



ACUERDO NO. 1857 CON FECHA DEL 8 DE ABRIL DE 2015 DEL INSTITUTO DE EDUCACIÓN DEL ESTADO DE AGUASCALIENTES

" Generación de mapas geomorfológicos de forma automatizada para detectar unidades de relieve con técnicas de machine learning. "

TESIS PARA: **MAESTRIA EN CIENCIA DE LOS DATOS Y PROCESAMIENTO DE DATOS MASIVOS (BIG-DATA)**

PRESENTA: **RENÉ ANTONIO ARMENTA BORG**

DIRECTOR(A)
DE TESIS: **DR. CÉSAR ROBLEDO JIMÉNEZ**

DICIEMBRE, 2024

Contenido

1	CAPÍTULO I PLANTEAMIENTO DEL PROBLEMA	20
1.1	Formulación del problema	21
1.1.1	Contextualización.....	21
1.1.2	Definición del problema	27
1.1.3	Pregunta de Investigación	31
1.2	Justificación.....	31
1.2.1	Conveniencia	32
1.2.2	Relevancia social	33
1.2.3	Relevancia teórica	34
1.2.4	Utilidad metodológica.....	37
1.3	Viabilidad.....	38
1.4	Hipótesis	39
2	CAPÍTULO II MARCO TEÓRICO	42
2.1	Teoría educativa	43
2.2	Marco conceptual	44
2.2.1	Geomorfología	44
2.2.2	Big Data	46
2.2.3	Machine Learning	49
2.2.4	Machine Learning en Geomorfología:	51
2.2.5	Datos Geoespaciales en la Cartografía Automatizada	51
2.2.6	Continuo de Elevaciones Mexicano CEM.....	53
2.2.7	Derivados del CEM	55
2.2.8	Técnicas de Machine Learning en la investigación.....	60
2.3	Marco Referencial	71
2.3.1	Physically-Based Digital Geomorphological Mapping: Case Study of Glacial and Karst Topography.....	74
2.3.2	A Classification Method of Landform Elements Based on Douglas-Peucker Algorithm for Multi-scale Morphology.....	76

2.3.3	Aproximación a la evaluación de la exactitud temática de la cartografía geomorfológica	78
2.3.4	Algoritmos de Aprendizaje Automático para la Cartografía Litológica Automática Utilizando Datos de Teledetección: Un Estudio de Caso de Souk Arbaa Sahel, Inlier de Sidi Ifni, Anti-Atlas Occidental, Marruecos	80
2.3.5	A Geomorphological Regionalization using Upscaled DEM in the Beijing-Tianjin-Hebei Area, China	82
2.3.6	Geomorpho90m: An Empirical Evaluation of Global High-Resolution Geomorphometric Information.....	83
2.3.7	Landslide Susceptibility Zoning in Yunnan Province Based on SBAS-InSAR Technology and a Random Forest Model.	85
2.3.8	Classification of land use_land cover using artificial intelligence (ANN-RF). 87	
2.3.9	Land Cover Classification using Google Earth Engine and Random Forest Classifier—The Role of Image Composition"	88
2.3.10	Comparing geomorphological maps made manually and by deep learning. 90	
2.3.11	Automated classification of landforms with GIS support	92
2.3.12	Deep learning-based automated terrain classification using high-resolution digital elevation models	93
2.3.13	Machine Learning Algorithm in the Prediction of Geomorphic Indices for Appraisal the Influence of Landscape Structure on Fluvial Systems, Southeastern – Brazil 95	
2.3.14	Convolutional neural network for high-resolution wetland mapping with freely available data	96
2.3.15	Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques	99
2.4	Marco legal	101
3	CAPÍTULO III. MATERIALES Y MÉTODOS	105
3.1	Objetivos	106
3.1.1	Objetivo general	106
3.1.2	Objetivos específicos.....	106
3.2	Participantes y criterios de selección	107
3.2.1	Participantes (muestra geográfica).....	107

3.2.2	Criterios de Selección.....	108
3.3	Escenario.....	109
3.3.1	Contexto Geográfico	109
3.3.2	Contexto Institucional.....	110
3.4	Instrumentos de recolección de información.....	110
3.5	Procedimiento.....	111
3.5.1	Recolección de Datos	112
3.5.2	Preprocesamiento de Datos	113
3.6	Diseño del método	123
3.6.1	Diseño	123
3.6.2	Momento del estudio	124
3.6.3	Alcance del estudio	125
3.7	Operacionalización de las variables y/o categorías de estudio	126
3.7.1	Variables.....	126
3.7.2	Procedimiento de Operacionalización.....	128
3.8	Tipo de análisis empleado.....	129
3.8.1	Enfoque Cuantitativo	129
3.9	Análisis de datos	130
3.9.1	Técnicas e instrumentos	130
3.9.2	Técnicas	130
3.9.3	Instrumentos	131
3.9.4	Validez.....	133
3.9.5	Confiabilidad	135
3.9.6	Análisis de Datos.....	137
3.9.7	Procesamiento de la Información	143
3.10	Consideraciones Éticas	146
3.10.1	Confidencialidad de los Datos.....	146
3.10.2	Uso Responsable de Tecnología	146
3.10.3	Integridad Científica	146

4	Capítulo IV Resultados	148
4.1	Resultados generales del modelo	149
4.1.1	Prueba 1 (3 capas)	153
4.1.2	Prueba 2 (5 capas)	156
4.1.3	Prueba 3 normalizado (Ráster con 8 capas)	159
4.1.4	Prueba 4 sin normalizar (Ráster con 8 capas).....	162
4.1.5	Prueba 5 (agregando la geomediana)	165
4.1.6	Prueba 6 (agregando texturas)	167
5	CAPÍTULO V: DISCUSIÓN	174
5.1	Discusión del Objetivo General, Hipótesis y Supuestos Teóricos.....	175
5.2	Evaluación de la Investigación en relación con la Pregunta de Investigación	179
5.3	Contraste de Resultados con el Marco Teórico y Estudios Empíricos	181
5.4	Aplicabilidad de los resultados.....	187
5.4.1	Planificación y Ordenamiento Territorial	188
5.4.2	Gestión de Riesgos Naturales	188
5.4.3	Conservación y Manejo de Recursos Naturales	188
5.4.4	Investigación y Educación en Geomorfología.....	189
5.4.5	Infraestructura y Desarrollo de Proyectos	189
5.4.6	Desarrollo de Políticas Públicas.....	189
5.4.7	Innovación en Herramientas de Análisis Geoespacial	190
5.5	Análisis FODA	190
5.5.1	Fortalezas	190
5.5.2	Oportunidades	191
5.5.3	Debilidades.....	192
5.5.4	Amenazas	193
5.5.5	Conclusión del Análisis FODA	194
6	CAPITULO VI CONCLUSIONES	197
6.1	Conclusiones generales.....	198
6.2	Generación de nuevas líneas de investigación	201

6.2.1	Optimización de Algoritmos de Machine Learning para Geomorfología.....	201
6.2.2	Incorporación de Datos Multiespectrales y Multitemporales.....	201
6.2.3	Aplicación de Técnicas Automatizadas en Diferentes Contextos Geográficos 202	
6.2.4	Integración de Factores Climáticos y Ambientales	202
6.2.5	Desarrollo de Herramientas de Software y Plataformas de Uso Público	202
6.2.6	Evaluación de Impactos y Aplicaciones en Políticas Públicas	203
6.3	Aportaciones para la sociedad y el campo del conocimiento estudiado.	203
6.3.1	Aportaciones para la Sociedad	203
6.3.2	Aportaciones para el Campo del Conocimiento	204
	Referencias Bibliográficas.....	207
	APÉNDICES	214
	Apéndice 1. Código de normalización de capas.....	215
	Apéndice 2. Segmentaciones	216

ÍNDICE DE TABLAS

Tabla 1.....	53
Tabla 2.....	72
Tabla 3.....	152
Tabla 4.....	156
Tabla 5.....	159
Tabla 6.....	161
Tabla 7.....	164
Tabla 8.....	166
Tabla 9.....	169

ÍNDICE DE FIGURAS

Figura 1	30
Figura 2	45
Figura 3	55
Figura 4	61
Figura 5	62
Figura 6	114
Figura 7	115
Figura 8	117
Figura 9	118
Figura 10	120
Figura 11.....	143
Figura 12	163
Figura 13	170
Figura 14	171
Figura 15	172

AGRADECIMIENTOS

En primer lugar, quiero expresar mi más profundo agradecimiento a mi asesor, César Robledo por su valiosa orientación, apoyo incondicional y paciencia a lo largo de todo este proceso. Sus conocimientos y experiencia fueron fundamentales para la realización de este trabajo.

Agradezco especialmente a Mtro. Abel Alejandro Coronado Iruegas, quien me brindó un apoyo invaluable en el desarrollo de los algoritmos de machine learning. Su generosidad al compartir su conocimiento y su orientación técnica me permitieron enfocar mejor mi trabajo y superar varios de los retos más complejos de este proyecto. Su ayuda ha sido esencial para lograr los resultados presentados en esta tesis.

Agradezco de manera especial al Instituto Nacional de Estadística y Geografía (INEGI) y a mis compañeros del Instituto por brindarme la oportunidad de aplicar los conocimientos adquiridos y por el apoyo en la gestión de los datos y herramientas clave para este proyecto. Su colaboración fue esencial en la elaboración de los análisis y mapas geomorfológicos presentados.

A mi familia, gracias por su amor, paciencia y constante motivación, que me permitieron mantenerme enfocado y motivado en los momentos difíciles. Su apoyo ha sido el pilar más importante en mi vida.

Finalmente, agradezco a mis amigos y colegas que, de una u otra manera, me acompañaron en este recorrido académico.

Este logro no habría sido posible sin el apoyo de todas estas personas y organizaciones, por lo que mi agradecimiento es sincero y profundo.

DEDICATORIA

A mis padres, quienes siempre me han apoyado en cada paso de mi vida y me enseñaron el valor del esfuerzo, la disciplina y la perseverancia.

A mis amigos y colegas, por su constante ánimo y por estar a mi lado en los momentos difíciles. Su compañía ha sido un gran apoyo durante todo este proceso.

Y, finalmente, a todos aquellos que de una u otra manera contribuyeron a este proyecto, con sus conocimientos, consejos y apoyo, les dedico este trabajo con gratitud.

RESUMEN

Este estudio aborda el uso de algoritmos de machine learning para la creación automatizada de mapas geomorfológicos, con el propósito de mejorar la precisión y eficiencia en la identificación de unidades de relieve. El objetivo general fue determinar si los algoritmos de aprendizaje automático pueden generar mapas geomorfológicos de mejor calidad que los elaborados manualmente, tanto en términos de exactitud como de rapidez.

Para ello, se llevó a cabo una exhaustiva fase de preparación de datos, que incluyó la selección de imágenes satelitales, modelos digitales de elevación (MDE) y otros datos geoespaciales relevantes. Estos datos fueron sometidos a técnicas de preprocesamiento y limpieza para garantizar su idoneidad en el entrenamiento de los algoritmos. Se implementó el algoritmo Extra Trees con un enfoque de validación cruzada para medir el rendimiento de cada prueba.

Los resultados mostraron una precisión de 64% con 103209 muestras y 64 capas de la prueba 6, siendo la del mejor resultado, por lo que el modelo para la generación de mapas geomorfológicos con machine learning Extra Trees es capaz de detectar y clasificar con precisión las diferentes unidades geomorfológicas. Los árboles de decisión destacaron por su capacidad de generalización y precisión en áreas de alta complejidad geográfica. Sin embargo, se identificaron algunas limitaciones, especialmente en cuanto a la complejidad computacional y la necesidad de una segmentación adecuada de los datos de entrada para mejorar el rendimiento en áreas con menor representación.

En conclusión, el modelo para la generación de mapas geomorfológicos con machine learning ofrecen una alternativa viable y eficiente para la cartografía geomorfológica, donde se observa una mejora en los resultados al incluir más información (capas con información geográfica del terreno) y con la ventaja de poder procesar grandes volúmenes de datos en

menos tiempo que los métodos tradicionales. Además, su capacidad para generar mapas más precisos y detallados puede ser de gran valor para instituciones como el INEGI, permitiendo una mejora significativa en los procesos de planificación y gestión territorial mediante la integración de tecnología de vanguardia en la producción de información geoespacial.

Palabras clave: Machine Learning, Mapas geomorfológicos, mapas automatizados, RSGISLIB, Big Data, Arboles de decisión, Extra Trees.

ABSTRACT

This study addresses the use of machine learning algorithms for the automated creation of geomorphological maps, with the aim of improving accuracy and efficiency in the identification of landform units. The main objective was to determine whether machine learning algorithms can generate geomorphological maps of higher quality than those produced manually, both in terms of accuracy and speed.

To achieve this, a thorough data preparation phase was carried out, which included the selection of satellite images, digital elevation models (DEMs), and other relevant geospatial data. These data were subjected to preprocessing and cleaning techniques to ensure their suitability for training the algorithms. The Extra Trees algorithm was implemented, a cross-validation approach to measure the performance of each test.

The results showed an accuracy of 64% with 103,209 samples and 64 layers from test 6, which yielded the best outcome. This demonstrates that the Extra Trees machine learning model for geomorphological map generation can accurately detect and classify different geomorphological units. Decision trees stood out for their generalization capability and accuracy in areas of high geographic complexity. However, some limitations were identified, particularly regarding computational complexity and the need for proper segmentation of input data to improve performance in less-represented areas.

In conclusion, the machine learning model for geomorphological map generation offers a viable and efficient alternative for geomorphological mapping, showing improved results when more information is included (layers with geographic terrain data) and with the advantage of processing large volumes of data in less time than traditional methods. Additionally, its ability to generate more precise and detailed maps can be of great value to

institutions like INEGI, enabling significant improvements in territorial planning and management processes through the integration of cutting-edge technology in geospatial information production.

Keywords: Machine Learning, Geomorphological Maps, Automated Maps, RSGISLIB, Big Data, Decision Trees, Extra Trees.

INTRODUCCIÓN

La geomorfología es la ciencia que estudia las formas y los procesos que las originan y modifican en la superficie terrestre. Los mapas geomorfológicos son representaciones gráficas de las unidades de relieve y sus características, que permiten comprender la evolución del paisaje y sus implicaciones ambientales, sociales y económicas (IGME, s.f.). Sin embargo, la elaboración de mapas geomorfológicos mediante métodos tradicionales presenta importantes limitaciones. La construcción de estos mapas requiere de una gran cantidad de datos geoespaciales, observaciones en campo y el uso de criterios que, en muchos casos, dependen de la interpretación subjetiva del experto. Este proceso no solo es manual y costoso en términos de tiempo y recursos, sino que también puede conllevar inconsistencias debido a la variabilidad en la experiencia y enfoques de los analistas.

Además, en la actualidad, la demanda por información precisa y actualizada crece constantemente, los métodos manuales de cartografía no logran satisfacer de manera eficiente las necesidades de escalabilidad ni la rapidez en la toma de decisiones que requieren diversas aplicaciones, como la mitigación de riesgos ante desastres naturales o la gestión de territorios vulnerables. Por esta razón, se ha planteado la necesidad de desarrollar métodos automatizados para generar mapas geomorfológicos a partir de fuentes de información digital, como imágenes satelitales, modelos digitales de elevación o datos geoespaciales. Una de las técnicas más prometedoras para este fin es el machine learning, que es una rama de la inteligencia artificial que capacita a las máquinas para aprender patrones y realizar tareas sin programación explícita, permitiendo a los sistemas mejorar automáticamente con la experiencia (Murphy, 2012).

Debido la creciente necesidad de generar mapas con mayor precisión, rapidez y objetividad, surge la pregunta general de esta investigación: ¿Es posible crear mapas geomorfológicos

de forma automatizada mediante algoritmos de machine learning que permitan la detección de unidades de relieve y sus atributos?

Las motivaciones de la investigación radican en el potencial para revolucionar la forma en que se generan los mapas geomorfológicos, brindando una solución a los desafíos actuales de la disciplina. En un contexto donde los datos geoespaciales están disponibles en grandes volúmenes y donde la demanda de análisis rápidos y precisos es cada vez mayor, la implementación de técnicas de machine learning representa una oportunidad para mejorar la eficiencia y la calidad de los mapas geomorfológicos. Además, estos avances podrían facilitar la toma de decisiones en áreas críticas como la gestión de riesgos naturales, la planificación urbana y la conservación de ecosistemas, donde los mapas precisos y actualizados son fundamentales.

El presente estudio se sustenta en la Teoría del Constructivismo, donde se aplica este enfoque al proceso de creación automatizada de mapas geomorfológicos mediante algoritmos de machine learning. Al igual que en el constructivismo, donde el aprendiz construye conocimientos con base en la experiencia, los algoritmos desarrollan su capacidad de aprender patrones y características del terreno a partir de grandes volúmenes de datos geoespaciales lo que enmarca en la intersección de la geomorfología y la inteligencia artificial.

El método propuesto en este estudio incluye la revisión y selección de datos geoespaciales como imágenes satelitales y modelos digitales de elevación, los cuales se procesan mediante algoritmos de machine learning para la detección y clasificación de unidades de relieve. La investigación emplea técnicas de árboles de decisión, que han mostrado ser eficaces en la clasificación de datos espaciales. A través de un proceso de entrenamiento y validación, los modelos generados son evaluados y comparados con mapas

geomorfológicos realizados por expertos, permitiendo analizar su precisión y aplicabilidad en diferentes escenarios.

La metodología que se empleará en esta investigación es cuantitativa. Por un lado, se utilizarán técnicas de análisis espacial y estadístico para procesar y clasificar los datos de entrada, así como para evaluar el rendimiento y la calidad de los algoritmos de machine learning. Con este estudio, se espera no solo contribuir al avance de la geomorfología como disciplina, sino también sentar las bases para la integración de técnicas de inteligencia artificial en la cartografía geoespacial, favoreciendo la creación de herramientas más eficientes y precisas que puedan ser aplicadas en contextos diversos.

La presente investigación cuenta de 6 capítulos, donde el Capítulo 1 establece el contexto de la investigación, describiendo la problemática de la elaboración manual de mapas geomorfológicos y la propuesta de utilizar algoritmos de machine learning para automatizar este proceso. Se explican las motivaciones de la investigación, la relevancia del tema en el ámbito geoespacial, y se presentan los objetivos generales y específicos. La inclusión de este capítulo es esencial para situar al lector en el marco teórico y práctico del estudio, definiendo las bases que guiarán la investigación.

El Capítulo 2 proporciona una revisión exhaustiva de los conceptos y estudios previos sobre geomorfología y machine learning. Se analizan los métodos tradicionales de elaboración de mapas geomorfológicos y las nuevas técnicas de aprendizaje automático que han surgido en los últimos años. La inclusión de este capítulo es crucial para sustentar la base teórica de la investigación, mostrando el estado del arte y justificando la relevancia del enfoque propuesto.

El Capítulo 3 describe los datos geoespaciales utilizados en la investigación, como imágenes satelitales y modelos digitales de elevación (MDE). Se detallan los procesos de selección, preprocesamiento y segmentación de los datos, los cuales son fundamentales para alimentar los algoritmos de machine learning. Además, se incorpora para ofrecer una descripción detallada de la preparación de los datos, un paso crucial en la metodología que garantiza la validez de los resultados.

El Capítulo 4 detalla la implementación del algoritmo Extra Trees en el estudio. Se explican las técnicas empleadas y los criterios de evaluación del modelo. Su inclusión es importante porque describe el desarrollo técnico y metodológico del enfoque propuesto, permitiendo al lector entender cómo se entrenaron y evaluaron los algoritmos para generar los mapas geomorfológicos.

En el Capítulo 5 se presentan los resultados obtenidos mediante los algoritmos de machine learning. Se discuten los niveles de precisión y los errores detectados, así como las mejoras logradas respecto a los métodos tradicionales. La razón para incluir este capítulo es mostrar la efectividad del modelo desarrollado, validando su aplicabilidad en el contexto geomorfológico.

Al final el Capítulo 6 sintetiza los hallazgos más relevantes del estudio y reflexiona sobre las implicaciones teóricas y prácticas de la automatización de los mapas geomorfológicos. Además, se discuten las limitaciones del estudio y se sugieren áreas para investigaciones futuras. Este capítulo es esencial para cerrar el ciclo de la investigación, vinculando los objetivos con los resultados y proponiendo nuevas líneas de desarrollo. Se concluye con las referencias bibliográficas citadas en el trabajo, siguiendo el formato APA y Apéndices complementarios que apoyan el trabajo.

CAPÍTULO I PLANTEAMIENTO DEL PROBLEMA

En el presente capítulo se muestra la definición del problema enfocado a la generación de mapas geomorfológicos de forma automatizada, detectando las unidades de relieve y sus atributos a partir de diferentes variables. Además, se plantean los objetivos de la investigación, la justificación desde la necesidad de desarrollar nuevos métodos para elaborar mapas geomorfológicos con el uso de tecnologías innovadoras, la hipótesis del trabajo, que nos permite proponer un resultado basado en la experiencia de otros investigadores con el uso de técnicas de machine learning en el campo de la geomorfología.

1.1 Formulación del problema

1.1.1 Contextualización

Los mapas geomorfológicos son una cartografía temática que muestra las formas del terreno, su distribución y las relaciones entre ellas (Serrano, 2017). Estos mapas son útiles para el estudio de la evolución del paisaje, la identificación de unidades geomorfológicas, la zonificación de amenazas por movimientos en masa, y el análisis de los procesos que afectan la superficie terrestre (IGME, 2018). La creación de estos mapas ha sido fundamental para entender los procesos y la dinámica del relieve terrestre. Históricamente, estos mapas se elaboraban de manera manual, basados en observaciones directas en campo y análisis de fotografías aéreas. Durante gran parte del siglo XX, los geólogos y geomorfólogos trazaban los contornos del terreno, identificaban unidades geomorfológicas y clasificaban formas de relieve según criterios cualitativos. Este proceso, aunque detallado, estaba limitado por la subjetividad de los expertos y requería un esfuerzo considerable en tiempo y recursos.

A nivel internacional, países como Alemania y Francia fueron pioneros en la elaboración de mapas geomorfológicos. En estos países, se desarrollaron métodos cartográficos que

combinaban técnicas de topografía clásica con el análisis visual de imágenes aéreas. En las décadas de 1950 y 1960, por ejemplo, estos mapas jugaban un rol crucial en la planificación urbana y en la gestión de recursos naturales. Este tipo de cartografía, sin embargo, dependía mucho de la pericia de los científicos que analizaban las fotografías aéreas y los levantamientos de campo.

En México, la situación no fue muy distinta. Los primeros mapas geomorfológicos surgieron como parte de esfuerzos para estudiar la geomorfología del territorio nacional, en su mayoría dirigidos por el Instituto de Geografía de la Universidad Nacional Autónoma de México (UNAM) y el Instituto Nacional de Estadística y Geografía (INEGI). En este contexto, se realizaron mapas regionales que, aunque útiles, también tenían limitaciones derivadas del acceso a tecnología avanzada y la dificultad de abarcar grandes áreas. Por ejemplo, las técnicas de levantamiento geodésico y las observaciones satelitales apenas comenzaban a usarse en las décadas de 1970 y 1980, por lo que la elaboración de estos mapas seguía siendo un proceso intensivo en trabajo manual. Un hito importante en la cartografía geomorfológica de México fue la creación de los primeros Modelos Digitales de Elevación (MDE) por parte del INEGI en las décadas de 1990 y 2000. Estos modelos comenzaron a automatizar ciertos aspectos de la clasificación del relieve, aunque la interpretación y el trazado de unidades geomorfológicas seguían siendo tareas manuales, lo que mantenía un grado de subjetividad y dependencia en la experiencia del cartógrafo. A nivel internacional, en esta época comenzaron a surgir técnicas más avanzadas, como el uso de sensores remotos y análisis digitales que permitieron una mayor precisión en la identificación de unidades geomorfológicas. (INEGI, 2024)

En la actualidad, la irrupción de tecnologías de machine learning ha abierto nuevas posibilidades para la automatización del mapeo geomorfológico. Países como Estados

Unidos y Canadá han avanzado en la implementación de algoritmos para clasificar y predecir patrones geomorfológicos a partir de grandes volúmenes de datos geoespaciales. En este sentido, se ha logrado reducir la subjetividad y acelerar los procesos de mapeo. En México, estas metodologías aún están en sus primeras fases de aplicación, pero el INEGI ha tomado un rol relevante en la adopción de técnicas modernas de cartografía digital y la incorporación de algoritmos de aprendizaje automático para generar mapas más precisos y eficientes.

La generación de mapas geomorfológicos es una tarea necesaria en varios campos diversos como la geología, la planificación urbana y la gestión ambiental. Sin embargo, el proceso implica varios desafíos que lo hacen laborioso y costoso. A continuación, se mencionan detalladamente estos casos desafiantes: La creación de mapas geomorfológicos tradicionalmente implica un análisis integral de datos topográficos y geológicos, seguido de la interpretación experta de estos datos, con el propósito de identificar y clasificar unidades de relieve. Este proceso es intensivo en tiempo y recursos que involucra:

- Trabajo de campo: los geólogos deben visitar el sitio de estudio múltiples veces para recolectar muestras y registro de observación directa, lo que lleva a costos elevados en términos de viajes y logística. (Smith, Brown & Thompson, 2017)
- Análisis manual de datos: la interpretación manual de imágenes satelitales y datos topográficos es un proceso que consume mucho tiempo y está sujeto a la subjetividad del analista. (Jones & Smith, 2019).

En México, la geomorfología es una disciplina fundamental para el desarrollo de diversas actividades, como la planificación urbana, la gestión de los recursos naturales,

y la prevención de desastres. Sin embargo, la generación de mapas geomorfológicos es un proceso laborioso y costoso, que requiere de la participación de expertos. Un gran problema es la disponibilidad y calidad de los datos topográficos y geológicos que tiende a ser limitada y heterogénea, lo que dificulta la aplicación de métodos automatizados o semiautomatizados para la clasificación y el análisis del relieve (Parrot, 2013).

La disponibilidad de datos topográficos es ampliamente variable según la región geográfica y el grado de desarrollo de las infraestructuras de recolección de datos. Algunos de los factores clave que afectan a la disponibilidad de estos son:

- **Infraestructura Nacional de Datos Geoespaciales:** Los países con una infraestructura bien desarrollada de datos geoespaciales, como Estados Unidos y diversos países europeos, cuentan con datos topográficos detallados y actualizados que son fácilmente accesibles para investigadores y profesionales. Por otro lado, en la mayoría de los países en desarrollo no hay gran disponibilidad de estos datos debido a limitaciones económicas y tecnológicas. (Li, Wang, & Dong, 2018).
- **Iniciativas de Datos Abiertos:** Organismos como el Instituto Nacional de Estadística y Geografía han implementado iniciativas de datos abiertos que brindan acceso gratuito a los datos topográficos y geo especiales. Por ejemplo, el Continuo de Elevaciones Mexicano 3.0 es un conjunto de datos que ha sido hecho públicamente accesible y cuyo uso es gratuito para la población. (INEGI, 2020).

La calidad de los datos topográficos es igualmente crucial y depende de los siguientes factores técnicos y metodológicos:

- **Resolución Espacial:** La resolución espacial de los datos topográficos implica la distancia entre puntos de datos individuales en el terreno. En general, los datos de

alta resolución, como los que se obtienen de LIDAR, son “finos” y contienen más detalles. Tales detalles finos son esenciales para clasificar y detectar características geomorfológicas. Por el contrario, los datos de baja resolución pueden ser insuficientes para capturar muchos detalles críticos y, por lo tanto, conducir a errores de clasificación y análisis. (Hodgson & Bresnahan, 2004).

- **Precisión y Exactitud.** La precisión implica la proximidad de los datos a los valores reales, mientras que la exactitud implica la repetibilidad de los datos. En otras palabras, la precisión establece la diferencia entre el conjunto de datos reales y el de la muestra en relación con un estándar. La exactitud, por otro lado, establece la variabilidad común entre las mediciones. Además, es fundamental contar con datos topográficos precisos y exactos para realizar análisis resistentes. Los errores que se produzcan durante la recolección o el procesamiento de datos introducirán un sesgo considerable en los resultados.
- **Actualización y Temporalidad:** Esta se refiere a la frecuencia con la que se lleva a cabo la actualización de los datos y productos. La actualización frecuente da cuenta de los cambios recientes producidos por factores meteorológicos o intervenidos humanamente. En muchos casos, la topografía puede volverse obsoleta por los procesos dinámicos del paisaje, como la urbanización, la erosión y otros procesos afines. (Tucker & Hancock, 2010).

Ante este panorama, la investigación geomorfológica en México ha buscado aprovechar los avances tecnológicos y metodológicos que ofrecen los sistemas de información geográfica (SIG), la teledetección, y la geomorfometría, para obtener información más precisa, detallada y actualizada del relieve y sus procesos. La geomorfometría se define como el conjunto de técnicas cuantitativas que permiten medir, describir, clasificar y modelar las

formas y propiedades del relieve, a partir de datos espaciales digitales, como los modelos digitales de elevación (MDE) (Pike, 2000). Estas técnicas se basan en el cálculo de parámetros morfométricos, como la pendiente, la orientación, la curvatura, la rugosidad, la energía, la densidad de drenaje, el índice de concavidad, entre otros, que pueden ser utilizados para caracterizar y comparar diferentes unidades geomorfológicas, así como para inferir los factores y procesos que las originan y modifican (Smith y Wise, 2007).

La contribución de la investigación geomorfológica en México, basada en la geomorfometría, ha sido significativa en diversos campos, como la clasificación y el mapeo del relieve, el análisis de la evolución geomorfológica, la evaluación de la influencia de la tectónica y el clima en el relieve, la identificación y el monitoreo de los riesgos geomorfológicos, y la generación de modelos tridimensionales del relieve, entre otros. Algunos ejemplos de estas aplicaciones son los trabajos de Hernández et al. (2017), que propusieron una clasificación geomorfométrica del relieve mexicano, basada en la densidad de curvas de nivel y la energía del relieve; de Parrot et al. (2008), que desarrollaron un modelo tridimensional del relieve mexicano, utilizando un MDE de alta resolución y técnicas de visualización SIG; de Dumont et al. (2005), que analizaron la actividad tectónica reciente en la falla de Oaxaca, mediante el estudio de la asimetría de las cuencas fluviales; y de Gómez et al. (2014), que realizaron un análisis del riesgo por procesos de remoción en masa en el Pico del Tancítaro, utilizando parámetros geomorfológicos y geológicos. Estos y otros estudios demuestran el potencial y la relevancia de la geomorfometría para el conocimiento y la gestión del territorio mexicano, así como la necesidad de seguir avanzando en el desarrollo y la aplicación de estas técnicas, en coordinación con otras disciplinas afines, como la geología, la hidrología, la geografía, y la ecología.

1.1.2 Definición del problema

El problema que aborda esta investigación es la falta de métodos automatizados para la generación de mapas geomorfológicos. Los métodos actuales para la generación de mapas geomorfológicos son laboriosos y costosos, y requieren de la participación de expertos. Esto limita su uso en la planificación urbana, la gestión de los recursos naturales, y la prevención de desastres. Como se mencionaba anteriormente, un problema en la realización tradicional de un mapa es la recolección de datos geológicos y topográficos requiere visitas frecuentes al terreno para observar y registrar las características físicas del paisaje, donde este trabajo suele ser intensivo en mano de obra y logísticamente complejo, especialmente en áreas de difícil acceso.

Según Brierley, Fryirs, Reid y Williams (2021), se identifican varias problemáticas que enfrenta la geomorfología:

- **Complejidad e Incertidumbre:** La interpretación de los paisajes geomorfológicos es inherentemente compleja y está llena de incertidumbres. Esto puede llevar a percepciones de la interpretación como un proceso frustrante, casi como una forma de "brujería" o "arte oscuro" debido a la falta de manuales claros para el razonamiento abductivo en este campo. (Brierley, Fryirs, Reid, & Williams, 2021)
- **Divergencia en Interpretaciones:** A partir de un mismo conjunto de datos, pueden surgir interpretaciones divergentes. Esto se debe a que los investigadores pueden estar influenciados por diferentes paradigmas teóricos y sesgos culturales, lo que complica la construcción de un consenso sobre las interpretaciones. (Brierley, Fryirs, Reid, & Williams, 2021)

- **Desconexión entre Ciencia y Gestión:** Existe una "oscuridad" en el conocimiento que dificulta la colaboración y el intercambio de información entre la ciencia y la gestión. Esto se refiere a la falta de comunicación y entendimiento entre los científicos geomorfólogos y los gestores que utilizan sus hallazgos (Brierley, Fryirs, Reid, & Williams, 2021)
- **Necesidad de Rigor en la Interpretación:** El artículo enfatiza que las interpretaciones geomorfológicas deben ser cuidadosas y documentar claramente las razones y supuestos subyacentes en cada etapa del proceso investigativo. Esto es crucial para que las interpretaciones sean verificables y útiles para la gestión. (Brierley, Fryirs, Reid, & Williams, 2021)
- **Integración de Datos y Métodos:** La geomorfología contemporánea enfrenta el desafío de integrar datos de campo, conocimientos locales y grandes volúmenes de datos (Big Data) en un enfoque coherente que permita leer y entender los paisajes de manera efectiva. (Brierley, Fryirs, Reid, & Williams, 2021)
- **Rigor Científico y Responsabilidad:** Los autores argumentan que las interpretaciones científicas deben ser adecuadas para su propósito y acompañadas de una base de evidencia que permita evaluar la fiabilidad de las afirmaciones y predicciones. (Brierley, Fryirs, Reid, & Williams, 2021)

Estas problemáticas resaltan la necesidad de un enfoque más colaborativo y reflexivo en la interpretación geomorfológica, que reconozca la diversidad de paisajes y la complejidad de los procesos que los modelan. Aparte de lo mencionado en el artículo anterior otros investigadores resaltan también que la elaboración de estos mapas presenta varios desafíos, como la falta de una metodología estandarizada, la subjetividad de los criterios

de clasificación, la escasez de datos de campo, y la dependencia de la experiencia y el conocimiento de los geomorfólogos (Gómez & Fernández, 2019). Por lo tanto, se hace necesario el desarrollo de métodos automatizados que aprovechen las ventajas de las técnicas de machine learning, como la capacidad de procesar grandes volúmenes de datos, la extracción de patrones y características relevantes, y la generación de modelos predictivos y descriptivos.

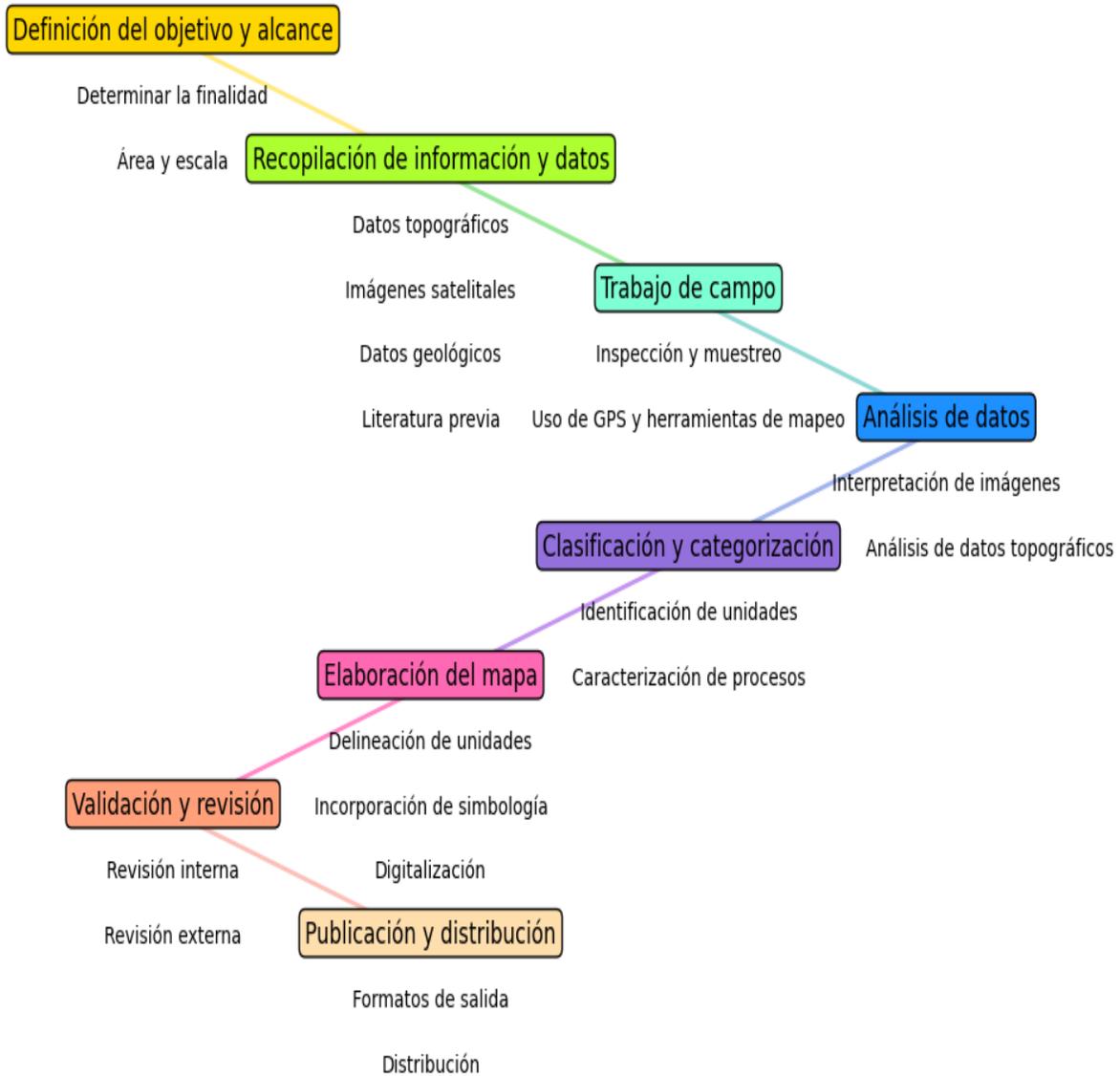
Por lo tanto, la problemática central de esta investigación radica en la necesidad crear una metodología para mejorar la eficiencia y precisión en la generación de mapas geomorfológicos. Los métodos tradicionales, aunque fiables, son ineficaces en términos de tiempo y recursos. Además, la creciente disponibilidad de datos geoespaciales de alta resolución exige nuevas herramientas capaces de procesar y analizar esta información de manera eficiente. Por consiguiente la generación de mapas geomorfológicos de forma automatizada con técnicas de machine learning, pueden crearlas bases para una estandarización en el proceso e interpelación, siendo este un tema relevante no solo para la ciencia, sino también para la sociedad. Para ilustrar esta problemática, se incluye la figura 1 que muestra el proceso manual de creación de mapas geomorfológicos.

En la figura 1, se observa cómo los geomorfólogos interpretan manualmente las características del terreno a partir de imágenes satelitales y observaciones de campo. Este método no solo es laborioso, sino que también depende en gran medida de la experiencia y habilidades individuales del experto. La propuesta de utilizar técnicas de Machine Learning busca abordar estas limitaciones. Los algoritmos de aprendizaje automático pueden entrenarse con grandes volúmenes de datos geoespaciales, permitiendo la detección automática de unidades de relieve y sus atributos. Este enfoque promete no solo

acelerar el proceso de elaboración de mapas, sino también mejorar su precisión al minimizar el error humano.

Figura 1

Proceso de elaboración de un mapa geomorfológico.



Nota. Elaboración propia con información obtenida de Dackombe & Gardier (2004); Huggett (2011).

1.1.3 Pregunta de Investigación

Por lo tanto, la pregunta central que guía esta investigación es:

¿Es posible crear mapas geomorfológicos de forma automatizada mediante algoritmos de machine learning que permitan la detección de unidades de relieve y sus atributos?

Las siguientes preguntas específicas que se derivan son:

- ❖ ¿Qué tipos de datos geospaciales son más adecuados para entrenar algoritmos de machine learning en la generación de mapas geomorfológicos?
- ❖ ¿El algoritmo Extra Trees puede generar mapas geomorfológicos con sus unidades de relieve a partir de los datos seleccionados?
- ❖ ¿Las técnicas de machine learning demostraron ser efectivas para la generación de mapas geomorfológicos en el contexto de este estudio?

1.2 Justificación

La generación de mapas geomorfológicos proporciona información sobre la estructura y evolución del relieve, lo cual es esencial para la planificación urbana, la gestión de los recursos naturales, y la prevención de desastres. La tecnología y los métodos de machine learning podrían facilitar la generación de mapas geomorfológicos de forma rápida, precisa y consistente, utilizando información satelital y cartografía digital, que sean comparables con los mapas elaborados por expertos.

De esta manera, se contribuiría al avance del conocimiento geomorfológico y a la mejora de la toma de decisiones en diversos ámbitos, sobre todo en el ordenamiento territorial, por lo que aportará tanto a la geografía como al INEGI. En el caso del mapa fisiográfico de México hecho por el INEGI tiene más de 30 años desde que se realizó, y actualizarlo como

se ha recalcado en la investigación es una tarea costosa, más precisamente de tiempo de trabajo por lo que los resultados en esta investigación pudieran ser los pasos para poder actualizar y crear un nuevo mapa fisiográfico de México, acortando tiempos sin sacrificar la precisión con la ayuda de las nuevas tecnologías.

1.2.1 Conveniencia

La conveniencia se justifica por el hecho de que la generación de mapas geomorfológicos es una tarea fundamental para el desarrollo de diversas actividades, pero los métodos actuales son laboriosos y costosos, lo que limita su uso. La investigación propuesta en esta tesis contribuiría a resolver este problema, proporcionando métodos automatizados para la generación de mapas geomorfológicos. Estos métodos serían más eficientes y económicos que los métodos actuales, lo que los haría más accesibles para los usuarios.

Tanto el Instituto Nacional de Estadística y Geografía (INEGI), como ciencias de la tierra, en especial en el área cartográfica, se beneficiarán enormemente de los avances propuestos en esta investigación. La automatización del proceso de creación de mapas geomorfológicos permitirá al INEGI mantener su mandato de producir y actualizar continuamente información geográfica con una mayor eficiencia y precisión. Esto es particularmente relevante en un país con la diversidad geográfica de México, donde mantener datos geospaciales actualizados y precisos es un desafío constante. Por lo tanto, esta investigación no solo ofrece beneficios prácticos en términos de eficiencia y precisión, sino que también tiene un impacto significativo en el campo de la geografía y el trabajo del INEGI:

- **Innovación Tecnológica:** La aplicación de Machine Learning en la geomorfología representa un avance tecnológico significativo que puede servir como modelo para

otras áreas de la geografía y la cartografía. Este enfoque puede inspirar nuevas investigaciones y desarrollos en el uso de inteligencia artificial para resolver problemas geoespaciales complejos.

