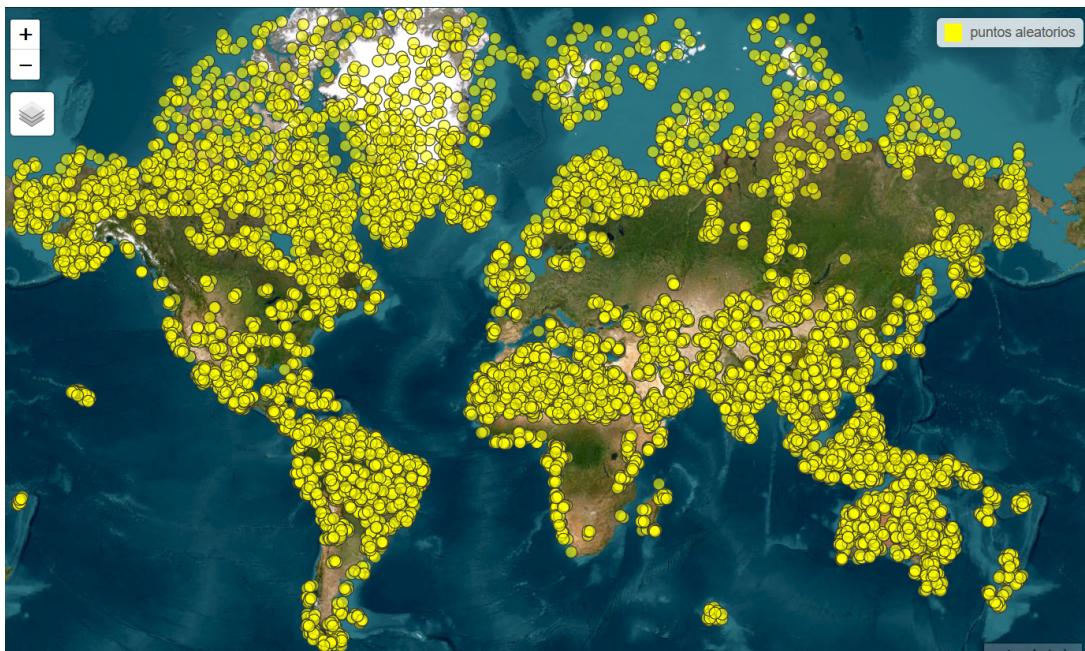
### **4.4.1 Obtención de datos de entrada**

**4.4.1.1 Selección por localización.** Para garantizar la eficacia y precisión de la investigación, se llevó a cabo una meticulosa selección de imágenes para el estudio. A continuación, se detalla el procedimiento de selección:

1. Como punto de partida, se generaron puntos de manera aleatoria alrededor del mundo, resultando en un total inicial de 20 000 puntos. Estos puntos sirvieron como centroides para definir buffers cuadrados con dimensiones de 30 720 metros por lado.

**Figura 19.**

*Generación de 20 000 puntos aleatorios a nivel mundial.*



*Nota.* Los puntos no excedieron los límites del WRS-2 (desde Landsat 4 hasta Landsat 9), que corresponden a las áreas o segmentos de captura de todas las imágenes Landsat, incluyendo los sensores más recientes.

2. Con los puntos geográficos establecidos, se procedió a seleccionar imágenes específicas. Se optó por imágenes MSS pertenecientes exclusivamente a las misiones Landsat 4 y 5, ya que coinciden con las imágenes obtenidas por el sensor TM de las mismas misiones durante el período de 1982 a 1999.
3. Una condición crucial en esta fase fue asegurar que el área definida por cada buffer estuviera íntegramente contenida dentro de al menos una imagen MSS y una imagen TM. Este criterio busca prevenir truncamientos o limitaciones en la amplitud de las imágenes.
4. Se excluyeron ciertas áreas geográficas del proceso: específicamente, regiones de la Antártida y zonas mayormente oceánicas, para asegurar la relevancia y calidad del conjunto de datos final.
5. Tras aplicar todos los criterios anteriores, de los 20 000 puntos iniciales, solo 19 546

satisfacían todas las condiciones y, por lo tanto, fueron retenidos para el estudio.

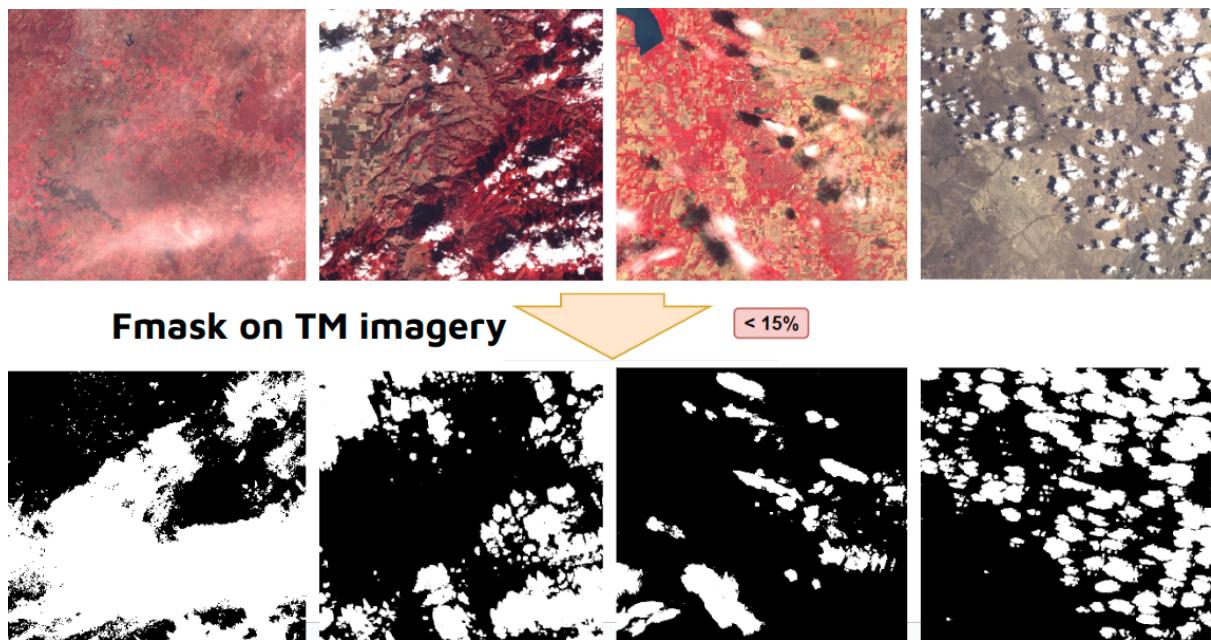
**4.4.1.2 Selección por tiempo y calidad.** Tras el filtro geográfico inicial, se procedió a una selección más detallada basada en criterios temporales y de calidad. A continuación se describe el procedimiento adoptado:

1. De los 20 000 puntos iniciales, solo 19 546 cumplían con los criterios de localización. Para estos puntos seleccionados, se llevó a cabo una comparación temporal entre las imágenes de ambos sensores.
2. La condición impuesta fue que las imágenes de ambos sensores debían tener un diferencial temporal no superior a 10 minutos. Es decir, por cada punto, se retuvieron solo aquellos pares que hubieran sido capturadas en un intervalo de tiempo menor o igual a 10 minutos entre ellas.
3. Es importante mencionar que este filtro temporal se aplicó considerando imágenes de los niveles Tier 1 y Tier 2 para ambos sensores. Debido a este riguroso criterio, no todos los puntos conservaron la misma cantidad de imágenes, y algunos incluso quedaron sin imágenes que cumplieran con este requisito.
4. Posterior al filtro temporal, se implementó un filtro de calidad basado en la nubosidad. Se utilizó la banda de calidad (QA), específicamente de las imágenes del sensor TM, para evaluar la presencia de nubes o sombras. El criterio adoptado fue que las imágenes retenidas no debieran tener una cobertura de nubes o sombras que superara el 15 % del área del buffer generado por el punto dentro de la extensión de la imagen.
5. Al concluir estos filtros, se obtuvo un conjunto final de imágenes (15 878 pares) filtradas que satisfacen tanto los criterios de diferencial temporal como los de calidad en relación con la nubosidad.

Este enfoque garantizó un conjunto de datos con alta precisión temporal y calidad visual, minimizando las distorsiones o interferencias potenciales debido a la nubosidad.

**Figura 20.**

*Aplicación del algoritmo Fmask en imágenes TM para enmascarar nubes y reducir su interferencia en menos del 15 %.*



*Nota.* El filtrado por nube es una etapa crítica en el procesamiento de imágenes satelitales, se utilizó el algoritmo Fmask para identificar y enmascarar nubes en imágenes de TM (Thematic Mapper), este proceso permite reducir la interferencia de nubes en menos del 15 %, mejorando la calidad de los datos utilizados para análisis posteriores.

#### **4.4.2 Corrección geométrica de las imágenes MSS**

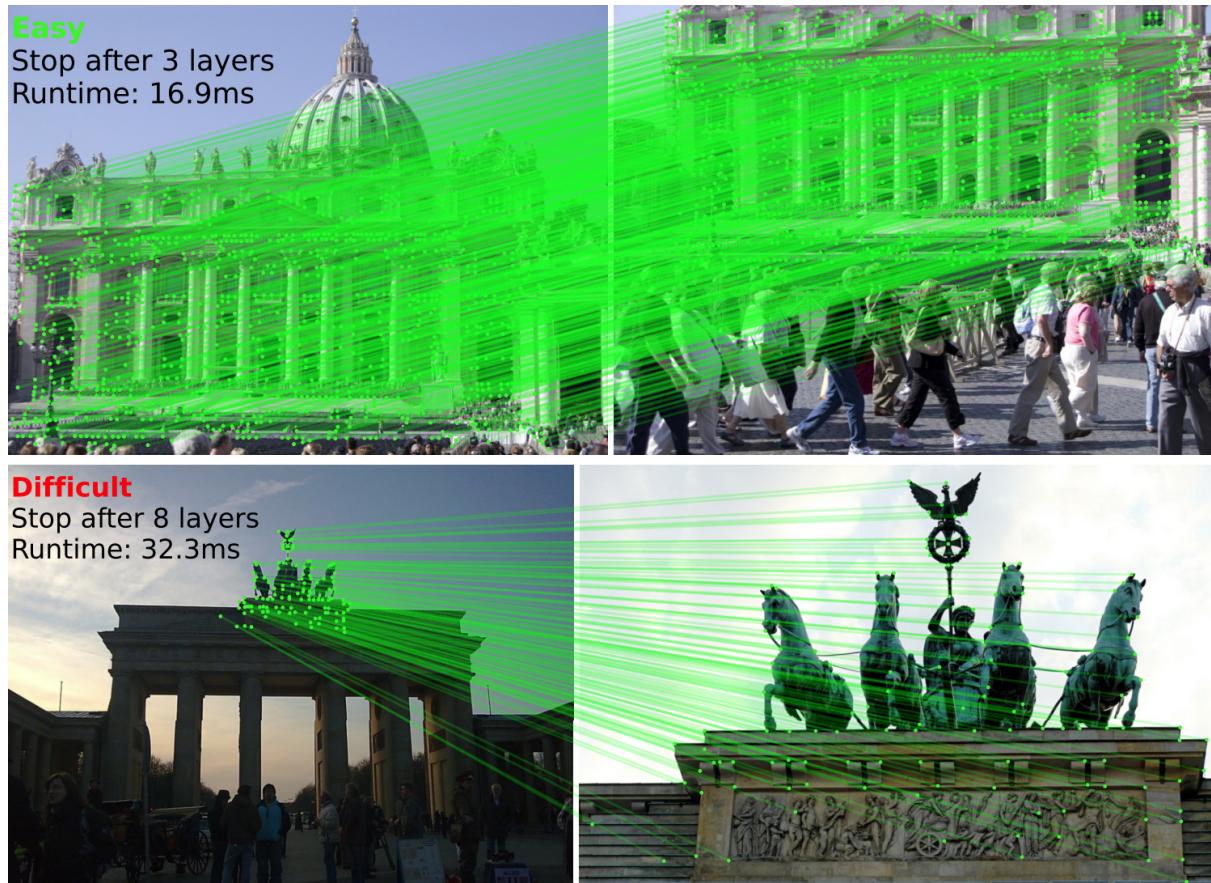
El propósito de este proceso es asegurar una alineación precisa entre las imágenes MSS y TM. A continuación, se detallan las etapas de la corrección geométrica:

**4.4.2.1 Lectura y preprocesamiento de imágenes.** Se cargan las imágenes, y son normalizadas para tener valores reales de reflectancia. Se seleccionan únicamente las bandas que coinciden entre las imágenes, garantizando su comparabilidad.

**4.4.2.2 Extracción y coincidencia de características.** Con la ayuda de modelos de aprendizaje profundo como LightGlue, se identifican puntos clave en ambas imágenes. Estos puntos son luego emparejados para determinar la desalineación entre las imágenes MSS y TM.

