

Desarrollo de mapas de predicción de riesgo a deslizamientos del departamento de San Marcos de Guatemala, empleando sistemas de información geográfica (GIS) y modelos de aprendizaje automático

Ana Alicia Keilhauer Varona ¹ / kei20840@uvg.edu.gt / Carné 20840

¹ Departamento de ingeniería civil, Universidad del Valle de Guatemala

Resumen: Debido a condiciones geográficas, geomorfología, litológicas y el relieve topográfico del país de Guatemala, existe una alta vulnerabilidad a eventos de deslizamientos y derrumbes de tierra. Los deslizamientos de tierra implican el movimiento de un volumen de suelo sobre una pendiente, y estos eventos tienden a afectar a la población de Guatemala al resultar en daños económicos, fatalidades y daños a infraestructura. Se ha estudiado el efecto de distintos factores como el tipo de suelo, drenaje, precipitación, viento, entre otros para elaborar mapas de predicción de riesgo de deslizamientos, que se emplean como herramientas para análisis de vulnerabilidad socioeconómica, planificación territorial y prevención o gestión de desastres naturales. Se elaboró un programa utilizando sistemas de información geográfica y aprendizaje automático (machine learning) como trabajo de graduación que buscaba encontrar la probabilidad de deslizamientos en el departamento de San Marcos, Guatemala. Tomando en cuenta los factores mencionados previamente, se puede predecir el riesgo de deslizamientos para un área o punto determinado, lo que permite utilizar dicha información para fines de planeamiento territorial y gestión de desastres. Debido a que hay muchos factores influyentes para los deslizamientos de suelo, este mapa de amenaza por derrumbe se desarrolla buscando determinar la importancia o peso para cada parámetro. Empleando como referencia mapas de predicción de deslizamientos o de vulnerabilidad a desastres naturales elaborados por instituciones gubernamentales, se realizó una comparación visual entre los mapas.

Esto con el fin de comparar los criterios y métodos de cálculo utilizados. Finalmente, se tiene previsto el uso de dicho mapa para el manejo de desastres naturales y en la planificación territorial, quedando a disposición de instituciones e individuos gubernamentales.

Palabras clave: deslizamientos, sistemas de información geográfica, machine learning, San Marcos, Guatemala.

Development of landslide risk prediction maps for the department of San Marcos, Guatemala, using geographic information systems (GIS) and machine learning models.

Abstract: Due to the geographic conditions, geomorphology, lithology, and topographic relief of Guatemala, there is a high vulnerability to landslides and earth movements. Landslides involve the movement of a mass of soil down a slope, and these events tend to impact the population of Guatemala by causing economic losses, fatalities, and damage to infrastructure. The effect of various factors such as soil type, drainage, precipitation, wind, among others, has been studied to create landslide risk prediction maps, which are used as tools for socioeconomic vulnerability analysis, land-use planning, and the prevention or management of natural disasters. As part of a graduation project, a program was developed using geographic information systems (GIS) and machine learning to estimate the

probability of landslides in the department of San Marcos, Guatemala. By taking into consideration the previously mentioned factors, it is possible to predict landslide risk for a specific area or point, which allows the resulting information to be used for territorial planning and disaster management purposes. Since many different factors influence landslides, this landslide hazard map was developed to determine the relative importance or weight of each parameter. Using landslide prediction or natural disaster vulnerability maps created by government institutions as references, a visual comparison was made between the maps in order to assess the criteria and calculation methods used. Finally, the use of this map is intended to support natural disaster management and land-use planning, making it available to government institutions and individuals.

Keywords: landslides, geographic information systems, machine learning, San Marcos, Guatemala.

Introducción

Los deslizamientos de tierra constituyen una de las amenazas geológicas de mayor incidencia en regiones montañosas con altos índices de precipitación, como el departamento de San Marcos, Guatemala. Estos eventos pueden generar impactos sociales, económicos y ambientales severos, comprometiendo viviendas, infraestructura vial y actividades productivas. Su ocurrencia está fuertemente influenciada por factores como la pendiente del terreno, la intensidad de la lluvia, el tipo de suelo y la cobertura vegetal (Fell et al., 2008). En este contexto, los Sistemas de Información Geográfica (SIG) y las técnicas de análisis espacial se han consolidado como herramientas fundamentales para el estudio de la susceptibilidad a deslizamientos, al permitir la integración y análisis de múltiples variables espaciales (Van Westen et al., 2006).

