

Prediction of Landslide-Susceptible Areas Using Artificial Neural Network Mathematical Model in the Upper Pamplonita River Basin

Predicción de Zonas Susceptibles a Movimientos en Masa Aplicando Modelo Matemático Red Neuronal Artificial en la Cuenca Alta Del Río Pamplonita

Sarmiento, C. A.¹; Estrada, J. J.²; Cantillo, J. R.³

¹**Ingeniería Civil Camila Andrea Sarmiento Sanguino. Maestrante en Ordenamiento Territorial; Universidad Santo Tomas; e-mail: camila.sarmiento@ustabuca.edu.co; ORCID <https://orcid.org/0009-0008-3921-8556>**

²**Ingeniero Geólogo Javier Javier Estrada Romero; email: ingenierogeologojavier@gmail.com; ORCID <https://orcid.org/0009-0008-6493-5231>**

³**Geólogo Es. Janer Rafael Cantillo Romero. Maestría en Ingeniería Ambiental; email: jrcantillo93@gmail.com; ORCID <https://orcid.org/0000-0001-6925-447X>**

Universidad Santo Tomas

Carrera 18 No. 9 -27; Bucaramanga, Santander, Colombia

Tel: (+57 7) 6985858.

e-mail: camila.sarmiento@ustabuca.edu.co

Recibido: 06/12/2024 / Aceptado: 24/09/2024

Resumen

La investigación se centró en predecir zonas susceptibles a movimientos en masa en la cuenca alta del Río Pamplonita, utilizando Redes Neuronales Artificiales (RNA). Se aplicó el Análisis de Componentes Principales (PCA) para reducir la multicolinealidad entre variables, resultando en la selección de tres componentes principales (PC1, PC2, y PC3) que capturaron el 87% y 91% de la varianza acumulada, respectivamente. Las variables consideradas incluyeron factores geométricos y ambientales como pendiente (Slope), curvatura, densidad de drenajes, y otros índices de vegetación.

La RNA implementada con la librería Keras de TensorFlow, configurada con cinco capas ocultas (82, 5, 125, 126 y 58 neuronas) y una tasa de abandono de 0.32837 para evitar el sobreentrenamiento. Se utilizaron algoritmos de optimización "Adam" y la función de activación "sigmoidea", fijando la tasa de aprendizaje en 0.00012. El modelo se entrenó durante 500

épocas, alcanzando valor de AUC de 0.98, indicando alta precisión en la predicción de zonas de susceptibilidad.

La evaluación de susceptibilidad mostró 50% del área de estudio tiene alta o muy alta susceptibilidad a movimientos en masa, con áreas clasificadas como muy alta cubriendo 1173 ha y áreas de alta susceptibilidad abarcando 716 ha. Estas áreas se caracterizan por pendientes pronunciadas y patrones diversos de flujo de agua. Las zonas de susceptibilidad moderada, baja y muy baja cubren el 15%, 16% y 19% del área, respectivamente, indicando la necesidad de enfoques de mitigación específicos para reducir la vulnerabilidad en las áreas más propensas a deslizamientos.

Palabras clave: Modelamiento, Red Neuronal Artificial, deslizamiento de tierra, cuenca rio Pamplonita.

Abstract

The research focused on predicting areas susceptible to landslides in the upper Pamplonita River basin using Artificial Neural Networks (RNA). Principal Component Analysis (PCA) was applied to reduce multicollinearity among variables, resulting in the selection of three principal components (PC1, PC2, and PC3) that captured 87% and 91% of the cumulative variance, respectively. The variables considered included geometric and environmental factors such as slope, curvature, drainage density, and various vegetation indices.

The RNA was implemented using the Keras library of TensorFlow, configured with five hidden layers (82, 5, 125, 126, and 58 neurons) and a dropout rate of 0.32837 to prevent overfitting. Optimization algorithms like Adam and the sigmoid activation function were utilized, with a learning rate set at 0.00012. The model was trained over 500 epochs, achieving an AUC value of 0.98, indicating high precision in predicting susceptibility areas.

The susceptibility evaluation showed that 50% of the study area has high or very high susceptibility to landslides, with areas classified as very high covering approximately 1173 ha and high susceptibility areas covering 716 ha. The moderate, low and very low susceptibility zones cover 15%, 16% and 19% of the area, respectively. This distribution highlights the need for specific mitigation approaches to reduce vulnerability in the most landslide-prone areas.

Keywords: Modeling, Artificial Neural Network, landslide, Pamplonita River basin.

1. INTRODUCCIÓN

“Los movimientos de masa son eventos geológicos que pueden desencadenarse tanto por fenómenos naturales como por actividades humanas, destacándose la precipitación y las condiciones hidrometeorológicas como su activación (Ordoñez, 2019). A nivel mundial, Asia lidera la incidencia de estos eventos representando el 41% del total de casos; en Sudamérica, Brasil y Colombia encabezan la lista con un 37% y 32% de

deslizamientos respectivamente (Ordoñez, 2019).

En Colombia, la topografía compleja y los procesos antrópicos caracterizada por el relieve accidentado de la cordillera de los Andes crean un escenario propicio para los movimientos en masa. Departamentos como Norte de Santander han registrado significativas pérdidas de vidas debido a estos eventos en las últimas décadas (Corporación OSSO et al., 2024). La problemática en la cuenca del río Pamplonita es un ejemplo de cómo los

movimientos de masa afectan a la población local e infraestructuras. Históricamente, esta región ha presentado numerosos deslizamientos, lo cual se agravó durante el fenómeno de La Niña en 2010-2011, 2022 activando nuevos movimientos en masa causando daños a la infraestructura vial primaria, secundaria y terciaria, interrumpiendo el tráfico vehicular y peatonal en múltiples ocasiones. Además, estos deslizamientos han ocasionado millonarias pérdidas en el sector agropecuario, ganadero, comercial y económico de la región, al arrasar cultivos y viviendas (Corporación OSSO et al., 2024). Estos impactos se han visto exacerbados por el rápido crecimiento urbano en áreas propensas a deslizamientos, intensificando aún más las consecuencias de estos eventos (Aristizábal et al., 2019; Wang et al., 2019).

Para abordar esta problemática, estudios y modelamientos se han convertido en herramientas útiles para especificar regiones con alta susceptibilidad a deslizamientos de tierra (Ayalew et al., 2005; Conforti et al., 2014), indicando la distribución espacial de probabilidad de que ocurran en área determinada, basándose en factores geológicos, antrópico y ambientales (Dai et al., 2002; Tekin & Çan, 2022). El proceso de mapeo de susceptibilidad a deslizamientos de tierra (LS) incluye varios enfoques cualitativos y cuantitativos (Fell et al., 2008; Soeters & Van Westen, 1996). Los métodos cualitativos están enfocados en la opinión de expertos (Corominas et al., 2014) como el método heurístico, mientras que los métodos cuantitativos, se fundamentan en estadísticas y probabilidades, siendo la base de los

métodos más sofisticados empleados (Park et al., 2013; Zhang et al., 2019).

Entre los métodos cuantitativos más avanzados se encuentran los de aprendizaje automático, como RNA (Yilmaz, 2009), máquinas de vectores de soporte (Kalantar et al., 2019), Naïve Bayes (Yi et al., 2019), 2019) y árboles de decisión (Hong et al., 2018). Las RNA, técnicas inspiradas en el funcionamiento del cerebro humano o herramientas de propósito general flexible usadas para modelar problemas complejos no lineales, son funcionales para resolver clasificaciones y regresiones (Cantillo Romero et al., 2023; Tekin & Çan, 2022). Aunque estos métodos difieren en sus fundamentos, ha demostrado ser eficaces en diversas aplicaciones. Algunos investigadores destacan la capacidad de RNA para modelar relaciones no lineales y su adaptabilidad a conjuntos de datos complejos (Hinton et al., 2012).