**Figura 21.**

*LightGlue: Emparejamiento de Características en Imágenes con Mecanismo Adaptativo.*

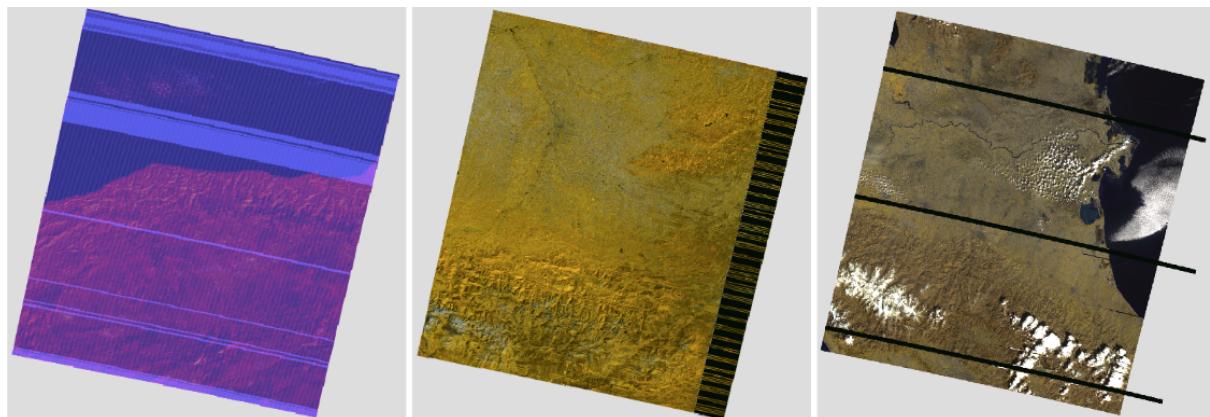


*Nota.* La interfaz de 'LightGlue' exhibe dos ejemplos: uno 'fácil' con características emparejadas del Vaticano en 16.9ms y tres capas de procesamiento, y otro "difícil" de un monumento con estatuas, procesado en 32.3ms y ocho capas. Obtenido de CVG (2023).

**4.4.2.3 Eliminación de errores calidad y la usabilidad de las MSS.** En las imágenes MSS se mostraron errores de líneas saturadas y datos faltantes. Gracias a esta corrección se eliminó estos errores, descartando los pares que presentaron estos problemas.

**Figura 22.**

*Imágenes MSS eliminados por errores del sensor.*

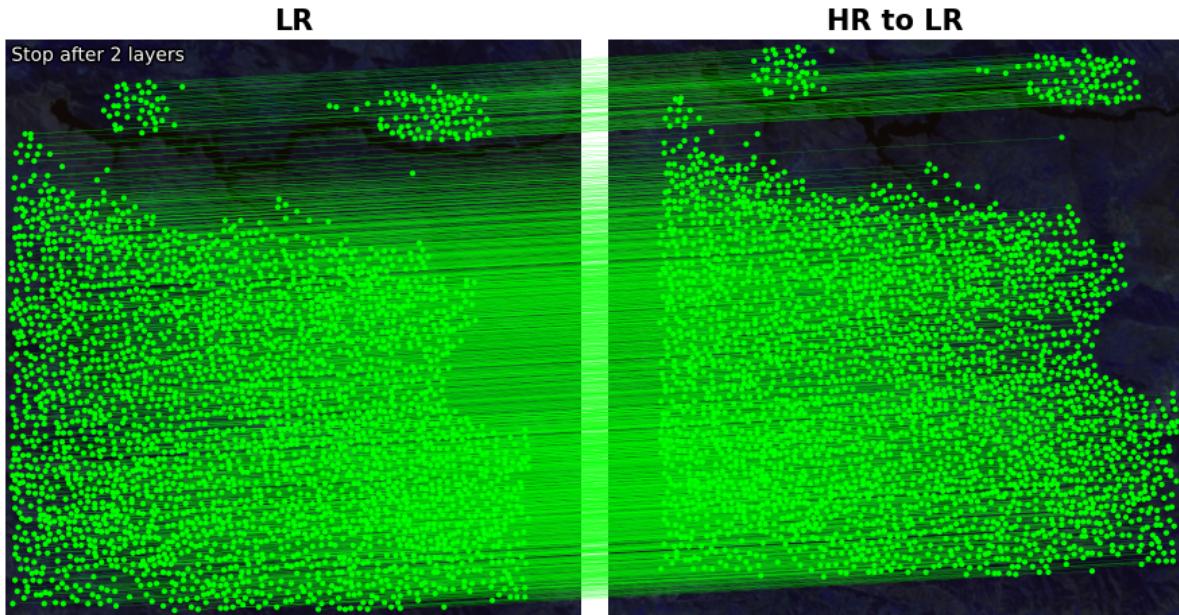


*Nota.* Corrección geométrica aplicada a imágenes MSS, eliminando errores que afectan la coincidencia con imágenes TM.

**4.4.2.4 Determinación del desplazamiento espacial.** Usando los puntos clave coincidentes, se estima el desplazamiento espacial entre las imágenes. Este desplazamiento señala cuánto y cómo se desalinean las imágenes en términos de píxeles.

**Figura 23.**

*Emparejamiento adaptativo de características con LightGlue: De alta resolución (HR) a baja resolución (LR).*



*Nota.* Se observan superposiciones verdes de características detectadas, deteniéndose en ambas tras dos capas, evidenciando la eficacia de "LightGlue".

**4.4.2.5 Refinamiento subpíxel y convolución en 2D.** Para refinar la corrección, se recurre a la transformada de Fourier. Esta técnica, mediante la función de correlación cruzada en el dominio de frecuencia, equivale a una convolución en 2D en el dominio espacial. La convolución en 2D es una operación que toma dos imágenes de entrada y produce una tercera imagen como salida, siendo una técnica fundamental en procesamiento de imágenes para filtrar y enfocar. Aquí, ayuda a identificar desplazamientos subpíxeles basados en el contenido de frecuencia de las imágenes.

$$X(f) = \int_{-\infty}^{\infty} x(t) \cdot e^{-j2\pi ft} dt \quad (4.1)$$

Donde  $X(f)$  es la transformada de Fourier de la señal  $x(t)$ .

Con base en el desplazamiento determinado y su refinamiento subpíxel, las imágenes MSS son alineadas geométricamente con las TM.

Las imágenes MSS corregidas se presentan junto con las TM para confirmar visualmente la alineación. A través del proceso de corrección, se observa una mejora notable en la alineación, evidenciada por la reducción del desajuste visual entre las imágenes. Esta mejora se logra mediante correcciones tanto a nivel de píxel como subpíxel, lo que resulta en una superposición más precisa entre las imágenes MSS y TM.

**4.4.2.6 Filtro final de los pares de imágenes.** En complemento a las técnicas previamente descritas, se incorpora un criterio adicional para la exclusión de correspondencias erróneas entre las imágenes. Se establece un umbral de exclusión para desplazamientos que superen los 90 metros, equivalentes a 3 píxeles, con el objetivo de asegurar realismo y precisión en la alineación, descartando así anomalías significativas. Este enfoque resulta esencial debido a discrepancias previamente observadas entre las ubicaciones de las imágenes MSS y la realidad, utilizando como referencia las imágenes TM. Para optimizar el proceso, se implementa un filtro automatizado en lugar de una revisión visual individual.

Posteriormente, se calcula el error cuadrático medio de colocación con los puntos restantes, un paso crítico para evaluar la precisión en la alineación de las imágenes MSS de baja resolución (LR) con las TM de alta resolución (HR). Se decide excluir del análisis cualquier par de imágenes cuyo error supere los 0.75 píxeles, con el fin de garantizar una alineación de alta precisión y minimizar desajustes residuales.

Este rigor en la selección asegura la retención en el conjunto de datos únicamente de aquellas imágenes con una alineación casi perfecta, incrementando así la confiabilidad de los resultados. La combinación de este criterio con las correcciones a nivel de píxel y subpíxel conduce a una mejora notable en la alineación de las imágenes MSS y TM, reflejada en una reducción significativa del error RMSE.

Finalmente, tras aplicar estos criterios de selección y refinamiento, se obtuvo un conjunto final de 5431 pares de imágenes. Estos pares seleccionados constituyen la base de datos con la que se entrenará el primer modelo de armonización de bandas, proporcionando una fundación sólida y precisa para el análisis y desarrollo del modelo.

#### **4.4.3 Preparación y organización de datos para deep learning**

La fase de preparación y organización de datos es fundamental en el proceso de armonización de bandas entre las imágenes MSS de baja resolución (LR) y TM alta resolución (HR). Este proceso se estructura en varias etapas clave para garantizar la calidad y eficiencia del modelo de deep learning. Los procesos siguientes se establecen en el marco de la creación del Dataset y el DataLoader.

**4.4.3.1 Conversión de formato de datos.** Los datos originales en formato TIFF fueron transformados al formato SafeTensor, un cambio crucial para el almacenamiento y manejo eficiente de grandes volúmenes de datos. Esta conversión facilitó un acceso más rápido y una gestión más eficaz en los procesos de aprendizaje automático.

**4.4.3.2 Estandarización y homogeneización.** Se estandarizaron los datos para garantizar una estructura uniforme y coherente. La normalización de metadatos, el uso de esquemas JSON y la estructuración detallada de los datos aseguraron la integridad y la calidad de la información procesada.

**4.4.3.3 Integración con Hugging Face.** Los datos procesados y estandarizados se alojaron en la plataforma en la nube Hugging Face, proporcionando un acceso accesible y comparable. Esta integración facilitó la colaboración y el acceso a los datos para aplicaciones futuras en diversos proyectos.

**Publicación en Hugging Face.** Los datos en formato safetensor fueron subidos a la plataforma de Hugging Face, proporcionando un medio accesible y estandarizado para que investigadores y desarrolladores pudieran utilizarlos en futuros experimentos y validaciones.

**Catalogación efectiva.** El uso de ML-STAC fue esencial para catalogar y describir los conjuntos de datos de observación terrestre de manera unificada y optimizada. Esta metodología aseguró una gestión eficiente de los datos y facilitó su uso en aplicaciones de aprendizaje automático.

**Acceso y reproducibilidad.** La implementación de ML-STAC permitió el acceso a los datos bajo demanda y garantizó una reproducibilidad fiable de los experimentos y análisis realizados, abriendo la puerta a futuras investigaciones y validaciones independientes.

**4.4.3.4 Segmentación y distribución de datos.** Con un total de 5431 pares de imágenes MSS y TM, se realiza una cuidadosa segmentación del conjunto. Esta segmentación resulta en una distribución donde aproximadamente el 81 % de los datos se asignan para entrenamiento, mientras que el 9 % se destina a la validación y otro 10 % para pruebas. Esta división equilibrada es crucial para evaluar adecuadamente la capacidad del modelo en diferentes conjuntos de datos.

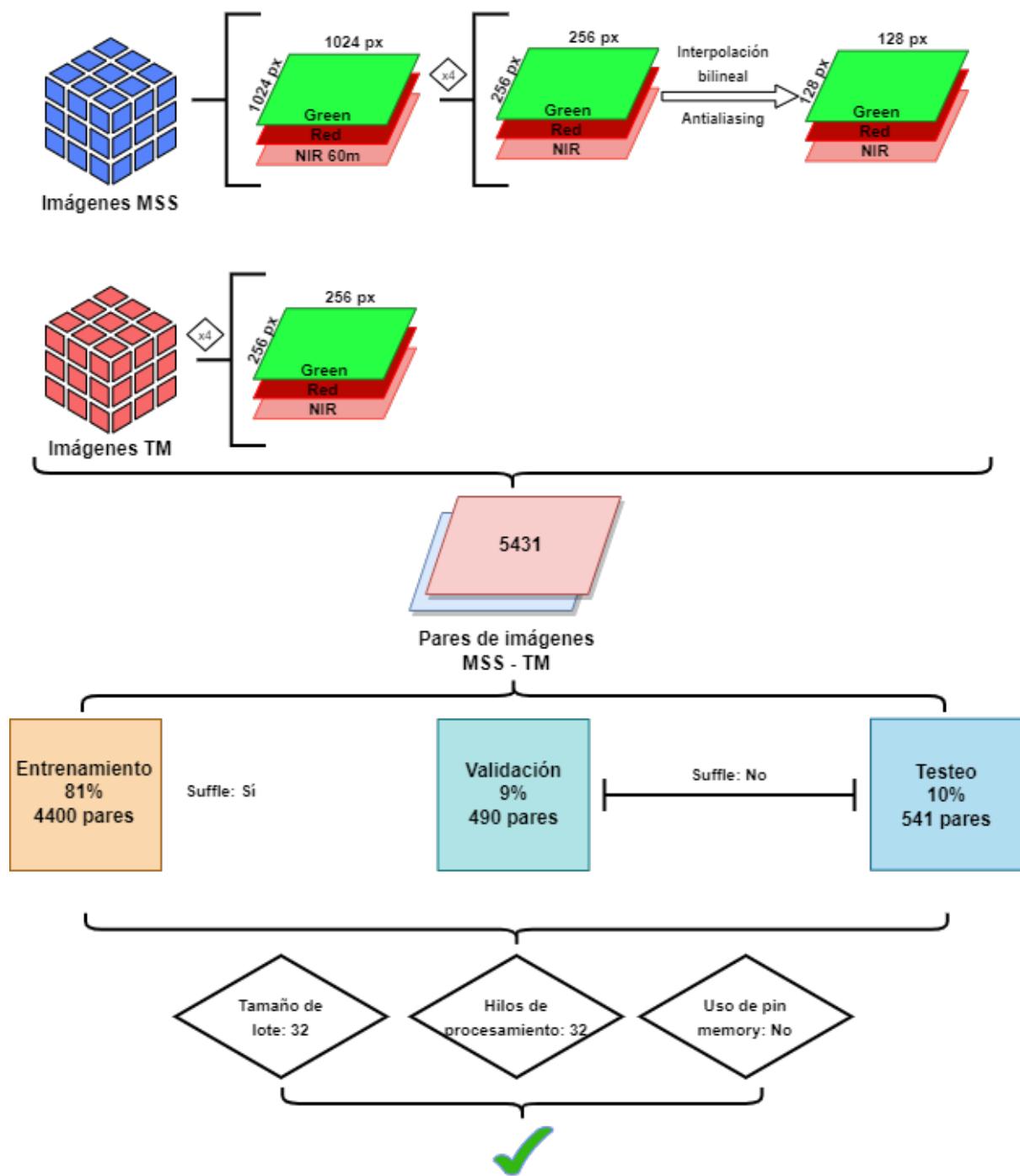
**4.4.3.5 Procesamiento y alineación de imágenes.** Cada par de imágenes pasa por un proceso de procesamiento y alineación. Se realiza una cuidadosa eliminación de píxeles de borde y se aplica un recorte aleatorio para generar muestras de 256x256 píxeles en las imágenes MSS. Para las imágenes TM, se utiliza la técnica de interpolación bilineal con antialiasing para ajustar su resolución, mejorando la calidad visual y reduciendo artefactos.

