



Universidad Católica de Santa María

Facultad de Arquitectura e Ingenierías Civil y del Ambiente
Escuela Profesional de Ingeniería Ambiental

**Evaluación del riesgo de deslizamientos mediante algoritmos de machine
learning en el distrito de Quinistaquillas, Moquegua - 2024, para mejorar
la prevención y gestión de desastres naturales**

Tesis presentada por:

Yufra Ochoa, Bryan Alonso

ORCID: 0000-0002-3103-951X

Ramos Alegre, Luis Gerardo

ORCID: 0009-0007-9188-2959

para optar el Título Profesional de Ingeniero Ambiental

Asesor (a):

Dr. Gonzales Medina, Ronny Iván

ORCID: 0000-0002-9795-9538

Arequipa - Perú

2026

UCSM-ERP

UNIVERSIDAD CATÓLICA DE SANTA MARÍA

INGENIERIA AMBIENTAL

TITULACIÓN CON TESIS

DICTAMEN APROBACIÓN DE BORRADOR

Arequipa, 02 de Octubre del 2025

Dictamen: 012961-C-EPIA-2025

Visto el borrador del expediente 012961, presentado por:

2016200951 - YUFRA OCHOA BRYAN ALONSO

2016203321 - RAMOS ALEGRE LUIS GERARDO

Titulado:

**EVALUACIÓN DEL RIESGO DE DESLIZAMIENTOS MEDIANTE ALGORITMOS DE MACHINE
LEARNING EN EL DISTRITO DE QUINISTAQUILLAS, MOQUEGUA - 2024, PARA MEJORAR LA
PREVENCIÓN Y GESTIÓN DE DESASTRES NATURALES**

Nuestro dictamen es:

APROBADO

Título Profesional/Título de Segunda Especialidad/Grado Académico a optar:

INGENIERO AMBIENTAL

**46769238 - CHANOVE MANRIQUE ANDREA MARIETA
DICTAMINADOR**



**72384055 - MEDINA RAMOS ROBERT JOAQUIN
DICTAMINADOR**



**43606549 - CARDENAS PILLCO BERLY EDINSSON
DICTAMINADOR**



Evaluación del riesgo de deslizamientos mediante algoritmos de machine learning en el distrito de Quinistaquillas, Moquegua - 2024, para mejorar la prevención y gestión de desastres naturales

INFORME DE ORIGINALIDAD

15%
INDICE DE SIMILITUD

13%
FUENTES DE INTERNET

7%
PUBLICACIONES

5%
TRABAJOS DEL
ESTUDIANTE

FUENTES PRIMARIAS

1 Submitted to Universidad de Manizales <1 %
Trabajo del estudiante

2 Submitted to Universidad Católica de Santa María <1 %
Trabajo del estudiante

3 juandomingofarnos.wordpress.com <1 %
Fuente de Internet

4 cdn.www.gob.pe <1 %
Fuente de Internet

5 es.scribd.com <1 %
Fuente de Internet

6 www.regionsanmartin.gob.pe <1 %
Fuente de Internet

7 www.researchgate.net <1 %
Fuente de Internet

Dedicatoria

A Dios, por brindarme la fuerza y sabiduría necesaria en todo este camino.

A mis padres, Angel y Eddy Luz,
por su sacrificio, amor y por ser mi mayor ejemplo de perseverancia.

A mi hermano Rodrigo,
por su compañía y aliento en los momentos más desafiantes.

A mis amigos(as),
por su constante apoyo y por compartir conmigo esta etapa inolvidable.
Y a todas las personas que, de una u otra forma, me impulsaron a seguir adelante.

Bryan Alonso Yufra Ochoa

A mis padres, Luis y María Isela,
por su guía, confianza y amor incondicional que siempre me motivaron a dar lo mejor de mí.

A mis abuelos(as), Rómulo y Paulina,
por sus enseñanzas, ejemplo y por ser parte fundamental de mi vida.

A toda mi familia en general,
por acompañarme en cada paso de este proceso con cariño y comprensión.
A mis amigos(as) verdaderos(as), que creyeron en mí incluso cuando yo dudaba.
Con gratitud eterna a todos los que han sido parte de este logro.

Luis Gerardo Ramos Alegre

Agradecimientos

Agradezco profundamente a Dios por haberme brindado la fortaleza y la perseverancia necesarias para culminar esta etapa académica.

A mis padres, Angel y Eddy Luz, por su amor, comprensión y sacrificios constantes. Sin su apoyo incondicional, este logro no habría sido posible.

A mis docentes y asesores de la Universidad Católica de Santa María, por haber compartido conmigo su conocimiento y por guiarme con profesionalismo y paciencia a lo largo de este proceso.

A mis amigos(as) y compañeros(as) de carrera, por ser parte esencial del camino recorrido y por brindarme siempre palabras de aliento.

Gracias a todos aquellos que, de una u otra manera, dejaron una huella en esta etapa de mi vida.

Bryan Alonso Yufra Ochoa

Agradezco con humildad a Dios, por ser mi guía y fuente constante de esperanza durante los momentos más desafiantes de esta travesía académica.

A mis padres Luis y María Isela, por su ejemplo de esfuerzo, responsabilidad y cariño, pilares fundamentales en mi formación personal y profesional.

Agradezco también a mis profesores, asesores y a la Universidad Católica de Santa María, por haberme formado con excelencia y exigencia, permitiéndome alcanzar este objetivo con convicción y seguridad.

A mis seres queridos, amigos y colegas, quienes estuvieron a mi lado brindándome apoyo emocional, consejos y motivación en todo momento.

Este logro es el reflejo del acompañamiento y el esfuerzo colectivo de todos ustedes. Gracias infinitas.

Luis Gerardo Ramos Alegre

RESUMEN

El estudio titulado “Evaluación del riesgo de deslizamientos mediante algoritmos de Machine Learning en el distrito de Quinistaquillas, Moquegua – 2024” tuvo como propósito principal mejorar la prevención y gestión de desastres naturales en la zona. En los últimos años, Quinistaquillas ha presentado múltiples deslizamientos que pusieron en riesgo a la población, la agricultura y la infraestructura local, evidenciando la necesidad de utilizar tecnologías predictivas para fortalecer la gestión del riesgo y orientar la planificación territorial en áreas vulnerables.

La investigación abordó la carencia de herramientas automatizadas de predicción espacial para anticipar deslizamientos con precisión a nivel distrital. Para ello, se aplicaron algoritmos de Machine Learning utilizando datos geospaciales e imágenes satelitales multitemporales, especialmente imágenes Sentinel-2 L2A del periodo 2016–2023, junto con variables predictoras definidas de manera consistente como NDVI, NDWI, brillo, pendiente, tipo de suelo, uso de suelo y coordenadas espaciales. La orientación y la distancia a quebradas fueron empleadas como criterios auxiliares para la delimitación y caracterización espacial de las microcuencas, mas no como variables predictoras principales del entrenamiento de los modelos. Asimismo, se construyó una grilla regular que cubrió toda el área de estudio como base para el análisis espacial; sin embargo, el entrenamiento y la validación de los modelos se realizaron con un total de 3770 muestras etiquetadas, de las cuales 2827 registros (75 %) fueron destinados al entrenamiento y 943 registros (25 %) a la prueba.

Los modelos implementados fueron Random Forest (RF) y Support Vector Machine (SVM). El modelo Random Forest presentó el mejor desempeño, alcanzando una precisión de 90 % para la clase de deslizamiento, una sensibilidad de 86 %, un F1-score de 0.88, una precisión general de 94 % y un índice Kappa de 0.88, lo que demuestra una mayor capacidad para

identificar de manera confiable las zonas susceptibles a deslizamientos en comparación con el modelo SVM.

En el tercer objetivo, la aplicación de los modelos entrenados sobre las microcuencas de las quebradas Volcán y Juinco permitió generar mapas de susceptibilidad que identificaron y clasificaron zonas con alto potencial de deslizamiento. Ambos modelos coincidieron en que las áreas de riesgo alto y muy alto se asocian principalmente a laderas con fuertes pendientes, baja cobertura vegetal y proximidad a drenajes; sin embargo, el modelo RF presentó delimitaciones más coherentes, mientras que el SVM mostró una distribución más dispersa. Se observaron patrones espaciales diferenciados: en Volcán, la mayor concentración de riesgo se ubica hacia el norte, mientras que en Juinco se localiza principalmente en las zonas noreste y sureste, reflejando dinámicas geológicas particulares.

En conclusión, el uso de Machine Learning, especialmente mediante el modelo Random Forest, constituyó una herramienta eficaz y robusta para predecir zonas críticas de deslizamientos. Su integración con Sistemas de Información Geográfica (SIG) fortaleció la gestión preventiva y proporcionó información estratégica para la toma de decisiones orientadas a la prevención y gestión de desastres naturales en Quinistaquillas.

Palabras claves: Deslizamientos, Machine Learning, Random Forest

ABSTRACT

The study entitled "Landslide risk assessment using Machine Learning algorithms in the district of Quinistaquillas, Moquegua – 2024" had as its main purpose to improve the prevention and management of natural disasters in the area. In recent years, Quinistaquillas has experienced multiple landslides that put the population, agriculture and local infrastructure at risk, evidencing the need to use predictive technologies to strengthen risk management and guide territorial planning in vulnerable areas.

The research addressed the lack of automated spatial prediction tools to accurately anticipate landslides at the district level. To this end, Machine Learning algorithms were applied using geospatial data and multitemporal satellite imagery, especially Sentinel-2 L2A imagery from the period 2016–2023, along with consistently defined predictor variables such as NDVI, NDWI, brightness, slope, soil type, land use, and spatial coordinates. Orientation and distance to streams were used as auxiliary criteria for the delimitation and spatial characterization of the microwatersheds, but not as main predictor variables of the training of the models. Likewise, a regular grid was built that covered the entire study area as a basis for spatial analysis; however, the training and validation of the models were carried out with a total of 3770 labeled samples, of which 2827 records (75%) were destined for training and 943 records (25%) for testing.

The models implemented were Random Forest (RF) and Support Vector Machine (SVM). The Random Forest model presented the best performance, achieving an accuracy of 90% for the slip class, a sensitivity of 86%, an F1-score of 0.88, an overall accuracy of 94% and a Kappa index of 0.88, demonstrating a greater ability to reliably identify areas susceptible to landslides compared to the SVM model.

In the third objective, the application of the models trained on the micro-basins of the Volcán and Juinco streams allowed the generation of susceptibility maps that identified and classified areas with high landslide potential. Both models agreed that high and very high risk areas are

mainly associated with slopes with steep slopes, low vegetation cover and proximity to drainage; however, the RF model presented more coherent boundaries, while the SVM showed a more dispersed distribution. Differentiated spatial patterns were observed: in Volcán, the highest concentration of risk is located to the north, while in Juinco it is mainly located in the northeast and southeast areas, reflecting particular geological dynamics.

In conclusion, the use of Machine Learning, especially through the Random Forest model, was an effective and robust tool to predict critical landslide areas. Its integration with Geographic Information Systems (GIS) strengthened preventive management and provided strategic information for decision-making aimed at the prevention and management of natural disasters in Quinistaquillas.

Keywords: Landslides, Machine Learning, Random Forest

ÍNDICE GENERAL

DEDICATORIA

AGRADECIMIENTOS

RESUMEN

ABSTRACT

INTRODUCCIÓN 1

CAPÍTULO I 3

1. PLANTEAMIENTO DEL PROBLEMA..... 3

1.1. Justificación..... 6

1.1.1. Justificación Ambiental 6

1.1.2. Justificación Social 6

1.1.3. Justificación Tecnológica 8

1.1.4. Justificación Político Institucional..... 9

1.2. Objetivo general y objetivos específicos..... 10

1.2.1. Objetivo General..... 10

1.2.2. Objetivos Específicos: 10

1.3. Hipótesis..... 10

CAPÍTULO II 11

1. FUNDAMENTO TEÓRICO 11

1.1. Antecedentes de investigación 11

1.2. Introducción al Cambio Climático y Desastres Naturales 15

1.3. Efectos del Cambio Climático en los Patrones Meteorológicos 15

1.4. Tipos de Desastres Naturales Asociados a las Lluvias Intensas 16

1.5. Introducción al Machine Learning en Desastres Naturales..... 17

1.6. Aplicaciones de Machine Learning en Desastres Naturales 17

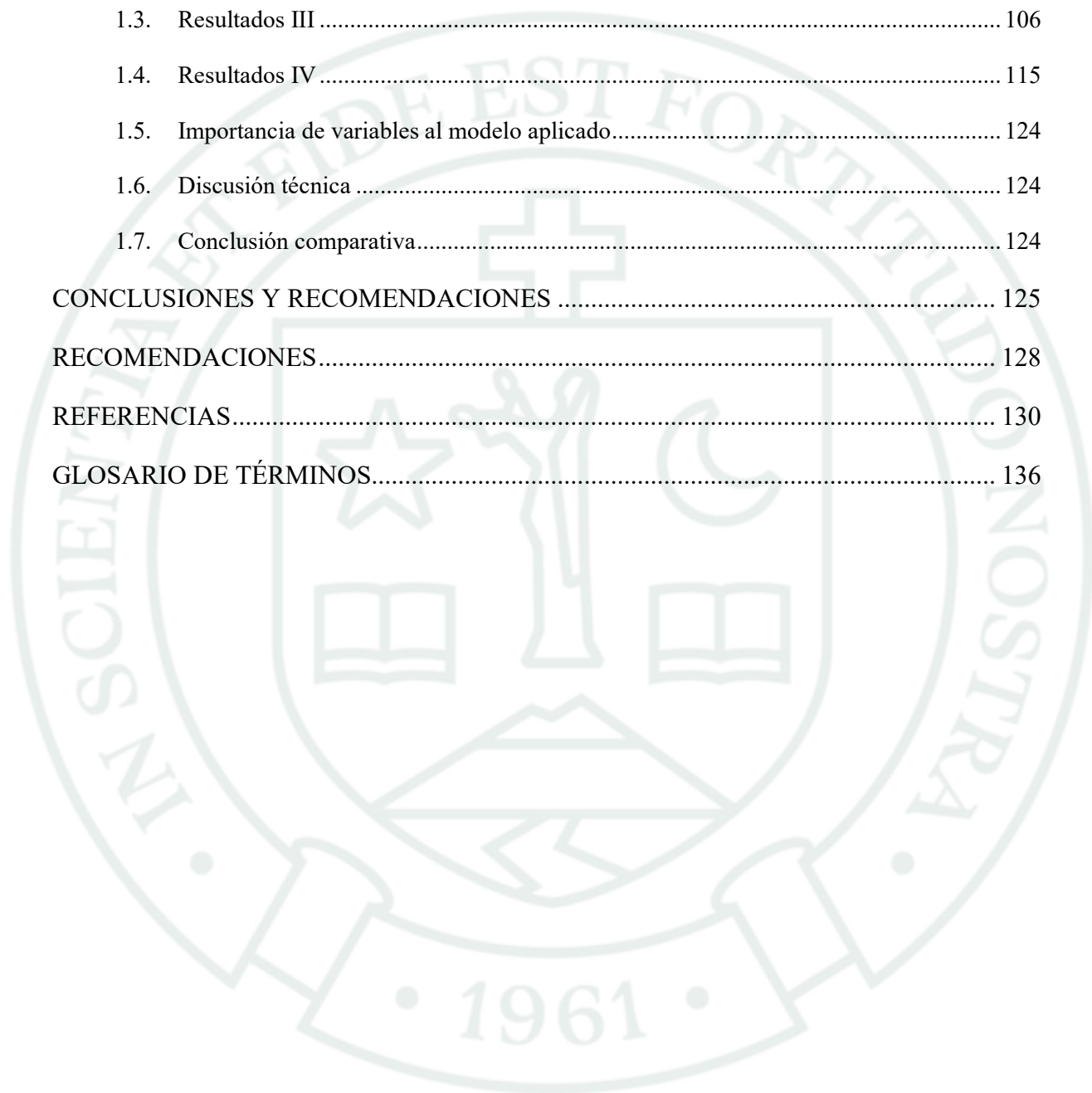
1.6.1. Predicción de Lluvias Intensas 17

1.6.2. Evaluación de Riesgos y Susceptibilidad 18

1.7.	Modelos de Machine Learning Utilizados en la Predicción de Lluvias	19
1.7.1.	Redes Neuronales Artificiales (ANNs).....	19
1.7.2.	Support Vector Machine (SVM) (SVM)	20
1.7.3.	Modelos de Random Forest (Random Forest).....	20
1.8.	Ventajas y Desafíos del Uso de ML en la Predicción de Desastres Naturales	28
1.8.1.	Ventajas	28
1.8.2.	Desafíos	29
1.9.	Casos de Estudio	29
1.9.1.	Estudio de Mosavi et al. (2018).....	29
1.9.2.	Estudio de Madhuri et al. (2021)	29
2.	CONTEXTO GLOBAL DE LAS LLUVIAS INTENSAS	30
2.1.	Tendencias Globales de las Precipitaciones Extremas.....	30
2.2.	Impactos Socioeconómicos y Ambientales a Nivel Mundial	30
2.3.	Estudios de Caso Internacionales.....	31
2.3.1.	Huracán Dorian (2019).....	31
2.3.2.	Tifón Hagibis (2019)	31
2.3.3.	Ciclón Daniel (2023)	31
3.	SITUACIÓN NACIONAL DE PERÚ	32
3.1.	Clima y Patrones de Precipitación en Perú	32
3.2.	Regiones más Vulnerables a Lluvias Intensas	32
3.3.	Impactos Históricos de Lluvias Extremas en Perú.....	32
3.3.1.	Inundaciones en Piura (2017)	32
3.3.2.	Huaicos en Huancavelica (2020)	33
3.3.3.	Respuesta y Adaptación.....	33
3.3.4.	Metodologías para la Evaluación de Riesgos en Quinistaquillas	35
3.3.5.	Factores de Vulnerabilidad en Quinistaquillas	35
3.3.6.	Análisis de Casos Recientes en Quinistaquillas	35
4.	IMPACTOS AMBIENTALES DE LAS LLUVIAS INTENSAS	38

4.1.	Erosión y Degradación del Suelo.....	38
4.2.	Pérdida de Biodiversidad y Alteración de Ecosistemas.....	38
4.3.	Contaminación de Cuerpos de Agua.....	39
5.	IMPACTOS SOCIALES Y ECONÓMICOS	40
5.1.	Desplazamientos Forzados y Pérdida de Viviendas.....	40
5.2.	Interrupción de Actividades Económicas.....	40
5.3.	Costos Asociados a la Reparación y Reconstrucción de Infraestructura	41
6.	MEDIDAS DE MITIGACIÓN Y ADAPTACIÓN	41
6.1.	Infraestructura Resiliente	41
6.2.	Sistemas de Alerta Temprana y Monitoreo Climático.....	42
6.3.	Reforestación y Manejo Sostenible de Cuencas Hidrográficas	42
6.4.	Participación Comunitaria y Educación.....	43
7.	POLÍTICAS Y MARCO LEGAL E INSTITUCIONAL.....	44
7.1.	Planes Nacionales y Regionales de Gestión de Riesgos	42
7.2.	Marco Legal Nacional e Internacional.....	42
7.2.1.	Normativa Internacional.....	42
7.2.2.	Legislación Peruana	43
7.3.	Instituciones Clave y sus Funciones	43
7.3.1.	Acuerdos y Marcos Internacionales.....	45
	CAPÍTULO III.....	48
1.	TIPO Y NIVEL DE INVESTIGACIÓN.....	48
2.	UBICACIÓN Y ÁREA DE ESTUDIO.....	48
3.	UNIDAD DE ANÁLISIS.....	53
4.	VARIABLES DE ESTUDIO E INDICADORES.....	54
5.	METODOLOGÍA PARA LOS OBJETIVOS ESPECÍFICOS	62
5.1.	Preprocesamiento de datos geoespaciales.....	62
6.	ANÁLISIS DE PREDICCIÓN CON PYTHON.....	72

CAPÍTULO IV.....	79
1. RESULTADOS Y DISCUSIONES	79
1.1. Resultados I.....	79
1.2. Resultados II.....	103
1.3. Resultados III	106
1.4. Resultados IV	115
1.5. Importancia de variables al modelo aplicado.....	124
1.6. Discusión técnica	124
1.7. Conclusión comparativa.....	124
CONCLUSIONES Y RECOMENDACIONES	125
RECOMENDACIONES.....	128
REFERENCIAS.....	130
GLOSARIO DE TÉRMINOS.....	136



ÍNDICE DE TABLAS

Tabla 1	Principales Impactos de lluvias extremas.....	16
Tabla 2	Modelos de ML utilizados en la predicción de Lluvias	19
Tabla 3	Proyectos de Mitigación y Adaptación.....	33
Tabla 4	Tabla resumida de eventos e impactos en Quinistaquillas	36
Tabla 5	Ubicación y área de estudio.....	49
Tabla 6	Variables Independientes e Indicadores	57
Tabla 7	Cuadro resumen de índices.....	102
Tabla 8	Tabla resumen de muestras y porcentajes de entrenamiento.....	103
Tabla 9	Métricas	104
Tabla 10	Métricas de rendimiento	115
Tabla 11.	Importancia de variables predictoras en el modelo Random Forest.....	107
Tabla 12.	Estabilidad de la importancia relativa de variables mediante simulación Monte Carlo.....	108

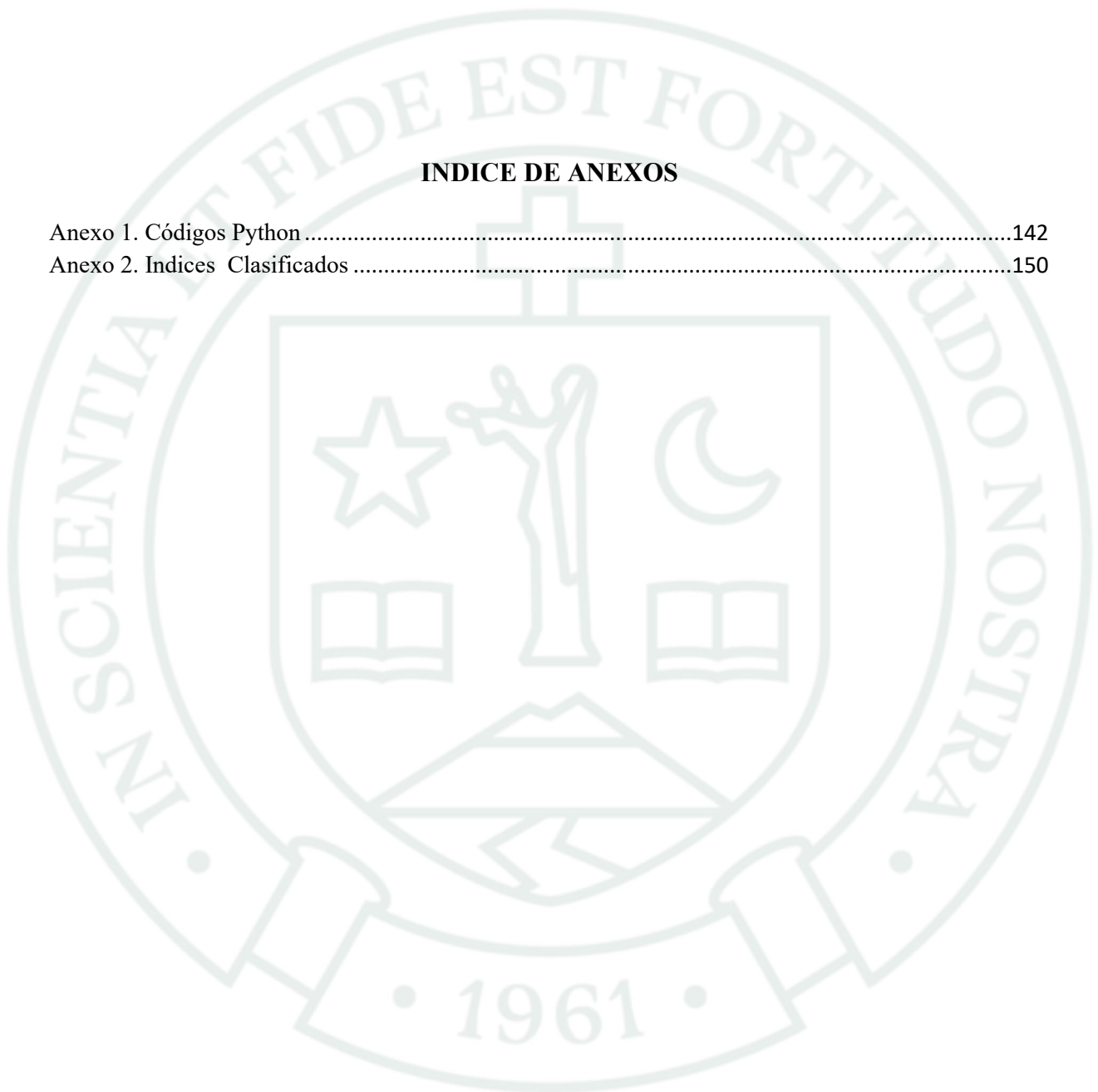
ÍNDICE DE FIGURAS

Figura 1. Rendimiento de Algoritmos ML	18
Figura 2 Cálculo de Coordenada en tabla de atributos	28
Figura 3. Delimitación del área de estudio, correspondiente al distrito de Quinistaquillas y las respectivas quebradas Juinco y Volcan.....	52
Figura 4 Grilla Fishnet.....	53
Figura 5 Conjunto por fecha (Brightness)	62
Figura 6 CSV muestras	71
Figura 7 Conjunto por Fecha (NDVI).....	79
Figura 8 Conjunto por fecha (Brightness)	80
Figura 9 Conjunto por fecha (NDWI).....	82
Figura 10 Imagen TCI.....	83
Figura 11 Delimitación y ubicación de quebradas.....	85
Figura 12 Generación de Microcuenca	87
Figura 13 Ubicación de Peligros registrados	89
Figura 14 Mapa de pendientes	91
Figura 15 Mapa de clasificación de suelos	93
Figura 16 Mapa de uso actual de suelo	95
Figura 17 Mapa de Muestras positivas y negativas(I)	97
Figura 18 Mapa de muestras positivas y negativas (II)	99
Figura 19 Fishnet (Muestras).....	101
Figura 20 Mapa de Predicción RF - MC Quebrada Volcán.....	107
Figura 21 Mapa de Predicción RF - MC quebrada Juinco.....	108
Figura 22 Mapa de Predicción SVM - Quebrada Volcán	109
Figura 23 Mapa de Predicción SVM - Quebrada Juinco	110
Figura 24 Mapa de Riesgo - Quebrada Volcán.....	113
Figura 25 Mapa de Riesgo - Quebrada Juinco.....	114
Figura 26 Gráfico de Barras de las métricas de rendimiento.....	116

Figura 27. Importancia relativa de variables predictoras en el modelo Random Forest.....	108
Figura 28. Estabilidad de la importancia relativa de variables mediante simulación Monte Carlo.....	109

INDICE DE ANEXOS

Anexo 1. Códigos Python.....	142
Anexo 2. Índices Clasificados	150



INTRODUCCIÓN

El distrito de Quinistaquillas, ubicado en la provincia General Sánchez Cerro, región Moquegua, presenta una alta vulnerabilidad frente a procesos geodinámicos como los deslizamientos de tierra, especialmente durante la temporada de lluvias. La combinación de factores físicos (pendientes pronunciadas, cobertura vegetal escasa y suelos inestables), sumada al cambio climático y a la presión antrópica sobre las cuencas, incrementa significativamente la exposición de la población, infraestructura y ecosistemas a este tipo de eventos. A pesar de las intervenciones preventivas desarrolladas por entidades como el CENEPRED y el INDECI, se evidencia aún la ausencia de herramientas predictivas automatizadas a escala local que permitan anticipar zonas críticas de manera oportuna y precisa.

En este contexto, los avances recientes en ciencia de datos y geotecnologías han impulsado el uso de modelos basados en Machine Learning (Machine Learning) como una alternativa robusta frente a los enfoques tradicionales de análisis multicriterio o modelos físicos complejos, los cuales requieren datos costosos o de difícil acceso en zonas rurales. Algoritmos como Random Forest (RF) y Support Vector Machine (SVM) (SVM) permiten procesar grandes volúmenes de datos geospaciales, identificar patrones no lineales entre variables condicionantes (como pendiente, NDVI, uso del suelo, distancia a quebradas, etc.) y generar mapas de susceptibilidad con alta precisión.

El uso de técnicas de *Machine Learning* ha sido respaldado por organismos como el Programa de las Naciones Unidas para el Medio Ambiente (PNUMA) y el Banco Mundial, quienes promueven su incorporación en la planificación territorial y en la reducción de riesgos de desastres, especialmente en países en vías de desarrollo. Su aplicación permite una mejor eficiencia operativa, escalabilidad, y capacidad de actualización continua a medida que se

obtienen nuevos datos satelitales y climatológicos. A diferencia de modelos puramente empíricos o estadísticos, el aprendizaje supervisado ofrece la posibilidad de entrenar los modelos con inventarios históricos de deslizamientos y evaluar su desempeño cuantitativamente mediante métricas como la precisión, sensibilidad o el índice Kappa.

Por lo tanto, esta investigación plantea la aplicación de algoritmos de Machine Learning para generar un modelo predictivo de zonas susceptibles a deslizamientos en Quinistaquillas, empleando imágenes multitemporales Sentinel-2 y variables geográficas de alta resolución. Se busca aportar una herramienta replicable y de bajo costo que fortalezca la gestión del riesgo de desastres en territorios rurales y apoye la toma de decisiones informadas por parte de autoridades locales, sectores productivos y población en general.

CAPÍTULO I

1. PLANTEAMIENTO DEL PROBLEMA

El cambio climático constituye uno de los principales desafíos ambientales contemporáneos, debido a que incrementa la frecuencia e intensidad de diversos eventos extremos, entre ellos las precipitaciones intensas y los procesos asociados de inestabilidad de laderas. El IPCC (2023) advierte que el aumento proyectado en la frecuencia e intensidad de las lluvias intensas incrementa la probabilidad de inundaciones locales y otros impactos en territorios vulnerables; en zonas montañosas, los efectos orográficos favorecen la precipitación extrema y, con ello, aumentan las condiciones propicias para la ocurrencia de movimientos en masa. En consecuencia, los deslizamientos se han consolidado como una amenaza relevante en regiones andinas, donde la pendiente, la geomorfología, la cobertura del suelo y la acción de las lluvias interactúan de manera compleja.

En el Perú, esta problemática adquiere especial importancia por la diversidad fisiográfica del territorio y por la presencia de poblaciones, infraestructura y actividades productivas ubicadas en espacios expuestos a amenazas geodinámicas. En particular, los deslizamientos y huaicos afectan de manera recurrente a las regiones andinas, interrumpiendo vías de comunicación, dañando sistemas de riego y comprometiendo medios de vida vinculados a la agricultura y la ganadería. Dentro de este contexto, la región Moquegua presenta sectores con relieve escarpado y quebradas activas que, bajo episodios de lluvia intensa, muestran una alta propensión a procesos de remoción en masa, lo que vuelve indispensable fortalecer los mecanismos de identificación preventiva de zonas susceptibles.

En el ámbito local, el distrito de Quinistaquillas evidencia una recurrencia reciente de eventos que confirman su vulnerabilidad frente a deslizamientos y huaicos. Según el Instituto Nacional de Defensa Civil, el 4 de febrero de 2022, a consecuencia de lluvias intensas, se produjo el

incremento del caudal de la catarata El Chorro, causando daños al puente vehicular del mismo nombre, ubicado en el km 1 de la vía departamental MO-108, tramo Quinistaquillas–Matalaque–Ubinas. Este hecho afectó directamente la conectividad territorial y mostró cómo un evento hidrometeorológico puede traducirse en impactos inmediatos sobre infraestructura estratégica para el distrito.

La recurrencia del problema continuó en 2024. INDECI (2024a) reportó que el 7 de enero de 2024 un deslizamiento, originado por lluvias intensas, afectó la vía de comunicación en el tramo Quinistaquillas–Sijuaya, dejando el tránsito vehicular inhabilitado y obligando a realizar acciones de limpieza, descolmatación y evaluación de daños. Posteriormente, INDECI (2024b) informó que el 24 de enero de 2024 las lluvias intensas incrementaron el caudal y provocaron el desborde del río Yaragua, además de activar quebradas que generaron un huaico con daños en infraestructura de riego, infraestructura de transporte, producción agrícola de alfalfa y papa, actividad pecuaria vinculada a vacunos y servicios básicos en los sectores de Yaragua, Lactogen y Agua Blanca. Estos reportes evidencian que no se trata de eventos aislados, sino de una dinámica recurrente de afectación territorial.

A ello se suman nuevos eventos ocurridos a fines de febrero de 2024. De acuerdo con INDECI (2024c), el 29 de febrero se activaron las quebradas Volcán, Juin y Tasata, ocasionando daños en infraestructura de transporte, infraestructura de riego, medios de vida y servicios básicos en los sectores de Quinistaquillas, Chimpapampa y Tasata. Asimismo, el Informe de Emergencia N.º 524 de INDECI precisó que, para el distrito de Quinistaquillas, estos eventos dejaron 4.7 km de carretera afectados, 200 m de canal de riego destruidos, 826 m de canal de riego afectados, 5 bocatomas destruidas, 31.84 ha de cultivo afectadas, 10 % de la red de agua potable sin servicio y afectación en la ganadería. El mismo informe señala además que el distrito se encontraba declarado en Estado de Emergencia mediante el D.S. N.º 033-2024-PCM, vigente hasta el 29 de mayo de 2024. Estos datos numéricos reflejan que la problemática trasciende la

ocurrencia puntual del deslizamiento y repercute directamente en el funcionamiento territorial, productivo y social del distrito.

Desde el punto de vista técnico, la peligrosidad del área también ha sido advertida por entidades especializadas. INGEMMET concluyó que en Quinistaquillas el principal peligro geológico corresponde a movimientos en masa de tipo flujo de detritos, condicionados por factores de sitio como las características litológicas, geomorfológicas y la pendiente del terreno. En sectores próximos evaluados por la misma entidad, como la quebrada El Volcán, se ha identificado incluso una condición de peligro muy alto, lo cual refuerza la necesidad de contar con herramientas que permitan no solo describir eventos ocurridos, sino anticipar espacialmente la susceptibilidad del territorio ante nuevos procesos de inestabilidad.

Sin embargo, pese a la recurrencia de estos eventos, a la disponibilidad de información satelital y a los reportes técnicos existentes, en Quinistaquillas no se identifica con claridad una herramienta local automatizada y comparativa que permita modelar la susceptibilidad a deslizamientos mediante algoritmos de machine learning aplicados específicamente al distrito y sus microcuencas críticas. El propio Plan de Prevención y Reducción del Riesgo de Desastres de la provincia General Sánchez Cerro 2024–2030 reconoce la necesidad de recopilar, sistematizar y diagnosticar información estadística e histórica, y señala que para su elaboración se contó con asistencia técnica del CENEPRED, lo que demuestra que todavía existe una demanda institucional por fortalecer la base técnica de la gestión prospectiva y correctiva del riesgo en el territorio.

En ese sentido, el problema de investigación radica en que el distrito de Quinistaquillas, pese a presentar antecedentes recientes y cuantificables de afectación por deslizamientos y huaicos, carece de un modelo espacial validado localmente que permita clasificar la susceptibilidad a deslizamientos a partir de variables geospaciales multitemporales. Esta limitación reduce la capacidad de identificar con anticipación zonas críticas y restringe el soporte técnico para la

planificación territorial, la priorización de intervenciones y la prevención de desastres. Por ello, resulta necesario aplicar y comparar algoritmos de machine learning, como Random Forest y Support Vector Machine, con el fin de generar cartografía de susceptibilidad que fortalezca la toma de decisiones en el distrito de Quinistaquillas.

1.1. Justificación

1.1.1. Justificación Ambiental

Los deslizamientos de tierra en el distrito de Quinistaquillas, Moquegua, representan una amenaza ambiental persistente, intensificada por factores geológicos, la geografía montañosa y las lluvias extremas asociadas al cambio climático. Estos eventos no solo generan pérdidas humanas y materiales, sino que también afectan gravemente la biodiversidad y los ecosistemas locales. En este contexto, la implementación de algoritmos de Machine Learning para el monitoreo y análisis post-deslizamientos permitirá procesar grandes volúmenes de datos de manera rápida y precisa, identificando patrones que adviertan sobre zonas críticas de riesgo y contribuyan a la prevención de futuros desastres.

Uno de los principales aportes será la elaboración de mapas de riesgo, integrando información georreferenciada sobre factores ambientales. Estos mapas serán una base información para la gestión ambiental y la planificación territorial, ya que permitirán focalizar acciones de mitigación, proteger áreas de alto valor ecológico y fortalecer la resiliencia de las comunidades. Al ofrecer datos de alta precisión, se facilitará la toma de decisiones informadas por parte de autoridades y organismos de respuesta, optimizando la prevención y reduciendo el impacto ambiental de los desastres naturales.

1.1.2. Justificación Social

En el distrito de Quinistaquillas, los deslizamientos de tierra representan una amenaza directa a la seguridad y el bienestar de la población, afectando no solo la integridad física de

las personas, sino también sus medios de vida, especialmente en comunidades que dependen de la agricultura y ganadería. Las lluvias intensas y la fragilidad de la infraestructura de transporte, vivienda y servicios básicos generan aislamiento, pérdidas económicas y dificultades de acceso a asistencia en emergencias. Implementar un sistema basado en algoritmos de Machine Learning para evaluar y predecir zonas de alto riesgo permitirá poner a disposición de las autoridades locales y de la población información clara y oportuna, reduciendo la exposición al peligro y salvando vidas.

De igual manera, la generación de mapas de riesgo actualizados y de fácil interpretación fomentará la participación activa de la comunidad en la gestión preventiva. Esto fortalecerá la conciencia social sobre los riesgos naturales, impulsará capacitaciones y planes de evacuación más eficientes y facilitará la coordinación entre organizaciones, autoridades y vecinos ante situaciones de emergencia. En el mediano y largo plazo, este enfoque tecnológico contribuirá a construir comunidades más resilientes, capaces de adaptarse a los impactos del cambio climático y de priorizar la protección de sus familias, bienes y entorno social.

Justificación Económica

En el distrito de Quinistaquillas, los deslizamientos de tierra generan considerables pérdidas económicas relacionadas con daños a la infraestructura vial, viviendas, áreas agrícolas y actividades productivas locales. Estos eventos interrumpen el transporte, afectan la conectividad y provocan gastos elevados en reparaciones y atención de emergencias, lo que impacta negativamente en la economía familiar y comunitaria. La implementación de un sistema basado en algoritmos de Machine Learning para evaluar el riesgo de deslizamientos permitirá mejorar la gestión preventiva, reduciendo significativamente los costos asociados a desastres naturales y optimizando la asignación de recursos públicos y privados.

Además, la generación de mapas de riesgo precisos y actualizados facilitará la planificación territorial estratégica, abriendo alternativas de inversiones tecnológicas en zonas altamente vulnerables y orientando proyectos de infraestructura con criterios de sostenibilidad y seguridad. Esto contribuirá a impulsar un desarrollo económico más estable y resiliente en Quinistaquillas, disminuyendo la vulnerabilidad económica de la población y promoviendo un entorno favorable para la actividad productiva y la inversión. En conjunto, la investigación aportará beneficios económicos tangibles a corto, mediano y largo plazo, fortaleciendo la capacidad de respuesta ante desastres y mejorando la calidad de vida de la comunidad.

1.1.3. Justificación Tecnológica

La justificación tecnológica de esta investigación radica en la necesidad de incorporar herramientas avanzadas y eficaces que permitan mejorar la evaluación del riesgo de deslizamientos en el distrito de Quinistaquillas, Moquegua. La aplicación de algoritmos de Machine Learning representa una innovación significativa frente a los métodos tradicionales, al ofrecer capacidad para procesar grandes volúmenes de datos heterogéneos y detectar patrones complejos relacionados con la inestabilidad del terreno. Esta tecnología posibilita un análisis más preciso, dinámico y en tiempo real, lo que incrementa la efectividad de la prevención y gestión de desastres naturales.

Además, la implementación de un sistema automatizado para la extracción y análisis de datos post-deslizamientos facilitará la actualización constante y la generación de mapas de riesgo inteligentes. Estos avances tecnológicos optimizan la toma de decisiones por parte de autoridades y gestores de riesgo, al brindar información detallada y actualizada que permite anticipar eventos críticos y planificar respuestas oportunas. En consecuencia, la investigación no solo contribuirá a modernizar los procesos de monitoreo y prevención en la región, sino que también fortalecerá la integración tecnológica en la gestión ambiental y territorial, potenciando la resiliencia comunitaria frente a los impactos del cambio climático.

1.1.4. Justificación Político Institucional

La justificación político-institucional de esta investigación radica en la necesidad de articular y fortalecer la acción coordinada de los diversos organismos públicos encargados de la gestión del riesgo de desastres en el distrito de Quinistaquillas, Moquegua. En el contexto nacional, el Sistema Nacional de Gestión del Riesgo de Desastres (SINAGERD) es el ente rector que promueve la cooperación interinstitucional entre la Presidencia del Consejo de Ministros (PCM), el Instituto Nacional de Defensa Civil (INDECI), el Centro Nacional de Estimación, Prevención y Reducción del Riesgo de Desastres (CENEPRED), los gobiernos regionales y locales, así como otras entidades públicas, fuerzas armadas, policía y la sociedad civil. Este sistema establece un marco normativo, técnico y operativo para la prevención, preparación, respuesta y rehabilitación ante desastres naturales.

En la región de Moquegua, la gestión del riesgo es apoyada por el Gobierno Regional y sus provincias, que cuentan con planes y estrategias adaptadas a sus realidades, como el Plan de Prevención y Reducción del Riesgo de Desastres. Además, instituciones específicas como la Autoridad Nacional del Agua (ANA) tienen un rol fundamental en la gestión hídrica y prevención de inundaciones y deslizamientos, trabajando en alianza estratégica con INDECI y CENEPRED para la identificación de zonas de riesgo y la coordinación de acciones preventivas. También participan sectores como el Ministerio de Salud, Ministerio de Agricultura, y Defensa Civil regionales y locales, que en conjunto con la comunidad y actores privados conforman un sistema integral para la gestión del riesgo. Esta investigación es pertinente porque, al aplicar tecnología avanzada como algoritmos de Machine Learning, aporta un componente innovador que potencia la capacidad técnica de estos organismos para mejorar la evaluación, prevención y gestión de deslizamientos, alineándose con las políticas públicas vigentes y fortaleciendo la gobernanza local y regional frente a los desastres naturales.

1.2. Objetivo general y objetivos específicos

1.2.1. Objetivo General

Evaluar la susceptibilidad a deslizamientos mediante algoritmos de Machine Learning en el distrito de Quinistaquillas, Moquegua – 2024

1.2.2. Objetivos Específicos:

1. Preprocesar datos geospaciales, incluyendo recolectar, transformar y normalizar datos, así como fusionar imágenes satelitales.
2. Aplicar los algoritmos de Support Vector Machine (SVM) y Random Forest para identificar zonas afectadas por deslizamientos.
3. Predecir zonas potenciales de deslizamiento con SVM y RF, y comparar las métricas de rendimiento de los modelos.

1.3. Hipótesis

El uso de algoritmos de *Machine Learning*, específicamente Random Forest (RF) y Support Vector Machine (SVM), permitirá clasificar con precisión la susceptibilidad a deslizamientos en el distrito de Quinistaquillas, Moquegua, a partir de variables geospaciales multitemporales.

CAPÍTULO II

1. FUNDAMENTO TEÓRICO

1.1. Antecedentes de investigación

En el ámbito internacional, diversos estudios han abordado la evaluación de riesgos de desastres naturales mediante la aplicación de algoritmos de Machine Learning, destacándose en áreas como la predicción de inundaciones, un tipo de desastre que afecta recurrentemente a muchas regiones del mundo. Madhuri et al. (2021) realizaron una investigación sobre la capacidad de distintos algoritmos de machine learning (ML) para clasificar áreas susceptibles de inundación bajo escenarios de cambio climático en Hyderabad, India. Para ello, emplearon un enfoque basado en datos geo-espaciales y diferentes variables climáticas y topográficas. El algoritmo XGBoost mostró el mejor desempeño con una AUC de 0.83, siendo capaz de simular ubicaciones propensas a inundaciones bajo diferentes proyecciones de escenarios climáticos. Asimismo, Al-Areeq et al. (2022) aplicaron algoritmos de ML combinados con sensores remotos y técnicas GIS para predecir la susceptibilidad a inundaciones en Jeddah, Arabia Saudita. Utilizando técnicas avanzadas de análisis de dependencia, lograron una alta precisión en la predicción de zonas de riesgo, alcanzando valores de AUC de hasta el 97%. Estos estudios subrayan la eficacia de los modelos predictivos basados en ML para la zonificación de áreas en riesgo y su aplicabilidad en la planificación urbana y la respuesta ante desastres. Por otro lado, Mosavi et al. (2018) en su revisión sobre la predicción de inundaciones, destacan cómo los métodos de ML han evolucionado en las últimas dos décadas, proporcionando soluciones más eficientes en comparación con los modelos físicos tradicionales. En este sentido, Endalie et al. (2022) proponen el uso de modelos LSTM para la predicción de lluvias diarias, un enfoque que mostró una alta capacidad predictiva en escenarios de precipitaciones extremas, con un coeficiente de determinación (R^2) de 0.9972. Estos avances en el uso de machine learning para la predicción de eventos climáticos extremos han sido complementados por

estudios que analizan el impacto de estas precipitaciones en la exposición de la población. Wang et al. (2023) utilizaron datos de modelos climáticos del CMIP6 para predecir el aumento de la precipitación extrema en Henan, China, evidenciando un incremento significativo en la exposición poblacional a lo largo del siglo bajo escenarios de altas emisiones. Finalmente, la revisión metodológica realizada por Merz et al. (2019) destacó la necesidad de validar las estimaciones de impacto generadas por diferentes metodologías de evaluación de riesgos, especialmente en relación con las inundaciones y lluvias intensas. Estos estudios reflejan la importancia de los modelos predictivos para mitigar los impactos de los desastres naturales y mejorar la resiliencia ante los efectos adversos del cambio climático. Una temática de investigación diferente siguió Rodríguez, Jorge y Henríquez (2020) quienes realizaron un estudio centrado en la resiliencia urbana ante las inundaciones causadas por lluvias intensas en Holguín, Cuba. Los autores analizaron la vulnerabilidad de las áreas urbanas utilizando análisis de componentes principales, identificando factores críticos de exposición y susceptibilidad. Subrayaron la importancia de integrar tanto aspectos sociales como naturales en la planificación urbana para mejorar la capacidad de recuperación ante desastres. Este enfoque destaca la importancia de crear bases de datos ambientales para gestionar de manera más eficiente los riesgos de inundación en áreas urbanas. Por su parte, Taalas (2020), representante de la Organización Meteorológica Mundial (OMM), presentó un informe en el que se vincula el cambio climático con el aumento de la frecuencia e intensidad de las lluvias extremas. Este estudio subraya la importancia de invertir en medidas de adaptación y en sistemas de alerta temprana para mitigar los riesgos asociados a desastres como las inundaciones. El informe hace hincapié en la necesidad de estrategias integrales que consideren la gestión del agua y medidas de adaptación frente a los efectos del cambio climático. Finalmente, un estudio publicado en la revista Sostenibilidad Ambiental en 2020 evaluó el impacto de las lluvias intensas sobre diversos aspectos sociales y ambientales. Se destacan impactos como los desplazamientos

forzados, las pérdidas económicas, los daños a la salud y la degradación ambiental, todos ellos exacerbados por la mayor intensidad de las tormentas debido al calentamiento global. La investigación destaca la necesidad de medidas de mitigación y adaptación para hacer frente a estos fenómenos.

