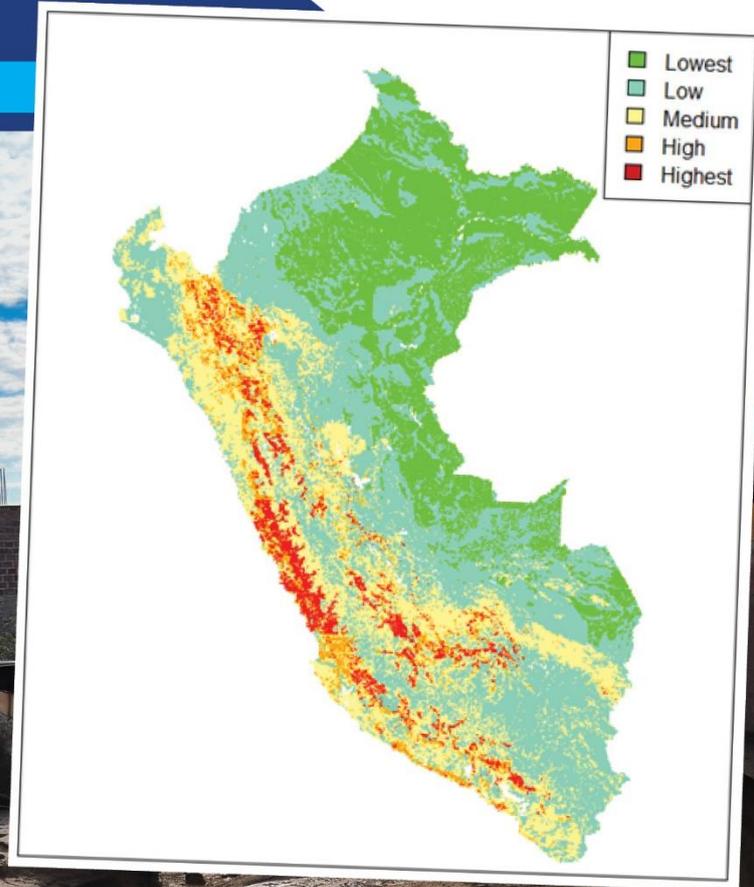


# MODELAMIENTO DE SUSCEPTIBILIDAD A MOVIMIENTOS EN MASA PRODUCTO DE LLUVIAS INTENSAS A NIVEL NACIONAL

ESTUDIO FINAL



**Citar:**

**Risco, E.; Lavado-Casimiro, W. y Felipe, O. (2022).** *Modelamiento de susceptibilidad a movimientos en masa producto de lluvias intensas a nivel nacional.* Estudio Final. Servicio Nacional de Meteorología e Hidrología del Perú –SENAMHI.

<https://repositorio.senamhi.gob.pe/handle/20.500.12542/1782>

**MODELAMIENTO DE SUSCEPTIBILIDAD A  
MOVIMIENTOS EN MASA  
PRODUCTO DE LLUVIAS INTENSAS A  
NIVEL NACIONAL**

**ESTUDIO FINAL**





## **DIRECTORIO**

PhD. Guillermo Antonio Baigorria Paz  
PRESIDENTE EJECUTIVO DEL SENAMHI

Ing. Oscar Gustavo Felipe Obando  
DIRECTOR DE HIDROLOGÍA DEL SENAMHI

PhD. Waldo Sven Lavado Casimiro  
SUBDIRECTOR DE ESTUDIOS E INVESTIGACIONES HIDROLÓGICAS DEL  
SENAMHI

## **REVISORES DEL ESTUDIO**

PhD. Waldo Sven Lavado Casimiro  
Ing. Oscar Gustavo Felipe Obando

## **AUTORES**

MSc. Eber Rostaing Risco Sence  
PhD. Waldo Sven Lavado Casimiro  
Ing. Oscar Gustavo Felipe Obando

**Lima-Perú**

**Mayo – 2022**

## Contenido

|   |   |    |
|---|---|----|
| 1 | Introducción.....   | 7  |
| 2 | Materiales y métodos.....   | 8  |
|   | a) Área de estudio.....   | 8  |
|   | b) Datos.....   | 8  |
|   | Variables Predictoras.....  | 8  |
|   | Datos de ocurrencia de remoción en masa producto de lluvias intensas..... | 14 |
|   | c) Metodología.....   | 15 |
|   | Modelo de Random Forest.....  | 15 |
|   | Implementación del modelo.....  | 16 |
| 3 | Resultados y discusión.....   | 16 |
|   | a) Análisis de sensibilidad de parámetros.....                            | 16 |
|   | b) Análisis explicativo de la importancia de los factores.....            | 17 |
| 4 | Conclusiones.....   | 23 |
| 5 | Referencias bibliográficas.....   | 28 |

## Contenido de figuras

|                   |   |    |
|-------------------|---|----|
| <b>Figura 1.</b>  | <i>Ubicación espacial de ocurrencias de movimientos en masa producto de lluvias intensas en el territorio peruano.....</i>  | 9  |
| <b>Figura 2.</b>  | <i>Variables explicativas en el desarrollo del modelo de RF.....</i>  | 11 |
| <b>Figura 3.</b>  | <i>Variables explicativas en el desarrollo del modelo de RF.....</i>  | 13 |
| <b>Figura 4.</b>  | <i>Precisión del modelo en función de los parámetros que se utilizaron en el modelo de RF (mtry y ntree son dos parámetros importantes del modelo de RF). .....</i>                       | 17 |
| <b>Figura 5.</b>  | <i>Importancia relativa de los factores de susceptibilidad a la ocurrencia de movimientos en masa producto de lluvias intensas. ....</i>  | 18 |
| <b>Figura 6.</b>  | <i>Análisis de eliminación de características recursivas (RFE). El punto en azul indica la cantidad óptima de factores para el modelo.....</i>  | 19 |
| <b>Figura 7.</b>  | <i>Mapa de susceptibilidad de ocurrencia de movimientos en masa producto de lluvias intensas. Izquierda: usando 26 predictores. Derecha: usando los 12 predictores seleccionados.....</i> | 20 |
| <b>Figura 8.</b>  | <i>Área bajo la curva para los datos de entrenamiento (80%) y datos de prueba (20%). .....</i>  | 21 |
| <b>Figura 9.</b>  | <i>Distribución de las ocurrencias de movimientos en masa producto de lluvias intensas según susceptibilidad obtenida. ....</i>   | 21 |
| <b>Figura 10.</b> | <i>Mapa de susceptibilidad de ocurrencia de movimientos en masa producto de lluvias intensas. ....</i>  | 22 |

## Contenido de tablas

|                 |   |    |
|-----------------|---|----|
| <b>Tabla 1:</b> | <i>Frecuencia de ocurrencia de remoción en masa producto por fuertes lluvias hasta el 2021. Fuente. SENAMHI- SIMPAD.....</i>  | 14 |
| <b>Tabla 2:</b> | <i>Clasificación de susceptibilidad a movimientos en masa producto de lluvias intensas.....</i>   | 24 |
| <b>Tabla 3:</b> | <i>Recopilación y sistematización de Estudios que han aplicado métodos de aprendizaje automático para la representación espacial de la susceptibilidad de deslizamiento de tierra e inundación en territorios de montañosos en Perú, América y el mundo. ....</i> | 25 |

# MAPA DE SUSCEPTIBILIDAD A MOVIMIENTOS EN MASA PRODUCTO DE LLUVIAS INTENSAS A NIVEL NACIONAL

## 1. Introducción

Los deslizamientos de tierra causan grandes pérdidas en todo el mundo en términos de recursos económicos, naturales y vidas humanas, dañando infraestructura como carreteras, puentes, líneas eléctricas, así como suelos y cultivos, provocando intensos procesos de erosión de suelo (Bălteanu et al., 2020; Orhan, Bilgilioglu, Kaya, Ozcan, & Bilgilioglu, 2020; Thi Ngo et al., 2021). La mayoría de los deslizamientos de tierra ocurren como resultado de eventos hidroclicmáticos, como lluvias prolongadas o intensas (Nhu et al., 2020). Ya sean provocados naturalmente o por actividades humanas, en particular los deslizamientos de tierra son recurrentes a lo largo de las carreteras y en pendientes abruptas en regiones montañosas.

En Perú, los deslizamientos de tierra representan una amenaza constante debido a la presencia de fuertes lluvias que se concentran en los meses de verano, más aún cuando se presenta el fenómeno del niño. El territorio peruano es una de las zonas más inestables de los Andes, siendo las causas de esta precaria condición el clima y la geología del territorio. Es de resaltar, que las evidencias históricas revelan que el Perú ha sido afectado por procesos geológicos como los movimientos en masa, los cuales alcanzaron a centros prehispánicos como Chavín (800 a.C. 200 a.C.) hasta la total devastación de ciudades andinas como Ranrahirca y Yungay (alud – avalancha de 1962 y 1970) (S. Villacorta, L. Fidel, 2012). Sin embargo, los complejos mecanismos y la falta de datos son dos obstáculos principales en el mapeo de movimientos en masa producidos por lluvias intensas dentro del territorio peruano.