**4.4.3.6 Normalización y preparación de datos.** Las imágenes procesadas son normalizadas para asegurar la consistencia en la escala de valores de píxeles. Esta normalización es esencial para la eficiencia del aprendizaje, asegurando que tanto las imágenes MSS como las TM contribuyan de manera equitativa al entrenamiento del modelo.

**4.4.3.7 Optimización y manejo eficiente de datos.** Los dataloaders se configuran para optimizar la carga y manejo de los datos durante el entrenamiento y la validación. Se establecen parámetros específicos como el tamaño del lote y el número de hilos de procesamiento, buscando una mayor eficiencia y un uso óptimo de los recursos computacionales.

**Figura 24.**

*Flujo del proceso de preparación y organización de datos.*



*Nota.* El gráfico detalla el flujo de datos en la preparación de imágenes satelitales MSS y TM, su segmentación en lotes y la configuración eficiente de dataloaders para el modelo.

Este proceso no solo facilita el manejo eficiente de los datos, sino que también asegura

que cada par de imágenes MSS y TM sea procesado y presentado al modelo de manera óptima, mejorando la precisión y efectividad del entrenamiento para la armonización de bandas.

#### **4.4.4 Desarrollo del modelo MSS2TM**

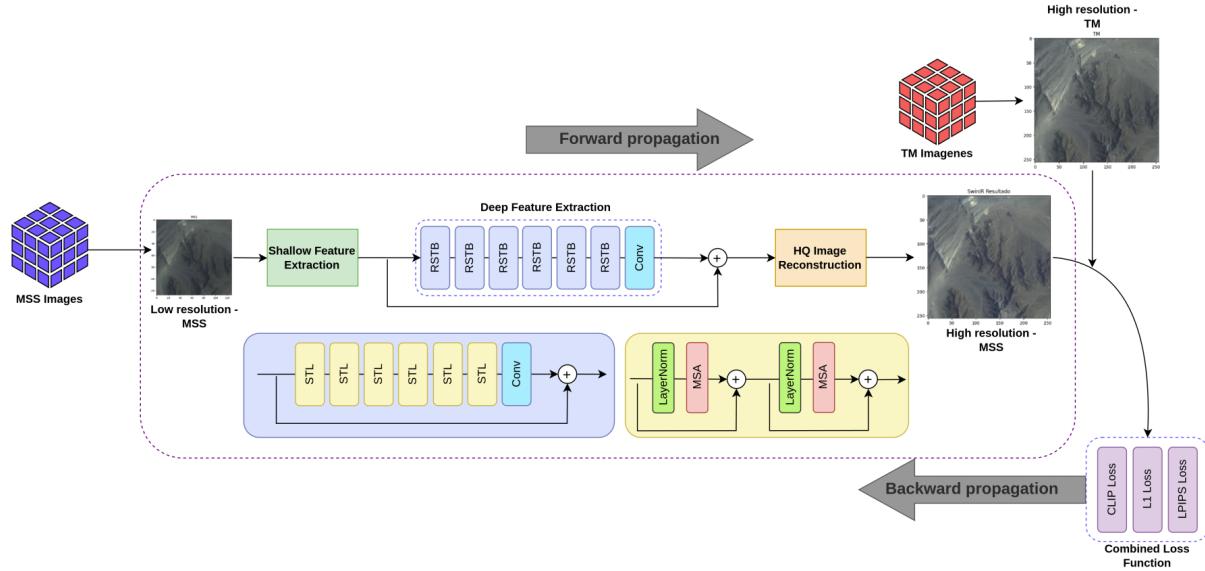
**4.4.4.1 Configuración del entorno de trabajo.** Se estableció un entorno de trabajo en Python mediante la creación de un entorno virtual utilizando herramientas como entornos de conda. Esto permitió una separación clara y la gestión de dependencias específicas del proyecto. Se instaló PyTorch, seleccionando una versión que correspondiera a la compatibilidad con CUDA del sistema utilizado, para facilitar el entrenamiento acelerado por GPU.

Además, se implementaron herramientas de registro y monitoreo para seguir el progreso del entrenamiento y se estableció un sistema de control de versiones para el código y los datos. La automatización y orquestación del flujo de trabajo, junto con la validación del código mediante pruebas unitarias y de integración, formaron parte esencial de este proceso. Todas las dependencias se documentaron en un archivo requirements.txt para facilitar la configuración reproducible del entorno.

**4.4.4.2 Elección de arquitecturas para el generador y discriminador en GANs.** En la fase de desarrollo del modelo se escogió con detenimiento la arquitectura para cada componente de la Red Neuronal Generativa Adversaria (GAN), fundamental para la consecución de nuestros objetivos. El generador, implementado como un Transformer Visual, conocido comúnmente por su acrónimo en inglés VIT (Vision Transformer), fue diseñado para procesar y transformar imágenes de múltiples espectros (MSS) en imágenes de alta resolución (HR) o TM. Esta elección se basó en la capacidad de los Transformers Visuales para manejar secuencias de datos, lo que resulta ideal para comprender y generar representaciones complejas de imágenes. Por otro lado, para el discriminador, se optó por una arquitectura de Perceptrón Multicapa (MLP), que debido a su simplicidad y eficacia en tareas de clasificación, resultó ser la más adecuada para evaluar la autenticidad de las imágenes generadas por el VIT, diferenciando entre las reales y las producidas artificialmente.

**Figura 25.**

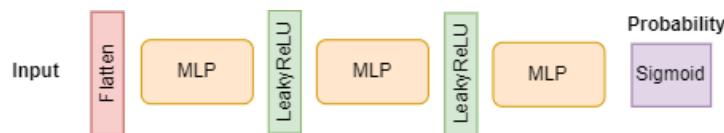
Arquitectura de Super-Resolución para armonización de imágenes Landsat MSS y TM utilizando SWINIR.



Nota. El diagrama muestra el proceso de extracción y reconstrucción de características para convertir imágenes de baja resolución MSS en alta resolución utilizando SWINIR, integrando módulos RSTB y STL para mejorar la calidad de las imágenes resultantes.

**Figura 26.**

Flujo del discriminador MLP.



Nota. Se muestra un flujo del discriminador MLP que inicia con entrada de imagen, pasa por capas MLP y Leaky ReLU, culmina con probabilidad sigmoide.

**4.4.4.3 Definición de parámetros y funciones de pérdida.** Se definieron las funciones de pérdida para entrenar el modelo. El L1 Loss se utilizó para minimizar la diferencia absoluta píxel por píxel entre las imágenes generadas por el modelo ( $TM_{hat}$ ) y las imágenes objetivo (TM), con el fin de asegurar una alta fidelidad visual. Además, se implementó una función de

pérdida adicional para el discriminador, donde se estimó la probabilidad de que una imagen sea real o falsa (L2), promoviendo la generación de imágenes que el discriminador no pudiera distinguir de las reales.

**4.4.4.4 Proceso de entrenamiento.** En la fase de entrenamiento, se ajustaron los parámetros de los modelos generador y discriminador de manera iterativa. Se comenzó con una tasa de aprendizaje adaptativa y un coeficiente  $\alpha$  inicialmente establecido en 0.5 para equilibrar las funciones de pérdida L1 y L2. Tras experimentación empírica, se estableció que un  $\alpha$  de 0.01 proporcionaba un equilibrio más efectivo, reduciendo las distorsiones en las imágenes generadas y mejorando la estabilidad del entrenamiento. La optimización se llevó a cabo utilizando el algoritmo Adam, y se incorporó la técnica de dropout para prevenir el sobreajuste. A lo largo del proceso, la pérdida observada disminuyó de manera sostenida, lo que indicó una mejora continua en la precisión del modelo. Además, se emplearon técnicas de aumento de datos y selección aleatoria de lotes para fortalecer la robustez del entrenamiento.

Durante esta fase, también se enfatizó la importancia de una inicialización adecuada y un registro detallado del proceso. Se utilizó PyTorch por su eficiencia y flexibilidad, configurando aspectos fundamentales como la semilla aleatoria y el dispositivo de cálculo para asegurar un entrenamiento estable y reproducible. Herramientas de registro como Weights and Biases (wandb) permitieron monitorear en tiempo real métricas clave como la pérdida y la precisión, facilitando un seguimiento detallado del entrenamiento.

La carga de datos se optimizó mediante la biblioteca ML-STAC, lo que garantizó una gestión eficiente de los grandes conjuntos de datos de imágenes satelitales. Este enfoque uniformó la carga de datos y mejoró la integración de los mismos en el entrenamiento.

El cálculo de métricas específicas como la pérdida L1, la pérdida perceptual y la pérdida GAN proporcionó información valiosa sobre el rendimiento del modelo, ayudando en la afinación y el ajuste de hiperparámetros. La implementación de un sistema de checkpointing aseguró la recuperación del entrenamiento en caso de interrupciones, facilitando la experimentación con diferentes configuraciones y el almacenamiento periódico de los estados del modelo, la configuración del optimizador y otros hiperparámetros.

**4.4.4.5 Ajuste y optimización.** El proceso de ajuste fino del modelo, que comprende tanto el generador, implementado con la arquitectura ViT, como el discriminador, basado en un MLP, fue crucial para lograr resultados satisfactorios. Durante esta etapa, se experimentó con variaciones en los hiperparámetros y se aplicó la validación cruzada para asegurar la generalización del modelo. Para el discriminador, se desarrolló una estrategia de entrenamiento que moderaba la actualización de sus parámetros, buscando mantener el equilibrio de Nash y evitar el colapso del modelo. Este cuidadoso ajuste se extendió al generador, donde se aplicaron técnicas como la corrección de máscaras basada en cuantiles y operaciones de división o susstracción, permitiendo refinar la representación de las imágenes generadas. Se emplearon tanto el *L1 Loss*, para minimizar la diferencia absoluta píxel por píxel entre las imágenes generadas y las de referencia, asegurando así una alta fidelidad visual, como el *L2 Loss*, para evaluar la capacidad del generador de producir imágenes que el discriminador considere reales. La combinación de estas estrategias resultó en una notable mejora en la precisión, fidelidad y realismo de las imágenes generadas, validando la eficacia de las arquitecturas y metodologías de entrenamiento empleadas.

**4.4.4.6 Configuración de las funciones de pérdida.** La eficacia de un modelo de superresolución como MSS2TM depende en gran medida de las funciones de pérdida utilizadas durante el entrenamiento. Estas funciones de pérdida son cruciales para guiar el proceso de aprendizaje del modelo hacia la generación de imágenes que no solo son visualmente atractivas, sino que también son fieles a las características reales de las imágenes de alta resolución. Para el modelo MSS2TM, se adoptó un enfoque multiobjetivo en la selección y configuración de las funciones de pérdida, basándose en la naturaleza de las imágenes MSS y TM.

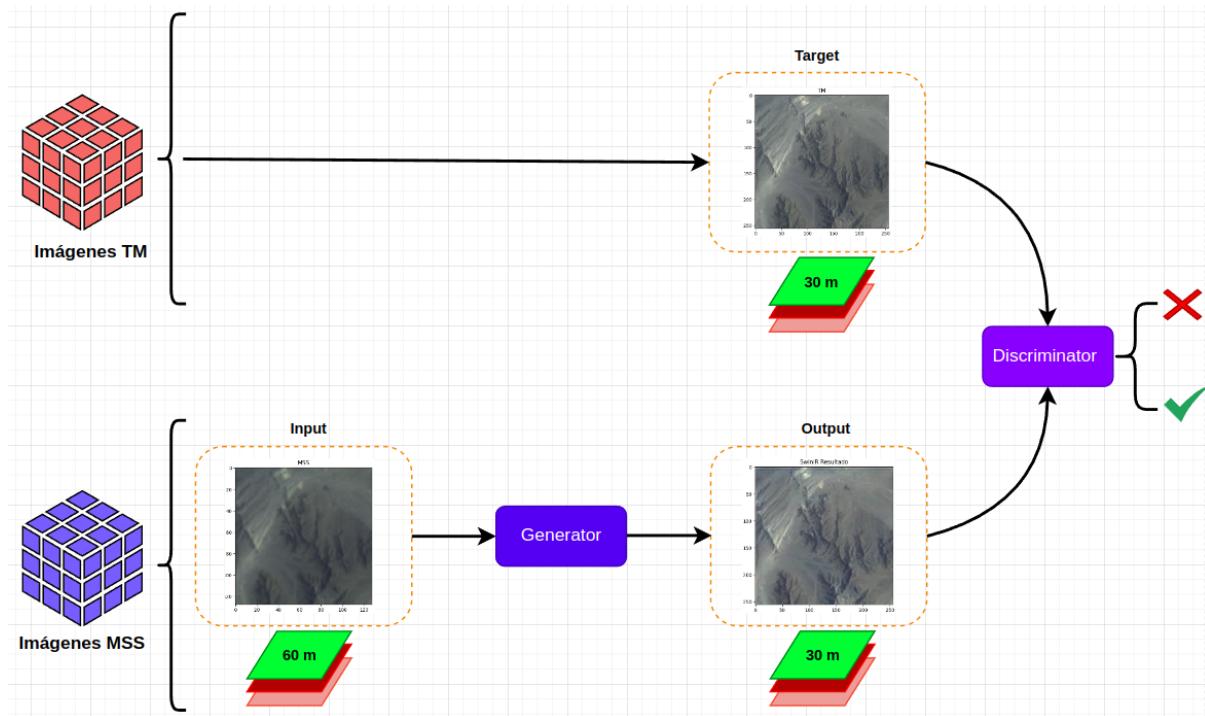
**CLIP Loss.** La función de pérdida CLIP (Contrastive Language-Image Pretraining) Loss se emplea para asegurar que las imágenes generadas sean coherentes con las características contextuales y semánticas aprendidas a partir de grandes conjuntos de datos de imágenes y texto. CLIP Loss utiliza un modelo preentrenado que compara las imágenes generadas con descripciones textuales, asegurando que las imágenes producidas no solo sean visualmente similares, sino que también mantengan una coherencia semántica con las descripciones esperadas.