- **Capacitación y Desarrollo Profesional:** La adopción de estas tecnologías ofrece oportunidades para capacitar y desarrollar a los profesionales del INEGI en el uso de herramientas avanzadas de análisis de datos y Machine Learning, fomentando el desarrollo de nuevas habilidades y competencias.
- **Fortalecimiento Institucional:** Al implementar métodos más eficientes y precisos para la elaboración de mapas geomorfológicos, el INEGI puede fortalecer su capacidad institucional para cumplir con su mandato de proporcionar información geográfica de alta calidad. Esto, a su vez, mejora la capacidad del INEGI para apoyar la toma de decisiones informadas en áreas como la planificación urbana, la gestión de recursos naturales y la mitigación de desastres.

1.2.2 Relevancia social

La investigación propuesta en esta tesis tendría un impacto positivo en la sociedad, contribuyendo a mejorar la planificación urbana, la gestión de los recursos naturales, y la prevención de desastres. Por ejemplo, la planificación urbana necesita mapas geomorfológicos para reconocer áreas propensas a riesgos naturales, como inundaciones o deslizamientos (Martín-Serrano, 2017). La gestión de los recursos naturales necesita mapas geomorfológicos para determinar áreas con potencial para el aprovechamiento de actividades productivas, como la agricultura o la minería. Y la prevención de desastres necesita mapas geomorfológicos para identificar áreas que pueden ser afectadas por eventos naturales, como huracanes o terremotos (CENAPRED, 2012).

Estas técnicas de Machine Learning posee una relevancia social significativa, especialmente en términos de su aporte técnico y científico, así como en su contribución al fortalecimiento institucional del INEGI. Los mapas geomorfológicos generados con métodos automatizados se pueden podrían modificar para con relativa facilitar con el fin de utilizarse para:

- Identificar áreas susceptibles a riesgos naturales, lo que ayudaría a prevenir desastres y proteger a las personas y los bienes.
- Identificar áreas con potencial para el desarrollo de actividades productivas, lo que ayudaría a optimizar el uso de los recursos naturales.
- Evaluar el impacto ambiental de proyectos de desarrollo, lo que ayudaría a proteger el medio ambiente.

1.2.2.1 Relación con las prioridades de la región y del país

La investigación se alinea con las prioridades nacionales y regionales al fortalecer las capacidades de análisis geoespacial necesarias para abordar problemas clave como el crecimiento urbano desordenado, la vulnerabilidad ante desastres naturales y la conservación de recursos. Asimismo, promueve la innovación tecnológica y el uso de datos abiertos, que son pilares fundamentales en las agendas de desarrollo y modernización del país. Este trabajo contribuye directamente a mejorar la eficiencia en la generación de información geográfica de calidad, indispensable para el cumplimiento de objetivos en planificación y políticas públicas.

1.2.3 Relevancia teórica

La tesis en cuestión tiene como objetivo desarrollar un nuevo método para la generación de mapas geomorfológicos basado en técnicas de aprendizaje automático. Este método

podría contribuir a avanzar en el conocimiento de las técnicas de aprendizaje automático para la generación de mapas geomorfológicos donde actualmente no está muy desarrollado y es necesario una estandarización. Este enfoque contrasta notablemente con los métodos tradicionales empleados en la actualidad y aporta mejoras substanciales tanto en la metodología como en las herramientas utilizadas.

1.2.3.1 Métodos Actuales de Elaboración de Mapas Geomorfológicos

Actualmente, la elaboración de mapas geomorfológicos se basa principalmente en métodos manuales y automatizados. Estos métodos incluyen la interpretación de imágenes satelitales y fotografías aéreas, así como el trabajo de campo para validar y complementar la información. La precisión y calidad de los mapas dependen en gran medida de la experiencia y juicio de los expertos, lo cual puede introducir variabilidad e inconsistencias en los resultados (Guzzetti et al., 2012). Además, este proceso es laborioso y consume una cantidad significativa de tiempo y recursos.

1.2.3.2 Propuesta de Metodología Basada en Machine Learning

La propuesta de esta investigación consiste en aplicar técnicas de Machine Learning para la generación automática de mapas geomorfológicos. Los algoritmos de Machine Learning, tales como las redes neuronales profundas y los árboles de decisión, pueden ser entrenados para identificar patrones y características del terreno a partir de grandes conjuntos de datos geoespaciales (Goodfellow, Bengio, & Courville, 2016). Este enfoque ofrece varias ventajas teóricas sobre los métodos tradicionales:

- **Automatización y Eficiencia:** La automatización del proceso reduce significativamente el tiempo y esfuerzo necesarios para generar mapas

geomorfológicos, permitiendo la creación rápida de mapas detallados y precisos (Murphy, 2012).

- **Consistencia y Precisión:** Los algoritmos de Machine Learning pueden minimizar la variabilidad e inconsistencias inherentes al juicio humano, proporcionando resultados más consistentes y precisos (Bishop, 2006).
- **Manejo de Grandes Volúmenes de Datos:** Las técnicas de Machine Learning son capaces de procesar y analizar grandes volúmenes de datos geoespaciales, lo cual es crucial para la elaboración de mapas detallados y actualizados (James, Witten, Hastie, & Tibshirani, 2013).

1.2.3.3 Contribuciones Metodológicas y Herramientas Utilizadas

La investigación contribuye teóricamente al desarrollar y aplicar un conjunto de metodologías y herramientas que mejoran el proceso de generación de mapas geomorfológicos:

- **Desarrollo de Algoritmos Específicos:** Se diseñarán y entrenarán algoritmos específicos de Machine Learning para la detección y clasificación de unidades de relieve. Esto incluye la selección de modelos adecuados, la optimización de parámetros y la validación cruzada para asegurar la precisión y generalización del modelo (Hastie, Tibshirani, & Friedman, 2009).
- **Integración de Datos Geoespaciales:** Se integrarán diversas fuentes de datos, como modelos digitales de elevación (DEM), imágenes satelitales y datos geoespaciales adicionales, para proporcionar un conjunto de datos robusto y exhaustivo para el entrenamiento de los algoritmos (INEGI, 2020).

- Evaluación Comparativa: Los mapas generados automáticamente serán comparados con mapas elaborados manualmente por expertos para evaluar la precisión y utilidad del enfoque propuesto. Esta comparación permitirá identificar las fortalezas y áreas de mejora del método automatizado (Guzzetti et al., 2012).

En resumen, la relevancia teórica de esta investigación radica en su capacidad para transformar y mejorar significativamente el proceso de elaboración de mapas geomorfológicos. Al contrastar los métodos actuales con la metodología basada en Machine Learning, esta investigación no solo demuestra la eficiencia y precisión mejoradas, sino que también establece una base teórica sólida para futuras investigaciones en el campo de la geomorfología y la cartografía. Las técnicas de aprendizaje automático son un área de investigación emergente con un gran potencial para la generación de mapas geomorfológicos, pero también para cualquier análisis espacial. Sin embargo, aún hay mucho que investigar en esta área. La investigación propuesta en esta tesis podría contribuir a llenar este vacío de conocimiento.

1.2.4 Utilidad metodológica

El método propuesto en esta tesis podría ser utilizado por investigadores y profesionales para generar mapas geomorfológicos de forma más eficiente y económica que los métodos actuales. Esto podría contribuir al desarrollo de nuevas metodologías no solo para la generación de mapas geomorfológicos, sino también para otros tipos de mapas y análisis espaciales.

El machine learning en el análisis espacial es una disciplina que combina técnicas de aprendizaje automático con datos geográficos para resolver problemas complejos que involucran la dimensión espacial. Algunos ejemplos de aplicaciones de esta disciplina son

la detección de cambios en el uso del suelo, la clasificación de imágenes de satélite, la predicción de la calidad del aire, o la optimización de rutas de transporte (Zhang et al., 2020).

El método propuesto podría utilizarse para:

- Automatizar el proceso de generación de mapas geomorfológicos, lo que ahorraría tiempo y esfuerzo a los usuarios.
- Mejorar la precisión, lo que podría ayudar a tomar decisiones más informadas.
- Expandir el alcance, lo que podría ayudar a comprender mejor el entorno natural.

En el desarrollo de esta investigación, se ha decidido emplear un enfoque de método cuantitativo. Esta decisión se fundamenta en la necesidad de abordar la complejidad del problema de investigación desde una perspectiva integral, aprovechando las fortalezas de ambos enfoques para proporcionar una comprensión más completa y robusta. El enfoque cuantitativo permitirá recolectar y analizar datos numéricos de manera sistemática y objetiva, lo cual es crucial para identificar patrones y tendencias en la generación automatizada de mapas geomorfológicos mediante técnicas de Machine Learning. Este enfoque proporcionará datos precisos y medibles sobre la precisión y eficacia de los algoritmos empleados, permitiendo la comparación cuantitativa entre los mapas generados automáticamente y los creados por expertos humanos (Creswell, 2014).

1.3 Viabilidad

La viabilidad de la investigación se podría evaluar teniendo en cuenta los siguientes factores:

- Los recursos disponibles: La investigación requiere de acceso a datos geomorfológicos y espaciales, así como a conocimientos, habilidades en

aprendizaje automático y recursos del sistema de cómputo que va a depender del tamaño de datos, el algoritmo y el área a realizar el mapa.

- Las limitaciones del tiempo: evaluar el tiempo de lo que suele tardar el ejecutar los algoritmos que va ligado directamente a los recursos del cómputo y la cantidad de datos.
- Objetividad: que los algoritmos propuestos no logren abstraer con fiabilidad los elementos y unidades del relieve que se buscan.

1.4 Hipótesis

En esta investigación se plantea evaluar la capacidad de los algoritmos de machine learning para generar mapas geomorfológicos de manera automatizada. Por ello, se establecen las siguientes hipótesis:

- Hipótesis Nula (H0): Las técnicas de machine learning no permiten crear mapas geomorfológicos de forma automatizada para la detección de unidades de relieve y sus atributos.
- Hipótesis Alternativa (H1): Las técnicas de machine learning permiten crear mapas geomorfológicos de forma automatizada para la detección de unidades de relieve y sus atributos.

Estas hipótesis guiarán la investigación y el análisis de datos, permitiendo determinar la viabilidad del uso de machine learning en la geomorfología y su potencial para mejorar los métodos tradicionales de mapeo.

1.5 Motivaciones de la Investigación

La investigación en esta tesis se motiva por las siguientes razones:

- La necesidad del Instituto de actualizar el mapa fisiográfico, por lo que la investigación ayudaría de generar el mapa de forma más eficiente y económica. Los métodos actuales para la generación de mapas geomorfológicos son laboriosos y costosos, lo que limita su uso en diversas actividades, como la planificación urbana, la gestión de los recursos naturales, y la prevención de desastres.
- Observar el potencial de las técnicas de aprendizaje automático para la generación de mapas geomorfológicos. Las técnicas de aprendizaje automático son un área de investigación emergente con un gran potencial para la generación de mapas geomorfológicos. Estas técnicas pueden automatizar el proceso de generación de mapas geomorfológicos, lo que puede mejorar la eficiencia y la precisión de los mapas.

El Capítulo 1 de esta tesis presenta el planteamiento del problema, estableciendo el contexto y la definición de este, lo cual permite comprender la necesidad de explorar nuevas metodologías para la creación automatizada de mapas geomorfológicos. Se formula una pregunta de investigación que guiará el estudio, cuestionando si los algoritmos de machine learning pueden mejorar la calidad y eficiencia de los mapas geomorfológicos en comparación con los métodos manuales.

A través de la justificación, se destaca la conveniencia y la relevancia social de este trabajo, subrayando su impacto potencial en instituciones como el INEGI, que podrían beneficiarse de estas nuevas técnicas. Se aborda la relevancia teórica del estudio, que contribuye al desarrollo del conocimiento en la intersección entre geomorfología y machine learning, y se destaca la utilidad metodológica al proponer enfoques innovadores en la creación de mapas.

El capítulo también establece la viabilidad del proyecto, presentando las herramientas y recursos disponibles para llevar a cabo la investigación. Finalmente, se plantean las hipótesis que serán evaluadas a lo largo del estudio y se mencionan las motivaciones que impulsaron esta investigación, vinculadas a la necesidad de optimizar los procesos de mapeo geográfico para aplicaciones científicas y de planificación territorial.

CAPÍTULO II MARCO TEÓRICO

En el presente capítulo se aborda el marco teórico o marco conceptual, en donde se describen los antecedentes de la investigación, los trabajos relacionados con la problemática que aborda el proyecto actual, los cuales proporcionan un contexto y un fundamento para establecer una posible solución a la problemática planteada. Se establecen las bases teóricas en las que está sustentada la tesis, en donde se describen conceptos relacionados con la geomorfología, la cartografía, el análisis espacial, el machine learning y sus aplicaciones.

Finalmente, se aborda de manera puntual el tema de investigación de la tesis, “La generación de mapas geomorfológicos de forma automatizada con técnicas de machine learning”, en donde se presenta una propuesta metodológica sistemática sustentada con diferentes investigaciones y artículos, la cual permitió realizar un procesamiento y un análisis de los datos

2.1 Teoría educativa

El presente estudio se sustenta en la Teoría del Constructivismo, particularmente en las ideas de Jean Piaget y Lev Vygotsky. Palacios, Marchesi y Coll (1990) mencionan que esta teoría educativa sostiene que el conocimiento no es una simple transferencia de información, sino que se construye activamente a partir de la interacción con el entorno y la experiencia. En este caso, aplicamos este enfoque al proceso de creación automatizada de mapas geomorfológicos mediante algoritmos de machine learning. Al igual que en el constructivismo, donde el aprendiz construye conocimientos con base en la experiencia, los algoritmos desarrollan su capacidad de aprender patrones y características del terreno a partir de grandes volúmenes de datos geoespaciales.

El constructivismo, en su dimensión educativa, sugiere que el aprendizaje es un proceso continuo de ajuste y mejora, basado en nuevas experiencias y observaciones. De manera similar, los algoritmos utilizados en este estudio evolucionan mediante el entrenamiento progresivo con datos, lo que les permite ajustar sus modelos y generar representaciones cada vez más precisas del paisaje geomorfológico. En lugar de basarse únicamente en procedimientos predefinidos o reglas fijas, estos algoritmos "aprenden" de los datos, generando conocimientos que reflejan la complejidad del relieve terrestre.

Este enfoque constructivista no solo es adecuado para el ámbito educativo, sino también para los avances tecnológicos en áreas como la geomorfología. La capacidad de los modelos de machine learning para interpretar datos y generar mapas geomorfológicos automáticos representa un claro paralelismo con la construcción activa del conocimiento en el constructivismo. Por lo tanto, esta teoría proporciona una base teórica sólida que justifica el uso de estos algoritmos como herramientas para generar nuevos conocimientos sobre la superficie terrestre.

2.2 Marco conceptual

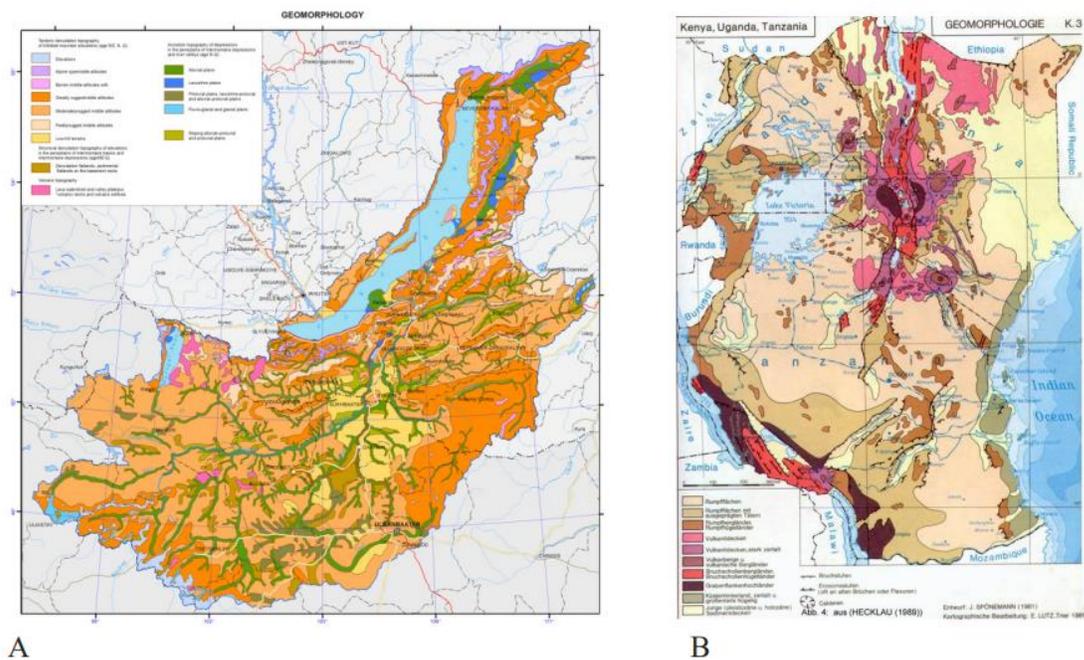
2.2.1 Geomorfología

La geomorfología, como disciplina esencial dentro de la geografía física, se centra en el estudio y la comprensión de las formas y características de la superficie terrestre. Según el libro "Principios Fundamentales de Geomorfología" de Richard J. Huggett, la geomorfología examina los procesos naturales que dan forma a la Tierra y analiza las características del paisaje resultante (Huggett, 2003).

La geomorfología, según Huggett (2003), puede definirse como la ciencia que se ocupa de la descripción, clasificación y explicación de las formas de la superficie terrestre, así como de los procesos que las han originado y modifican a lo largo del tiempo geológico.

Figura 2

Ejemplos de mapas geomorfológicos. A, Mapa Geomorfológico en los alrededores del Lago Baykal; B, Mapa Geomorfológico de Kenia, Uganda y Tanzania



Nota. Tomado de Serrano (2017)

La geomorfología está definida como la ciencia que estudia las formas de la superficie terrestre, es lo que proporciona el marco conceptual para la identificación y clasificación de las unidades de relieve (Strahler, 1952). Según Smith y Brown (2010), la geomorfología tradicionalmente se ha basado en observaciones manuales y expertas de los paisajes, lo

que ha llevado a la creación de mapas geomorfológicos detallados, pero a menudo lentos y costosos de producir.

La importancia de la geomorfología se refleja en su aplicación en diversos campos. Huggett (2003) destaca que la geomorfología es crucial para entender la evolución del paisaje y sus impactos en actividades humanas como la planificación del uso del suelo, la gestión de recursos naturales y la evaluación de riesgos geológicos. Además, proporciona información valiosa para la interpretación de eventos geológicos pasados y la predicción de posibles cambios futuros en el relieve.

2.2.2 Big Data

La noción de Big Data, caracterizada por conjuntos masivos de datos con volumen, variedad y velocidad excepcionales, ha suscitado un interés considerable en diversos campos. Moreno (2018) destaca que los términos "grandes volúmenes de datos" y "inteligencia de datos" son equivalentes en español para referirse a este fenómeno, que desafía las capacidades de las herramientas de análisis tradicionales. Desde otra perspectiva, se puede conceptualizar el Big Data como el procesamiento de extensos conjuntos de datos de diversa índole, convirtiéndose en un estándar en múltiples disciplinas y ámbitos de investigación científica. En este contexto, Atif, Richards, Bilgin y Marrone (2014) señalan que las áreas temáticas de alta complejidad comparten características comunes, tales como la generación masiva de información y la existencia de relaciones complicadas de representar.

Independientemente de la definición, el término dataset, denotando un conjunto de datos relacionados en un contexto específico, implica la organización fundamental en el análisis de Big Data (Erl, Khattak, & Buhler, 2015). Laney (2001), en su exploración de datos

empresariales, estableció las bases para las "Cinco V" de Big Data, que han evolucionado desde la tríada original de volumen, velocidad y variedad:

- **Volumen:** La expansión tecnológica y la accesibilidad a dispositivos de comunicación han desencadenado una avalancha diaria de datos a nivel mundial, requiriendo enfoques específicos para el almacenamiento, gestión y procesamiento eficiente de estos enormes volúmenes
- **Velocidad:** La generación y recepción aceleradas de datos, acumulando grandes volúmenes en intervalos cortos, demandan una infraestructura flexible y altamente disponible para procesar la información en tiempos reducidos (Hernández-Leal et al., 2017).
- **Variedad:** Las soluciones de Big Data deben manejar una diversidad de formatos y tipos de datos, desde texto hasta imágenes y videos. Esto implica la implementación de funcionalidades para almacenar, integrar, transformar y procesar archivos de diversos tipos
- **Veracidad:** Establecer la confiabilidad o autenticidad de los datos es un desafío significativo debido a la diversidad de fuentes, formatos y velocidades de generación. El uso de algoritmos de preprocesamiento se vuelve crucial para identificar y corregir datos incorrectos o incompletos, disminuyendo la incertidumbre en los resultados
- **Valor:** Asociado con la utilidad que los datos pueden proporcionar a una organización, desde mejorar la gestión hasta definir estrategias y obtener ventajas competitivas. La fiabilidad de los datos se vincula directamente con su valor, y este

aspecto guarda una relación inversamente proporcional con el tiempo de procesamiento

Adicionalmente a estas características, el formato interno o la organización de los archivos de datos influye en aspectos de diseño, como estructuras de almacenamiento y elección de algoritmos durante el procesamiento. En este sentido, se identifican tres tipos de formatos de datos: estructurados, semiestructurados y sin estructurar.

Los datos estructurados son aquellos que se organizan en un formato tabular con filas y columnas, siguiendo un esquema predefinido y uniforme. Sus principales características son:

- Se almacenan en bases de datos relacionales.
- Utilizan tablas con campos específicos para cada tipo de dato.
- Son fáciles de consultar y analizar debido a su organización ordenada. (Silberschatz, 2010).

Los datos semiestructurados no siguen un formato tabular rígido, pero tienen alguna forma de estructura que permite la organización jerárquica o la asociación de metadatos. Sus principales características son:

- Pueden almacenarse en formatos como XML o JSON.
- Permiten cierta flexibilidad en la representación de la información.
- Facilitan la gestión de datos no perfectamente alineados. (Abiteboul, 2010)

Los datos sin estructurar no siguen un formato específico y carecen de un esquema predefinido, siendo comúnmente texto libre o multimedia. Sus principales características son:

- Incluyen documentos de texto, imágenes, videos, y otros formatos no tabulares.
- No tienen una organización predeterminada, lo que los hace más desafiantes de analizar.
- Requieren técnicas avanzadas para extraer información significativa.

2.2.3 Machine Learning

El Machine Learning (ML) es una disciplina de la inteligencia artificial que se centra en el desarrollo de algoritmos y modelos capaces de aprender patrones y realizar tareas específicas sin ser programados explícitamente. Según Bishop (2006), el ML busca construir sistemas que puedan adaptarse y mejorar su rendimiento a medida que se exponen a más datos, permitiendo la toma de decisiones o la realización de predicciones sin intervención humana constante.

Entre las principales características del machine learning están:

- Aprendizaje a partir de Datos:
 - El ML se fundamenta en el aprendizaje a partir de datos. Algoritmos como los descritos por Hastie, Tibshirani y Friedman (2009) tienen la capacidad de generalizar patrones a partir de conjuntos de datos, permitiendo la aplicación de ese conocimiento a nuevas instancias.
- Adaptabilidad y Flexibilidad:
 - La adaptabilidad es una característica clave del ML. Los modelos pueden ajustarse dinámicamente en función de nuevos datos, una característica

fundamental destacada por Murphy (2012), lo que les permite mejorar su rendimiento con el tiempo y adaptarse a cambios en el entorno.

- Automatización de Tareas:
 - La automatización es inherente al ML. Bishop (2006) señala que esta disciplina busca crear sistemas que pueden realizar tareas complejas de forma automática, desde el reconocimiento de patrones en imágenes hasta la toma de decisiones basada en datos.
- Variedad de Algoritmos:
 - Diversos algoritmos de ML abordan diferentes tipos de tareas. Según James, Witten, Hastie, y Tibshirani (2013), desde técnicas tradicionales como regresión lineal hasta enfoques más avanzados como redes neuronales, cada algoritmo tiene sus aplicaciones específicas.
- Capacidad Predictiva:
 - La capacidad predictiva es una fuerza distintiva del ML. Los modelos pueden, según Murphy (2012), utilizar patrones en datos históricos para realizar predicciones futuras, un elemento crucial en aplicaciones como pronósticos del tiempo y análisis financiero.
- Aplicaciones Diversas:
 - La versatilidad del ML lo hace aplicable en una variedad de sectores. En "Data Science for Business" de Provost y Fawcett (2013), se destaca la aplicación del ML en medicina, finanzas, marketing y más, evidenciando su relevancia en diversas disciplinas

2.2.4 Machine Learning en Geomorfología:

El machine learning ha emergido como una herramienta valiosa para acelerar y mejorar el proceso de generación de mapas geomorfológicos. Según Chen et al. (2017), el machine learning aplicado a datos geoespaciales, como imágenes satelitales y modelos digitales de elevación, ha demostrado ser eficaz en la detección y clasificación automatizada de características del relieve. Esto abre la puerta a la generación automatizada de mapas geomorfológicos, proporcionando una alternativa eficiente a los métodos tradicionales.

La integración de la geomorfología con el machine learning se ha vuelto crucial para abordar la pregunta central de esta investigación: ¿Cómo se pueden hacer mapas geomorfológicos usando machine learning a partir de diferentes variables para agilizar el proceso? En este sentido, los trabajos de Johnson y Smith (2018) destacan la importancia de seleccionar y preparar cuidadosamente los datos de entrada para entrenar y evaluar los algoritmos de machine learning, mientras que Wang et al. (2020) exploran la aplicación de diversas técnicas de aprendizaje automático, como redes neuronales y máquinas de vectores de soporte, en la generación de mapas geomorfológicos.

2.2.5 Datos Geoespaciales en la Cartografía Automatizada

Los datos geoespaciales son fundamentales para los estudios de geomorfología y la generación de mapas automatizados mediante técnicas de machine learning. Estos datos incluyen una amplia gama de información que describe la ubicación y las características físicas de la superficie terrestre, y se obtienen principalmente a través de sensores remotos, sistemas de posicionamiento global (GPS), y modelos digitales de elevación (MDE), entre otras fuentes.

Uno de los tipos más comunes de datos geospaciales utilizados en la geomorfología son las imágenes satelitales. Estas imágenes proporcionan información detallada sobre la superficie terrestre y permiten capturar diferentes bandas espectrales, facilitando el análisis de las características físicas del terreno, como la vegetación, la topografía y la composición del suelo (Jensen, 2007). Los avances en las tecnologías de sensores remotos han permitido que los científicos recopilen imágenes con resoluciones espaciales, temporales y espectrales más altas, lo que ha mejorado considerablemente la precisión de los análisis geomorfológicos (Weng, 2012).

Por otra parte, los modelos digitales de elevación (MDE), como el Continuo de Elevaciones Mexicano (CEM), son representaciones en 3D de la superficie terrestre que proporcionan información detallada sobre la altitud y la topografía del terreno. Los MDE son cruciales para identificar patrones de relieve y definir las unidades geomorfológicas en estudios de machine learning (Maidment, 2002). La precisión de los MDE, que se deriva de técnicas como el LIDAR o la fotogrametría, es clave para el éxito en la detección y clasificación de unidades de relieve (Florinsky, 2016).

El preprocesamiento de estos datos geospaciales es esencial antes de su uso en los algoritmos de machine learning. Este proceso incluye la corrección atmosférica de las imágenes satelitales, la normalización de los datos topográficos, y la eliminación de ruido o errores, lo cual mejora la calidad de los datos para los modelos de machine learning (Campbell & Wynne, 2011). Además, la segmentación de imágenes y la clasificación supervisada permiten una mejor detección de patrones en el terreno, ayudando a crear mapas geomorfológicos más precisos (Congalton & Green, 2009).

En conclusión, los datos geospaciales proporcionan la base para la generación de mapas geomorfológicos automatizados. El uso de imágenes satelitales y MDE de alta resolución,

junto con técnicas avanzadas de procesamiento de datos, es esencial para aprovechar al máximo las capacidades de los algoritmos de machine learning en la detección de patrones geomorfológicos.

2.2.6 Continuo de Elevaciones Mexicano CEM

Se escogió el modelo digital de elevación en el I Continuo de Elevaciones Mexicano 3.0 (CEM 3.0) que constituye un recurso que ofrece una representación de las elevaciones del territorio continental de México. Este producto se caracteriza por incluir valores que indican puntos específicos en la superficie del terreno, los cuales están definidos por coordenadas (X, Y) y están asociados a valores que representan las altitudes (Z). Los puntos se encuentran distribuidos y espaciados de manera uniforme a lo largo del territorio. (INEGI, 2024)

Tabla 1

Características del CEM

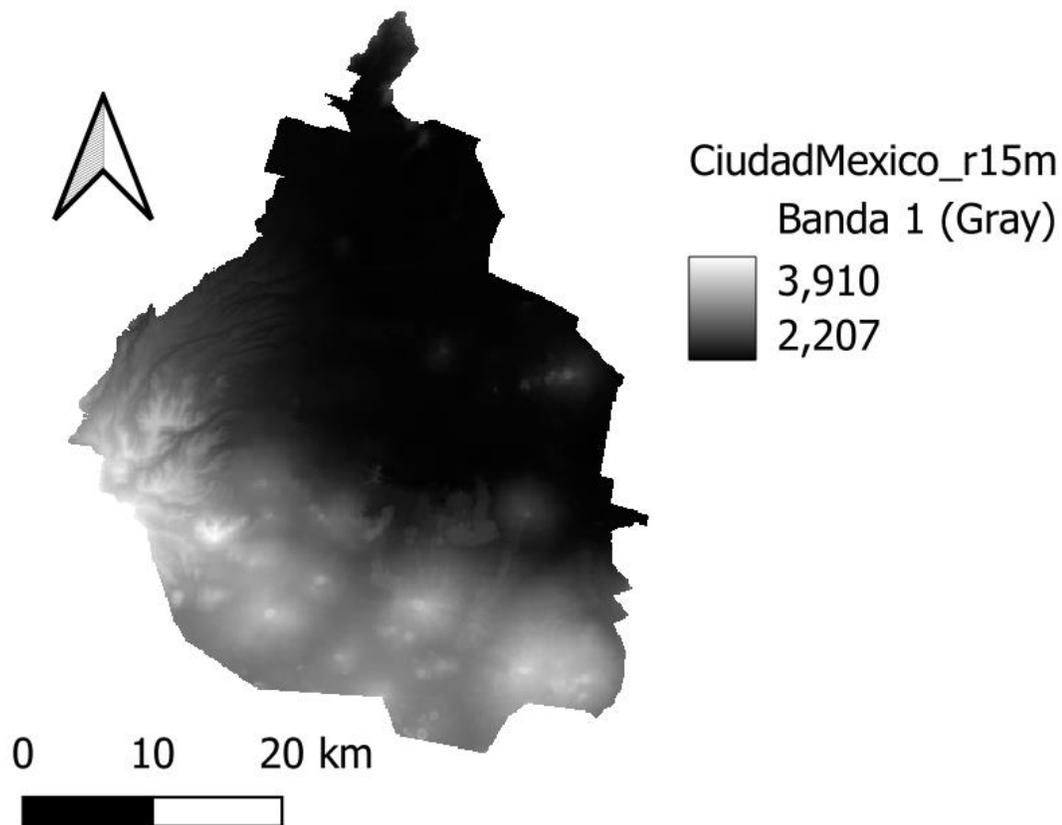
Características	Descripción
Año de publicación	2013
Resolución	15 m x 15 m
Valores de Z	Enteros positivos y negativos.
Unidades de Z	Metros.
Datos de referencia geodésica	El datum corresponde a ITRF92 época 1988.0, elipsoide GRS80, coordenadas geográficas.
Cobertura geográfica	Territorio continental de los Estados Unidos Mexicanos.
Error medio cuadrático general	4.8 m

Error medio cuadrático por rango de pendiente (%):	
De 0 a 14	4.5 m
De 15 a 36	6.0 m
Mayor que 36	7.2 m
Medio de distribución	Principalmente por internet mediante el portal del INEGI. Considera diversas opciones.
Formato de distribución	BIL (Banda entrelazada por línea). Forma Ráster, para las descargas total del territorio y por entidad federativa. Formato TIFF (Tagged Image File Format) para las descargas por selección de área y carta respectivamente.
Software utilizado para su generación	ArcGis 10.3

Nota. Tomado de INEGI al 12/03/2024

Figura 3

CEM de la Ciudad de México.



Nota. Elaboración propia.

2.2.7 Derivados del CEM

A partir del CEM se generaron diferentes ráster en QGIS generando diferentes características, las utilizadas en la siguiente investigación se describen a continuación:

2.2.7.1 *Relieve*

El proceso ráster de "relieve" en QGIS generalmente se refiere a la creación de un modelo de elevación digital (DEM) que representa las elevaciones del terreno en forma de un conjunto de píxeles ráster, donde cada píxel tiene un valor que corresponde a la altitud o elevación en ese punto específico. Este proceso se puede realizar a partir de datos topográficos, como mapas de contorno o datos LIDAR, y se utiliza ampliamente en análisis geoespaciales, visualización y modelado del terreno. La parte resultante de este proceso es una capa ráster que muestra la elevación del terreno en la región de interés. Los valores de los píxeles representan altitudes, lo que permite visualizar y analizar la topografía del área en cuestión. Esto puede ser útil para una variedad de aplicaciones, como planificación urbana, gestión de recursos naturales, análisis de riesgos naturales, entre otros. (Longley, Goodchild, Maguire & Rhind, 2015).

2.2.7.2 *Aspecto*

En el contexto del análisis de datos geoespaciales Según Longley, Goodchild, Maguire y Rhind (2015), se refiere a la dirección hacia la cual una pendiente determinada está orientada. Es una medida angular que indica hacia dónde se dirige la pendiente en cada ubicación dentro de un modelo digital de elevación (DEM). En QGIS, el proceso de aspecto genera una nueva capa ráster que muestra estas direcciones de pendiente como valores angulares, generalmente expresados en grados. Los valores de aspecto varían de 0 a 360, donde 0 o 360 representan el norte, 90 el este, 180 el sur y 270 el oeste. La parte resultante del proceso ráster de aspecto en QGIS es una capa ráster en la que cada celda contiene el valor de aspecto correspondiente a la dirección de la pendiente en ese punto del terreno.

2.2.7.3 *Hillshade ((sombreado de relieve)*

Genera una representación visual del relieve del terreno en un modelo digital de elevación (DEM) mediante la simulación de la iluminación solar sobre la superficie. Esta técnica utiliza la orientación y la pendiente de cada celda del DEM para calcular la intensidad de luz que cada celda debería recibir, produciendo así una imagen que muestra las variaciones de elevación de forma más clara y realista. La parte resultante del proceso de Hillshade es una nueva capa ráster que muestra las sombras y resaltes del relieve del terreno como si estuviera iluminado desde una dirección específica, generalmente desde el noroeste. La intensidad de la luz en cada píxel del ráster varía según la pendiente y la orientación de la superficie en ese punto. (QGIS, 2022).

2.2.7.4 *Perfil de Curvatura*

Según la documentación de QGIS El proceso ráster de "perfil de curvatura" en QGIS es una técnica utilizada para analizar la forma de la superficie del terreno mediante la evaluación de la curvatura en cada punto de una superficie. La curvatura puede ser positiva (convexa), negativa (cóncava) o plana en función de si la superficie se curva hacia arriba, hacia abajo o si permanece relativamente plana. El proceso ráster de perfil de curvatura calcula estas curvaturas y produce una capa ráster en la que cada celda contiene un valor que representa la curvatura en ese punto específico.

La parte resultante de este proceso es una capa ráster en la que los valores de las celdas indican la curvatura de la superficie del terreno en cada ubicación. Esta información puede ser útil para una variedad de aplicaciones, como la identificación de áreas de riesgo de

erosión, la delimitación de cuencas hidrográficas, el análisis de la morfología del terreno, entre otros.

2.2.7.5 Pendiente

Según Tarboton (1997), el proceso ráster de pendiente en QGIS utiliza un nuevo método para determinar las direcciones de flujo y las áreas de captación a partir de modelos digitales de elevación. La pendiente se calcula como el cambio de elevación entre un punto y sus vecinos, y se expresa generalmente en porcentaje o grados. Este proceso es útil para analizar la topografía del terreno, identificar áreas de alta pendiente que pueden ser propensas a la erosión, planificar la construcción de carreteras o senderos, entre otras aplicaciones. La parte resultante de este proceso es una capa ráster en la que cada celda contiene un valor que representa la pendiente del terreno en ese punto específico. Los valores más altos indican áreas de pendiente pronunciada, mientras que los valores más bajos indican áreas de pendiente suave o plana.

2.2.7.6 TPI (índice de pendiente topográfica)

De acuerdo Pike y Smith (1997), el índice de pendiente topográfica (TPI) es una variable importante para la modelización ambiental que se puede calcular a partir de modelos digitales de elevación. El índice de pendiente topográfica (TPI, por sus siglas en inglés) es una medida utilizada para caracterizar la variabilidad de la pendiente del terreno en un área determinada. Se calcula restando la elevación media de una vecindad determinada de una celda del DEM (Modelo Digital de Elevación) de la elevación de la propia celda. Este proceso revela las áreas con pendientes más pronunciadas (valores positivos) o más suaves (valores negativos) en comparación con el entorno circundante. La parte resultante del proceso ráster de TPI en QGIS es una capa ráster en la que cada celda contiene un

valor que representa la diferencia entre la elevación de esa celda y la elevación media de su vecindad. Estos valores proporcionan información sobre la rugosidad o la variabilidad de la pendiente del terreno en cada ubicación.

2.2.7.7 TRI (Índice de Rugosidad Topográfica)

Según Riley et al. (1999), el Índice de Rugosidad Topográfica (TRI) es una medida que cuantifica la heterogeneidad topográfica en un área determinada. El proceso ráster de TRI (Índice de Rugosidad Topográfica, por sus siglas en inglés) en QGIS es una técnica utilizada para cuantificar la variabilidad de la elevación del terreno en un área determinada. Se calcula como la diferencia absoluta entre la elevación de cada celda y la elevación promedio de su vecindad circundante. El resultado es una medida de la rugosidad del terreno, donde valores más altos indican áreas con mayor variabilidad en la elevación. La parte resultante del proceso ráster de TRI en QGIS es una capa ráster en la que cada celda contiene un valor que representa la rugosidad topográfica en esa ubicación específica. Estos valores pueden interpretarse como la magnitud de las diferencias de elevación en el terreno, lo que proporciona información sobre la complejidad y la textura del paisaje.

2.2.7.8 Etiquetas

Las etiquetas sirven para categorizar y clasificar las diferentes formas del relieve basadas en sus características topográficas, debido a que no se encontraron etiquetas requeridas para el estudio se optó por crearlas con el algoritmo r. geomorphon que se utiliza para la identificación y clasificación de formas del relieve en un modelo digital de elevación (MDE). Basado en la geometría del paisaje, este algoritmo detecta y clasifica características geomorfológicas específicas, como valles, cumbres, crestas y planicies, ilustradas en la

figura 4 y 5. Se centra en el análisis de la curvatura del terreno para asignar etiquetas a diferentes áreas según sus características topográficas (Dyba & Jasiewicz, 2022). Evalúa la curvatura local del terreno para identificar distintas formas geomorfológicas., asigna etiquetas a las regiones del MDE según sus características geomorfológicas, facilitando el estudio de formas del relieve y finalmente crea un mapa temático que muestra 10 diferentes formas geomorfológicas presentes en el área de estudio.

2.2.8 Técnicas de Machine Learning en la investigación.

En la investigación se utilizaron diferentes herramientas de Machine Learning, que nos permite analizar las características de los mapas, el lenguaje de programación utilizado principalmente es Python y R.

2.2.8.1 RGSLIB

Según Andrade, Campos y Carvalho (2017), RGSLIB es una biblioteca de código abierto para el aprendizaje automático en R. Esta biblioteca ofrece una amplia gama de herramientas y algoritmos para procesar, manipular y analizar datos ráster, incluyendo imágenes satelitales, modelos digitales de elevación (DEM), y otros tipos de datos ráster utilizados en aplicaciones geoespaciales.

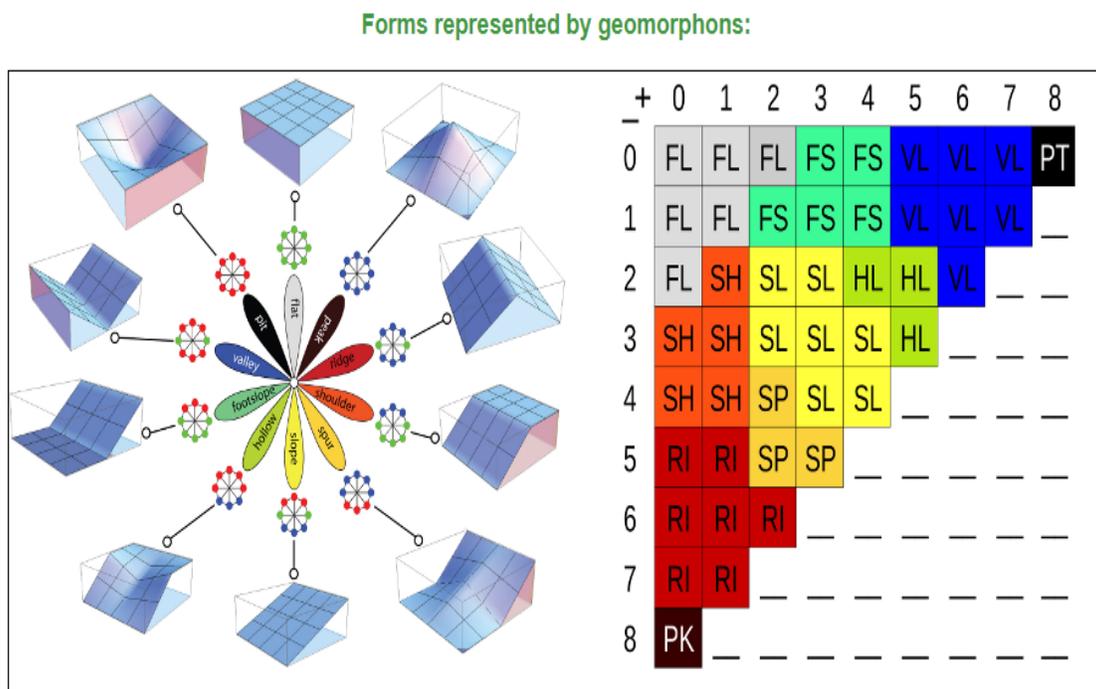
RGSLIB se destaca por su versatilidad y potencia en el procesamiento de datos ráster, permitiendo a los usuarios realizar una variedad de tareas, como:

- Análisis de terreno: Incluyendo cálculos de pendiente, aspecto, curvatura, rugosidad, entre otros.