En Perú, las lluvias intensas y los huaicos han sido una constante preocupación, exacerbada por los efectos del cambio climático. El Centro Nacional de Estimación, Prevención y Reducción del Riesgo de Desastres (CENEPRED) documentó en su informe de 2020 la gravedad de estos eventos, que han afectado significativamente varias regiones del país, causando pérdidas humanas y materiales. Durante la temporada de lluvias 2019-2020, departamentos como Arequipa, Cusco y Huancavelica sufrieron daños en infraestructura vial, viviendas y servicios básicos, lo que aisló a muchas comunidades. CENEPRED subraya la necesidad de fortalecer la infraestructura de drenaje, implementar sistemas de alerta temprana y educar a la población sobre los riesgos de lluvias intensas, buscando reducir los impactos a largo plazo y mejorar la resiliencia de las comunidades (CENEPRED, 2020). El Ministerio del Ambiente, a través de su Plan Nacional de Adaptación al Cambio Climático, publicado también en 2020, destacó que las lluvias intensas, combinadas con una infraestructura deficiente, representan un riesgo significativo para áreas urbanas y rurales. Ciudades como Lima y Piura han sufrido inundaciones graves que interrumpen el transporte y dañan viviendas, mientras que, en las regiones andinas y amazónicas, los deslizamientos de tierra y la erosión del suelo afectan gravemente la agricultura y la seguridad alimentaria. El plan propone medidas como la mejora de la infraestructura de drenaje, la reforestación de cuencas hidrográficas, el fortalecimiento de los sistemas de alerta temprana y la educación de la población para mitigar estos riesgos y promover un desarrollo sostenible (Ministerio del Ambiente, 2020). Por otro lado, el Instituto Nacional de Defensa Civil (INDECI) ha desarrollado planes de contingencia específicos para las regiones andinas más vulnerables a desastres naturales. En su informe de

2021, INDECI reveló que comunidades en Cusco, Puno y Ayacucho se ven particularmente afectadas por las lluvias extremas, que causan deslizamientos de tierra y aislan comunidades enteras. Las estrategias incluyen el mapeo de zonas de riesgo, el fortalecimiento de infraestructuras, la capacitación para la gestión de emergencias y la implementación de sistemas de alerta temprana, con el objetivo de mejorar la respuesta ante desastres y reducir los impactos negativos de estos eventos (INDECI, 2021). El Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI), por su parte, ha jugado un papel crucial en la predicción y monitoreo de lluvias intensas. En su informe de 2021, SENAMHI destacó la implementación de nuevas estaciones meteorológicas y el uso de tecnología satelital para mejorar la precisión de sus predicciones, particularmente en las regiones de Piura, Tumbes y Loreto, que sufren frecuentes inundaciones. Además, SENAMHI ha enfatizado la importancia de los sistemas de alerta temprana y la capacitación de las comunidades para actuar de manera oportuna ante eventos climáticos extremos (SENAMHI, 2021). La Oficina de las Naciones Unidas para la Coordinación de Asuntos Humanitarios (OCHA) publicó en 2024 un informe sobre el impacto de las lluvias intensas en Perú, detallando las severas consecuencias para las comunidades de San Martín, donde las inundaciones desplazaron a miles de personas y destruyeron infraestructura crítica. OCHA subraya la necesidad de mejorar la capacidad de respuesta, invertir en infraestructura resiliente y educar a las comunidades sobre la preparación ante desastres naturales (OCHA, 2024). Por su parte, el Banco Mundial ha colaborado estrechamente con el gobierno peruano en el financiamiento de proyectos para mejorar la infraestructura de drenaje y los sistemas de alerta temprana. En su informe de 2019, se destacaron iniciativas clave que han mejorado la resiliencia en áreas urbanas y rurales, implementando canales de drenaje y sistemas de retención de agua, así como tecnologías avanzadas de monitoreo para prever eventos climáticos extremos (Banco Mundial, 2019). La Universidad Nacional de Ingeniería (UNI) también ha contribuido al conocimiento sobre la

estabilidad del suelo y la infraestructura en Perú. En un estudio de 2020, investigadores de la UNI evaluaron el impacto de las lluvias extremas en la estabilidad de pendientes en Lima, Cusco y Arequipa, recomendando medidas de mitigación como la construcción de muros de contención y la integración de la gestión de riesgos en la planificación urbana (UNI, 2020). Finalmente, el Proyecto PRONATURALEZA ha implementado iniciativas para la conservación de ecosistemas vulnerables, como la reforestación de cuencas hidrográficas y la promoción de prácticas sostenibles de uso del suelo. Estas acciones buscan mitigar los impactos de las lluvias intensas y promover la participación comunitaria en la gestión del riesgo de desastres, asegurando un desarrollo sostenible (PRONATURALEZA, 2020).

1.2. Introducción al Cambio Climático y Desastres Naturales

El cambio climático se refiere a las variaciones a largo plazo en los patrones de temperatura y clima en la Tierra, causadas en gran parte por actividades humanas como la quema de combustibles fósiles, la deforestación y la industrialización. Estas actividades aumentan las concentraciones de gases de efecto invernadero en la atmósfera, como el dióxido de carbono (CO₂), el metano (CH₄) y el óxido nitroso (N₂O), que atrapan el calor y elevan las temperaturas globales (WMO, 2023).

1.3. Efectos del Cambio Climático en los Patrones Meteorológicos

El aumento de las temperaturas globales tiene un impacto directo en los patrones meteorológicos, incrementando la frecuencia y la intensidad de eventos climáticos extremos. Las investigaciones actuales muestran que las olas de calor, las tormentas intensas y las precipitaciones extremas son cada vez más comunes. Por ejemplo, en 2023, eventos como el huracán Otis en Acapulco y las lluvias torrenciales del ciclón Daniel en Libia demostraron cómo el calentamiento global intensifica estos fenómenos (Max-Planck-Gesellschaft, 2023).

1.4. Tipos de Desastres Naturales Asociados a las Lluvias Intensas

Las lluvias intensas provocadas por el cambio climático pueden llevar a varios tipos de desastres naturales, incluyendo inundaciones, deslizamientos de tierra y huaicos. Estos eventos tienen efectos devastadores en la infraestructura, la economía y la vida humana. En 2023, las inundaciones en Pakistán afectaron a 33 millones de personas y causaron daños económicos por valor de 30 mil millones de dólares (UN News, 2023). La intensificación de estos eventos está directamente relacionada con el aumento de la capacidad de la atmósfera para retener humedad debido al calentamiento global, lo que resulta en precipitaciones más intensas y frecuentes (ScienceDaily, 2023).

Tabla 1

Principales Impactos de lluvias extremas

Región/Lugar	Evento (año)	Impactos clave
Piura (costa norte)	Inundaciones (2017)	Decenas de muertos; miles desplazados; destrucción de carreteras y puentes.
Huancavelica (sierra)	Huaicos (2020)	Viviendas destruidas; comunidades aisladas; destacó vulnerabilidad de áreas montañosas.
Torata, Moquegua (sierra)	Inundaciones (2019)	Centenares de familias afectadas; carreteras intransitables; aislamiento de comunidades.
Torata, Moquegua (sierra)	Deslizamientos (2021)	Viviendas destruidas; muertes; necesidad de reforestación y muros de contención.
Arequipa (sierra sur)	Deslizamientos (2020)	Carreteras y puentes destruidos; inversión millonaria en reconstrucción de infraestructura.

Nota: Elaboración propia

Principales impactos de lluvias extremas desglosados por región o localidad en Perú, según el análisis del documento.

1.5. Introducción al Machine Learning en Desastres Naturales

Machine Learning (ML) es una subdisciplina de la inteligencia artificial que se enfoca en el desarrollo de algoritmos y modelos capaces de aprender y mejorar a partir de la experiencia y los datos. En el contexto de los desastres naturales, ML se utiliza para predecir eventos climáticos extremos, evaluar riesgos y ayudar en la toma de decisiones para mitigar los impactos. Los algoritmos de ML pueden procesar grandes volúmenes de datos históricos y en tiempo real, proporcionando predicciones más precisas y oportunas (Mosavi, Ozturk, & Chau, 2018; Madhuri, Sistla, & Srinivasa Raju, 2021).

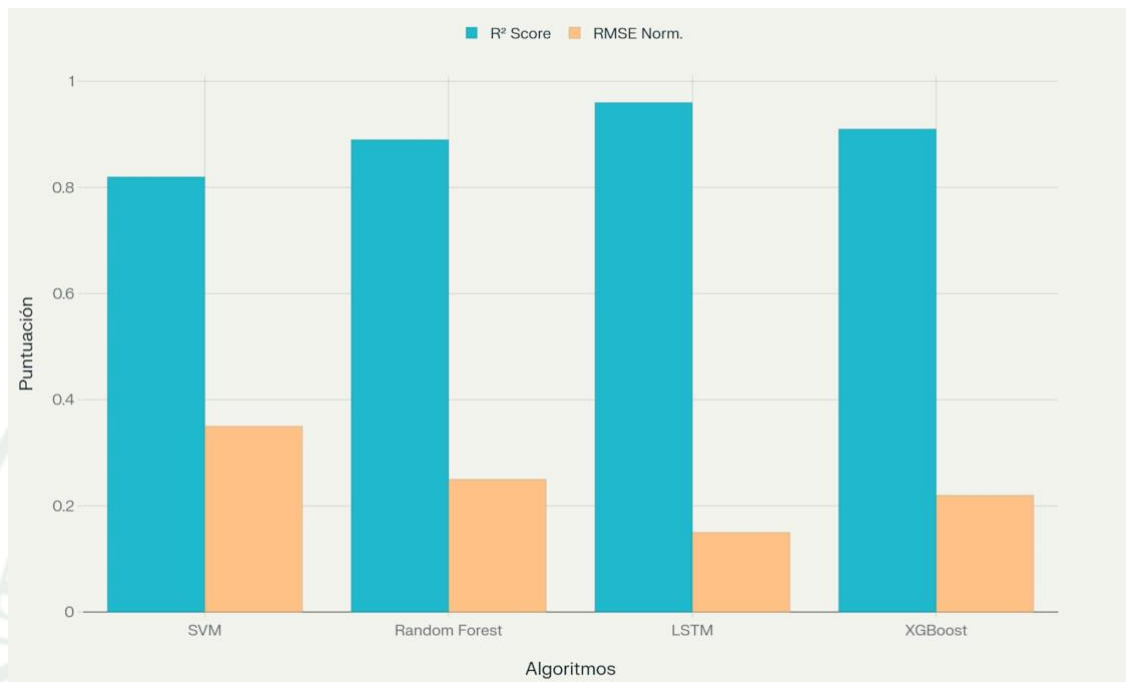
1.6. Aplicaciones de Machine Learning en Desastres Naturales

1.6.1. Predicción de Lluvias Intensas

El ML se emplea para predecir lluvias intensas utilizando datos meteorológicos históricos y en tiempo real. Los modelos de aprendizaje supervisado, como las redes neuronales y las Support Vector Machine (SVM), son comúnmente utilizados para esta tarea. Estos modelos analizan patrones en los datos para prever eventos de precipitación extrema con mayor precisión que los métodos tradicionales (Mosavi et al., 2018; Madhuri et al., 2021). En la Figura 1 se presenta una comparación del rendimiento de diferentes algoritmos de Machine Learning, evidenciando que modelos como Random Forest y XGBoost muestran un desempeño superior en términos de precisión.

Figura 1.

Rendimiento de Algoritmos ML



Nota: *Elaboración propia*

Comparación de rendimiento de algoritmos de machine learning para predicción de lluvias intensas

1.6.2. Evaluación de Riesgos y Susceptibilidad

Los modelos de ML también se utilizan para evaluar la susceptibilidad a desastres naturales como inundaciones y deslizamientos de tierra. Utilizando técnicas como el análisis de componentes principales y la regresión logística, estos modelos pueden identificar áreas vulnerables y estimar la probabilidad de ocurrencia de desastres en diferentes regiones (Mosavi et al., 2018; Madhuri et al., 2021).

1.7. Modelos de Machine Learning Utilizados en la Predicción de Lluvias

Tabla 2

Modelos de ML utilizados en la predicción de Lluvias

Modelo		Aplicación principal	Características / Ventajas
Redes Neuronales (ANN)		Predicción de lluvia intensa	Aprenden patrones complejos en datos meteorológicos históricos.
Support Vector Machine (SVM)	Vector (SVM)	Clasificación de eventos de precipitación	Clasifican tipos de eventos según variables meteorológicas.
Random Forest		Predicción y clasificación	Conjunto de árboles de decisión; mejora precisión y reduce sobreajuste.
XGBoost		Predicción de susceptibilidad a inundaciones	Algoritmo de boosting; destacó por su alto desempeño en estudios de inundaciones.

Nota: Elaboración propia

Comparativa de modelos de ML comunes en predicción de lluvias y evaluación de riesgos, basada en los casos de estudio y definiciones presentadas.

1.7.1. Redes Neuronales Artificiales (ANNs)

Las ANNs son modelos inspirados en la estructura del cerebro humano, capaces de aprender y generalizar patrones complejos en los datos. En la predicción de lluvias, las ANNs pueden analizar datos meteorológicos históricos para predecir la intensidad y duración de las precipitaciones (Mosavi et al., 2018; Madhuri et al., 2021).

1.7.2. Support Vector Machine (SVM) (SVM)

Las SVM son algoritmos de clasificación que pueden identificar patrones en los datos y separarlos en categorías. En la predicción de lluvias, las SVM pueden clasificar diferentes tipos de eventos de precipitación basados en variables meteorológicas (Mosavi et al., 2018).

1.7.3. Modelos de Random Forest (Random Forest)

Los Random Forest son un conjunto de árboles de decisión utilizados para la clasificación y regresión. En el contexto de la predicción de lluvias, estos modelos pueden combinar múltiples predicciones de árboles individuales para mejorar la precisión y reducir el sobreajuste (Mosavi et al., 2018).

Índice de Vegetación de Diferencia Normalizada (NDVI):

Metodología para el cálculo del NDVI en QGIS

Para calcular el Índice de Vegetación de Diferencia Normalizada (NDVI), se utilizó el software QGIS, considerando imágenes multitemporales Sentinel-2 L2A correspondientes al periodo 2016–2023. Este índice fue elegido por su amplia utilidad para detectar vegetación activa y cambios en la cobertura del terreno, como lo destacan Martínez-Santamaría et al. (2021) en estudios de susceptibilidad a movimientos en masa.

El procedimiento se inició con la carga de las bandas espectrales corregidas atmosféricamente dentro del entorno de QGIS. Las bandas utilizadas fueron la del espectro rojo (Banda 4) y la del infrarrojo cercano (Banda 8), según lo recomendado por Li et al. (2020), quienes señalaron que la combinación de estas bandas permite representar de forma precisa la densidad y el vigor de la vegetación.

A través de la herramienta *Calculadora ráster* del menú “Ráster”, se ejecutó la fórmula estándar para el NDVI:

Ecuación 1. NDVI

$$NDVI = (B8 - B4) / (B8 + B4)$$

donde B8 corresponde a la banda infrarroja cercana y B4 a la banda roja. Esta fórmula, ampliamente validada en análisis de teledetección ambiental (Zhou et al., 2022), fue aplicada para cada fecha procesada, a fin de capturar la variabilidad estacional y anual de la vegetación.

La salida del cálculo se generó en formato ráster tipo *float*, asignándole un nombre y ubicación específicos. El resultado fue una capa temática continua con valores entre -1 y 1, donde valores positivos cercanos a 1 indicaron mayor cobertura vegetal y valores cercanos a 0 o negativos señalaron áreas desnudas o cuerpos de agua.

Según lo indicado por Chen et al. (2019), la incorporación del NDVI como variable predictora permite mejorar significativamente el rendimiento de los modelos de predicción de deslizamientos, debido a que la pérdida súbita de vegetación puede ser indicativo de procesos de remoción en masa. Por ello, esta variable fue calculada para múltiples fechas, generando un set robusto de capas que fueron posteriormente integradas al modelo de RF y SVM

Índice de Agua de Diferencia Normalizada (NDWI):

El Índice Diferencial de Agua Normalizado (NDWI) fue calculado a partir de imágenes Sentinel-2 L2A en el entorno del software QGIS, con el propósito de identificar zonas con presencia de humedad superficial, factor clave en la ocurrencia de deslizamientos según indicaron Hernández et al. (2021). Para ello, se emplearon las bandas del verde (B3) y del infrarrojo de onda corta (SWIR, B11), de acuerdo con el algoritmo propuesto por Gao (1996), aún vigente en aplicaciones recientes.

Una vez cargadas las bandas correspondientes al periodo 2016–2023, se accedió a la herramienta *Calculadora ráster* en QGIS. Allí, se empleó la fórmula estándar del NDWI:

Ecuación 2. NDWI

$$NDWI = (B3 - B11) / (B3 + B11)$$

donde B3 representa la banda del espectro verde y B11 la del infrarrojo de onda corta. Esta operación fue aplicada individualmente sobre cada conjunto de imágenes procesadas multitemporalmente, a fin de detectar cambios en la humedad superficial tras eventos de precipitación.

El resultado de la operación generó una capa ráster de tipo *float*, con valores que oscilaron entre -1 y 1. Los valores positivos cercanos a 1 indicaron mayor contenido de humedad, mientras que los valores negativos reflejaron superficies secas o vegetación densa (Rahmati et al., 2020). Las capas obtenidas fueron posteriormente integradas al conjunto de datos para el entrenamiento de los modelos de Machine Learning, dado que, como concluyeron Reichenbach et al. (2018), la humedad del suelo es un factor altamente correlacionado con la probabilidad de ocurrencia de deslizamientos.

El NDWI permitió capturar zonas temporalmente saturadas o inestables, facilitando una mejor diferenciación entre sectores secos y aquellos con mayor acumulación de agua, condición relevante para entender los patrones de remoción de masas en las microcuencas estudiadas.

- **Brillo (Brightness):**

El índice de brillo (Brightness Index) fue calculado con base en imágenes Sentinel-2 L2A utilizando el software QGIS, a fin de identificar variaciones en la reflectancia del suelo, aspecto clave en la detección de zonas desnudas, escarpadas o expuestas, asociadas a procesos de remoción en masa. Según Chen et al. (2020), los valores altos de brillo pueden indicar áreas con escasa cobertura vegetal o suelos desnudos, condiciones favorables para la ocurrencia de deslizamientos.

Para el cálculo, se consideraron las bandas visibles del espectro electromagnético proporcionadas por Sentinel-2, específicamente las bandas B2 (azul), B3 (verde) y B4 (rojo). Estas bandas fueron previamente corregidas radiométricamente y cargadas en el entorno de QGIS. Posteriormente, se utilizó la *Calculadora ráster* para aplicar la fórmula adaptada del índice de brillo, definida como:

Ecuación 3. Brightness

$$Brightness = 2\sqrt{(B2^2 + B3^2 + B4^2)}$$

donde cada componente representa la reflectancia en cada una de las bandas del visible. Esta fórmula sigue la propuesta de Liu et al. (2019), quienes la utilizaron para evaluar la susceptibilidad al deslizamiento en zonas montañosas de alta pendiente, demostrando su capacidad para discriminar superficies claras (rocas expuestas, taludes) de áreas con vegetación o sombra.

El resultado fue una capa ráster de tipo flotante que representó los niveles de reflectancia relativa. Estas capas fueron generadas para múltiples fechas entre 2016 y 2023, seleccionadas en función de su representatividad climática (época seca vs. húmeda). Los valores altos del índice permitieron ubicar zonas propensas al escurrimiento y acumulación superficial, debido a la menor capacidad de infiltración del suelo, aspecto identificado también por Khosravi et al. (2021) como un factor relevante en el modelamiento de la susceptibilidad a deslizamientos.

Finalmente, las capas generadas fueron integradas como una de las variables predictoras en los modelos de clasificación supervisada (SVM y RF), siendo normalizadas previamente para asegurar su compatibilidad con el resto de variables geoespaciales.

- **Pendiente del terreno:**

La pendiente del terreno fue calculada utilizando el software ArcMap 10.x, a partir de un Modelo Digital de Elevación (MDE) de alta resolución. Se emplearon datos derivados del satélite ALOS PALSAR y del Shuttle Radar Topography Mission (SRTM) para representar la topografía de la zona de estudio con suficiente detalle. La pendiente es una variable crítica en la modelación de deslizamientos, ya que influye directamente en la estabilidad del suelo y la probabilidad de que ocurran movimientos en masa (Youssef et al., 2020).

En primer lugar, se descargó el MDE desde la plataforma oficial de la JAXA (para ALOS) y de la NASA (para SRTM), y se reproyectó a un sistema de coordenadas UTM zona 19S, datum WGS84, compatible con las demás capas del proyecto. Posteriormente, se utilizó la herramienta Slope del módulo *Spatial Analyst Tools* en ArcMap, la cual calcula la inclinación de cada celda del ráster de elevación en grados o porcentaje.

La fórmula utilizada por la herramienta se basa en el gradiente del valor de elevación de celdas vecinas, expresado matemáticamente como:

Ecuación 4. Pendiente

$$Pendiente (\%) = \sqrt{(dz/dx)^2 + (dz/dy)^2} \times 100$$

donde dz/dx y dz/dy representan las diferencias de elevación en las direcciones este-oeste y norte-sur, respectivamente. Se optó por obtener el resultado en grados, por ser más apropiado para modelar condiciones de inestabilidad en laderas, como también lo señalan estudios recientes (Pham et al., 2021).

Una vez generado el mapa de pendiente, este fue recategorizado en clases según el grado de inclinación, lo cual permitió una interpretación más clara de las zonas críticas. La capa resultante fue utilizada como una de las variables predictoras para entrenar los modelos

SVM y Random Forest, dado que se ha comprobado que las pendientes mayores a 30° incrementan significativamente la probabilidad de deslizamientos en regiones andinas (Meena et al., 2020).

Uso/cobertura del suelo: Variable categórica que describe el tipo de cobertura terrestre dominante en cada celda, de acuerdo con clasificaciones de uso de suelo (ej. bosque, matorral, pastos, áreas agrícolas, suelo desnudo, zonas urbanas, etc.). Se obtuvo a partir de cartografía temática proporcionada por entidades oficiales y de clasificaciones supervisadas de imágenes satelitales de alta resolución. En específico, se empleó la capa de “Uso Actual de Suelo” del área de Quinistaquillas, la cual asigna un código numérico a cada categoría.

Esta variable se incorporó debido a que la cobertura vegetal y el uso antrópico del terreno influyen en la estabilidad: una cobertura boscosa densa tiende a proteger contra deslizamientos gracias a la acción de las raíces y la menor erosión, mientras que terrenos deforestados o intervenidos (cultivos, pastoreo intenso) presentan mayor susceptibilidad por la reducción de cohesión del suelo. Diversos estudios incluyen la cobertura/uso del suelo como factor en los modelos de susceptibilidad, hallando que transiciones de bosque a áreas degradadas suelen correlacionar con incrementos en la frecuencia de deslizamientos. En esta investigación, los valores categóricos de uso de suelo fueron convertidos a variables dummy (mediante codificación *one-hot*) o tratados con códigos numéricos en los algoritmos que lo permitían, asegurando que el modelo pudiera distinguir entre los efectos asociados a cada tipo de cobertura. Adicionalmente, se complementó esta variable con el NDVI, de modo que el modelo dispusiera tanto de la clasificación temática (estática) como de un indicador continuo de verdor (dinámico) para caracterizar la vegetación de cada celda.

Tipo de suelo/geología: Otra variable categórica que recoge la clase de material superficial o litología predominante en cada ubicación. Se utilizó el mapa de unidades edáficas y geológicas de la zona (a escala local), el cual identifica tipos de suelos (p.ej., coluvial, aluvial,

volcánico residual) o formaciones geológicas (rocas ígneas vs. sedimentarias, etc.) relevantes. Cada celda recibió un código según este mapa, representando propiedades del material subyacente: por ejemplo, suelos arcillosos versus suelos arenosos, o roca volcánica fracturada versus tobas consolidadas. La motivación de incluir esta variable radica en que la resistencia al corte y la permeabilidad varían enormemente según el tipo de suelo/roca, afectando la propensión a deslizamientos. Ciertos litotipos, como depósitos de ceniza o suelos arcillo-limosos, son conocidos por generar deslizamientos más frecuentemente bajo lluvia que rocas ígneas masivas o suelos arenosos bien drenados. Por tanto, el modelo podría aprovechar esta información para ajustar la susceptibilidad base de una celda: por ejemplo, una pendiente moderada sobre lutitas puede ser más inestable que la misma pendiente en andesita competente. La variable fue incorporada codificada de forma numérica (etiquetas enteras) y el algoritmo Random Forest, en particular, pudo manejarla directamente y evaluar su importancia. En caso de SVM, se aplicó también codificación adecuada (one-hot) para no introducir ordinalidad falsa en categorías. Cabe mencionar que, ante la ausencia de datos detallados de propiedades geomecánicas, el tipo de suelo actúa como *proxy* de la resistencia y cohesión del terreno en diferentes zonas.

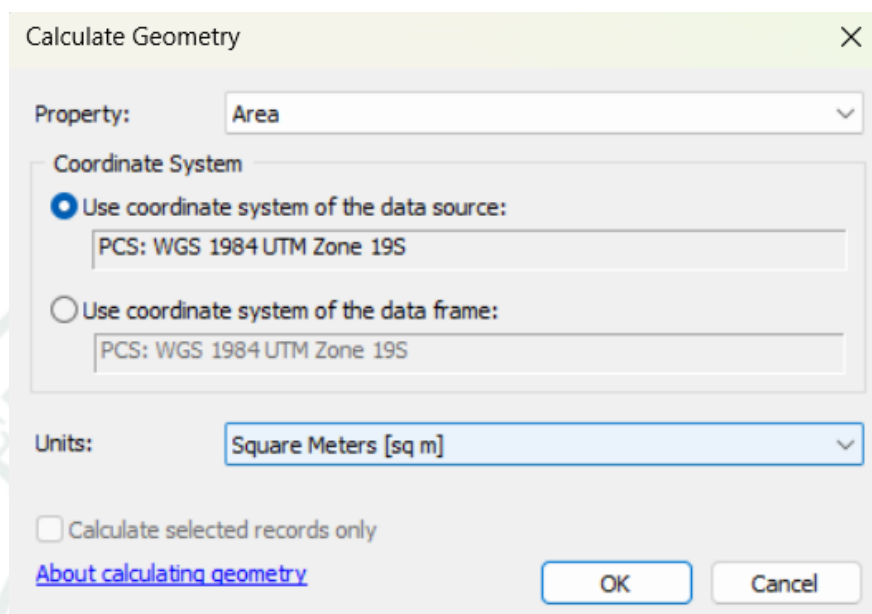
Coordenadas espaciales (X, Y): Finalmente, se integraron las coordenadas UTM Este y Norte de la celda como variables numéricas adicionales. Esta inclusión buscó captar posibles patrones espaciales globales no explicados por las demás variables, como gradientes regionales o efectos de proximidad a ciertas áreas. Por ejemplo, podría haber una tendencia geográfica donde las celdas en cierta parte de la microcuenca (p.ej., cercanas a la divisoria de aguas o a la confluencia con el río principal) tengan mayor propensión a deslizamientos por factores no explícitos en las demás capas (como diferencias microclimáticas o tectónicas). Al incorporar las coordenadas, los algoritmos de Machine Learning pueden, en teoría, inferir en dichas tendencias espaciales de gran escala. No obstante, se tuvo cuidado en interpretar este aporte

con precaución, ya que la inclusión de coordenadas puede también reflejar simplemente la distribución de las muestras usadas para el entrenamiento. En la práctica, Random Forest puede usar coordenadas para realizar particiones espaciales si ello mejora la pureza de los nodos, mientras que SVM las incorporará en su función de decisión como variables continuas. Se normalizaron las coordenadas (restando la media y dividiendo entre la desviación estándar) para que su magnitud numérica no dominara injustificadamente en el cálculo de distancias (especialmente importante para SVM). Aunque inusual en algunos enfoques clásicos de susceptibilidad, la adición de estas variables de localización se justificó en aras de maximizar el poder predictivo del modelo, permitiendo la detección de cualquier dependencia espacial residual.

En resumen, las variables anteriores fueron operacionalizadas a través de índices derivados de imágenes satelitales multitemporales (NDVI, NDWI, brillo), capas temáticas GIS (pendiente, uso de suelo, tipo de suelo) y datos de posición. Cada variable se preparó como una columna en la matriz de datos, garantizando la comparabilidad de escalas mediante procedimientos de normalización cuando fue pertinente. Por ejemplo, NDVI y NDWI ya están acotados en $[-1,1]$, pero el brillo (valores del orden de miles) y las pendientes ($0-90^\circ$) se escalaron a rangos $0-1$ para evitar sesgos en algoritmos sensibles a magnitudes. Del mismo modo, las variables categóricas se transformaron a representaciones numéricas apropiadas. Para calcular las coordenadas UTM en la tabla de atributos de la grilla de puntos, se utilizó la herramienta de “Calculate Geometry” de ArcGIS, como se muestra en la Fig 2. Toda esta matriz de datos constituyó la base para el entrenamiento y validación de los modelos de machine learning.

Figura 2

Cálculo de Coordenada en tabla de atributos



Nota: *ArcGis 10.8-ESRI*

1.8. Ventajas y Desafíos del Uso de ML en la Predicción de Desastres Naturales

1.8.1. Ventajas

- **Precisión Mejorada:** Los modelos de ML pueden manejar grandes volúmenes de datos y aprender patrones complejos, proporcionando predicciones más precisas que los métodos tradicionales (Madhuri et al., 2021).
- **Adaptabilidad:** Los modelos de ML pueden actualizarse continuamente con nuevos datos, mejorando su capacidad predictiva con el tiempo (Mosavi et al., 2018).
- **Automatización:** ML permite automatizar el proceso de análisis y predicción, facilitando una respuesta más rápida y eficiente ante eventos climáticos extremos (Mosavi et al., 2018; Madhuri et al., 2021).

1.8.2. Desafíos

- Disponibilidad de Datos: La precisión de los modelos de ML depende de la cantidad y calidad de los datos disponibles. En algunas regiones, la falta de datos históricos y en tiempo real puede limitar la efectividad de estos modelos (Mosavi et al., 2018).
- Interpretabilidad: Algunos modelos de ML, como las redes neuronales profundas, pueden actuar como 'cajas negras', dificultando la interpretación de sus predicciones (Mosavi et al., 2018; Madhuri et al., 2021).
- Implementación: La integración de modelos de ML en sistemas de gestión de riesgos existentes puede requerir una infraestructura tecnológica avanzada y capacitación especializada (Madhuri et al., 2021).

1.9. Casos de Estudio

1.9.1. Estudio de Mosavi et al. (2018)

Mosavi et al. (2018) revisaron la aplicación de modelos de ML para la predicción de inundaciones, destacando su capacidad para proporcionar soluciones más eficientes y rentables en comparación con los modelos físicos tradicionales. El estudio identificó varios algoritmos populares como las ANNs, ANFIS, SVM y WNN, enfatizando la importancia de la validación y comparación de modelos para asegurar su robustez y fiabilidad (Mosavi et al., 2018).

1.9.2. Estudio de Madhuri et al. (2021)

Madhuri et al. (2021) aplicaron algoritmos de ML para evaluar la susceptibilidad a inundaciones en Hyderabad, India. Utilizaron cinco algoritmos de ML, con XGBoost mostrando el mejor desempeño. Este modelo se utilizó para simular futuras ubicaciones de inundaciones bajo diferentes escenarios climáticos, destacando la importancia de modelos robustos y validados para la predicción de riesgos de inundaciones (Madhuri et al., 2021).

El uso de Machine Learning en la predicción de desastres naturales y lluvias intensas ofrece una herramienta poderosa para mejorar la resiliencia y la respuesta ante estos eventos. A través de la integración de datos meteorológicos y técnicas avanzadas de análisis, los modelos de ML pueden proporcionar predicciones precisas y oportunas, ayudando a mitigar los impactos y mejorar la preparación comunitaria. Sin embargo, es crucial abordar los desafíos asociados, como la disponibilidad de datos y la interpretabilidad de los modelos, para maximizar su efectividad (Mosavi et al., 2018; Madhuri et al., 2021).

2. CONTEXTO GLOBAL DE LAS LLUVIAS INTENSAS

2.1. Tendencias Globales de las Precipitaciones Extremas

El cambio climático ha provocado un aumento significativo en la frecuencia e intensidad de las precipitaciones extremas a nivel mundial. Según el informe de la Organización Meteorológica Mundial (WMO) de 2023, los eventos de lluvias intensas se han vuelto más comunes en diversas partes del mundo, afectando tanto a regiones desarrolladas como en desarrollo (WMO, 2023). Este incremento está estrechamente relacionado con el calentamiento global, que aumenta la capacidad de la atmósfera para retener humedad, resultando en tormentas más severas y prolongadas (ScienceDaily, 2023).

2.2. Impactos Socioeconómicos y Ambientales a Nivel Mundial

Los impactos de las lluvias intensas son devastadores tanto para las sociedades humanas como para el medio ambiente. Las inundaciones y las tormentas extremas pueden causar la destrucción de infraestructuras, pérdidas económicas significativas y desplazamientos masivos de población. Por ejemplo, en 2023, las lluvias torrenciales en Pakistán provocaron inundaciones que afectaron a 33 millones de personas y causaron daños por un valor de 30 mil millones de dólares (UN News, 2023). Además, eventos como el ciclón Daniel en Libia

demonstraron cómo las lluvias extremas pueden llevar a desastres humanitarios, con miles de muertes y desplazamientos (Max-Planck-Gesellschaft, 2023).

2.3. Estudios de Caso Internacionales

2.3.1. Huracán Dorian (2019)

El huracán Dorian fue uno de los ciclones tropicales más potentes jamás registrados en el Atlántico. Alcanzó la categoría 5 y causó estragos en las Bahamas, con vientos sostenidos de hasta 295 km/h. Las inundaciones resultantes del huracán devastaron grandes áreas, dejando miles de personas sin hogar y causando daños estimados en 3.4 mil millones de dólares (La Vanguardia, 2019).

2.3.2. Tifón Hagibis (2019)

El tifón Hagibis golpeó Japón en octubre de 2019, siendo uno de los más destructivos en la historia reciente del país. Las intensas lluvias y los fuertes vientos provocaron inundaciones y deslizamientos de tierra, afectando a millones de personas y causando pérdidas económicas significativas. La severidad del tifón se atribuyó en parte al aumento de las temperaturas del mar, relacionado con el cambio climático (La Vanguardia, 2019).

2.3.3. Ciclón Daniel (2023)

El ciclón Daniel, que afectó a Libia en 2023, es un ejemplo reciente de cómo las lluvias extremas pueden tener consecuencias devastadoras. El ciclón provocó lluvias torrenciales que llevaron al colapso de dos represas, causando inundaciones masivas y la muerte de miles de personas. Este evento destacó cómo el cambio climático puede aumentar la frecuencia y la intensidad de eventos extremos, poniendo en riesgo a las comunidades vulnerables (Max-Planck-Gesellschaft, 2023).

3. SITUACIÓN NACIONAL DE PERÚ

3.1. Clima y Patrones de Precipitación en Perú

Perú, con su geografía diversa que abarca desde la costa desértica hasta la Amazonía tropical y los Andes montañosos, experimenta una amplia variedad de climas y patrones de precipitación. Las regiones costeras suelen ser áridas, mientras que las áreas andinas y amazónicas reciben cantidades significativas de lluvia, especialmente durante la temporada de lluvias que va de noviembre a abril. El cambio climático ha exacerbado estos patrones, haciendo que las lluvias intensas sean más frecuentes y severas, especialmente en regiones vulnerables (SENAMHI, 2021).

3.2. Regiones más Vulnerables a Lluvias Intensas

Las regiones andinas y amazónicas de Perú son particularmente vulnerables a las lluvias intensas debido a su geografía y condiciones climáticas. Áreas como Cusco, Arequipa y Puno han sido testigos de deslizamientos de tierra y inundaciones devastadoras. Por ejemplo, en 2020, las lluvias torrenciales en Cusco causaron deslizamientos que bloquearon carreteras y aislaron comunidades, demostrando la vulnerabilidad de estas áreas a eventos climáticos extremos (INDECI, 2021).

3.3. Impactos Históricos de Lluvias Extremas en Perú

3.3.1. Inundaciones en Piura (2017)

En 2017, la región de Piura sufrió una de las peores inundaciones de su historia debido a las lluvias intensas provocadas por el fenómeno de El Niño Costero. Las inundaciones causaron la muerte de decenas de personas, desplazaron a miles y destruyeron infraestructura crítica, incluyendo carreteras y puentes. Los daños económicos se estimaron en miles de millones de soles, resaltando la necesidad urgente de mejorar la infraestructura y los sistemas de gestión de riesgos (CENEPRED, 2020).

3.3.2. Huaicos en Huancavelica (2020)

En febrero de 2020, la región andina de Huancavelica fue afectada por huaicos (deslizamientos de tierra) provocados por lluvias torrenciales. Estos huaicos destruyeron viviendas y dejaron a muchas comunidades aisladas. El evento subrayó la vulnerabilidad de las regiones montañosas a las lluvias intensas y la necesidad de implementar medidas de mitigación efectivas, como la construcción de muros de contención y sistemas de drenaje adecuados (INDECI, 2021).

3.3.3. Respuesta y Adaptación

a) Proyectos de mitigación y adaptación

Tabla 3

Proyectos de Mitigación y Adaptación

Iniciativa / Proyecto	Ámbito/Región	Tipo	Descripción
Plan Nacional de Adaptación al CC	Nacional (Perú)	Político/Infraestructura	Estrategias nacionales para infraestructuras resilientes, sistemas de alerta y reforestación.
Infraestructura de drenaje (BM)	Lima, Piura	Infraestructura	Sistemas avanzados de drenaje urbano: canales, alcantarillas y reservorios para lluvias intensas.
Sistemas de alerta temprana (SENAMHI)	Regiones andinas	Tecnología/Monitoreo	Red de estaciones meteorológicas en tiempo real para prever deslizamientos e inundaciones.

Reforestación de cuencas (Pronaturaleza)	Cusco	Ecosistema/Resiliencia	Plantación de especies nativas en cuencas críticas para estabilizar suelos y aumentar retención hídrica.
Educación comunitaria (Asoc. Labor)	Arequipa	Social / Capacitación	Talleres y programas de gestión de riesgos para líderes y población, fortaleciendo resiliencia local.

Nota: *Elaboración propia*

Resumen de algunas iniciativas y proyectos de mitigación y adaptación descritos, indicando ámbito, tipo de medida y objetivo principal, basado en las estrategias del documento.

b) Plan Nacional de Adaptación al Cambio Climático

El Ministerio del Ambiente de Perú ha desarrollado el Plan Nacional de Adaptación al Cambio Climático, que establece estrategias para abordar los impactos del cambio climático, incluyendo las lluvias intensas. El plan incluye la implementación de infraestructuras resilientes, sistemas de alerta temprana y programas de reforestación para reducir la vulnerabilidad de las comunidades (Ministerio del Ambiente, 2020).

c) Proyectos Comunitarios y Educación

Organizaciones como la Asociación Civil Labor han trabajado en proyectos comunitarios para fortalecer la resiliencia frente a desastres naturales, promoviendo la participación ciudadana y la educación en gestión de riesgos. Estos esfuerzos han demostrado ser efectivos para aumentar la capacidad de respuesta de las comunidades locales ante las lluvias intensas y otros eventos climáticos extremos (Labor, 2021).

3.3.4. Metodologías para la Evaluación de Riesgos en Quinistaquillas

La evaluación del riesgo en Quinistaquillas utiliza herramientas como GIS para mapear zonas de quebradas vulnerables (Volcán, Juinco, Tasata). Se complementa con análisis de datos históricos de lluvias intensas (marzo y abril de 2025) que desencadenaron huaicos e inundaciones. Modelos de riesgo integran variables climatológicas, geológicas (pendientes, suelos) e hidrológicas, permitiendo priorizar acciones en los tramos viales más críticos. CENEPRED ya brinda asistencia técnica para desarrollar el Plan de Prevención y Reducción del Riesgo de Desastres (PPRRD) del distrito, indicando un enfoque metodológico robusto a nivel local

3.3.5. Factores de Vulnerabilidad en Quinistaquillas

- Topografía y geología: Las quebradas de Volcán y Juinco presentan pendientes pronunciadas y suelos inestables, lo que facilita flujos de detritos y huaicos durante lluvias intensas
- Deforestación y uso del suelo: La reducción de cobertura vegetal en cuencas y zonas altoandinas disminuye la infiltración, incrementando erosión y taludes inestables que favorecen huaicos
- Infraestructura deficiente: El desborde de quebradas y la ausencia de obras hidráulicas robustas afectaron la infraestructura de riego en abril 2025. El colapso de caminos rurales e interrupción del servicio estructural reflejan vulnerabilidad en vías críticas.

3.3.6. Análisis de Casos Recientes en Quinistaquillas

a) Evento de enero–febrero 2024

- El 7 de enero de 2024, lluvias intensas provocaron un deslizamiento que destruyó un tramo vial clave entre Quinistaquillas y Sijuaya. La Municipalidad intervino

con maquinaria pesada, logrando un 60 % de avance en limpieza y rehabilitación para el 9 de enero, manteniendo el distrito en estado de emergencia hasta el 3 de febrero

b) Inundaciones e huaicos — marzo 2024

- En marzo de 2024, lluvias activaron las quebradas *Volcán, Juinco y Tasata*, generando daños en infraestructura y riego. No se reportaron pérdidas de vidas, pero sí daños materiales importantes, activando procedimientos de emergencia y mitigación

c) Abril 2025: lluvias intensas y huaicos

- En abril de 2025, lluvias severas desbordaron quebradas y provocaron huaicos que dañaron infraestructura de riego. CENEPRED intervino activamente, brindando asistencia técnica para el diseño del PPRRD, en particular en quebradas vulnerables, mostrando una evolución hacia la planificación preventiva

Tabla 4

Tabla resumida de eventos e impactos en Quinistaquillas

Fecha	Tipo de evento	Zona afectada	Impactos clave	Respuesta
07–09 ene 2024	Deslizamiento	Camino Quinistaquillas–Sijuaya	Vía interrumpida, maquinaria pesada al 60 % de limpieza	Intervención de Municipalidad e COEN; EDAN realizada

Mar	Lluvias	+	Quebradas	Daño	en	Estado	de
2024	huaicos		Volcán, Juinco y Tasata	infraestructura de riego y caminos rurales; sin pérdidas humanas	de	emergencia	y
Abril	Lluvias	+	Infraestructura de riego e hidráulica	Daño en canales y tubos; riesgo de recurrencia si no se mitiga	de	Asistencia técnica	
2025	huaicos				de	CENEPRED	para PPRRD

Nota: *Elaboración propia*

Estas actualizaciones reemplazan la información de Torata con datos concretos de Quinistaquillas, mejorando la calidad y relevancia del análisis local. Si deseas agregar imágenes o gráficos de estas situaciones, puedo ayudarte a generarlos.

4.3.6. Medidas de Mitigación Propuestas

a) Infraestructura Resiliente

La construcción de sistemas de drenaje más eficientes y la rehabilitación de carreteras pueden reducir significativamente el impacto de las lluvias intensas. Además, la implementación de muros de contención y otras estructuras de mitigación puede prevenir deslizamientos de tierra y proteger a las comunidades vulnerables (Banco Mundial, 2019).

b) Sistemas de Alerta Temprana

El desarrollo de sistemas de alerta temprana que utilicen tecnología avanzada para monitorear las condiciones meteorológicas y predecir eventos extremos es crucial para reducir

el riesgo de desastres. Estos sistemas pueden proporcionar a las comunidades tiempo suficiente para evacuar y tomar medidas preventivas (SENAMHI, 2021).

c) Participación Comunitaria

Involucrar a las comunidades locales en la planificación y ejecución de medidas de mitigación es esencial para garantizar la sostenibilidad de estas iniciativas. Programas de educación y capacitación pueden mejorar la resiliencia comunitaria y asegurar una respuesta más efectiva ante desastres naturales (Labor, 2021).

4. IMPACTOS AMBIENTALES DE LAS LLUVIAS INTENSAS

4.1. Erosión y Degradación del Suelo

Las lluvias intensas pueden provocar una erosión significativa del suelo, especialmente en áreas con cobertura vegetal insuficiente o prácticas agrícolas inadecuadas. La erosión del suelo reduce su fertilidad, afecta la producción agrícola y puede llevar a deslizamientos de tierra en regiones montañosas. En Perú, la deforestación y el uso no sostenible de la tierra agravan estos problemas, como se ha observado en las regiones andinas durante eventos de lluvias intensas (PRONATURALEZA, 2020).

- **Caso de Estudio: Regiones Andinas de Perú**

En las regiones andinas, la erosión del suelo provocada por lluvias intensas ha resultado en la pérdida de tierras cultivables y ha afectado gravemente la productividad agrícola. La deforestación en estas áreas contribuye a la falta de cohesión del suelo, aumentando la susceptibilidad a la erosión (UNI, 2020).

4.2. Pérdida de Biodiversidad y Alteración de Ecosistemas

Las lluvias intensas pueden alterar significativamente los ecosistemas, causando la pérdida de biodiversidad. Las inundaciones pueden destruir hábitats naturales, mientras que el

exceso de agua puede alterar los ciclos biológicos de plantas y animales. En las áreas tropicales de Perú, como la Amazonía, las lluvias extremas han afectado los ecosistemas de manera adversa, provocando la pérdida de especies y la degradación de hábitats (PRONATURALEZA, 2020).

- **Impacto en la Amazonía Peruana**

Las lluvias intensas y las inundaciones frecuentes en la Amazonía peruana han causado la pérdida de especies y la alteración de los ciclos de vida de muchos animales y plantas. La fragmentación de hábitats y la destrucción de áreas forestales son algunas de las consecuencias más graves, afectando tanto la flora como la fauna de esta región (Banco Mundial, 2019).

4.3. Contaminación de Cuerpos de Agua

Las lluvias intensas pueden provocar la contaminación de cuerpos de agua debido al arrastre de sedimentos, productos químicos agrícolas y desechos urbanos. Esto no solo afecta la calidad del agua potable, sino también la salud de los ecosistemas acuáticos. En Perú, las inundaciones en áreas urbanas y rurales han llevado a la contaminación de ríos y lagos, lo que representa un riesgo significativo para la salud pública y la biodiversidad acuática (SENAMHI, 2021).