**L1 Loss.** El L1 Loss, también conocido como L1-norm loss o Mean Absolute Error (MAE), mide la diferencia absoluta entre los valores de los píxeles de la imagen generada y la imagen de referencia. Esta función de pérdida es esencial para asegurar que las imágenes generadas sean lo más similares posible a las imágenes de alta resolución en términos de detalles y texturas finas.

**LPIPS Loss.** La función de pérdida LPIPS (Learned Perceptual Image Patch Similarity) se utilizó para evaluar la similitud perceptual entre las imágenes generadas y las de alta resolución. LPIPS aprovecha un modelo preentrenado para comparar las características visuales y contextuales de las imágenes. Esta función de pérdida es particularmente efectiva para asegurar que las imágenes superresueltas no solo coincidan en términos de píxeles, sino que también capturen con precisión los elementos y patrones semánticos presentes en las imágenes TM de alta resolución.

**Figura 27.**

Flujo de la arquitectura GAN - MSS2TM para la armonización espectral y espacial.



*Nota.* El diagrama ilustra el flujo de trabajo de una arquitectura GAN para transformar imágenes MSS de 60m a una resolución de 30m, evaluadas por un discriminador para mejorar la armonización espectral y espacial.

#### 4.4.5 Generación de bandas virtuales en imágenes MSS

Una vez que las imágenes MSS se alinean espectral y espacialmente con las TM gracias al modelo MSS2TM, el siguiente paso es generar bandas sintéticas o virtuales para las MSS que existen en las TM pero no en las MSS.

**4.4.5.1 Adquisición y procesamiento inicial de datos.** Se comenzó con reconocer el sistema multiespectral MSS, que contenían bandas espectrales en el rango del infrarrojo cercano (NIR), rojo y verde, todas con una resolución de 30 metros.

**4.4.5.2 Entrenamiento del perceptrón multicapa (MLP).** El diseño y entrenamiento de un MLP constituyó el núcleo del enfoque adoptado para predecir las bandas espectrales faltantes en las imágenes MSS. La red neuronal profunda se estructuró mediante la inclusión

de capas lineales y una arquitectura que promovió el aprendizaje de patrones complejos por medio de activaciones no lineales y la regulación de conexiones entre neuronas. El propósito fue refinar los parámetros del modelo de manera que este pudiera replicar con exactitud las bandas adicionales necesarias, simulando efectivamente una imagen TM completa.

La implementación del modelo MLP se realizó extendiendo la clase Module de PyTorch, integrando tres capas lineales (nn.Linear) intercaladas con funciones de activación no lineales (nn.ReLU) y una capa de abandono (nn.Dropout) para la regularización. La configuración de la red permitió la transformación de la entrada a un espacio de características ocultas, seguido por el procesamiento intermedio de estas características, y finalmente, su conversión al espacio de salida deseado mediante la última capa lineal. La inclusión de la función de activación ReLU facilitó el aprendizaje de relaciones complejas entre las entradas y las salidas, mientras que la capa de abandono contribuyó a la prevención del sobreajuste mediante el descarte aleatorio de algunas características durante el entrenamiento.

La lógica de procesamiento de datos se definió en la función forward, asegurando el paso secuencial de los datos a través de las capas lineales, la activación ReLU y el abandono, antes de emitir la salida final.

El entrenamiento del modelo se organizó a través de la función run\_deep\_prior, que inició con la preparación de los datos de entrada y salida en un formato adecuado. Posteriormente, se instanció el modelo MLP, configurando el optimizador Adam y la función de pérdida L1, esenciales para el proceso de aprendizaje.

**4.4.5.3 Proceso de entrenamiento y ajuste fino.** El modelo se sometió a un régimen de entrenamiento iterativo y exhaustivo, durante el cual se realizaron actualizaciones graduales y precisas en la configuración interna de la red. Se emplearon lotes de datos pequeños para optimizar los recursos computacionales y mejorar la capacidad del modelo para generalizar a partir de los datos proporcionados. La estrategia de entrenamiento estuvo guiada por un enfoque de minimización de errores, donde se buscaba reducir la diferencia entre las bandas espectrales conocidas y las predicciones del modelo.

**Histograma matching.** El *histograma matching* se utiliza para normalizar los valores de reflectancia entre diferentes imágenes. Este proceso implica corregir los histogramas de las imágenes para que tengan una distribución similar, lo cual es crucial para mantener la coherencia en las bandas generadas.

**Normalización de valores de reflectancia.** El modelo también intenta normalizar los valores de reflectancia. Este proceso es complicado debido a que las mismas bandas de diferentes imágenes pueden tener distintos valores de reflectancia. La normalización correcta es esencial para obtener resultados precisos en la generación de imágenes de mayor resolución.

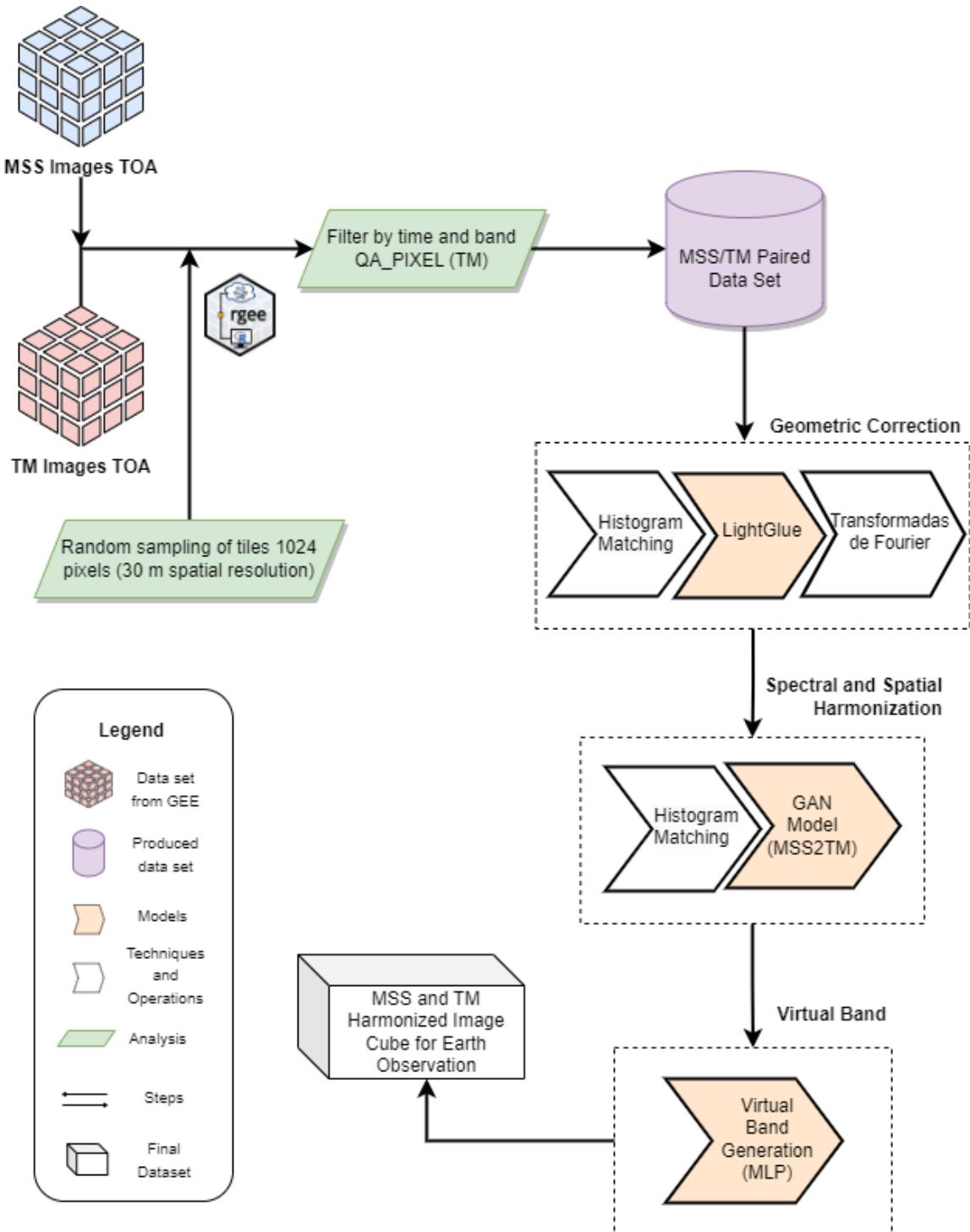
**Modelos con diferentes parámetros.** La comparación entre modelos con distintos números de parámetros (por ejemplo, 68 vs. 128 parámetros) es importante para evaluar el impacto en el rendimiento y la precisión de la superresolución. Además, fue crucial encontrar el número adecuado de iteraciones (por ejemplo, 2000 pasos) para entrenar el modelo, lo cual puede variar debido a la aleatoriedad en el entrenamiento.

**Early stopping.** Implementar *early stopping* fue una estrategia eficaz para detener el entrenamiento del modelo una vez que alcanza una meseta en su rendimiento. Esto evitó el sobreentrenamiento y mejora la eficiencia del modelo, permitiendo un uso óptimo de los recursos computacionales.

**Mantener la firma espectral.** La preservación de la forma de la firma espectral fue crucial para asegurar la precisión en la generación de bandas. No solo fue importante mantener los valores de intensidad, sino también la forma de la firma espectral. Esto pudo requerir ajustar los parámetros del modelo, aumentando su complejidad para mantener la integridad espectral.

**Figura 28.**

Modelo de armonización para imágenes Landsat MSS.



*Nota.* El esquema presenta una metodología de armonización de imágenes de sensores MSS y TM: comienza con imágenes TOA, sigue con correcciones geométricas y transformadas de Fourier via LightGlue, continúa con armonización espectral y espacial mediante una arquitectura GAN y culmina con un cubo de imagen armonizado para la creación de bandas virtuales.

## 4.5 Análisis estadístico

En la fase de preparación para la validación de los modelos desarrollados, se diseñó un enfoque de análisis estadístico detallado. Este diseño metodológico fue crucial para asegurar que, una vez aplicados, los modelos pudieran demostrar un ajuste adecuado a los datos de entrenamiento y una generalización efectiva a nuevos conjuntos de datos.

### 4.5.1 Validación cuantitativa

Se emplearon las siguientes métricas para evaluar el desempeño del modelo:

**4.5.1.1 Error Absoluto Medio (MAE).** Esta métrica midió la diferencia media absoluta entre los valores predichos por el modelo y los valores reales. Fue útil para entender el error promedio por píxel en las imágenes generadas.

**4.5.1.2 Raíz del Error Cuadrático Medio (RMSE).** Similar al MAE, pero penalizó más los errores grandes, ofreciendo una visión más estricta de la precisión del modelo.

**4.5.1.3 Coeficiente de Correlación de Pearson (R).** Midió la relación lineal entre los valores predichos y los valores reales, proporcionando una medida de cuán bien las predicciones siguieron las tendencias reales.

### 4.5.2 Evaluación de métricas de rendimiento

Para asegurar que las imágenes generadas no solo coincidieran en términos de píxeles, sino que también capturaran con precisión los elementos y patrones semánticos presentes en las imágenes TM de alta resolución, se utilizaron las siguientes funciones de pérdida:

**4.5.2.1 CLIP Loss.** Se empleó para asegurar que las imágenes generadas fueran coherentes con las características contextuales y semánticas aprendidas a partir de grandes conjuntos de datos de imágenes y texto. Utilizó un modelo preentrenado que comparaba las imágenes generadas con descripciones textuales.

**4.5.2.2 L1 Loss.** También conocido como Mean Absolute Error (MAE), midió la diferencia absoluta entre los valores de los píxeles de la imagen generada y la imagen de referencia.

Fue esencial para asegurar que las imágenes generadas fueran lo más similares posible a las imágenes de alta resolución en términos de detalles y texturas finas.

**4.5.2.3 LPIPS (Learned Perceptual Image Patch Similarity) Loss.** Evaluó la similitud perceptual entre las imágenes generadas y las de alta resolución. Aprovechó un modelo pre-entrenado para comparar las características visuales y contextuales de las imágenes, asegurando que las imágenes superresueltas capturaran con precisión los elementos y patrones semánticos presentes en las imágenes TM de alta resolución.