La Cuenca del río Pamplonita una zona con características diferenciadas, geoformas altamente susceptibles a movimientos en masa debido a la topografía escarpada y montañosa, pendientes predominantes, relieve colinado y procesos de meteorización de suelos residuales poco cohesivos. La presente investigación tiene como objetivo predecir las zonas susceptibles a deslizamientos de tierra en la cuenca del río Pamplonita, Norte de Santander (Colombia) aplicando RNA, generando una capacidad para predecir futuros deslizamientos con alta precisión.

2. LOCALIZACIÓN

La cuenca alta del río Pamplonita se localiza en la parte sur del departamento de Norte de Santander, en la cordillera Oriental. Esta cuenca se encuentra

ubicada entre los municipios de Pamplona

y Pamplonita (ver *Figura 1*).

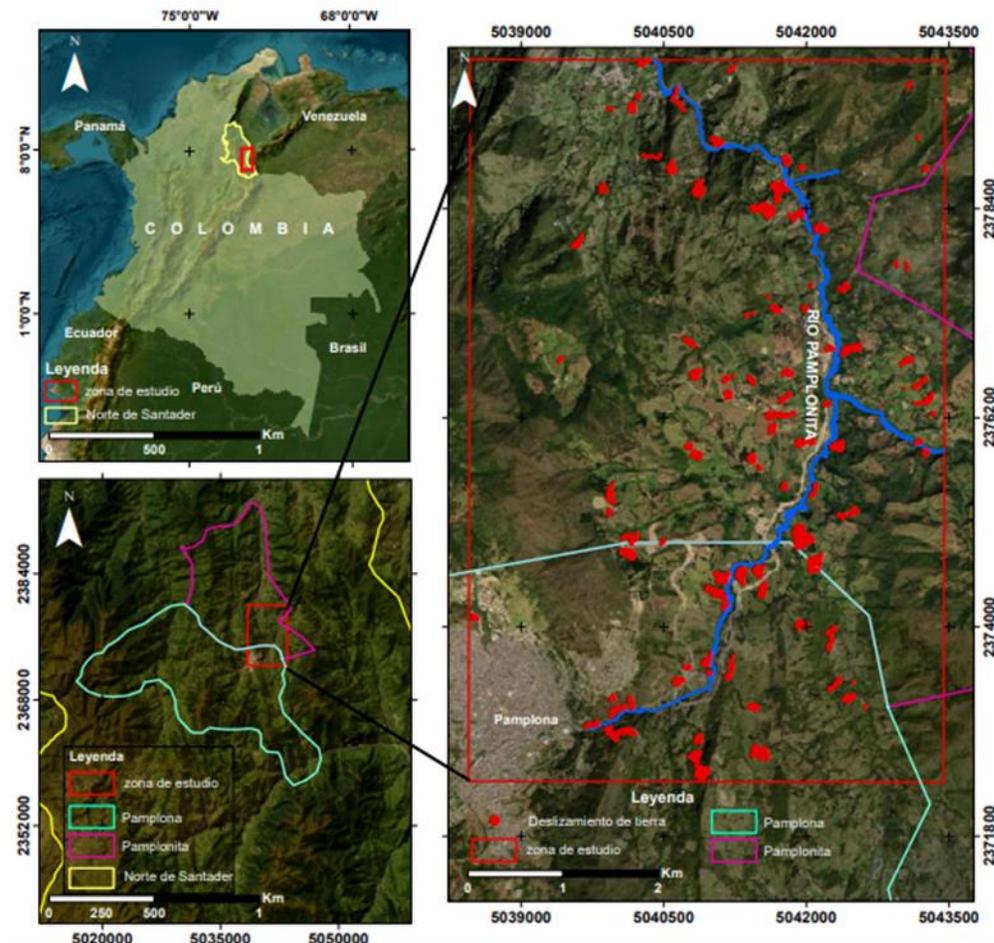


Figura 1 Localización del área de estudio e inventario de deslizamientos de tierra del área. **Fuente:** Autores.

3. METODOLOGIA

Para llevar a cabo esta investigación, se realizó una exhaustiva revisión histórica documental en diversas bases de datos académicas y científicas; posteriormente, fases de campo para reconocimiento detallado de la zona, inventario de movimiento en masa según los formatos del (Rodríguez *et al.*, 2017) y cartografía geológica superficial escala 1:25.000. El inventario de movimientos en masa se efectuó mediante la combinación de dos fuentes de información: trabajo de campo y análisis de imágenes satelitales a través de la fotografía aérea tomada por autores.

Este proceso involucró caracterización y documentación de 159 eventos de movimientos en masa.

Se seleccionaron 23 parámetros de acondicionamiento que se agrupan en 4 grandes grupos:

- Variables categóricas: Geología (UGS), Geomorfología (GMF) y cobertura vegetal (cob).
- Variables de proximidad: Densidad de drenajes.
- Variables geométricas: Slope, Slope2sen, LS, Aspect, Curvatura, Modelo Elevación Digital (DEM), Area Surface, Convergencia, Valley Dep, Vector Terrain Ruggedness, Terrain Ruggedness, Plan Curve,

Geomorphon, Multi Scale, Topographi, Relative Surface

- Índices de vegetación: NDVI y EVI.

Estas variables relacionadas con las características y mecanismos regionales de los deslizamientos de tierra. Las bases de datos generadas de los parámetros corresponden a los conjuntos de datos de variable categórica, proximidad, geométricas (continuas) e índices de vegetación y conjunto de datos del inventario de deslizamientos.

Los mapas fueron digitalizados en ArcGIS Pro-3.2 en formato ráster con resolución 2 m x 2 m y 2 m el tamaño del píxel para un total de 950000 pixeles. Para mejorar la precisión y/o eficiencia de la técnica de aprendizaje automático se discretizaron los datos (categóricos) en formato vectorial. Para discretizar los parámetros categóricos se crearon variables ficticias o codificando y clasificando las clases según el porcentaje relativo del área afectada por deslizamientos de tierra, utilizando el método de cálculo de pesos de evidencias (WoE) conocido como método bivariado (Rodríguez *et al.*, 2017).

El DEM se obtuvo a través de una fotogrametría realizada por autores. A partir del DEM se generaron variables geométricas usando operaciones de filtro local o fórmulas matemáticas mediante el uso de la herramienta Spatial Analyst Tool del software ArcGis pro-3.2. Se evidenció alta correlación o multicolinealidad entre las variables de entrada a partir de métodos estadísticos (matriz de correlación y matriz de dispersión); posteriormente, se descorrelacionaron aplicando el método estadístico Análisis de Componentes Principales (PCA) (Huang *et al.*, 2022), el cual permitió identificar los principales espacios ortogonales que se pueden proyectar los

componentes de grupo de datos, evitando multicolinealidad y minimización de dimensionalidad sin perder la información que describe el sistema (Baeza & Corominas, 2001; Correa-Muñoz & Higido-Castro, 2017; Gwelo, 2019; Lei *et al.*, 2011). Para este estudio se eligió tres componentes principales (PC1) cuya varianza acumulada fue 87%. y (PC2 y PC3) cuya varianza acumulada fue 91%.