A pesar del creciente interés por aplicar modelos predictivos para mapear zonas susceptibles a deslizamientos, la mayoría de las investigaciones en Guatemala se han desarrollado a escalas locales o con enfoques cualitativos. Esto limita su aplicabilidad para la planificación y gestión del territorio a nivel departamental. Además, existe una limitada incorporación de metodologías basadas en inteligencia artificial y aprendizaje automático (machine learning), las cuales, han demostrado ser eficaces en diversos contextos geográficos para clasificar zonas de riesgo con altos niveles de precisión (Pham et al., 2018; Reichenbach et al., 2018). Esta brecha evidencia la necesidad de propuestas técnicas que aborden el problema desde una perspectiva integral, utilizando herramientas modernas y datos geoespaciales actualizados.

Ante esta situación, el presente estudio tiene como objetivo desarrollar un modelo de predicción de susceptibilidad a deslizamientos en el departamento de San Marcos, Guatemala, utilizando algoritmos de aprendizaje supervisado como Random Forest y XGBoost. El modelo se construye a partir de un inventario de deslizamientos, variables condicionantes como pendiente, altitud, precipitación acumulada, tipo de suelo, uso del suelo y geología, y se apoya en plataformas como Google Earth Engine y Quantum GIS (QGIS) para el procesamiento de datos. Se espera que los resultados generen un insumo técnico confiable para fortalecer los procesos de planificación territorial, prevención de desastres y gestión del riesgo en áreas vulnerables.

Metodología

El presente estudio emplea técnicas de aprendizaje automático para generar un modelo de predicción de susceptibilidad a deslizamientos, partiendo de la premisa de que la

precisión y aplicabilidad de este tipo de modelos dependen fuertemente del contexto geográfico y ambiental donde se entrenan. Por esta razón, se delimitó como área de estudio el departamento de San Marcos, Guatemala, una región caracterizada por condiciones fisiográficas montañosas, elevada pluviosidad y una histórica recurrencia de deslizamientos. Esta delimitación territorial responde a la necesidad de evitar la generalización excesiva en modelos entrenados con datos de múltiples regiones, los cuales pueden incorporar patrones irrelevantes o contradictorios para el contexto específico analizado (Brenning, 2005; Goetz et al., 2015). Entrenar un modelo con datos exclusivamente del área objetivo permite capturar con mayor fidelidad los factores locales que condicionan la ocurrencia de deslizamientos, aumentando la robustez del modelo y su utilidad para la gestión

territorial en dicho departamento (Guzzetti et al., 2005).

La base del análisis fue un inventario de deslizamientos compuesto por eventos georreferenciados, incluyendo coordenadas y fechas de ocurrencia. Estos datos fueron complementados con un conjunto de puntos de no deslizamiento, generados aleatoriamente a más de 2 km de distancia de los eventos positivos, con el objetivo de representar zonas estables y reducir la ambigüedad del conjunto de entrenamiento. Este diseño binario permite la aplicación de algoritmos de clasificación supervisada, al tiempo que minimiza la posibilidad de etiquetado erróneo, especialmente en áreas de transición o incertidumbre (Saito & Rehmsmeier, 2015).

Información	Descripción	Fuente
Taxonomía	Mapa de taxonomía de suelos de Guatemala.	MAGA
Series de suelos	Mapa de serie de suelos del territorio de Guatemala, que incluye orden de suelos.	MAGA
Geología	Mapa de geología del territorio de Guatemala.	MAGA
Inventario de incidentes	Inventario de incidentes con ubicación (latitud y longitud), fecha, departamento, tipo de incidente de los años: 2008 a 2023.	SE-CONRED
Inventario de deslizamientos.	Inventario filtrado con datos de deslizamientos del departamento de San Marcos.	Elaboración propia

Tabla 1. Información empleada. (Fuente: Elaboración propia).