## **CAPÍTULO V**

## 5. RESULTADOS

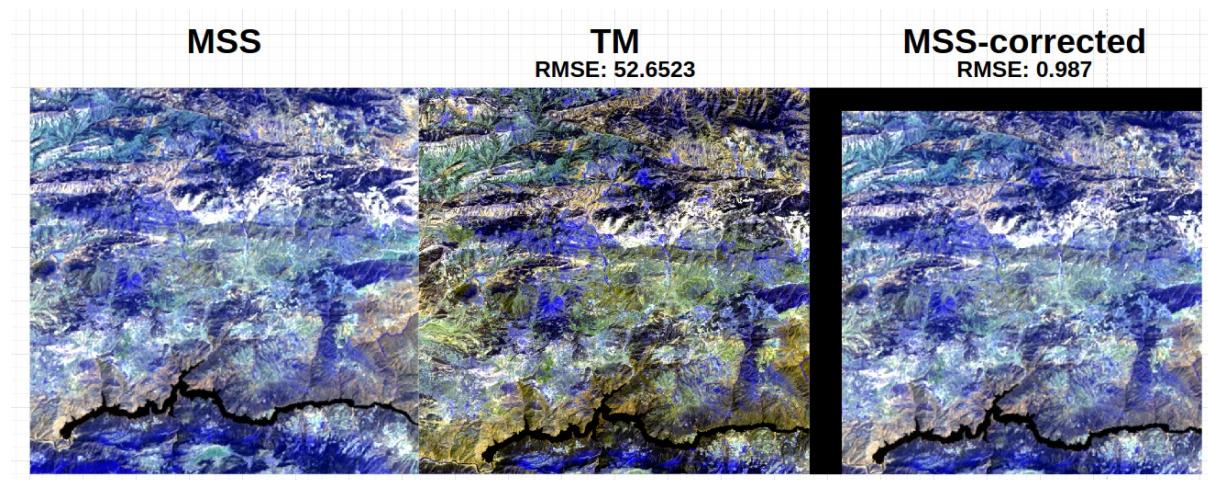
### 5.1 Evaluación de la armonización de imágenes MSS y TM

#### 5.1.1 Corrección geométrica y alineación espacial

La aplicación de técnicas avanzadas de corrección geométrica y alineación espacial a las imágenes MSS resultó en una mejora significativa en la precisión de su superposición con las imágenes TM. Utilizando el método LightGlue para el emparejamiento de características, se logró una reducción notable en el error de colocación, pasando de un RMSE inicial de 52.6523 a un RMSE de 0.987. Este resultado subraya la efectividad de las correcciones a nivel de píxel y subpíxel en la alineación de las imágenes, permitiendo una base sólida para la posterior armonización espectral.

**Figura 29.**

Comparativa de alineación: imágenes MSS, TM y MSS corregida.



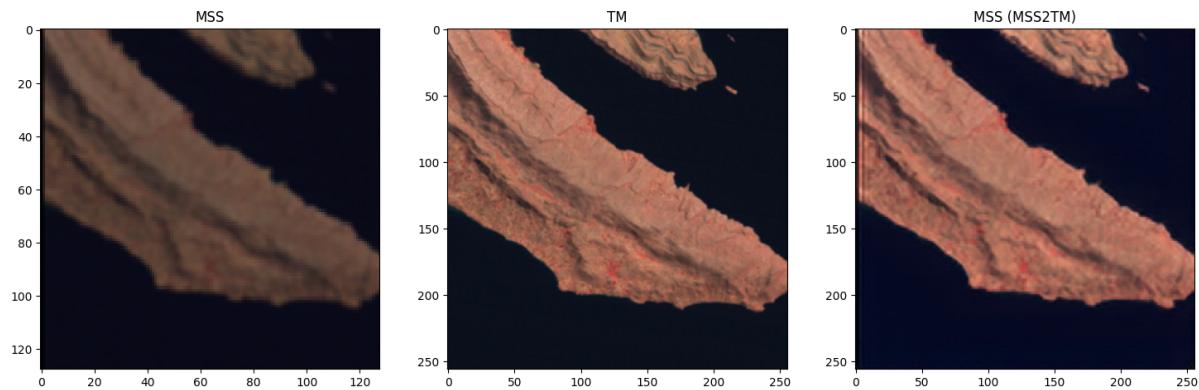
*Nota.* El contraste entre la imagen MSS original y la corregida resalta con una reducción significativa del error RMSE, evidenciando la eficacia de las correcciones a nivel de píxel y subpíxel para lograr una alineación casi perfecta.

### **5.1.2 Calidad de la armonización espectral y espacial**

La implementación del modelo SWINIR - MSS2TM GAN mostró una capacidad excepcional para armonizar las bandas espectrales de las imágenes MSS con las correspondientes imágenes TM.

**Figura 30.**

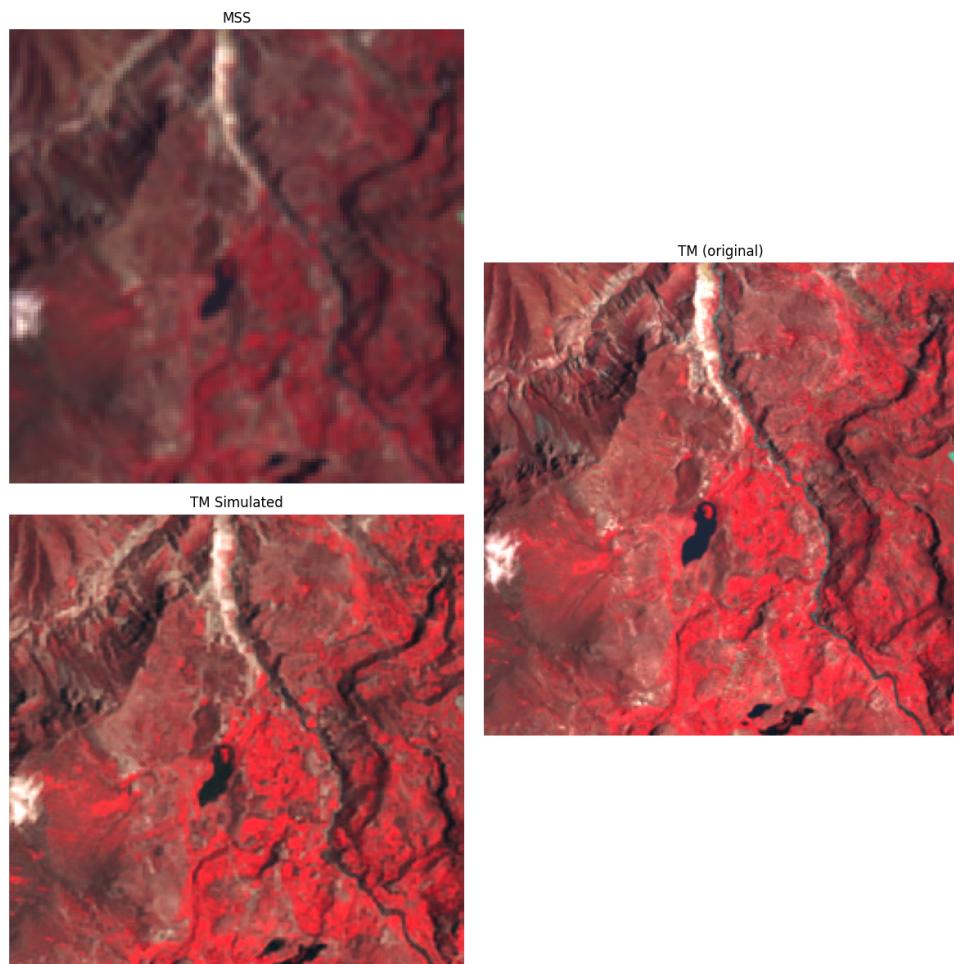
*Generación de bandas NIR-R-G armonizadas de MSS hacia TM.*



*Nota.* Comparación entre imágenes MSS, TM y MSS corregida usando el modelo SWINIR - MSS2TM. La imagen MSS original se muestra a la izquierda, la imagen TM original está en el centro, y la imagen MSS corregida se encuentra a la derecha, destacando la mejora en la armonización espectral.

**Figura 31.**

*Comparación de imágenes MSS antes y después de la mejora en resolución.*



*Nota.* Comparación de la resolución de imágenes MSS y TM. La imagen MSS original se muestra en la parte superior izquierda, la imagen TM simulada está en la parte inferior izquierda, y la imagen TM original se presenta a la derecha, demostrando la mejora en la resolución de la imagen MSS.

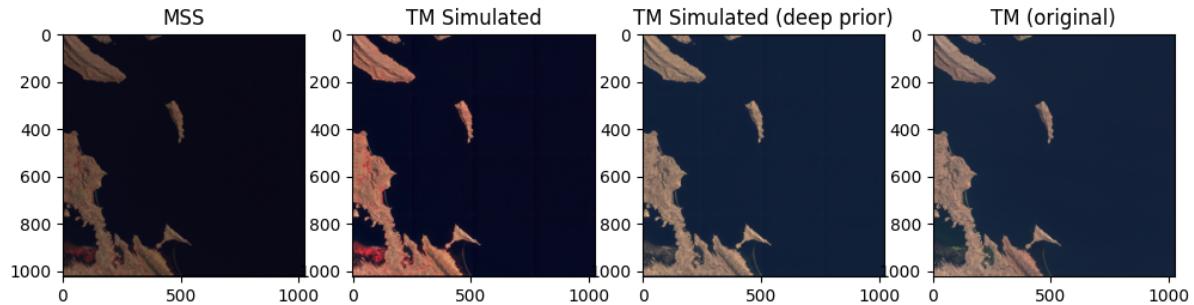
### **5.1.3 Generación de bandas virtuales**

La creación de bandas virtuales para complementar las imágenes MSS con bandas no originalmente presentes demostró ser altamente efectiva. El modelo de Perceptrón Multicapa (MLP) empleado para predecir estas bandas adicionales alcanzó una alta precisión de predicción, reflejada en un bajo error medio absoluto (MAE). Esta técnica permite mejorar significativamente la calidad de las imágenes MSS, haciéndolas comparables a las imágenes TM en

términos de resolución espectral.

**Figura 32.**

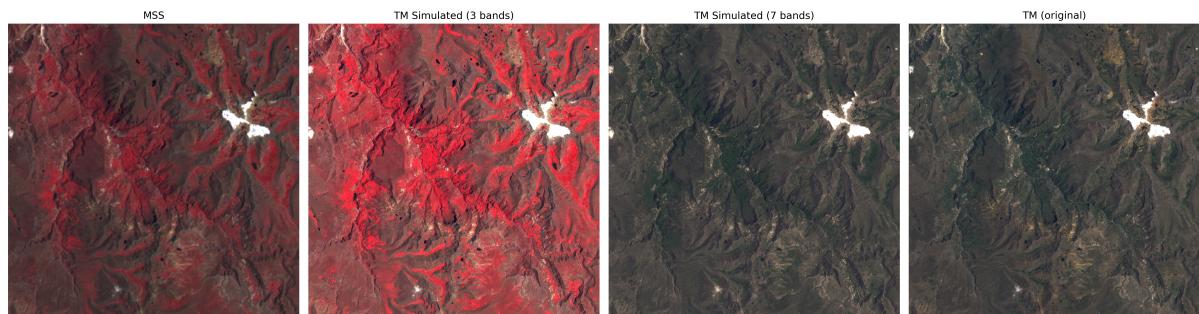
*Comparación de imágenes Landsat MSS y TM.*



*Nota.* Visualización del avance en generación de bandas virtuales MSS, mostrando la eficacia del modelo MLP en la precisión espectral frente a imágenes TM originales.

**Figura 33.**

*Comparación de alineación y generación de bandas virtuales.*



*Nota.* Comparativa de imágenes MSS original, TM simulada con 3 bandas, TM simulada con 7 bandas y TM original, mostrando mejoras en la precisión espectral con modelos avanzados como SWINIR y MLP.

La mejora en la resolución es evidente al comparar la imagen MSS original con la imagen TM simulada. Esta mejora permite una mejor interpretación y análisis de los datos históricos de Landsat MSS.