En los últimos años, el avance de la tecnología permite la captura de información a partir de sensores remotos que representan variable climáticas, relieve del terreno o las condiciones del suelo de un lugar en particular (Sobie, 2020). A consecuencia del rápido crecimiento de los datos debido al desarrollo de la tecnología de la información y la comunicación (TIC) ha estimulado la demanda de minería de datos, aprendizaje automático, redes neuronales que generalmente se consideran como las herramientas analíticas más útiles en el análisis de grandes conjuntos de datos (S. Lee, Lee, & Jung, 2017).

Las áreas propensas a deslizamientos de tierra se pueden identificar y evaluar con base en análisis científicos. Por lo tanto, las áreas propensas a deslizamientos de tierra generalmente se identifican mediante modelos de susceptibilidad a deslizamientos de tierra (J. H. Lee, Sameen, Pradhan, & Park, 2018). Dichos métodos han progresado en los últimos años, como modelos propuestos mediante herramientas de sistemas de información geográfica (S. Villacorta, L. Fidel, 2012; Wang et al., 2018), en los últimos

años muchos investigadores han propuesto y aplicado numerosas técnicas como algoritmos de aprendizaje profundo como la red neuronal recurrente y la red neuronal convolucional (Thi Ngo et al., 2021), aprendizaje automático (Nhu et al., 2020; Pham et al., 2019; Wang et al., 2018), especialmente el algoritmo de Random Forest (Catani, Lagomarsino, Segoni, & Tofani, 2013; de Oliveira, Ruiz, Guasselli, & Haetinger, 2019; Orhan et al., 2020; Park & Kim, 2019; Zhao, Pang, Xu, Yue, & Tu, 2018), técnicas de minería de datos (J. H. Lee et al., 2018; S. Lee et al., 2017), modelo aditivos generalizados (Bordoni et al., 2020). Cada método tiene sus propias ventajas y desventajas según la base conceptual del modelo, los conjuntos de datos utilizados y la topografía del área de estudio, algunos estudios muestran que el algoritmo de Random Forest presenta ventajas sobre otros métodos de minería de datos y aprendizaje automático (Park & Kim, 2019; Zhao et al., 2018).

El presente estudio propone el desarrollo de un mapa de susceptibilidad a movimientos en masa producto de lluvias intensas en el territorio peruano y proporciona una herramienta de gestión del territorio para la mitigación de eventos de remoción en masa potencialmente desastrosos. Para ello, se ha realizado un trabajo sistemático, que ha permitido recopilar datos in situ de los movimientos en masa producto de las lluvias intensas, que han servido para el ajuste y validación de un modelo de predicción de probabilidad de ocurrencia del evento de remoción en masa producto de lluvias intensas mediante técnicas de aprendizaje automático (*machine learning*) usando el algoritmo de bosques aleatorios (*Random Forest*).

## 2. Materiales y métodos

### a) Área de estudio

Perú está ubicado en la costa oeste del continente sudamericano. Tiene una superficie de 1.285.220 km<sup>2</sup> y una población de aproximadamente 32.5 millones de personas. Limita al oeste con el Océano Pacífico, al norte con Ecuador y Colombia, y al sureste con Brasil, Bolivia y Chile. La cordillera de los Andes crea una topografía compleja e introduce variabilidad climática a lo largo de sus tres regiones hidrográficas: Pacífico, Atlántico y Titicaca. Esta barrera orográfica natural atrapa la humedad atmosférica del Atlántico, produciendo altas precipitaciones en la región andino-amazónica y tierras bajas amazónicas (lado este) y escasas precipitaciones en la costa (lado occidental) (Lavado Casimiro, Labat, Guyot, & Ardoin-Bardin, 2011). Estas fuertes precipitaciones cuando son continuas generan movimientos en masa que se concentran a lo largo de esta barrera orográfica natural como se puede apreciar en la Figura 1.

### b) Datos

#### VARIABLES PREDICTORAS

El principal desafío en la evaluación de la susceptibilidad a movimientos en masa producto de las lluvias intensas es establecer relaciones multivariadas y no lineales

### UBICACIÓN ESPACIAL DE OCURRENCIAS DE REMOCIÓN EN MASA PRODUCTO DE FUERTES LLUVIAS

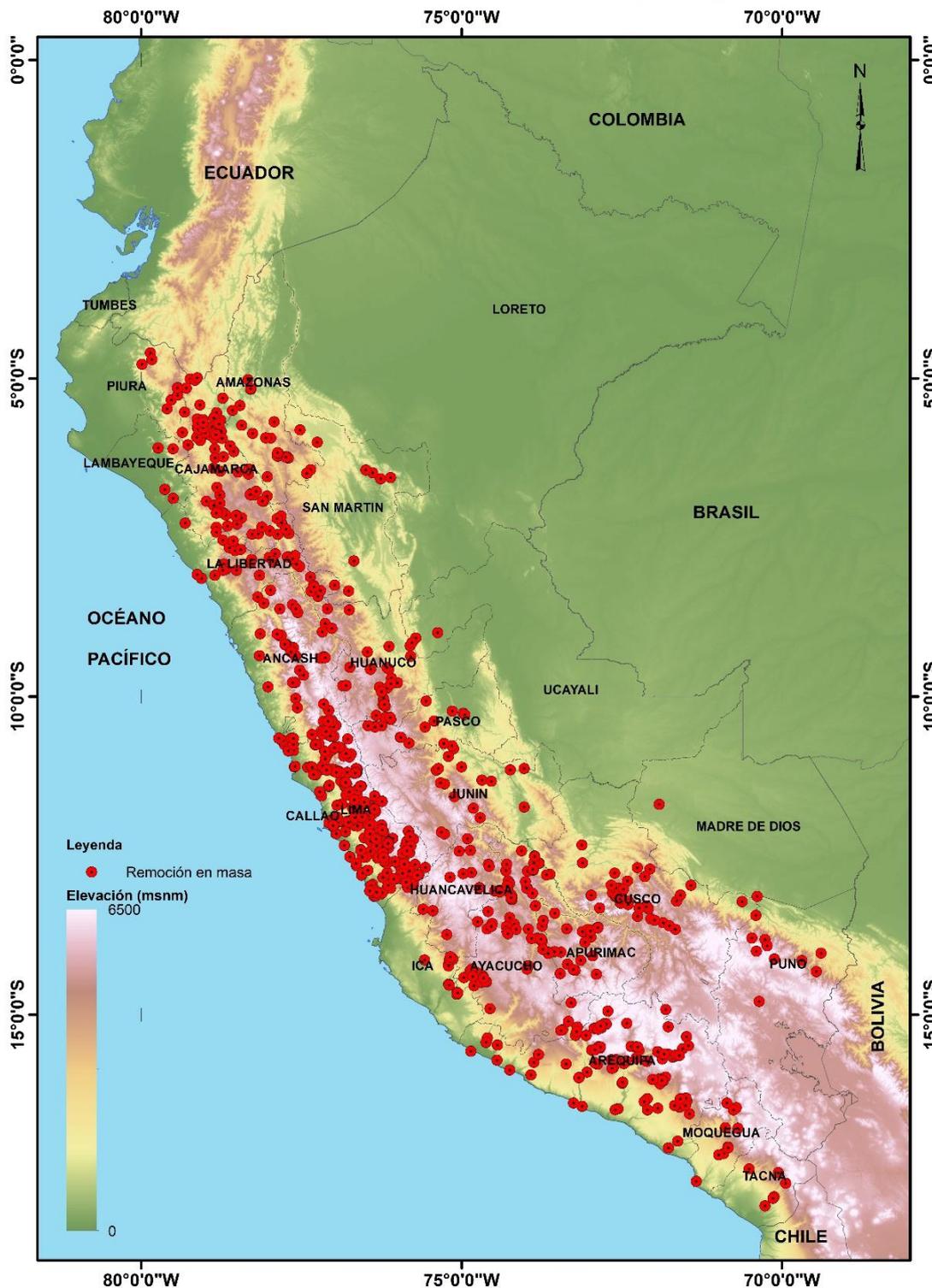


Figura 1. Ubicación espacial de ocurrencias de movimientos en masa producto de lluvias intensas en el territorio peruano.

entre los factores explicativos y las frecuencias de ocurrencia. Identificar estas variables explicativas es el primer paso importante. Un total de 26 variables explicativas de ubicación geográfica, topografía del terreno, meteorológicas, hidrológicas, de cobertura y suelo fueron seleccionados y aplicados en este estudio con base en una investigación de registros históricos de ocurrencia de movimientos en masa producto de lluvias intensas: longitud (LO) , latitud (LA), elevación (EV), pendiente (SL), aspecto (AS), elevación relativa (RE), densidad de drenaje (DD), número de curva ( CN), índice topográfico de humedad (TWI), índice de potencia de las corrientes (SPI), índice de transporte de sedimentos (STI), humedad del suelo promedio de noviembre-abril (SM), humedad del subsuelo promedio de noviembre-abril (SSM), Factor LS de USLE (LS), precipitación diaria máxima promedio anual (AP), precipitación diaria extrema (EP), lluvia antecedente con una ventana de 08 días considerando los meses de noviembre-abril: máxima y promedio (MAP, AAP), frecuencia de tormentas fuertes (Umbral = percentil de lluvia 95 (noviembre-abril)) (FP), índice de vegetación de diferencia normalizada (NDVI), litología (LT), cobertura de suelo – MODIS (SC), geología (GEOL), geomorfología (GEOM), textura del suelo (a 0, 10, 30 cm).