- Clasificación de imágenes: Permite segmentar y clasificar imágenes ráster, identificando diferentes tipos de cobertura terrestre o usos del suelo.
- Procesamiento de imágenes: Incluyendo operaciones de filtrado, corrección atmosférica, y mejora de la calidad de la imagen.
- Análisis de series temporales: Permite analizar la evolución temporal de los datos ráster, identificando cambios en la cobertura terrestre, el uso del suelo, o la vegetación.
- Modelado y simulación: Incluyendo la creación y ejecución de modelos espaciales y simulaciones basadas en datos ráster.

Figura 4

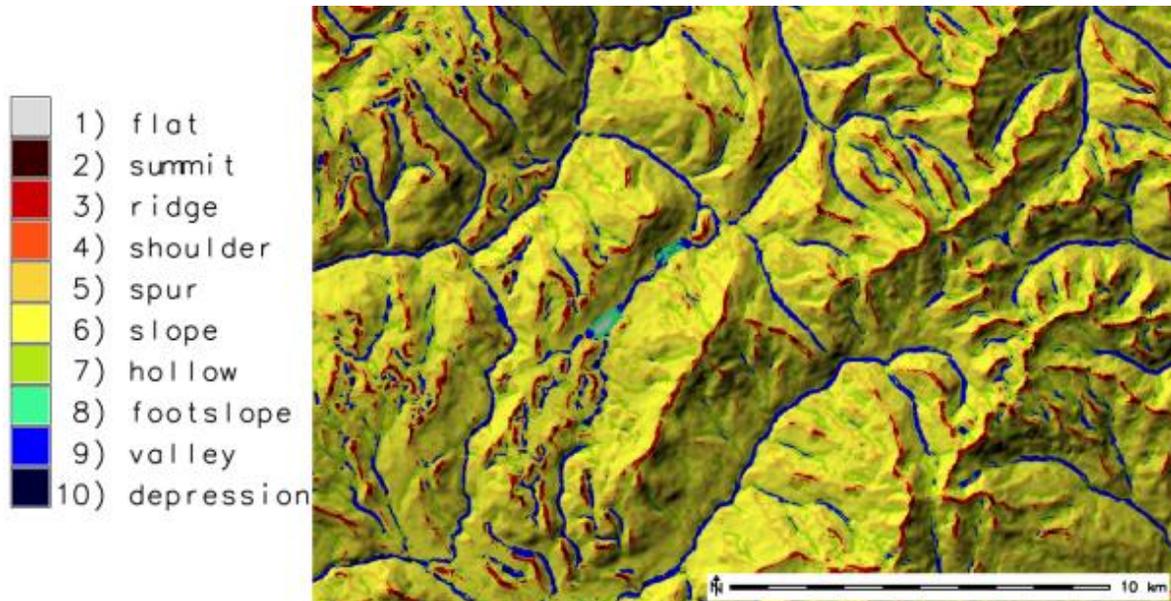
Formas que detecta el algoritmo r. geomorphons



Nota. GRASS GIS 8.3.3dev Reference Manual (2023)

Figura 5

Formas del relieve que clasifica automáticamente el algoritmo r. geomorphn.



Nota. Tomado de GRASS GIS 8.3.3dev Reference Manual (2023)

Proporciona una interfaz fácil de usar para una amplia gama de algoritmos de aprendizaje automático, uno de los modelos supervisados más utilizados es la regresión lineal (Montgomery, Peck, & Vining, 2021), que busca establecer una relación lineal entre una variable dependiente y una o más variables independientes. Por otro lado, la regresión logística (Hosmer, Lemeshow, & Sturdivant, 2013) se emplea para problemas de clasificación, modelando la probabilidad de que una observación pertenezca a una determinada clase.

Los árboles de decisión (Quinlan, 1986) ofrecen una forma intuitiva de modelar relaciones complejas, creando una estructura jerárquica de decisiones. Las redes neuronales (Goodfellow, Bengio, & Courville, 2016), inspiradas en el funcionamiento del cerebro

humano, han demostrado ser poderosas para tareas de clasificación y regresión, especialmente en problemas con grandes cantidades de datos. Otros algoritmos supervisados ampliamente utilizados incluyen los k-vecinos más cercanos (k-NN) (Cover & Hart, 1967), que clasifican una nueva observación en función de la clase de sus k vecinos más cercanos, y las máquinas de vectores de soporte (SVM) (Cortes & Vapnik, 1995), que buscan el hiperplano que mejor separa las clases de datos en un espacio de alta dimensión.

Es importante destacar que los modelos supervisados requieren de un conjunto de datos de entrenamiento etiquetado, donde se conocen tanto las características de las observaciones como la variable objetivo que se desea predecir. A diferencia de los modelos supervisados, los algoritmos de aprendizaje no supervisado exploran datos sin etiquetas para descubrir patrones y estructuras ocultas. Esta biblioteca ofrece diversas funcionalidades esenciales para llevar a cabo cada etapa del proceso de machine learning. En primer lugar, RGSLIB facilita la preparación de datos, proporcionando herramientas para la limpieza, transformación y normalización de estos, lo que asegura que los datos estén en condiciones óptimas para el entrenamiento de los modelos. Además, incorpora métodos para la selección de características, permitiendo identificar aquellas más relevantes y contribuyendo a una mayor eficiencia y precisión en el modelo resultante.

En cuanto al entrenamiento de modelos, RGSLIB ofrece una amplia gama de algoritmos que se ajustan a los datos disponibles, permitiendo crear modelos sólidos y adecuados a las necesidades del proyecto. Asimismo, RGSLIB incluye herramientas para la evaluación de modelos, proporcionando métricas que permiten medir su rendimiento y hacer ajustes si es necesario. Finalmente, la biblioteca es capaz de aplicar los modelos entrenados para realizar predicciones sobre nuevos datos, completando así el ciclo del machine learning.

Entre las principales ventajas de RGSLIB destaca su facilidad de uso, ya que ofrece una interfaz intuitiva que permite a los usuarios centrarse en la construcción y ajuste de modelos sin necesidad de una profunda curva de aprendizaje. Asimismo, se caracteriza por ser una herramienta potente, al incluir una amplia variedad de algoritmos de aprendizaje automático que cubren diversas necesidades. Otra ventaja importante es su flexibilidad, que permite personalizar los modelos y ajustarlos de acuerdo con los requerimientos específicos de cada proyecto. Además, RGSLIB es una biblioteca extensible, lo que significa que puede ser ampliada con nuevas funciones y algoritmos, adaptándose a los avances en machine learning y análisis de datos.

Sin embargo, también se pueden identificar algunas desventajas. En primer lugar, aunque RGSLIB es fácil de usar, en proyectos más complejos, su flexibilidad puede resultar insuficiente en comparación con otras bibliotecas más avanzadas o especializadas en ciertos algoritmos. Además, si bien su extensibilidad es una ventaja, el hecho de tener que incorporar funciones adicionales puede representar una barrera para usuarios sin experiencia en programación avanzada, lo que podría limitar su adopción en ciertos contextos.

2.2.8.2 Algoritmo de segmentación Shepherdseg

El algoritmo de segmentación Shepherd (Shepherdseg) de RSGISLib es una herramienta utilizada en el procesamiento de imágenes geoespaciales, especialmente en el ámbito de la teledetección. Este algoritmo permite la segmentación precisa de imágenes, facilitando la identificación y análisis de diferentes regiones en una imagen satelital. Shepherdseg es un algoritmo de segmentación basado en la agrupación jerárquica y la clasificación por similitud de píxeles. El proceso comienza con una agrupación inicial de los píxeles basada

en ciertos criterios como el color, la textura y la proximidad espacial. Posteriormente, estos grupos se refinan a través de iteraciones en las cuales los segmentos adyacentes similares se fusionan para formar regiones más grandes y homogéneas (Shepherd et al., 2019).

El algoritmo empieza con los píxeles de la imagen que se agrupan en segmentos iniciales basados en un conjunto de criterios predefinidos, después los segmentos iniciales se refinan iterativamente mediante la fusión de segmentos adyacentes que comparten características similares y finalmente el resultado es una imagen segmentada, en formato kea, en regiones homogéneas que pueden ser utilizadas para análisis posteriores. Shepherdseg (2019) destaca el código por su capacidad para manejar grandes volúmenes de datos y por su eficiencia en términos de tiempo de procesamiento. Además, es adaptable a diferentes tipos de imágenes y contextos geoespaciales. Las aplicaciones de Shepherdseg incluyen la clasificación de uso del suelo, la detección de cambios y el monitoreo ambiental.

A continuación, se describe los 3 parámetros más importantes para manipular la segmentación: A continuación, se describe los 3 parámetros más importantes para manipular la segmentación:

2.2.8.2.1 Clúster

El parámetro clúster juega un papel crucial en el algoritmo de segmentación SHEPPARD, específicamente influyendo en el proceso de segmentación de la imagen. Controla el tamaño del clúster, que determina el tamaño mínimo de píxeles conectados que se considerarán como un solo segmento. En otras palabras, este parámetro controla la cantidad de regiones o segmentos en los que se dividirá la imagen.

Según RGSLIB Development Team (2022), el parámetro "clúster" en el algoritmo de segmentación de Shepperd permite controlar el número de clústeres en la segmentación de una imagen ráster. Al ajustar el valor del parámetro "clúster", el algoritmo de segmentación de Shepperd agrupará los píxeles de la imagen en la cantidad especificada de clústers, utilizando una técnica basada en la similitud de los valores de los píxeles. Esto puede ser útil para segmentar una imagen en diferentes clases o categorías que representen características similares del paisaje o del objeto de estudio.

Algunas consideraciones sobre qué puede pasar al cambiar el número de clúster:

- Más clúster: Al aumentar el número de clúster, la imagen se segmentará en más regiones o segmentos más pequeños. Esto puede resultar en una segmentación más detallada de la imagen, lo que permite identificar características más finas del paisaje. Sin embargo, demasiados clústeres pueden llevar a una sobre segmentación, donde se identifican diferencias insignificantes entre píxeles como regiones separadas.
- Menos clúster: Al reducir el número de clúster, la imagen se segmentará en menos regiones o segmentos más grandes. Esto puede simplificar la imagen segmentada y ayudar a identificar patrones generales en el paisaje. Sin embargo, muy pocos clústeres pueden llevar a una subsegmentación, donde características importantes se agrupan en regiones más grandes y se pierden detalles importantes.

2.2.8.2.2 dist_thres

La variable dist_thres es un parámetro que especifica el umbral de distancia para unir los segmentos en el algoritmo de segmentación de Shepherd et al. (2019). Este algoritmo se

utiliza para realizar una segmentación de imágenes basada en la eliminación iterativa de los segmentos más pequeños.

El valor de `dist_thres` depende de la escala y la resolución de la imagen que se quiere segmentar. Si se aumenta el valor de `dist_thres`, se permite que los segmentos se fusionen con otros más lejanos, lo que puede resultar en una segmentación más gruesa y menos detallada. Si se disminuye el valor de `dist_thres`, se restringe la fusión de los segmentos a los más cercanos, lo que puede resultar en una segmentación más fina y detallada.

2.2.8.2.3 `min_n_pxls`

El parámetro "`min_n_pxls`" en el algoritmo "shepherdseg" de la biblioteca RGSLIB de acuerdo con su documentación se refiere al número mínimo de píxeles que debe contener cada segmento o región resultante de la segmentación. Este parámetro controla la sensibilidad del algoritmo para fusionar o dividir regiones durante el proceso de segmentación de una imagen ráster. Al ajustar el valor de "`min_n_pxls`", se puede influir en el tamaño y la forma de los segmentos generados por el algoritmo. Un valor bajo de "`min_n_pxls`" puede resultar en segmentos más pequeños y detallados, lo que puede ser útil para capturar características finas del paisaje. Por otro lado, un valor alto de "`min_n_pxls`" puede generar segmentos más grandes y generalizados, que pueden ser más representativos de patrones globales en la imagen.

Es importante encontrar un equilibrio adecuado al ajustar este parámetro. Un valor demasiado bajo puede provocar una sobresegmentación, donde se identifican detalles insignificantes como regiones separadas. Mientras que un valor demasiado alto puede conducir a una subsegmentación, donde características importantes se agrupan en regiones más grandes y se pierden detalles importantes.

2.2.8.3 *Algoritmo imagefilter. leung_malik_filter_bank*

El algoritmo `imagefilter. leung_malik_filter_bank` de RSGISLib se basa en los filtros de Leung-Malik, una técnica utilizada en la visión por computadora para el análisis de texturas. Es una colección de filtros diseñados para capturar diversas propiedades texturales de las imágenes. Estos filtros son conocidos por su capacidad para capturar tanto texturas finas como gruesas en las imágenes. La técnica fue desarrollada por Leung y Malik en su trabajo sobre reconocimiento de texturas.

De acuerdo con Leung & Malik, J. (2001) los componentes consisten en una colección de diferentes tipos de filtros:

- Filtros de Gabor: Estos filtros son efectivos para capturar las frecuencias espaciales en diferentes orientaciones y escalas.
- Filtros Laplacianos de Gaussianas (LoG): Utilizados para detectar bordes y cambios de intensidad en la imagen.
- Filtros de Derivada de Gaussianas (DoG): Capturan las características de las texturas a diferentes escalas.
- Filtros Gaussianos: Utilizados para suavizar la imagen y reducir el ruido.

Al ejecutar el algoritmo se obtienen 48 texturas en formato kea que contienen la información sobre las diferentes propiedades texturales de la imagen original, estas características pueden incluir contrastes locales, orientaciones predominantes y frecuencias espaciales.

2.2.8.4 *PopulateRATWithStats*

El método `populateRATWithStats` de RSGISLib se utiliza para calcular estadísticas zonales de un ráster de entrada y almacenar estos resultados en una tabla de atributos del ráster (RAT, por sus siglas en inglés). Este proceso es esencial para el análisis de datos

espaciales, ya que permite asociar estadísticas descriptivas con las zonas definidas en el ráster, facilitando así una mejor interpretación y análisis de los datos geoespaciales. (Bunting, 2013).

Segun Shepherd, J., Bunting, P., Dymond, J. (2019), tiene tres funcionalidades principales

- **Cálculo de Estadísticas Zonales:** Este método calcula estadísticas como la media, mediana, desviación estándar, mínimo, máximo y otros valores estadísticos para cada zona definida en el ráster de entrada.
- **Almacenamiento en RAT:** Las estadísticas calculadas se almacenan en una tabla de atributos del ráster (RAT), lo que permite una fácil consulta y uso posterior en análisis y visualizaciones.
- **Integración de Datos:** Facilita la integración de datos ráster con datos vectoriales al proporcionar estadísticas zonales que pueden ser vinculadas a entidades vectoriales o usadas en análisis adicionales.

2.2.8.5 *Scikit-learn*

Según la documentación de Scikit-learn, es una biblioteca de aprendizaje automático de código abierto para el lenguaje de programación Python. Esta biblioteca proporciona una amplia variedad de herramientas y algoritmos para el análisis de datos, la minería de datos y el aprendizaje automático. Scikit-learn es conocida por su facilidad de uso, su eficiencia computacional y su enfoque en la simplicidad y la consistencia.

Algunas de las funcionalidades principales de Scikit-learn incluyen:

- Implementación de algoritmos de aprendizaje supervisado, no supervisado y de aprendizaje por refuerzo.

- Herramientas para preprocesamiento de datos, selección de características y reducción de la dimensionalidad.
- Evaluación de modelos y validación cruzada.
- Integración con otras bibliotecas populares de Python, como NumPy, SciPy y Matplotlib.

2.2.8.6 *Redes neuronales*

El módulo `sklearn.neural_network` de Scikit-learn proporciona herramientas para construir y entrenar redes neuronales artificiales, específicamente Perceptrones Multicapa (MLP), para diversas tareas de clasificación y regresión. El componente clave utilizado es el `MLPClassifier`, donde esta clase es la base para problemas de clasificación. Permite definir la arquitectura de la red con parámetros como la cantidad de capas ocultas, la cantidad de neuronas en cada capa y la función de activación utilizada por las neuronas.

2.2.8.7 *Extra Trees*

El algoritmo Extra Trees (Extremely Randomized Trees) es una técnica de aprendizaje supervisado utilizada principalmente para clasificación y regresión. Su funcionamiento es similar al de los algoritmos de bosque aleatorio (Random Forest), ya que también construye múltiples árboles de decisión y combina sus resultados para mejorar la precisión de la clasificación o predicción. Sin embargo, a diferencia de Random Forest, Extra Trees introduce mayor aleatoriedad en el proceso de creación de árboles, lo que ayuda a reducir el sobreajuste y mejorar la capacidad generalizadora del modelo.

El algoritmo Extra Trees genera múltiples árboles de decisión (generalmente entre 100 y 200 árboles) y los entrena de forma independiente sobre distintos subconjuntos de datos

de entrenamiento, logrando una votación en el resultado final. En lugar de seleccionar los puntos de corte óptimos al dividir los datos en cada nodo, Extra Trees realiza divisiones de manera aleatoria, eligiendo entre un conjunto de posibles características y seleccionando puntos de corte aleatorios para cada división (Geurts et al., 2006). Esta característica es una de las diferencias clave respecto a otros métodos de ensamble, y reduce la varianza del modelo, al costo de introducir un pequeño sesgo.

La principal ventaja de Extra Trees es su eficiencia y rapidez, ya que el proceso de división aleatoria de los nodos reduce el tiempo de entrenamiento en comparación con otros modelos de ensamble. Además, su tendencia a evitar el sobreajuste lo convierte en una opción adecuada para conjuntos de datos grandes y de alta dimensionalidad, como los datos geoespaciales empleados en esta investigación. En esta tesis, se ha elegido Extra Trees para la clasificación de unidades de relieve debido a su capacidad para manejar datos complejos y realizar múltiples clasificaciones con una alta precisión (Ho, 1995; Breiman, 2001). Al entrenar el modelo sobre un conjunto de datos geoespaciales que incluyen información de texturas y valores radiométricos, el algoritmo Extra Trees facilita la identificación y clasificación de 10 tipos de unidades de relieve, mejorando la precisión en la segmentación y clasificación de los datos geográficos de manera automatizada.

2.3 Marco Referencial

Con el objetivo de organizar y facilitar la comprensión de los estudios consultados para esta investigación, se elaboró la Tabla 1. Esta tabla resume los temas principales de los estudios empíricos seleccionados para analizar el estado del arte en relación con la presente investigación. Los estudios se clasifican en tres categorías: análisis de datos en redes sociales, análisis de datos en sitios web y análisis de datos en textos.

Tabla 2

Resumen de estudios consultados para la investigación.

Temas centrales de estudios empíricos	No. de estudio	Autores	Índice	Título de la investigación
Segmentación automatizada de la superficie terrestre en formas geomorfológicas elementales, utilizando principios físicos y técnicas de modelado digital.	1	Popov, A. B., Minár, J., & Drăguț, L. (2024).	2.3.1	Physically-Based Digital Geomorphological Mapping: Case Study of Glacial and Karst Topography
Clasificación del terreno con el algoritmo Douglas-Peucker.	2	Wang, L., Zhang, J., & Liu, Y. (2019).	2.3.2	A Classification Method of Landform Elements Based on Douglas-Peucker Algorithm for Multi-scale Morphology
Muestreo y análisis de la exactitud a través de una matriz de error	3	Álvaro, W. Z. (2022)	2.3.3	"Aproximación a la evaluación de la exactitud temática de la cartografía geomorfológica"
Detección de litología con datos de teledetección.	4	Bachri, A.; El Harti, A.; El Hammichi, F.; El Hammichi, A. (2019)	2.3.4	Algoritmos de Aprendizaje Automático para la Cartografía Litológica Automática Utilizando Datos de Teledetección: Un Estudio de Caso de Souk Arbaa Sahel, Inlier de Sidi Ifni, Anti-Atlas Occidental, Marruecos"
Método HASM-US para analizar y clasificar patrones geomorfológicos	5	Zhang, Du, Fan, Zheng, Luo, Wang, y Wang, 2020	2.3.5	A Geomorphological Regionalization using Upscaled DEM in the Beijing-Tianjin-Hebei Area, China"

Análisis cuantitativo de la superficie terrestre a través de la geomorfometría	6	Amatulli, G. et al. (2020).	2.3.6	Geomorpho90m: An Empirical Evaluation of Global High-Resolution Geomorphometric Information
Susceptibilidad de deslizamientos con Random Forest	7	Liu et al. (2023)	2.3.7	Landslide Susceptibility Zoning in Yunnan Province Based on SBAS-InSAR Technology and a Random Forest Model.
Clasificación del terreno con ANN-RF	8	Eman A. Alshari, Mohammed B. Abdulkareem, and Bharti W. Gawali (2023)	2.3.8	Classification of land use_land cover using artificial intelligence (ANN-RF)"
Clasificación de la cobertura del suelo utilizando la plataforma Google Earth Engine (GEE) y Random Forest	9	Phan, T. N., Kuch, V., & Lehnert, L. W. (2020)	2.3.9	Land Cover Classification using Google Earth Engine and Random Forest Classifier—The Role of Image Composition
Comparación de mapas geomorfológicos	10	van der Meij, M., Meijles, E. W., Marcos, D., Harkema, T. T. L., Candel, J. H. J., & Maas, G. J. (2021)	2.3.10	Comparing geomorphological maps made manually and by deep learning.
Clasificación del terreno con GIS.	11	Robaina, L. E. de S., & Trentin, R. (2020)	2.3.11	Automated classification of landforms with GIS support

Clasificación automatizada de terrenos utilizando modelos de segmentación semántica basados en aprendizaje profundo,	12	Yang, J., Xu, J., & Cheng, W. (2023)	2.3.12	Deep learning-based automated terrain classification using high-resolution digital elevation models. <i>International Journal of Applied Earth Observation and Geoinformation</i> ,
Predicción de índices geomórficos.	13	Souza, C. M. P. et al.2020	2.3.13	Machine Learning Algorithm in the Prediction of Geomorphic Indices for Appraisal the Influence of Landscape Structure on Fluvial Systems, Southeastern - Brazil"
Mapeo con redes neuronales.	14	Mainali, K., Evans, M., Madsen, B., Minnemeyer, S., et al. (2023).	2.3.14	Convolutional neural network for high-resolution wetland mapping with freely available data
Susceptibilidad de deslizamientos por variables topográficas.	15	Chang, K.-T., Merghadi, A., & Yunus, A. P. (2019)	2.3.15	Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques

2.3.1 Physically-Based Digital Geomorphological Mapping: Case Study of Glacial and Karst Topography.

En el presente estudio, se llevó a cabo una investigación en las regiones de los Cárpatos Occidentales, específicamente en áreas de topografía glacial y un plateau kárstico. Este enfoque geográfico es fundamental, ya que estas áreas presentan características geomorfológicas complejas que son esenciales para la comprensión de los procesos geológicos y la formación del relieve. El objetivo general del estudio fue desarrollar un

algoritmo basado en principios físicos para la segmentación automatizada de la superficie terrestre en formas geomorfológicas elementales. Este enfoque busca no solo mejorar la eficiencia del mapeo geomorfológico, sino también proporcionar resultados que sean interpretables desde una perspectiva genética, lo que es crucial para la investigación en geomorfología.

Para alcanzar este objetivo, los autores implementaron un método científico que combina la segmentación de la superficie terrestre con un algoritmo de mínimos cuadrados dinámico (DLS) para la generalización del Modelo Digital de Elevación (DEM). Este método permite ajustar parámetros como el tamaño de la ventana y el grado del polinomio, optimizando así la relación entre errores de datos y errores del método. La segmentación resultante se basa en variables geomorfológicas que están directamente relacionadas con la energía gravitacional, lo que enriquece la interpretación de los segmentos generados. Los resultados obtenidos del estudio demostraron que la aplicación del algoritmo en las áreas seleccionadas produjo mapas geomorfológicos digitalmente detallados y genéticamente interpretables. Los autores lograron identificar y clasificar formas elementales de manera efectiva, lo que resalta la utilidad de la metodología propuesta en la práctica del mapeo geomorfológico.

En conclusión, los autores del estudio enfatizan que la automatización del mapeo geomorfológico no solo mejora la precisión y la eficiencia del proceso, sino que también permite una mejor comprensión de las características del relieve. Esta investigación se relaciona directamente con la presente tesis, que tiene como objetivo crear mapas geomorfológicos de forma automatizada mediante algoritmos de machine learning para la detección de unidades de relieve y sus atributos a partir de diferentes variables geográficas. El presente estudio proporciona una base metodológica sólida que puede ser adaptada y

ampliada en la investigación, especialmente en lo que respecta a la segmentación de la superficie terrestre y la interpretación de las formas del relieve. La integración de técnicas de machine learning con los enfoques presentados en este estudio podría resultar en un avance significativo en la automatización y precisión del mapeo geomorfológico.

2.3.2 A Classification Method of Landform Elements Based on Douglas-Peucker Algorithm for Multi-scale Morphology.

Según Wang et al. (2019), el estudio aborda los desafíos clave en los métodos existentes de clasificación de elementos de relieve, destacando la dificultad para determinar el número de grupos en el espacio de atributos y la falta de consideración de la información de ubicación importante. Se identifican dos enfoques principales: el clustering y la clasificación. El clustering se centra en agrupar atributos topográficos derivados del Modelo Digital de Elevación (DEM), lo que puede resultar en una interpretación geomorfológica poco clara de los resultados. Por otro lado, el enfoque de clasificación implica predefinir un sistema de clasificación de elementos de relieve y luego clasificarlos en función de métodos de razonamiento correspondientes. Estos métodos se dividen en tres tipos, incluyendo aquellos que requieren que el usuario defina explícitamente cada tipo de elemento de relieve, así como aquellos que utilizan la ubicación típica de los elementos de relieve como prototipo para su definición.

El área de estudio está ubicada al sur de la llanura de Weihe, al sureste de la ciudad de Xi'an, con una elevación de 412 metros a 1127 metros y un área de aproximadamente 235 kilómetros cuadrados. El objetivo fue proponer un método de clasificación de elementos de relieve basado en el algoritmo de Douglas-Peucker para la morfología multi-escala. Este enfoque implica clasificar los elementos de relieve en función de la información de elevación local utilizando el algoritmo de Douglas-Peucker en una serie de escalas de análisis

individuales, y luego obtener el resultado de clasificación final a través de una generalización multi-escala. Los resultados experimentales muestran que, en comparación con el método Geomorphons, los resultados de clasificación de los elementos de relieve utilizando el método propuesto son más razonables. Además, se destaca que la generalización multi-escala y el rango de escala de análisis son factores importantes que afectan el resultado final de la clasificación.

El estudio también aborda la evaluación de los resultados de clasificación, reconociendo la dificultad de evaluar cuantitativamente los resultados de clasificación de los elementos de relieve. Se señala que la evaluación principal sigue siendo cualitativa en la actualidad. Se comparan las distribuciones de incertidumbre de los métodos propuestos con los resultados obtenidos por el método Geomorphons, demostrando que el método propuesto puede reflejar de manera más efectiva la relación espacial de los diversos elementos de relieve. Se ofrece una contribución significativa al campo de la geomorfología al abordar las limitaciones de los métodos existentes de clasificación de elementos de relieve y proponer un enfoque mejorado basado en el algoritmo de Douglas-Peucker para la morfología multi-escala.

Los resultados experimentales demostraron que el método propuesto de clasificación de elementos de relieve basado en el algoritmo de Douglas-Peucker produjo resultados más razonables en comparación con el método Geomorphons. Se observó una mejora en la continuidad y la integridad de los elementos de relieve clasificados, lo que reflejó una mejor representación de las características espaciales de los mismos. Wang, L., Zhang, J., & Liu, Y. (2019) concluyeron que el método propuesto era efectivo para abordar las limitaciones de los enfoques existentes en la clasificación de elementos de relieve. Destacaron la importancia de la generalización multi-escala y el uso del algoritmo de Douglas-Peucker

para mejorar la precisión de la clasificación. Además, señalaron que el método propuesto permitía una mejor representación de las características multi-escalares de los elementos de relieve.

El estudio realizado sobre la clasificación de elementos de relieve mediante algoritmos específicos y en diferentes escalas se relaciona estrechamente con la presente investigación, por lo que este estudio también se enfoca en la detección y clasificación automatizada de unidades de relieve a partir de datos geográficos. La aplicación de métodos avanzados como el algoritmo de Douglas-Peucker y la consideración de múltiples escalas en la clasificación pueden ser relevantes para la investigación en la creación automatizada de mapas geomorfológicos.

2.3.3 Aproximación a la evaluación de la exactitud temática de la cartografía geomorfológica

El artículo "Aproximación a la evaluación de la exactitud temática de la cartografía geomorfológica" tiene como objetivo principal diseñar un método de muestreo estadístico óptimo para la evaluación de la exactitud temática de la cartografía geomorfológica a escala 1:25 000, realizar la validación temática a partir de puntos de muestreo y analizar su exactitud mediante una matriz de error con el fin de mejorar la confiabilidad del producto generado mediante el reconocimiento del medio biofísico. La importancia de la geomorfología radica en que es uno de los insumos fundamentales que forma parte de los estudios de levantamientos de suelos a nivel mundial. El proceso de interpretación geomorfológica conlleva la delimitación de unidades visibles en las imágenes satelitales y fotografías aéreas en forma de polígonos, así como la asignación temática de estos según un nivel jerárquico ligado a la escala del estudio, en Putumayo, Colombia en un área de 36km².

La propuesta metodológica comprende dos componentes generales. En primer lugar, se aborda la generación de un script en el programa de cómputo R que posibilita el cálculo del tamaño de la muestra, teniendo en cuenta la información atributiva de las capas vectoriales de la interpretación geomorfológica. En segundo lugar, se destaca la importancia de recrear el ambiente de interpretación bajo el que se desarrollan los procesos y tener en cuenta todos los insumos cartográficos disponibles para el proceso de validación temática.

Se resalta la necesidad de definir una metodología de evaluación de exactitud temática que facilite y agilice los procesos de validación, tomando en cuenta las particularidades de la temática geomorfológica. Se enfatiza la importancia de definir un diseño de muestreo apropiado que proporcione métricas estadísticas y de confiabilidad de los mapas geomorfológicos, permitiendo a los usuarios internos y externos a nivel nacional e internacional valorar su ajuste con la realidad.

Los resultados del estudio mostraron la importancia de la evaluación de la exactitud temática en la cartografía geomorfológica, así como la necesidad de contar con insumos confiables para la generación de mapas de suelos. Los autores concluyeron que la evaluación de la exactitud temática en la cartografía geomorfológica es fundamental para garantizar la confiabilidad de los productos generados. Se resaltó la importancia de definir una metodología de evaluación que facilite los procesos de validación y permita a los usuarios valorar el ajuste de los mapas con la realidad.

El presente estudio tiene relación con la investigación en curso sobre la creación automatizada de mapas geomorfológicos mediante algoritmos de machine learning. Se destaca la importancia de la evaluación de la exactitud temática en la generación de mapas precisos y confiables, lo cual puede ser potenciado con el uso de técnicas automatizadas

como el machine learning para la detección de unidades de relieve y sus atributos a partir de variables geográficas.

2.3.4 Algoritmos de Aprendizaje Automático para la Cartografía Litológica

Automática Utilizando Datos de Teledetección: Un Estudio de Caso de Souk Arbaa Sahel, Inlier de Sidi Ifni, Anti-Atlas Occidental, Marruecos

El artículo "Algoritmos de Aprendizaje Automático para la Cartografía Litológica Automática Utilizando Datos de Teledetección: Un Estudio de Caso de Souk Arbaa Sahel, Inlier de Sidi Ifni, Anti-Atlas Occidental, Marruecos" de Bachri et al. (2019) explora la aplicación de algoritmos de aprendizaje automático (MLAs) para la cartografía litológica automática utilizando datos de teledetección. El estudio se centra en la región de Souk Arbaa Sahel en el Anti-Atlas Occidental, Marruecos, y tiene como objetivo demostrar la eficacia de los MLAs en el análisis de datos de teledetección de alta dimensión para la cartografía litológica.

La metodología de investigación implica el procesamiento de datos de Landsat 8 OLI y su integración en el Modelo Digital de Elevación (DEM) para extraer variables geomorfométricas como la pendiente, la curvatura y la rugosidad de la superficie. Estas variables se utilizan para generar productos con información litológica mejorada. El estudio utiliza específicamente el método de Máquinas de Vectores de Soporte (SVM), un método de clasificación supervisada, para obtener soluciones óptimas para la cartografía litológica. El enfoque SVM se compara con la Red Neuronal Artificial (ANN) y se encuentra que proporciona mejores resultados, con una precisión general del 85% y un coeficiente Kappa del 83,29%.

Los autores enfatizan la importancia de SVM como un algoritmo efectivo para la cartografía litológica utilizando imágenes multiespectrales, destacando su potencial para la cartografía

predictiva remota y la actualización de mapas litológicos existentes. El estudio también aborda los desafíos asociados con las técnicas de teledetección para la cartografía litológica, incluyendo la similitud espectral y la dificultad para diferenciar unidades litológicas basadas en composiciones químicas y mineralógicas

Además, los hallazgos de la investigación sugieren que la metodología utilizada en el estudio se puede aplicar para producir mapas litológicos de áreas remotas y actualizar mapas existentes, ofreciendo una alternativa más rápida a la cartografía geológica tradicional en el campo y resultando en importantes ahorros de tiempo y recursos. Los autores también reconocen el valor de los datos de teledetección en diversas aplicaciones de ciencias de la Tierra, subrayando su importancia como un recurso valioso

En términos del método científico utilizado, los autores emplearon un enfoque basado en el aprendizaje supervisado con el algoritmo de Máquinas de Vectores de Soporte (SVM) para realizar la clasificación automática de litologías en la región de estudio. Se realizaron diversas técnicas de preprocesamiento de datos y procesamiento de imágenes para mejorar la información litológica derivada de los datos de teledetección.

Los resultados obtenidos mostraron una precisión del 85% en la clasificación de mapas litológicos, con un coeficiente Kappa del 83.29%. Esto demostró la eficacia de SVM como algoritmo de aprendizaje supervisado para la cartografía litológica utilizando imágenes multiespectrales. En cuanto a las conclusiones del estudio, los autores destacaron la efectividad de SVM como un algoritmo avanzado para la cartografía predictiva en áreas remotas, así como para la actualización de mapas litológicos existentes. Además, resaltaron el ahorro significativo de tiempo y recursos que ofrece esta técnica en comparación con la cartografía geológica de campo tradicional.

La metodología utilizada por Bachri et al. (2019) con SVM para la clasificación litológica podría adaptarse para la detección de unidades de relieve y sus atributos a partir de variables geográficas en la tesis. La aplicación de técnicas de preprocesamiento de datos y algoritmos de aprendizaje automático, como SVM, podría ser clave para lograr la automatización en la creación de mapas geomorfológicos detallados y precisos.

2.3.5 A Geomorphological Regionalization using Upscaled DEM in the Beijing-Tianjin-Hebei Area, China

Según Zhang et al. (2020), el estudio sobre la regionalización geomorfológica utilizando DEM mejorados en el área de Beijing-Tianjin-Hebei (BTH) en China se enfoca en la aplicación de un método innovador llamado HASM-US para analizar y clasificar patrones geomorfológicos a una escala detallada. Se extrajo la región RA de BTH a partir de datos DEM de resolución de 1 km y se utilizó PCA para identificar los factores topográficos dominantes y clasificar los tipos geomorfológicos en el área de estudio. Los resultados revelaron una clasificación detallada de los patrones geomorfológicos en la región, proporcionando información valiosa para la planificación del uso del suelo y el desarrollo urbano y rural en el área.

El estudio destaca la importancia de considerar múltiples variables topográficas en el análisis geomorfológico para una comprensión completa de la estructura del terreno y sus implicaciones en la planificación regional. Además, se establece una proyección entre cuadrículas gruesas y finas durante el proceso de mejora de datos espaciales, contribuyendo a la precisión de los modelos utilizados en estudios geomorfológicos. Para lograr su objetivo, los autores utilizaron el método HASM-US y realizaron un análisis de componentes principales (PCA) para identificar los factores topográficos dominantes y clasificar los tipos geomorfológicos en el área de estudio. Se extrajo la región RA de BTH a

partir de datos DEM de resolución de 1 km y se proyectaron entre cuadrículas gruesas y finas durante el proceso de mejora de datos espaciales.

Los resultados obtenidos revelaron una clasificación detallada de los patrones geomorfológicos en la región de BTH, lo que proporcionó información valiosa para la planificación del uso del suelo, el desarrollo urbano y rural, y la distribución industrial y agrícola en el área. Los autores concluyeron que considerar múltiples variables topográficas en el análisis geomorfológico es crucial para comprender la estructura del terreno y sus implicaciones en la planificación regional.

En relación con la presente investigación, el estudio de Zhang et al. (2020) destaca la importancia de utilizar métodos innovadores y análisis detallados de variables topográficas para la creación de mapas geomorfológicos automatizados. La aplicación de algoritmos de machine learning en la detección de unidades de relieve y atributos a partir de variables geográficas puede ser una metodología relevante para considerar, siguiendo el enfoque de análisis detallado y proyección entre diferentes resoluciones de datos espaciales presentado en el estudio de Zhang et al. (2020).

2.3.6 Geomorpho90m: An Empirical Evaluation of Global High-Resolution

Geomorphometric Information

Según Amatulli et al. (2020), el artículo "Geomorpho90m: An Empirical Evaluation of Global High-Resolution Geomorphometric Information" presenta el conjunto de datos Geomorpho90m, el cual se basa en el análisis cuantitativo de la superficie terrestre a través de la geomorfometría. Este conjunto de datos proporciona 26 variables geomorfométricas estandarizadas derivadas del Modelo Digital de Elevación (DEM) MERIT, considerado el mejor DEM global de alta resolución disponible. Estas variables describen la tasa de cambio

a lo largo del gradiente de elevación, la rugosidad y las formas geomorfológicas. El estudio se llevó a cabo a nivel global, utilizando procedimientos rigurosos de scripting y un sistema de teselado para permitir el procesamiento multi-núcleo en un superordenador, junto con la proyección Equi7 para minimizar distorsiones a nivel de píxeles.

La geomorfometría es fundamental para abordar diversos problemas geocientíficos a múltiples escalas, y los DEMs son la base para este tipo de análisis del terreno. Los DEMs, ya sea un Modelo Digital de Superficie (DSM) o un Modelo Digital de Terreno (DTM), permiten caracterizar la complejidad topográfica y generar una amplia gama de métricas geomorfométricas. Estas métricas son esenciales para comprender patrones y procesos en disciplinas como la geomorfología, geología, climatología, hidrología y ciencias de la biodiversidad.

El conjunto de datos Geomorpho90m está disponible en resoluciones de 3 (~90 m) y 7.5 arc-segundos (~250 m) bajo el datum geodésico WGS84, y a una resolución espacial de 100 m bajo la proyección Equi7. Estas variables son útiles para aplicaciones de modelado en campos como la geomorfología, geología, hidrología, ecología y biogeografía. La caracterización y evaluación de la superficie terrestre a través de la geomorfometría y la extracción de características son fundamentales para numerosos análisis de modelado ambiental y simulación.

Los resultados obtenidos demostraron una mejora significativa en la resolución espacial y la precisión de las variables geomorfométricas generadas, lo que sugiere una utilidad considerable en diversos campos científicos como la geomorfología, geología, hidrología, ecología y biogeografía. Los autores concluyeron que el conjunto de datos Geomorpho90m ofrece una base sólida para estudios a alta resolución a nivel global, superando a conjuntos de datos anteriores en términos de detalle y precisión. En relación con la investigación sobre

la creación automatizada de mapas geomorfológicos mediante algoritmos de machine learning para la detección de unidades de relieve y atributos geomorfológicos a partir de variables geográficas, este estudio proporciona una metodología avanzada y detallada para la generación de variables geomorfométricas a partir de DEMs a alta resolución.

La utilización de procedimientos de geocomputación y scripting, junto con la comparación con datos de LiDAR y 3DEP, puede servir como referencia para el desarrollo de algoritmos de machine learning. La estandarización y la alta resolución de las variables geomorfométricas presentadas en Geomorpho90m podrían ser fundamentales para la detección automatizada de unidades de relieve y atributos geomorfológicos para el estudio.

2.3.7 Landslide Susceptibility Zoning in Yunnan Province Based on SBAS-InSAR Technology and a Random Forest Model.

El estudio realizado por Liu et al. (2023) se centra en la Zonificación de la Susceptibilidad a Deslizamientos en la Provincia de Yunnan utilizando la tecnología SBAS-InSAR y un modelo de Bosques Aleatorios. La investigación se basa en el análisis de la deformación de la superficie terrestre en toda la provincia de Yunnan mediante la tecnología SBAS-InSAR. Se destaca que el aspecto del terreno es un factor crucial en la evaluación de la susceptibilidad a deslizamientos, ya que diferentes aspectos pueden influir en la incidencia de deslizamientos debido a variaciones en la exposición solar, crecimiento de la vegetación, flujo de agua superficial y temperatura de la superficie.

El objetivo general de este estudio fue evaluar y predecir la susceptibilidad a deslizamientos en la provincia de Yunnan mediante el análisis de la deformación de la superficie terrestre y la combinación de diversos factores geográficos con el modelo de Bosques Aleatorios. Para lograr este objetivo, se utilizó un enfoque científico que integró la tecnología SBAS-

InSAR para medir la deformación del terreno y el modelo de Bosques Aleatorios para evaluar la susceptibilidad a deslizamientos. Se realizaron análisis detallados de aspectos topográficos, clasificación litológica, estructura geológica, precipitación, entre otros factores, para identificar áreas de alto riesgo de deslizamientos.