Se identificó las zonas de baja susceptibilidad relativa por un estudio realizado por Cantillo, 2024 en la cual aplico la metodología de modelo discriminante, donde se ubicó de manera aleatoria 2000 puntos correspondientes a no procesos de remoción en masa (no_PR) y proporción 1:1 conforme a la metodología propuesta por (Lee *et al.*, 2003). La relación equilibrada se implementó para evitar que la red neuronal prefiera por una clase específica durante el entrenamiento. En este contexto, los 2000 puntos de procesos de remoción en masa (PR) se etiquetaron como "1" (indicando alta susceptibilidad u ocurrencia de movimientos en masa), mientras que 2000 puntos de baja susceptibilidad se catalogaron como "0" (indicando la ausencia de movimientos en masa). Posteriormente, se procedió a unir todos los puntos correspondientes a alta susceptibilidad (PR) y baja susceptibilidad (No_Pr), para un total de 4000 puntos.

Estos puntos se utilizaron como máscara para extraer valores de los factores condicionantes en formato "dbf", consolidándose en una tabla de datos de entrenamiento como base para el desarrollo de las técnicas de aprendizaje automático. Además, se utilizó la librería Python (GDAL) (Warmerdam *et al.*, 2023) para la creación de ráster virtuales permitiendo combinar múltiples mosaicos

en único archivo, utilizándose 16 variables causantes (PC1, PC2, PC3, UGS, GMF, cob, densidad de drenajes, aspect, DEM, convergencia, vector ruggedness, geomorphon, topographi, relative surface, NDVI y EVI), permitiendo consolidar valores de píxeles que se emplearon en predicción de ocurrencia de procesos de remoción de masa.

Una vez establecidas los parámetros de entradas, puntos de PR y no_PR, se procedió a determinar los parámetros de entrenamiento de los algoritmos. Se usaron los 4000 puntos para entrenar y validar técnicas de aprendizaje automático. Una vez definidos los parámetros de entrenamiento se generó el código en Colab y se procedió a correr la red, registrándose los hiperparámetros permitiendo validar y conocer valores predictivos de los modelos (Remondo *et al.*, 2003). Para la validación de los resultados, se realizó la curva de características operativas del receptor (ROC) y se calculó el valor del área bajo la curva (AUC) para el modelo. Este método ha sido ampliamente utilizado para proporcionar predicciones del rendimiento del modelo (Beguiría, 2006; Hosmer & Lemeshow, 2002). El análisis ROC se considera un método interesante para la validación del desempeño de los modelos LS (Frattini *et al.*, 2010; Nandi & Shakoore, 2006).

Para realizar esta evaluación, se creó un conjunto de datos de validación utilizando muestras seleccionadas al azar de áreas identificadas como deslizamientos de tierra, junto con zonas catalogadas como seguras. Como alternativa, se optó por seleccionar aleatoriamente un porcentaje (15%) de muestras del conjunto de entrenamiento existente para evaluar el

desempeño de los mapas de susceptibilidad a deslizamientos de tierra. Esta elección se hizo en ausencia de suficientes datos de deslizamientos de tierra para realizar una validación independiente y asegurar la fiabilidad de los modelos.

Para el mapeo de susceptibilidad a deslizamientos se aplicaron pruebas basadas en histograma asociados a métodos de clasificación automatizados disponibles en ArcGIS Pro-3.2 (ESRI, 2018), para encontrar la clasificación que se adapta a la escala de la probabilidad calculada. La probabilidad obtenida de RNA se dividió por método de clasificación "intervalo geométrico" debido a que valores de los datos indicaron asimetría negativa; esta clasificación se adaptó a la asimetría positiva o negativa de los valores de los datos. Las probabilidades calculadas se dividieron en 5 clases discreta: muy baja, baja, media, alta y muy alta; bajo y muy bajo indican el área clasificada es zona estable, mientras que alto y muy alto indica zona inestable. La clasificación media podría clasificarse como zona amarilla donde no se puede definir claramente estable o inestable.

4. ANALISIS Y DISCUSIÓN DE RESULTADOS

Parámetros de Acondicionamiento de Deslizamientos de tierra (Figura 3, Figura 4, Figura 5, Figura 6, Figura 7, Figura 8, Figura 9).