Variables de ubicación: Longitud: generado como una interpolación lineal a partir de las longitudes dentro del área de estudio (territorio peruano) a 90 m. Latitud: generado como una interpolación lineal a partir de las latitudes dentro del área de estudio (territorio peruano) a 90 m.

Variables de terreno: Elevación: obtenido del catálogo de datos de la plataforma Google Earth Engine (GEE): MERIT Hydro: global hydrography datasets a una resolución espacial de 90 m. Pendiente: obtenido como la primera derivada del modelo digital de elevaciones. Aspecto: obtenido del catálogo de datos de la plataforma Google Earth Engine (GEE): MERIT Hydro: global hydrography datasets a una resolución espacial de 90 m. Elevación relativa: Se seleccionó la elevación relativa (*RE*) para revelar el grado de cambio de elevación dentro de una celda en una ventana de 3x3. RE se calculó de acuerdo a la siguiente ecuación:

$$RE_i = E_i - E_{min}$$

$E_i$  representa la elevación promedio de la cuadrícula  $i$  y  $E_{min}$  representa la elevación mínima de las 8 cuadrículas alrededor de  $i$ . Geomorfología: obtenido a partir del catálogo de datos del Ministerio del Ambiente (MINAM), en la cual presente 54 unidades geomorfológicas en el territorio peruano.

Variables hidrológicas: Densidad de drenaje: La densidad de drenaje se define como la longitud total de la corriente por unidad de área, que se puede calcular como se muestra en la siguiente ecuación:

$$DD = \frac{1}{S} \sum_i L_i^S$$

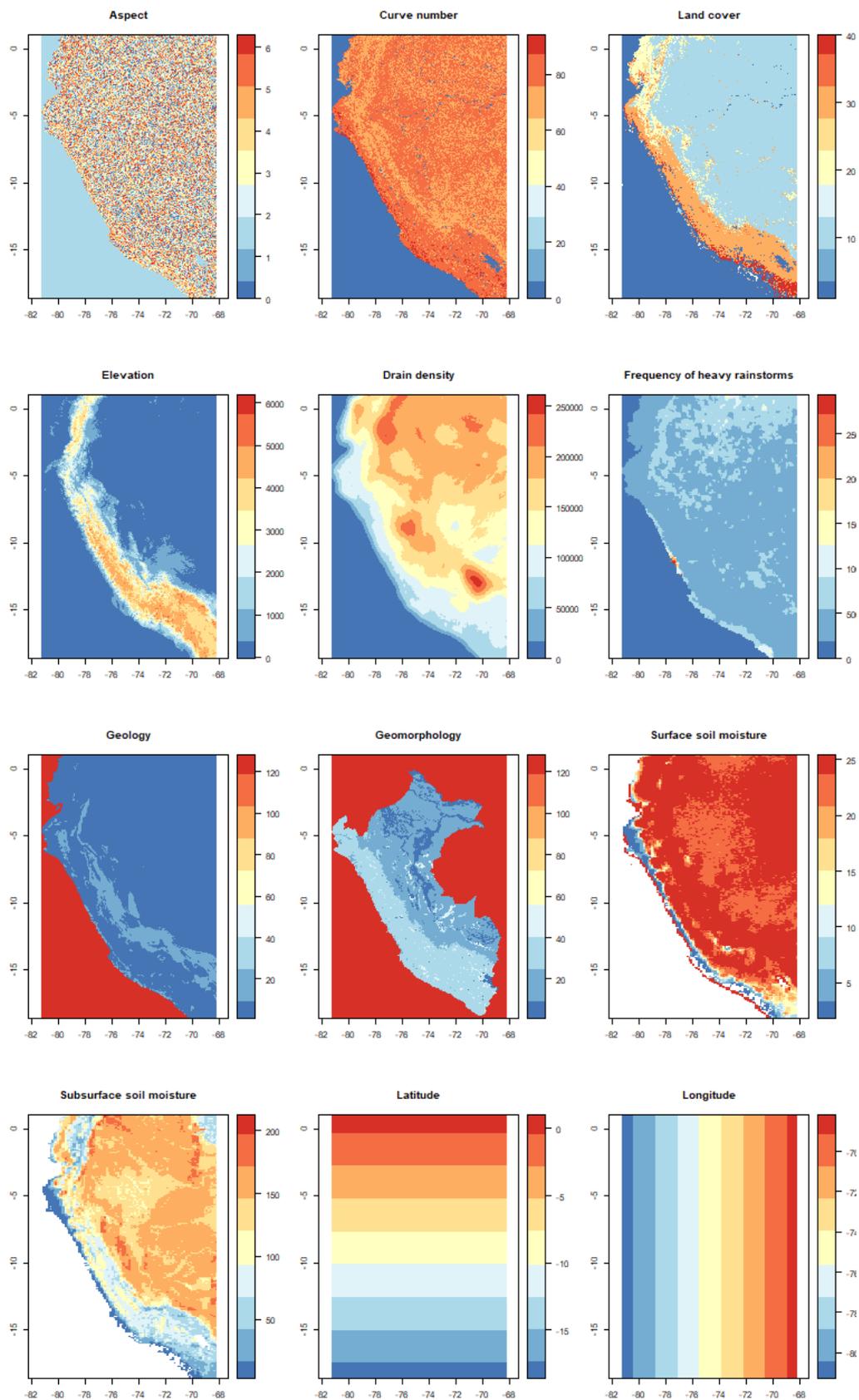


Figura 2. Variables explicativas en el desarrollo del modelo de RF.

Donde  $S$  representa el área de la cuadrícula y  $L_i^S$  representa la longitud del río  $i$  dentro de la cuadrícula. Teniendo en cuenta la influencia de los ríos dentro de las celdas de la cuadrícula vecinas, el área de la cuadrícula  $S$  incluía la celda actual y ocho celdas vecinas. Número de curva: Fue obtenido a partir de la fuente de datos *HYSOGs250m, global gridded hydrologic soil groups for curve-number-based runoff modeling*, disponible en <https://www.nature.com/articles/sdata201891>. Índice topográfico de humedad: El  $TWI$  combina la contribución al escurrimiento de un área local drenada y la pendiente de la misma, y es comúnmente usado para cuantificar el control topográfico sobre los procesos hidrológicos y está definido como (Sørensen, Zinko, & Seibert, 2006):

$$TWI = \ln \left[ \frac{A}{\tan(\beta)} \right]$$

$A$  es el área local drenada para un punto de cálculo, y  $\beta$ , es la pendiente direccional de la celda de interés (y de las 8 vecinas en el caso de utilizar un algoritmo D8). Índice de potencia de las corrientes: Se basa en una predicción de la tasa de incisión en el lecho rocoso como una función de la potencia del flujo de agua (Finlayson et al., 2002; Tucker y Whipple, 2002.). La fórmula general es:

$$e = k \cdot A_s^m \cdot S^n$$

$e$  refleja la tasa de incisión local,  $A_s$  es el área de drenaje (desagüe) aguas arriba (empleada como una aproximación para el caudal),  $S$  es el gradiente de la pendiente local,  $m$ ,  $n$  son constantes;  $k$  representa la dureza de la roca (relacionado a la litología), Whipple (2001). Índice de transporte de sedimentos: Es un índice utilizado para estimar el potencial topográfico para la erosión o deposición por medio de una expresión que representa el cambio en la capacidad de transporte de sedimentos en la dirección del flujo (Martínez 1999).

$$STC = \left( \frac{A_s}{22.13} \right)^{0.6} \cdot \left( \frac{\sin(\beta)}{0.0896} \right)^{1.3}$$

$STC$ =índice de capacidad de transporte de sedimentos,  $A_s$ =área de drenaje específica,  $\beta$  = ángulo de la pendiente. Humedad del suelo (noviembre-abril): obtenido como el promedio de los meses de noviembre a abril 2015-2020 de la *NASA-USDA Enhanced SMAP Global Soil Moisture Data*, disponible en el catálogo de datos de GEE. Humedad del subsuelo (noviembre-abril): obtenido como el promedio de los meses de noviembre a abril 2015-2020 de la *NASA-USDA Enhanced SMAP Global Soil Moisture Data*, disponible en el catálogo de datos de GEE. Factor LS de USLE: obtenido a partir del modelo digital de elevaciones de acuerdo a la ecuación de pérdida de suelo.