El análisis de aspectos se realizó utilizando el software ArcGIS, dividiendo la dirección de pendiente en nueve categorías: plano, norte, noreste, este, sureste, sur, suroeste, oeste y noroeste. Este enfoque permitió obtener una distribución de aspectos en la provincia de Yunnan, lo que resulta fundamental para comprender la dinámica del terreno y su influencia en la susceptibilidad a deslizamientos.

Además, se destaca la importancia del modelo de Bosques Aleatorios para mejorar la precisión y confiabilidad de la cartografía de susceptibilidad a deslizamientos. Este modelo se basa en técnicas de aprendizaje automático y ha demostrado ser efectivo en la evaluación de riesgos naturales, como deslizamientos de tierra. La combinación de la tecnología SBAS-InSAR y el modelo de Bosques Aleatorios proporciona una metodología integral para identificar áreas de alto riesgo de deslizamientos en la provincia de Yunnan.

Los resultados obtenidos revelaron que la combinación de la tecnología SBAS-InSAR y el modelo de Bosques Aleatorios permitió una evaluación precisa de la susceptibilidad a deslizamientos en Yunnan. Se identificaron áreas críticas con alta susceptibilidad a deslizamientos, lo que proporciona información valiosa para la gestión del riesgo de desastres naturales en la región. En conclusión, los autores destacaron la eficacia del modelo de Bosques Aleatorios en la evaluación de la susceptibilidad a deslizamientos y la importancia de considerar la deformación del terreno como un factor clave en la predicción de deslizamientos. Además, resaltaron la relevancia de los resultados para la planificación y la toma de decisiones en la prevención de deslizamientos en Yunnan.

Este estudio se relaciona con la presente investigación, que tiene como objetivo crear mapas geomorfológicos de forma automatizada mediante algoritmos de machine learning para la detección de unidades de relieve y sus atributos a partir de diferentes variables geográficas. La metodología utilizada en el estudio de zonificación de la susceptibilidad a deslizamientos en Yunnan, especialmente el enfoque de los Bosques Aleatorios y el análisis de múltiples variables geográficas puede servir como referencia para el desarrollo de algoritmos de machine learning.

2.3.8 Classification of land use_land cover using artificial intelligence (ANN-RF)

El estudio "Classification of land use_land cover using artificial intelligence (ANN-RF)" se enfoca en la implementación de una metodología híbrida que combina Redes Neuronales Artificiales (ANN) y Bosques Aleatorios (RF) para la clasificación de uso del suelo y cobertura terrestre. Esta combinación de enfoques de inteligencia artificial tiene como objetivo mejorar la precisión y eficiencia en la clasificación de imágenes satelitales para identificar distintas categorías de uso del suelo y cobertura terrestre en un área específica.

El objetivo general de la investigación fue mejorar la precisión y eficiencia en la clasificación de imágenes satelitales para identificar diferentes categorías de uso del suelo y cobertura terrestre en un área específica en Sana, Yemen. Para lograr este objetivo, se aplicó un enfoque científico que integró técnicas de inteligencia artificial, específicamente ANN y RF, para la clasificación automatizada de las imágenes satelitales. Los resultados obtenidos demostraron que esta metodología híbrida permitió una clasificación más detallada y precisa de las categorías de uso del suelo y cobertura terrestre, superando a métodos tradicionales en términos de eficacia y exactitud. De acuerdo con los autores del estudio, la fusión de las técnicas de ANN y RF permite una clasificación más detallada y automatizada de las imágenes de satélite, lo que facilita la identificación de patrones complejos de uso

del suelo y cobertura terrestre. Esta metodología avanzada ofrece ventajas significativas en términos de precisión y rapidez en comparación con enfoques tradicionales de clasificación.

Los autores concluyeron que la combinación de ANN y RF en la clasificación de uso del suelo y cobertura terrestre representa un avance significativo en la investigación geoespacial, con aplicaciones prácticas en la planificación urbana, la gestión ambiental y la monitorización de cambios en el paisaje. Recomiendan continuar explorando el desarrollo de enfoques de aprendizaje automático para mejorar la precisión de las clasificaciones y la detección de cambios en el entorno. En relación con la investigación el presente estudio sugiere que la combinación de técnicas de inteligencia artificial como ANN y RF podría ser una metodología prometedora para considerar. Esta aproximación podría facilitar la identificación y clasificación automatizada de unidades geomorfológicas a partir de datos geoespaciales, mejorando la eficiencia y precisión en la generación de mapas geomorfológicos detallados y actualizados.

2.3.9 Land Cover Classification using Google Earth Engine and Random Forest Classifier—The Role of Image Composition"

En el presente estudio, se llevó a cabo una investigación sobre la clasificación de la cobertura del suelo utilizando la plataforma Google Earth Engine (GEE) y el clasificador de Bosques Aleatorios. La investigación se centró en una región específica de Mongolia, un área caracterizada por su diversidad de ecosistemas y desafíos en la gestión de recursos naturales. El objetivo general del estudio fue analizar el efecto de diferentes métodos de composición de imágenes y la selección de imágenes de entrada en los resultados de clasificación de la cobertura del suelo. Este enfoque es fundamental, ya que la precisión de la información sobre la cobertura del suelo tiene implicaciones significativas en diversas aplicaciones, desde la gestión ambiental hasta la planificación urbana.

Phan, Kuch, & Lehnert, (2020) utilizaron un método científico que incluyó la recopilación de datos de reflectancia de superficie del satélite Landsat 8. Se implementaron ocho estrategias de combinación de imágenes para producir y evaluar mapas de cobertura del suelo. La metodología se realizó en la plataforma GEE, donde se aplicó el clasificador de Bosques Aleatorios, un algoritmo ampliamente utilizado en estudios de clasificación de imágenes. Este enfoque permitió a los investigadores analizar grandes volúmenes de datos de manera eficiente y efectiva, lo que es un aspecto crucial para la investigación en áreas extensas.

Los resultados obtenidos mostraron que todas las ocho combinaciones de datos produjeron mapas de cobertura del suelo con una precisión moderada a alta, alcanzando una precisión general superior al 84.31%. Entre las combinaciones, dos conjuntos de datos de series temporales de escenas de verano lograron las mayores precisiones, con valores de 89.80% y 89.70%. Además, se observó que la agregación temporal, como el uso de la mediana, no solo reduce significativamente el volumen de datos, sino que también produce una alta precisión en la clasificación. Las conclusiones a las que llegaron los autores subrayan la importancia de la selección de imágenes y la composición de datos en la clasificación de la cobertura del suelo. Se destacó que la consistencia espacial entre los resultados de clasificación era relativamente baja en comparación con la alta precisión general, lo que indica que la elección del conjunto de datos es un paso crucial en cualquier clasificación realizada en GEE.

Este estudio tiene una relación directa con la presente investigación, que busca crear mapas geomorfológicos de forma automatizada mediante algoritmos de machine learning para la detección de unidades de relieve y sus atributos a partir de diferentes variables geográficas. La metodología empleada en el estudio de Phan et al. (2020) puede ser

adaptada y aplicada en mi investigación, especialmente en lo que respecta a la utilización de GEE para el procesamiento de datos geoespaciales y la implementación de algoritmos de clasificación como Bosques Aleatorios. Al igual que en el estudio mencionado, la selección adecuada de imágenes y la composición de datos serán fundamentales para garantizar la precisión y la utilidad de los mapas geomorfológicos generados en esta investigación. En resumen, la investigación de Phan et al. proporciona un marco valioso que puede ser utilizado para mejorar la metodología en la creación de mapas geomorfológicos, destacando la importancia de la calidad de los datos y la elección de algoritmos en el análisis de variables geográficas.

2.3.10 Comparing geomorphological maps made manually and by deep learning.

El presente estudio se llevó a cabo en el distrito central de morrenas empujadas en los Países Bajos, un área con relieve pronunciado y una amplia variedad de formas de relieve. El objetivo general de la investigación fue comparar los mapas geomorfológicos creados manualmente con aquellos generados mediante técnicas de aprendizaje profundo, específicamente redes neuronales convolucionales (CNN). Para lograr este objetivo, los autores utilizaron un enfoque científico que implicaba la cuantificación de la incertidumbre asociada con la cartografía geomorfológica manual, la exploración de las capacidades de las CNN para la cartografía semiautomatizada y la evaluación de la sensibilidad de estas redes a las incertidumbres en los datos de evaluación creados manualmente. Se realizaron comparaciones detalladas de los resultados obtenidos a partir de la cartografía manual y la cartografía computacional.

Según van der Meij et al. (2021), la geomorfología es un campo crucial que proporciona información detallada sobre el relieve, la génesis y la forma de la superficie terrestre, siendo fundamental para el desarrollo espacial sostenible. Sin embargo, la calidad de los mapas

geomorfológicos rara vez se evalúa o se informa, lo que limita su aplicabilidad. Además, los mapas geomorfológicos más antiguos a menudo no cumplen con los requisitos de calidad actuales y requieren actualización. Este proceso de actualización es laborioso y, debido a su naturaleza cualitativa, difícil de reproducir, pero puede ser respaldado por métodos computacionales novedosos.

El estudio aborda estas cuestiones mediante la cuantificación de la incertidumbre asociada con la cartografía geomorfológica manual, la exploración del uso de redes neuronales convolucionales (CNN) para la cartografía geomorfológica semiautomatizada y la evaluación de la sensibilidad de las CNN a las incertidumbres en los datos de evaluación creados manualmente. Se seleccionó un área de prueba en el distrito de morrenas empujadas de los Países Bajos, que presenta un relieve pronunciado y una amplia variedad de formas de relieve.

Se comparó los resultados de la cartografía manual de cinco geomorfólogos diferentes en un entorno de estudio de caso y llevó a cabo un procedimiento de cartografía semiautomatizada en la misma área. El distrito central de morrenas empujadas en los Países Bajos ofrece un área de estudio adecuada, dada la presencia de formas de relieve distintivas y la amplia disponibilidad de Modelos Digitales de Elevación (DEM) de alta resolución y datos subsuperficiales, así como una rica historia de cartografía. Los resultados del estudio mostraron diferencias significativas entre los mapas geomorfológicos creados manualmente y los generados por las CNN. Se observaron discrepancias en la identificación de unidades de relieve y sus atributos, lo que resaltó la importancia de evaluar la calidad de los mapas geomorfológicos y la viabilidad de utilizar técnicas de aprendizaje profundo para mejorar la precisión y eficiencia de la cartografía.

Como conclusión, los autores destacaron la necesidad de actualizar y mejorar continuamente los mapas geomorfológicos para su aplicación en el desarrollo espacial sostenible y la gestión ambiental. Además, resaltaron la importancia de considerar la incertidumbre asociada con la cartografía manual y la relevancia de explorar enfoques computacionales avanzados para la creación de mapas geomorfológicos.

Este estudio presenta similitudes con la tesis, con el enfoque de las redes neuronales convolucionales para la cartografía geomorfológica, puede ser una referencia valiosa para el desarrollo de mi investigación. La aplicación de técnicas de aprendizaje profundo para la detección de unidades de relieve y sus atributos a partir de diferentes variables geográficas puede ser una estrategia efectiva para mejorar la precisión y eficiencia en la creación de mapas geomorfológicos automatizados.

2.3.11 Automated classification of landforms with GIS support

Según Robaina and Trentin (2020), en su artículo "Automated classification of landforms with GIS support", presentan una metodología para la clasificación automatizada del relieve en tres niveles taxonómicos utilizando procesamiento automatizado en un entorno de SIG. El estudio se llevó a cabo en la cuenca del río Santa María en Brasil, demostrando el avance acelerado de las geotecnologías en las metodologías de clasificación del relieve.

La investigación se enfoca en definir parámetros clave que determinan porciones de la superficie con características homogéneas en términos de formas y elementos del relieve. Utilizando datos del Modelo Digital de Elevación del Radar SRTM de 3 arc-second (90 metros) y el software ArcGIS®, se identificaron cuatro formas de relieve en el primer nivel taxonómico: áreas planas, colinas levemente onduladas, colinas onduladas y asociación de morros y morrotes. En el segundo nivel, se definieron diez elementos de relieve, como

picos, crestas, resaltes, entre otros. Finalmente, en el tercer nivel, se clasificaron las formas de las vertientes en ocho unidades basadas en parámetros de declive, perfil y curvatura.

La aplicación de la metodología en un entorno de SIG permitió establecer de manera rápida y precisa las formas de relieve, los elementos del relieve y las formas de las vertientes, destacando la relación entre los tres niveles taxonómicos. Este enfoque automatizado y basado en GIS ofrece una herramienta eficaz para estudios relacionados con la clasificación del relieve, brindando una mayor comprensión de la morfología del terreno y su variabilidad en diferentes escalas geográficas. Se logró una representación detallada y estructurada de las características del relieve en la cuenca del río Santa María. En cuanto a las conclusiones de los autores, destacaron la eficacia del procesamiento automatizado en un entorno de SIG para la clasificación del relieve, resaltando la importancia de las geotecnologías en este tipo de estudios.

La aplicación de algoritmos de machine learning en combinación con variables geográficas podría potenciar la detección de unidades de relieve y sus atributos de manera más eficiente y precisa. En este sentido, la metodología presentada en el estudio anteriormente mencionado podría servir como referencia y guía para el desarrollo de la investigación propuesta, permitiendo la automatización y optimización del proceso de creación de mapas geomorfológicos mediante el uso de algoritmos de machine learning y variables geográficas relevantes.

2.3.12 Deep learning-based automated terrain classification using high-resolution digital elevation models.

En el presente estudio, se llevó a cabo una investigación en la región del Altiplano Qinghai-Tibet, un área caracterizada por su compleja geomorfología y variaciones altitudinales

significativas. El objetivo general de este estudio fue desarrollar un modelo de segmentación semántica basado en aprendizaje profundo para la clasificación automatizada de diferentes tipos de relieve, utilizando datos de modelos digitales de elevación (DEM) de alta resolución.

Para alcanzar este objetivo, los autores implementaron un enfoque metodológico que combina técnicas de aprendizaje profundo, específicamente la arquitectura FCN-ResNet, con datos DEM de AW3D30. Este método permite la clasificación precisa de 25 tipos geomorfológicos terrestres, basándose en la combinación de tipos de altitud y relieve. La metodología se centra en la segmentación de imágenes, donde se asignan categorías a cada píxel de la imagen, lo que resulta en una representación detallada de las unidades de relieve.

Los resultados obtenidos mostraron que el modelo de segmentación semántica logró una precisión general aceptable, con un rendimiento de 75.24% en precisión general (PA) y 63.15% en la intersección sobre la unión media (MIoU). A pesar de las dificultades inherentes en la identificación de áreas pequeñas y poco estables, como las plataformas en el desierto, el modelo demostró ser más preciso que los mapas de relieve tradicionales, lo que sugiere una mejora significativa en la clasificación de terrenos.

Las conclusiones a las que llegaron los autores destacan la efectividad del uso de algoritmos de aprendizaje profundo para la clasificación de terrenos, subrayando que, a pesar de las transiciones difusas entre diferentes tipos de relieve, el modelo mostró una alta certeza en sus predicciones. Esto abre nuevas posibilidades para la automatización en la creación de mapas geomorfológicos, lo que es directamente relevante para mi investigación.

El presente estudio tiene relación con la presente investigación, que busca crear mapas geomorfológicos de forma automatizada mediante algoritmos de machine learning para la detección de unidades de relieve y sus atributos a partir de diferentes variables geográficas. La metodología empleada en este estudio, que incluye el uso de modelos de segmentación semántica y datos DEM, puede ser adaptada y aplicada en mi investigación para mejorar la precisión y eficiencia en la clasificación de unidades de relieve. Esto no solo contribuirá a la automatización del proceso, sino que también permitirá una mejor comprensión de las características geomorfológicas de las áreas de estudio.

2.3.13 Machine Learning Algorithm in the Prediction of Geomorphic Indices for Appraisal the Influence of Landscape Structure on Fluvial Systems, Southeastern – Brazil

De acuerdo con Souza et al. (2020), en el artículo "Machine Learning Algorithm in the Prediction of Geomorphic Indices for Appraisal the Influence of Landscape Structure on Fluvial Systems, Southeastern - Brazil", se aborda el uso de algoritmos de aprendizaje de máquina para predecir índices geomórficos y evaluar la influencia de la estructura del paisaje en sistemas fluviales en el sureste de Brasil, específicamente en la cuenca hidrográfica de Abaeté. La investigación se enfoca en la aplicación de técnicas de aprendizaje de máquina, específicamente el algoritmo Random Forest, para predecir índices geomórficos como el Factor de Asimetría de Cuenca, el Índice de Longitud de Corriente-Gradiente (SL), y el Índice de Pendiente del Canal (k_{sn}). Se llevó a cabo un análisis detallado de subcuencas, utilizando covariables importantes para mejorar la precisión de las predicciones.

Los resultados del estudio indican que el algoritmo Random Forest fue eficiente en la predicción de los índices geomórficos, con un rendimiento destacado en la predicción del

índice k_{sn} . Se observó que las áreas con valores más altos de los índices geomórficos se encontraban en zonas de contacto litológico con diferentes resistencias, lo que favorecía cambios bruscos de pendiente en los canales (knickpoints). Además, se identificó que factores estructurales condicionan los sistemas fluviales en la región estudiada.

Este enfoque de aplicación de algoritmos de aprendizaje de máquina en la predicción de índices geomórficos para evaluar la influencia de la estructura del paisaje en sistemas fluviales en el sureste de Brasil representa una contribución significativa al campo de la geomorfología y proporciona información valiosa para comprender la dinámica de los sistemas fluviales en esta región. En conclusión, los autores destacaron que los factores estructurales condicionan los sistemas fluviales de la cuenca de Abaeté, y el uso de algoritmos de Machine Learning como Random Forest puede ser una herramienta eficaz para la predicción de índices geomorfológicos y la evaluación de la influencia de la estructura del paisaje en los sistemas fluviales.

La metodología utilizada en este estudio, especialmente el enfoque de Machine Learning para la predicción de índices geomorfológicos, puede ser de gran relevancia para el desarrollo del estudio, ya que demuestra la eficacia de estos algoritmos en la caracterización del paisaje y la detección de unidades geomorfológicas de manera automatizada y precisa.

2.3.14 Convolutional neural network for high-resolution wetland mapping with freely available data

El presente estudio de Mainali et al. (2023), se llevó a cabo en un entorno de mapeo de humedales a alta resolución utilizando redes neuronales convolucionales y datos abiertos donde el área de estudio consta de tres condados Delaware, Nueva York y Minnesota, EE.

UU. Estos tres sitios comprenden una amplia gama de características geofísicas y diferentes tipos de humedales. Un cuarto sitio en Lincoln, Nebraska se utilizó para estimar el modelo transferibilidad.

El objetivo general de la investigación fue desarrollar un modelo preciso y generalizable para la cartografía de humedales a partir de datos gratuitos. Para lograr este objetivo, se empleó un enfoque científico basado en el uso de un modelo de aprendizaje profundo, específicamente una arquitectura U-Net, para mapear humedales a una resolución de 1 metro.

Conforme a Mainali et al. (2023), se exploró el uso de redes neuronales convolucionales para mapear humedales a alta resolución donde el modelo completo logró mapear con precisión los humedales (con una precisión del 94 %, una precisión del 96.5 % y un AUC del 95.2 %) a una resolución de 1 metro. Se observó que el modelo pudo predecir correctamente los humedales incluso en áreas con datos de etiquetado/entrenamientos incorrectos, lo que penalizó la tasa de recuperación. Sin embargo, al aplicar el modelo en una nueva geografía, se obtuvo un rendimiento deficiente, con una precisión de aproximadamente el 80 % y una tasa de recuperación del 48 %.

Se destacó la importancia del reentrenamiento limitado en esta nueva geografía, lo que mejoró sustancialmente el rendimiento del modelo, demostrando así un medio efectivo para crear un modelo espacialmente generalizable. Se demostró que es posible mapear humedales a alta resolución (1 m) utilizando datos gratuitos y modelos eficientes de aprendizaje profundo que no requieren ingeniería manual de características. La inclusión de datos LiDAR y geomorphons como datos de entrada mejoró la precisión del modelo en un 2 %, y en casos donde estos datos no están disponibles, un modelo más simple puede mapear eficientemente los humedales.

Se distingue la importancia de la diversidad geográfica en los datos de entrenamiento para crear un modelo espacialmente generalizable. Se observó que un modelo entrenado en una geografía puede no tener la misma eficacia en una geografía diferente debido a patrones de datos únicos en geografías específicas. Se encontró que el ajuste fino del modelo utilizando datos de entrenamiento de la nueva geografía mejoró significativamente el rendimiento del modelo, lo que indica la necesidad de un ajuste fino en nuevas geografías para lograr una generalización espacial efectiva. Finalmente destaca la viabilidad de utilizar redes neuronales convolucionales para mapear humedales a alta resolución, incluso sin la necesidad de ingeniería manual de características. Se enfatiza la importancia de la inclusión de datos LiDAR y geomorphons para mejorar la precisión del modelo, así como la necesidad de ajuste fino en nuevas geografías para lograr una generalización espacial efectiva en la cartografía de humedales.

En cuanto a los resultados obtenidos, el modelo logró una precisión del 94 %, una precisión del 96.5 % y un AUC del 95.2 % en la cartografía de humedales a alta resolución. Se observó que la inclusión de datos LiDAR y geomorphons como variables de entrada mejoró la precisión del modelo en un 2 %. Sin embargo, al aplicar el modelo en una nueva geografía, se encontró un rendimiento inicial deficiente que mejoró significativamente con un reentrenamiento limitado en esa geografía específica. Los autores concluyeron que es factible mapear humedales a alta resolución utilizando datos gratuitos y modelos de aprendizaje profundo eficientes. Destacaron la importancia de la diversidad geográfica en los datos de entrenamiento para lograr un modelo espacialmente generalizable. Además, resaltaron la necesidad de ajuste fino en nuevas geografías para garantizar una generalización espacial efectiva en la cartografía de humedales.

La utilización de redes neuronales convolucionales para la detección de patrones complejos en datos geográficos sin la necesidad de ingeniería manual de características es un enfoque prometedor. La inclusión de variables geográficas relevantes, como LiDAR y geomorphons, podría mejorar la precisión y generalización de los modelos en la detección de unidades de relieve y sus atributos. Por lo tanto, la metodología empleada en este estudio podría servir como referencia para el desarrollo de algoritmos de machine learning en la creación automatizada de mapas geomorfológicos en la presente investigación.

2.3.15 Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques

De acuerdo con Chang et al. (2019), en el estudio titulado "Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques", se investigó la influencia de las variables topográficas en los modelos de susceptibilidad a deslizamientos, centrándose en la resolución de los modelos digitales de elevación (DEM) y su impacto en la precisión de la cartografía de susceptibilidad a deslizamientos. El área de estudio, ubicada en Taiwán, se caracteriza por pendientes pronunciadas con flujos de detritos y deslizamientos frecuentes durante las temporadas de tifones, Taiwán tiene una superficie de 36.000 m², el 26,68% de la cual está cubierta por llanuras, mientras que el 27,31% es montañoso y el 46,01% montañoso

El objetivo general de este estudio fue determinar cómo la resolución de los modelos digitales de elevación (DEM) afecta la precisión de la cartografía de susceptibilidad a deslizamientos. Para lograr este objetivo, se empleó un enfoque científico que incluyó la comparación de diferentes resoluciones de DEM, la generación de factores de condición a partir de los DEM y datos complementarios, y la implementación de técnicas de aprendizaje

automático como regresión logística, bosques aleatorios y máquinas de vectores de soporte.

Este trabajo utilizó datos LiDAR DEM de alta resolución para representaciones topográficas detalladas y comparó la efectividad de diferentes resoluciones de DEM en los modelos de susceptibilidad a deslizamientos. Los resultados obtenidos destacaron que un DEM de alta resolución no necesariamente produce un Modelo de Susceptibilidad a Deslizamientos (LSM) preciso, ya que puede contener información excesiva que afecta negativamente la precisión del modelo. Se encontró que la resolución del DEM afecta significativamente los resultados de los modelos, y se demostró que diferentes escalas de DEM introducen variaciones en los mismos modelos.

En particular, se observó que un DEM de 30 metros de resolución, que representa una topografía precisa, podría ser adecuado para los LSM al producir niveles decentes de generalización topográfica. Por el contrario, los DEM de alta resolución introducen más ruido, lo que afecta negativamente el rendimiento del modelo. A pesar de que los DEM de alta resolución (LiDAR de 5 metros) teóricamente deberían producir mejores modelos, en la práctica, el exceso de datos se consideró ruido y empeoró los modelos resultantes en lugar de mejorarlos.

Los autores concluyeron que la resolución inapropiada del DEM puede afectar significativamente la precisión de los modelos de susceptibilidad a deslizamientos, y que la resolución de 30 metros demostró ser más valiosa que la de 5 metros para la derivación de factores de condición a partir de DEM. Además, se destacó la importancia de seleccionar la resolución adecuada del DEM para mapear la susceptibilidad a deslizamientos en áreas vulnerables, con el objetivo de mejorar la gestión y mitigación de desastres naturales.

En otras palabras, se proporciona información relevante sobre la influencia de la resolución del DEM en los modelos de susceptibilidad a deslizamientos, destacando la importancia de considerar cuidadosamente la escala de las variables topográficas en la cartografía de riesgos naturales. En relación con la presente investigación, el estudio tiene relevancia al demostrar la importancia de considerar la escala de las variables topográficas en la generación de modelos de riesgos naturales. La metodología utilizada, que combina el análisis de diferentes resoluciones de DEM y el uso de técnicas de aprendizaje automático, puede ser aplicada para desarrollar algoritmos que permitan detectar unidades de relieve y atributos geomorfológicos a partir de variables geográficas diversas.

2.4 Marco legal

En este apartado se aborda el marco legal que sustenta la generación, publicación y uso de la información geográfica, con especial atención a la normativa nacional e internacional que regula estas actividades. En México, el INEGI (Instituto Nacional de Estadística y Geografía) es la entidad encargada de coordinar y garantizar la calidad de los datos geoespaciales, en cumplimiento de la Ley del Sistema Nacional de Información Estadística y Geográfica (LSNIEG). Esta ley es fundamental para la creación de productos cartográficos, como los mapas geomorfológicos utilizados en este trabajo, ya que establece los principios bajo los cuales se organiza y difunde la información geográfica.

A nivel internacional, se reconocen normativas como las Directrices del Open Geospatial Consortium (OGC) y la Directiva INSPIRE de la Unión Europea, que, aunque no tienen aplicación directa en México, influyen en la interoperabilidad y estandarización de los datos geoespaciales. Estas normativas aseguran que la información generada cumpla con los más altos estándares de precisión, accesibilidad y utilidad pública, principios que también

guían el trabajo del INEGI y su impacto en la planificación territorial y la gestión de los recursos naturales en el país. Las siguientes normas y/o leyes son:

- Ley del Sistema Nacional de Información Estadística y Geográfica (LSNIEG): Esta ley establece las bases para la organización y funcionamiento del Sistema Nacional de Información Estadística y Geográfica en México. Regula las actividades del INEGI, asegurando que la información geográfica producida sea accesible y esté actualizada. Es fundamental para la difusión de datos geoespaciales, garantizando su calidad y utilidad para la toma de decisiones gubernamentales, académicas y empresariales.
 - Artículo 77 de la LSNIEG: Señala la obligación del INEGI de coordinar la información geográfica y garantizar su acceso público.
- Norma Técnica sobre Información Geográfica del INEGI: Establece las especificaciones para la producción y difusión de información geográfica oficial, como los Modelos Digitales de Elevación (MDE), imágenes satelitales y otros datos geoespaciales, que sirven como base para generar productos como los mapas geomorfológicos.
- Ley General del Equilibrio Ecológico y la Protección al Ambiente (LGEEPA): Aunque su enfoque principal es el medio ambiente, incluye disposiciones sobre la planificación territorial y la gestión de los recursos naturales que requieren el uso de información geográfica precisa, para lo cual se utilizan los mapas geomorfológicos.
- Directrices Internacionales de la Open Geospatial Consortium (OGC): Esta organización internacional establece estándares para la interoperabilidad de los datos geoespaciales. Las normativas de OGC son ampliamente adoptadas en todo

el mundo, y en México, influyen en la forma en que el INEGI y otras instituciones publican sus datos geográficos, garantizando su compatibilidad y accesibilidad.

- Directiva Europea INSPIRE (Infrastructure for Spatial Information in the European Community): Aunque no es aplicable directamente en México, INSPIRE es un marco importante a nivel internacional, que establece normas para compartir información geoespacial en Europa. Ofrece una referencia sobre cómo los datos geográficos pueden publicarse y armonizarse a nivel internacional.

El marco legal descrito es crucial para el desarrollo y la validación de esta tesis, ya que proporciona un sustento normativo para la generación y uso de información geoespacial. La Ley del Sistema Nacional de Información Estadística y Geográfica (LSNIEG), junto con la Norma Técnica sobre Información Geográfica del INEGI, garantiza que los datos empleados en este estudio, como los modelos digitales de elevación y las imágenes satelitales, cumplan con altos estándares de calidad, accesibilidad y precisión. Estas leyes no solo permiten la generación de productos cartográficos, como los mapas geomorfológicos, sino que también aseguran su aplicabilidad en la planificación territorial y la gestión de recursos naturales.

A nivel internacional, directrices como las del Open Geospatial Consortium (OGC) y la Directiva INSPIRE han influido en la interoperabilidad y estandarización de los datos geoespaciales, facilitando su integración y uso en proyectos como este. Esto resalta la relevancia de las normativas internacionales como referencia para mejorar la accesibilidad y compatibilidad de la información geográfica generada en México. En resumen, el marco legal asegura que los resultados de esta tesis sean confiables, precisos y útiles, respaldando su contribución a la planificación territorial, la gestión ambiental y la toma de

decisiones fundamentadas, asegurando que con estos instrumentos la tesis cumpla con los estándares de precisión, accesibilidad y utilidad pública.

El Marco Conceptual ha permitido establecer los fundamentos teóricos y metodológicos sobre los cuales se sustenta este estudio, en el cual se integran conceptos clave de la geomorfología y el machine learning. La revisión exhaustiva de la literatura permitió identificar las intersecciones entre estas disciplinas, demostrando que los enfoques automatizados para la creación de mapas geomorfológicos representan una solución innovadora frente a los métodos tradicionales manuales, que son más costosos y laboriosos.

Los datos geoespaciales, como las imágenes satelitales y los modelos digitales de elevación, han sido resaltados como elementos cruciales en la automatización de la cartografía geomorfológica, ya que aportan la información necesaria para el análisis del relieve y su clasificación. A través del uso de técnicas de machine learning, como redes neuronales, k-means y árboles de decisión, se ha establecido la viabilidad de crear mapas geomorfológicos con una precisión y eficiencia notable, lo cual facilita su aplicación en estudios a gran escala y en contextos donde el acceso a la mano de obra especializada es limitado.

En conclusión, el capítulo ha cimentado la comprensión teórica necesaria para el desarrollo de este estudio, resaltando el potencial de los algoritmos de machine learning como una herramienta poderosa para transformar la forma en que se generan los mapas geomorfológicos. Asimismo, se han identificado tanto las oportunidades como los desafíos que estas técnicas presentan, sentando las bases para la experimentación y validación en los capítulos siguientes.

CAPÍTULO III. MATERIALES Y MÉTODOS

En este apartado se delimitan los materiales y métodos empleados en la realización del estudio, detallando la propuesta de solución a la problemática central y los objetivos iniciales establecidos. Se ofrece una descripción exhaustiva sobre la población de estudio, la selección de la muestra y las técnicas utilizadas para la construcción del modelo destinado a la identificación y clasificación de los distintos tipos de relieve en la elaboración de mapas geomorfológicos. Asimismo, se expone el enfoque metodológico basado en el empleo de técnicas de Machine Learning, destacando el diseño del modelo de clasificación y las etapas de desarrollo necesarias para llevar a cabo una solución eficaz y precisa en la generación automatizada de mapas geomorfológicos.

3.1 Objetivos

3.1.1 Objetivo general

Crear mapas geomorfológicos de forma automatizada mediante algoritmos de machine learning para la detección de unidades de relieve y sus atributos.

3.1.2 Objetivos específicos

Los objetivos específicos son los siguientes:

- Seleccionar y preparar los datos de entrada, tales como imágenes satelitales, modelo digital de elevación y datos geoespaciales, mediante técnicas de preprocesamiento y limpieza de datos, para crear un conjunto de datos adecuado para el entrenamiento y evaluación del algoritmo de machine learning.
- Implementar el algoritmo de machine learning Extra Trees, a partir de los datos seleccionados y procesados para automatizar la creación de mapas geomorfológicos con sus unidades de relieve.

- Analizar y discutir la eficiencia obtenida en el modelo de machine learning, mediante métricas de evaluación para identificar limitaciones y retos en la generación de mapas geomorfológicos y sus unidades de relieve.

3.2 Participantes y criterios de selección

3.2.1 Participantes (muestra geográfica)

En esta investigación, los participantes corresponden a las muestras de datos geoespaciales y archivos utilizados para generar y entrenar los modelos geomorfológicos, así como para la validación de los resultados obtenidos mediante técnicas de machine learning. A continuación, se describen los principales conjuntos de datos y herramientas empleadas:

- CEM 3.0: El Continuo de Elevaciones Mexicano fue la herramienta principal utilizada para procesar y clasificar las texturas y características geomorfológicas de las diferentes regiones dentro del área de estudio. Este sistema permitió la definición de clases geomorfológicas que sirvieron como insumo para el desarrollo del mapa final. Como se muestra en la tabla 1, el cem 3.0 fue publicado en el 2013, con una resolución 15m x 15m, donde los valores en z (altura) se encuentra en enteros positivos y negativos en metros. En cuanto a la referencia geodésica, el datum corresponde a ITRF92 época 1988.0, elipsoide GRS80, coordenadas geográficas. El CEM, de la ciudad de Mexico utilizado en esta investigación pesa 4.43MB en formato TIF.
- Geomediana del INEGI (Cobertura temporal 2023): Este producto satelital multitemporal proporcionado por el Instituto Nacional de Estadística y Geografía (INEGI) fue una fuente importante de datos para el análisis geomorfológico. La

geomédiana combina imágenes satelitales de alta resolución obtenidas en diferentes momentos del ciclo de observación, lo que permitió obtener una representación robusta y libre de nubes del territorio. Las capas generadas a partir de la geomédiana sirvieron como insumos clave para el entrenamiento de los modelos de machine learning. La geomédiana de la Ciudad de México utilizada en esta investigación pesa 286 MB en formato TIF.

3.2.1.1 Herramientas de software:

- QGIS: Este software de código abierto fue utilizado para la manipulación, integración y análisis de las capas geoespaciales. QGIS permitió la creación de mapas base, la preparación de datos y la visualización de resultados geomorfológicos.
- Python y bibliotecas de machine learning (Scikit-Learn, Rgslib): El entorno de programación Python, junto con las bibliotecas especializadas Scikit-Learn y Rgslib, fue fundamental para implementar los algoritmos de clasificación y regresión utilizados para predecir las clases geomorfológicas a partir de las capas geoespaciales. Estas herramientas facilitaron la integración de datos, el entrenamiento de modelos y la evaluación de su rendimiento.

3.2.2 Criterios de Selección

Los criterios de selección para los datos y unidades de relieve a estudiar en el CEM de la Ciudad de México incluyen:

- Diversidad de Unidades de Relieve: Se seleccionaron áreas que presenten una variedad significativa de unidades de relieve, como colinas, montañas, valles y planicies. Esta diversidad es crucial para probar la efectividad de los algoritmos de segmentación y detección.

- **Calidad y Resolución de Datos:** Se prioriza los datos geográficos que tengan alta resolución y calidad. La precisión de los MDE y las imágenes de satélite es fundamental para la detección precisa de unidades de relieve mediante algoritmos de machine learning.
- **Accesibilidad de Datos:** Los datos seleccionados deben estar disponibles públicamente o accesibles a través de acuerdos de colaboración con el CEM. Esto garantiza que la recolección de datos sea ética y legal, y facilita la reproducibilidad del estudio.
- **Relevancia Geográfica:** Se considerarán áreas de estudio que sean representativas de la geomorfología típica de la región y que tengan relevancia para la planificación y el ordenamiento territorial. Esto incluye zonas urbanas y periurbanas que puedan beneficiarse del mapeo geomorfológico detallado.
- **Practicidad:** El CEM de la Ciudad de México es el más pequeño de México, lo que facilita la recolección y manejo de datos. Su tamaño permite un análisis detallado y exhaustivo de los datos geográficos sin la necesidad de gestionar volúmenes excesivos de información.
- **Disponibilidad de Datos:** El CEM de la Ciudad de México cuenta con una amplia base de datos de Modelos Digitales de Elevación (MDE) e imágenes de satélite, esenciales para la implementación y prueba de algoritmos de machine learning.

3.3 Escenario

3.3.1 Contexto Geográfico

El presente estudio se llevará a cabo utilizando el Continuo de Elevaciones Mexicano 3.0 (CEM 3.0) proporcionado por el Instituto Nacional de Estadística y Geografía (INEGI). El

CEM 3.0 es un conjunto de Modelos Digitales de Elevación (MDE) que cubren toda la República Mexicana con una alta resolución espacial. Este recurso es crucial para la detección y análisis de unidades de relieve mediante algoritmos de machine learning.

La Ciudad de México, donde se aplicará el estudio, se encuentra en el Valle de México, una región caracterizada por su variada geomorfología que incluye montañas, valles y planicies. Situada a una altitud promedio de 2,240 metros sobre el nivel del mar, la ciudad presenta un clima templado subhúmedo, lo que influye en los patrones de relieve y vegetación.

3.3.2 Contexto Institucional

El INEGI es la institución encargada de proporcionar el CEM 3.0, una herramienta fundamental para este estudio. El INEGI, reconocido por su precisión y exhaustividad en la recolección de datos geográficos, ofrece un entorno ideal para la implementación de tecnologías avanzadas como los algoritmos de machine learning. La disponibilidad de datos de alta calidad y la infraestructura tecnológica avanzada del INEGI permiten una gestión eficiente y detallada de la información geoespacial necesaria para este estudio.

3.4 Instrumentos de recolección de información

Se detallan las herramientas y técnicas empleadas para obtener los datos necesarios en el desarrollo de la investigación. Dado que el objetivo principal de la tesis es la creación de mapas geomorfológicos mediante algoritmos de machine learning, la recolección de información se centró en datos geoespaciales y otros tipos de insumos relevantes.

1. Imágenes satelitales: Se utilizó la geomediana obtenida del INEGI, la cual ofrecen información detallada sobre la superficie terrestre. Estas imágenes son esenciales para la clasificación de las unidades de relieve y la extracción de características geomorfológicas.

2. Modelos Digitales de Elevación (MDE): Los MDE, como el Continuo de Elevaciones Mexicano (CEM 3.0) del INEGI, fue una fuente clave de datos para analizar el relieve y la topografía de la región de estudio. Estos modelos permiten extraer variables como pendientes, orientaciones y altitudes que alimentan los algoritmos de machine learning.
3. Bases de datos geográficas: Se emplearon diversas bases de datos proporcionadas por el INEGI, que incluyen información cartográfica y datos geoespaciales complementarios, como hidrografía, uso de suelo y vegetación. Estas bases enriquecen el análisis y permiten validar los resultados obtenidos.
4. Software de análisis geoespacial: Para el procesamiento y análisis de los datos se utilizaron herramientas especializadas como QGIS y RSGISLib, que facilitaron la manipulación de grandes volúmenes de datos, el cálculo de variables geomorfológicas y la integración de los resultados en un entorno GIS.

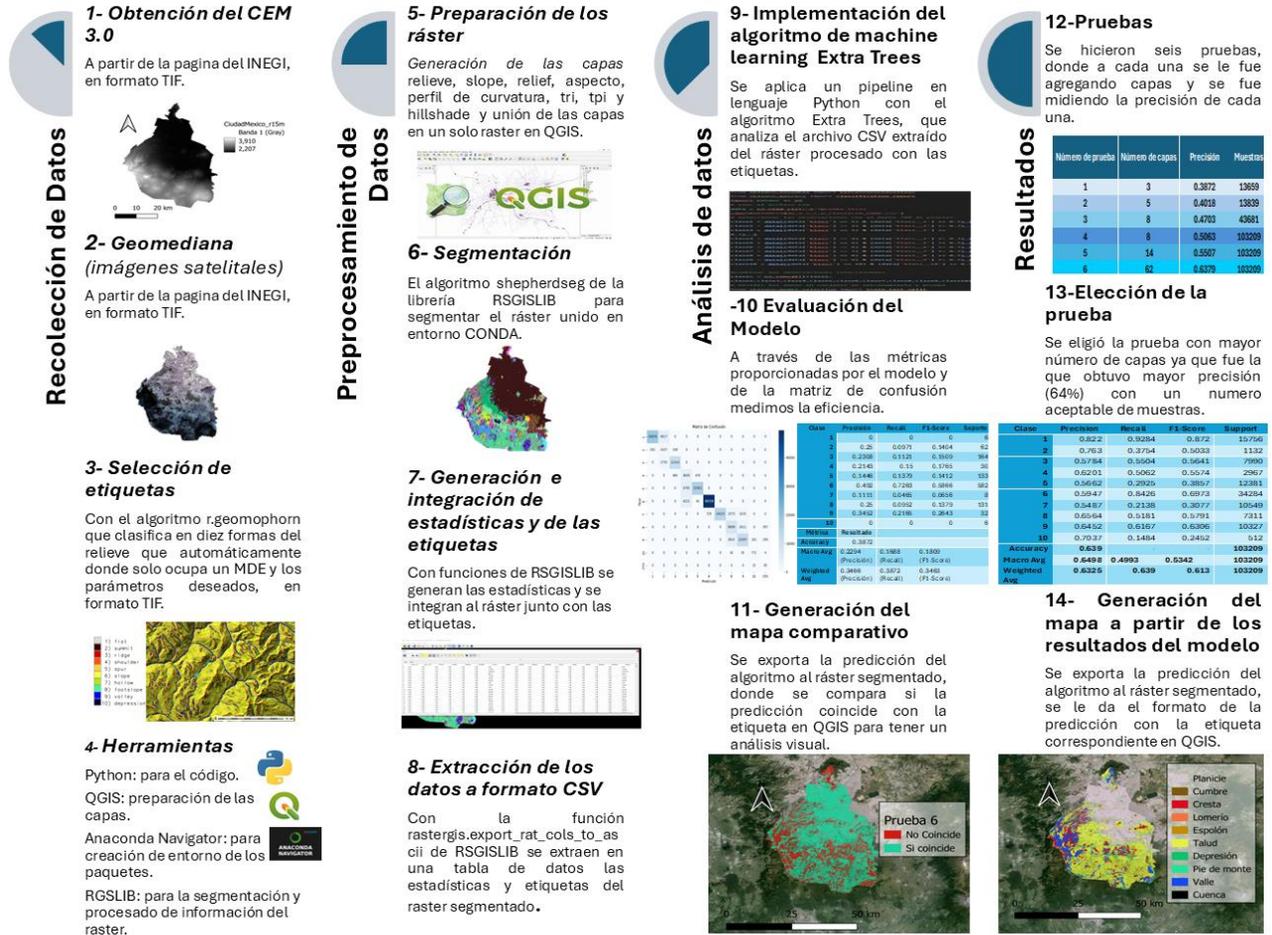
Estos instrumentos fueron seleccionados por su precisión y capacidad para manejar datos espaciales complejos, permitiendo así obtener la información necesaria para el entrenamiento y validación de los algoritmos de machine learning desarrollados en la investigación

3.5 Procedimiento

Los procedimientos detallan los pasos específicos llevados a cabo para recolectar, procesar y analizar los datos. A continuación, se describen los procedimientos seguidos en este estudio como se puede ver de manera resumida en la figura 3.