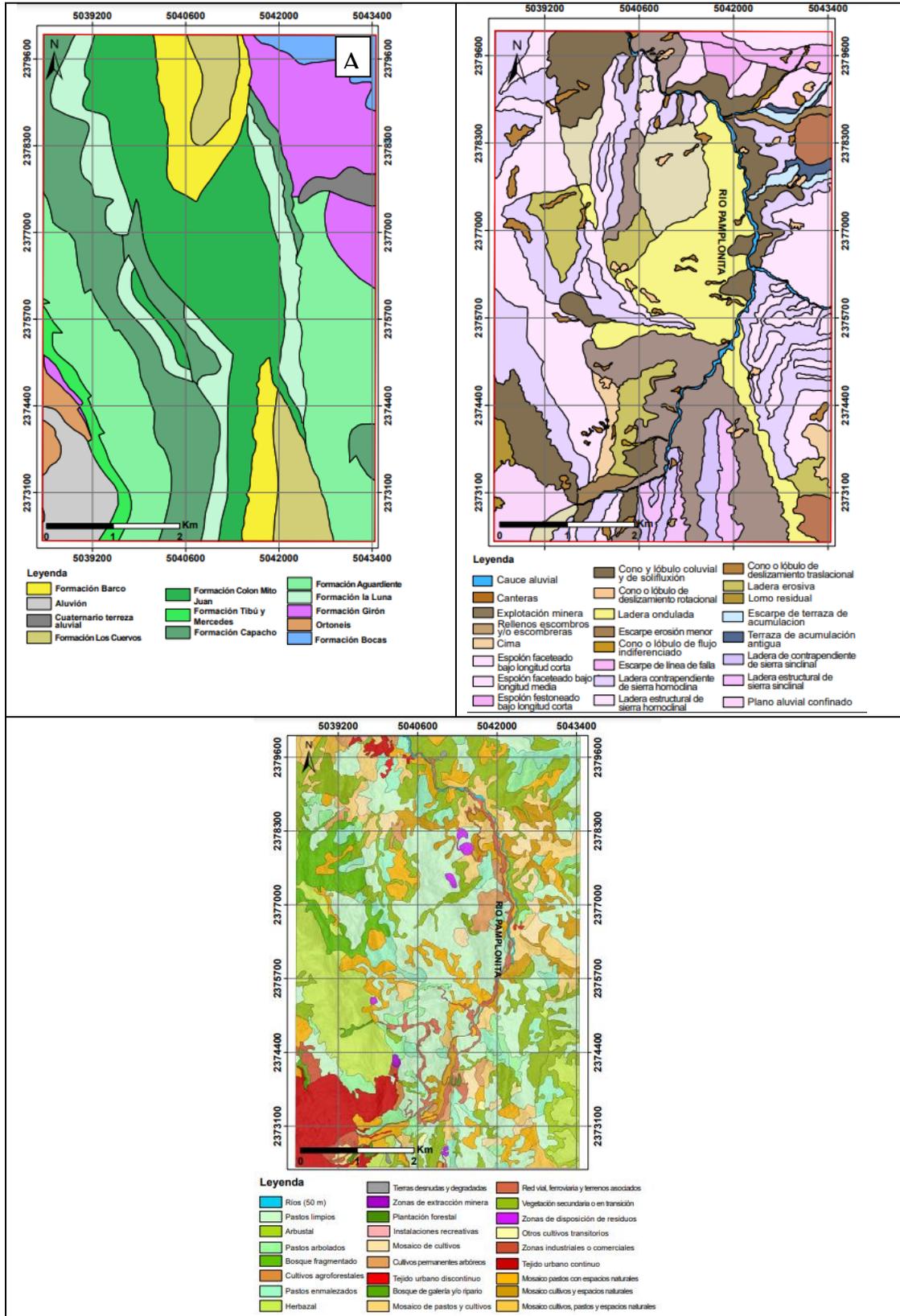


Figura 2 Variables categóricas, A) UGS, B) GMF, C) COB. Fuente: Autores

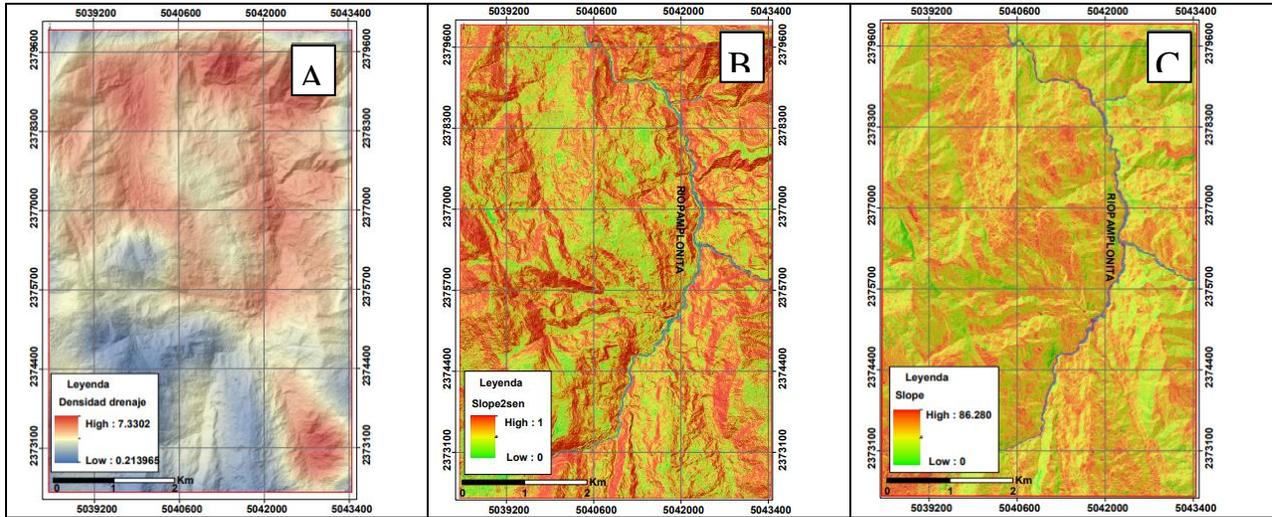


Figura 3 Variables de proximidad, variables geométricas. A) Densidad de drenajes, B) Slope, C) Slope2sen. Fuente: Autores

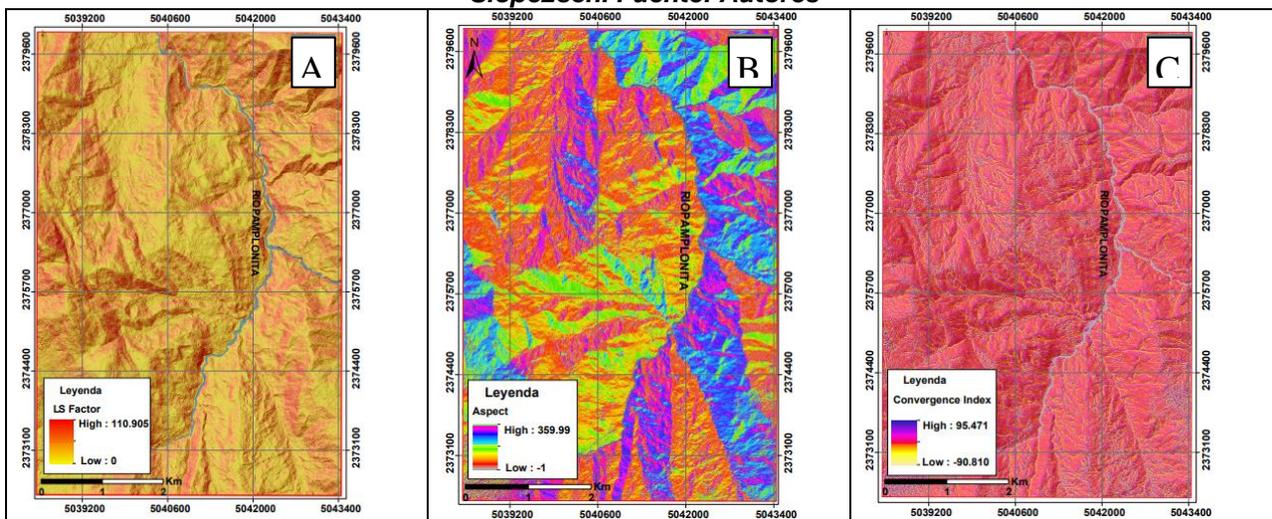


Figura 4 variables geométricas. A) LS Factor, B) Aspect, C) Convergence Index. Fuente: Autores

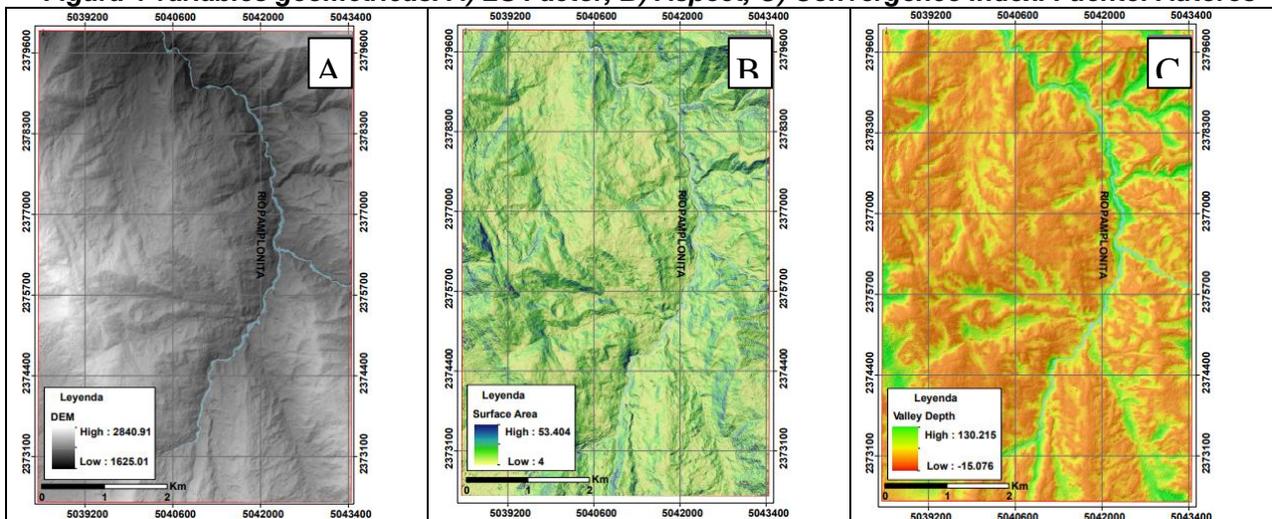


Figura 5 Variables geométricas A) DEM, B) Surface area, C) Valley Dep. Fuente: Autores