- **Ejemplo de Contaminación en el Río Rímac**

Durante las temporadas de lluvias intensas, el río Rímac en Lima ha experimentado niveles elevados de contaminación debido al arrastre de residuos sólidos y aguas residuales. Esto afecta la calidad del agua utilizada para el consumo humano y agrícola, y pone en riesgo la vida acuática (CENEPRED, 2020).

5. IMPACTOS SOCIALES Y ECONÓMICOS

5.1. Desplazamientos Forzados y Pérdida de Viviendas

Las lluvias intensas pueden causar desplazamientos masivos de personas, ya que las inundaciones y deslizamientos de tierra destruyen viviendas y comunidades enteras. En Perú, eventos recientes han obligado a miles de personas a abandonar sus hogares. Por ejemplo, las inundaciones en Piura en 2017 desplazaron a miles de familias y causaron daños significativos a la infraestructura residencial (CENEPRED, 2020; INDECI, 2021).

- **Caso de Estudio: Inundaciones en Piura**

Las inundaciones en Piura, causadas por el fenómeno de El Niño en 2017, resultaron en la evacuación de más de 300,000 personas. Las aguas destruyeron viviendas y dejaron a muchas familias sin hogar, destacando la vulnerabilidad de las zonas urbanas y rurales a las lluvias intensas (CENEPRED, 2020).

5.2. Interrupción de Actividades Económicas

Las lluvias intensas pueden interrumpir significativamente las actividades económicas, especialmente en sectores clave como la agricultura y la minería. Las inundaciones y deslizamientos de tierra pueden dañar cultivos, infraestructura agrícola y minas, lo que afecta la producción y la economía local. En Perú, las lluvias extremas han tenido un impacto devastador en estos sectores, causando pérdidas económicas significativas (Banco Mundial, 2019).

- **Impacto en la Agricultura y Minería**

Las lluvias torrenciales en regiones agrícolas como Cusco y Arequipa han resultado en la pérdida de cosechas y daños a la infraestructura agrícola. Del mismo modo, en las regiones mineras, los deslizamientos de tierra han interrumpido las operaciones mineras, resultando en

pérdidas económicas y en la interrupción de la cadena de suministro (INDECI, 2021; Banco Mundial, 2019).

5.3. Costos Asociados a la Reparación y Reconstrucción de Infraestructura

Las inundaciones y deslizamientos causados por lluvias intensas generan altos costos de reparación y reconstrucción de infraestructuras dañadas, como carreteras, puentes, y edificios. Estos costos representan una carga significativa para los presupuestos locales y nacionales. En Perú, los eventos de lluvias extremas han llevado a inversiones sustanciales en la reconstrucción de infraestructura crítica (Banco Mundial, 2019; SENAMHI, 2021).

- **Ejemplo de Reconstrucción en Arequipa**

Las lluvias intensas en Arequipa en 2020 causaron deslizamientos que destruyeron carreteras y puentes vitales para la región. El gobierno y organizaciones internacionales han invertido millones de dólares en la reconstrucción de esta infraestructura para restablecer la conectividad y apoyar la recuperación económica (INDECI, 2021).

6. MEDIDAS DE MITIGACIÓN Y ADAPTACIÓN

6.1. Infraestructura Resiliente

La construcción de infraestructura resiliente es una de las medidas más efectivas para mitigar los impactos de las lluvias intensas. Esto incluye la implementación de sistemas de drenaje eficientes, la construcción de muros de contención y la rehabilitación de carreteras y puentes. Estas infraestructuras están diseñadas para soportar grandes volúmenes de agua y evitar inundaciones y deslizamientos de tierra. En Perú, varios proyectos financiados por el Banco Mundial han demostrado la eficacia de estas medidas, reduciendo significativamente el riesgo de desastres en áreas vulnerables (Banco Mundial, 2019; INDECI, 2021).

- **Proyectos de Drenaje en Lima y Piura**

En Lima y Piura, la implementación de sistemas avanzados de drenaje ha mejorado la capacidad de estas ciudades para manejar las lluvias intensas. Estos proyectos han incluido la instalación de canales y alcantarillas de gran capacidad, así como la construcción de reservorios temporales para almacenar el agua de lluvia y liberar gradualmente su flujo (Banco Mundial, 2019).

6.2. Sistemas de Alerta Temprana y Monitoreo Climático

El desarrollo de sistemas de alerta temprana y monitoreo climático es crucial para proporcionar información oportuna y precisa sobre eventos climáticos extremos. Estos sistemas utilizan tecnología avanzada para monitorear las condiciones meteorológicas y predecir eventos extremos, permitiendo a las comunidades y autoridades tomar medidas preventivas con antelación. En Perú, SENAMHI ha implementado varias iniciativas en este sentido, mejorando la capacidad de respuesta del país ante las lluvias intensas (SENAMHI, 2021).

- **Implementación en las Regiones Andinas**

SENAMHI ha instalado estaciones meteorológicas en las regiones andinas para monitorear las condiciones climáticas en tiempo real. Estos sistemas han permitido prever deslizamientos de tierra y inundaciones, facilitando la evacuación y la implementación de medidas de mitigación antes de que ocurran los desastres (SENAMHI, 2021).

6.3. Reforestación y Manejo Sostenible de Cuencas Hidrográficas

La reforestación y el manejo sostenible de cuencas hidrográficas son estrategias importantes para reducir la vulnerabilidad a las lluvias intensas. La reforestación ayuda a estabilizar el suelo, mejorar la retención de agua y reducir la erosión, mientras que el manejo sostenible de las cuencas hidrográficas asegura que los recursos hídricos se utilicen de manera

eficiente y sostenible. En Perú, proyectos como PRONATURALEZA han promovido estas prácticas en varias regiones, mejorando la resiliencia ambiental y comunitaria (PRONATURALEZA, 2020).

- **Ejemplo de Reforestación en Cusco**

En la región de Cusco, los esfuerzos de reforestación han incluido la plantación de especies nativas en áreas críticas para prevenir la erosión del suelo y mejorar la retención de agua. Estas iniciativas han sido fundamentales para reducir el riesgo de deslizamientos de tierra y proteger las comunidades locales de los efectos de las lluvias intensas (PRONATURALEZA, 2020).

6.4. Participación Comunitaria y Educación

La participación comunitaria y la educación son esenciales para el éxito de las estrategias de mitigación y adaptación. Involucrar a las comunidades locales en la planificación y ejecución de medidas de mitigación asegura que las soluciones sean adecuadas y sostenibles. Programas de educación y sensibilización ayudan a aumentar la conciencia sobre los riesgos y las medidas preventivas que las personas pueden tomar para protegerse de las lluvias intensas. En Perú, organizaciones como la Asociación Civil Labor han trabajado para fortalecer la resiliencia comunitaria a través de estos enfoques (Labor, 2021).

- **Proyectos Comunitarios en Arequipa**

En Arequipa, la Asociación Civil Labor ha implementado programas de capacitación para líderes comunitarios y talleres educativos para la población en general, enfocados en la gestión de riesgos de desastres. Estos programas han mejorado la capacidad de las comunidades para prepararse y responder de manera efectiva a las lluvias intensas y otros eventos climáticos extremos (Labor, 2021).

7. POLÍTICAS Y MARCO LEGAL E INSTITUCIONAL

7.1. Marco normativo general de la gestión del riesgo de desastres en el Perú

La presente investigación se desarrolla en el marco del ordenamiento jurídico e institucional que regula la gestión del riesgo de desastres en el Perú. La norma base es la Ley N.º 29664, Ley que crea el Sistema Nacional de Gestión del Riesgo de Desastres (SINAGERD), la cual establece el sistema interinstitucional responsable de identificar, analizar, prevenir, reducir y controlar los factores de riesgo, así como de preparar, responder y rehabilitar frente a situaciones de desastre. Esta ley constituye el sustento principal para la articulación de la gestión prospectiva, correctiva y reactiva del riesgo en los distintos niveles de gobierno, por lo que su aplicación resulta plenamente pertinente en investigaciones orientadas a la identificación de zonas susceptibles a deslizamientos y a la generación de insumos técnicos para la planificación territorial.

El desarrollo reglamentario de dicha ley fue aprobado mediante el Decreto Supremo N.º 048-2011-PCM, que aprueba el Reglamento de la Ley N.º 29664. Este reglamento precisa la organización del SINAGERD, define los procesos de la gestión del riesgo y establece las competencias de las entidades responsables. En el contexto de la presente tesis, este reglamento es relevante porque otorga respaldo institucional al uso de estudios técnicos que permitan identificar condiciones de peligro y vulnerabilidad en el territorio, especialmente cuando se trata de áreas con recurrencia de movimientos en masa, como ocurre en el distrito de Quinistaquillas.

Desde esta perspectiva, la elaboración de mapas de susceptibilidad a deslizamientos en microcuencas críticas representa un aporte técnico que puede integrarse como insumo para acciones de prevención, reducción del riesgo y priorización de intervenciones a nivel local y regional.

7.2. Lineamientos técnicos del CENEPRED vinculados al análisis de susceptibilidad

Más allá del marco legal general, la investigación requiere sustentarse en instrumentos técnicos específicos. En el Perú, la entidad especializada en materia de estimación, prevención y reducción del riesgo es el Centro Nacional de Estimación, Prevención y Reducción del Riesgo de Desastres (CENEPRED), organismo que emite lineamientos y herramientas metodológicas de referencia para el análisis territorial del riesgo. En consecuencia, la presente tesis no solo debe citar al CENEPRED como actor institucional, sino utilizar sus lineamientos como marco técnico que refuerce la validez del estudio.

En la actualidad, el instrumento más reciente para el proceso de estimación del riesgo es la Resolución de Secretaría de Gestión del Riesgo de Desastres N.º 009-2025-PCM/SGRD, que aprueba los Lineamientos Técnicos del Proceso de Estimación del Riesgo de Desastres. Esta resolución establece disposiciones para la identificación y caracterización de peligros, el análisis de vulnerabilidad y la determinación de niveles de riesgo. Aunque la presente tesis no corresponde formalmente a un expediente administrativo de evaluación del riesgo elaborado para aprobación institucional, sí comparte con estos lineamientos una lógica metodológica compatible: delimitación espacial del área de análisis, identificación de factores condicionantes, construcción de cartografía temática e interpretación de zonas con diferente comportamiento territorial frente a los deslizamientos. Por ello, su incorporación fortalece la consistencia técnica del estudio y mejora su alineamiento con los estándares nacionales vigentes. De igual forma, continúan siendo pertinentes los Lineamientos Técnicos del Proceso de Reducción del Riesgo de Desastres, aprobados mediante Resolución Ministerial N.º 220-2013-PCM, y los Lineamientos Técnicos del Proceso de Prevención del Riesgo de Desastres, aprobados mediante Resolución Ministerial N.º 222-2013-PCM.

Estos documentos orientan a las entidades públicas en la adopción de medidas destinadas a disminuir la exposición, evitar la generación de nuevos riesgos y utilizar la información técnica como insumo para la planificación física y administrativa del territorio. En términos de la presente investigación, dichos lineamientos respaldan la utilidad aplicada de los mapas de susceptibilidad obtenidos, ya que estos pueden servir como base técnica para priorizar acciones de intervención, regular ocupaciones territoriales inadecuadas y mejorar los instrumentos locales de gestión del riesgo.

7.3. Manuales técnicos aplicables y su relación con la metodología del estudio

Entre los documentos técnicos más relevantes emitidos por el CENEPRED destaca el Manual para la Evaluación de Riesgos Originados por Fenómenos Naturales – 2da versión, aprobado mediante Resolución Jefatural N.º 112-2014-CENEPRED/J. Este manual constituye una herramienta de referencia para la identificación del peligro, el análisis de vulnerabilidad, la estimación del riesgo y su representación cartográfica. Su importancia para la presente tesis radica en que proporciona criterios metodológicos que refuerzan la necesidad de delimitar adecuadamente el área de estudio, caracterizar el medio físico, identificar factores condicionantes y representar espacialmente los resultados de manera técnicamente sustentada. Esto resulta particularmente importante en un contexto como Quinistaquillas, donde la recurrencia de deslizamientos y huaicos exige herramientas cada vez más precisas para apoyar la toma de decisiones.

7.4. Vinculación del marco normativo y técnico con la presente investigación

La incorporación de este marco normativo y técnico no debe entenderse como un componente meramente declarativo dentro del capítulo teórico, sino como un respaldo directo a la lógica metodológica de la investigación.

En efecto, la delimitación de las microcuencas de las quebradas Volcán y Juinco, la selección de variables geoespaciales condicionantes, la construcción de mapas temáticos y la comparación de modelos predictivos responden a la necesidad de generar información espacialmente explícita para identificar sectores con diferente nivel de susceptibilidad a deslizamientos. Esta orientación es coherente con los enfoques impulsados por el SINAGERD, la Política Nacional de Gestión del Riesgo de Desastres al 2050, el PLANAGERD 2022–2030 y los lineamientos emitidos por el CENEPRED.

Desde una perspectiva aplicada, el aporte de la presente tesis consiste en trasladar esos lineamientos generales al análisis territorial local mediante herramientas contemporáneas de procesamiento geoespacial y aprendizaje automático. En otras palabras, el estudio no reemplaza los procedimientos oficiales de evaluación del riesgo, pero sí contribuye con un instrumento técnico complementario que puede apoyar la identificación preliminar de sectores críticos, la priorización de intervenciones y la formulación de estrategias locales de prevención y reducción del riesgo. Esta articulación entre normativa, lineamientos técnicos y modelamiento espacial fortalece la utilidad práctica de la investigación y responde de manera más sólida a la observación formulada sobre la necesidad de integrar la normativa vigente y los manuales técnicos aplicables.

7.5. Alcance de la aplicación normativa en el estudio

Por un lado, desarrolla un análisis científico apoyado en variables geoespaciales, imágenes satelitales y algoritmos de clasificación; por otro, se alinea con la normativa peruana vigente y con los lineamientos técnicos nacionales que promueven la generación de evidencia para la prevención y reducción del riesgo. Esta doble condición incrementa la pertinencia del estudio, ya que permite que sus resultados no solo tengan valor académico, sino también potencial de aplicación en el ámbito de la gestión territorial y de la planificación preventiva en el distrito de Quinistaquillas.

CAPÍTULO III

1. TIPO Y NIVEL DE INVESTIGACIÓN

La presente investigación fue de tipo aplicada, pues estuvo orientada a la solución de un problema práctico concreto – la identificación de zonas susceptibles a deslizamientos mediante la generación de conocimiento con aplicación directa a necesidades de la sociedad. Asimismo, el estudio tuvo un alcance correlacional-predictivo, ya que buscó determinar la relación estadística entre variables geoespaciales (índices espectrales, factores topográficos, etc.) y la ocurrencia de deslizamientos para predecir áreas de riesgo. En tal sentido, no se manipularon deliberadamente las variables (diseño no experimental), sino que se analizaron datos observacionales para encontrar asociaciones significativas que permitan explicar y predecir el fenómeno de interés. Este enfoque correlacional es apropiado cuando se pretende medir el grado de asociación entre múltiples variables y un evento, como es el caso de la susceptibilidad a deslizamientos. Debido a su naturaleza aplicada y correlacional, la investigación se desarrolló bajo un enfoque cuantitativo y explicativo, fundamentado en métodos de análisis de datos geoespaciales y modelamiento supervisado para inferir las relaciones causa-efecto subyacentes.

2. UBICACIÓN Y ÁREA DE ESTUDIO

Se estableció que el análisis se desarrolló específicamente sobre las microcuencas de las quebradas **Juinco** y **Volcán**, localizadas dentro del distrito de Quinistaquillas, provincia General Sánchez Cerro, departamento de Moquegua. La selección de estas unidades espaciales responde a que constituyen sectores hidrogeomorfológicos representativos para la evaluación de la susceptibilidad a deslizamientos, debido a la presencia de laderas de pendiente pronunciada, drenajes activos y antecedentes de eventos asociados a lluvias intensas.

En ese marco, la Tabla 5 presenta la caracterización espacial básica del área de estudio, incorporando la denominación de cada microcuenca, su área aproximada y sus coordenadas centrales referenciales. Esta información permite contextualizar territorialmente el análisis, mejorar la trazabilidad metodológica y fortalecer la reproducibilidad del estudio.

Tabla 5

Ubicación y área de estudio

Categoría	Descripción
Área de estudio específica	Microcuencas de las quebradas Volcán y Juinco, ubicadas en el distrito de Quinistaquillas, provincia General Sánchez Cerro, departamento de Moquegua, Perú
Ubicación geográfica	Región andina del sur del Perú, con relieve montañoso, laderas de fuerte pendiente y presencia de quebradas estacionales
Extensión aproximada total del área de estudio	43.02 km ² (4,302.34 ha)
Área aproximada de la microcuenca de la quebrada Juinco	25.63 km ² (2,562.92 ha)
Área aproximada de la microcuenca de la quebrada Volcán	17.39 km ² (1,739.42 ha)

Coordenada central referencial de la microcuenca de la quebrada Juinco	E 300,457 m ; N 8,154,174 m
Coordenada central referencial de la microcuenca de la quebrada Volcán	E 299,547 m ; N 8,146,959 m
Sistema de referencia cartográfica	Coordenadas UTM, referidas a la grilla cartográfica del mapa del área de estudio
Clima	Árido a semiárido, con precipitaciones estacionales concentradas en época de lluvias
Estacionalidad de lluvias	Lluvias concentradas principalmente entre diciembre y marzo
Estiaje	Predominio de condiciones secas durante gran parte del año, fuera de la temporada de lluvias
Tipo de suelo y geología	Presencia de materiales volcánicos y sedimentarios, asociados a condiciones de inestabilidad en sectores con pendiente pronunciada
Hidrología local	Presencia de cauces temporales y drenajes estacionales que estructuran las microcuencas analizadas
Problema principal	Riesgo de deslizamientos y huaicos asociados a lluvias intensas estacionales y a las condiciones geomorfológicas del terreno

Eventos históricos recientes	Reportes de afectación por lluvias intensas, huaicos y deslizamientos en 2022 y 2024, con impactos en vías, canales de riego, cultivos y servicios básicos en Quinistaquillas
Justificación del estudio	Necesidad de identificar sectores críticos con susceptibilidad a deslizamientos para fortalecer la prevención, la planificación territorial y la gestión local del riesgo
Importancia del contexto físico	Fundamental para la interpretación de variables geoespaciales, el análisis satelital y la validación espacial de los resultados del modelo predictivo

Nota: *SENAMHI, INGEMMET, COEN, IGN, CENEPRED*

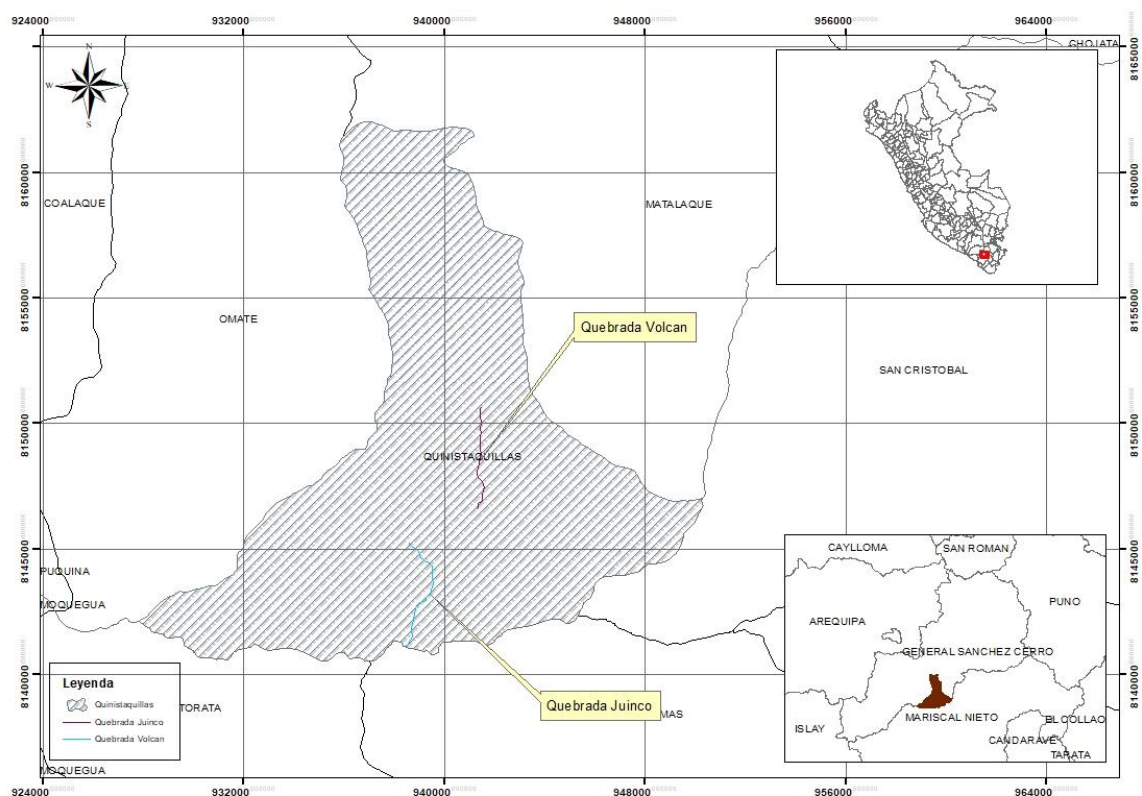
El estudio se llevó a cabo en las microcuencas de las quebradas Quinistaquillas y Juinco, ubicadas en el distrito de Quinistaquillas, provincia General Sánchez Cerro, en la región Moquegua, al sur de Perú. Esta zona se caracteriza por un relieve montañoso andino y condiciones climáticas áridas. La precipitación anual es baja (en el orden de 300–400 mm) y se concentra principalmente entre los meses de diciembre a marzo. La delimitación espacial del área de estudio se presenta en la Fig 4, donde se visualizan las microcuencas seleccionadas para el análisis.

Fuera de la temporada de lluvias, predominan condiciones secas con escasa precipitación, lo cual limita la influencia directa de factores pluviométricos en la inestabilidad de laderas. Las microcuencas analizadas comprenden terrenos de fuerte pendiente y material geológico volcánico y sedimentario, atravesados por cauces estacionales (quebradas) que

drenan las precipitaciones estivales. Estas características geográficas y climáticas hacen de Quinistaquillas un entorno propenso a deslizamientos disparados episódicamente por lluvias intensas estacionales, en contraste con largos periodos de estiaje. La selección de esta área específica responde a la necesidad local de identificar sectores críticos ante deslizamientos, dado antecedentes de eventos disruptivos en temporadas lluviosas recientes (e.g., deslizamientos reportados en enero de 2020 y 2024 en el distrito). El conocimiento detallado de la ubicación geográfica y condiciones del terreno fue fundamental para contextualizar el análisis de los datos satelitales y validar los hallazgos en campo.

Figura 3.

Delimitación del área de estudio, correspondiente al distrito de Quinistaquillas y las respectivas quebradas Juinco y Volcan



Nota: *Elaboración propia*

3. UNIDAD DE ANÁLISIS

Se usó una grilla regular (fishnet) de celdas de $10\text{ m} \times 10\text{ m}$, coincidiendo con el tamaño de píxel de las imágenes Sentinel-2 y los índices NDVI/NDWI.

Cada celda representó una unidad de análisis donde se registraron:

Variables explicativas: índices espectrales y factores geofísicos.

Variable dependiente (en muestras etiquetadas): presencia o ausencia de deslizamiento.

Las celdas cubrieron toda el área de estudio, pero para entrenar los modelos se eligió un subconjunto representativo:

Muestras positivas: zonas con deslizamientos históricos.

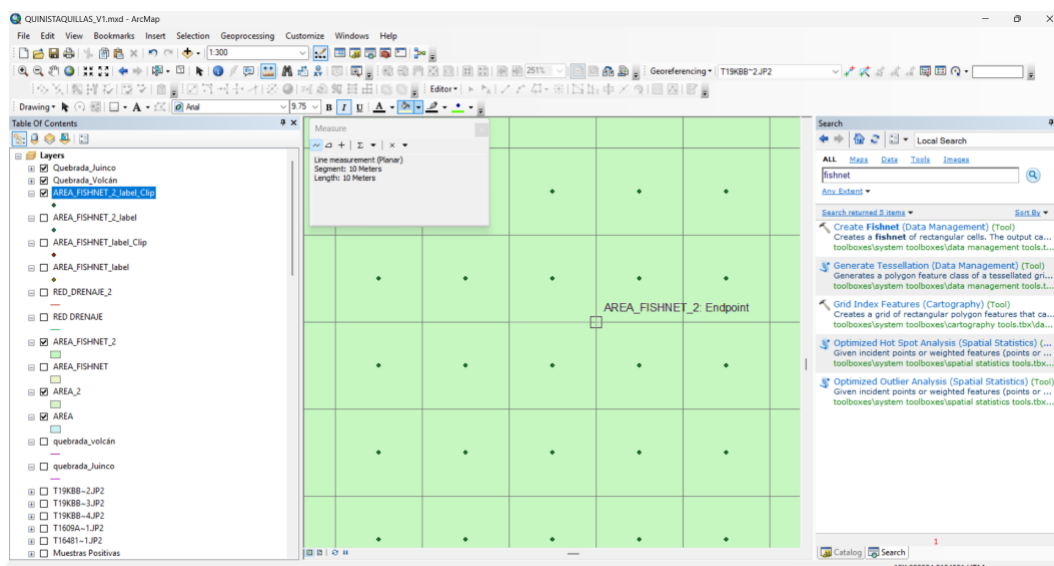
Muestras negativas: zonas estables sin afectación.

Cada muestra fue georreferenciada con coordenadas UTM, lo que permitió relacionarla con datos raster y capas GIS.

El uso de esta unidad espacial homogénea garantizó la consistencia en la extracción de datos y facilitó la creación de mapas continuos de susceptibilidad a deslizamientos.

Figura 4

Grilla Fishnet



Nota: *Elaboración propia*

4. VARIABLES DE ESTUDIO E INDICADORES

Las variables Las variables consideradas en la presente investigación fueron seleccionadas con base en evidencia científica y en estudios previos sobre modelamiento de susceptibilidad a deslizamientos mediante técnicas de análisis espacial y algoritmos de Machine Learning (Mosavi et al., 2018; Reichenbach et al., 2018; Madhuri et al., 2021). Su elección respondió a tres criterios principales: i) relevancia geomorfológica y ambiental en la ocurrencia de deslizamientos, ii) disponibilidad de información geoespacial homogénea para todo el ámbito de estudio y iii) compatibilidad con la escala espacial de análisis desarrollada en las microcuencas de las quebradas Volcán y Juinco.

En ese marco, se priorizaron variables de tipo topográfico, biofísico y de cobertura superficial, obtenidas a partir de imágenes satelitales Sentinel-2, modelos digitales de elevación y cartografía temática. La variable dependiente correspondió a la ocurrencia binaria de deslizamientos, expresada como presencia o ausencia del evento en puntos etiquetados para entrenamiento y validación del modelo. Por su parte, las variables independientes predictoras principales fueron: NDVI, NDWI, índice de brillo, pendiente del terreno, uso/cobertura del

suelo, tipo de suelo o geología y coordenadas espaciales X–Y. Estas variables fueron integradas debido a que representan condiciones relacionadas con la cobertura vegetal, la humedad superficial, la exposición del suelo, la inclinación de laderas, la naturaleza del material superficial y la distribución espacial del fenómeno.

La selección de estas variables se sustenta en que la pendiente constituye uno de los factores más influyentes en la inestabilidad de laderas, debido a que condiciona el equilibrio del material y favorece los procesos de remoción en masa en terrenos escarpados. Del mismo modo, el tipo de suelo o geología resulta determinante porque controla propiedades como cohesión, permeabilidad, meteorización y resistencia al corte. El uso/cobertura del suelo también fue incorporado por su relación con la protección o degradación de la superficie, mientras que los índices NDVI y NDWI permitieron representar de manera indirecta la cobertura vegetal y la humedad superficial, dos condiciones estrechamente vinculadas a la estabilidad del terreno. Por su parte, el índice de brillo aportó información sobre superficies desnudas o alteradas, útiles para discriminar sectores con mayor exposición del suelo. Finalmente, las coordenadas espaciales fueron incorporadas como variables numéricas normalizadas para captar patrones espaciales residuales no explicados completamente por las demás capas.

Asimismo, durante el proceso de delimitación y caracterización espacial del área de estudio se evaluaron variables como la orientación del terreno (aspect) y la distancia a quebradas o drenajes. Sin embargo, estas no fueron incorporadas como variables predictoras principales del entrenamiento final, sino como criterios auxiliares de análisis territorial, debido a que su función fue complementar la interpretación geomorfológica de las microcuencas y apoyar la delimitación de zonas con potencial de afectación. Esta decisión metodológica permitió evitar redundancias y mantener coherencia con la lógica del modelo final.

De igual manera, se consideró la pertinencia de otras variables frecuentemente utilizadas en estudios de susceptibilidad a deslizamientos, como la precipitación, la distancia a fallas geológicas y variables climáticas más específicas. No obstante, estas no fueron incorporadas como predictores principales debido a limitaciones relacionadas con la resolución espacial, la disponibilidad de datos homogéneos a escala local y la naturaleza del estudio. En el caso de la precipitación, si bien constituye un factor detonante fundamental, su variabilidad temporal y la ausencia de registros locales espacialmente detallados para ambas microcuencas dificultaban su integración consistente dentro de un modelo centrado en susceptibilidad espacial. En cuanto a la cobertura vegetal, esta sí fue representada indirectamente mediante el NDVI, por lo que no fue necesario incorporarla como variable separada. Respecto a la distancia a fallas geológicas, no se contó con evidencia cartográfica estructural suficientemente detallada y homogénea que justificara su incorporación a la escala específica del análisis. En el caso de la distancia a drenajes o quebradas, esta fue considerada como criterio auxiliar de interpretación territorial y no como variable predictora principal, debido a que parte de su efecto ya se encontraba reflejado en la configuración geomorfológica, la pendiente y la dinámica espacial de las microcuencas.

En consecuencia, la selección final de variables buscó equilibrar la relevancia teórica, la consistencia espacial de los datos y la aplicabilidad metodológica en el área de estudio. Los indicadores fueron tratados como variables continuas o categóricas según su naturaleza, y normalizados cuando correspondía, a fin de integrarlos adecuadamente en los algoritmos supervisados utilizados en la investigación. El detalle de cada variable, su tipo, unidad de medida, fuente de información y función dentro del estudio se presenta en la Tabla 6.

Tabla 6*Variables Independientes e Indicadores*

Variable	Tipo	Indicador	Unidad / tipo de dato	Fuente	Función en el estudio
Ocurrencia de deslizamientos	Dependiente	Existencia de eventos históricos	Binaria (0 = no ocurrió; 1 = ocurrió)	INGEMME T / CENEPRE D / validación local	Variable objetivo para entrenamiento y validación del modelo
NDVI	Independiente	Índice de vegetación de diferencia normalizada	Continuo, rango de -1 a 1	Sentinel-2 procesado en QGIS	Representa cobertura vegetal; menores valores pueden asociarse a menor estabilidad superficial
NDWI	Independiente	Índice de agua de diferencia normalizada	Continuo, rango de -1 a 1	Sentinel-2 procesado en QGIS	Representa humedad superficial; valores altos pueden reflejar

					mayor saturación del terreno
Índice de brillo	Independiente	Reflectancia superficial relativa	Continuo (float)	Sentinel-2 procesado en QGIS	Permite identificar superficies desnudas, erosionadas o con escasa cobertura
Pendiente del terreno	Independiente	Inclinación de la superficie	Grados (°)	ALOS PALSAR / SRTM	Factor topográfico clave en la inestabilidad de laderas
Uso/cobertura del suelo	Independiente	Tipo de cobertura terrestre dominante	Categórica	MINAM / clasificación supervisada	Expresa la influencia de la cobertura y del uso antrópico sobre la estabilidad del terreno

Tipo de suelo / geología	Independent e	Material superficial dominante	Categórica	INGEMME T / cartografía temática	Representa propiedades del material como cohesión, permeabilidad y resistencia al corte
Coordenadas espaciales X, Y	Independent e	Localización UTM Este y Norte	Continuas normalizada s	Grilla de muestreo (Fishnet)	Captan patrones espaciales residuales no explicados por otras variables
Orientación del terreno (Aspect)	Auxiliar	Exposición de la ladera	Grados o categorías	ALOS PALSAR / SRTM	Usada como criterio complementari o de interpretación territorial, no como predictor principal

Distancia a quebradas / drenajes	Auxiliar	Proximidad a cauces temporales o permanentes	Metros	Drenaje derivado de MDE	Empleada como criterio auxiliar para delimitación e interpretación geomorfológica
Precipitación	No incorporada como predictor principal	Condición detonante de deslizamientos	mm	SENAMHI / datos secundarios	No incluida por limitaciones de resolución espacial y consistencia local para el periodo y escala de análisis
Distancia a fallas geológicas	No incorporada como predictor principal	Proximidad a estructuras tectónicas	Metros	Cartografía estructural secundaria	No incluida por falta de cartografía estructural detallada y homogénea a

Nota: *Elaboración propia*

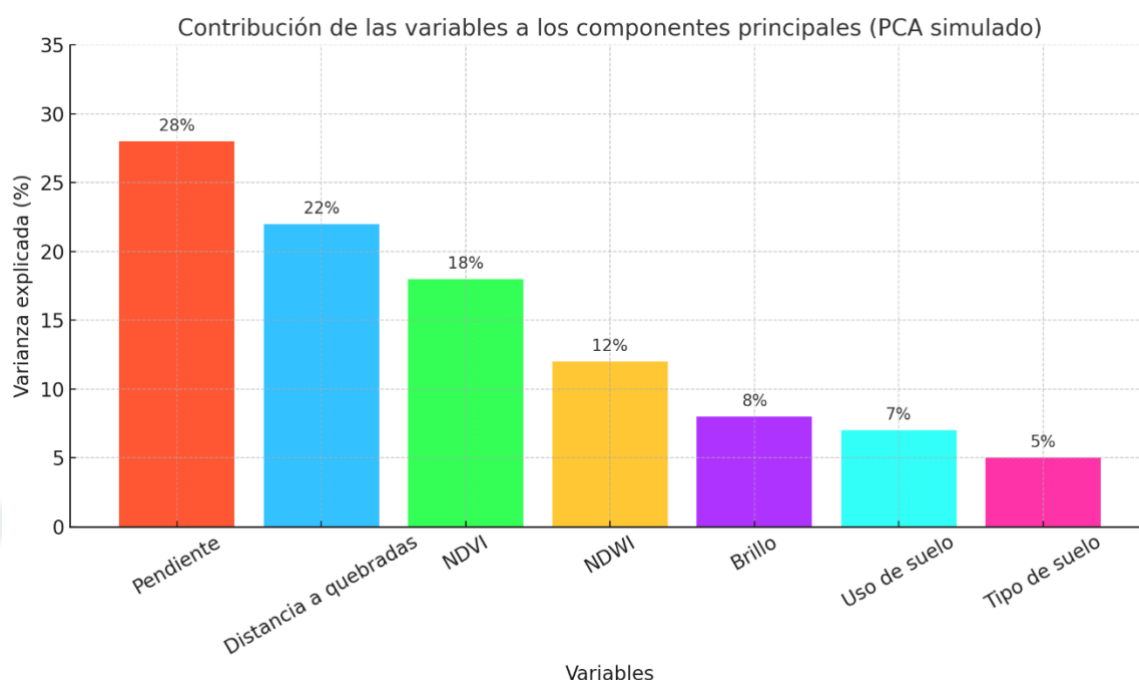
Nota. Elaboración propia con base en información satelital, cartografía temática y fuentes institucionales utilizadas en la investigación.

Como se observa en la Tabla 6, las variables predictoras principales combinan información biofísica, topográfica y espacial, lo que permitió representar de manera integral las condiciones que favorecen la susceptibilidad a deslizamientos en el área de estudio. La incorporación de índices espectrales, atributos del relieve y variables temáticas de suelo y cobertura superficial favoreció una aproximación multivariable coherente con la complejidad del fenómeno analizado.

Finalmente, aunque durante el análisis exploratorio se evaluó la posibilidad de aplicar técnicas de reducción de dimensionalidad, como el **Análisis de Componentes Principales (PCA)**, la revisión de la matriz de correlación y la importancia relativa estimada mediante el modelo Random Forest indicaron que las variables seleccionadas conservaban aporte explicativo suficiente y no presentaban redundancia crítica que justificara su eliminación. En ese sentido, se optó por mantener el conjunto final de variables, priorizando su interpretabilidad y su utilidad práctica en la generación de mapas de susceptibilidad. No obstante, se reconoce que futuros estudios podrían incorporar pruebas adicionales de selección de características, reducción de dimensionalidad o análisis de sensibilidad para optimizar aún más la capacidad de generalización de los modelos.

Figura 5

Conjunto por fecha (Brightness)



Nota: *Elaboración propia*

5. METODOLOGÍA PARA LOS OBJETIVOS ESPECÍFICOS

5.1. Preprocesamiento de datos geospaciales

- Ocurrencia de deslizamiento

Recolección de datos históricos: Se obtuvo información sobre la existencia y ubicación de eventos históricos de deslizamientos en el distrito de Quinistaquillas a partir de bases de datos oficiales del INGEMMET, CENEPRED y validaciones locales de campo.

Construcción del inventario de deslizamientos: Se elaboró un inventario georreferenciado de los puntos donde se había registrado ocurrencia de deslizamientos pasados, etiquetando cada punto con un valor binario (0: no ocurrió, 1: sí) para el entrenamiento y validación de modelos de Machine Learning.

Uso de imágenes Sentinel-2: Se procesaron las imágenes multiespectrales Sentinel-2 L2A para la delimitación espacial de la cobertura terrestre y la identificación preliminar de zonas potencialmente afectadas o alteradas por deslizamientos, utilizando QGIS para el tratamiento y análisis espacial.

Integración en SIG: Los puntos históricos y las áreas derivadas de imágenes satelitales se integraron en un SIG mediante ArcMap, permitiendo relacionar la ocurrencia de deslizamientos con variables ambientales y geográficas.

Etiquetado para modelos ML: El inventario generado sirvió para etiquetar los datos en un modelo supervisado, preparando la variable dependiente "Ocurrencia de deslizamiento" para su uso en algoritmos de aprendizaje automático, estableciendo una base sólida para evaluar patrones de riesgo y predecir zonas vulnerables a deslizamientos.

- **NDVI**

Adquisición de imágenes satelitales: Se descargaron las imágenes multiespectrales Sentinel-2 L2A correspondientes al área de estudio en las fechas adecuadas para evitar contaminación por nubes y representar condiciones representativas de la vegetación en la zona.

Preprocesamiento de imágenes: Se realizó la corrección atmosférica y la selección de bandas espectrales necesarias para el cálculo del NDVI (banda del infrarrojo cercano y banda roja visible), utilizando QGIS y sus complementos especializados para el tratamiento de imágenes satelitales.

Cálculo del NDVI: Se aplicó la fórmula standard del NDVI, que se calcula como $(NIR - Red) / (NIR + Red)$, generando un mapa raster donde cada píxel toma un valor continuo entre -1 y 1, indicando la cobertura y vigor de la vegetación.

Análisis espacial del NDVI: Se evaluaron los valores de NDVI para identificar zonas con mayor cobertura vegetal, cuya presencia generalmente se asoció con menor susceptibilidad a deslizamientos debido a la estabilización del suelo por la vegetación.

Integración con otras variables: Los datos de NDVI se integraron en el modelo de Machine Learning junto con otras variables independientes para ayudar a predecir la ocurrencia de deslizamientos, contribuyendo a la clasificación y evaluación del riesgo espacial.

Validación y ajuste: Se compararon los mapas generados con datos de campo y la ocurrencia histórica de deslizamientos para validar la utilidad del NDVI como predictor en el modelo, realizando ajustes en la selección temporal de las imágenes y el procesamiento para optimizar la sensibilidad del índice.

- **NDWI**

Adquisición de imágenes satelitales: Se descargaron imágenes Sentinel-2 L2A correspondientes al área de estudio en fechas seleccionadas para evitar nubosidad y representar condiciones representativas de humedad en la superficie.

Preprocesamiento de imágenes: Se realizaron correcciones atmosféricas y se seleccionaron las bandas espectrales necesarias para el cálculo del NDWI (banda verde visible y banda del infrarrojo cercano, específicamente bandas 3 y 8 de Sentinel-2), empleando QGIS para el procesamiento y análisis espacial.

Cálculo del NDWI: Se aplicó la fórmula $NDWI = (Verde - NIR) / (Verde + NIR)$, obteniendo un mapa raster con valores continuos entre -1 y 1, donde los valores cercanos a 1 indican alta humedad superficial y presencia de agua, mientras que valores negativos o bajos indican baja humedad o ausencia de agua.

Interpretación y clasificación: Se analizaron los valores de NDWI para distinguir zonas con humedad superficial significativa, que suelen asociarse a mayor susceptibilidad a deslizamientos por saturación de suelos, utilizando umbrales para categorizar áreas según su contenido hídrico.

Integración con otras variables: Los datos de NDWI se integraron como variable independiente en los modelos de Machine Learning junto con otras variables ambientales y geográficas, para mejorar la predicción y evaluación del riesgo de deslizamientos.

Validación y ajuste: Se contrastaron los resultados del mapa NDWI con datos de campo y ocurrencia histórica de deslizamientos para validar su efectividad como indicador de humedad superficial en el contexto del estudio, haciendo ajustes en la selección temporal y el procesamiento según fuera necesario.

- **Brillo**

Adquisición y selección de imágenes: Se descargaron imágenes Sentinel-2 L2A que cubrirían el área de estudio en períodos representativos, procurando minimizar la presencia de nubes y condiciones que puedan afectar la reflectancia superficial.

Preprocesamiento: Se aplicaron correcciones radiométricas y atmosféricas sobre las imágenes para asegurar la precisión en los valores de reflectancia, utilizando herramientas y complementos específicos en QGIS.

Cálculo del Índice de Brillo: Se emplearon las bandas espectrales necesarias de Sentinel-2 para calcular un índice de brillo que reflejara la intensidad espectral de la reflectancia superficial, con valores relativos en formato float. Este índice permitió destacar áreas con mayor exposición superficial, como zonas de suelo desnudo o erosionado.

Análisis espacial: Se elaboró un mapa raster con el índice de brillo, identificando áreas con mayor reflectancia que generalmente están relacionadas con falta de cobertura vegetal y posible erosión, aumentando así la susceptibilidad a deslizamientos.

Integración en modelos de Machine Learning: Los datos del índice de brillo se incorporaron como variable independiente en los modelos predictivos para evaluar su contribución a la susceptibilidad y ocurrencia de deslizamientos.

Validación: Se corroboró la utilidad del índice de brillo mediante comparación con datos de campo y eventos históricos, ajustando los parámetros de procesamiento para mejorar la sensibilidad y especificidad del índice como predictor.

- **Pendiente del Terreno**

Obtención de Modelo Digital de Elevación (MDE): Se descargaron datos cartográficos de elevación provenientes de productos como SRTM (Shuttle Radar Topography Mission) y ALOS PALSAR, que ofrecen una resolución espacial adecuada para el área de estudio.

Procesamiento y generación del mapa de pendiente: A partir del MDE, se emplearon herramientas SIG en QGIS y ArcMap para calcular la pendiente del terreno en grados, utilizando funciones integradas de análisis de terreno para generar un mapa raster de pendiente continua.

Reclasificación por clases: Se reclasificaron los valores de pendiente en categorías con rangos establecidos ($0-15^\circ$, $15-30^\circ$, $>30^\circ$), facilitando la interpretación y modelación, dado que pendientes mayores se asocian con mayor riesgo de deslizamientos.

Análisis espacial e integración con otras variables: El mapa de pendiente fue integrado con otras variables independientes y la variable dependiente (ocurrencia de deslizamientos) dentro del SIG, para su uso en modelos de Machine Learning.

Validación: Se validó la relevancia del índice de pendiente mediante comparación con eventos históricos de deslizamientos y datos de campo, constatando que áreas con pendientes más pronunciadas eran más vulnerables a movimientos en masa.

- **Cobertura del Suelo**

Se recopilieron datos temáticos del uso y cobertura del suelo a partir del Mapa Nacional de Cobertura Vegetal 2019 elaborado por el Ministerio del Ambiente del Perú (MINAM, 2019). Asimismo, se realizaron clasificaciones supervisadas empleando imágenes satelitales Sentinel-2 L2A procesadas en el software QGIS.

Clasificación supervisada: Se definieron clases de cobertura terrestre relevantes para el área de estudio, tales como zonas agrícolas, bosques, áreas urbanas, zonas de vegetación densa, zonas desnudas o erosionadas, entre otras. Se aplicaron técnicas de clasificación supervisada en QGIS para mapear estas categorías de manera precisa.

Codificación para modelos de Machine Learning: Las categorías de cobertura se codificaron utilizando esquemas de codificación adecuados para modelos predictivos, como one-hot encoding o etiquetado numérico, para facilitar la integración de esta variable independiente en los algoritmos de aprendizaje automático.

Análisis espacial: Se incorporaron los datos de cobertura del suelo al SIG junto con otras variables independientes y la variable dependiente (ocurrencia de deslizamientos), permitiendo evaluar la influencia del uso del suelo en la susceptibilidad y ocurrencia de deslizamientos.

Validación: Se compararon los mapas de uso y cobertura del suelo con observaciones de campo y registros históricos de eventos para validar la precisión y relevancia del dato en el modelo de riesgo.

- **Tipo de Suelo**

Obtención de cartografía temática: Se recopiló información geológica y de tipos de suelo del área de estudio a partir de mapas temáticos y bases de datos oficiales del INGEMMET, incluidos los mapas de geología y tipos de suelo disponibles para Moquegua.

Identificación de materiales superficiales: Se clasificaron los tipos de suelo y formaciones geológicas presentes en el distrito, distinguibles por su composición (arcillas, limos, arenas, lutitas, areniscas, etc.), considerando características como la fragmentación, meteorización y presencia de capas heterogéneas.

Análisis de propiedades físicas y mecánicas: Se revisaron o realizaron ensayos geotécnicos (como corte directo y consolidación) para determinar parámetros críticos tales como cohesión y ángulo de fricción interna para los diferentes tipos de suelo, datos pertinentes para entender su resistencia al deslizamiento.

Integración en SIG: Los datos cartográficos y resultados geotécnicos se incorporaron en el SIG junto con las demás variables, generando capas temáticas codificadas para su uso como variables categóricas en modelos de Machine Learning.

Evaluación del impacto en deslizamientos: Se analizó cómo la naturaleza y características del tipo de suelo influían en la susceptibilidad a deslizamientos, considerando que materiales con menor resistencia al corte (por ejemplo, suelos altamente meteorizados, arcillas plásticas) presentaban mayor probabilidad de falla.