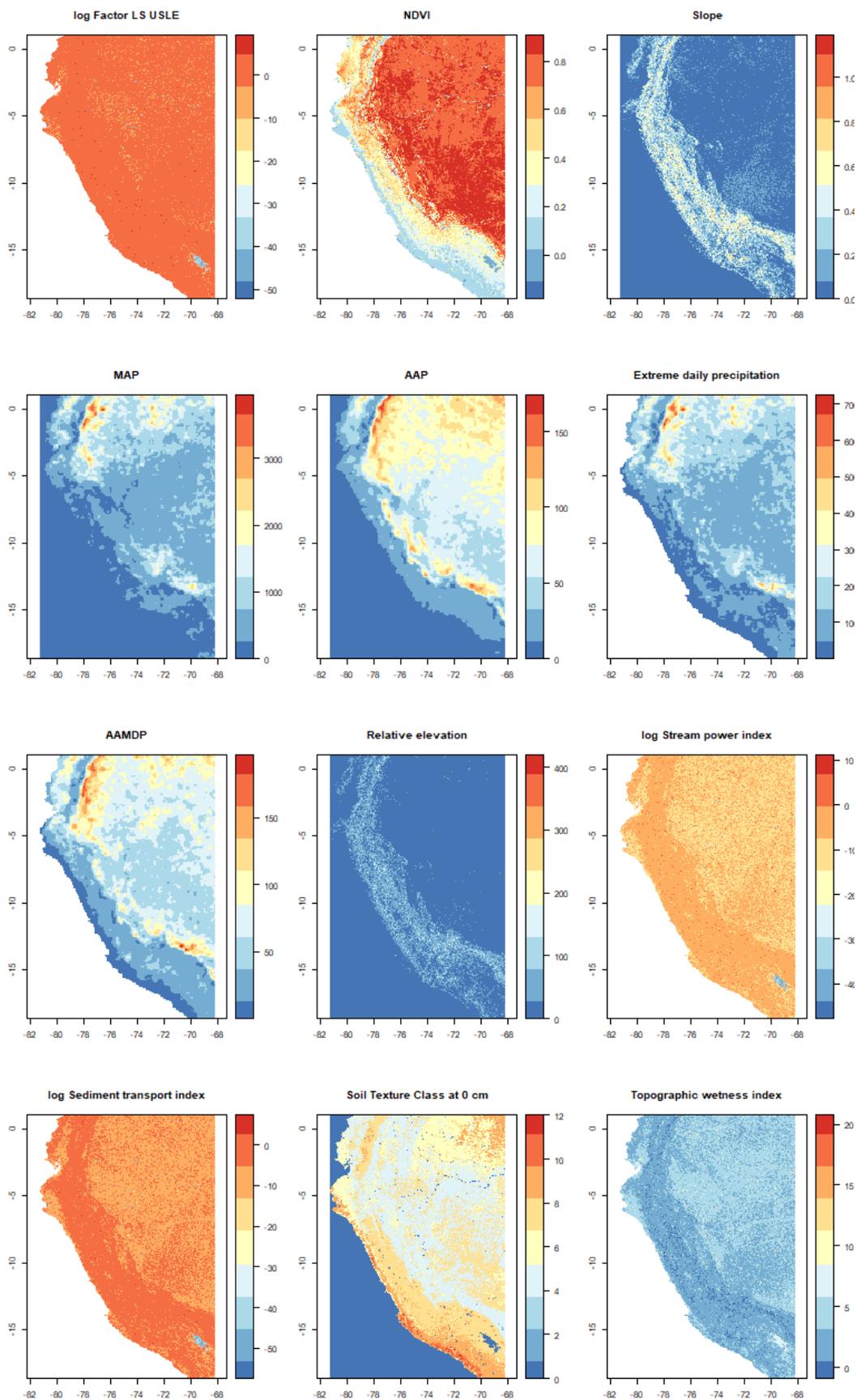


Figura 3. Variables explicativas en el desarrollo del modelo de RF

Variables meteorológicas: Precipitación diaria máxima promedio anual: obtenido a partir del producto PISCO - precipitación diaria como el promedio de todos los máximos diarios anuales (1981-2020). Precipitación diaria extrema: obtenido a partir del producto PISCO - precipitación diaria como el máximo extremo diario (1981-2020). Lluvia antecedente (ventana de 08 días (noviembre-abril): obtenido a partir del producto PISCO - precipitación diaria como la sumatoria de precipitación diaria de los últimos ocho días, considerando solo los meses de noviembre a abril (1981-2020). Frecuencia de lluvias fuertes (Umbral = percentil de lluvia 95 (noviembre-abril)): obtenido a partir del producto PISCO - precipitación diaria como las veces que se sobrepasa el umbral definido como el percentil 95 considerando solo los meses de noviembre a abril (1981-2020).

Variables de cobertura y de suelo: Índice de vegetación de diferencia normalizada (NDVI): obtenido del producto MOD13A2.006 *Terra Vegetation Indices 16-Day Global* 1km, disponible en el catálogo de datos del GEE. Textura: obtenido de la USDA y disponible en el catálogo de GEE a 0, 10 y 30 cm. Cobertura de suelo (MODIS): obtenido del producto MCD12Q1.006 MODIS *Land Cover Type Yearly Global* 500m, disponible en el catálogo de datos del GEE. Geología: obtenido de INGEMMET.

### Datos de ocurrencia de remoción en masa producto de lluvias intensas

Los datos de ocurrencia de los eventos de remoción en masa producto de lluvias intensas son los proporcionados por el Servicio Nacional de Meteorología e Hidrología del Perú- SENAMHI y el Sistema de Información Nacional para la Respuesta y Rehabilitación - SIMPAD al año 2021. La Frecuencia de ocurrencia por departamento se muestra en la Tabla 1, donde se aprecia que Lima y Arequipa presentan el mayor registro de ocurrencia de dichos eventos.

**Tabla 1:** Frecuencia de ocurrencia de remoción en masa producto por fuertes lluvias hasta el 2021. Fuente. SENAMHI- SIMPAD.

| Departamento | Movimientos | Departamento  | Movimientos |
|--------------|-------------|---------------|-------------|
| Lima         | 545         | Pasco         | 15          |
| Arequipa     | 111         | Huancavelica  | 14          |
| Ayacucho     | 80          | Ica           | 12          |
| Cajamarca    | 80          | Puno          | 11          |
| La Libertad  | 44          | Moquegua      | 10          |
| Cusco        | 43          | Piura         | 9           |
| Ancash       | 35          | San Martin    | 8           |
| Huánuco      | 33          | Tacna         | 6           |
| Apurímac     | 22          | Lambayeque    | 5           |
| Amazonas     | 20          | Ucayali       | 3           |
| Junín        | 18          | Madre De Dios | 1           |

## c) Metodología

### Modelo de Random Forest

Random Forest (RF) es una extensión del método del árbol de clasificación y regresión (CART) propuesto por Breiman en 2001. En comparación con el método CART tradicional que utiliza conjuntos de datos completos, el RF entrena cada CART individual en muestras de arranque del conjunto de datos total. En lugar de utilizar todas las funciones, la RF utiliza una selección aleatoria de funciones para dividir cada nodo. La mejor división se elige entre un subconjunto de variables de entrada seleccionado al azar en cada nodo. Luego, todos los CART se calculan al tamaño máximo sin podar y el resultado final es el promedio de las predicciones de esos árboles. Los árboles con aumento de gradiente generalmente funcionan mejor que la RF, pero es mejor para aplicarla en diferentes situaciones.

RF realiza una validación cruzada incorporada en paralelo al proceso de entrenamiento mediante el uso de las llamadas muestras fuera de la bolsa (OOB). Los datos de entrenamiento se muestrean con reemplazos durante el arranque, por lo que algunos datos de entrenamiento se “dejan fuera” de la muestra. Estas muestras "excluidas" constituyen las muestras OOB (Breiman, 2001; Svetnik et al., 2003). Una vez que se construye un árbol mediante una muestra de entrenamiento extraída al azar, las muestras OOB se pueden usar para probar su precisión. El error de aprendizaje se obtiene promediando el error de predicción (error cuadrático medio) de cada árbol individual con sus muestras OOB (EOOB), como se muestra en la siguiente ecuación:

$$E_{OOB} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_i)^2$$

Donde  $n$  es el número total de muestras OOB;  $y_i$  es la salida observada de muestras OOB;  $\bar{y}_i$  es la salida de RF que corresponde a los factores introducidos OOB.

La importancia de cada variable se puede evaluar en función de cuánto peor será la predicción si la variable se permuta al azar. Los errores de predicción de las muestras OOB de cada árbol (denominado  $E_{OOB1}$ ) se calculan durante el procedimiento de entrenamiento. Al mismo tiempo, cada variable de entrada en las muestras OOB se permuta una a la vez. El árbol también predice estos conjuntos de datos modificados (denominados  $E_{OOB2}$ ). Al final del procedimiento de entrenamiento, la importancia de cada variable se obtiene promediando la diferencia entre  $E_{OOB1}$  y  $E_{OOB2}$ , que luego se normaliza por la desviación estándar de todos los valores de importancia de cada variable.

En este estudio, se implementó el modelo de regresión de Random Forest con el paquete “Random Forest” basado en el lenguaje de programación R. En el procedimiento de entrenamiento de RF están presentes dos parámetros

configurables por el usuario. El primero representa el número de variables utilizadas por árbol (*mtry*) (Gislason et al., 2004). El otro es el número de árboles (*ntree*).

### Implementación del modelo

Las diferentes capas de datos de los factores explicativos se convirtieron al mismo tamaño de celda utilizando técnicas de interpolación a 90 metros de resolución espacial. Algunos investigadores consideraron innecesario realizar una validación cruzada separada para RF debido a su procedimiento de validación cruzada incorporado con muestras OOB (Touw et al., 2013; Boucekine et al., 2013). Preferimos utilizar un subconjunto separado del inventario de ocurrencias de movimientos en masa producto de las fuertes lluvias para comparar los resultados de RF con los de otros métodos y justificar el rendimiento del modelo de RF, a pesar de que este proceso es computacionalmente exigente. Los datos del inventario de ocurrencias de movimientos en masa producto de fuertes lluvias se dividieron en subconjuntos de 80% y 20% (principio de Pareto) para ser utilizados para calibración y validación, respectivamente. Las ubicaciones de calibración (302) se seleccionaron al azar y las ubicaciones de inundación restantes (765) se utilizaron para la validación. También se utilizaron dos conjuntos de datos independientes para validar la precisión del modelo de RF y el mapa de susceptibilidad a las ocurrencias de movimientos en masa producto de las fuertes lluvias. El conjunto de inventarios de ocurrencias corresponde a la recopilación de datos realizado por el Servicio Nacional de Meteorología e Hidrología del Perú – SENAMHI.