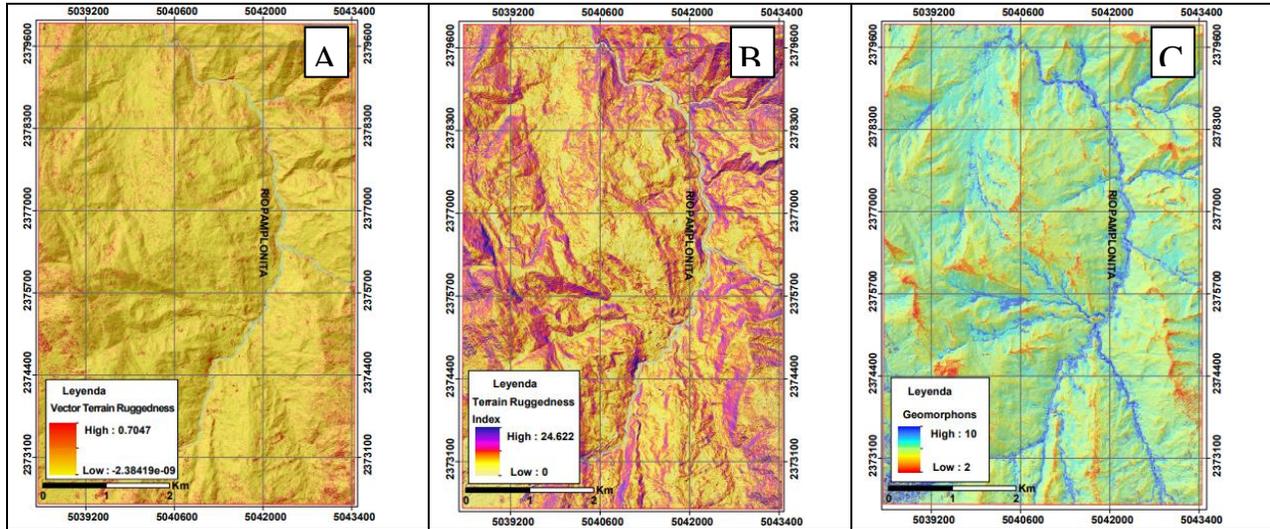


Figura 6 variables geométricas A) Vector Ter, B) Terrain Ruggedness, C) Geomorphons. Fuentes: Autores.

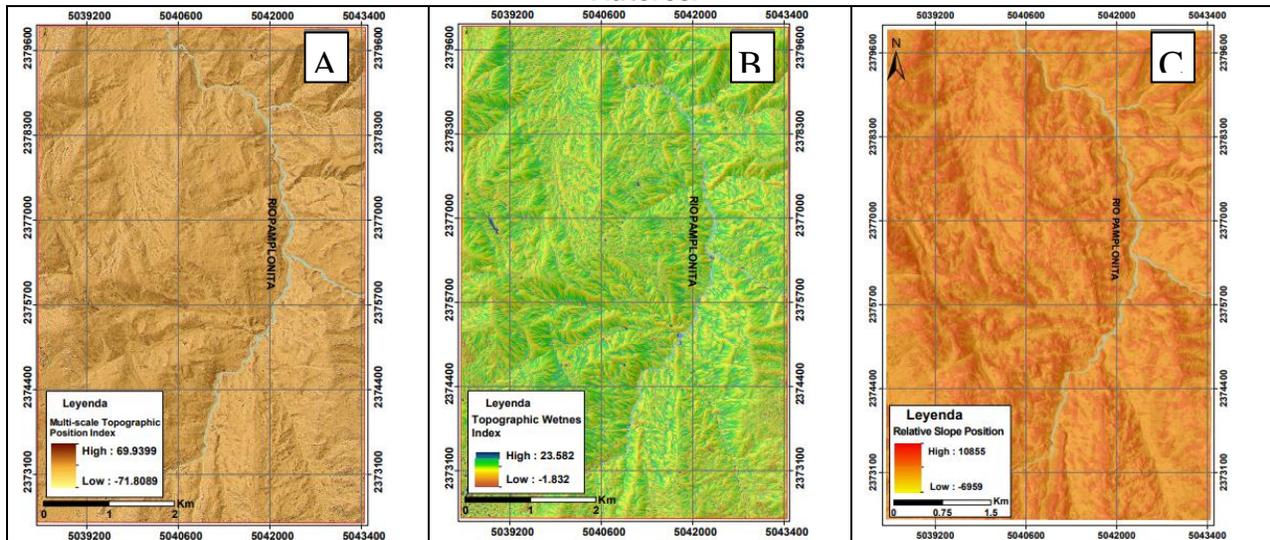


Figura 7 variables geométricas A) Multi Scale, B) Topographic, C) Relative Slope. Fuentes: Autores.

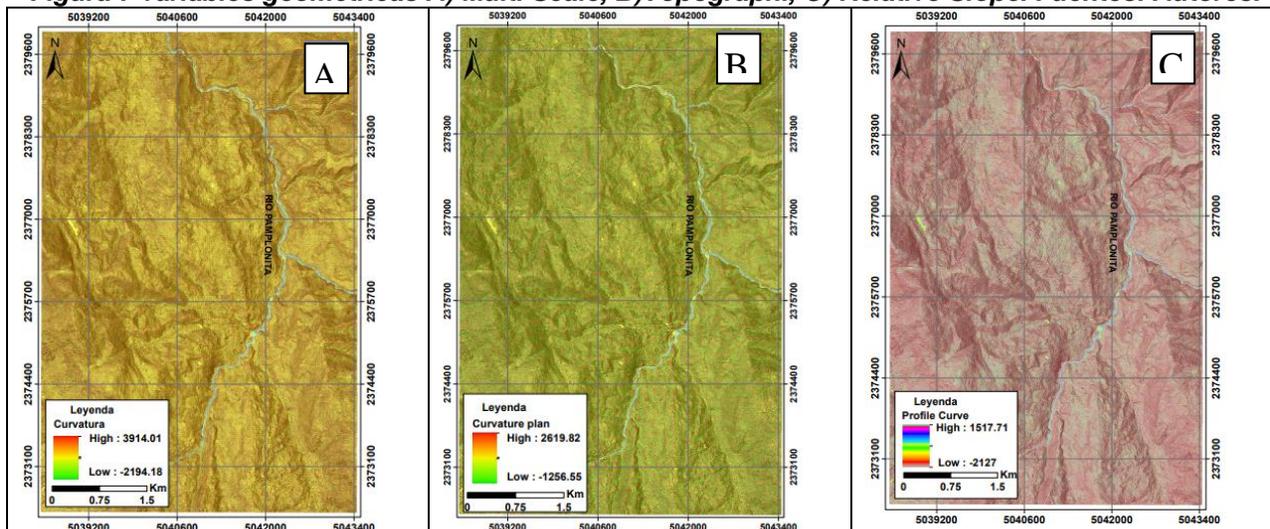


Figura 8 Variable geométrica A) Curvature, B) Plan curvature, C) Profile curvature. Fuente: Autores.