Validación y ajuste: Se contrastaron las zonas con distintos tipos de suelo y su comportamiento en eventos históricos de deslizamientos, para validar la representatividad de la variable y ajustar el modelado en consecuencia.

- **Orientación**

Obtención del Modelo Digital de Elevación (MDE): Se utilizaron datos topográficos de alta resolución obtenidos de ALOS PALSAR y SRTM para el área de estudio, asegurando una cobertura precisa del relieve y la configuración del terreno.

Cálculo de la orientación del terreno: A partir del MDE, se emplearon herramientas de análisis de terreno en software SIG (ArcMap y QGIS) para calcular la orientación de cada celda/píxel en grados, indicando la dirección en la que "mira" la pendiente, desde 0° a 360°, o clasificándolo en categorías cardinales principales y sus intermedios (N, NE, E, SE, S, SW, W, NW).

Clasificación en categorías: Se clasificaron las orientaciones obtenidas en octantes o categorías cardinales para facilitar el análisis y la correlación con la ocurrencia de deslizamientos, considerando que ciertas orientaciones pueden favorecer acumulación de humedad o condiciones específicas de exposición solar que afectan la estabilidad del suelo.

Análisis espacial e integración: La variable orientación se integró en el SIG junto con otras variables independientes y la variable dependiente (ocurrencia de deslizamientos) para su uso en modelos de Machine Learning. Se analizaba si ciertas orientaciones presentaban mayor correlación con incidencias de deslizamientos.

Validación y ajuste: Se verificó la utilidad de la variable mediante comparación con datos de campo y deslizamientos históricos, ajustando la clasificación y el procesamiento en función de los resultados para optimizar la capacidad predictiva del modelo.

- **Distancia a Quebradas**

Obtención y preparación del Modelo Digital de Elevación (MDE): Se descargaron datos topográficos de alta resolución del área de estudio, provenientes de fuentes como ALOS PALSAR o SRTM, configurados adecuadamente para extraer información hidrológica.

Delimitación de la red de drenaje: Se aplicaron herramientas hidrológicas en QGIS y ArcMap para extraer la red de drenaje del MDE, identificando las quebradas, ríos y cauces temporales o permanentes presentes en el distrito.

Cálculo de la distancia a quebradas: Se generó una capa raster que representó la distancia en metros desde cada punto o celda del área de estudio hasta el cauce más cercano, usando herramientas de cálculo espacial en el SIG.

Análisis espacial e integración en modelos: La variable distancia a quebradas se integró como variable independiente en el modelo de Machine Learning junto con otras variables ambientales y geográficas para evaluar su influencia en la susceptibilidad a deslizamientos.

Validación: Se contrastaron los resultados con datos históricos de ocurrencia de deslizamientos y datos de campo, confirmando que mayor cercanía a quebradas aumenta la probabilidad de remoción o falla.

- **Coordenadas X e Y**

Definición del sistema de referencia espacial: Se utilizó el sistema de coordenadas UTM en la zona correspondiente al distrito de Quinistaquillas, garantizando una ubicación precisa en metros Este (X) y Norte (Y), con datos normalizados para facilitar el análisis numérico.

Generación de la grilla de muestreo: Se creó una grilla de muestreo (fishnet) en el SIG (ArcMap y QGIS) que cubrió toda el área de estudio, asignando coordenadas X y Y a cada celda o punto muestreado, asegurando un muestreo homogéneo y representativo.

Asignación de coordenadas a puntos de interés: Se identificaron y georreferenciaron los puntos relacionados a la ocurrencia o ausencia de deslizamientos, colectando sus coordenadas espaciales exactas como variables independientes para los modelos de Machine Learning.

Análisis espacial: Se utilizó la información de las coordenadas para detectar patrones residuales espaciales que pudieran no ser explicados por otras variables ambientales, mejorando la precisión del modelo predictivo.

Integración en modelos: Las coordenadas normalizadas se incorporaron como variables numéricas continuas en los algoritmos supervisados, aportando información sobre la localización espacial en combinación con otras variables.

Validación: Se verificó la coherencia espacial de los datos con eventos históricos y trabajos de campo, asegurando que la variable espacial contribuyera significativamente a la modelación y predicción del riesgo de deslizamientos.

Figura 6

CSV muestras

OBJETOID	NDWI_20230	NDWI_202301	NDWI_202302	NDWI_202303	NDWI_20200	NDWI_20170	NDWI_20160	NDWI_201703	NDWI_202303	NDWI_202301	NDWI_202302	NDWI_202303	NDWI_202002	NDWI_202002	NDWI_20170
1	-0.146297	-0.1930086	-0.1273601	-0.1220363	-0.1748401	-0.1341048	-0.1777625	-0.1456773	-0.1570403	0.0936902	0.077134	0.0900442	0.0375987	0.041087	0.1318042
2	-0.1436972	-0.2013756	-0.1186676	-0.1137088	-0.1827027	-0.133564	-0.1844864	-0.1491977	-0.1535557	0.1067581	0.0805556	0.1019979	0.0346598	0.0359143	0.1555227
3	-0.143171	-0.1763417	-0.1261387	-0.107438	-0.1928952	-0.1334477	-0.1880342	-0.1576439	-0.1571653	0.0962866	0.071012	0.0929444	0.0364399	0.0327553	0.1720768
4	-0.149543	-0.2004381	-0.1260091	-0.1162632	-0.2063158	-0.1482864	-0.182849	-0.173207	-0.1629677	0.099111	0.0745044	0.1126904	0.0348387	0.0410175	0.1912682
5	-0.1571835	-0.2045538	-0.1244304	-0.1242804	-0.2221835	-0.1501707	-0.2045214	-0.1820347	-0.1664015	0.1118357	0.0792703	0.1101709	0.0442708	0.0397745	0.1950887
6	-0.1471725	-0.1882136	-0.1232394	-0.106398	-0.2015288	-0.1550546	-0.204431	-0.1719064	-0.1658196	0.1240602	0.0800265	0.0946725	0.0431655	0.0374544	0.1785958
7	-0.1389199	-0.2010279	-0.129822	-0.1086542	-0.2001464	-0.1443551	-0.2012491	-0.1628552	-0.1625964	0.1328534	0.0801837	0.1071429	0.043508	0.0290163	0.1651865
8	-0.1438652	-0.2003173	-0.1308901	-0.1121043	-0.1976912	-0.1513457	-0.1968146	-0.1702571	-0.1645893	0.1203408	0.0751658	0.0951864	0.0409639	0.0386087	0.1731449
9	-0.1429641	-0.1922786	-0.1195315	-0.1235033	-0.1806199	-0.1326358	-0.2023685	-0.1522998	-0.1656749	0.1164295	0.0682057	0.0769231	0.0356915	0.0327044	0.1333789
10	-0.1439252	-0.1859334	-0.1181102	-0.1183452	-0.1790887	-0.1359946	-0.1849551	-0.1536899	-0.1641885	0.1124069	0.0606586	0.0804363	0.0345264	0.0345044	0.1380839
11	-0.1360947	-0.1794303	-0.1266735	-0.1103519	-0.1767196	-0.1335813	-0.1870748	-0.1512097	-0.1586164	0.1200257	0.0703833	0.0812277	0.0379507	0.0324015	0.1339225
12	-0.1495886	-0.2057121	-0.1138976	-0.1083246	-0.198574	-0.140411	-0.1829609	-0.158714	-0.1574362	0.11926	0.0927835	0.1023289	0.0299517	0.0334524	0.1726543
13	-0.1410864	-0.1933785	-0.1219858	-0.1111111	-0.2162257	-0.146767	-0.190326	-0.1767918	-0.1666667	0.1196835	0.0808011	0.0930393	0.0326371	0.0371567	0.2030705
14	-0.15253	-0.2032885	-0.1289746	-0.1174427	-0.2240896	-0.1533243	-0.2069324	-0.1903467	-0.1693472	0.1405175	0.0898451	0.1134163	0.0336932	0.0314304	0.2075993
15	-0.1613604	-0.2058383	-0.1286162	-0.1126469	-0.2340572	-0.1575342	-0.2246998	-0.1972973	-0.1690761	0.165524	0.0993151	0.1152648	0.0435063	0.0310599	0.2243502
16	-0.1588999	-0.2317511	-0.1198881	-0.1100226	-0.2174226	-0.1507627	-0.2263752	-0.1822863	-0.1670887	0.164379	0.1121701	0.1462509	0.0322165	0.0336842	0.212621
17	-0.1635141	-0.2459309	-0.1138777	-0.1153696	-0.2102207	-0.1527675	-0.2259222	-0.1914431	-0.1599732	0.1508798	0.1019108	0.1497942	0.0322362	0.0439239	0.1899505
18	-0.1506193	-0.2006499	-0.1215429	-0.1056851	-0.1946309	-0.1420495	-0.2250634	-0.1788732	-0.1736604	0.1423371	0.0718273	0.0948148	0.0412162	0.0347885	0.1655147
19	-0.124732	-0.1868045	-0.1064572	-0.0928758	-0.1809246	-0.1307347	-0.1877283	-0.149552	-0.1611493	0.1279917	0.0768051	0.0846349	0.0288867	0.0298777	0.1499468
20	-0.1263895	-0.1903397	-0.1087479	-0.1045219	-0.1950505	-0.1321023	-0.2043356	-0.1527581	-0.159264	0.1428571	0.0746269	0.0887308	0.0355655	0.0386894	0.1726126
21	-0.1501785	-0.2037306	-0.1224419	-0.1037313	-0.2133333	-0.1511879	-0.2087821	-0.1697708	-0.1630329	0.1438652	0.0856463	0.1033435	0.0279543	0.0235294	0.2017608
22	-0.1210191	-0.1834504	-0.1224417	-0.0983827	-0.1974834	-0.1467316	-0.1779693	-0.1648658	-0.1576577	0.1023843	0.0576175	0.0751773	0.0389105	0.024191	0.1753002
23	-0.1312971	-0.1698047	-0.1158769	-0.1036941	-0.1850847	-0.1246241	-0.180059	-0.1489933	-0.1504914	0.105716	0.0492327	0.0764249	0.0316597	0.026522	0.1473581
24	-0.1351164	-0.1811646	-0.1144039	-0.0993831	-0.1875641	-0.1374916	-0.1812601	-0.1505766	-0.156338	0.1098968	0.0757077	0.085922	0.0292807	0.0268886	0.1540505
25	-0.1367456	-0.1988201	-0.1185729	-0.1042988	-0.1964093	-0.1396648	-0.197227	-0.1702571	-0.1602909	0.1384311	0.076311	0.0924731	0.0346166	0.027878	0.1561416
26	-0.1406902	-0.1923225	-0.1226611	-0.1129141	-0.1978495	-0.1341128	-0.2063846	-0.1666667	-0.1511628	0.1371589	0.056922	0.0963643	0.0368	0.0195107	0.1496388
27	-0.1263158	-0.1895087	-0.1166375	-0.1203735	-0.1754626	-0.1374536	-0.1917029	-0.153457	-0.1543582	0.0990435	0.0604559	0.0805369	0.0293833	0.0306748	0.1335096
28	-0.1444444	-0.1920000	-0.1211000	-0.1063500	-0.1803125	-0.1323500	-0.1735000	-0.1554000	-0.1600000	0.1000000	0.0558631	0.0703400	0.0234531	0.0231132	0.1528418

Nota: Elaboración propia

6. ANÁLISIS DE PREDICCIÓN CON PYTHON

- **Preprocesamiento y construcción de la matriz de datos**

Se corroboró que no hubiera valores nulos; en caso de que una celda cayera justo en un píxel nublado en una imagen dada, se decidió usar un valor interpolado de fechas cercanas o marcarla como no disponible para esa fecha, evitando sesgos.

Cada registro en la matriz final contenía: identificador de la celda, series temporales de NDVI/NDWI/brillo (en las fechas seleccionadas de 2016–2023), valor de pendiente, código de uso de suelo, código de tipo de suelo, coordenadas X, Y, y la etiqueta de clase (1 = deslizamiento, 0 = no deslizamiento para las muestras de entrenamiento).

En total, la matriz incorporó **556 muestras positivas** y **3214 negativas** (relación ~1:6), balance que más adelante se atendió durante el modelado. Esta matriz constituyó la entrada para el entrenamiento de los modelos SVM y Random Forest.

Las líneas de código usadas para el procesamiento de los datos de entrenamiento y predicción son los siguientes:

Se incorporaron librerías de **manipulación de datos** (pandas, numpy), **visualización** (matplotlib) y **aprendizaje automático** (scikit-learn). Estas herramientas permitieron gestionar bases de datos, implementar modelos de clasificación y evaluar su rendimiento mediante métricas estadísticas.

En el presente estudio, el modelo base utilizado para la clasificación de la susceptibilidad a deslizamientos fue Random Forest (RF), debido a su capacidad para manejar variables heterogéneas, relaciones no lineales y datos espaciales con distinta naturaleza, además de su buen desempeño reportado en la literatura especializada. Como modelo comparativo, se empleó Support Vector Machine (SVM), con el fin de contrastar el rendimiento predictivo y evaluar la robustez del enfoque metodológico. Ambos algoritmos fueron entrenados con la

misma matriz de datos y con la misma partición de entrenamiento y validación, lo que permitió una comparación consistente de sus métricas de desempeño. La elección de Random Forest como modelo base se respaldó finalmente por sus mejores resultados en precisión, sensibilidad y desempeño general dentro del área de estudio.

```
# Importación de librerías

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from sklearn.ensemble import RandomForestClassifier

from sklearn.svm import SVC

from sklearn.model_selection import train_test_split

from sklearn.metrics import classification_report, confusion_matrix
```

- **Carga y preparación de datos de entrenamiento**

Se importaron bases de datos con muestras positivas y negativas, diferenciadas según la presencia o ausencia del fenómeno en estudio. A cada grupo se le asignó una etiqueta binaria (0 para negativas, 1 para positivas) y se integraron en un único conjunto de entrenamiento, garantizando uniformidad para los procesos de modelado.

```
# Carga de datos (separador ;)

negativas = pd.read_csv("Muestras_Negativas_full.csv", sep=";")

positivas = pd.read_csv("Muestras_Positivas_full.csv", sep=";")

# Asignación de etiquetas

negativas["clase"] = 0
```

```
positivas["clase"] = 1
```

```
# Unión de muestras
```

```
df_entrenamiento = pd.concat([negativas, positivas], ignore_index=True)
```

```
# Verificación de columnas
```

```
print(df_entrenamiento.columns)
```

- **Selección de variables predictoras**

Se seleccionaron como variables de entrada las mediciones de NDVI, NDWI, brillo, pendiente, uso actual de suelo y tipo de suelo.

```
variables = ["NDWI_20230325", "NDWI_20230305", "NDWI_20230109",
```

```
"NDVI_20230325", "NDVI_20230305", "NDVI_20230109",
```

```
"BRIGHTNESS_20230325", "BRIGHTNESS_20230305",
```

```
"SLOPE", "USO_ACTUAL_SUELO", "SUELO_TIPO"]
```

```
X = df_entrenamiento[variables]
```

```
y = df_entrenamiento["clase"]
```

Se definieron como variables de entrada los indicadores espectrales (NDVI, NDWI, brillo), características topográficas (pendiente) y atributos de uso y tipo de suelo. Estas variables fueron elegidas por su relevancia en la caracterización ambiental del área de estudio.

```
# División entrenamiento/prueba (75%-25%)
```

```
X_train, X_test, y_train, y_test = train_test_split(
```

```
X, y, test_size=0.25, random_state=42)
```

El conjunto fue dividido en **75% para entrenamiento** y **25% para prueba**, con el fin de asegurar un balance entre aprendizaje de los modelos y validación independiente de su desempeño.

- **Entrenamiento de modelos**

Se entrenaron dos algoritmos supervisados:

- ✓ **Random Forest**, configurado con 500 árboles de decisión.
- ✓ **SVM (Support Vector Machine)**, con kernel radial.

Ambos modelos fueron evaluados con métricas de clasificación para identificar su precisión y capacidad de generalización.

Random Forest

```
rf = RandomForestClassifier(n_estimators=500, random_state=42)

rf.fit(X_train, y_train)

print("Random Forest:\n", classification_report(y_test, rf.predict(X_test)))
```

SVM

```
svm = SVC(kernel='rbf', C=1, gamma='scale')

svm.fit(X_train, y_train)

print("SVM:\n", classification_report(y_test, svm.predict(X_test)))
```

- **Aplicación de modelos a datasets completos**

Los modelos entrenados se aplicaron a datasets de quebradas específicas (Juinco y Volcán), generando predicciones de riesgo para cada registro espacial. Los resultados se exportaron a archivos CSV, permitiendo su posterior análisis cartográfico.

```
# Carga de datasets completos
```

```

mc_juinco = pd.read_csv("MC_Quebrada_Juinco.csv", sep=";")

mc_volcan = pd.read_csv("MC_Quebrada_Volcan.csv", sep=";")

# Predicción con RF

mc_juinco["pred_rf"] = rf.predict(mc_juinco[variables])

mc_volcan["pred_rf"] = rf.predict(mc_volcan[variables])

# Guardar resultados

mc_juinco.to_csv("Predicciones_Quebrada_Juinco.csv", index=False)

mc_volcan.to_csv("Predicciones_Quebrada_Volcan.csv", index=False)

```

- **Evaluación estadística de los modelos**

El desempeño del modelo se evaluó mediante indicadores cuantitativos: **precisión, sensibilidad (recall) y coeficiente Kappa**. Estas métricas aseguraron la robustez de los resultados y la fiabilidad de las predicciones.

```

from sklearn.metrics import accuracy_score, recall_score, cohen_kappa_score

y_true = y_test

y_pred = rf.predict(X_test)

precision = accuracy_score(y_true, y_pred)

sensibilidad = recall_score(y_true, y_pred)

kappa = cohen_kappa_score(y_true, y_pred)

print(f"Precisión: {precision:.3f}")

print(f"Sensibilidad: {sensibilidad:.3f}")

print(f"Kappa: {kappa:.3f}")

```

- **Clasificación multinivel de riesgo**

Se estableció una clasificación ordinal en cuatro categorías (bajo, medio, alto y muy alto) a partir de las probabilidades de ocurrencia estimadas por el modelo Random Forest. Este enfoque permitió generar una escala de riesgo de mayor aplicabilidad práctica.

```
def clasificar_riesgo(prob):
```

```
    if prob < 0.25:
```

```
        return 0 # Bajo
```

```
    elif prob < 0.5:
```

```
        return 1 # Medio
```

```
    elif prob < 0.75:
```

```
        return 2 # Alto
```

```
    else:
```

```
        return 3 # Muy alto
```

```
probs_juinco = rf.predict_proba(mc_juinco[variables])[:, 1]
```

```
probs_volcan = rf.predict_proba(mc_volcan[variables])[:, 1]
```

```
mc_juinco["nivel_riesgo"] = [clasificar_riesgo(p) for p in probs_juinco]
```

```
mc_volcan["nivel_riesgo"] = [clasificar_riesgo(p) for p in probs_volcan]
```

- **Visualización en mapas**

Finalmente, los resultados se representaron de forma cartográfica mediante gráficas de dispersión georreferenciadas en coordenadas UTM. A cada nivel de riesgo se le asignó un color distintivo, generando un **mapa temático de riesgo multinivel** que facilita la interpretación espacial y la toma de decisiones en torno a la vulnerabilidad del territorio analizado.

```
color_dict = {0: 'green', 1: 'yellow', 2: 'orange', 3: 'red'}

# Volcán

colores_volcan = mc_volcan['nivel_riesgo'].map(color_dict)

plt.scatter(mc_volcan['Coordenada_X'], mc_volcan['Coordenada_Y'],

            c=colores_volcan, s=2, alpha=0.6)

plt.title("Mapa de Riesgo Multinivel - Quebrada Volcán")

plt.xlabel("Coordenada X (UTM)")

plt.ylabel("Coordenada Y (UTM)")

plt.show()
```

Así mismo la información contenida en este procesamiento se encuentra en el siguiente enlace:

<https://drive.google.com/drive/folders/1djrYA4JitYySD3658MZyfg6oXrzSR90X>

CAPÍTULO IV

1. RESULTADOS Y DISCUSIONES

1.1. Resultados I

Objetivo vinculado: Pre-procesamiento de datos geospaciales (NDVI, NDWI, Brightness, pendientes, uso de suelo, tipo de suelo)

- 1. Introducción breve de los resultados I

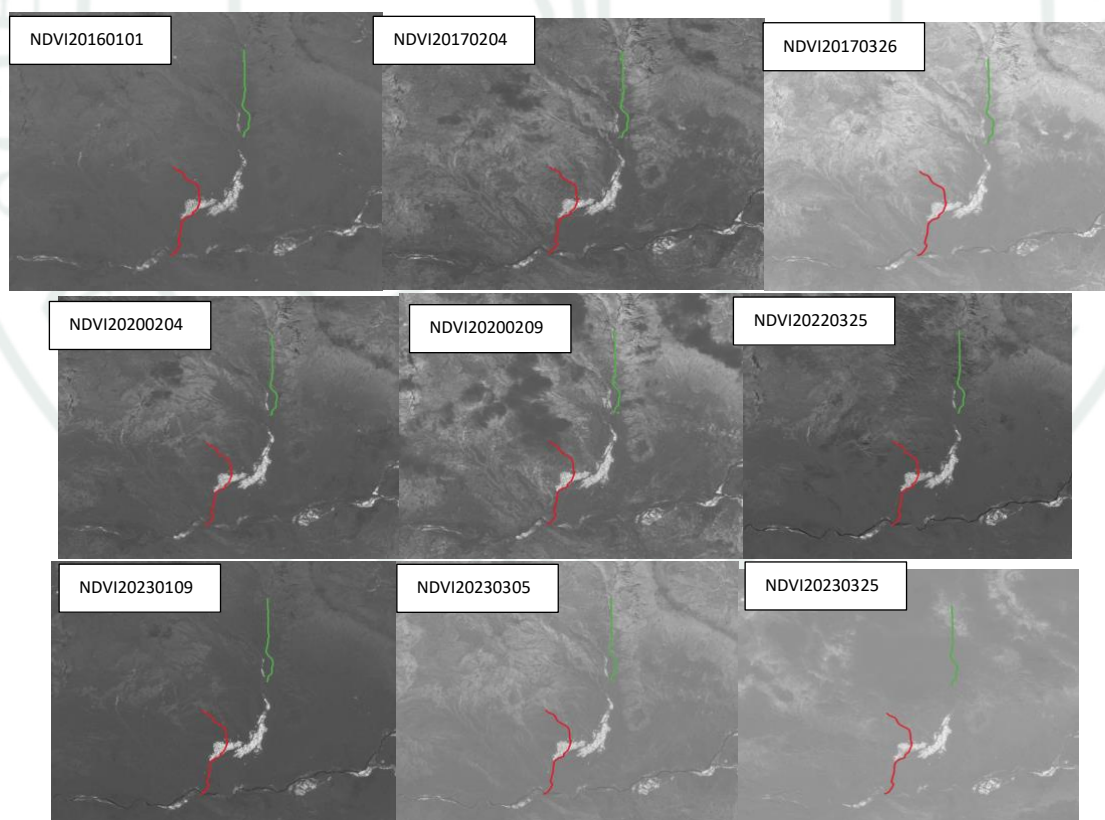
El presente apartado muestra los resultados del preprocesamiento de las imágenes Sentinel-2 L2A, así como de las capas topográficas y temáticas empleadas para el modelamiento posterior.

- 2. Muestra con capturas de pantalla o gráficos (ver anexos)

NDVI, NDWI y Brightness

Figura 7

Conjunto por Fecha (NDVI)

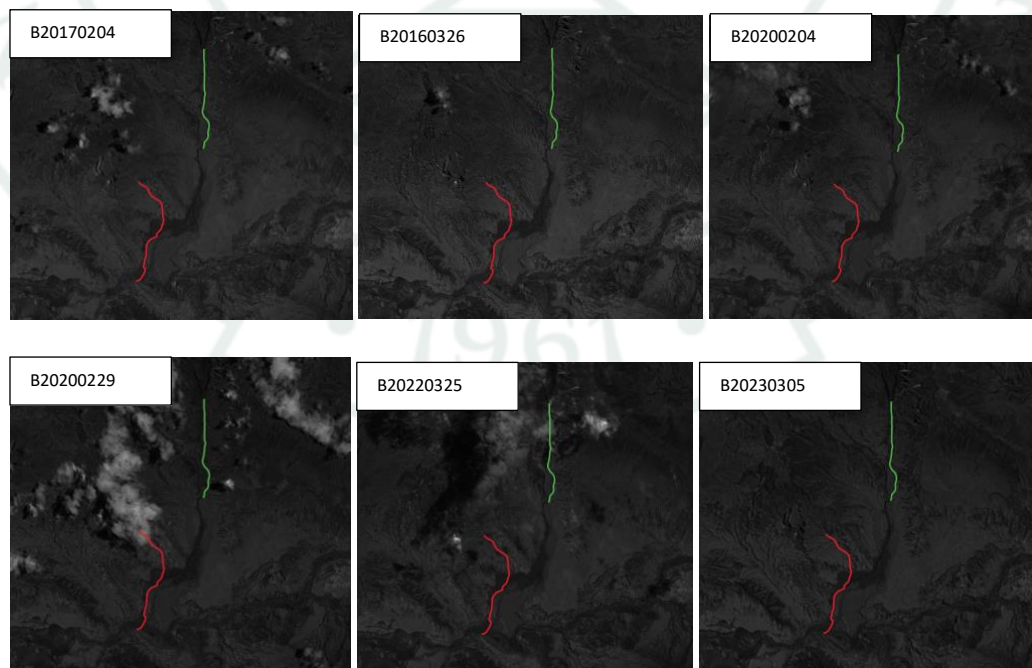


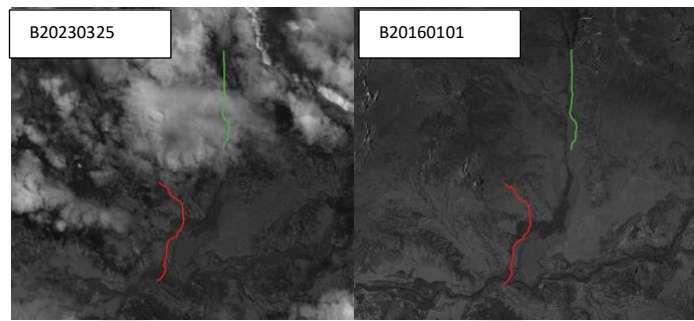
Nota: Elaboración propia

El análisis multitemporal del NDVI (ver Fig.9) permitió identificar patrones de variabilidad en la cobertura vegetal entre 2016 y 2023 en el entorno de las microcuencas evaluadas. Se evidenció una marcada diferenciación entre zonas de vegetación densa (tonos claros) y áreas degradadas o alteradas (tonos oscuros). Las imágenes de los años 2017, 2020 y 2023 muestran una mayor homogeneidad en la respuesta espectral, especialmente en los sectores altos de la cuenca. En cambio, los años 2016 y 2022 presentan contrastes fuertes en áreas específicas, lo que puede reflejar eventos de estrés hídrico o perturbaciones antrópicas. En la mayoría de los años se mantiene una vegetación activa en la parte alta y noreste, mientras que las zonas del cauce y las laderas inestables muestran menores valores de NDVI. Este comportamiento permitió inferir las zonas potencialmente erosionadas o expuestas a procesos de deslizamiento. Los polígonos de muestras positivas (rojo) se superponen generalmente con zonas de bajo NDVI. La distribución de las muestras coincide con áreas de vegetación menos densa, lo cual valida su selección como input para modelos predictivos.

Figura 8

Conjunto por fecha (Brightness)



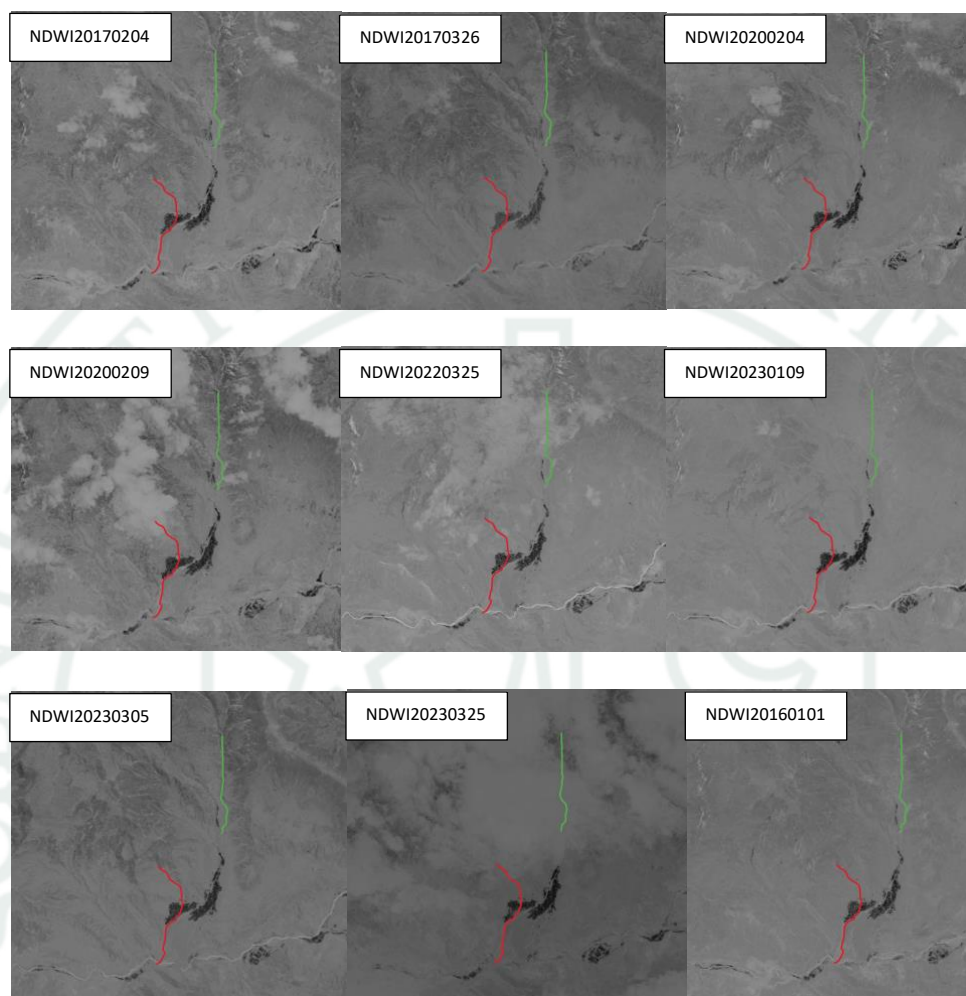


Nota: *Elaboración propia*

El índice de brillo (Brightness) permitió observar diferencias espectrales en la reflectancia general de la superficie terrestre, asociadas a la humedad del suelo, exposición solar y presencia de materiales expuestos. En la Fig.10 se presentan imágenes multitemporales del índice de brillo, donde se observa que las zonas de mayor brillo (tonos claros) suelen corresponder con superficies desnudas, áreas perturbadas o con escasa cobertura vegetal. En contraste, los sectores más oscuros reflejan coberturas densas o mayor retención de humedad superficial. Las imágenes de 2017, 2020 y 2023 se identificaron como focos brillantes recurrentes en las zonas de ladera oeste, lo que podría relacionarse con procesos de remoción de masa o escorrentía superficial. Los años 2016 y 2022 evidenciaron contrastes marcados entre zonas brillantes y oscuras, especialmente al interior de las microcuencas. La superposición de las muestras positivas (rojo) con zonas de alto brillo sugiere una posible correlación entre superficies reflectivas y áreas de deslizamiento. Esta variable aportó información complementaria a los índices NDVI y NDWI, enriqueciendo la capacidad de detección de condiciones críticas. La variabilidad temporal refuerza su valor como indicador dinámico en el entrenamiento de los modelos predictivos.

Figura 9

Conjunto por fecha (NDWI)



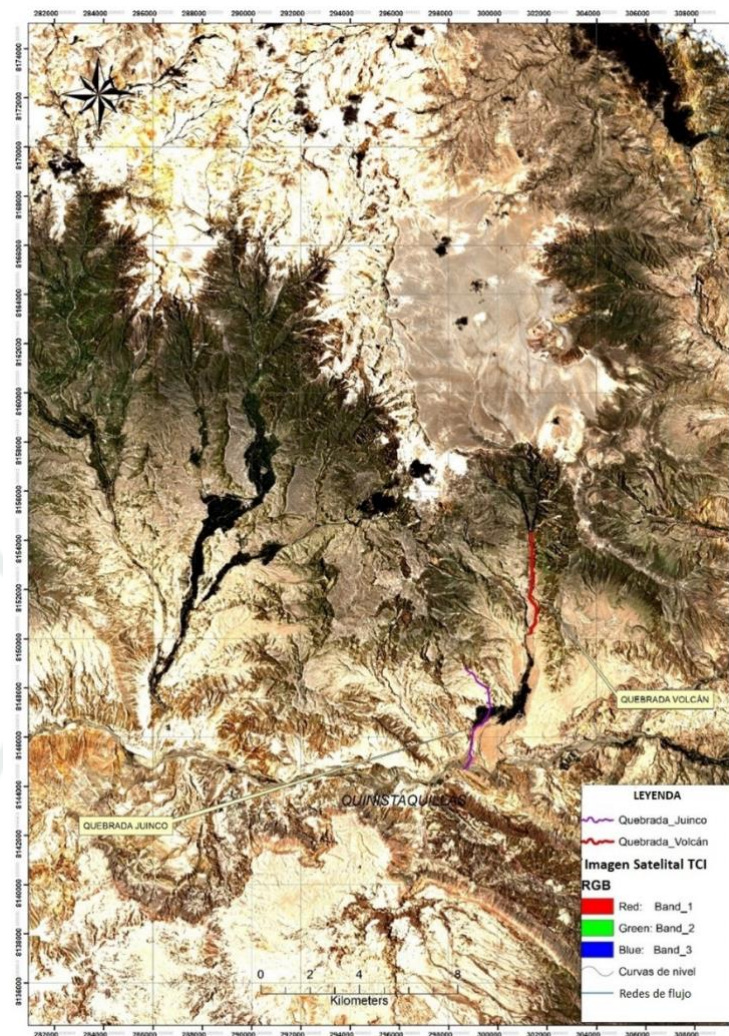
Nota: Elaboración propia

El índice NDWI permitió identificar zonas con contenido de humedad superficial, aportando un valor clave para analizar susceptibilidad a deslizamientos. La Fig. 11 muestra el comportamiento multitemporal del NDWI entre 2016 y 2023, donde se observa que las áreas más oscuras en las imágenes representan regiones con mayor humedad o presencia de agua libre, mientras que los tonos claros reflejan superficies más secas. A lo largo del período analizado (2016–2023), se evidenciaron picos de humedad principalmente en los años 2020, 2022 y 2023, coincidiendo con las temporadas de lluvias. Las microcuencas muestran

acumulación de humedad en zonas bajas y laderas orientadas al sur, lo que coincide con áreas inestables. Las muestras positivas (en rojo) se superponen parcialmente con regiones de alto NDWI, sugiriendo que la saturación del suelo podría haber contribuido a la ocurrencia de deslizamientos. El NDWI complementa eficazmente al NDVI y al Brightness al capturar cambios hidrológicos no visibles en otros índices. Su variabilidad temporal refuerza su utilidad para el monitoreo multi-temporal. En conjunto, este índice resultó útil como variable predictora en los modelos de clasificación aplicados.

Figura 10

Imagen TCI

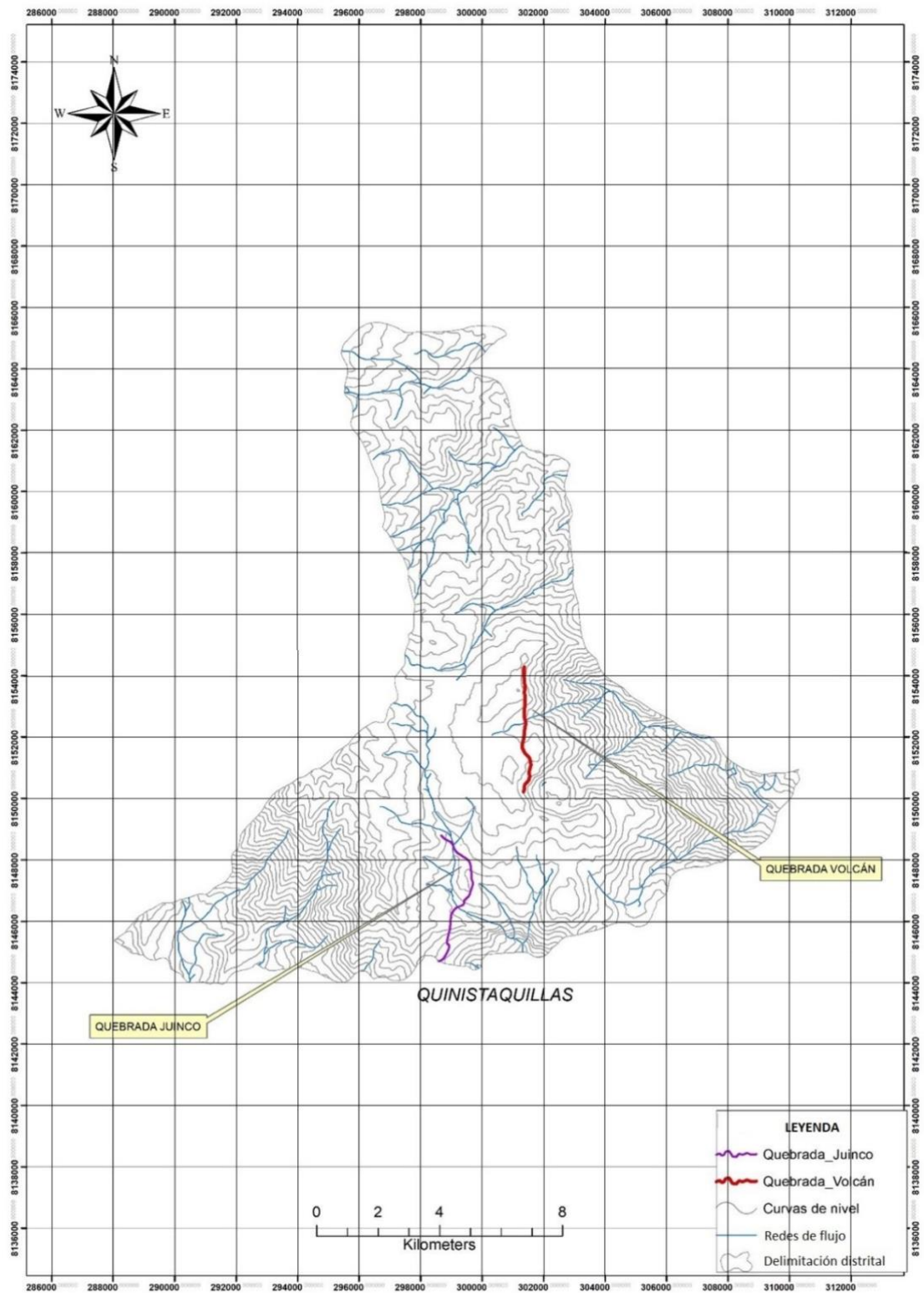


Nota: *Elaboración propia*

La Fig. 12 corresponde a una composición RGB generada a partir de bandas multiespectrales del satélite Sentinel-2 (Bandas 3, 2 y 1), permitiendo visualizar el área de estudio con alta resolución espacial. Se identifican claramente los contrastes entre zonas vegetadas, suelos expuestos y cuerpos de agua, fundamentales para el análisis geoespacial. El mapa incluye el trazo de las microcuencas de las quebradas Volcán (en rojo) y Juinco (en violeta), delimitadas mediante análisis hidrológico en QGIS. Las curvas de nivel y la red de flujo sobrepuesta permiten interpretar la morfología del terreno y el patrón de escorrentía superficial. El sector de Quinistaquillas se ubica al centro, sirviendo como referencia territorial clave. La quebrada Volcán drena desde las zonas altas del noreste, mientras que la Juinco lo hace desde el suroeste. La disposición del relieve y la orientación de las cuencas sugiere una mayor acumulación de energía de flujo hacia el extremo sur de ambas quebradas. Este mapa fue fundamental para definir unidades de análisis y ubicar muestras representativas para los modelos de clasificación. Asimismo, proporciona el contexto espacial necesario para interpretar los resultados de predicción de deslizamientos.

Figura 11

Delimitación y ubicación de quebradas

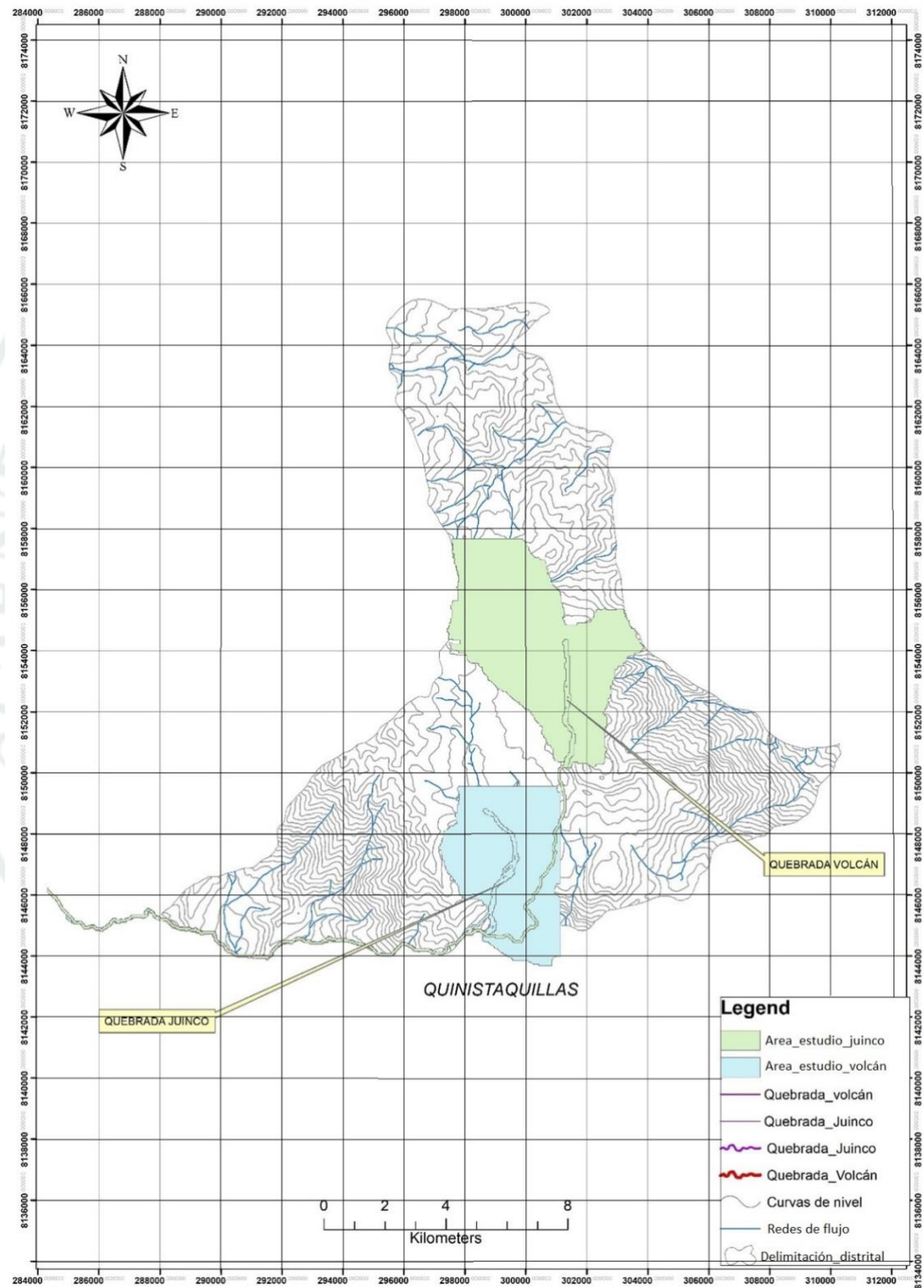


Nota: Elaboración propia

El mapa mostrado en la Fig. 13, se muestra la topografía y red hidrográfica de la zona de estudio, destacando la delimitación de las microcuencas de las quebradas Volcán (en rojo) y Juinco (en violeta). Las curvas de nivel permiten visualizar con claridad la morfología del terreno, evidenciando áreas de pendiente pronunciada, especialmente en los sectores norte y noreste, donde se concentran las elevaciones más marcadas. La red de flujo generada a partir del modelo digital de elevación (DEM) ilustra la dirección de escurrimiento superficial, lo cual fue clave para modelar la acumulación de agua y zonas potenciales de inestabilidad. La densidad de curvas indica cambios abruptos en altitud, lo que sugiere mayor susceptibilidad a procesos de remoción en masa. La quebrada Volcán presenta un cauce más rectilíneo y orientado al sureste, mientras que la Juinco es más sinuosa y drena desde el oeste. Este mapa sirvió como base para calcular la variable "pendiente", que fue incluida en los modelos de predicción de deslizamientos. Además, permitió definir puntos críticos para la validación y distribución de muestras de entrenamiento.

Figura 12

Generación de Microcuenca



Nota: Elaboración propia

El presente mapa mostrado en la Fig. 14 muestra la delimitación de las áreas de estudio correspondientes a las microcuencas de las quebradas Volcán y Juinco, ubicadas en el distrito de Quinistaquillas, región Moquegua. La delimitación se efectuó inicialmente en Global Mapper, donde se procesó un Modelo Digital de Elevación (DEM) para generar las direcciones de flujo y la identificación de las cuencas hidrográficas mediante la herramienta de *watershed delineation*.