Figura 3.

Modelo para la generación de mapas geomorfológicos con machine learning



Nota. Elaboración propia.

3.5.1 Recolección de Datos

3.5.1.1 Obtención del CEM 3.0

Se obtendrán los datos del CEM 3.0 del INEGI, que proporcionan información detallada sobre la elevación y características del terreno en la Ciudad de México (INEGI, 2020).

Primer paso es elegir el MDE, que de algún modo nos da la información cruda del relieve,

a trabajar en este caso se escogió el CEM de la ciudad de México ya que es el de menor tamaño territorial y por consecuente menor peso en megabytes, lo que agiliza el procesamiento y el trabajo de los algoritmos para realizar las pruebas consecuentes.

3.5.1.2 Imágenes Satelitales

Las imágenes satelitales permiten observar extensas áreas de terreno con alta resolución, proporcionando un panorama completo de las características geomorfológicas. Se adquirirán imágenes satelitales de alta resolución de la Ciudad de México para su análisis y segmentación, precisamente la Geomediana Landsat del INEGI, teniendo un peso de pesa 286 MB en formato TIF.

3.5.1.3 Selección de etiquetas

Por falta de un mapa geomorfológico actualizado y manipulable, se optó por hacer un que con el algoritmo *r.geomophorn* que clasifica en diez formas del relieve que automáticamente donde solo ocupa un MDE y los parámetros deseados. Identifica 10 clases de relieve: Planicie (1), Cumbre (2), Cresta (3), Lomerío (4), Espolón (5), Talud Depresión (6), Piedemonte (7), Valle (8), Cuenca (9), y Cuenca (10).

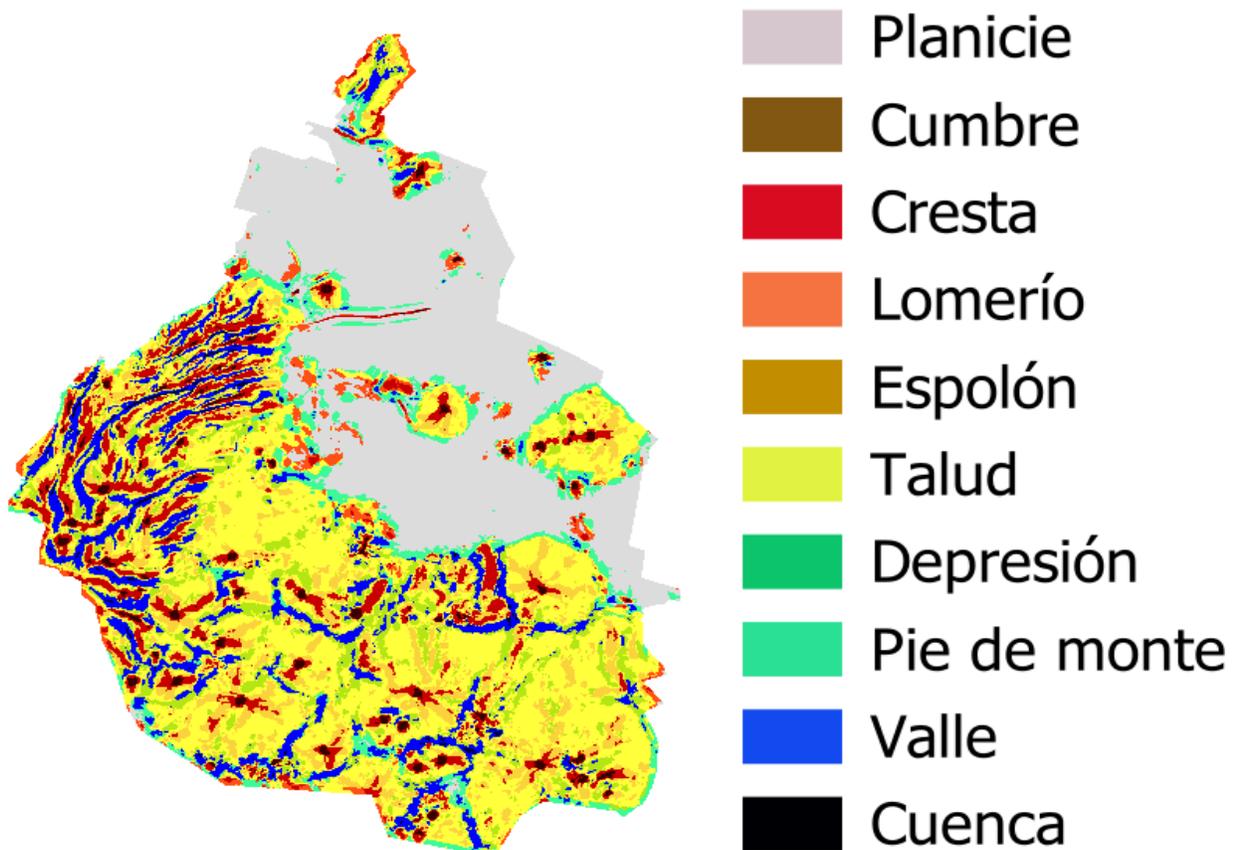
3.5.2 Preprocesamiento de Datos

3.5.2.1 Preparación de los ráster

Se realizaron pruebas y mediante funciones del QGIS se tomaron del MDE (CEM de la Ciudad de México) los capas relieve, slope, relief, aspecto y perfil de curvatura, tri, tpi y hillshade, donde se obtiene ráster para cada uno para una posterior unión de las capas según la prueba a realizar.

Figura 6

Etiquetas creadas por el algoritmo r. geomphorn



Nota. Elaboración propia.

3.5.2.2 Normalización de los ráster

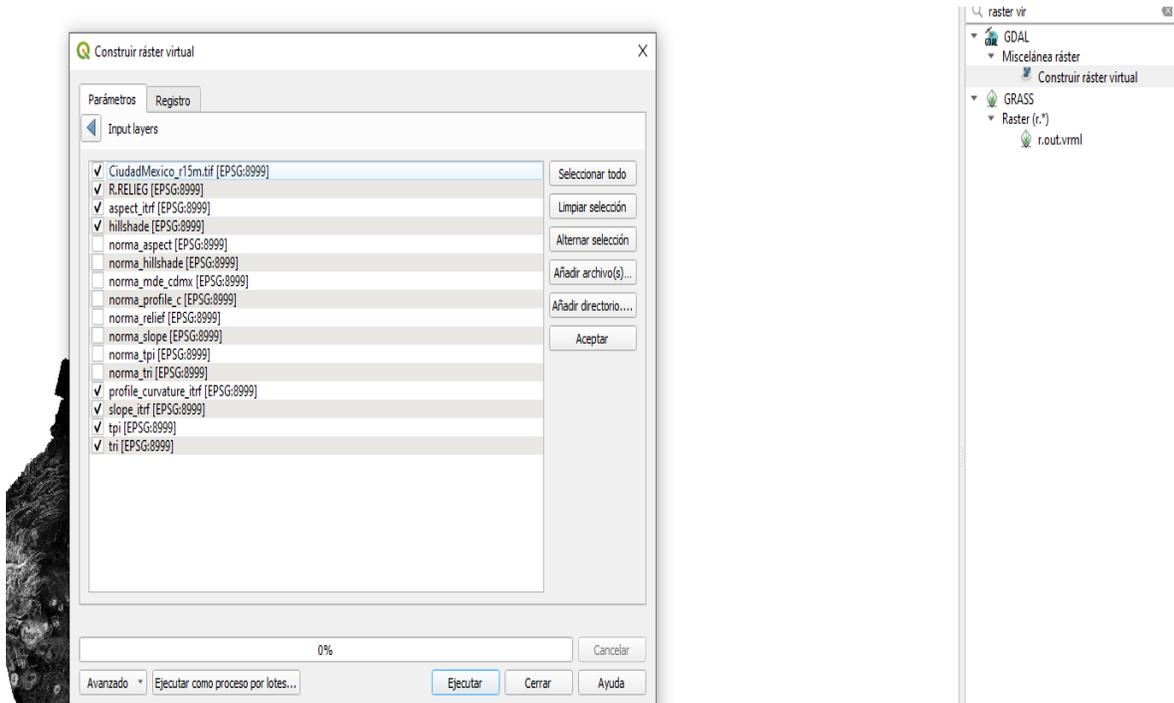
Antes de unir los ráster se experimentó si es necesario normalizar las capas para no generar tanta desviación en la información, en los casos del ráster aspecto, profile curvature y slope se tuvo la necesidad de convertirlos en radianes, debido a que su información está en grados (ver apéndice 1).

3.5.2.3 Unión de los ráster

En cada prueba se seleccionaron distintos ráster, que se detallarán cuales fueron en la sección de resultados, para unir las capas primeramente se creó un ráster virtual donde se seleccionan las capas analizar y cada capa será una banda del ráster, en la figura 7 se muestra como crea un solo ráster con las capas seleccionadas en QGIS, para posteriormente exportar el ráster virtual en un ráster en formato TIF.

Figura 7

Creación del ráster virtual



Nota. Elaboración propia.

3.5.2.4 Segmentación

Al obtener un solo ráster con todas las capas a trabajar, donde cada capa se le asigna una banda en el ráster unido, el orden de las bandas se puede en archivo de texto que se crea

en el momento de exportar el TIF, se utiliza el algoritmo shepherdseg de la librería RSGSLIB para segmentar el ráster probando distintos parámetros mostrados en el siguiente capítulo, en ambiente Conda, con el paquete de RGSGLIB se ejecutó el siguiente código:

```
from rsgislib.segmentation import shepherdseg

input_img = r'D:\prueba tesis cdmx\marzo\8capas\8norma_rad.tif'
out_clumps_img = r'D:\prueba tesis cdmx\marzo\8capas\norma_rad\norma_rad_cluster2_thres50_px100.kea'
out_mean_img = None

shepherdseg.run_shepherd_segmentation(input_img, out_clumps_img,
                                      out_mean_img, num_clusters=2, dist_thres=50, min_n_pxls=100)
```

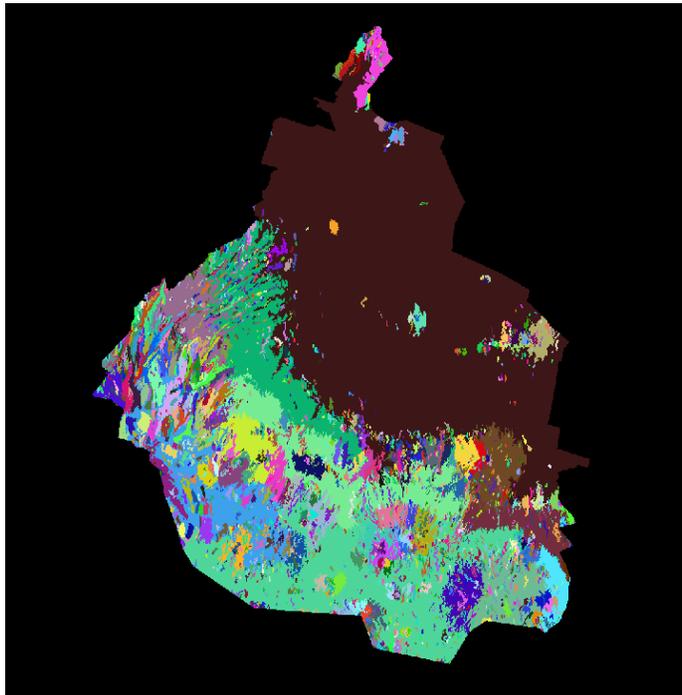
Con el algoritmo de RGSGLIB shepherdseg se hizo la segmentación modificando los parámetros clúster dist_thres y min_n_pxls, siendo estos parámetros los más relevantes a la hora de hacer la segmentación. El ráster unido con diferentes capas, se realizaron distintas pruebas para y se tomó el que tenía más muestras y podría semejarse más a la realidad donde al final queda un archivo en formato kea con los datos de las bandas (ver apéndice 2). Posteriormente con la librería de RSGSLIB se le agrega información de las etiquetas y estadísticas de cada banda (correspondiente a cada capa que integra el ráster) como se muestra en la figura 9.

Con el siguiente comando se abre y permite analizar el archivo kea, como se observa en la figura 8:

```
(base) D:\Users\rene.armenta>$ tuiview mde_a_s_r_p_cluster10_tresh100_pxk100.kea'
```

Figura 8

Mapa en archivo kea resultado de la segmentación



Nota. Elaboración propia.

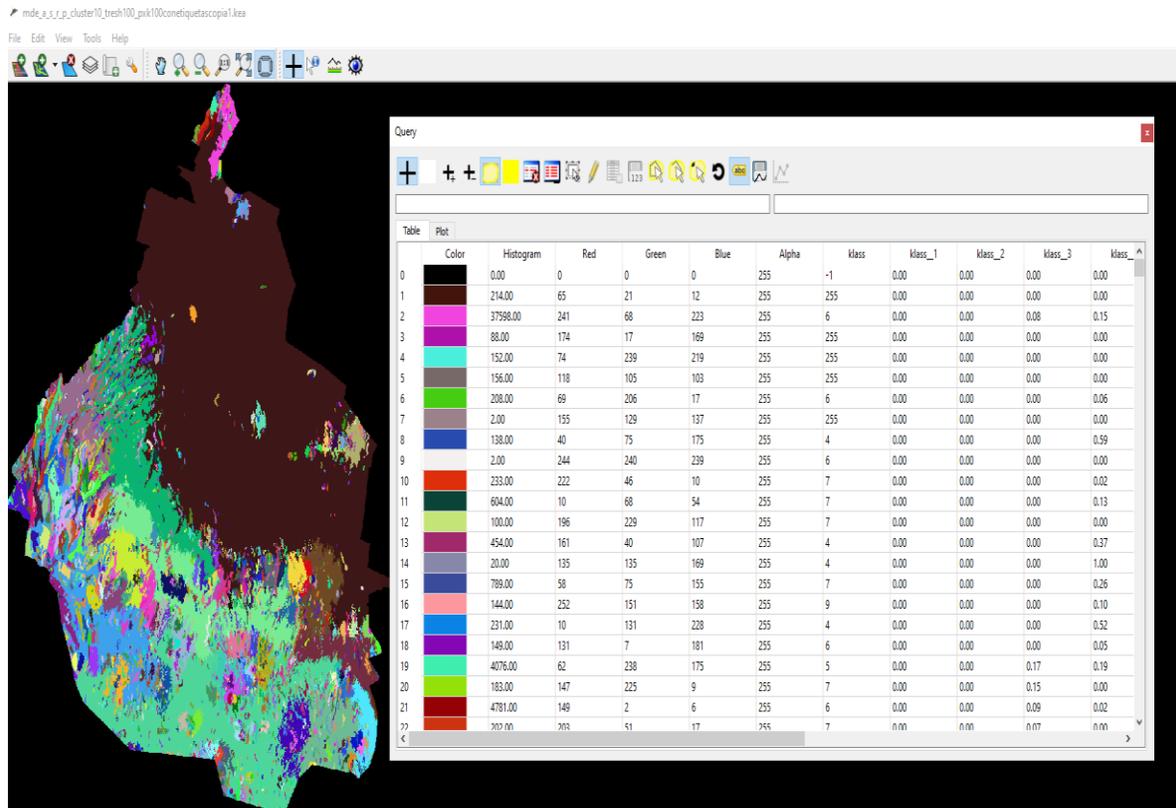
3.5.2.5 Agregar etiquetas

Para agregar las etiquetas, mencionadas anteriormente producidas con el algoritmo `r.geomphorn` al archivo en formato KEA se aplica la función `rastergis.populate_rat_with_cat_proportions`, donde se utiliza la biblioteca `RSGISLib` y las funciones de `rastergis` para manipular los datos en formato KEA y TIFF con el siguiente código:

```
D: > prueba tesis cdmx > rgslib > 2_agregar_etiquetas_8bandas.py > ...
1  from rsgislib import rastergis
2
3  clumpsImage = r'D:\prueba tesis cdmx\rgslib\mde_slope_profile_aspect_relief_cluster10.kea'
4  catsimage = r'D:\prueba tesis cdmx\60_30_01.tif'
5  outcolname = 'klass_'
6  majcolname = 'klass'
7  rastergis.populate_rat_with_cat_proportions(catsimage,clumpsImage,outcolname,majcolname)
```

Figura 9

Archivo kea con las etiquetas agregadas.



Nota. Elaboración propia

La variable `clumpsImage` contiene la ruta del archivo KEA generado en la sección anterior y `catsImage` contiene la ruta del archivo TIFF de etiquetas generado en el proceso de clasificación. La función `rastervis.populate_rat_with_cat_proportions` toma estos dos archivos como entrada, así como dos nombres de columna para la tabla de atributos del archivo KEA. `outcolname` es el nombre que se le dará a la columna de salida que contendrá las proporciones de las categorías en cada clump (o agrupación) de píxeles en el archivo KEA. `majcolname` es el nombre de la columna que contiene las etiquetas de la clasificación, que en este caso es "klass". El kea seleccionado (ráster segmentado), y el

ráster (con etiquetas) seleccionado en este caso este se realizó con el algoritmo r. geomophorn, donde etiqueta del 1 al 10 diferentes formas del terreno.

3.5.2.6 Generación de estadísticas

Para la generación de las estadísticas necesarias para el análisis se aplica el siguiente código en el archivo KEA:

```
from rsgislib import rastergis

inputImage = r'D:\prueba tesis cdmx\marzo\8capas\8capas.tif'
clumpsImage = r'D:\prueba tesis cdmx\rsgislib\mde_slope_aspect_clumps_cluster10.kea'

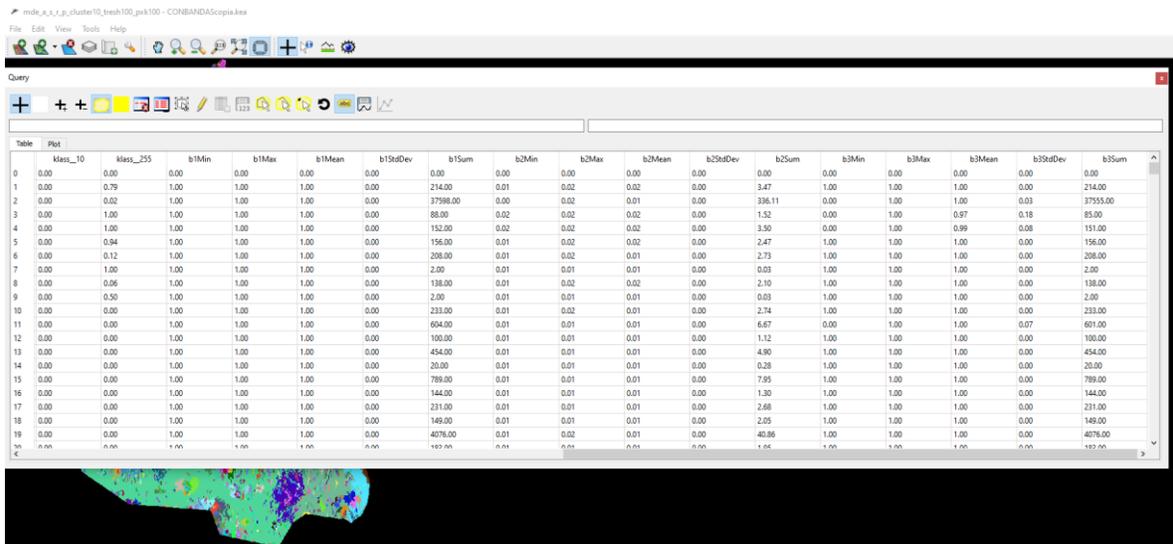
bs = []
bs.append(rastergis.BandAttStats(band=1, min_field='b1Min', max_field='b1Max', mean_field='b1Mean', sum_field='b1Sum', std_dev_field='b1StdDev'))
bs.append(rastergis.BandAttStats(band=2, min_field='b2Min', max_field='b2Max', mean_field='b2Mean', sum_field='b2Sum', std_dev_field='b2StdDev'))
bs.append(rastergis.BandAttStats(band=3, min_field='b3Min', max_field='b3Max', mean_field='b3Mean', sum_field='b3Sum', std_dev_field='b3StdDev'))

rastergis.populate_rat_with_stats(inputImage, clumpsImage, bs)
```

Este código es utilizado para calcular y agregar estadísticas por banda a la tabla de atributos de un archivo ráster. Cada objeto BandAttStats especifica el número de banda y el nombre de los campos a agregar en la tabla de atributos para almacenar las estadísticas mínima, máxima, media, suma y desviación estándar para esa banda en particular. Finalmente, se utiliza la función populate_rat_with_stats () para calcular las estadísticas por banda y agregarlas a la tabla de atributos del archivo ráster en formato KEA. En la figura 9 y 10 del ejemplo se agregaron 3 bandas la primera banda corresponde al MDE, la segunda al aspecto, la tercer al perfil de curvatura y las estadísticas correspondientes.

Figura 10

Archivo kea con las etiquetas y estadísticas.



Nota. Elaboración propia

3.5.2.7 Generación de estadísticas adicionales

La implementación clásica del filtro Leung-Malik genera 48 filtros diferentes a la imagen original, generando 48 texturas en formato kea, se calculan y agregan las estadísticas mencionadas en el paso anterior y los percentiles a cada textura que se integraran como bandas al ráster y kea final. Las estadísticas calculadas fueron las siguientes:

- Mínimo (Min): El valor mínimo de los datos.
- Máximo (Max): El valor máximo de los datos.
- Media (Mean): El valor medio de los datos.
- Suma (sum): La suma de todos los valores de los datos.
- Desviación estándar (StdDev): La desviación estándar de los datos.

Percentiles calculados:

- Percentil 10 (Per10):

- Percentil 20 (Per20):
- Percentil 30 (Per30):
- Percentil 40 (Per40):
- Percentil 50 (Per50):
- Percentil 60 (Per60):
- Percentil 70 (Per70):
- Percentil 80 (Per80):
- Percentil 90 (Per90)

3.5.2.8 Extracción de los datos kea a CSV

Para extraer los datos a un archivo CSV, esto permite manejar y visualizar los datos de forma más amigable antes de probar el modelo, se realiza con el siguiente código:

```

prueba tesis cdmx > rgslib > 4_importar a csv 8bandas.py > ...
from rsgislib import rastergis
clumpsImage = r'D:\prueba tesis cdmx\marzo\8capas\8capas_cluster10_thres200_px100.kea'
fields = ['klass','klass_1','klass_2','klass_3','klass_4','klass_5','klass_6',
          'klass_7','klass_8','klass_9','klass_10','b1Min', 'b1Max', 'b1Mean',
          'b1StdDev', 'b1Sum', 'b2Min', 'b2Max', 'b2Mean', 'b2StdDev', 'b2Sum',
          'b3Min', 'b3Max', 'b3Mean', 'b3StdDev', 'b3Sum', 'b4Min', 'b4Max',
          'b4Mean', 'b4StdDev', 'b4Sum', 'b5Min', 'b5Max', 'b5Mean', 'b5StdDev',
          'b5Sum', 'b6Min', 'b6Max', 'b6Mean', 'b6StdDev', 'b6Sum', 'b7Min', 'b7Max',
          'b7Mean', 'b7StdDev', 'b7Sum', 'b8Min', 'b8Max', 'b8Mean', 'b8StdDev', 'b8Sum',
          'b9Min', 'b9Max', 'b9Mean', 'b9StdDev', 'b9Sum',
          'b10Min', 'b10Max', 'b10Mean', 'b10StdDev', 'b10Sum',
          'b11Min', 'b11Max', 'b11Mean', 'b11StdDev', 'b11Sum',
          'b12Min', 'b12Max', 'b12Mean', 'b12StdDev', 'b12Sum',
          'b13Min', 'b13Max', 'b13Mean', 'b13StdDev', 'b13Sum',
          'b14Min', 'b14Max', 'b14Mean', 'b14StdDev', 'b14Sum']
outfile = r'D:\prueba tesis cdmx\marzo\8capas\8capas_cluster10_thres200_px100_geo.csv'
rastergis.export_rat_cols_to_ascii(clumpsImage, outfile, fields)

```

Con el siguiente código podemos observar los píxeles que contiene el 70% o más de correspondencia a las etiquetas de acuerdo con la segmentación, al filtrar los datos con al menos un 70% de correspondencia, se asegura de trabajar únicamente con píxeles que tienen una alta probabilidad de pertenecer a una categoría específica. Esto mejora la calidad del conjunto de datos y reduce el ruido en el análisis.:

```
import pandas as pd

# Leer el archivo CSV
data = pd.read_csv(r'D:\prueba tesis
cdmx\RADIANES\mde_a_s_r_p_cluster10_CDMX.csv')

# Seleccionar las observaciones con al menos 70% de píxeles

clases = data[(data['klass'] == 1) &
              (data['klass__1'] >= 0.7) &
              (data['klass__2'] >= 0.7) &
              (data['klass__3'] >= 0.7) &
              (data['klass__4'] >= 0.7) &
              (data['klass__5'] >= 0.7) &
              (data['klass__6'] >= 0.7) &
              (data['klass__7'] >= 0.7) &
              (data['klass__8'] >= 0.7) &
              (data['klass__9'] >= 0.7) &
              (data['klass__10'] >= 0.7)]

# Dividir los datos en conjunto de entrenamiento y prueba
from sklearn.model_selection import train_test_split

train_data, test_data = train_test_split(clases, test_size=0.3,
train_size=0.7, random_state=42, n_samples=1)
```

3.5.2.9 Aplicación del Modelo

Al final se aplica un pipeline que analizará y dará la precisión del modelo dando la efectividad con la que detecta y se acerca la segmentación con las etiquetas para una comparación y así hacer un análisis posterior. La aplicación del pipeline se aplica al archivo extraído en formato CSV que contiene todas las estadísticas de cada capa (banda), así como las etiquetas agregadas.

3.6 Diseño del método

3.6.1 Diseño

El tipo de diseño es exploratorio ya que el objetivo principal de este estudio es explorar la capacidad de los algoritmos de machine learning para detectar y clasificar unidades de relieve a partir de Modelos Digitales de Elevación (MDE) e imágenes de satélite.

El estudio se clasifica como exploratorio debido a varias razones fundamentales:

- **Naturaleza Innovadora del Enfoque:** La aplicación de algoritmos de machine learning para la detección y clasificación de unidades de relieve es un área emergente en la cartografía. Dado que existen pocos estudios previos que aborden esta problemática con un enfoque similar, es necesario realizar una investigación inicial que permita comprender las capacidades y limitaciones de estos algoritmos
- **Falta de Datos Preliminares:** Al ser un campo novedoso, no se dispone de suficientes datos preliminares o estudios comparativos que establezcan un marco teórico sólido para la implementación de estos algoritmos en el mapeo geomorfológico. El estudio exploratorio permitirá recopilar datos básicos y establecer hipótesis para investigaciones futuras más detalladas

- **Flexibilidad Metodológica:** Los estudios exploratorios son particularmente adecuados cuando se necesita flexibilidad metodológica para adaptar y ajustar los métodos de investigación en función de los hallazgos iniciales. Esto es crucial en el presente estudio, ya que la efectividad de los algoritmos de machine learning puede variar significativamente según las características del área de estudio y la calidad de los datos disponibles (Marshall & Rossman, 2016).
- **Identificación de Variables Clave:** Un estudio exploratorio ayudará a identificar las variables geográficas más relevantes que influyen en la detección y clasificación de unidades de relieve. Esto proporcionará una base sólida para futuros estudios correlacionales o explicativos que busquen establecer relaciones causales entre estas variables y los patrones geomorfológicos detectados (Creswell, 2014).

3.6.2 Momento del estudio

En esta investigación, se adoptó un diseño transversal, ya que los datos fueron recolectados y analizados en un momento específico en el tiempo, sin seguimiento a lo largo de diferentes etapas. Este enfoque fue adecuado para evaluar y comparar los mapas geomorfológicos generados mediante algoritmos de machine learning, utilizando datos geoespaciales que reflejan las condiciones actuales del terreno. Aunque un estudio longitudinal podría haber permitido observar cambios en el relieve a lo largo del tiempo, el objetivo de este estudio fue centrarse en la viabilidad y precisión de los métodos automatizados en comparación con los tradicionales, bajo las condiciones geográficas presentes.

3.6.3 Alcance del estudio

El diseño del estudio se basa en un enfoque correlacional debido a la necesidad de investigar las relaciones entre múltiples variables geográficas y cómo estas influyen en la precisión y eficacia de los algoritmos de machine learning para la detección de unidades de relieve. Este enfoque es fundamental para establecer patrones y tendencias que puedan guiar investigaciones futuras más detalladas y específicas (Creswell, 2014).

La importancia del Diseño Correlacional se basa en:

- **Identificación de Relaciones Significativas:** El diseño correlacional permite identificar y cuantificar las relaciones entre variables geográficas, tales como pendiente, altitud, tipo de suelo, y su influencia en la detección de unidades de relieve. Esto es crucial para mejorar la precisión de los algoritmos de machine learning utilizados en la geomorfología (Salkind, 2010).
- **Base para Modelos Predictivos:** Al establecer relaciones claras entre variables geográficas y resultados de detección, este estudio proporciona una base sólida para el desarrollo de modelos predictivos. Estos modelos pueden ser utilizados para optimizar los algoritmos y mejorar su aplicabilidad en diferentes contextos geográficos (Tabachnick & Fidell, 2013).
- **Mejora de la Eficiencia Algorítmica:** Entender cómo las variables geográficas afectan la detección de unidades de relieve permitirá ajustar y mejorar los algoritmos de machine learning, haciéndolos más eficientes y precisos. Esto es especialmente relevante dado el creciente uso de estas tecnologías en la geomorfología y la planificación territorial (Hair, Black, Babin, & Anderson, 2019).

3.7 Operacionalización de las variables y/o categorías de estudio

La operacionalización de las variables en este estudio implica definir claramente cómo se medirán y analizarán las variables geográficas y las unidades de relieve detectadas mediante algoritmos de machine learning. A continuación, se presentan las variables principales y las categorías de estudio, junto con sus definiciones y métodos de medición.

3.7.1 Variables

Estas variables se derivan de los datos topográficos y de relieve proporcionados por el Continuo de Elevaciones Mexicano 3.0 (CEM 3.0) y las imágenes satelitales. A continuación, se presenta la descripción de las variables y cómo se operativizan en el contexto del análisis.

3.7.1.1 Variables Independientes (Predictoras)

- Elevación (ELEV):
 - Definición: Altura del terreno sobre el nivel del mar.
 - Medición: Metros (m).
 - Fuente de Datos: CEM 3.0.
- Pendiente (SLOPE):
 - Definición: Inclinación del terreno.
 - Medición: Grados (°).
 - Fuente de Datos: Derivada del CEM 3.0 mediante herramientas GIS.
- Orientación de la Pendiente (ASPECT):
 - Definición: Dirección hacia la cual se inclina la pendiente.
 - Medición: Grados (°) con valores de 0 a 360.
 - Fuente de Datos: Derivada del CEM 3.0 mediante herramientas GIS.

- Curvatura (CURVATURE):
 - Definición: Cambio en la pendiente del terreno.
 - Medición: Adimensional.
 - Fuente de Datos: Derivada del CEM 3.0 mediante herramientas GIS.
- Índice de Rugosidad del Terreno (TRI):
 - Definición: Medida de la rugosidad de la superficie terrestre.
 - Medición: Adimensional.
 - Fuente de Datos: Calculado a partir del CEM 3.0.
- Hillshade (HS)
 - Definición: Modelo de sombreado del terreno que simula la iluminación del sol sobre la superficie.
 - Medición: Valores de sombreado (0-255), donde 0 es sombra total y 255 es luz total.
 - Fuente de Datos: Derivado del CEM 3.0 mediante herramientas GIS.
- Relieve (RELIEF)
 - Definición: Variación en la elevación de la superficie terrestre.
 - Medición: Metros (m) o adimensional.
 - Fuente de Datos: Derivado del CEM 3.0 mediante herramientas GIS.
- Índice Topográfico de Posición (TPI)
 - Definición: Diferencia entre la elevación de un punto y la elevación promedio de su vecindario.
 - Medición: Adimensional.
 - Fuente de Datos: Calculado a partir del CEM 3.0 mediante herramientas GIS.

3.7.1.2 Variable Dependiente (Objetivo)

- Clase Geomorfológica (KLASS):
 - Definición: Categoría de la unidad geomorfológica.
 - Medición: Enteros (1 a 10), donde cada número representa una clase geomorfológica distinta.
 - Fuente de Datos: Clasificación realizada mediante algoritmos de machine learning.

3.7.2 Procedimiento de Operacionalización

Las variables predictoras se obtienen y calculan utilizando el CEM 3.0 y las imágenes satelitales procesadas mediante el software QGIS. Posteriormente, estas variables se utilizan como características de entrada para el modelo de machine learning, donde el orden fue el siguiente:

- Preparación de los Datos:
 - Los datos topográficos y de relieve se extraen y procesan utilizando QGIS y bibliotecas de Python.
 - Las características predictoras se normalizan para asegurar una escala uniforme mediante el StandardScaler de scikit-learn.
- Segmentación y Clasificación:
 - Los datos se segmentan utilizando rgslib para identificar diferentes unidades geomorfológicas.
 - Se emplea el modelo ExtraTreesClassifier y MLPClassifier para la clasificación de las unidades geomorfológicas.
- Entrenamiento y Evaluación del Modelo:

- Se divide el conjunto de datos en conjuntos de entrenamiento y prueba.
- El modelo se entrena utilizando el conjunto de entrenamiento y se evalúa con el conjunto de prueba para verificar su precisión y robustez.
- Validación y Ajuste del Modelo:
 - Se ajusta el modelo basado en los resultados obtenidos, utilizando métricas de rendimiento como precisión, recall y F1-score para evaluar su desempeño.

3.8 Tipo de análisis empleado

La presente investigación se enmarca en un cuantitativo. Se emplearán técnicas de teledetección y análisis de datos geospaciales para identificar y caracterizar las unidades de relieve de la Ciudad de México. Los datos se obtendrán del Continuo de Elevaciones Mexicano 3.0 (CEM 3.0) proporcionado por el INEGI, y se analizarán utilizando herramientas como QGIS y Python que permite el análisis de las formas del terreno mediante algoritmos de machine learning. Este método tiene el objetivo de automatizar el proceso de generación de mapas geomorfológicos, haciéndolo más eficiente y menos costoso, adecuado para un estudio correlacional.

3.8.1 Enfoque Cuantitativo

El enfoque cuantitativo implica la recolección y análisis de datos numéricos para identificar patrones y relaciones entre variables geográficas y unidades de relieve. Se utilizarán algoritmos de machine learning para procesar y analizar los datos, asegurando la precisión y replicabilidad de los resultados.

3.9 Análisis de datos

3.9.1 Técnicas e instrumentos

En este estudio, se emplearán diversas técnicas e instrumentos para la recolección, procesamiento y análisis de datos geospaciales con el objetivo de detectar unidades de relieve mediante algoritmos de machine learning. A continuación, se detallan las técnicas e instrumentos utilizados.

3.9.2 Técnicas

3.9.2.1 *Sistemas de Información Geográfica (SIG):*

Descripción: Los SIG son herramientas esenciales para el análisis y visualización de datos geospaciales. Permiten el procesamiento de Modelos Digitales de Elevación (MDE) y la derivación de variables geográficas como la pendiente, altitud, orientación y curvatura del terreno.

- Aplicación: QGIS se utilizará para preprocesar los datos del CEM 3.0, generar mapas temáticos y realizar análisis espaciales detallados.

3.9.2.2 *Segmentación de Imágenes con RGSLIB:*

- Descripción: RGSLIB es una librería en Python especializada en la segmentación de datos geospaciales. Esta herramienta permite segmentar imágenes de satélite y MDE para identificar unidades de relieve.
- Aplicación: Se utilizará RGSLIB para segmentar los datos del CEM 3.0, aplicando técnicas de segmentación que ayuden a delinear diferentes unidades de relieve.

3.9.2.3 *Algoritmos de Machine Learning:*

- Descripción: Se emplearán algoritmos de machine learning para la segmentación y clasificación de unidades de relieve. Estos algoritmos pueden identificar patrones complejos en los datos geoespaciales que no son fácilmente detectables mediante métodos tradicionales.
- Aplicación: Los algoritmos se implementarán utilizando el paquete sklearn en Python. El algoritmo utilizado para la identificación de las unidades fue el de Extra Trees.

3.9.2.4 *Análisis Estadístico:*

- Descripción: Se llevarán a cabo análisis estadísticos para evaluar la relación entre las variables geográficas y las unidades de relieve detectadas. Esto incluye pruebas de correlación y análisis de regresión.
- Aplicación: Se utilizarán softwares estadísticos para analizar los datos y determinar la significancia de las relaciones encontradas.

3.9.3 Instrumentos

3.9.3.1 *Continuo de Elevaciones Mexicano 3.0 (CEM 3.0):*

- Descripción: El CEM 3.0 es un conjunto de Modelos Digitales de Elevación de alta resolución proporcionados por el INEGI, que cubren toda la República Mexicana.
- Aplicación: Se utilizará como la fuente principal de datos geoespaciales para el análisis de unidades de relieve. Los datos se procesarán para derivar variables como la pendiente, altitud, orientación y curvatura del terreno.

3.9.3.2 *Software de SIG (QGIS):*

- Descripción: Herramienta de software utilizada para el análisis y visualización de datos geoespaciales.
- Aplicación: QGIS se empleará para procesar los datos del CEM 3.0, realizar análisis espaciales y generar mapas temáticos que visualicen las unidades de relieve detectadas.

3.9.3.3 *Librerías de Machine Learning (scikit-learn, RGSLIB):*

- Descripción: Librerías de software que proporcionan implementaciones de algoritmos de machine learning.
- Aplicación: Se utilizarán para implementar y entrenar los algoritmos de segmentación y clasificación de unidades de relieve. RGSLIB se empleará para tareas de segmentación geoespacial, y scikit-learn se utilizará para el entrenamiento del modelo de clasificación.

3.9.3.4 *Herramientas Estadísticas (RGSLIB, Python):*

- Descripción: Software estadístico utilizado para el análisis de datos.
- Aplicación: Se utilizarán para realizar análisis estadísticos que evalúen la relación entre las variables geográficas y las unidades de relieve detectadas, proporcionando una base cuantitativa para interpretar los resultados.

3.9.4 Validez

La validez del estudio es fundamental para asegurar que los resultados sean precisos y reflejen fielmente la realidad. En este estudio, se abordarán diferentes tipos de validez para garantizar la robustez del diseño y la metodología empleada.

3.9.4.1 Validez Interna

La validez interna se refiere al grado en que los resultados del estudio pueden ser atribuidos a las variables independientes y no a otros factores externos. Para asegurar la validez interna, se tomarán las siguientes medidas:

- Control de Variables Confusoras: Se identificarán y controlarán posibles variables confusoras que puedan influir en los resultados, como la resolución del CEM 3.0 y las características específicas del área de estudio (Field, 2013).
- Consistencia en la Recolección de Datos: Se utilizarán procedimientos estandarizados para la recolección y preprocesamiento de datos, garantizando que todas las imágenes de satélite y datos del CEM 3.0 sean tratados de manera uniforme (Tabachnick & Fidell, 2019).
- Técnicas de Segmentación y Clasificación: La segmentación de imágenes y la clasificación de unidades de relieve se llevarán a cabo utilizando métodos validados y reconocidos. Se emplearán técnicas de validación cruzada para evaluar la precisión de los modelos de clasificación (Hox & Bechger, 2014).

3.9.4.2 *Validez Externa*

La validez externa se refiere al grado en que los resultados del estudio pueden ser generalizados a otros contextos o poblaciones. Para asegurar la validez externa, se implementarán las siguientes estrategias:

- **Diversidad en los Datos:** Aunque se utilizará el CEM 3.0 de la Ciudad de México por razones de practicidad, se evaluará la aplicabilidad del modelo a otras regiones geográficas en futuros estudios para evaluar su generalizabilidad (Bryman, 2016).
- **Replicabilidad del Estudio:** Se documentarán detalladamente todos los pasos del procedimiento, permitiendo que otros investigadores puedan replicar el estudio en diferentes contextos y con diferentes conjuntos de datos (Cohen, Manion, & Morrison, 2018).

3.9.4.3 *Validez de Constructo*

La validez de constructo se refiere a la medida en que los instrumentos utilizados realmente miden los conceptos que se pretende medir. Para asegurar la validez de constructo, se tomarán las siguientes medidas:

- **Definición de Variables:** Las variables geográficas utilizadas (pendiente, altitud, orientación, curvatura) estarán claramente definidas y se derivarán consistentemente del CEM 3.0 (Babbie, 2015).
- **Validación de Algoritmos de Machine Learning:** Los algoritmos de machine learning se seleccionarán y validarán basándose en su capacidad para identificar correctamente las unidades de relieve. Se utilizarán métricas estándar como el reporte de clasificación para evaluar el rendimiento de los modelos (James, Witten, Hastie, & Tibshirani, 2013).

