

**UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO**  
**FACULTAD DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA,**  
**INFORMÁTICA Y MECÁNICA**  
**ESCUELA PROFESIONAL DE INGENIERÍA INFORMÁTICA Y DE**  
**SISTEMAS**



**TESIS**

**MODELO PREDICTIVO DEL USO Y COBERTURA DEL SUELO  
EN LA PROVINCIA DEL CUSCO, INTEGRANDO  
APRENDIZAJE AUTOMÁTICO Y SIMULACIÓN ESPACIAL**

**PRESENTADO POR:**

Br. MARCELO FABIAN CHOQUE NAVARRO

**PARA OPTAR AL TÍTULO PROFESIONAL  
DE INGENIERO INFORMÁTICO Y DE  
SISTEMAS**

**ASESOR:**

Mgt. IVAN CESAR MEDRANO VALENCIA

**CUSCO - PERÚ**  
**2026**



# Universidad Nacional de San Antonio Abad del Cusco

## INFORME DE SIMILITUD

(Aprobado por Resolución Nro. CU-321-2025-UNSAAC)

El que suscribe, el Asesor Mgt. IVAN CESAR MEDRANO VALENCIA  
..... quien aplica el software de detección de similitud al  
trabajo de investigación/tesis titulada: MODELO PREDICTIVO DEL USO Y  
COBERTURA DEL SUELO EN LA PROVINCIA DEL CUSCO,  
INTEGRANDO APRENDIZAJE AUTOMÁTICO Y SIMULACIÓN ESPACIAL

Presentado por: MARCELO FABIAN CHOQUE NAVARRO DNI N° 70395931;  
presentado por: ..... DNI N°: .....  
Para optar el título Profesional/Grado Académico de INGENIERO INFORMÁTICO  
Y DE SISTEMAS

Informo que el trabajo de investigación ha sido sometido a revisión por 2 veces, mediante el  
Software de Similitud, conforme al Art. 6° del **Reglamento para Uso del Sistema Detección de**  
**Similitud en la UNSAAC** y de la evaluación de originalidad se tiene un porcentaje de 1 %.

### Evaluación y acciones del reporte de coincidencia para trabajos de investigación conducentes a grado académico o título profesional, tesis

Porcentaje	Evaluación y Acciones	Marque con una (X)
Del 1 al 10%	No sobrepasa el porcentaje aceptado de similitud.	<input checked="" type="checkbox"/>
Del 11 al 30 %	Devolver al usuario para las subsanaciones.	<input type="checkbox"/>
Mayor a 31%	El responsable de la revisión del documento emite un informe al inmediato jerárquico, conforme al reglamento, quien a su vez eleva el informe al Vicerrectorado de Investigación para que tome las acciones correspondientes; Sin perjuicio de las sanciones administrativas que correspondan de acuerdo a Ley.	<input type="checkbox"/>

Por tanto, en mi condición de Asesor, firmo el presente informe en señal de conformidad y adjunto las primeras páginas del reporte del Sistema de Detección de Similitud.

Cusco, 21 de ENERO de 2026

[Firma]  
Firma

Post firma IVAN CESAR MEDRANO VALENCIA

Nro. de DNI 73881501

ORCID del Asesor 0000-0003-0623-090X

### Se adjunta:

- Reporte generado por el Sistema Antiplagio.
- Enlace del Reporte Generado por el Sistema de Detección de Similitud: oid: 27259:547769691

# Marcelo Fabian Choque Navarro

## Modelo Predictivo Cusco.pdf

 Universidad Nacional San Antonio Abad del Cusco

### Detalles del documento

Identificador de la entrega

trn:oid:::27259:547769691

Fecha de entrega

20 ene 2026, 6:21 p.m. GMT-5

Fecha de descarga

20 ene 2026, 6:27 p.m. GMT-5

Nombre del archivo

Modelo Predictivo Cusco.pdf

Tamaño del archivo

8.4 MB

202 páginas

46.775 palabras

265.700 caracteres




# 1% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

## Filtrado desde el informe

- Bibliografía
- Texto citado
- Texto mencionado
- Coincidencias menores (menos de 20 palabras)

## Fuentes principales

- 0%  Fuentes de Internet
- 0%  Publicaciones
- 1%  Trabajos entregados (trabajos del estudiante)

## Marcas de integridad

### N.º de alertas de integridad para revisión

No se han detectado manipulaciones de texto sospechosas.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.

## Dedicatoria

*A mi mamá Irma y mi papá Marcelino  
por ser el pilar de mi vida.*

*Por su amor incondicional y por cada  
sacrificio que hicieron para que yo pudiera  
llegar hasta aquí.*

*A mi hermana Lucía y a mi hermano Jeshua,  
por su alegría y por recordarme siempre  
que el esfuerzo compartido es más ligero.*

*A mi tía Rosa,  
por su cariño invaluable y por ser un apoyo  
en cada etapa de mi formación.*

*A la memoria de mis abuelas, Lucía e Irene,  
por su legado y por permanecer presentes  
en nuestro recuerdo.*

## **Agradecimiento**

Deseo expresar mi gratitud a las personas e instituciones que hicieron posible el desarrollo y culminación de esta investigación.

Al Mgt. Ivan Medrano Valencia, asesor de esta tesis, por el espacio brindado y su confianza para la ejecución del proyecto. Al Dr. Javier Arturo Rozas Huacho, primer dictaminante, cuya orientación y sugerencias fueron determinantes para elevar la calidad de este trabajo. Asimismo, agradezco a la Mgt. Vanessa Maribel Choque Soto, segunda dictaminante, por sus precisas observaciones y aportes durante el proceso de revisión.

A las instituciones que facilitaron el acceso a la información primaria: al Instituto Nacional de Estadística e Informática (INEI), por la disposición de datos geoespaciales clave, y a la Municipalidad Distrital de San Sebastián, por facilitar la información catastral necesaria para el análisis local.

Gracias a mi familia, por su apoyo incondicional; a mis compañeros y amigos, por el aliento constante.

A mi casa de estudios, la Universidad Nacional de San Antonio Abad del Cusco, y a la Escuela Profesional de Ingeniería Informática y de Sistemas, por la formación profesional y las herramientas que me permitieron desarrollar este trabajo.

## Resumen

Las transformaciones en el Uso y Cobertura del Suelo (LULC) son un desafío ambiental de escala global; por lo que su análisis y predicción son necesarios para una planificación territorial sostenible y la mitigación de impactos ecológicos. La provincia del Cusco, ha experimentado un proceso de urbanización que ha transformado la cobertura natural del suelo. Este estudio tiene como objetivo analizar los cambios históricos en el LULC durante los periodos 2004-2014 y 2014-2024 mediante aprendizaje automático, utilizando plataformas de procesamiento en la nube. Se utilizaron imágenes satelitales de acceso abierto, adecuadas para el análisis multitemporal propuesto. La precisión de la clasificación se optimizó mediante el uso de variables derivadas, incluyendo índices espectrales y capas topográficas. Para la proyección de escenarios futuros, se aplicó un modelo híbrido de simulación que integra enfoques de dinámica espacial con técnicas de aprendizaje automático. Se validó el modelo mediante la proyección del año 2024 y comparándolo con datos reales. Para 2034, se prevé un incremento de las zonas urbanas en  $16.67 \text{ km}^2$ , pasando de representar el 9.48 % al 12.63 % del área total. La arquitectura híbrida propuesta presenta un desempeño predictivo consistente, respaldado por una Exactitud Global de 0.889,  $K_{histo} \approx 0.98$  y Análisis Multiescala (960 m)  $> 78 \%$ . Estos resultados indican la capacidad del modelo para capturar patrones no lineales complejos y evidencian su potencial para generar información para la gestión sostenible de los recursos naturales y el ordenamiento territorial en la provincia de Cusco.

*Palabras clave:* Aprendizaje Automático, Modelos Predictivos, Teledetección, Uso y Cobertura del Suelo.

## Abstract

Land Use and Land Cover (LULC) transformations represent a global environmental challenge; therefore, their analysis and prediction are essential for sustainable land-use planning and the mitigation of ecological impacts. The province of Cusco has undergone a significant urbanization process that has transformed its natural land cover. This study aims to analyze historical LULC changes during the 2004–2014 and 2014–2024 periods using machine learning techniques and cloud-based processing platforms. Open-access satellite imagery suitable for the proposed multitemporal analysis was employed. Classification accuracy was enhanced through the incorporation of derived variables, including spectral indices and topographic layers. To project future scenarios, a hybrid simulation model integrating spatial dynamics approaches with machine learning techniques was implemented. Model validation was performed by projecting conditions for 2024 and comparing the results with observed data. By 2034, an increase in urban areas of 16.67 km<sup>2</sup> is projected, rising from 9.48 % to 12.63 % of the total area. The proposed hybrid architecture demonstrates robust predictive performance, supported by an Overall Accuracy of 0.889,  $K_{histo} \approx 0.98$ , and a Multiscale Analysis (960 m) exceeding 78 %. These results indicate the model's ability to capture complex non-linear patterns and highlight its potential to support sustainable natural resource management and land-use planning in the province of Cusco.

*Keywords:* Machine Learning, Predictive Models, Remote Sensing, Land Use and Land Cover.



## Índice general

Dedicatoria . . . . .	I
Agradecimientos . . . . .	II
Resumen . . . . .	III
Abstract . . . . .	IV
Lista de Tablas . . . . .	IX
Lista de Figuras . . . . .	XI
Lista de Abreviaturas . . . . .	XV
I Introducción . . . . .	1
1.1 Generalidades . . . . .	1
1.2 Justificación . . . . .	3
1.3 Planteamiento del problema . . . . .	3
1.3.1 Formulación del problema . . . . .	5
1.3.2 Problema general . . . . .	5
1.3.3 Problemas específicos . . . . .	6
1.4 Alcances y Limitaciones . . . . .	6
1.4.1 Alcances del Estudio . . . . .	6
1.4.2 Limitaciones . . . . .	7
1.5 Objetivos . . . . .	8
1.5.1 Objetivo General . . . . .	8
1.5.2 Objetivos Específicos . . . . .	9
1.6 Antecedentes . . . . .	9
II Marco teórico . . . . .	16
2.1 La Luz y el Espectro Electromagnético . . . . .	16

2.1.1	Bandas Espectrales . . . . .	16
2.2	Sistemas de Referencia Espacial . . . . .	17
2.2.1	WGS 84 (Sistema Geodésico Mundial de 1984) . . . . .	17
2.2.2	Sistemas de Referencia Espacial en Perú (EPSG) . . . . .	18
2.3	Teledetección . . . . .	19
2.3.1	Sensores Satelitales . . . . .	20
2.3.2	Programas Satelitales . . . . .	21
2.3.3	Programa Landsat . . . . .	21
2.3.4	Programa Sentinel . . . . .	22
2.3.5	Escenas (Tiles) . . . . .	23
2.3.6	Niveles de Procesamiento de Datos Satelitales . . . . .	24
2.3.7	Geospatial Big Data (GBD) . . . . .	26
2.4	Sistemas de Información Geográfica . . . . .	26
2.4.1	Google Earth Engine (GEE) . . . . .	26
2.4.2	Editor de Código de Earth Engine . . . . .	27
2.4.3	QGIS . . . . .	27
2.4.4	Google Earth Pro . . . . .	28
2.4.5	Google Colab . . . . .	29
2.5	Tecnologías de Procesamiento Geoespacial y Científico . . . . .	30
2.5.1	Lenguajes de Programación . . . . .	30
2.5.2	Librerías de Análisis y Procesamiento Científico . . . . .	30
2.5.3	Librerías de Análisis Geoespacial y Aprendizaje Automático . . . . .	31
2.6	Uso y Cobertura del Suelo (LULC) . . . . .	31
2.6.1	Enmascaramiento y Eliminación de Nubes . . . . .	31
2.6.2	Remuestreo . . . . .	32
2.6.3	Índices Espectrales . . . . .	34
2.6.4	Variables Topográficas . . . . .	36
2.6.5	Análisis de Cambios Multitemporal . . . . .	37
2.7	Aprendizaje Automático . . . . .	39

2.7.1	Métodos de Muestreo . . . . .	40
2.7.2	Clasificación Supervisada de Coberturas del Suelo . . . . .	40
2.7.3	Modelos ML para Predicción de Transiciones LULC . . . . .	43
2.7.4	Métodos de Búsqueda de Hiperparámetros (HPO) . . . . .	45
2.8	Modelos Híbridos para Simulación Espacio - Temporal . . . . .	46
2.8.1	Modelo Híbrido CA-Markov . . . . .	46
2.8.2	Componente Temporal: Cadenas de Markov y Matrices de Transición .	47
2.8.3	Componente Espacial: Autómatas Celulares (CA) . . . . .	47
2.8.4	Variables Predictoras . . . . .	48
2.8.5	Análisis de Colinealidad . . . . .	49
2.9	Métricas de Evaluación de Modelos . . . . .	50
2.9.1	Métricas de Evaluación de la Clasificación . . . . .	50
2.9.2	Métricas de Evaluación de Simulación Espacio-Temporal . . . . .	54
2.10	Inteligencia Artificial Explicable (XAI) en la Modelización de Cobertura de Suelo	56
2.10.1	Análisis de Dominancia de Variables mediante Permutación . . . . .	56
2.10.2	Gráficos de Dependencia Parcial . . . . .	57
III	Metodología de la investigación . . . . .	58
3.1	Adquisición y Preprocesamiento de Datos Geoespaciales . . . . .	59
3.2	Clasificación Supervisada de Coberturas (LULC) . . . . .	60
3.3	Análisis de Cambios Multitemporal . . . . .	60
3.4	Simulación Híbrida (MLP-CA-MC) . . . . .	61
3.5	Proyección del Escenario Futuro 2034 . . . . .	63
IV	Desarrollo del proyecto . . . . .	64
4.1	Área de Estudio . . . . .	64
4.2	Datos Geoespaciales . . . . .	65
4.2.1	Delimitación Temporal . . . . .	65
4.2.2	Sensores Satelitales . . . . .	67
4.2.3	Preprocesamiento de Imágenes Satelitales . . . . .	68
4.2.4	Variables Auxiliares . . . . .	75

4.3	Clasificación Supervisada de Coberturas (LULC)	81
4.3.1	Definición de Clases de Cobertura para el Estudio	81
4.3.2	Recopilación de Datos de Entrenamiento y Validación	83
4.3.3	Optimización de Hiperparámetros y Entrenamiento del Modelo	86
4.3.4	Exportación de Mapas Clasificados y Análisis de Cambios	88
4.4	Análisis de Cambios Multitemporal	89
4.5	Simulación Híbrida (MLP-CA-MC)	90
4.5.1	Modelado de la Simulación Híbrida (MLP-CA-MC)	90
4.5.2	Validación de la Precisión del Modelo de Simulación	100
4.5.3	Interpretabilidad del Modelo (XAI)	103
4.6	Proyección del Escenario Futuro 2034	103
4.6.1	Fases Preparatorias	104
4.6.2	Proceso de Simulación Geoespacial 2024 - 2034	104
4.6.3	Generación del Mapa LULC Proyectado para 2034	105
V	Resultados y Discusión	107
5.1	Resultados	107
5.1.1	Desempeño de la Clasificación Supervisada	107
5.1.2	Cambios Espaciales y Temporales del LULC (2004-2024)	116
5.1.3	Desempeño y Validación del Modelo Predictivo MLP-CA-MC	119
5.1.4	Proyección del LULC para el Periodo 2024–2034	130
5.2	Discusión	135
5.2.1	Fiabilidad de la Clasificación y Confusión Espectral	135
5.2.2	Análisis de la Dinámica Temporal y Secuencialidad del Cambio	137
5.2.3	Validez del Modelo Predictivo MLP-CA-MC	138
5.2.4	Dinámica Territorial 2034	142
	Conclusiones	145
	Recomendaciones	146
	Referencias	147
	Apéndices	173

## Índice de tablas

1	Características principales del sistema EPSG:32719 . . . . .	19
2	Bandas Espectrales de Sensores Landsat. . . . .	22
3	Bandas espectrales del instrumento MSI de Sentinel-2 . . . . .	23
4	Índices Espectrales (IE) utilizados en el estudio. . . . .	35
5	Resumen de Imágenes Satelitales Seleccionadas por Año. . . . .	68
6	Fórmulas de Índices Espectrales por Sensor. . . . .	78
7	Comparativa de rangos dinámicos: Landsat 5 (2004) y Landsat 8 (2014). . . . .	80
8	Bandas y rangos Sentinel-2 (2024). . . . .	81
9	Distribución de los puntos de muestra recolectados . . . . .	85
10	Hiperparámetros MLP en estudios LULC. . . . .	95
11	Arquitecturas de capas ocultas reportadas en estudios LULC con MLP. . . . .	96
12	Comparación de configuraciones MLP. . . . .	97
13	Precisión global e hiperparámetros óptimos . . . . .	109
14	Matriz de confusión de la clasificación LULC 2004. . . . .	110
15	Métricas de precisión de la clasificación LULC 2004. . . . .	110
16	Matriz de confusión de la clasificación LULC 2014. . . . .	110
17	Métricas de precisión de la clasificación LULC 2014. . . . .	111
18	Matriz de confusión de la clasificación LULC 2024. . . . .	111
19	Métricas de precisión de la clasificación LULC 2024. . . . .	111
20	Cambios netos de área por clase LULC durante el período 2004 - 2014. . . . .	116
21	Cambios netos de área por clase LULC durante el período 2014 - 2024. . . . .	116
22	Matriz de probabilidad de transición estimada para el período 2004 - 2014. . . . .	117
23	Matriz de probabilidad de transición estimada para el período 2014 - 2024. . . . .	117

24	Matriz de confusión LULC 2024. . . . .	121
25	Métricas de precisión LULC 2024. . . . .	121
26	Validación multiescala LULC 2024. . . . .	123
27	Vegetación por orientación de ladera (2024). . . . .	125
28	Ranking de Importancia (Permutación) . . . . .	127
29	Proyección de área por clase LULC 2024-2034. . . . .	130
30	Matriz de probabilidad de transición utilizada para la proyección 2024 - 2034. .	132
31	Distribución proyectada de la vegetación según orientación (2034). . . . .	134
32	Fuentes de Atractivos Turísticos en la Provincia de Cusco . . . . .	179

## Índice de figuras

1	Cusco: densidad poblacional por provincia 2023–2024. . . . .	4
2	El espectro electromagnético. . . . .	16
3	División del espectro electromagnético en bandas. . . . .	17
4	Representación del sistema de coordenadas WGS 84. . . . .	18
5	Sistema de referencia EPSG:32719. . . . .	19
6	Línea de Tiempo del Programa Landsat. . . . .	22
7	Ejemplo de Escena (Tile) Sentinel-2 (27-06-2024). . . . .	24
8	Niveles de procesamiento de Datos Satelitales. . . . .	25
9	Google Earth Engine. . . . .	27
10	Interfaz del Editor de Código de Google Earth Engine. . . . .	28
11	QGIS. . . . .	28
12	Modelo Digital de Elevación (DEM). . . . .	37
13	Ejemplo clasificación de Coberturas del Suelo. . . . .	41
14	Arquitectura general de una Red Neuronal Artificial. . . . .	44
15	Ejemplos de Variables Predictoras de Proximidad. . . . .	48
16	Flujo Metodológico General de la Investigación. . . . .	58
17	Flujo de Adquisición y Preprocesamiento de Datos. . . . .	59
18	Flujo Metodológico: Fase de Clasificación Supervisada. . . . .	60
19	Proceso de Análisis de Cambios Multitemporal. . . . .	61
20	Simulación Híbrida (MLP-CA-MC). . . . .	62
21	Flujo de Proyección y Análisis del Escenario 2034. . . . .	63
22	Ubicación Geográfica del Área de Estudio. . . . .	64

23	Disponibilidad de Imágenes Satelitales Invierno Austral por Sensor. . . . .	66
24	Ejemplo de una escena “norte” de Sentinel-2. . . . .	69
25	Ejemplo de una escena “sur” adyacente de Sentinel-2. . . . .	70
26	Imagen Landsat-5 del 21 de agosto de 2004. . . . .	70
27	Imagen Landsat-8 del 02 de septiembre de 2014. . . . .	71
28	Imagen Sentinel-2 del 22 de julio de 2024. . . . .	71
29	Imagen Landsat-5 enmascarada del 21 de agosto de 2004. . . . .	73
30	Imagen Landsat-8 enmascarada del 02 de septiembre de 2014. . . . .	73
31	Imagen Sentinel-2 enmascarada del 22 de julio de 2024. . . . .	74
32	Compuesto de mediana (Landsat 5, invierno austral 2004). . . . .	75
33	Compuesto de mediana (Landsat 8, invierno austral 2014). . . . .	76
34	Compuesto de mediana (Sentinel-2, invierno austral 2024). . . . .	76
35	Distribución espacial del Índice de Vegetación de Diferencia Normalizada (NDVI) a lo largo del periodo de estudio. . . . .	77
36	Modelo de Elevación Digital (DEM) de la provincia de Cusco. . . . .	78
37	Mapa de Pendientes de la Provincia de Cusco. . . . .	79
38	Distribución espacial de los puntos de muestra 2004. . . . .	85
39	Puntos de muestra sobre Mapa de Peligros (2004). . . . .	86
40	Distribución espacial de los puntos de muestra 2014. . . . .	86
41	Distribución espacial de los puntos de muestra 2024. . . . .	87
42	Mapa LULC clasificado (2004). . . . .	89
43	Mapa LULC clasificado (2014). . . . .	89
44	Mapa LULC clasificado (2024). . . . .	90
45	Esquema general de clases. . . . .	94
46	Potencial para Ser de Cobertura Urbana (2014-2024). . . . .	99
47	Mapa LULC simulado (2024). . . . .	100
48	Potencial para Ser de Cobertura Urbana (2024-2034). . . . .	105
49	Mapa LULC proyectado (2034). . . . .	106
50	Resultados de optimización de hiperparámetros (2004). . . . .	107



51	Mapa de calor de precisión de validación (OA) 2004. . . . .	107
52	Resultados de optimización de hiperparámetros (2014). . . . .	108
53	Mapa de calor de precisión de validación (OA). . . . .	108
54	Resultados de optimización de hiperparámetros (2024). . . . .	108
55	Mapa de calor de precisión de validación (OA). . . . .	109
56	Mapa LULC clasificado (Sector Oeste Cusco 2004). . . . .	112
57	Mapa LULC clasificado (Sector Este Cusco 2004). . . . .	112
58	Mapa LULC clasificado (Sector Oeste Cusco 2014). . . . .	113
59	Mapa LULC clasificado (Sector Este Cusco 2014). . . . .	113
60	Mapa LULC clasificado (Sector Oeste Cusco 2024). . . . .	114
61	Mapa LULC clasificado (Sector Este Cusco 2024). . . . .	114
62	Importancia relativa de variables predictoras (2004). . . . .	115
63	Importancia relativa de variables predictoras (2014). . . . .	115
64	Importancia relativa de variables predictoras (2024). . . . .	115
65	Tendencia de evolución de coberturas. . . . .	116
66	Diagrama de Sankey de la evolución de coberturas del suelo. . . . .	117
67	Cambio de cobertura de suelo desnudo a urbana (2004-2014). . . . .	118
68	Cambio de cobertura de suelo desnudo a urbana (2014-2024). . . . .	118
69	Matriz de correlación de variables predictoras (2004). . . . .	119
70	Matriz de correlación de variables predictoras (2014). . . . .	120
71	Matriz de correlación de variables predictoras (2024). . . . .	120
72	Curva ROC Global (Macro-Average) del modelo MLP-CA-MC. . . . .	122
73	Validación multiescala. . . . .	123
74	Mapa LULC Simulado 2024 (Oeste) superpuesto en Alta Resolución. . . . .	124
75	Mapa LULC Simulado 2024 (Este) superpuesto en Alta Resolución. . . . .	124
76	Diagramas de frecuencia direccional de la cobertura vegetal (2024). . . . .	125
77	Mapa de Potencial de Vegetación (2014-2024). . . . .	126
78	Ranking de importancia de variables basado en el impacto en el AUC. . . . .	126
79	Impacto de la permutación de variables en Curvas ROC. . . . .	127

80	Curvas de respuesta PDP . . . . .	128
80	Curvas de respuesta PDP (continuación) . . . . .	129
81	Tendencia de evolución de coberturas (Proyección 2034). . . . .	131
82	Simulación de cambio de cobertura de suelo desnudo a urbana (2024 - 2034). .	131
83	Diagrama de Sankey de la evolución de coberturas del suelo (2020 - 2034). . .	132
84	Mapa LULC Proyectado (Oeste) 2034. . . . .	133
85	Mapa LULC Proyectado (Este) 2034. . . . .	133
86	Diagrama de frecuencia direccional proyectada de la vegetación (2034). . . . .	134
87	Mapa de Potencial de Vegetación (Proyección 2024-2034). . . . .	135
88	Mapa de Peligros de la Ciudad del Cusco 2004 . . . . .	174
89	Diferencias porcentuales entre el área calculada (Shapefile) y la documentada por fuentes oficiales. . . . .	176
90	Mapas de Predictores Espaciales utilizados en el modelo. . . . .	177

## Abreviaturas

- ANN: Artificial Neural Network (Red Neuronal Artificial)
- AUC: Area Under the Curve (Área Bajo la Curva)
- BOA: Bottom of Atmosphere (Reflectancia en el fondo de la atmósfera)
- BSI: Bare Soil Index (Índice de Suelo Desnudo)
- CA: Cellular Automata (Autómatas Celulares)
- CART: Classification and Regression Trees (Árboles de Clasificación y Regresión)
- DEM: Digital Elevation Model (Modelo Digital de Elevación)
- EPSG: European Petroleum Survey Group (Sistema de referencia geoespacial)
- EVI: Enhanced Vegetation Index (Índice de Vegetación Mejorado)
- F1 Score: F1 Measurement (Medida armónica de precisión y exhaustividad)
- FN: False Negative (Falso Negativo)
- FP: False Positive (Falso Positivo)
- GEE: Google Earth Engine
- GIS: Geographic Information System (Sistema de Información Geográfica)
- HPO: Hyperparameter Optimization (Optimización de Hiperparámetros)
- IA: Inteligencia Artificial
- IIEE: Instituciones Educativas
- INEI: Instituto Nacional de Estadística e Informática
- LULC: Land Use and Land Cover (Uso y Cobertura del Suelo)

- MC: Markov Chain (Cadenas de Markov)
- ML: Machine Learning (Aprendizaje Automático)
- MLP: Multilayer Perceptron (Perceptrón Multicapa)
- MLP-CA-MC: Multilayer Perceptron - Cellular Automata - Markov Chain (Modelo híbrido de Perceptrón Multicapa, Autómatas Celulares y Cadenas de Markov)
- MNDWI: Modified Normalized Difference Water Index (Índice Diferencial de Agua Normalizado Modificado)
- MSI: MultiSpectral Instrument (Instrumento Multiespectral - Sensor Sentinel-2)
- NDBI: Normalized Difference Built-up Index (Índice de Diferencia Normalizada de Área Construida)
- NDMI: Normalized Difference Moisture Index (Índice de Humedad de Diferencia Normalizada)
- NDVI: Normalized Difference Vegetation Index (Índice de Vegetación de Diferencia Normalizada)
- NIR: Near Infrared (Infrarrojo Cercano)
- OA: Overall Accuracy (Precisión Global)
- OLI: Operational Land Imager (Sensor Landsat 8)
- PA: Producer Accuracy (Precisión del Productor / Exhaustividad)
- PDP: Partial Dependence Plot (Gráfico de Dependencia Parcial)
- QGIS: Quantum GIS
- RF: Random Forest (Bosque Aleatorio)
- RGB: Red-Green-Blue (Rojo, Verde, Azul)
- ROC: Receiver Operating Characteristic (Curva Característica Operativa del Receptor)

- SAVI: Soil Adjusted Vegetation Index (Índice de Vegetación Ajustado al Suelo)
- SGD: Stochastic Gradient Descent (Descenso de Gradiente Estocástico)
- SR: Surface Reflectance (Reflectancia Superficial)
- SVM: Support Vector Machine (Máquina de Vectores de Soporte)
- SWIR: Short-Wave Infrared (Infrarrojo de Onda Corta)
- TM: Thematic Mapper (Sensor Landsat 5)
- TN: True Negative (Verdadero Negativo)
- TOA: Top of Atmosphere (Tope de la Atmósfera)
- TP: True Positive (Verdadero Positivo)
- TPM: Transition Probability Matrix (Matriz de Probabilidad de Transición)
- UA: User Accuracy (Precisión del Usuario / Precisión)
- UI: Urban Index (Índice Urbano)
- UTM: Universal Transverse Mercator (Sistema de coordenadas Universal Transversal de Mercator)
- WGS 84: World Geodetic System 1984 (Sistema Geodésico Mundial de 1984)
- XAI: Explainable Artificial Intelligence (Inteligencia Artificial Explicable)

## **CAPÍTULO I**

### **Introducción**

#### **1.1 Generalidades**

La detección y predicción de cambios en el uso y cobertura del suelo (LULC) se ha consolidado recientemente como un tema de gran relevancia dentro del ámbito de la teledetección, captando el interés tanto de investigadores como de planificadores territoriales. Este interés surge debido al impacto que las transformaciones en LULC ejercen sobre la cobertura natural, el cambio climático y otros problemas ambientales asociados. A escala global, diversas clases de cobertura terrestre han experimentado modificaciones considerables como resultado de actividades antropogénicas crecientes, entre las que destacan la deforestación, la expansión agrícola, la urbanización y la minería. Bajo este escenario, la actualización periódica de mapas de LULC resulta un requisito básico para comprender y gestionar la naturaleza dinámica del territorio.

Para el mapeo espacio-temporal de LULC, las imágenes multispectrales obtenidas mediante sensores remotos son una fuente de información esencial. Entre las plataformas satelitales más utilizadas se encuentran MODIS, SPOT, SAR, la serie Landsat, las misiones Sentinel-2 y RapidEye. De estas, las series Landsat y Sentinel-2 han sido ampliamente empleadas en estudios de LULC debido a su disponibilidad y características técnicas. Se recomienda particularmente el uso de composiciones libres de nubes, generadas a partir de series temporales de imágenes, en lugar de depender de imágenes individuales, para mitigar distorsiones atmosféricas. No obstante, los métodos tradicionales para la búsqueda, filtrado, enmascaramiento de nubes, composición, descarga y clasificación de estos grandes volúmenes de datos demandan una considerable capacidad computacional y de almacenamiento. Frente a estos desafíos, la plataforma Google Earth Engine (GEE) permite procesar datos de teledetección a gran escala de manera óptima.

En el entorno de GEE se encuentran disponibles diversos algoritmos de clasificación supervisada, tales como Árbol de Decisión (DT), Bosque Aleatorio (RF), Naive Bayes

(NB), Distancia Mínima (MD) y CART. Entre ellos, el algoritmo RF ha sido extensamente recomendado en la literatura científica por ofrecer una alta exactitud en la clasificación y por su capacidad intrínseca para manejar datos de alta dimensionalidad con un número reducido de parámetros a ajustar. Asimismo, se ha demostrado que la incorporación de factores topográficos (como elevación y pendiente) e índices espectrales derivados puede mejorar la precisión de la clasificación LULC.

Además de la clasificación histórica de LULC, la predicción de cambios futuros resulta relevante para simular escenarios proyectados del paisaje y comprender los factores impulsores de la transformación del suelo. Existen diversos modelos de predicción, categorizados como temporales y espaciales, así como enfoques híbridos que combinan elementos de ambos. Recientemente, los modelos híbridos, como el modelo CA-Markov, han demostrado ser especialmente efectivos al integrar las capacidades de simulación espacial del CA con la predicción temporal de la Cadena de Markov. Numerosos investigadores han aplicado el modelo CA-Markov para predecir la dinámica futura de LULC en diversos contextos geográficos.

Este estudio analiza la dinámica histórica del uso y cobertura del suelo (LULC) en la provincia del Cusco para los años 2004, 2014 y 2024, empleando el algoritmo Random Forest dentro de Google Earth Engine para generar clasificaciones. A partir de estas cartografías, se identifican y cuantifican los cambios ocurridos en el territorio a lo largo de las dos décadas evaluadas. Además, se desarrolla una proyección para el año 2034 mediante un modelo híbrido que combina Perceptrón Multicapa, Cadenas de Markov y Autómatas Celulares, incorporando variables topográficas (altitud, pendiente y aspecto) junto con factores de proximidad (red hidrográfica, Instituciones Educativas, vías nacionales y provinciales, atractivos turísticos, poblaciones dispersas, poblaciones grandes). La propuesta metodológica integra información espectral y topográfica para incrementar la precisión de las clasificaciones y por aplicar un enfoque predictivo avanzado para anticipar la evolución espacio-temporal del paisaje.

La presente investigación tiene como finalidad aportar información cuantitativa y espacialmente explícita para comprender la magnitud y los patrones del crecimiento urbano y su impacto en las coberturas naturales de la provincia del Cusco. Se aspira a que los resultados obtenidos, incluyendo los mapas históricos, mapas de transición espacial, mapas de potencial

y el escenario proyectado, constituyan una base de información para el análisis y la planificación urbana en el marco de la gestión sostenible del territorio.

## **1.2 Justificación**

La justificación principal para emprender esta investigación reside en la necesidad de gestionar la tierra y los recursos naturales de manera sostenible ante la compleja dinámica del Uso y Cobertura del Suelo (LULC) (Lin & Peng, 2024; Setiawan, 2024; Taloor, 2024). Se justifica el desarrollo de modelos predictivos y análisis espaciales para anticipar y mitigar riesgos ambientales y prevenir el deterioro ecológico. Este estudio busca ofrecer un sustento técnico riguroso para que los planificadores urbanos y los responsables de políticas públicas puedan formular estrategias de desarrollo sostenible y gestionar recursos no renovables (Kondum et al., 2024). La cuantificación del LULC y su predicción son determinantes para optimizar los patrones de uso del suelo con el fin de contribuir a los objetivos de reducción de emisiones de carbono y protección de ecosistemas (H. D. Nguyen et al., 2024; Wang et al., 2024).

## **1.3 Planteamiento del problema**

El acelerado proceso de expansión urbana, particularmente marcado en los países en desarrollo, constituye un fenómeno complejo impulsado principalmente por la migración hacia las ciudades en busca de mejores oportunidades, actuando como un motor demográfico determinante (Huaraca Yucra & Surco Vega, 2019). Si bien esta migración puede dinamizar la economía mediante la incorporación de una fuerza laboral flexible, también introduce desafíos que afectan negativamente la sostenibilidad ambiental y urbana (Huarcaya Fernandez Baca & Villalba Velasque, 2016; Sharma et al., 2024). En paralelo, el incremento del turismo intensifica la presión sobre el territorio, generando una mayor demanda de infraestructura y servicios urbanos, lo que a menudo excede la capacidad planificada (Condori Juarez, 2019; B. Roy, 2021).

La expansión urbana no planificada produce impactos adversos en el uso del suelo y la vivienda (Trujillo, 2019), profundizando desigualdades socioeconómicas al limitar el acceso equitativo a servicios básicos, vivienda y oportunidades de empleo (Gündüz, 2025; Trujillo,

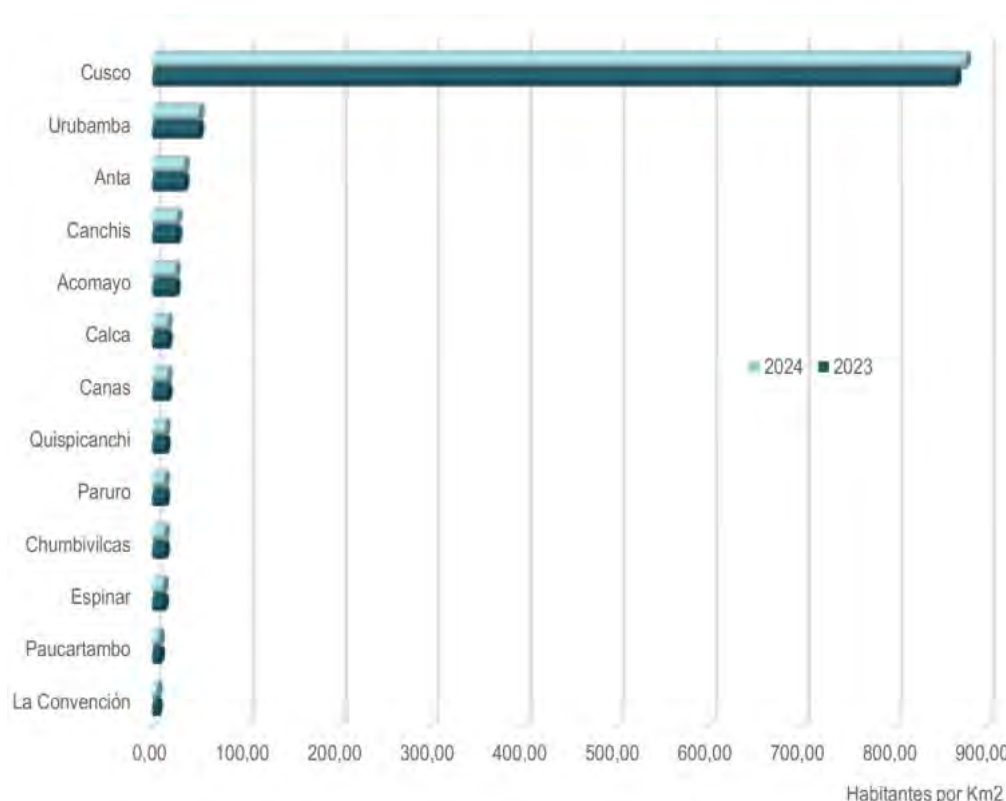


2019). Como consecuencia, los procesos de urbanización no regulada generan transformaciones aceleradas del territorio que comprometen tanto la calidad ambiental como la cohesión social.

La provincia de Cusco presenta la mayor concentración poblacional del Departamento homónimo, superando ampliamente en densidad al resto, como se observa en la Figura 1. Este dinamismo demográfico y turístico ha desencadenado una ocupación del territorio caracterizada por múltiples problemáticas.

### Figura 1

*Cusco: densidad poblacional por provincia 2023–2024.*



*Nota. Fuente: Instituto Nacional de Estadística e Informática (INEI), 2024.*

Se evidencia un impacto territorial y ambiental severo, la expansión descontrolada ocupa zonas no aptas, afectando áreas de cultivo, bosques y paisajes (Trujillo, 2019). La construcción informal avanza sobre laderas inestables, bordes de ríos y terrenos expuestos a amenazas de inundaciones y aluviones (Huaraca Yucra & Surco Vega, 2019; Mamani & Cutipa, 2024; Trujillo, 2019).

Adicionalmente, se observa una degradación del patrimonio cultural y urbano. Las ocupaciones informales han invadido zonas arqueológicas y áreas de importancia histórica (Mamani & Cutipa, 2024; Trujillo, 2019), favorecidas por la ausencia de lineamientos de

protección (Mamani & Cutipa, 2024).

La gestión del suelo presenta altos niveles de informalidad, con proliferación de transferencias de “derechos y acciones” sin habilitación urbana ni títulos de propiedad (Berrio Gomez, 2017), incluyendo ocupaciones sobre terrenos estatales y privados. Estas nuevas áreas urbanas suelen carecer de servicios básicos, sistemas de saneamiento adecuados, espacios públicos y áreas verdes (Yañe Zuñiga, 2019), ocasionando sobrecarga de la infraestructura existente. De manera paralela, la alta densidad no planificada incrementa la demanda eléctrica sin considerar riesgos de incendio o explosión (Huaraca Yucra & Surco Vega, 2019).

Finalmente, la movilidad urbana presenta disfunciones. El planeamiento ha priorizado el enfoque centrado en la vialidad, relegando el espacio público. Paradójicamente, la ampliación de la oferta vial ha inducido mayor congestión (Trujillo, 2019), mientras que la pérdida de densidad habitacional en ciertas zonas aumenta los costos de mantenimiento de la infraestructura urbana (Berrio Gomez, 2017).

Todos estos procesos, además de alterar la configuración urbana y social, comprometen la sostenibilidad de los servicios ecosistémicos, esenciales para la estabilidad económica y ambiental a largo plazo. En este escenario, la detección eficiente del estado actual y los cambios futuros en el uso y cobertura del suelo (LULC) adquiere un papel estratégico.

La falta de información actualizada y técnicamente sólida limita la toma de decisiones, lo que puede derivar en una planificación urbana inadecuada y comprometer la sostenibilidad del territorio. Por ello, las estrategias de monitoreo basadas en aprendizaje automático y simulación espacial representan herramientas clave, capaces de generar proyecciones detalladas que permitan apoyar el diseño de políticas y estrategias de desarrollo sostenible.

### ***1.3.1 Formulación del problema***

### ***1.3.2 Problema general***

¿En qué medida se proyectan los cambios en el uso y cobertura del suelo mediante la implementación de un modelo predictivo, con el fin de generar información para la gestión sostenible de los recursos naturales?

### **1.3.3 Problemas específicos**

1. ¿Con qué nivel de precisión se puede clasificar el uso y cobertura del suelo a partir de imágenes satelitales históricas?
2. ¿Cuál es la magnitud de los cambios entre clases de cobertura del suelo, y cuáles son las transiciones más relevantes durante el periodo de análisis?
3. ¿Puede un modelo basado en tendencias históricas reproducir cambios reales de uso y cobertura de suelo?

## **1.4 Alcances y Limitaciones**

### **1.4.1 Alcances del Estudio**

- Área de estudio y periodo temporal: El análisis se centra en la provincia de Cusco, evaluando la dinámica de cambio de la Cobertura y Uso del Suelo (LULC) durante la temporada de invierno de los años 2004, 2014 y 2024. El objetivo principal es identificar transformaciones históricas y proyectar escenarios futuros.
- Fuentes de datos satelitales: Se usan datos multisatélite de teledetección, utilizando Reflectancia Superficial (SR) de las misiones Landsat 5 (TM) y Landsat 8 (OLI), así como Reflectancia en la Base de la Atmósfera (BOA) de Sentinel-2 (MSI). El acceso, filtrado y preprocesamiento se realizó a través de la plataforma Google Earth Engine ([GEE](#)). Con el fin de reducir interferencias por nubosidad, se priorizaron las capturas correspondientes a la temporada de invierno.
- Infraestructura y procesamiento: El procesamiento masivo de datos (Big Data) se realizará mediante plataformas de análisis geoespacial en la nube, específicamente Google Earth Engine (GEE) y Google Colab.
- Clasificación de cobertura terrestre: La clasificación LULC fue realizada mediante el algoritmo Random Forest (RF), empleando bandas espectrales, índices espectrales (NDVI, NDBI, NDMI, UI, BSI, SAVI) y variables topográficas (altitud y pendiente). La

generación de muestras se basa en imágenes de alta resolución disponibles en Google Earth Pro para los años 2002, 2014 y 2024 y un Mapa de Peligros del 2004 (Apéndice A). Se considerarán tres clases principales: cobertura urbana/asentamientos, vegetación y suelo desnudo.

- **Modelamiento y predicción espacial:** La simulación de dinámicas de cambio se llevará a cabo utilizando Cadenas de Markov (MC) y Autómatas Celulares (CA), integrados con una red neuronal Perceptrón Multicapa (MLP). Se incluirán variables derivadas del modelo digital de elevación (elevación, pendiente, aspecto), así como factores de accesibilidad a elementos naturales (distancia a la red hidrológica), a sitios de interés turístico y a infraestructura y servicios (distancia a poblaciones dispersas y centros urbanos, IIEE y carreteras) para modelar los potenciales de transición y la distribución espacial futura del LULC.

#### **1.4.2 Limitaciones**

- **Enfoque Disciplinario:** El estudio se desarrolla principalmente desde una perspectiva técnica, orientada por los principios y métodos del aprendizaje automático y el modelamiento computacional. En consecuencia, el análisis se concentra en la implementación y evaluación de herramientas de procesamiento y modelamiento de datos espaciales, sin integrar en profundidad enfoques provenientes de otras áreas del conocimiento.
- **Recursos y Reproducibilidad:** El estudio emplea exclusivamente herramientas *open-source* o con planes de acceso gratuito (GEE, Google Colab), lo que garantiza su reproducibilidad. Esta elección implica trabajar con los recursos computacionales disponibles en dichas plataformas y con imágenes satelitales de acceso libre.
- **Calidad y Disponibilidad de Datos:** El estudio enfrenta restricciones relacionadas con los datos satelitales. La resolución espacial de Landsat (30 m), aunque estándar para análisis históricos, puede influir en la precisión del mapeo LULC en áreas heterogéneas (Li et al., 2024; Setiawan, 2024). Además, la disponibilidad limitada de imágenes en

algunos años puede afectar las composiciones generadas.

- **Generalización y Dependencia Histórica del Modelo:** Los algoritmos de Machine Learning (RF y ANN) basan su aprendizaje en patrones históricos. Esto puede limitar su capacidad de generalización si emergen dinámicas de cambio futuras completamente nuevas o no representadas en los datos de entrenamiento (Bojer, 2024; H. D. Nguyen et al., 2024).
- **Exclusión de Factores Dinámicos:** Como consecuencia del enfoque disciplinario, los modelos espaciales híbridos dependen en gran medida de variables biofísicas y de proximidad. Esto restringe la incorporación de factores socioeconómicos y políticos dinámicos (cambios en la zonificación, establecimiento de áreas restringidas o protegidas, fluctuaciones del mercado inmobiliario) que influyen directamente en las proyecciones futuras (M'Barek, 2024; Zafar, Zubair, Zha, Mehmood et al., 2024).
- **Disponibilidad de Variables Predictoras:** La mayoría de los rásteres utilizados como variables predictoras (aspecto, pendiente, elevación, distancia a la red hidrológica, poblaciones dispersas, centros urbanos, carreteras y sitios de interés turístico) se mantuvieron constantes debido a la ausencia de información geoespacial correspondiente a cada año. Solo las capas de distancia a Instituciones Educativas (IIEE) pudieron desagregarse temporalmente. Esta limitación reduce la sensibilidad del modelo para capturar cambios asociados a variaciones reales en la infraestructura y la accesibilidad.

## **1.5 Objetivos**

### ***1.5.1 Objetivo General***

Cuantificar los cambios proyectados en el uso y cobertura del suelo mediante un modelo predictivo, con el fin de generar información para la gestión sostenible de los recursos naturales.

### 1.5.2 *Objetivos Específicos*

1. Determinar el nivel de precisión de la clasificación de uso y cobertura del suelo a partir de imágenes satelitales.
2. Cuantificar los patrones espaciales de cambio y las transiciones más relevantes entre clases de cobertura del suelo.
3. Validar la precisión del modelo contrastando sus resultados con datos reales correspondientes a un año clave dentro del periodo de análisis.

## 1.6 *Antecedentes*

(Gündüz, 2025), *Land-Use Land-Cover Dynamics and Future Projections Using GEE, ML, and QGIS-MOLUSCE: A Case Study in Manisa*, Department of Geomatics Engineering, Aksaray University, 68100 Aksaray, Turkey.

- Se compararon cuatro algoritmos de aprendizaje automático (*RF*, *SVM*, *KNN* y *CART*) en la plataforma *Google Earth Engine* para la generación de mapas LULC, donde *Random Forest* obtuvo un *OA* de 98 % y un índice  $\kappa$  de 0.97, superando el desempeño de los modelos basados en árboles de decisión simples y vecindad.
- Se aplicó un post-procesamiento mediante operaciones de moda focal para reducir el ruido tipo *salt-and-pepper* en las clasificaciones, mejorando la continuidad espacial de las categorías de cobertura terrestre.
- La arquitectura CA-ANN (usando el plugin MOLUSCE) alcanzó una corrección proporcional del 92 % ( $K_h = 0.99$ ), permitiendo proyectar para el año 2030 un incremento del 23.67 % en áreas urbanas y una reducción del 3.16 % en cuerpos de agua.

Comentario: El flujo de trabajo evidencia que los algoritmos de aprendizaje por conjuntos (RF) presentan una menor tasa de error en zonas de transición espectral en comparación con CART

y KNN, los cuales muestran mayor confusión en clases de suelo desnudo y áreas periurbanas debido a limitaciones en la optimización de hiperparámetros en GEE.

(B. Roy, 2021), *A machine learning approach to monitoring and forecasting spatio-temporal dynamics of land cover in Cox's Bazar district, Bangladesh from 2001 to 2019*, Department of Geography and Environmental Science, Begum Rokeya University, Bangladesh.

- Se desarrolló un flujo de trabajo de clasificación basado en el algoritmo *Random Forest* aplicado a series temporales de sensores Landsat 4-5 TM y Landsat 8 OLI, alcanzando métricas de precisión entre el 87 % y el 99.8 %.
- La validación del modelo predictivo mediante matrices de confusión y estadísticos Kappa demostró que la integración de RF con CAS permite proyectar escenarios futuros con una fiabilidad del 93.2 %.

Comentario: La integración de clasificadores de aprendizaje por conjuntos (RF) con sistemas de autómatas celulares parametrizados mediante redes neuronales (CA-ANN) permite la captura de dependencias espaciales no lineales. El uso de RF para la generación de mapas LULC reduce el ruido en la clasificación de datos, lo cual disminuye los sesgos para el modelamiento de transición temporal.

(Belay et al., 2024), *Scenario-Based Land Use and Land Cover Change Detection and Prediction Using the Cellular Automata-Markov Model in the Gumara Watershed, Upper Blue Nile Basin, Ethiopia*.

- Se ejecutó una clasificación multitemporal utilizando el algoritmo *Random Forest* (RF) integrado en la plataforma *Google Earth Engine* (GEE), procesando 10 variables que incluyen bandas espectrales de Landsat, índices de vegetación (NDVI, SAVI) y factores topográficos.
- El análisis de importancia de variables identificó a la elevación como el factor de mayor peso en la discriminación de clases. Se obtuvo una precisión general de entre 91.13 % y 94.84 % y coeficientes Kappa ( $\kappa$ ) de 0.88 a 0.94.

- Para la fase predictiva (2035-2065), se implementó el modelo de Autómatas Celulares-Markov, validado con métricas de concordancia espacial  $K_{no}$  (0.89) y  $K_{standard}$  (0.94), evaluando escenarios de tendencia inercial (BAU) y de gobernanza ambiental (GOV).

Comentario: El diseño computacional destaca por la integración de variables topográficas y espectrales para lograr una mejor clasificación.

(Badshah et al., 2024), *The role of random forest and Markov chain models in understanding metropolitan urban growth trajectory.*

- La metodología hace uso de imágenes satelitales Landsat (1991-2021) a través de la plataforma *Google Earth Engine*, utilizando el algoritmo de aprendizaje automático *Random Forest* para la clasificación de coberturas terrestres e integrando variables topográficas y de proximidad.
- Para la simulación de la trayectoria de crecimiento urbano hacia los años 2031, 2041 y 2051, se implementó el modelo híbrido Cadenas de Markov y Autómatas Celulares mediante el *Land Change Modeler* (LCM) en el software TerrSet.
- Este sistema emplea redes neuronales de Perceptrón Multicapa para generar mapas de potencial de transición, permitiendo modelar interacciones complejas entre factores socioeconómicos y cambios ambientales en el área metropolitana.
- La validación de la precisión espacial del modelo predictivo alcanzó un valor AUC-ROC de 0.88.
- El clasificador RF demostró una alta estabilidad con una exactitud general superior al 90 % y un coeficiente Kappa de aproximadamente 0.88.

Comentario: El estudio sugiere la eficiencia de integrar técnicas de *machine learning* con modelos probabilísticos para entender la dinámica territorial.

(Kamran et al., 2024), *Application of Cellular Automata and Markov Chain model for urban green infrastructure in Kuala Lumpur, Malaysia.*



- La análisis multitemporal se desarrolló a partir de imágenes satelitales Landsat (1990, 2005 y 2021) para monitorear la infraestructura verde urbana (UGI) en Kuala Lumpur, empleando tanto el clasificador de Máxima Verosimilitud como máquinas de vectores de soporte para categorizar el suelo en seis clases principales.
- La validación de los mapas clasificados mostró una Exactitud General que progresó del 91.06 % en 1990 al 98.28 % en 2021 y coeficientes Kappa entre 0.8997 y 0.9626.
- Para modelar las dinámicas de cambio y proyectar el estado del territorio hacia el año 2050, se aplicó el modelo híbrido Cadenas de Markov y Autómatas Celulares.
- Se identificó que las áreas construidas (*built-up area*) presentaron una probabilidad de 0.2736 de transicionar hacia suelo desnudo (*bare ground*), mientras que la probabilidad de permanecer como área construida fue de 0.6381.

Comentario: La integración de CA-Markov resulta fundamental para la planificación urbana, ya que permite anticipar cómo la expansión de superficies impermeables y suelo desnudo incrementa riesgos ambientales como las inundaciones, proporcionando una base científica para fortalecer la resiliencia de los servicios ecosistémicos en ciudades de rápido crecimiento.

(Vahid & Aly, 2025), *A Comprehensive Systematic Review of Machine Learning Applications in Assessing Land Use/Cover Dynamics and Their Impact on Land Surface Temperatures*.

- La metodología comprende una revisión sistemática de 144 publicaciones (período 2014-2024) sobre el uso de algoritmos de aprendizaje automático (*Machine Learning*) para monitorear cambios en el uso y cobertura del suelo (LULC) y su impacto directo en la temperatura de la superficie terrestre (LST).
- Los hallazgos identifican a *Random Forest* y *Support Vector Machines* como los clasificadores más utilizados y precisos para la obtención de mapas de LULC, gracias a su alta capacidad para procesar datos no lineales y multitemporales.

- El análisis destaca la superioridad de los modelos híbridos, especialmente la combinación de Autómatas Celulares con Redes Neuronales Artificiales o algoritmos de optimización, para la simulación precisa de dinámicas térmicas y territoriales futuras.
- Se observa una transición técnica hacia el *Deep Learning*, lo que permite manejar la complejidad de los datos satelitales de alta resolución y mejorar la predictibilidad de los servicios ecosistémicos ante el cambio climático.

Comentario: Esta revisión sistematiza la evolución de las herramientas computacionales en la gestión territorial, evidenciando que la integración de inteligencia artificial es determinante para modelar la relación entre urbanización y microclima. La transición hacia modelos predictivos ofrece una base científica para el diseño de estrategias de adaptación urbana que prioricen la restauración térmica mediante infraestructura verde.

(Toscan et al., 2025), *Impact of nature-based solutions (NBS) on urban surface temperatures and land cover changes using remote sensing and machine learning*.

- La metodología se fundamenta en el procesamiento de series temporales de imágenes Landsat (1984-2023) a través de la plataforma *Google Earth Engine*, utilizando el algoritmo de aprendizaje automático *Random Forest* para la clasificación supervisada de la cobertura terrestre y la recuperación de la temperatura de la superficie terrestre (LST).
- El flujo de trabajo técnico integra el cálculo de índices espectrales como el NDVI y el NDBI como variables predictoras para modelar la correlación térmica entre la expansión de superficies impermeables y la pérdida de vegetación.
- Para la simulación de la dinámica espacial hacia los años 2030, 2040 y 2050, se implementó el modelo estocástico CA-Markov, permitiendo proyectar escenarios de cambio de uso de suelo bajo la influencia de soluciones basadas en la naturaleza (NBS).
- El modelo alcanzó un valor de validación Kappa de 0.86, los datos predichos se compararon con la cobertura real de 2023, resultando en una exactitud general del 84.78 % y un valor Kappa general de 0.88.

Comentario: La investigación sugiere la viabilidad de integrar modelos de *machine learning* y simulaciones probabilísticas para la planificación urbana climáticamente inteligente. Al cuantificar el impacto térmico de la urbanización, el estudio proporciona un marco metodológico útil para evaluar escenarios de planificación urbana bajo criterios climáticos.

(Ahmad et al., 2025), *How do land use changes affect temperature and groundwater in urban areas? An integrated remote sensing, and machine learning approach.*

- Para la modelización de escenarios futuros, se aplicó el sistema híbrido CA-Markov, que simula las transiciones espaciales basándose en la probabilidad histórica de cambio de suelo.
- La validación del modelo de simulación espacial demostró una alta fiabilidad técnica, obteniendo un valor de kappa histogram: 0.9495.
- El análisis de los resultados revela una correlación crítica donde la sustitución de áreas agrícolas por zonas construidas no solo eleva la LST, sino que reduce la recarga de los acuíferos, proyectando un descenso significativo en los niveles freáticos para el año 2038.

Comentario: Al integrar modelos predictivos de LULC con variables hidrogeológicas, la investigación proporciona una herramienta para la gestión sostenible de recursos hídricos, permitiendo a los planificadores urbanos anticipar y mitigar los efectos del estrés térmico y la escasez de agua en megaciudades de rápido crecimiento.

(K. C. Roy et al., 2024), *Land-use/cover change and future prediction by integrating the ML techniques of random forest and CA-Markov chain model of the Ganges alluvial tract of Eastern India*, Weste University.

- Se propone una arquitectura que fusiona RF para la extracción de características espectrales con modelos de simulación espacio-temporal (CA-Markov).
- El modelo incorpora “variables predictoras” (*drivers*) espaciales, como la distancia euclidiana a redes hídricas y viales, procesadas mediante lógica difusa o reescalado para alimentar las reglas de transición del autómata.

- La predicción a 10 y 20 años se genera iterando las probabilidades de la cadena de Markov, ajustada por la idoneidad del terreno.

Comentario: El estudio ejemplifica la aplicación de *Soft Computing* en geoinformática, donde el aprendizaje automático (RF) no solo clasifica, sino que mejora la calidad de los datos de entrada para los modelos determinísticos y estocásticos subsiguientes, refinando la precisión de las simulaciones a largo plazo.

(Bendechou et al., 2024), *Monitoring and Predicting Land Use/Land Cover Dynamics in Djelfa City, Algeria, using Google Earth Engine and a Multi Layer Perceptron Markov Chain Model*, Weste University.

- Se aplicó el algoritmo *Support Vector Machine* en GEE para la clasificación supervisada, aprovechando su capacidad de generalización en espacios de características de alta dimensión mediante funciones kernel.
- Para la predicción, se utilizó una red neuronal *Multi-Layer Perceptron* integrada con Cadenas de Markov (MLP-Markov). El MLP se entrena para modelar el potencial de transición de cada celda basándose en variables explicativas, capturando relaciones no lineales complejas.
- La validación del modelo híbrido MLP-Markov alcanzó una precisión del 83.96 %.