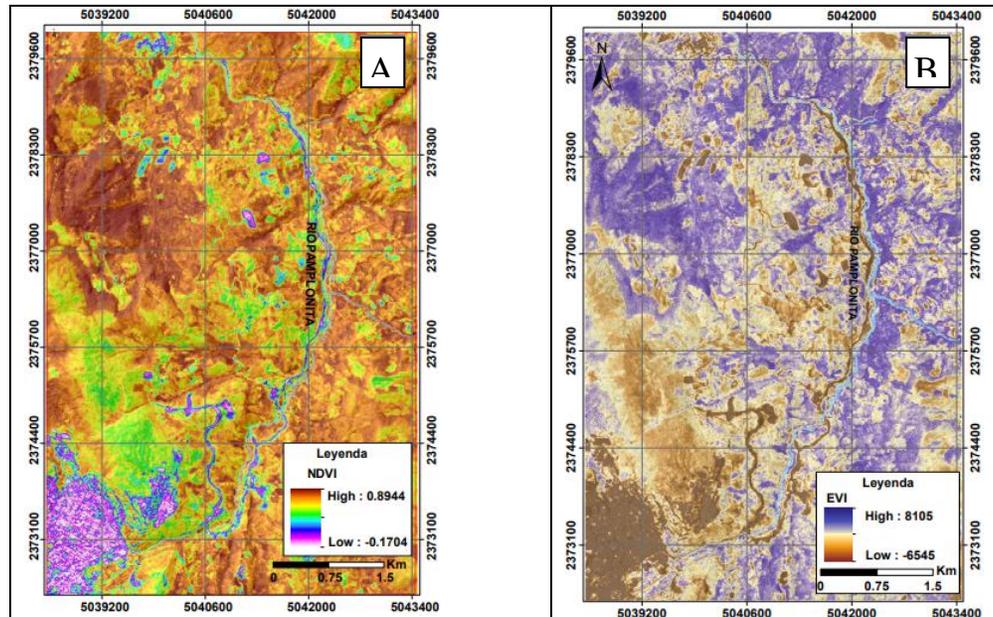


Figura 9 Variables Índices de Vegetación. A) NVDI, B) EVI. Fuente: Autores

4.2 Análisis Entre las Variables de Entrada

Los análisis estadísticos de las variables incluyó estandarización, dispersión y correlación, permitió identificar factores que afectan el rendimiento de los algoritmos computacionales para tomar medidas correctivas; además, se evaluó la correlación o multicolinealidad entre las variables de entrada, indicando que la multicolinealidad existe cuando se encuentran correlaciones significativas entre variables predictoras (Goyes-Peñafiel & Hernandez-Rojas, 2021), es decir, cuando los datos están relacionados linealmente entre sí y por lo tanto tienen coeficientes de correlación altos (Tabla 1). La Figura 10 se observó diversas correlaciones entre las variables estudiadas en la cuenca del Río Pamplonita. En lo cual según la clasificación de Mindrila & Balentyne (2017), las correlaciones con valores absolutos mayores a 0.57 se consideran fuertes correlaciones, como slope y slope2sen (0.92) sugiere que ambas

variables proporcionan información similar sobre la pendiente del terreno. de manera similar, variables terrain ruggedness y surface area presentó alta correlación (0.95), indicando redundancia significativa en la información capturada sobre la superficie del terreno.

Las Correlaciones Moderadas, se encuentran entre 0.5 y 0.7 según Mindrila & Balentyne (2017), muestran relaciones sustanciales, pero no perfectas, como factor LS con slope (0.83), slope2sen (0.77), y surface area (0.71). Esto sugiere que el factor LS está relacionado con la erosión del suelo, influenciado por la pendiente y otros aspectos topográficos del terreno. Además, se observó correlación moderada entre convergencia y multi scale (0.31), lo que sugiere una relación entre convergencia del flujo de agua y escala multiresolución.

Seguidamente las correlaciones débiles, con valores menores a 0.3, indican relaciones ligeras entre las variables, como curvatura y slope (0.02), y entre DEM y terrain ruggedness (-0.08), estas

correlaciones sugieren que estas variables no están fuertemente relacionadas. La correlación negativa fuerte entre plan curve y profile curve (-0.64) indicó que estas variables capturaron información opuesta o complementaria sobre la curvatura del terreno. De manera similar, la correlación negativa entre densidad drenaje y DEM (-0.55) sugiere que, a mayor altitud, la densidad de drenaje tiende a ser menor, indicando una distribución diferencial de la densidad de drenaje en función de la altitud.

Esta observación permitió identificar redundancias y relaciones complejas entre las variables, facilitando la toma de decisiones informadas para mejorar el rendimiento de los algoritmos computacionales. Posteriormente, se aplicó la técnica de reducción de dimensionalidad lineal "PCA" (Huang et al., 2022) para mitigar la multicolinealidad y simplificar la estructura de los datos.

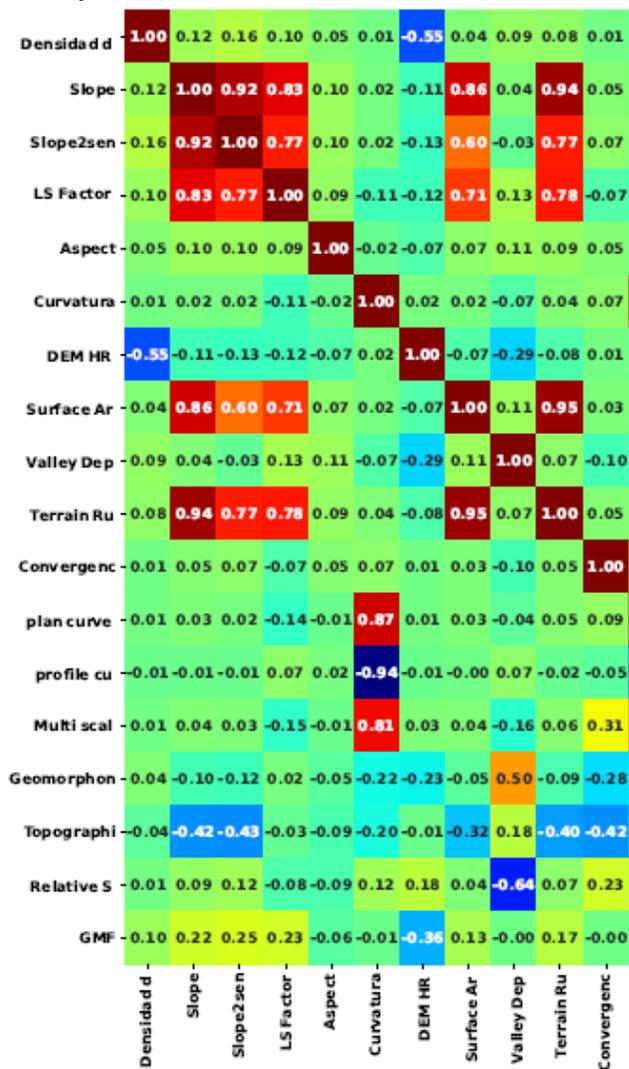


Figura 10 Matriz de correlación de Pearson. Fuente: Autores

Tabla 1 Interpretación de los valores del coeficiente de correlación de Pearson. **Fuente:** (Mindrila & Balentyne, 2017)

Valor de r	Fuerza de la correlación
$ r < 0.3$	Ninguna o débil
$0.3 < r < 0.5$	Débil
$0.5 < r < 0.7$	Moderada
$r > 0.7$	Fuerte

Se aplicó PCA a las variables con correlaciones moderadas y fuertes (>0.5). Se realizaron las siguientes agrupaciones: Primer PCA, se seleccionó variables como slope, slope2sen, LS factor, terrain ru y surface Ar, las cuales presentaban correlaciones moderadas o fuertes entre sí. Este grupo de variables se redujo a un componente principal, llamado PC1, que capturó el 88% de la varianza acumulada explicada en los datos. Segundo PCA a otro conjunto de variables que tenía una correlación moderada y/o fuerte entre sí, como curvature, plan curve, profile curve, valley depth y multi scale. Para estas variables, se seleccionaron dos componentes principales, PC2 y PC3, que capturaron el 91% de la varianza acumulada explicada en los datos.