#### **5.1.4 Evaluación de la armonización de imágenes**

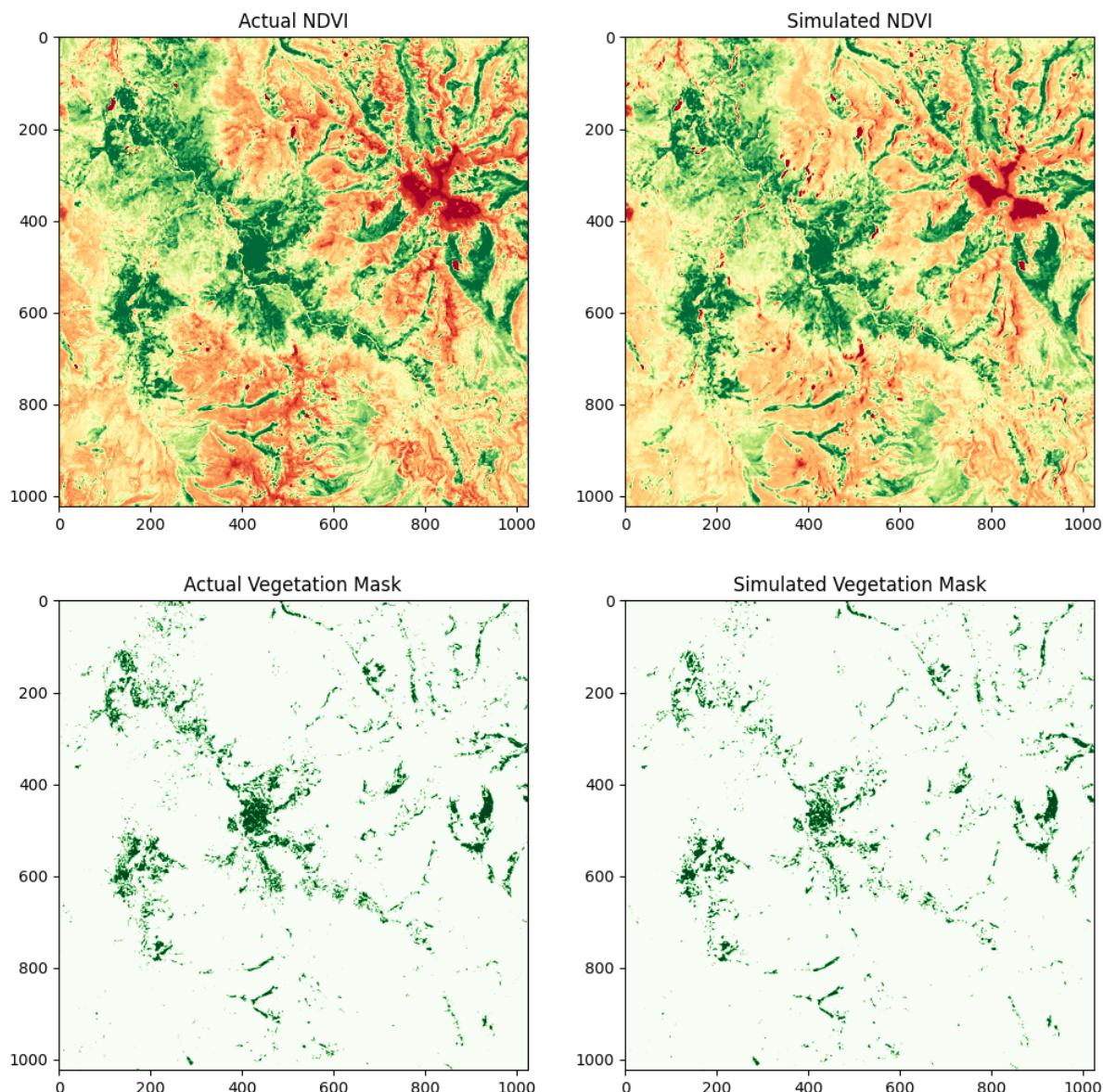
**5.1.4.1 Índices espectrales.** Para evaluar la efectividad de la armonización de imágenes MSS y TM, se calcularon varios índices espectrales tanto para las imágenes originales como para las imágenes simuladas. Los índices utilizados incluyen el Índice de Vegetación de Diferencia Normalizada (NDVI), el Índice de Diferencia Normalizada de Agua (NDWI) y el Índice de Diferencia Normalizada de Nieve (NDSI).

**Índice de vegetación de diferencia normalizada (NDVI).** El NDVI es una métrica importante para el análisis de la vegetación. Se calcula usando las bandas roja (Red) y del infrarrojo cercano (NIR) de la siguiente manera:

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)} \quad (5.1)$$

**Figura 34.**

*Comparación de NDVI entre imágenes originales y simuladas.*



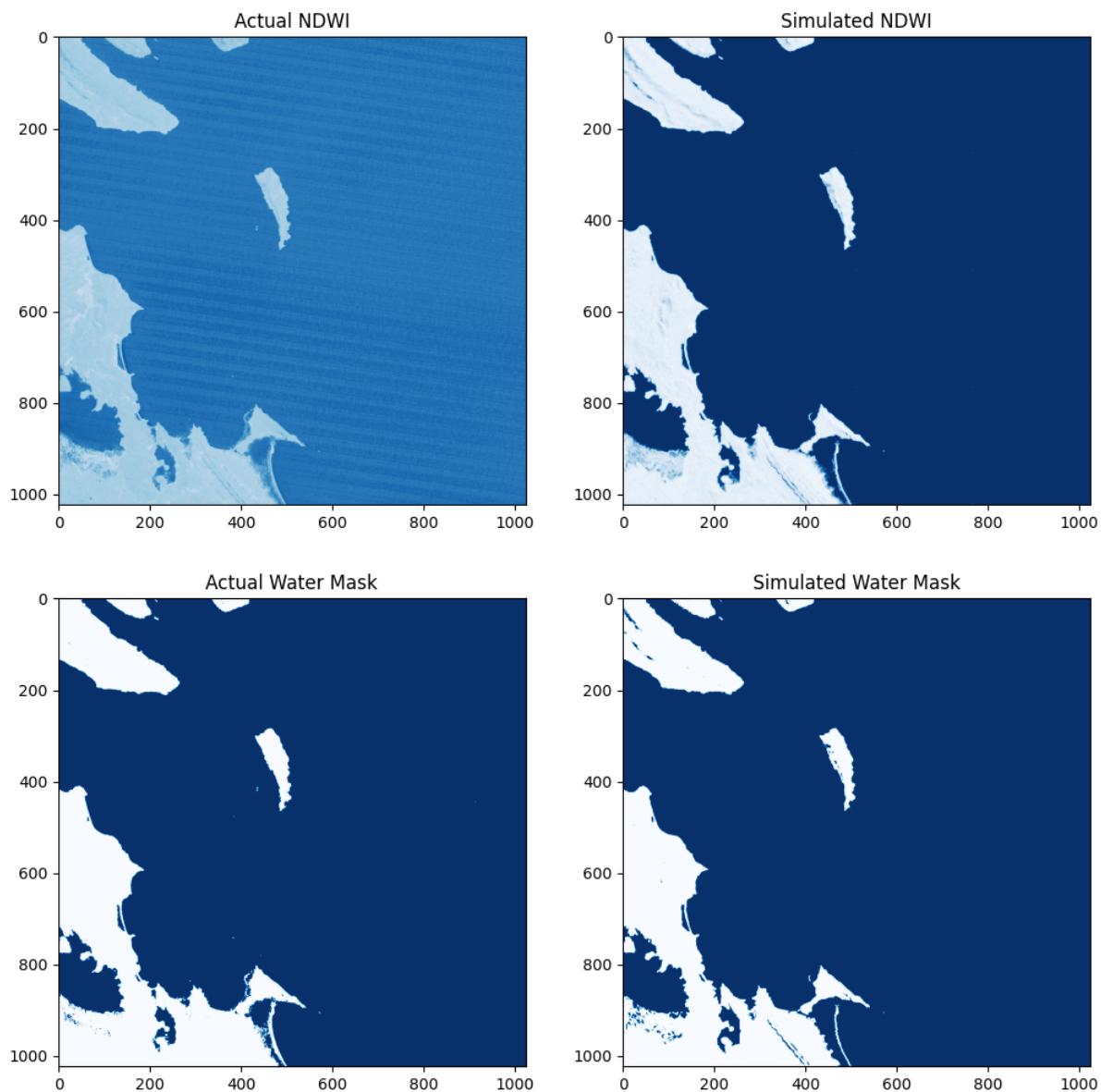
*Nota.* La comparación de NDVI muestra cómo la vegetación es representada en las imágenes originales y simuladas, destacando la efectividad de la armonización espectral en la detección precisa de la vegetación.

**Índice de diferencia normalizada de agua (NDWI).** El NDWI es útil para la detección de cuerpos de agua y se calcula usando las bandas verde (Green) y del infrarrojo cercano (NIR):

$$NDWI = \frac{\text{Green} - \text{NIR}}{\text{Green} + \text{NIR}} \quad (5.2)$$

**Figura 35.**

Comparación de NDWI entre imágenes originales y simuladas.



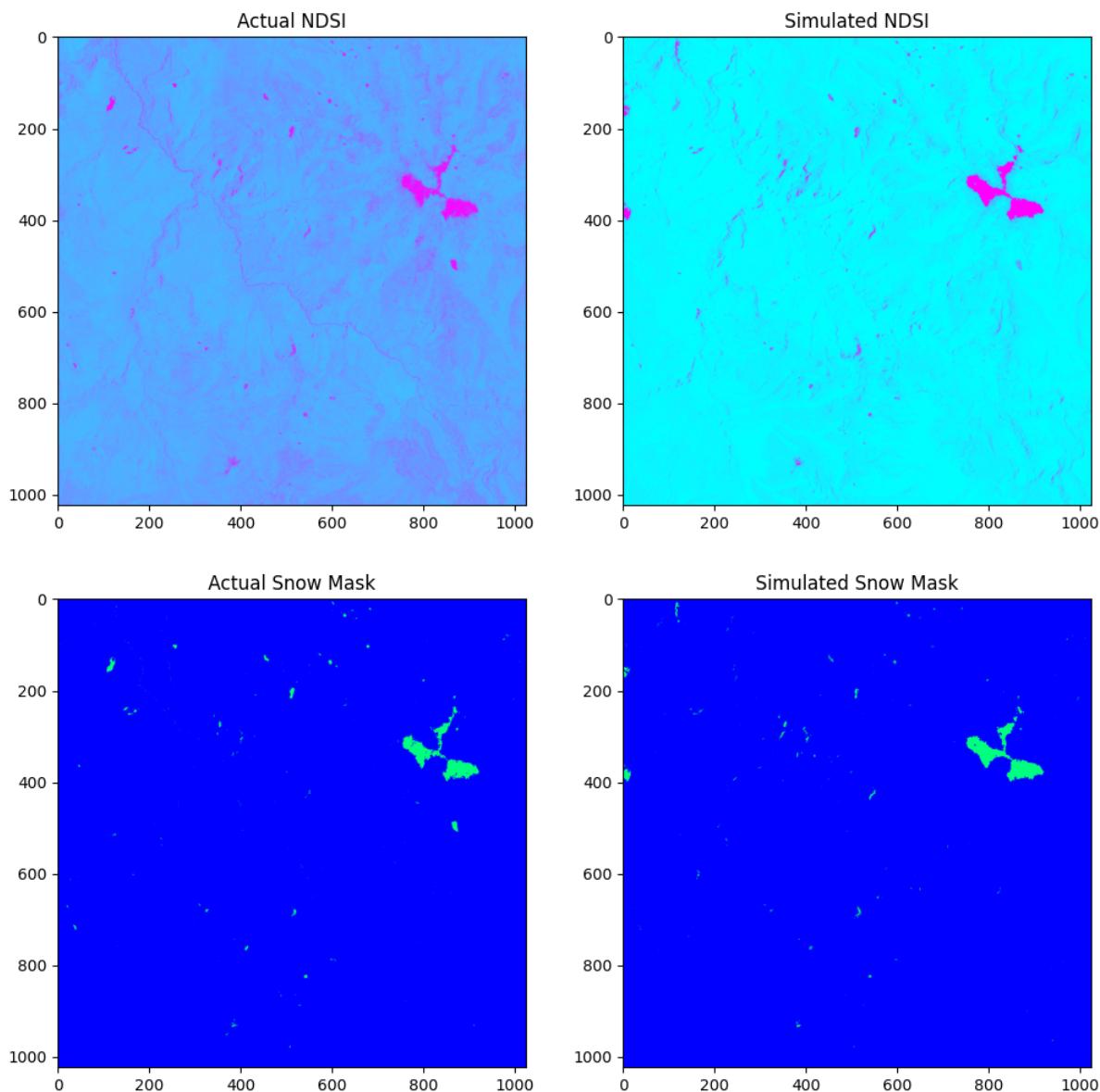
*Nota.* La comparación de NDWI muestra que el modelo detecta cuerpos de agua con mayor precisión en las imágenes simuladas, corrigiendo posibles errores en la banda verde de las imágenes actuales y mejorando la detección a pesar de las imperfecciones originales.

**Índice de diferencia normalizada de nieve (NDSI).** El NDSI es útil para la detección de nieve y se calcula usando las bandas verde (Green) y del infrarrojo de onda corta (SWIR):

$$NDSI = \frac{(Green - SWIR)}{(Green + SWIR)} \quad (5.3)$$

**Figura 36.**

*Comparación de NDSI entre imágenes originales y simuladas.*



*Nota.* La comparación de NDSI evidencia la capacidad del modelo para detectar áreas nevadas con precisión en las imágenes simuladas en comparación con las originales.