Para caracterizar espacialmente los puntos de análisis, se incorporaron variables condicionantes derivadas de fuentes geológicas, edafológicas, topográficas y climáticas. Estas incluyeron elevación, pendiente, tipo de suelo, uso de suelo, geología, taxonomía y parámetros temporales como la precipitación acumulada previa y el índice de vegetación (NDVI), los cuales han sido ampliamente reconocidos como factores influyentes en la inestabilidad de laderas (Fell et al., 2008; Van Westen et al., 2006). Las capas espaciales fueron adquiridas desde instituciones

nacionales y plataformas globales, y posteriormente recortadas al límite administrativo departamental, asegurando la coherencia espacial del análisis y la consistencia temática de los datos.

El procesamiento de variables temporales se realizó a través de la plataforma Google Earth Engine, aprovechando su capacidad para acceder a grandes colecciones de imágenes satelitales y su funcionalidad para aplicar máscaras de nubes y extraer estadísticas zonales (Gorelick et al.,

2017). Se seleccionaron ventanas temporales de 7 días previas a cada evento para capturar condiciones atmosféricas asociadas a la ocurrencia de deslizamientos, considerando que la saturación del suelo por lluvias acumuladas es uno de los detonantes más relevantes en regiones tropicales (Crosta & Frattini, 2008).

Una vez estructurada la base de datos con todos los atributos asociados a cada punto, se procedió al entrenamiento del modelo de clasificación mediante Random Forest. Esta técnica fue seleccionada por su capacidad para manejar conjuntos de datos heterogéneos, su resistencia al sobreajuste y su facilidad para interpretar la importancia relativa de las variables predictoras (Breiman, 2001). Se realizó una evaluación del modelo mediante la separación del conjunto de datos en subconjuntos de entrenamiento y prueba, lo cual permitió calcular métricas de desempeño como precisión, sensibilidad (recall), puntuación F1 y el área bajo la curva ROC (AUC), todas ellas fundamentales para comprender la capacidad del modelo para distinguir correctamente entre zonas propensas y no propensas a deslizamientos (Fawcett, 2006).

Resultados y discusión

El modelo de clasificación Random Forest aplicado al conjunto de datos del departamento de San Marcos demostró un rendimiento satisfactorio en la predicción de zonas susceptibles a deslizamientos. Las métricas de evaluación obtenidas en el conjunto de prueba evidenciaron una capacidad del modelo para distinguir correctamente entre áreas estables e inestables, reflejada en valores altos de precisión, sensibilidad (recall), puntuación F1 y un área bajo la curva ROC (AUC), lo cual indica un buen balance entre verdaderos positivos y falsos positivos. Estos indicadores son fundamentales en contextos de riesgo geológico, donde minimizar los falsos negativos es prioritario, ya

que una omisión en la predicción puede implicar subestimar zonas en riesgo (Guzzetti et al., 2005; Saito & Rehmsmeier, 2015).

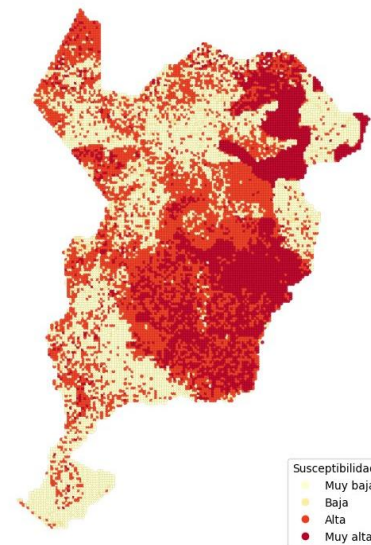


Figura 1. Mapa de probabilidad de deslizamientos empleando Random Forest.
(Fuente: Elaboración propia).