## 3. Resultados y discusión

### a) Análisis de sensibilidad de parámetros

El entrenamiento y la validación de modelos son fundamentales para la aplicación exitosa de los modelos de RF, *mtry* y *ntree* son los dos parámetros más importantes que afectan significativamente la precisión del modelo. Por tanto, la validación cruzada se basó en el modelo de RF con 26 factores para analizar la sensibilidad de los dos parámetros.

La Figura 4 muestra la verificación del *mtry* y *ntree* óptimos comparando la precisión del modelo con los datos de entrenamiento (80 % de los datos). Como se muestra en la Figura 4, la precisión aumenta con la cantidad de *mtry*, teniéndose un valor óptimo con 20, el valor de *ntree* no muestra una variación significativa, teniendo un valor óptimo con 2500, estos parámetros se utilizarán para el modelo RF.

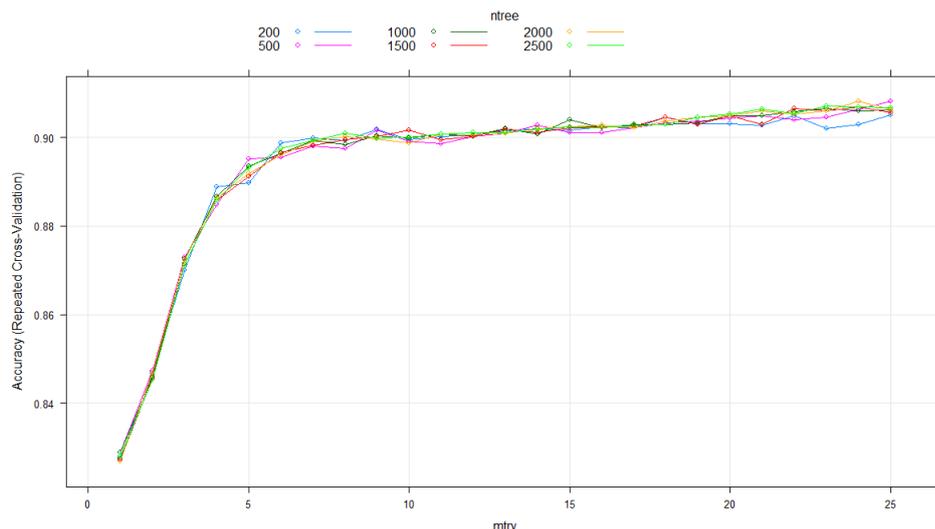
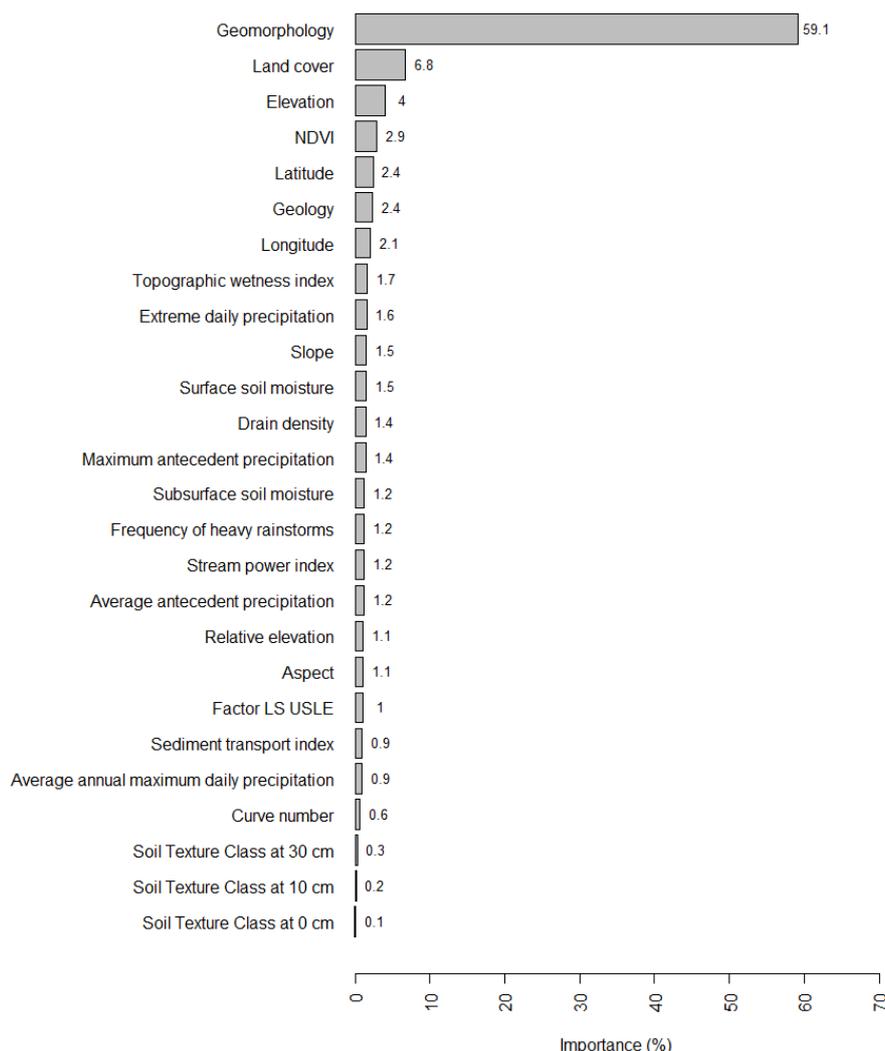


Figura 4. Precisión del modelo en función de los parámetros que se utilizaron en el modelo de RF (mtry y ntree son dos parámetros importantes del modelo de RF).

**b) Análisis explicativo de la importancia de los factores**

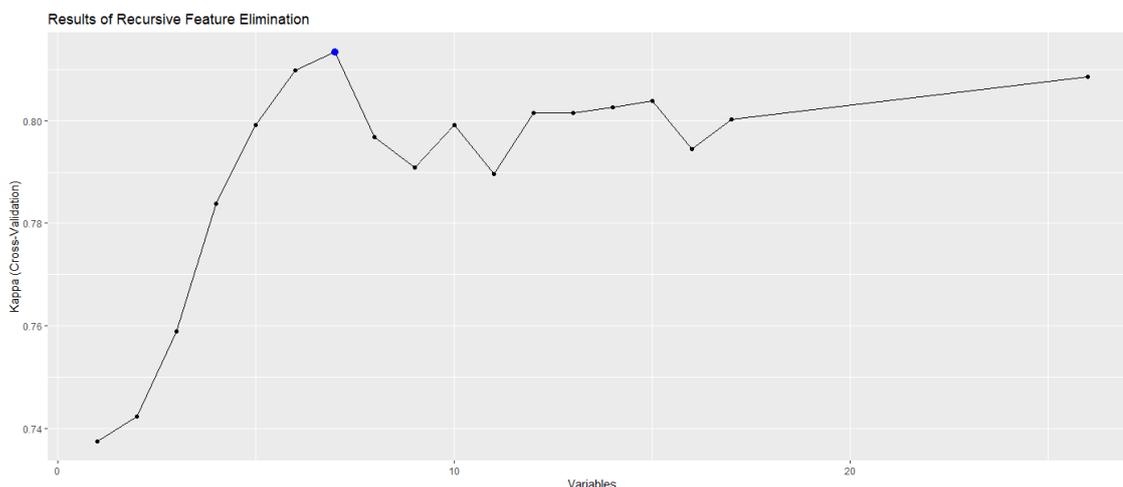
El modelo RF puede medir la importancia de una característica e incorporarla a los resultados finales. El modelo puede evaluar la contribución relativa de cada factor permutando los valores de cada característica en cada observación en el conjunto de datos y midiendo cuánto peor se vuelve el MSE después de la permutación. La Figura 5 muestra las contribuciones relativas de cada variable de entrada a la susceptibilidad a las inundaciones.

La Figura 5 muestra que estos factores explicativos se pueden dividir ampliamente en tres grupos de acuerdo con los resultados de la evaluación.



**Figura 5.** Importancia relativa de los factores de susceptibilidad a la ocurrencia de movimientos en masa producto de lluvias intensas.