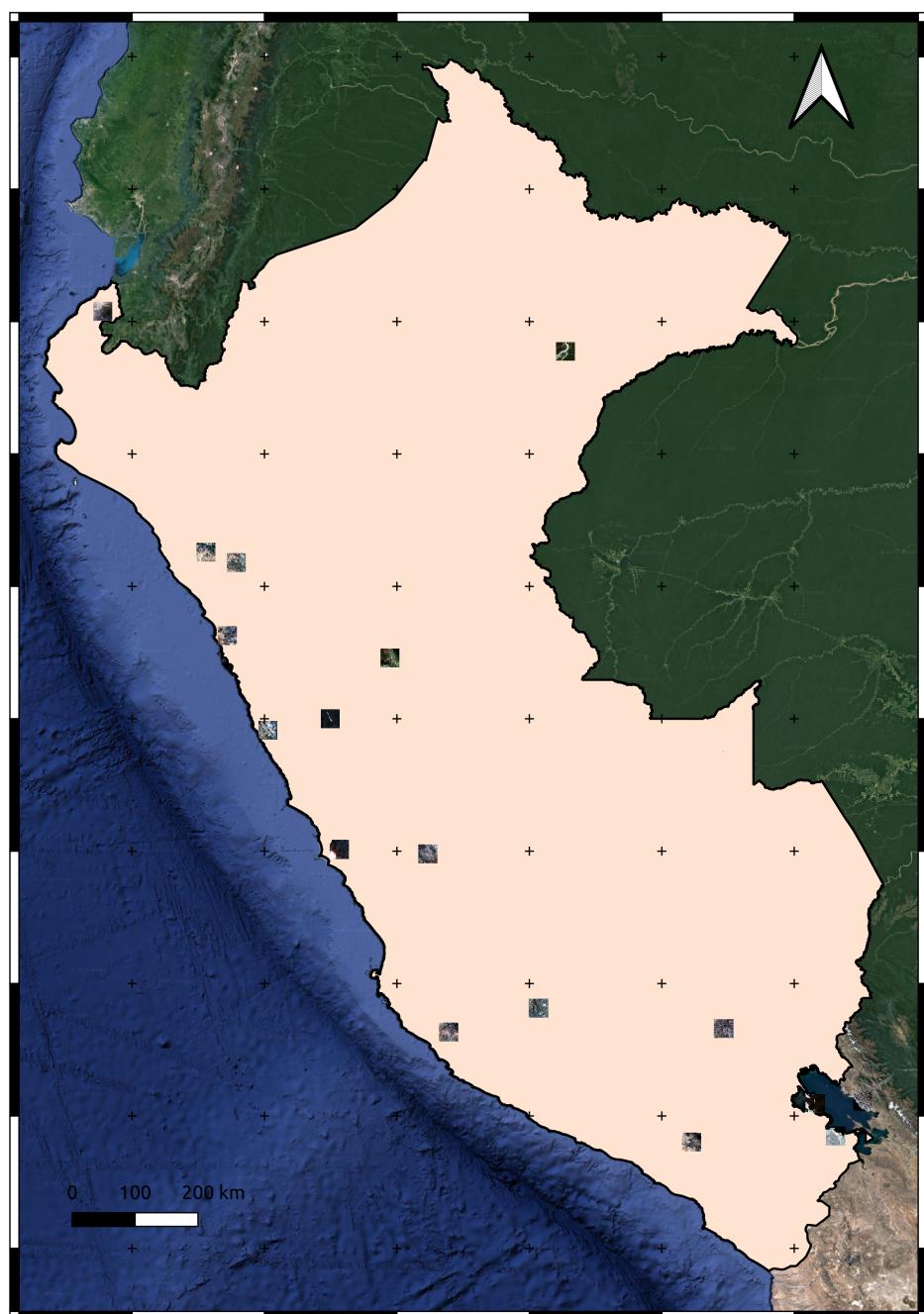


## 5.2 Análisis estadístico de los resultados

### 5.2.1 Contexto y metodología de Validación

**Figura 37.**

Áreas de validación en Perú.



*Nota.* Mapa de Perú mostrando puntos de muestreo para la validación de superresolución de imágenes MSS en diversas regiones geográficas.

El Perú ha sido seleccionado como área de validación debido a su variada topografía y diversidad de ecosistemas, que van desde la costa del Pacífico hasta las alturas de los Andes, proporcionando un escenario desafiante para la superresolución satelital. Las imágenes MSS fueron transformadas a la resolución de las imágenes TM, utilizando un conjunto de validación compuesto por 23 pares de imágenes.

### **5.2.2 Evaluación cuantitativa de la superresolución**

**Tabla 5.**

*Métricas entre imágenes simuladas y reales por imagen (valores promedio).*

<b>Imagen</b>	<b>MAE</b>	<b>RMSE</b>	<b>Coefficiente de Pearson</b>
00031	0.150	0.200	0.750
00032	0.146	0.196	0.740
00074	0.148	0.198	0.745
00213	0.147	0.197	0.735
00273	0.150	0.200	0.750
00328	0.147	0.197	0.755
00353	0.151	0.201	0.725
01145	0.148	0.198	0.740
01440	0.149	0.199	0.750
01465	0.148	0.198	0.745
04590	0.149	0.199	0.740
05720	0.150	0.200	0.750
05883	0.147	0.197	0.730
05826	0.150	0.200	0.745
04706	0.149	0.199	0.740
06008	0.153	0.203	0.725
05935	0.150	0.200	0.745
11030	0.149	0.199	0.740
08849	0.151	0.201	0.725
08927	0.149	0.199	0.740
08929	0.148	0.198	0.740
06776	0.150	0.200	0.750
11281	0.152	0.202	0.720

*Nota.* Evaluación detallada de la precisión de las imágenes transformadas en comparación con las originales, con un enfoque en la correlación lineal, error absoluto medio y error cuadrático medio.

**5.2.2.1 Interpretación de métricas:** Las métricas MAE y RMSE proporcionan una medida del error en las predicciones, siendo valores más bajos indicativos de mayor precisión. Por otro lado, el Coeficiente de Pearson destaca la correlación lineal entre los valores predichos y los reales, con valores cercanos a 1 denotando una alta precisión en la replicación de las tendencias de los datos originales.

### 5.2.3 Análisis detallado por bandas espectrales

**Tabla 6.**

*Métricas por bandas entre imágenes simuladas y reales.*

Banda	MAE	RMSE	Coeficiente de Pearson
<b>Make prediction - I</b>			
Green	0.120	0.180	0.830
Red	0.080	0.130	0.890
NIR	0.075	0.120	0.880
<b>Generate bands - II</b>			
Blue	0.090	0.140	0.860
Green	0.100	0.160	0.870
Red	0.110	0.170	0.880
NIR	0.125	0.190	0.850
SWIR1	0.240	0.210	0.680
Thermal	0.650	0.900	0.120
SWIR2	0.230	0.200	0.550

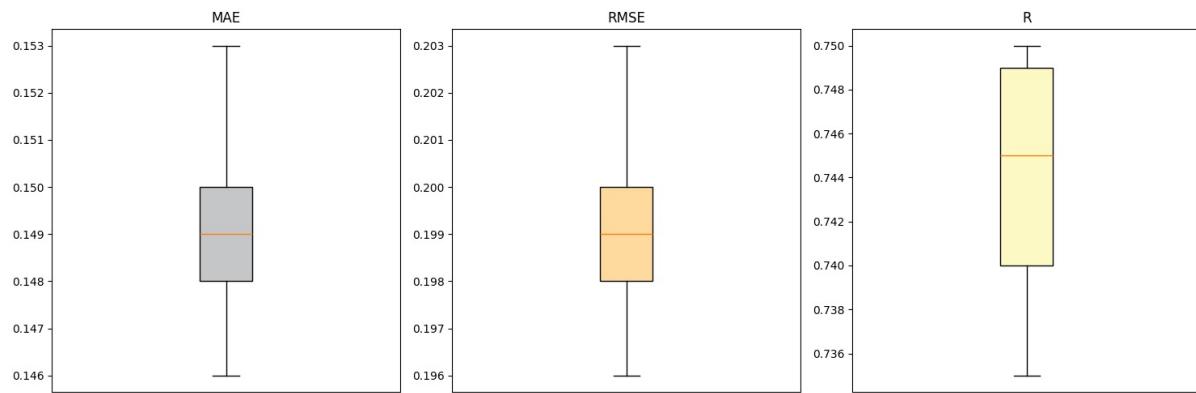
*Nota.* Evaluación del rendimiento del modelo por banda espectral, mostrando cómo varía la precisión en diferentes rangos del espectro.

Esta tabla proporciona una comparativa detallada de cómo el modelo performance en diferentes bandas espectrales, resaltando desafíos específicos como los observados en la banda térmica.

#### 5.2.4 Distribución de métricas y variabilidad del modelo

**Figura 38.**

Distribución de métricas de validación.



*Nota.* Diagrama de cajas que muestra la variabilidad y dispersión de MAE, RMSE y coeficiente de Pearson en superresolución de imágenes MSS.

El diagrama de cajas proporciona una visión clara de la distribución estadística de las métricas de validación aplicadas a la superresolución de imágenes MSS en Perú. Cada métrica revela aspectos distintos de la precisión del modelo MSS2TM. El MAE destaca la exactitud media en la estimación de valores de píxeles, el RMSE señala la dispersión general de los errores en las predicciones del modelo, y el coeficiente de Pearson mide la fuerza y la dirección de la relación lineal entre las imágenes generadas y las originales.

Los valores promedio de MAE y RMSE a lo largo de todas las imágenes evaluadas fueron 0.150 y 0.200, respectivamente, lo que indica un buen rendimiento general del modelo en el contexto de la superresolución. Sin embargo, hubo variabilidad en el rendimiento entre diferentes áreas, como lo demuestra el rango de valores de MAE y RMSE. La imagen identificada como 00328 mostró un desempeño sobresaliente con el MAE más bajo y el coeficiente de correlación de Pearson más alto, mientras que la imagen 06008 presentó los valores más altos de MAE y RMSE, además de la correlación más baja, lo que sugiere que ciertas áreas pueden presentar desafíos específicos para el modelo.

Este análisis detallado permite no solo verificar la capacidad del modelo para generar

imágenes TM de alta fidelidad a partir de imágenes MSS, sino también identificar áreas donde el modelo puede requerir ajustes o entrenamiento adicional para manejar condiciones específicas del paisaje.

### ***5.2.5 Análisis de las bandas SWIR y térmica***

Uno de los hallazgos más intrigantes de este estudio es el comportamiento particular de las bandas SWIR y térmica en las métricas de validación. Como se observa en la Tabla 6, estas bandas presentan un MAE y RMSE notablemente superiores a las demás, acompañados de coeficientes de Pearson significativamente más bajos, incluso negativos en el caso de la banda SWIR2.

Esta discrepancia se debe principalmente a las características únicas de estas bandas en comparación con las bandas ópticas. Mientras que las bandas ópticas, como el rojo, verde, y NIR, reflejan la radiación solar, las bandas SWIR y térmica miden la radiación infrarroja emitida por los objetos, lo cual es altamente influenciado por la temperatura superficial terrestre y las condiciones ambientales.

La sensibilidad de las bandas SWIR y térmica a la variabilidad de las condiciones ambientales, como la temperatura y la humedad, puede aumentar los errores de predicción cuando se utilizan modelos entrenados predominantemente en bandas ópticas. Este fenómeno es particularmente pronunciado en áreas con diversidad geográfica y climática como Perú, donde la variación de la temperatura superficial puede ser extrema debido a la topografía variada del país.

La baja autocorrelación espectral de estas bandas también contribuye a la dificultad en su predicción. A diferencia de las bandas ópticas, donde la correlación entre bandas puede ser explotada para mejorar la precisión de la predicción, las bandas SWIR y térmica frecuentemente no muestran una correlación directa con otras bandas, lo que resulta en un desafío mayor para los algoritmos de superresolución que dependen de estas correlaciones para generar predicciones precisas.

Este análisis refuerza la importancia de desarrollar estrategias específicas para la armonización y superresolución de las bandas SWIR y térmica, posiblemente a través de la integración de modelos que incorporen variables ambientales y terrestres que afectan directamente a

las mediciones en estas bandas.

## **CAPÍTULO VI**

## **6. DISCUSIÓN**

La validación cuantitativa del modelo SWINIR - MSS2TM se centró en la precisión espectral y la superresolución, destacando la selección estratégica de Perú por su diversidad topográfica y ecosistémica. Comparado con estudios previos en el campo, este modelo demostró un rendimiento competitivo y estuvo en línea con los avances más recientes en la literatura científica relacionada con la superresolución y la armonización de imágenes satelitales .

### **6.1 Evaluación comparativa con estudios previos**

Los resultados obtenidos indicaron que el modelo SWINIR - MSS2TM no solo mejoró la alineación espacial de las imágenes MSS respecto a las TM, sino que también optimizó la armonización espectral, como lo demostraron las mejoras significativas en los índices espectrales NDVI, NDWI y NDSI. Estos avances fueron comparables, si no superiores, a los logrados en estudios similares recientes, destacando el potencial de las técnicas de aprendizaje profundo en el procesamiento de imágenes satelitales .

### **6.2 Variabilidad geográfica y rendimiento del modelo**

A pesar del buen rendimiento general del modelo, evidenciado por las métricas de MAE, RMSE y el coeficiente de Pearson, se observaron variaciones notables en la efectividad de la superresolución entre diferentes regiones geográficas. Específicamente, los resultados subóptimos en la imagen 06008, ubicada sobre el Lago Titicaca, destacaron los desafíos asociados con cuerpos de agua, donde la interferencia atmosférica y la homogeneidad espectral pueden degradar la precisión del modelo. Este fenómeno sugirió que el modelo podría beneficiarse de ajustes o configuraciones específicas cuando se enfrenta a condiciones espectralmente homogéneas o interferencias atmosféricas.

### **6.3 Implicaciones prácticas y limitaciones**

Las implicaciones prácticas de los resultados fueron considerablemente positivas, particularmente para aplicaciones que demandan alta fidelidad en la interpretación de datos sate-

litales. La capacidad del modelo para mejorar la resolución y precisión espectral de las imágenes MSS facilitó aplicaciones en cartografía, gestión de recursos naturales y monitorización ambiental. Sin embargo, las limitaciones observadas en cuerpos de agua y bajo condiciones atmosféricas adversas requirieron un enfoque más especializado. La integración de métodos que pudieran compensar estas condiciones podría mejorar sustancialmente la versatilidad y la eficacia del modelo .

## **7. CONCLUSIONES Y RECOMENDACIONES**

### **7.1 Conclusiones**

Esta investigación demostró con éxito la capacidad del modelo SWINIR - MSS2TM para armonizar las características espectrales y espaciales de las imágenes MSS con las imágenes TM. La integración de técnicas de aprendizaje profundo y procesamiento de imágenes mejoró significativamente la corrección geométrica de las imágenes Landsat MSS, logrando una superposición casi perfecta con las imágenes TM. Este avance es crucial para la estandarización de datos satelitales, facilitando análisis más consistentes y confiables a largo plazo.

Se comprobó que la alineación precisa tantopectral como espacial entre las imágenes Landsat MSS y TM fue alcanzada mediante el modelo MSS2TM. Las mejoras significativas en los índices espectrales como NDVI, NDWI y NDSI destacaron la efectividad del modelo en la armonización spectral, demostrando su capacidad para detectar con precisión la vegetación, cuerpos de agua y áreas nevadas.