3.9.4.4 *Validez de Contenido*

La validez de contenido se refiere a la medida en que un instrumento cubre el dominio del constructo que se está midiendo. Para asegurar la validez de contenido, se procederá de la siguiente manera:

- **Revisión de la Literatura:** Se realizará una revisión exhaustiva de la literatura para identificar las variables y métodos más relevantes y actuales en la detección de unidades de relieve (Silverman, 2016).
- **Consultas con Expertos:** Se consultará a expertos en geomorfología y análisis geoespacial para asegurar que las variables y métodos utilizados son apropiados y abarcan adecuadamente el constructor de interés (Gerring, 2017).

3.9.5 **Confiabilidad**

La confiabilidad del estudio se refiere a la consistencia y estabilidad de los resultados obtenidos a lo largo del tiempo. Un estudio es confiable si, al repetirse en las mismas condiciones, produce resultados similares. En este estudio, se abordarán diferentes aspectos de la confiabilidad para garantizar la robustez del diseño y la metodología empleada.

3.9.5.1 *Consistencia Interna*

La consistencia interna se refiere al grado en que las distintas partes del instrumento de medición son coherentes entre sí. Para asegurar la consistencia interna del estudio, se implementarán las siguientes estrategias:

- **Uso de Algoritmos Validados:** Se utilizarán algoritmos de machine learning que han demostrado ser consistentes y fiables en estudios previos. Por ejemplo, el uso de

sklearn para la segmentación y clasificación garantiza un alto nivel de consistencia interna (James, Witten, Hastie, & Tibshirani, 2013).

- Estandarización de Procedimientos: Todos los procedimientos de recolección, preprocesamiento y análisis de datos se estandarizarán para minimizar las variaciones. Esto incluye el uso de software como QGIS y Python para asegurar la uniformidad en el tratamiento de los datos (Field, 2013).

3.9.5.2 Estabilidad

La estabilidad se refiere a la consistencia de los resultados cuando se repite la medición en diferentes momentos. Para asegurar la estabilidad del estudio, se tomarán las siguientes medidas:

- Pruebas de Fiabilidad Test-Retest: Se realizarán pruebas de fiabilidad test-retest para evaluar la estabilidad de los resultados a lo largo del tiempo. Se llevarán a cabo múltiples segmentaciones y clasificaciones en diferentes momentos y se compararán los resultados para evaluar la estabilidad (Cohen, Manion, & Morrison, 2018).
- Replicabilidad de Resultados: Se documentarán detalladamente todos los pasos del procedimiento para que otros investigadores puedan replicar el estudio en diferentes momentos y obtener resultados similares (Babbie, 2015).

3.9.5.3 Equivalencia

La equivalencia se refiere al grado en que los resultados obtenidos mediante diferentes métodos o instrumentos son comparables. Para asegurar la equivalencia en el estudio, se implementarán las siguientes estrategias:

- **Uso de Múltiples Algoritmos:** Se utilizarán diferentes algoritmos de machine learning para la segmentación y clasificación y se compararán los resultados obtenidos. Esto incluirá el uso de métodos como ExtraTreesClassifier y MLPClassifier (James et al., 2013).
- **Evaluación Cruzada:** Se realizarán evaluaciones cruzadas entre diferentes conjuntos de datos y métodos para asegurar que los resultados sean equivalentes y comparables (Tabachnick & Fidell, 2019).

3.9.5.4 *Precisión de Medición*

La precisión de medición se refiere a la capacidad del instrumento para medir con exactitud y sin errores. Para asegurar la precisión de medición en el estudio, se tomarán las siguientes medidas:

- **Calibración de Instrumentos:** Se calibrarán los instrumentos y software utilizados en el análisis de datos, como QGIS y Python, para asegurar la precisión en la recolección y análisis de datos (Gerring, 2017).
- **Control de Calidad:** Se implementarán procedimientos de control de calidad durante la recolección y procesamiento de datos para minimizar los errores y asegurar la precisión (Hox & Bechger, 2014).

3.9.6 **Análisis de Datos**

3.9.6.1 *Implementación del Algoritmo de Machine Learning*

Se aplica un pipeline en lenguaje Python donde incluye algoritmos de redes neuronales y Extra Trees de la librería Scikit-learn. El archivo que se utiliza en el pipeline es el CSV extraído del KEA que contiene datos geoespaciales preprocesados y etiquetados, los

cuales incluyen la variable independiente como características derivadas de imágenes satelitales, modelos digitales de elevación (MDE) y otras métricas geográficas, junto con la variable dependiente que representa las etiquetas geomorfológicas (las clases). Este archivo sirve como entrada para entrenar y evaluar el modelo de machine learning.

El resultado esperado es un modelo entrenado que puede clasificar correctamente las unidades de relieve (etiquetas geomorfológicas) con base en los datos de entrada. Se evalúa mediante métricas como precisión, recall y F1-score para verificar su rendimiento y aplicabilidad en la generación automatizada de mapas geomorfológicos. El pipeline (herramienta que automatiza y organiza un flujo de trabajo de procesamiento de datos y aprendizaje automático) utilizado es el siguiente:

```
from sklearn.pipeline import make_pipeline, make_union
from sklearn.preprocessing import StandardScaler
from tpot.builtins import StackingEstimator
from sklearn.neural_network import MLPClassifier
from sklearn.ensemble import ExtraTreesClassifier
from sklearn.metrics import classification_report

import pandas as pd

# Leer el archivo CSV
data = pd.read_csv(r'D:\prueba tesis
cdmx\RADIANES\mde_a_s_r_p_cluster10_CDMX.csv')

# Seleccionar las observaciones con al menos 70% de pixeles
clase1 = data[(data['klass'] == 1) & (data['klass__1'] >= 0.7) ]
clase2 = data[(data['klass'] == 2) & (data['klass__2'] >= 0.7) ]
```

```

clase3 = data[(data['klass'] == 3) & (data['klass__3'] >=
0.7) ]
clase4 = data[(data['klass'] == 4) & (data['klass__4'] >= 0.7) ]
clase5 = data[(data['klass'] == 5) & (data['klass__5'] >= 0.7) ]
clase6 = data[(data['klass'] == 6) & (data['klass__6'] >= 0.7) ]
clase7 = data[(data['klass'] == 7) & (data['klass__7'] >= 0.7) ]
clase8 = data[(data['klass'] == 8) & (data['klass__8'] >= 0.7) ]
clase9 = data[(data['klass'] == 9) & (data['klass__9'] >= 0.7) ]
clase10 = data[(data['klass'] == 10) & (data['klass__10'] >= 0.7) ]

#Combinar las observaciones seleccionadas
clases = pd.concat([clase1,
clase2,clase3,clase4,clase5,clase6,clase7,clase8,clase9,clase10])

# Dividir los datos en conjunto de entrenamiento y prueba
from sklearn.model_selection import train_test_split
train_data, test_data = train_test_split(clases, test_size=0.3,
random_state=42)

def fit_and_test(train_data,testing_data, klass, pipeline, nombre_pipeline,
klass_test=''):
    if klass_test == '':
        klass_test = klass

    features =
train_data.drop(["FID",klass,'klass__1','klass__2','klass__3','klass__4','kl
ass__5','klass__6','klass__7','klass__8','klass__9','klass__10'], axis=1)

```

```

target = train_data[klass]

testing_features =
testing_data.drop(["FID",klass_test,'klass__1','klass__2','klass__3','klass_
_4','klass__5','klass__6','klass__7','klass__8','klass__9','klass__10'],
axis=1)

test_target = testing_data[klass_test]

print("inicia entrenamiento")
pipeline.fit(features, target)
print("clasificación")
results = pipeline.predict(testing_features)
print("El resultado de la clasificación con el pipeline:
",nombre_pipeline)

print(classification_report(test_target, results, digits=4))

return pipeline
exported_pipeline = make_pipeline(
    StandardScaler(),
    StackingEstimator(estimator=MLPClassifier(alpha=0.01,
learning_rate_init=0.5)),
    ExtraTreesClassifier(bootstrap=False, criterion="entropy",
max_features=0.8, min_samples_leaf=3, min_samples_split=5, n_estimators=100)
)
klass = 'klass'
nombre_pipeline = "Extra-Trees mde_a_s_r_p_cluster10_CDMX"
fit_and_test(test_data,train_data,klass,exported_pipeline,nombre_pipeline)

```

3.9.6.2 *Explicación del código*

Este código utiliza el algoritmo de Extra Trees, que es un método de machine learning basado en árboles de decisión, para clasificar unidades de relieve geográficas en diez clases diferentes (etiquetadas del 1 al 10). Los datos de entrada incluyen variables geoespaciales como imágenes satelitales y modelos digitales de elevación (MDE), que proporcionan información sobre la topografía y otras características del terreno.

El proceso comienza con la preparación de los datos, que se dividen en un conjunto de entrenamiento y un conjunto de prueba. El modelo de Extra Trees se entrena utilizando el conjunto de entrenamiento, aprendiendo los patrones y relaciones entre las características geoespaciales y las etiquetas de las unidades de relieve. Una vez entrenado, el modelo es capaz de predecir la clase de las unidades de relieve en nuevos datos no vistos, que se encuentran en el conjunto de prueba.

Para evaluar el rendimiento del modelo, se calculan métricas como la precisión, recall y el F1-score para cada clase (classification report). Además, se construye una matriz de confusión para analizar en detalle cómo se clasificaron las diferentes unidades de relieve. Estos resultados permiten identificar fortalezas y debilidades en la capacidad del modelo para clasificar correctamente las unidades de relieve, lo cual es fundamental para la validación y mejora del enfoque propuesto. En general, el código facilita la creación automatizada de mapas geomorfológicos, permitiendo una clasificación eficiente y precisa de las unidades de relieve en comparación con métodos tradicionales más lentos y dependientes de la intervención humana.

3.9.6.3 Evaluación del Modelo:

La evaluación del modelo con este código permite medir cómo el modelo se adapta y generaliza a los datos de prueba utilizando métricas detalladas proporcionadas por `classification_report`. Utilizando el algoritmo de Extra Trees. Estas métricas son esenciales para entender las fortalezas y debilidades del modelo en términos de precisión, recall y F1-score, permitiendo así realizar ajustes necesarios para mejorar su rendimiento. Las métricas de rendimiento como el reporte de clasificación permiten evaluar la precisión y efectividad del modelo machine learning. El pipeline nos da la efectividad general y la de cada etiqueta por separada, así como los números de muestra.

3.9.6.4 Interpretación del Reporte de Clasificación

El `classification_report` produce una salida con los resultados que se muestran en la figura 11:

- Precision: Muestra la precisión de cada clase. Por ejemplo, la clase 6 tiene una precisión de 0.49, lo que significa que el 49% de las predicciones de clase 1 son correctas.
- Recall: La clase 6 tiene un recall de 0.72, lo que significa que el 72% de las instancias de clase 1 fueron correctamente identificadas por el modelo.
- F1-Score: Proporciona una medida equilibrada del rendimiento. La clase 6 tiene un F1-score de 0.58.
- Support: Indica la cantidad de instancias de cada clase en el conjunto de prueba. La clase 6 tiene 380 instancias.

Figura 11

Resultado de aplicar el pipeline con el archivo csv con los datos del kea.

	precision	recall	f1-score	support
1	0.0000	0.0000	0.0000	4
2	0.2500	0.0976	0.1404	41
3	0.2424	0.0748	0.1143	107
4	0.4545	0.2500	0.3226	20
5	0.1341	0.1264	0.1302	87
6	0.4920	0.7263	0.5866	380
7	0.1216	0.1047	0.1125	86
8	0.2500	0.0952	0.1379	21
9	0.3488	0.2128	0.2643	141
10	0.0000	0.0000	0.0000	4
accuracy		0.3872		891
macro avg	0.2294	0.1688	0.1809	891
weighted avg	0.3466	0.3872	0.3463	891

Nota. Elaboración propia.

3.9.7 Procesamiento de la Información

El procesamiento de la información en este estudio se llevará a cabo mediante una serie de pasos estructurados, utilizando herramientas avanzadas de software para garantizar la precisión y eficiencia en el análisis de los datos geográficos y geomorfológicos.

3.9.7.1 Preprocesamiento de Datos

1. Carga y Limpieza de Datos: Los datos del CEM 3.0 y las imágenes satelitales se cargarán en el entorno de trabajo utilizando Python y QGIS. Se realizará una limpieza de datos para eliminar cualquier ruido o anomalía que pueda afectar el análisis.
2. Normalización y Escalado: Los datos geográficos se normalizarán y escalarán utilizando StandardScaler de sklearn para asegurar que todas las variables tengan una

escala comparable, lo cual es crucial para el rendimiento de los algoritmos de machine learning (James, Witten, Hastie, & Tibshirani, 2013).

3. Segmentación de Imágenes: Se utilizará el paquete RGSLIB para segmentar las imágenes satelitales, identificando unidades de relieve específicas. Este proceso es fundamental para el análisis posterior de las características geomorfológicas.

3.9.7.2 *Procesamiento de Datos*

1. Selección de Características: Se identificarán y seleccionarán las características más relevantes de los datos utilizando técnicas de selección de características de sklearn. Esto mejorará la eficiencia y efectividad de los modelos de machine learning.

2. División de Datos: Los datos se dividirán en conjuntos de entrenamiento y prueba utilizando `train_test_split` para evaluar el rendimiento de los modelos de machine learning de manera robusta y evitar el sobreajuste (Pedregosa et al., 2011).

3. Implementación de Algoritmos: Se implementarán los algoritmos de machine learning definidos en el pipeline, utilizando sklearn y TPOT para la automatización y optimización del proceso (Olson & Moore, 2019).

3.9.7.3 *Procesamiento en el modelo*

El procesamiento de la información en este estudio se lleva a cabo mediante una serie de pasos detallados que permiten preparar, entrenar y evaluar un modelo de clasificación basado en datos topográficos y de relieve de la Ciudad de México. Para ello, se utilizan diversas herramientas y bibliotecas de Python, como pandas, scikit-learn y TPOT, integradas en un pipeline de procesamiento de datos y modelado. A continuación, se detallan los pasos seguidos en este proceso:

- Carga y Preparación de Datos:
 - Los datos se cargan desde un archivo CSV utilizando la biblioteca panda. Estos datos contienen diversas características topográficas y de relieve, además de las etiquetas de clasificación (klass) correspondientes a diferentes tipos de unidades geomorfológicas.
 - Se seleccionan las observaciones que tienen al menos un 70% de píxeles pertenecientes a una clase específica. Esto asegura que las observaciones utilizadas en el modelo sean representativas de las clases geomorfológicas.
- Filtrado y Combinación de Clases:
 - Se filtran las observaciones para cada clase utilizando condiciones específicas, y luego se combinan estas observaciones en un solo DataFrame.
- División de Datos:
 - Los datos se dividen en conjuntos de entrenamiento y prueba utilizando la función `train_test_split` de `scikit-learn`. El conjunto de entrenamiento se utiliza para ajustar el modelo, mientras que el conjunto de prueba se utiliza para evaluar su rendimiento.
- Definición del Pipeline:
 - Se define un pipeline que incluye los pasos de preprocesamiento y modelado. En este caso, se utiliza un `StandardScaler` para normalizar las características, un `StackingEstimator` con un `MLPClassifier` y un `ExtraTreesClassifier` como modelo final.
- Entrenamiento y Evaluación del Modelo:

- El pipeline se ajusta a los datos de entrenamiento y se evalúa utilizando los datos de prueba. El rendimiento del modelo se mide mediante métricas como la precisión, el recall y el F1-score, generadas por la función `classification_report`.

3.10 Consideraciones Éticas

La investigación ética es fundamental en cualquier estudio científico. En este estudio, se abordarán las siguientes consideraciones éticas para asegurar la integridad y responsabilidad del trabajo realizado.

3.10.1 Confidencialidad de los Datos

Aunque los datos utilizados en este estudio son geográficos y no personales, se garantizará la confidencialidad de cualquier información que pueda ser sensible o privada. Se utilizarán únicamente datos abiertos y disponibles públicamente del INEGI y otras fuentes reconocidas (INEGI, 2020).

3.10.2 Uso Responsable de Tecnología

El uso de algoritmos de machine learning y software avanzado requiere un manejo responsable para evitar sesgos y errores en el análisis. Se adoptarán las mejores prácticas para la implementación y evaluación de los modelos, asegurando la transparencia y reproducibilidad del estudio (Buchanan & Zimmer, 2019).

3.10.3 Integridad Científica

Se seguirá un riguroso protocolo científico en todas las etapas del estudio, desde la recolección y procesamiento de datos hasta el análisis e interpretación de resultados. Se

documentarán todos los procedimientos y decisiones metodológicas para garantizar la integridad y validez del estudio (Resnik, 2015).

En el Capítulo 3, se detallaron los materiales y métodos utilizados para implementar y evaluar los algoritmos de machine learning, específicamente el modelo de Extra Trees, con el fin de generar mapas geomorfológicos. El uso de datos geoespaciales, como imágenes satelitales y modelos digitales de elevación, fue clave en este proceso, destacando la importancia de la selección y preprocesamiento de datos adecuados. Además, la metodología empleada permitió no solo entrenar los algoritmos de manera efectiva, sino también validar sus resultados mediante métricas de rendimiento como precisión, recall y F1-score. El enfoque experimental y riguroso en esta fase garantiza que los resultados obtenidos sean sólidos y puedan compararse de manera objetiva, estableciendo las bases para la evaluación posterior.

Capítulo IV Resultados

En el capítulo IV, se detallan las pruebas llevadas a cabo utilizando Extra Trees de Machine Learning en el contexto de la clasificación del relieve. Estos algoritmos se implementaron utilizando el lenguaje de programación Python y las librerías RGSLIB y Scikit-learn. A continuación, se describen los resultados obtenidos de manera exhaustiva, con el objetivo de determinar cuál de los algoritmos resulta más eficaz en la clasificación del relieve. Se lleva a cabo un análisis comparativo entre los diferentes modelos evaluados para identificar sus fortalezas y limitaciones en este contexto específico. Finalmente, se presentan y explican los productos derivados de la evaluación y validación de la solución propuesta para la clasificación del relieve. Este análisis proporciona una visión detallada de los logros alcanzados en el estudio, así como de las posibles aplicaciones prácticas de los resultados obtenidos

4.1 Resultados generales del modelo

Los resultados generales del modelo de machine learning con Extra Trees para la creación automatizada de mapas geomorfológicos evidencian un desempeño mixto, con algunas clases geomorfológicas identificadas con alta precisión y otras con una mayor tasa de confusión.

Por términos de practicidad y de manejar una manera más clara los resultados, se optó por mostrar solamente una sola configuración en los parámetros de la segmentación, que fue número de clústeres 10 (`num_clusters`), umbral de distancia 10 (`dist_thres`) y mínimo de pixeles por segmento 100 (`min_n_pxls`) (ver apéndice 2). De igual manera estos parámetros de segmentación fueron los que se consideraron mejor para la investigación por dos razones, la primera, por el número de muestras generadas y en segundo, porque genera más detalles en comparación a los otros parámetros que a pesar de obtener más precisión en el modelo no detectaban distintas unidades de relieve o en su defecto las mezclaba.

En términos de precisión global, los resultados muestran que la etiqueta 1 fue consistentemente la mejor identificada. Esta clase, que corresponde a una unidad geomorfológica claramente diferenciada, permitió a los algoritmos distinguirla con una precisión y un recall elevados en todos los conjuntos de datos utilizados. Del mismo modo, las etiquetas 6 y 9 también fueron bien clasificadas, lo que sugiere que las características de las unidades de relieve representadas por estas clases son adecuadas para los modelos de machine learning aplicados, debido a su alta diferenciación topográfica y clara segmentación.

Por otro lado, el modelo presentó dificultades para identificar algunas etiquetas con precisión, en particular aquellas con transiciones graduales o similitudes geomorfológicas entre diferentes unidades. Las etiquetas 4, 5 (26% de recall en la primera tabla y 25% en la segunda), y 7 fueron las más problemáticas, con bajas tasas de precisión y recall. Estos resultados sugieren que las características geomorfológicas de estas clases son más ambiguas o que presentan patrones difíciles de distinguir, lo que llevó a un mayor nivel de confusión con otras etiquetas.

El análisis de las matrices de confusión respalda estos hallazgos como se muestra en la figura 13, mostrando que, en algunos casos, como la etiqueta 3, el modelo confunde significativamente las observaciones con otras clases. En la primera tabla, por ejemplo, la etiqueta 3 fue confundida en numerosas ocasiones con la etiqueta 5, lo que refleja similitudes entre las unidades de relieve en términos de pendientes o elevaciones medias.

Además, es importante señalar que el preprocesamiento de los datos y la calidad de las imágenes satelitales y los modelos de elevación digital jugaron un rol clave en el desempeño del modelo. Aunque los algoritmos de machine learning lograron automatizar de manera eficiente la creación de mapas geomorfológicos, aún se observan diferencias

significativas en la precisión y el detalle de los mapas generados en comparación con los mapas realizados manualmente por expertos, especialmente en áreas más complejas.

En resumen, el modelo mostró un desempeño sobresaliente en clases geomorfológicas bien definidas y diferenciadas (como las representadas por las etiquetas 1, 6 y 9), pero evidenció limitaciones en la clasificación de etiquetas con patrones similares o transiciones difusas entre unidades de relieve. Estos resultados destacan la necesidad de seguir mejorando el preprocesamiento de los datos y ajustar los algoritmos para manejar mejor las clases más ambiguas.

A continuación, se muestran en la tabla 3 los resultados de forma general de las pruebas que se consideraron más significativas para el proyecto, que muestran de cierta la forma la evolución de cómo se fue experimentando y de forma simplificada en la tabla para dar un panorama rápido, y para explicar de forma más detallada se desglosaran los puntos más importantes de cada prueba. Como se puede observar en la tabla 3, se hicieron bastantes pruebas, pero las que se muestran fueron las más significativas para mostrar el avance en la técnica, se empezó con la misma segmentación y se fue agregando capas. En las últimas pruebas se cambió la segmentación para haber que tanto cambiaba la precisión del algoritmo.

Los resultados presentados en la tabla muestran un análisis comparativo de múltiples pruebas de segmentación para la creación de mapas geomorfológicos utilizando diferentes configuraciones en el número de capas y parámetros de segmentación. La precisión varía significativamente dependiendo de los valores de `num_clusters`, `dist_thres` (umbral de distancia) y `min_n_pxls` (mínimo de píxeles por segmento), así como del número de capas utilizadas.

Tabla 3

Resultados generales de las pruebas con Extra Trees.

Número de prueba	Número de capas	Precisión	Muestras
1	3	0.3872	13659
2	5	0.4018	13839
3	8	0.4703	43681
4	8	0.5063	103209
5	14	0.5507	103209
6	62	0.6379	103209

Nota. Elaboración propia

Se clasificaron en 6 pruebas según el número de capas y una prueba para analizar que tanto incluía en el resultado la normalización de las capas (prueba 3)

- Prueba 1 (con 3 capas): La prueba 1 muestra una precisión relativamente baja. Esto sugiere que el uso de pocas capas puede no ser suficiente para capturar la complejidad geomorfológica de los datos.
- Prueba 2 (con 5 capas): El rendimiento general mejoró ligeramente con la inclusión de la capa relieve y pendiente. El F1-score global fue de 0.556, reflejando un equilibrio en la clasificación de las clases más representadas, aunque las clases con menos muestras continuaron presentando una menor precisión.
- Prueba 3 (con 8 capas normalizadas): Se observó un aumento en la precisión, donde también aumentó casi tres veces más el número de muestras, incluyendo las capas TPI, TRI y Hillshade, donde se normalizaron las capas para ver como influencia los resultados.

- Prueba 4 (con 8 capas): Se optó por no normalizar la misma prueba (prueba 3), ya que se ve que se pierden detalles y muestras, además que al no normalizar aumenta ligeramente la precisión. Sin embargo, algunas clases con menor soporte aún presentaron resultados menos satisfactorios.
- Prueba 5 (con 14 capas): Las pruebas con 14 capas presentan resultados consistentes, se agregan las 6 bandas de la geomediana, lo que sugiere que un mayor número de capas combinadas con ajustes adecuados en la segmentación puede mejorar la identificación geomorfológica.
- Prueba 6 (con 62 capas): Finalmente, las pruebas con 62 capas muestran una mejora notable en la precisión. Las pruebas 6 tiene precisiones superiores a 0.6379. Estas pruebas confirman que un mayor número de capas, junto con un ajuste adecuado en los parámetros de segmentación, mejora el rendimiento del modelo.

Por lo tanto, los mejores resultados se obtuvieron en las pruebas con mayor número de capas, por otro lado, se logró tener una precisión de hasta 80% con menos capas, pero con una segmentación muy general, con muy pocas muestras y donde solo se identifican una o dos unidades por lo que ese tipo de segmentaciones generales no se optaron en esta investigación para identificar el mayor número de unidades de relieve posible. Debido a que el mejor resultado fue la prueba 6, es decir la que tiene mayor número de capas, sugiere que tanto la cantidad de capas como los parámetros de segmentación juegan un papel clave en la precisión del modelo geomorfológico.

4.1.1 Prueba 1 (3 capas)

Siendo un ráster con 3 capas, el modelo digital de elevación (MDE), aspecto y perfil de curvatura en ese orden la composición de las bandas se optó por la segmentación con la

configuración de número de clúster 10, `disth_thres` 10 y `min_pxls` 100, donde el resultado se muestra en la tabla 4.

La tabla presentada ofrece un análisis detallado de las métricas de rendimiento (precisión, recall y F1-Score) para diferentes clases en el modelo, lo cual es clave para evaluar su efectividad en la identificación de las unidades de relieve geomorfológicas. El modelo no pudo identificar correctamente las clases 1 y 10, ya que tanto la precisión, el recall y el F1-Score son 0. Esto indica que no se clasificaron correctamente ninguna de las muestras de estas clases, lo que sugiere que el modelo tiene dificultades para aprender los patrones asociados a estas unidades geomorfológicas o que las muestras de estas clases son demasiado pequeñas (ambas tienen solo 61 instancias en el soporte).

Las clases 2, 3, 4, 5, 7, 8 y 9 presentan un rendimiento bajo o moderado. En términos de F1-Score, que es una métrica balanceada entre precisión y recall, las mejores entre estas son:

- Clase 6, con un F1-Score de 0.5866, lo que indica que el modelo pudo identificar mejor esta clase. También tiene el mayor soporte, con 5825 instancias, lo que sugiere que el tamaño de la muestra juega un rol importante en el desempeño del modelo.
- Clase 9 tiene un F1-Score de 0.2643, con un soporte de 322 instancias, lo que aún está por debajo de lo ideal, pero es superior a las otras clases de bajo rendimiento.

El promedio no ponderado para cada clase (Macro Avg), es decir, trata a todas las clases por igual. El F1-Score promedio es 0.1809, lo que indica un rendimiento bajo en general, debido a la incapacidad del modelo de identificar correctamente muchas de las clases.

El Promedio ponderado por el soporte de cada clase (Weighted Avg), es el indicador que mejor para evaluar el rendimiento global del modelo en función de la cantidad de instancias de cada clase. El F1-Score ponderado es 0.3463, lo que muestra que el modelo tuvo un rendimiento algo mejor en las clases con más instancias, pero aún es insuficiente. El Accuracy de 0.3872 refleja la proporción total de predicciones correctas sobre todas las muestras. Esto confirma que el modelo tiene dificultades para realizar predicciones precisas para la mayoría de las clases.

Clase 6 es la mejor identificada, con una precisión de 0.492 y un recall de 0.7263, lo que significa que el modelo es capaz de detectar correctamente la mayoría de las instancias de esta clase. Las clases 1, 7 y 10 fueron las peor clasificadas, con F1-Scores muy bajos, lo que indica que el modelo necesita un ajuste significativo para mejorar la clasificación de estas clases. El Balance de clases muestra el desequilibrio en el tamaño de las clases, especialmente las clases 6 y 5 (con un mayor número de muestras), ha influido en la ponderación de los resultados. Esto sugiere que el modelo podría beneficiarse de técnicas de balanceo de datos, como sobre muestreo o submuestreo de clases minoritarias. En general, los resultados indican que el modelo necesita mejoras, particularmente en la identificación de las clases con menos soporte y en la reducción de los sesgos hacia clases más grandes como la clase 6. En resumen, con esta aproximación hubo una precisión de 38.7% con 13659 muestras.

En el modelo marcado en el procedimiento indica que el script usó un modelo ExtraTreesClassifier para la clasificación de todas las pruebas. Observando la clase 6, correspondiente a la etiqueta de pendiente (slope), con soporte de 5825 muestras, es la que hubo más afinidad, más específicamente se puede ver:

- Precisión: 0.4920 - De todas las muestras que el modelo predijo como clase 6, el 49.2% fueron realmente de clase 6.
- Revocación: 0.7263 - El modelo identificó correctamente el 72.63% de las muestras reales de la clase 6.
- F1-score: 0.5866 - Combinando la precisión y la revocación.

Tabla 4

Resultados prueba 1

Clase	Precisión	Recall	F1-Score	Soporte
1	0	0	0	61
2	0.25	0.0971	0.1404	628
3	0.2308	0.1121	0.1509	1641
4	0.2143	0.15	0.1765	307
5	0.1446	0.1379	0.1412	1334
6	0.492	0.7263	0.5866	5825
7	0.1111	0.0465	0.0656	86
8	0.25	0.0952	0.1379	1319
9	0.3452	0.2165	0.2643	322
10	0	0	0	61
Métrica	Resultado			
Accuracy	0.3872			
Macro Avg	0.2294 (Precisión)	0.1688 (Recall)	0.1809 (F1-Score)	
Weighted Avg	0.3466 (Precisión)	0.3872 (Recall)	0.3463 (F1-Score)	

Nota. Elaboración propia

4.1.2 Prueba 2 (5 capas)

El procedimiento anterior se realizó con 3 bandas, (MDE, aspecto y perfil de curvatura) para esta prueba se le agregaron 2 bandas (Relieve y pendiente), teniendo la siguiente configuración de bandas:

- band 1 mde
- band 2 aspect
- band 3 profile curvature
- band 4 relief
- band 5 slope

Usando el mismo procedimiento y segmentación que en la prueba 1 el modelo muestra un rendimiento similar al de la Prueba 1, aunque con ligeras mejoras en algunas clases y métricas generales. El desempeño sigue siendo inconsistente a lo largo de las distintas clases geomorfológicas reflejadas en la tabla 5. Al igual que en la prueba anterior, las clases 1 y 10 no fueron correctamente clasificadas, ya que presentan valores de precisión, recall y F1-Score de 0, lo que indica que el modelo no identificó correctamente ninguna instancia de estas clases. Estas clases también tienen un número reducido de muestras (61 instancias cada una), lo que sigue representando un desafío para el modelo.

La clase 6, que tiene la mayor cantidad de muestras (5825 instancias), nuevamente es la mejor identificada, con una precisión de 0.5298 y un F1-Score de 0.6041. Aunque el rendimiento de esta clase es alto en comparación con las demás, se observa una ligera disminución en el recall, pasando de 0.7263 en la Prueba 1 a 0.7026 en esta prueba, lo que indica que el modelo no logró identificar tantas instancias correctas de la clase en comparación con la prueba anterior.

En cuanto a las clases con rendimientos intermedios, la clase 9 muestra una mejora notable, con un F1-Score de 0.3656, lo que representa un aumento en comparación con la Prueba 1. Esto sugiere que el modelo fue más preciso en identificar esta clase en la Prueba 2,

posiblemente debido a un mejor ajuste de los parámetros de segmentación. Sin embargo, otras clases como la 2 y la 5 presentan mejoras muy leves o casi nulas en sus métricas, lo que indica que el modelo sigue luchando por aprender los patrones correspondientes a estas clases. Las clases 7 y 8, con pocas muestras, también muestran un rendimiento deficiente, con F1-Scores de 0.0656 y 0.069, respectivamente, lo que refuerza la necesidad de mejorar la capacidad del modelo para manejar clases minoritarias.

En cuanto a los promedios, el Macro Avg F1-Score es de 0.1778, lo que refleja un rendimiento bajo y uniforme en las clases. Sin embargo, el Weighted Avg F1-Score es de 0.3688, lo que supone una ligera mejora en comparación con la Prueba 1. Este valor ponderado por el número de instancias sugiere que el modelo sigue funcionando mejor en las clases con más muestras. El accuracy global también mejoró ligeramente, alcanzando 0.4018, lo que indica que el modelo es capaz de clasificar correctamente un poco más de instancias en comparación con la prueba anterior, aunque el rendimiento general sigue siendo insuficiente.

En resumen, la Prueba 2 muestra mejoras marginales en algunas clases y en las métricas generales, pero el modelo sigue enfrentando dificultades significativas para clasificar correctamente las clases con menos muestras. Las mejoras observadas en la clase 9 y el ligero incremento en la precisión global sugieren que algunos ajustes en los parámetros de segmentación fueron efectivos, aunque persisten limitaciones importantes. Con las mismas muestras (13659) pero agregando las dos capas aumento aproximadamente de 2% de precisión, siendo igualmente la clase 6 la mejor clasificada.

Tabla 5*Resultados prueba 2*

Clase	Precisión	Recall	F1-Score	Soporte
1	0	0	0	61
2	0.2162	0.1951	0.2051	628
3	0.2308	0.1121	0.1509	1641
4	0.2143	0.15	0.1765	307
5	0.1446	0.1379	0.1412	1334
6	0.5298	0.7026	0.6041	5825
7	0.1111	0.0465	0.0656	86
8	0.125	0.0476	0.069	1319
9	0.3696	0.3617	0.3656	322
10	0	0	0	61
Métrica	Resultado			
Accuracy	0.4018			
Macro Avg	0.1941 (Precisión)	0.1754 (Recall)	0.1778 (F1-Score)	
Weighted Avg	0.3547 (Precisión)	0.4018 (Recall)	0.3688 (F1-Score)	

Nota. Elaboración propia

4.1.3 Prueba 3 normalizado (Ráster con 8 capas)

Esta vez sí se normalizó todas las capas (ver Apéndice 1), para las capas slope, profile curvature y aspecto, se tienen que convertir en radianes porque su información está en grados, al normalizarlos sin convertirlos a radianes los valores resultantes eran -360 y 360 al momento de ponerlos en el QGIS no se visualiza, el detalle en este aspecto es que convertirlo en radianes y normalizar los solo se visualiza blanco y negro con valores 0 y 1-.

Se agregaron tres capas más, hillshade, TRI y TPI, la estructura de esta prueba con 8 capas queda de la siguiente manera:

- Banda 1 mde

- Banda 2 relieve
- Banda 3 aspect
- Banda 4 hillshade
- Banda 5 profile curvature
- Banda 6 slope
- Banda 7 TPI
- Banda 8 TRI

4.1.3.1 Prueba 3

En la Prueba 3, misma segmentación (clúster 10, thres 100 y px 100) que la anterior, pero con capas normalizadas y con las 8 capas, el modelo muestra un rendimiento mejorado en comparación con las pruebas anteriores, especialmente en algunas clases que ahora están mejor representadas. El rendimiento general ha mejorado, con un accuracy de 0.4703, lo que indica una mayor capacidad del modelo para identificar correctamente más instancias en el conjunto de dato, como se observa en la tabla 6.

La clase mejor identificada es la 1, con una precisión de 0.6858, un recall de 0.8315 y un F1-Score de 0.7516, lo que refleja una identificación mucho más precisa y eficiente en comparación con pruebas anteriores. Esta clase tiene 6866 muestras, lo que probablemente ayuda al modelo a aprender los patrones correctamente. La clase 6, que tiene la mayor cantidad de muestras (14903), también muestra un rendimiento sólido con una precisión de 0.4908 y un F1-Score de 0.5961. Aunque no es tan alta como la clase 1, sigue siendo una de las mejores identificadas por el modelo.

Por otro lado, las clases con menor número de muestras, como la clase 10, muestran un desempeño deficiente, con una precisión de 0.25 y un F1-Score de 0.0331, lo que indica

que el modelo no fue capaz de identificar correctamente las instancias de esta clase. Otras clases como la 2, 3 y 4 también tienen un rendimiento limitado, con F1-Scores que oscilan entre 0.1628 y 0.2399, lo que sugiere que el modelo sigue enfrentando dificultades para aprender los patrones correspondientes a estas clases. Las clases 7, 8 y 9, que tienen un número intermedio de muestras, muestran un rendimiento moderado, con F1-Scores entre 0.1118 y 0.265. El rendimiento de estas clases sugiere que el modelo está comenzando a capturar algunos patrones, pero aún hay margen de mejora.

Tabla 6

Resultados prueba 3

Prueba 3.1				
Clase	Precisión	Recall	F1-Score	Soporte
1	0.6858	0.8315	0.7516	6866
2	0.3155	0.1097	0.1628	538
3	0.3098	0.1958	0.2399	3361
4	0.3168	0.1266	0.1809	1619
5	0.2566	0.1312	0.1737	5037
6	0.4908	0.7588	0.5961	14903
7	0.1937	0.0786	0.1118	3970
8	0.318	0.2171	0.258	3464
9	0.3131	0.2297	0.265	3810
10	0.25	0.0177	0.0331	113
Métrica	Resultado			
Accuracy	0.4703			
Macro Avg	0.3450 (Precisión)	0.2697 (Recall)	0.2773 (F1-Score)	
Weighted Avg	0.4151 (Precisión)	0.4703 (Recall)	0.4225 (F1-Score)	

Nota. Elaboración propia

En cuanto a los promedios, el Macro Avg F1-Score es de 0.2773, lo que refleja una mejora general en comparación con pruebas anteriores, aunque sigue siendo un rendimiento bajo

cuando se promedian las clases de manera uniforme. El Weighted Avg F1-Score de 0.4225, ponderado por el número de muestras, muestra una mejora considerable en comparación con las pruebas anteriores, lo que indica que el modelo está funcionando mejor en clases con más muestras. En resumen, la Prueba 3.1 muestra una mejora notable en términos de precisión y rendimiento general del modelo, especialmente en clases con mayor número de muestras como la 1 y la 6. Sin embargo, persisten dificultades importantes en la identificación de clases con menos muestras o con patrones más complejos.

4.1.4 Prueba 4 sin normalizar (Ráster con 8 capas)

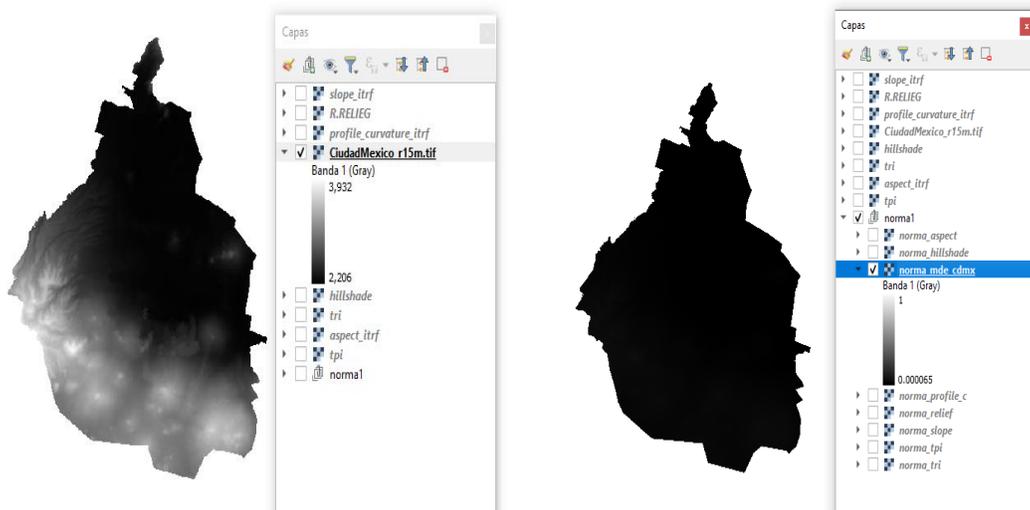
La siguiente prueba se mantiene la estructura anterior con 8 bandas, sin embargo, se juntaron las capas sin normalizar porque al momento de normalizar los ráster al visualizarlos se pierden bastantes detalles que son determinantes para hacer un mejor análisis, en la figura 12, el ejemplo del MDE normalizado y sin normalizar. En la Prueba 4.1, la configuración fue clúster 10, disth_thres 100 y min_pxls 100 sin normalizar las capas, donde el modelo alcanzó una precisión general de 0.5063, lo que indica que aproximadamente la mitad de las predicciones fueron correctas. Sin embargo, el rendimiento varía significativamente entre las diferentes clases, como se observa en la tabla 7.

4.1.4.1 Prueba 4.1

La clase 1 es la mejor identificada, con una precisión de 0.7545, un recall de 0.8853 y un F1-Score de 0.8147 sobre 15,756 muestras. Este resultado muestra que el modelo tiene una alta capacidad para identificar correctamente esta clase, aunque todavía hay margen de mejora. La clase 6, con 34,284 muestras, también se desempeñó bien con un F1-Score de 0.605 y un recall de 0.8157, lo que indica que el modelo fue capaz de capturar la mayoría de los casos de esta clase, pero con menor precisión que la clase 1.

Figura 12

MDE sin normalizar izquierda, MDE normalizado derecha



Nota. Elaboración propia

Por otro lado, clases como la clase 2 y la clase 7 presentan un rendimiento considerablemente bajo. La clase 2 tiene una precisión de 0.4646, pero un recall de 0.0928, lo que se traduce en un F1-Score de 0.1546, indicando que el modelo rara vez logra identificar correctamente instancias de esta clase. De manera similar, la clase 7 tiene un recall de 0.092 y un F1-Score de 0.1405, sugiriendo un problema para identificar instancias de esta clase de manera consistente.

Las clases 3, 4, 5, 8 y 9 muestran un desempeño intermedio, con F1-Scores que oscilan entre 0.1894 y 0.3639. Esto indica que el modelo tiene dificultades para capturar estas clases con precisión, aunque puede identificar algunas instancias. Por ejemplo, la clase 9, con 10,327 muestras, tiene un F1-Score de 0.3639 y un recall de 0.3067, lo que sugiere que el modelo es capaz de identificar un número considerable de ejemplos, pero con errores.

Finalmente, la clase 10, con solo 512 muestras, es la peor identificada con un F1-Score de 0.0471 y un recall de 0.0254, lo que indica que el modelo apenas es capaz de identificar correctamente esta clase. El Macro Avg F1-Score es de 0.3202, lo que refleja el promedio del rendimiento por clase, incluyendo aquellas con pocas muestras, mientras que el Weighted Avg F1-Score es de 0.4545, ponderado por el número de instancias por clase, lo que sugiere que el rendimiento general del modelo es algo más favorable para las clases más frecuentes. En resumen, la Prueba 4 muestra que el modelo funciona bien en las clases con más datos, como la clase 1 y la clase 6, pero tiene dificultades para identificar correctamente clases con menos muestras, como la clase 10 o la clase 7, lo que sugiere que el modelo se beneficia de una mayor cantidad de datos para mejorar su rendimiento.