Comentario: El uso de perceptrones multicapa permite ponderar la influencia de múltiples variables conductoras, superando las limitaciones de las matrices de probabilidad estáticas propias de los modelos markovianos tradicionales.

## CAPÍTULO II

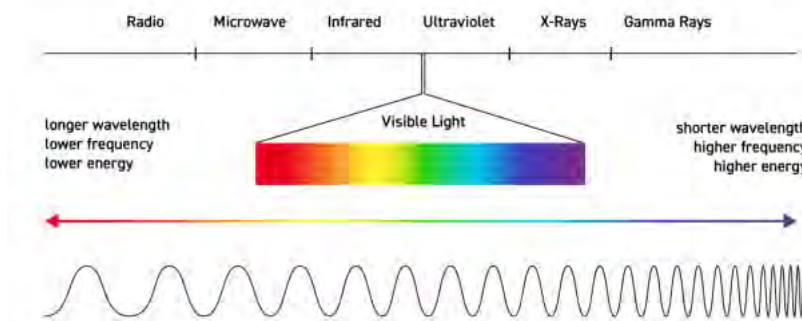
### Marco teórico

#### 2.1 La Luz y el Espectro Electromagnético

La luz es una forma de energía que se manifiesta como una onda electromagnética. Aunque la percepción visual humana se limita a una fracción reducida de este fenómeno, el espectro visible, la luz forma parte de un continuo más amplio denominado espectro electromagnético. El espectro abarca desde las ondas de radio, caracterizadas por longitudes de onda muy largas, hasta los rayos gamma, con longitudes de onda extremadamente cortas. Cada tipo de radiación electromagnética se define por su longitud de onda y frecuencia. La interacción de esta energía con los objetos permite obtener información sobre ellos a distancia (Khan Academy, 2024).

#### Figura 2

*El espectro electromagnético.*



*Nota. Fuente: National Institute of Standards and Technology (NIST), 2024.*

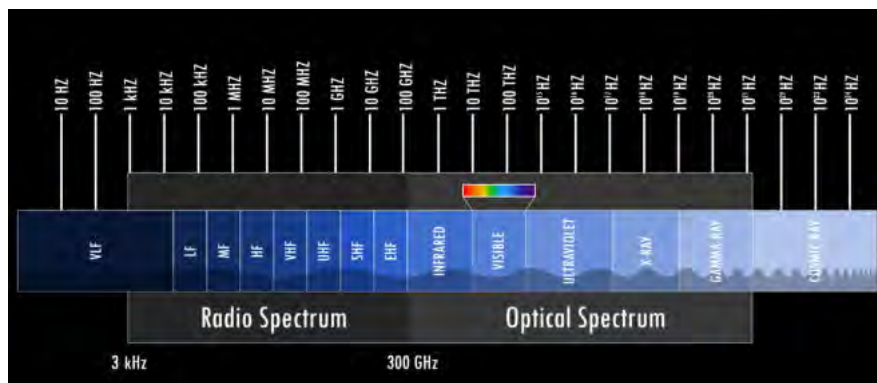
##### 2.1.1 Bandas Espectrales

Una banda espectral se define como un intervalo específico y estrecho de longitudes de onda dentro del cual un sensor multicanal detecta la radiación electromagnética. La capacidad de un sensor para definir y discriminar estos canales se conoce como su resolución espectral (Ezekiel, 2017).

Existen numerosas formas de designar estas bandas, como la de Institute of Electrical and Electronics Engineers (IEEE), que son un estándar para radar y comunicaciones (Manning, 2018).

### Figura 3

*División del espectro electromagnético en bandas.*



*Nota. Fuente: Manning, 2018.*

## 2.2 Sistemas de Referencia Espacial

### 2.2.1 WGS 84 (Sistema Geodésico Mundial de 1984)

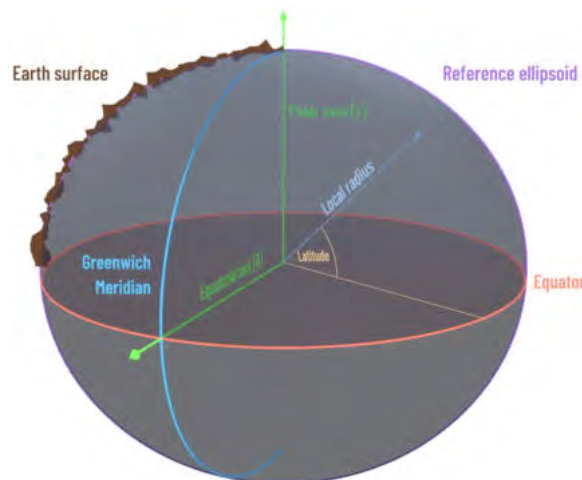
El WGS 84 (World Geodetic System 1984) es el sistema de referencia geodésico tridimensional estándar, ampliamente adoptado y utilizado a nivel global para aplicaciones de geoposicionamiento y navegación. Define un marco para especificar posiciones mediante latitud, longitud y altitud elipsoidal. Ha sido adoptado por organismos internacionales como la Organización de Aviación Civil Internacional (OACI, ICAO por sus siglas en inglés) (National Geospatial-Intelligence Agency, 2025).

Este sistema fue desarrollado por la *National Geospatial-Intelligence Agency* (NGA) de los Estados Unidos. Se usa un elipsoide de referencia geocéntrico (el origen se ubica en el centro de masas de la Tierra) y define un sistema de coordenadas cartesiano tridimensional fijo a la Tierra (ECEF - Earth-Centered, Earth-Fixed). En este sistema, el eje Z apunta hacia el Polo Norte geodésico convencional, el eje X interseca el punto donde el ecuador se cruza con el meridiano de Greenwich (longitud 0°), y el eje Y completa un sistema ortogonal dextrógiro (orientado a la derecha), como se ilustra en la Figura 4. WGS 84 es el datum de referencia oficial para el Sistema de Posicionamiento Global (GPS) y es el estándar de facto en cartografía

moderna, navegación satelital y Sistemas de Información Geográfica (SIG) (United Nations Office for Outer Space Affairs (UNOOSA), 2012).

#### Figura 4

*Representación del sistema de coordenadas WGS 84.*



*Nota. Fuente: GeneSys Elektronik GmbH, 2024.*

#### 2.2.2 Sistemas de Referencia Espacial en Perú (EPSG)

De acuerdo con (Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI), 2015), los sistemas de referencia espacial, identificados mediante códigos asignados por el EPSG (European Petroleum Survey Group, actualmente IOGP Geomatics Committee), permiten definir de manera unívoca cómo se representan las coordenadas geográficas y proyectadas en los mapas. Para el territorio peruano, los códigos EPSG:32717, EPSG:32718 y EPSG:32719 corresponden a sistemas de coordenadas proyectadas bajo la Proyección Universal Transversa de Mercator (UTM) para las zonas 17 Sur, 18 Sur y 19 Sur, respectivamente. Todos ellos utilizan el datum WGS 84. La adopción de estos sistemas estandarizados garantiza la interoperabilidad, consistencia y precisión en la visualización, análisis e intercambio de datos geoespaciales en el país.

**EPSG:32719.** El sistema de referencia de coordenadas EPSG:32719 (Figura 5a) corresponde a la proyección UTM Zona 19 Sur sobre el datum WGS 84. Es el sistema que cubre la mayor extensión de la provincia de Cusco y se utiliza comúnmente para la cartografía en la región oriental del Perú (Figura 5b).

Las principales características del sistema se resumen en la Tabla 1.

**Tabla 1**

*Características principales del sistema EPSG:32719*

Característica	Descripción
Nombre del sistema	WGS 84 / UTM zone 19S
Límites geográficos (WGS 84)	Longitudes entre $-72.0^{\circ}$ y $-66.0^{\circ}$ ; latitudes entre $-80.0^{\circ}$ y $0.0^{\circ}$
Área de uso	Región comprendida entre los meridianos $72^{\circ}$ O y $66^{\circ}$ O en el hemisferio sur, desde el ecuador hasta $80^{\circ}$ S; aplicable en áreas terrestres y marinas
Países incluidos	Argentina, Bolivia, Brasil, Chile, Colombia y Perú
Sistema de referencia geográfico base	EPSG:4326 (WGS 84)

*Nota. Fuente: (Esri Support, 2025).*

**Figura 5**

*Sistema de referencia EPSG:32719.*



*Nota. Fuente de la subfigura (a): EPSG.io, 2025a. Fuente de la subfigura (b): EPSG.io, 2025b.*

## 2.3 Teledetección

La teledetección (Remote Sensing, RS) se define como el conjunto de técnicas y procedimientos utilizados para obtener información sobre un objeto, área o fenómeno sin establecer contacto físico directo con él (M. Khan et al., 2023). En el contexto de las ciencias de la Tierra, la teledetección es una disciplina esencial para la captura, procesamiento y análisis de imágenes digitales de la superficie terrestre, obtenidas principalmente a través de sensores



montados en satélites artificiales (Valladares Herrera, 2023).

Actualmente, la teledetección se utiliza ampliamente para adquirir una comprensión detallada de los patrones del paisaje, registrar la energía electromagnética emitida o reflejada por la superficie, y proveer un contexto histórico fundamental para el estudio de los cambios en la cobertura y uso del suelo (LULC) (M. Khan et al., 2023). Misiones satelitales históricas, como la serie Landsat, y programas más recientes, como Sentinel, han sido instrumentales para el avance de esta disciplina.

### **2.3.1 Sensores Satelitales**

La función primordial de un sensor satelital en teledetección es detectar y medir la radiación electromagnética (EM) reflejada o emitida por la superficie terrestre. Esta información capturada se transforma en datos digitales que pueden ser posteriormente procesados y analizados para extraer información temática (Ezekiel, 2017).

Existen dos categorías principales de sensores utilizados en la teledetección satelital, clasificadas según la fuente de energía que utilizan para la observación (Ezekiel, 2017):

#### **1. Sensores Pasivos (Ópticos/Infrarrojos):**

- Son los sensores más comunes para la Observación de la Tierra (EO - Earth Observation).
- Detectan la energía electromagnética natural, principalmente la radiación solar reflejada por los objetos en la superficie terrestre, o la radiación térmica emitida por ellos.
- La información contenida en las imágenes ópticas depende de las propiedades de reflectividad (o emisividad) espectral del objeto observado en las longitudes de onda específicas que el sensor es capaz de registrar (bandas espectrales).
- La adquisición de datos se realiza muestreando la energía en puntos discretos a lo largo de los segmentos óptico e infrarrojo del espectro electromagnético, almacenando esta información en diferentes bandas.

- La calidad de los datos de los sensores ópticos pasivos puede verse afectada por las condiciones de iluminación solar (hora del día, estación, latitud), la presencia de sombras, y las condiciones atmosféricas (aerosoles, nubes, vapor de agua).

## 2. Sensores Activos (Radar/LiDAR):

- Estos sensores emiten su propia fuente de energía (generalmente microondas o pulsos láser) hacia la superficie y miden la señal que retorna tras interactuar con los objetos.

### 2.3.2 *Programas Satelitales*

Un programa satelital se refiere a una iniciativa coordinada, generalmente a largo plazo, que involucra el diseño, desarrollo, lanzamiento y operación de uno o más satélites con objetivos específicos, comúnmente orientados al monitoreo sistemático y la observación de la Tierra desde el espacio (Ezekiel, 2017).

### 2.3.3 *Programa Landsat*

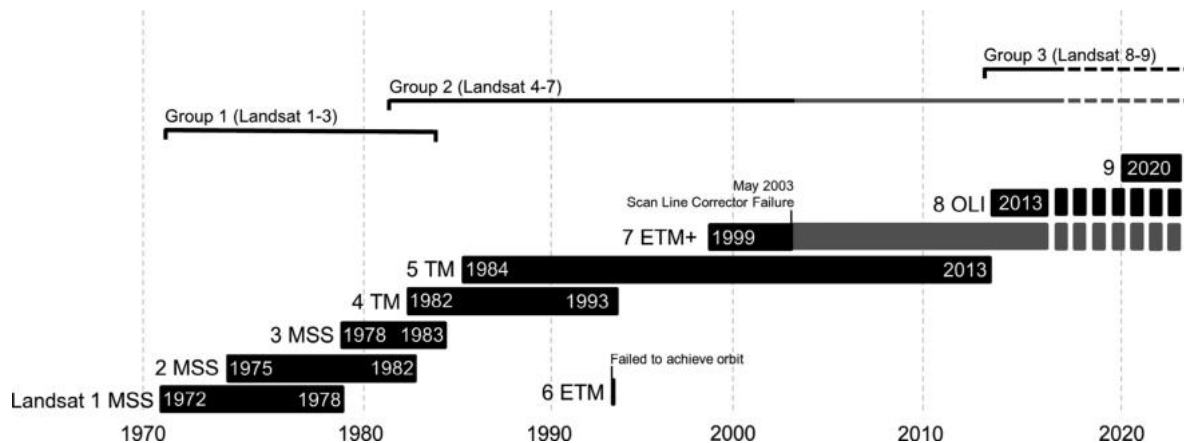
El programa Landsat es una de las misiones de observación de la Tierra de mayor continuidad temporal, gestionado conjuntamente por la Administración Nacional de Aeronáutica y el Espacio (NASA) y el Servicio Geológico de los Estados Unidos (USGS) (F. Zhao et al., 2024). Este programa proporciona el archivo histórico más extenso de imágenes satelitales de resolución moderada disponible a nivel global, con registros ininterrumpidos desde 1972 (Laar et al., 2024).

Las imágenes Landsat, con una resolución espacial nominal de 30 metros para la mayoría de sus bandas espectrales, son un recurso fundamental para estudios ecológicos, análisis de cambios de cobertura y uso del suelo (LULC), y gestión de recursos naturales. Aunque no permiten captar detalles muy finos, su resolución es adecuada para estudios a escala de paisaje, evitando al mismo tiempo los altos requerimientos de almacenamiento y procesamiento asociados a sensores de muy alta resolución. Además, la constelación Landsat tiene una frecuencia de revisita de aproximadamente 16 días, permitiendo el análisis temporal y fenológico del territorio (Young et al., 2017).

Los satélites Landsat se agrupan en tres generaciones principales según la evolución de sus plataformas e instrumentos (Figura 6). La Tabla 2 resume las bandas espectrales, rangos y resoluciones de los sensores más representativos del programa.

**Figura 6**

*Línea de Tiempo del Programa Landsat.*



*Nota. Fuente: Young et al., 2017.*

**Tabla 2**

*Bandas Espectrales de Sensores Landsat.*

Banda	LS 1–5 MSS	LS 4–5 TM	LS 7 ETM+	LS 8–9 OLI/TIRS	Pixel (m)
Coastal				B1 (0.43–0.45)	30
Blue		B1 (0.45–0.52)	B1 (0.45–0.52)	B2 (0.45–0.51)	30
Green	B1 (0.50–0.60)	B2 (0.52–0.60)	B2 (0.52–0.60)	B3 (0.53–0.59)	30 (60 <sup>†</sup> MSS)
Red	B2 (0.60–0.70)	B3 (0.63–0.69)	B3 (0.63–0.69)	B4 (0.64–0.67)	30 (60 <sup>†</sup> MSS)
NIR	B4 (0.80–1.10)	B4 (0.76–0.90)	B4 (0.77–0.90)	B5 (0.85–0.88)	30 (60 <sup>†</sup> MSS)
SWIR 1		B5 (1.55–1.75)	B5 (1.55–1.75)	B6 (1.57–1.65)	30
SWIR 2		B7 (2.08–2.35)	B7 (2.09–2.35)	B7 (2.11–2.29)	30
Thermal		B6 (10.40–12.50)	B6 <sup>‡</sup> (10.40–12.50)	B10/B11 (10.6–12.5)	30 <sup>†</sup>
Pan			B8 (0.52–0.90)	B8 (0.50–0.68)	15
Cirrus				B9 (1.36–1.38)	30

*Nota. Fuente: (Young et al., 2017). Los valores † indican resolución nativa de MSS.*

### 2.3.4 Programa Sentinel

El programa Sentinel constituye la principal constelación de observación de la Tierra de la Agencia Espacial Europea (ESA) dentro del marco del programa Copernicus (Gascon et al., 2015). A diferencia de Landsat, Sentinel se estructura en misiones especializadas por tipo de

aplicación y sensor.

**Sentinel-2.** La misión Sentinel-2 consta de dos satélites en operación (Sentinel-2A y Sentinel-2B), equipados con el instrumento multiespectral MSI, mediante el cual se registra información en 13 bandas distribuidas en el visible, infrarrojo cercano y SWIR. Las resoluciones espaciales de estas bandas son de 10 m, 20 m y 60 m, permitiendo una observación detallada de la estructura y dinámica de la vegetación. La Tabla 3 resume las características principales de estas bandas.

**Tabla 3**

*Bandas espectrales del instrumento MSI de Sentinel-2*

Banda	Escala	Resolución (m)	Longitud de onda (nm)	Descripción
B1	0.0001	60	444	Aerosoles
B2	0.0001	10	495	Azul
B3	0.0001	10	560	Verde
B4	0.0001	10	665	Rojo
B5	0.0001	20	704	Borde rojo 1
B6	0.0001	20	740	Borde rojo 2
B7	0.0001	20	780	Borde rojo 3
B8	0.0001	10	835	NIR (ancho)
B8A	0.0001	20	865	NIR (estrecho)
B9	0.0001	60	945	Vapor de agua
B11	0.0001	20	1610	SWIR 1
B12	0.0001	20	2190	SWIR 2

*Nota. Fuente: (European Union/ESA/Copernicus and Google Earth Engine, n.d.).*

### 2.3.5 Escenas (Tiles)

En el contexto del procesamiento de datos satelitales, el término “tile” se refiere a una subdivisión geográfica regular de la órbita del satélite. Es una unidad de archivo y procesamiento de datos geográficos, generalmente rectangular, que forma parte de un sistema de mosaico predefinido (tiling scheme), el cual cubre la superficie terrestre (Guth et al., 2021).

Características de los tiles:

1. Límites Geográficos: Los límites de los tiles suelen definirse en un sistema de coordenadas proyectado (como UTM) y forman una cuadrícula regular sobre el globo.
2. Nomenclatura Estándar: Los archivos de datos distribuidos suelen nombrarse utilizando un código que identifica unívocamente al tile, facilitando su catalogación y recuperación.

3. Contenido y Estructura: Un tile contiene los datos de reflectancia (u otras mediciones) para todas las bandas espectrales adquiridas por el sensor sobre esa área geográfica específica, organizados como una matriz o array de píxeles. Facilitan la gestión y el procesamiento distribuido de grandes volúmenes de datos.

La función principal de dividir los datos en tiles es optimizar la descarga, el almacenamiento, el procesamiento y la visualización de los datos, permitiendo trabajar con unidades manejables en lugar de franjas orbitales completas (Guth et al., 2021). La Figura 7 muestra un ejemplo visual de un tile de Sentinel-2 sobre el área de estudio.

### **Figura 7**

*Ejemplo de Escena (Tile) Sentinel-2 (27-06-2024).*



*Nota. Fuente: Elaboración propia.*

#### **2.3.6 Niveles de Procesamiento de Datos Satelitales**

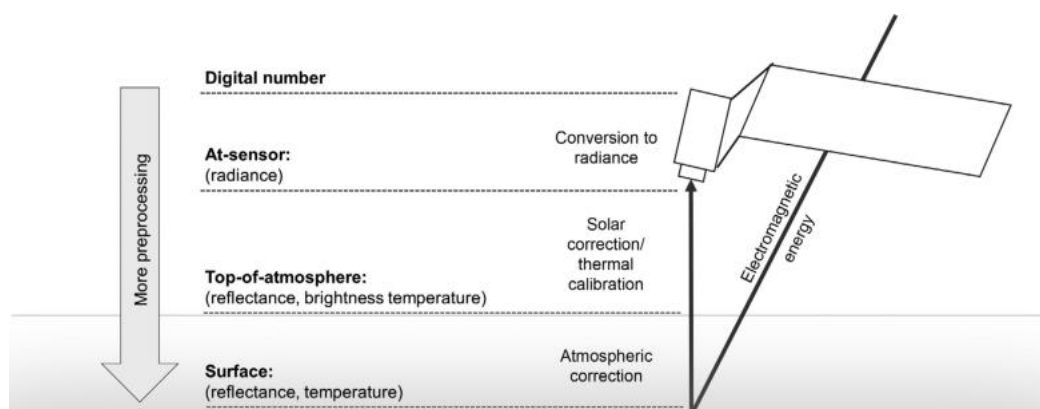
Los datos adquiridos por los sensores satelitales pasan por diferentes etapas de procesamiento antes de estar listos para su análisis temático. Estos niveles estandarizados indican el grado de corrección y calibración aplicado a los datos brutos. A continuación, se describen los principales niveles de procesamiento relevantes para imágenes ópticas (Liang et al., 2024):

- Nivel-0 (L0): Corresponde a los datos brutos (*raw data*) transmitidos desde el satélite. Están descomprimidos y formateados, pero no presentan correcciones radiométricas ni geométricas.
- Nivel-1 (L1): Incluye los datos L0 a los que se han aplicado correcciones radiométricas (conversión de niveles digitales a radiancia espectral) y correcciones geométricas (georreferenciación basada en parámetros orbitales y, ocasionalmente, puntos de control). Los productos L1 se entregan normalmente en Radiancia o Reflectancia en el Tope de la Atmósfera (TOA), lo que significa que aún incluyen los efectos atmosféricos como aerosoles, vapor de agua y dispersión.
- Nivel-2 (L2) y superiores: Derivan de los productos L1 tras aplicar corrección atmosférica y, en muchos casos, transformaciones adicionales para obtener variables biofísicas o geofísicas. Incluyen la Reflectancia Superficial (SR o BOA), que representa la reflectancia real de la superficie terrestre, libre de interferencias atmosféricas. Los productos L2 (como Sentinel-2 L2A o Landsat Collection 2 Nivel-2) son los más adecuados para análisis cuantitativos, clasificación LULC, cálculo de índices espectrales y estudios multitemporales debido a su mayor consistencia entre fechas.

La Figura 8 ilustra esquemáticamente los niveles de procesamiento.

**Figura 8**

*Niveles de procesamiento de Datos Satelitales.*



*Nota. Fuente: (Young et al., 2017)..*

### **2.3.7 Geospatial Big Data (GBD)**

El concepto de Geospatial Big Data (GBD) se aplica a datos geoespaciales que destacan por su elevado volumen, rapidez de producción, heterogeneidad y requerimientos computacionales avanzados para su procesamiento y análisis (J. Wu et al., 2024). Entre ellos se encuentran los archivos de imágenes satelitales multitemporales Landsat y Sentinel.

## **2.4 Sistemas de Información Geográfica**

Un Sistema de Información Geográfica (SIG) es un sistema informático diseñado para capturar, almacenar, manipular, analizar, administrar y presentar todo tipo de datos geográficamente referenciados (U.S. Geological Survey, 2023). Su función principal es conectar los datos descriptivos (atributos) a su ubicación espacial en un mapa, integrando la información geográfica con diversos tipos de información temática. Esta integración proporciona la base para la elaboración de mapas, el análisis espacial y la modelización, aplicable en una vasta gama de disciplinas científicas e industrias. Mediante los SIG se visualizan y comprenden patrones espaciales, relaciones geográficas y el contexto territorial de fenómenos diversos. Sus beneficios incluyen la mejora en la comunicación de información espacial, el aumento de la eficiencia en la gestión de datos territoriales y el apoyo a una toma de decisiones más informada (Esri, n.d.).

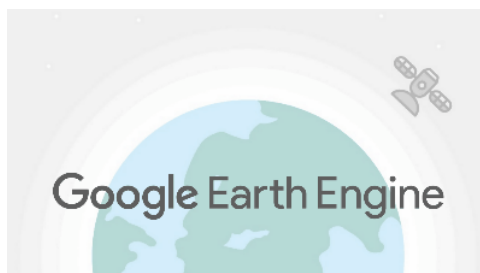
### **2.4.1 Google Earth Engine (GEE)**

Google Earth Engine (GEE), ilustrado en la Figura 9, es una plataforma basada en la nube que combina un catálogo de datos geoespaciales de escala planetaria con capacidades de cómputo para el análisis científico y la visualización. Es ampliamente utilizada por la comunidad académica, organizaciones sin fines de lucro, entidades gubernamentales y el sector privado. GEE aloja un extenso archivo de datos públicos, incluyendo series temporales históricas de imágenes satelitales que abarcan más de cuarenta años. Nuevas imágenes se incorporan continuamente, poniéndolas a disposición para análisis y minería de datos a escala global. Además del catálogo de datos, GEE ofrece Interfaces de Programación de Aplicaciones (APIs) y un entorno de desarrollo para facilitar el análisis de estos grandes conjuntos de datos

geoespaciales (Gorelick et al., 2017).

### **Figura 9**

*Google Earth Engine.*



*Nota. Fuente: Google Developers, 2023.*

#### **2.4.2 Editor de Código de Earth Engine**

El Editor de Código de Google Earth Engine es un entorno de desarrollo integrado (IDE) basado en la web, diseñado para escribir, ejecutar y depurar scripts utilizando la API de JavaScript de Earth Engine. Facilita el desarrollo interactivo de flujos de trabajo geoespaciales complejos de manera ágil y visual (Google Developers, 2023).

Por defecto, Google Earth Engine realiza los cálculos internamente utilizando el sistema de coordenadas geográficas WGS 84 (EPSG:4326), aunque permite la reproyección a otros sistemas según sea necesario, garantizando así la compatibilidad con una amplia gama de datos espaciales globales (Google Earth Engine Developers, 2025).

#### **2.4.3 QGIS**

QGIS es un Sistema de Información Geográfica (SIG) de escritorio, libre y de código abierto, disponible para múltiples plataformas (Windows, macOS y Linux) (Figura 11). Se utiliza en distintas etapas del flujo de trabajo geoespacial, abarcando tareas de preprocesamiento como la preparación de capas vectoriales (Khosravi, 2025), la manipulación y análisis de datos ráster (Fu et al., 2024), y la extracción de información puntual a partir de capas ráster (Buthelezi et al., 2024). Asimismo, permite la gestión y edición de datos vectoriales, la digitalización y la elaboración de mapas temáticos, soporta análisis espaciales avanzados y la recopilación de datos en campo (Haripavan et al., 2025).



**Figura 10**

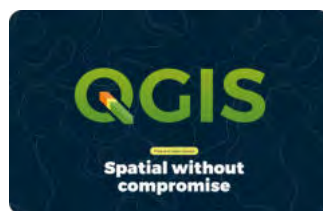
*Interfaz del Editor de Código de Google Earth Engine.*



*Nota. Fuente: Google Developers, 2023.*

**Figura 11**

*QGIS.*



*Nota. Fuente: U.S. Geological Survey, 2023.*

#### 2.4.4 Google Earth Pro

Google Earth Pro es una aplicación de escritorio que permite explorar un globo terráqueo virtual tridimensional (3D), construido a partir de imágenes satelitales, aéreas y datos SIG. Es ampliamente utilizado como herramienta de geovisualización y, en el contexto de la clasificación LULC supervisada, como fuente de datos de referencia (K. C. Roy et al., 2024). A diferencia de los SIG 2D tradicionales que utilizan proyecciones cartográficas planas, Google Earth emplea una proyección en perspectiva dinámica que simula la vista desde un punto elevado, actualizándose interactivamente a medida que el usuario navega (desplaza, inclina, acerca/aleja) (Sweet, 2011).

Su aplicación principal en flujos de trabajo LULC se centra en:

- **Obtención de Muestras de Entrenamiento:** Las imágenes de alta resolución histórica disponibles en Google Earth Pro sirven como base para la identificación visual y digitalización manual de puntos o polígonos de entrenamiento (*ground truth*). Los investigadores interpretan visualmente la cobertura del suelo en estas imágenes para generar los datos etiquetados necesarios para entrenar algoritmos de clasificación supervisada como Random Forest (Ganjirad, 2024; Ishtiaque et al., 2021).
- **Validación y Evaluación de Precisión:** De manera similar, las imágenes de alta resolución de Google Earth Pro se utilizan como fuente independiente de datos de referencia para validar la precisión de los mapas LULC generados a partir de imágenes satelitales de menor resolución. Se comparan puntos o áreas del mapa clasificado con la interpretación visual en Google Earth para construir la matriz de confusión y calcular las métricas de precisión (Mazroa et al., 2024).

Aunque la interfaz principal es 3D, la base de datos de imágenes subyacente de Google Earth utiliza internamente coordenadas geográficas (latitud/longitud) referidas al datum WGS 84 (EPSG:4326) (Sweet, 2011).

#### **2.4.5 Google Colab**

Google Colaboratory, comúnmente conocido como Google Colab, es un servicio gratuito basado en la nube que proporciona entornos de ejecución de cuadernos Jupyter (*Jupyter notebooks*). Se utiliza ampliamente en la comunidad científica y educativa para la implementación, ejecución y compartición de código, especialmente en los campos del Aprendizaje Automático (ML) y el Aprendizaje Profundo (Deep Learning, DL) (Lukas et al., 2024).

Está diseñado para facilitar tareas de ciencia de datos, investigación y educación, ofreciendo acceso gratuito (con ciertas limitaciones) a recursos computacionales de alto rendimiento, como Unidades de Procesamiento Gráfico (GPUs) y Unidades de Procesamiento Tensorial (TPUs), que son esenciales para entrenar modelos computacionalmente exigentes (Burke, 2023; Vidhya, 2020).

Características principales (Burke, 2023; Vidhya, 2020):

1. Integración con el Ecosistema de Google: Se integra de forma nativa con Google Drive para el almacenamiento y acceso a cuadernos y datos. Permite la carga de datos desde diversas fuentes.
2. Colaboración: Facilita la colaboración en tiempo real, permitiendo que varios usuarios editen el mismo cuaderno simultáneamente, de manera similar a Google Docs o Sheets.
3. Accesibilidad y Configuración Mínima: Al ser un servicio basado en la nube, es accesible desde cualquier dispositivo con un navegador web y conexión a internet, eliminando la necesidad de instalar software complejo o disponer de hardware potente localmente (Google Research, 2025). Proporciona entornos preconfigurados con librerías comunes de ciencia de datos (NumPy, Pandas, Scikit-learn, TensorFlow, PyTorch, etc.).

## 2.5 Tecnologías de Procesamiento Geoespacial y Científico

El desarrollo del presente estudio requirió la integración de lenguajes de programación de alto nivel y bibliotecas especializadas en cálculo numérico, análisis espacial y aprendizaje automático. Estas herramientas permitieron la orquestación de flujos de trabajo complejos, desde el preprocesamiento de imágenes satelitales hasta la modelización predictiva.

### 2.5.1 Lenguajes de Programación

**Python.** Python es un lenguaje de programación interpretado, de alto nivel y propósito general, es actualmente el lenguaje más utilizado en la ciencia de datos y la computación científica moderna. Su diseño prioriza la legibilidad del código y ofrece una vasta colección de bibliotecas para el manejo de matrices y datos geoespaciales (Python Software Foundation, 2025).

**JavaScript.** JavaScript es un lenguaje ligero, interpretado y orientado a objetos, conocido principalmente por su ejecución en entornos web (Mozilla Developer Network, 2025).

### 2.5.2 Librerías de Análisis y Procesamiento Científico

- NumPy: Es la biblioteca base para la computación científica en Python. Proporciona objetos de matriz multidimensionales (*arrays*) de alto rendimiento y herramientas para

trabajar con estos arreglos.

- SciPy: Construida sobre NumPy, esta librería ofrece algoritmos para optimización, integración, interpolación, problemas de valores propios y otras tareas avanzadas.
- Pandas: Proporciona estructuras de datos de alto nivel (DataFrames) para la manipulación y análisis de datos tabulares.

### **2.5.3 Librerías de Análisis Geoespacial y Aprendizaje Automático**

- GDAL (Geospatial Data Abstraction Library): Es una biblioteca de traducción para formatos de datos geoespaciales ráster y vectoriales.
- Scikit-learn: Es una biblioteca de aprendizaje automático que integra herramientas simples y eficientes para minería de datos y análisis.
- Matplotlib, Seaborn y Plotly: Conjunto de bibliotecas para la visualización de datos. *Matplotlib* y *Seaborn* se utilizaron para la generación de gráficos estáticos y estadísticos de alta calidad, mientras que *Plotly* integró capacidades interactivas para la exploración dinámica de los resultados.

## **2.6 Uso y Cobertura del Suelo (LULC)**

La cobertura del suelo (Land Cover) se refiere a los elementos biofísicos observados sobre la superficie terrestre, tales como la vegetación (bosques, pastizales), cuerpos de agua, suelo desnudo, nieve/hielo o infraestructura construida (Saoum & Sarkar, 2024). Por otro lado, el uso del suelo (Land Use) describe las actividades humanas o las funciones socioeconómicas que se desarrollan sobre un área determinada, implicando a menudo la gestión, transformación o alteración de la cobertura biofísica para la producción de bienes y servicios (agricultura, urbanización, silvicultura, conservación, etc.).

### **2.6.1 Enmascaramiento y Eliminación de Nubes**

El enmascaramiento de nubes es una etapa fundamental en el preprocesamiento de imágenes satelitales ópticas, ya que la presencia de nubes y sus sombras altera la señal espectral,

reduce la observabilidad de la superficie e introduce sesgos en los productos derivados (Garcia et al., 2025; Liang et al., 2024). Las técnicas utilizadas para este fin pueden agruparse en dos enfoques principales: (i) métodos basados en metadatos de calidad (QA/QC) y (ii) métodos multitemporales y de composición.

### **Métodos Multi-temporales y de Composición.**

- **Bandas de calidad (QA/QC):** Muchos productos satelitales incluyen bandas especializadas para la identificación de píxeles como nubes, sombras, nieve, agua o saturación. En sensores como Landsat, estas bandas codifican información binaria o multibit (QA\_PIXEL, QA\_RADSAT), permitiendo enmascarar nubes, sombras y píxeles saturados mediante operaciones lógicas (Liang et al., 2024).
- **Índices probabilísticos de claridad (Cloud Score+):** Cloud Score+ es un método basado en la combinación de múltiples métricas espectrales, estadísticas y físicas para obtener un valor continuo de “claridad”. Los píxeles con valores altos se consideran confiables, permitiendo generar máscaras más flexibles que las basadas únicamente en clasificaciones discretas.
- **Composición temporal (mediana o percentiles):** Es una de las técnicas más utilizadas, ya que reduce valores atípicos asociados a nubes y sombras, y genera representaciones más estables y realistas de la superficie terrestre (Garcia et al., 2025; Karurung et al., 2025).

#### **2.6.2 Remuestreo**

El remuestreo (*resampling*) constituye una etapa de preprocesamiento ampliamente empleada en teledetección cuando se integran múltiples conjuntos de datos con distintas resoluciones espaciales (Luo & Chen, 2025). Este procedimiento permite unificar la resolución espacial de las variables de entrada, con el fin de garantizar la coherencia analítica y la compatibilidad espacial entre las distintas fuentes (Rotich et al., 2025a).

### **Técnicas de Remuestreo.**

- Vecino más cercano (*Nearest Neighbor*): Es un método ampliamente utilizado por su fácil implementación y por la conservación de la información espectral original (Ettehad et al., 2019).
- Bilineal (*Bilinear Interpolation*): Utiliza el valor medio ponderado de los cuatro píxeles circundantes. Esto da como resultado píxeles suavizados (Guth et al., 2021).
- Bicúbica (*Bicubic Interpolation*): Es una técnica de orden superior que ofrece una calidad de imagen mejorada en comparación con las anteriores (Guria et al., 2024).

**Consecuencias del Cambio de Resolución.** El remuestreo influye directamente en la calidad de los datos:

- Introducción de distorsiones: El remuestreo de un mapa casi siempre resulta en distorsiones debido al desplazamiento de píxeles y ajustes de tamaño (Guth et al., 2021; Zambrano-Luna et al., 2025).
- Pérdida de detalle (*Downscaling - Aumento de tamaño de píxel*):
  - La conversión de una resolución fina a una más gruesa provoca una pérdida de detalles espaciales y un aumento en el número de píxeles mixtos (Ettehad et al., 2019), simplifica el análisis, pero conlleva la pérdida de detalles más finos (Pande et al., 2024).
  - El aumento del tamaño de la cuadrícula conduce a la pérdida de valores extremos (como elevaciones altas y valles), acercando los valores a la media (Guth et al., 2021).
- Mejora potencial del detalle (*Upscaling - Reducción de tamaño de píxel*):
  - Generalmente, se prefieren píxeles más pequeños porque tienen el potencial de acomodar un detalle más fino (Guth et al., 2021).
  - Se ha sugerido que el *downscaling* (hacia una resolución más fina) es más adecuado para la clasificación de cobertura terrestre, ya que utiliza por completo la información detallada de las bandas de alta resolución (Ettehad et al., 2019).

- Sin embargo, el remuestreo posterior a una resolución más pequeña (hacer un *upscaling* por interpolación) siempre hará que el tamaño del píxel y la resolución espacial intrínseca de los datos diverjan (Guth et al., 2021).

### 2.6.3 Índices Espectrales

Los Índices Espectrales (IE) son herramientas cuantitativas adimensionales, derivadas de datos de teledetección, diseñadas para realzar y caracterizar propiedades biofísicas específicas de la superficie terrestre (Ezekiel, 2017; Patle, 2024). Estos índices se calculan típicamente mediante operaciones algebraicas que combinan los valores de reflectancia superficial (SR) medidos en dos o más bandas espectrales (Ezekiel, 2017). Su incorporación en los procesos de clasificación supervisada contribuye en la mejora de la precisión de los mapas temáticos de Uso y Cobertura del Suelo (LULC) (K. C. Roy et al., 2024).

Su propósito es realzar fenómenos o características específicas de la superficie terrestre que no son evidentes en las bandas individuales, sirviendo como indicadores útiles para modelar o inferir dinámicas biofísicas (Patle, 2024). En el contexto de la clasificación LULC, la inclusión de índices espectrales como variables predictoras adicionales ha demostrado consistentemente mejorar la precisión de los mapas temáticos resultantes (Ganjirad, 2024).

La mayoría de los índices espectrales comúnmente utilizados se construyen bajo la forma de un Índice de Diferencia Normalizada (NDI), calculado como el cociente entre la diferencia y la suma de la reflectancia en dos bandas seleccionadas. Esta normalización produce valores que generalmente oscilan en el rango de -1 a +1 (Dahal et al., 2024). Un valor cercano a +1 suele indicar una alta probabilidad de presencia o una fuerte expresión de la característica biofísica que el índice está diseñado para detectar, mientras que valores bajos o negativos indican ausencia o baja expresión (Ezekiel, 2017). En la Tabla 4 se detallan los índices espectrales seleccionados por su capacidad para discriminar las coberturas de interés en el área de estudio.

***Índice de Vegetación de Diferencia Normalizada (NDVI):*** . Evalúa la presencia, vigor y densidad de vegetación fotosintéticamente activa (K. C. Roy et al., 2024).

- Valores altos: vegetación densa y saludable (Wijayanto et al., 2025).
- Valores negativos: superficies sin vegetación (suelo desnudo, roca, agua, nieve) (Pham

**Tabla 4**  
*Índices Espectrales (IE) utilizados en el estudio.*

IE	Nombre Completo	Fórmula
NDVI	Índice de Vegetación de Diferencia Normalizada	$\frac{NIR-RED}{NIR+RED}$ (Shrestha et al., 2023)
SAVI	Índice de Vegetación Ajustado al Suelo	$\frac{(NIR-RED)}{(NIR+RED+L)} \times (1 + L)$ (Belay et al., 2024)
NDMI	Índice de Humedad de Diferencia Normalizada	$\frac{NIR-SWIR1}{NIR+SWIR1}$ (Youssef, 2024)
NDBI	Índice de Diferencia Normalizada de Área Construida	$\frac{SWIR1-NIR}{SWIR1+NIR}$ (Mhangara et al., 2024)
UI	Índice Urbano	$\frac{SWIR2-NIR}{SWIR2+NIR}$ (M. S. Khan et al., 2021)
BSI	Índice de Suelo Desnudo (Normalizado)	$\frac{(RED+SWIR1)-(NIR+BLUE)}{(RED+SWIR1)+(NIR+BLUE)}$ (Ganjirad, 2024)

*Nota. NIR: Infrarrojo Cercano; RED: Rojo; BLUE: Azul; SWIR1: Infrarrojo de Onda Corta 1; SWIR2: Infrarrojo de Onda Corta 2; L: Factor de ajuste del suelo para SAVI (usualmente 0.5).*

& Ali, 2024).

**Índice de Vegetación Ajustado al Suelo (SAVI):** . Variante del NDVI que reduce la influencia del suelo en áreas con vegetación escasa. Incluye un factor de corrección  $L$  (típicamente  $L = 0.5$ ) (Belay et al., 2024).

**Índice de Diferencia Normalizada de Área Construida (NDBI):** . Utilizado para identificar zonas urbanas o construidas (*built-up*) (Dahal et al., 2024; K. C. Roy et al., 2024).

- Valores altos: áreas edificadas (Pham & Ali, 2024).
- Valores negativos: cuerpos de agua o vegetación densa (Dahal et al., 2024).

**Índice de Humedad de Diferencia Normalizada (NDMI):** . Informa sobre el contenido de humedad en la vegetación y el suelo (Ishtiaque et al., 2021; Kandulna et al., 2025).

- Valores altos: alta humedad (vegetación vigorosa, agua).
- Valores bajos: sequedad (suelo desnudo, vegetación estresada, zonas construidas).

**Índice de Suelo Desnudo (BSI):** . Destaca áreas desprovistas de vegetación (suelo expuesto, superficies erosionadas, caminos de tierra) (Shrestha et al., 2023).

**Índice Urbano (UI):** . Resalta características espectrales de zonas urbanas utilizando la diferencia entre SWIR2 y NIR (Hidalgo-García & Arco-Díaz, 2022).



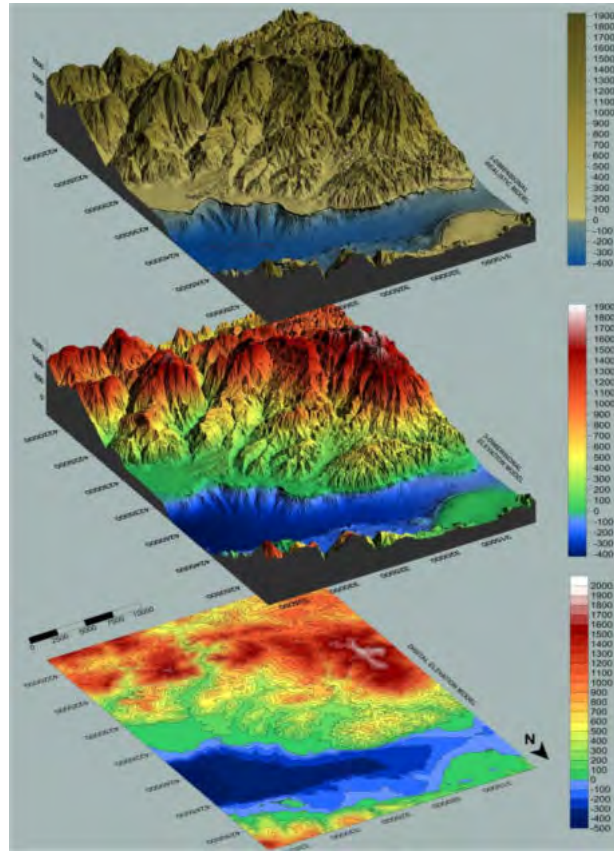
#### 2.6.4 Variables Topográficas

Las variables topográficas describen la forma y las características del relieve terrestre. Son factores importantes que influyen en diversos procesos ambientales y en la distribución espacial de las coberturas del suelo.

**Modelo Digital de Elevación (DEM).** Un Modelo Digital de Elevación (DEM - Digital Elevation Model) es una representación digital y rasterizada de la topografía de una superficie (Figura 12). Consiste en una matriz regular de celdas (píxeles), donde cada celda almacena un valor que representa la elevación (altitud, Z) en una ubicación geográfica específica (X, Y). Específicamente, un DEM representa la elevación del “terreno desnudo” (*bare earth*), es decir, la superficie del suelo sin incluir elementos sobrepuestos como vegetación (árboles, arbustos) o estructuras artificiales (edificios) (Equator Studios, 2023). Los DEM son utilizados en una amplia gama de disciplinas, incluyendo geomorfología, hidrología, planificación territorial, gestión de riesgos naturales y teledetección (Guth et al., 2021). Son producidos por diversas entidades, como agencias cartográficas nacionales, instituciones científicas y proveedores comerciales de datos geospaciales (Guth et al., 2021). Proporciona directamente la *Elevación* y se utiliza para derivar la *Pendiente (Slope)* y el *Aspecto (Aspect)* (Shen et al., 2024).

**ALOS World 3D-30m (ALOSDEM).** El ALOSDEM, proveniente del satélite Advanced Land Observing Satellite, es una fuente de datos topográficos de alta resolución, frecuentemente citado con una resolución de  $12.5 \times 12.5$  metros (Shen et al., 2024). Los datos del sensor ALOS PALSAR (banda L) se han integrado con datos LiDAR y climáticos para mejorar los mapas de biomasa aérea sobre bosques secos tropicales, lo que demuestra su valor en el monitoreo forestal de LULC (Singh et al., 2024).

**Copernicus Global DEM (CGDEM).** Los productos de elevación del Programa Copernicus se utilizan tanto como fuente de variables topográficas como para fines de validación de otros productos de cobertura terrestre, con una resolución de 30 metros (Z. Xu et al., 2025).

**Figura 12***Modelo Digital de Elevación (DEM).**Nota. Fuente: Simou et al., 2014.*

### 2.6.5 Análisis de Cambios Multitemporal

El análisis de cambios multitemporal, también conocido como detección de cambios (change detection), es una metodología central en teledetección y SIG que consiste en analizar imágenes o datos geospaciales adquiridos sobre la misma área geográfica en diferentes momentos para identificar, cuantificar y caracterizar las transformaciones ocurridas entre esas fechas (Duan et al., 2025). Su propósito principal es determinar qué ha cambiado, dónde ha cambiado y cuál ha sido la magnitud de dicho cambio en un período específico (Tadesse, 2024). Este análisis permite comprender la dinámica espacio-temporal de los procesos de transformación del paisaje (Sarif & Gupta, 2024). Además, el análisis histórico de cambios constituye la base para la calibración y validación de modelos de simulación y predicción de escenarios futuros de LULC (Shrestha et al., 2023).

**Matriz de Probabilidad de Transición.** La Matriz de Probabilidad de Transición es una herramienta cuantitativa fundamental derivada del análisis de cambios LULC entre dos fechas, utilizada para describir la dinámica de las transformaciones observadas. Es un componente clave en los modelos de Cadenas de Markov aplicados a la simulación LULC (Toscan et al., 2025).

1. La matriz de probabilidad de transición representa las probabilidades estimadas de que un píxel que pertenecía a la clase de cobertura  $i$  en el tiempo inicial ( $t_1$ ) transite a la clase  $j$  en el tiempo final ( $t_2$ ) (Mansour et al., 2025).

2. Estructura e Interpretación:

- Es una matriz cuadrada donde las filas representan las clases LULC en el tiempo inicial ( $t_1$ ) y las columnas representan las mismas clases en el tiempo final ( $t_2$ ) (Toscan et al., 2025).
- Entradas Diagonales ( $P_{ii}$ ): Los elementos a lo largo de la diagonal principal representan la probabilidad de persistencia o estabilidad de cada clase. Un valor cercano a 1 indica que una alta proporción del área de esa clase en  $t_1$  permaneció sin cambios en  $t_2$  (Pham & Ali, 2024).
- Entradas Fuera de la Diagonal ( $P_{ij}, i \neq j$ ): Representan las probabilidades de transición de la clase  $i$  a la clase  $j$ . Indican la proporción del área de la clase  $i$  en  $t_1$  que se transformó a la clase  $j$  en  $t_2$  (Duan et al., 2025).
- Propiedad Matemática: Cada elemento  $P_{ij}$  es una probabilidad, por lo que su valor está comprendido entre 0 y 1 ( $0 \leq P_{ij} \leq 1$ ). Además, la suma de las probabilidades a lo largo de cada fila debe ser igual a 1 ( $\sum_{j=1}^n P_{ij} = 1$ ), indicando que toda el área de la clase  $i$  en  $t_1$  debe asignarse a alguna clase (incluida ella misma) en  $t_2$  (Mansour et al., 2025).

**Consideraciones sobre la Escala Temporal.** La elección de los intervalos de tiempo para el análisis de cambios y la derivación de matrices de transición es un aspecto metodológico

importante. Diferentes escalas temporales pueden revelar distintos aspectos de la dinámica LULC, y una elección inadecuada puede afectar la interpretación de los resultados:

- **Intervalos Cortos:** Un período de tiempo muy corto entre observaciones puede capturar fluctuaciones estacionales o cambios menores, pero podría no reflejar tendencias estructurales a largo plazo. Además, puede magnificar el impacto del “ruido” o errores de clasificación entre fechas (Guo & Shen, 2024).
- **Intervalos Largos:** Un período de tiempo muy extenso puede promediar u ocultar cambios intermedios importantes y dificultar la atribución de los cambios a factores causales específicos (Guo & Shen, 2024).
- **Práctica Común:** En estudios de modelado y predicción LULC, es común utilizar intervalos de análisis histórico que oscilan entre 5 y 15 años para calibrar los modelos de transición que luego se usan para proyectar cambios futuros (Bendeckou et al., 2024).

## 2.7 Aprendizaje Automático

El Aprendizaje Automático (Machine Learning, ML) es un subcampo de la Inteligencia Artificial (IA) (Qamar & Zardari, 2023). Los modelos de Aprendizaje Automático comprenden un conjunto de metodologías computacionales impulsadas por datos, diseñadas para identificar patrones, relaciones y estructuras subyacentes en conjuntos de datos multidimensionales (Amindin et al., 2024).

Existen diversas categorías de aprendizaje automático, siendo las principales:

- **Aprendizaje Supervisado:** En esta modalidad, el algoritmo se entrena utilizando un conjunto de datos previamente etiquetado, donde cada ejemplo de entrada está asociado a una salida o etiqueta conocida. El objetivo es que el modelo aprenda una función de mapeo que pueda generalizar a nuevos datos no vistos.
- **Aprendizaje No Supervisado:** En contraste, este enfoque utiliza datos no etiquetados. El algoritmo explora los datos de manera autónoma para descubrir patrones, agrupaciones o estructuras inherentes sin guía explícita sobre las salidas correctas.

Para tareas de clasificación, como la identificación de coberturas del suelo, la clasificación supervisada generalmente presenta resultados con mayor precisión (accuracy) en comparación con los métodos no supervisados (Xiang, 2024).

### ***2.7.1 Métodos de Muestreo***

Existen diversas estrategias para la selección de datos de muestra, entre ellas el Muestreo Aleatorio Simple (Simple Random Sampling, SRS), el Muestreo Aleatorio Estratificado (Stratified Random Sampling, STRAT) y el Muestreo Sistemático (Systematic Sampling, SYSTEM).

**Muestreo Aleatorio Estratificado (Stratified Random Sampling, STRAT).** El muestreo aleatorio estratificado (STRAT) resulta especialmente adecuado cuando se requiere garantizar una representación equilibrada de todas las categorías presentes en el área de estudio. Al dividir la población en grupos homogéneos reduce el sesgo, mejora la robustez de la evaluación de la precisión y asegura una distribución proporcional de los puntos de muestra, incrementando la fiabilidad del proceso de validación (Wijayanto et al., 2025).

- Se utiliza ampliamente en mapeo LULC para la generación de muestras de entrenamiento y prueba (Khosravi, 2025).
- La estratificación asegura una representación equilibrada de todas las clases temáticas, reduciendo el sesgo y mejorando la precisión de las evaluaciones (Wijayanto et al., 2025).
- Al asignar puntos de muestra de forma proporcional a cada categoría, se incrementa la consistencia y fiabilidad de la validación de clasificaciones (Ahmad et al., 2025; Wijayanto et al., 2025).

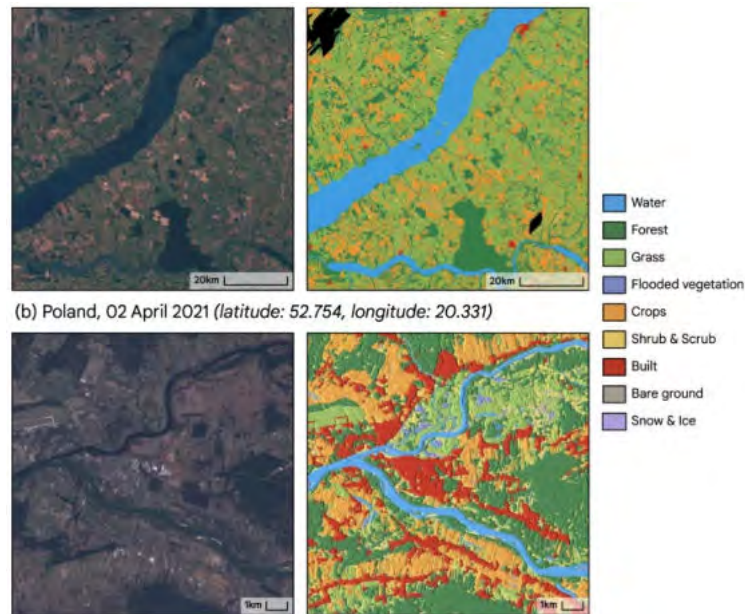
### ***2.7.2 Clasificación Supervisada de Coberturas del Suelo***

La clasificación supervisada es un método ampliamente usado en teledetección y ML para el análisis cuantitativo y la segmentación temática de imágenes satelitales (Patel & Vyas,

2024). Los mapas resultantes de la clasificación LULC proporcionan información que permite identificar y cuantificar el tipo, la ubicación, el uso y la extensión de las diversas características biofísicas (naturales) y antrópicas (creadas por el ser humano) sobre la superficie terrestre (Figura 13).

### Figura 13

*Ejemplo clasificación de Coberturas del Suelo.*



*Nota. Fuente: Brown et al., 2022.*

El proceso metodológico de la clasificación supervisada se fundamenta en las siguientes etapas:

1. Se selecciona un conjunto de áreas o puntos de muestra representativos dentro de la imagen, correspondientes a las distintas categorías LULC de interés (clases). La calidad y representatividad de estas muestras son críticas para el desempeño del clasificador (Xiang, 2024).
2. Utilizando las muestras de entrenamiento etiquetadas, el algoritmo clasificador de ML aprende a identificar los patrones característicos (espectrales, texturales, contextuales) asociados a cada clase (Sarif & Gupta, 2024).
3. Una vez entrenado, el modelo establece una función discriminante  $f(\mathbf{x})$  que se aplica a cada píxel  $\mathbf{x}$  de la imagen completa. Esta función asigna automáticamente cada píxel a

una de las  $m$  categorías predefinidas  $C_k$ , segmentando así la imagen en un mapa temático LULC (M. Khan et al., 2023):

$$\text{Píxel } \mathbf{x} \in \mathbb{R}^n \rightarrow C_k \in \{C_1, C_2, \dots, C_m\} \quad (1)$$

donde  $\mathbf{x}$  es el vector de características del píxel (valores de reflectancia en  $n$  bandas) y  $C_k$  es la clase asignada.

**Algoritmo Bosque Aleatorio (Random Forest, RF).** El Bosque Aleatorio (Random Forest, RF) es un algoritmo de aprendizaje automático supervisado, no paramétrico, propuesto por Breiman en 2001 (K. C. Roy et al., 2024). Es un método de aprendizaje por conjuntos (*ensemble learning*) que construye y combina las predicciones de múltiples árboles de decisión (Decision Trees, DTs) individuales para mejorar la precisión general, la robustez y la estabilidad de las predicciones (Ganjirad, 2024; Sharma et al., 2024).

Durante el entrenamiento, se crean múltiples subconjuntos de datos mediante muestreo aleatorio con reemplazo a partir del conjunto de entrenamiento original. Cada árbol de decisión del bosque se entrena de forma independiente utilizando uno de estos subconjuntos (Chen, 2024). Además, en cada nodo de cada árbol, la división óptima se busca solo sobre un subconjunto aleatorio de las variables predictoras disponibles, lo que introduce una segunda fuente de aleatoriedad y ayuda a decorrelacionar los árboles. La predicción final del bosque se obtiene agregando las predicciones de todos los árboles individuales mediante el voto mayoritario en tareas de clasificación, o calculando la media en tareas de regresión (Arunab & Mathew, 2024; Sharma et al., 2024). Esta estrategia combinada confiere al RF una notable versatilidad para abordar una amplia gama de problemas (DataScientest, 2025).

La estructura conceptual de un Bosque Aleatorio se puede representar como un conjunto de árboles:

$$\{h(\mathbf{x}, \Theta_k), k = 1, \dots, K\} \quad (2)$$

donde  $h(\mathbf{x}, \Theta_k)$  representa el  $k$ -ésimo árbol de decisión entrenado con un vector de parámetros aleatorios  $\Theta_k$  (que incluye el subconjunto de datos bootstrap y las variables consideradas en

cada nodo),  $\mathbf{x}$  es el vector de variables de entrada para un píxel, y  $K$  es el número total de árboles en el bosque (K. C. Roy et al., 2024). La predicción final se basa en la agregación de las salidas  $h(\mathbf{x}, \Theta_k)$  para  $k = 1, \dots, K$ .