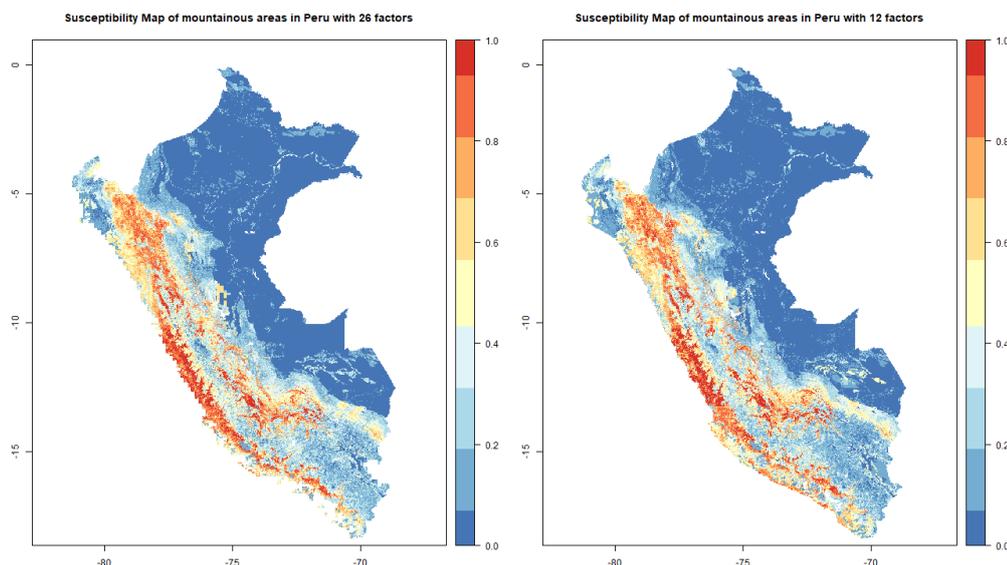
En el primer grupo, los factores corresponden a la estructura y cobertura del terreno, representados por los factores de geomorfología, cobertura de suelo, elevación, NDVI, geología, en el cual el factor geomorfológico es el más importante en el modelo con un aporte de 59.1%. El segundo grupo de factores más importante fueron los de ubicación y meteorológicos. Teniendo a los factores de latitud, longitud, precipitación extrema como los más resaltantes con un aporte de 2.4%, 2.1% y 1.6% respectivamente. El tercer grupo en importancia corresponden a las condiciones de humedad y características hidrológicas del terreno, se tiene a los factores humedad subsuperficial, densidad de drenaje, máxima precipitación antecedente, frecuencia de tormentas, como los más importantes con 1.5%, 1.4%, 1.4%, 1.2% de aporte en el modelo respectivamente.



**Figura 6.** Análisis de eliminación de características recursivas (RFE). El punto en azul indica la cantidad óptima de factores para el modelo

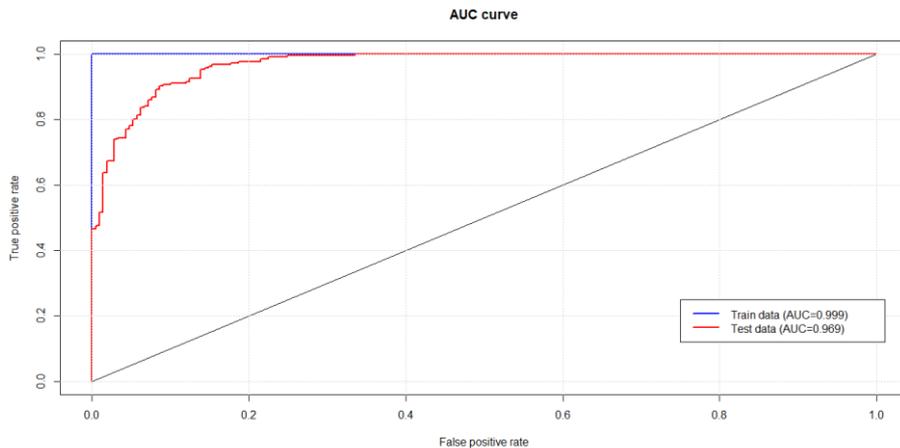
Luego del análisis de importancia de los factores, se realizó el análisis de eliminación de características recursivas (RFE), para identificar a cantidad de predictores necesarios en el modelo, este análisis muestra que la cantidad de factores adecuada es de siete, los cuales serían: geomorfología, cobertura, elevación, NDVI, latitud, geología, longitud. Sin embargo, se seleccionaron doce predictores en el cual se incluyan las variables meteorológicas que resultaron más importantes en el análisis de importancia de factores, teniendo finalmente los siguientes factores: geomorfología, elevación, cobertura, NDVI, latitud, longitud, densidad de drenaje, pendiente, TWI, precipitación extrema, máxima precipitación antecedente y geología.

El mapa de susceptibilidad a remoción en masa producto de fuertes lluvias utilizando el modelo de RF se muestra en la Figura 7. El mapa de susceptibilidad muestra la probabilidad de ocurrencia del evento, siendo 1 los valores en rojo los más probables o de mayor susceptibilidad, mientras los valores cercanos a cero en color azul lo menos susceptibles de ocurrir.

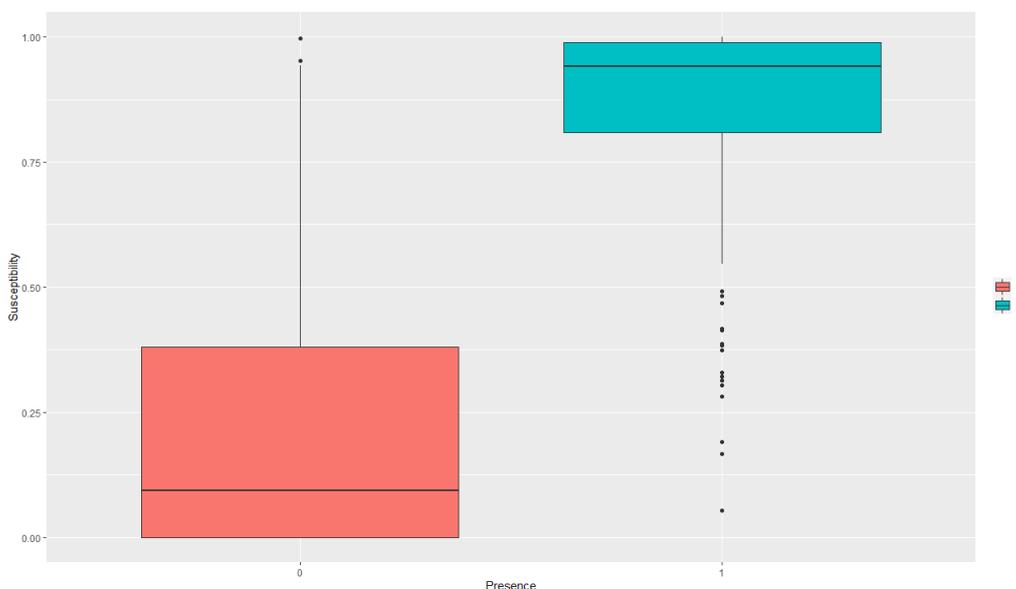


**Figura 7.** Mapa de susceptibilidad de ocurrencia de movimientos en masa producto de lluvias intensas. Izquierda: usando 26 predictores. Derecha: usando los 12 predictores seleccionados

Luego, se evaluó el mapa de susceptibilidad ocurrencia de movimientos en masa producto de lluvias intensas utilizando los dos conjuntos de datos (entrenamiento y prueba). El área bajo la curva (AUC), se ha aplicado ampliamente para examinar la eficiencia del mapeo de susceptibilidad, el método divide el mapa de susceptibilidad en categorías de áreas iguales y las clasifica de un valor mínimo a un máximo. La curva de tasa de predicción se crea trazando el porcentaje acumulado de áreas susceptibles desde la probabilidad más alta hasta las más bajas en el eje x, y el porcentaje acumulado de eventos de inundación en el eje y. Una curva más pronunciada representa que más eventos de inundación caen en las áreas más susceptibles con un área bajo la curva más alta (AUC). Un AUC de 0.5 es una clasificación por azar y un AUC de 1 es una clasificación perfecta. Los resultados de la evaluación con dos conjuntos de datos independientes se muestran en la Figura 8. Los resultados mostraron que el área bajo las curvas de tasa de predicción para los dos conjuntos de datos son 0.999 para los datos entrenamiento y 0.969 para los datos de prueba, lo que revela que el mapa de susceptibilidad a las ocurrencias de movimientos en masa producto de lluvias intensas logró una precisión satisfactoria.