Para verificar la efectividad de la técnica PCA y asegurarse que cumplió con su objetivo de reducir la multicolinealidad, se examinó la matriz de correlación y matriz de dispersión de las variables resultantes. Este análisis confirmó la aplicación PCA fue exitosa en la reducción de la multicolinealidad, simplificando estructura de los datos y mejorando así la eficiencia e interpretabilidad del modelo al utilizar un

conjunto de variables menos correlacionadas e independientes.

Los coeficientes de correlación de Pearson máximos obtenidos fueron 0.5 para las variables resultante (Figura 11). Estos valores indicaron correlación débil a ninguna o débil (Tabla 1), estableciendo que la multicolinealidad ha sido eliminada. Los tres componentes principales corresponden a nuevos mapas predictores de las capas de entrada para el cálculo de probabilidades.

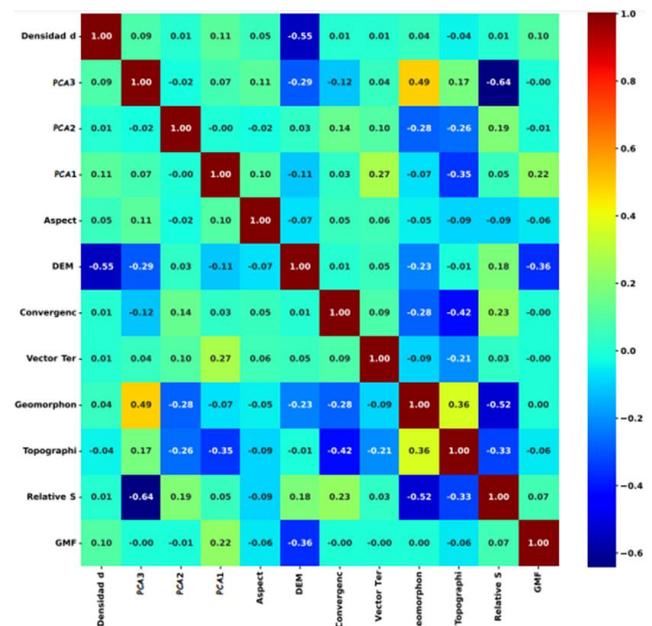


Figura 11 Matriz de correlación de las variables predictoras resultante. **Fuente:** Autores

La Figura 12 presentó la matriz de dispersión para las variables de entrada, se observó amplia dispersión de los puntos de datos a lo largo de los ejes, excepto Geomorphon y GMF, aunque presentó una

dispersión, mostró comportamiento lineal debido a la naturaleza de sus datos categóricos discretizados. Esta dispersión considerable indicó falta de patrón definido o agrupamiento entre las variables; Sugiriendo baja correlación entre las variables, respaldado por los valores consistentemente bajos del coeficiente de correlación de Pearson, todos por debajo de 0.57. Esta falta de relaciones fuertes entre

las variables pudo simplificar la tarea del algoritmo al separar los datos mediante superficie de decisión, debido a que las variables no están fuertemente correlacionadas y, por lo tanto, no presentan multicolinealidad significativa.

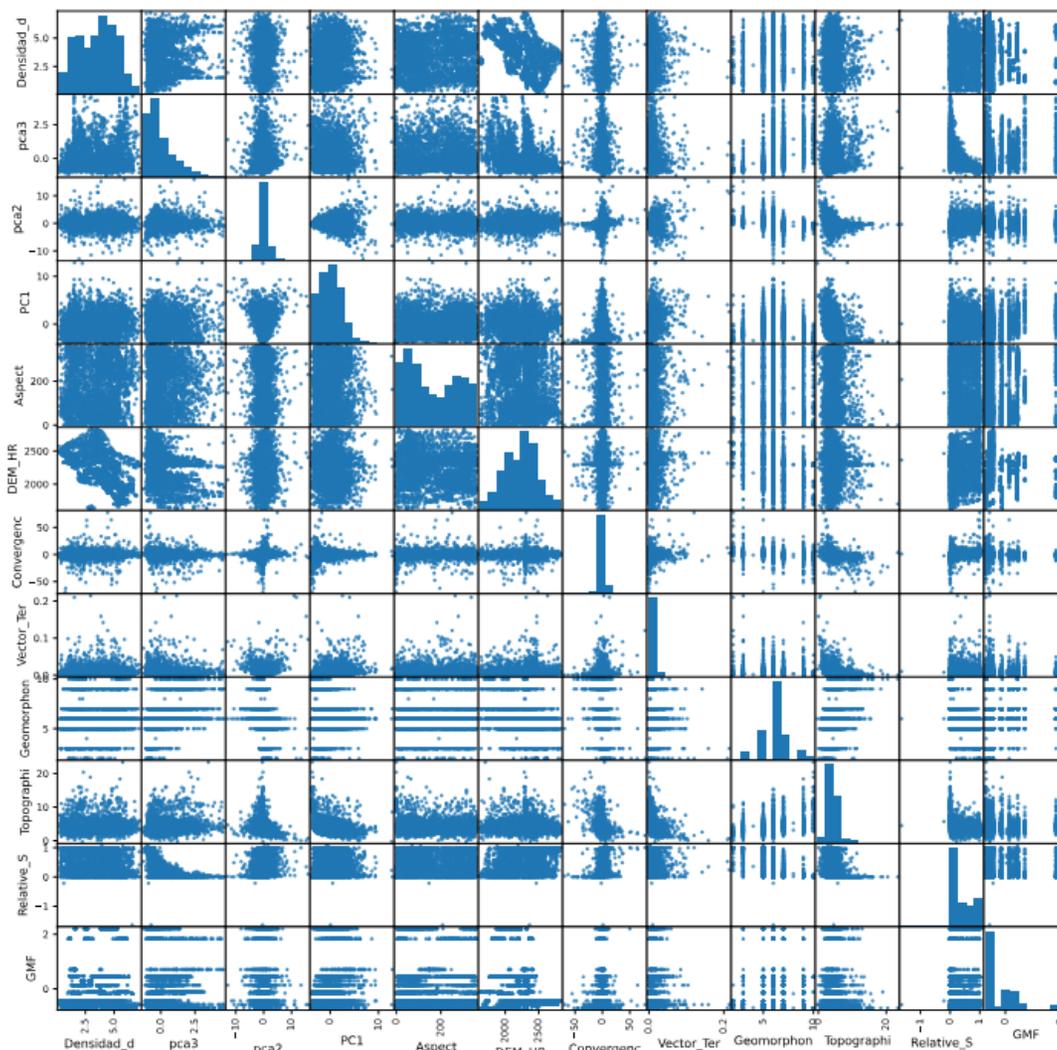


Figura 12 Matriz de dispersión para las variables entrante. En la diagonal se muestra el histograma para cada variable. **Fuente:** Autores

4.2 Diseño de la Red Neuronal Artificial

Red Neuronal Artificial (RNA)

El modelo RNA implementó librería Keras de TensorFlow, configurando los

hiperparámetros resultando una arquitectura compuesta por:

- Primera capa oculta: 82 unidades de neuronas.