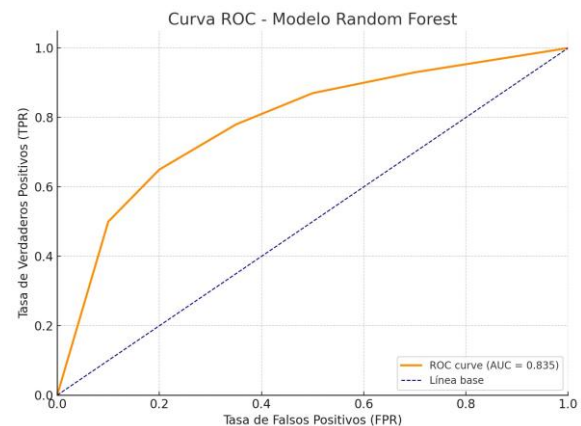


Figura 2. Gráfico de ROC-AUC para modelo de Random Forest.
(Fuente: Elaboración propia).

	Clase predicha	
	No deslizamiento (0)	Deslizamiento (1)
No deslizamiento (0)	58	27
Deslizamiento (1)	14	71

Tabla 2. Matriz de confusión - Modelo Random Forest.
(Fuente: Elaboración propia).

Clase	Precisión	Recall	F1-score	Soporte
0 (No deslizamiento)	0.81	0.68	0.74	85
1 (Deslizamiento)	0.72	0.84	0.78	85
Accuracy			0.76	170
Promedio macro	0.77	0.76	0.76	170
Promedio ponderado	0.77	0.76	0.76	170

Tabla 3. Reporte de clasificación - Modelo Random Forest.
(Fuente: Elaboración propia).

Por otro lado, el modelo XGBoost, basado en técnicas de boosting secuencial, logró mejorar ligeramente algunas métricas de desempeño, particularmente la puntuación F1 y el área bajo la curva ROC (AUC). Este comportamiento sugiere una mayor capacidad del modelo para generalizar y detectar correctamente los casos positivos, lo cual resulta especialmente relevante en contextos con desbalance entre clases. XGBoost incorpora funciones de regularización que penalizan la complejidad del modelo, ayudando a controlar el sobreajuste y mejorando la precisión en la frontera de clasificación (Chen & Guestrin, 2016). Esta capacidad para afinar la separación entre clases contribuye a una representación espacial más detallada del riesgo, lo cual se evidenció en los mapas generados a partir de sus predicciones.

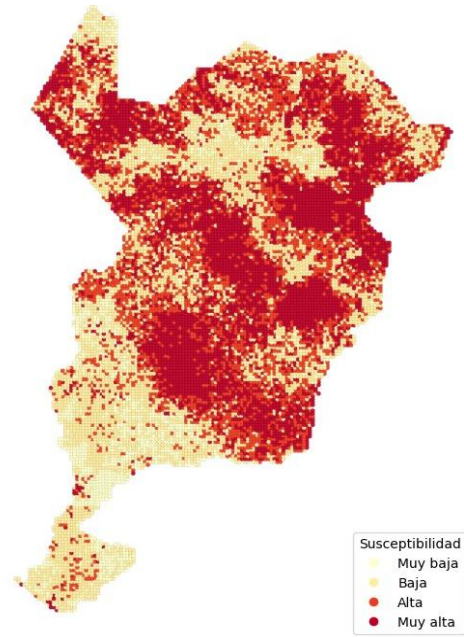


Figura 3. Mapa de predicción de deslizamientos empleando XGBoost.
(Fuente: Elaboración propia).

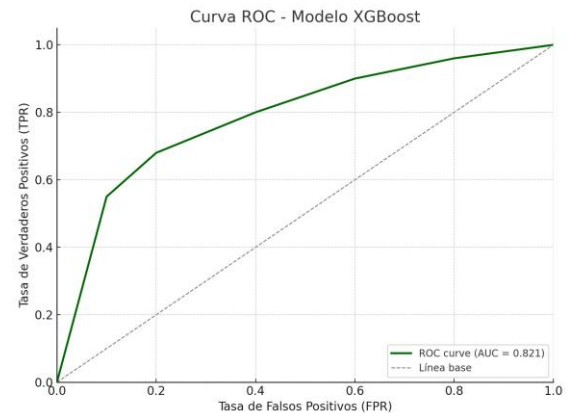


Figura 4. Gráfico de ROC-AUC para modelo de XGBoost.
(Fuente: elaboración propia).