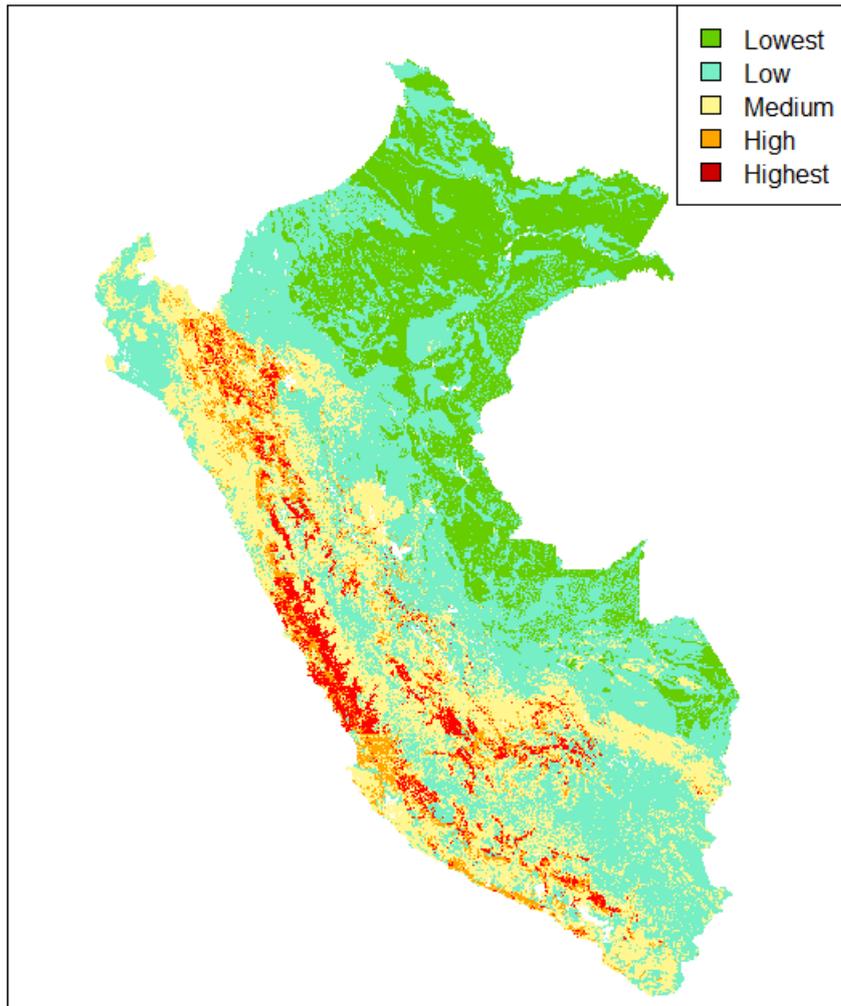
**Figura 8.** Área bajo la curva para los datos de entrenamiento (80%) y datos de prueba (20%).



**Figura 9.** Distribución de las ocurrencias de movimientos en masa producto de lluvias intensas según susceptibilidad obtenida.

Para clasificar el mapa de susceptibilidad generado se realizó el análisis de distribuciones de los datos de entrada, donde el valor 0 identifica zonas donde no se han registrado eventos de movimientos en masa producto de lluvias intensas, mientras el valor 1 representa las zonas donde si existen registros. La Figura 9 nos muestra que las ocurrencias de eventos se dan a partir del nivel de 0.75 de susceptibilidad, mientras las ausencias de eventos se dan por debajo de 0.3 de susceptibilidad. Por lo tanto, se plantea la siguiente clasificación en base al análisis de las distribuciones de los datos, teniendo así: de 0 - 0.30 muy bajo, 0.3 - 0.5 bajo, 0.5 – 0.75 medio, 0.75 – 0.9 alto y 0.9 – 1.0 muy alto.

### Susceptibility Map of mountainous areas in Peru



*Figura 10. Mapa de susceptibilidad de ocurrencia de movimientos en masa producto de lluvias intensas.*

El mapa de susceptibilidad a movimientos en masa producto de lluvias intensas utilizando el modelo de RF se muestra en la figura 10. El mapa de susceptibilidad se clasificó de acuerdo a la distribución de los valores de ocurrencia de los eventos según la susceptibilidad obtenida (ver Tabla 2). Un total del 19.54% de las áreas montañosas (cordillera andina) eran áreas propensas a movimientos en masa producto de lluvias intensas. El área de mayor susceptibilidad comprende el 3.76%, que debería priorizarse para la mitigación de movimientos en masa.

## 4. Conclusiones

Los movimientos en masa generados por lluvias intensas son uno de los fenómenos más catastróficos que ocurren en entornos montañosos, especialmente en Perú, particularmente en los años de evento del niño. Se necesita con urgencia un mapeo de susceptibilidad de dicho evento a escala nacional para que la gestión de desastres reduzca las graves pérdidas de vidas y propiedades. El objetivo de este estudio fue desarrollar un método confiable de mapeo de susceptibilidad a movimientos en masa producto de lluvias intensas para regiones de gran escala y con escasos datos. Con base a los registros de movimientos en masa que cuenta SENAMHI y el SIMPAD hasta el 2021, la susceptibilidad espacial a los movimientos en masa producto de lluvias intensas en las áreas montañosas de Perú se identificaron utilizando técnicas de RF. La precisión del mapa se valida mediante dos conjuntos de datos históricos independientes. Los puntos clave de este estudio se resumen a continuación:

- (1) El modelo de RF caracterizó con éxito la relación entre las variables predictoras y las ocurrencias de eventos de movimientos en masa producto de lluvias intensas. El RF mostró muy buena eficiencia del modelo tanto en el período de calibración como en el de validación. En este estudio, se plantearon inicialmente 26 factores, finalmente se establecieron en 12 los factores explicativos de acuerdo a la performance y al criterio de parsimonia.
- (2) Se evaluaron las contribuciones de los doce factores explicativos a través del modelo RF. Generalmente, los factores que corresponden a la estructura y cobertura del terreno fueron los más importantes, representados por los factores de geomorfología, cobertura de suelo, elevación, NDVI, geología, en el cual el factor geomorfológico es el más importante en el modelo con un aporte de 59.1%. El segundo grupo de factores más importante fueron los de ubicación y meteorológicos. Teniendo a los factores de latitud, longitud, precipitación extrema como los más resaltantes con un aporte de 2.4%, 2.1% y 1.6% respectivamente.
- (3) La susceptibilidad a los movimientos en masa producto de lluvias intensas se clasificó en cinco niveles según la distribución de las ocurrencias registradas (susceptibilidad más alta, alta y media) en Perú. El área de mayor susceptibilidad comprende el 3.76% del área de estudio y se debe priorizar para el manejo de

cobertura y suelo para la reducción de movimientos en masa. Dos conjuntos de datos de eventos históricos de movimientos en masa producto de lluvias intensas validaron la precisión del mapa, y la mayoría de los eventos históricos se ubicaron en áreas propensas a inundaciones, lo que reveló que el modelo de RF logró una precisión satisfactoria.

**Tabla 2:** Clasificación de susceptibilidad a movimientos en masa producto de lluvias intensas.

| ID | Nivel    | Porcentaje |
|----|----------|------------|
| 1  | Muy alto | 3.76%      |
| 2  | Alto     | 5.00%      |
| 3  | Medio    | 11.08%     |
| 4  | Bajo     | 14.02%     |
| 5  | Muy bajo | 66.14%     |

**Tabla 3:** Recopilación y sistematización de Estudios que han aplicado métodos de aprendizaje automático para la representación espacial de la susceptibilidad de deslizamiento de tierra e inundación en territorios de montañosos en Perú, América y el mundo.

| Estudio                             | Región                         | Método   | Predictores   | Datos de entrenamiento y prueba      | Principales resultados   |
|-------------------------------------|--------------------------------|--|---|--------------------------------------|--|
| (S. Villacorta, L. Fidel, 2012)     | Perú                           | modelo heurístico multivariado (Algebra de mapas con herramientas de Sistemas de Información Geográfica) | Elevación, geología, geomorfología, cobertura y uso de suelo, hidrogeología   | -                                    | La mayor cantidad de movimientos en masa se agrupa en las categorías de alta y muy alta susceptibilidad (86 % de los eventos inventariados). El mapa final generado, se elaboró a escala 1:2.000.000   |
| (Lai, Chen, Wang, Xu, & Yang, 2018) | Cuenca Dongjiang, Sur de China | Random Forest  | Precipitación máxima diaria, elevación, pendiente, factor de erodabilidad del suelo, coeficiente de escorrentía, NDVI.  | Entrenamiento = 67%<br>Prueba = 33 % | La elevación y la pendiente representaron los predictores más importantes, seguido por el NDVI y la precipitación máxima diaria.   |
| (S. Lee et al., 2017)               | Corea del Sur                  | Support vector machine (SVM) y artificial neural network (ANN)   | Pendiente, aspecto de la pendiente, curvatura, índice topográfico de humedad (TWI), índice de potencia de la corriente (SPI), factor de longitud de pendiente, altura estandarizada, profundidad del valle. | Entrenamiento = 70%<br>Prueba = 30%  | El mapa creado en este estudio mostró que la susceptibilidad a los deslizamientos de tierra disminuyó con el gradiente de la pendiente. Debido a que los deslizamientos de tierra son causados por diferencias en la gravedad debido a la forma del terreno, la ocurrencia se ve afectada por la estructura geográfica del área. |
| (de Oliveira et al., 2019)          | Cuenca Fão , Sur de Brasil     | Random forest (RF) y artificial neural network (ANN)   | Elevación, pendiente, TWI, curvatura perfil, curvatura en planta, factor LS, profundidad de valle, distancia vertical a una red de canales,   | Entrenamiento = 80%<br>Prueba = 20%  | El mayor aporte de este trabajo fue mostrar la importancia de analizar el patrón espacial de los mapas de susceptibilidad y la relevancia de los atributos del terreno. Este estudio demostró el gran potencial de las técnicas de aprendizaje automático de   |