El algoritmo RF presenta varias características que explican su amplio uso en clasificación LULC:

- Frecuentemente alcanza precisiones de clasificación superiores a otros algoritmos como SVM o CART, logrando a menudo Precisiones Globales (OA) por encima del 90 % y coeficientes Kappa ( $\kappa$ ) superiores a 0.85 en aplicaciones LULC (Zafar, Zubair, Zha, Fahd & Nadeem, 2024).
- Es inherentemente resistente al ruido en los datos, a la presencia de valores atípicos (*outliers*) y exhibe una baja tendencia al sobreajuste (*overfitting*) comparado con un único árbol de decisión, lo que mejora la capacidad de generalización del modelo a datos no vistos (Amindin et al., 2024; IBM, 2025; Vahid & Aly, 2025).
- Puede procesar eficazmente conjuntos de datos con un gran número de variables predictoras (alta dimensionalidad), como los derivados de imágenes multiespectrales con múltiples índices y variables auxiliares. Tolera adecuadamente la multicolinealidad entre variables (Amindin et al., 2024; Zafar, Zubair, Zha, Fahd & Nadeem, 2024).
- Es paralelizable y generalmente requiere menos ajuste fino de hiperparámetros en comparación con otros métodos como SVM o redes neuronales profundas (Badshah et al., 2024; Wahdatyar, 2024; Wahdatyar et al., 2024).
- RF proporciona una medida intrínseca de la importancia relativa de cada variable predictora en el modelo, lo cual es útil para la interpretación del modelo y la selección de características (A et al., 2024; Chen, 2024).

### 2.7.3 Modelos ML para Predicción de Transiciones LULC

**Redes Neuronales Artificiales (ANN).** Actualmente se utiliza con frecuencia el aprendizaje automático para derivar las reglas de transición de manera empírica a partir de datos históricos



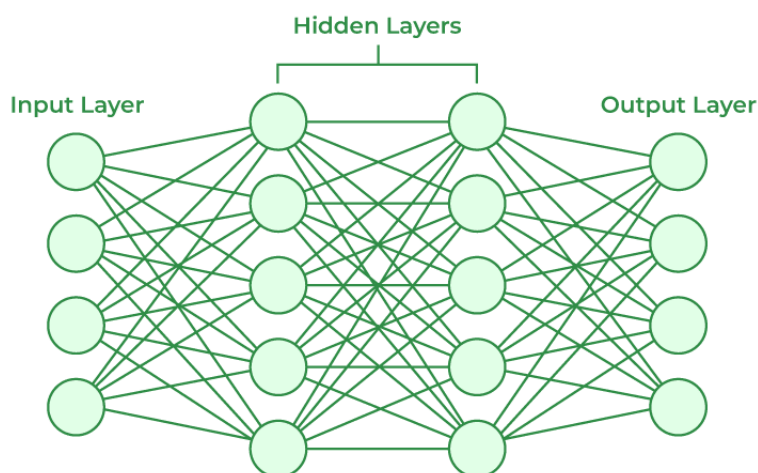
de cambios en la cobertura y uso del suelo (LULC) y de sus factores impulsores. Dentro de estos métodos, las Redes Neuronales Artificiales (ANN) resultan idóneas para modelar, debido a su capacidad para modelar relaciones no lineales y de alta complejidad entre variables predictoras y la probabilidad de transición LULC (Suthar et al., 2024; M. Zhang et al., 2024).

**Perceptrón Multicapa (MLP).** El Perceptrón Multicapa (MLP) es un modelo representativo dentro de las Redes Neuronales Artificiales de tipo feed-forward (Bhuyan et al., 2024; G. Liu et al., 2024; Xiang et al., 2024). Se caracteriza por su estructura sencilla, su facilidad de entrenamiento, un costo computacional moderado y tiempos de predicción eficientes, lo que ha favorecido su aplicación en tareas como la simulación de la expansión urbana (T. Xu et al., 2022).

El objetivo principal del MLP es modelar relaciones complejas y no lineales entre variables para la generación de mapas de potencial de transición entre clases de cobertura y uso del suelo (Bendeckou et al., 2024). Cada mapa expresa, para cada unidad espacial de análisis, la aptitud relativa para transicionar hacia una clase futura específica. Estos mapas actúan como el conjunto de reglas espaciales ( $f$ ) que guían la asignación de cambios realizada por el Autómata Celular durante la simulación del período siguiente ( $t_2 \rightarrow t_3$ ) (Naeem et al., 2025). La Figura 14 presenta una arquitectura general de una ANN.

**Figura 14**

*Arquitectura general de una Red Neuronal Artificial.*



*Nota. La red está compuesta por capas de entrada, ocultas y de salida. Fuente: geeksforgeeks, 2024.*

El entrenamiento de un MLP se realiza habitualmente mediante el algoritmo de retropropagación (backpropagation), que ajusta los pesos de la red para minimizar una función de error (Ishtiaque et al., 2021; K. C. Roy et al., 2024).

El optimizador es responsable de la actualización de los pesos durante el entrenamiento. El Descenso de Gradiente Estocástico (SGD) es uno de los métodos más utilizados (Chaulagain et al., 2025). La incorporación del término de momento ( $\alpha$ ) permite acelerar la convergencia del algoritmo en direcciones coherentes del gradiente y reducir oscilaciones, lo que mejora la eficiencia del entrenamiento (Ishtiaque et al., 2021; K. C. Roy et al., 2024; Shen et al., 2024).

La función de activación introduce la no linealidad necesaria para que el modelo pueda capturar relaciones complejas entre las variables (Hussain et al., 2025). La función tangente hiperbólica (Tanh) se emplea con frecuencia en las capas ocultas, dado que transforma los valores en el intervalo entre -1 y 1 (Alam & Maiti, 2025; Fu et al., 2024), lo que puede contribuir a estabilizar el entrenamiento. Diversos estudios han reportado resultados favorables del uso de Tanh en arquitecturas ANN para predicciones de transición LULC (Tasan et al., 2025).

La función Softmax se utiliza comúnmente en la capa de salida de modelos de clasificación multiclase, ya que transforma un vector de valores no normalizados en una distribución de probabilidad normalizada (contributors, 2025). Esta función garantiza que todas las probabilidades se encuentren en el rango  $[0, 1]$  y que su suma sea igual a 1:

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (3)$$

El tamaño del lote determina cuántas muestras del conjunto de entrenamiento se procesan antes de actualizar los pesos (Hussain et al., 2025). Un tamaño de lote igual a 1 corresponde al aprendizaje estocástico. Este enfoque realiza una actualización de los pesos después de cada muestra individual, lo que introduce variabilidad en el proceso de optimización y permite una adaptación más fina de los parámetros del modelo.

#### **2.7.4 Métodos de Búsqueda de Hiperparámetros (HPO)**

La búsqueda de hiperparámetros es un paso esencial en el aprendizaje automático (Machine Learning - ML) para lograr una alta precisión y un rendimiento óptimo de los

modelos (B. Roy, 2021). Los hiperparámetros son propiedades o parámetros del modelo que se configuran antes del entrenamiento y controlan el proceso de aprendizaje supervisado, afectando directamente la precisión (B. Roy, 2021).

### **Búsqueda en Cuadrícula (Grid Search).**

- Es un enfoque sistemático diseñado para identificar los mejores hiperparámetros. Su método consiste en evaluar cada combinación posible de valores de parámetros que han sido predefinidos dentro de un rango específico (Hanh et al., 2025).
- A pesar de su simplicidad, el método de *Grid Search* a menudo proporciona resultados fiables (Hanh et al., 2025).
- Este método evalúa el rendimiento del modelo (generalmente la precisión) mediante el uso de validación cruzada (cross-validation) (Maddah et al., 2025). Después de probar todas las combinaciones, identifica el conjunto de hiperparámetros que proporciona el mayor rendimiento (Maddah et al., 2025).

## **2.8 Modelos Híbridos para Simulación Espacio - Temporal**

La simulación espacio-temporal busca representar explícitamente tanto la evolución temporal como la distribución espacial de los cambios LULC (K. C. Roy et al., 2024).

Para capturar la complejidad de las dinámicas LULC, que involucran interacciones no lineales entre factores biofísicos y socioeconómicos a diferentes escalas, los modelos híbridos han demostrado ser particularmente efectivos. Estos modelos integran las fortalezas de diferentes enfoques algorítmicos (Risma, 2019). Un enfoque híbrido prominente y ampliamente utilizado es la combinación de Autómatas Celulares (CA) y Cadenas de Markov (MC), a menudo potenciado con técnicas de aprendizaje automático para definir las reglas de transición espacial.

### **2.8.1 Modelo Híbrido CA-Markov**

El modelo CA-Markov es un método eficaz y frecuentemente empleado que fusiona dos componentes metodológicos complementarios para simular la dinámica LULC de manera integrada en el espacio y el tiempo (Duan et al., 2025).

### 2.8.2 *Componente Temporal: Cadenas de Markov y Matrices de Transición*

Una Cadena de Markov es un proceso estocástico que modela la probabilidad de transición entre diferentes estados a lo largo del tiempo, asumiendo que el estado futuro depende únicamente del estado actual (propiedad Markoviana). Aplicado al análisis LULC entre dos fechas ( $t_1$  y  $t_2$ ), el análisis de Markov cuantifica empíricamente la probabilidad de que una unidad de área (píxel) perteneciente a la clase  $i$  en  $t_1$  transite a la clase  $j$  en  $t_2$  (Pham & Ali, 2024; M. Zhang et al., 2024). El resultado principal de este análisis es la Matriz de Probabilidad de Transición ( $P_{ij}$ ), la cual resume estas probabilidades y define la cantidad total de cambio esperado (“demanda de transición”) entre cada par de clases para el siguiente intervalo de tiempo (Mansour et al., 2025). Sin embargo, el modelo de Markov intrínsecamente carece de capacidad espacial, predice cuánto cambio ocurrirá, pero no dónde ocurrirá, asumiendo implícitamente una distribución espacial aleatoria de las transiciones (Kashani et al., 2025).

### 2.8.3 *Componente Espacial: Autómatas Celulares (CA)*

Para superar la limitación espacial del modelo de Markov, se integra un Autómata Celular. Un CA es un modelo dinámico espacialmente explícito, basado en una cuadrícula regular de celdas (píxeles). El estado futuro de una celda en el tiempo  $t+1$  ( $S_{t+1}$ ) está determinado por su propio estado actual ( $S_t$ ) y el estado de sus celdas vecinas ( $N$ ), de acuerdo con un conjunto predefinido de reglas de transición locales ( $f$ ) (Z. Yang et al., 2023; X. Zhang et al., 2023). Formalmente, su funcionamiento se puede describir como:

$$S_{t+1} = f(S_t, N) \quad (4)$$

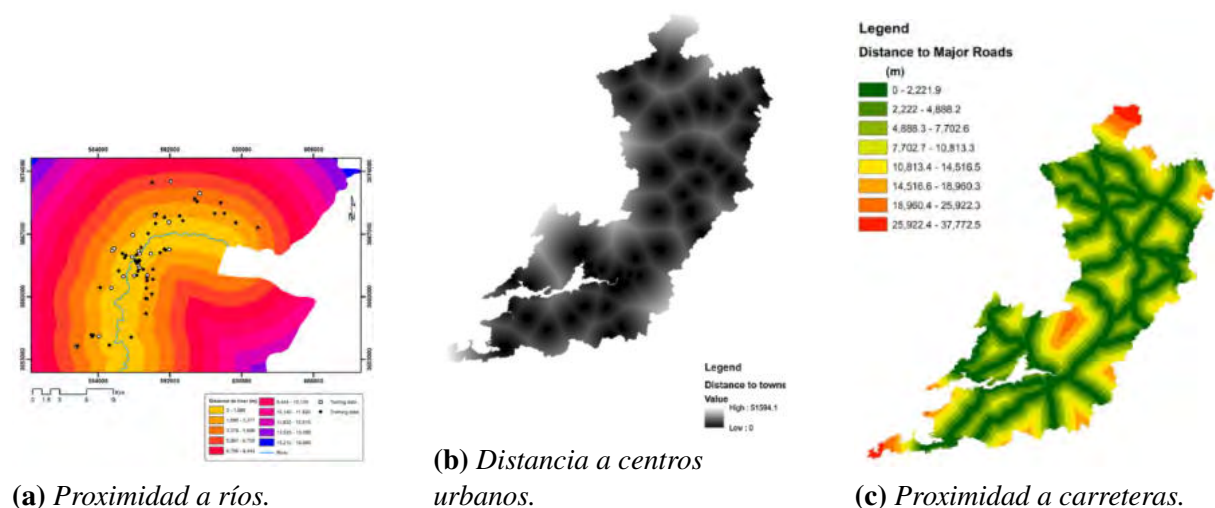
En el modelo híbrido CA-Markov, el CA actúa como el motor de asignación espacial. Utiliza la cantidad total de cambio por transición calculada por la Cadena de Markov (la “demanda”) y la distribuye espacialmente sobre la cuadrícula. Esta asignación no es aleatoria, sino que se basa en la “aptitud” o “idoneidad” de cada celda para experimentar una transición específica, la cual se determina mediante las reglas de transición ( $f$ ). Estas reglas suelen incorporar información sobre factores locales y la influencia del vecindario (Belay et al., 2024; Kashani et al., 2025).

### 2.8.4 Variables Predictoras

Las variables predictoras, también denominadas factores impulsores (*driving factors* o *drivers*), son las variables espaciales que representan los factores biofísicos y socioeconómicos. Estas variables sirven como entradas para entrenar el modelo de aprendizaje automático que genera los mapas de potencial de transición (Alqadhi et al., 2024; Xiang, 2024).

- Factores Topográficos: Derivados de un DEM, como la elevación, la pendiente y el aspecto (orientación de la ladera).
- Factores de Proximidad o Accesibilidad: Distancia euclidiana a elementos lineales (carreteras, ríos) o puntuales (centros urbanos). La Figura 15 muestra ejemplos visuales de mapas de proximidad.
- Factores Socioeconómicos: Densidad de población, precio del suelo, zonificación, políticas de uso del suelo.
- Factores Ambientales: Tipo de suelo, precipitación, temperatura.

**Figura 15**  
*Ejemplos de Variables Predictoras de Proximidad.*



*Nota. Mapas basados en distancia euclidiana. Fuente (a) y (c): Al-Abadi, 2015. Fuente (b): Gharbia et al., 2016.*

### 2.8.5 *Análisis de Colinealidad*

Se refiere al fenómeno estadístico donde dos o más variables predictoras exhiben una fuerte correlación (Z. Xu et al., 2025), lo que introduce información redundante en el modelo (D. Zhao et al., 2025).

El objetivo principal de este análisis es asegurar la validez de la predicción (Hanh et al., 2025). Los modelos de aprendizaje automático son sensibles a la multicolinealidad, la cual puede distorsionar la precisión (Alam & Maiti, 2025), conducir al sobreajuste (overfitting) (Buthelezi et al., 2024) y dificultar la interpretación de la contribución individual de cada variable (Z. Xu et al., 2025). Si se detecta una alta correlación, es necesario eliminar una de las variables en conflicto para evitar resultados sesgados (Danso et al., 2025).

**Coeficiente de Correlación de Pearson (r).** El Coeficiente de Correlación de Pearson (PCC) es uno de los métodos más usados para medir la relación lineal entre dos variables (Hanh et al., 2025; Naeem et al., 2025). Se considera que existe una fuerte correlación cuando el valor absoluto  $|r| > 0.7$  o  $|r| > 0.8$  (Buthelezi et al., 2024; Hanh et al., 2025; Tayyab et al., 2024).

El coeficiente se calcula dividiendo la covarianza de las dos variables por el producto de sus desviaciones estándar (Badavath & Sahoo, 2025)

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Donde  $r$  es el coeficiente de correlación,  $x_i$  y  $y_i$  son los valores de las variables  $x$  e  $y$ ,  $\bar{x}$  y  $\bar{y}$  representan los valores promedio de cada variable,  $n$  representa el número de observaciones para  $x$  y  $y$  (Naeem et al., 2025).

## 2.9 Métricas de Evaluación de Modelos

### 2.9.1 Métricas de Evaluación de la Clasificación

**Matriz de Confusión (*Confusion Matrix*, MC).** La Matriz de Confusión es la herramienta más utilizada para evaluar el desempeño de un modelo de clasificación supervisada (Patel & Vyas, 2024).

Propósito y Estructura:

- Es una matriz de tamaño  $n \times n$ , donde  $n$  es el número de clases temáticas. Compara las clases reales o de referencia (verdad de terreno, usualmente en las filas) con las clases asignadas por el modelo (predicciones, usualmente en las columnas).
- Los elementos a lo largo de la diagonal principal ( $MC_{ii}$ ) representan el número de muestras (píxeles) que fueron clasificadas correctamente, es decir, donde la clase predicha coincide con la clase real (Ezekiel, 2017). Los elementos fuera de la diagonal ( $MC_{ij}, i \neq j$ ) representan las clasificaciones erróneas (confusiones entre clases).

La matriz de confusión permite derivar cuatro conteos básicos (Anzalone et al., 2024):

1. Verdadero Positivo (VP o TP - *True Positive*): Número de muestras positivas que fueron correctamente clasificadas como positivas.
2. Verdadero Negativo (VN o TN - *True Negative*): Número de muestras negativas que fueron correctamente clasificadas como negativas.
3. Falso Positivo (FP - *False Positive*): Número de muestras negativas que fueron incorrectamente clasificadas como positivas. También se conoce como Error de Tipo I o Error de Comisión.
4. Falso Negativo (FN - *False Negative*): Número de muestras positivas que fueron incorrectamente clasificadas como negativas. También se conoce como Error de Tipo II o Error de Omisión.

**Métricas de Precisión Derivadas de la Matriz de Confusión.** La Matriz de Confusión permite evaluar el desempeño de un clasificador (Kondum et al., 2024). Entre las más utilizadas se encuentra la Precisión Global (*Overall Accuracy*, OA), que resume la proporción total de muestras correctamente clasificadas (Tola & Deyassa, 2024; X. Wu et al., 2024). Aunque es sencilla e intuitiva, puede resultar insuficiente cuando existe desbalance entre clases (Ezekiel, 2017).

**Precisión Global (*Overall Accuracy*, OA).** La OA se obtiene dividiendo la suma de los elementos de la diagonal principal de la matriz por el número total de muestras evaluadas (Ishtiaque et al., 2021). Su valor oscila entre 0 (clasificación completamente incorrecta) y 1 (clasificación perfecta), o entre 0 % y 100 %.

$$OA = \frac{\sum_{i=1}^n MC_{ii}}{N} = \frac{TP + TN}{TP + FP + TN + FN} \quad (\times 100 \%) \quad (5)$$

donde  $MC_{ii}$  son los elementos diagonales de la matriz,  $N$  es el total de muestras, y la expresión alternativa corresponde al caso binario (Gündüz, 2025; Shrestha et al., 2023).

**Precisión del Usuario (*User Accuracy*, UA) o Precisión (*Precision*).** La UA mide la fiabilidad del mapa desde la perspectiva del usuario (Gündüz, 2025). Para una clase  $i$ , representa la proporción de muestras clasificadas como esa clase que realmente pertenecen a ella (Ezekiel, 2017; K. C. Roy et al., 2024). Refleja la probabilidad de que un píxel etiquetado como clase  $i$  sea correcto, y está inversamente asociada al *Error de Comisión* (falsos positivos) (Anzalone et al., 2024).

$$UA \text{ (Precision)} = \frac{MC_{ii}}{\sum_{k=1}^n MC_{ki}} = \frac{TP}{TP + FP} \quad (\times 100 \%) \quad (6)$$

donde  $\sum_k MC_{ki}$  corresponde al total de muestras clasificadas como clase  $i$  (total de la columna).

**Precisión del Productor (*Producer Accuracy*, PA) o Exhaustividad (*Recall / Sensibilidad*).** La PA evalúa qué tan bien el mapa representa la realidad para una clase específica (Gündüz, 2025). Para la clase  $i$ , es la proporción de muestras de referencia de esa clase que fueron correctamente identificadas en el mapa (K. C. Roy et al., 2024). Está inversamente asociada al *Error de Omisión* (falsos negativos) (Anzalone et al., 2024).



$$PA \text{ (Recall)} = \frac{MC_{ii}}{\sum_{j=1}^n MC_{ij}} = \frac{TP}{TP + FN} \quad (\times 100 \%) \quad (7)$$

donde  $\sum_j MC_{ij}$  corresponde al total de muestras reales de la clase  $i$  (total de la fila).

**Puntuación F1 (F1-score).** El F1 es la media armónica entre la Precisión (UA) y la Exhaustividad (PA). Proporciona una métrica equilibrada que integra simultáneamente los errores de comisión y omisión (Anzalone et al., 2024). Resulta especialmente útil cuando las clases están desbalanceadas o cuando se requiere ponderar por igual ambos tipos de error (Rotich et al., 2025b).

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

**Coeficiente Kappa (K).** El Coeficiente Kappa ( $K$ ) es una métrica que evalúa el grado de concordancia entre un mapa clasificado y los datos de referencia, corrigiendo el acuerdo que podría ocurrir por azar (B. Roy, 2021). A diferencia de la Precisión Global, Kappa incorpora explícitamente la probabilidad de coincidencia aleatoria, lo que lo convierte en un indicador más robusto cuando las clases son desbalanceadas.

**Rango e interpretación.** Kappa expresa el nivel de acuerdo más allá del azar (K. C. Roy et al., 2024). Su valor oscila típicamente entre -1 y +1:

- $K > 0.80$ : Acuerdo casi perfecto o excelente (Tola & Deyassa, 2024).
- $0.61 \leq K \leq 0.80$ : Acuerdo sustancial (Mhangara et al., 2024).
- $0.41 \leq K \leq 0.60$ : Acuerdo moderado (Mhangara et al., 2024).
- $0.21 \leq K \leq 0.40$ : Acuerdo regular o justo (Mhangara et al., 2024).
- $K \leq 0.20$ : Acuerdo leve o pobre (Mhangara et al., 2024).

Un valor de  $K = 1$  indica concordancia perfecta;  $K = 0$  implica que el acuerdo observado es equivalente al esperado por azar; y valores negativos reflejan un acuerdo inferior al azar.

**Cálculo del coeficiente.** El Kappa se basa en la proporción de acuerdo observado ( $P_o$ , equivalente a la OA) y la proporción de acuerdo esperado por azar ( $P_e$ ) (Rotich et al., 2025b).

1. Fórmula conceptual:

$$K = \frac{P_o - P_e}{1 - P_e} \quad (9)$$

donde  $P_o = \sum MC_{ii}/N$  es la exactitud observada y  $P_e$  es la probabilidad de coincidencia aleatoria, calculada como:

$$P_e = \frac{\sum_{i=1}^n (MC_{i+} \cdot MC_{+i})}{N^2} \quad (10)$$

2. Fórmula matricial (Kappa de Cohen): Para el cálculo directo a partir de la matriz de confusión, se emplea la siguiente expresión expandida:

$$K = \frac{N \sum_{i=1}^n MC_{ii} - \sum_{i=1}^n (MC_{i+} \cdot MC_{+i})}{N^2 - \sum_{i=1}^n (MC_{i+} \cdot MC_{+i})} \quad (11)$$

donde:  $N$  es el número total de muestras;  $n$  es el número de clases;  $MC_{ii}$  es el valor diagonal (aciertos);  $MC_{i+}$  es la suma de la fila  $i$  (Total Referencia/Real);  $MC_{+i}$  es la suma de la columna  $i$  (Total Clasificado/Mapa).

**Variantes del coeficiente Kappa.** Dado que el Kappa estándar penaliza tanto los errores de ubicación como los de cantidad, se emplean variantes para descomponer el acuerdo (Pontius, 2000). Para el presente estudio se calculó el Kappa de Histograma ( $K_{histo}$ ), el cual evalúa el acuerdo máximo posible dadas las proporciones marginales (totales por clase) de la matriz.

El  $K_{histo}$  utiliza la concordancia máxima posible ( $P_{m\acute{a}x}$ ) en lugar de la observada ( $P_o$ ):

$$K_{histo} = \frac{P_{m\acute{a}x} - P_e}{1 - P_e} \quad (12)$$

donde  $P_{m\acute{a}x}$  se calcula sumando los mínimos de los totales marginales para cada clase:

$$P_{m\acute{a}x} = \frac{\sum_{i=1}^n \min(MC_{i+}, MC_{+i})}{N} \quad (13)$$

**Área Bajo la Curva ROC (AUC-ROC).** El análisis de la Curva de Característica Operativa del Receptor (ROC) y el Área Bajo la Curva (AUC) constituye una métrica para evaluar la precisión y fiabilidad de los modelos que simulan o predicen cambios en LULC (Badshah et al., 2024; Shrestha et al., 2023).

A diferencia de las métricas anteriores (como OA o Kappa) que evalúan una clasificación final “dura”, el AUC-ROC determina qué tan efectivamente un modelo predice la distribución continua de una variable booleana (probabilidad de transición), cuantificando la capacidad de discriminar entre áreas de cambio y no cambio independientemente del umbral de corte seleccionado (Badshah et al., 2024; Shrestha et al., 2023).

**Interpretación de la Curva y el Área.** La curva ROC expone el rendimiento contrastando la Tasa de Verdaderos Positivos (Sensibilidad) en el eje Y contra la Tasa de Falsos Positivos (1 - Especificidad) en el eje X. El valor del AUC resume este rendimiento en un único escalar:

- El valor oscila teóricamente entre 0 y 1. Un  $AUC = 0.5$  indica un rendimiento aleatorio (sin capacidad predictiva).
- Cuanto más se aproxime el valor a 1, mayor será la capacidad del modelo para separar correctamente las clases de transición.

**Relevancia frente a otras métricas.** El coeficiente Kappa presenta limitaciones para diferenciar entre errores de cuantificación y errores de ubicación (Badshah et al., 2024). El AUC-ROC complementa esta evaluación al validar la idoneidad de las *probabilidades de transición* generadas por modelos como MLP-CA-MC, siendo una herramienta estándar en módulos de validación de software geoespacial avanzado (Badshah et al., 2024).

## 2.9.2 Métricas de Evaluación de Simulación Espacio-Temporal

Las métricas de evaluación estándar, como la Precisión Global (OA) y el Coeficiente Kappa, operan mediante una comparación estricta píxel a píxel (Ovejero-Campos, 2021). Esta aproximación es rigurosa en el análisis geoespacial, ya que penaliza de igual manera todos los errores sin considerar el contexto (Mas et al., 2019). Estos “casi aciertos” (*near miss*) son

comunes en paisajes heterogéneos, zonas de transición (ecotonos) o al trabajar con resoluciones moderadas (Liang et al., 2024; Robalino Trujillo, 2023).

Para obtener una evaluación más realista de la concordancia estructural, se emplean métricas de precisión difusa (*fuzzy accuracy*) que relajan el criterio estricto de coincidencia, introduciendo tolerancia espacial, temática o de escala.

**Precisión Difusa por Vecindad.** Este enfoque aborda la incertidumbre de localización (Hagen, 2003; Visser, 2004). En lugar de una comparación directa píxel a píxel, introduce una tolerancia a errores de localización reconociendo el acuerdo dentro de una vecindad espacial definida, usualmente mediante una “ventana móvil” (Esri, 2024).

Un píxel del mapa simulado se considera un “acierto difuso” si la clase correcta (según los datos de referencia) se encuentra presente *en cualquier lugar* dentro de la ventana de vecindad definida (Mas et al., 2019; Robalino Trujillo, 2023). Esta métrica cuantifica la proporción de aciertos y “casi aciertos”, validando la capacidad del modelo para simular patrones espaciales generales, incluso si la localización exacta a nivel de píxel presenta ligeros desplazamientos (Ovejero-Campos, 2021).

**Análisis de Validación Multiescala.** El Análisis de Validación Multiescala es un enfoque alternativo esencial para evaluar modelos LULC, dada la heterogeneidad del paisaje a diferentes niveles de detalle (Fassnacht et al., 2014; Tiamgne et al., 2025).

El mecanismo consiste en agregar espacialmente los mapas simulado y de referencia a resoluciones progresivamente más gruesas (H. H. Nguyen et al., 2025) y recalcular las métricas de concordancia en cada nueva escala.

Se espera que, si el modelo captura correctamente los patrones espaciales generales, la concordancia aumente a medida que la escala se vuelve más gruesa, ya que los pequeños errores de localización a nivel de píxel se promedian y el acuerdo en la estructura general del paisaje se hace evidente (Marey et al., 2025).

## 2.10 Inteligencia Artificial Explicable (XAI) en la Modelización de Cobertura de Suelo

La Inteligencia Artificial Explicable (XAI) comprende procesos y métodos diseñados para proporcionar claridad y capacidad de interpretación a los sistemas de IA, contrarrestando la naturaleza de “caja negra” típica de modelos complejos como las redes neuronales profundas (Mehdiyev et al., 2025a, 2025b). En el contexto de la gestión ambiental y la cartografía de susceptibilidad, la XAI es necesaria para fomentar la confianza y validar la consistencia ecológica de las predicciones (Pradhan et al., 2023; X. Yang et al., 2025).

### 2.10.1 Análisis de Dominancia de Variables mediante Permutación

Para interpretar el funcionamiento interno de este modelo de “caja negra”, se requiere valorar la aportación específica de cada variable predictora. La Importancia de Características por Permutación (PFI) es una técnica “agnóstica del modelo” que evalúa la influencia de cada factor en la estabilidad predictiva (Medianovsky et al., 2023).

El método opera bajo una premisa de *ruptura de asociación*, se mide la degradación en el rendimiento del modelo cuando la relación estadística entre una variable predictora específica y la variable objetivo es destruida mediante un reordenamiento aleatorio (*shuffling*) de sus valores, manteniendo inalterada su distribución marginal (Altmann et al., 2010). Esto permite identificar interacciones no lineales complejas entre los predictores (Díaz-Uriarte & Alvarez de Andrés, 2006).

Se define como Altmann et al. (2010):

$$VI_j = S_{base} - S_{perm,j} \quad (14)$$

Donde  $VI_j$  es la importancia de la variable  $j$ ,  $S_{base}$  es el AUC-ROC del modelo con los datos originales, y  $S_{perm,j}$  es el AUC-ROC calculado tras permutar aleatoriamente el vector de la variable  $j$ . Una disminución apreciable en el AUC ( $VI_j \gg 0$ ) indica que el modelo depende de dicha variable para discriminar las transiciones de uso de suelo.

### 2.10.2 Gráficos de Dependencia Parcial

Las Curvas de Respuesta, materializadas en los Gráficos de Dependencia Parcial (PDP), explican cómo dichas variables influyen en la predicción (X. Yang et al., 2025). Un PDP visualiza la dependencia marginal entre una característica de interés y la variable objetivo (susceptibilidad de cobertura de suelo), promediando el efecto de todas las demás variables del modelo (Medianovskyi et al., 2023; Z. Yang et al., 2024). Estos gráficos son esenciales para investigar la relación entre un factor ambiental y la cobertura de suelo, identificando umbrales y comportamientos no lineales que los análisis de importancia global por sí solos no pueden detectar (Mondal et al., 2023; Nam et al., 2025).

Formalmente, sea  $S$  el subconjunto de características de interés y  $C$  el subconjunto complemento que contiene todas las demás variables predictoras. La función de dependencia parcial  $\hat{f}_S$  se define teóricamente como la esperanza matemática de la salida del modelo  $\hat{f}$  sobre la distribución marginal de las características en  $C$ :

$$\hat{f}_S(x_S) = E_{X_C} [\hat{f}(x_S, X_C)] = \int \hat{f}(x_S, X_C) d\mathbb{P}(X_C) \quad (15)$$

En la práctica, dado que la distribución real de los datos es desconocida, esta función se estima empíricamente utilizando el promedio de las predicciones sobre el conjunto de datos de entrenamiento. Para un valor específico de la variable de interés  $x_S$ , la estimación parcial se calcula promediando las salidas del modelo al forzar  $x_S$  en todas las observaciones, manteniendo inalterados los valores de las otras variables ( $x_C^{(i)}$ ):

$$\hat{f}_S(x_S) \approx \frac{1}{n} \sum_{i=1}^n \hat{f}(x_S, x_C^{(i)}) \quad (16)$$

donde  $n$  es el número total de muestras en el conjunto de validación y  $x_C^{(i)}$  representa los valores reales de las variables restantes para la muestra  $i$ . De esta manera, la ecuación marginaliza la influencia de  $X_C$ , aislando el efecto exclusivo de  $x_S$  sobre la probabilidad de transición simulada.

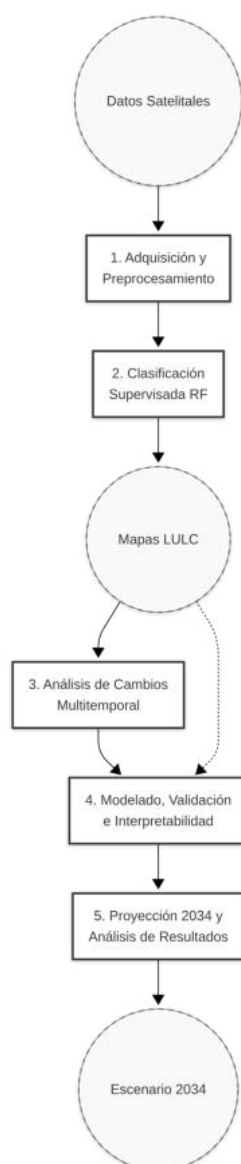
## CAPÍTULO III

### Metodología de la investigación

La metodología del presente estudio se ha estructurado en cuatro etapas secuenciales que abarcan desde el procesamiento de datos satelitales hasta la proyección del escenario tendencial para el año 2034. A continuación, se describe cada etapa (Figura 16).

**Figura 16**

*Flujo Metodológico General de la Investigación.*



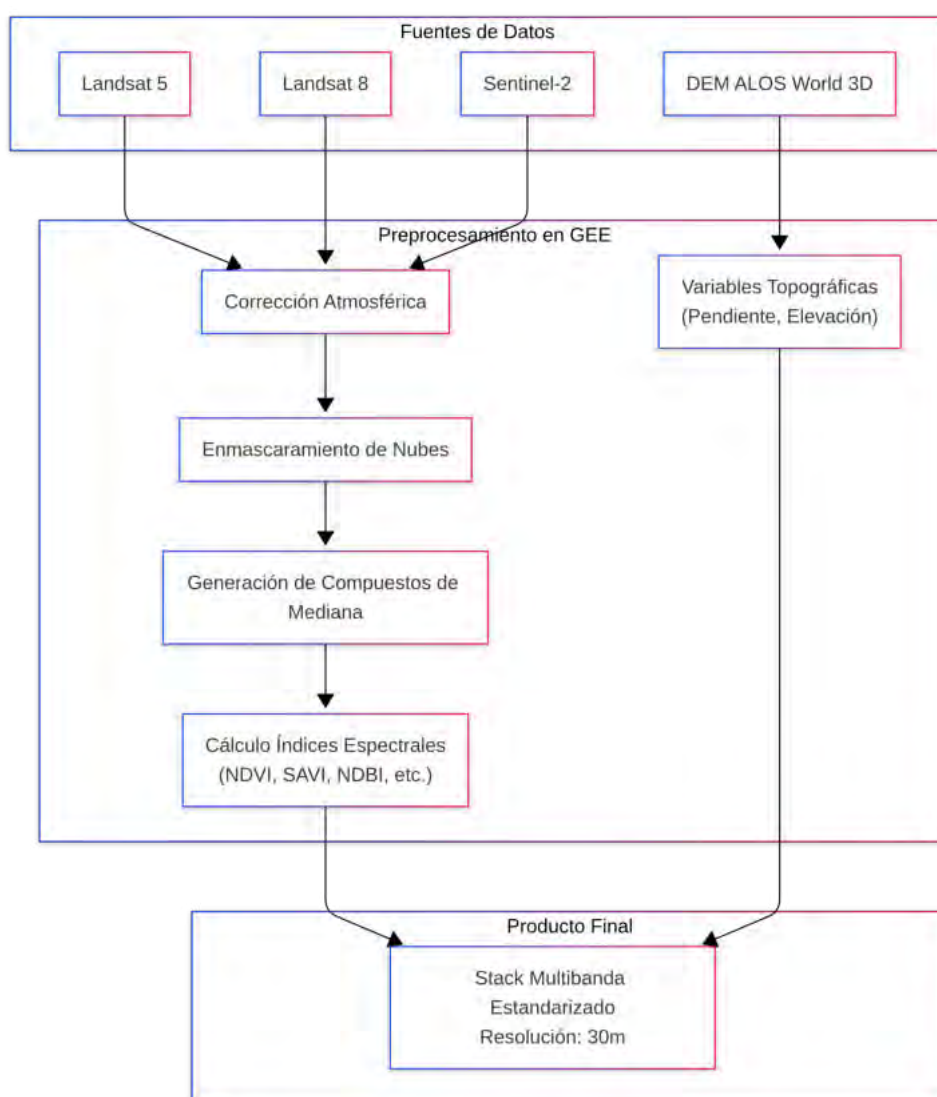
*Nota. El diagrama resume la secuencia operativa desde la adquisición de datos hasta el análisis del escenario futuro. Fuente: Elaboración propia.*

### 3.1 Adquisición y Preprocesamiento de Datos Geospaciales

Se centra en la selección de sensores ópticos (Landsat 5, 8 y Sentinel-2) y la delimitación temporal en la estación de invierno austral para minimizar la nubosidad. El preprocesamiento incluye la corrección atmosférica, enmascaramiento de nubes, generación de compuestos de mediana, incorporación de índices espectrales y variables topográficas, así como la estandarización a una resolución común de 30 metros (Figura 17).

**Figura 17**

*Flujo de Adquisición y Preprocesamiento de Datos.*



*Nota. Fuente: Elaboración propia.*

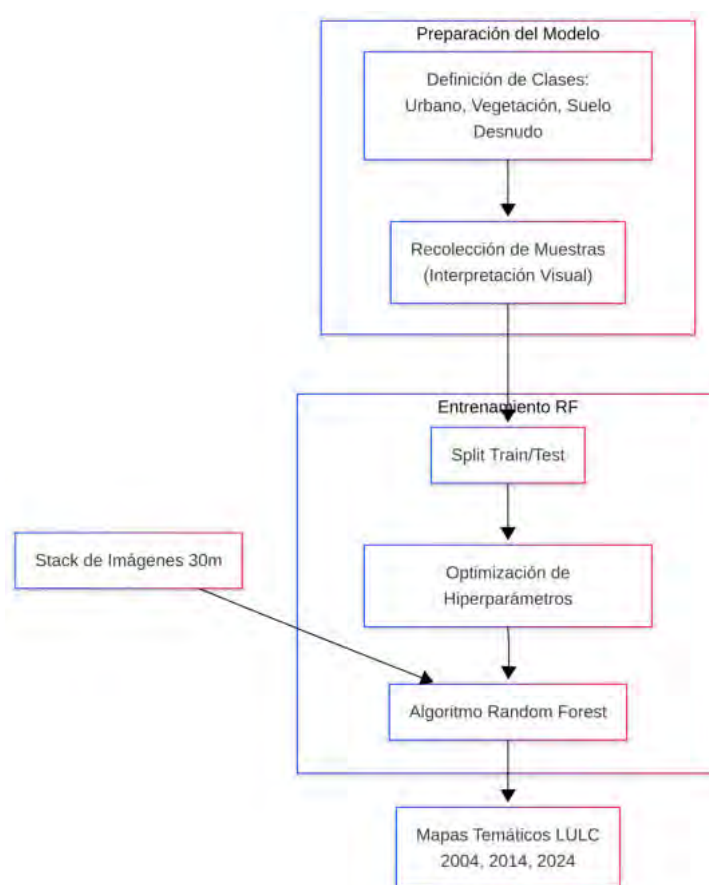


### 3.2 Clasificación Supervisada de Coberturas (LULC)

Se implementa el algoritmo de aprendizaje automático *Random Forest* (RF) para generar los mapas temáticos de los años 2004, 2014 y 2024. Esta fase incluye la definición de clases (Urbano, Vegetación, Suelo Desnudo), la recolección de muestras de entrenamiento mediante interpretación visual y la optimización de hiperparámetros para maximizar la precisión global del clasificador (Figura 18).

**Figura 18**

*Flujo Metodológico: Fase de Clasificación Supervisada.*



*Nota. Fuente: Elaboración propia.*

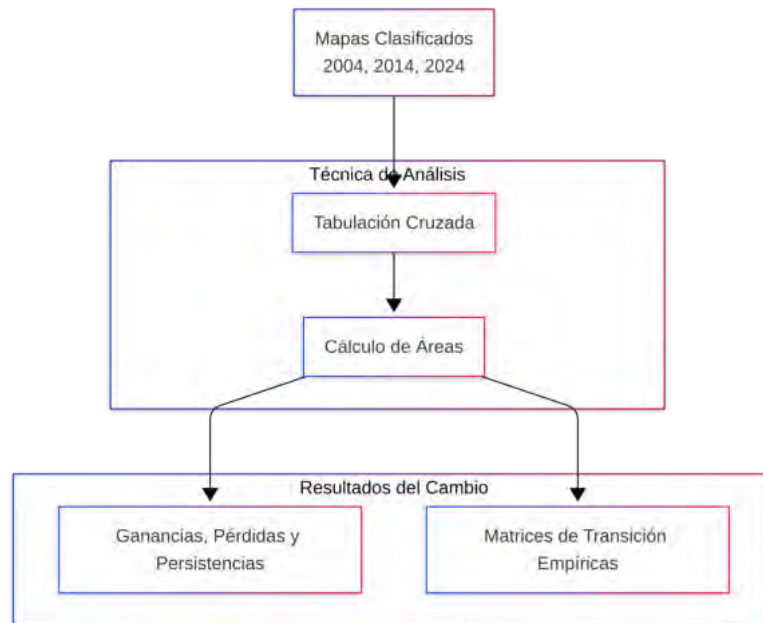
### 3.3 Análisis de Cambios Multitemporal

Una vez generados los mapas clasificados, se aplica una técnica de tabulación cruzada para cuantificar las ganancias, pérdidas y persistencias de cada categoría. Este análisis no solo describe la magnitud del cambio, sino que genera las matrices de transición empíricas que

alimentarán el componente temporal del modelo predictivo (Figura 19).

**Figura 19**

*Proceso de Análisis de Cambios Multitemporal.*

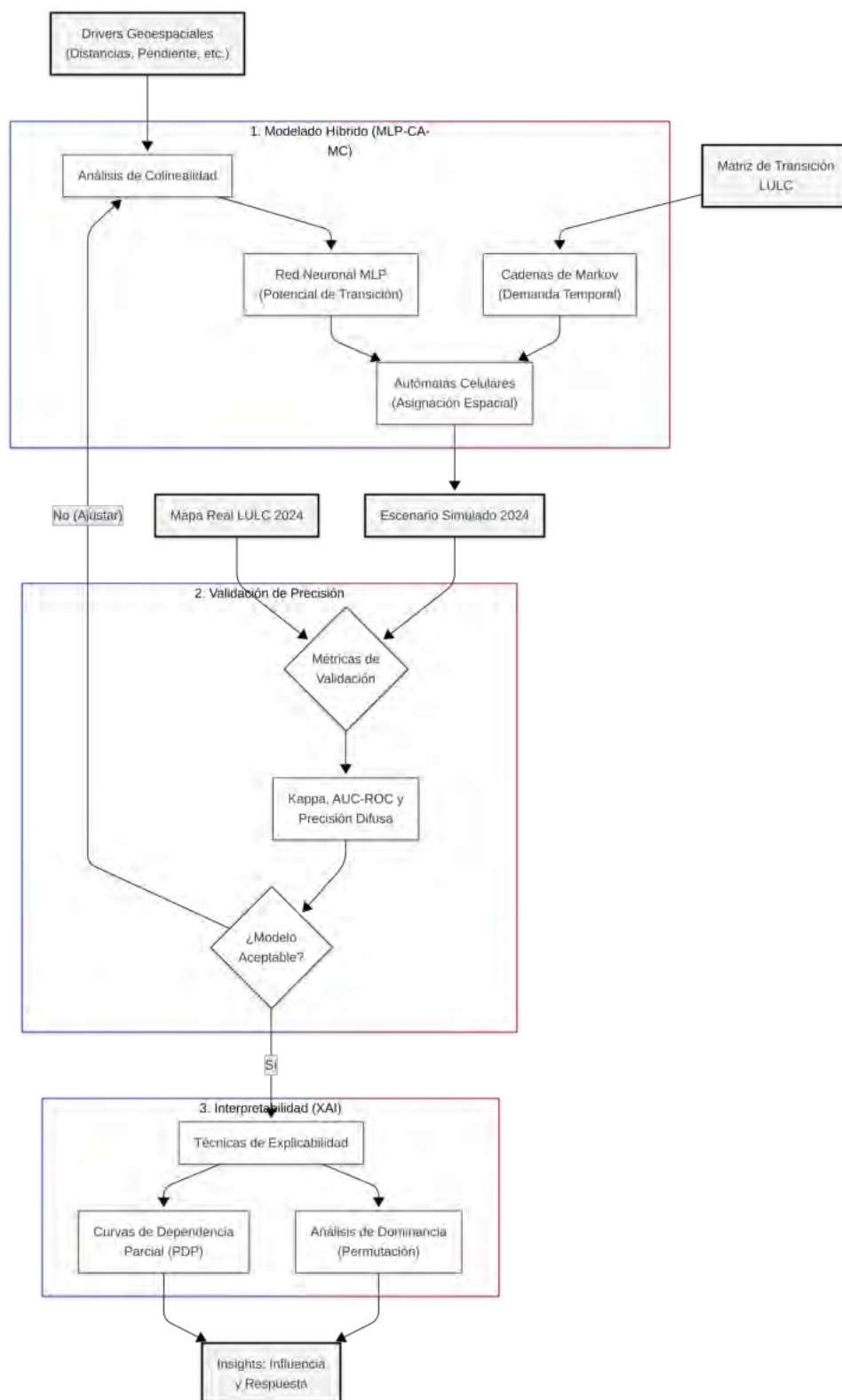


*Nota. Fuente: Elaboración propia.*

### 3.4 Simulación Híbrida (MLP-CA-MC)

Se integra una red neuronal de Perceptrón Multicapa (MLP) para aprender las reglas de transición espacial basadas en variables conductoras (*drivers*), acoplada con Cadenas de Markov (MC) para la demanda temporal y Autómatas Celulares (CA) para la asignación espacial. El proceso incluye el análisis de colinealidad de predictores, la búsqueda y optimización de hiperparámetros (épocas, muestras, tasa de aprendizaje, momento y arquitectura de capas) para calibrar la red, y el entrenamiento final del modelo. Se evalúa la fiabilidad del simulador comparando el escenario proyectado al 2024 con el mapa real. Se emplean métricas categóricas estándar, métricas de potencial (AUC-ROC), métricas espaciales (Precisión Difusa y Validación Multiescala) y validación espacial de la clase vegetación. Posteriormente, se aplican técnicas de Inteligencia Artificial Explicable mediante análisis de dominancia por permutación y curvas de dependencia parcial (PDP), se identifica qué variables influyen más en la urbanización y cómo responde el modelo ante cambios en factores como la pendiente o la distancia a servicios (Figura 20).

**Figura 20**  
**Simulación Híbrida (MLP-CA-MC).**



*Nota. Fuente: Elaboración propia.*

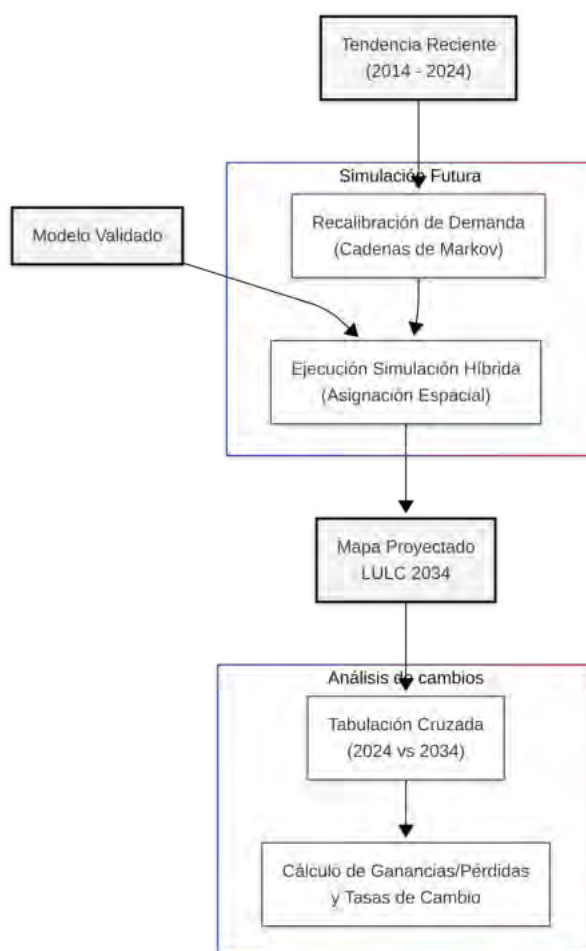
### 3.5 Proyección del Escenario Futuro 2034

Esta etapa final inicia con la recalibración del componente temporal del modelo. Tras validar la simulación histórica, se actualiza la matriz de probabilidad de transición utilizando la tendencia más reciente observada (periodo 2014-2024) y se ejecuta la simulación híbrida (MLP-CA-MC) hacia el año 2034. El resultado es un mapa ráster que espacializa la distribución probable de las coberturas bajo un escenario tendencial.

Posteriormente, se procede al análisis de cambios post-simulación. Utilizando el mapa real de 2024 y el escenario proyectado de 2034, se realiza una nueva tabulación cruzada para cuantificar las ganancias y pérdidas netas esperadas (Figura 21).

**Figura 21**

*Flujo de Proyección y Análisis del Escenario 2034.*



*Nota. Fuente: Elaboración propia.*

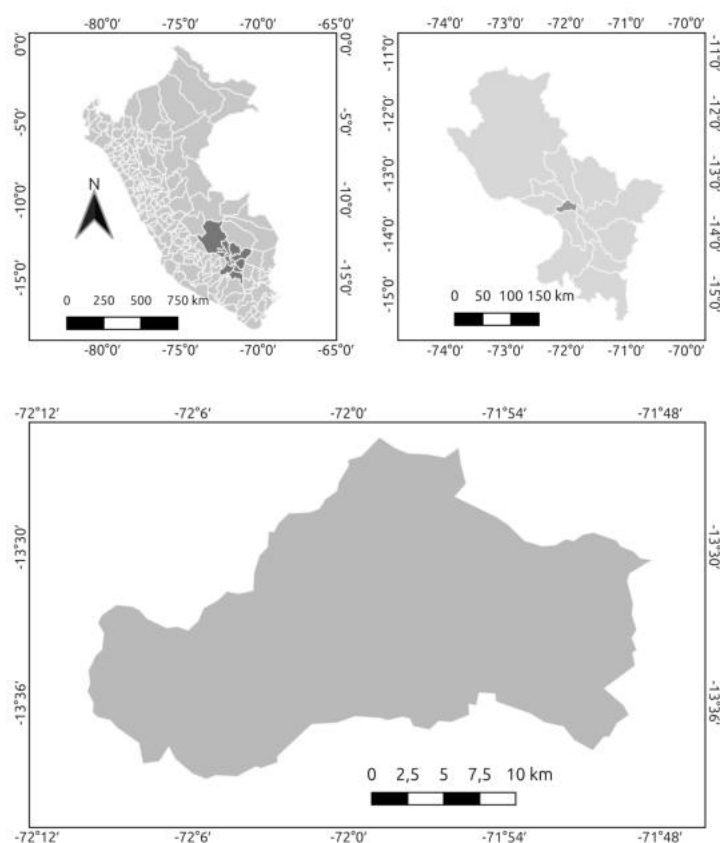
## CAPÍTULO IV

### Desarrollo del proyecto

#### 4.1 Área de Estudio

La provincia del Cusco, una de las trece que conforman el departamento homónimo (Gaona Obando, 2019a), se ubica en la zona central de la región sur andina del Perú (Tito Oca & Mescoco Pumasupa, 2020). Geográficamente, se localiza aproximadamente en las coordenadas  $13^{\circ} 30' 45''$  de latitud Sur y  $71^{\circ} 58' 33''$  de longitud Oeste, emplazada principalmente en el valle del río Huatanay. La altitud dentro de la provincia varía considerablemente, desde los 3,128 m s. n. m. hasta los 4,641 m s. n. m. (Figura 22).

**Figura 22**  
*Ubicación Geográfica del Área de Estudio.*



*Nota. Fuente: Elaboración propia.*

La provincia abarca una superficie oficial de 529.21 km<sup>2</sup>, según la validación realizada en el Apéndice B. Sus límites territoriales colindan al norte con las provincias de Calca y

Urubamba, al este con la provincia de Quispicanchi, al sur con la provincia de Paruro, y al oeste con la provincia de Anta (M. Huaman Conza & Huaman Gaspar, 2019). Administrativamente, la provincia se subdivide en ocho distritos: Cusco (Cercado), Ccorca, Poroy, San Jerónimo, San Sebastián, Santiago, Saylla y Wanchaq (Lulichac Cayhuari & Miranda Mango, 2017).

## **4.2 Datos Geoespaciales**

Para la elaboración de los mapas de Uso y Cobertura del Suelo (LULC) de la provincia de Cusco, se utilizaron colecciones de imágenes satelitales ópticas multiespectrales, procesadas a nivel de Reflectancia Superficial (SR - Surface Reflectance para Landsat) o equivalente (BOA - Bottom of Atmosphere para Sentinel-2), asegurando así mediciones consistentes de la energía reflejada por la superficie terrestre tras la corrección de efectos atmosféricos. Se accedió a estos conjuntos de datos a través de la plataforma de computación en la nube Google Earth Engine (GEE), con la cual se realizó el filtrado, preprocesamiento y análisis de las series temporales de imágenes según el área de interés y los criterios temporales definidos.

A continuación, se exponen los criterios y fundamentos para la selección del período de estudio, la ventana estacional de análisis y los sensores remotos específicos empleados.

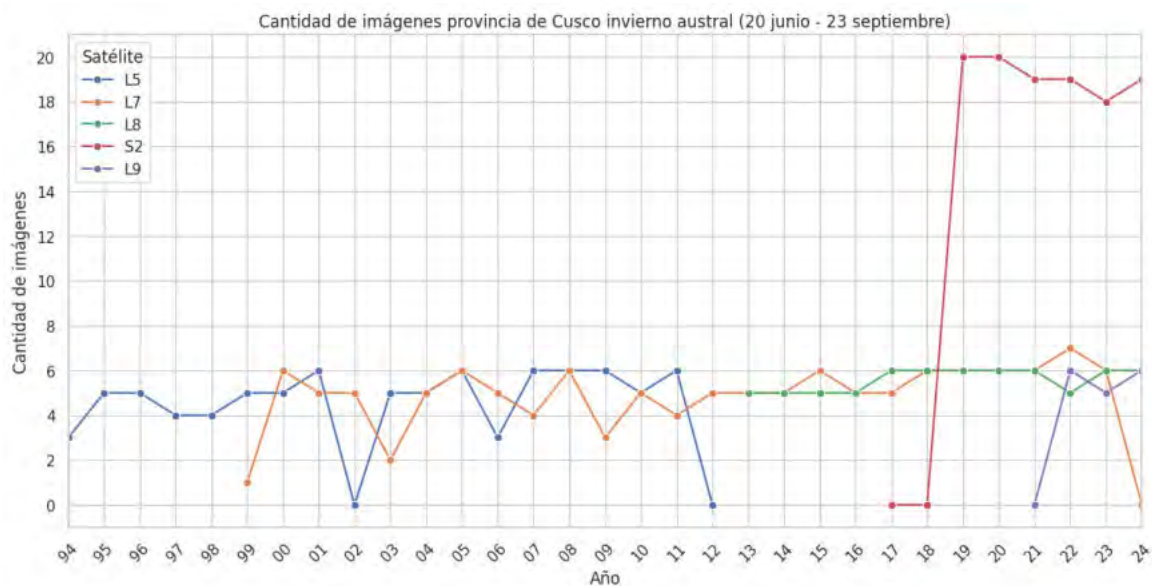
### **4.2.1 Delimitación Temporal**

El análisis multitemporal se estructuró en intervalos decenales, seleccionando los años 2004, 2014 y 2024 como puntos clave para la clasificación LULC. Esta periodicidad de diez años es una práctica recurrente en estudios de dinámica LULC, ya que permite observar transformaciones acumuladas en el paisaje y calibrar modelos predictivos (M. Khan et al., 2023). La elección específica de estos años se fundamentó en la disponibilidad y calidad de las imágenes satelitales dentro de la plataforma GEE para la ventana estacional seleccionada, buscando garantizar una cantidad suficiente de imágenes para generar compuestos completos, en contraste con otras combinaciones de años que presentaban menor cobertura de imágenes útiles (Figura 23).

Con el fin de asegurar la comparabilidad interanual de los datos y minimizar los efectos de la variabilidad fenológica y atmosférica estacional, el estudio se restringió temporalmente

**Figura 23**

*Disponibilidad de Imágenes Satelitales Invierno Austral por Sensor.*



*Nota. Fuente: Elaboración propia.*

a la estación de Invierno Austral (aproximadamente del 20 de junio al 23 de septiembre). Esta decisión metodológica se justifica por las siguientes razones:

1. **Condiciones Atmosféricas Favorables:** El invierno en la región andina del Cusco coincide con la temporada de estiaje o seca, caracterizada por una mínima precipitación pluviométrica (Carbajal Coronado, 2024) y baja humedad relativa atmosférica (Castillo Alire & Gutierrez Kancha, 2019). Estas condiciones meteorológicas reducen la probabilidad de cobertura nubosa, siendo la más baja del año entre mayo y septiembre (Aiquipa Alosilla & Soncco Mamani, 2019), y minimizan la interferencia atmosférica (dispersión y absorción por vapor de agua). Esto maximiza la obtención de imágenes satelitales con cielos despejados.
2. **Homogeneidad Ambiental Interanual:** Al centrar el análisis en la misma ventana estacional cada año, se busca asegurar que las condiciones ambientales generales (temperatura, régimen hídrico, estado fenológico predominante) sean lo más consistentes posible entre los diferentes años de observación (Ayma Quispe, 2022), facilitando la atribución de las diferencias observadas a cambios reales en LULC en lugar de a variaciones estacionales.

3. Contexto del Ciclo Agrícola Regional: El período invernal representa una fase de baja actividad agrícola para los cultivos predominantes de secano en la región. Corresponde típicamente al período posterior a la cosecha principal (post-cosecha) de cultivos andinos como la papa y la quinua (Salamanca Oviedo Cabrera, 2016) y anterior a la preparación del terreno y siembra (pre-siembra) que usualmente comienza con las primeras lluvias de primavera (septiembre-octubre) (Arce Quispe & Chavez Cahuana, 2022). Durante el invierno, muchas tierras agrícolas se encuentran en descanso o barbecho (Quispe Quispe, 2021), lo que facilita la discriminación espectral entre áreas agrícolas inactivas (clasificadas como suelo desnudo o con vegetación residual seca) y áreas con vegetación natural persistente o plantaciones forestales activas.