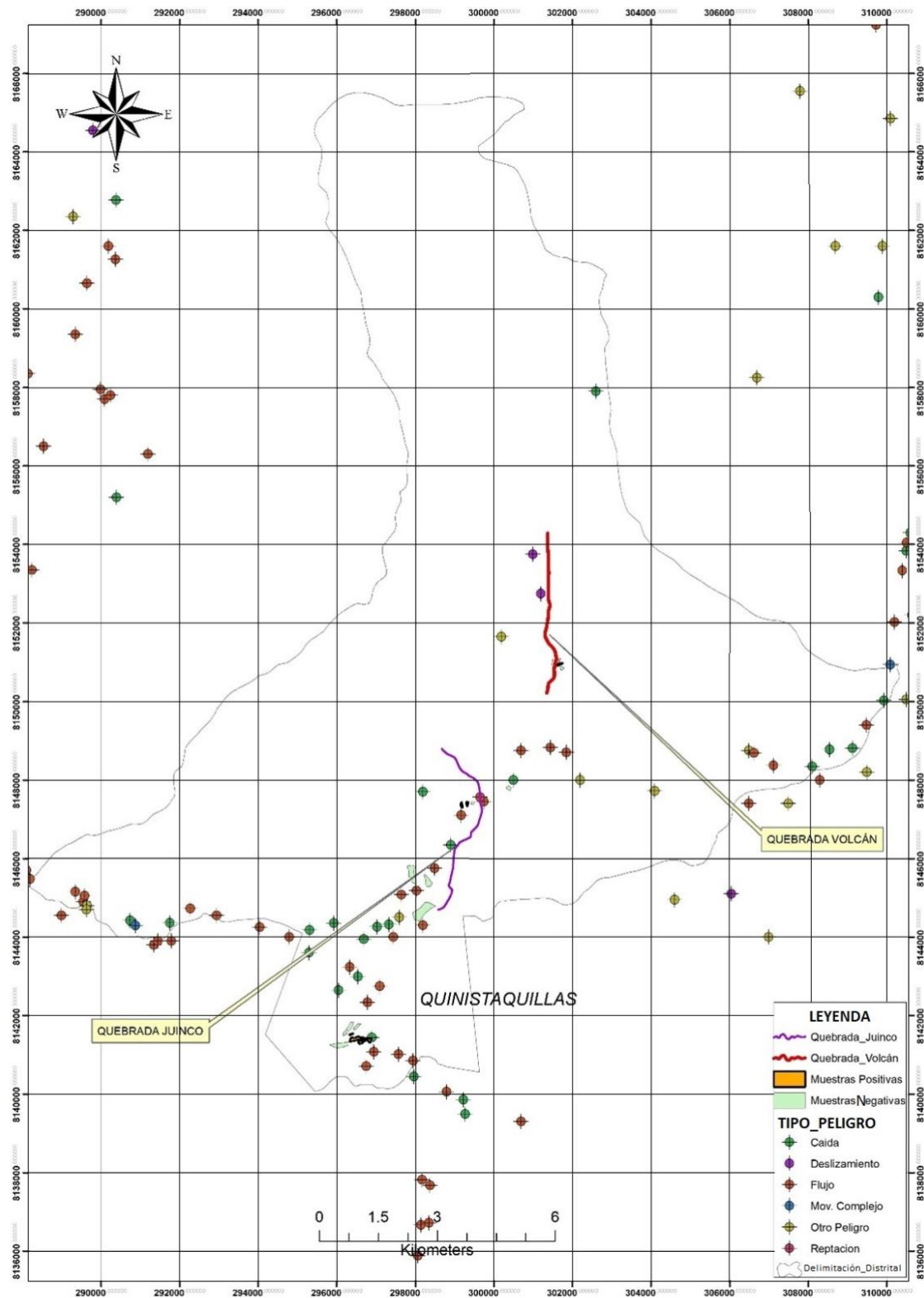
Posteriormente, la información fue exportada y refinada en ArcMap, donde se ajustaron los polígonos de las microcuencas y se superpusieron elementos complementarios como las curvas de nivel y las redes de drenaje generadas automáticamente. Esta metodología permitió una adecuada representación del relieve y de los patrones de escurrimiento superficial en cada unidad hidrográfica.

Las áreas coloreadas en verde y celeste representan las microcuencas seleccionadas para el análisis: la quebrada Volcán al noreste y la quebrada Juinco al suroeste. Estas zonas fueron elegidas por su alta pendiente, morfología inestable y evidencia previa de movimientos en masa.

Este mapa fue clave para definir los límites espaciales del estudio “Machine Learning para identificación de zonas potenciales de deslizamiento en Quinistaquillas – 2024”, y permitió estructurar el análisis geoespacial, la generación de grillas de muestreo y la aplicación de algoritmos predictivos como SVM y Random Forest.

Figura 13

Ubicación de Peligros registrados



Nota: Elaboración propia

El mapa mostrado en la Fig. 15 muestra la distribución espacial de las muestras utilizadas en el entrenamiento de modelos de clasificación para la identificación de zonas potenciales de deslizamiento en el distrito de Quinistaquillas. Se han representado tanto las muestras positivas (zonas donde ocurrieron deslizamientos u otros movimientos en masa) como las muestras negativas (zonas estables sin evidencia de inestabilidad).

Las muestras fueron georreferenciadas y categorizadas según el tipo de peligro reportado, distinguiéndose fenómenos como deslizamientos, caídas, flujos, movimientos complejos, reptación y otros. Esta clasificación fue fundamental para construir un dataset robusto y representativo del comportamiento del terreno en la zona de estudio.

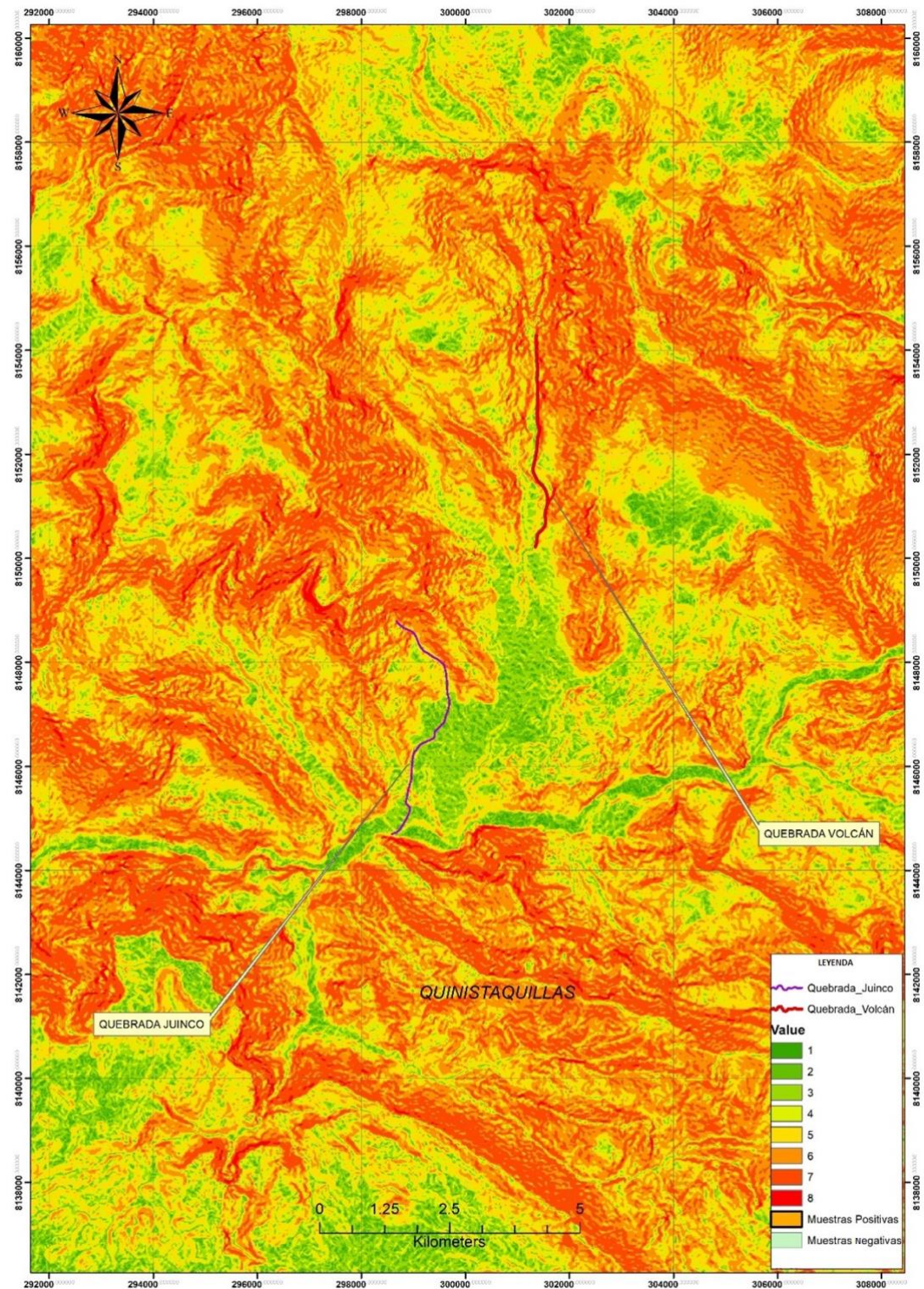
Los puntos se distribuyen mayormente en torno a las microcuencas quebrada Volcán y quebrada Juinco, seleccionadas por su morfología accidentada y presencia previa de eventos de remoción en masa. En el mapa se observa una concentración de eventos en la franja centro-sur, coincidiendo con áreas de fuerte pendiente y coberturas vegetales perturbadas.

Este conjunto de datos fue empleado para entrenar los algoritmos de Vectores de Soporte (SVM) y Random Forest (Random Forest), estableciendo así las bases para los modelos predictivos desarrollados en esta investigación. Las muestras negativas se seleccionaron de manera controlada en zonas distantes a registros históricos de deslizamiento, priorizando la variabilidad de condiciones topográficas y de uso del suelo.

El uso de este mapa permitió validar visual y espacialmente la distribución de muestras en función de su proximidad a elementos geomorfológicos críticos, mejorando la precisión de los modelos y su capacidad de generalización.

Figura 14

Mapa de pendientes

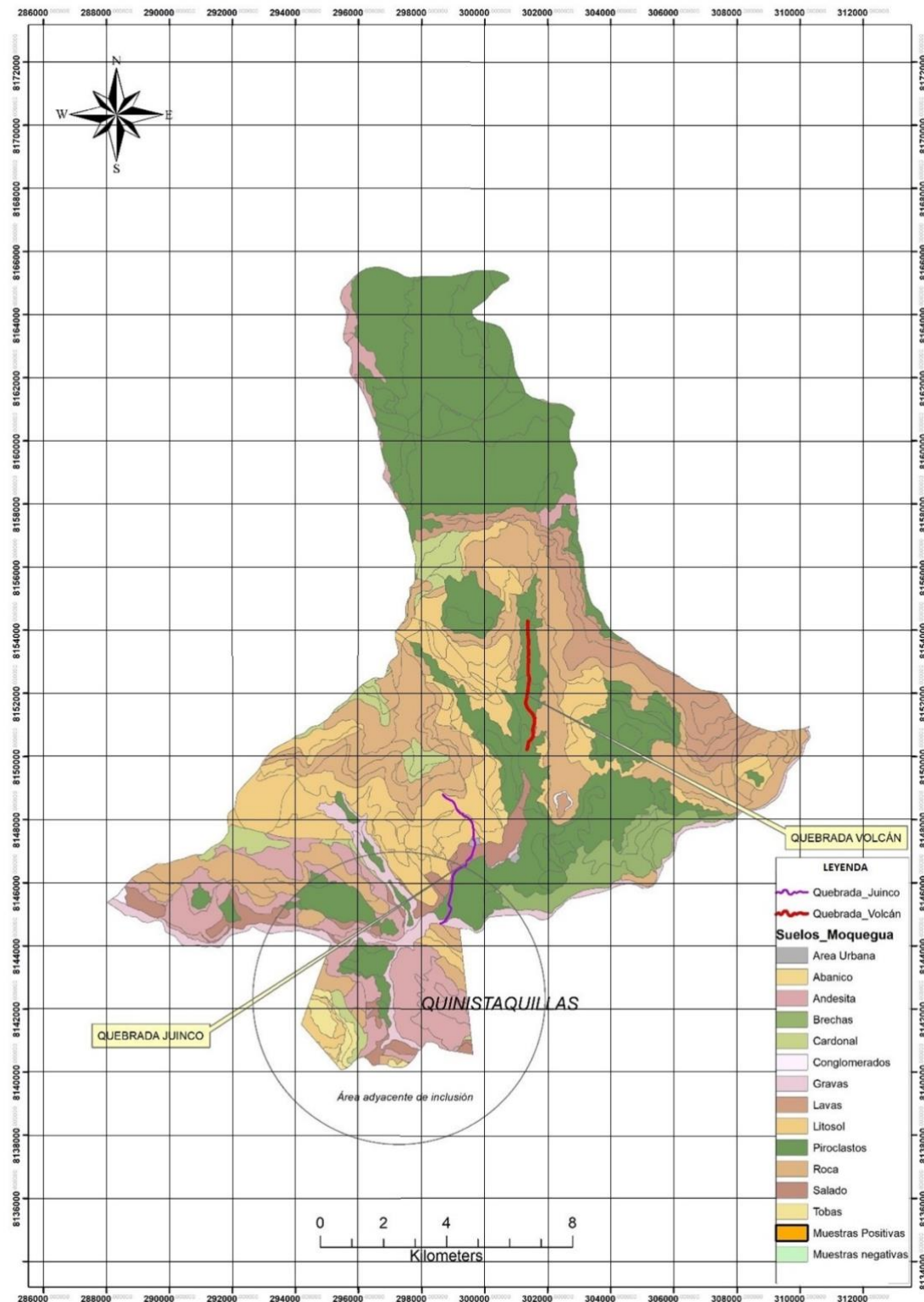


Nota: Elaboración propia

- El mapa mostrado en la Fig. 16 representa el análisis espacial del gradiente de pendiente en el distrito de Quinistaquillas, elaborado a partir de un modelo digital de elevación (DEM). Este insumo resulta esencial para el estudio de susceptibilidad a deslizamientos, ya que la pendiente es uno de los factores más determinantes en la inestabilidad de laderas.
- La leyenda clasifica los valores de pendiente en una escala ordinal de nueve rangos, desde zonas planas o suavemente onduladas (valor 1, en verde claro) hasta pendientes extremadamente abruptas (valor 9, en rojo intenso). Esta gradación facilita la interpretación visual de las zonas topográficamente más vulnerables.
- Las quebradas Volcán y Juinco, representadas en rojo y morado respectivamente, atraviesan zonas con alta variabilidad altitudinal. Particularmente en las zonas medias y altas de ambas microcuencas se concentran áreas con pendientes superiores al 40%, identificadas en tonos anaranjados y rojos. Estas áreas son propensas a procesos gravitacionales como deslizamientos y caídas de roca.
- También se observa una concordancia entre las ubicaciones de las muestras positivas (eventos reales de deslizamiento) y las zonas de mayor pendiente, lo cual valida la relevancia del mapa en el proceso de entrenamiento de modelos predictivos.
- Este mapa fue empleado como una de las variables independientes en los algoritmos de clasificación SVM y Random Forest, contribuyendo significativamente a la precisión del modelo al capturar el comportamiento morfodinámico del terreno.

Figura 15

Mapa de clasificación de suelos



Nota: *Elaboración propia*

El mapa mostrado en la Fig.17 representa la distribución espacial de las unidades edafológicas predominantes en el distrito de Quinistaquillas, extraídas de la cartografía temática de suelos de Moquegua. Esta información es crucial para el estudio de deslizamientos, ya que cada tipo de suelo presenta distinta susceptibilidad mecánica y drenaje interno.

Se identificaron unidades como gravas, tobas, piroclastos, lavas, conglomerados y rocas, que dominan el paisaje montañoso del centro y noreste del distrito. Estos materiales volcánicos e ignimbríticos, especialmente las tobas y piroclastos, poseen baja cohesión y alta porosidad, lo que incrementa su vulnerabilidad a procesos de remoción en masa bajo condiciones de saturación o perturbación antrópica.

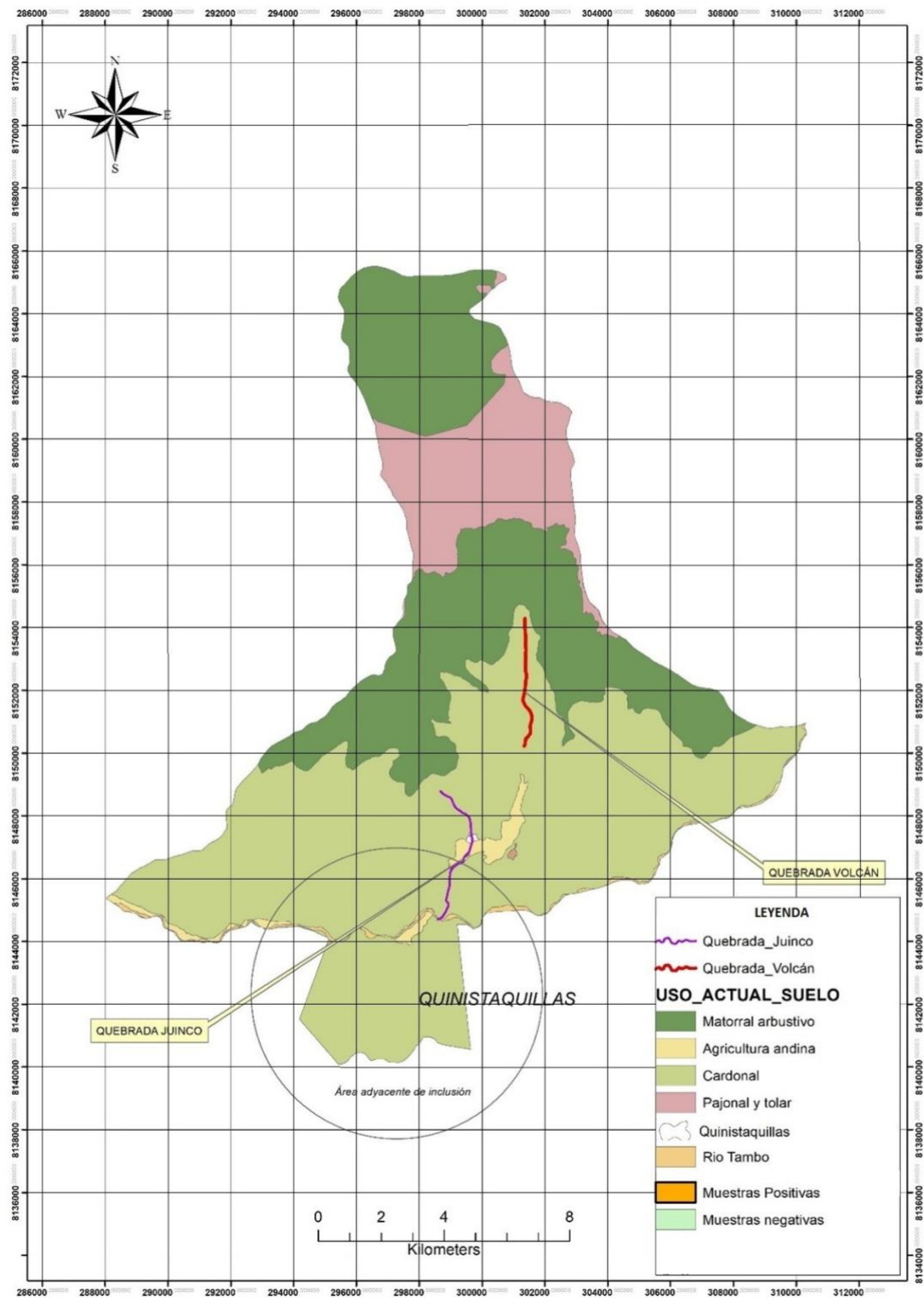
En las zonas bajas y planas, próximas al valle agrícola y al centro poblado de Quinistaquillas, predominan suelos litólicos, salados, y fluvisoles desarrollados en depósitos coluvio-aluviales. Aunque más estables, estas áreas pueden actuar como zonas de acumulación de material deslizado.

La superposición de las muestras positivas sobre estas unidades edafológicas permite verificar que una gran parte de los eventos se localiza sobre materiales de origen volcánico poco consolidados (tobas, piroclastos y lavas alteradas), validando su influencia en la ocurrencia de deslizamientos.

La integración de esta variable fue fundamental para el entrenamiento de los modelos SVM y Random Forest, permitiendo discriminar zonas de alto riesgo basadas en características litológicas.

Figura 16

Mapa de uso actual de suelo



Nota: Elaboración propia

El mapa mostrado en la Fig. 18 representa la distribución espacial del uso actual del suelo dentro de la unidad territorial de Quinistaquillas, con énfasis en las microcuencas de las quebradas Volcán y Juinco. Esta información fue extraída y clasificada a partir de imágenes satelitales Sentinel L2A y validada con información catastral disponible.

Entre las principales coberturas identificadas, destaca el matorral arbustivo, que cubre extensas zonas del centro y norte del distrito, seguido por la agricultura andina, concentrada en las terrazas y valles del sureste y suroeste. Estas zonas agrícolas suelen presentar movimientos de masa inducidos por intervención antrópica y mal manejo del riego.

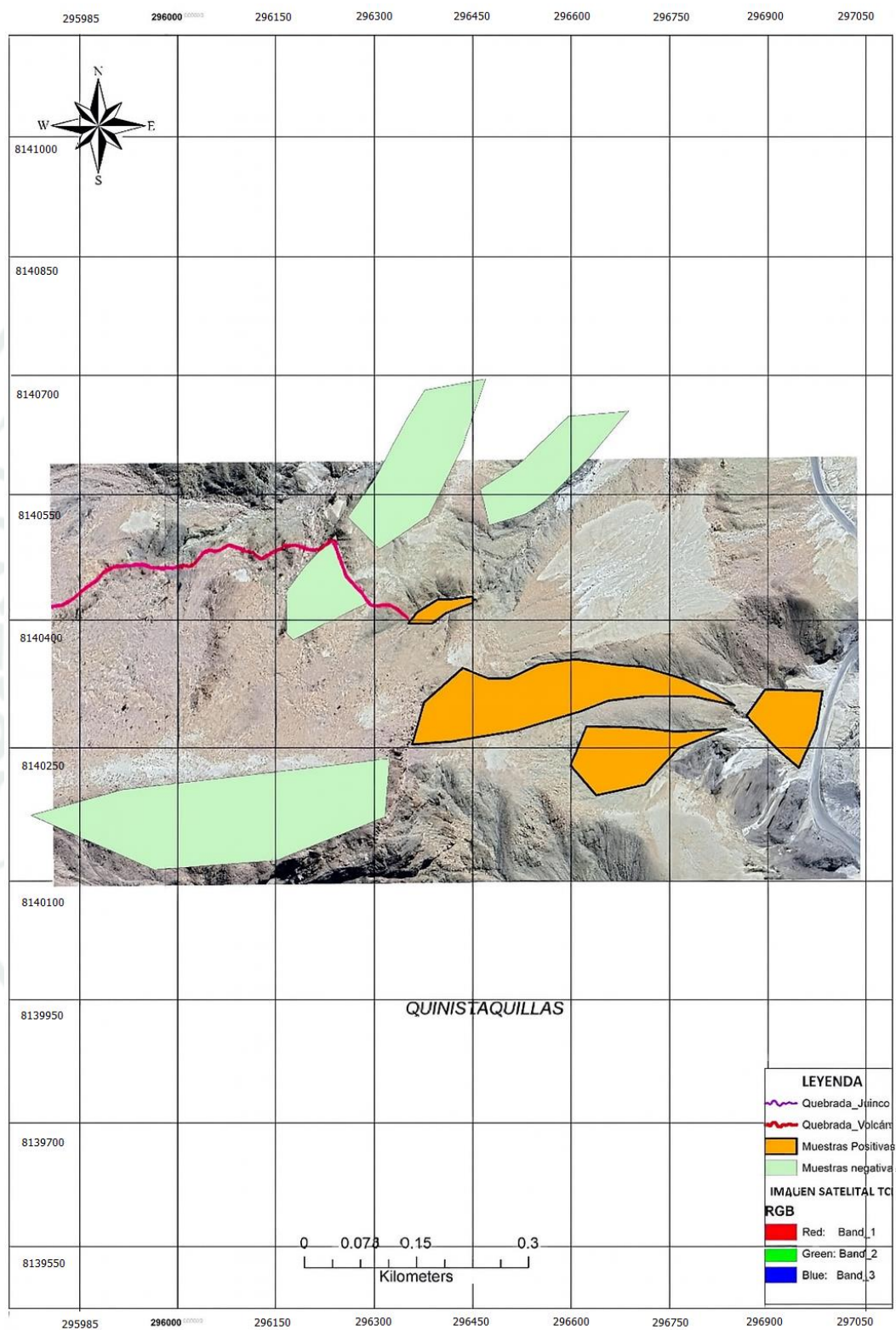
Asimismo, las unidades de pajonal y tolar, características de la alta montaña, se distribuyen en las partes altas, mientras que coberturas como cardonal y áreas específicas como San José de Chimpapampa y el centro poblado de Quinistaquillas han sido cartografiadas de forma diferenciada.

Se puede observar que las muestras positivas de deslizamientos se localizan mayormente en áreas con vegetación rala o intervenida, como el matorral y zonas de agricultura extensiva. Esto refuerza la relación entre cobertura vegetal reducida y mayor susceptibilidad a deslizamientos, lo cual fue clave para alimentar los modelos predictivos.

La variable “uso actual del suelo” fue normalizada y empleada como predictor categórico tanto para el modelo SVM como Random Forest, siendo relevante en la discriminación de áreas de riesgo.

Figura 17

Mapa de Muestras positivas y negativas(I)



Nota: Elaboración propia

El presente mapa correspondiente a la Fig. 19, corresponde a una validación visual sobre una imagen satelital de alta resolución TCI (True Color Index) en la zona adyacente al centro poblado de Quinistaquillas. Se han superpuesto las muestras de deslizamientos clasificadas durante el proceso de modelado predictivo (muestras positivas y negativas) para evaluar la coherencia espacial de la predicción respecto al terreno observado.

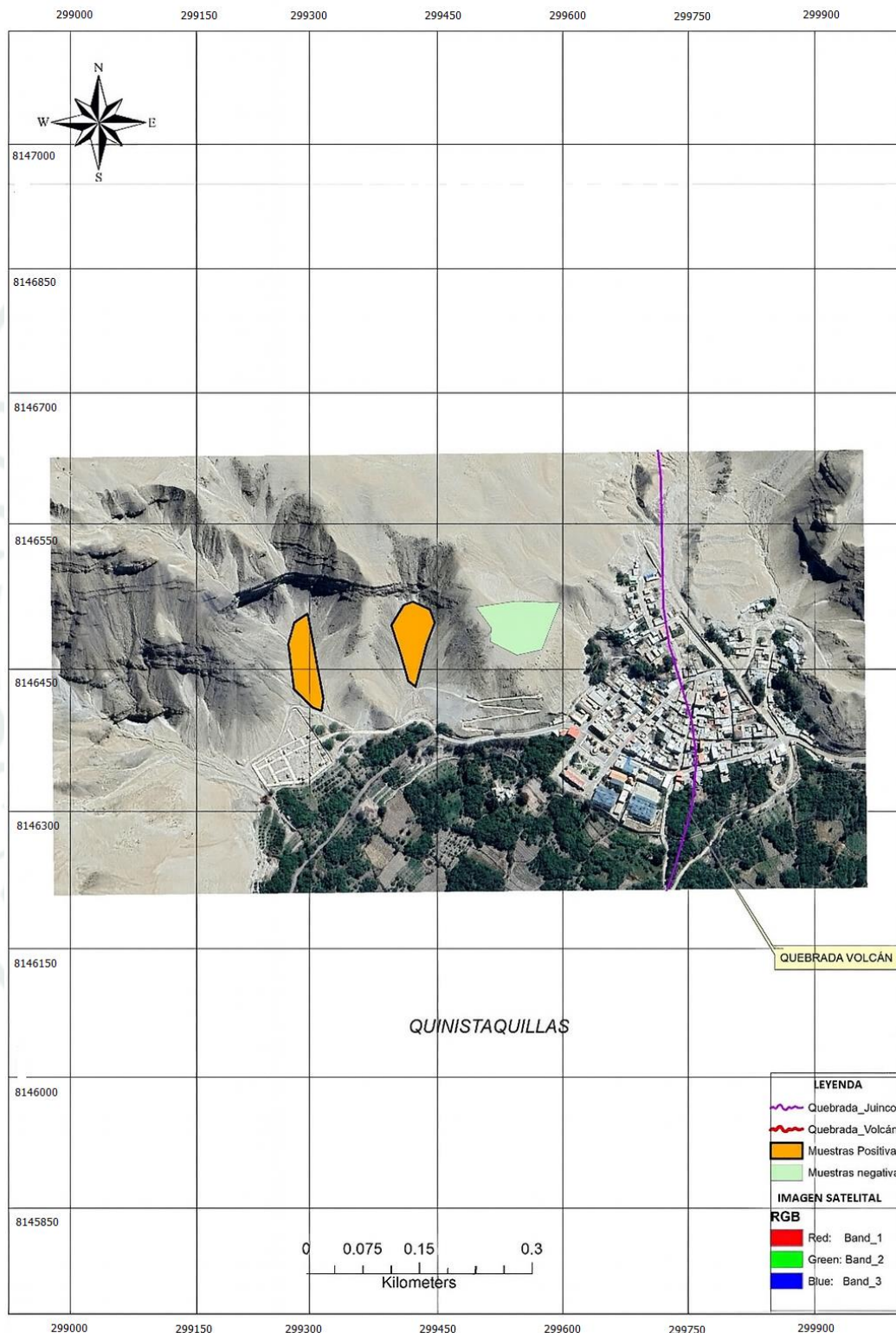
La imagen RGB base combina las bandas 1 (rojo), 2 (verde) y 3 (azul) y permite identificar con claridad áreas perturbadas, depósitos aluviales, taludes cortados y pendientes expuestas, lo que facilita contrastar las zonas detectadas por los algoritmos con las características topográficas reales.

Las muestras positivas (naranja) se concentran en áreas erosionadas, cercanas a cortes de carretera, cauces activos y laderas degradadas, lo que confirma la eficacia del modelo al identificar zonas inestables. Por su parte, las muestras negativas (verde claro) se ubican predominantemente en zonas con cobertura edáfica o vegetación natural, sin evidencias visuales de deslizamientos recientes.

Esta validación visual, aunque cualitativa, constituye un importante respaldo al análisis cuantitativo, reforzando la fiabilidad de los modelos SVM y Random Forest en la detección de zonas susceptibles a movimientos en masa en Quinistaquillas.

Figura 18

Mapa de muestras positivas y negativas (II)



Nota: Elaboración propia

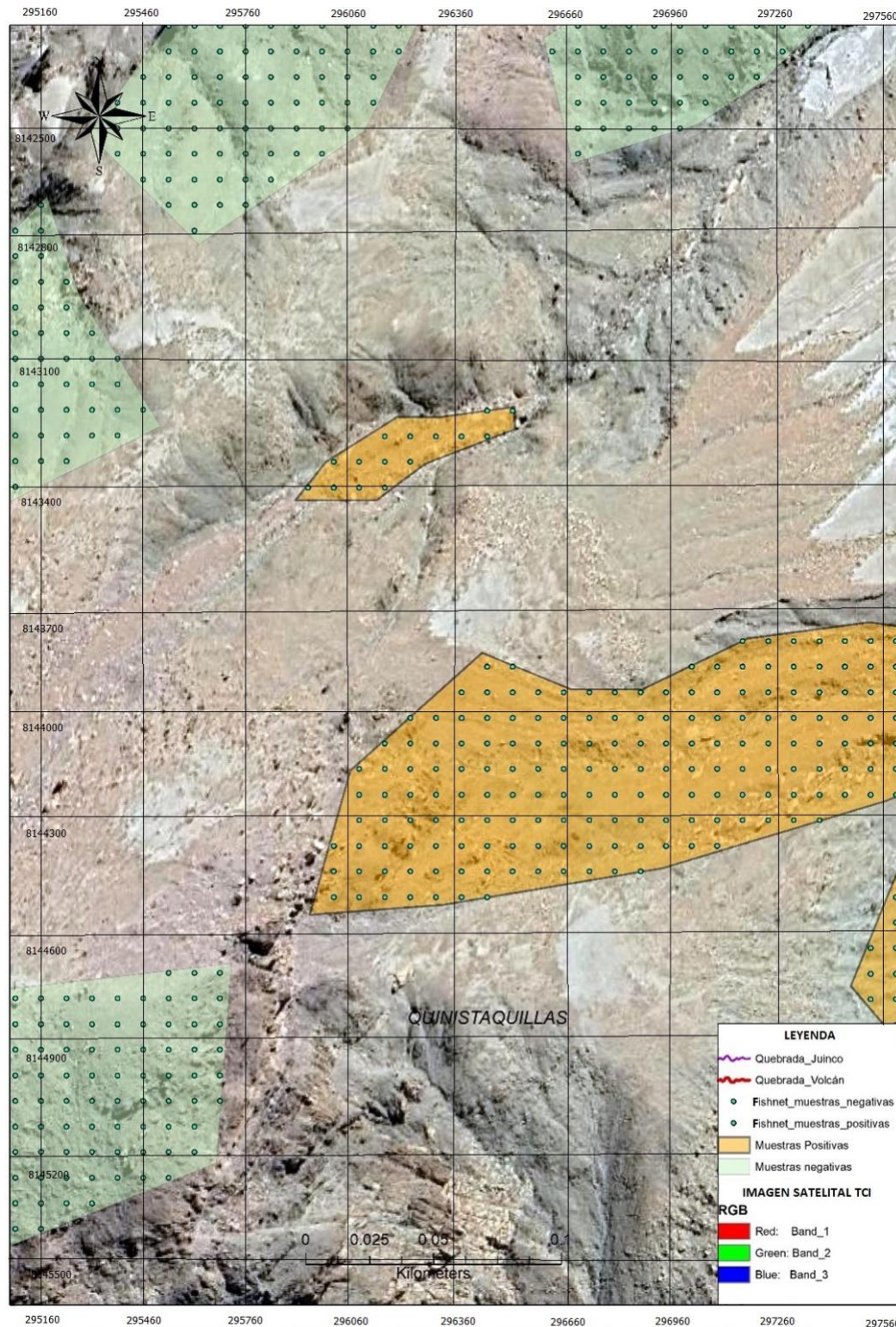
La Fig. 18 muestra las muestras positivas (naranja) se localizan sobre laderas abruptas ubicadas en el flanco occidental del poblado, asociadas a cortes casi verticales y zonas de material suelto, indicando una alta susceptibilidad a deslizamientos. Por otro lado, las muestras negativas (verde claro) aparecen en sectores planos o agrícolas, donde no se evidencia inestabilidad visible.

La trayectoria de la quebrada Volcán (trazo púrpura) bordea el casco urbano, lo cual representa un riesgo potencial en escenarios de lluvias extremas o sobresaturación del terreno. Esta validación visual refuerza la pertinencia del modelo predictivo aplicado, demostrando una adecuada correlación espacial entre las áreas de riesgo y la realidad del terreno.

El análisis confirma que las zonas críticas se encuentran en contacto directo con infraestructura urbana, resaltando la necesidad de acciones preventivas orientadas al monitoreo y estabilización de taludes en el área de estudio.

Figura 19

Fishnet (Muestras)



Nota: Elaboración propia

Este mapa correspondiente a la Fig. 21 representa un análisis de validación fina sobre un sector crítico de la quebrada Volcán en Quinistaquillas, integrando una malla de puntos generada mediante el algoritmo *fishnet* para evaluar con mayor resolución espacial la cobertura y distribución de las muestras positivas (en naranja) y negativas (en verde claro).

La imagen satelital de fondo (DESZ1_2.jpg) compuesta por las bandas RGB permite identificar las características geomorfológicas de las laderas con mayor detalle. La superposición de las mallas con las muestras permite verificar gráficamente la coherencia del etiquetado realizado y la efectividad del modelo predictivo entrenado.

Las zonas de alta pendiente, con acumulación de detritos o sin cobertura vegetal significativa, concentran la mayor cantidad de puntos positivos (riesgo alto), mientras que los negativos aparecen sobre laderas estables o zonas planas.

Este enfoque visual y cuantitativo refuerza la calidad del entrenamiento del modelo de machine learning, mostrando una correcta correspondencia entre predicciones y condiciones reales del terreno, además de ser una herramienta eficaz para el control de precisión de futuros mapas de susceptibilidad.

Tabla 7

Cuadro resumen de índices

Variable	Rango observado	Unidades	Fuente
NDVI_2023	-0.21 a 0.72	adimensional	Sentinel-2 L2A (Copernicus)
Brillo_2020	0.04 a 0.87	normalizado	Calculado (banda visible+NIR)
Pendiente	0.5° a 42°	grados	MDE (30 m)

Nota: *Elaboración propia*

1.2. Resultados II

Objetivo vinculado: Aplicación de SVM y Random Forest para identificar zonas deslizadas conocidas.

1. Incluye un cuadro con cantidad de muestras positivas y negativas, porcentaje de entrenamiento y test, etc.

Tabla 8

Tabla resumen de muestras y porcentajes de entrenamiento

Descripción	Valor
Total datos de muestras positivas (clase 1)	556
Total de datos de muestras negativas (clase 0)	3214
Total de datos de muestras generales	3770
Proporción entrenamiento	75 % (2827 datos)
Proporción prueba	25 % (943 datos)
Algoritmos evaluados	Random Forest (RF), SVM (RBF)
Variables utilizadas	NDVI, NDWI, Brightness, Pendiente, Tipo de suelo, Uso de suelo, Coordenadas
Herramientas	Python (Jupyter Notebook), scikit-learn

Nota: Elaboración propia a partir de datos Sentinel-2 L2A y catastro SIG, 2024.

2. Matrices de confusión

Una vez entrenados los modelos de aprendizaje supervisado, se procedió a evaluar su rendimiento clasificatorio mediante la generación de matrices de confusión. Estas matrices permiten observar el comportamiento predictivo del modelo sobre el conjunto de validación, identificando verdaderos positivos (TP), verdaderos negativos (TN), falsos positivos (FP) y falsos negativos (FN).

3. Describe los valores de precisión, sensibilidad y F1 obtenidos (por modelo)

Usa una tabla comparativa:

Después del entrenamiento y validación de los modelos Random Forest (RF) y de Vectores de Soporte (SVM), se evaluó el rendimiento de ambos utilizando métricas clave: precisión, sensibilidad (recall), especificidad y F1-score. Estas métricas permiten medir la eficacia de los algoritmos en la clasificación binaria de zonas con y sin deslizamiento, aplicando criterios de balance entre verdaderos positivos y falsos negativos.

Los resultados obtenidos para el conjunto de prueba (25 % de los datos) se detallan en la siguiente tabla comparativa:

Tabla 9

Métricas

Métrica	SVM	Random Forest (RF)
Precisión (Clase 1)	62 %	90 %
Sensibilidad (Recall)	44 %	86 %
Especificidad	96 %	98 %
F1-score (Clase 1)	0.52	0.88
Accuracy (Precisión gral)	77 %	94 %
Índice Kappa (κ)	0.54	0.88

Nota: Elaboración propia en Python con `classification_report()` de scikit-learn, 2024.

Análisis comparativo

Los resultados evidencian que el modelo Random Forest obtuvo un rendimiento sustancialmente superior en las métricas clave de clasificación. Presentó un F1-score de 0.88 para la clase deslizada, superando ampliamente al modelo SVM (0.52). Asimismo, alcanzó una precisión de 90 % para la clase 1, frente al 62 % del modelo SVM, así como una precisión general (accuracy) de 94 %, superior al 77 % obtenido por SVM. Además, Random Forest mostró una sensibilidad de 86 %, lo cual es especialmente relevante en estudios de deslizamientos, donde resulta prioritario minimizar la omisión de zonas potencialmente inestables.

Por su parte, el modelo SVM, aunque mantuvo una especificidad alta de 96 %, presentó una sensibilidad reducida de 44 %, lo que indica una subdetección importante de eventos reales de deslizamiento. En contraste, Random Forest alcanzó una especificidad de 98 %, mostrando además un mejor equilibrio global entre identificación de casos positivos y negativos. Este comportamiento confirma que RF ofrece una capacidad predictiva más robusta para la clasificación de zonas susceptibles a deslizamientos en el área de estudio.

Este resultado coincide con estudios recientes que destacan la eficacia de Random Forest para el análisis geoespacial con datos heterogéneos y multivariantes, al superar a otros algoritmos en tareas de clasificación de susceptibilidad a deslizamientos (Youssef et al., 2023; Prakash et al., 2022).

4. Analisis del rendimiento

El análisis del rendimiento de los algoritmos aplicados evidenció una ventaja considerable del modelo Random Forest sobre la Máquina de Vectores de Soporte (SVM), especialmente en términos de sensibilidad, F1-score, accuracy e índice Kappa. Random Forest alcanzó una precisión general del 94 %, un F1-score de 0.88 en la clase de deslizamientos y un índice Kappa de 0.88, demostrando una alta capacidad para capturar patrones complejos entre

las variables espaciales y espectrales utilizadas. Este desempeño puede atribuirse a la robustez del algoritmo frente a datos heterogéneos, su tolerancia a relaciones no lineales y su buen comportamiento en contextos con múltiples variables predictoras.

En contraste, el modelo SVM presentó una precisión general de 77 %, un F1-score de 0.52 y un índice Kappa de 0.54, además de una sensibilidad de 44 %, lo que evidencia limitaciones para detectar correctamente los eventos reales de deslizamiento. Aunque su especificidad fue alta (96 %), su menor capacidad para reconocer la clase positiva representa una desventaja importante en estudios de gestión del riesgo, donde es preferible reducir los falsos negativos. Estos resultados sugieren que, en este caso, SVM fue menos eficaz para modelar la complejidad espacial y temática de los datos empleados.

Además, la capacidad de Random Forest para manejar múltiples atributos predictivos, como índices espectrales, pendiente, tipo de suelo, uso de suelo y coordenadas espaciales, le otorgó una ventaja operativa dentro del flujo metodológico aplicado. En términos prácticos, se concluye que Random Forest constituye la alternativa más confiable para la clasificación de zonas potencialmente inestables en Quinistaquillas, aportando una base más sólida para la toma de decisiones en planificación territorial y gestión de desastres.

1.3. Resultados III

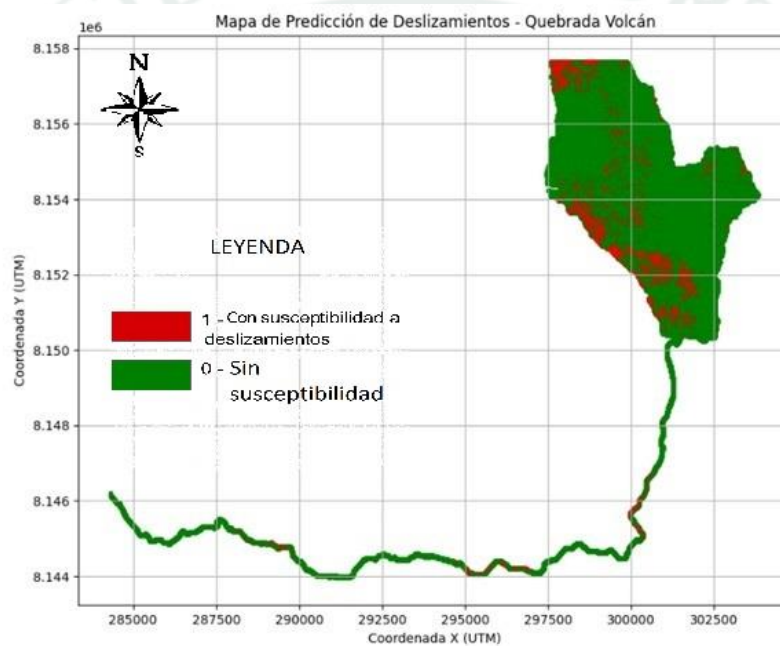
Objetivo vinculado: Predicción de zonas potenciales de deslizamiento con SVM y RF

1. Presenta los 4 mapas en secciones separadas

- Mapa 1: PREDICCIÓN RF – Quebrada Volcán

Figura 20

Mapa de Predicción RF - MC Quebrada Volcán



Nota: Elaboración propia

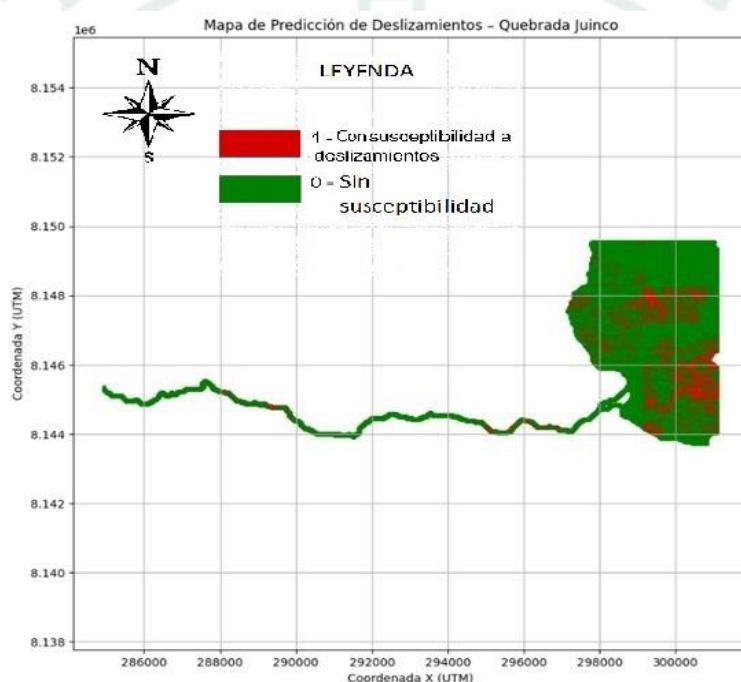
El modelo Random Forest identificó zonas críticas de deslizamiento a lo largo de la ladera este de la microcuenca Volcán. La Fig. 22 muestra el mapa resultante de esta predicción, donde se observa una clara concentración de áreas en rojo (clase 1) en sectores de alta pendiente y vegetación perturbada, principalmente en la parte norte y centro. Las predicciones son espacialmente coherentes, sin ruido disperso, y siguen el patrón morfológico del terreno. La segmentación es precisa, especialmente en las inmediaciones del cauce principal. El patrón de predicción sugiere que el modelo fue sensible a los gradientes de vegetación y pendiente. Las zonas sin riesgo (verde) dominan el mapa, lo que concuerda con la proporción general de

muestras negativas. Se aprecia buena definición en la frontera entre clases. La consistencia topológica refuerza la calidad del entrenamiento con los datos. Se nota buena concordancia con la hidrografía local.