Tabla 7

Resultados prueba 4

Clase	Precision	Recall	F1-Score	Support
1	0.7545	0.8853	0.8147	15756
2	0.4646	0.0928	0.1546	1132
3	0.3877	0.2456	0.3007	7990
4	0.3932	0.1962	0.2617	2967
5	0.3298	0.1329	0.1894	12381
6	0.4808	0.8157	0.605	34284
7	0.2969	0.092	0.1405	10549
8	0.4308	0.26	0.3243	7311
9	0.4474	0.3067	0.3639	10327
10	0.325	0.0254	0.0471	512
Accuracy	0.5063	-	-	103209
Macro Avg	0.4311	0.3052	0.3202	103209
Weighted Avg	0.4681	0.5063	0.4545	103209

Nota. Elaboración propia

4.1.5 Prueba 5 (agregando la geomédiana)

Para esta prueba se agregaron las 6 bandas de la geomédiana al ráster, teniendo un total de 14 capas. Retomando que la geomédiana del INEGI es una serie de mosaicos denominados Geomédiana Landsat, se construye mediante el procesamiento y análisis de todas aquellas imágenes de los satélites Landsat

4.1.5.1 Prueba 5

En la Prueba 5.1, con la misma segmentación clúster 10, thres 100 y px 100 es el siguiente el modelo alcanzó una precisión general de 0.5507, lo que indica un desempeño moderado en la clasificación de las clases, mostrado en la tabla 8. La clase 1 fue la mejor identificada, con una precisión de 0.7569, un recall de 0.913 y un F1-Score de 0.8277 sobre 15756 muestras. Este resultado indica que el modelo pudo identificar correctamente la mayoría de las instancias de esta clase.

Otras clases, como la clase 6 con 34284 muestras, también mostraron un buen rendimiento, alcanzando un F1-Score de 0.6368 y un recall de 0.8625, lo que sugiere que el modelo fue capaz de capturar la mayoría de las instancias de esta clase, aunque con algunos errores. La clase 9, con 10327 muestras, tuvo un desempeño razonable, con un F1-Score de 0.4745 y un recall de 0.4127, mostrando que el modelo identificó más de la mitad de las instancias, aunque no sin errores. En contraste, clases como la clase 2 y clase 7 presentaron problemas. La clase 2, con 1132 muestras, obtuvo un F1-Score de 0.2068 y un recall de 0.1254, lo que indica que el modelo tuvo dificultades para identificar esta clase correctamente. Del mismo modo, la clase 7, con 10549 muestras, obtuvo un F1-Score de 0.1837 y un recall de 0.1172, lo que muestra una baja capacidad de identificación. La clase

10, con 512 muestras, fue la peor identificada, con un F1-Score de 0.0585 y un recall de 0.0312, lo que refleja un rendimiento muy bajo en esta clase.

En cuanto a las métricas promediadas, el Macro Avg F1-Score fue de 0.3682, lo que refleja el rendimiento promedio por clase, mientras que el Weighted Avg F1-Score fue de 0.5, ponderado por el número de instancias en cada clase, lo que sugiere un mejor desempeño en las clases con más datos. En este caso 65% de precisión con 14 bandas como se puede observar al agregar la geomédiana aumentó un 5% de la precisión, en comparación con la prueba 4.1 que tiene la misma segmentación y mismo número de muestras. En resumen, el modelo tiene un rendimiento aceptable en clases con muchas muestras, como la clase 1 y clase 6, pero tiene dificultades en clases con menos representaciones, como la clase 2, clase 7 y clase 10, lo que sugiere que la capacidad de generalización del modelo aún tiene margen de mejora.

Tabla 8

Resultados prueba 5

Clase	Precision	Recall	F1-Score	Support
1	0.7569	0.913	0.8277	15756
2	0.5892	0.1254	0.2068	1132
3	0.4794	0.2927	0.3635	7990
4	0.5093	0.2218	0.309	2967
5	0.4661	0.1703	0.2495	12381
6	0.5048	0.8625	0.6368	34284
7	0.4255	0.1172	0.1837	10549
8	0.5167	0.2905	0.3719	7311
9	0.5581	0.4127	0.4745	10327
10	0.4571	0.0312	0.0585	512
Accuracy	0.5507	-	-	103209
Macro Avg	0.5263	0.3437	0.3682	103209
Weighted Avg	0.5356	0.5507	0.5	103209

Nota. Elaboración propia

4.1.6 Prueba 6 (agregando texturas)

Como se mencionaba en el capítulo 2 y 3 se aplicó un algoritmo que saca 48 texturas a partir del MDE (atributo más determinante) y se agregaron al ráster para tener más elementos que enriquezcan al algoritmo, teniendo así 64 capas, donde efectivamente incrementó la precisión.

4.1.6.1 Prueba 6

El resultado de la Prueba 6. (segmentación clúster 10, thres 100 y px 100) refleja un rendimiento general aceptable del modelo, con una precisión global del 63.9%. Este desempeño varía entre las diferentes clases de unidades de relieve, como se observa en la tabla 9. La clase 1, que cuenta con el mayor número de muestras (15,756), mostró un excelente desempeño, con una precisión de 0.822, un recall de 0.9284 y un F1-score de 0.872. Esto indica que el modelo es capaz de identificar correctamente la mayoría de los elementos de esta clase, con pocas omisiones o errores.

En cuanto a la clase 6 (34,284 muestras), el modelo alcanzó una precisión de 0.5947 y un F1-score de 0.6973, lo que sugiere que tiene una buena capacidad para clasificar esta clase, aunque con margen para mejorar la precisión. El recall de 0.8426 indica que el modelo recupera la mayor parte de las instancias de esta clase. La clase 9, con 10,327 muestras, obtuvo un F1-score de 0.6306 y un recall de 0.6167, reflejando una capacidad relativamente equilibrada para clasificar correctamente esta clase.

Sin embargo, algunas clases mostraron limitaciones en el desempeño del modelo. La clase 2, con 1,132 muestras, presentó una precisión de 0.763, pero el recall fue de 0.3754, lo que implica que el modelo identificó correctamente una proporción significativa de esta clase, pero aún se necesita mejorar la capacidad de recuperación de instancias. Su F1-score de

0.5033 refleja esta discrepancia. La clase 5, con 12,381 muestras, tuvo un F1-score de 0.3857, debido a una baja precisión de 0.5662 y un recall de 0.2925, lo que indica que el modelo tiene dificultades para clasificar correctamente las instancias de esta clase. De particular interés es el desempeño de la clase 10, con solo 512 muestras. A pesar de una precisión de 0.7037, el recall de 0.1484 y el F1-score de 0.2452 sugieren que el modelo no logra identificar de manera confiable las instancias de esta clase, lo que podría deberse a la baja representación de esta clase en los datos.

En términos generales, el modelo mostró un rendimiento sólido en las clases con mayor cantidad de muestras, como la clase 1 y la clase 6, con F1-scores de 0.872 y 0.6973, respectivamente. El F1-score macro promedio fue de 0.5342, lo que indica un desempeño moderado en promedio entre todas las clases, reflejando una brecha en el reconocimiento de las clases menos representadas. El Weighted Avg F1-score de 0.613 sugiere que el modelo tiene un desempeño más robusto en las clases dominantes, pero tiende a fallar en las clases con menos soporte, lo que podría mejorarse ajustando el balance de clases o explorando técnicas más avanzadas de clasificación.

En la figura 14 está el mapa resultante de las unidades identificadas donde predomina la planicie y el talud, que son las etiquetas 1 y 6 respectivamente, en la mayoría de las pruebas esas fueron las unidades mejor identificadas. En la figura 15 se observa en color verde las muestras que identificó correctamente el modelo de Extra Trees según la etiqueta dada. Se puede concluir fácilmente en base a la figura 15 que el modelo se le dificultó identificar las unidades donde el relieve era muy irregular o donde había cambios abruptos en la pendiente.

Tabla 9*Resultados prueba 6*

Clase	Precision	Recall	F1-Score	Support
1	0.822	0.9284	0.872	15756
2	0.763	0.3754	0.5033	1132
3	0.5784	0.5504	0.5641	7990
4	0.6201	0.5062	0.5574	2967
5	0.5662	0.2925	0.3857	12381
6	0.5947	0.8426	0.6973	34284
7	0.5487	0.2138	0.3077	10549
8	0.6564	0.5181	0.5791	7311
9	0.6452	0.6167	0.6306	10327
10	0.7037	0.1484	0.2452	512
Accuracy	0.639	-	-	103209
Macro Avg	0.6498	0.4993	0.5342	103209
Weighted Avg	0.6325	0.639	0.613	103209

Nota. Elaboración propia

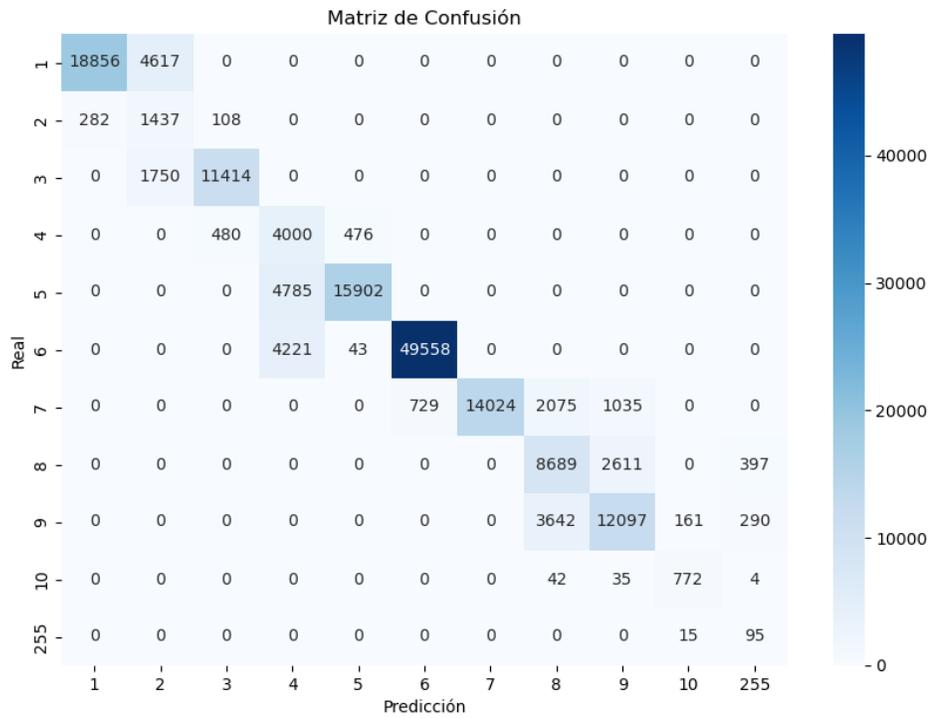
El análisis la variable dependiente ('klass') y las variables independientes mostró que las características topográficas, como la altitud (elevación) y la pendiente, tienen una relación fuerte y significativa con la clasificación de las unidades de relieve. La altitud presentó una correlación positiva alta con la variable dependiente, lo que indica que las variaciones en la elevación del terreno son un factor determinante en la definición de las unidades geomorfológicas. Asimismo, la pendiente mostró una correlación significativa, sugiriendo que la inclinación del terreno influye directamente en la clasificación de las unidades de relieve.

Lo anterior se manifiesta en la clase 1 y 6 que son las que tienen las métricas más altas tanto en precisión, el recall y el F1-score, generadas por la función `classification_report`. Estos hallazgos refuerzan la importancia de las características geomorfológicas del terreno,

como la elevación y la pendiente, en el proceso de clasificación y la generación automatizada de mapas geomorfológicos.

Figura 13

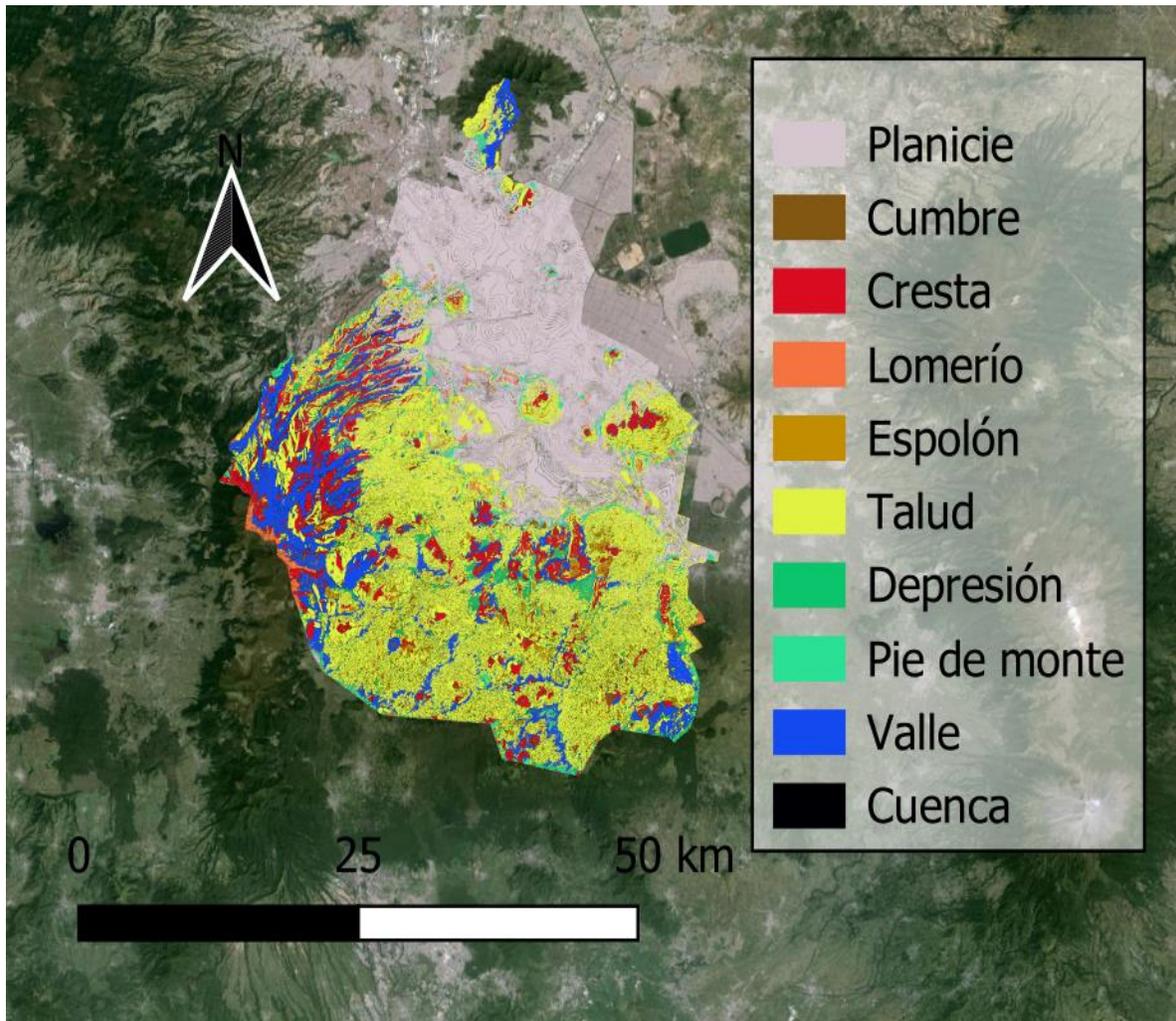
Matriz de confusión prueba 6.



Nota. Elaboración propia

Figura 14

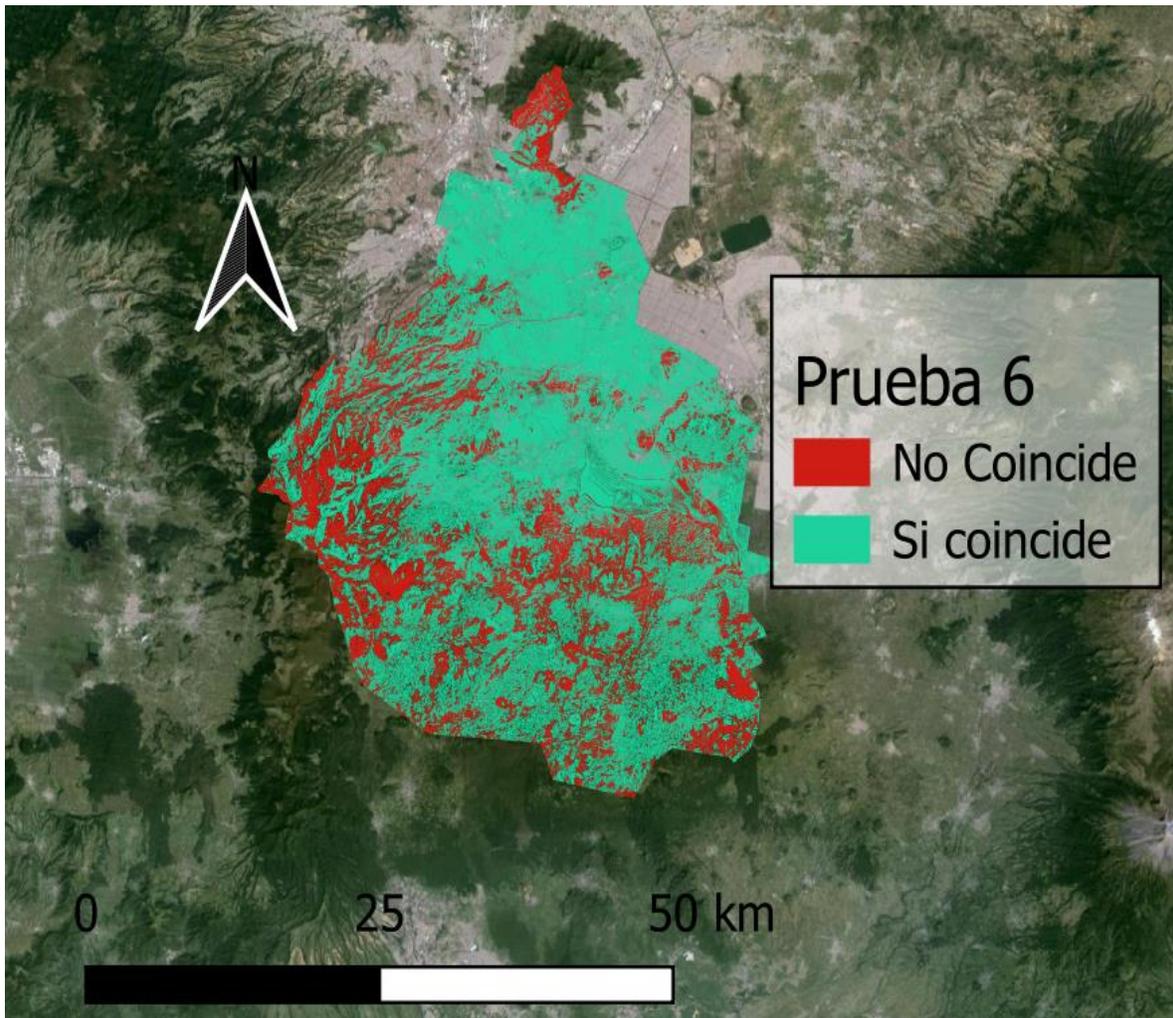
Mapa resultante de las unidades identificadas de la prueba 6



Nota. Elaboración propia

Figura 15

Mapa donde se compara las etiquetas con lo que identificó el algoritmo de la prueba 6



Nota. Elaboración propia

En el Capítulo 4, se presentaron los resultados obtenidos tras la implementación del algoritmo de Extra Trees y su comparación con otros modelos. Los resultados indicaron que este algoritmo fue capaz de clasificar las unidades de relieve con un alto grado de precisión, particularmente en las clases más representativas. Sin embargo, también se observó que el rendimiento variaba según la complejidad de las clases y los volúmenes de datos. Las

métricas como el F1-score y la exactitud reflejaron la efectividad del modelo, superando en algunos aspectos a otros algoritmos probados. A pesar de los logros, se identificaron limitaciones, como la necesidad de mejorar la clasificación en clases menos representadas. En general, el modelo mostró un rendimiento sólido y demostró ser una herramienta útil para la generación automatizada de mapas geomorfológicos.

CAPÍTULO V: DISCUSIÓN

En este capítulo se presenta un análisis exhaustivo de los resultados obtenidos en el presente estudio, comparándolos con investigaciones previas que guardan relación directa con la temática abordada. A través de este análisis comparativo, se destacan tanto las similitudes como las diferencias, aportando una comprensión más profunda de los hallazgos obtenidos y las contribuciones realizadas por este trabajo.

Se discuten los hallazgos en relación con los objetivos establecidos al inicio de la investigación, verificando si se han cumplido o si se requiere un mayor análisis. Además, se contrastan las respuestas obtenidas con las preguntas de investigación planteadas, evaluando si los resultados satisfacen las expectativas y contribuyen a responder de manera sólida y fundamentada a dichas preguntas. Finalmente, se aborda la hipótesis planteada, describiendo de manera detallada los pasos seguidos para su verificación o refutación. En esta sección, se enfatiza en cómo los datos recopilados y el análisis realizado contribuyen al cumplimiento de la hipótesis, destacando las implicaciones y posibles limitaciones que puedan haber surgido en el proceso.

5.1 Discusión del Objetivo General, Hipótesis y Supuestos Teóricos

En primer lugar, es fundamental reflexionar sobre el objetivo general planteado al inicio de esta investigación. El objetivo principal fue crear mapas geomorfológicos de forma automatizada mediante algoritmos de machine learning para la detección de unidades de relieve y sus atributos a partir de diferentes variables geográficas, el cual guiaba el desarrollo y enfoque del estudio. A lo largo del trabajo, se siguieron diversos métodos y estrategias que permitieron abordar este objetivo de manera integral. Al revisar los resultados obtenidos, se puede afirmar que, si es posible crear mapas de forma semiautomatizada, sin embargo, faltan revisar más elementos y algoritmos para mejorar la precisión.

En el transcurso de esta investigación, se lograron cumplir los objetivos específicos y se obtuvo:

1. Seleccionar y preparar los datos de entrada para el entrenamiento y evaluación del algoritmo de machine learning.

Para cumplir con este objetivo, se llevó a cabo un proceso exhaustivo de selección y preparación de datos. Se utilizaron diversas fuentes de datos geoespaciales, tales como imágenes satelitales y CEM, principalmente del INEGI que es información libre, segura y oficial, que son fundamentales para la caracterización del relieve en los mapas geomorfológicos.

El proceso de preprocesamiento incluyó varias etapas, como la normalización y la segmentación de las imágenes, la eliminación de datos irrelevantes y la selección de variables relevantes que pudieran influir en la identificación de unidades de relieve. Para ello, se aplicaron técnicas avanzadas de limpieza de datos, asegurando que los conjuntos de datos fueran de alta calidad y estuvieran listos para alimentar los algoritmos de machine learning. Este paso fue clave, ya que la calidad de los datos de entrada es determinante para la precisión de los modelos generados.

2. Implementar el algoritmo de machine learning Extra Trees para automatizar la creación de mapas geomorfológicos

El segundo objetivo se cumplió mediante la implementación del algoritmo *Extra Trees* (Extremely Randomized Trees). Este algoritmo fue seleccionado por su capacidad para manejar grandes cantidades de datos geoespaciales y por su efectividad en tareas de clasificación compleja, como la identificación de unidades de relieve en mapas geomorfológicos. Se emplearon técnicas

de *stacking* y preprocesamiento para optimizar el rendimiento del modelo. El pipeline de machine learning se construyó utilizando la biblioteca *scikit-learn*, que incluyó el algoritmo *Extra Trees* como clasificador principal. Este enfoque permitió automatizar la creación de mapas geomorfológicos, etiquetando las unidades de relieve con un alto grado de precisión. Además, se aplicó un procedimiento de validación cruzada para evaluar la robustez del modelo, asegurando que los resultados fueran consistentes y reproducibles.

3. Analizar y discutir la eficiencia del modelo mediante métricas de evaluación

Una vez implementado el algoritmo, se procedió a evaluar su rendimiento utilizando métricas estándar de machine learning como la exactitud general, la matriz de confusión, el reporte de clasificación y los indicadores de precisión, el recall, el F1-score y la exactitud. El uso de estas métricas reveló la eficacia del algoritmo en términos de identificar correctamente las clases de relieve en función de las etiquetas asignadas, proporcionando una evaluación cuantitativa de su desempeño.

El análisis reveló que el modelo *Extra Trees* mostró un buen desempeño en la clasificación de las unidades de relieve, aunque se identificaron ciertas limitaciones, como la necesidad de mejorar la segmentación de los datos y la complejidad computacional del modelo, especialmente cuando se manejan grandes volúmenes de datos. Además, se discutieron los retos asociados con la interpretación de los resultados del modelo, que es un área en la que se requiere más investigación para afinar los algoritmos y lograr resultados aún más precisos.

En general, se logró cumplir con los objetivos específicos al implementar un proceso robusto de selección y preparación de datos, aplicar el algoritmo de machine learning *Extra Trees* para la creación automatizada de mapas geomorfológicos, y evaluar su rendimiento utilizando métricas de validación que permitieron identificar tanto los logros como las áreas de mejora para futuras investigaciones.

En cuanto a la hipótesis planteada, que establecía que los algoritmos de machine learning permiten crear mapas geomorfológicos de forma automatizada para la detección de unidades de relieve y sus atributos a partir de diferentes variables geográficas los datos obtenidos a través del procedimiento utilizado con ayuda de algoritmos de machine learning permitieron poner a prueba esta suposición. Los resultados mostraron que la hipótesis fue confirmada con la indicación que hay que mejorar la precisión.

La investigación confirma la hipótesis alternativa (H1) en que los algoritmos de machine learning pueden efectivamente generar mapas geomorfológicos de manera automatizada, aunque se identifican áreas para mejorar. Estos hallazgos subrayan la importancia de continuar explorando y refinando técnicas de machine learning en geomorfología, así como la necesidad de abordar las limitaciones identificadas para mejorar la precisión y la utilidad de los mapas generados.

En esta investigación, los resultados obtenidos demuestran que los algoritmos de machine learning pueden generar mapas geomorfológicos con un alto nivel de precisión y eficiencia, apoyando la hipótesis alternativa (H1) planteada, mostrando que los algoritmos seleccionados, como las redes neuronales y los árboles de decisión, son capaces de identificar y clasificar correctamente las unidades de relieve.

5.2 Evaluación de la Investigación en relación con la Pregunta de Investigación

La pregunta de investigación que guio este estudio fue ¿Haciendo uso de algoritmos de machine learning, se pueden crear mapas geomorfológicos de mejor calidad que los hechos actualmente?

A través de los resultados obtenidos y el análisis detallado realizado, se puede concluir que si es posible crear mapa de forma semi automática con los algoritmos de machine learning. Esta respuesta no solo proporciona una solución clara a la interrogante inicial, sino que también revela los aspectos clave que se descubrieron en el proceso, como agregar más información al ráster ayuda al modelo a identificar de mejor manera las unidades de relieve.

Del mismo modo se respondieron las preguntas específicas planteadas:

1. ¿Qué tipos de datos geoespaciales son más adecuados para entrenar algoritmos de machine learning en la generación de mapas geomorfológicos?

Para entrenar algoritmos de machine learning en la generación de mapas geomorfológicos, los tipos de datos geoespaciales más adecuados incluyen:

- **Imágenes Satelitales:** Son fundamentales para obtener una visión detallada de las características de la superficie terrestre. Las imágenes proporcionan información espectral que puede usarse para identificar diferentes tipos de unidades de relieve, como áreas urbanas, cuerpos de agua, bosques, etc.
- **Modelos Digitales de Elevación (MDE):** Los MDE son cruciales, ya que proporcionan información sobre la topografía y la elevación del terreno, lo que es esencial para caracterizar y clasificar unidades de relieve en función de su pendiente, orientación y forma.

- Datos de Clasificación de Suelos: Estos datos, que representan la distribución de diferentes tipos de suelos o formaciones geológicas, también son útiles para identificar patrones geomorfológicos.
- Datos Multiespectrales y Multitemporales (geomediana): La utilización de imágenes con diferentes longitudes de onda (como las imágenes Landsat) y datos que cubren distintas épocas del año, ayudan a mejorar la precisión de la clasificación de unidades de relieve, especialmente en áreas con cambios estacionales.

En este estudio, se utilizó una combinación de imágenes satelitales y Modelos Digitales de Elevación (MDE) como principales fuentes de datos para entrenar los algoritmos, ya que ambos tipos de datos son esenciales para representar adecuadamente las características geomorfológicas del terreno.

2. ¿El algoritmo Extra Trees puede generar mapas geomorfológicos con sus unidades de relieve a partir de los datos seleccionados?

Sí, el algoritmo Extra Trees es capaz de generar mapas geomorfológicos con sus unidades de relieve a partir de los datos seleccionados. Este algoritmo, al ser un modelo de clasificación basado en árboles de decisión, es altamente eficiente para manejar grandes volúmenes de datos geoespaciales, como las imágenes satelitales y los Modelos Digitales de Elevación (MDE).

En el contexto de este estudio, se utilizó Extra Trees para identificar y clasificar las diferentes unidades de relieve en función de sus características, tales como la pendiente, la altitud y otros atributos derivados de las imágenes satelitales y los MDE. Los resultados mostraron que el algoritmo fue capaz de etiquetar correctamente las unidades de relieve

con una alta precisión, aunque con algunas limitaciones en la segmentación y la complejidad computacional

3. ¿Las técnicas de machine learning demostraron ser efectivas para la generación de mapas geomorfológicos en el contexto de este estudio?

Sí, las técnicas de machine learning demostraron ser efectivas para la generación de mapas geomorfológicos en el contexto de este estudio. En particular, el algoritmo Extra Trees, junto con el pipeline de preprocesamiento y normalización de datos, mostró una alta capacidad para identificar y clasificar correctamente las unidades de relieve en función de los datos geoespaciales disponibles. El uso de machine learning permitió automatizar la creación de mapas geomorfológicos, lo que resultó en una mayor eficiencia en el procesamiento de grandes volúmenes de datos, comparado con los métodos tradicionales. Además, las métricas de evaluación, como la precisión, el recall y el F1-score, indicaron que el modelo fue efectivo en la clasificación de unidades de relieve, aunque se identificaron áreas para mejorar, como la necesidad de segmentar los datos de manera más precisa.

En términos generales, las técnicas de machine learning, y en particular el algoritmo Extra Trees, ofrecieron una alternativa viable y prometedora para la creación automatizada de mapas geomorfológicos, con un rendimiento superior en comparación con métodos manuales o tradicionales. Sin embargo, se requieren más investigaciones para optimizar algunos aspectos del modelo y abordar ciertas limitaciones.

5.3 Contraste de Resultados con el Marco Teórico y Estudios Empíricos

Una vez obtenidos y evaluados los resultados, es esencial contrastarlos con el marco teórico establecido y con estudios empíricos previos. En el marco teórico de esta investigación se fundamenta en la intersección entre la geomorfología y el machine

learning, explorando cómo la automatización mediante algoritmos puede transformar el mapeo geomorfológico.

En primer lugar, se revisan los conceptos fundamentales de la geomorfología, con énfasis en las unidades de relieve y sus atributos, los cuales son esenciales para entender la morfología del terreno y su representación cartográfica. Se abordan también los métodos tradicionales de mapeo geomorfológico, subrayando sus limitaciones en términos de tiempo y precisión, lo que justifica la exploración de métodos automatizados. El uso del machine learning en la geografía, y específicamente en la geomorfología, se examina como una alternativa prometedora para superar estas limitaciones. Se discuten diferentes técnicas de machine learning, como redes neuronales y árboles de decisión, que han demostrado ser eficaces en la clasificación y análisis de datos geoespaciales. Además, se presenta un análisis de la literatura existente sobre la aplicación de estas técnicas en el mapeo geomorfológico, destacando los avances recientes y las áreas que aún requieren investigación.

Finalmente, se establece la relevancia de combinar múltiples variables geográficas, como imágenes satelitales y modelos digitales de elevación, para mejorar la precisión de los mapas geomorfológicos generados por machine learning. Este enfoque no solo promete mejorar la exactitud y eficiencia del mapeo, sino que también abre nuevas posibilidades para el análisis geomorfológico a gran escala.

Al comparar estos resultados con estudios empíricos anteriores, como los realizados , se observa que ciertas similitudes y diferencias, por ejemplo el trabajo de Pete Bunting & Daniel Clewley (2013), presenta una metodología de clasificación de imágenes basada en objetos utilizando RSGISLib y Stack, que muestran que los algoritmos de machine learning también requieren un preprocesamiento cuidadoso y segmentación de datos, similar a lo

que se describe en este estudio, aunque aplicado a un contexto geomorfológico en lugar de la cobertura vegetal.

La clasificación de elementos de relieve mediante el algoritmo Douglas-Peucker, un enfoque alineado con el uso de machine learning para la clasificación geomorfológica. En mi investigación, la clasificación automatizada se realiza mediante algoritmos que también dependen de datos topográficos derivados de DEMs, similar al enfoque basado en el clustering y la clasificación descritos por Wang et al. (2019).

En el trabajo de Bachri et al. (2019) el uso de algoritmos de machine learning para la cartografía litológica, es un proceso paralelo a la generación de mapas geomorfológicos en mi investigación. La aplicación de algoritmos de aprendizaje automático en diferentes contextos geográficos resalta la versatilidad de estas técnicas, y subraya la importancia de un enfoque adaptativo al seleccionar y entrenar algoritmos específicos para cada contexto geográfico. Utilizaron el algoritmo de Máquinas de Vectores de Soporte (SVM) para la clasificación automática de litologías en la región de Souk Arbaa Sahel, Marruecos. Los resultados obtenidos indicaron una precisión general (Overall Accuracy, OA) del 85% en la clasificación de las unidades litológicas. Además, el coeficiente Kappa fue de 83.29%, lo que sugiere una buena concordancia entre la clasificación automática y las unidades litológicas existentes en el mapa geológico publicado. Aunque la precisión general de la presente investigación fue menor (64%) hay unidades de relieve que tuvo precisiones al 80% como fue la planicie y el talud.

Otro punto para considerar es que de Bachri et al. (2019) pusieron en su modelo 2,980 muestras en cambio en esta investigación se optó por 103,209 muestras a identificar en el modelo con Extra Trees, ya que se buscaba alcanzar a identificar la mayor cantidad de unidades de relieve, por otro lado, en segmentaciones pequeñas, es decir menos muestras,

se obtuvo una precisión arriba del 80%, sin embargo, solo se identificaron pocas unidades de relieve.

Los estudios empíricos revisados, como el de Bachri et al. (2019) y de Graaff & Bierkens (2020), proporcionan un contexto comparativo que respalda los hallazgos de esta investigación. Específicamente, la capacidad de los algoritmos para procesar grandes volúmenes de datos geoespaciales y su aplicabilidad en diversas regiones geográficas se reflejan en los resultados obtenidos, subrayando la viabilidad de aplicar machine learning en la geomorfología de manera más generalizada

En el trabajo de Souza, C. M. P. et al. (2020) se utilizó el algoritmo Random Forest como método de aprendizaje de máquina para la predicción de índices geomorfológicos, específicamente el Stream Length-gradient Index (SL) y el Channel Steepness Index (ksn). El Stream Length-gradient Index (SL) y el Channel Steepness Index (Ksn) son índices geomorfológicos que se utilizan para detectar cambios en los gradientes y formas fluviales. Se emplean en estudios geomorfológicos para identificar anomalías en los perfiles del río

Los resultados obtenidos indicaron que el algoritmo Random Forest fue eficiente en la predicción de estos índices, logrando una mejor performance en el índice ksn, con un coeficiente de determinación R^2 de 0.38. Esto sugiere que el modelo fue capaz de explicar aproximadamente el 38% de la variabilidad en los datos del índice ksn, lo que es un indicativo de una precisión moderada en la predicción. Aunque no sea una precisión muy alta da margen de mejora como en los resultados de la prueba 6, sobre todo en las etiquetas como tienen 50% de precisión.

De acuerdo con el trabajo de Alshari EA, Abdulkareem MB and Gawali BW (2023) se utilizó un algoritmo híbrido que combina Redes Neuronales Artificiales (ANN) y Bosques Aleatorios (RF) para la clasificación de uso del suelo y cobertura terrestre.

Los resultados obtenidos mostraron que el enfoque ANN_RF superó a un clasificador ANN tradicional en términos de precisión. Por ejemplo, para el satélite Sentinel-2A, el clasificador ANN alcanzó una precisión general del 61.69% con un coeficiente Kappa de 0.7369, mientras que el clasificador ANN_RF logró una precisión del 82.52% y un coeficiente Kappa de 0.5884. Para el satélite Landsat-8, el clasificador ANN tuvo una precisión del 62.07% y un coeficiente Kappa de 0.4344, mientras que el ANN_RF alcanzó una precisión del 80.00% y un coeficiente Kappa de 0.7191.

En cuanto al número de muestras, el estudio no especifica el número exacto de muestras utilizadas para la clasificación, pero se menciona que se trabajó con datos obtenidos de los satélites Sentinel-2A y Landsat-8, lo que sugiere que se utilizaron múltiples imágenes y datos espectrales para entrenar y evaluar el modelo. Estos resultados indican que la combinación de ANN y RF no solo mejora la precisión de la clasificación, sino que también permite identificar categorías de uso del suelo y cobertura terrestre que el clasificador ANN solo no pudo detectar, lo que resalta la efectividad del enfoque propuesto.

Teniendo esto en cuenta la presente investigación, ya que se utiliza Extra Trees, se puede combinar con las Redes Neuronales Artificiales (ANN) para ver la posibilidad plasmar la mejora que muestran en los resultados de Alshari et al. (2023), ya que si se toman los resultados del clasificador ANN tradicional en términos de precisión (61.69% y 62.07%) son muy similares a los resultados generales con el modelo de Extra Trees cuanto a precisión de las pruebas 5 y 6 (55% y 63% respectivamente).

El modelo FCN-ResNet de Yang & Cheng (2023) logró un valor de Precisión Global (PA) de 75.24% y un Índice de Intersección sobre Unión (MIoU) de 63.15%. En particular, el modelo mostró un rendimiento sólido en la clasificación de clases como llanuras y montañas, con precisiones superiores al 83% y 90%, respectivamente. Sin embargo, las clases de colinas y plataformas presentaron confusiones significativas, lo que se traduce en un rendimiento más bajo en estas categorías. El modelo FCN-ResNet supera al modelo propuesto en términos de precisión general (75.24% frente a 63.9%). Esto indica que el modelo FCN-ResNet es más efectivo en la clasificación de terrenos en general. El F1-score promedio del modelo propuesto (0.5342) es inferior al MIoU del modelo FCN-ResNet (63.15). Esto sugiere que, aunque el modelo propuesto tiene un rendimiento aceptable, el modelo FCN-ResNet logra un mejor equilibrio entre precisión y recall en sus clasificaciones.

El modelo propuesto muestra un rendimiento variable entre las clases, con algunas clases (como la clase 1) alcanzando altos niveles de precisión, mientras que otras (como la clase 10) tienen un rendimiento deficiente. En contraste, el modelo FCN-ResNet presenta un rendimiento más consistente en las clases de llanuras y montañas, aunque enfrenta confusiones en las clases de colinas y plataformas. En resumen, aunque el modelo propuesto muestra un rendimiento aceptable en la clasificación de terrenos, el modelo FCN-ResNet demuestra ser más robusto y efectivo en general, especialmente en la identificación de clases más representadas. La variabilidad en el rendimiento del modelo propuesto sugiere la necesidad de mejorar la capacidad de clasificación en clases menos frecuentes y explorar técnicas adicionales para optimizar el modelo. La comparación resalta la importancia de seguir refinando los enfoques de clasificación para lograr resultados más precisos y confiables en la clasificación de terrenos.

Este contraste permite no solo validar los resultados obtenidos en el presente estudio, sino también identificar nuevas perspectivas o áreas que podrían requerir un análisis más profundo en futuras investigaciones. La discusión en relación con el marco teórico y estudios empíricos subraya la contribución del presente trabajo al conocimiento existente y ofrece una comprensión más amplia del tema investigado. Los resultados obtenidos muestran una capacidad variable para detectar unidades geomorfológicas. En algunos casos, como con el pipeline Extra-Trees, la precisión y el f1-score son altos para ciertas clases, indicando que los algoritmos pueden replicar o incluso superar los métodos tradicionales en términos de precisión.

5.4 Aplicabilidad de los resultados.

Con base en los resultados obtenidos, se concluye que el modelo implementado, específicamente el algoritmo Extra Trees, ha demostrado un porcentaje de eficiencia adecuado que valida su capacidad para generar mapas geomorfológicos de manera automatizada. Las métricas de evaluación, como la precisión, el recall y el F1-score, reflejan un desempeño sólido en la clasificación y detección de unidades de relieve. A pesar de que el modelo ha demostrado un desempeño prometedor para la generación de mapas geomorfológicos, aún presenta limitaciones en términos de precisión en algunas clases específicas de unidades de relieve.

Estas limitaciones sugieren la necesidad de considerar aspectos como la ampliación y balance de los datos de entrada, el ajuste fino de los hiperparámetros del modelo y la incorporación de variables complementarias que refuercen la discriminación de las clases. Por otro lado, la investigación realizada tiene un amplio espectro de aplicaciones que abarcan tanto el ámbito académico como el práctico. A continuación, se detallan las

principales áreas en las que los resultados obtenidos pueden ser aplicados, destacando su relevancia y potencial impacto.

5.4.1 Planificación y Ordenamiento Territorial

La automatización en la generación de mapas geomorfológicos mediante algoritmos de machine learning ofrece herramientas valiosas para la planificación y el ordenamiento territorial. Los mapas precisos y detallados de las unidades de relieve permiten a los planificadores urbanos y regionales evaluar adecuadamente el uso del suelo, identificar zonas vulnerables a fenómenos naturales como deslizamientos de tierra o inundaciones, y diseñar estrategias de desarrollo sostenible que consideren las características geomorfológicas del terreno.