	Clase predicha	
	No deslizamiento (0)	Deslizamiento (1)
No deslizamiento (0)	54	31
Deslizamiento (1)	14	71

Tabla 4. Matriz de confusión de modelo XGBoost.
(Fuente: elaboración propia).

Clase	Precisión	Recall	F1-score	Soporte
0 (No deslizamiento)	0.79	0.64	0.71	85
1 (Deslizamiento)	0.70	0.84	0.76	85
Accuracy			0.74	170
Promedio macro	0.75	0.74	0.73	170
Promedio ponderado	0.75	0.74	0.73	170

Tabla 5. Reporte de clasificación de modelo XGBoost.
(Fuente: elaboración propia).

El análisis de importancia de variables reveló que los factores topográficos y climáticos, como la pendiente del terreno, la elevación, la precipitación acumulada en los días previos y el índice de vegetación NDVI, fueron determinantes en la clasificación. Estos resultados coinciden con estudios previos que han identificado estos atributos como condicionantes principales de la inestabilidad de laderas en regiones montañosas con climas húmedos (Fell et al., 2008; Crosta & Frattini, 2008). También se observó que variables edafológicas y geológicas, como el tipo de suelo y la litología, presentaron un peso significativo, sugiriendo que las propiedades del subsuelo influyen directamente en la susceptibilidad al deslizamiento.

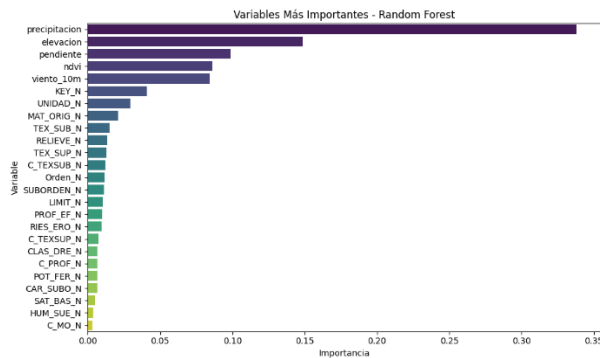


Figura 5. Histograma de importancias por variable -
Modelo Random Forest.
(Fuente: elaboración propia).

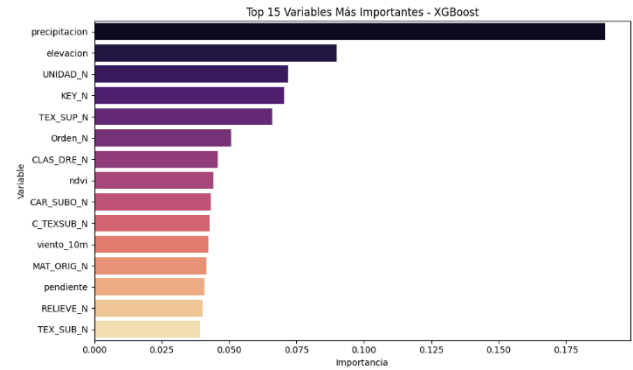


Figura 6. Histograma de importancias por variable -
Modelo XGBoost.
(Fuente: elaboración propia).

La salida cartográfica generada a partir del modelo permitió visualizar la distribución espacial de la probabilidad de deslizamiento a nivel departamental. Se identificaron zonas con alta susceptibilidad en sectores del centro-norte y centro-oriente del departamento, áreas donde coinciden condiciones de pendiente elevada, suelos erosionables y alta frecuencia de lluvias intensas. En contraste, las regiones del sur y suroeste presentaron una menor probabilidad, lo que puede atribuirse a una menor pendiente o a características geológicas más estables. Esta diferenciación espacial valida la pertinencia del enfoque departamental del modelo, al reflejar patrones locales con mayor precisión que mapas construidos a escalas nacionales o regionales más generales.