#### **4.2.2 Sensores Satelitales**

La elección de los sensores satelitales específicos para cada año del estudio se basó en la disponibilidad histórica de datos corregidos atmosféricamente (Nivel-2) dentro de GEE, las capacidades técnicas de cada instrumento y la necesidad de mantener la mayor consistencia posible a lo largo del tiempo:

- Para el año 2004: Se utilizaron imágenes del sensor Thematic Mapper (TM) a bordo del satélite Landsat 5. En el momento del análisis, esta era la fuente principal de datos de reflectancia superficial con cobertura adecuada y consistente en GEE para ese año, siendo indispensable para establecer la línea base histórica del estudio.
- Para el año 2014: Se emplearon datos del sensor Operational Land Imager (OLI) a bordo del satélite Landsat 8. Se prefirió Landsat 8 sobre su predecesor contemporáneo, Landsat 7 (sensor ETM+), debido a la falla permanente del Scan Line Corrector (SLC) de Landsat 7 desde mayo de 2003, que introduce franjas sin datos (gaps) en sus imágenes (M'Barek, 2024). Landsat 8, lanzado en 2013, asegura una cobertura completa y una calidad radiométrica y espectral mejorada para el período de análisis de 2014.
- Para el año 2024: Se optó por utilizar datos del sensor Multispectral Instrument (MSI) a bordo de la constelación Sentinel-2 (Sentinel-2A y Sentinel-2B). La principal ventaja de



Sentinel-2 para el año más reciente es su alta frecuencia de revisita (aproximadamente 5 días con ambos satélites operativos (Gascon et al., 2015)), mayor que la de Landsat 8/9 (16 días (Ganjirad, 2024)). Esta mayor frecuencia temporal aumenta considerablemente la probabilidad de adquirir imágenes libres de nubes dentro de la ventana estacional definida.

### 4.2.3 *Preprocesamiento de Imágenes Satelitales*

Se aplicó un flujo de preprocesamiento estandarizado a las colecciones de imágenes seleccionadas para cada año dentro de la plataforma GEE, con el objetivo de obtener imágenes compuestas, libres de nubes y espacialmente consistentes.

**Filtrado Inicial de Colecciones.** Para cada año (2004, 2014, 2024) y sensor correspondiente (Landsat 5 TM, Landsat 8 OLI, Sentinel-2 MSI), se filtró inicialmente el catálogo de GEE para seleccionar únicamente las imágenes de Nivel-2 (Reflectancia Superficial) que intersectaban el área de estudio (provincia de Cusco) y cuya fecha de adquisición estuviera comprendida dentro de la ventana temporal definida (20 de junio al 23 de septiembre). El número de imágenes resultantes para cada colección se detalla en la Tabla 5.

**Tabla 5**

*Resumen de Imágenes Satelitales Seleccionadas por Año.*

Sensor/Satélite	Año	Periodo de Adquisición <sup>a</sup>	Nº de Imágenes
Landsat 5 / TM	2004	20 de junio - 23 de septiembre	5
Landsat 8 / OLI	2014	20 de junio - 23 de septiembre	5
Sentinel-2 / MSI	2024	20 de junio - 23 de septiembre	19

*Nota.*<sup>a</sup> El período corresponde al intervalo que abarca las fechas variables del invierno astronómico en el hemisferio sur (20-21 de junio a 22-23 de septiembre) (Gobierno del Perú, 2025).

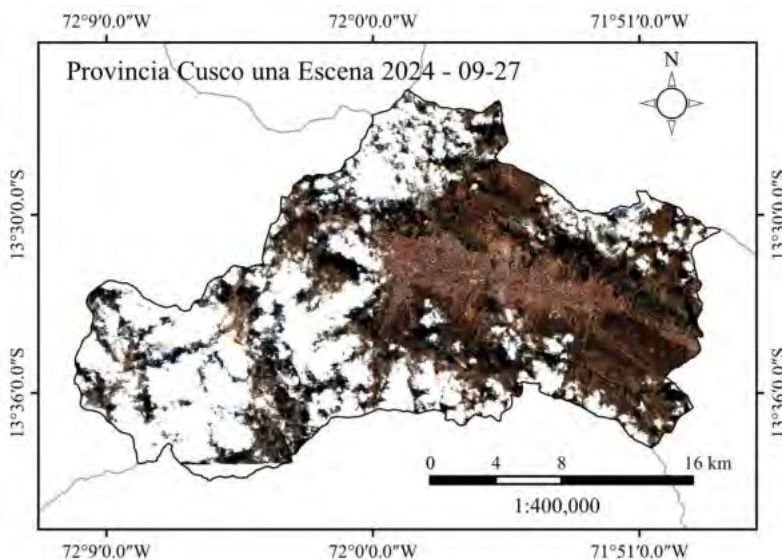
**Composición de Mosaico para Imágenes Sentinel-2.** A diferencia de las escenas individuales de Landsat que cubren completamente la provincia de Cusco, las escenas de Sentinel-2 requieren combinar múltiples escenas adyacentes para lograr una cobertura espacial completa del área de estudio en una fecha dada. Por lo tanto, se implementó un paso adicional de preprocesamiento para la colección Sentinel-2:

1. Identificación de Escenas: Para cada fecha de adquisición dentro de la colección filtrada, se identificaron las escenas específicas de Sentinel-2 necesarias para cubrir la totalidad de la provincia.
2. Generación de Mosaico por Fecha: Se utilizó la función `mosaic()` de GEE para combinar los tiles correspondientes a cada fecha en una única imagen compuesta (mosaico).
3. Recorte al Área de Estudio: Cada mosaico diario fue recortado espacialmente (`clip()`) utilizando el polígono vectorial que define los límites de la provincia de Cusco, asegurando que todos los análisis posteriores se realizaran estrictamente dentro del área de interés. Las Figuras 24 y 25 ilustran tiles individuales antes del mosaico.

Este procedimiento se aplicó a las 19 fechas disponibles en la colección de 2024, generando una nueva colección de imágenes diarias mosaico/recortadas.

#### **Figura 24**

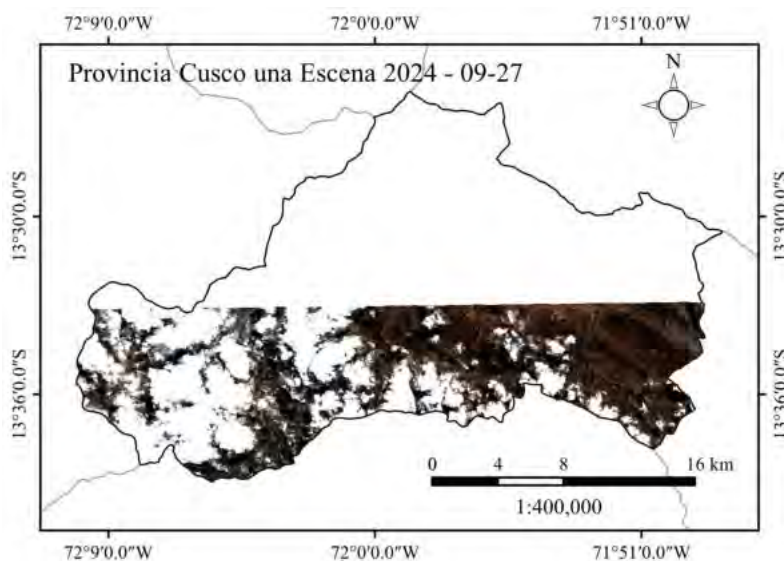
*Ejemplo de una escena “norte” de Sentinel-2.*



*Nota. Elaboración propia a partir de datos Sentinel-2 obtenidos de Google Earth Engine.*

**Figura 25**

*Ejemplo de una escena “sur” adyacente de Sentinel-2.*

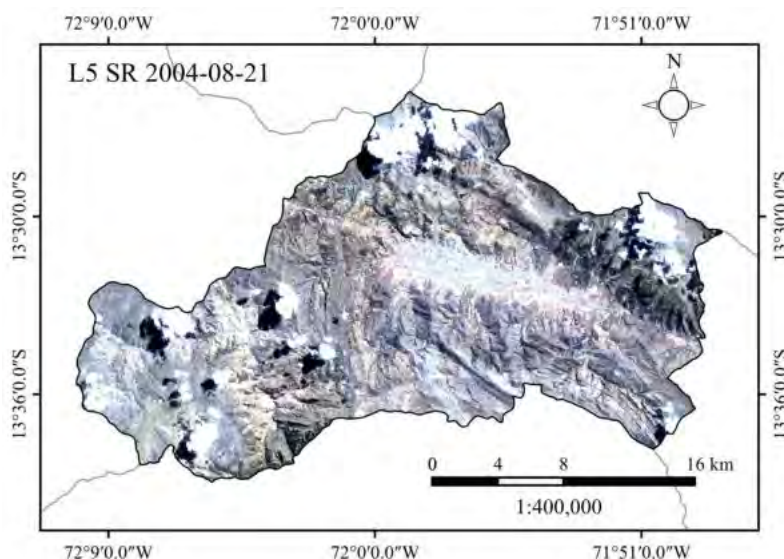


*Nota. Elaboración propia a partir de datos Sentinel-2 obtenidos de Google Earth Engine.*

A efectos de concisión, se presenta a continuación una imagen representativa para cada año de análisis: Landsat-5 para 2004 (Figura 26), Landsat-8 para 2014 (Figura 27) y Sentinel-2 para 2024 (Figura 28). El catálogo completo de imágenes satelitales procesadas se encuentra disponible en el repositorio digital descrito en el Apéndice E.

**Figura 26**

*Imagen Landsat-5 del 21 de agosto de 2004.*

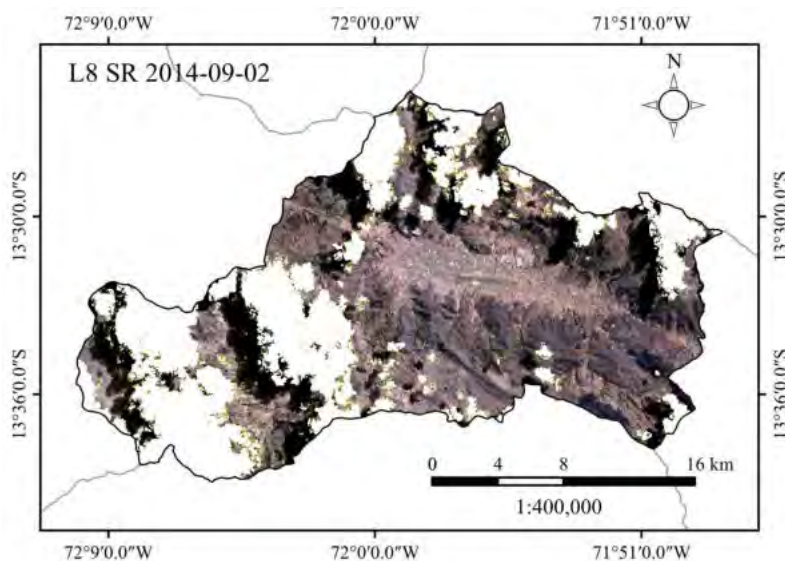


*Nota. Fuente: Adaptado a partir de datos de Google Earth.*

Si bien es habitual seleccionar únicamente imágenes con un porcentaje reducido de

**Figura 27**

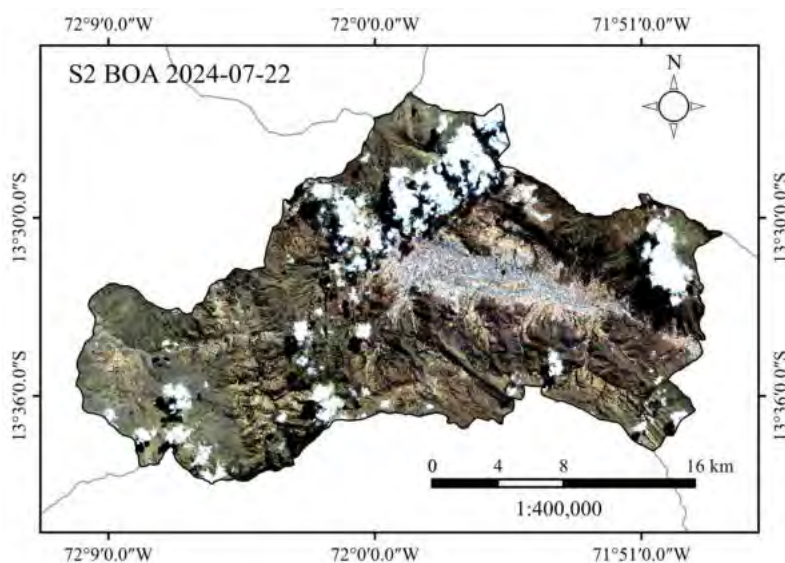
*Imagen Landsat-8 del 02 de septiembre de 2014.*



*Nota. Fuente: Adaptado a partir de datos de Google Earth.*

**Figura 28**

*Imagen Sentinel-2 del 22 de julio de 2024.*



*Nota. Fuente: Adaptado a partir de datos de Google Earth.*

nubosidad, en este caso no se aplicó dicho criterio debido a la limitada disponibilidad de datos. Por ello, se decidió trabajar con la totalidad del conjunto de imágenes, quedando el tratamiento y corrección de nubosidad como una etapa posterior del análisis.

**Enmascaramiento de Nubes y Sombras.** Para eliminar píxeles contaminados por nubes, sombras de nubes u otros efectos atmosféricos o radiométricos, se aplicaron algoritmos de

enmascaramiento específicos para cada sensor, utilizando las bandas de calidad proporcionadas con los datos de Nivel-2.

***Landsat 5 TM y Landsat 8 OLI:*** Se aplicó un procedimiento de enmascaramiento idéntico y consistente a ambas colecciones (2004 y 2014) basado en las bandas de calidad Pixel Quality Assessment (QA\_PIXEL) y Radiometric Saturation Quality Assessment (QA\_RADSAT) de la Colección 2 Nivel-2 de USGS:

1. Enmascaramiento de Nubes y Sombras: Se utilizó la banda QA\_PIXEL y máscaras de bits específicas para identificar y marcar como inválidos los píxeles clasificados con alta confianza como Nube (Cloud) o Sombra de Nube (Cloud Shadow). El objetivo fue lograr un equilibrio entre la eliminación efectiva de contaminantes y la preservación de la mayor cantidad posible de píxeles válidos.
2. Enmascaramiento de Saturación Radiométrica: Se utilizó la banda QA\_RADSAT para identificar y enmascarar píxeles donde uno o más de los sensores ópticos alcanzaron su límite de detección (saturación), evitando así el uso de valores de reflectancia potencialmente incorrectos.

***Sentinel-2 MSI:*** Para la colección Sentinel-2 de 2024, se empleó un enfoque basado en la banda de probabilidad de nubes derivada del algoritmo S2 Cloud Score+:

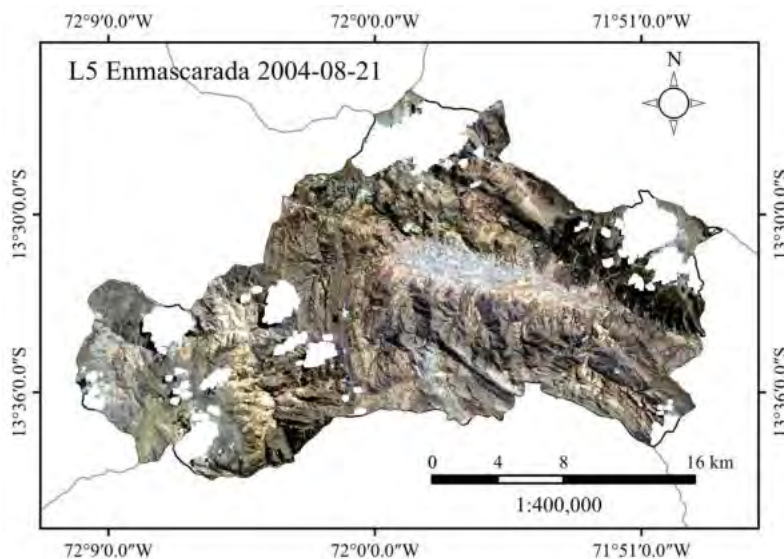
1. Vinculación con Cloud Score+: Cada imagen de la colección fue vinculada con su correspondiente capa de probabilidad de nube (cs) de la colección GEE GOOGLE/CLOUD\_SCORE\_PLUS/V1/S2\_SR.
2. Enmascaramiento Probabilístico: Se aplicó un umbral de probabilidad, conservando únicamente aquellos píxeles con una probabilidad estimada de estar libres de nubes superior a un valor determinado (50 %). Este enfoque busca ser comparable en rigurosidad al método basado en bits de Landsat.

Se muestran a continuación ejemplos representativos del resultado del enmascaramiento de nubes y sombras para cada sensor: Landsat-5 en 2004 (Figura 29), Landsat-8 en 2014 (Figura 30) y Sentinel-2 en 2024 (Figura 31). La colección completa de imágenes procesadas y enmascaradas se encuentra disponible para su consulta en el repositorio digital (Apéndice E).



**Figura 29**

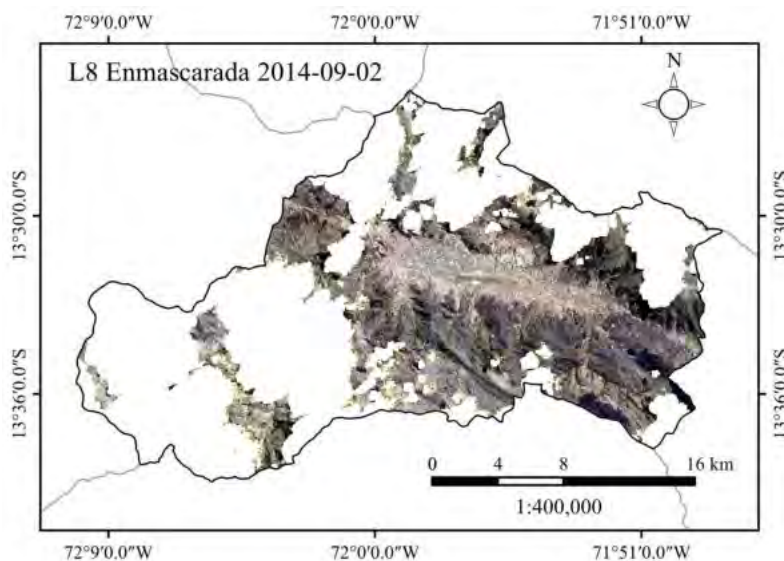
*Imagen Landsat-5 enmascarada del 21 de agosto de 2004.*



*Nota. Fuente: Adaptado a partir de datos de Google Earth Engine.*

**Figura 30**

*Imagen Landsat-8 enmascarada del 02 de septiembre de 2014.*



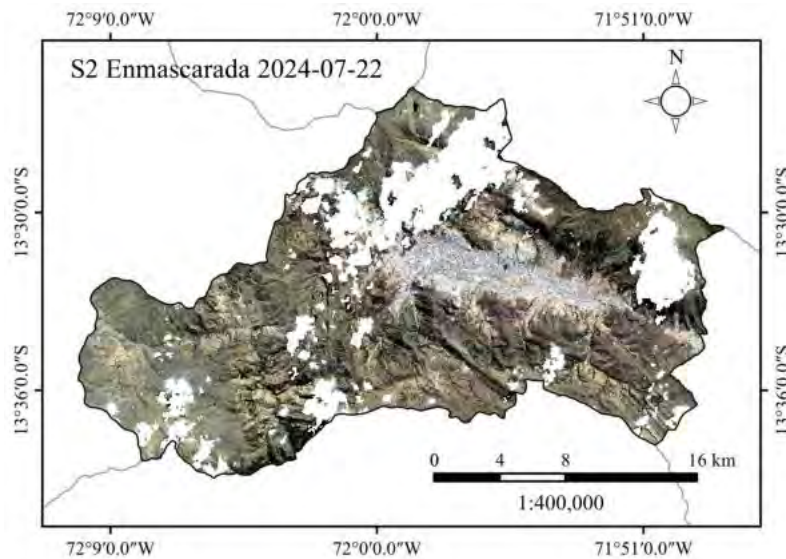
*Nota. Fuente: Adaptado a partir de datos de Google Earth Engine.*

**Generación de Compuestos de Mediana.** Tras enmascarar las colecciones de imágenes para cada año (2004, 2014, 2024), el siguiente paso fue reducir cada colección multitemporal a una única imagen compuesta, representativa de las condiciones promedio del invierno austral.

Se utilizó para ello la función de agregación temporal por mediana (`.median()`) disponible en GEE. Esta técnica calcula, para cada banda espectral y cada ubicación de píxel, el valor de la mediana de todos los píxeles válidos (no enmascarados) correspondientes a esa

**Figura 31**

*Imagen Sentinel-2 enmascarada del 22 de julio de 2024.*



*Nota. Fuente: Adaptado a partir de datos de Google Earth Engine.*

ubicación en la colección de imágenes del año. La mediana resulta preferible al promedio, ya que se considera una medida más eficiente en presencia de datos espaciales ruidosos (Dahal et al., 2024).

**Relleno de Vacíos en el Compuesto de 2004.** Una inspección cuantitativa de los píxeles del compuesto de mediana generado para el año 2004 (Landsat 5) reveló la presencia de un único píxel sin datos (vacío o *gap*) dentro del área de estudio. Aunque pequeño, este vacío podría afectar la consistencia espacial de análisis posteriores o la aplicación de algoritmos de clasificación.

Para corregir esta discontinuidad, se implementó un procedimiento de relleno de vacíos (*gap filling*) basado en interpolación por vecindad:

1. Localización del Píxel Vacío: Se identificaron las coordenadas exactas del píxel sin datos.
2. Cálculo de Valor Interpolado: Para cada una de las bandas espectrales del compuesto, se calculó el valor promedio de los píxeles vecinos válidos.
3. Aplicación del Relleno: Se creó una imagen temporal de un solo píxel con los valores promedio calculados. Se utilizó la función `unmask()` de GEE para fusionar esta imagen

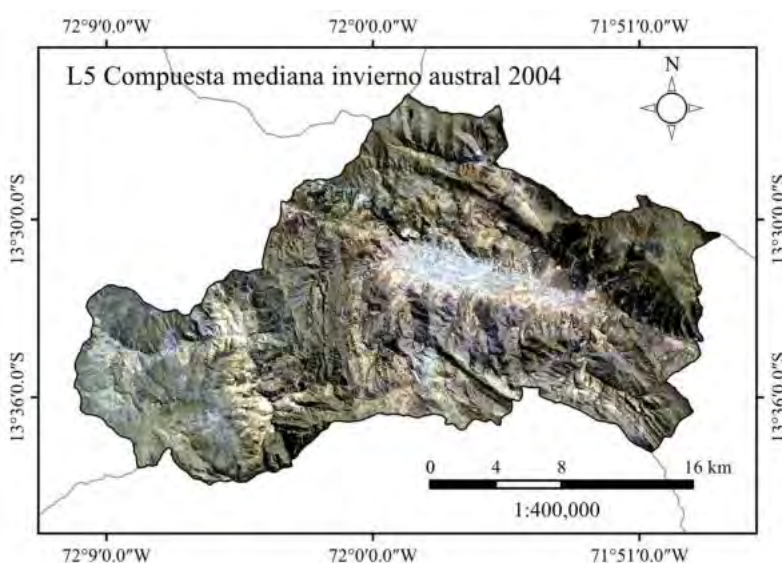
de relleno con el compuesto original. La función mencionada reemplaza únicamente los píxeles enmascarados (vacíos) con los valores correspondientes de la imagen de relleno, dejando intactos todos los píxeles que ya tenían datos válidos.

Este procedimiento resultó en una imagen compuesta final para el año 2004 que era espacialmente completa dentro del área de estudio. Con esto, se obtuvieron tres imágenes compuestas finales, espacialmente completas y listas para la adición de variables auxiliares y la clasificación.

Las Figuras 32, 33 y 34 muestran la visualización en color real (RGB) de estas tres imágenes compuestas finales. Las imágenes utilizadas en el análisis contienen el conjunto completo de bandas espectrales seleccionadas: para Landsat 5, [SR\_B1, SR\_B2, SR\_B3, SR\_B4, SR\_B5, SR\_B7]; para Landsat 8, [SR\_B1, SR\_B2, SR\_B3, SR\_B4, SR\_B5, SR\_B6, SR\_B7]; y para Sentinel-2, [B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12].

### Figura 32

*Compuesto de mediana (Landsat 5, invierno austral 2004).*



*Nota. Elaboración propia a partir de datos Sentinel-2 obtenidos de Google Earth Engine.*

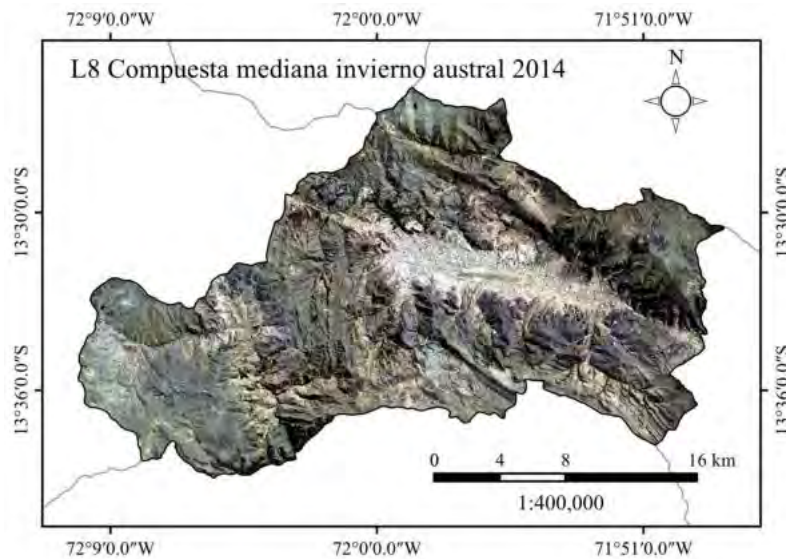
#### 4.2.4 Variables Auxiliares

Para enriquecer el conjunto de datos espectrales base y mejorar la capacidad discriminativa de los algoritmos de aprendizaje automático entre las diferentes clases LULC,



**Figura 33**

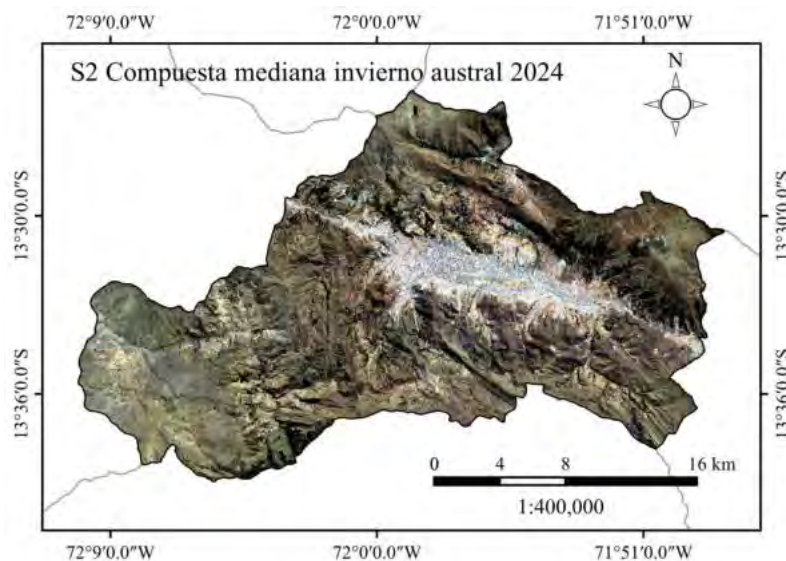
*Compuesto de mediana (Landsat 8, invierno austral 2014).*



*Nota. Elaboración propia a partir de datos Sentinel-2 obtenidos de Google Earth Engine.*

**Figura 34**

*Compuesto de mediana (Sentinel-2, invierno austral 2024).*



*Nota. Elaboración propia a partir de datos Sentinel-2 obtenidos de Google Earth Engine.*

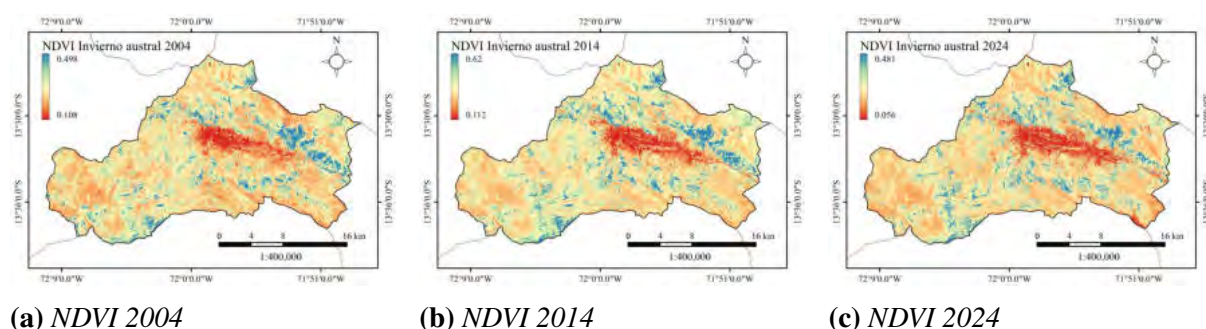
se generaron y añadieron variables auxiliares a cada una de las imágenes compuestas. Estas variables se categorizan en índices espectrales y variables topográficas.

**Índices Espectrales.** Para la caracterización de las coberturas, se seleccionó un conjunto de seis índices espectrales. Esta selección se fundamenta en su recurrencia y efectividad documentada en la literatura científica reciente para discriminar las clases de interés (vegetación, suelo

desnudo y área construida) (Vahid & Aly, 2025).

Específicamente, se priorizaron índices que pudieran ser derivados consistentemente a partir de las bandas ópticas y del infrarrojo de onda corta (SWIR) disponibles en los tres sensores utilizados (Landsat 5, Landsat 8 y Sentinel-2). Dado que el sensor MSI de Sentinel-2 carece de bandas térmicas, se descartaron índices dependientes de la temperatura superficial directa, optándose por indicadores como el NDBI y el UI que han demostrado una fuerte correlación con la dinámica de urbanización y las variaciones térmicas asociadas (Vahid & Aly, 2025).

Las fórmulas específicas utilizadas para calcular estos índices, adaptadas a las bandas correspondientes de cada sensor (Landsat 5 TM, Landsat 8 OLI, Sentinel-2 MSI), se presentan en la Tabla 6. A modo de ilustración del procesamiento espectral realizado, se muestran en la Figura 35 los mapas resultantes del Índice de Vegetación de Diferencia Normalizada (NDVI) para los tres periodos de análisis. La colección completa de visualizaciones para el resto de índices (SAVI, NDBI, NDMI, BSI, UI) se encuentra disponible en el repositorio digital del proyecto (Apéndice E).



**Figura 35**

*Distribución espacial del Índice de Vegetación de Diferencia Normalizada (NDVI) a lo largo del periodo de estudio.*

**Variables Topográficas.** Las variables topográficas, como la elevación y la pendiente derivadas de un Modelo Digital de Elevación (DEM), se incorporan frecuentemente en la modelización y clasificación LULC. Se consideran factores ambientales estáticos determinantes que pueden mejorar la precisión del análisis, especialmente en áreas con relieve complejo o accidentado, al ayudar a discriminar clases con firmas espectrales similares pero diferente distribución altitudinal o de pendiente (Belay et al., 2024). En este estudio se utilizaron:

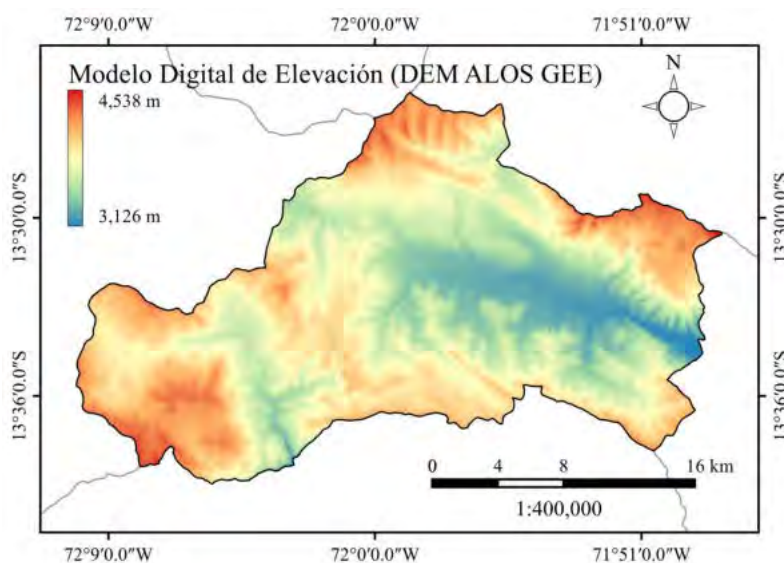
**Tabla 6***Fórmulas de Índices Espectrales por Sensor.*

Índice	L5 2004 (TM)	L8 2014 (OLI)	S2 2024 (MSI)
NDVI	$\frac{SR\_B4 - SR\_B3}{SR\_B4 + SR\_B3}$	$\frac{SR\_B5 - SR\_B4}{SR\_B5 + SR\_B4}$	$\frac{B8 - B4}{B8 + B4}$
SAVI	$\frac{(SR\_B4 - SR\_B3) \times 1.5}{SR\_B4 + SR\_B3 + 0.5}$	$\frac{(SR\_B5 - SR\_B4) \times 1.5}{SR\_B5 + SR\_B4 + 0.5}$	$\frac{(B8 - B4) \times 1.5}{B8 + B4 + 0.5}$
NDBI	$\frac{SR\_B5 - SR\_B4}{SR\_B5 + SR\_B4}$	$\frac{SR\_B6 - SR\_B5}{SR\_B6 + SR\_B5}$	$\frac{B11 - B8}{B11 + B8}$
NDMI	$\frac{SR\_B4 - SR\_B5}{SR\_B4 + SR\_B5}$	$\frac{SR\_B5 - SR\_B6}{SR\_B5 + SR\_B6}$	$\frac{B8 - B11}{B8 + B11}$
BSI	$\frac{(SR\_B5 + SR\_B3) - (SR\_B4 + SR\_B1)}{(SR\_B5 + SR\_B3) + (SR\_B4 + SR\_B1)}$	$\frac{(SR\_B6 + SR\_B4) - (SR\_B5 + SR\_B2)}{(SR\_B6 + SR\_B4) + (SR\_B5 + SR\_B2)}$	$\frac{(B11 + B4) - (B8 + B2)}{(B11 + B4) + (B8 + B2)}$
UI	$\frac{SR\_B7 - SR\_B4}{SR\_B7 + SR\_B4}$	$\frac{SR\_B7 - SR\_B5}{SR\_B7 + SR\_B5}$	$\frac{B12 - B8}{B12 + B8}$

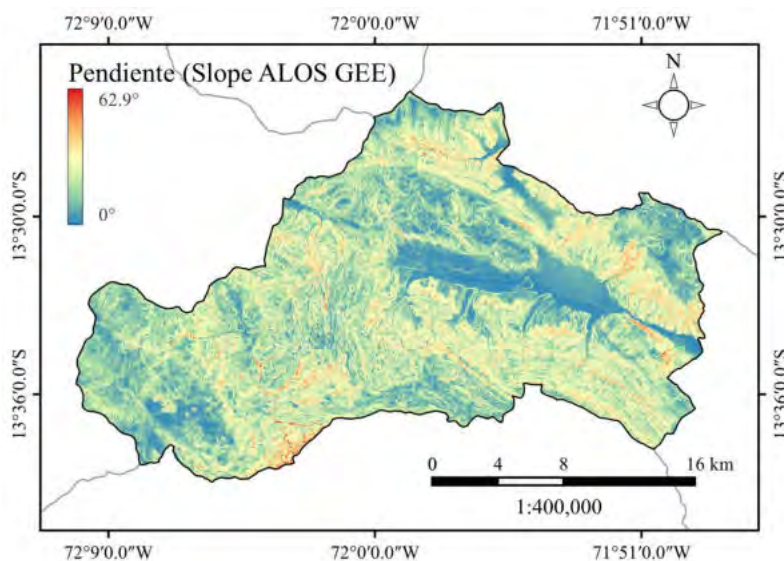
*Nota:* Las fórmulas generales se encuentran en la Tabla 4.

- Elevación: Altitud sobre el nivel del mar, obtenida directamente del DEM (Figura 36).
- Pendiente: Grado de inclinación del terreno, calculado a partir del DEM y expresado en grados (Figura 37).

Ambas variables se derivaron del DEM ALOS World 3D - 30m (AW3D30), disponible en Google Earth Engine.

**Figura 36***Modelo de Elevación Digital (DEM) de la provincia de Cusco.*

*Nota. Fuente:* ALOS AW3D30 vía GEE

**Figura 37***Mapa de Pendientes de la Provincia de Cusco.**Nota. Derivado del DEM ALOS AW3D30.*

**Normalización de Variables.** La normalización de las variables predictoras es un procedimiento habitual en aprendizaje automático para asegurar que todas las variables contribuyan de manera equitativa al modelo, independientemente de sus unidades o rangos originales. En este estudio:

- **Índices Espectrales:** Los índices calculados (NDVI, SAVI, NDBI, NDMI, BSI, UI), por su formulación como diferencias normalizadas, producen valores en un rango acotado  $[-1, +1]$ . Esta normalización intrínseca reduce la sensibilidad a variaciones en las condiciones de iluminación (efectos solares o topográficos) y facilita la comparación entre diferentes sensores y fechas. Por lo tanto, no se aplicó una normalización adicional a estos índices.
- **Variables Topográficas:** Las variables de elevación (metros) y pendiente (grados) se incorporaron al conjunto de datos con sus unidades y rangos físicos originales. No se aplicó escalado o normalización adicional a estas variables, dado que el algoritmo clasificador seleccionado, Random Forest, es inherentemente insensible a la escala de las variables de entrada.

**Características de las Imágenes Compuestas Finales.** Tras la adición de los índices espectrales y las variables topográficas a las imágenes compuestas de mediana, se obtuvieron los conjuntos de datos finales (imágenes multibanda). Las Tablas 7 y 8 resumen las bandas incluidas en cada compuesto final y sus respectivos rangos de valores (mínimo y máximo observados dentro del área de estudio).

Un aspecto metodológico importante para asegurar la validez y consistencia del análisis multitemporal fue la estandarización de la resolución espacial. Todas las bandas de entrada para la clasificación (bandas espectrales originales, índices calculados y variables topográficas) fueron homogeneizadas a una resolución espacial común de 30 metros.

Esta decisión se basó en la resolución nativa de 30 metros de los sensores históricos clave utilizados (Landsat 5 TM y Landsat 8 OLI). Por consiguiente, las bandas de Sentinel-2, que originalmente poseen resoluciones más finas (10 m y 20 m), fueron remuestreadas a 30 metros antes de ser combinadas con las demás variables. Este paso asegura que la comparación de LULC entre los diferentes años se realice sobre una base espacial coherente.

**Tabla 7**

*Comparativa de rangos dinámicos: Landsat 5 (2004) y Landsat 8 (2014).*

Banda / Índice	Landsat 5 (2004)		Landsat 8 (2014)	
	Mínimo	Máximo	Mínimo	Máximo
SR_B1	0.0051	0.3291	-0.0854	0.3769
SR_B2	0.0063	0.3879	-0.0630	0.4484
SR_B3	0.0113	0.4233	0.0031	0.5365
SR_B4	0.0135	0.4925	0.0007	0.6190
SR_B5	0.0086	0.6138	0.0108	0.6731
SR_B6	-	-	0.0086	1.0236
SR_B7	0.0030	0.5849	0.0056	1.0272
NDVI	-0.2656	0.8484	-0.4064	0.9650
SAVI	-0.0356	0.6555	-0.0662	0.7050
NDBI	-0.6207	0.5014	-0.7878	0.4667
NDMI	-0.5014	0.6207	-0.4667	0.7878
BSI	-0.4632	0.4576	-0.6716	0.4257
UI	-0.8736	0.4895	-0.8362	0.4566
Elevación (m)	2967	4642	2967	4642
Pendiente (grados)	0.0	70.8944	0.0	70.8944

Nota. Fuente: Elaboración propia.

**Tabla 8***Bandas y rangos Sentinel-2 (2024).*

Banda / Índice	Valor Mínimo	Valor Máximo
B2	55	3559
B3	393	6242
B4	271	7147
B5	279	5829
B6	247	5996
B7	251	5843
B8	229	7068
B8A	219	5719
B11	173.5	8770.5
B12	115	9708
NDVI	-0.1810	0.7531
SAVI	-0.2714	1.1295
NDBI	-0.5388	0.4283
NDMI	-0.4283	0.5388
BSI	-0.4653	0.3854
UI	-0.7960	0.4357
Elevación (m)	2967	4642
Pendiente (grados)	0.0	70.8944

*Nota. Valores de compuesta invierno austral 2024. Fuente: Elaboración propia.*

### 4.3 Clasificación Supervisada de Coberturas (LULC)

#### 4.3.1 Definición de Clases de Cobertura para el Estudio

Para el presente estudio, se definió un esquema de clasificación compuesto por tres clases temáticas principales: Urbano, Vegetación y Suelo Desnudo. Esta agrupación responde a dos consideraciones principales: (i) la marcada estacionalidad del régimen climático en la provincia de Cusco, caracterizada por una estación seca pronunciada (Mamani & Cutipa, 2024), y (ii) la semejanza espectral que presentan algunas coberturas durante dicho periodo, lo que dificulta su separación precisa en sensores ópticos. El propósito es obtener mapas LULC comparables entre años, priorizando la estabilidad espectral de las clases.

#### 1. Vegetación

Clase que agrupa coberturas con actividad fotosintética sostenida durante el invierno austral, orientada a capturar formaciones vegetales con alta estabilidad espacial y

espectral entre los años analizados, y con persistencia estructural a lo largo del ciclo anual.

Durante la estación seca, los cultivos de secano permanecen en barbecho (Cahuana & Pariguana, 2023), y la vegetación herbácea estacional (como pajonales altoandinos) reduce su actividad fotosintética por limitación hídrica y temperatura (Carlos Rosales, 2025). Esto permite distinguir espectralmente la vegetación persistente, que mantiene mayor estabilidad.

## 2. Suelo Desnudo

Clase que representa superficies sin vegetación activa. Diversos estudios agrupan tanto suelos expuestos naturales como superficies agrícolas en descanso dentro de esta categoría (Rotich et al., 2025b; Yadav et al., 2024).

Incluye:

- Zonas erosionadas o rocosas.
- Caminos sin pavimentar.
- Terrenos agrícolas en barbecho.
- Ecosistemas herbáceos o arbustivos que reducen fuertemente su verdor en la estación seca (Pajonal de Puna Húmeda, Matorral Andino) (Ministerio del Ambiente (MINAM), 2015).

Estas coberturas presentan firmas espectrales similares a la cobertura urbana, lo cual constituye una fuente recurrente de confusión en la clasificación LULC (Ettehadi et al., 2019; Ganjirad, 2024). Su agrupación contribuye a reducir la ambigüedad espectral y mejorar la precisión del proceso de clasificación.

## 3. Urbano

Clase que agrupa superficies impermeables y construidas asociadas a asentamientos humanos.

- Edificaciones residenciales, comerciales e industriales.

- Infraestructura vial pavimentada.
- Otras superficies artificiales (plazas, losas deportivas, etc.).

**Exclusión de la Clase “Agua”.** Aunque existen cuerpos de agua (principalmente ríos) dentro de la provincia de Cusco, se tomó la decisión metodológica de no incluir una clase específica “Agua” en el esquema de clasificación. Esta exclusión se fundamenta principalmente en la resolución espacial de los datos satelitales utilizados (30 metros para Landsat, homogeneizado para Sentinel-2). Los cauces de los ríos principales en la provincia se reducen durante la temporada seca (invierno austral), no superando los 30 metros de ancho. Además, la presión urbana y la ocupación de las fajas marginales han contribuido a un angostamiento considerable de algunos cauces en las últimas décadas (Calvo Mamani & Polo Dolmos, 2017; Paredes Catunta, 2019).

Como consecuencia, muchos tramos de los ríos representan elementos lineales cuyo ancho es frecuentemente inferior a la dimensión de un píxel de 30 metros, convirtiéndolos en características sub-píxel. Intentar clasificar estos elementos como una clase separada resultaría en una representación cartográfica inconsistente y fragmentada. Además, la presencia de píxeles mixtos (agua/suelo/vegetación) a lo largo de los bordes de los ríos podría introducir ruido espectral y afectar negativamente la precisión de la clasificación de las otras tres clases principales. Por estas razones, se optó por no mapear explícitamente la clase Agua.

#### **4.3.2 Recopilación de Datos de Entrenamiento y Validación**

La selección de muestras de referencia (datos de entrenamiento y validación), también conocida como obtención de verdad terreno (*ground truth*), es una etapa importante en cualquier proceso de clasificación supervisada de imágenes de teledetección para cartografía LULC (K. C. Roy et al., 2024). La calidad, cantidad y representatividad de estas muestras influye directamente en el rendimiento del clasificador entrenado. La mayoría de los estudios LULC se basan en la interpretación visual de imágenes de referencia para delimitar espacialmente áreas homogéneas representativas de cada clase de interés (Sarif & Gupta, 2024).

Para la recolección de las muestras en este estudio, se utilizó como fuente principal



de referencia las imágenes históricas de alta resolución disponibles en la plataforma Google Earth Pro. Esta herramienta es ampliamente utilizada en la literatura científica para la selección manual interactiva de puntos o polígonos de entrenamiento y validación en estudios LULC (Ganjirad, 2024; Gündüz, 2025; Zafar, Zubair, Zha, Fahd & Nadeem, 2024). El proceso de muestreo se adaptó a la disponibilidad temporal de estas imágenes de alta resolución para cada uno de los años de estudio:

- Año 2024: Se utilizaron imágenes de alta resolución correspondientes a mayo de 2024, próximas al período de invierno austral y coherentes con la ventana temporal de análisis.
- Año 2014: Se emplearon imágenes de alta resolución disponibles para junio de 2014, igualmente alineadas con la estación seca y adecuadas para la comparación interanual.
- Año 2004: Debido a la ausencia de imágenes de alta resolución directamente asociadas al año 2004 en Google Earth Pro, se utilizó como referencia principal el mosaico más cercano disponible, correspondiente a agosto de 2002. Para garantizar la validez temporal de la interpretación, especialmente en áreas urbanas en expansión, esta referencia visual fue complementada con el Mapa de Peligros de la ciudad del Cusco del 2004, que documenta la configuración urbana hasta 2004 (y PNUD, 2004). Este procedimiento de validación cruzada con cartografía auxiliar es consistente con estrategias adoptadas en estudios LULC con limitaciones similares en la disponibilidad de datos históricos (Pham & Ali, 2024).

Con el objetivo de mitigar los problemas de clasificación errónea (*misclassification*) asociados al fenómeno del píxel mixto, donde un único píxel de la imagen satelital abarca múltiples tipos de cobertura del suelo (Ganjirad, 2024), se adoptó una estrategia de muestreo conservadora. El etiquetado de las muestras (asignación de la clase LULC) se realizó aplicando criterios estrictos, seleccionando únicamente puntos ubicados en el centro de áreas que presentaban una alta homogeneidad espectral y textural en la imagen de alta resolución de referencia (Brown et al., 2022). Este enfoque, centrado en identificar píxeles “puros”, es particularmente importante en zonas de transición entre coberturas o en áreas conocidas

por su alta confusión espectral, buscando garantizar que cada punto de muestra represente inequívocamente una única clase LULC (Youssef, 2024).

Mediante un procedimiento de interpretación visual y selección en Google Earth Pro, se recolectó un conjunto de puntos de muestra para las tres clases definidas: Urbano, Vegetación y Suelo Desnudo. La cantidad de puntos recolectados para cada clase se presentan en la Tabla 9 y en las figuras 38, 39, 40 y 41.

**Tabla 9**

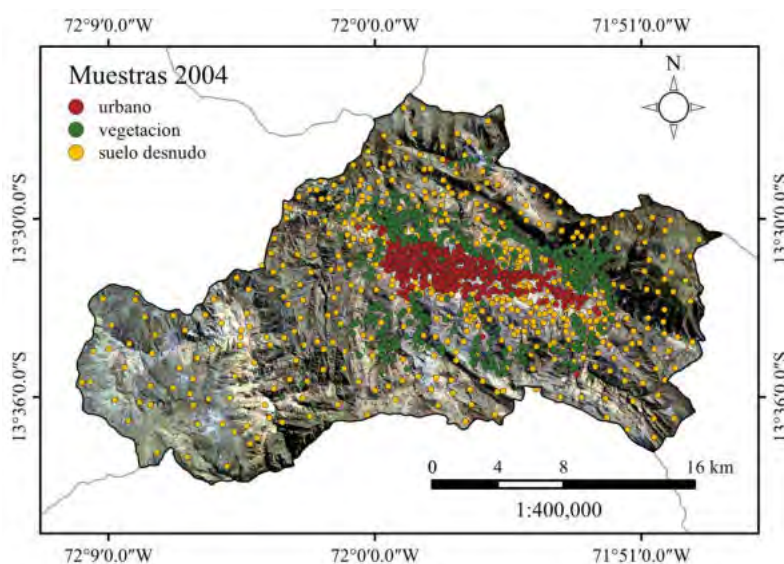
*Distribución de los puntos de muestra recolectados*

Clase	2004	2014	2024	Total por clase
Urbano	500	500	500	1500
Vegetación	500	500	500	1500
Suelo desnudo	511	510	500	1521
Total por año	1511	1510	1500	

*Nota.* El área de estudio contiene un total de 589,049 píxeles.

**Figura 38**

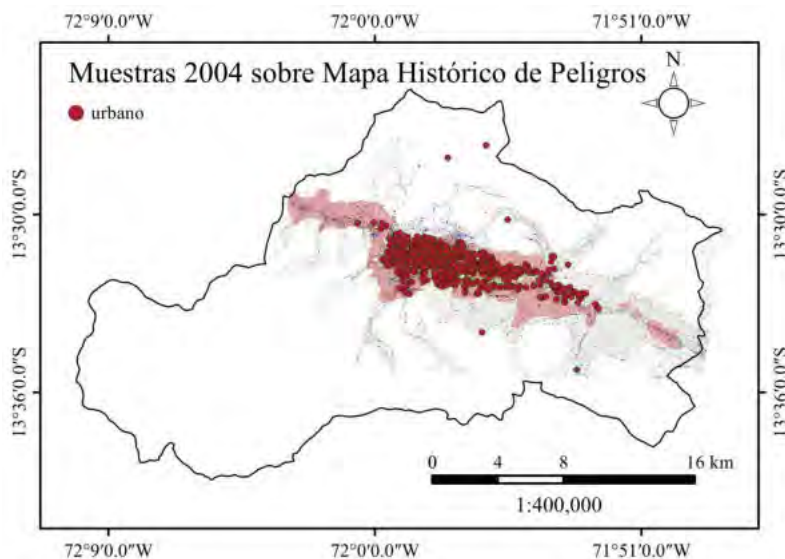
*Distribución espacial de los puntos de muestra 2004.*



*Nota. Fuente: Elaboración propia.*

**Figura 39**

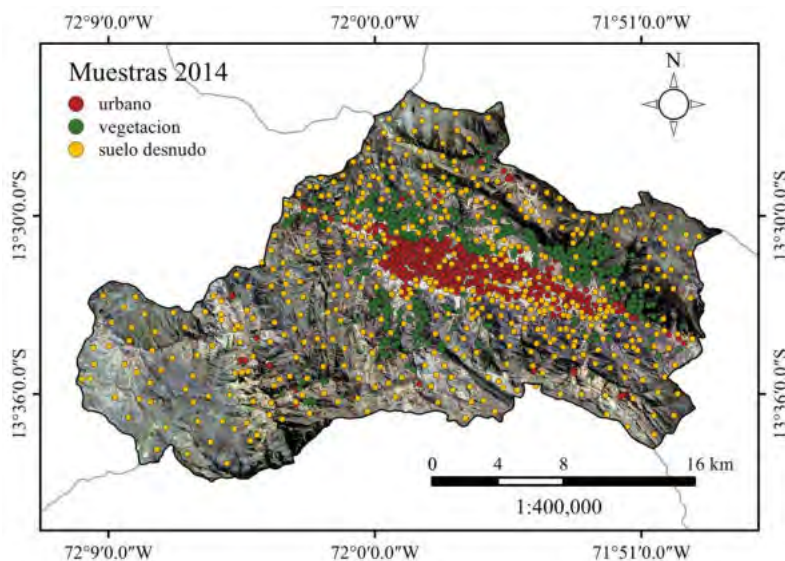
*Puntos de muestra sobre Mapa de Peligros (2004).*



*Nota. Fuente: Elaboración propia.*

**Figura 40**

*Distribución espacial de los puntos de muestra 2014.*



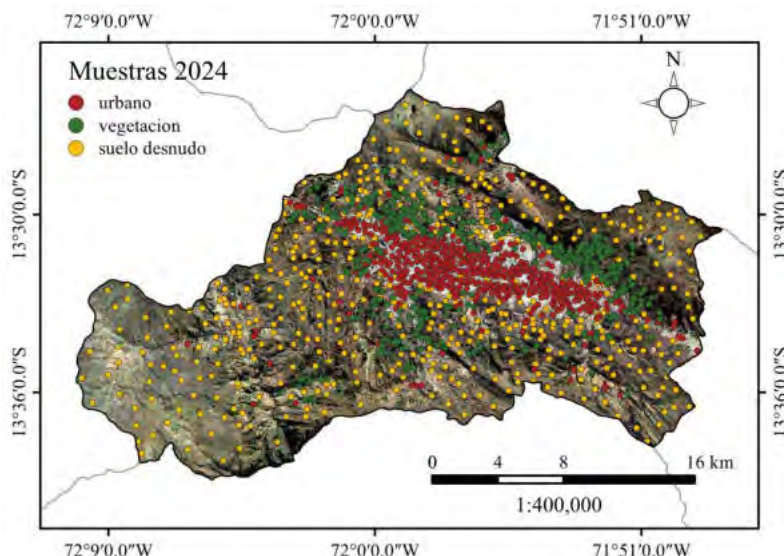
*Nota. Fuente: Elaboración propia.*

#### **4.3.3 Optimización de Hiperparámetros y Entrenamiento del Modelo**

Para la generación de los mapas LULC correspondientes a los años 2004, 2014 y 2024, se implementó un flujo de trabajo metodológico consistente dentro de la plataforma Google Earth Engine (GEE). A continuación, se detalla el proceso aplicado de forma independiente para cada año.

**Figura 41**

*Distribución espacial de los puntos de muestra 2024.*



*Nota. Fuente: Elaboración propia.*

**Preparación de Datos de Entrenamiento y Validación.** Se integraron en GEE los conjuntos de puntos de muestra para cada año. El total de puntos se dividió aleatoriamente en dos subconjuntos mutuamente excluyentes: el 70 % se utilizó para entrenar el clasificador y el 30 % restante para la validación independiente de la precisión. Esta proporción (70/30) es una práctica estándar en estudios de clasificación supervisada en teledetección (Badshah et al., 2024; Yasin, 2024; Zafar, Zubair, Zha, Fahd & Nadeem, 2024).

Luego, tanto para los puntos del conjunto de entrenamiento como para los de validación, se extrajeron los valores de todas las variables predictoras (bandas espectrales, índices espectrales y variables topográficas) directamente desde la imagen compuesta correspondiente a cada año. Estos valores constituyeron la matriz de atributos utilizada como entrada para el modelo Random Forest.

**Optimización de Hiperparámetros del Clasificador Random Forest.** Los hiperparámetros de un modelo de aprendizaje automático son parámetros de configuración que no se aprenden directamente de los datos durante el entrenamiento, sino que deben establecerse previamente. Controlan aspectos importantes del proceso de aprendizaje y pueden influir en el rendimiento (precisión) y la complejidad del modelo resultante (B. Roy, 2021). Para el algoritmo Random Forest (RF), dos de los hiperparámetros con mayor influencia son el número de árboles en el

bosque y la fracción de muestras utilizada para entrenar cada árbol individual.

Para identificar la configuración de hiperparámetros que ofreciera el mejor rendimiento predictivo para cada año, se implementó un procedimiento de ajuste sistemático (búsqueda en cuadrícula). Se definió un espacio de búsqueda para ambos hiperparámetros: el número de árboles se varió desde 10 hasta 150 (en incrementos de 10), y la fracción de muestreo por bolsa se varió desde 0.1 hasta 0.9 (en incrementos de 0.1).

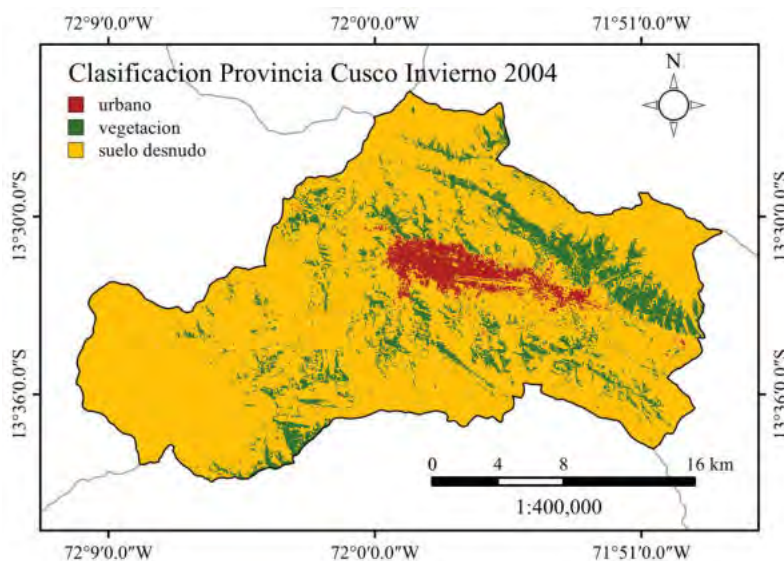
Para cada combinación posible de (*numTrees*, *bagFraction*) dentro de esta cuadrícula, se entrenó un modelo RF utilizando el conjunto de entrenamiento y se evaluó su precisión (Overall Accuracy) utilizando el conjunto de validación independiente.