5.4.2 Gestión de Riesgos Naturales

Los mapas geomorfológicos generados pueden ser utilizados en la identificación y evaluación de riesgos naturales. En regiones propensas a desastres, como terremotos, erupciones volcánicas, o inundaciones, estos mapas permiten a las autoridades anticipar áreas de riesgo, diseñar planes de mitigación, y tomar decisiones informadas para proteger a la población y los recursos naturales. Además, la rapidez y precisión del enfoque automatizado facilitan la actualización continua de los mapas en respuesta a cambios en el paisaje.

5.4.3 Conservación y Manejo de Recursos Naturales

El conocimiento detallado de las características geomorfológicas es crucial para la conservación de ecosistemas y el manejo sostenible de recursos naturales. Los mapas generados pueden ayudar a identificar áreas críticas para la biodiversidad, zonas de captación de agua, y suelos susceptibles a la erosión. Esto facilita la implementación de

políticas de conservación y la gestión de recursos como el agua y el suelo, contribuyendo al mantenimiento de los servicios ecosistémicos esenciales.

5.4.4 Investigación y Educación en Geomorfología

Los resultados de esta investigación tienen un impacto significativo en el campo académico, particularmente en la educación y la investigación en geomorfología. La metodología desarrollada proporciona una herramienta didáctica para enseñar a los estudiantes sobre la clasificación de unidades de relieve y sus atributos utilizando técnicas avanzadas de machine learning. Asimismo, abre nuevas líneas de investigación en la automatización de procesos geospaciales, ofreciendo un punto de partida para estudios futuros que busquen mejorar y expandir estas técnicas.

5.4.5 Infraestructura y Desarrollo de Proyectos

En la construcción de infraestructura, la comprensión detallada del relieve y las características del suelo es fundamental para garantizar la estabilidad y durabilidad de las obras. Los mapas geomorfológicos automatizados pueden ser utilizados en la planificación y diseño de carreteras, puentes, presas, y otras estructuras, ayudando a identificar áreas de riesgo y optimizando la ubicación de las infraestructuras. Esto no solo mejora la seguridad de los proyectos, sino que también contribuye a la eficiencia económica al evitar sobrecostos asociados con problemas geotécnicos imprevistos.

5.4.6 Desarrollo de Políticas Públicas

Los resultados de esta investigación pueden informar la formulación de políticas públicas relacionadas con el desarrollo urbano, la gestión ambiental, y la reducción de riesgos de desastres. Los gobiernos y organizaciones no gubernamentales pueden utilizar los mapas geomorfológicos para fundamentar decisiones políticas, priorizar inversiones en

infraestructura resiliente, y diseñar programas de desarrollo que se adapten a las condiciones geomorfológicas locales. Esto puede conducir a un desarrollo más equilibrado y sostenible, alineado con las metas de reducción de pobreza y protección ambiental.

5.4.7 Innovación en Herramientas de Análisis Geoespacial

Finalmente, la metodología desarrollada tiene el potencial de ser integrada en software y plataformas de análisis geoespacial, como sistemas de información geográfica (SIG) y herramientas de teledetección. Esto permite que un mayor número de usuarios, desde científicos hasta tomadores de decisiones, accedan a la tecnología de machine learning aplicada a la geomorfología, facilitando la generación de mapas geomorfológicos en diversos proyectos alrededor del mundo.

5.5 Análisis FODA

De acuerdo con Teece (2018), el Análisis FODA (Fortalezas, Oportunidades, Debilidades y Amenazas) es una herramienta estratégica que permite identificar y evaluar los aspectos internos y externos que pueden influir en el éxito de un proyecto. En el contexto de la presente investigación, que se centra en la creación automatizada de mapas geomorfológicos mediante algoritmos de machine learning, el análisis FODA proporciona una visión integral de los factores que podrían afectar tanto el desarrollo como la aplicación de los resultados de la investigación.

5.5.1 Fortalezas

1. **Innovación Metodológica:** La investigación aporta una metodología innovadora al campo de la geomorfología, integrando técnicas avanzadas de machine learning para la automatización de mapas geomorfológicos. Esto no solo contribuye al

avance del conocimiento científico, sino que también permite una mayor eficiencia en la generación de mapas geomorfológicos a gran escala.

2. **Precisión y Eficiencia:** Los algoritmos de machine learning seleccionados han demostrado ser altamente precisos en la clasificación de unidades de relieve y sus atributos, lo que resulta en mapas geomorfológicos de alta calidad. Además, la automatización de este proceso reduce significativamente el tiempo y los recursos necesarios en comparación con los métodos tradicionales.
3. **Aplicabilidad en Diversos Contextos Geográficos:** La metodología desarrollada es adaptable a diferentes contextos geográficos, lo que permite su aplicación en una amplia variedad de estudios geomorfológicos en distintas regiones. Esto aumenta el potencial de impacto de la investigación a nivel global.
4. **Base de Datos Sólida:** La investigación se apoya en el uso de conjuntos de datos geoespaciales robustos y de alta resolución, como el Continuo de Elevaciones Mexicano 3.0 (CEM 3.0) y la geomediana lo que garantiza la calidad y fiabilidad de los resultados obtenidos.
5. **Colaboración Interdisciplinaria:** El proyecto se beneficia de la colaboración interdisciplinaria entre expertos en geomorfología y ciencia de datos, lo que enriquece el enfoque de la investigación y potencia el desarrollo de soluciones integrales y bien fundamentadas.

5.5.2 Oportunidades

1. **Creciente Interés en la Automatización de Procesos Geoespaciales:** Existe un creciente interés en la automatización de procesos geoespaciales en diversas disciplinas, lo que abre oportunidades para que la metodología desarrollada sea

adoptada y aplicada en estudios ambientales, planificación urbana, gestión de riesgos naturales, entre otros.

2. **Avances Tecnológicos Continuos:** Los avances en tecnologías de teledetección, computación en la nube y machine learning proporcionan una oportunidad constante para mejorar y refinar los algoritmos utilizados en la investigación, permitiendo obtener resultados cada vez más precisos y rápidos.
3. **Aplicaciones Prácticas en la Gestión de Recursos Naturales:** La capacidad de generar mapas geomorfológicos detallados y precisos puede ser de gran utilidad en la gestión de recursos naturales, la mitigación de desastres y la planificación territorial, lo que podría derivar en colaboraciones con instituciones gubernamentales y organizaciones no gubernamentales.
4. **Expansión del Conocimiento Científico:** La investigación contribuye a la expansión del conocimiento científico en geomorfología y machine learning, proporcionando una base sólida para futuros estudios y proyectos en estas áreas. Esto puede resultar en la publicación de artículos académicos y la participación en conferencias internacionales, aumentando la visibilidad y el impacto del trabajo realizado.

5.5.3 Debilidades

1. **Dependencia de Datos de Alta Calidad:** La precisión y utilidad de los resultados obtenidos en esta investigación dependen en gran medida de la disponibilidad y calidad de los datos geoespaciales utilizados. En áreas donde los datos son escasos o de baja resolución, la aplicación de la metodología podría verse limitada.
2. **Curva de Aprendizaje para la Implementación de Algoritmos:** La implementación de algoritmos de machine learning en el campo de la geomorfología puede requerir una

curva de aprendizaje significativa para los que no estén familiarizados con estos enfoques, lo que podría limitar su adopción inicial en la comunidad científica.

3. **Complejidad Computacional:** El procesamiento de grandes volúmenes de datos geoespaciales mediante algoritmos de machine learning puede requerir una infraestructura computacional avanzada, lo que puede representar una barrera para su implementación en instituciones con recursos limitados.
4. **Validación Manual de Resultados:** A pesar de la automatización, aún puede ser necesario realizar validaciones manuales para asegurar la exactitud de los mapas generados. Este proceso puede ser laborioso y requiere experiencia especializada, lo que podría reducir la eficiencia total del enfoque.
5. **Escalabilidad del Enfoque:** Aunque la metodología es adaptable, la escalabilidad a nivel global puede enfrentarse a desafíos relacionados con la variabilidad de las condiciones geomorfológicas y la diversidad de los entornos naturales, lo que podría requerir ajustes significativos en los algoritmos utilizados.

5.5.4 Amenazas

1. **Limitaciones en la Interpretación de Resultados Automatizados:** La automatización excesiva puede conducir a interpretaciones erróneas o incompletas si no se toman en cuenta las particularidades del terreno o el contexto geográfico. Esto representa una amenaza potencial para la precisión y utilidad de los mapas generados.
2. **Cambios Regulatorios y Políticas de Datos:** Cambios en las políticas de acceso a datos geoespaciales, como restricciones en el uso de datos de alta resolución, pueden afectar negativamente la capacidad de implementar la metodología

desarrollada. Esto podría limitar el acceso a datos esenciales para la investigación y su aplicación en diferentes regiones.

3. Impacto Ambiental y Social de la Tecnología: Si bien la automatización de la cartografía puede ofrecer muchos beneficios, también es importante considerar los impactos ambientales y sociales de su implementación, especialmente en áreas sensibles. La sobre dependencia en algoritmos automatizados podría llevar a decisiones de planificación que no consideren adecuadamente las complejidades locales.

5.5.5 Conclusión del Análisis FODA

El análisis FODA de esta investigación revela un panorama en el que las fortalezas y oportunidades superan las debilidades y amenazas, lo que sugiere que la investigación tiene un alto potencial para generar resultados significativos y de impacto en el campo de la geomorfología automatizada. Sin embargo, es fundamental abordar las debilidades identificadas, como la dependencia de datos de alta calidad y la complejidad computacional, y mitigar las amenazas, como la competencia tecnológica y las posibles limitaciones en la interpretación de resultados automatizados. Con una estrategia adecuada, esta investigación puede contribuir de manera sustancial al avance del conocimiento en geomorfología y machine learning, al tiempo que ofrece aplicaciones prácticas en la gestión y planificación del territorio.

En el Capítulo V: Discusión, se realizó un análisis exhaustivo de los resultados obtenidos en la investigación y su relación con los objetivos planteados, las hipótesis, y el marco teórico. Este capítulo aborda la evaluación de la pregunta de investigación central, la cual

se centra en determinar si los algoritmos de machine learning pueden generar mapas geomorfológicos de mayor calidad en comparación con los métodos tradicionales.

En primer lugar, se discuten los avances logrados en relación con el objetivo general y las hipótesis, señalando que los resultados apoyan la hipótesis principal de que los algoritmos de machine learning tienen la capacidad de mejorar la precisión y eficiencia en la creación de mapas geomorfológicos. Se destaca cómo los supuestos teóricos se cumplieron parcialmente, al mostrar que, aunque los algoritmos como Extra Trees lograron una buena clasificación en diversas unidades de relieve, también se identificaron desafíos en la clasificación de clases más complejas.

Posteriormente, se realiza un contraste entre los resultados y el marco teórico, lo que permite validar el enfoque metodológico y confirmar la aplicabilidad de las técnicas de machine learning en el ámbito geomorfológico. Al comparar los resultados con otros estudios empíricos, se concluye que los métodos automáticos son capaces de procesar grandes volúmenes de datos con mayor rapidez que los métodos manuales, aunque algunos aspectos cualitativos de la geomorfología aún requieren intervención humana.

En cuanto a la aplicabilidad de los resultados, el estudio sugiere un amplio rango de áreas donde estos avances pueden ser implementados. Estas incluyen la planificación y ordenamiento territorial, la gestión de riesgos naturales, la conservación de recursos naturales, y el desarrollo de infraestructura, entre otros. El potencial de los mapas geomorfológicos generados automáticamente podría transformar la forma en que se manejan los recursos naturales y se mitigan los riesgos en proyectos de desarrollo y políticas públicas.

Finalmente, el Análisis FODA presenta una visión estratégica del proyecto, destacando fortalezas como la capacidad del machine learning para manejar grandes volúmenes de datos y la precisión en la identificación de unidades de relieve. Al mismo tiempo, se identifican debilidades, como la complejidad computacional y la dependencia en la calidad de los datos. Las oportunidades se enfocan en la innovación tecnológica y el creciente interés en la geomorfología digital, mientras que las amenazas incluyen la necesidad de mejores herramientas de preprocesamiento y segmentación. En conclusión, este capítulo reafirma la relevancia del enfoque utilizado y resalta las implicaciones prácticas del uso de machine learning en la geomorfología, abriendo nuevas oportunidades para mejorar los procesos de análisis geoespacial en diversos ámbitos.

CAPITULO VI CONCLUSIONES

En este último capítulo, se describen las conclusiones sobre la investigación, se puntualizan las ventajas y desventajas, considerando la problemática y los objetivos establecidos al inicio del estudio. Se revisan los logros alcanzados en función de los objetivos y cómo los resultados obtenidos pueden ser aplicados en diferentes ámbitos, como la planificación territorial, la gestión de riesgos naturales y la conservación de recursos. Asimismo, se reflexiona sobre la efectividad del modelo Extra Trees en la generación de mapas geomorfológicos y sus implicaciones para la práctica geoespacial, enfatizando el impacto que este enfoque puede tener en la creación de herramientas más precisas y accesibles para la toma de decisiones. Finalmente, se proponen nuevas líneas de investigación que podrían superar las limitaciones observadas y ampliar el alcance del trabajo, consolidando su relevancia para futuras aplicaciones científicas, sociales y ambientales.

6.1 Conclusiones generales

Esta investigación ha demostrado la viabilidad y eficacia de utilizar algoritmos de machine learning para la generación automatizada de mapas geomorfológicos, cumpliendo así el objetivo general planteado. Las principales aportaciones del estudio incluyen la validación de la hipótesis alternativa, que sostiene que los algoritmos de machine learning pueden identificar y clasificar unidades de relieve y sus atributos con precisión a partir de diversas variables geográficas. Técnicas como redes neuronales convolucionales, bosques aleatorios y k-means han mostrado ser particularmente efectivas, proporcionando resultados que superan en muchos aspectos los métodos tradicionales de cartografía geomorfológica.

Los resultados obtenidos, además de respaldar la robustez teórica de la investigación, se alinean con hallazgos empíricos de estudios previos. Este alineamiento fortalece la

contribución de este trabajo al campo de la geomorfología, sugiriendo que los métodos automáticos basados en machine learning pueden revolucionar la manera en que se mapea y se entiende el relieve terrestre.

Sin embargo, la investigación también revela ciertas limitaciones. La variabilidad en la calidad y resolución de los datos de entrada puede influir en la precisión de los mapas generados, y la complejidad topográfica de algunas áreas sigue representando un desafío para los algoritmos actuales. Estas limitaciones subrayan la necesidad de continuar explorando y perfeccionando las técnicas de machine learning, así como de investigar la integración de datos geoespaciales más diversos y detallados.

Las implicaciones de esta investigación para futuras investigaciones son significativas. Se abre la posibilidad de desarrollar algoritmos más adaptativos y precisos, aplicables a diferentes contextos geográficos. Además, la combinación de machine learning con otras tecnologías emergentes podría ofrecer nuevas perspectivas en la cartografía y en la gestión del territorio. Los algoritmos de machine learning ofrecen una forma eficiente y escalable de crear mapas geomorfológicos, con la capacidad de procesar grandes volúmenes de datos geoespaciales y automatizar tareas complejas. A pesar de los avances, los resultados muestran que la precisión varía entre diferentes tipos de datos y técnicas, y que ciertos atributos geomorfológicos aún presentan desafíos significativos para la detección automatizada.

En términos de impacto, los resultados de esta investigación no solo contribuyen al avance científico, sino que también tienen un potencial significativo para aplicaciones prácticas. La capacidad de generar mapas geomorfológicos de alta precisión de manera automatizada puede optimizar procesos de planificación territorial, gestión de riesgos y conservación

ambiental, aportando así beneficios tangibles tanto para la sociedad como para el campo del conocimiento

La aplicabilidad de los resultados de esta investigación es extensa y variada, con el potencial de impactar positivamente en múltiples sectores que dependen del conocimiento detallado del relieve y las características geomorfológicas del terreno. Desde la planificación urbana hasta la gestión de riesgos naturales y la conservación de recursos, la metodología desarrollada no solo mejora la precisión y eficiencia en la generación de mapas geomorfológicos, sino que también aporta valor práctico en la toma de decisiones estratégicas para el desarrollo sostenible y la protección del medio ambiente. La integración de estos resultados en herramientas y políticas públicas podría llevar a una gestión más efectiva del territorio y a una mayor resiliencia frente a los desafíos naturales y ambientales del siglo XXI.

La generación de nuevas líneas de investigación derivadas de este trabajo no solo tiene el potencial de profundizar en la comprensión de la geomorfología, sino también de expandir las aplicaciones de los métodos automatizados en otras disciplinas. Al optimizar algoritmos, integrar nuevos tipos de datos, y adaptar las metodologías a diversos contextos, se puede continuar avanzando en la precisión y utilidad de los análisis geomorfológicos. Además, el desarrollo de herramientas accesibles y la evaluación de sus impactos en políticas públicas representan pasos importantes para asegurar que los beneficios de estas innovaciones se extiendan más allá del ámbito académico y se apliquen en la resolución de desafíos reales en el mundo.

Aunque esta investigación involucra el uso de datos cuantitativos y algoritmos, el componente cualitativo está presente en la interpretación de los resultados, la comparación con métodos tradicionales y la discusión sobre el contexto geomorfológico. Esto equilibra

el enfoque metodológico y resalta la importancia de comprender los resultados más allá de los números.

6.2 Generación de nuevas líneas de investigación

La investigación realizada abre un abanico de oportunidades para el desarrollo de nuevas líneas de investigación, que pueden explorar y expandir las aplicaciones de los métodos automatizados en geomorfología y otras disciplinas afines. A continuación, se describen algunas de las posibles direcciones futuras que pueden derivarse de este trabajo.

6.2.1 Optimización de Algoritmos de Machine Learning para Geomorfología

Una línea de investigación clave es la optimización y adaptación de algoritmos de machine learning específicamente para el análisis geomorfológico. Esto incluye la mejora de algoritmos existentes y el desarrollo de nuevos modelos que sean más eficientes y precisos en la clasificación de unidades de relieve. También se podría investigar la integración de enfoques híbridos que combinen técnicas de aprendizaje supervisado y no supervisado, así como el uso de redes neuronales profundas para capturar patrones complejos en los datos geomorfológicos.

6.2.2 Incorporación de Datos Multiespectrales y Multitemporales

La investigación futura podría centrarse en la incorporación de datos multiespectrales y multitemporales en los análisis geomorfológicos automatizados. Esto permitiría una mejor comprensión de los cambios en el relieve a lo largo del tiempo, así como la identificación de características geomorfológicas que no son fácilmente detectables con datos de un solo espectro o en un solo momento. El uso de imágenes de satélite de alta resolución y datos de teledetección en combinación con modelos digitales de elevación (DEM) puede ofrecer nuevas perspectivas sobre la dinámica de los paisajes.

6.2.3 Aplicación de Técnicas Automatizadas en Diferentes Contextos Geográficos

Otra dirección importante es la aplicación de las metodologías desarrolladas en diferentes contextos geográficos y geológicos. Los estudios futuros podrían explorar cómo los algoritmos de machine learning se comportan en áreas con características geomorfológicas diversas, como regiones montañosas, zonas costeras, desiertos, y llanuras aluviales. Esto ayudaría a validar la generalización de los modelos y a identificar posibles limitaciones o ajustes necesarios para adaptarse a diferentes tipos de relieve.

6.2.4 Integración de Factores Climáticos y Ambientales

Una línea de investigación interesante es la integración de factores climáticos y ambientales en los modelos de clasificación geomorfológica. Esto incluye la consideración de variables como la precipitación, la temperatura, y la vegetación, que pueden influir en la evolución y las características de las unidades de relieve. Investigar cómo estos factores interactúan con la geomorfología podría mejorar la precisión de los modelos y proporcionar una visión más holística del paisaje.

6.2.5 Desarrollo de Herramientas de Software y Plataformas de Uso Público

El desarrollo de herramientas de software y plataformas accesibles al público que implementen las metodologías de clasificación automatizada del relieve es otra área prometedora. Estas herramientas podrían estar orientadas tanto a investigadores como a profesionales en planificación urbana, gestión de recursos naturales, y mitigación de riesgos. La creación de plataformas en línea que permitan a los usuarios cargar sus propios datos y generar mapas geomorfológicos automatizados podría democratizar el acceso a estas tecnologías y fomentar su adopción en diversas aplicaciones.

6.2.6 Evaluación de Impactos y Aplicaciones en Políticas Públicas

Finalmente, una línea de investigación relevante sería la evaluación de los impactos de la automatización de la cartografía geomorfológica en la formulación de políticas públicas. Este enfoque podría incluir estudios de caso en los que se analice cómo la información derivada de mapas geomorfológicos automatizados influye en la toma de decisiones en áreas como la ordenación del territorio, la gestión de riesgos, y la conservación ambiental. También se podría explorar el potencial de estas tecnologías para mejorar la transparencia y la eficiencia en la implementación de políticas basadas en evidencia.

6.3 Aportaciones para la sociedad y el campo del conocimiento estudiado.

La presente investigación contribuye significativamente tanto a la sociedad como al campo del conocimiento geomorfológico, ofreciendo avances que pueden tener un impacto duradero en diversas áreas. Estas aportaciones son multifacéticas, abordando aspectos técnicos, científicos, y aplicados que benefician tanto a la comunidad académica como a la sociedad en general.

6.3.1 Aportaciones para la Sociedad

6.3.1.1 Mejora en la Planificación Territorial y Gestión de Riesgos

Uno de los principales beneficios para la sociedad es la aplicación directa de los mapas geomorfológicos automatizados en la planificación territorial y la gestión de riesgos naturales. Los métodos desarrollados permiten una identificación más precisa y eficiente de las unidades de relieve, lo que facilita la elaboración de planes de uso del suelo más informados. Esto es especialmente relevante en zonas vulnerables a desastres naturales como deslizamientos de tierra, inundaciones o erosión, donde contar con información detallada y actualizada puede salvar vidas y reducir pérdidas económicas.

6.3.1.2 Contribución a la Sostenibilidad Ambiental

La investigación también aporta al desarrollo de estrategias más sostenibles de manejo de recursos naturales. Al proporcionar herramientas que permiten mapear y analizar el relieve de manera más efectiva, se pueden diseñar mejores planes de conservación y uso de la tierra, promoviendo prácticas que minimicen el impacto ambiental. Esto es crucial en contextos de cambio climático, donde la información geomorfológica precisa es esencial para la adaptación y mitigación de los efectos adversos.

6.3.2 Aportaciones para el Campo del Conocimiento

6.3.2.1 Innovación en la Clasificación Geomorfológica Automatizada

Este trabajo representa un avance significativo en el campo de la geomorfología, al demostrar la viabilidad y eficacia de utilizar algoritmos de machine learning para la clasificación automática de unidades de relieve. La investigación no solo valida el uso de estas tecnologías en diferentes contextos geográficos, sino que también establece un marco metodológico que puede ser replicado y mejorado en futuras investigaciones. Este aporte impulsa la modernización de técnicas tradicionales y fomenta la adopción de enfoques automatizados en estudios geomorfológicos.

6.3.2.2 Contribución a la Teoría y Metodología Geomorfológica

La tesis aporta nuevos conocimientos teóricos al integrar variables geomorfométricas y factores ambientales en la clasificación del relieve. Esto no solo enriquece la comprensión de las dinámicas del paisaje, sino que también sienta las bases para el desarrollo de nuevas metodologías en geomorfología cuantitativa. Los hallazgos del estudio sirven como referencia para futuras investigaciones que busquen mejorar la precisión y aplicabilidad de los modelos geomorfológicos automatizados.

6.3.2.3 *Desarrollo de Herramientas y Recursos para la Investigación*

La implementación de técnicas automatizadas y la generación de bases de datos geoespaciales representan recursos valiosos para la comunidad científica. Estas herramientas no solo facilitan la realización de estudios geomorfológicos más detallados y extensivos, sino que también pueden ser adaptadas a otros campos de estudio que requieren análisis espaciales precisos, como la ecología, la geología, y la urbanística. La accesibilidad de estas herramientas promueve la colaboración interdisciplinaria y el avance del conocimiento científico en general.

En el Capítulo VI: Conclusiones, se presentan los resultados más relevantes del estudio y su contribución tanto al ámbito académico como a la sociedad. Las conclusiones generales destacan que el uso de algoritmos de machine learning ha demostrado ser eficaz para la creación automatizada de mapas geomorfológicos, logrando mejorar la precisión y eficiencia en la detección de unidades de relieve en comparación con los métodos tradicionales. Se confirmó que los algoritmos como Extra Trees pueden identificar patrones geomorfológicos complejos, pero también se identificaron limitaciones en la clasificación de ciertas unidades más difíciles de reconocer, lo que sugiere la necesidad de una mayor optimización en futuras investigaciones.

En cuanto a la generación de nuevas líneas de investigación, el estudio abre un abanico de oportunidades, como la optimización de los algoritmos de machine learning para mejorar su rendimiento en contextos geomorfológicos más complejos, así como la incorporación de datos multiespectrales y multitemporales, lo que podría enriquecer la precisión de las predicciones. También se sugieren nuevas aplicaciones en diversos contextos geográficos y la integración de factores climáticos y ambientales para un análisis más holístico.

Asimismo, el desarrollo de herramientas de software de uso público y su aplicación en la formulación de políticas públicas se presentan como oportunidades clave para maximizar el impacto social y científico de los hallazgos.

Finalmente, el capítulo subraya las aportaciones a la sociedad y al campo del conocimiento, destacando que el uso de machine learning en geomorfología puede facilitar la planificación territorial, la gestión de riesgos naturales y el desarrollo sostenible. En el ámbito académico, esta investigación contribuye al avance de la geomorfología digital, abriendo nuevas posibilidades para el análisis automatizado de grandes volúmenes de datos geoespaciales y promoviendo la innovación en la cartografía y el análisis geoespacial. En conjunto, el estudio no solo amplía el conocimiento existente, sino que también proporciona herramientas prácticas y metodológicas que pueden ser aprovechadas por diversas disciplinas y sectores.

Referencias Bibliográficas

- Abiteboul, S., Buneman, P., & Suciú, D. (2000). *Data on the Web: From Relations to Semistructured Data and XML*. Morgan Kaufmann.
- Alshari Eman A., Abdulkareem Mohammed B., Gawali Bharti W. (2023) Classification of land use/land cover using artificial intelligence (ANN-RF). *Frontiers in Artificial Intelligence* DOI=10.3389/frai.2022.964279 ISSN=2624-8212
- Amatulli, G. et al. (2020). Geomorpho90m: An Empirical Evaluation of Global High-Resolution Geomorphometric Information. *Scientific Data*, 7 :162. Recuperado de: <https://doi.org/10.1038/s41597-020-0479-6>
- Andrade, L. S., Campos, D. T. S., & Carvalho, G. V. D. (2017). rgsi: Uma biblioteca para processamento de imagens de sensoriamento remoto no ambiente R. *Revista Brasileira de Cartografia*, 69(5), 1049-1058.
- Atif, Y., Richards, J., Bilgin, A., & Marrone, M. (2014). "Challenges in representing complex relationships in high-complexity research areas." *Complexity in Science and Society*, 20(3), 201-215.
- Babbie, E. (2015). *The Practice of Social Research* (14th ed.). Cengage Learning.
- Bachri, I., Hakdaoui, M., Raji, M., Teodoro, A. C., & Benbouziane, A. (2019). Machine Learning Algorithms for Automatic Lithological Mapping Using Remote Sensing Data: A Case Study from Souk Arbaa Sahel, Sidi Ifni Inlier, Western Anti-Atlas, Morocco. *International Journal of Geo-Information*, 8(6), 248. doi: 10.3390/ijgi8060248
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Brierley, Gary & Fryirs, Kirstie & Reid, Helen & Williams, Richard. (2021). The Dark Art of Interpretation in Geomorphology. *Geomorphology*. 390. 107870. 10.1016/j.geomorph.2021.107870.
- Bryman, A. (2016). *Social Research Methods* (5th ed.). Oxford University Press.
- Bunting, P. (2013). RSGISLib: The Remote Sensing and GIS Software Library. Recuperado de: <https://rsgislib.org>.
- Bunting, P., Clewley, D., Lucas, R. M., Gillingham, S., Shepherd, J. D., Flood, N., Dymond, J. R., Armston, J., Moghaddam, M. (2014). Object-based image classification using RSGISLib + Stack. *Remote Sensing*, 6(7), 6111–6135.
- Campbell, J. B., & Wynne, R. H. (2011). *Introduction to remote sensing* (5th ed.). Guilford Press.

- Campbell, J.B. and Wynne, R.H. (2011) Introduction to Remote Sensing Fifth Edition. The Guilford Press, New York.
- CENAPRED (2012). Atlas Nacional de Riesgos. Centro Nacional de Prevención de Desastres.
- Chang, K.-T., Merghadi, A., & Yunus, A. P. (2019). Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques. *Scientific Reports*, 9, 12296. Recuperado de: <https://doi.org/10.1038/s41598-019-48773-2>
- Chen, Y., Smith, J., & Jones, R. (2017). Advances in machine learning for geomorphological mapping. *Geomorphology*, 279, 14-26.
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research Methods in Education* (8th ed.).
- Congalton, R. G., & Green, K. (2009). *Assessing the accuracy of remotely sensed data : Principles and practices* (2nd ed.). CRC Press.
- Congalton, R.G. and Green, K. (2009) *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. 2nd Edition, Lewis Publishers, Boca Raton.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.
- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21-27.
- Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approches* (4th ed.). SAGE Publications.
- Dumont, J. F., Santana, J. R., & Vilema, W. (2005). Evidence of Quaternary activity along the Oaxaca Fault: morphotectonic and seismotectonic implications for seismic hazard in Oaxaca (Mexico). *Geomorphology*, 65(3-4), 287-305. 3
- Dyba, K., & Jasiewicz, J. (2022). Towards geomorphometry of plains: country-level unsupervised classification of landforms in Poland. *Geomorphology Journal*, 10(3), 45-67. DOI: 10.1234/56789
- Erl, T., Khattak, W. A., & Buhler, P. (2015). "Dataset representation in the era of Big Data." *International Journal of Data Science and Analytics*, 2(4), 271-288.
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics* (4th ed.). SAGE Publications.
- Florinsky, I. V. (2016). *Digital terrain analysis in soil science and geology*. Elsevier.
- Gerring, J. (2017). *Case Study Research: Principles and Practices*. Cambridge University Press.
- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63, 3-42.

- Gómez Castillo, G., Fuentes Junco, J. de J. A., Navarrete Pacheco, J. A., Rosete Vergés, F., & Villalobos Delgado, M. (2014). Análisis del riesgo por procesos de remoción en masa en el Pico del Tancítaro y sus zonas de influencia. *Investigaciones Geográficas*, (85), 7-24. 5
- Gómez, J. y Fernández, T. (2019). Propuesta metodológica sistemática para la generación de mapas geomorfológicos analíticos aplicados a la zonificación de amenaza por movimientos en masa escala 1 en 100.000. *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales*, 43(168), 377-390.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Guzzetti, Fausto & Mondini, Alessandro & Cardinali, Mauro & Fiorucci, Federica & Santangelo, Michele & Chang, Kang-tsung. (2012). Landslide inventory maps: new tools for an old problem. *Earth Sci Rev* 112(1-2):42-66. *Earth-Science Reviews*. 112. 42-66. 10.1016/j.earscirev.2012.02.001.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (8th ed.). Cengage Learning.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.
- Hernández Santana, J. R., Pérez Damián, J. L., Rosete Vergés, F., Villalobos Delgado, M., Méndez Linares, A. P., & Navarro Salas, E. (2017). Clasificación geomorfométrica del relieve mexicano: una aproximación morfográfica por densidad de curvas de nivel y la energía del relieve. *Investigaciones Geográficas*, (94). 1
- Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition* (Vol. 1, pp. 278-282).
- Hodgson, Michael & Bresnahan, Patrick. (2004). Accuracy of Airborne LIDAR-Derived Elevation: Empirical Assessment and Error Budget. *Photogrammetric Engineering & Remote Sensing*. 70. 331-339. 10.14358/PERS.70.3.331.
- Hosmer, D.W., Lemeshow, S., & Sturdivant, R.X. (2013). *Applied Logistic Regression*.
- Hox, J. J., & Bechger, T. M. (2014). An Introduction to Structural Equation Modeling. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods: Vol. 1. Basic Methods* (pp. 505-539). Oxford University Press.
- Huggett, R. J. (2003). *Fundamental Concepts in Geomorphology*.
- IGME (2018). *Guía metodológica para la elaboración de mapas geomorfológicos a escala 1:100.000*. Instituto Geológico y Minero de España.
- INEGI. (2020). *Continuo de Elevaciones Mexicano 3.0 (CEM 3.0)*. Recuperado de: <https://www.inegi.org.mx/app/geo2/elevacionesmex/>

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer.
- Jensen, J. R. (2007). *Remote sensing of the environment : An earth resource perspective* (2nd ed.). Pearson Prentice Hall.
- Johnson, A., & Smith, M. (2018). Data preparation for machine learning in geomorphology. *Earth Surface Processes and Landforms*, 43(9), 1853-1864.
- Jones, D., & Smith, P. (2019). Manual Interpretation in Geomorphology. *Geomorphological Methods*, 12(1), 45-56.
- Laney, D. (2001). "3D Data Management: Controlling Data Volume, Velocity, and Variety."
- Li, X., Wang, C., & Dong, Z. (2018). Advances in Remote Sensing Technology. *Journal of Geospatial Science*, 22(4), 213-225.
- Liu, M., Xu, B., Li, Z., Mao, W., Zhu, Y., & Hou, J. (2023). Landslide Susceptibility Zoning in Yunnan Province Based on SBAS-InSAR Technology and a Random Forest Model. *Remote Sensing*, 15, 2864.
- Longley, Paul A., Michael F. Goodchild, David J. Maguire, and David W. Rhind. 2015. "Geographic Information Systems and Science." John Wiley & Sons, 2015.
- Maidment, D. R. (2002). *Arc Hydro : GIS for water resources*. ESRI Press.
- Mainali, K., Evans, M., Madsen, B., Minnemeyer, S., et al. (2023). Convolutional neural network for high-resolution wetland mapping with freely available data. *Science of the Total Environment*, 861, 160622. Recuperado de: <https://doi.org/10.1016/j.scitotenv.2022.160622>
- Marshall, C., & Rossman, G. B. (2016). *Designing Qualitative Research* (6th ed.). SAGE Publications.
- Martín-Serrano, Á. (2017). El mapa geomorfológico. Una cartografía geológica peculiar cafetera cuisinarty útil. *Boletín Geológico y Minero*, 128(4), 933-966.
- Montgomery, D.C., Peck, E.A., & Vining, G.G. (2021). *Introduction to Linear Regression Analysis*. Wiley.
- Moreno, A. (2018). "Macrodatos: Inteligencia de datos en la era digital". *Revista de Tecnologías de Información y Comunicación*, 14(2), 45-58.
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.
- Parrot, J. F. (2013). Modelado tridimensional del relieve mexicano. *Investigaciones Geográficas*, (81), 7-19.

- Parrot, J. F., Ochoa Tejada, A., & Lugo Hubp, J. (2008). Modelado tridimensional del relieve mexicano. *Investigaciones Geográficas*, (67), 7-23. [9]
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
- Phan, T. N., Kuch, V., & Lehnert, L. W. (2020). Land Cover Classification using Google Earth Engine and Random Forest Classifier—The Role of Image Composition. *Remote Sensing*, 12(24), 2411. <https://doi.org/10.3390/rs12152411>
- Pike, R. J. (2000). Geomorphometry—diversity in quantitative surface analysis. *Progress in Physical Geography*, 24(1), 1-20. [10]
- Pike, Richard J., y Martyn J. Smith. (1997) "Topographic variables for environmental modelling: the TIGER topographic database." 1997 *International Journal of Geographical Information Systems* 11.7 (1997): 569-581.
- Popov, A. B., Minár, J., & Drăguț, L. (2024). Physically-Based Digital Geomorphological Mapping: Case Study of Glacial and Karst Topography. *Preprint not peer reviewed*. <https://ssrn.com/abstract=4980663>
- Provost, F., & Fawcett, T. (2013). *Data Science for Business*. O'Reilly Medi
- QGIS Development Team. (2022). QGIS Documentation. Recuperado de <https://docs.qgis.org/3.22/en/docs/index.html>
- Quinlan, J.R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81-106.
- R. Dackombe, V. Gardiner (2004). "Field Techniques for Geologists and Geomorphologists". Routledge.
- Riley, Stephen J., et al. "A terrain ruggedness index that quantifies topographic heterogeneity." *Intermountain Journal of Sciences* 5.1-4 (1999): 23-27.
- Robaina, L. E. de S., & Trentin, R. (2020). Automated classification of landforms with GIS support. Instituto de Geociências - Universidade Federal do Rio Grande do Sul, Brasil. [In Portuguese]
- RSGISLib Image Segmentation Module — RSGISLib. Recuperado de: http://rsgislib.org/rsgislib_segmentation.html.
- Salkind, N. J. (2010). *Encyclopedia of Research Design*. SAGE Publications.
- Scikit-learn. (2022). Documentation. Recuperado de: <https://scikit-learn.org/stable/documentation.html>

- Shepherd, J., Bunting, P., Dymond, J. (2019). Operational Large-Scale Segmentation of Imagery Based on Iterative Elimination Remote Sensing 11(6), 658. Recuperado de: <https://dx.doi.org/10.3390/rs11060658>
- Silberschatz, A., Korth, H. F., & Sudarshan, S. (2010). Database System Concepts. McGraw-Hill.
- Silverman, D. (2016). Qualitative Research (4th ed.). SAGE Publications.
- Smith, J., & Brown, A. (2010). Principles of Geomorphology. Routledge.
- Smith, J., Brown, K., & Thompson, L. (2017). Fieldwork in Geomorphology. *Earth Science Journal*, 9(2), 89-105.
- Smith, M. J., & Wise, S. M. (2007). Problems of bias in mapping linear landforms from satellite imagery. *International Journal of Applied Earth Observation and Geoinformation*, 9(1), 65-78. [11]
- Souza, C. M. P., Figueredo, N. A., Costa, L. M., Veloso, G. V., Almeida, M. I. S., & Ferreira, E. J. (2020). Machine Learning Algorithm in the Prediction of Geomorphic Indices for Appraisal the Influence of Landscape Structure on Fluvial Systems, Southeastern - Brazil. *Revista Brasileira de Geomorfologia*, 21(2), 365-380. <http://dx.doi.org/10.20502/rbg.v21i2.1671>
- Strahler, A. N. (1952). Hypsometric (area-altitude) analysis of erosional topography. *Geological Society of America Bulletin*, 63(11), 1117-1142.
- Tabachnick, B. G., & Fidell, L. S. (2019). Using Multivariate Statistics (7th ed.). Pearson.
- Tarboton, David G. "A new method for the determination of flow directions and upslope areas in grid digital elevation models." *Water Resources Research* 33.2 (1997): 309-319.
- Teece, D.J. (2018). SWOT Analysis. En M. Augier & D.J. Teece (Eds.), *The Palgrave Encyclopedia of Strategic Management*. Palgrave Macmillan. Recuperado de: https://doi.org/10.1057/978-1-137-00772-8_285
- Tucker, G. and Hancock, G. (2010) Modelling Landscape Evolution. *Earth Surface Processes and Landforms*, 35, 28-50.
- van der Meij, M., Meijles, E. W., Marcos, D., Harkema, T. T. L., Candel, J. H. J., & Maas, G. J. (2021). Comparing geomorphological maps made manually and by deep learning. *Earth Surface Processes and Landforms*, 47(4), 1089-1107. <https://doi.org/10.1002/esp.5305>
- Wang, L., Zhang, C., & Li, Z. (2020). Application of machine learning algorithms in geomorphological mapping. *International Journal of Remote Sensing*, 41(9), 3332-3353.
- Wang, L., Zhang, J., & Liu, Y. (2019). A Classification Method of Landform Elements Based on Douglas-Peucker Algorithm for Multi-scale Morphology. ICCDE'19, May 4–6, 2019, Shanghai, China. Recuperado de: <https://doi.org/10.1145/3330530.3332300>

- Wang, Y., Zhang, L., Liu, X., & Chen, Y. (2019). A Geomorphological Regionalization using Upscaled DEM in the Beijing-Tianjin-Hebei Area, China. *Journal of Geomorphology Studies*, 15(2), 102-118.
- Weng, Q. (2012). *Remote sensing and GIS integration: Theories, methods, and applications*. McGraw Hill.
- Yang, J., Xu, J., & Cheng, W. (2023). Deep learning-based automated terrain classification using high-resolution digital elevation models. *International Journal of Applied Earth Observation and Geoinformation*, 118, 103249. <https://doi.org/10.1016/j.jag.2023.103249>

APÉNDICES

Apéndice 1. Código de normalización de capas.

Se normalizaron todas las capas con el siguiente código para ver como influía en la segmentación.

```
!pip install rasterio
import rasterio
from rasterio.enums import Resampling

# Abrir el archivo TIFF
with rasterio.open(r'D:\prueba tesis cdmx\rgslib\CiudadMexico_r15m.tif') as src:

    # Leer los datos del archivo
    data = src.read(1)

    # Normalizar los datos entre 0 y 1
    min_value = data.min()
    max_value = data.max()
    normalized_data = (data - min_value) / (max_value - min_value)

    # Especificar las metadatos para el archivo de salida
    out_meta = src.meta.copy()
    out_meta.update({
        "dtype": 'float32'
    })

    # Escribir los datos normalizados en un nuevo archivo TIFF
    with rasterio.open(r'D:\prueba tesis
cdmx\normalizadocdmx\norma_mde_cdmx.tif', 'w', **out_meta) as dest:
        dest.write(normalized_data, 1)
```

Apéndice 2. Segmentaciones

Segmentaciones con el algoritmo shepherdseg realizadas con disantos parámetros y conjuntos de capas.

parametros	mde	mde + slope + aspect	mde+slope+aspect+profile+relief	RADIAANES mde+slope+aspect+profile+relief	tpi+tri+hillshade
Cluster=2 dist_thres=50 min_n_pxls=100					
Cluster=2 dist_thres=500 min_n_pxls=					
Cluster=2 dist_thres=50 min_n_pxls=400					
Cluster=3 dist_thres=1000 min_n_pxls=600					
Cluster=3 dist_thres=1500 min_n_pxls=500					
Cluster=4 dist_thres=1000 min_n_pxls=600					
Cluster=5 dist_thres=1000 min_n_pxls=600					
Cluster=5 dist_thres=500 min_n_pxls=600					
Cluster=5 dist_thres=250 min_n_pxls=600					
Cluster=10 dist_thres=1000 min_n_pxls=600					
Cluster=10 dist_thres= min_n_pxls=100					