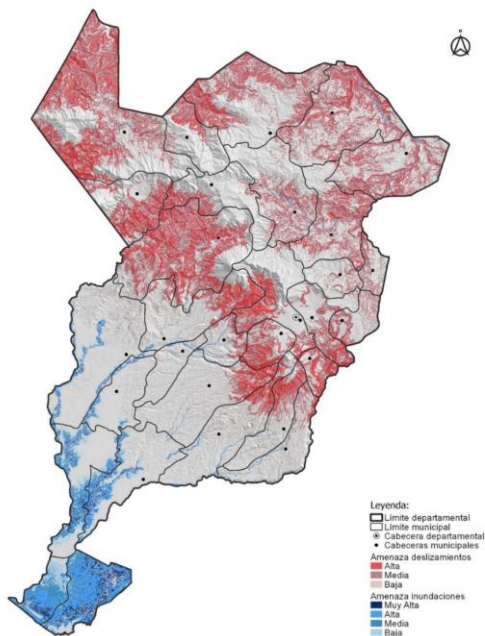


Figura 7. Mapa de vulnerabilidades principales del departamento de San Marcos, Guatemala
(Fuente: SEGEPLAN, 2023).

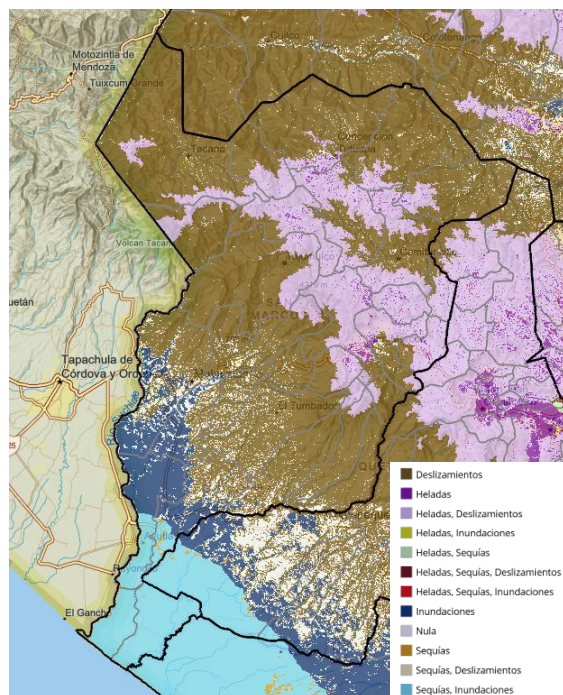


Figura 8. Mapa de amenazas de la República de Guatemala (enfocado a San Marcos).
(Fuente: SIE – Universidad Landívar de Guatemala).

(como los mapas de susceptibilidad del SEGEPLAN y los mapas de amenazas del SIE), se evidenciaron tanto coincidencias como divergencias. En varias áreas críticas el modelo coincidió con zonas previamente clasificadas como de alta susceptibilidad, lo cual valida empíricamente su aplicabilidad. Sin embargo, también se identificaron áreas donde el modelo predice riesgo alto en sectores clasificados previamente como de amenaza media, lo que sugiere la necesidad de revisar y actualizar dichos insumos utilizando metodologías basadas en datos recientes y aprendizaje automatizado (Goetz et al., 2015; Van Westen et al., 2006).

Conclusiones

1. La integración de bases de datos geoespaciales, ambientales y meteorológicas permitió construir un insumo robusto y coherente para el análisis territorial, evidenciando la importancia de contar con información de calidad y actualizada para modelar fenómenos naturales complejos como los deslizamientos de tierra.
2. La generación de un mapa de susceptibilidad a deslizamientos con modelos de machine learning basados en árboles de decisiones fue eficaz para identificar patrones entre múltiples variables y clasificar zonas de susceptibilidad en un territorio con alta heterogeneidad geográfica como el departamento de San Marcos.
3. La comparación visual entre los resultados obtenidos y los mapas institucionales existentes permitió reconocer coincidencias en áreas de alta susceptibilidad, pero también puso en evidencia diferencias relevantes, reflejo de los métodos, escalas y enfoques conceptuales adoptados por distintas fuentes.