**Selección y Entrenamiento del Modelo RF Óptimo.** Para cada año, se seleccionó la combinación de hiperparámetros (*numTrees*, *bagFraction*) que produjo la mayor Precisión Global (OA) en el conjunto de validación durante la fase de optimización (detallada en la Sección 5.1.1). Utilizando estos parámetros óptimos identificados, se procedió a entrenar el modelo Random Forest final para cada año, empleando el conjunto de entrenamiento. Este procedimiento sigue las prácticas recomendadas en la literatura para la calibración y selección de modelos de clasificación supervisada (Gündüz, 2025).

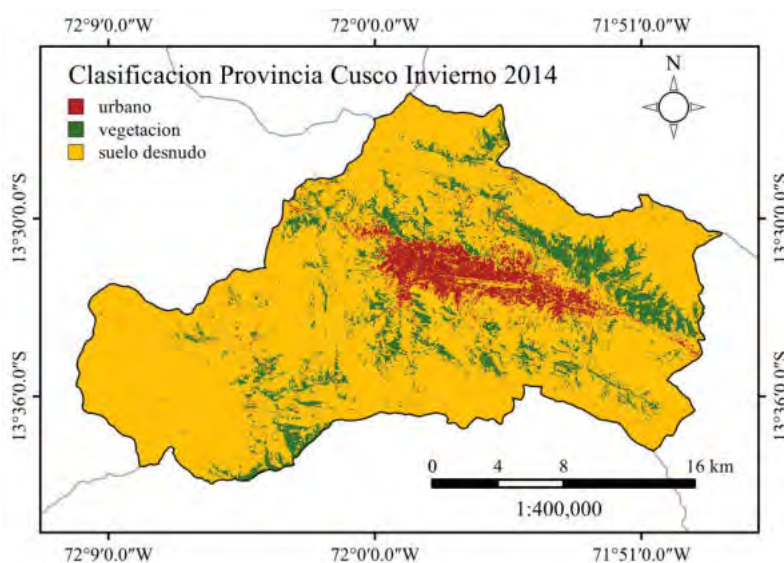
**Generación y Validación de los Mapas LULC.** El modelo RF óptimo entrenado para cada año fue aplicado a la imagen compuesta correspondiente (incluyendo todas las bandas predictoras) para generar los mapas clasificados LULC. La calidad y fiabilidad de cada uno de estos mapas clasificados se evaluó cuantitativamente utilizando el conjunto de validación independiente. Se construyó la matriz de confusión para cada año y se calcularon las métricas de precisión estándar derivadas: Precisión Global (OA), Coeficiente Kappa ( $\kappa$ ), Precisión del Productor (PA o Recall) por clase, y Precisión del Usuario (UA o Precision) por clase.

#### **4.3.4 Exportación de Mapas Clasificados y Análisis de Cambios**

Los mapas LULC clasificados 42, 43 y 44 fueron exportados desde Google Earth Engine en formato raster. Las imágenes (.tif) se encuentran disponibles en el repositorio digital (Apéndice E).

**Figura 42***Mapa LULC clasificado (2004).*

*Nota. Mapa de Coberturas y Usos del Suelo (LULC) clasificado para el año 2004. Fuente: Elaboración propia.*

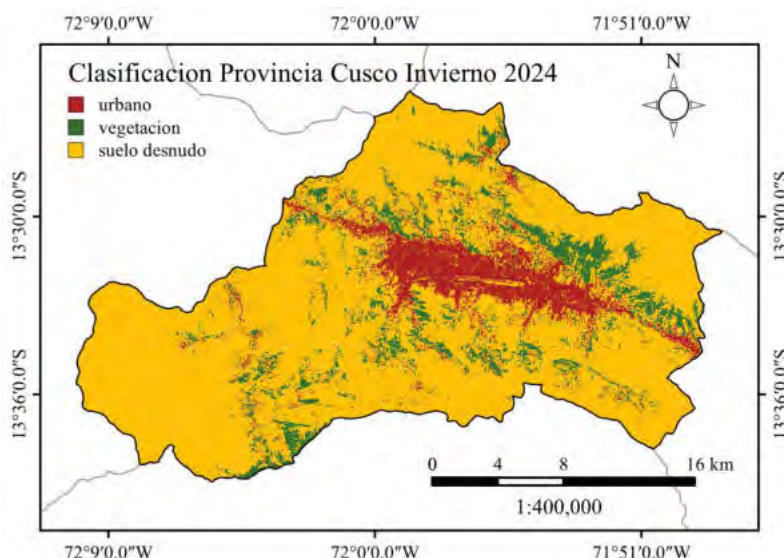
**Figura 43***Mapa LULC clasificado (2014).*

*Nota. Mapa de Coberturas y Usos del Suelo (LULC) clasificado para el año 2014. Fuente: Elaboración propia.*

#### **4.4 Análisis de Cambios Multitemporal**

El análisis cuantitativo de los cambios en la cobertura del suelo permite identificar las transformaciones ocurridas en el área de estudio. El método consiste en procesar los mapas de cobertura del suelo clasificados (Figuras 42, 43 y 44) mediante tabulación cruzada (*cross-*



**Figura 44***Mapa LULC clasificado (2024).*

*Nota. Mapa de Coberturas y Usos del Suelo (LULC) clasificado para el año 2024. Fuente: Elaboración propia.*

*tabulation).*

Este procedimiento permite cuantificar las transiciones entre clases, identificando para cada período las áreas de ganancia, pérdida y persistencia. Los resultados cuantitativos de este análisis se detallan en la Sección 5.1.2.

## **4.5 Simulación Híbrida (MLP-CA-MC)**

### **4.5.1 Modelado de la Simulación Híbrida (MLP-CA-MC)**

**Preparación de Variables Predictoras.** La selección de las variables predictoras (factores impulsores) es una etapa determinante que define la capacidad del modelo para simular la dinámica territorial. Las variables empleadas en este estudio fueron seleccionadas con base en su uso documentado en la literatura. Se incorporan factores topográficos y biofísicos, como la pendiente (Belay et al., 2024; Bendechou et al., 2024), el aspecto (Maddah et al., 2025; Rotich et al., 2025b), la elevación (Alam & Maiti, 2025; Chaulagain et al., 2025) y la proximidad a la red hidrológica (Danso et al., 2025; Rotich et al., 2025a). Asimismo, se consideran los principales impulsores de accesibilidad y atracción antrópica, incluyendo la distancia a carreteras (Al-Kordi et al., 2025; Y. Liu et al., 2024b), vías férreas (G. Liu et al., 2024; Zhou et al., 2025), el

aeropuerto, poblaciones dispersas (Ishtiaque et al., 2021; Khan Academy, 2024), poblaciones grandes (centros urbanos) (Y. Liu et al., 2024b; Sarfo et al., 2024), centros educativos (T. Xu et al., 2022) y hospitales (T. Xu et al., 2022). Finalmente, se incluyó la distancia a atractivos turísticos. Aunque otros estudios mencionan su influencia (M. Khalid et al., 2023; Senthilkumar et al., 2025), en esta investigación se agrega un ráster de proximidad para capturar el rol potencial del turismo como un factor impulsor de cambio.

Para la preparación de las variables espaciales predictoras se aplicó un flujo de trabajo estandarizado en QGIS. Este procedimiento garantizó la homogeneidad espacial de todos los datos, asegurando una misma resolución (30 metros), sistema de referencia cartográfica (WGS 84 / UTM Zona 19S, EPSG:32719) y extensión espacial. La extensión fue definida por un polígono vectorial, y fue configurada explícitamente para todos los rásteres resultantes con la siguiente definición espacial: 157140, 196380, 8488950, 8513400 [EPSG:32719].

El proceso comprendió la recopilación, delimitación y estandarización de los datos vectoriales y ráster originales, así como la generación de variables derivadas. Adicionalmente, se realizó un proceso de depuración y filtrado de datos para asegurar la calidad de los factores impulsores para las capas asociadas a instituciones educativas y centros de salud. Se partió de bases tabulares con información a nivel nacional, de las cuales se seleccionaron únicamente los registros ubicados dentro del departamento y la provincia del Cusco, verificando la existencia de datos de coordenadas (latitud y longitud) y la disponibilidad de la fecha de inicio de actividades. Este proceso permitió construir tres versiones temporales de cada conjunto de datos (hasta 2004, hasta 2014 y hasta 2024), preservando la consistencia temporal requerida para el modelado. Las etapas de limpieza y estructuración de la base de datos se llevaron a cabo en Google Colab usando la biblioteca Pandas, para posteriormente integrarse en la generación de mapas de proximidad en QGIS.

El procedimiento metodológico detallado para la construcción de las variable predictora se presenta en el repositorio virtual Apéndice E, mientras que la visualización de los mapas ráster finales se muestra en el Apéndice C.

***Análisis de Multicolinealidad.*** El análisis de multicolinealidad es una etapa esencial en la preparación de variables predictoras para modelos estadísticos y de aprendizaje automático,



incluyendo aquellos utilizados para modelar la susceptibilidad a cambios o la predicción LULC (Hanh et al., 2025). Su propósito es identificar la presencia de altas correlaciones lineales entre pares o grupos de variables predictoras independientes (Z. Xu et al., 2025). La multicolinealidad severa puede introducir inestabilidad en la estimación de los parámetros del modelo, dificultar la interpretación de la influencia individual de cada variable y reducir la capacidad de generalización del modelo (Hanh et al., 2025).

En este estudio, se evaluó de colinealidad para el siguiente conjunto inicial de variables predictoras:

1. Aspecto (orientación de la ladera, derivado del DEM Copernicus de 30 m)
2. Pendiente (grado de inclinación, derivada del DEM Copernicus de 30 m)
3. Distancia a la red hidrológica
4. Distancia a atractivos turísticos más ofertados (Apéndice D).
5. Distancia a poblaciones dispersas
6. Distancia a poblaciones grandes (centros urbanos)
7. Elevación (DEM)
8. Distancia a Instituciones Educativas (IIEE) (2004, 2014, 2024)
9. Distancia a Establecimientos de Salud (RENIPRESS) (2004, 2014, 2024)
10. Distancia a Vías Férreas
11. Distancia al Aeropuerto
12. Distancia a Vías (Carreteras)

El método seleccionado fue el cálculo de las matrices de correlación de Pearson ( $\rho$ ) entre todas las variables predictoras para cada período temporal.

Se estableció como criterio de decisión que cualquier par de variables que presentara una alta correlación, definida por un umbral de  $|\rho| \geq 0.8$  (Buthelezi et al., 2024; Tayyab et al.,

2024), sería evaluado para la exclusión de una de las variables. El objetivo de este procedimiento metodológico es reducir la redundancia de información.

Los resultados visuales se presentan en las Figuras 69, 70 y 71.

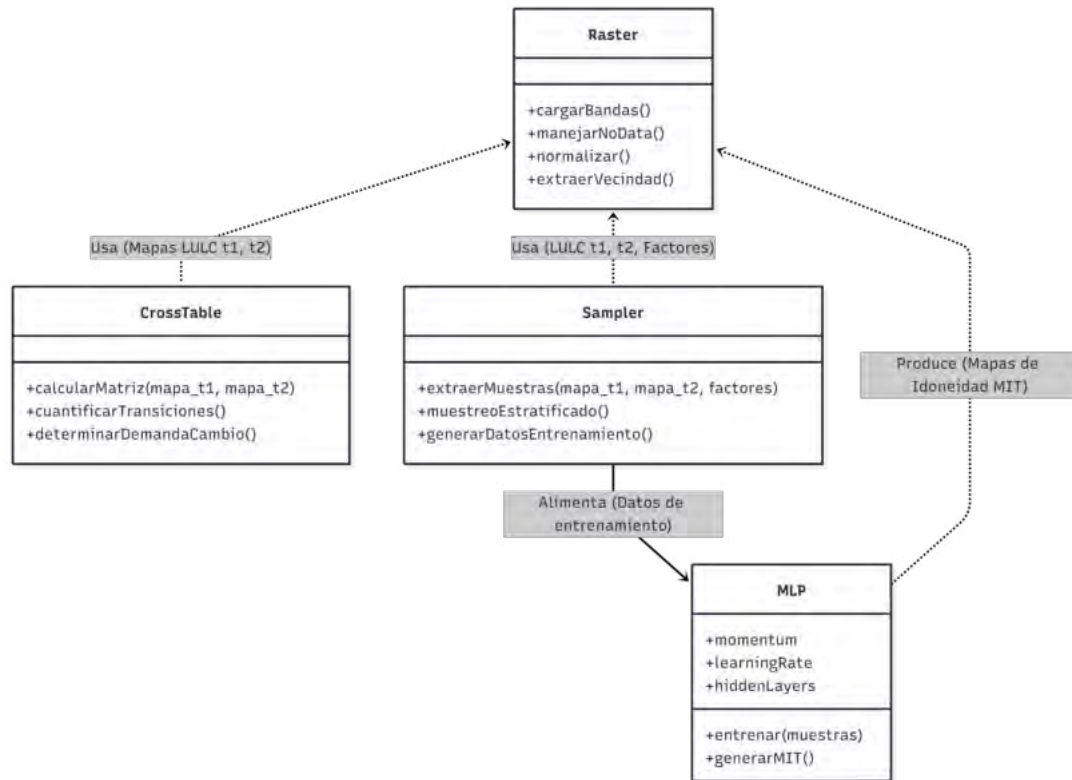
- Exclusión de variables de transporte: Se detectó una alta multicolinealidad ( $|\rho| > 0.83$ ) entre *distancia a vías*, *distancia al aeropuerto* y *distancia a las vías férreas*. Se decidió excluir las variables de aeropuerto y vías férreas, conservando únicamente distancia a vías como el indicador principal de accesibilidad.
- Exclusión de variables de servicios: Se identificó una correlación elevada ( $\rho \approx 0.82$ ) entre *Instituciones Educativas* y *RENIPRESS*. Se optó por mantener Instituciones Educativas como el indicador de servicios.

**Arquitectura de Clases Implementada en Colab.** Todas las clases utilizadas en el procesamiento geoespacial, el análisis de transición temporal, el aprendizaje de reglas espaciales y simulación fueron implementadas en Google Colab. La Figura 45 resume la arquitectura general del sistema, mostrando las clases principales y sus relaciones.

**Procesamiento de Datos Geoespaciales.** La clase *Raster* se encarga de las operaciones de entrada/salida y manipulación de los datos ráster utilizados por el modelo. Sus funciones incluyen la carga de bandas, manejo de NoData, normalización y extracción de vecindades espaciales.

**Análisis de Transición Temporal.** La clase *CrossTable* implementa el núcleo del análisis de Markov. A partir de dos mapas LULC de fechas distintas ( $t_1$  y  $t_2$ ), calcula una matriz de tabulación cruzada que cuantifica el número de píxeles que han transicionado de cada clase en  $t_1$  a cada clase en  $t_2$ . Esta matriz empírica es la base para determinar la demanda de cambio que se utilizará en la simulación del AC.

**Aprendizaje de Reglas de Transición Espacial.** Para el modelado de la dinámica espacial, se optó por una implementación explícita de la arquitectura Perceptrón Multicapa utilizando librerías de cálculo numérico (*NumPy*). Esta decisión prioriza la transparencia algorítmica

**Figura 45***Esquema general de clases.**Nota. Fuente: Elaboración propia.*

sobre la abstracción de frameworks de alto nivel, garantizando una trazabilidad total del flujo de información.

- El **Sampler** extrae datos de entrenamiento del periodo histórico. Para cada píxel, genera una muestra que contiene: el estado de su vecindad en el mapa inicial ( $t_1$ ), los valores de los factores impulsores en su vecindad en  $t_1$ , y la clase resultante de ese píxel en el mapa final ( $t_2$ ). Además, debido a que la vecindad utilizada es de  $3 \times 3$  píxeles, el vector de entrada del MLP se compone de 9 valores por cada factor raster y 18 valores adicionales provenientes del estado (codificado con *one-hot* truncado, donde cada píxel aporta solo 2 variables). El tamaño total del vector de entrada es 99 componentes.
- Para estimar la matriz de transición histórica se emplearon los mapas clasificados de 2004 y 2014, lo que permitió identificar la magnitud de los cambios ocurridos entre ambas fechas y calcular el componente Markov del modelo.
- Para representar los factores impulsores del cambio se usó el conjunto de

variables disponibles para 2004, ya que reflejan las condiciones socioeconómicas y geográficas iniciales que explican las transformaciones observadas una década después.

- El Perceptrón Multicapa (MLP), se entrenó utilizando un conjunto de muestras obtenidas mediante un muestreo estratificado (*Stratified Sampling*). Este método de muestreo asegura una representación proporcional de las diferentes clases o transiciones en los datos de entrenamiento. El objetivo del entrenamiento es capacitar al modelo para discernir la relación no lineal entre las características espaciales locales (estado inicial de cobertura del suelo en la vecindad y valores de los factores impulsores) y la idoneidad de una celda para transicionar a una clase de cobertura futura específica.
- Para optimizar el rendimiento del MLP, se llevó a cabo un proceso de ajuste de hiperparámetros, explorando diferentes valores para los coeficientes de momentum y tasa de aprendizaje, el tamaño del conjunto de muestras, el número de épocas de entrenamiento y la arquitectura de las capas ocultas. La selección de los rangos y valores específicos para esta hiperparametrización se fundamentó en configuraciones y resultados documentados en la literatura científica previa sobre modelado de cambios de cobertura del suelo (Tabla 10).
- El modelo resultante es capaz de generar Mapas de Idoneidad para la Transición (MIT) para cada posible clase de destino.

**Tabla 10**  
*Hiperparámetros MLP en estudios LULC.*

Estudio	Capas Ocultas	Tasa Aprend.	Momento	Épocas	Tamaño Vecindad	Muestras
M. Khan et al., 2023	3	0.1	-	-	-	-
Gündüz, 2025	10	0.01	0.001	1,000	3x3	-
Mazroa et al., 2024	10	0.001	0.05	10,000	3x3	-
S. K. Roy et al., 2023	10	0.1	0.5	1,000	3x3	-

*Nota. La Tabla resume las configuraciones de hiperparámetros utilizadas en estudios de redes neuronales MLP aplicados a LULC. Fuente: Elaboración propia.*

Una limitación frecuente en la literatura es que, si bien se reporta el número total de capas ocultas, a menudo no se detalla la distribución de neuronas en cada una de ellas. Esta omisión dificulta la replicación exacta de las arquitecturas y la comparación de su complejidad (Blissag, Bilal et al., 2024; Kondum et al., 2024; Penfound & Vaz, 2024; S. K. Roy et al., 2023; Xiang, 2024). La Tabla 11 muestra arquitecturas reportadas.

**Tabla 11**

*Arquitecturas de capas ocultas reportadas en estudios LULC con MLP.*

Estudio	Nº de Capas Ocultas	Neuronas por Capa
Lukas et al., 2024	2	[100, 100]
Haydar et al., 2024	2	[100, 50]
Buthelezi et al., 2024	2	[10, 5]
Tasan et al., 2025	2	[50, 50]
Tehrani et al., 2024	2	[70, 30]

*Nota. Fuente: Elaboración propia.*

El proceso de búsqueda de hiperparámetros óptimos se realizó en dos etapas. En la primera, se exploró un amplio rango de valores para identificar regiones prometedoras del espacio de búsqueda. Posteriormente, se llevó a cabo una segunda etapa de búsqueda más refinada (búsqueda fina) enfocada en esas regiones. En ambos procesos de búsqueda se ejecutó cada configuración 5 veces para promediar los resultados de precisión. En total, se evaluaron más de 4700 combinaciones, de las cuales dos configuraciones destacaron (Tabla 12). Los registros tabulares completos de rendimiento para todas las configuraciones evaluadas en ambas fases se encuentran disponibles en el repositorio de datos (Apéndice E)

La Configuración A logró la mayor precisión media, pero presentó una desviación estándar más alta, lo que indica mayor variabilidad entre ejecuciones. La Configuración B obtuvo una precisión media ligeramente menor, con una desviación estándar más baja, reflejando un rendimiento más estable.

Dado que la diferencia en precisión media fue pequeña ( $\Delta\mu_{acc} \approx 0.0078$ ), se eligió la Configuración B para entrenar el modelo MLP final en las simulaciones posteriores.

**Entrenamiento Final con la Configuración Óptima.** Una vez determinada la configuración óptima de hiperparámetros mediante el proceso de búsqueda sistemática descrito previamente,

**Tabla 12**  
*Comparación de configuraciones MLP.*

Hiperparámetro	Configuración A	Configuración B
Vecindad (radio)	1 (3x3)	1 (3x3)
Muestras Entrenamiento	9000	9000
Épocas Entrenamiento	11000	11000
Arquitectura Capas Ocultas	[50, 50]	[100, 100]
Tasa de Aprendizaje ( $\alpha$ )	0.0007	0.0005
Momento ( $\beta$ )	0.001	0.006
Precisión Media ( $\mu_{acc}$ )	0.8613	0.8535
Desviación Estándar ( $\sigma_{acc}$ )	0.0039	0.0006

*Nota. Fuente: Elaboración propia.*

se procedió a entrenar el modelo final.

- Se emplearon distintas particiones del conjunto de datos para entrenamiento y validación, con el fin de recoger la variabilidad del dataset y evitar que el rendimiento dependa de una única división.
- Se utilizaron diferentes inicializaciones aleatorias de los pesos, considerando que el proceso de optimización en redes neuronales es sensible al punto de partida en el espacio de parámetros.

Este procedimiento permitió generar un conjunto de modelos entrenados bajo las mismas condiciones de arquitectura e hiperparámetros, pero sujetos a variaciones aleatorias controladas. Para la selección del modelo final, se comparó el desempeño de todas las ejecuciones sobre el conjunto de validación. Conforme a prácticas comunes en la literatura, se seleccionó la ejecución con el mejor rendimiento en validación, garantizando que la instancia final utilizada en el proyecto representara el mayor nivel de desempeño alcanzado por la configuración óptima de la red neuronal. La tabla comparativa con las métricas de precisión de las distintas ejecuciones candidatas se encuentra disponible para su consulta en el repositorio de datos (Apéndice E).

El modelo final utilizado para la generación de resultados y productos derivados corresponde a la ejecución que alcanzó una exactitud del 85.70 %, valor ligeramente superior al rendimiento promedio obtenido durante la etapa de búsqueda de hiperparámetros (85.35 %). Esta diferencia mínima evidencia la estabilidad del comportamiento de la red neuronal y la

consistencia de los parámetros óptimos identificados.

**Simulación de Validación - Año 2024.** Una vez entrenado y optimizado el modelo MLP (Sección 4.5.1), el siguiente paso consiste en ejecutar una simulación para el período 2014-2024. El objetivo de este proceso es generar un mapa LULC simulado para el año 2024 ( $L_{sim}(t_2)$ ), utilizando el mapa real de 2014 ( $L_{real}(t_1)$ ) como punto de partida.

Este mapa simulado se utiliza para la validación del modelo, ya que su posterior comparación con el mapa LULC *real* de 2024 (obtenido a partir de la clasificación satelital) permitirá cuantificar la precisión predictiva del modelo.

**Carga del Modelo Predictivo (MLP).** Se carga el modelo MLP que encapsula las reglas de idoneidad espacial (la relación entre los factores impulsores y la probabilidad de cambio) aprendidas a partir de la dinámica histórica 2004-2014.

**Cuantificación de la Demanda de Cambio (Markov).** Se carga la matriz de transición histórica (Tabla 22), calculada en la Sección 4.5.1. Esta matriz define la *cantidad* o “demanda” de píxeles  $D_{ij}$  que deben transicionar de una clase  $i$  a una clase  $j$  para replicar la tasa de cambio observada en el período 2004-2014, bajo el supuesto de inercia tendencial.

**Carga de Datos de Partida (Año 2014).** Se cargan y preparan los datos que servirán como estado inicial para la simulación:

- Mapa LULC Inicial: El mapa LULC clasificado del año 2014 ( $L(t_1 = 2014)$ ).
- Factores Impulsores: El conjunto de factores geoespaciales y socioeconómicos correspondientes al año 2014 ( $\mathbf{V}(t_1 = 2014)$ ). Estos factores se normalizan para coincidir con la escala de los datos de entrenamiento.

**Proceso de Simulación Geoespacial (2014 a 2024).** La clase `MlpSimulator` ejecuta la simulación combinando la lógica del MLP y el CA.

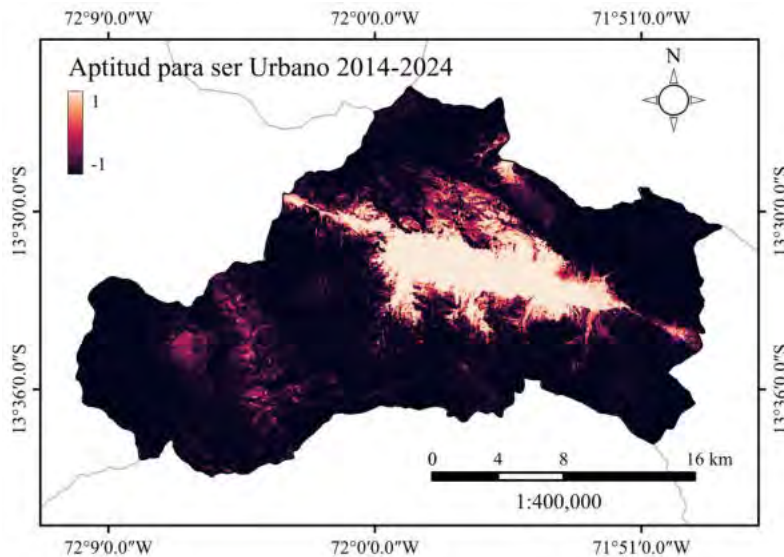
**Generación de Mapas de Idoneidad para 2024.** El modelo MLP entrenado  $f(\mathbf{x})$  se aplica a *cada píxel*  $p$  del mapa LULC de 2014. Para cada píxel, el modelo evalúa su estado, el de su vecindad y los valores de los factores impulsores de 2014 ( $\mathbf{V}(t_1 = 2014)$ ).

Este proceso genera un conjunto de Mapas de Potencial (Aptitud)  $\mathbf{S}_k$ , uno para cada clase destino  $k$ . Estos mapas cuantifican la aptitud espacial intrínseca para el período 2014-

2024 y son la base desde la cual se derivan los Mapas de Idoneidad de Transición. La Figura 46 muestra la distribución espacial del potencial para la clase Urbana. El conjunto completo de mapas de aptitud (para Vegetación y Suelo Desnudo) y los mapas de idoneidad de transición específicos se encuentran disponibles en el repositorio digital del proyecto (Apéndice E).

**Figura 46**

*Potencial para Ser de Cobertura Urbana (2014-2024).*



*Nota. Fuente: Elaboración propia.*

**Asignación Espacial de Transiciones (CA).** El Autómata Celular (CA) distribuye la demanda de cambio  $D_{ij}$  (obtenida de la matriz de Markov 2004-2014) sobre la cuadrícula, utilizando los mapas de idoneidad  $S_k$  como guía. Para cada transición  $iaj$ :

1. Se identifica la demanda total  $D_{ij}$ .
2. Se localizan todos los píxeles candidatos .
3. Dentro de este conjunto de candidatos, se seleccionan los  $D_{ij}$  píxeles que tienen los valores más altos en el mapa de idoneidad para la clase destino  $j$  ( $S_j(p)$ ).
4. Se cambia el estado de estos  $D_{ij}$  píxeles de  $i$  a  $j$  en el nuevo mapa simulado.

Este proceso se repite para todas las transiciones definidas en la matriz de demanda.

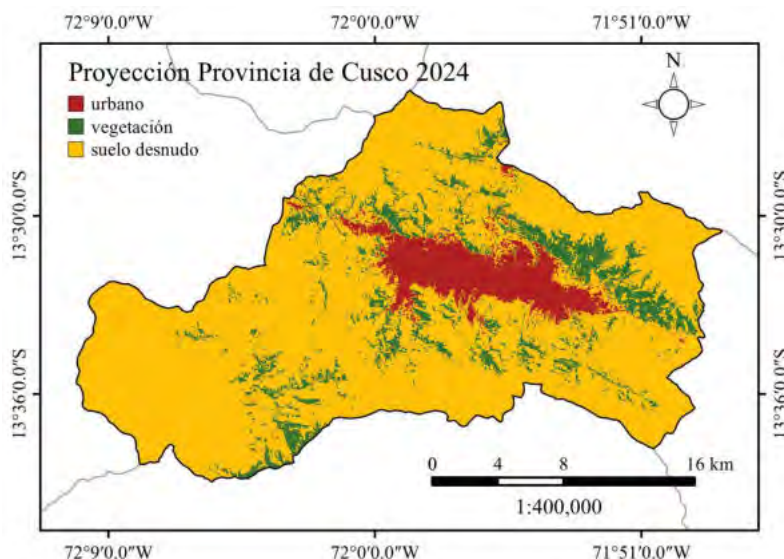
**Generación del Mapa LULC Simulado para 2024.** El mapa resultante  $L_{sim}(t_2 = 2024)$  representa el escenario LULC simulado para el año 2024. Este mapa (presentado en



la Figura 47) es el producto final de la simulación de validación y la base para el análisis de precisión que se detalla en la siguiente sección.

### Figura 47

*Mapa LULC simulado (2024).*



*Nota. Mapa de Coberturas y Usos del Suelo (LULC) simulado para la Provincia de Cusco, año 2024. Fuente: Elaboración propia.*

#### 4.5.2 Validación de la Precisión del Modelo de Simulación

La validación se realizó comparando el mapa LULC simulado para el año 2024 (Figura 47) con el mapa LULC real (clasificado a partir de imágenes satelitales) para el mismo año 2024 (Figura 44). Se emplearon diversas métricas para evaluar la concordancia entre ambos mapas: a nivel categórico (píxel a píxel), mediante AUC-ROC para evaluar el desempeño del modelo en la estimación del potencial de transición, y a nivel espacial (considerando la vecindad y agregación).

**Métricas de Precisión Categórica Estándar.** Se construyó una matriz de confusión, que permitió calcular métricas de precisión categórica píxel a píxel, incluyendo la exactitud del productor, la exactitud del usuario y el F1-score por clase, así como métricas globales de desempeño tales como la exactitud global, el acuerdo esperado por azar ( $P_e$ ) y el coeficiente Kappa estándar. Adicionalmente, se incorporó la descomposición del Kappa mediante el Kappa de histograma ( $K_{histo}$ ) para evaluar la correspondencia en términos de cantidad entre las clases

simuladas y observadas.

**Validación del Potencial de Transición mediante AUC-ROC.** Para determinar la capacidad del modelo MLP-CA-MC de discriminar correctamente entre categorías de transición, se implementó un procedimiento de evaluación basado en la Curva ROC. A diferencia de las métricas de clasificación dura (como Kappa), esta etapa se enfoca en medir la calidad de las probabilidades de transición generadas por la red neuronal.

***Generación del Conjunto de Validación.***

1. Muestreo Estratificado: Para garantizar la representatividad estadística de todas las categorías de transición, se configuró un muestreo estratificado (*Stratified Sampling*). Se fijó una semilla aleatoria ( $SEED = 42$ ) para asegurar la reproducibilidad exacta de la selección de muestras.
2. Construcción de la Matriz de Características ( $X$ ): Para cada píxel muestreado, se reconstruyó el vector de entrada concatenando el estado de la vecindad y los valores de los factores conductores (drivers) normalizados (DEM, Pendiente, Distancias, etc.).

***Cálculo de Probabilidades y Métrica AUC Global.*** Una vez estructurada la matriz de entrada  $X$  y el vector de etiquetas reales  $y$ , se ejecutó el siguiente flujo de cálculo:

1. Se ingresó la matriz  $X$  completa al modelo MLP pre-entrenado, obteniendo los valores de salida no normalizados de la última capa neuronal.
2. Dado que el cálculo del AUC requiere probabilidades, las salidas del modelo fueron transformadas mediante la función Softmax. Esta transformación permitió obtener una distribución de probabilidad por píxel, necesaria para la evaluación del rendimiento.
3. Debido a la naturaleza multiclase del problema, se implementó el cálculo del Área Bajo la Curva ROC para cada clase y se promediaron sus áreas para obtener una métrica global unificada ( $AUC_{base}$ ).

**Análisis de la Precisión Posicional y Espacial.** Para cuantificar y caracterizar la naturaleza de los errores de localización espacial, se aplicaron métricas de validación espacialmente explícitas, diseñadas para gestionar los “casi aciertos” inherentes al modelado espacial:

***Precisión Difusa (Fuzzy Accuracy).*** Esta métrica evalúa la concordancia permitiendo una tolerancia espacial. Se definió una ventana de vecindad móvil de  $3 \times 3$  píxeles. Bajo esta lógica, si el píxel central del mapa simulado no coincide con el real, pero la clase correcta se encuentra dentro de esa ventana de  $3 \times 3$  en el mapa simulado, se contabiliza como un “acierto difuso”. Esta métrica permite cuantificar los “casi aciertos”, píxeles clasificados correctamente en tipo (clase LULC), pero con un ligero desplazamiento espacial (error de localización  $\leq 30$  metros).

***Análisis de Validación Multiescala.*** Finalmente, se aplicó una técnica de validación multiescala para evaluar cómo cambia la concordancia al agregar espacialmente los mapas simulado y real a resoluciones progresivamente más gruesas (Tabla 26). El objetivo es observar si los errores de localización a escala fina se anulan, lo que indicaría que los patrones espaciales generales son correctos. Se analizó la tendencia del Coeficiente Kappa (Figura 73) y la descomposición de los errores (en cantidad y localización) en cada escala de agregación.

***Análisis de Patrón Clase Vegetación .*** La generación de los diagramas de rosa de los vientos se realizó mediante un procedimiento de análisis que integra la información de orientación de ladera con la distribución espacial de la clase Vegetación. En primer lugar, se extrajo el raster de aspecto y se discretizó en ocho clases direccionales (N, NE, E, SE, S, SW, W y NW), cada una asociada a un rango angular específico. Posteriormente, se generó una máscara binaria para la clase Vegetación, tanto para el mapa real como para el mapa simulado, identificando únicamente los píxeles pertenecientes a dicha cobertura.

Sobre esta base, el algoritmo contabilizó la frecuencia de píxeles de vegetación en cada clase de orientación, normalizando los conteos para obtener porcentajes relativos. Finalmente, estos valores se representaron gráficamente mediante diagramas polares tipo rosa de los vientos, permitiendo una comparación directa entre la distribución direccional observada y la simulada para el año 2024.

### 4.5.3 Interpretabilidad del Modelo (XAI)

Dado que las Redes Neuronales Artificiales, como el Perceptrón Multicapa, son consideradas tradicionalmente como “cajas negras” debido a la complejidad de sus pesos internos, se implementó un módulo de Inteligencia Artificial Explicable (XAI). El objetivo fue identificar los factores conductores (drivers) del cambio de uso de suelo y entender la relación funcional entre estos y la probabilidad de cada cobertura de estudio.

**Análisis de Dominancia de Variables.** Para identificar los predictores con mayor influencia en el desempeño del modelo, se aplicó el método de Importancia por Permutación (Permutation Feature Importance). Este enfoque permite evaluar la sensibilidad del modelo ante la perturbación de cada variable sin necesidad de reentrenar la red neuronal.

El procedimiento aplicado consistió en:

1. Calcular el AUC de referencia ( $AUC_{base}$ ) con el conjunto de validación original.
2. Permutar aleatoriamente los valores de un predictor  $X_j$ , rompiendo su relación con la variable objetivo.
3. Generar nuevas predicciones con el conjunto permutado y obtener el desempeño degradado ( $AUC_{perm}$ ).
4. Cuantificar la importancia de la variable como la disminución del AUC respecto al valor base.

**Curvas de Respuesta.** Para analizar el comportamiento no lineal de cada factor, se generaron Gráficos de Dependencia Parcial (PDP). Esta técnica aísla el efecto de una variable  $X_j$  sobre la probabilidad de salida, manteniendo constantes el resto de las variables en su valor promedio. Se simulaban píxeles sintéticos variando  $X_j$  a lo largo de su rango de distribución observado, permitiendo visualizar si la relación con las clases de estudio es positiva o negativa.

## 4.6 Proyección del Escenario Futuro 2034

Una vez validado el modelo y analizadas sus reglas internas para interpretar el funcionamiento del MLP, se procedió a la proyección del escenario futuro de LULC para

el año 2034. Este proceso utiliza el mismo modelo MLP entrenado (que contiene las reglas de idoneidad), y lo aplica al estado más reciente (2024) y utiliza la matriz de demanda de cambio más reciente (2014-2024).

#### 4.6.1 Fases Preparatorias

**Carga del Modelo Predictivo (MLP).** Se carga el modelo MLP  $f(\mathbf{x})$  entrenado y optimizado que se utilizó en la simulación de validación.

**Cuantificación de la Demanda de Cambio (Markov).** Para definir la magnitud de las transiciones esperadas entre 2024 y 2034, se calcula la matriz de demanda basada en el período histórico más reciente (2014 a 2024).

1. Se cargaron los mapas LULC clasificados para los años  $t_1 = 2014$  y  $t_2 = 2024$ .
2. Se calculó la matriz de tabulación cruzada ( $\mathbf{D}'$ ) y la matriz de probabilidad de transición ( $\mathbf{P}'$ , mostrada en la Tabla 23) para el período 2014 - 2024, siendo esta la “demanda de cambio” para la simulación 2024 - 2034.

#### 4.6.2 Proceso de Simulación Geoespacial 2024 - 2034

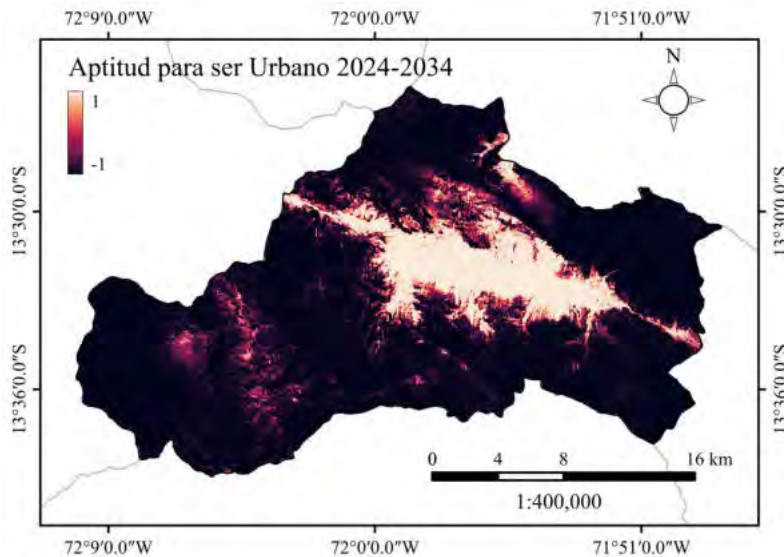
**Generación de Mapas de Potencial e Idoneidad para 2034.** Para la proyección al año 2034, se determina la aptitud espacial de cada píxel a partir del estado de cobertura del año 2024. El procedimiento es el siguiente:

1. Se emplea el mapa LULC clasificado  $L(t_2 = 2024)$  como estado inicial, junto con el conjunto correspondiente de factores impulsores  $\mathbf{V}(t_2 = 2024)$ .
2. Los factores impulsores  $\mathbf{V}(t_2)$  son normalizados.
3. El modelo MLP  $f(\mathbf{x})$  se aplica a cada píxel  $p$  del área de estudio. Esta operación genera un conjunto de Mapas de Potencial/Aptitud  $\mathbf{S}'_k$ , uno para cada clase  $k$ .
4. A partir de estos mapas de potencial, se derivan los Mapas de Idoneidad de Transición.

La Figura 48 presenta el mapa de potencial para la cobertura Urbana (2024-2034). La colección completa de mapas de aptitud (para Vegetación y Suelo Desnudo) y los mapas de idoneidad para todas las transiciones específicas se encuentran disponibles en el repositorio digital del proyecto (Apéndice E).

**Figura 48**

*Potencial para Ser de Cobertura Urbana (2024-2034).*



*Nota. Fuente: Elaboración propia.*

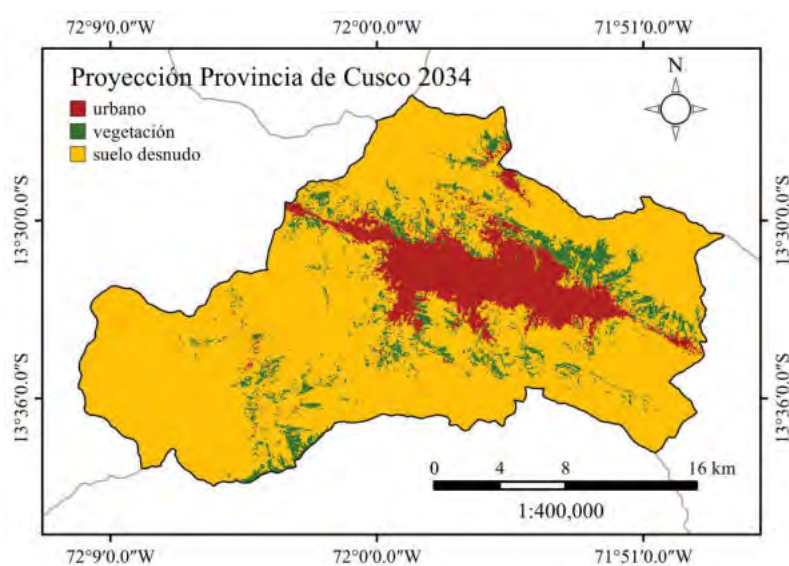
**Asignación Espacial de Transiciones (CA).** Utilizando los mapas de idoneidad  $S'$  y la matriz de demanda  $D'$ , el simulador CA asigna espacialmente las transiciones proyectadas.

1. Se itera a través de las transiciones  $i$  a  $j$ .
2. Para cada transición, se identifica la demanda  $D'_{ij}$  (de la matriz 2014-2024).
3. Se seleccionaron los  $D'_{ij}$  píxeles con los valores más altos en el mapa de idoneidad  $S'_j$ .
4. Se cambió de  $i$  a  $j$  el estado de estos píxeles seleccionados.

#### **4.6.3 Generación del Mapa LULC Proyectado para 2034**

Una vez completada la asignación espacial para todas las transiciones definidas por la demanda  $D'$ , el mapa resultante representa el escenario LULC proyectado para el año 2034. Este mapa final se presenta en la Figura 49.

**Figura 49**  
*Mapa LULC proyectado (2034).*



*Nota. Mapa de Coberturas y Usos del Suelo (LULC) proyectado para la Provincia de Cusco, año 2034, bajo un Escenario Tendencial. Fuente: Elaboración propia.*

## CAPÍTULO V

### Resultados y Discusión

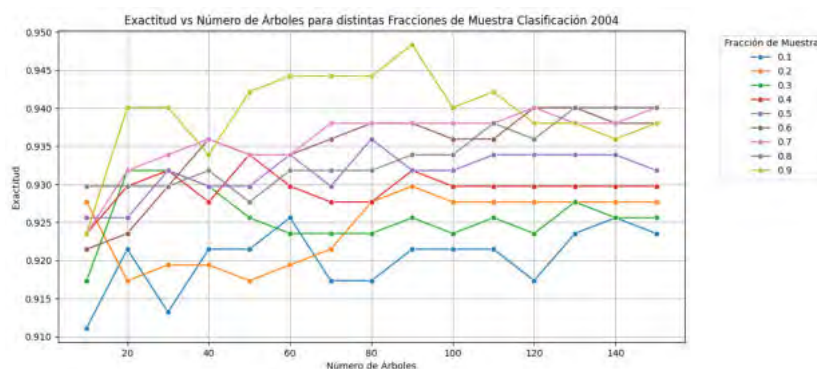
#### 5.1 Resultados

##### 5.1.1 Desempeño de la Clasificación Supervisada

**Configuración Óptima del Modelo de Clasificación.** Las Figuras 50 a 55 muestran los mapas de rendimiento obtenidos durante la optimización de hiperparámetros.

**Figura 50**

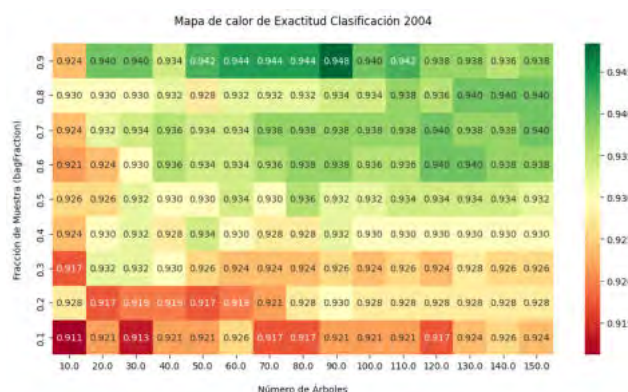
*Resultados de optimización de hiperparámetros (2004).*



*Nota. Precisión de validación (OA) en función del número de árboles para diferentes fracciones de muestreo para el año 2004. Fuente: Elaboración propia.*

**Figura 51**

*Mapa de calor de precisión de validación (OA) 2004.*

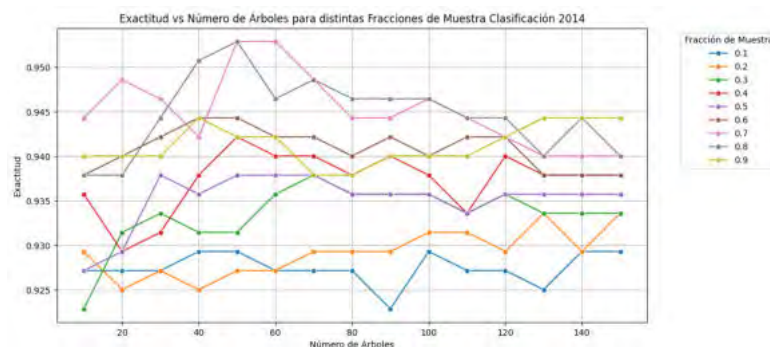


*Nota. Mapa de calor de precisión de validación (OA) para cada combinación de número de árboles y fracción de muestreo, resultado de la optimización de hiperparámetros para el año 2004. Fuente: Elaboración propia.*



**Figura 52**

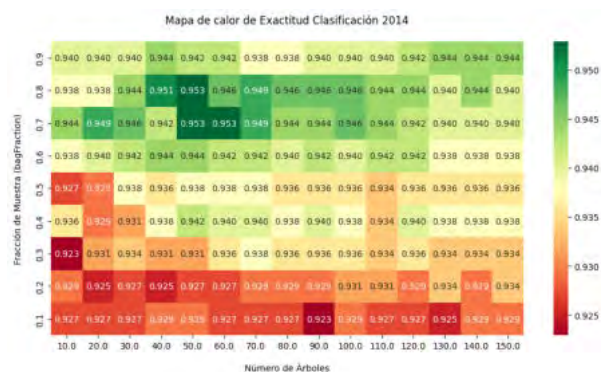
*Resultados de optimización de hiperparámetros (2014).*



*Nota. Precisión de validación (OA) en función del número de árboles para diferentes fracciones de muestreo para el año 2014. Fuente: Elaboración propia.*

**Figura 53**

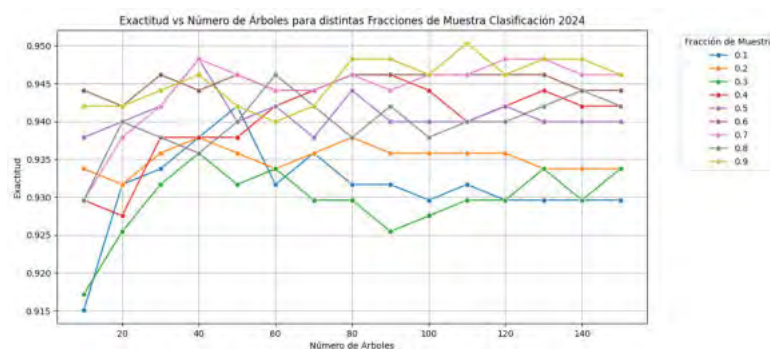
*Mapa de calor de precisión de validación (OA).*



*Nota. Mapa de calor de precisión de validación (OA) para cada combinación de número de árboles y fracción de muestreo, resultado de la optimización de hiperparámetros para el año 2014. Fuente: Elaboración propia.*

**Figura 54**

*Resultados de optimización de hiperparámetros (2024).*



*Nota. Precisión de validación (OA) en función del número de árboles para diferentes fracciones de muestreo para el año 2024. Fuente: Elaboración propia.*

**Figura 55**

*Mapa de calor de precisión de validación (OA).*



*Nota. Mapa de calor de precisión de validación (OA) para cada combinación de número de árboles y fracción de muestreo, resultado de la optimización de hiperparámetros para el año 2024. Fuente: Elaboración propia.*

Los hiperparámetros seleccionados para cada periodo y la precisión global (OA) resultante se resumen en la Tabla 13.

**Tabla 13**

*Precisión global e hiperparámetros óptimos*

Año	Precisión Global (OA)	Nº Óptimo de Árboles	Fracción de Muestreo Óptima
2004	0.948	90	0.9
2014	0.953	50	0.7
2024	0.950	110	0.9

*Nota. Fuente: Elaboración propia.*

### Métricas de Precisión de la Clasificación.

**Precisión de Clasificación – Año 2004 (Landsat 5 TM).** El modelo de clasificación para el año 2004 alcanzó una precisión global (OA) del 94.83 % y un coeficiente Kappa de 0.922 (Tablas 14 y 15).

**Precisión de Clasificación – Año 2014 (Landsat 8 OLI).** La clasificación de 2014 reportó una precisión global de 95.29 % y un coeficiente Kappa de 0.929 (Tablas 16 y 17).

**Precisión de Clasificación – Año 2024 (Sentinel-2 MSI).** En el periodo 2024, la precisión global fue de 95.03 % con un coeficiente Kappa de 0.925 (Tablas 18 y 19).

Las Figuras 56 a 61 muestran la clasificación LULC. Los datos .tif e imágenes complementarios están disponibles en el repositorio del proyecto (Apéndice E).

**Tabla 14***Matriz de confusión de la clasificación LULC 2004.*

Clase de Referencia	Clase Clasificada (Mapa)			Total Ref.
	Urbano	Vegetación	Suelo	
Urbano	163	0	13	176
Vegetación	0	145	1	146
Suelo	6	5	151	162
Total Clasificado	169	150	165	484

*Nota. La matriz muestra la validación de la clasificación supervisada para el año 2004.**Fuente: Elaboración propia.***Tabla 15***Métricas de precisión de la clasificación LULC 2004.*

Clase LULC	Exactitud del Productor	Exactitud del Usuario	F1-Score
1. Urbano	92.61 %	96.45 %	0.9449
2. Vegetación	99.32 %	96.67 %	0.9797
3. Suelo	93.21 %	91.52 %	0.9235
Métricas Globales			
Exactitud Global ( <i>Overall Accuracy</i> )			94.83 %
Acuerdo esperado por azar ( $P_e$ )			0.3346
Coeficiente Kappa Estándar ( $K_{standard}$ )			0.9224
Kappa de Histograma ( $K_{histo}$ )			0.9783

*Nota. Resumen de las métricas de precisión por clase y globales obtenidas del proceso de clasificación. Fuente: Elaboración propia.***Tabla 16***Matriz de confusión de la clasificación LULC 2014.*

Clase de Referencia	Clase Clasificada (Mapa)			Total Ref.
	Urbano	Vegetación	Suelo	
Urbano	119	0	15	134
Vegetación	0	162	2	164
Suelo	4	1	164	169
Total Clasificado	123	163	181	467

*Nota. La matriz muestra la validación de la clasificación supervisada para el año 2014.**Fuente: Elaboración propia.*

**Tabla 17***Métricas de precisión de la clasificación LULC 2014.*

Clase LULC	Exactitud del Productor	Exactitud del Usuario	F1-Score
1. Urbano	88.81 %	96.75 %	0.9261
2. Vegetación	98.78 %	99.39 %	0.9908
3. Suelo	97.04 %	90.61 %	0.9371
Métricas Globales			
Exactitud Global ( <i>Overall Accuracy</i> )			95.29 %
Acuerdo esperado por azar ( $P_e$ )			0.3384
Coeficiente Kappa Estándar ( $K_{standard}$ )			0.9288
Kappa de Histograma ( $K_{histo}$ )			0.9612

*Nota. Resumen de las métricas de precisión por clase y globales obtenidas del proceso de clasificación. Fuente: Elaboración propia.*

**Tabla 18***Matriz de confusión de la clasificación LULC 2024.*

Clase de Referencia	Clase Clasificada (Mapa)			Total Ref.
	Urbano	Vegetación	Suelo	
Urbano	134	0	16	150
Vegetación	0	168	0	168
Suelo	8	0	157	165
Total Clasificado	142	168	173	483

*Nota. La matriz muestra la validación de la clasificación supervisada para el año 2024. Fuente: Elaboración propia.*

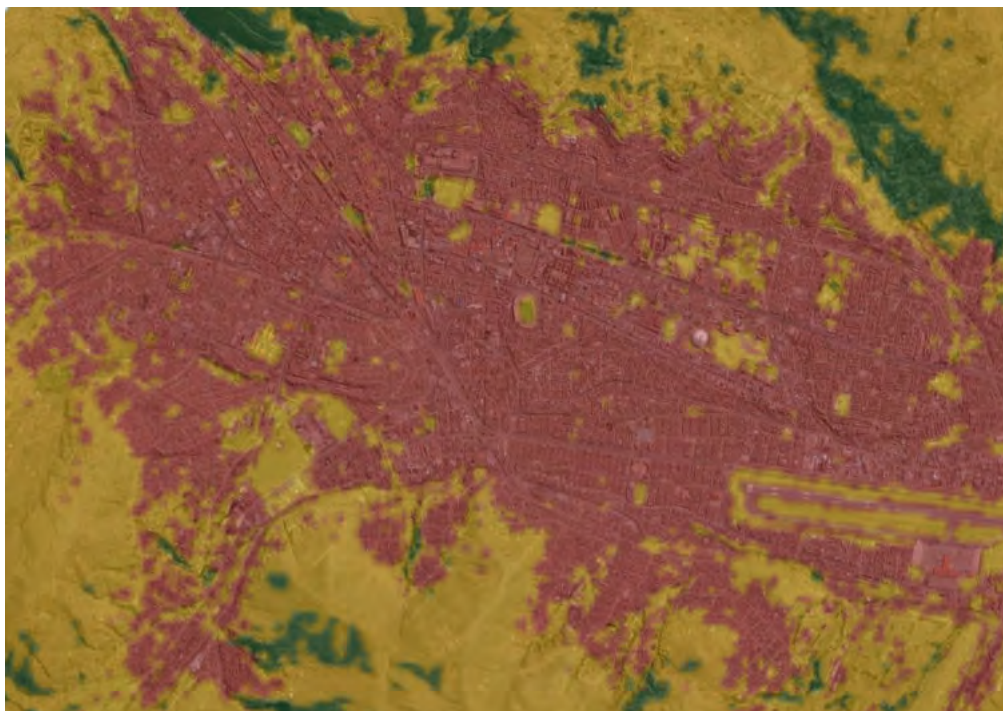
**Tabla 19***Métricas de precisión de la clasificación LULC 2024.*

Clase LULC	Exactitud del Productor	Exactitud del Usuario	F1-Score
1. Urbano	89.33 %	94.37 %	0.9178
2. Vegetación	100.00 %	100.00 %	1.0000
3. Suelo	95.15 %	90.75 %	0.9290
Métricas Globales			
Exactitud Global ( <i>Overall Accuracy</i> )			95.03 %
Acuerdo esperado por azar ( $P_e$ )			0.3346
Coeficiente Kappa Estándar ( $K_{standard}$ )			0.9253
Kappa de Histograma ( $K_{histo}$ )			0.9751

*Nota. Resumen de las métricas de precisión por clase y globales obtenidas del proceso de clasificación. Fuente: Elaboración propia.*

**Figura 56**

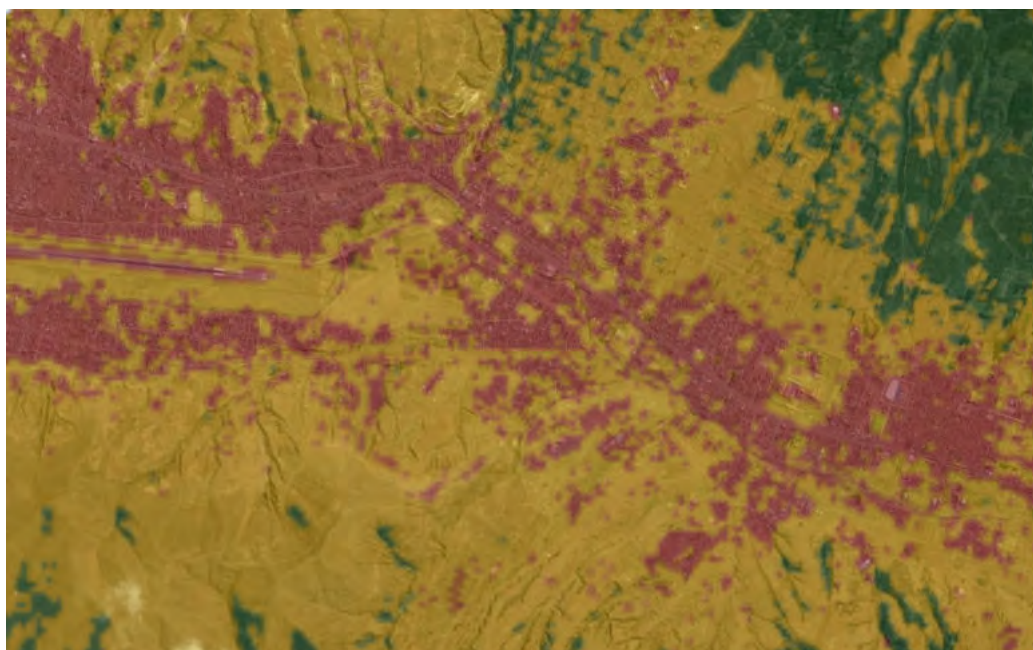
*Mapa LULC clasificado (Sector Oeste Cusco 2004).*



*Nota. Superposición del mapa LULC clasificado del año 2004 sobre la imagen de referencia de alta resolución (Mayo 2002). Fuente: Elaboración propia.*

**Figura 57**

*Mapa LULC clasificado (Sector Este Cusco 2004).*



*Nota. Superposición del mapa LULC clasificado del año 2004 sobre la imagen de referencia de alta resolución (Mayo 2002). Fuente: Elaboración propia.*



**Figura 58**

*Mapa LULC clasificado (Sector Oeste Cusco 2014).*



*Nota. Superposición del mapa LULC clasificado del año 2014 sobre la imagen de referencia de alta resolución. Fuente: Elaboración propia.*

**Figura 59**

*Mapa LULC clasificado (Sector Este Cusco 2014).*



*Nota. Superposición del mapa LULC clasificado del año 2014 sobre la imagen de referencia de alta resolución. Fuente: Elaboración propia.*



**Figura 60**

*Mapa LULC clasificado (Sector Oeste Cusco 2024).*



*Nota. Superposición del mapa LULC clasificado del año 2024 sobre la imagen de referencia de alta resolución. Fuente: Elaboración propia.*

**Figura 61**

*Mapa LULC clasificado (Sector Este Cusco 2024).*

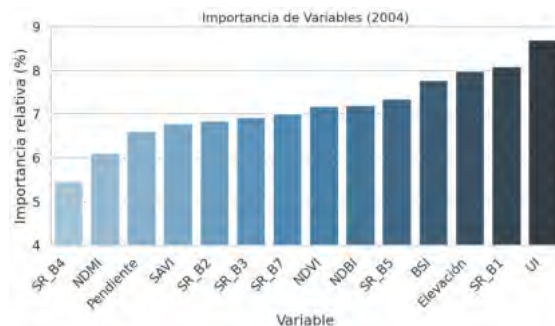


*Nota. Superposición del mapa LULC clasificado del año 2024 sobre la imagen de referencia de alta resolución. Fuente: Elaboración propia.*

**Importancia Relativa de las Variables Predictoras.** La importancia relativa de las variables predictoras para los tres periodos se presenta en las Figuras 62, 63 y 64.