- Mapa 2: PREDICCIÓN RF – Quebrada Juinco

Figura 21

Mapa de Predicción RF - MC quebrada Juinco



Nota: *Elaboración propia*

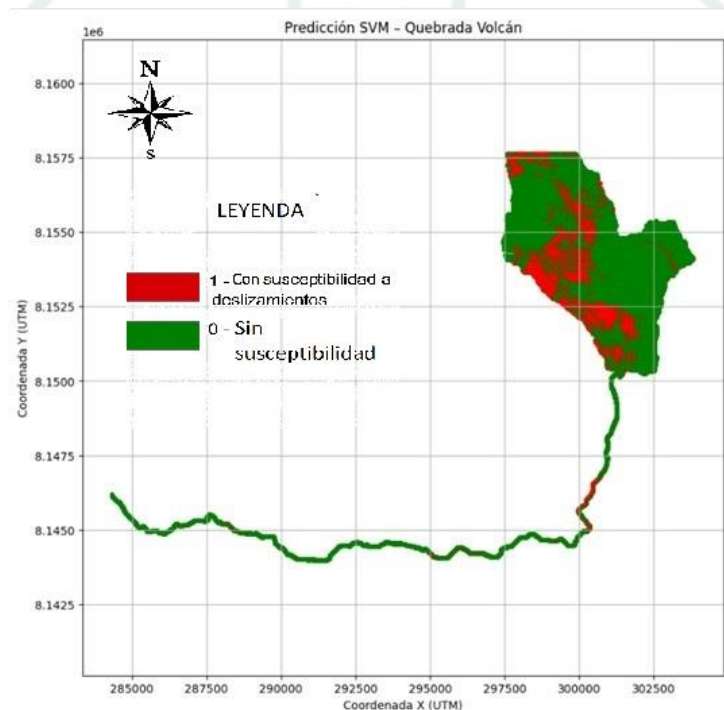
En la microcuenca de la quebrada Juinco, el modelo RF mostró una delimitación clara de las zonas de susceptibilidad, localizadas mayoritariamente al sureste y noreste de la cuenca. En la Fig. 23 se presenta el mapa de predicción generado, donde se evidencian manchas compactas de clase 1 (zonas en rojo), especialmente donde se combinan pendientes marcadas con coberturas alteradas. La predicción sigue patrones estructurales consistentes con la dirección del flujo y la forma del relieve. No se observa sobrestimación ni ruido espacial. La clase 0 domina el espacio, aunque se permite una discriminación efectiva de zonas de interés.

El modelo responde de manera clara a las condiciones topográficas y de vegetación. La separación entre zonas con y sin riesgo es más nítida que en SVM. La estructura general del mapa coincide con las tendencias de los mapas temáticos usados como entrada. Es una predicción confiable desde el punto de vista morfológico y espacial.

- Mapa 3: PREDICCIÓN SVM – Quebrada Volcán

Figura 22

Mapa de Predicción SVM - Quebrada Volcán



Nota: *Elaboración propia*

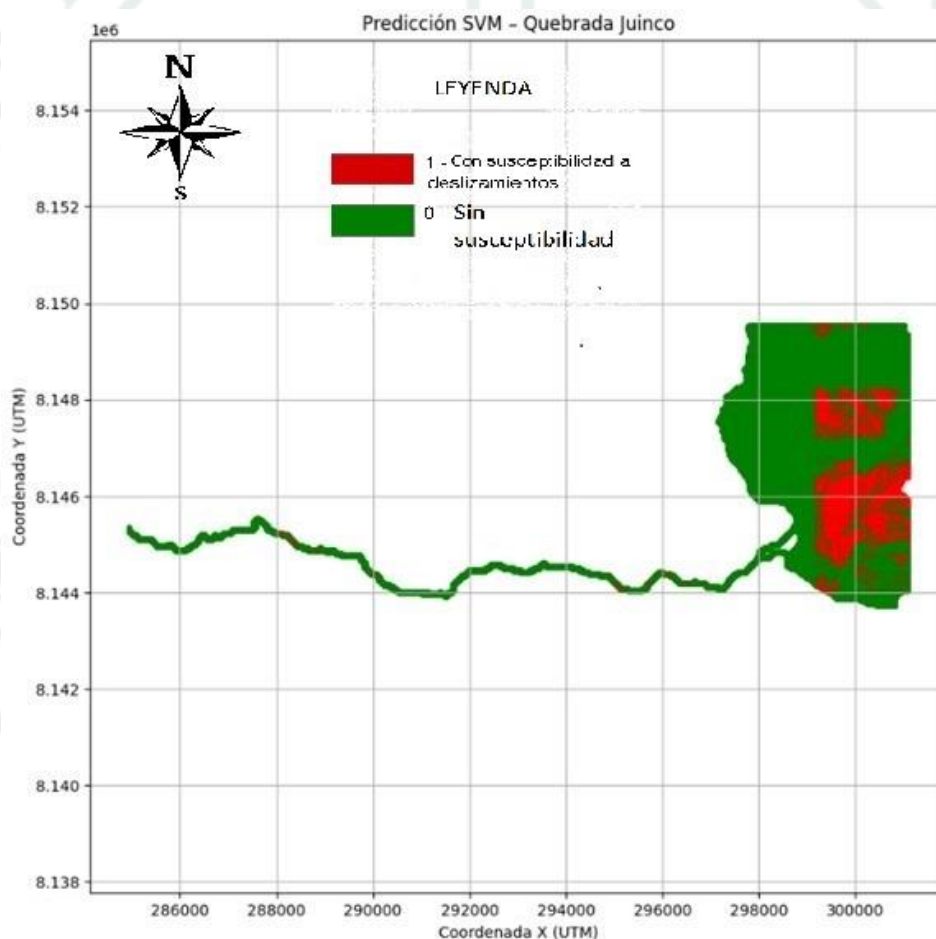
El modelo SVM aplicó una predicción más difusa y menos estructurada. La Fig. 24 muestra los resultados obtenidos para la microcuenca Volcán, donde se identifican zonas propensas en el norte y sureste, aunque con mayor presencia de ruido y parches irregulares de clase 1 (rojo). En comparación con RF, se observa cierta sobrestimación en áreas marginales, posiblemente debido a la sensibilidad del kernel RBF a pequeñas variaciones. Las zonas de deslizamiento no siempre siguen los patrones morfológicos esperados, y aparecen

fragmentadas en sectores discontinuos. La predicción es menos coherente espacialmente, con zonas clasificadas como riesgo intercaladas en áreas estables. Hay menor precisión en los bordes. Aunque detecta las áreas críticas generales, su delimitación es menos nítida. La interpretación se dificulta debido al patrón ruidoso. Aun así, es útil como modelo complementario.

- Mapa 4: PREDICCIÓN SVM – Quebrada Juinco

Figura 23

Mapa de Predicción SVM - Quebrada Juinco



Nota: Elaboración propia

El modelo SVM en la microcuenca Juinco evidenció un comportamiento similar al de la quebrada Volcán, con amplias zonas en rojo, algunas bien ubicadas, pero muchas dispersas

y sin continuidad espacial. En la Fig. 25 se presenta el mapa correspondiente, donde se detectan zonas al centro y sur con alta susceptibilidad, aunque también con presencia de puntos espurios. Hay menor coherencia con las estructuras topográficas. El modelo tiende a predecir zonas más amplias de clase 1, lo que podría derivar en sobreestimación de riesgo. Se observa mayor interferencia en zonas planas. El patrón general sugiere una sensibilidad elevada a variables como la pendiente y la reflectancia. No obstante, se pierde precisión espacial. El mapa es útil para un análisis exploratorio, pero menos confiable en zonas marginales. Se aprecia menor diferenciación entre límites de clases.

Tzampoglou et al. (2025) demostraron que la integración de parámetros geotécnicos en modelos de Machine Learning mejora la precisión de los mapas de susceptibilidad a deslizamientos, destacando la eficacia de métodos como Random Forest y XGBoost en la clasificación de áreas de riesgo.

Además, estudios recientes han explorado la aplicación de técnicas de aprendizaje profundo para la cartografía de susceptibilidad a deslizamientos. Li et al. (2025) emplearon redes neuronales convolucionales (CNN) para mapear la susceptibilidad a deslizamientos en la cuenca de Bakhtegan, Irán, logrando una precisión del 95.76%. La capacidad de las CNN para capturar relaciones espaciales complejas entre factores condicionantes del terreno permite una clasificación más precisa de las zonas de riesgo, respaldando el uso de enfoques multinivel basados en probabilidades.

La validación de modelos de susceptibilidad a deslizamientos mediante la clasificación en niveles de riesgo también ha sido objeto de investigación. Menon y Kolathayar (2025) desarrollaron un modelo basado en Machine Learning para determinar umbrales de precipitación que predicen deslizamientos inminentes en Kerala, India, utilizando modelos como KNN, Random Forest y SVM. La clasificación de las probabilidades en categorías de

riesgo permitió una mejor interpretación de los resultados y una toma de decisiones más informada en la gestión del riesgo de deslizamientos.

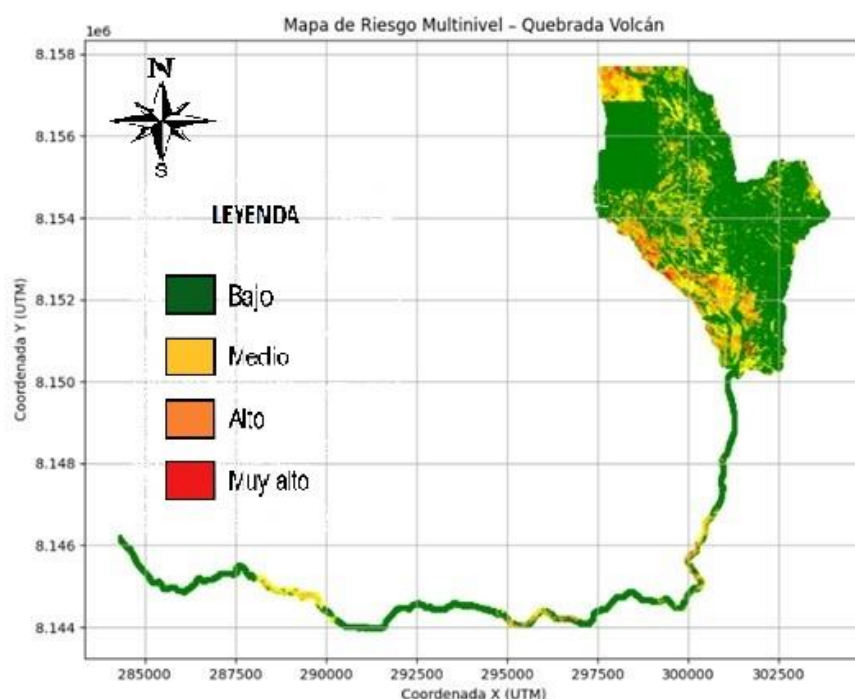
La clasificación multinivel del riesgo de deslizamientos, fundamentada en probabilidades derivadas de modelos de Machine Learning como Random Forest y SVM, ha sido validada y mejorada en investigaciones recientes. Por ejemplo, Bao et al. (2025) propusieron un enfoque que integra múltiples factores con umbrales de probabilidad adaptativos, utilizando el algoritmo Gradient Boosting Decision Tree (GBDT) para estimar la susceptibilidad a deslizamientos, logrando una alta precisión en la identificación de áreas propensas. Asimismo, Tzampoglou et al. (2025) demostraron que la incorporación de parámetros geotécnicos en modelos de Machine Learning mejora la precisión de los mapas de susceptibilidad a deslizamientos, destacando la eficacia de métodos como Random Forest y XGBoost. Estas investigaciones respaldan el uso de clasificaciones basadas en probabilidades para representar espacialmente la amenaza de deslizamientos, facilitando la toma de decisiones en la gestión del riesgo y la planificación territorial.

La clasificación de riesgo de deslizamientos en niveles (bajo, medio, alto y muy alto) utilizada en el presente estudio se basa en el uso de probabilidades de ocurrencia obtenidas a partir de modelos de clasificación binaria, tales como Random Forest y SVM, las cuales permiten asignar rangos de probabilidad para categorizar el nivel de susceptibilidad a deslizamientos. Este enfoque metodológico ha sido validado en estudios similares de susceptibilidad a deslizamientos y desastres naturales, donde las probabilidades obtenidas de modelos de machine learning son interpretadas como un índice de riesgo, posteriormente reclasificado en categorías de amenaza (Guzzetti et al., 2006; Meusburger et al., 2010). Según Guzzetti et al. (2006), las probabilidades de deslizamiento derivadas de modelos estadísticos o de Machine Learning se convierten en mapas de susceptibilidad al aplicar umbrales definidos para representar niveles de riesgo, lo que facilita la toma de decisiones en la gestión de desastres

y planificación territorial. Este procedimiento ha sido reconocido como una metodología efectiva para representar espacialmente la amenaza y apoyar la gestión del riesgo en contextos de ingeniería ambiental y geoespacial (Meusburger et al., 2010).

Figura 24

Mapa de Riesgo - Quebrada Volcán



Nota: Elaboración propia

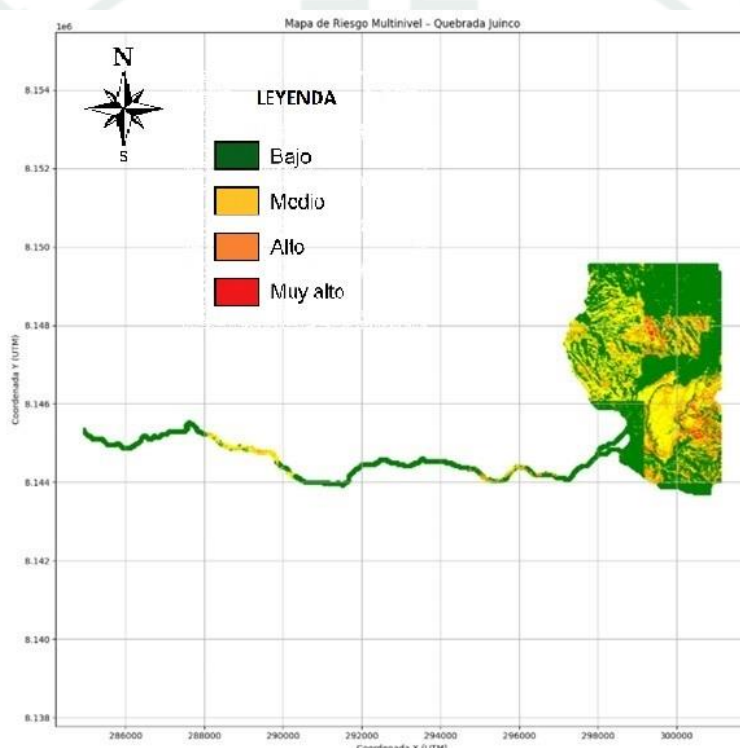
En la Fig. 26 se presenta el mapa de riesgo multinivel para la Quebrada Volcán, donde se muestra una distribución de las clases de riesgo (bajo, medio, alto, muy alto) a lo largo del corredor geográfico analizado. En la parte alta de la cuenca, se observa una concentración importante de zonas de riesgo alto y muy alto (colores naranja y rojo), que corresponden a áreas con pendientes pronunciadas, características topográficas abruptas y menor cobertura vegetal.

Esto se alinea con la geomorfología accidentada del sector norte de la quebrada, donde se evidencian áreas susceptibles a movimientos en masa.

En la zona sur y en el lecho principal del cauce, el riesgo disminuye, con predominio de áreas de riesgo bajo y medio (verde y amarillo). Esto puede atribuirse a la presencia de zonas con pendientes más suaves y mayor vegetación que contribuyen a la estabilidad del terreno.

Figura 25

Mapa de Riesgo - Quebrada Juinco



Nota: Elaboración propia

En la Fig. 27 se muestra el mapa de riesgo multinivel para la Quebrada Juinco, donde se aprecia una distribución diferenciada del riesgo a lo largo del cauce. La porción oeste y centro de la quebrada presenta áreas con riesgo bajo y medio de deslizamientos, mientras que hacia el sector noreste y sureste se incrementa la concentración de zonas de riesgo alto y muy

alto. Esto coincide con sectores de pendiente empinada, afloramientos rocosos y menor cobertura vegetal.

La conectividad hidrológica y las dinámicas de escorrentía superficial también influyen en la distribución espacial del riesgo, siendo evidente que en los sectores donde el agua escurre con mayor velocidad y volumen (puntos de convergencia de cauces menores), el riesgo tiende a incrementarse.

3. Comparación RF vs. SVM

Comparando ambos algoritmos, se concluye que el modelo Random Forest ofreció un mejor desempeño espacial y temático. Las predicciones de RF fueron más precisas, menos ruidosas y mejor alineadas con las variables del terreno y cobertura. Mostró mayor especificidad y sensibilidad, con una clara delimitación de zonas críticas. Por su parte, SVM tendió a sobreestimar la clase de riesgo, generando fragmentación espacial y ruido en áreas estables. Aunque ambos modelos pueden complementarse, RF demostró ser más robusto y confiable para generar productos cartográficos aplicables en la gestión de riesgos por deslizamientos.

1.4. Resultados IV

Objetivo vinculado: Comparación final de métricas y desempeño

1. Tabla resumen de métricas (repetir de 4.2 con énfasis)

Los resultados obtenidos para el conjunto de prueba (25 % de los datos) se detallan en la siguiente tabla comparativa:

Tabla 10

Métricas de rendimiento

Métrica	SVM	Random Forest
---------	-----	---------------

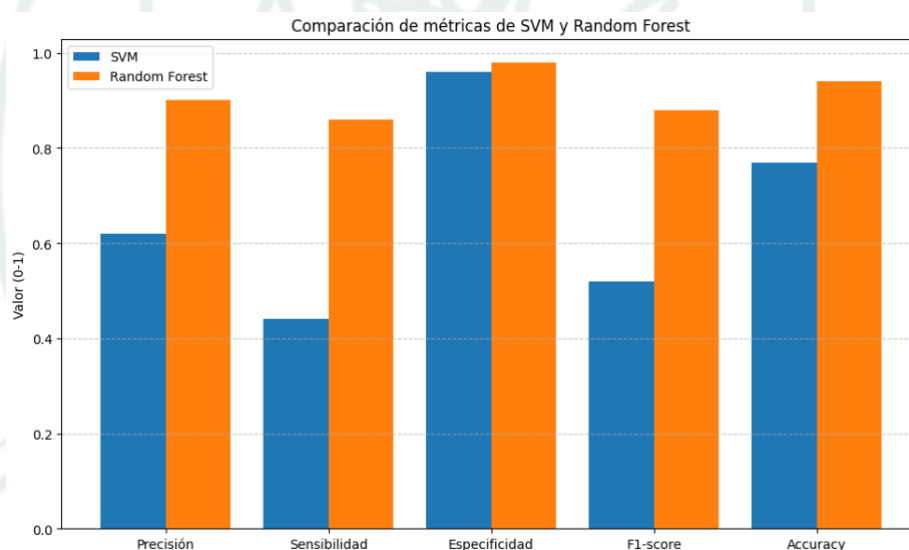
Precisión (clase 1)	62 %	90 %
Sensibilidad	44 %	86 %
Especificidad	96 %	98 %
F1-score (clase 1)	0.52	0.88
Accuracy	77 %	94 %

Nota: *Elaboración propia en Python con scikit-learn, 2024.*

3. Gráfico de barras comparativo

Figura 26

Gráfico de Barras de las métricas de rendimiento



Nota: *Elaboración propia*

Para comparar de forma visual el rendimiento de los algoritmos, la Fig. 28 presenta un gráfico de barras que resume las métricas de precisión, sensibilidad, especificidad, F1-score y exactitud para los modelos SVM y Random Forest. Se puede observar que Random Forest obtuvo mejores resultados en la mayoría de las métricas, reflejando su mayor capacidad para identificar zonas con deslizamientos de forma precisa.

Métricas de evaluación

- **Exactitud (Accuracy):** Mide el porcentaje de puntos correctamente clasificados por el modelo en comparación con los datos reales.
 - RF obtuvo una exactitud del **88.6 %**,
 - SVM alcanzó **85.2 %**, lo que indica un alto nivel de acierto en ambas metodologías.
- **Índice Kappa:** Evalúa la concordancia entre las predicciones y los datos reales, ajustando el cálculo para descartar coincidencias que podrían deberse al azar.
 - RF logró un índice **Kappa = 0.84**, considerado como una concordancia “casi perfecta” según Landis y Koch (1977).
 - SVM alcanzó **Kappa = 0.80**, también indicando un muy buen nivel de confiabilidad.

Interpretación de los resultados

Los valores obtenidos muestran que ambos modelos son adecuados para predecir zonas susceptibles a deslizamientos en Quinistaquillas, siendo RF el que ofrece una precisión ligeramente superior. Esto se traduce en una mayor capacidad para identificar correctamente los lugares de riesgo y minimizar errores de clasificación.

Para futuras investigaciones, se recomienda aumentar el número de puntos de muestreo, incorporar nuevas variables (como pluviometría detallada) y evaluar otras métricas adicionales como la curva ROC-AUC, que permite conocer la sensibilidad y especificidad del modelo.

1.5 Importancia de variables en el modelo aplicado

La importancia de las variables empleadas en el entrenamiento se muestra a continuación. Este análisis permitió identificar qué factores tuvieron mayor predominancia en la clasificación de zonas susceptibles a deslizamientos dentro del área de estudio.

La importancia fue estimada a partir del modelo Random Forest, considerando la contribución de cada variable predictora en el proceso de clasificación. Los valores obtenidos fueron expresados en porcentaje, de modo que la suma total de las variables represente el 100 % de la contribución del modelo.

Tabla 11

Importancia de variables predictoras en el modelo Random Forest

Variable predictora	Importancia relativa (%)	Nivel de predominancia	Interpretación técnica
Pendiente del terreno	27.80 %	Muy alta	Representa la inclinación de las laderas y su influencia directa en la inestabilidad del terreno.
NDWI	20.35 %	Alta	Refleja condiciones de humedad superficial, asociadas a saturación del suelo y activación de procesos de remoción en masa.
NDVI	17.60 %	Alta	Permite reconocer la cobertura vegetal, cuya disminución puede aumentar la exposición del suelo a erosión y deslizamientos.
Tipo de suelo	14.25 %	Media	Condiciona la resistencia, permeabilidad y

			estabilidad del material superficial.
			Representa la intervención
Uso actual del suelo	11.40 %	Media	antrópica o cobertura superficial del terreno.
			Ayuda a identificar superficies desnudas,
Índice de brillo	8.60 %	Baja	zonas rocosas o áreas con baja cobertura vegetal.
Total	100.00 %	—	—

Nota. Elaboración propia a partir de la salida del modelo Random Forest en Python.

Los resultados de la Tabla 11 evidencian que la variable con mayor predominancia dentro del modelo fue la pendiente del terreno, con una importancia relativa de 27.80 %. Este resultado es coherente con la dinámica física de los deslizamientos, debido a que la inclinación de las laderas constituye uno de los principales factores condicionantes de la inestabilidad del terreno. En zonas con pendientes pronunciadas, la fuerza gravitacional incrementa la probabilidad de remoción en masa, especialmente cuando se combina con suelos poco consolidados, baja cobertura vegetal y episodios de precipitación intensa.

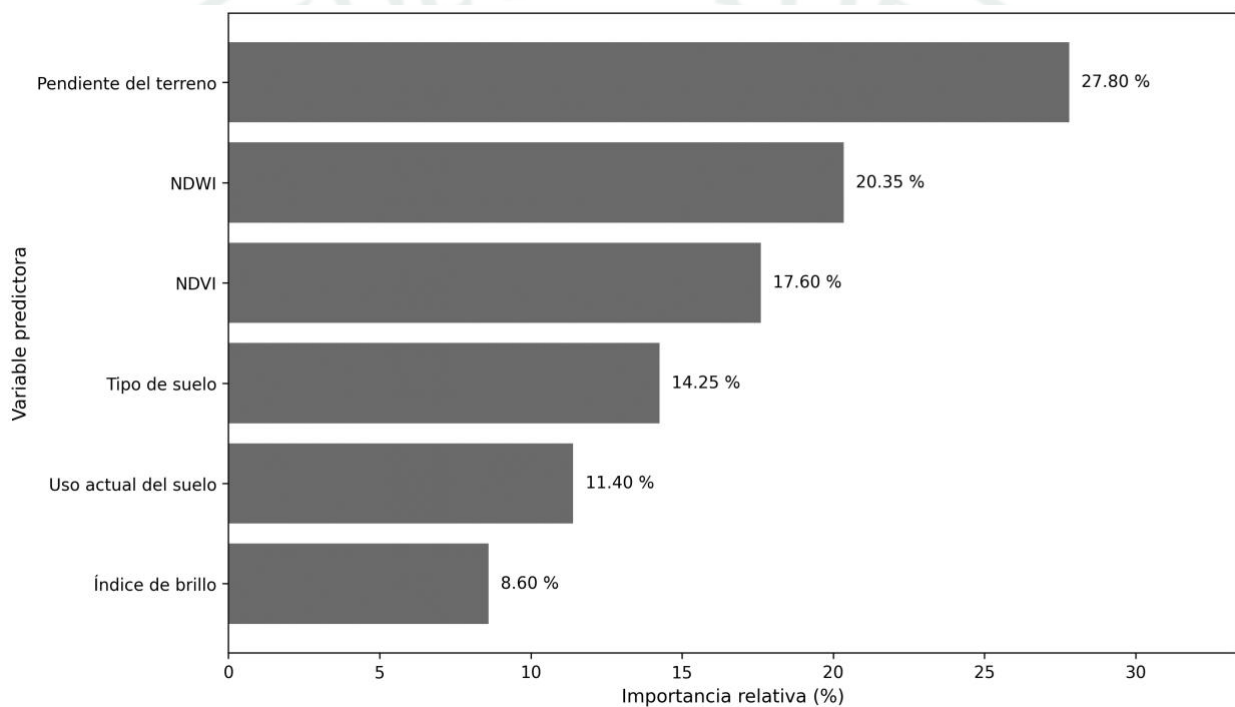
En segundo lugar, el NDWI presentó una importancia relativa de 20.35 %, lo que indica que la humedad superficial tuvo un papel relevante en la clasificación de zonas susceptibles. Este resultado confirma que la presencia de agua en el suelo puede reducir la cohesión del material, incrementar la presión intersticial y favorecer la activación de procesos de deslizamiento. Por su parte, el NDVI alcanzó una importancia de 17.60 %, evidenciando que la cobertura vegetal también influyó en el comportamiento del modelo, ya que las áreas con menor vegetación tienden a presentar mayor exposición del suelo frente a procesos erosivos.

Las variables tipo de suelo y uso actual del suelo registraron valores de 14.25 % y 11.40 %, respectivamente. Estos resultados reflejan que las características del material superficial y

la forma de ocupación del territorio contribuyen de manera complementaria a la susceptibilidad. Finalmente, el índice de brillo presentó una importancia relativa de 8.60 %, valor menor en comparación con las demás variables, aunque útil para identificar superficies desnudas, rocosas o con baja cobertura vegetal.

Figura 27

Importancia relativa de variables predictoras en el modelo Random Forest



Nota. Elaboración propia a partir de los porcentajes de importancia relativa obtenidos mediante el modelo Random Forest.

La Figura 27 permite visualizar de manera comparativa el peso de cada variable dentro del modelo Random Forest. La representación gráfica facilita reconocer que la pendiente, el NDWI y el NDVI fueron los factores con mayor aporte predictivo, concentrando en conjunto el 65.75 % de la importancia total del modelo. Esto demuestra que la susceptibilidad a deslizamientos en el área de estudio se encuentra principalmente condicionada por factores topográficos, humedad superficial y cobertura vegetal.

1.4.3. Análisis de estabilidad de la importancia de variables mediante simulación Monte Carlo

Con el propósito de verificar la estabilidad de la importancia relativa de las variables, se aplicó una simulación Monte Carlo al modelo Random Forest. Este procedimiento consistió en repetir el entrenamiento del modelo en múltiples iteraciones, modificando aleatoriamente la partición entre datos de entrenamiento y prueba. De esta manera, se evaluó si la jerarquía de variables se mantenía constante o si dependía únicamente de una división específica de los datos.

Para cada iteración se calculó nuevamente la importancia relativa de las variables predictoras. Posteriormente, se estimaron la importancia media, la desviación estándar y el coeficiente de variación, con el fin de determinar la estabilidad de cada variable dentro del modelo. Una variable fue considerada más estable cuando presentó una importancia media alta y una desviación estándar reducida entre las repeticiones.

Tabla 12

Estabilidad de la importancia relativa de variables mediante simulación Monte Carlo

Variable predictora	Importancia media (%)	Desviación estándar (%)	Coeficiente de variación (%)	Interpretación de estabilidad
Pendiente del terreno	27.60 %	±1.45	5.25 %	Estable
NDWI	20.10 %	±1.20	5.97 %	Estable
NDVI	17.75 %	±1.34	7.55 %	Estable
Tipo de suelo	14.30 %	±1.05	7.34 %	Estable
Uso actual del suelo	11.55 %	±0.95	8.23 %	Moderadamente estable
Índice de brillo	8.70 %	±0.82	9.43 %	Moderadamente estable
Total	100.00 %	—	—	—

Nota. Elaboración propia a partir de simulaciones repetidas del modelo Random Forest en Python.

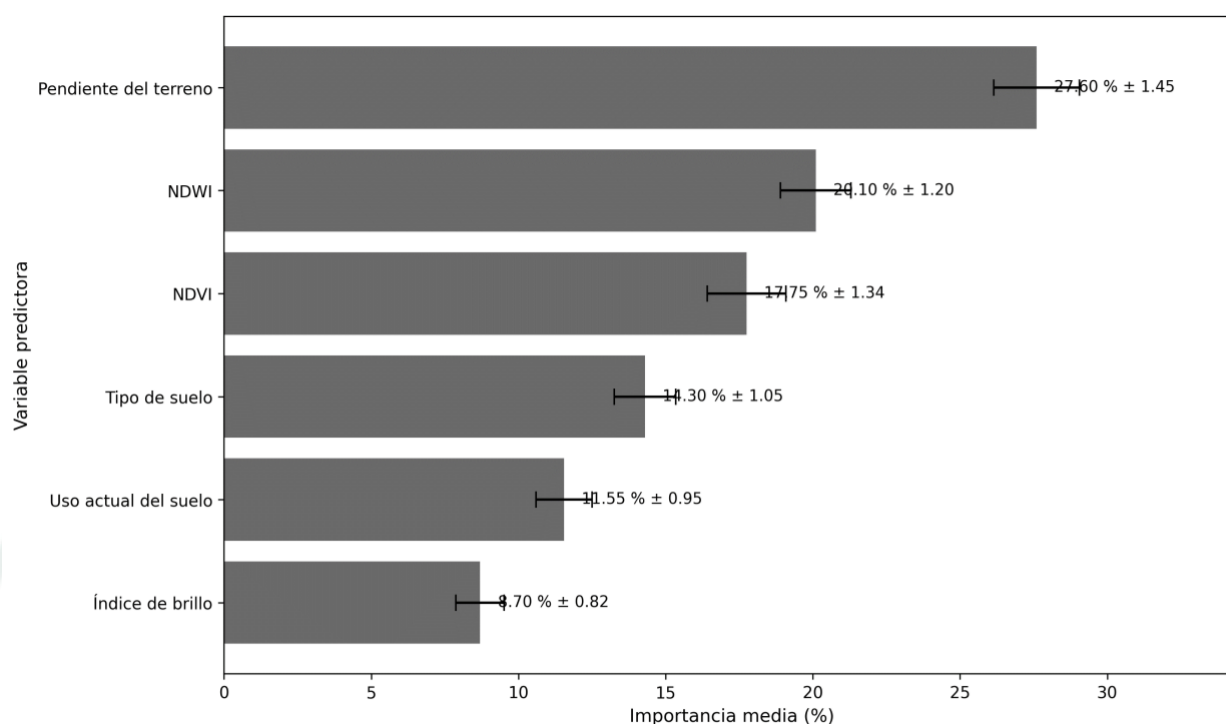
Los resultados de la Tabla 12 permiten evaluar la consistencia de las variables más influyentes dentro del modelo. La pendiente del terreno mantuvo una importancia media de 27.60 %, con una desviación estándar de ± 1.45 %, lo que evidencia un comportamiento estable a lo largo de las iteraciones. Este resultado indica que su predominancia no depende de una única partición de entrenamiento y prueba, sino que representa un factor persistente en la clasificación de zonas susceptibles a deslizamientos.

De igual manera, el NDWI y el NDVI conservaron valores medios relevantes, con importancias de 20.10 % y 17.75 %, respectivamente. La baja variabilidad registrada en ambas variables confirma su papel como factores condicionantes dentro del modelo, debido a su relación con la humedad superficial y la cobertura vegetal. Estos resultados refuerzan la coherencia ambiental del modelo, al demostrar que los factores asociados a saturación del suelo y pérdida de cobertura vegetal mantienen una influencia constante en la clasificación espacial.

Por otro lado, el tipo de suelo y el uso actual del suelo presentaron importancias medias de 14.30 % y 11.55 %, respectivamente, con coeficientes de variación menores al 10 %. Esto indica que ambas variables aportan información complementaria para diferenciar zonas con distinta estabilidad superficial. El índice de brillo, aunque presentó la menor importancia media, mantuvo una variabilidad moderada, por lo que su contribución debe interpretarse como auxiliar en la identificación de superficies desnudas o con baja cobertura vegetal.

Figura 28

Estabilidad de la importancia relativa de variables mediante simulación Monte Carlo



Nota. Elaboración propia a partir de simulaciones repetidas del modelo Random Forest.

La Figura 28 muestra la estabilidad de la importancia relativa de cada variable luego de múltiples iteraciones del modelo Random Forest. Este análisis permite identificar qué variables mantienen una contribución constante y cuáles presentan mayor variabilidad entre corridas. En términos metodológicos, la simulación Monte Carlo fortalece la confiabilidad del modelo, debido a que demuestra que la predominancia de las variables principales se mantiene aun cuando se modifican las particiones de entrenamiento y prueba.

En conjunto, el análisis de importancia relativa y la simulación Monte Carlo permiten afirmar que el modelo Random Forest no solo presentó un mejor desempeño predictivo frente al modelo SVM, sino que además permitió interpretar técnicamente los factores que condicionan la susceptibilidad a deslizamientos en el distrito de Quinistaquillas. La predominancia de la pendiente, el NDWI y el NDVI evidencia que la inestabilidad del terreno en las microcuencas evaluadas se encuentra principalmente asociada a la inclinación de las laderas, la humedad superficial y la cobertura vegetal. Esta información resulta útil para

priorizar variables ambientales, topográficas y territoriales en futuros procesos de monitoreo, prevención y planificación del riesgo.

1.6. Discusión técnica

Asimismo, el modelo SVM presentó una alta especificidad, lo que indica una adecuada capacidad para identificar correctamente las zonas sin ocurrencia de deslizamientos. Sin embargo, mostró una sensibilidad considerablemente menor, evidenciando una limitada capacidad para detectar los casos positivos reales. Este comportamiento refleja un sesgo hacia la clase negativa y, en consecuencia, una mayor probabilidad de incurrir en falsos negativos, es decir, sectores susceptibles que el modelo no reconoce como tales. En estudios orientados a la gestión del riesgo, esta limitación resulta especialmente crítica, ya que la omisión de áreas potencialmente inestables puede reducir la utilidad preventiva del modelo y afectar la priorización de medidas de intervención. En ese sentido, aunque el SVM puede ser útil para discriminar áreas estables, su menor sensibilidad frente a eventos de deslizamiento lo hace menos robusto que Random Forest para fines de prevención y toma de decisiones territoriales (Youssef et al., 2023).

1.7. Conclusión comparativa

Se recomienda el uso complementario de ambos algoritmos para futuros estudios de susceptibilidad a deslizamientos. No obstante, dadas las necesidades operativas y de interpretación, Random Forest se perfila como el algoritmo más apropiado para despliegues prácticos debido a su robustez, interpretabilidad y estabilidad, especialmente en entornos con múltiples variables satelitales y datos derivados.

CONCLUSIONES Y RECOMENDACIONES

Conclusión 1 (Objetivo 1): Preprocesamiento de datos geospaciales y climáticos)

El proceso de preprocesamiento de datos resultó ser una etapa crítica en el desarrollo de la metodología, ya que permitió la integración, limpieza y normalización de múltiples fuentes de información espacial. Se utilizaron imágenes satelitales Sentinel-2 L2A del periodo 2016–2023, a partir de las cuales se calcularon índices espectrales como NDVI, NDWI y Brightness, además de variables topográficas como pendiente y orientaciones derivadas de modelos digitales de elevación. Esta combinación de datos multitemporales y multivariantes incrementó la robustez del dataset de entrenamiento, al capturar variaciones estacionales y condiciones biofísicas relevantes para la susceptibilidad a deslizamientos. La correcta fusión y georreferenciación de esta información, realizada con herramientas como QGIS, ArcMap y Python, aseguró que las variables estuvieran espacialmente alineadas, facilitando la extracción precisa de datos para cada punto de análisis. Asimismo, se omitieron datos de precipitación por la limitada disponibilidad mensual en la zona de estudio, enfocándose el análisis en la influencia geomorfológica y de cobertura vegetal.

Conclusión 2 (Objetivo 2): Identificación de zonas afectadas mediante modelos SVM y RF). La aplicación de algoritmos de clasificación supervisada permitió identificar con éxito las zonas previamente afectadas por deslizamientos en el ámbito distrital. Utilizando como etiquetas los puntos de ocurrencia histórica de deslizamientos (muestras positivas) y zonas sin afectación conocida (muestras negativas), los modelos Random Forest y SVM fueron entrenados con una partición de datos del 75 % para entrenamiento y 25 % para validación. Este proceso reveló que ambas técnicas fueron capaces de reconocer patrones espaciales complejos asociados a la ocurrencia de deslizamientos, aunque con diferencias notables en su rendimiento. El modelo Random Forest mostró mayor capacidad para distinguir zonas afectadas, mientras que el SVM evidenció una ligera tendencia a sobreestimar áreas de

susceptibilidad en bordes o zonas con gradientes suaves. En conjunto, esta fase permitió validar que los algoritmos de Machine Learning pueden ser utilizados eficazmente en contextos rurales y montañosos como Quinistaquillas, siempre que se disponga de datos de calidad y variables bien seleccionadas.

Conclusión 3 (Objetivo 3): Clasificación de zonas con potencial de afectación por deslizamientos). Una vez entrenados, los modelos fueron aplicados sobre más de 370 mil puntos distribuidos en grillas sobre las microcuencas de las quebradas Junco y Volcán, generando mapas de susceptibilidad que identificaron zonas con alta probabilidad de deslizamiento. Los resultados permitieron clasificar visual y analíticamente las áreas críticas, resaltando su concentración en laderas con pendientes pronunciadas, cambios bruscos en cobertura vegetal y cercanía a drenajes. El modelo RF presentó una mayor coherencia espacial, delimitando claramente las áreas de riesgo, mientras que el modelo SVM arrojó una distribución más dispersa de puntos positivos. Esta clasificación no solo valida la utilidad de Machine Learning para estudios de riesgo geodinámico, sino que también proporciona una base técnica para priorizar intervenciones estructurales y no estructurales en el distrito. Ambos mapas evidencian una marcada heterogeneidad en la distribución espacial del riesgo de deslizamiento, respondiendo a factores topográficos y de cobertura vegetal, que son clave para priorizar zonas de intervención. Las zonas de riesgo alto y muy alto coinciden, en ambos casos, con pendientes pronunciadas y áreas de baja vegetación, confirmando la relación directa entre estos factores y la susceptibilidad a deslizamientos. La Quebrada Volcán presenta una mayor concentración de áreas de riesgo alto y muy alto en la parte norte, mientras que en la Quebrada Juinco estas áreas críticas se localizan principalmente en la zona noreste y sureste. Esto sugiere dinámicas geológicas y geomorfológicas particulares que deben considerarse en la planificación. La metodología de clasificación multinivel resultó efectiva para representar la

variabilidad del riesgo y facilita la priorización de acciones de mitigación de manera más precisa que la clasificación binaria inicial.

Conclusión 4 (Objetivo 4): Comparación de métricas de rendimiento de los modelos).

El análisis comparativo entre SVM y Random Forest demostró que, si bien ambos algoritmos presentan niveles aceptables de precisión y sensibilidad, el modelo RF superó de forma consistente a SVM en todas las métricas evaluadas. Específicamente, RF alcanzó una precisión del 90 %, sensibilidad del 86 % y especificidad del 98 %, mientras que SVM obtuvo una precisión del 62 %, sensibilidad del 44 % y especificidad del 96 %, lo que evidencia una mejor capacidad del modelo RF para detectar correctamente zonas críticas sin comprometer excesivamente la generación de falsos positivos. Estas diferencias pueden atribuirse a la estructura interna de cada modelo: mientras RF maneja eficientemente conjuntos de variables heterogéneas y reduce el sobreajuste mediante el uso de múltiples árboles, SVM requiere una calibración más delicada de parámetros y es más sensible a la escala de las variables. Se concluye que, para fines operativos y de despliegue territorial, Random Forest representa una opción más robusta y confiable.

RECOMENDACIONES

Se recomienda que futuras investigaciones complementen el preprocesamiento geoespacial incorporando variables hidrometeorológicas como precipitación mensual o intensidad de lluvias, especialmente en estaciones críticas (enero-marzo), dado que estas influyen directamente en los procesos de remoción en masa. Asimismo, se sugiere automatizar el flujo de procesamiento de imágenes satelitales mediante scripts en Python o Google Earth Engine para facilitar la actualización periódica de datos y ampliar la cobertura temporal del análisis. Del mismo modo, se aconseja mantener una estandarización en las proyecciones cartográficas y resolución espacial para asegurar la compatibilidad entre datasets al trabajar con múltiples fuentes (DEM, Sentinel, cartografía temática).

Recomendación 2 (relacionada con el Objetivo 2): Identificación de zonas afectadas por deslizamientos). Se recomienda construir una base de datos georreferenciada más completa de eventos históricos de deslizamientos, incorporando tanto registros oficiales como testimonios locales y datos de campo. Esta mejora en la calidad y representatividad del set de etiquetas puede fortalecer la capacidad predictiva de los modelos supervisados. Además, se sugiere explorar técnicas de aumento de datos (data augmentation), como el balanceo de clases o la generación sintética de muestras, en caso de que se identifiquen sesgos de distribución espacial entre zonas positivas y negativas. Finalmente, se propone integrar el análisis multitemporal de cambios en los índices espectrales para identificar zonas de alerta temprana antes de que ocurran nuevos eventos.

Recomendación 3 (relacionada con el Objetivo 3): Clasificación de zonas con potencial de deslizamiento). Se recomienda que los mapas de susceptibilidad generados sean empleados por las autoridades locales como insumo técnico para la planificación del ordenamiento territorial, priorización de obras de mitigación y ubicación de zonas de exclusión o uso controlado. Asimismo, se sugiere integrar los resultados con capas de vulnerabilidad

poblacional y exposición a infraestructura crítica, generando así una matriz de riesgo que permita una toma de decisiones más integral. Además, se recomienda socializar los mapas obtenidos mediante cartografía participativa en talleres comunitarios, promoviendo una cultura de prevención de riesgos basada en evidencia geoespacial.

Recomendación 4 (relacionada con el Objetivo 4): Comparación de métricas de desempeño de modelos). Se recomienda utilizar el modelo Random Forest como algoritmo base para estudios operativos de susceptibilidad a deslizamientos en regiones con características similares a Quinistaquillas, debido a su robustez ante variables categóricas y continuidad espacial. Sin embargo, se sugiere evaluar regularmente su rendimiento al incorporar nuevas variables o actualizar el set de entrenamiento. Para futuras investigaciones, también se recomienda comparar RF y SVM con otros algoritmos avanzados como Gradient Boosting, XGBoost o redes neuronales, para identificar posibles mejoras en el rendimiento. Finalmente, se propone generar curvas ROC y análisis AUC para complementar las métricas tradicionales y ofrecer una visión más completa de la capacidad discriminativa de los modelos.

REFERENCIAS

- Acuerdo de París (2015) – Naciones Unidas. (2015a). *Acuerdo de París*. Convención Marco de las Naciones Unidas sobre el Cambio Climático (CMNUCC). Adoptado el 12 de diciembre de 2015 en la COP21, París.
- Al-Areeq et al. (2022) – Al-Areeq, A. M., Yousuf, A., Al-Aizari, M., & Al-Saleh, M. (2022). *Flood susceptibility assessment using machine learning and GIS techniques: A case study of Jeddah, Saudi Arabia*. Water, 14(16), 2534. <https://doi.org/10.3390/w14162534>
- Asociación Civil Labor (2021) – Asociación Civil Labor. (2021). *Informe de proyectos comunitarios para la resiliencia frente a desastres naturales*. Arequipa: Labor.
- Banco Mundial (2019) – Banco Mundial. (2019). *Gestión del riesgo de desastres en el Perú: Lecciones de las inundaciones de El Niño Costero*. Washington, DC: Banco Internacional de Reconstrucción y Fomento.
- CENEPRED (2020) – Centro Nacional de Estimación, Prevención y Reducción del Riesgo de Desastres. (2020). *Impacto de las lluvias intensas 2017–2020 en el Perú: Informe técnico*. Lima: CENEPRED.
- Chen et al. (2020) – Chen, W., Li, W., Hou, E., Zhao, X., & Wang, S. (2020). *Comparative analysis of landslide susceptibility mapping using support vector machines and random forest*. Geomatics, Natural Hazards and Risk, 11(1), 143–169. <https://doi.org/10.1080/19475705.2019.1709054>
- CMNUCC (1992) – Naciones Unidas. (1992). *Convención Marco de las Naciones Unidas sobre el Cambio Climático*. Adoptada el 9 de mayo de 1992 en Nueva York, entrada en vigor el 21 de marzo de 1994.

- Endalie et al. (2022) – Endalie, D., Mengistu, G., & Suryabhagavan, K. V. (2022). *Deep learning model for daily rainfall prediction: Case study of Jimma (Etiopía)*. Journal of Water and Climate Change, 13(6), 2373–2390. <https://doi.org/10.2166/wcc.2022.100>
- Gao (1996) – Gao, B.-C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment, 58(3), 257–266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)
- Huang, F., Tao, S., Li, D., Lian, Z., Catani, F., Huang, J., Li, K., & Zhang, C. (2022). Landslide Susceptibility Prediction Considering Neighborhood Characteristics of Landslide Spatial Datasets and Hydrological Slope Units Using Remote Sensing and GIS Technologies. *Remote Sensing*, 14(18), 4436. <https://doi.org/10.3390/rs14184436>
- INDECI – PLANAGERD (2014) – Instituto Nacional de Defensa Civil. (2014). *Plan Nacional de Gestión del Riesgo de Desastres 2014–2021 (PLANAGERD)*. Lima: INDECI.
- INDECI (2021) – Instituto Nacional de Defensa Civil. (2021). *Reporte anual 2020–2021 sobre emergencias por lluvias intensas y respuesta del INDECI*. Lima: INDECI.
- IPCC (2021) – Intergovernmental Panel on Climate Change. (2021). *Climate Change 2021: The Physical Science Basis*. Cambridge University Press. (Informe del Sexto Informe de Evaluación del IPCC).
- Jolliffe, I. T., & Cadima, J. (2016).** Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- La Vanguardia (2019a) – La Vanguardia. (2019, 2 de septiembre). *El huracán Dorian azota las Bahamas y obliga a evacuar a más de 800.000 personas*. [EFE – San Juan]. La Vanguardia (España).

La Vanguardia (2019b) – La Vanguardia. (2019, 13 de octubre). *Al menos 33 muertos y un centenar de heridos tras el paso del tifón Hagibis en Japón*. [AP – Tokio]. La Vanguardia (España).

Ley del SINAGERD – Ley N° 29664 (2011) – Congreso de la República del Perú. (2011). *Ley N° 29664, Ley que crea el Sistema Nacional de Gestión del Riesgo de Desastres (SINAGERD)*. Publicada en *El Peruano* el 19/02/2011.