|                        |   |   |   |                                     |   |
|------------------------|---|---|---|-------------------------------------|---|
|                        |   |   | distancia horizontal a las crestas, distancia horizontal a una red de canales.  |                                     | ANN y RF en el mapeo de susceptibilidad a deslizamientos de tierra, utilizando solo variables extraídas del SRTM DEM como atributos predictivos   |
| (Park & Kim, 2019)     | Zona montañosa de Woomyeon, Corea del Sur | Random forest (RF) y boosted regression tree (BRT)  | Elevación, pendiente, aspecto, curvatura de perfil, curvatura en planta, distancia a las corrientes, TWI, SPI, índice de transporte de sedimentos (STI), índice de rugosidad del terreno, características forestales (tipo, diámetro, edad y densidad de la madera) | Entrenamiento = 70%<br>Prueba = 30% | Los mapas de susceptibilidad de deslizamiento producidos en este estudio resultan útiles para los tomadores de decisiones, planificadores e ingenieros en la planificación de desastres para minimizar las pérdidas económicas y las víctimas. En un estudio futuro, la precisión de los mapas generados podría mejorarse seleccionando variables explicativas de deslizamientos más óptimas y resolviendo el problema del sobreajuste. |
| (Thi Ngo et al., 2021) | Irán                                      | Algoritmos recurrent neural network (RNN) y convolutional neural network (CNN),   | Elevación, pendiente, orientación, curvatura en planta, curvatura de perfil, distancia al río, distancia a la carretera, distancia a fallas, lluvia, geología, uso de la tierra.  | Entrenamiento = 70<br>Prueba = 30%  | En general, el algoritmo RNN mostró un rendimiento un 3,4% más alto en comparación con CNN. Los resultados indicaron que la pendiente es la variable más importante en casos de deslizamientos de tierra en Irán, seguida de la geología, el uso de la tierra, la distancia a una falla, la curvatura del plano, la precipitación, la distancia a la carretera, la altitud, la curvatura de perfil, la distancia al río y el aspecto    |
| (Orhan et al., 2020)   | Cuenca Arhavi-Kabisre, Turquía            | Artificial neural network (ANN), logistic regression (LR), support vector machine (SVM), random forest (RF), classification | Elevación, pendiente, aspecto, curvatura en planta, mapa de suelo, litología, cobertura y uso del suelo,  | Entrenamiento = 70%<br>Prueba = 30% | Se puede mencionar que todos los modelos mostraron rendimientos aceptables para el mapeo de susceptibilidad a deslizamientos. Sin embargo, el modelo ANN presentó la  |

|                     |                                |  |  |                                      |   |
|---------------------|--------------------------------|--|--|--------------------------------------|---|
|                     |                                | and regression tree (CART)   | distancia a las vías, distancia a los ríos, TWI, densidad de drenaje.  |                                      | capacidad de predicción más alta para el mapeo de susceptibilidad a deslizamientos de tierra en esta región. En conclusión, los resultados del estudio pueden ser útil para los tomadores de decisiones y la planificación del uso de la tierra en áreas propensas a deslizamientos de tierra.  |
| (Wang et al., 2018) | Región húmeda del sur de China | Random forest (RF)   | Elevación, pendiente, NDVI, distancia a una falla, capacidad de resistencia al corte, capacidad de almacenamiento de agua disponible, escorrentía, precipitación máxima en un día  | Entrenamiento = 67 %<br>Prueba = 33% | RF clasifica las variables de elevación, pendiente, precipitación máxima en un día y distancia a la falla como las más importantes. Este estudio muestra el potencial de proporcionar un nuevo método de determinación del peso para la evaluación de la susceptibilidad a los deslizamientos de tierra.  |
| (Zhao et al., 2018) | China                          | Random Forest (RF), artificial neural network (ANN) y support vector machine (SVM) | Latitud, Longitud, elevación, densidad de drenaje, precipitación diaria máxima promedio anual, precipitación diaria extrema, frecuencia de tormentas fuertes, humedad del suelo, número de curva, cobertura de vegetación, pendiente y elevación relativa. | Entrenamiento = 80%<br>Prueba = 20%  | El modelo de RF caracterizó con éxito la relación entre los factores explicativos y las frecuencias de ocurrencia de eventos de inundación y superó a los modelos ANN y SVM. El RF mostró la mayor eficiencia del modelo tanto en la calibración y periodos de validación. Además, se observó que la eficiencia del modelo de RF aumentó continuamente al considerar más factores explicativos. |

## 5. Referencias bibliográficas

- Bălțeanu, D., Micu, M., Jurchescu, M., Malet, J. P., Sima, M., Kucsicsa, G., ... Senzaconi, F. (2020). National-scale landslide susceptibility map of Romania in a European methodological framework. *Geomorphology*, 371. <https://doi.org/10.1016/j.geomorph.2020.107432>
- Bordoni, M., Galanti, Y., Bartelletti, C., Persichillo, M. G., Barsanti, M., Giannecchini, R., ... Meisina, C. (2020). The influence of the inventory on the determination of the rainfall-induced shallow landslides susceptibility using generalized additive models. *Catena*, 193(March), 104630. <https://doi.org/10.1016/j.catena.2020.104630>
- Catani, F., Lagomarsino, D., Segoni, S., & Tofani, V. (2013). Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues. *Natural Hazards and Earth System Sciences*, 13(11), 2815-2831. <https://doi.org/10.5194/nhess-13-2815-2013>
- de Oliveira, G. G., Ruiz, L. F. C., Guasselli, L. A., & Haetinger, C. (2019). Random forest and artificial neural networks in landslide susceptibility modeling: a case study of the Fão River Basin, Southern Brazil. *Natural Hazards*, 99(2), 1049-1073. <https://doi.org/10.1007/s11069-019-03795-x>
- Lai, C., Chen, X., Wang, Z., Xu, C. Y., & Yang, B. (2018). Rainfall-induced landslide susceptibility assessment using random forest weight at basin scale. *Hydrology Research*, 49(5), 1363-1378. <https://doi.org/10.2166/nh.2017.044>
- Lavado Casimiro, W. S., Labat, D., Guyot, J. L., & Ardoin-Bardin, S. (2011). Assessment of climate change impacts on the hydrology of the Peruvian Amazon-Andes basin. *Hydrological Processes*, 25(24), 3721-3734. <https://doi.org/10.1002/hyp.8097>
- Lee, J. H., Sameen, M. I., Pradhan, B., & Park, H. J. (2018). Modeling landslide susceptibility in data-scarce environments using optimized data mining and statistical methods. *Geomorphology*, 303, 284-298. <https://doi.org/10.1016/j.geomorph.2017.12.007>
- Lee, S., Lee, M. J., & Jung, H. S. (2017). Data mining approaches for landslide susceptibility mapping in Umyeonsan, Seoul, South Korea. *Applied Sciences (Switzerland)*, 7(7). <https://doi.org/10.3390/app7070683>
- Nhu, V. H., Mohammadi, A., Shahabi, H., Ahmad, B. Bin, Al-Ansari, N., Shirzadi, A., ... Nguyen, H. (2020). Landslide susceptibility mapping using machine learning algorithms and remote sensing data in a tropical environment. *International Journal of Environmental Research and Public Health*, 17(14), 1-23. <https://doi.org/10.3390/ijerph17144933>
- Orhan, O., Bilgilioglu, S. S., Kaya, Z., Ozcan, A. K., & Bilgilioglu, H. (2020). Assessing and mapping landslide susceptibility using different machine learning methods. *Geocarto International*, 0(0), 000. <https://doi.org/10.1080/10106049.2020.1837258>
- Park, S., & Kim, J. (2019). Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance. *Applied Sciences (Switzerland)*, 9(5). <https://doi.org/10.3390/app9050942>
- Pham, B. T., Shirzadi, A., Shahabi, H., Omidvar, E., Singh, S. K., Sahana, M., ... Lee, S. (2019). Landslide susceptibility assessment by novel hybrid machine learning algorithms. *Sustainability (Switzerland)*, 11(16), 1-25. <https://doi.org/10.3390/su11164386>

- S. Villacorta, L. Fidel, B. Z. (2012). Mapa de susceptibilidad por movimientos en masa del Perú. *Revista de la asociación geológica argentina*, 3(69), 393-399.
- Sobie, S. R. (2020). Future changes in precipitation-caused landslide frequency in British Columbia. *Climatic Change*, 162(2), 465-484. <https://doi.org/10.1007/s10584-020-02788-1>
- Sørensen, R., Zinko, U., & Seibert, J. (2006). On the calculation of the topographic wetness index: Evaluation of different methods based on field observations. *Hydrology and Earth System Sciences*, 10(1), 101-112. <https://doi.org/10.5194/hess-10-101-2006>
- Thi Ngo, P. T., Panahi, M., Khosravi, K., Ghorbanzadeh, O., Kariminejad, N., Cerda, A., & Lee, S. (2021). Evaluation of deep learning algorithms for national scale landslide susceptibility mapping of Iran. *Geoscience Frontiers*, 12(2), 505-519. <https://doi.org/10.1016/j.gsf.2020.06.013>
- Wang, P., Bai, X., Wu, X., Yu, H., Hao, Y., & Hu, B. X. (2018). GIS-based random forestweight for rainfall-induced landslide susceptibility assessment at a humid region in Southern China. *Water (Switzerland)*, 10(8). <https://doi.org/10.3390/w10081019>
- Zhao, G., Pang, B., Xu, Z., Yue, J., & Tu, T. (2018). Mapping flood susceptibility in mountainous areas on a national scale in China. *Science of the Total Environment*, 615, 1133-1142. <https://doi.org/10.1016/j.scitotenv.2017.10.037>

**Datos disponibles en:**

<https://figshare.com/account/home#/upload>