**Figura 62**

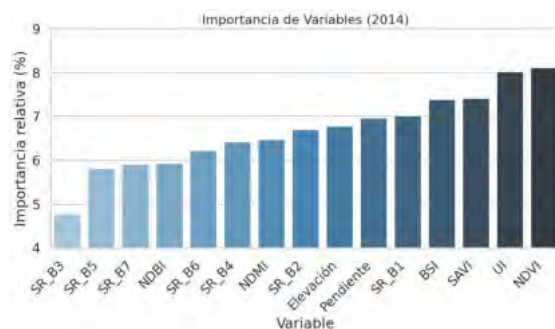
*Importancia relativa de variables predictoras (2004).*



*Nota. Fuente: Elaboración propia.*

**Figura 63**

*Importancia relativa de variables predictoras (2014).*



*Nota. Fuente: Elaboración propia.*

**Figura 64**

*Importancia relativa de variables predictoras (2024).*



*Nota. Fuente: Elaboración propia.*



### 5.1.2 Cambios Espaciales y Temporales del LULC (2004-2024)

**Evolución de Superficies y Cambios Netos de Cobertura.** La distribución de las coberturas de suelo estudiadas en la provincia de Cusco para los años 2004, 2014 y 2024 se presenta en las Tablas 20, 21 y la Figura 65.

**Tabla 20**

*Cambios netos de área por clase LULC durante el período 2004 - 2014.*

Clase	2004 (km <sup>2</sup> )	2014 (km <sup>2</sup> )	$\Delta$ (km <sup>2</sup> )	2004 (%)	2014 (%)	$\Delta$ (%)
Urbano	21.13	33.61	+12.48	3.99	6.34	+2.35
Vegetación	70.62	62.80	-7.82	13.32	11.85	-1.47
Suelo Desnudo	438.39	433.73	-4.66	82.69	81.81	-0.88

*Nota. Fuente: Elaboración propia.*

**Tabla 21**

*Cambios netos de área por clase LULC durante el período 2014 - 2024.*

Clase	2014 (km <sup>2</sup> )	2024 (km <sup>2</sup> )	$\Delta$ (km <sup>2</sup> )	2014 (%)	2024 (%)	$\Delta$ (%)
Urbano	33.61	50.28	+16.67	6.34	9.48	+3.14
Vegetación	62.80	51.96	-10.85	11.85	9.80	-2.05
Suelo Desnudo	433.73	427.91	-5.82	81.81	80.72	-1.10

*Nota. Fuente: Elaboración propia.*

**Figura 65**

*Tendencia de evolución de coberturas.*



*Nota. La figura muestra la tendencia de la evolución de las coberturas de suelo del 2004 hasta el año 2024. Fuente: Elaboración propia.*

**Matrices de Transición y Probabilidades de Markov.** Las matrices de probabilidad de transición (derivadas del análisis de Markov, Tablas 22, 23 y Figura 66) contienen las tasas de estabilidad y conversión bruta entre clases.

**Tabla 22**

*Matriz de probabilidad de transición estimada para el período 2004 - 2014.*

De	A		
	Urbano	Vegetación	Suelo Desnudo
Urbano	0.8645	0.0017	0.1337
Vegetación	0.0179	0.6931	0.2890
Suelo Desnudo	0.0321	0.0315	0.9364

*Nota. Fuente: Elaboración propia.*

**Tabla 23**

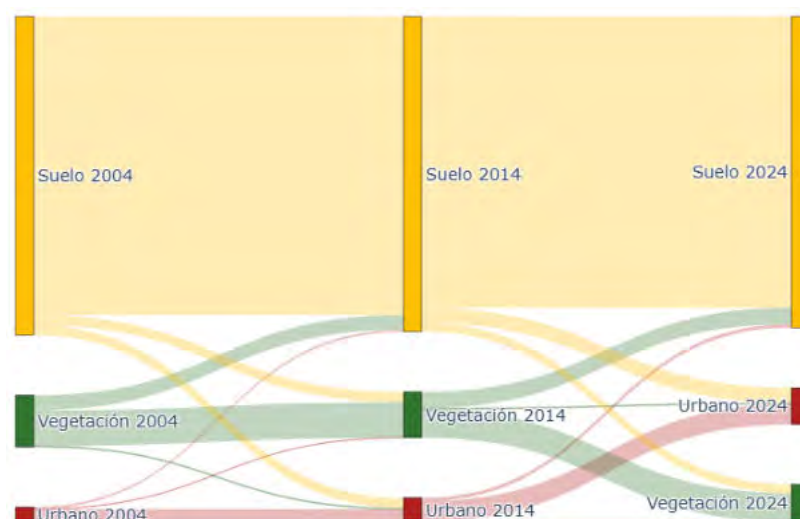
*Matriz de probabilidad de transición estimada para el período 2014 - 2024.*

De	A		
	Urbano	Vegetación	Suelo Desnudo
Urbano	0.8278	0.0272	0.1451
Vegetación	0.0219	0.6223	0.3558
Suelo Desnudo	0.0486	0.0276	0.9238

*Nota. Fuente: Elaboración propia.*

**Figura 66**

*Diagrama de Sankey de la evolución de coberturas del suelo.*

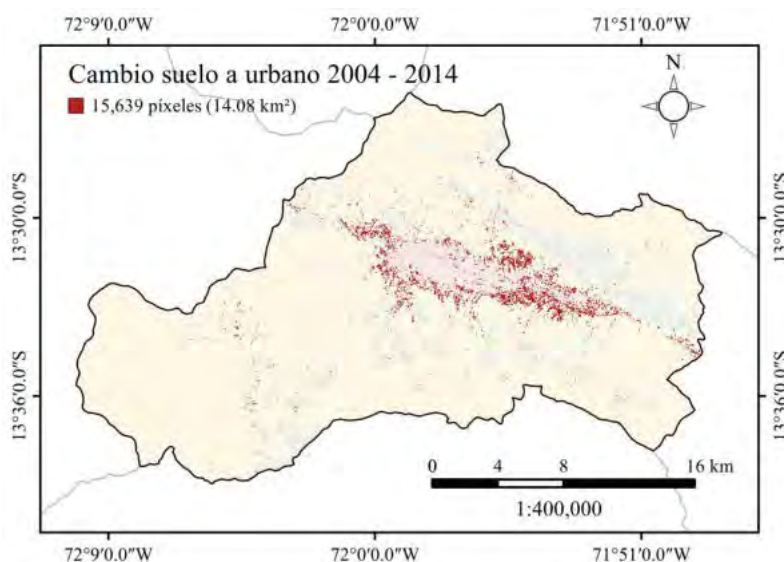


*Nota. El diagrama de Sankey ilustra la evolución de las coberturas del suelo (LULC) en el período estudiado. Fuente: Elaboración propia.*

**Distribución Espacial de los Cambios de Cobertura.** Las Figuras 67 y 68 presentan un ejemplo de la transición de Suelo Desnudo a Urbano para los periodos 2004-2014 y 2014-2024, respectivamente. El catálogo completo de mapas de cambio para todas las transiciones y clases se encuentra organizado en el repositorio digital del proyecto (Apéndice E).

**Figura 67**

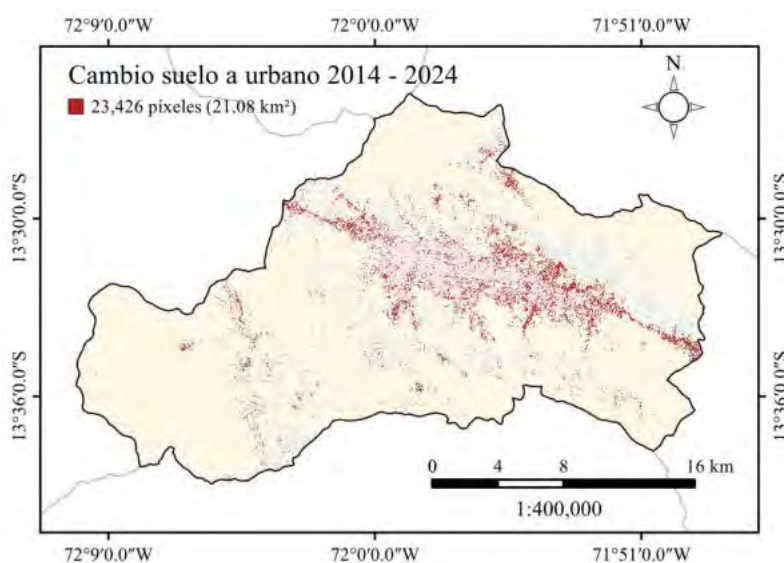
*Cambio de cobertura de suelo desnudo a urbana (2004-2014).*



*Nota. Mapa de cambio de la cobertura de suelo desnudo a urbana entre los años 2004 y 2014.  
Fuente: Elaboración propia.*

**Figura 68**

*Cambio de cobertura de suelo desnudo a urbana (2014-2024).*



*Nota. Mapa de cambio de la cobertura de suelo desnudo a urbana entre los años 2014 y 2024.  
Fuente: Elaboración propia.*

### 5.1.3 Desempeño y Validación del Modelo Predictivo MLP-CA-MC

**Colinealidad entre Variables Predictoras.** Las matrices de correlación de Pearson para los años 2004, 2014 y 2024 se presentan en las Figuras 69, 70 y 71.

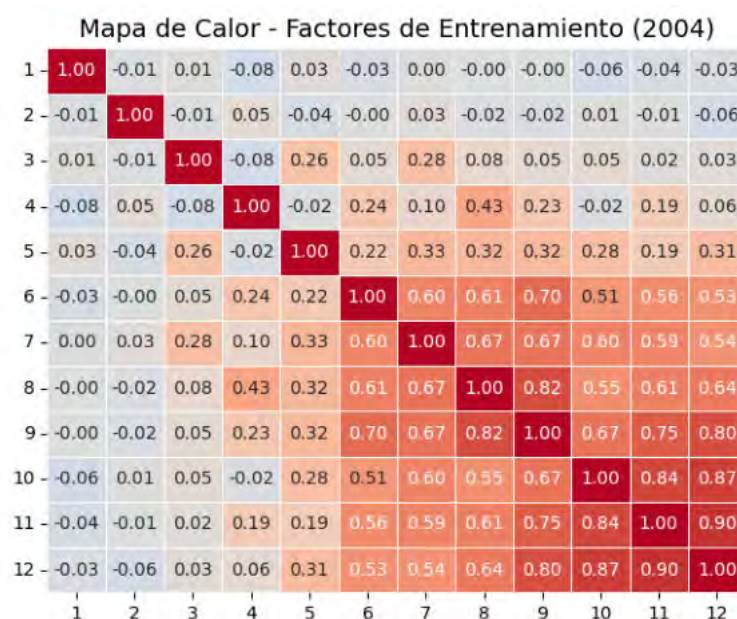
Las variables de distancia a infraestructura (*vías*, *aeropuerto* y *rieles*) se situaron en el rango de 0.84 a 0.90 durante el periodo analizado. En el bloque de equipamiento urbano, las variables de *educación* y *salud* presentaron valores de 0.82, 0.81 y 0.84 para los años 2004, 2014 y 2024, respectivamente.

En el intervalo comprendido entre 2004 y 2024, el coeficiente entre la *distancia a instituciones educativas* y la *distancia a vías* pasó de 0.64 a 0.73 y la relación entre la variable *educación* y la *distancia al aeropuerto* varió de 0.61 a 0.74.

Finalmente, los factores topográficos de *aspecto* y *pendiente* mantuvieron coeficientes inferiores a 0.1 en relación con el resto de los predictores en toda la serie temporal.

**Figura 69**

*Matriz de correlación de variables predictoras (2004).*

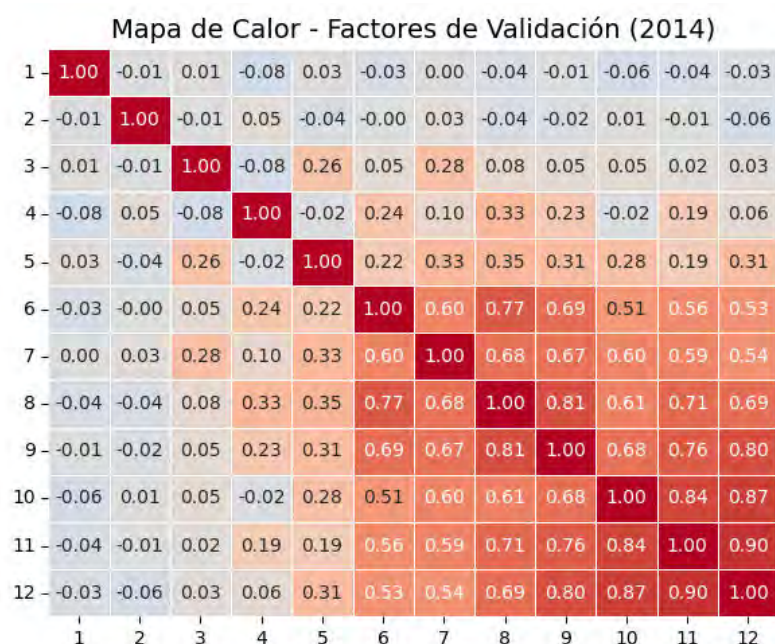


*Nota. La figura muestra la matriz de correlación entre las variables predictoras consideradas para el período 2004. Fuente: Elaboración propia.*



**Figura 70**

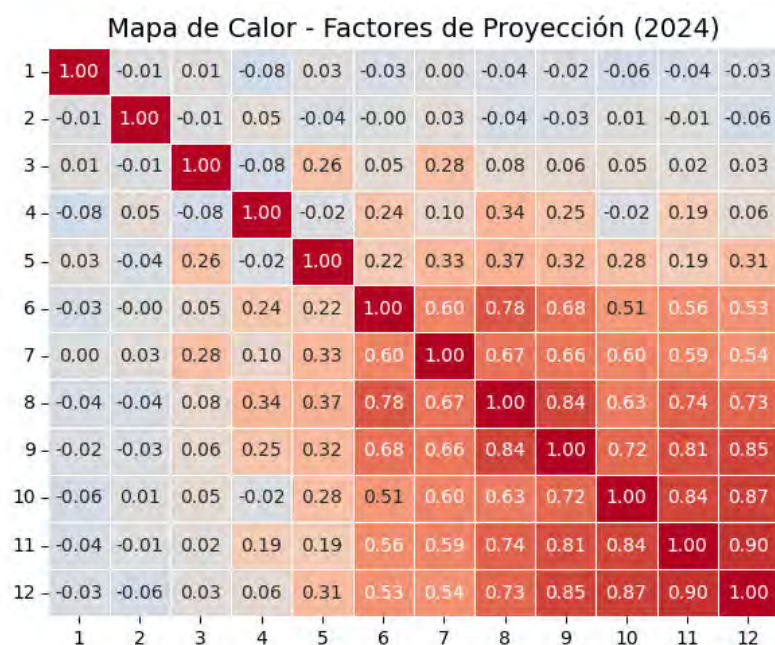
*Matriz de correlación de variables predictoras (2014).*



*Nota. La figura muestra la matriz de correlación entre las variables predictoras consideradas para el período 2014. Fuente: Elaboración propia.*

**Figura 71**

*Matriz de correlación de variables predictoras (2024).*



*Nota. La figura muestra la matriz de correlación entre las variables predictoras consideradas para el período 2024. Fuente: Elaboración propia.*

### Métricas de Validación del Escenario Simulado 2024.

**Exactitud Categórica.** La validación del modelo MLP-CA-MC para el año 2024 se efectuó mediante la comparación entre el mapa simulado y clasificado. Se registró una Exactitud Global ( $OA$ ) de 88.90 % y un coeficiente  $K_{standard}$  de 0.6620. Respecto al F1-Score por categoría, se obtuvieron valores de 0.9340 para la clase Suelo, 0.7160 para Urbano y 0.6841 para Vegetación (Tabla 25). En cuanto a los componentes del acuerdo, se cuantificó un valor de  $P_{quantity}$  de 0.9921 y un  $K_{histo}$  de 0.9759. El acuerdo esperado por azar ( $P_e$ ) fue de 0.6717.

**Tabla 24**

*Matriz de confusión LULC 2024.*

Clase Real (Referencia)	Clase Simulada (Predicción)			Total Real
	Urbano	Vegetación	Suelo	
Urbano	38,331	1,224	16,308	55,863
Vegetación	1,354	40,644	15,734	57,732
Suelo	11,522	19,229	444,703	475,454
Total Simulado	51,207	61,097	476,745	589,049

*Nota. La matriz muestra la comparación entre clases simuladas y reales para el año 2024.*

*Fuente: Elaboración propia.*

**Tabla 25**

*Métricas de precisión LULC 2024.*

Clase LULC	Exactitud del Productor	Exactitud del Usuario	F1-Score
1. Urbano	68.62 %	74.85 %	0.7160
2. Vegetación	70.40 %	66.52 %	0.6841
3. Suelo	93.53 %	93.28 %	0.9340
Métricas Globales			
Exactitud Global ( <i>Overall Accuracy</i> )			88.90 %
Acuerdo esperado por azar ( $P_e$ )			0.6717
Coeficiente Kappa Estándar ( $K_{standard}$ )			0.6620
Kappa de Histograma ( $K_{histo}$ )			0.9759

*Nota. La Tabla resume las métricas de precisión por clase y las métricas globales del modelo de simulación. Fuente: Elaboración propia.*

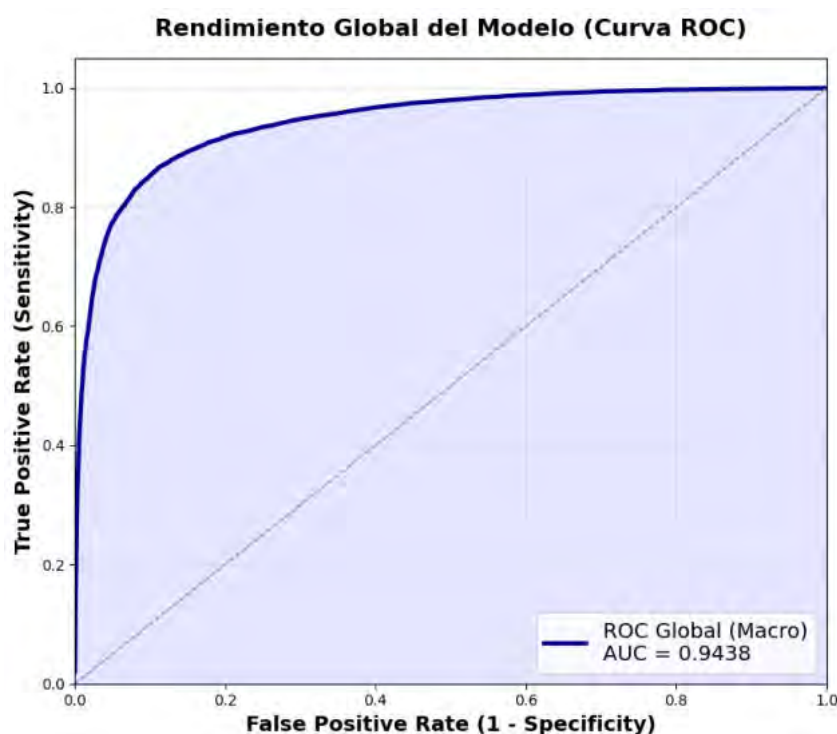
La matriz de confusión para el año 2024 clasificado y la simulación muestra un intercambio entre las coberturas naturales (*Suelo Desnudo* y *Vegetación*) de 34 963 píxeles de diferencia. En este bloque, se registraron 19 229 píxeles de *Suelo Desnudo* asignados a la clase *Vegetación* y 15 734 píxeles de *Vegetación* asignados a *Suelo Desnudo*.

Respecto a la categoría *Urbano*, se cuantificó una omisión de 16 308 píxeles ( $14.7 \text{ km}^2$ ), los cuales figuran en la simulación como *Suelo Desnudo*. Asimismo, la comisión en esta clase fue de 11 522 píxeles, correspondientes a áreas de *Suelo Desnudo* en el mapa de referencia que se encuentran integradas a la clase *Urbano* en el mapa simulado.

**Capacidad Predictiva del Potencial de Transición (AUC-ROC).** La evaluación de las probabilidades de transición mediante el análisis ROC multiclase resultó en un valor de área bajo la curva (AUC) de 0.9438. El trazado de la Figura 72 se ubica hacia el sector superior izquierdo del espacio de la gráfica.

**Figura 72**

*Curva ROC Global (Macro-Average) del modelo MLP-CA-MC.*



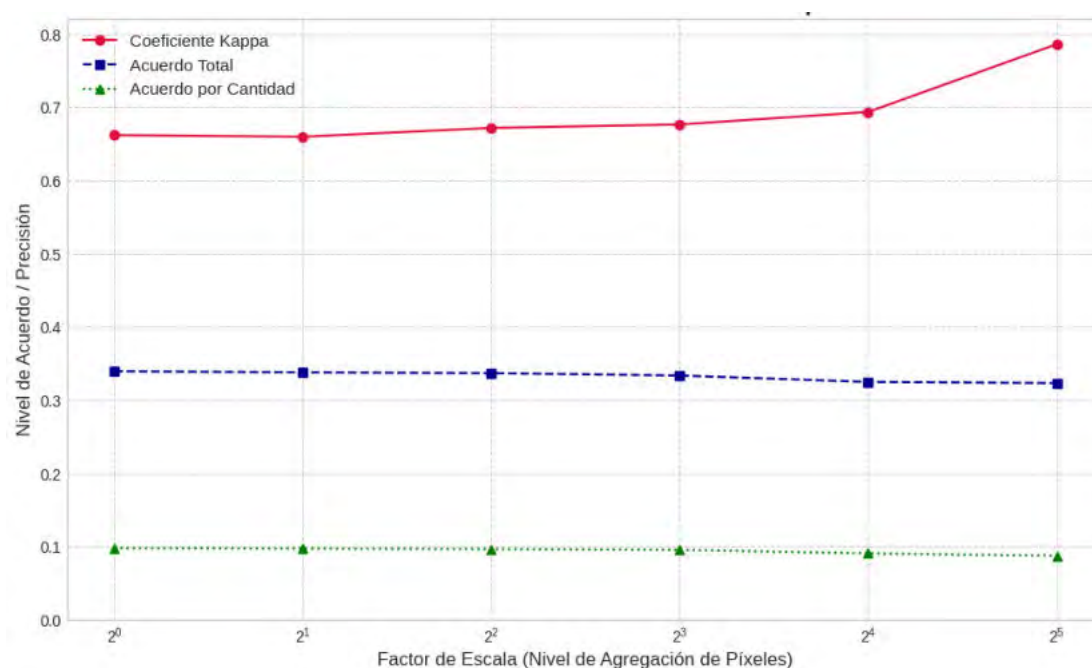
*Nota. El desempeño del modelo (línea azul) se compara con un clasificador aleatorio (línea punteada gris). El AUC resultante es de 0.9438. Fuente: Elaboración propia.*

**Precisión Posicional y Multiescala.** La precisión global con una ventana de tolerancia de  $3 \times 3$  píxeles fue de 97.05 %, frente al 88.90 % obtenido en la escala original. El coeficiente  $K_{standard}$  registró valores de 0.6620 en la resolución de 30m y de 0.7861 en el nivel de agregación de 960m (Tabla 26). El desacuerdo por cantidad se mantuvo en un valor de 0.09 en todas las escalas evaluadas. Por su parte, el desacuerdo por localización varió desde 0.3394 en la resolución nativa hasta 0.3233 en la escala de 960m.

**Tabla 26***Validación multiescala LULC 2024.*

Iteración	Escala (Agregación)	Kappa ( $K_{standard}$ )	Desacuerdo por Cantidad	Desacuerdo por Localización
0	1x (30m)	0.6620	0.0979	0.3394
1	2x (60m)	0.6597	0.0973	0.3380
2	4x (120m)	0.6718	0.0966	0.3369
3	8x (240m)	0.6765	0.0955	0.3336
4	16x (480m)	0.6935	0.0907	0.3249
5	32x (960m)	0.7861	0.0874	0.3233

*Nota. La Tabla muestra los resultados de la validación multiescala comparando mapas simulados y reales de 2024 a distintos niveles de agregación espacial. Fuente: Elaboración propia.*

**Figura 73***Validación multiescala.*

*Nota. Fuente: Elaboración propia.*

Las Figuras 74 y 75 presentan las vistas comparativas de la simulación LULC 2024 versus la realidad, para los sectores Oeste y Este, respectivamente.



**Figura 74**

*Mapa LULC Simulado 2024 (Oeste) superpuesto en Alta Resolución.*



*Nota. Fuente: Elaboración propia.*

**Figura 75**

*Mapa LULC Simulado 2024 (Este) superpuesto en Alta Resolución.*

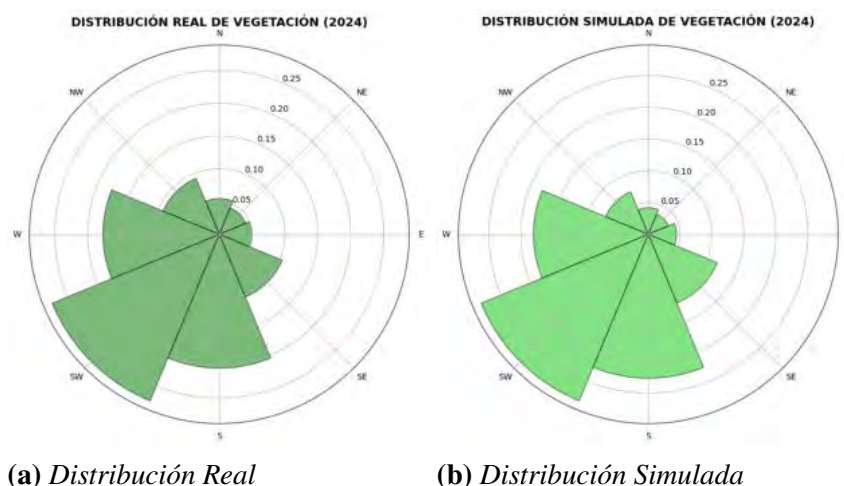


*Nota. Fuente: Elaboración propia.*

**Patrones Espaciales de la Clase Vegetación.** La distribución de la vegetación por orientación de ladera para el año 2024 se detalla en la Tabla 27 y Figura 76. En la orientación Suroeste (SW) se registró un valor de 27.50 % en el mapa real y de 28.37 % en la simulación. Las diferencias porcentuales entre ambos mapas se situaron en un rango de -2.01 % (NW) a +2.23 % (S).

**Figura 76**

*Diagramas de frecuencia direccional de la cobertura vegetal (2024).*



*Nota. Los diagramas de Rosa de los Vientos muestran la orientación predominante de la cobertura vegetal observada (Real) frente a la simulada por el modelo (Simulada) para el año 2024. Fuente: Elaboración propia.*

**Tabla 27**

*Vegetación por orientación de ladera (2024).*

Orientación	% Real (2024)	% Simulado (2024)	Diferencia
Norte (N)	5.45 %	4.17 %	-1.28 %
Noreste (NE)	4.35 %	3.43 %	-0.92 %
Este (E)	5.00 %	4.44 %	-0.56 %
Sureste (SE)	10.34 %	11.70 %	+1.36 %
Sur (S)	20.42 %	22.65 %	+2.23 %
Suroeste (SW)	27.50 %	28.37 %	+0.87 %
Oeste (W)	17.70 %	18.01 %	+0.31 %
Noroeste (NW)	9.24 %	7.23 %	-2.01 %

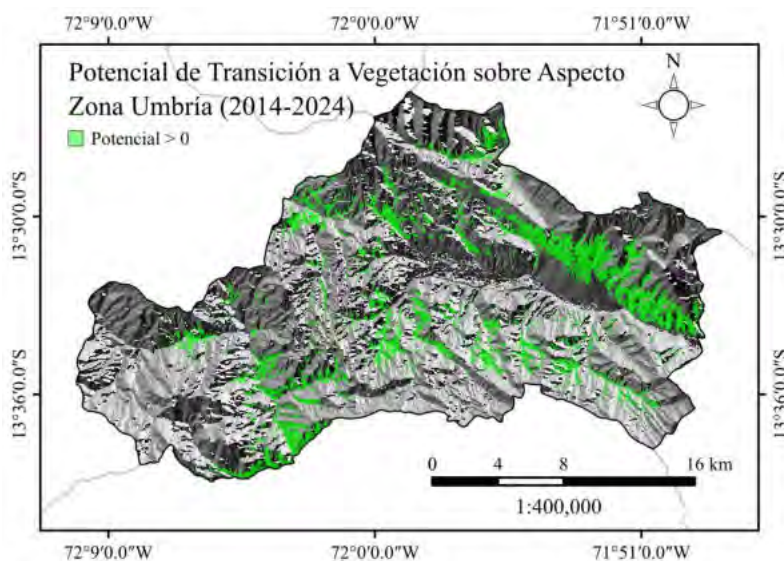
*Nota. La Tabla muestra la comparación porcentual de la vegetación según orientación de ladera en 2024. Fuente: Elaboración propia.*

**Potencial de Cobertura Vegetación.** Mapa de potencial de ser de cobertura Vegetación generados por el modelo híbrido (Figura 77).



**Figura 77**

*Mapa de Potencial de Vegetación (2014-2024).*



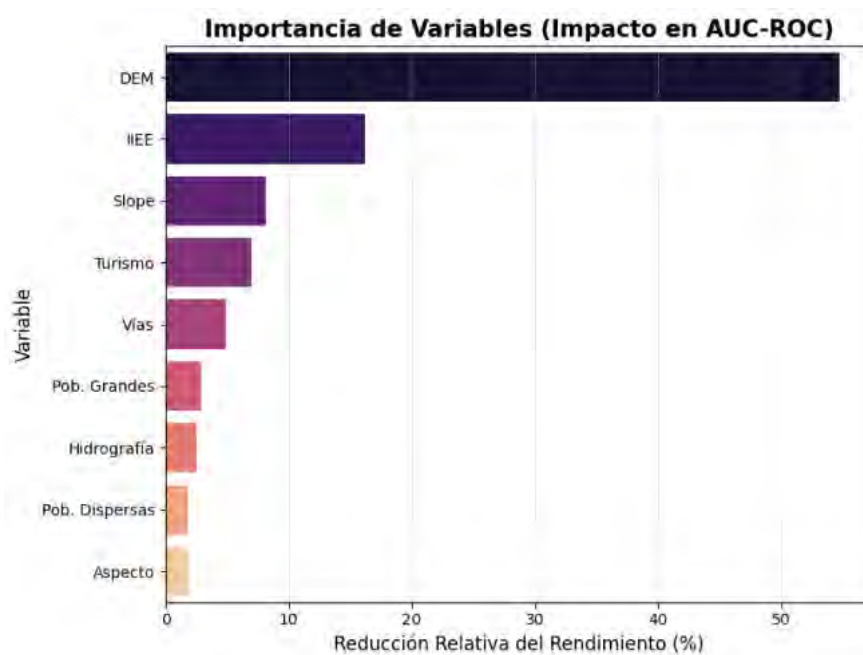
*Nota. Potencial de Vegetación para el período de validación (2014-2024). Fuente: Elaboración propia.*

### **Factores Impulsores del Cambio (XAI).**

**Contribución Relativa de los Predictores.** Los valores obtenidos en el análisis de importancia por permutación se detallan en las Figuras 78, 79 y Tabla 28.

**Figura 78**

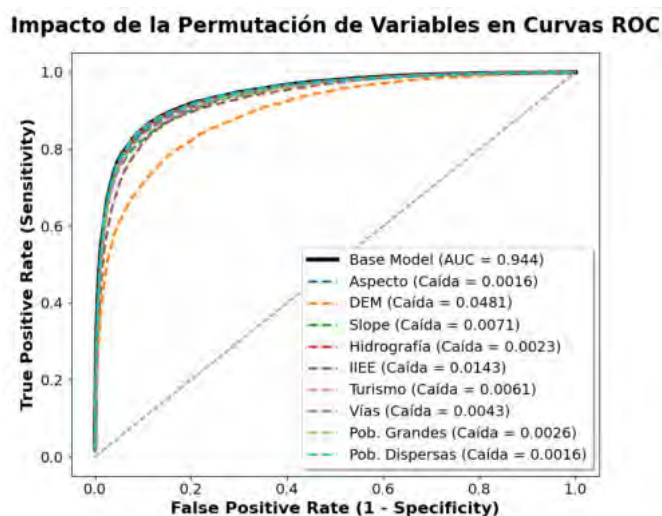
*Ranking de importancia de variables basado en el impacto en el AUC.*



*Nota. Importancia de las variables predictoras. Fuente: Elaboración propia.*

**Figura 79**

*Impacto de la permutación de variables en Curvas ROC.*



*Nota. Se ilustra el impacto de la permutación de las variables predictoras en las Curvas ROC generadas por el modelo. Fuente: Elaboración propia.*

**Tabla 28**

*Ranking de Importancia (Permutación)*

Variable Predictora	AUC Permutado	Caída ( $\Delta$ AUC)	Imp. Relativa (%)
1. Elevación (DEM)	0.8957	0.0481	54.70
2. Inst. Educativas (IIEE)	0.9296	0.0143	16.21
3. Pendiente (Slope)	0.9367	0.0071	8.12
4. Turismo	0.9377	0.0061	6.97
5. Vías	0.9396	0.0043	4.84
6. Pob. Grandes	0.9413	0.0026	2.93
7. Hidrografía	0.9416	0.0023	2.56
8. Pob. Dispersas	0.9422	0.0016	1.84
9. Aspecto	0.9422	0.0016	1.82
AUC Base del Modelo:			0.9438

*Nota. Fuente: Elaboración propia.*

La aleatorización de la variable Elevación (*DEM*) resultó en un *AUC* permutado de 0.8957, lo que representa una variación de  $\Delta = 0.0481$  respecto al valor base y una importancia relativa del 54.70 %. Para la variable *Instituciones Educativas*, se registró un *AUC* de 0.9296 ( $\Delta = 0.0143$ ) y un peso del 16.21 %. En el caso de las variables *Pendiente*, *Turismo* y *Vías*, los valores se situaron en 0.0071, 0.0061 y 0.0043, respectivamente. Los factores restantes (*Poblaciones*, *Hidrografía* y *Aspecto*) presentaron variaciones en el *AUC* inferiores a 0.003, con valores de importancia relativa entre el 1.82 % y el 2.93 %.

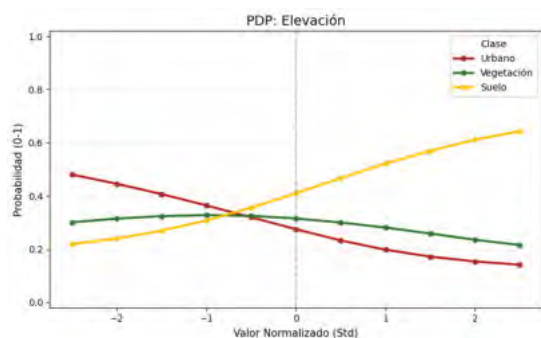
**Dinámica de Respuesta Espacial (PDP).** Caracteriza la relación marginal entre los predictores y la probabilidad estimada de transición (Figura 80). Las tablas detalladas se encuentran disponibles en el repositorio digital del estudio (Apéndice E).

- Elevación (DEM). Los valores de probabilidad urbana se sitúan entre 0.4801 y 0.1412, mientras que el Suelo Desnudo oscila entre 0.2189 y 0.6433.
- Pendiente (Slope). La probabilidad urbana varía de 0.4187 a 0.2287.
- Aspecto e Hidrografía. El rango de respuesta para el aspecto se ubica entre 0.3801 y 0.2960. El intervalo de probabilidad urbana se mantiene entre 0.3389 y 0.3278.
- Instituciones Educativas (IIEE). Se registraron valores de 0.4189 y 0.2159 para la clase urbana, con un cambio en la probabilidad de Suelo hasta 0.5133.
- Turismo. La probabilidad de transición urbana se sitúa en 0.3996 en el origen y en 0.2294 en la distancia máxima evaluada.
- Vías. El vector de salida para la clase urbana varía entre 0.3666 y 0.2783, mientras que la Vegetación se ubica entre 0.2928 y 0.3990.
- Poblaciones Grandes. La probabilidad urbana registra un cambio de 0.3286 a 0.2987.
- Poblaciones Dispersas. Se obtuvo un rango de 0.2865 a 0.3317 para la categoría urbana.

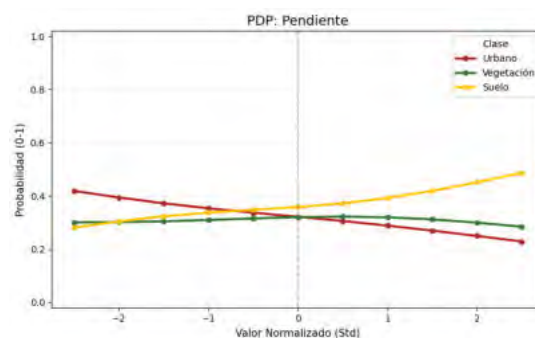
**Figura 80**

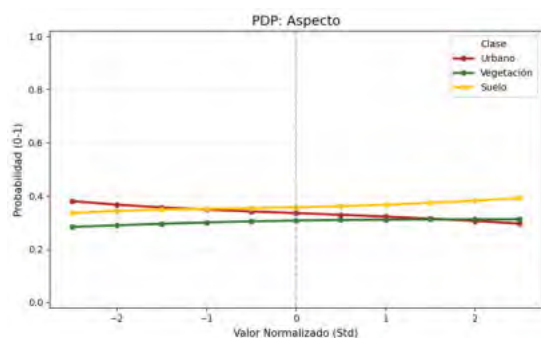
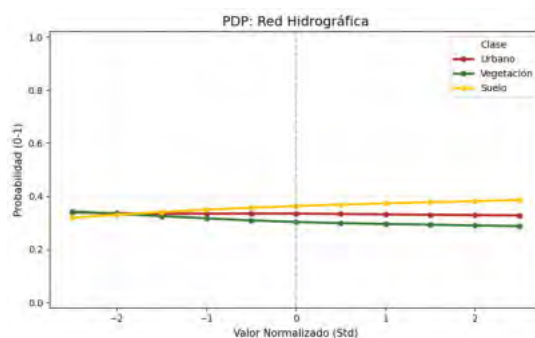
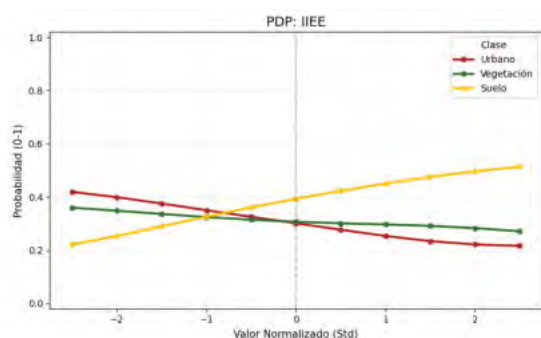
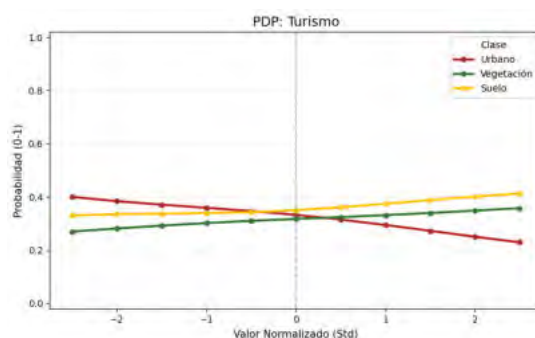
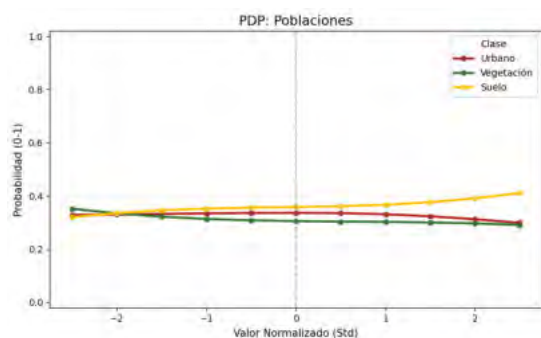
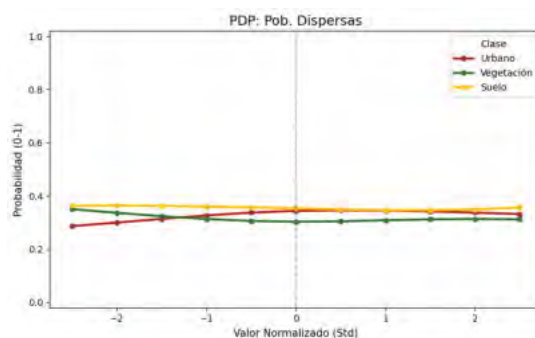
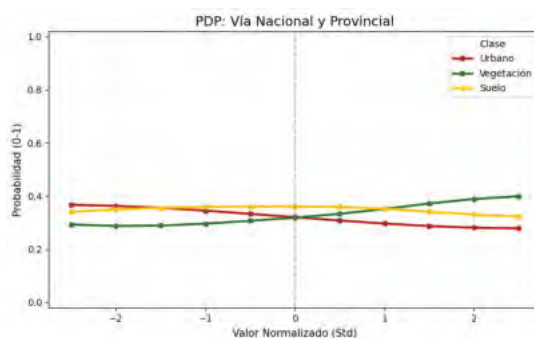
*Curvas de respuesta PDP*

**(a) Elevación (DEM).**



**(b) Pendiente.**



**Figura 80***Curvas de respuesta PDP (continuación)***(c) Aspecto.****(d) Red Hidrográfica.****(e) Instituciones Educativas.****(f) Turismo.****(g) Poblaciones Grandes.****(h) Poblaciones Dispersas.****(i) Red Vial.**

*Nota. Las curvas de dependencia parcial (PDP) muestran la influencia marginal de cada variable predictora en el modelo. Fuente: Elaboración propia.*

#### 5.1.4 Proyección del LULC para el Periodo 2024–2034

**Cuantificación de Cambios Proyectados.** La Tabla 29 detalla los cambios netos proyectados para cada clase al año 2034. Complementariamente, la Figura 81 muestra la continuidad de estas tendencias respecto a la serie histórica 2004 - 2024.

- La cobertura Urbana alcanzó una superficie proyectada de  $66.95 \text{ km}^2$  representando un incremento neto de  $16.67 \text{ km}^2$  y el 12.63 % del área total del estudio.
- La clase Vegetación se proyectó en  $41.11 \text{ km}^2$ , estableciendo una reducción neta de  $10.85 \text{ km}^2$  respecto al año 2024 y una pérdida de 2.05 % en la configuración del paisaje
- El Suelo Desnudo presentó una transición de  $427.91 \text{ km}^2$  a  $422.09 \text{ km}^2$  durante el horizonte proyectado, con un diferencial negativo de  $5.82 \text{ km}^2$  y una ocupación final del 79.62 % del territorio

**Tabla 29**

*Proyección de área por clase LULC 2024-2034.*

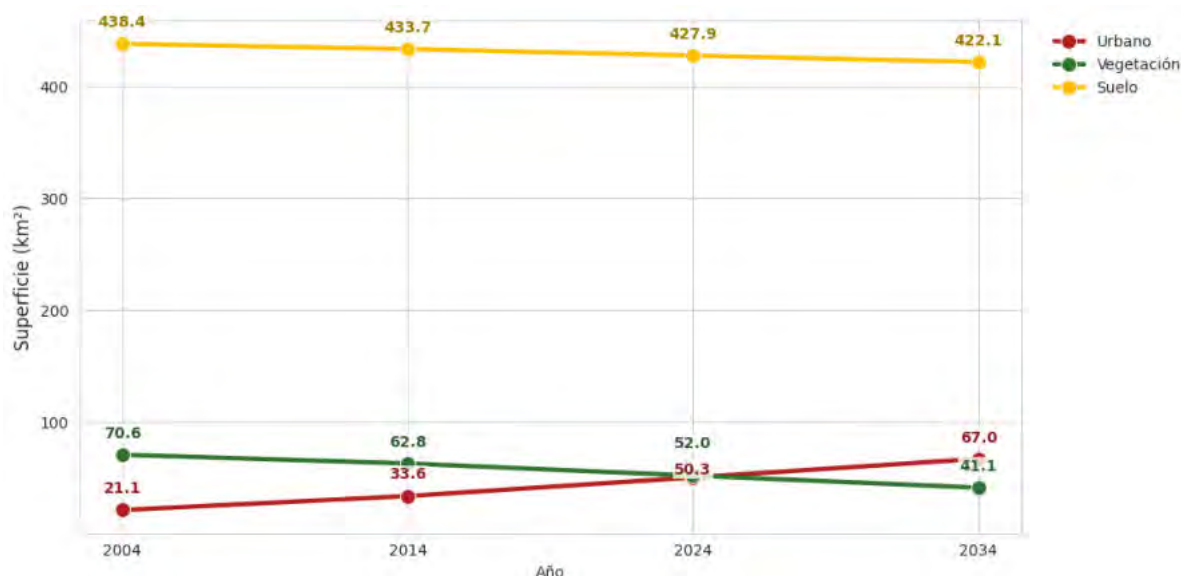
Clase	2024 ( $\text{km}^2$ )	2034 ( $\text{km}^2$ )	$\Delta$ ( $\text{km}^2$ )	2024 (%)	2034 (%)	$\Delta$ (%)
Urbano	50.28	66.95	+16.67	9.48	12.63	+3.14
Vegetación	51.96	41.11	-10.85	9.80	7.76	-2.05
Suelo Desnudo	427.91	422.09	-5.82	80.72	79.62	-1.10

*Nota. Estadísticas de área proyectadas por clase LULC para el período 2024–2034. Fuente: Elaboración propia.*

La Figura 82 ilustra la transición de Suelo Desnudo a Urbano proyectada para el periodo 2024-2034, mostrando las zonas que el modelo proyecta expansión. El catálogo completo de mapas de cambio simulados, incluyendo todas las transiciones y persistencias proyectadas, se encuentra organizado en el repositorio digital del proyecto (Apéndice E).

**Figura 81**

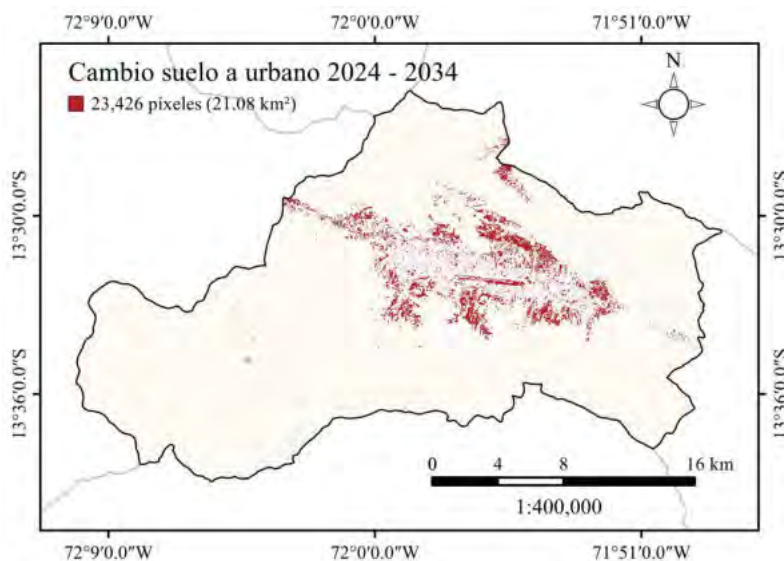
*Tendencia de evolución de coberturas (Proyección 2034).*



*Nota. Tendencia de la evolución de las coberturas de suelo proyectada para el período 2024 a 2034. Fuente: Elaboración propia.*

**Figura 82**

*Simulación de cambio de cobertura de suelo desnudo a urbana (2024 - 2034).*



*Nota. Fuente: Elaboración propia.*

**Dinámica de Transiciones Proyectada.** La dinámica de cambios esperada para el periodo 2024-2034 se muestra en la Tabla 30 y el diagrama de Sankey (Figura 83).

- La clase Suelo Desnudo (Clase 3) presenta una probabilidad de permanencia del 95.07 % ( $P_{33}$ ), la cobertura Urbana con 88.49 % ( $P_{11}$ ) y la Vegetación con 79.13 % ( $P_{22}$ ).



- La transición de Suelo Desnudo a Urbano ( $P_{31}$ ) se cuantifica en 4.93 %, mientras que la conversión de Vegetación a Urbano ( $P_{21}$ ) alcanza el 2.64 %.
- La cobertura vegetal muestra una probabilidad de conversión hacia Suelo Desnudo ( $P_{23}$ ) del 18.23 %.
- Se registró una probabilidad nula ( $P_{32} = 0.00$ ) para la transición de Suelo Desnudo hacia Vegetación. Por otro lado, la transición de Urbano a Suelo Desnudo ( $P_{13}$ ) fue de 11.51 %.

**Tabla 30**

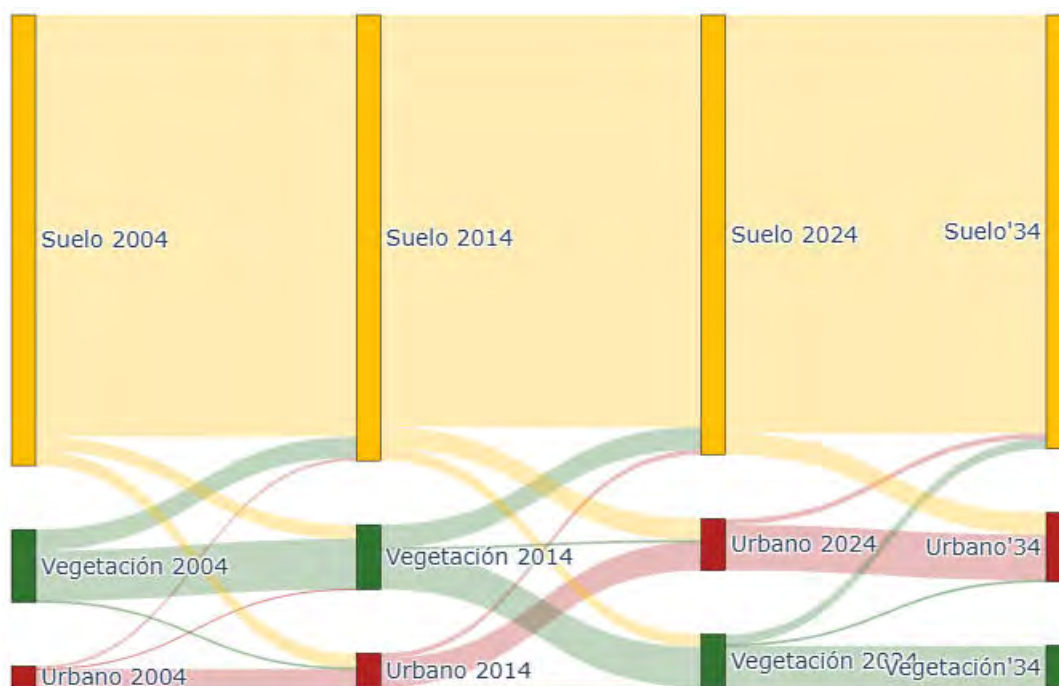
*Matriz de probabilidad de transición utilizada para la proyección 2024 - 2034.*

De (Clase)	A (Clase)		
	Urbano	Vegetación	Suelo Desnudo
Urbano	0.8849	0.0000	0.1151
Vegetación	0.0264	0.7913	0.1823
Suelo Desnudo	0.0493	0.0000	0.9507

*Fuente: Elaboración propia a partir del análisis de cambio 2014 - 2024.*

**Figura 83**

*Diagrama de Sankey de la evolución de coberturas del suelo (2020 - 2034).*



*Nota. El diagrama ilustra la evolución y estabilidad de las coberturas del suelo (LULC) en la proyección al 2034. Fuente: Elaboración propia.*

Las Figuras 84 y 85 presentan la configuración espacial de la cobertura LULC proyectada al año 2034 para los sectores oeste y este de la provincia. El archivo en formato raster (.tif) correspondiente a la proyección completa para el área de estudio se encuentra disponible en el repositorio digital del proyecto (Apéndice E).

**Figura 84**

*Mapa LULC Proyectado (Oeste) 2034.*



*Nota. Fuente: Elaboración propia.*

**Figura 85**

*Mapa LULC Proyectado (Este) 2034.*

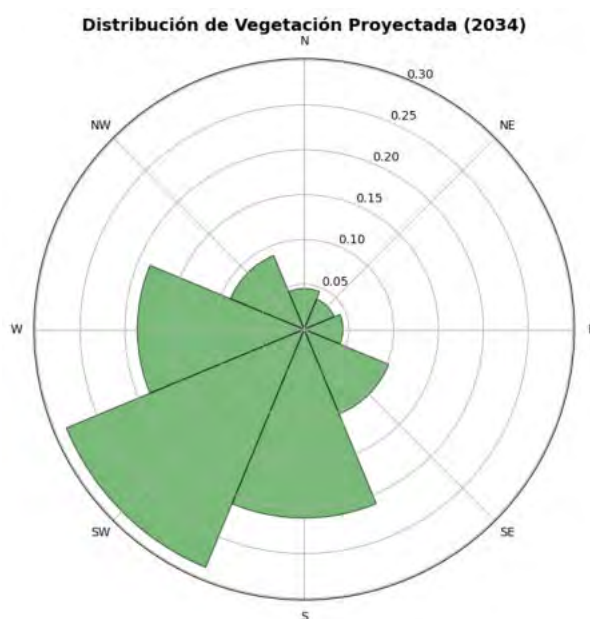


*Nota. Fuente: Elaboración propia.*

**Consistencia de la Distribución Clase Vegetación Proyectada.** La clase Vegetación proyectada al 2034 se distribuye en las laderas de umbría (S, SE, SW), las cuales integran más del 60 % de la superficie total de esta cobertura (Figura 86). La orientación Suroeste (SW) registra el 28.73 % de la superficie vegetal, mientras que las orientaciones de solana (N, NE) presentan valores inferiores al 5 % (Tabla 31).

**Figura 86**

*Diagrama de frecuencia direccional proyectada de la vegetación (2034).*



*Nota. La Rosa de los Vientos muestra la orientación predominante de la cobertura vegetal proyectada por el modelo para el año 2034. Fuente: Elaboración propia.*

**Tabla 31**

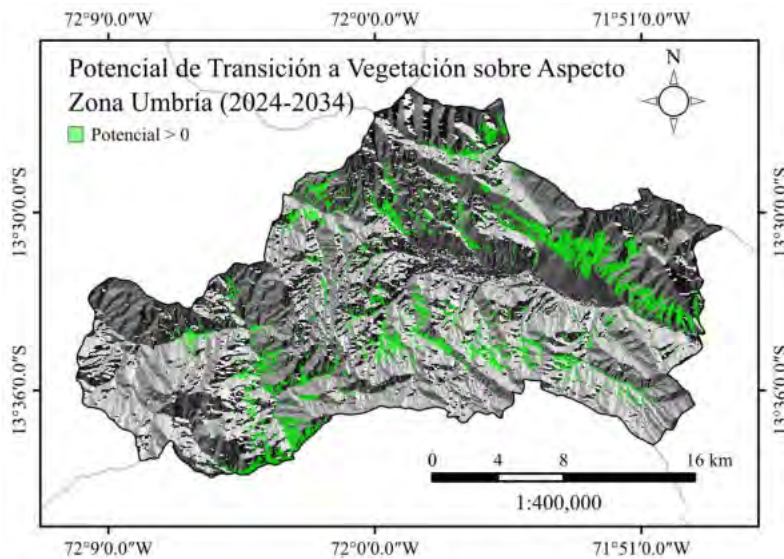
*Distribución proyectada de la vegetación según orientación (2034).*

Orientación	% Vegetación 2034
Norte (N)	4.54 %
Noreste (NE)	3.64 %
Este (E)	4.33 %
Sureste (SE)	10.22 %
Sur (S)	21.04 %
Suroeste (SW)	28.73 %
Oeste (W)	18.62 %
Noroeste (NW)	8.90 %

*Nota. Fuente: Elaboración propia.*

**Figura 87**

*Mapa de Potencial de Vegetación (Proyección 2024-2034).*



*Nota. El mapa muestra el Potencial de Vegetación proyectado por el modelo para el período 2024-2034. Fuente: Elaboración propia.*

## 5.2 Discusión

### 5.2.1 Fiabilidad de la Clasificación y Confusión Espectral

**Interpretación del Desempeño de Clasificación.** Las precisiones globales obtenidas (94.8 %-95.3 %) y los coeficientes Kappa (0.922-0.929) se sitúan dentro de los rangos reportados en la literatura reciente. El estudio de K. C. Roy et al. (2024) se limitó al uso de índices espectrales obteniendo un OA máximo de 90.73 %, la inclusión de variables topográficas (elevación y pendiente) en el trabajo de Belay et al. (2024) permitió elevar la precisión global hasta un 94.84 %. Esto sugiere que estos datos topográficos son factores discriminantes que mejoran la clasificación de coberturas. Sin embargo, los resultados fueron inferiores a los reportados por Gündüz (2025) (OA de 98 %). Este resultado se atribuye principalmente a la resolución espacial. Mientras que en esta investigación los sensores fueron homogenizados a 30 m, Gündüz trabajó exclusivamente con la resolución nativa de Sentinel-2, lo que sugiere que una mayor resolución espacial favorece la separación de coberturas.