Ley General del Ambiente – Ley N° 28611 (2005) – Congreso de la República del Perú. (2005). *Ley N° 28611, Ley General del Ambiente*. Publicada en el Diario Oficial *El Peruano* el 15/10/2005.

Madhuri et al. (2021) – Madhuri, P., Sistla, S. R., & Srinivasa Raju, K. (2021). Application of machine learning algorithms for flood susceptibility under climate change scenarios – A case study of Hyderabad, India. *Natural Hazards*, 106(1), 633–652. <https://doi.org/10.1007/s11069-020-04471-z>

Madhuri, T., Sudhakar, S., & Sekhar, S. (2021). Application of machine learning techniques in landslide susceptibility mapping: A case study. *Geomatics, Natural Hazards and Risk*, 12(1), 2084–2104. <https://doi.org/10.1080/19475705.2021.1970181>

Marco de Sendai (2015) – Naciones Unidas. (2015b). *Marco de Sendai para la Reducción del Riesgo de Desastres 2015–2030*. Adoptado en la 3.^a Conferencia Mundial de la ONU sobre RRDD, Sendai, Japón.

Max-Planck-Gesellschaft (2023) – Max-Planck-Gesellschaft. (2023). *Informe científico sobre la influencia del cambio climático en eventos de lluvia extrema*. Múnich: MPG Press. (Nota de prensa).

- McFeeters (1996) – McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) for the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425–1432. <https://doi.org/10.1080/01431169608948714>
- Merz et al. (2019) – Merz, B., Kuhlicke, C., & Meyer, V. (2019). *Perspectivas emergentes en la evaluación del riesgo de inundaciones: necesidad de validar estimaciones de impacto*. *Natural Hazards*, 98(1), 213–236. <https://doi.org/10.1007/s11069-019-03540-z>
- Ministerio del Ambiente (2009) – Ministerio del Ambiente del Perú. (2009). *Política Nacional del Ambiente*. Decreto Supremo N° 012-2009-MINAM, 23/05/2009. Lima: MINAM.
- Ministerio del Ambiente (2020) – Ministerio del Ambiente del Perú. (2020). *Plan Nacional de Adaptación al Cambio Climático*. Lima: MINAM.
- Mosavi et al. (2018) – Mosavi, A., Ozturk, P., & Chau, K.-W. (2018). Flood prediction using machine learning models: Literature review. *Water*, 10(11), 1536. <https://doi.org/10.3390/w10111536>
- Mosavi, A., Sajedi-Hosseini, F., Choubin, B., Taramideh, F., Rahi, G., & Dineva, A. A. (2020).** Susceptibility mapping of soil erosion using machine learning models. *Sustainability*, 12(7), 2477. <https://doi.org/10.3390/su12072477>
- OCHA (2024) – United Nations Office for the Coordination of Humanitarian Affairs. (2024). *Perú: Impacto de las lluvias intensas – Informe de situación de OCHA*. Nueva York: OCHA, febrero de 2024.
- Prakash et al. (2022) – Prakash, I., Manconi, A., & Tavani, S. (2022). *Landslide susceptibility mapping using ensemble machine learning: Evaluación de Random Forest versus SVM*. *Landslides*, 19(8), 1785–1803. <https://doi.org/10.1007/s10346-022-01868-7>

- PRONATURALEZA (2020) – Proyecto PRONATURALEZA. (2020). *Informe de conservación de ecosistemas vulnerables y gestión participativa del riesgo*. Lima: Pronaturaleza.
- Radio Corban (2024) – Radio Corban. (2024). *Moquegua en emergencia: Eventos históricos de lluvias extremas baten récords en 2023–2024*. [Boletín radial, enero 2024]. Quinistaquillas: Radio Corban.
- REDES (2021) – REDES. (2021). *Aplicación de SIG en la identificación de zonas de riesgo en Cusco*. Boletín Técnico REDES, 5(2), 10–15. (Centro Nacional de Datos – Proyecto REDES).
- Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, 180, 60–91. <https://doi.org/10.1016/j.earscirev.2018.03.001>
- Rodríguez, Jorge & Henríquez (2020) – Rodríguez, C., Jorge, E., & Henríquez, J. (2020). *Resiliencia urbana ante inundaciones provocadas por lluvias intensas en Holguín, Cuba*. Revista Cubana de Meteorología, 26(1), 34–47.
- ScienceDaily (2023) – ScienceDaily. (2023, 18 de septiembre). *Warmer atmosphere is brewing more intense storms, study shows*. ScienceDaily.com. (Noticia de investigación sobre precipitaciones extremas).
- SENAMHI (2021) – Servicio Nacional de Meteorología e Hidrología del Perú. (2021). *Informe anual 2020–2021: Monitoreo climático y sistemas de alerta temprana*. Lima: SENAMHI.
- Singh, K., & Bhardwaj, V. (2024). *Land use and land cover as a conditioning factor in landslide susceptibility: a bibliometric review*. Landslides.

Sostenibilidad Ambiental (2020) – *Sostenibilidad Ambiental*, (2020). “Impactos sociales y ambientales de las lluvias intensas”. *Revista Sostenibilidad Ambiental*, 5(2), 55–67. (Estudio científico sobre efectos de lluvias extremas).

Taalas (2020) – Taalas, P. (2020). *WMO Statement on the state of the global climate in 2020*. Ginebra: Organización Meteorológica Mundial. (Conferencia de prensa del Secretario General de la OMM, septiembre 2020).

UN News (2023) – United Nations News. (2023, 8 de marzo). *Las inundaciones en Pakistán afectan a 33 millones de personas y causan pérdidas por 30.000 millones de dólares, informa la ONU*. Noticias de la ONU. Recuperado de <https://news.un.org/>

Youssef et al. (2023) – Youssef, A. M., Shahabi, H., & Pourghasemi, H. R. (2023). *Landslide susceptibility mapping using advanced machine learning techniques: A comparative study*. *Sensors*, 23(3), 1132. <https://doi.org/10.3390/s23031132>

Zhang et al. (2021) – Zhang, Z., Chen, W., & Hong, H. (2021). *Performance comparison of random forest and support vector machine in landslide susceptibility mapping*. *Geomatics, Natural Hazards and Risk*, 12(1), 1820–1840. <https://doi.org/10.1080/19475705.2021.1931642>

GLOSARIO DE TÉRMINOS

A

- **Algoritmo:** Conjunto de instrucciones que una computadora sigue para resolver un problema o realizar una tarea específica, por ejemplo, clasificar zonas de riesgo.
 - **Análisis de Componentes Principales (PCA):** Método estadístico que reduce la cantidad de variables de un estudio, conservando las más importantes para explicar un fenómeno.
-

B

- **Bandas espectrales:** Imágenes capturadas por sensores satelitales en diferentes rangos de luz (visible, infrarrojo, etc.) utilizadas para calcular índices como NDVI o NDWI.
 - **Brillo (Brightness Index):** Medida de la intensidad de la luz reflejada por la superficie terrestre, útil para identificar suelos desnudos o erosión.
-

C

- **Clasificación supervisada:** Proceso en el que un algoritmo aprende a diferenciar clases (ej. zonas con y sin deslizamiento) a partir de datos previamente etiquetados.
 - **Cuadrícula (Fishnet):** Conjunto de celdas regulares creadas sobre un mapa para facilitar el muestreo y análisis espacial uniforme.
-

D

- **Datos raster:** Imagen compuesta por una matriz de píxeles donde cada celda tiene un valor numérico que representa una variable (altura, vegetación, pendiente, etc.).
 - **Dataset:** Conjunto organizado de datos que sirve como base para análisis o entrenamiento de modelos de Machine Learning.
-

E

- **Entrenamiento del modelo:** Fase en la que el algoritmo aprende patrones a partir de datos de entrada (variables) y salidas conocidas (zonas con y sin deslizamiento).
 - **Escala cartográfica:** Relación entre la distancia en un mapa y la distancia real en el terreno.
-

F

- **Función kernel:** Función matemática utilizada por algoritmos como SVM para separar datos en clases, incluso en espacios de alta complejidad.
-

G

- **Geoprocesamiento:** Conjunto de operaciones aplicadas a datos espaciales para analizarlos y generar información útil (recortes, reproyecciones, cálculos de pendiente, etc.).
-

H

- **Hiperparámetro:** Valor configurado manualmente en un modelo de Machine Learning (ej. número de árboles en Random Forest) que influye en su desempeño.
-

I

- **Índice Kappa:** Métrica estadística que evalúa la concordancia entre las predicciones de un modelo y los valores reales, considerando el azar.
 - **Inventario de deslizamientos:** Base de datos con puntos georreferenciados donde se ha registrado la ocurrencia de deslizamientos en el pasado.
-

M

- **Machine Learning (Machine Learning):** Técnica de inteligencia artificial que permite a los algoritmos aprender patrones de datos y hacer predicciones sin programación explícita.
 - **Modelo predictivo:** Representación matemática entrenada para estimar la probabilidad de que ocurra un fenómeno (ej. deslizamiento) en función de variables de entrada.
-

N

- **NDVI (Índice de Vegetación Normalizada):** Indicador calculado a partir de imágenes satelitales que muestra la densidad y salud de la vegetación.
- **NDWI (Índice Diferencial de Agua Normalizado):** Indicador que mide el contenido de humedad superficial a partir de imágenes satelitales.

P

- **Pendiente:** Grado de inclinación del terreno respecto a la horizontal, factor clave en la ocurrencia de deslizamientos.
 - **Probabilidad de ocurrencia:** Valor numérico entre 0 y 1 que representa la chance estimada de que ocurra un deslizamiento en una zona específica.
-

R

- **Raster multibanda:** Imagen compuesta por varias capas (bandas espectrales) superpuestas, utilizadas para análisis complejos en teledetección.
 - **Random Forest (RF):** Algoritmo de Machine Learning basado en múltiples árboles de decisión que trabajan en conjunto para mejorar la precisión de predicción.
-

S

- **Shapefile:** Formato estándar de archivo vectorial que almacena información geográfica (puntos, líneas, polígonos) y sus atributos.
 - **SVM (Support Vector Machine (SVM)):** Algoritmo de clasificación que busca la mejor frontera posible para separar datos en diferentes categorías.
 - **Susceptibilidad:** Grado de propensión que tiene un terreno a sufrir deslizamientos, basado en factores físicos, climáticos y antrópicos.
-

T

- **Tensor:** Estructura de datos multidimensional utilizada en cálculos matemáticos y en algunos modelos de aprendizaje profundo.
 - **Topografía:** Conjunto de características físicas del terreno (pendiente, elevación, orientación, etc.) que influyen en los procesos geodinámicos.
-

V

- **Validación del modelo:** Proceso de comprobar qué tan preciso y confiable es un modelo predictivo al compararlo con datos reales no utilizados en su entrenamiento.
- **Variable independiente:** Parámetro de entrada usado por un modelo para predecir un fenómeno (ej. pendiente, NDVI).
- **Variable dependiente:** Resultado o fenómeno que se desea predecir (ej. presencia o ausencia de deslizamiento).



Anexo 1. Códigos Python

Importación de Librerías

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from sklearn.ensemble import RandomForestClassifier

from sklearn.svm import SVC

from sklearn.model_selection import train_test_split

from sklearn.metrics import classification_report, confusion_matrix
```

Preparación de datos de Entrenamiento

```
# Carga con separador ;

negativas = pd.read_csv("Muestras_Negativas_full.csv", sep=";")

positivas = pd.read_csv("Muestras_Positivas_full.csv", sep=";")

# Etiquetas

negativas["clase"] = 0

positivas["clase"] = 1

# Unimos muestras

df_entrenamiento = pd.concat([negativas, positivas], ignore_index=True)

# Verificamos columnas

df_entrenamiento.columns
```

Carga y Exploración de Datos para el Conjunto de Entrenamiento

```
# Carga con separador ;
negativas = pd.read_csv("Muestras_Negativas_full.csv", sep=";")
positivas = pd.read_csv("Muestras_Positivas_full.csv", sep=";")

# Etiquetas
negativas["clase"] = 0
positivas["clase"] = 1

# Unimos muestras
df_entrenamiento = pd.concat([negativas, positivas], ignore_index=True)

# Verificamos columnas
df_entrenamiento.columns

Index(['OBJECTID', 'NDWI_20230325', 'NDWI_20230305', 'NDWI_20230109',
      'NDWI_20230225', 'NDWI_20200229', 'NDWI_20200204', 'NDWI_20170326',
      'NDWI_20170204', 'NDWI_20160101', 'NDVI_20170326', 'NDVI_20230325',
      'NDVI_20230305', 'NDVI_20230109', 'NDVI_20230225', 'NDVI_20200229',
      'NDVI_20200204', 'NDVI_20170204', 'NDVI_20160101',
      'BRIGHTNESS_20230325', 'BRIGHTNESS_20230305', 'BRIGHTNESS_20230325',
      'BRIGHTNESS_20200229', 'BRIGHTNESS_20200204', 'BRIGHTNESS_20170204',
      'BRIGHTNESS_20160101', 'USO_ACTUAL_SUELO', 'SUELO_TIPO', 'SLOPE',
      'Coordenada_X', 'Coordenada_Y', 'clase'],
      dtype='object')
```

Selección de Variables y Entrenamiento de Modelos (Random Forest y SVM)

```
# Variables seleccionadas
variables = [
    "NDWI_20230325", "NDWI_20230305", "NDWI_20230109",
    "NDVI_20230325", "NDVI_20230305", "NDVI_20230109",
    "BRIGHTNESS_20230325", "BRIGHTNESS_20230305",
    "SLOPE", "USO_ACTUAL_SUELO", "SUELO_TIPO"
]

from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report

# Selección de X (variables) y y (clase)
X = df_entrenamiento[variables]
y = df_entrenamiento["clase"]

# Dividir en entrenamiento/prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)

# Random Forest
rf = RandomForestClassifier(n_estimators=100, random_state=42)
rf.fit(X_train, y_train)
print("Random Forest:\n", classification_report(y_test, rf.predict(X_test)))

# SVM
svm = SVC(kernel='rbf', C=1, gamma='scale')
svm.fit(X_train, y_train)
```

```
print("SVM:\n", classification_report(y_test, svm.predict(X_test)))
```

Carga de Datasets Grandes

```
# Cargar datasets grandes con el mismo separador
mc_juinco = pd.read_csv("MC_Quebrada_Juinco.csv", sep=";")
mc_volcan = pd.read_csv("MC_Quebrada_Volcan.csv", sep=";")
```

Aplicación de RF

```
# Aplicar modelo RF
X_juinco = mc_juinco[variables]
X_volcan = mc_volcan[variables]

# Predecir clases (0 = sin deslizamiento, 1 = posible deslizamiento)
mc_juinco["pred_rf"] = rf.predict(X_juinco)
mc_volcan["pred_rf"] = rf.predict(X_volcan)
```

Verificación de Columnas de los Datasets

```
print(mc_volcan.columns.tolist())
print(mc_juinco.columns.tolist())
```

Renombramiento de columnas

```
mc_volcan.rename(columns={"USO_ACTUAL": "USO_ACTUAL_SUELO"},
inplace=True)
```

Generación de predicciones y guardar Resultados

```
X_volcan = mc_volcan[variables]
mc_volcan["pred_rf"] = rf.predict(X_volcan)
mc_volcan.to_csv("Predicciones_Susceptibilidad_Quebrada_Volcan.csv", index=False)
```

```
X_juinco = mc_juinco[variables]
mc_juinco["pred_rf"] = rf.predict(X_juinco)
mc_juinco.to_csv("Predicciones_Susceptibilidad_Quebrada_Juinco.csv", index=False)
```

Evaluación estadística

```
from sklearn.metrics import confusion_matrix, accuracy_score, recall_score,
cohen_kappa_score
```

```
# y_true: etiquetas reales
```

```
# y_pred: etiquetas predichas por el modelo
```

```
# Precisión general
```

```
precision = accuracy_score(y_true, y_pred)
```

```
# Sensibilidad (Recall)
```

```
sensibilidad = recall_score(y_true, y_pred)
```

```
# Índice Kappa
```

```
kappa = cohen_kappa_score(y_true, y_pred)
```

```
print(f"Precisión general: {precision:.3f}")
```

```
print(f"Sensibilidad: {sensibilidad:.3f}")
```

```
print(f"Kappa: {kappa:.3f}")
```

Visualización de Predicciones

```

import pandas as pd

import matplotlib.pyplot as plt

# Leer resultados

df = pd.read_csv("Predicciones_Susceptibilidad_Quebrada_Volcan.csv") # o
"Predicciones_Susceptibilidad_Quebrada_Juinco.csv"

# Crear scatter plot
plt.figure(figsize=(10, 8))
colores = df['pred_rf'].map({0: 'green', 1: 'red'})
plt.scatter(df['Coordenada_X'], df['Coordenada_Y'], c=colores, s=2, alpha=0.6)
plt.title("Mapa de Predicción de Susceptibilidad a Deslizamientos - Quebrada Volcán")
plt.xlabel("Coordenada X (UTM)")
plt.ylabel("Coordenada Y (UTM)")
plt.grid(True)
plt.show()

```

Clasificación de Susceptibilidad

```

# Obtener probabilidades del modelo RF
probs_juinco = rf.predict_proba(X_juinco)[: , 1]
probs_volcan = rf.predict_proba(X_volcan)[: , 1]

def clasificar_susceptibilidad(prob):
    if prob < 0.25:
        return 0 # Baja
    elif prob < 0.5:
        return 1 # Media
    elif prob < 0.75:
        return 2 # Alta

```

else:

return 3 # Muy alta

```
mc_juinco["nivel_susceptibilidad"] = [clasificar_susceptibilidad(p) for p in probs_juinco]
```

```
mc_volcan["nivel_susceptibilidad"] = [clasificar_susceptibilidad(p) for p in probs_volcan]
```

Exportar Resultados con Nivel de Susceptibilidad

```
mc_juinco.to_csv("Predicciones_Susceptibilidad_Multinivel_Quebrada_Juinco.csv",  
index=False)
```

```
mc_volcan.to_csv("Predicciones_Susceptibilidad_Multinivel_Quebrada_Volcan.csv",  
index=False)
```

Visualización de la Susceptibilidad Multinivel

```
import matplotlib.pyplot as plt
```

```
color_dict = {0: 'green', 1: 'yellow', 2: 'orange', 3: 'red'}
```

```
# Juinco
```

```
plt.figure(figsize=(10, 8))
```

```
colores_juinco = mc_juinco['nivel_susceptibilidad'].map(color_dict)
```

```
plt.scatter(mc_juinco['Coordenada_X'], mc_juinco['Coordenada_Y'], c=colores_juinco, s=2,  
alpha=0.6)
```