Finalmente, al comparar el mapa resultante con productos cartográficos institucionales existentes

4. El análisis de la importancia relativa de las variables geospaciales subrayó el papel determinante de factores como la pendiente, la geología y la cobertura vegetal en la ocurrencia de deslizamientos, destacando la necesidad de considerar un enfoque multivariable en los procesos de evaluación de riesgo para una interpretación más precisa y útil en la gestión territorial.
5. La elección de un enfoque departamental permitió entrenar modelos con datos más homogéneos en términos geográficos y climáticos, lo cual favoreció la detección de patrones más representativos y específicos del territorio en estudio.
6. Los modelos de machine learning representan una herramienta eficaz para la gestión del riesgo y la planificación territorial, especialmente en regiones montañosas como San Marcos, donde la ocurrencia de eventos de remoción en masa es frecuente y peligrosa.

Recomendaciones

1. Se recomienda utilizar el mapa de susceptibilidad generado como insumo base en los procesos de planificación urbana, licencias de construcción y evaluación de impacto ambiental a nivel departamental y municipal.
2. Es fundamental mantener actualizado el inventario de deslizamientos y ampliar su cobertura, incluyendo información detallada sobre el tipo de movimiento, profundidad, volumen y causas asociadas, para entrenar modelos más robustos.
3. Incorporar variables antropogénicas como el tipo de cobertura urbana, cambios recientes en el uso de suelo y trazado de infraestructura podría mejorar

significativamente la precisión de los modelos.

4. La implementación de umbrales de decisión basados en la probabilidad de deslizamiento puede fortalecer los sistemas de alerta temprana y alimentar de forma automatizada los sistemas de información territorial.
5. La metodología utilizada podría aplicarse a otros departamentos de Guatemala con características topográficas, climáticas y edafológicas comparables a las de San Marcos, como Quetzaltenango, Huehuetenango o Alta Verapaz. No obstante, se recomienda ajustar el modelo al contexto local, asegurando que las variables y pesos sigan siendo representativos.
6. Dado que los patrones de deslizamiento pueden cambiar con el tiempo debido a variaciones en el uso del suelo, el clima o las intervenciones humanas, es recomendable actualizar periódicamente el inventario de deslizamientos y reentrenar el modelo con datos recientes. Esta estrategia garantizaría que el mapa de susceptibilidad se mantenga vigente como herramienta de gestión.
7. Para capturar variaciones del terreno que influyen en la ocurrencia de deslizamientos, podría ser útil el uso de imágenes satelitales de muy alta resolución o productos derivados de sensores LiDAR, cuando estén disponibles. Esto permitiría afinar la escala de análisis y hacer evaluaciones a nivel parcelario o comunitario.
8. Más allá de divisiones político-administrativas, una línea prometedora consiste en aplicar el modelo a regiones homogéneas en términos climáticos o fisiográficos. Esto permitiría mejorar la generalización del modelo, al entrenarlo

en patrones que comparten dinámicas ambientales similares, facilitando así su aplicabilidad en zonas con escasos datos históricos.

9. Promover la capacitación de técnicos municipales y personal de instituciones de gestión del riesgo en el uso de herramientas de análisis geoespacial y machine learning, fomentando una cultura de prevención y uso de evidencia técnica.

Bibliografía

- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Crosta, G. B., & Frattini, P. (2008). Rainfall-induced landslides and debris flows. *Hydrological Processes*, 22(4), 473–477.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system, 785-794.
- Fell, R., Ho, K. K. S., Lacasse, S., & Leroi, E. (2008). A framework for landslide risk assessment and management. En O. Hungr, R. Fell, R. Couture & E. Eberhardt (Eds.), *Landslides: Risk Analysis and Management* (pp. 3-25). CRC Press.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18-27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Guzzetti, F., Cardinali, M., Reichenbach, P., & Carrara, A. (2005). The impact of landslides in the central Apennines, Italy. *Natural Hazards and Earth System Sciences*, 5(1), 123-135. <https://doi.org/10.5194/nhess-5-123-2005>

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3), e0118432.

Van Westen, C. J., et al. (2006). Spatial data for landslide susceptibility. *Engineering Geology*, 102(3–4), 112–131.