**Interpretación de los Patrones de Acierto.** La exactitud del productor en el presente análisis varía entre 88.81 % y 100 %, mientras que la exactitud del usuario se sitúa en el rango de 90.61 % a 100 %. Estos valores son comparables con los reportados por estudios previos,

donde la exactitud del productor oscila entre 91.1 % y 100 % y la exactitud del usuario entre 99.1 % y 100 % (B. Roy, 2021). Si bien en este trabajo se registran ligeras disminuciones en la exactitud del usuario para las clases Urbano y Suelo en determinados años, los resultados globales confirman una elevada fiabilidad de la clasificación.

La clase Vegetación fue la mejor identificada en los tres años. Se observa una mejora progresiva en su Recall (99.3 % en 2004, 98.8 % en 2014, 100 % en 2024), lo que muestre la efectividad de las variables topográficas e índices espectrales para discriminar vegetación persistente activa durante el invierno austral.

Los modelos RF para cada año registraron precisiones globales (OA) superiores a 0.91 incluso con configuraciones de 10 árboles y una fracción de muestreo de 0.1. Este desempeño, alcanzado con configuraciones simples, contrasta con redes neuronales (CNN o ANN), las cuales, a pesar de capturar relaciones no lineales complejas, demandan mayores recursos de procesamiento y volúmenes de datos más extensos (Hussain et al., 2025; Khosravi, 2025).

#### ***Análisis de los Patrones de Error.***

1. Las matrices de transición muestran confusión entre las categorías Urbano (Clase 1) y Suelo Desnudo (Clase 3), con valores de transición ( $P_{13}$ ) que oscilan entre el 0.13 y 0.14 en ambos periodos. Este patrón de error se debe a la similitud espectral entre ciertos materiales de construcción como arcilla y adobe en viviendas (Yañe Zuñiga, 2019) y las superficies de suelo desnudo en el espectro óptico (Ettehadi et al., 2019; Krivoguz, 2024). Dicha dificultad se ve acentuada por la presencia de píxeles mixtos, los cuales integran contribuciones de asfalto, vegetación y estructuras dentro de una misma unidad (Tasan et al., 2025). La literatura científica valida estos hallazgos, reportando tasas de confusión similares en rangos de 0.07 a 0.27 (Badshah et al., 2024; Bendeckhou et al., 2024; Duan et al., 2025; Kamran et al., 2024), lo que confirma que esta imprecisión es un desafío inherente y esperado en la clasificación LULC.
2. La clase Suelo Desnudo presentó la Precisión del Usuario más baja de las tres clases (91.5 % en 2004, 90.6 % en 2014, 90.7 % en 2024). Esto indica que los píxeles etiquetados como "Suelo Desnudo" en los mapas clasificados tienen una mayor probabilidad de incluir errores de comisión, principalmente píxeles que en realidad son



Urbanos pero que el modelo no logró identificar correctamente.

3. La clase Urbano tuvo los valores más bajos de Precisión del Productor (92.6 % en 2004, 88.8 % en 2014, 89.3 % en 2024). Esto significa que el principal fallo del modelo es la omisión de una fracción de los píxeles urbanos reales, confundiéndolos predominantemente con Suelo Desnudo.

La comparación del rendimiento entre los diferentes sensores sugiere que las variaciones en las características radiométricas y espectrales entre Landsat 5 TM, Landsat 8 OLI y Sentinel-2 MSI no introdujeron sesgos, ni degradaron el desempeño del clasificador RF. Esto respalda la viabilidad de utilizar estos sensores de forma combinada para análisis multitemporales, consecuente con estudios que usaron distintos sensores (Buthelezi et al., 2024; Khosravi, 2025)

**Sinergia de Variables Espectrales y Topográficas.** Los resultados de importancia relativa indican que la clasificación no depende de un único predictor, sino de la interacción entre datos espectrales y físicos. El peso de la topografía es el factor más determinante, lo que responde a la configuración geográfica del valle del Cusco. En este entorno andino, la altitud y la inclinación del terreno dictan la distribución de las coberturas, mientras el crecimiento urbano se concentra en el fondo del valle, la vegetación y los suelos descubiertos ocupan principalmente las laderas (Figuras 80b, 80a). Mientras que los índices espectrales capturan la firma de los materiales (Vahid & Aly, 2025).

### ***5.2.2 Análisis de la Dinámica Temporal y Secuencialidad del Cambio***

La expansión urbana en la provincia de Cusco muestra una aceleración entre los dos periodos analizados. El cambio neto de la superficie urbana subió de 12.48 km<sup>2</sup> en el primer decenio a 16.67 km<sup>2</sup> en el segundo. Este ritmo más rápido coincide con el aumento en la probabilidad de transición desde el suelo desnudo ( $P_{31}$ ), que pasó de 3.21 % a 4.86 %, siendo el principal flujo de suelo hacia cobertura urbana.

El análisis de las matrices de Markov revela un proceso de cambio secuencial en dos etapas. En la primera, la Vegetación actúa como la fuente primaria de cambio hacia el Suelo Desnudo. Esta transición ( $P_{23}$ ) es la más alta y se intensificó del 28.9 % al 35.6 % entre

periodos. En la segunda etapa, este suelo desnudo (nuevo y preexistente) funciona como la reserva inmediata para la urbanización final. Esta secuencia explica por qué la Vegetación registra la mayor pérdida neta de área natural, a pesar de que el Suelo Desnudo transfiere mas superficie a la clase urbana.

La vulnerabilidad de las coberturas también es desigual. Mientras que el Suelo Desnudo es la clase más estable del paisaje ( $P_{33} > 92\%$ ), la Vegetación es la más inestable ( $P_{22} = 69.3\%$  y  $62.2\%$ ). Las transiciones directas de Vegetación a Urbano ( $P_{21} < 2.2\%$ ) son mínimas, lo que confirma que el proceso de urbanización en Cusco no ocurre directamente sobre vegetación, sino sobre terrenos despejados.

### 5.2.3 Validez del Modelo Predictivo MLP-CA-MC

**Implicancias de la Colinealidad en el Modelamiento.** La fuerte correlación (0.84-0.90) entre las distancias a vías, aeropuerto y rieles evidencia que el crecimiento se concentra sobre un eje común. Este patrón se extiende a los servicios públicos, donde la asociación entre centros de salud e instituciones educativas se mantuvo estable ( $\rho = 0.82$ ,  $\rho = 0.81$  y  $\rho = 0.84$ ), indicando una concentración del equipamiento urbano en sectores específicos de la provincia.

Se cuantificó un incremento en la correlación entre el acceso educativo y las variables de transporte. La asociación con la *distancia a vías* ascendió de 0.64 a 0.73, mientras que con la *distancia al aeropuerto* subió de 0.61 a 0.74 entre 2004 y 2024.

La elevación (*DEM*) mostró coeficientes moderados ( $\rho \approx 0.67$ ) respecto a las variables de educación y salud. Asimismo, se halló una correlación de  $\rho \approx 0.60$  entre la elevación y las poblaciones de mayor densidad.

La independencia estadística de la pendiente y el aspecto ( $|\rho| < 0.1$ ) respecto a los conductores antrópicos valida su inclusión como predictores, permitiendo al modelo capturar restricciones físicas no ligadas al desarrollo de infraestructura.

### Interpretación de las Métricas de Acuerdo.

**Acuerdo Categórico.** La diferencia observada entre la Exactitud Global ( $OA = 88.90\%$ ) y el coeficiente  $K_{standard}$  (0.6620) se explica por la estructura del paisaje en el área de estudio, donde la clase Suelo Desnudo ocupa aproximadamente el 80% del territorio.

Esta marcada dominancia incrementa el acuerdo esperado por azar ( $P_e$ ), reduciendo el margen de mejora del índice Kappa. El valor obtenido indica que el modelo MLP-CA-MC explica el 66.2 % del potencial de cambio más allá de una asignación aleatoria.

La precisión global (OA) obtenida supera a la reportada por Pham y Ali (2024) (79.58 %), cuyo modelo se basó en un conjunto más restringido de variables predictoras, principalmente topográficas y algunos índices espectrales. Esta diferencia sugiere que la incorporación de un conjunto más amplio y diverso de predictores contribuye a mejorar el desempeño del modelo. Asimismo, los resultados son comparables a los reportados por W. Khalid et al. (2024) (85.39 %), lo cual puede explicarse por la similitud en el enfoque metodológico, ya que ambos trabajos integran variables topográficas junto con factores antrópicos y de accesibilidad. Sin embargo, la precisión alcanzada es inferior a la reportada por Blissag, Bilal et al. (2024) (92.63 %), quien incorporó variables dinámicas como la densidad de población. Este contraste indica que la inclusión de factores dinámicos con fuerte incidencia en los procesos de cambio del uso del suelo resulta determinante para alcanzar mayores niveles de desempeño predictivo.

Los valores de  $P_{quantity} = 0.9921$  y  $K_{histo} = 0.9759$  validan el uso de las matrices de transición históricas. La simulación reproduce las magnitudes totales de superficie, lo que indica que la demanda de cambio calculada mediante Cadenas de Markov es precisa y no induce sesgos. Estos resultados concuerdan con investigaciones previas que reportaron valores de  $K_{histo}$  de 0.95 (Ahmad et al., 2025) y 0.9 (Al Kafy et al., 2024).

El análisis del F1-Score evidencia diferencias en el desempeño del modelo de predicción según la cobertura simulada. La clase Suelo presenta un valor elevado ( $F1 = 0.93$ ), asociado a su alta representatividad espacial y a la adecuada reproducción de su patrón de ocupación. En contraste, las clases Urbano y Vegetación registran valores más moderados ( $F1 = 0.72$  y  $0.68$ , respectivamente), lo que indica que la Exactitud Global se encuentra influenciada por la predominancia de la clase mayoritaria.

El menor F1-Score de la clase Vegetación refleja dificultades en la asignación espacial de las transiciones entre vegetación y suelo desnudo. La mayor incertidumbre de las proyecciones se concentra en áreas periféricas sometidas a presión urbana, donde los modelos presentan limitaciones para simular crecimientos fragmentados o espontáneos en comparación con



patrones de expansión concéntrica (Duan et al., 2025). Esta dinámica irregular dificulta la predicción precisa de la localización de nuevos desarrollos urbanos, afectando la exactitud espacial del mapeo de cambios (Y. Liu et al., 2024a; Tiamgne et al., 2025).

La subestimación de la expansión construida se manifiesta en la asignación de píxeles urbanos como *Suelo Desnudo*, lo que reduce la exactitud del productor para la clase Urbano y conduce a una representación de menor extensión de las áreas de transición urbano-periurbanas respecto a la realidad observada (Duan et al., 2025). De forma complementaria, la comisión de píxeles en la clase *Urbano* se asocia a probabilidades de transición superiores a la ocupación real registrada en el mapa de referencia, lo que afecta la exactitud del usuario.

**Capacidad Discriminante de Potencial de Cambio.** El valor de *AUC* obtenido (0.9438) se sitúa en el rango calificado como excelente por la literatura especializada ( $AUC > 0.90$ ), lo que ratifica que la combinación de variables conductoras y el modelo híbrido explican la complejidad de las dinámicas de cambio en la provincia de Cusco. Este resultado es equivalente al 0.88 reportado en estudios previos sobre modelado de crecimiento (Badshah et al., 2024).

*AUC* es superior a la Exactitud Global (*OA*), esto sugiere que, aunque existan errores de asignación final en píxeles con probabilidades muy próximas, el modelo ordena y prioriza correctamente las zonas con mayor susceptibilidad de cambio. En consecuencia, se valida el uso de estas superficies de probabilidad como insumo base para los Autómatas Celulares, asegurando que la asignación espacial de la expansión urbana proyectada se fundamenta en reglas de transición estadísticamente sólidas.

**Sensibilidad Espacial y Multiescala.** El aumento de la precisión global al 97.05 % bajo una tolerancia difusa de  $3 \times 3$  indica que las discrepancias de asignación representan, predominantemente, desplazamientos espaciales inferiores a 30 metros respecto a la referencia. Al obtenerse un coeficiente  $K_{Fuzzy}$  superior al de los datos base, se confirma la validez técnica del modelo de simulación (Hagen, 2003). Este comportamiento se refuerza con los resultados de la validación multiescala, el incremento sostenido del  $K_{standard}$  conforme aumenta la unidad de agregación (alcanzando 0.7861 a 960m) demuestra que los errores de localización a escala de celda se compensan al evaluar el sistema en escalas geográficas mayores. El uso de escalas

tan diversas en la literatura técnica dificulta la comparación directa entre estudios (Marey et al., 2025).

***Coherencia Simulación de la Clase Vegetación.*** La Tabla 27 muestra una correspondencia entre los valores reales y simulados para la distribución porcentual por orientación. Las diferencias se mantienen dentro de un rango con desviaciones negativas en las orientaciones N, NE, E y NW, y desviaciones positivas en SE, S, SW y W.

La orientación suroeste presenta el mayor porcentaje tanto en los datos reales (27.50 %) como en los simulados (28.37 %), con una diferencia reducida de 0.87 %. Mientras que la mayor subestimación corresponde a la orientación noroeste (-2.01 %).

Los datos muestran que la simulación reproduce la distribución porcentual con errores reducidos, lo que indica una buena concordancia entre los valores reales y los simulados.

Los mapas de potencial de ser de cobertura Vegetación generados por el MLP (Figura 77) refuerzan esta interpretación. Las zonas clasificadas con potencial positivo coinciden con las laderas de menor exposición solar, mientras que las áreas más secas son penalizadas.

### **Interpretación de los Factores Impulsores del Cambio.**

***Jerarquía y Contribución de las variables predictoras.*** A partir del análisis de importancia por permutación (Tabla 28), se evidencia una contribución desigual de las variables al desempeño del modelo (AUC base = 0.9438).

En el modelo MLP la elevación (DEM) es la variable más influyente, con una caída del AUC de 0.0481, equivalente al 54.7 % de la importancia relativa. En un segundo nivel se sitúan las instituciones educativas (IIEE), con una reducción del AUC de 0.0143 (16.21 %).

Las variables pendiente, turismo y vías presentan una importancia intermedia, con disminuciones del AUC entre 0.0043 y 0.0071. En contraste, poblaciones grandes, hidrografía, poblaciones dispersas y aspecto muestran efectos marginales, con caídas del AUC inferiores a 0.003 y aportes relativos menores al 3 %.

***Análisis de Dependencia no Lineal de las Variables Predictoras.*** La elevación y la pendiente presentan el impacto más claro sobre las probabilidades. La clase urbana es máxima en terrenos bajos y planos, disminuyendo de forma constante a medida que aumenta la altitud y la inclinación. El aspecto, por su parte, muestra una influencia menor con cambios leves en la

tendencia.

La probabilidad de expansión urbana es más alta en áreas cercanas a instituciones educativas, vías y zonas turísticas, reduciéndose progresivamente conforme aumenta la distancia. En el caso de las vías, este descenso urbano se compensa con un incremento en la probabilidad de vegetación en zonas alejadas.

La cercanía a poblaciones grandes favorece la probabilidad de la clase urbana. En contraste, el efecto de las poblaciones dispersas es moderado y muestra un comportamiento donde la probabilidad urbana se estabiliza después de alcanzar valores intermedios.

La hidrografía presenta probabilidades constantes en todo su rango. Esto indica que la cercanía a ríos no actúa como un factor que permita diferenciar entre las coberturas de suelo, urbano o vegetación.

#### **5.2.4 Dinámica Territorial 2034**

**Análisis de Cambios Proyectados.** La proyección al año 2034 confirma la persistencia de las tendencias observadas en las últimas dos décadas. El incremento del 33 % en la cobertura urbana muestra una aceleración en la presión sobre el territorio, sugiriendo que los procesos de consolidación y densificación continuarán en torno a los núcleos existentes y sus periferias inmediatas.

La contracción proyectada de la Vegetación ( $-10.85 \text{ km}^2$ ) resulta significativa, siendo la clase con la pérdida más alta.

**Dinámica de Transiciones al 2034.** El análisis de las probabilidades de transición proyectadas revela un patrón de cambio territorial, el cual guarda una estrecha coherencia con las dinámicas históricas analizadas en fases previas de esta investigación. Esta lógica de cambio se explica a través de los siguientes ejes interpretativos:

***Presión y Degradación de la Vegetación Remanente.*** El modelo muestra una persistencia de 79.1 % para la Vegetación ( $P_{22}$ ), este valor aunque mayor a de los anteriores años de estudio, se aplica sobre una superficie cada vez mas pequeña (9.8 %). La secuencia de transformación observada en periodos anteriores continúa, donde la vegetación es sometida primero a una conversión a Suelo Desnudo ( $P_{23} = 18.2 \%$ ), y el suelo desnudo nuevo y

preexistente transita hacia cobertura Urbana ( $P_{21} = 2.64 \%$ ). La estabilidad del Suelo Desnudo ( $P_{33} = 0.95$ ), evidencia que el modelo incorpora restricciones de aptitud que frenan una transformación masiva de las laderas del valle, manteniendo la expansión urbana confinada a los lugares de mayor potencial.

***Irreversibilidad de las Tendencias Proyectadas (Vegetación).*** Se proyecta una probabilidad nula de regeneración de vegetación a partir de suelo desnudo ( $P_{32} = 0.0$ ) bajo las tendencias actuales. La ausencia de este flujo en la simulación implica que cada hectárea de vegetación perdida se considera una reducción definitiva del patrimonio natural dentro del horizonte de proyección al 2034.

***Interpretación de Anomalías Metodológicas y Confusión Espectral.*** El escenario proyectado al 2024 y 2034 presenta dos anomalías espaciales localizados que requieren una interpretación detallada.

***La “Desurbanización” (Efecto  $P_{13}$ ):.***

1. El ruido de clasificación (píxeles que “oscilan” entre Urbano y Suelo Desnudo) se cuantifica en la matriz de transición como un cambio real. Este fenómeno genera un incremento artificial en la probabilidad  $P_{13}$  (Urbano a Suelo Desnudo), asignándole un valor que es lógicamente improbable para una ciudad en crecimiento.
2. La cadena de Markov determina la probabilidad de cambio  $P_{13}$  a Suelo Desnudo.
3. El MLP, al entrenarse con los datos ruidosos, aprende a identificar las características de los píxeles que “cambiaron” de Urbano a Suelo asignando potencial de transición  $P_{13}$ .
4. El autómata celular (CA) aplica la demanda de Markov ( $P_{13}$ ) en las ubicaciones de mayor potencial del MLP, resultando en una aparente desurbanización.

***La “Urbanización” de Intersticios Urbanos (Efecto  $P_{31}$ ).*** Esta anomalía sigue una lógica de modelo diferente, relacionada con la transición  $P_{31}$  (Suelo Desnudo a Urbano), la cual constituye un proceso real en la dinámica de la provincia.

1. La conversión de Suelo Desnudo a Urbano ( $P_{31} = 4.93 \%$ ) es la principal dinámica de crecimiento esperada por el modelo. El error observado no radica en la magnitud de

la transición, sino en su asignación espacial preferente hacia el interior de la ciudad consolidada.

2. El MLP asigna un potencial de transición ( $P_{31}$ ) alto al suelo desnudo ubicados dentro de la cobertura urbana (intersticios). Esto ocurre porque dichos espacios son ideales para las variables jerárquicamente dominantes, se ubican en el rango de Elevación predilecto (54.7 %) y maximizan la cercanía a Instituciones Educativas (16.2 %).
3. Al ejecutar la simulación, el Autómata Celular (CA) canaliza parte de la demanda de crecimiento ( $P_{31}$ ) hacia estos vacíos internos, generando un efecto de densificación. Esto resulta en una mancha urbana simulada mucho más compacta que la real y produce fenómenos funcionalmente inviables, como la urbanización alrededor de las pistas de aterrizaje.

**Coherencia Proyección de la Clase Vegetación.** El modelo proyecta la mayor densidad de la clase Vegetación en las laderas Sur y Suroeste, esto demuestra que el modelo híbrido reproduce correctamente las restricciones biofísicas de la provincia. Al observar los graficas de influencia del aspecto (Figuras 80c, 78) en el MLP, se puede inferir que este fenómeno no proviene del predictor aspecto, si no de las tendencias historicas.

## Conclusiones

1. La proyección del escenario tendencial al 2034 cuantifica una expansión urbana neta de 16.67 km<sup>2</sup>, sostenida por la reducción simultánea de las coberturas de Vegetación (-10.85 km<sup>2</sup>) y Suelo Desnudo (-5.82 km<sup>2</sup>). La simulación estima una probabilidad nula de regeneración de Vegetación a partir de Suelo Desnudo ( $P_{32} = 0.0$ ), estableciendo la irreversibilidad de la pérdida de cobertura vegetal bajo las tendencias históricas.
2. El clasificador Random Forest demostró una alta precisión y consistencia para los tres períodos (2004, 2014 y 2024), alcanzando precisiones globales superiores al 94.8 % y coeficientes Kappa mayores a 0.92. Se concluye que la metodología es robusta para la clasificación LULC. La principal fuente de error residual identificada es la confusión espectral sistemática entre las clases Suelo Desnudo y Urbano.
3. Se cuantificó una expansión urbana acelerada impulsada por un proceso secuencial. El área urbana pasó de 21.13 km<sup>2</sup> en 2004 a 50.28 km<sup>2</sup> en 2024. El análisis de transiciones identificó una dinámica encadenada como el proceso dominante de transformación: la Vegetación es convertida a Suelo Desnudo y esta reserva de Suelo Desnudo (nuevo y preexistente) actúa como la fuente bruta inmediata para la expansión final de la clase Urbano.
4. La simulación alcanzó una Exactitud Global (*Overall Accuracy*) de 0.8890, respaldada por una capacidad discriminante del análisis AUC-ROC = 0.9438. Su principal fortaleza es la alta fidelidad para simular la cantidad de cambio ( $Kh_{isto} \approx 0.98$ ) y los patrones espaciales generales (Precisión Difusa > 97 % y Análisis Multiescala (960 m) > 78 %). Su debilidad principal es la precisión de la localización exacta a nivel de píxel.

### Recomendaciones

- Se sugiere explorar el uso de imágenes satelitales de mayor resolución espacial para los periodos más recientes disponibles. Esto permitiría mejorar la segmentación de superficies construidas y reducir la incertidumbre en la estimación de las tasas de cambio, refinando la precisión de la clasificación base.
- Dado que el modelo actual se fundamenta en variables estáticas, se recomienda enriquecer el espacio de características mediante la integración de datos socioeconómicos dinámicos, como la densidad poblacional o la valoración del suelo, para capturar con mayor fidelidad la complejidad de los factores que impulsan la urbanización.
- Se propone desarrollar un flujo de trabajo automatizado que incorpore y procese nuevas imágenes satelitales de forma periódica. Esto permitiría el reentrenamiento automático de los pesos del Perceptrón Multicapa y la actualización de las matrices de transición, transformando el modelo estático en un sistema de monitoreo continuo y adaptable a nuevas tendencias.
- Se sugiere la implementación de un módulo de software para la identificación de la transición de vegetación a suelo desnudo. Este sistema permitiría monitorear la aparición de nuevas áreas expuestas y generar notificaciones automáticas sobre zonas con alta probabilidad de expansión urbana no planificada.
- Aprovechando las proyecciones generadas para el horizonte 2034, se recomienda el desarrollo de algoritmos de optimización de ubicación-asignación (Location-Allocation Algorithms) que utilicen la distribución simulada de la población como dato de entrada. Estos algoritmos permitirían calcular las coordenadas óptimas para nuevos nodos de servicios, buscando minimizar los costos de desplazamiento y maximizar la cobertura de atención a la población.

## Referencias

- A, S., Mazroa, A. A., Maashi, M., T.M, N., & V, P. (2024). Remote sensing and machine learning approach for zoning of wastewater drainage system. *Desalination and Water Treatment*, 319, 100549. <https://doi.org/10.1016/j.dwt.2024.100549>
- Ahmad, S., Farooq, R., Waseem, M., & Kohnová, S. (2025). How do land use changes affect temperature and groundwater in urban areas? An integrated remote sensing, and machine learning approach. *Advances in Space Research*, 76, 3963-3987. <https://doi.org/10.1016/j.asr.2025.08.021>
- Aiquipa Alosilla, A., & Soncco Mamani, A. R. (2019). Terrapuerto Inter-Regional del Eje Sur - Cusco [Facultad de Arquitectura e Ingeniería Civil, Escuela Profesional de Arquitectura. Asesores: Mgt. Arq. Wilbert Sany Salazar Muñiz y Arq. Mario Ivan Gallegos Gongora].
- Al Kafy, A., Dey, N. N., Saha, M., Altuwaijri, H. A., Fattah, M. A., Rahaman, Z. A., Kalaivani, S., Bakshi, A., & Rahaman, S. N. (2024). Leveraging machine learning algorithms in dynamic modeling of urban expansion, surface heat islands, and carbon storage for sustainable environmental management in coastal ecosystems. *Journal of Environmental Management*, 370, 122427. <https://doi.org/10.1016/j.jenvman.2024.122427>
- Al-Abadi, A. (2015). Modeling of groundwater productivity in northeastern Wasit Governorate, Iraq using frequency ratio and Shanno's entropy models. *Applied Water Science*, 7. <https://doi.org/10.1007/s13201-015-0283-1>
- Alam, S. A., & Maiti, R. (2025). Deciphering sediment pathways: Novel integrated approaches for sediment source identification and vulnerability prediction by machine learning models in major dam catchments in Chota Nagpur plateau, India. *Physics and Chemistry of the Earth*, 140, 103974. <https://doi.org/10.1016/j.pce.2025.103974>
- Al-Kordi, H., Al-Amri, A., & Govinda raju, G. (2025). Landslide susceptibility mapping using geospatial, analytical hierarchy process (AHP), and binary logistic regression (BLR) techniques - A study of Wadi Habban Basin, Shabwah, Yemen. *Results in Earth Sciences*, 3, 100103. <https://doi.org/10.1016/j.rines.2025.100103>



- Alqadhi, S., Bindajam, A. A., Mallick, J., Rahman, A., & Talukdar, S. (2024). Applying deep learning to manage urban ecosystems in arid Abha, Saudi Arabia: Remote sensing-based modelling for ecological condition assessment and decision-making. *Heliyon*, 10, e25731. <https://doi.org/10.1016/j.heliyon.2024.e25731>
- Altmann, A., Tološi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: a corrected feature importance measure. *Bioinformatics*, 26(10), 1340-1347. <https://doi.org/10.1093/bioinformatics/btq134>
- Amindin, A., Siamian, N., Kariminejad, N., Clague, J. J., & Pourghasemi, H. R. (2024). An integrated GEE and machine learning framework for detecting ecological stability under land use/land cover changes. *Global Ecology and Conservation*, 53, e03010. <https://doi.org/10.1016/j.gecco.2024.e03010>
- Anzalone, A., Pagliaro, A., & Tutone, A. (2024). An Introduction to Machine and Deep Learning Methods for Cloud Masking Applications. *Applied Sciences*, 14(7). <https://doi.org/10.3390/app14072887>
- Arce Quispe, B. A., & Chavez Cahuana, A. E. (2022). Plan de Gestión de Riesgos en la Ejecución de la Obra Mejoramiento del Servicio de Transitabilidad Vehicular y Peatonal en la Av. Pachacutec (Tramo: Óvalo Pachacutec - Óvalo Libertadores), Distrito de Wanchaq - Cusco - Cusco [Facultad de Ingeniería Civil y Arquitectura, Escuela Profesional de Ingeniería Civil. Asesor: Dr. Ing. Elenko E. Huarhua Huarhua].
- Arunab, K., & Mathew, A. (2024). Exploring spatial machine learning techniques for improving land surface temperature prediction. *Kuwait Journal of Science*, 51, 100242. <https://doi.org/10.1016/j.kjs.2024.100242>
- Ayma Quispe, J. C. (2022). Diagnóstico y Propuesta de Mejora de Procesos en la Gestión de Almacén en una Empresa de Cusco, 2021 [Facultad de Ingeniería de Procesos, Escuela Profesional de Ingeniería Química. Asesora: Mgt. Ysabel Almanza Loaiza].
- Badavath, N., & Sahoo, S. (2025). Geospatial assessment and integrated multi-model approach for landslide susceptibility mapping in Meghalaya, India. *Advances in Space Research*, 75(3), 2764-2791. <https://doi.org/https://doi.org/10.1016/j.asr.2024.11.052>

- Badshah, M. T., Hussain, K., Rehman, A. U., Mehmood, K., Muhammad, B., Wiarta, R., Silamon, R. F., Khan, M. A., & Meng, J. (2024). The role of random forest and Markov chain models in understanding metropolitan urban growth trajectory. *Front. For. Glob. Chang.*, 7.
- Belay, H., Melesse, A. M., & Tegegne, G. (2024). Scenario-Based Land Use and Land Cover Change Detection and Prediction Using the Cellular Automata-Markov Model in the Gumara Watershed, Upper Blue Nile Basin, Ethiopia. *Land*, 13(3). <https://doi.org/10.3390/land13030396>
- Bendechou, H., Akakba, A., Issam, K., Mohammed, & Salem, H. (2024). Monitoring and predicting land use/land cover dynamics in Djelfa city, Algeria, using Google Earth Engine and a Multi Layer Perceptron Markov Chain model. *Geogr. Pannonica*, 28(1), 1-20.
- Berrio Gomez, R. A. (2017). Modelo de Gestión para la Promoción de Viviendas Populares de Alta Densidad como Producto Turístico en la Región del Cusco.
- Bhuyan, N., Sajjad, H., Saha, T. K., Roshani, Sharma, Y., Masroor, M., Rahaman, M. H., & Ahmed, R. (2024). Assessing landscape ecological vulnerability to riverbank erosion in the Middle Brahmaputra floodplains of Assam, India using machine learning algorithms. *Catena*, 234, 107581. <https://doi.org/10.1016/j.catena.2023.107581>
- Blissag, Bilal, Yebdri, Djilali & Kessar, Cherif. (2024). Spatiotemporal change analysis of LULC using remote sensing and CA-ANN approach in the Hodna basin, NE of Algeria. *Physics and Chemistry of the Earth*, 133, 103535. <https://doi.org/10.1016/j.pce.2023.103535>
- Bojer, A. K. (2024). Machine learning and remote sensing based time series analysis for drought risk prediction in Borena Zone, Southwest Ethiopia. *Journal of Arid Environments*, 222, 105160. <https://doi.org/10.1016/j.jaridenv.2024.105160>
- Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., & Tait, A. M. (2022). Dynamic World, Near Real-Time Global 10 m Land Use Land Cover Mapping. *Scientific Data*, 9(1), 251. <https://doi.org/10.1038/s41597-022-01307-4>

- Burke, J. (2023). Why and how to use Google Colab [Accessed: 2025-10-26]. *TechTarget*. <https://www.techtarget.com/searchenterpriseai/tutorial/Why-and-how-to-use-Google-Colab>.
- Buthelezi, M. N. M., Lottering, R. T., Peerbhay, K. Y., & Mutanga, O. (2024). A machine learning approach to mapping suitable areas for forest vegetation in the eThekweni municipality. *Remote Sensing Applications: Society and Environment*, 35, 101208. <https://doi.org/10.1016/j.rsase.2024.101208>
- Cahuana, W. B. A., & Pariguana, O. Q. (2023). Modelamiento Hidrogeológico para Determinar la Influencia del Flujo Subterráneo a los Acuíferos de la Cuenca Huanacahuire, San Jerónimo - Cusco.
- Calvo Mamani, J., & Polo Dolmos, Z. (2017). *Evaluación de la Contaminación del Río Huatanay - Provincias de Cusco y Quispicanchi* [Tesis para optar el Título Profesional de Biólogo]. Universidad Nacional de San Antonio Abad del Cusco [Asesor: M. Sc. Luciano Julián Cruz Miranda].
- Carbajal Coronado, Y. S. (2024). Determinación de los parámetros para calcular el nivel de peligro por deslizamiento en la quebrada Ayahuayco, del distrito de Cusco, provincia del Cusco, departamento del Cusco, 2022 [Facultad de Ingeniería Geológica, Minas y Metalúrgica. Asesor: Mgt. Ing. José Dionicio Cárdenas Roque].
- Carlos Rosales, R. (2025). MICROORGANISMOS EFICIENTES EN LA PRODUCCIÓN DE AVENA FORRAJERA VARIEDAD INIA 908 MELLICERA (Avena sativa L.) EN K'AYRA, SAN JERONIMO - CUSCO.
- Castillo Alire, S. M., & Gutierrez Kancha, D. W. (2019). Nueva Sede - Dirección Desconcentrada de Cultura - Cusco [Facultad de Arquitectura e Ingeniería Civil, Escuela Profesional de Arquitectura. Asesores: Arqto. Luis Alberto Palacio Valenzuela y Arqto. Nancy Elena Paredes V.D.T.].
- Ccopa Barrionuevo, P. S., Dina y Mamani Sihuyaro. (2019). Conocimiento de Primeros Auxilios en Actividades Deportivas Recreativas de los Estudiantes de Primaria de la Institución Educativa de San Francisco de Borja del Cusco, 2017 [Tesis para optar al Título Profesional de Licenciadas en Educación Secundaria, Especialidad Educación Física, Universidad Nacional de San Antonio Abad del Cusco]. *unsaac*.

- Chaulagain, S., Lamichhane, M., & Chaulagain, U. (2025). A review of current trends, challenges, and future perspectives in machine learning applications to water resources in Nepal. *Journal of Hazardous Materials Advances*, 18, 100678. <https://doi.org/10.1016/j.hazadv.2025.100678>
- Chen, G. (2024). Urban inundation rapid prediction method based on multi-machine learning algorithm and rain pattern analysis. *Journal of Hydrology*, 633, 131059. <https://doi.org/10.1016/j.jhydrol.2024.131059>
- Condori Juarez, F. M., Yovana y Visa Quispe. (2019). Brecha en la calidad del servicio turístico caso: Agencias de viaje de la ciudad del Cusco, periodo 2017 [Tesis para optar el Título Profesional de Economista, Universidad Nacional de San Antonio Abad del Cusco]. *unsaac*.
- contributors, G. (2025). Softmax Activation Function in Neural Networks [última actualización: 17 Nov, 2025]. <https://www.geeksforgeeks.org/deep-learning/the-role-of-softmax-in-neural-networks-detailed-explanation-and-applications/>
- Dahal, S., Dangi, B., B.C., M., & Bhattarai, R. (2024). Land use and Land Cover Classification for Dang District Nepal using Satellite Imagery and Machine Learning on Google Earth Engine. *Journal of Geography, Environment and Earth Science International*, 28, 52-66. <https://doi.org/10.9734/jgeesi/2024/v28i12848>
- Danso, S. Y., Ma, Y., & Addo, I. Y. (2025). Delineating flood susceptibility zones using novel ensemble models - An application of evidential belief function, relative frequency ratio, and Shannon entropy. *International Journal of Applied Earth Observation and Geoinformation*, 141, 104669. <https://doi.org/10.1016/j.jag.2025.104669>
- DataScientest. (2025). Random Forest: Bosque aleatorio. Definición y funcionamiento [Último acceso: 18 de febrero de 2025]. <https://datascientest.com/es/random-forest-bosque-aleatorio-definicion-y-funcionamiento>
- Díaz-Uriarte, R., & Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7(1), 3. <https://doi.org/10.1186/1471-2105-7-3>

- Duan, X., Haseeb, M., Tahir, Z., Mahmood, S. A., & Tariq, A. (2025). Analyzing and predicting land use and land cover dynamics using multispectral high-resolution imagery and hybrid CA-Markov modeling. *Land Use Policy*, 157, 107655. <https://doi.org/10.1016/j.landusepol.2025.107655>
- EPSG.io. (2025a). EPSG:32719 – WGS 84 / UTM zone 19S. Consultado el 23 de julio de 2025, desde <https://epsg.io/32719>
- EPSG.io. (2025b). Mapa interactivo EPSG:32719 – WGS 84 / UTM zona 19S [Mapa interactivo basado en el sistema de referencia EPSG:32719 (UTM zona 19 Sur), con visualización callejera. Coordenadas centrales: Este 181562.69, Norte 8501457.67. Zoom nivel 9. Reproyectado desde WGS 84.]. Consultado el 23 de julio de 2025, desde <https://epsg.io/map%5C#srs=32719%5C&x=181562.687660%5C&y=8501457.671807%5C&z=9%5C&reproject=1%5C&layer=streets>
- Equator Studios. (2023). What is a Digital Elevation Model (DEM)? [Accessed: 2025-07-27]. <https://equatorstudios.com/what-is-a-digital-elevation-model-dem/>
- Esri. (2024). Vista general del conjunto de herramientas Vecindad [Accedido el 1 de noviembre de 2025]. <https://desktop.arcgis.com/es/arcmap/latest/tools/spatial-analyst-toolbox/an-overview-of-the-neighborhood-tools.htm>
- Esri. (n.d.). What is GIS? [Accessed: 2025-08-03]. <https://www.esri.com/en-us/what-is-gis/overview>
- Esri Support. (2025). EPSG code [EPSG es un acrónimo de European Petroleum Survey Group. Representa un identificador asignado a parámetros geodésicos, sistemas de referencia, datums, transformaciones y unidades de medida, mantenido por el IOGP Geomatics Committee. Cada parámetro recibe un código único EPSG y una representación en WKT según el diccionario GIS de Esri.]. Consultado el 23 de julio de 2025, desde <https://support.esri.com/es-es/gis-dictionary/epsg-code>
- Ettehadi, P., Kaya, S., Sertel, E., & Alganci, U. (2019). Separating Built-Up Areas from Bare Land in Mediterranean Cities Using Sentinel-2A Imagery. *Remote Sensing*, 11, 345. <https://doi.org/10.3390/rs11030345>

- European Space Agency. (2021, marzo). Copernicus Global DEM GLO-30 [Accessed: 2025-10-15]. <https://portal.opentopography.org/raster?opentopoID=OTSDEM.032021.4326.3>
- European Union/ESA/Copernicus and Google Earth Engine. (n.d.). COPERNICUS\_S2\_SR\_HARMONIZED: Harmonized Sentinel-2 MSI Level-2A Surface Reflectance [Consultado el 26 de julio de 2025].
- Ezekiel, M. O. (2017). AUTOMATIC EXTRACTION OF URBAN AREAS FROM SATELLITE IMAGES IN NIGERIA. *M.Sc. Dissertation*.
- Fassnacht, F. E., Vohland, M., Vierling, L. A., & Rahlf, J. (2014). Forest structure and biomass in complex landscapes: Integrating remote sensing and forest inventory. *Remote Sensing of Environment*, 141, 25-37. <https://doi.org/10.1016/j.rse.2013.10.024>
- Fu, S., Wang, L., Khalil, U., Cheema, A. H., Ullah, I., Aslam, B., Tariq, A., Aslam, M., & Alarifi, S. S. (2024). Prediction of surface urban heat island based on predicted consequences of urban sprawl using deep learning: A way forward for a sustainable environment. *Physics and Chemistry of the Earth*, 135, 103682. <https://doi.org/https://doi.org/10.1016/j.pce.2024.103682>
- Ganjirad, M. (2024). Google Earth Engine-based mapping of land use and land cover for weather forecast models using Landsat 8 imagery. *Ecological Informatics*, 80, 102498. <https://doi.org/10.1016/j.ecoinf.2024.102498>
- Gaona Obando, O. J. (2019a). SATISFACCIÓN DE LOS CLIENTES EN LA PRESTACIÓN DE SERVICIOS EN EL BANCO DE LA NACIÓN AGENCIA CUSCO, AÑO 2019. *Tesis de Maestría, Universidad Nacional San Antonio Abad del Cusco*, 1-67.
- Gaona Obando, O. J. (2019b). Satisfacción de los clientes en la prestación de servicios en el Banco de la Nación Agencia Cusco, año 2019 [Tesis para optar al Grado Académico de Maestro en Economía con mención en Proyectos de Inversión, Universidad Nacional de San Antonio Abad del Cusco]. *unsaac*.
- Garcia, A. D. B., Islam, M. S., Prudente, V. H. R., Sanches, I. D., & Cheng, I. (2025). Irrigated rice-field mapping in Brazil using phenological stage information and optical and microwave remote sensing. *Applied Computing and Geosciences*, 25, 100223. <https://doi.org/10.1016/j.acags.2025.100223>

- Gascon, F., Hoersch, C., Sy, O., Livens, T., D'Oria, M., Hoersch, B., Mavrocordatos, C., Martimort, P., & Bezy, J.-L. (2015). Sentinel-2: The Optical Imaging Mission for Copernicus. *ESA Bulletin*, (161).
- geeksforgeeks. (2024). Artificial Neural Networks and its Applications [Último acceso: 05 de marzo de 2025]. <https://www.geeksforgeeks.org/artificial-neural-networks-and-its-applications/>
- GeneSys Elektronik GmbH. (2024). Position information and their coordinate systems WGS84 and ETRS [Última actualización: 25 de marzo de 2024]. Consultado el 23 de julio de 2025, desde <https://genesys-offenburg.de/support/application-aids/gnss-basics/position-information/>
- GeoGPS Perú. (2015, septiembre). MTC - Red Vial Nacional, Departamental y Vecinal (Actualizado) [Accessed: 2025-10-15]. <https://www.geogpsperu.com/2015/09/mtc-red-vial-nacional-descarga-gratis.html>
- GeoGPS Perú. (2020, octubre). Ríos del Perú (Red Hidrográfica) - Descargar Shapefile [Accessed: 2025-10-15]. <https://www.geogpsperu.com/2020/10/rios-del-peru-red-hidrografica.html>
- GeoGPS Perú. (2021, agosto). Descargar Colegios a Nivel Nacional del Perú (Shapefile) [Accessed: 2025-10-15]. <https://www.geogpsperu.com/2021/08/descargar-colegios-nacionalparticular.html>
- Gharbia, S., Abd, S., Gill, L., Johnston, P., & Pilla, F. (2016). Land use scenarios and projections simulation using an integrated GIS cellular automata algorithms. *Modeling Earth Systems and Environment*, 2, 151. <https://doi.org/10.1007/s40808-016-0210-y>
- Gobierno del Perú. (2025). Fechas de las estaciones astronómicas en el Perú [Último cambio: 14 enero 2024].
- Gobierno Regional del Cusco. (2022). Plan de Desarrollo Regional Concertado de Cusco al 2033 (PDRC 2033) - Tomo 2: Propuesta de Visión, Objetivos, Indicadores y Acciones Estratégicas [Documento de Planificación Estratégica Regional]. *unsaac*.
- Google Developers. (2023). Google Earth Engine Playground [Consultado el 28 de julio de 2025]. <https://developers.google.com/earth-engine/guides/playground?hl=es-419>

- Google Earth Engine Developers. (2025). Earth Engine: Projections [Accessed: 2025-10-26].
- Google Research. (2025). Colab: a hosted Jupyter Notebook service [Accessed: 2025-10-26].
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2017.06.031>
- Gündüz, H. İ. (2025). Land-Use Land-Cover Dynamics and Future Projections Using GEE, ML, and QGIS-MOLUSCE: A Case Study in Manisa. *Sustainability*, 17(4). <https://doi.org/10.3390/su17041363>
- Guo, J., & Shen, B. (2024). Past dynamics and future prediction of the impacts of land use cover change and climate change on landscape ecological risk across the Mongolian plateau. *Journal of Environmental Management*, 355, 120365. <https://doi.org/10.1016/j.jenvman.2024.120365>
- Guria, R., Mishra, M., da Silva, R. M., Mishra, M., & Santos, C. A. G. (2024). Predicting forest fire probability in Similipal Biosphere Reserve (India) using Sentinel-2 MSI data and machine learning. *Remote Sensing Applications: Society and Environment*, 36, 101311. <https://doi.org/10.1016/j.rsase.2024.101311>
- Guth, P. L., Van Niekerk, A., Grohmann, C. H., Muller, J.-P., Hawker, L., Florinsky, I. V., Gesch, D., Reuter, H. I., Herrera-Cruz, V., Riazanoff, S., López-Vázquez, C., Carabajal, C. C., Albinet, C., & Strobl, P. (2021). Digital Elevation Models: Terminology and Definitions. *Remote Sensing*, 13(18). <https://doi.org/10.3390/rs13183581>
- Hagen, A. (2003). A fuzzy-set approach to assessing similarity of categorical maps. *International Journal of Geographical Information Science*, 17(3), 235-249. <https://doi.org/10.1080/13658810210157822>
- Hanh, L. N., Lang, L. P. C., Hang, P. A., An, N. V., & Son, N. H. (2025). A novel approach in comparing the performance of bivariate statistical methods, boosting, and stacking models in flood susceptibility assessment. *Journal of Environmental Management*, 387, 125670. <https://doi.org/10.1016/j.jenvman.2025.125670>



- Haripavan, N., Dey, S., & Chandana, C. H. M. (2025). Integration of geospatial techniques and machine learning in land parcel prediction. *Geosystems and Geoenvironment*, 4, 100371. <https://doi.org/10.1016/j.geogeo.2025.100371>
- Haydar, M., Rafi, A. H., Sadia, H., & Hossain, M. T. (2024). Data driven forest fire susceptibility mapping in Bangladesh. *Ecological Indicators*, 166, 112264. <https://doi.org/10.1016/j.ecolind.2024.112264>
- Hidalgo-García, D., & Arco-Díaz, J. (2022). Modeling the Surface Urban Heat Island (SUHI) to study of its relationship with variations in the thermal field and with the indices of land use in the metropolitan area of Granada (Spain). *Sustainable Cities and Society*, 87, 104166. <https://doi.org/10.1016/j.scs.2022.104166>
- Huaman Conza, M. (2019). El Plan de Negocios y su Incidencia en la Comercialización de Queso Fresco de la Asociación de Productores Agropecuarios de Cusipata Provincia de Quispicanchi 2018 [Tesis para optar el Título Profesional de Licenciada en Administración, Universidad Nacional de San Antonio Abad del Cusco]. *unsaac*.
- Huaman Conza, M., & Huaman Gaspar, R. (2019). DINERO ELECTRONICO COMO MECANISMO DE INCLUSION FINANCIERA DE LAS MYPES DE LA PROVINCIA DEL CUSCO - 2017. *Tesis de Titulo Profesional, Universidad Nacional de San Antonio Abad del Cusco*, 1-240.
- Huaman Conza, R., Mirian y Huaman Gaspar. (2019). Dinero Electrónico como Mecanismo de Inclusión Financiera de las MYPES de la Provincia del Cusco - 2017 [Tesis para optar el Título Profesional de Licenciados en Administración, Universidad Nacional de San Antonio Abad del Cusco]. *unsaac*.
- Huaman Gaspar, R., & Visa Quispe, F. M. (2019). Valoración de la Inclusión Financiera de los Comerciantes de las Ferias Itinerantes de la Ciudad del Cusco - 2017 [Tesis para optar el Título Profesional de Licenciados en Administración, Universidad Nacional de San Antonio Abad del Cusco]. *unsaac*.
- Huaraca Yucra, I., & Surco Vega, E. (2019). Estudio de los Riesgos Eléctricos en el Área Urbana de la Ciudad del Cusco.

- Huarcaya Fernandez Baca, P. O., & Villalba Velasque, J. (2016). DETERMINACIÓN DE LOS ASPECTOS HIDROLÓGICOS E HIDROGEOLÓGICOS DE LA CUENCA DEL RIO CCOCHAPUNQO EN LA PROVINCIA DE ANTA.
- Hussain, S., Qaisrani, S. A., Tariq, A., Mubeen, M., & Ullahd, S. (2025). Assessing past, present, and simulated future prediction of land use land cover changes using CA-Markov chain models with Satellite data. *Results in Engineering*, 26, 105585. <https://doi.org/10.1016/j.rineng.2025.105585>
- IBM. (2025). ¿Qué es el random forest? [Último acceso: 18 de febrero de 2025]. <https://www.ibm.com/mx-es/topics/random-forest>
- Instituto Nacional de Estadística e Informática (INEI). (2024, mayo). Compendio Estadístico Regional Cusco 2024 [Elaborado por la Oficina Departamental de Estadística e Informática Cusco, con la participación del Comité de Coordinación Departamental de Estadística - CCOIED-Cusco]. <https://cdn.www.gob.pe/uploads/document/file/7596201/6445207-compendio-estadistico-cusco-2024.pdf>
- Ishtiaque, A., Murthy, M. S. R., Mir, M. A. H., & Khan, M. D. U. H. (2021). Land use/cover change and future prediction by integrating the ML and CA-Markov model in the Bagerhat District of Bangladesh. *Environmental Monitoring and Assessment*, 193(252). <https://doi.org/10.1007/s10661-021-09022-z>
- Kamran, J. G., Faizah, C. R., & Shuib, R. (2024). Application of Cellular Automata and Markov Chain model for urban green infrastructure in Kuala Lumpur, Malaysia. *Regional Sustainability*, 5(4), 100179.
- Kandulna, W., Jain, M. K., Chugh, Y. P., & Agarwal, S. (2025). Spatial Variability of Land Surface Temperature of a Coal Mining Region Using a Geographically Weighted Regression Model: A Case Study. *Land*, 14(4), 696. <https://doi.org/10.3390/land14040696>
- Karurung, W. S., Lee, K., & Lee, W. (2025). Assessment of forest fire vulnerability prediction in Indonesia: Seasonal variability analysis using machine learning techniques. *International Journal of Applied Earth Observation and Geoinformation*, 138, 104435. <https://doi.org/10.1016/j.jag.2025.104435>

- Kashani, S. B., Fadavi, M., & Azizifar, V. (2025). Presenting the urban physical construction development model using fuzzy multi-criteria decision-making and CA-Markov supported by remote sensing techniques. *Environmental Challenges*, 19, 101128. <https://doi.org/10.1016/j.envc.2025.101128>
- Khalid, M., Ahmad, M., Javed, M., & Ahmad, S. (2023). Modeling future urban network capacity and land use/land cover simulation using GEE and remote sensing data. *Arabian Journal of Geosciences*. <https://doi.org/10.1007/s12517-023-11801-7>
- Khalid, W., Shamim, S. K., & Ahmad, A. (2024). Synergistic approach for land use and land cover dynamics prediction in Uttarakhand using cellular automata and Artificial neural network. *Geomatica*, 76(2), 100017. <https://doi.org/https://doi.org/10.1016/j.geomat.2024.100017>
- Khan, M., Qasim, M., Tahir, A. A., & Farooqi, A. (2023). Machine learning-based assessment and simulation of land use modification effects on seasonal and annual land surface temperature variations. *Heliyon*, 9(12). <https://doi.org/10.1016/j.heliyon.2023.e23043>
- Khan, M. S., Ullah, S., & Chen, L. (2021). Comparison on Land-Use/Land-Cover Indices in Explaining Land Surface Temperature Variations in the City of Beijing, China. *Land*, 10(10), 1018. <https://doi.org/10.3390/land10101018>
- Khan Academy. (2024). *Light and the electromagnetic spectrum*. Consultado el 18 de octubre de 2025, desde <https://www.khanacademy.org/science/in-in-class-12th-physics-india/in-in-electromagnetic-waves/in-in-phy-cbse-introduction-to-electromagnetic-waves/a/light-and-the-electromagnetic-spectrum>
- Khosravi, I. (2025). Advancements in crop mapping through remote sensing: A comprehensive review of concept, data sources, and procedures over four decades. *Remote Sensing Applications: Society and Environment*, 38, 101527. <https://doi.org/10.1016/j.rsase.2025.101527>
- Kondum, F., Rowshon, M., Luqman, C., Hasfalina, C., & Zakari, M. (2024). Change analyses and prediction of land use and land cover changes in Bernam River Basin, Malaysia. *Remote Sensing Applications: Society and Environment*, 36, 101281. <https://doi.org/10.1016/j.rsase.2024.101281>

- Krivoguz, D. (2024). Geo-spatial analysis of urbanization and environmental changes with deep neural networks: Insights from a three-decade study in Kerch peninsula. *Ecological Informatics*, 80, 102513. <https://doi.org/10.1016/j.ecoinf.2024.102513>
- Laar, C., Annan, K. B. K., Gibrilla, A., Kusi-Afrakoma, Z., & Korkor-Asante, O. (2024). Assessing the impact of land use and land cover change on the Densu Delta wetland using Markov chain modeling and artificial neural networks. *Environmental Challenges*, 17, 101018. <https://doi.org/10.1016/j.envc.2024.101018>
- Li, J., Yan, Y., & Stouffs, R. (2024). Simulation and prediction of daytime surface urban heat island intensity under multiple scenarios via fully connected neural network. *Sustainable Cities and Society*, 116, 105922. <https://doi.org/10.1016/j.scs.2024.105922>
- Liang, S., He, T., Huang, J., Jia, A., Zhang, Y., Cao, Y., Chen, X., Chen, X., Cheng, J., Jiang, B., Jin, H., Li, A., Li, S., Li, X., Liu, L., Liu, X., Ma, H., Ma, Y., Song, D.-X., . . . Song, L. (2024). Advancements in high-resolution land surface satellite products: A comprehensive review of inversion algorithms, products and challenges. *Science of Remote Sensing*, 10, 100152. <https://doi.org/10.1016/j.srs.2024.100152>
- Lin, Z., & Peng, S. (2024). Patterns of change, driving forces and future simulation of LULC in the Fuxian Lake Basin based on the IM-RF-Markov-PLUS framework. *Sustainable Futures*, 8, 100289. <https://doi.org/10.1016/j.sftr.2024.100289>
- Liu, G., Liu, H., Liu, Y., & Lv, Z. (2024). Exploring the drivers and dynamics of urban waters: A case study of Wuhan from 1980 to 2060. *Ecological Indicators*, 167, 112625. <https://doi.org/https://doi.org/10.1016/j.ecolind.2024.112625>
- Liu, Y., An, Z., & Ming, Y. (2024a). Simulating influences of land use/land cover composition and configuration on urban heat island using machine learning. *Sustainable Cities and Society*, 108, 105482. <https://doi.org/https://doi.org/10.1016/j.scs.2024.105482>
- Liu, Y., An, Z., & Ming, Y. (2024b). Simulating influences of land use/land cover composition and configuration on urban heat island using machine learning. *Sustainable Cities and Society*, 108, 105482. <https://doi.org/10.1016/j.scs.2024.105482>
- Lukas, P., Melesse, A. M., & Kenea, T. T. (2024). Predicting reservoir sedimentation using multilayer perceptron - Artificial neural network model with measured and forecasted

- hydrometeorological data in Gibe-III reservoir, Omo-Gibe River basin, Ethiopia. *Journal of Environmental Management*, 359, 121018. <https://doi.org/10.1016/j.jenvman.2024.121018>
- Lulichac Cayhuari, Y. J., & Miranda Mango, L. W. (2017). IMPACTO EN LA COMUNICACION INTERPERSONAL Y EFECTO EN LA GRAMATICA PRODUCIDOS POR EL USO DE LA MENSAJERIA MOVIL EN ESTUDIANTES DEL NIVEL SECUNDARIO DE LA I E P N P ARTURO PALOMINO RODRIGUEZ. *Tesis de Licenciatura, Universidad Nacional de San Antonio Abad del Cusco*, 1-175.
- Luo, T., & Chen, M. (2025). Advancements in supervised machine learning for outdoor thermal comfort: A comprehensive systematic review of scales, applications, and data types. *Energy & Buildings*, 329, 115255. <https://doi.org/10.1016/j.enbuild.2024.115255>
- Maddah, S., Khosravi, K., Jun, C., Bateni, S. M., Kim, D., & Liang, S. (2025). Assessment of a family of recurrent neural network models for flood susceptibility Mapping: An explainable glass-box approach. *Engineering Applications of Artificial Intelligence*, 160, 111867. <https://doi.org/10.1016/j.engappai.2025.111867>
- Mamani, E. M., & Cutipa, A. J. L. (2024). Instrumento Técnico Normativo de la Zona de Reglamentación Especial del Sector de Patapata - Distrito de San Jerónimo.
- Manning, C. G. (2018, septiembre). What are the spectrum band designators and bandwidths? [Page Last Updated: Jul 25, 2023]. *National Aeronautics, Space Administration (NASA)*.
- Mansour, S., Alotaibi, O., & Alnasrallah, M. (2025). Geospatial modelling of urban expansion effects on the land ecosystems in Kuwait using random forest and cellular automata. *Journal of Urban Management*. <https://doi.org/10.1016/j.jum.2025.05.011>
- Marey, A., Zou, J., Goubran, S., Wang, L. L., & Gaur, A. (2025). Urban morphology impacts on urban microclimate using artificial intelligence: a review. *City and Environment Interactions*, 28, 100221. <https://doi.org/10.1016/j.cei.2024.100221>
- Mas, J.-F., Ochoa-Gaona, O., Palma-Prieto, M., González-Mora, M., López-Sánchez, A., et al. (2019). Análisis Espacial de Vecindad Continua y regresión logística en el modelado espacial de probabilidad de ocurrencia de deslizamientos. *Investigaciones Geográficas*,

2019(98). [https://www.scielo.org.mx/scielo.php?script=sci\\_arttext%5C&pid=S0188-46112019000100006](https://www.scielo.org.mx/scielo.php?script=sci_arttext%5C&pid=S0188-46112019000100006)

Mazroa, A. A., Maashi, M., Kouki, F., Othman, K. M., & Salih, N. (2024). An analysis of urban sprawl growth and prediction using remote sensing and machine learning techniques. *Journal of South American Earth Sciences*, 142, 104988. <https://doi.org/10.1016/j.jsames.2024.104988>

M'Barek, S. A. (2024). The combined impact of climate change scenarios and land use changes on water resources in a semi-arid watershed. *Scientific African*, 25, e02319. <https://doi.org/10.1016/j.sciaf.2024.e02319>

Medianovskyi, K., Malakauskas, A., Lakstutiene, A., & Yahia, S. B. (2023). Interpretable machine learning for heterogeneous treatment effect estimators with Double ML: a case of access to credit for SMEs. *Procedia Computer Science*, 225, 2163-2172. <https://doi.org/10.1016/j.procs.2023.10.207>

Mehdiyev, N., Majlatow, M., & Fettke, P. (2025a). Integrating permutation feature importance with conformal prediction for robust Explainable Artificial Intelligence in predictive process monitoring. *Engineering Applications of Artificial Intelligence*, 149, 110363. <https://doi.org/10.1016/j.engappai.2025.110363>

Mehdiyev, N., Majlatow, M., & Fettke, P. (2025b). Integrating permutation feature importance with conformal prediction for robust Explainable Artificial Intelligence in predictive process monitoring. *Engineering Applications of Artificial Intelligence*, 149, 110363. <https://doi.org/10.1016/j.engappai.2025.110363>

Mesones La Rosa, R. A. (2022). Generación y Gestión de Residuos Sólidos Domiciliarios en la Municipalidad Provincial del Cusco 2015-2018 [Tesis para optar el Grado Académico de Maestro en Administración, Universidad Nacional de San Antonio Abad del Cusco]. *unsaac*.