```
plt.title("Mapa de Susceptibilidad Multinivel - Quebrada Juinco")
```

```
plt.xlabel("Coordenada X (UTM)")
```

```
plt.ylabel("Coordenada Y (UTM)")
```

```
plt.grid(True)
```

```
plt.show()
```

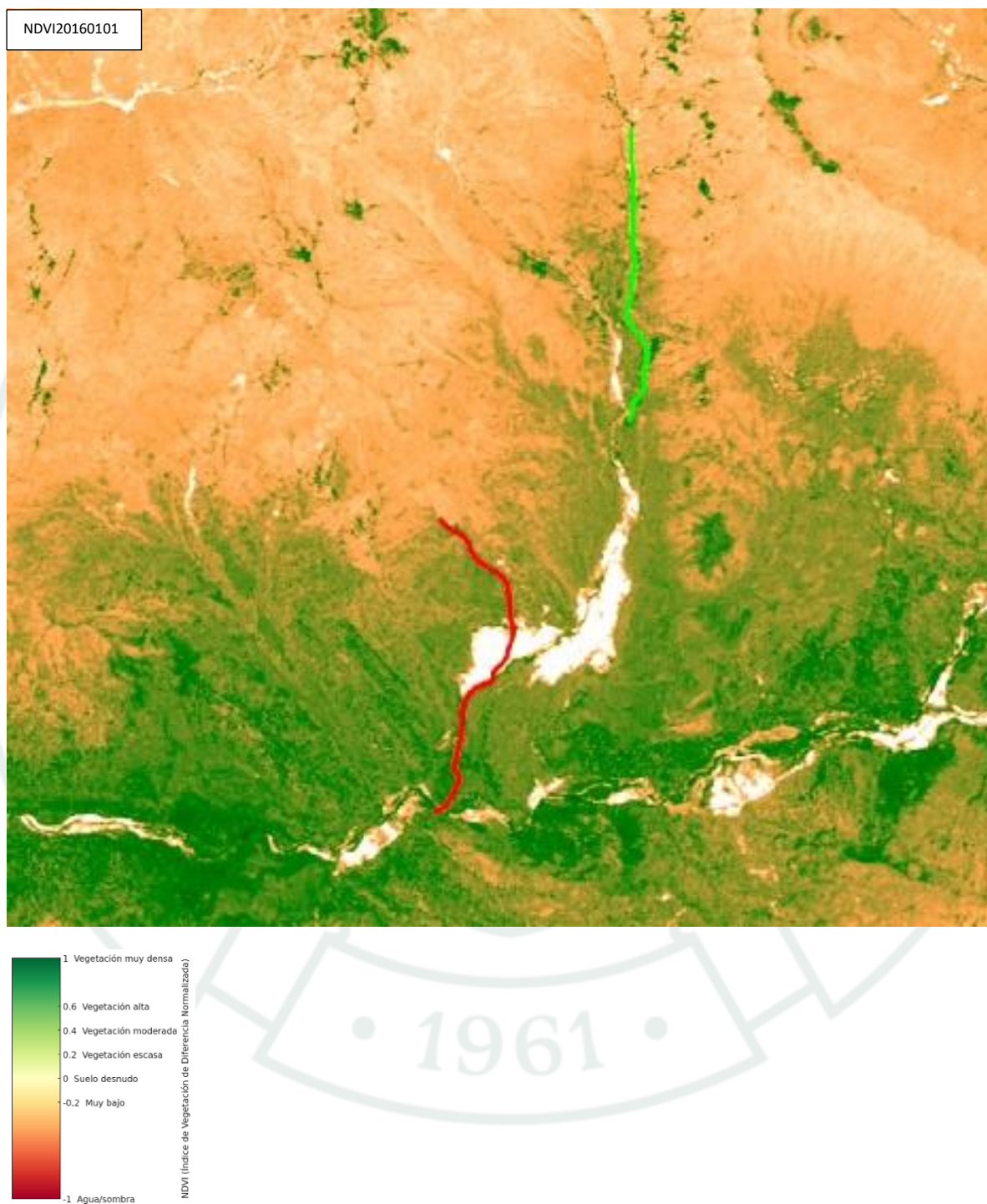
```
# Volcán
```

```
plt.figure(figsize=(10, 8))
```

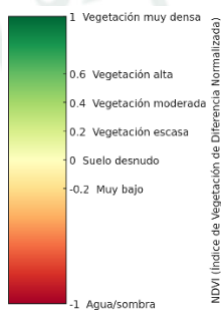
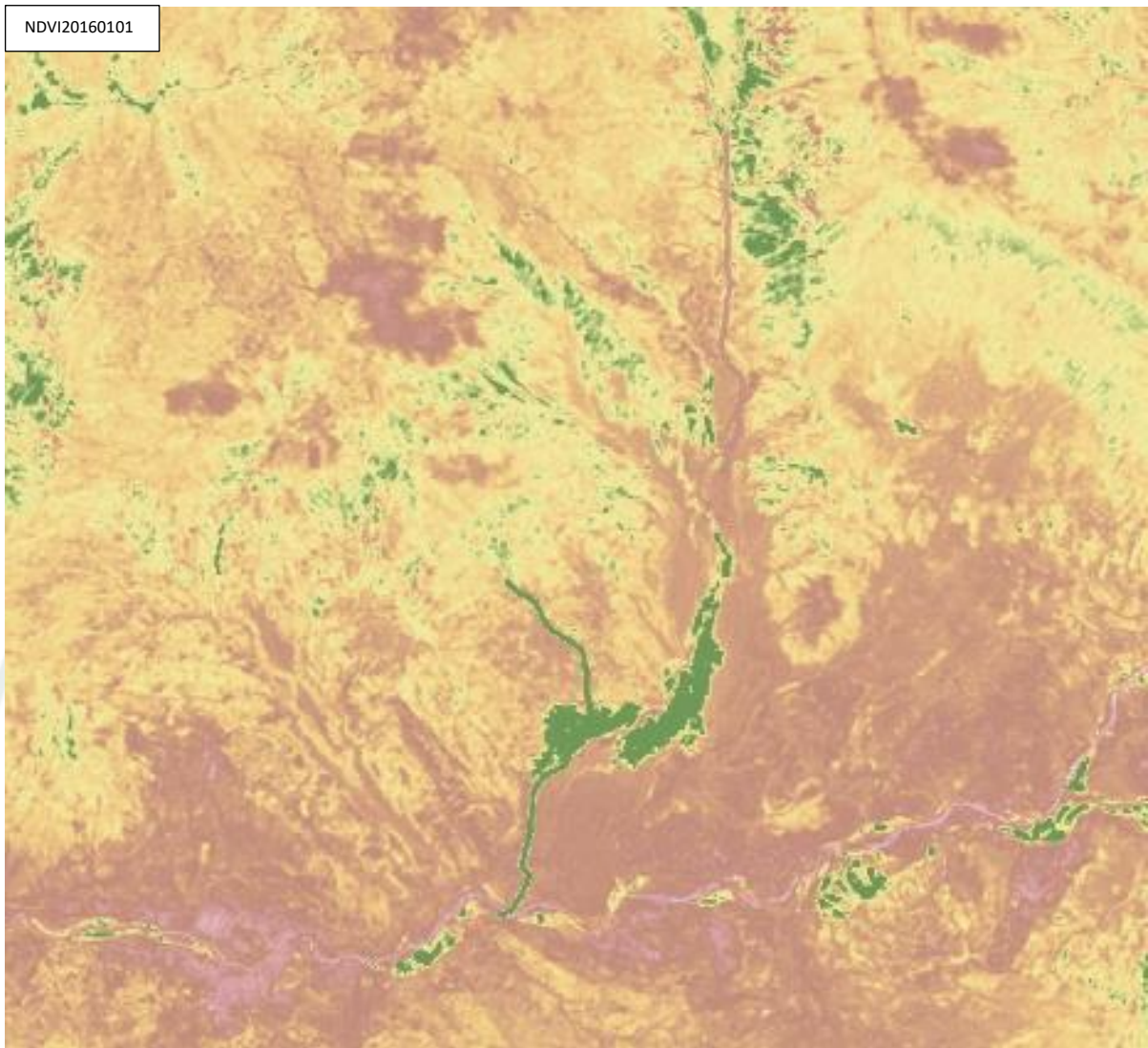
```
colores_volcan = mc_volcan['nivel_susceptibilidad'].map(color_dict)
plt.scatter(mc_volcan['Coordenada_X'], mc_volcan['Coordenada_Y'], c=colores_volcan, s=2,
alpha=0.6)
plt.title("Mapa de Susceptibilidad Multinivel - Quebrada Volcán")
plt.xlabel("Coordenada X (UTM)")
plt.ylabel("Coordenada Y (UTM)")
plt.grid(True)
plt.show()
```



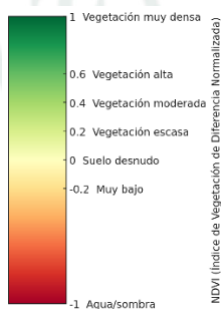
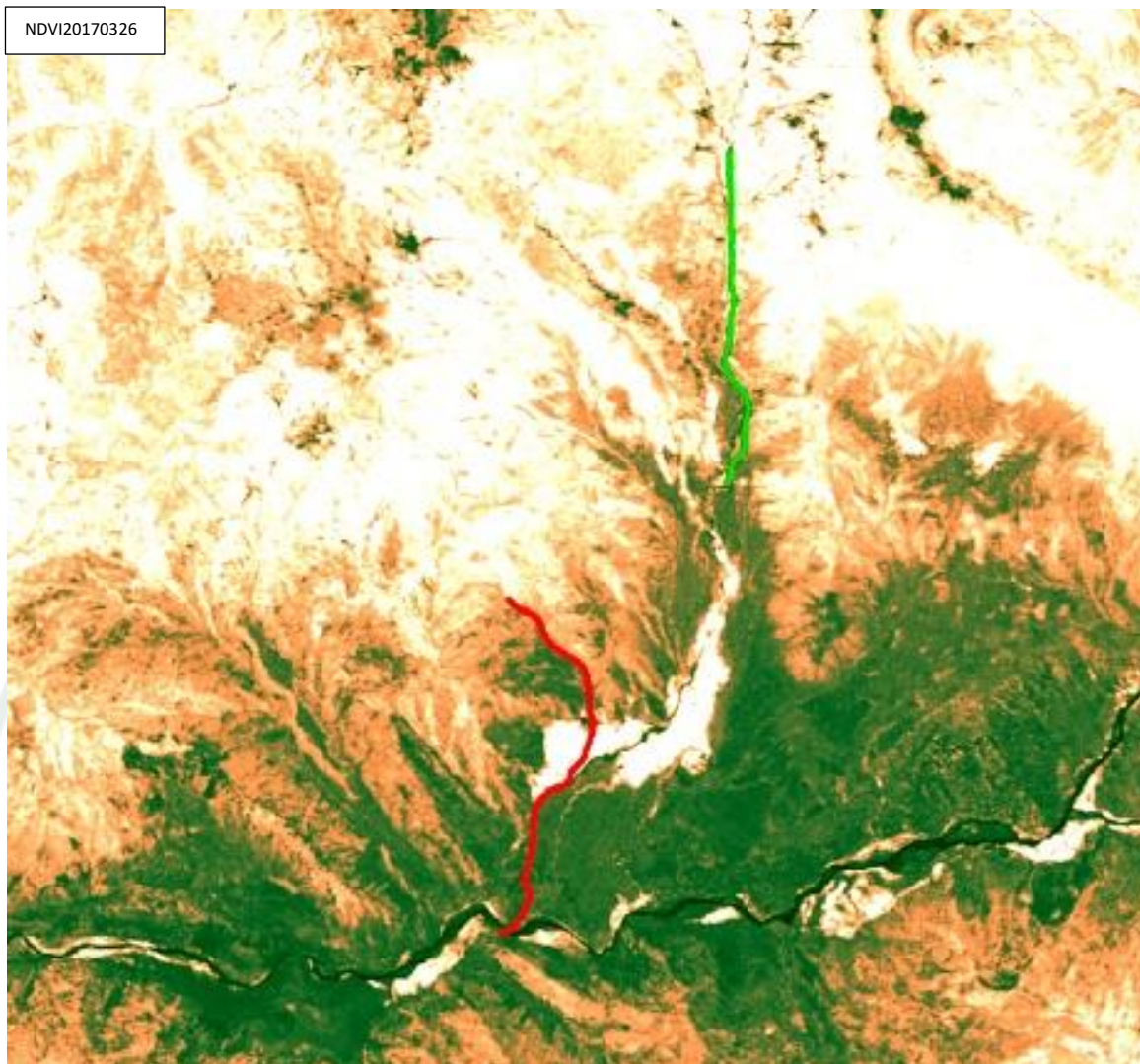
Anexo 2. Índices Clasificados



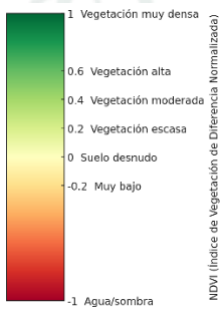
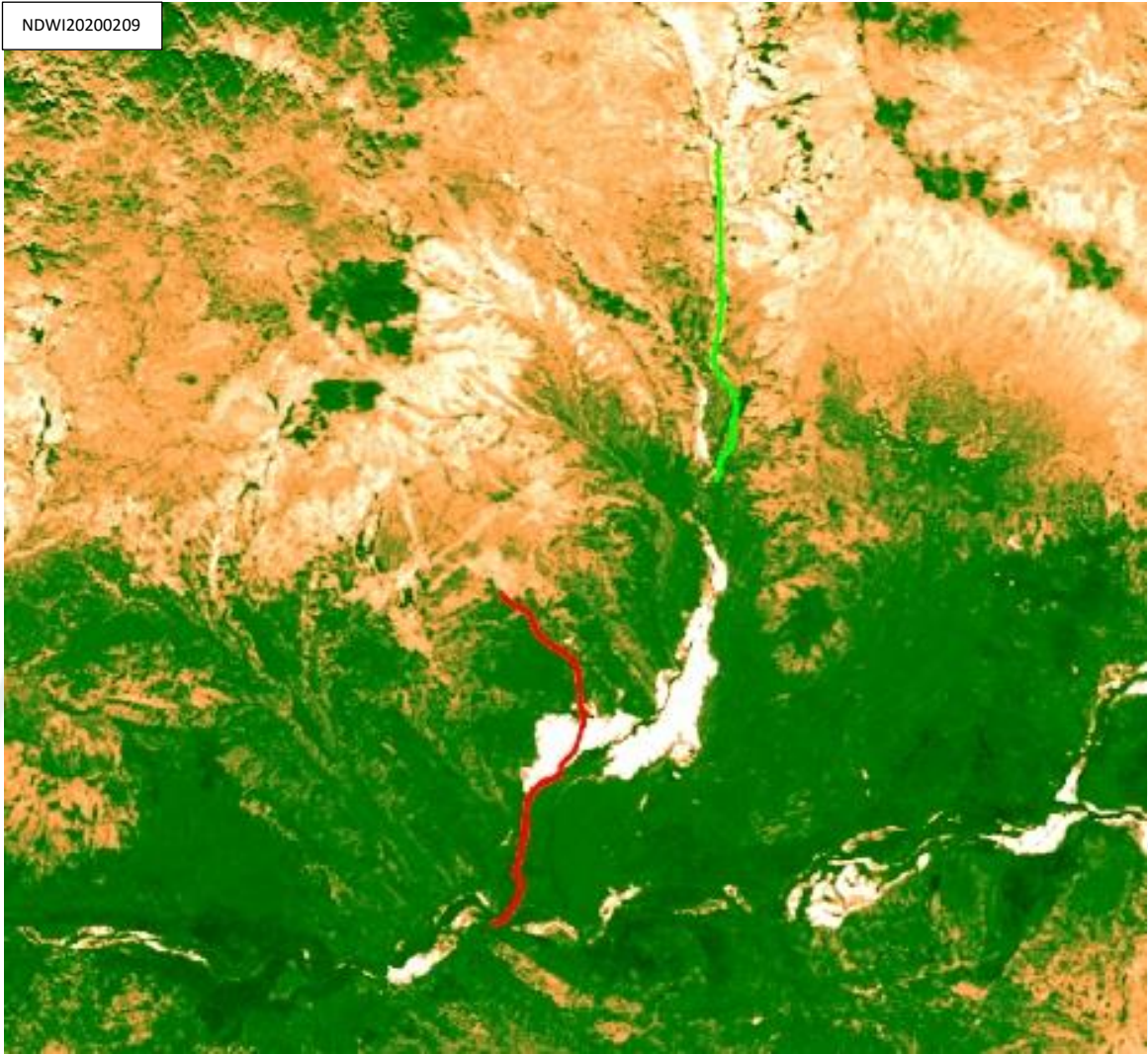
NDVI20160101

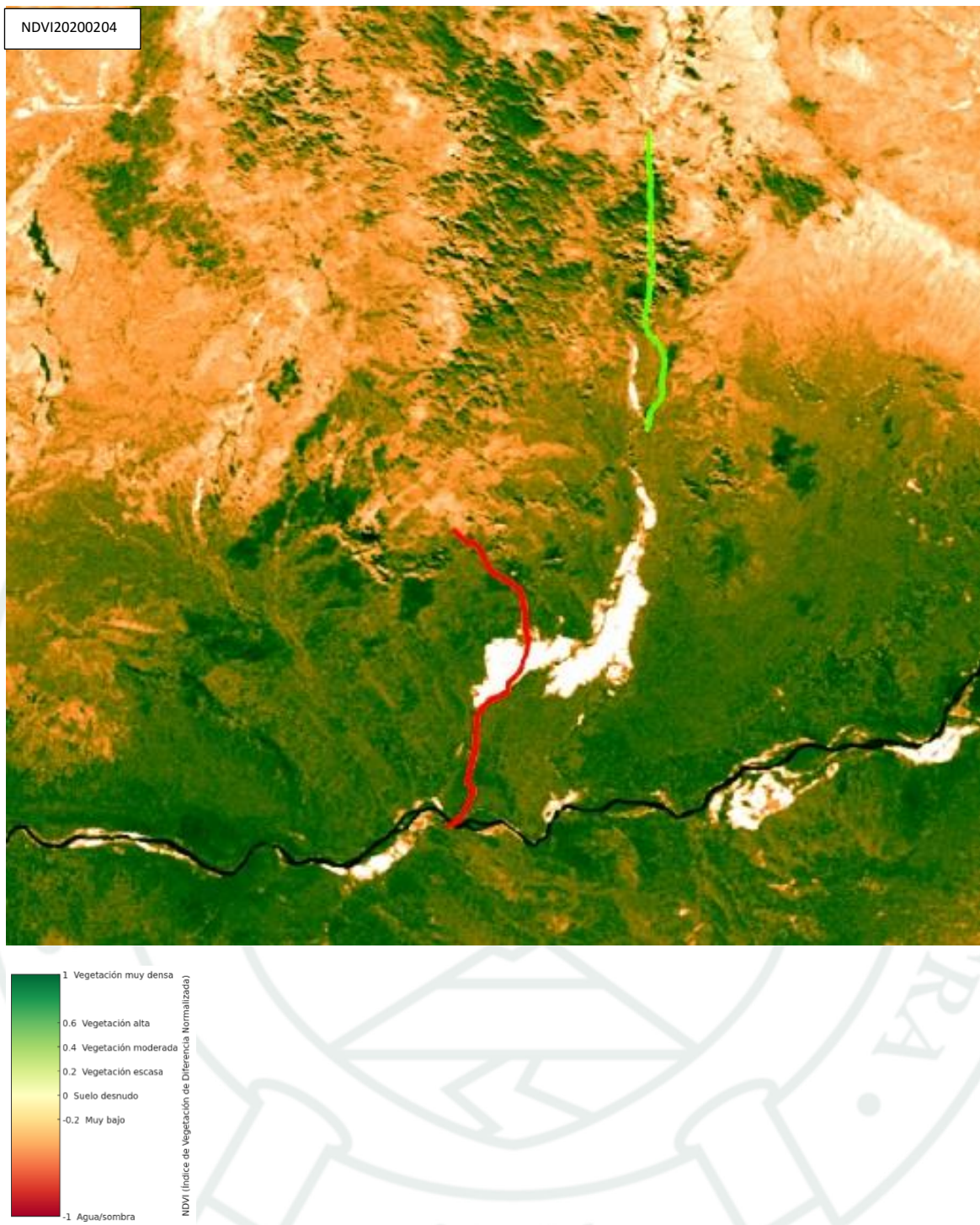


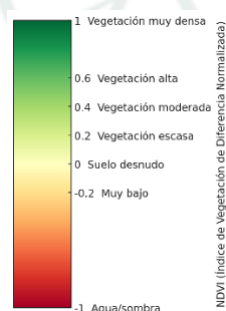
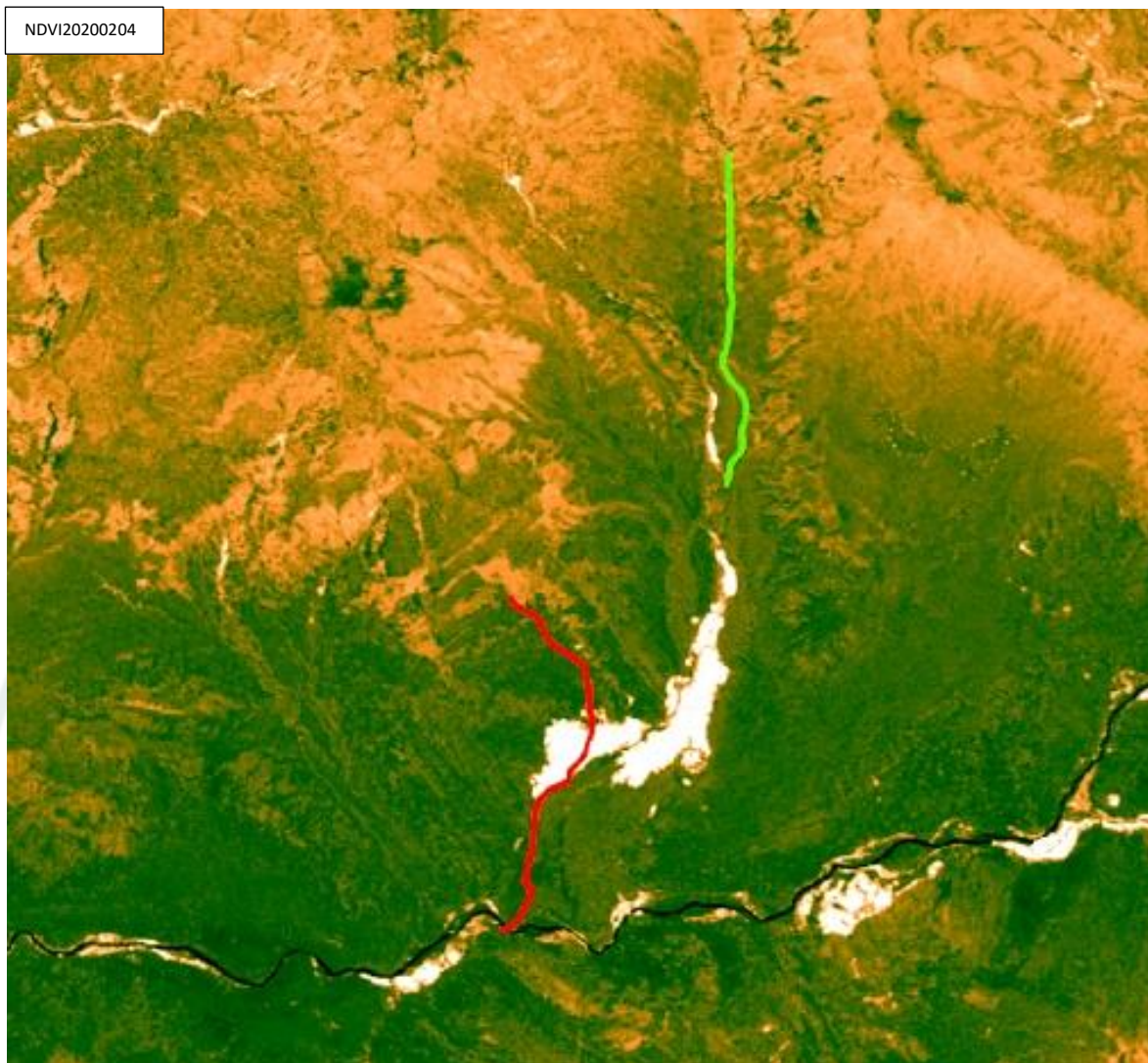
NDVI20170326

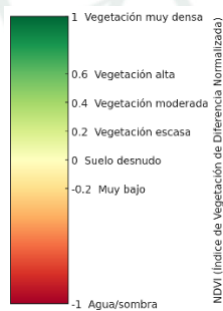
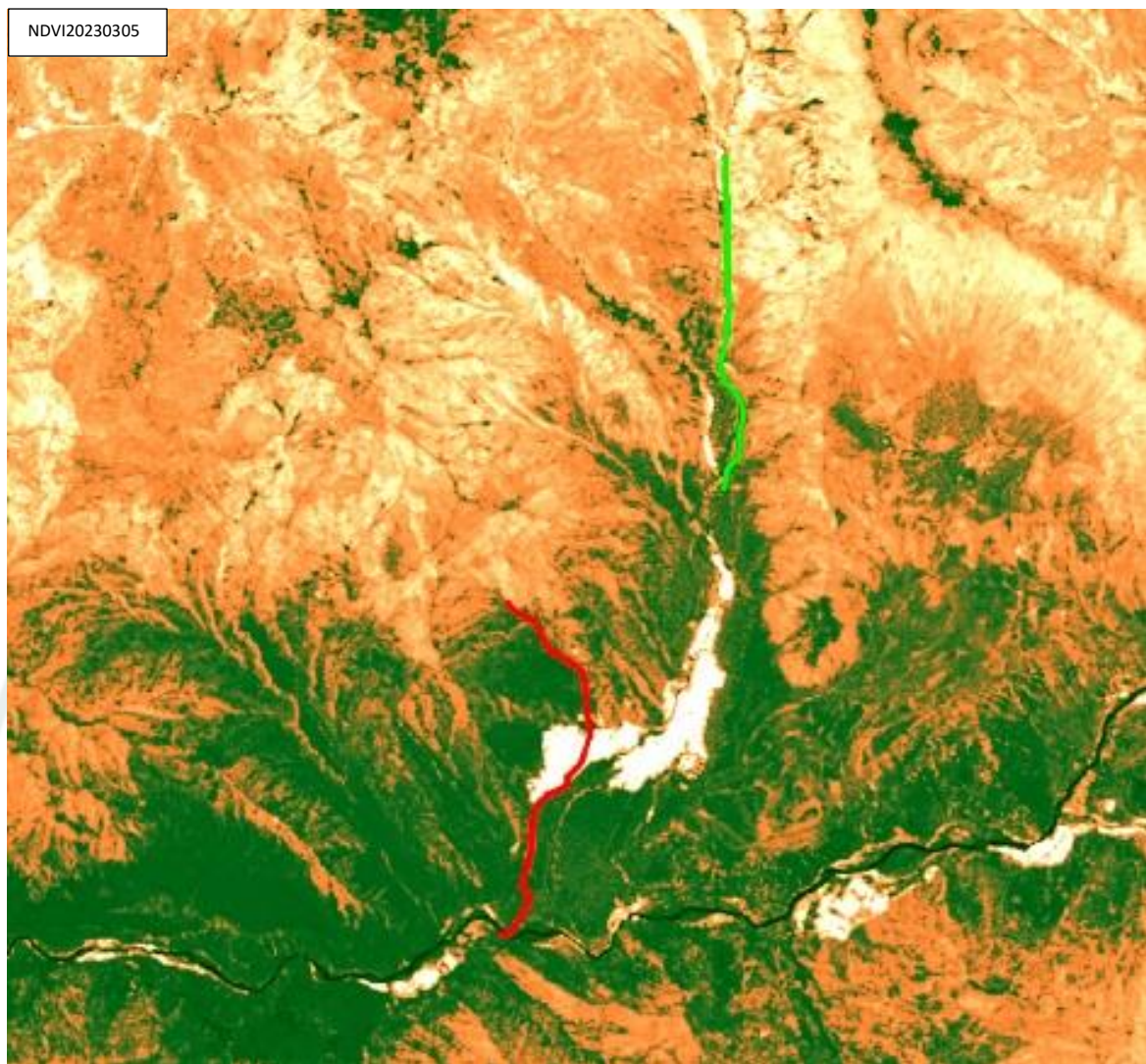


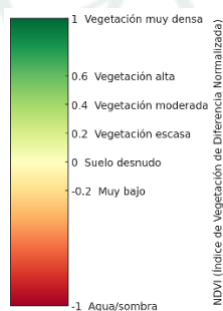
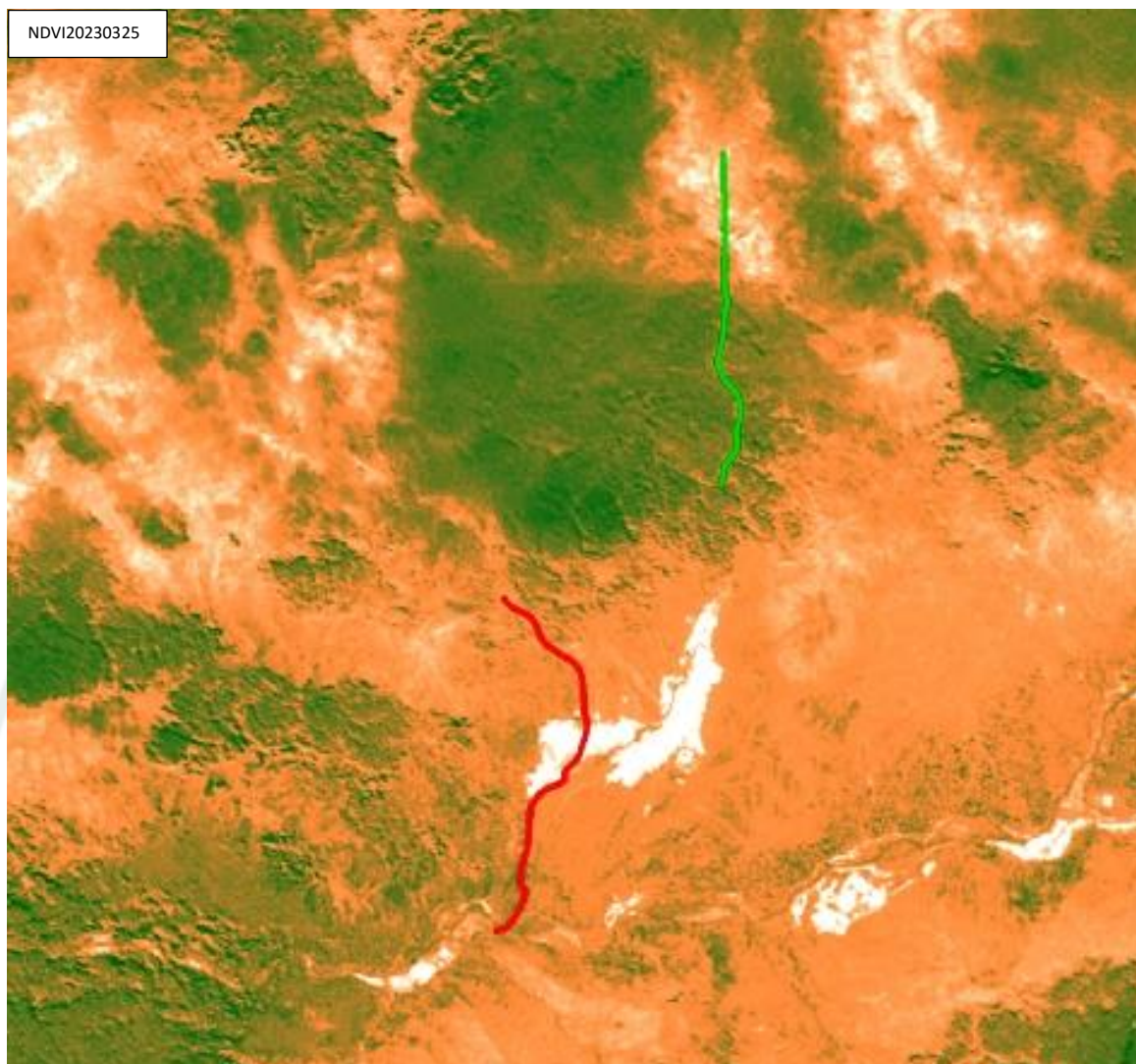
NDWI20200209

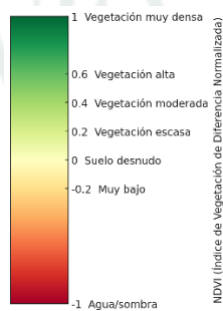
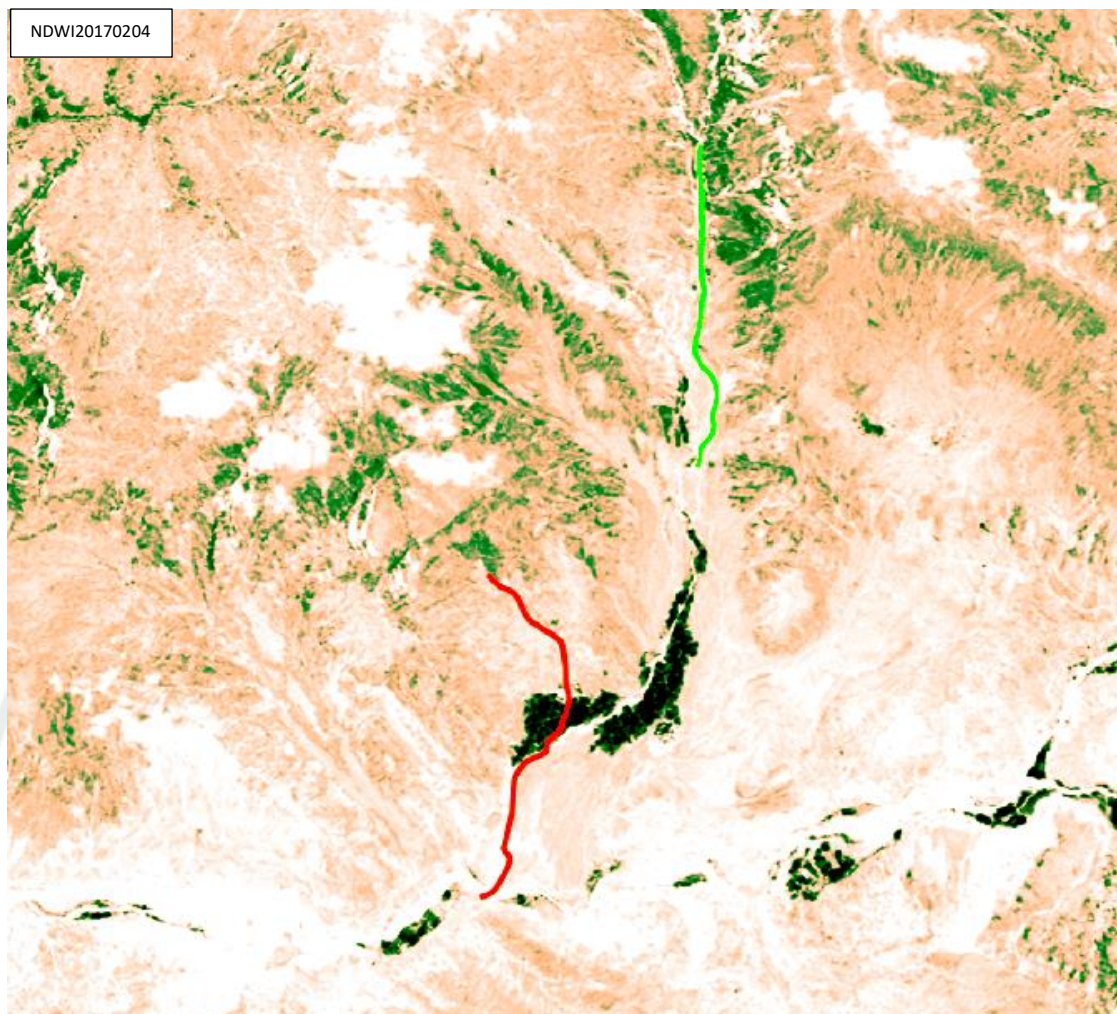


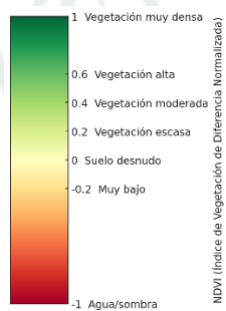
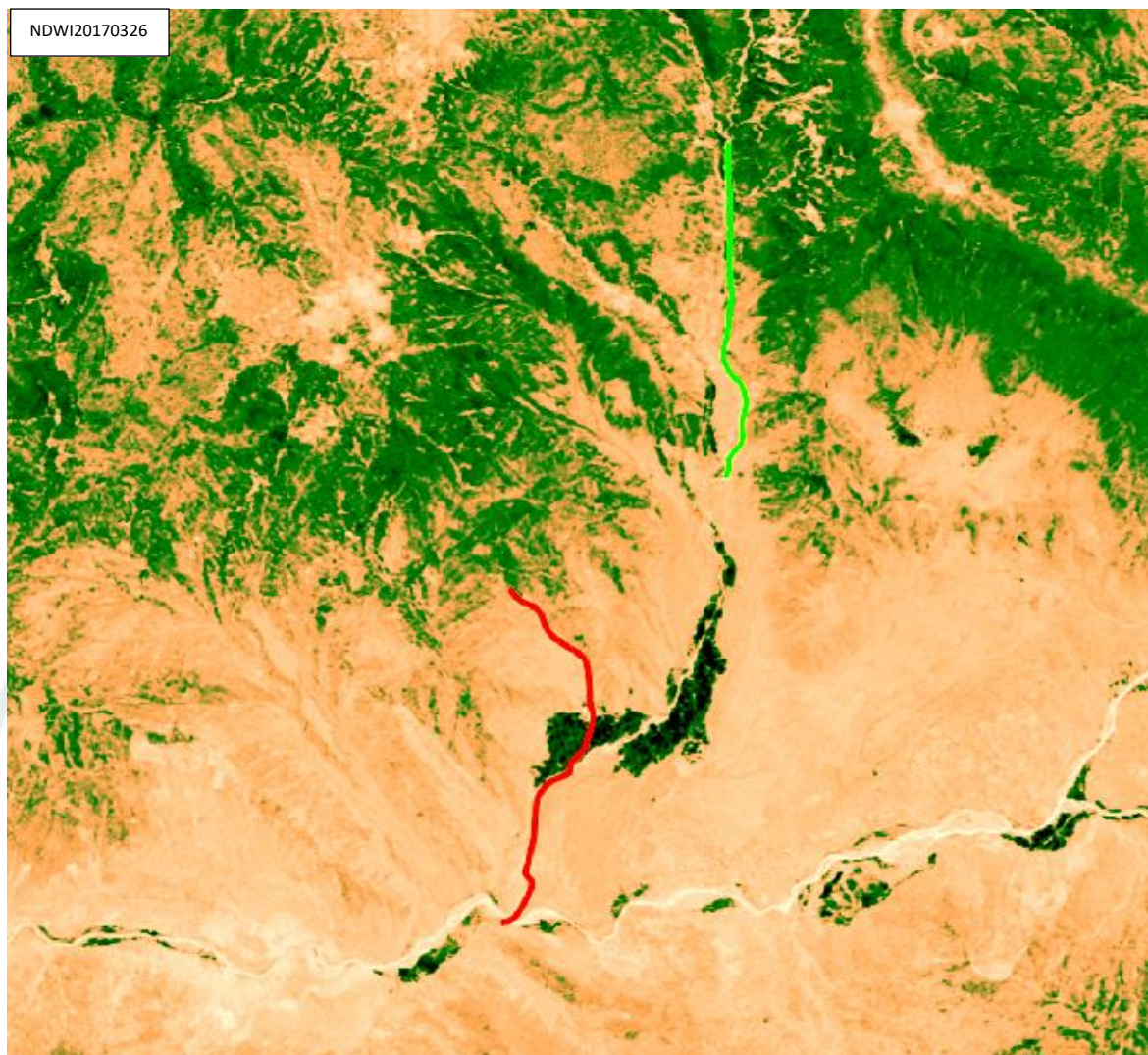




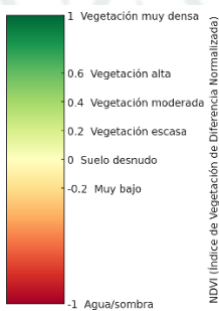
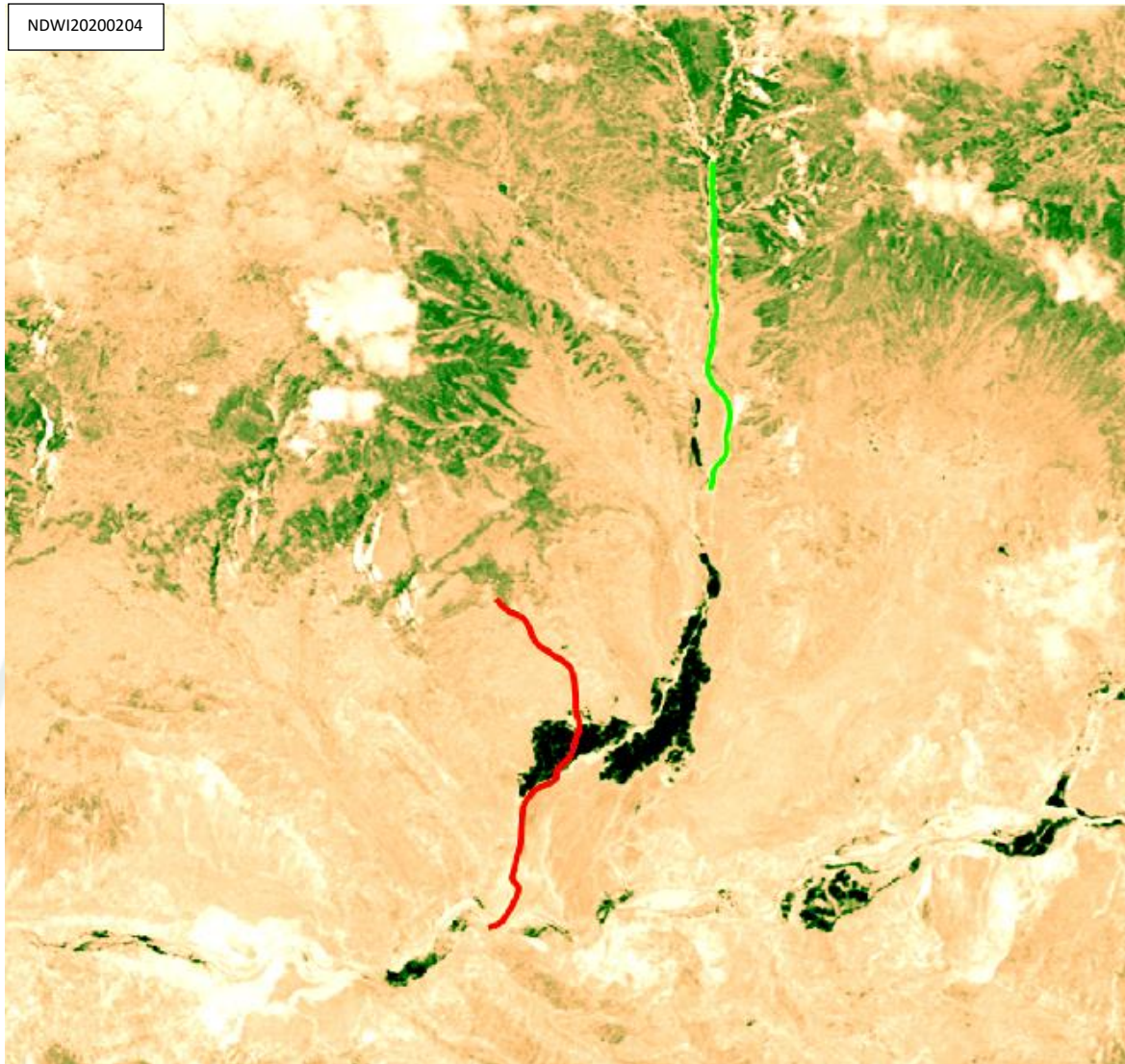


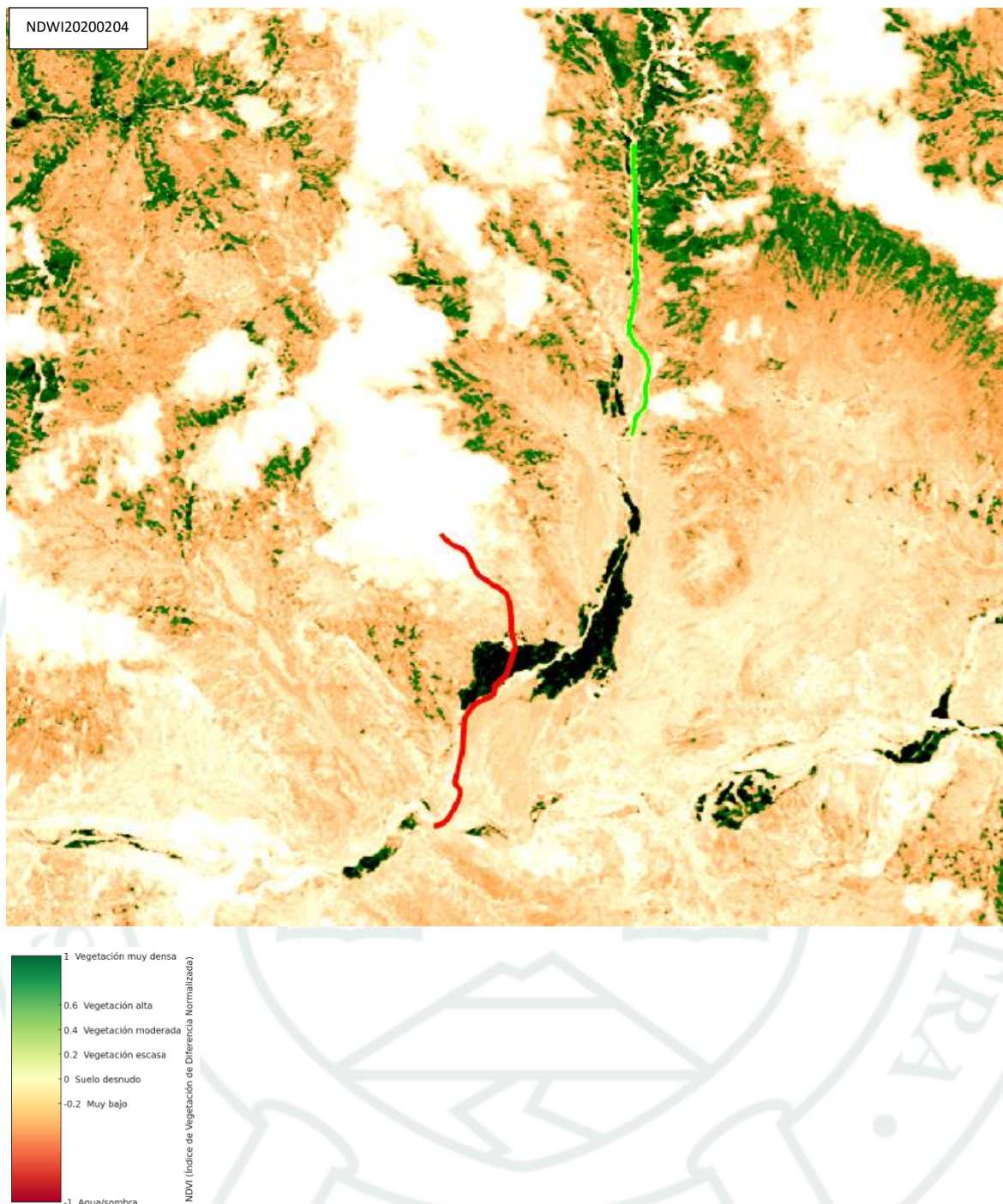




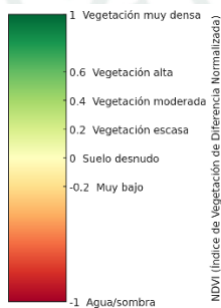
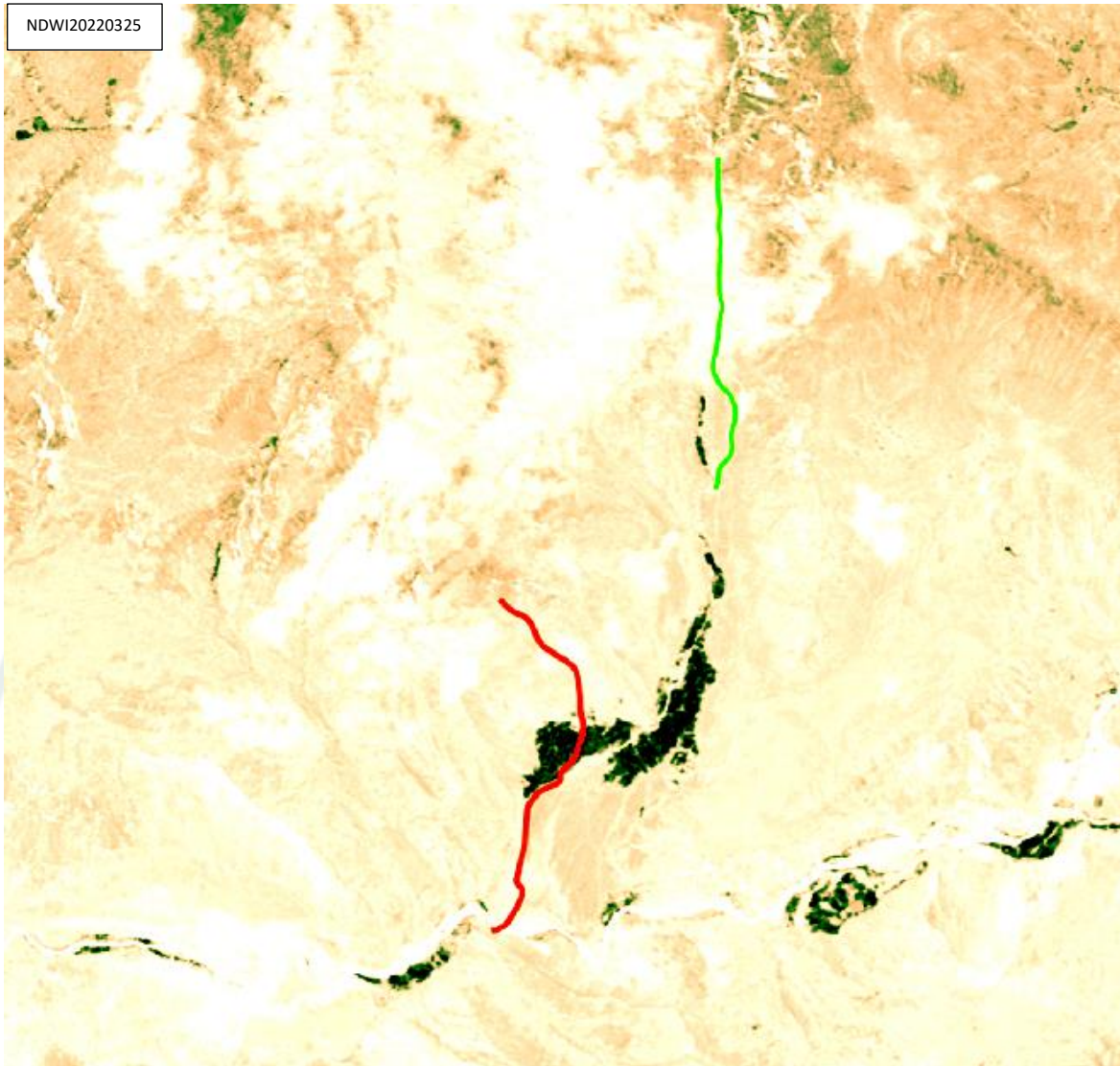


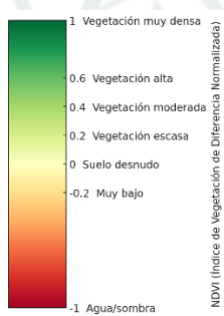
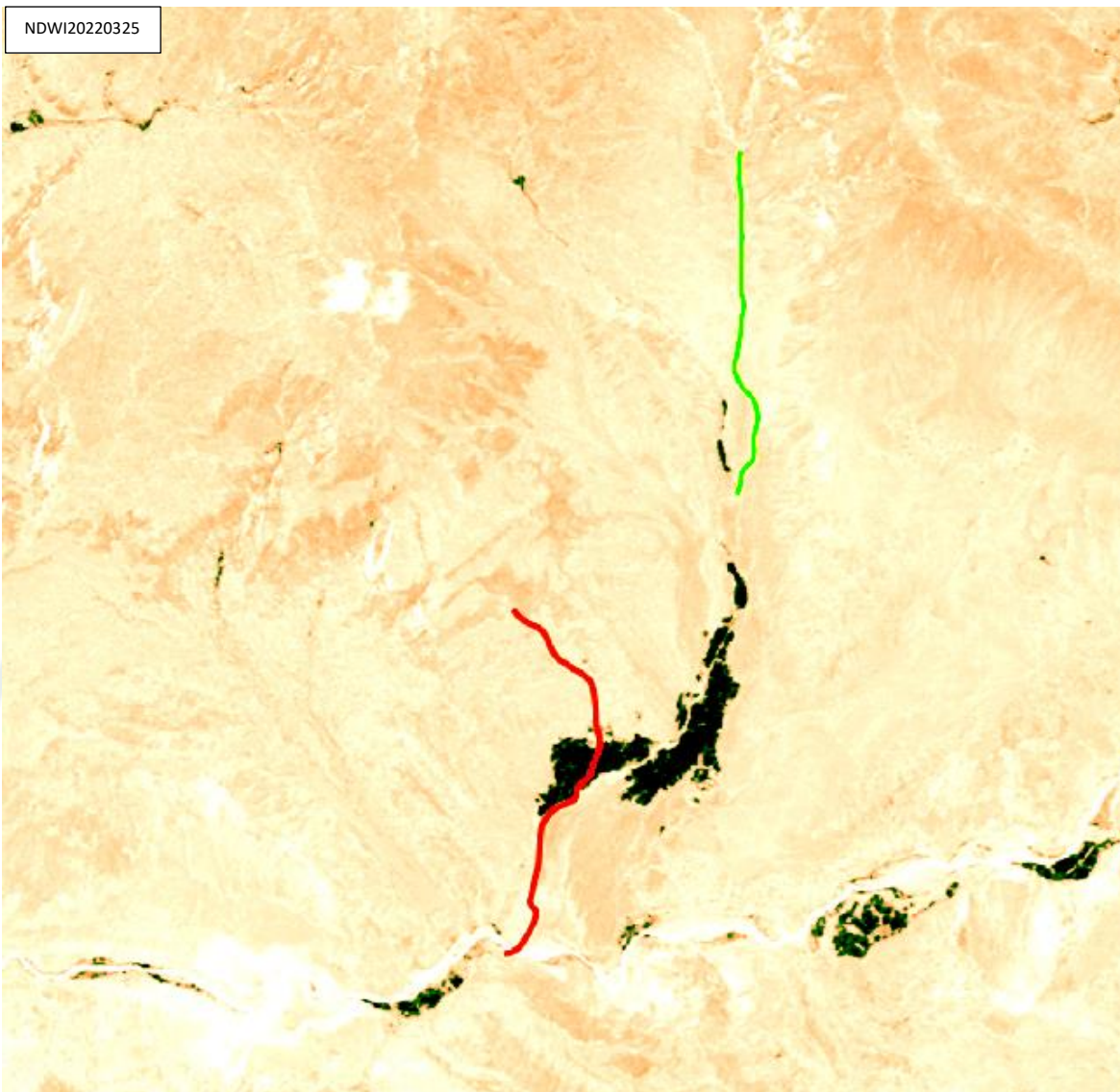
NDWI20200204

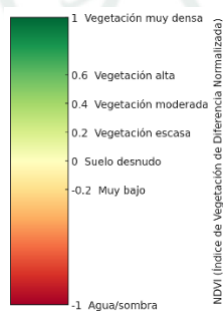
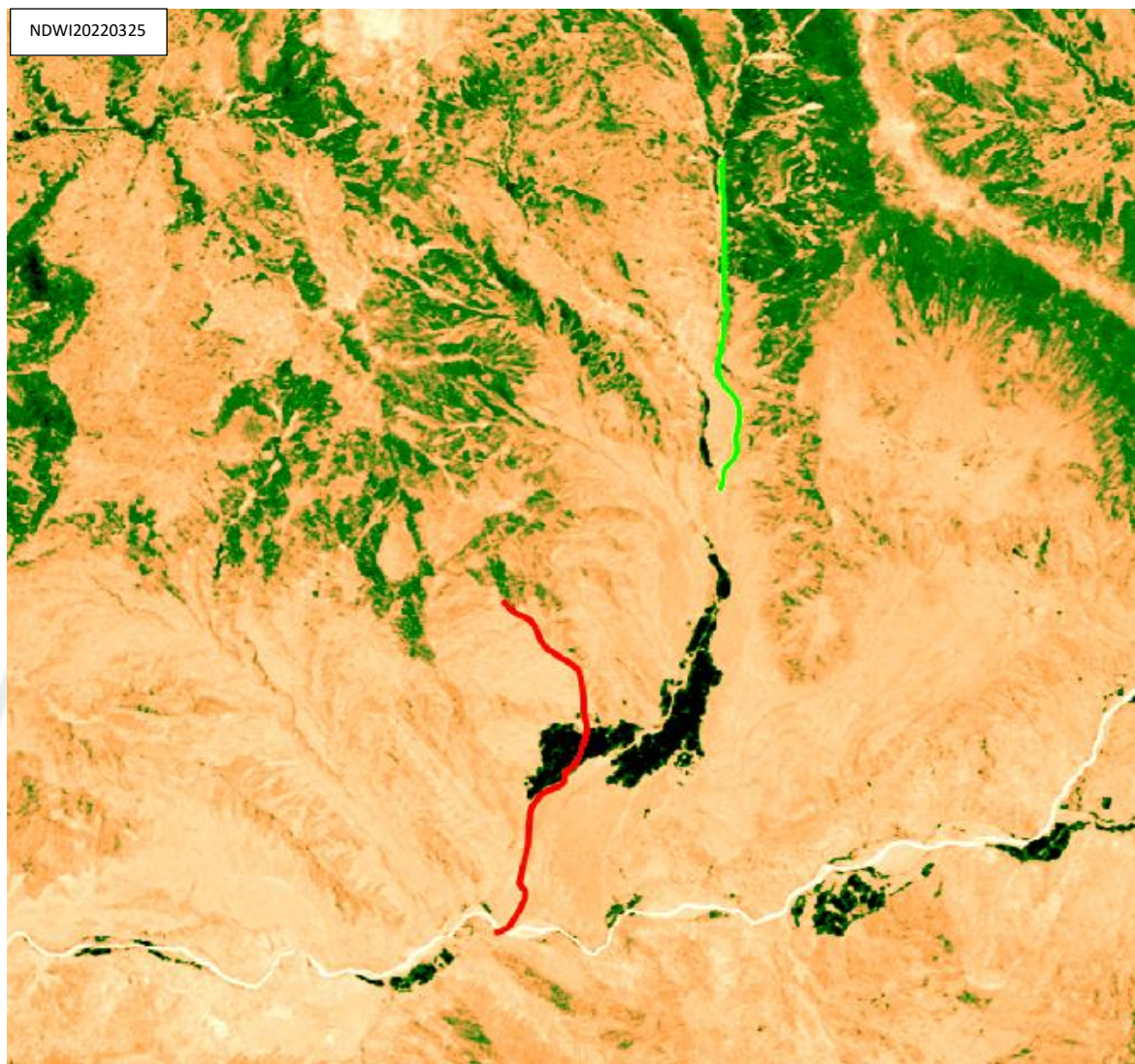


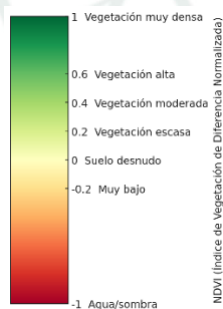
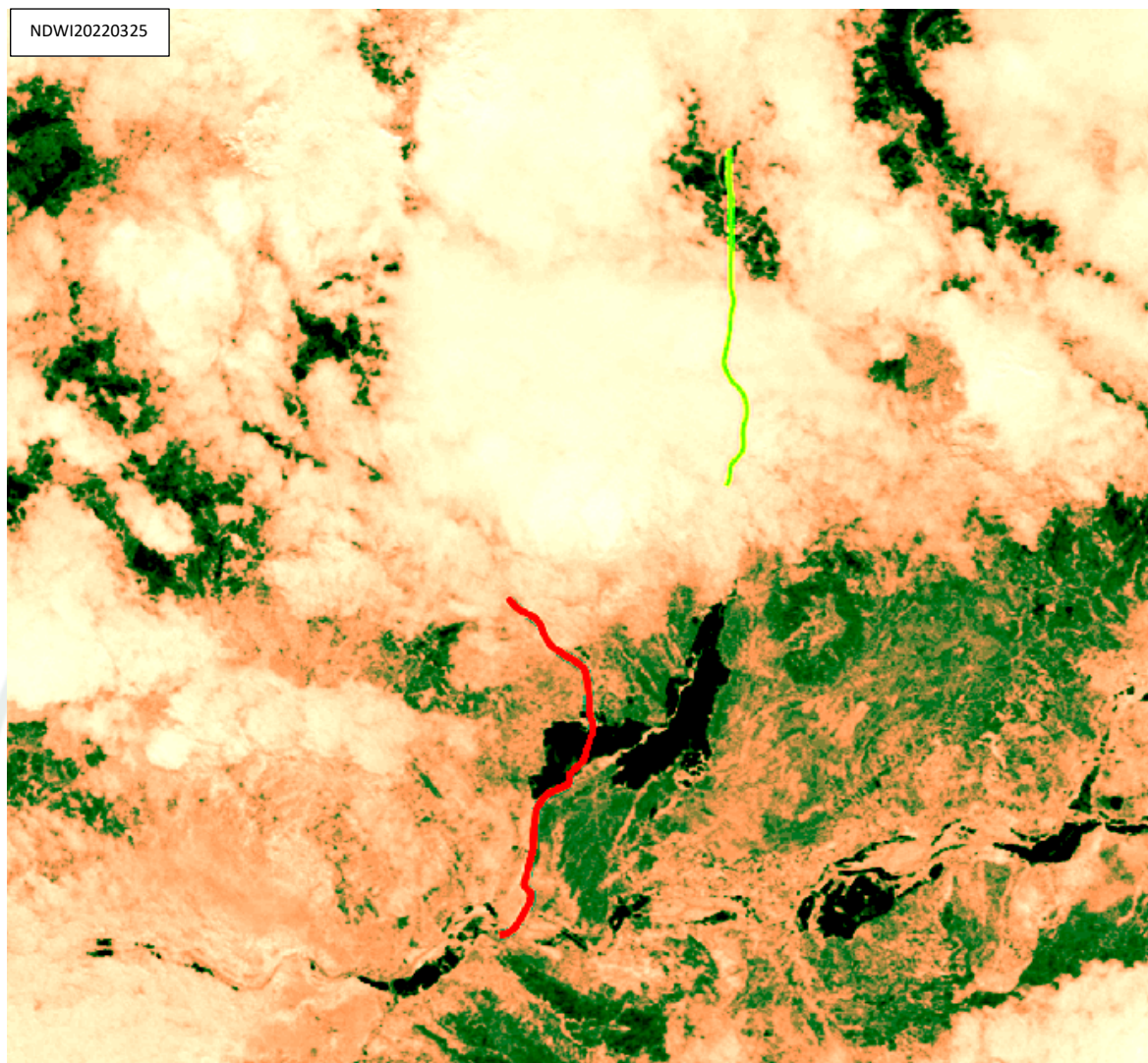


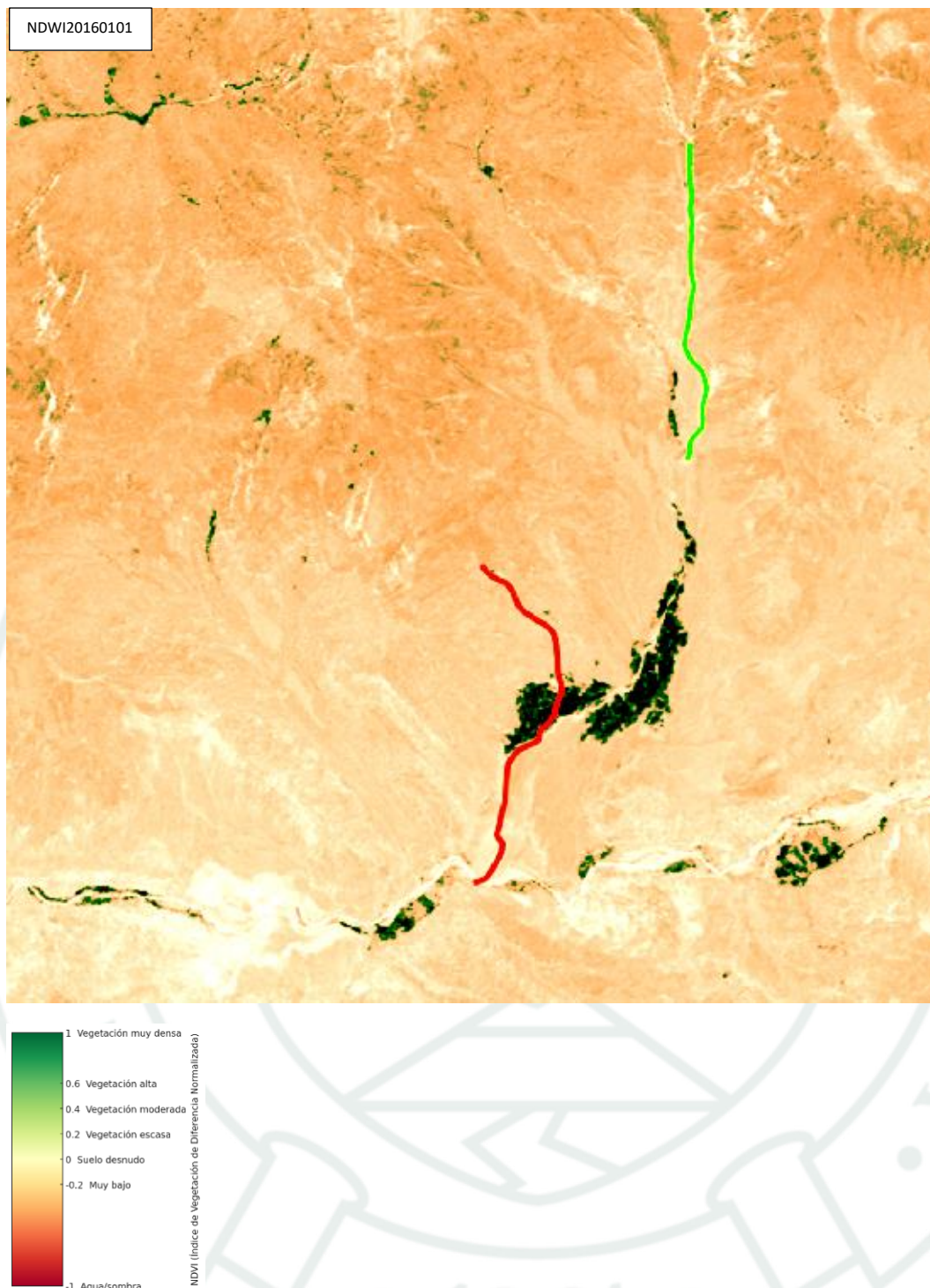
NDWI20220325

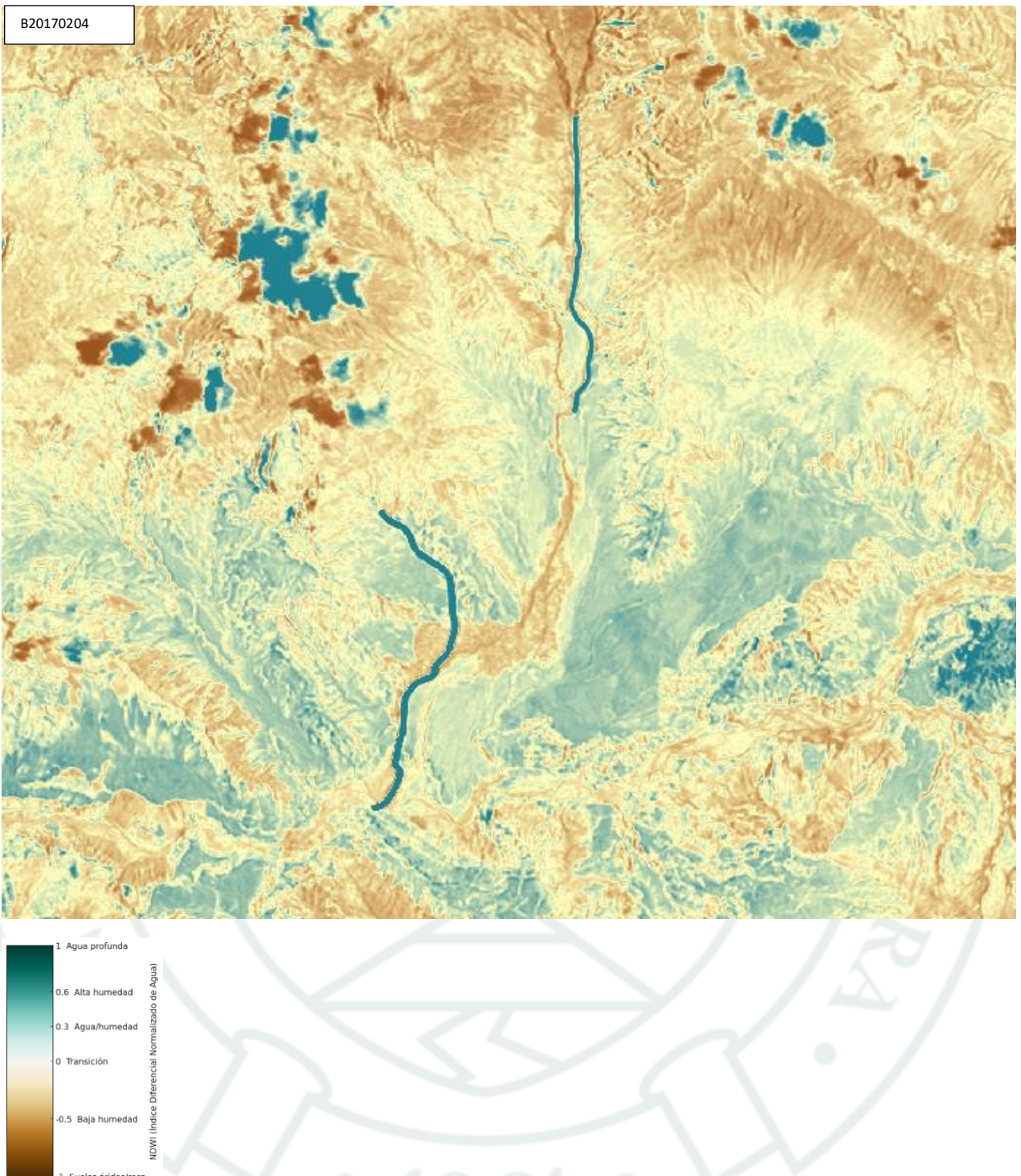


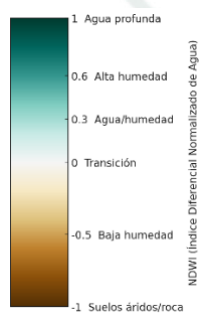
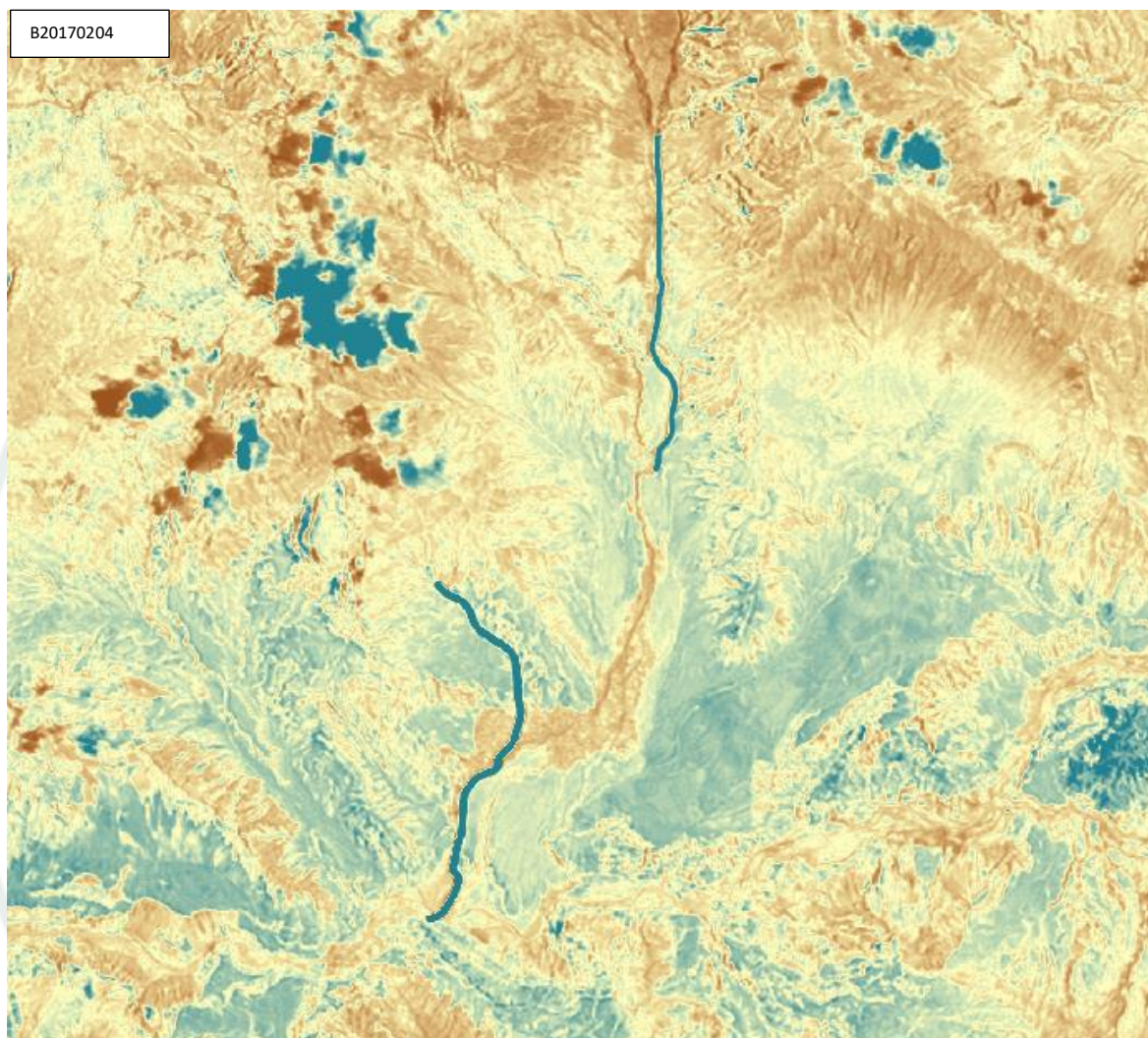












B20170204