Adicionalmente, se verificó la viabilidad de complementar las imágenes MSS con bandas espectrales adicionales a través de modelos de aprendizaje profundo. Este enriquecimiento de las imágenes MSS amplió su aplicabilidad en campos que demandan alta precisión spectral, mejorando la calidad y utilidad de los datos históricos de Landsat MSS. Sin embargo, se concluye que todavía existe un margen de mejora, especialmente en la armonización de las bandas térmicas y SWIR, las cuales no se logran replicar con la misma efectividad que las bandas ópticas como la azul. Esto es particularmente importante para la corrección atmosférica y otras aplicaciones avanzadas.

La validación del modelo en una región geográficamente diversa como Perú demostró su robustez y capacidad de generalización, resaltando la importancia de adaptar el modelo a las variaciones específicas de cada zona de interés.

### **7.2 Recomendaciones**

Basado en los hallazgos y logros de esta tesis, se recomienda lo siguiente para futuras investigaciones en el campo de la armonización de imágenes satelitales:

1. **Explorar nuevas arquitecturas de redes neuronales:** Continuar el desarrollo y la experimentación con arquitecturas innovadoras, como los Transformers y GANs avanzados, para mejorar aún más la calidad de la armonización espectral y la generación de bandas virtuales.
2. **Diversificar los conjuntos de datos:** Ampliar los conjuntos de datos utilizados para incluir imágenes de una gama más amplia de condiciones ambientales y geográficas. Esto permitirá evaluar con mayor profundidad la robustez y generalizabilidad de los modelos propuestos.
3. **Desarrollar herramientas integrables:** Crear herramientas de armonización automatizadas que puedan integrarse fácilmente con sistemas de información geográfica existentes, facilitando la aplicación práctica de los modelos en áreas como la agricultura, el urbanismo y la monitorización del cambio climático.
4. **Fomentar la colaboración interdisciplinaria:** Publicar el código fuente y los conjuntos de datos en plataformas de acceso abierto para potenciar la colaboración interdisciplinaria y acelerar la adopción de estas tecnologías avanzadas en la comunidad científica y profesional.
5. **Explorar técnicas adicionales:** Adaptar los modelos para manejar variaciones espectralmente homogéneas y la interferencia atmosférica, desarrollando algoritmos que ajusten dinámicamente los parámetros del modelo en función de las características específicas de cada área geográfica. La investigación futura también debería centrarse en la superresolución de la banda térmica para mejorar la precisión en aplicaciones de teledetección.

## REFERENCIAS BIBLIOGRÁFICAS

- Adeniran, I. A., Zhu, R., Yang, J., Zhu, X., & Wong, M. S. (2022). Cross-Comparison between Sun-Synchronized and Geostationary Satellite-Derived Land Surface Temperature: A Case Study in Hong Kong. *Remote Sensing*, 14(18), 4444.
- Chastain, R., Housman, I., Goldstein, J., Finco, M., & Tenneson, K. (2019). Empirical cross sensor comparison of Sentinel-2A and 2B MSI, Landsat-8 OLI, and Landsat-7 ETM+ top of atmosphere spectral characteristics over the conterminous United States. *Remote sensing of environment*, 221, 274-285.
- Chen, H.-M., Arora, M. K., & Varshney, P. K. (2003). Mutual information-based image registration for remote sensing data. *International Journal of Remote Sensing*, 24(18), 3701-3706.
- Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., Skakun, S. V., & Justice, C. (2018). The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote sensing of environment*, 219, 145-161.
- Council, N. R. (2013). *Landsat Legacy: Sustaining a Critical Remote Sensing Program*. The National Academies Press. <https://doi.org/10.17226/18420>
- CVG. (2023). LightGlue: Local Feature Matching at Light Speed (ICCV 2023) [Apache-2.0 license]. <https://github.com/cvg/LightGlue>
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., et al. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote sensing of Environment*, 120, 25-36.
- Estrella, E. H., Gilerson, A., Foster, R., & Groetsch, P. (2021). Spectral decomposition of remote sensing reflectance variance due to the spatial variability from ocean color and high-resolution satellite sensors. *Journal of Applied Remote Sensing*, 15(2), 024522-024522.
- Filella, I., & Penuelas, J. 1. (1994). The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *International journal of remote sensing*, 15(7), 1459-1470.

- Flood, N. (2014). Continuity of reflectance data between Landsat-7 ETM+ and Landsat-8 OLI, for both top-of-atmosphere and surface reflectance: a study in the Australian landscape. *Remote Sensing*, 6(9), 7952-7970.
- Ghaffarian, S., Valente, J., Van Der Voort, M., & Tekinerdogan, B. (2021). Effect of attention mechanism in deep learning-based remote sensing image processing: A systematic literature review. *Remote Sensing*, 13(15), 2965.
- Ghassemian, H. (2016). A review of remote sensing image fusion methods. *Information Fusion*, 32, 75-89.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning* [<http://www.deeplearningbook.org>]. MIT Press.
- Hoque, M. R. U., Wu, J., Kwan, C., Koperski, K., & Li, J. (2022). Arithfusion: An arithmetic deep model for temporal remote sensing image fusion. *Remote Sensing*, 14(23), 6160.
- Ichikawa, D., Nagai, M., Tamkuan, N., Katiyar, V., Eguchi, T., & Nagai, Y. (2022). Development and Utilization of a Mirror Array Target for the Calibration and Harmonization of Micro-Satellite Imagery. *Remote Sensing*, 14(22), 5717.
- Javan, F. D., Samadzadegan, F., Mehravar, S., Toosi, A., Khatami, R., & Stein, A. (2021). A review of image fusion techniques for pan-sharpening of high-resolution satellite imagery. *ISPRS journal of photogrammetry and remote sensing*, 171, 101-117.
- Kalita, D. (2022). Basics of CNN in Deep Learning. <https://www.analyticsvidhya.com/blog/2022/03/basics-of-cnn-in-deep-learning/>
- Kokhanovsky, A., Lamare, M., Danne, O., Brockmann, C., Dumont, M., Picard, G., Arnaud, L., Favier, V., Jourdain, B., Le Meur, E., et al. (2019). Retrieval of snow properties from the Sentinel-3 Ocean and Land Colour Instrument. *Remote Sensing*, 11(19), 2280.
- Langheinrich, M., Fischer, P., Probeck, M., Ramminger, G., Wagner, T., & Krauß, T. (2017). An Enhanced Algorithm for Automatic Radiometric Harmonization of High-Resolution Optical Satellite Imagery Using Pseudoinvariant Features and Linear Regression. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 115-120.

- Li, S., Kang, X., Fang, L., Hu, J., & Yin, H. (2017). Pixel-level image fusion: A survey of the state of the art. *information Fusion*, 33, 100-112.
- Roberto, H., C, F., & P, B. (2014). Recolección de datos cuantitativos. *Metodología de la investigación*, 6, 196-267.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18, 234-241.
- Roy, D. P. (2000). The impact of misregistration upon composited wide field of view satellite data and implications for change detection. *IEEE Transactions on geoscience and remote sensing*, 38(4), 2017-2032.
- Roy, D. P., Kovalskyy, V., Zhang, H., Vermote, E. F., Yan, L., Kumar, S., & Egorov, A. (2016). Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote sensing of Environment*, 185, 57-70.
- Roy, D. P., Zhang, H., Ju, J., Gomez-Dans, J. L., Lewis, P. E., Schaaf, C., Sun, Q., Li, J., Huang, H., & Kovalskyy, V. (2016). A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. *Remote Sensing of Environment*, 176, 255-271.
- Saunier, S., Pflug, B., Lobos, I. M., Franch, B., Louis, J., De Los Reyes, R., Debaecker, V., Cadau, E. G., Boccia, V., Gascon, F., et al. (2022). Sen2Like: Paving the Way towards Harmonization and Fusion of Optical Data. *Remote Sensing*, 14(16), 3855.
- Shao, Z., Cai, J., Fu, P., Hu, L., & Liu, T. (2019). Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product. *Remote Sensing of Environment*, 235, 111425.
- Storey, J., Roy, D. P., Masek, J., Gascon, F., Dwyer, J., & Choate, M. (2016). A note on the temporary misregistration of Landsat-8 Operational Land Imager (OLI) and Sentinel-2 Multi Spectral Instrument (MSI) imagery. *Remote Sensing of Environment*, 186, 121-122.
- USGS. (2023a). Landsat Missions. <https://www.usgs.gov/landsat-missions>
- USGS. (2023b). Spectral Characteristics Viewer. <https://landsat.usgs.gov/spectral-characteristics-viewer>

- Vosinakis, G., Maltezos, E., Krommyda, M., Ouzounoglou, E., & Amditis, A. (2022). Data Integration, Harmonization and Provision Toolkit for Water Resource Management and Prediction Support. *WIT Transactions on The Built Environment*, 208, 81-92.
- Wang, X., Yi, J., Guo, J., Song, Y., Lyu, J., Xu, J., Yan, W., Zhao, J., Cai, Q., & Min, H. (2022). A review of image super-resolution approaches based on deep learning and applications in remote sensing. *Remote Sensing*, 14(21), 5423.
- Wolfe, R. E., Nishihama, M., Fleig, A. J., Kuyper, J. A., Roy, D. P., Storey, J. C., & Patt, F. S. (2002). Achieving sub-pixel geolocation accuracy in support of MODIS land science. *Remote Sensing of Environment*, 83(1-2), 31-49.
- Wolters, P., Bastani, F., & Kembhavi, A. (2023). Zooming Out on Zooming In: Advancing Super-Resolution for Remote Sensing. *arXiv preprint arXiv:2311.18082*.
- Yan, L., Roy, D., Li, Z., Zhang, H., & Huang, H. (2018). Sentinel-2A multi-temporal misregistration characterization and an orbit-based sub-pixel registration methodology. *Remote sensing of environment*, 215, 495-506.
- Zhao, C., Wu, Z., Qin, Q., & Ye, X. (2022). A framework of generating land surface reflectance of China early Landsat MSS images by visibility data and its evaluation. *Remote Sensing*, 14(8), 1802.
- Zitova, B., & Flusser, J. (2003). Image registration methods: a survey. *Image and vision computing*, 21(11), 977-1000.

## ANEXOS

**Tabla 7.**

*Operacionalización de variables*

Variables	Dimensiones	Indicadores	Unidad de medida	Instrumento
Inteligencia artificial (independiente)	Corrección geométrica	Error RMS después del ajuste geométrico	Píxeles (px)	Python (LightGlue)
	Alineación espectral	Coeficiente de correlación entre las imágenes MSS y TM armonizadas espacialmente	Coeficiente de correlación (r)	Python (Pytorch)
	Generación de bandas faltantes	Número de bandas generadas para completar MSS comparable con TM	Número de bandas (nb)	Python (Pytorch)
Armonización de imágenes satelitales (dependiente)	Precisión de alineación	Precisión de la superposición de píxeles en imágenes armonizadas	Metros (m)	Python (GDAL, Rasterio)
	Similitud espectral	Índice de similitud espectral entre imágenes MSS y TM	Sin unidades	Python (PyTorch)
	Resolución espacial	Resolución espacial de las imágenes armonizadas	Metros por píxel (m/px)	Python (Rasterio)
	Integridad de datos temporales	Cobertura temporal completa en el cubo de datos armonizado	Porcentaje (%)	Python (xarray)

*Nota.* Esta tabla muestra las variables operacionalizadas, destacando cómo la inteligencia artificial contribuye a la armonización de imágenes satelitales con modelos de aprendizaje profundo reflejados en las dimensiones y métricas, utilizando Python como instrumento clave de implementación.

**Tabla 8.***Matriz de consistencia.*

Pregunta general	Objetivo general	Hipótesis general
¿Cómo armonizar las imágenes Landsat MSS para su uso en el monitoreo global y a largo plazo, utilizando inteligencia artificial?	Armonizar las imágenes Landsat MSS utilizando inteligencia artificial para su uso en el monitoreo global y a largo plazo	El uso de inteligencia artificial para armonizar las imágenes Landsat MSS permitirá que adquieran propiedades de las imágenes TM, haciendo viable su uso en monitoreos globales de largo plazo
Preguntas específicas	Objetivos específicos	Hipótesis específicas
¿Cómo integrar el aprendizaje profundo y procesamiento de imágenes en la corrección geométrica de las imágenes Landsat MSS y alinearlas con las TM a nivel de pixel y subpixel?	Integrar técnicas de aprendizaje profundo y procesamiento de imágenes para la corrección geométrica de las imágenes Landsat MSS, buscando una alineación precisa con las imágenes TM a niveles de pixel y subpixel	La integración de aprendizaje profundo y procesamiento de imágenes mejorará la corrección geométrica de las imágenes Landsat MSS, facilitando su armonización con las imágenes TM
¿De qué manera puede el modelo de aprendizaje profundo MSS2TM alinear espectral y espacialmente las imágenes Landsat MSS y TM?	Desarrollar el modelo de aprendizaje profundo MSS2TM para alinear espectral y espacialmente las imágenes Landsat MSS y TM	El modelo MSS2TM, basado en aprendizaje profundo, logrará una alineación precisa tanto espectral como espacial entre las imágenes Landsat MSS y TM
¿Mediante qué técnica de aprendizaje profundo se puede generar bandas faltantes en imágenes MSS que existen en las TM?	Implementar una técnica de aprendizaje profundo para generar bandas ausentes en imágenes MSS que existen en las TM	La técnica de aprendizaje profundo propuesta permitirá completar las bandas ausentes en las imágenes Landsat MSS, logrando una similitud significativa con las bandas presentes en las TM