Mhangara, P., Gidey, E., & Manjoo, R. (2024). Analysis of urban sprawl dynamics using machine learning, CA-Markov chain, and the Shannon entropy model: a case study in Mbombela City, South Africa. *Environmental Systems Research*, 13(1), 17. <https://doi.org/10.1186/s40068-024-00348-5>

- Ministerio del Ambiente (MINAM). (2015). Mapa Nacional de Cobertura Vegetal: Memoria Descriptiva. <https://www.minam.gob.pe/patrimonio-natural/wp-content/uploads/sites/6/2013/10/MAPA-NACIONAL-DE-COBERTURA-VEGETAL-FINAL.compressed.pdf>
- Mondal, K., Bandyopadhyay, S., & Karmakar, S. (2023). Framework for global sensitivity analysis in a complex 1D-2D coupled hydrodynamic model: Highlighting its importance on flood management over large data-scarce regions. *Journal of Environmental Management*, 332, 117312. <https://doi.org/10.1016/j.jenvman.2023.117312>
- Mozilla Developer Network. (2025). *JavaScript | MDN Web Docs* [Documentación oficial. Accedido: 21 de noviembre de 2025]. Mozilla. <https://developer.mozilla.org/en-US/docs/Web/JavaScript>
- Naeem, M., Zhang, Y., Tian, X., Miao, P., Li, C., Xu, Z., Wang, L., Mumtaz, F., Tang, Z., & He, S. (2025). Assessing and predicting Bojiang lake area and LULC changes from 2000 to 2045. *Journal of Hydrology: Regional Studies*, 58, 102216. <https://doi.org/10.1016/j.ejrh.2025.102216>
- Nam, K., Lee, Y., Lee, S., Kim, S., & Zhang, S. (2025). Explainable Artificial Intelligence (XAI) for Flood Susceptibility Assessment in Seoul: Leveraging Evolutionary and Bayesian AutoML Optimization. *Remote Sensing*, 17(13), 2244. <https://doi.org/10.3390/rs17132244>
- National Geospatial-Intelligence Agency. (2025). World Geodetic System 1984 (WGS 84) [Última modificación: 23 de junio de 2025]. Consultado el 23 de julio de 2025, desde <https://earth-info.nga.mil/index.php?dir=wgs84%5C&action=wgs84>
- National Institute of Standards and Technology (NIST). (2024). The electromagnetic spectrum [Image credit: N. Hanacek/NIST]. <https://www.nist.gov/image/electromagnetic-spectrum>
- Nguyen, H. H., Kim, H., Crow, W., Yueh, S., Wagner, W., Lei, F., Wigneron, J.-P., & Colliander. (2025). From theory to hydrological practice: Leveraging CYGNSS data over seven years for advanced soil moisture monitoring. *Remote Sensing of Environment*, 316, 114509.



- Nguyen, H. D., Nguyen, Q.-H., Dang, D. K., Van, C. P., & Truong, Q. H. (2024). A novel flood risk management approach based on future climate and land use change scenarios. *Science of the Total Environment*, 921, 171204. <https://doi.org/10.1016/j.scitotenv.2024.171204>
- Ovejero-Campos, A. C. (2021). Modelado de expansión urbana mediante autómatas celulares y redes neuronales [DOI: <https://doi.org/10.4995/raet.2021.16113>]. *Revista Universitaria de Geografía*, (59), 39-51. <https://www.redalyc.org/journal/3832/383271068004/html/>
- Pande, C. B., Egbueri, J. C., Costache, R., Sidek, L. M., Wang, Q., Alshehri, F., Din, N. M., Gautam, V. K., & Pal, S. C. (2024). Predictive modeling of land surface temperature (LST) based on Landsat-8 satellite data and machine learning models for sustainable development. *Journal of Cleaner Production*, 444, 141035. <https://doi.org/10.1016/j.jclepro.2024.141035>
- Paredes Catunta, J. C. (2019). *Derecho a Gozar de un Ambiente Equilibrado y Adecuado para el Desarrollo de la Vida. Estudio del Caso Rio Huatanay. Cusco-2016* [Tesis de Maestría en Derecho mención Derecho Constitucional y Procesal Constitucional]. Universidad Nacional de San Antonio Abad del Cusco.
- Patel, A., & Vyas, D. (2024). Novel approach for the LULC change detection using GIS & Google Earth Engine through spatiotemporal analysis to evaluate the urbanization growth of Ahmedabad city. *Results in Engineering*, 21, 101788. <https://doi.org/10.1016/j.rineng.2024.101788>
- Patle, S. (2024). Urban fragmentation approach for assessing thermal environment dynamics: A case study of semi-arid city from a comfort perspective. *Urban Climate*, 53, 101784. <https://doi.org/10.1016/j.uclim.2023.101784>
- Penfound, E., & Vaz, E. (2024). Modelling future wetland loss with land use landcover change simulation in the Greater Toronto and Hamilton Area: The importance of continued greenbelt development restrictions. *Habitat International*, 143, 102974. <https://doi.org/10.1016/j.habitatint.2023.102974>
- Pham, Q. B., & Ali, S. A. (2024). Multi-spectral remote sensing and GIS-based analysis for decadal land use land cover changes and future prediction using random forest tree and



- artificial neural network. *Advances in Space Research*, 74, 17-47. <https://doi.org/10.1016/j.asr.2024.03.027>
- Pontius, R. (2000). Quantification Error versus Location Error in Comparison of Categorical Maps. *Photogrammetric Engineering & Remote Sensing*, 66, 1011-1016.
- Pradhan, B., Lee, S., Dikshit, A., & Kim, H. (2023). Spatial flood susceptibility mapping using an explainable artificial intelligence (XAI) model. *Geoscience Frontiers*, 14(4), 101625. <https://doi.org/10.1016/j.gsf.2023.101625>
- Python Software Foundation. (2025). *Python 3.14.0 Documentation* [Accedido: 21 de noviembre de 2025]. Python Software Foundation. <https://docs.python.org/3/>
- Qamar, R., & Zardari, B. (2023). Artificial Neural Networks: An Overview. *Mesopotamian Journal of Computer Science*, 2023, 130-139. <https://doi.org/10.58496/MJCSC/2023/015>
- Quispe Quispe, F. d. M. (2021). Diagnóstico y Propuesta de Mejora de los Procesos del Área de Logística en la Empresa Constructora Sudamericana S.A.C., Cusco - 2021 [Facultad de Ingeniería de Procesos, Escuela Profesional de Ingeniería Química. Asesora: Mgt. Ysabel Almanza Loaiza].
- Risma, e. a. (2019). Prediction of land use and land cover (LULC) changes using CA-Markov model in Mamuju Subdistrict. *Journal of Physics: Conference Series*, 1341, 082033. <https://doi.org/10.1088/1742-6596/1341/8/082033>
- Robalino Trujillo, E. J. (2023). *DISEÑO DE OPERADORES DE VENTANA BASADOS EN MATEMÁTICA DIFUSA PARA LA SEGMENTACIÓN DE IMÁGENES MÉDICAS* [Tesis de Bioingeniería]. Universidad Nacional de Mar del Plata. Facultad de Ingeniería. <https://rinfi.fi.mdp.edu.ar/bitstream/handle/123456789/904/EJRobalinoTrujillo-TD-Bio-2023.pdf>
- Rotich, B., Ahmed, A., Kinyili, B., & Kipkulei, H. (2025a). Historical and projected forest cover changes in the Mount Kenya Ecosystem: Implications for sustainable forest management. *Environmental and Sustainability Indicators*, 26, 100628. <https://doi.org/10.1016/j.indic.2025.100628>

- Rotich, B., Ahmed, A., Kinyili, B., & Kipkulei, H. (2025b). Historical and Projected Forest Cover Changes in the Mount Kenya Ecosystem: Implications for Sustainable Forest Management. *Environmental and Sustainability Indicators*, 26C, 100628. <https://doi.org/10.1016/j.indic.2025.100628>
- Roy, B. (2021). A machine learning approach to monitoring and forecasting spatio-temporal dynamics of land cover in Cox's Bazar district, Bangladesh from 2001 to 2019. *Environmental Challenges*, 5, 100237. <https://doi.org/https://doi.org/10.1016/j.envc.2021.100237>
- Roy, K. C., Soren, D. D. L., & Biswas, B. (2024). Land-use/cover change and future prediction by integrating the ML techniques of random forest and CA-Markov chain model of the Ganges alluvial tract of Eastern India. *Environment, Development and Sustainability*, 1-28. <https://doi.org/10.1007/s10668-024-05545-x>
- Roy, S. K., Alam, M. T., Mojumder, P., Mondal, I., Kafy, A. .-. A., Dutta, M., Ferdous, M. N., Mamun, M. A. A., & Mahtab, S. B. (2023). Dynamic assessment and prediction of land use alterations influence on ecosystem service value: A pathway to environmental sustainability. *Environmental and Sustainability Indicators*, 21, 100319. <https://doi.org/10.1016/j.indic.2023.100319>
- Salamanca Oviedo Cabrera, V. (2016). ANALISIS DEL MERCADO Y CANALES DE COMERCIALIZACION DE PAPAS NATIVAS DEL DISTRITO DE VILCABAMBA, LA CONVENCION - CUSCO.
- Saoum, M. R., & Sarkar, S. K. (2024). Monitoring mangrove forest change and its impacts on the environment. *Ecological Indicators*, 159, 111666. <https://doi.org/10.1016/j.ecolind.2024.111666>
- Sarfo, I., Qiao, J., Yeboah, E., Pupilampu, D. A., Kwang, C., Fynn, I. E. M., Batame, M., Appea, E. A., Hagan, D. F. T., Ayelazuno, R. A., Boamah, V., & Sarfo, B. A. (2024). Meta-analysis of land use systems development in Africa: Trajectories, implications, adaptive capacity, and future dynamics. *Land Use Policy*, 144, 107261. <https://doi.org/10.1016/j.landusepol.2024.107261>

- Sarif, M. O., & Gupta, R. D. (2024). Predicting Prayagraj's Urbanization Trajectory using CA-ANN Modelling: Population Pressures and Land Use Dynamics. *Journal of Environmental Management*, 370, 122469. <https://doi.org/10.1016/j.jenvman.2024.122469>
- Senthilkumar, C., Alsolai, H., Allafi, R., & Arasid, M. A. (2025). Integrated image segmentation techniques for high-resolution coastal habitat mapping: Advancing remote sensing for coastal ecosystem assessment. *Journal of South American Earth Sciences*, 159, 105526. <https://doi.org/10.1016/j.jsames.2024.105526>
- Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI). (2015). Directiva N° 001-2015/SENAMHI-PREJ-OGEI: Directiva para el uso y actualización del catálogo de metadatos geográficos [Aprueba lineamientos para generar, registrar y publicar metadatos geográficos en GeoNetwork con base en normas ISO 19115/19139, Dublin Core y Perfil Básico de Metadatos del Perú]. Consultado el 23 de julio de 2025, desde <https://www.senamhi.gob.pe/load/file/00701SENA-196.pdf>
- Setiawan, O. (2024). Unraveling land use land cover change, their driving factors, and implication on carbon storage through an integrated modelling approach. *The Egyptian Journal of Remote Sensing and Space Sciences*, 27, 615-627. <https://doi.org/10.1016/j.ejrs.2024.08.002>
- Sharma, R., Pradhan, L., Kumari, M., Bhattacharya, P., Mishra, V. N., & Kumar, D. (2024). Spatio-Temporal Assessment of Urban Carbon Storage and Its Dynamics Using InVEST Model. *Land*, 13(9). <https://doi.org/10.3390/land13091387>
- Shen, Z., Wang, D., Arabameri, A., Santosh, M., Egbueri, J. C., & Arora, A. (2024). Landslide susceptibility assessment and mapping using new ensemble model. *Advances in Space Research*, 74(7), 2859-2882. <https://doi.org/https://doi.org/10.1016/j.asr.2024.06.018>
- Shrestha, M., Mitra, C., Rahman, M., & Marzen, L. (2023). Mapping and Predicting Land Cover Changes of Small and Medium Size Cities in Alabama Using Machine Learning Techniques. *Remote Sensing*, 15(1). <https://doi.org/10.3390/rs15010106>
- Sihue Huamani, J. C., & Choque Sanga, L. K. (2019). Valoración de la Percepción de la Informalidad de los Vendedores Minoristas de los Mercados de la Ciudad del

- Cusco, 2017 [Tesis para optar el Título Profesional de Licenciados en Administración, Universidad Nacional de San Antonio Abad del Cusco]. *unsaac*.
- Simou, E., Nomikou, P., Lykousis, V., Papanikolaou, D., & Vassilakis, E. (2014). Coastal hazard related to landslide distribution derived from morphotectonic analysis (SW Gulf of Corinth, Greece).
- Singh, R. K., Biradar, C. M., Behera, M. D., Prakash, A. J., Das, P., Mohanta, M. R., Krishna, G., Dogra, A., Dhyani, S. K., & Rizvi, J. (2024). Optimising carbon fixation through agroforestry: Estimation of aboveground biomass using multi-sensor data synergy and machine learning. *Ecological Informatics*, 79, 102408. <https://doi.org/10.1016/j.ecoinf.2023.102408>
- Soto Oscco, R. J., & Chalico Solis, R. E. (2017). Calidad de Servicios Turísticos en la Agencia de Viajes Limatours Cusco [Tesis para optar el Título Profesional de Licenciados en Turismo, Universidad Nacional de San Antonio Abad del Cusco]. *unsaac*.
- Suthar, G., Kaul, N., Khandelwal, S., & Singh, S. (2024). Predicting land surface temperature and examining its relationship with air pollution and urban parameters in Bengaluru: A machine learning approach. *Urban Climate*, 53, 101830. <https://doi.org/10.1016/j.uclim.2024.101830>
- Sweet, M. (2011, marzo). Google Earth versus Google Maps projections [Consultado el 26 de octubre de 2025]. *Montana Association of Geographic Information Professionals (MAGIP)*.
- Tadesse, T. (2024). Impacts of land use and cover changes on ecosystem service values from 1992 to 2052 in Gena District, Southwest Ethiopia. *Scientific African*, 24, e02244. <https://doi.org/10.1016/j.sciaf.2024.e02244>
- Taloor, A. K. (2024). Land use land cover simulations using integrated CA-Markov model in the Tawi Basin of Jammu and Kashmir India. *Geosystems and Geoenvironment*, 3, 100268. <https://doi.org/10.1016/j.geogeo.2024.100268>
- Tasan, M., Dąbrowska, J., Michałowska, K., & Uciechowska-Grakowicz, A. (2025). Long-term dynamics of urban heat island and hot spots in Wrocław: A 25-year satellite-based

- analysis using machine learning. *Sustainable Cities and Society*, 132, 106797. <https://doi.org/https://doi.org/10.1016/j.scs.2025.106797>
- Tayyab, M., Hussain, M., Zhang, J., Ullah, S., Tong, Z., Rahman, Z. U., Al-Aizari, A. R., & Al-Shaibah, B. (2024). Leveraging GIS-based AHP, remote sensing, and machine learning for susceptibility assessment of different flood types in peshawar, Pakistan. *Journal of Environmental Management*, 371, 123094. <https://doi.org/10.1016/j.jenvman.2024.123094>
- Tehrani, A. A., Veisi, O., kia, K., Delavard, Y., Bahrami, S., Sobhaninia, S., & Mehan, A. (2024). Predicting urban Heat Island in European cities: A comparative study of GRU, DNN, and ANN models using urban morphological variables. *Urban Climate*, 56, 102061. <https://doi.org/10.1016/j.uclim.2024.102061>
- Tiamgne, X. T., Domingo, D., Kalaba, F. K., Nyirenda, V. R., & Gómez, C. (2025). Multi-scenario land use simulation for fast human growth regions: A case study of Zambia's Solwezi mining district. *Resources Policy*, 107, 105661. <https://doi.org/10.1016/j.resourpol.2025.105661>
- Tola, B., & Deyassa, G. (2024). A modeling approach for evaluating and predicting the impacts of land use land cover changes on groundwater recharge in Walga Watershed, Upper Omo Basin, Central Ethiopia. *Journal of Hydrology: Regional Studies*, 51, 101659. <https://doi.org/10.1016/j.ejrh.2024.101659>
- Toscan, P. C., Seong, K., Jiao, J., Ribeiro, C. A. L. R., Carvalho, F. A. C., Oliveira, M. L., & Pereira, E. B. (2025). Impact of nature-based solutions (NBS) on urban surface temperatures and land cover changes using remote sensing and machine learning. *Remote Sensing Applications: Society and Environment*, 39, 101721. <https://doi.org/10.1016/j.rsase.2025.101721>
- Trujillo, G. T. (2019). El Espacio Público en la Ciudad del Cusco.
- Ttito Ocsa, N., & Mescoco Pumasupa, A. (2020). FACTORES QUE INFLUYEN EN LA DECISIÓN DE COMPRA DE VIVIENDA DEL SECTOR INMOBILIARIO EN LA CIUDAD DEL CUSCO, 2019 [El año de estudio es 2019, pero la fecha de publicación

- de la tesis es 2020.]. *Tesis de Título Profesional, Universidad Nacional de San Antonio Abad del Cusco*, 1-125.
- United Nations Office for Outer Space Affairs (UNOOSA). (2012). World Geodetic System 1984 (WGS ,84) [Accedido en julio de 2025]. [https://www.unoosa.org/pdf/icg/2012/template/WGS\\_84.pdf](https://www.unoosa.org/pdf/icg/2012/template/WGS_84.pdf)
- U.S. Geological Survey. (2023). What is a Geographic Information System (GIS)? [Consulted: 27 de julio de 2025]. <https://www.usgs.gov/faqs/what-a-geographic-information-system-gis>
- Vahid, R., & Aly, M. H. (2025). A Comprehensive Systematic Review of Machine Learning Applications in Assessing Land Use/Cover Dynamics and Their Impact on Land Surface Temperatures. *Urban Science*, 9, 234. <https://doi.org/10.3390/urbansci9070234>
- Valladares Herrera, J. P. (2023). MONITOREO Y PROYECCIÓN DEL CAMBIO DE COBERTURA FORESTAL POR IMPACTO ANTRÓPICO, APLICANDO TELEDETECCIÓN EN INAMBARI - MADRE DE DIOS AÑOS 2000 AL 2030. *Tesis para optar el Título Profesional de Ingeniero Geógrafo*.
- Vidhya, A. (2020). Google Colab Features, Best Practices Guide [Accessed: 2025-10-26]. *Analytics Vidhya Blog*. <https://www.analyticsvidhya.com/blog/2020/03/google-colab-machine-learning-deep-learning/>.
- Visser, H. (2004). *The MAP COMPARISON KIT: methods, software and applications* (inf. téc. N.º 550002005/2004). RIVM / Netherlands Environmental Assessment Agency. Bilthoven, The Netherlands.
- Vizcarra Olarte, E., & Chambi Palomino, W. (2019). Programa de Intervención para la Disminución del Estrés Laboral en los Docentes del Nivel Primario de la Institución Educativa José Antonio Encinas de la Ciudad del Cusco - 2018 [Tesis para optar el Título Profesional de Licenciados en Educación, Especialidad Educación Física, Universidad Nacional de San Antonio Abad del Cusco]. *unsaac*.

- Wahdatyar, R. (2024). Exploring the dynamics and future projections of land use land cover changes by exploiting geospatial techniques; A case study of the Kabul River Basin. *Heliyon*, 10, e39020. <https://doi.org/10.1016/j.heliyon.2024.e39020>
- Wahdatyar, R., Khokhar, M. F., Ahmad, S., Rahil, M. U., Stanikzai, M. A., Khan, J. A., & Kamran. (2024). Exploring the dynamics and future projections of land use land cover changes by exploiting geospatial techniques; A case study of the Kabul River Basin. *Heliyon*, 10(20). <https://doi.org/10.1016/j.heliyon.2024.e39020>
- Wang, N., Yue, Z., Tong, Z., Liu, Y., & Liu, Y. (2024). A multi-objective optimization framework for regional land-use allocation: Fully utilizing terrestrial vegetation to mitigate carbon emissions. *Journal of Cleaner Production*, 479, 144058. <https://doi.org/10.1016/j.jclepro.2024.144058>
- Wijayanto, A. W., Zalukhu, B. V., Putri, S., Wilantika, N., Yuniarto, B., Kurniawan, R., & Pratama, A. (2025). Deep Learning and Remote Sensing for Agricultural Land Use Monitoring: A Spatio-Multitemporal Analysis of Rice Field Conversion using Optical Satellite Images. *International Journal of Advances in Data and Information Systems*, 6, 45-62. <https://doi.org/10.59395/ijadis.v6i2.1385>
- Wu, J., Gan, W., & Yu, P. (2024). Geospatial Big Data: Survey and Challenges. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, PP, 1-14. <https://doi.org/10.1109/JSTARS.2024.3438376>
- Wu, X., Zhao, X., Chen, P., Zhu, B., & Cai, W. (2024). Assessing the effects of combined future climate and land use/cover changes on streamflow in the Upper Fen River Basin, China. *Journal of Hydrology: Regional Studies*, 53, 101853. <https://doi.org/10.1016/j.ejrh.2024.101853>
- Xiang, X. (2024). Modelling future land use land cover changes and their impacts on urban heat island intensity in Guangzhou, China. *Journal of Environmental Management*, 366, 121787. <https://doi.org/10.1016/j.jenvman.2024.121787>
- Xiang, X., Zhai, Z., Fan, C., Ding, Y., Ye, L., & Li, J. (2024). Modelling future land use land cover changes and their impacts on urban heat island intensity in Guangzhou,

- China. *Journal of Environmental Management*, 366, 121787. <https://doi.org/10.1016/j.jenvman.2024.121787>
- Xu, T., Zhou, D., & Li, Y. (2022). Integrating ANNs and Cellular Automata-Markov Chain to Simulate Urban Expansion with Annual Land Use Data. *Land*, 11(7). <https://doi.org/10.3390/land11071074>
- Xu, Z., Li, J., Cheng, S., Rui, X., Guan, H., Sharma, A., Erxleben, M., Zhao, Y., He, H., Chang, R., & Xu, L. L. (2025). Deep learning for wildfire risk prediction: Integrating remote sensing and environmental data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 227, 632-677. <https://doi.org/10.1016/j.isprsjprs.2025.06.002>
- y PNUD, I. (2004). Mapa de tipos de suelos de la ciudad del Cusco [En línea; accedido el 21 de octubre de 2025]. <https://sigrid.cenepred.gob.pe/sigridv3/documento/4244>
- Yadav, P. K., Mishra, V. N., Kumari, M., Kumar, A., Kumar, P., & Bhatlad, R. (2024). Spatially explicit simulation and forecasting of urban growth using weights of evidence based cellular automata model in a millennium city of India. *Physics and Chemistry of the Earth, Parts A/B/C*, 136, 103739. <https://doi.org/10.1016/j.pce.2024.103739>
- Yang, X., Hao, Y., Ding, H., Yu, C., Liu, J., Li, L., & Chen, J. (2025). Explainable Artificial Intelligence (XAI) framework using XGBoost and SHAP for assessing urban fire risk based on spatial distribution features. *International Journal of Disaster Risk Reduction*, 129, 105798. <https://doi.org/https://doi.org/10.1016/j.ijdr.2025.105798>
- Yang, Z., Dai, X., Lu, H., Liu, C., Nie, R., Zhang, M., Ma, L., Li, N., Liu, T., He, Y., Yang, Z., Qu, G., Li, W., & Wang, Y. (2023). Evaluation and prediction of water conservation of the Yellow river basin in Sichuan Province, China, based on Google Earth Engine and CA-Markov. *Heliyon*, 9(7). <https://doi.org/10.1016/j.heliyon.2023.e17903>
- Yang, Z., Xia, H., Guo, Z., Xie, Y., Liao, Q., Yang, W., Li, Q., Dong, C., & Si, M. (2024). Development and application of machine learning models for prediction of soil available cadmium based on soil properties and climate features. *Environmental Pollution*, 355, 124148. <https://doi.org/10.1016/j.envpol.2024.124148>
- Yañe Zuñiga, L. (2019). La construcción de viviendas y su incidencia en la Economía de la Ciudad Metropolitano Cusco Periodo 2010 - 2016.



- Yasin, K. H. (2024). Characterizing and modeling spatiotemporal trends in rangelands: Prosopis juliflora impact in middle Awash Basin, Ethiopia. *Journal of Environmental Management*, 371, 123336. <https://doi.org/10.1016/j.jenvman.2024.123336>
- Young, N., Anderson, R., Chignell, S., Vorster, A., Lawrence, R., & Evangelista, P. (2017). A survival guide to Landsat preprocessing. *Ecology*, 98, 920-932. <https://doi.org/10.1002/ecy.1730>
- Youssef, Y. M. (2024). Insight into land cover dynamics and water challenges under anthropogenic and climatic changes in the eastern Nile Delta: Inference from remote sensing and GIS data. *Science of the Total Environment*, 913, 169690. <https://doi.org/10.1016/j.scitotenv.2023.169690>
- Zafar, Z., Zubair, M., Zha, Y., Fahd, S., & Nadeem, A. A. (2024). Performance assessment of machine learning algorithms for mapping of land use/land cover using remote sensing data. *The Egyptian Journal of Remote Sensing and Space Sciences*, 27, 216-226. <https://doi.org/10.1016/j.ejrs.2024.03.003>
- Zafar, Z., Zubair, M., Zha, Y., Mehmood, M. S., Rehman, A., Fahd, S., & Nadeem, A. A. (2024). Predictive modeling of regional carbon storage dynamics in response to land use/land cover changes: An InVEST-based analysis. *Ecological Informatics*, 82, 102701. <https://doi.org/10.1016/j.ecoinf.2024.102701>
- Zambrano-Luna, B., Milne, R., & Wang, H. (2025). Cyanobacteria hot spot detection integrating remote sensing data with convolutional and Kolmogorov-Arnold networks. *Science of the Total Environment*, 960, 178271. <https://doi.org/10.1016/j.scitotenv.2024.178271>
- Zhang, M., Tan, S., Liang, J., Zhang, C., & Chen, E. (2024). Predicting the impacts of urban development on urban thermal environment using machine learning algorithms in Nanjing, China. *Journal of Environmental Management*, 356, 120560. <https://doi.org/10.1016/j.jenvman.2024.120560>
- Zhang, X., Qi, Y., Li, H., Wang, X., & Yin, Q. (2023). Assessing the response of non-point source nitrogen pollution to land use change based on SWAT model. *Ecological Indicators*, 158, 111391. <https://doi.org/10.1016/j.ecolind.2023.111391>

- Zhao, D., Miao, F., Chen, Y., Wu, Q., Feng, G., Chang, B., Su, H., Ren, P., Hao, C., Li, Z., Li, X., & Cai, J. (2025). Enhanced groundwater potential mapping using a GIS based chaotic sparrow search algorithm optimized weighted broad learning system: A case study of the Guozhuang spring region, northern China. *Journal of Hydrology: Regional Studies*, 61, 102708. <https://doi.org/10.1016/j.ejrh.2025.102708>
- Zhao, F., Yang, P., Gao, Q., Xia, L., Fan, L., & Hu, M. (2024). Effects of the long-term rice expansion on ecosystem carbon budget in the typical agricultural area of Northeast China. *Sustainable Production and Consumption*, 52, 613-623. <https://doi.org/10.1016/j.spc.2024.11.027>
- Zhou, Z., Xiao, J., Huang, L., Song, R., & Ren, P. (2025). Analysis of land use and carbon storage dynamics in the Panxi region, Southwest Sichuan, China: Spatiotemporal evolution and driving forces in multiple scenarios. *Environmental and Sustainability Indicators*, 28, 100914. <https://doi.org/10.1016/j.indic.2025.100914>

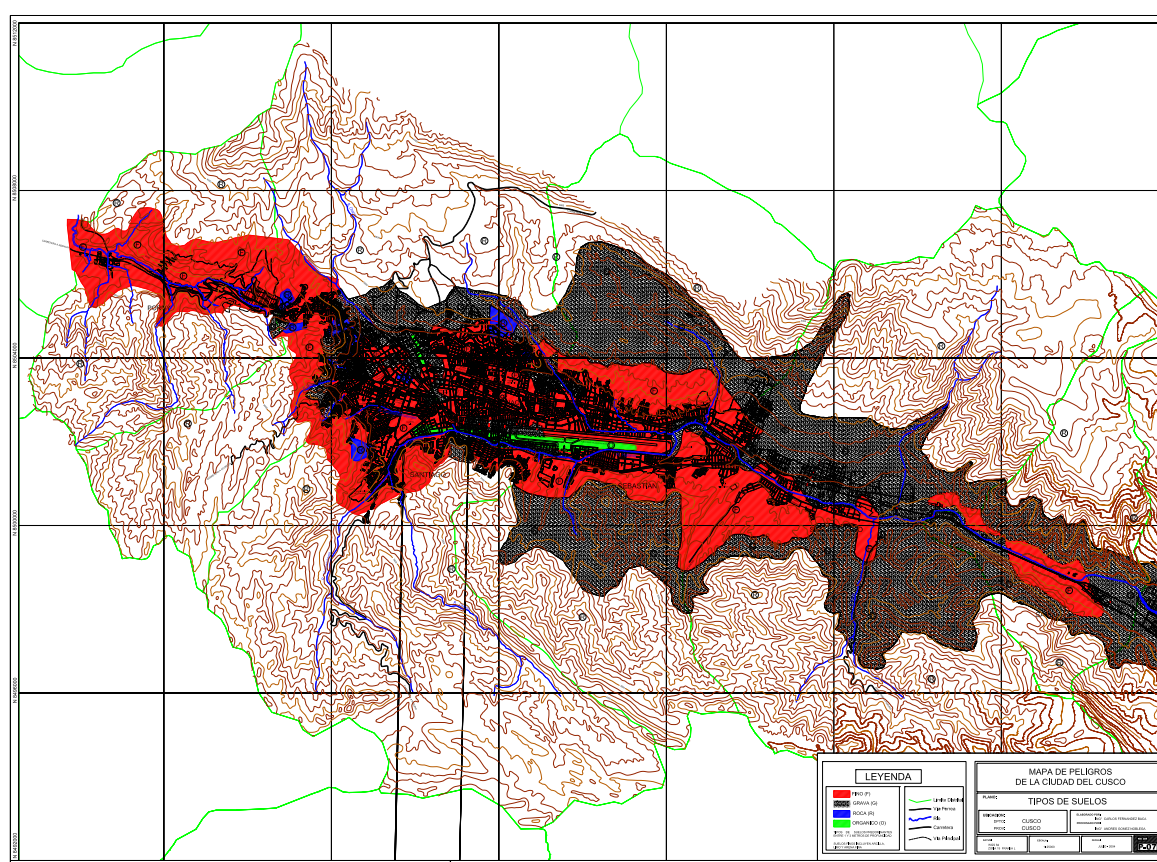
## Apéndices

### Apéndice A

#### Mapa de Peligros de la Ciudad del Cusco 2004

**Figura 88**

*Mapa de Peligros de la Ciudad del Cusco 2004*



## Apéndice B

### Validación de Fuente de Datos Geográficos: El Caso de la Provincia de Cusco

La determinación de la superficie oficial de la Provincia del Cusco presenta inconsistencias notables en la literatura técnica y académica. Si bien la cifra de 617 km<sup>2</sup> es la más frecuente en compendios estadísticos del INEI y antecedentes académicos (Ccopa Barrionuevo, 2019; Condori Juarez, 2019; Gaona Obando, 2019a, 2019b; M. Huaman Conza & Huaman Gaspar, 2019; R. Huaman Conza, 2019; Ttito Ocsa & Mescoco Pumasupa, 2020), existen reportes divergentes. Diversas investigaciones sitúan el área en 719 km<sup>2</sup> (M. Huaman Conza, 2019; Huaman Gaspar & Visa Quispe, 2019; Sihue Huamani & Choque Sanga, 2019; Vizcarra Olarte & Chambi Palomino, 2019), mientras que otros documentos técnicos sugieren valores inferiores como 543.08 km<sup>2</sup> (Mesones La Rosa, 2022; Soto Oscco & Chalico Solis, 2017), 529.21 km<sup>2</sup> (Gobierno Regional del Cusco, 2022) e incluso 523 km<sup>2</sup> (Castillo Alire & Gutierrez Kancha, 2019).

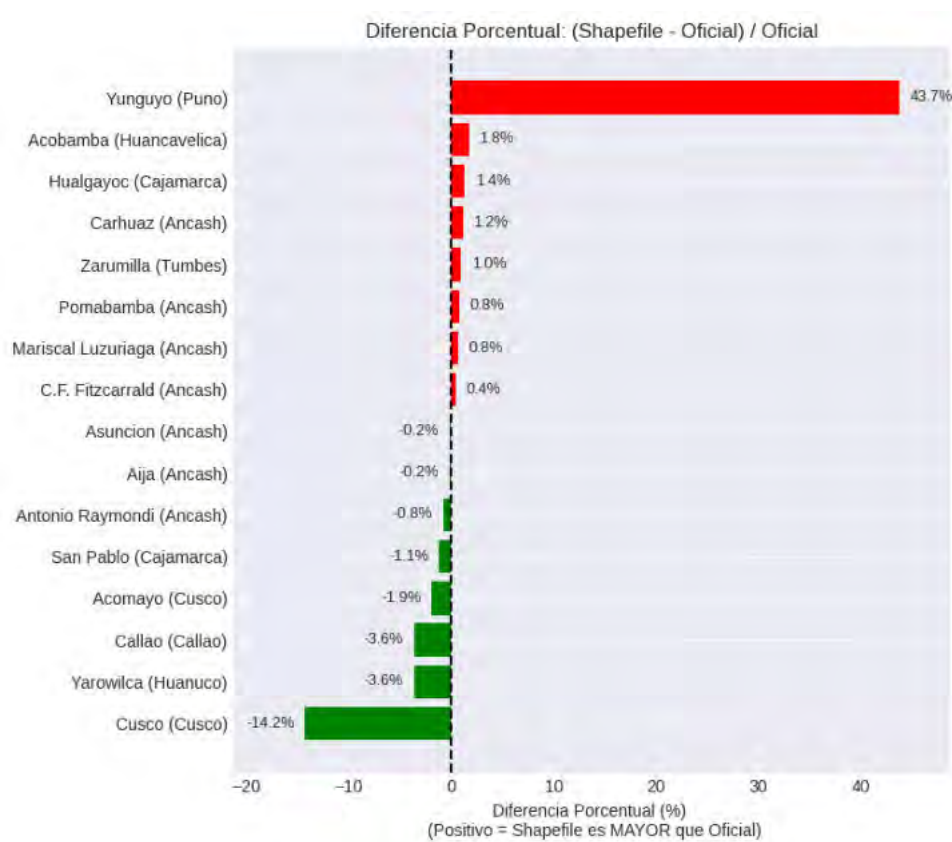
Ante esta variabilidad, se procedió a validar la fiabilidad de la fuente de datos primarios (*shapefile*) mediante un análisis comparativo. Se contrastó el área geométrica calculada del *shapefile* frente a datos oficiales para una muestra de 16 provincias de similar extensión (< 1000 km<sup>2</sup>).

Los resultados, resumidos en la Figura 89, evidencian una alta concordancia en la mayoría de los casos (13 de 16 provincias presentan diferencias menores al  $\pm 2\%$ ), lo que valida la precisión general de la cartografía vectorial utilizada.

Sin embargo, el análisis detectó anomalías críticas en las provincias de Yunguyo y Cusco. En el caso específico de Cusco, el valor documental frecuentemente citado (617 km<sup>2</sup>) excede en un 14.2 % al área geométrica real. Dado que el *shapefile* demostró ser preciso en los casos de control, se concluye que los datos documentales históricos contienen errores o desactualizaciones. Por consiguiente, esta investigación adopta el valor calculado de 529.208 km<sup>2</sup> como la superficie oficial del área de estudio, cifra que además es consistente con reportes técnicos recientes de zonificación regional (Gobierno Regional del Cusco, 2022).

**Figura 89**

*Diferencias porcentuales entre el área calculada (Shapefile) y la documentada por fuentes oficiales.*



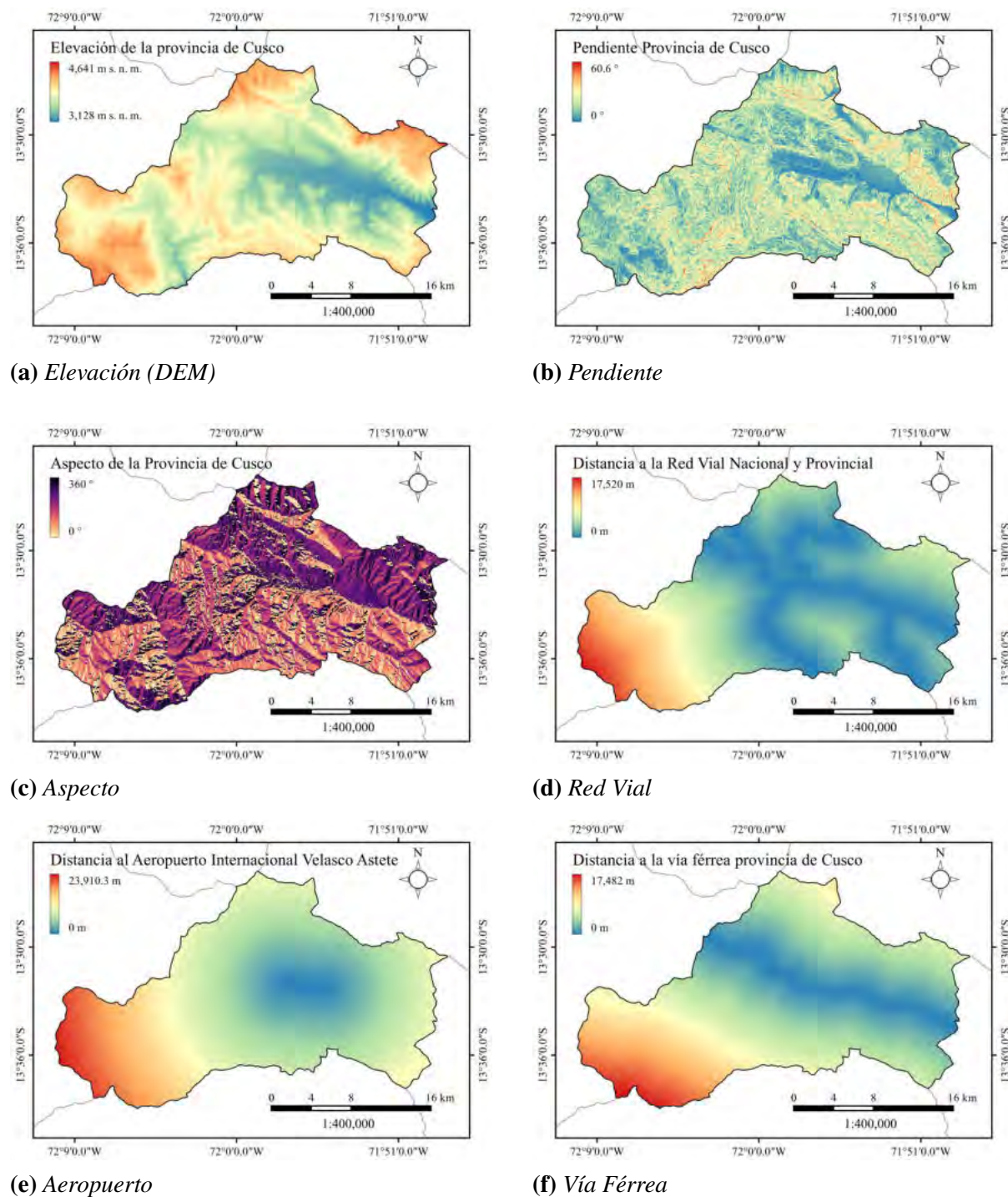
*Nota. Fuente: Elaboración propia.*

## Apéndice C

### Mapas de Predictores Espaciales (Drivers de Cambio)

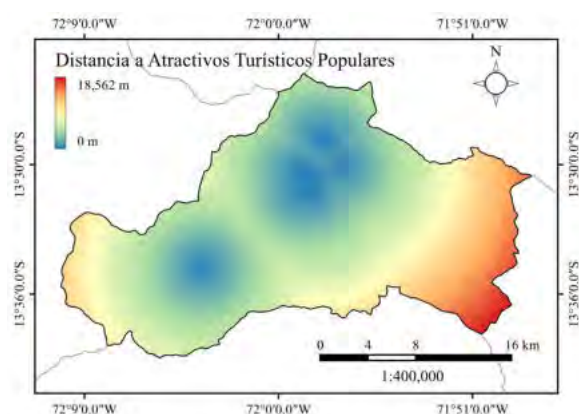
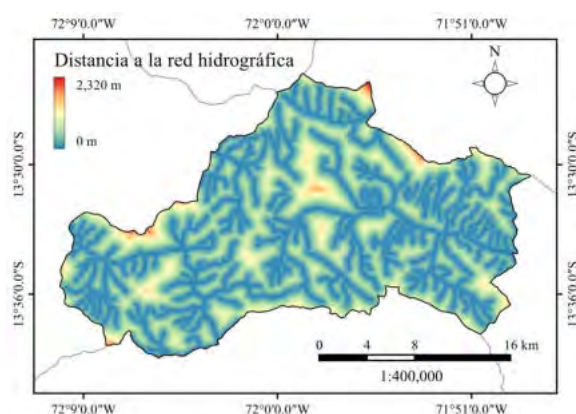
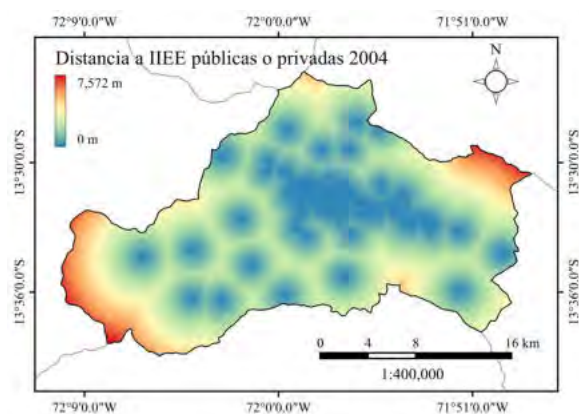
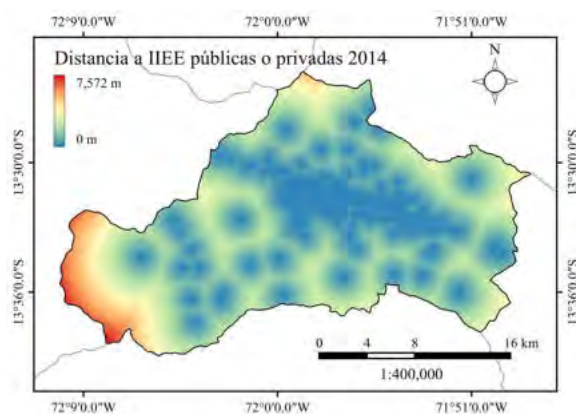
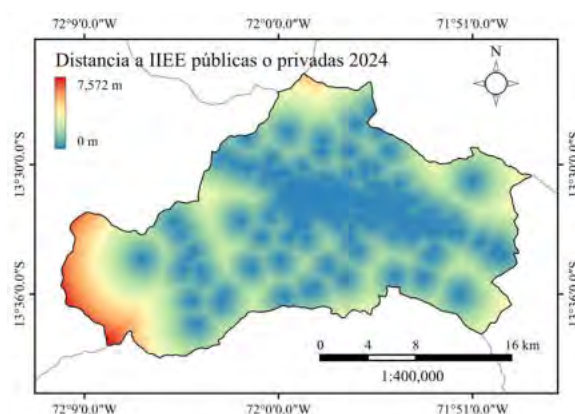
**Figura 90**

*Mapas de Predictores Espaciales utilizados en el modelo.*



(Continúa en la siguiente página...)



**Figura 90***Mapas de Predictores Espaciales (Continuación).***(g) Atractivos Turísticos****(h) Red Hidrográfica****(i) IIEE (2004)****(j) IIEE (2014)****(k) IIEE (2024)**

*Nota. Fuente: Elaboración propia a partir de datos del Modelo Digital de Elevación Copernicus GLO-30 (European Space Agency, 2021), red vial e infraestructura de GeoGPS Perú (GeoGPS Perú, 2015, 2021) y datos hidrográficos (GeoGPS Perú, 2020).*

## Apéndice D

### Principales Sitios de Interés Ofertados por Empresas en la Provincia de Cusco

**Tabla 32**

*Fuentes de Atractivos Turísticos en la Provincia de Cusco*

Nombre del Sitio	Enlaces de las Fuentes
Plaza de Armas del Cusco	<a href="http://inkanmilkyway.com">inkanmilkyway.com</a> , <a href="http://machu-picchu.org">machu-picchu.org</a> , <a href="http://expedia.com">expedia.com</a> , <a href="http://voyageperou.info">voyageperou.info</a> , <a href="http://viajeroscallejeros.com">viajeroscallejeros.com</a> , <a href="http://audiala.com">audiala.com</a>
Sacsayhuamán	<a href="http://inkanmilkyway.com">inkanmilkyway.com</a> , <a href="http://machu-picchu.org">machu-picchu.org</a> , <a href="http://expedia.com">expedia.com</a> , <a href="http://cosituc.gob.pe">cosituc.gob.pe</a> , <a href="http://boletomachupicchu.com">boletomachupicchu.com</a> , <a href="http://viajeroscallejeros.com">viajeroscallejeros.com</a>
Barrio de San Blas	<a href="http://inkanmilkyway.com">inkanmilkyway.com</a> , <a href="http://machu-picchu.org">machu-picchu.org</a> , <a href="http://expedia.com">expedia.com</a> , <a href="http://voyageperou.info">voyageperou.info</a> , <a href="http://viajeroscallejeros.com">viajeroscallejeros.com</a> , <a href="http://boletomachupicchu.com">boletomachupicchu.com</a>
Qorikancha (Templo del Sol)	<a href="http://expedia.com">expedia.com</a> , <a href="http://boletomachupicchu.com">boletomachupicchu.com</a> , <a href="http://freewalkingtoursperu.com">freewalkingtoursperu.com</a> , <a href="http://boletomachupicchu.com">boletomachupicchu.com</a> , <a href="http://voyageperou.info">voyageperou.info</a> , <a href="http://viajeroscallejeros.com">viajeroscallejeros.com</a>
Mercado Central de San Pedro	<a href="http://inkanmilkyway.com">inkanmilkyway.com</a> , <a href="http://machu-picchu.org">machu-picchu.org</a> , <a href="http://expedia.com">expedia.com</a> , <a href="http://voyageperou.info">voyageperou.info</a> , <a href="http://viajeroscallejeros.com">viajeroscallejeros.com</a>
Catedral del Cusco	<a href="http://machu-picchu.org">machu-picchu.org</a> , <a href="http://expedia.com">expedia.com</a> , <a href="http://boletomachupicchu.com">boletomachupicchu.com</a>
Cristo Blanco	<a href="http://machu-picchu.org">machu-picchu.org</a> , <a href="http://youtube.com">youtube.com</a> , <a href="http://voyageperou.info">voyageperou.info</a> , <a href="http://viajeroscallejeros.com">viajeroscallejeros.com</a>
Piedra de los doce ángulos	<a href="http://expedia.com">expedia.com</a> , <a href="http://voyageperou.info">voyageperou.info</a> , <a href="http://viajeroscallejeros.com">viajeroscallejeros.com</a>
Templo de San Blas	<a href="http://machu-picchu.org">machu-picchu.org</a> , <a href="http://expedia.com">expedia.com</a>
Templo de la Compañía de Jesús	<a href="http://machu-picchu.org">machu-picchu.org</a> , <a href="http://voyageperou.info">voyageperou.info</a>

*Continúa en la siguiente página...*



...viene de la página anterior

Nombre del Sitio	Enlaces de las Fuentes
Convento de Santa Catalina y Museo de Arte	<a href="http://boletomachupicchu.com">boletomachupicchu.com</a> , <a href="http://freewalkingtoursperu.com">freewalkingtoursperu.com</a>
Museo Inka	<a href="http://boletomachupicchu.com">boletomachupicchu.com</a> , <a href="http://freewalkingtoursperu.com">freewalkingtoursperu.com</a>
Museo de Arte Precolombino	<a href="http://boletomachupicchu.com">boletomachupicchu.com</a> , <a href="http://freewalkingtoursperu.com">freewalkingtoursperu.com</a>
Tambomachay	<a href="http://cosituc.gob.pe">cosituc.gob.pe</a> , <a href="http://viajeroscallejeros.com">viajeroscallejeros.com</a>
Puka Pukara	<a href="http://cosituc.gob.pe">cosituc.gob.pe</a> , <a href="http://viajeroscallejeros.com">viajeroscallejeros.com</a>
Qenqo	<a href="http://viajeroscallejeros.com">viajeroscallejeros.com</a>
Museo de Arte Religioso (Palacio Arzobispal)	<a href="http://boletomachupicchu.com">boletomachupicchu.com</a> , <a href="http://freewalkingtoursperu.com">freewalkingtoursperu.com</a>
Museo Histórico Regional de Cusco	<a href="http://boletomachupicchu.com">boletomachupicchu.com</a>
Acueducto Colonial de Sapan-tiana	<a href="http://machu-picchu.org">machu-picchu.org</a>
Mirador de San Cristóbal	<a href="http://machu-picchu.org">machu-picchu.org</a>
Museo de Arte Contemporáneo	<a href="http://boletomachupicchu.com">boletomachupicchu.com</a>
Iglesia y Convento de la Merced	<a href="http://freewalkingtoursperu.com">freewalkingtoursperu.com</a>
Museo Machu Picchu de la Casa Concha	<a href="http://boletomachupicchu.com">boletomachupicchu.com</a>
Centro Qosqo de Arte Nativo	<a href="http://cosituc.gob.pe">cosituc.gob.pe</a>
Farallones de Tecsecocha	<a href="https://www.scribd.com">scribd.com</a> , <a href="http://mincetur.gob.pe">mincetur.gob.pe</a>

## **Apéndice E**

### **Repositorio de Datos**

Debido a la extensión del material gráfico generado, se ha dispuesto un repositorio digital complementario. Este repositorio tiene como finalidad garantizar la reproducibilidad técnica del estudio y permitir al lector la inspección detallada de los resultados intermedios y finales en alta resolución, los cuales se presentan de forma sintetizada en este documento.

#### **Acceso al Repositorio y Visualización Web**

Todo el material suplementario se encuentra alojado públicamente y puede ser consultado a través de las siguientes plataformas:

##### **1. Repositorio de Archivos (Google Drive)**

Almacena los mapas iniciales, mosaicos libres de nubes, índices espectrales, variables predictoras y mapas de potencial de transición para las fases de validación y proyección. Contiene los resultados de clasificación histórica, validación (2024) y proyección (2034) en vistas de área completa y detalles ampliados, incluido los archivo GeoTIFF (.tif)

[https://drive.google.com/drive/folders/1WZxv24N5wEuXPruSdiYXY-Sp\\_dER06Du?usp=sharing](https://drive.google.com/drive/folders/1WZxv24N5wEuXPruSdiYXY-Sp_dER06Du?usp=sharing)

##### **2. Plataforma Web de Visualización**

Se ha desarrollado un sitio web complementario que permite la visualización de todos los mapas generados sin necesidad de descarga previa.

<https://lulc-cusco-2034.vercel.app/>



*Escanee para acceder a la visualización web.*

## Estructura de Directorios

La información en el repositorio se ha organizado siguiendo la secuencia lógica de los objetivos de la tesis, distribuida en las siguientes carpetas:

### 00\_Metodologia\_y\_Flujos

Contiene los diagramas de flujo detallados de los procesos de clasificación supervisada y modelado predictivo.

### 01\_Insumos\_Satelitales

Muestra las imágenes compuestas finales (mosaicos libres de nubes) utilizadas para los años 2004, 2014 y 2024 en color real (RGB), junto con ejemplos del proceso de enmascaramiento.

### 02\_Indices\_Espectrales

Colección completa de los mapas de índices calculados (NDVI, SAVI, NDBI, NDMI, BSI, UI) para cada año de estudio, utilizados como variables predictoras.

### 03\_Predictores\_Espaciales

Almacena los mapas ráster de las variables conductoras. Se incluye el manual visual detallado de los geoprocesos.

Nota sobre las Fuentes de Datos: Los insumos base fueron obtenidos mediante la solicitud de acceso a la información pública (Código: 79zh4rpnl) dirigida al INEI. Según la respuesta formal (Correo N°5757-2025-INEI/OTD-OEIN), la información utilizada (RENIPRESS y Directorio de Entidades) corresponde a registros administrativos de carácter público y portales de datos abiertos, de acuerdo con el DS N° 043-2001-PCM, garantizando la transparencia y disponibilidad para fines de investigación.

#### ■ Insumos Base y Correspondencia:

- *INEI*: Correo electrónico con información de: *Registro Nacional de Instituciones Prestadoras de Servicios de Salud - RENIPRESS.xlsx* y *entidades.xlsx*.

- *Municipalidad de San Sebastián - Cusco*: Información cartográfica técnica en formato CAD: *PLANO BASE REFERENCIAL SAN SEBASTIAN.dwg*.

■ Variables Generadas:

- *Topografía*: Elevación (DEM), Pendiente y Aspecto.
- *Accesibilidad*: Distancia a vías, aeropuerto y red ferroviaria.
- *Socioeconómico*: Distancia a instituciones educativas (temporal), centros de salud y atractivos turísticos.
- *Hidrografía*: Distancia a la red hídrica.

#### **04\_Resultados\_LULC\_Clasificados**

Mapas finales de Uso y Cobertura del Suelo (Urbano, Vegetación, Suelo Desnudo) resultantes de la clasificación Random Forest para 2004, 2014 y 2024, en formato de imagen de alta calidad. Estos incluyen:

- Mapas de clasificación superpuestos sobre imágenes de referencia de alta resolución (Google Earth Pro), utilizados para la verificación visual cualitativa de la precisión espacial.
- Mapas de clasificación superpuestos sobre las imágenes compuestas multiespectrales de resolución media (utilizadas por el modelo Random Forest en la etapa de clasificación), lo que permite contrastar y evaluar visualmente la coherencia espacial de los resultados.
- Mapas de clasificación temáticos (versiones “puras”), sin fondo de referencia, para observar claramente la distribución espacial y la continuidad de las clases.
- Vistas ampliadas (zoom) de zonas de interés específicas (urbanas, vegetadas, suelo desnudo) para cada año, presentadas igualmente con diferentes fondos de referencia (alta resolución, compuesto mediano).
- Mapa de Peligros para 2004 debido a la falta de imagen de alta resolución para este año.

- Archivos GeoTIFF (.tif) de las clasificaciones, optimizados para ser abiertos directamente en Google Earth Pro y permitir una visualización detallada e interactiva.

## **05\_Analisis\_de\_Cambios**

Mapas detallados de las transiciones específicas entre clases para los periodos 2004-2014 y 2014-2024.

## **06\_Simulacion\_y\_Validacion**

Contiene los insumos y resultados del proceso de modelado híbrido:

- Archivos tabulares (.csv) con las métricas de rendimiento de las combinaciones evaluadas durante la búsqueda de hiperparámetros (fases amplia y fina) del modelo MLP.
- Documento (.pdf) que detalla el rendimiento comparativo de las ejecuciones realizadas bajo la configuración óptima, utilizado para seleccionar la instancia definitiva de la red neuronal.
- Colección de mapas de Idoneidad de Transición generados por el modelo calibrado para los periodos 2014-2024 y 2024-2034.
- Mapa simulado del año 2024 utilizado para el cálculo de métricas de precisión frente al mapa real.
- Mapas de validación (2024) en formato GeoTIFF (.tif), optimizados para su visualización detallada en Google Earth Pro.
- Resultados del Análisis de Dependencia Parcial (PDP).

## **07\_Proyeccion\_Futura\_2034**

Mapa del escenario tendencial proyectado al año 2034 y mapas de los cambios simulados para la década 2024-2034. Se incluye también el mapa proyectado al 2034 en formato GeoTIFF (.tif), optimizado para su apertura y exploración en Google Earth Pro.