- Segunda capa oculta: 5 unidades de neuronas.
- Tercera capa oculta: 125 unidades de neuronas.
- Cuarta capa oculta: 126 unidades de neuronas.
- Quinta capa oculta: 58 unidades de neuronas.

Además, se utilizó “dropout rate” del 0.32837 para prevenir sobreentrenamiento, desconectando aleatoriamente unidades de la red neuronal y eliminándolas temporalmente del modelo junto con todas sus conexiones entrantes y salientes (Mantilla & Goyes-Peñañiel, 2023).

La función de activación utilizada fue sigmoidea (logística), generó valores entre cero y uno para cada una de las entradas. La tasa de aprendizaje se fijó en 0.00012, un parámetro crucial que determinó el tamaño del paso en cada iteración. Cabe resaltar que no existe una regla general para seleccionar la tasa de aprendizaje adecuada, sino que debe determinarse experimentalmente para cada problema específico (Yesilnacar & Topal, 2005).

El proceso de entrenamiento RNA implicó el uso de algoritmos como “optimizador Adam” en conjunto con backpropagation, para resolver problemas de optimización en grandes conjuntos de datos, requiriéndose 500 épocas para marcar el final de la fase de aprendizaje de RNA.

4.3 Evaluación de Precisión

La evaluación del modelo a partir de valores del área bajo la curva (AUC) y características operativas del receptor presentó valor AUC de 0.98 para RNA según (Fawcett, 2006; Yilmaz, 2009) presento ajuste perfecto (Figura 13). Esto indica la estimación de la curva ROC de RNA reflejó con precisión los resultados y evaluación de la función de pérdida durante el entrenamiento y validación de RNA. El

algoritmo de entrenamiento mostró un buen desempeño utilizando doce variables de entrada, alcanzando precisión cercana al 100% para los datos de validación.

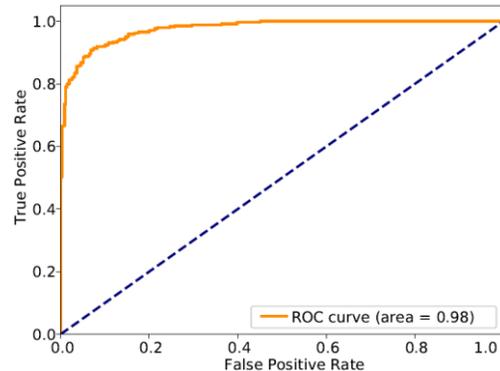


Figura 13 Evaluación de la capacidad de predicción: ROC CURVE. **Fuente:** Autores.

4.4 Evaluación de la Susceptibilidad

La Figura 14 se evidencia las zonas de susceptibilidad muy alta y alta que cubre 50% del área, principalmente en partes medias y altas. Morfológicamente, estas áreas se caracterizan por relieves moderado a muy altos, con pendientes abruptas y patrones de flujo de agua asociados a materiales con erosión moderada convirtiéndolas propensas a movimientos en masa. Estas susceptibilidades se evidenciaron en coberturas vegetal como herbáceas, ríos, mosaicos de pastos y cultivos, cultivos permanentes, zonas de disposición de escombros, pastos limpios, red vial y ferroviaria, y terrenos asociados. Además, predominan en formaciones geológicas como la Formación Colon, Formación Mito Juan, Formación Barco, Formación La Luna, Formación Girón, Formación Los Cuervos, Formación Capacho y Bocas, y subunidades geomorfológicas como conos y lóbulos coluviales y de solifluxión, laderas onduladas, laderas colinadas, laderas altamente denudadas, espolones festoneados bajos de longitud corta, espolones facetados bajos de longitud

corta, y laderas de contrapendiente de sierra sinclinal.

La susceptibilidad moderada cubre alrededor del 15% del área total y comparte similitudes con la susceptibilidad alta y muy alta. Estas áreas presentaron pendientes inclinadas a moderadamente abruptas, flujo de agua y áreas de escorrentía lenta, con materiales relativamente blandos y alta erosión. Morfológicamente, están vinculadas a subunidades geomorfológica como lomos residuales, conos y lóbulos coluviales y de solifluxión, laderas estructurales de sierra homoclinal, y escarpes de erosión menor. Estas áreas incluyen suelos residuales de la Formación Giron, la Formación Aguardiente, la Formación La Luna, Ortoneis, y las formaciones Tibú y Mercedes, con coberturas vegetales como tejido urbano discontinuo, zonas industriales o comerciales, vegetación secundaria o en transición, pastos enmalezados y mosaicos de cultivos.

La susceptibilidad baja y muy baja abarcó alrededor 35% del área total. Estas áreas se caracterizan por pendientes plana a suave y áreas de escorrentía lenta, asociadas con materiales blandos y relieves moderados. Morfológicamente, se relacionan con subunidades geomorfológicas como laderas estructurales de sierra homoclinal y planos aluviales confinados, predominando algunas unidades geológicas como la Formación Aguardiente y aluviones. La cobertura vegetal en estas áreas incluye pastos enmalezados, pastos limpios y tejido urbano.

La distribución de las áreas en las clases de susceptibilidad (Figura 14), indica que la zona presenta principalmente susceptibilidades muy altas con área de 1173 ha, alta susceptibilidad con area de

716 ha, susceptibilidad muy baja con área 727 ha, bajas con área 624 ha, moderadas con área de 557 ha.

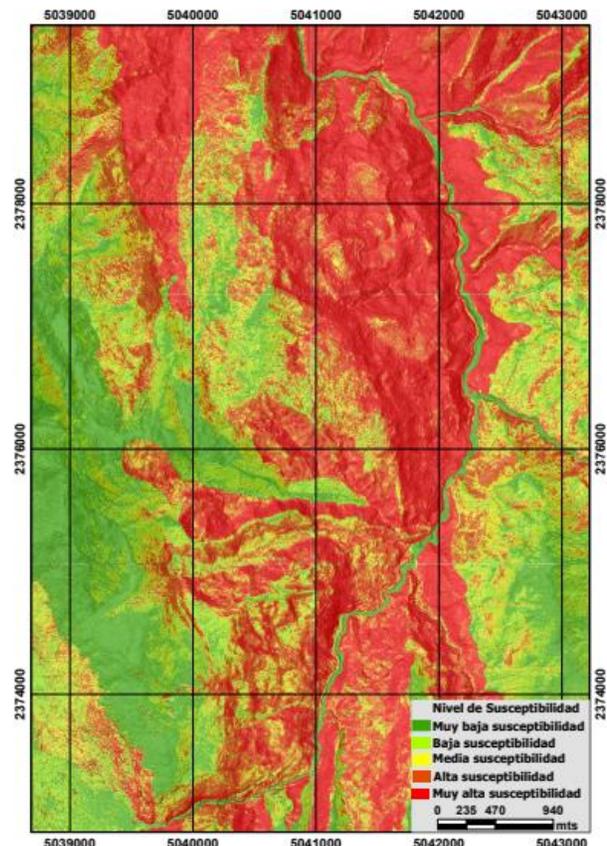


Figura 14 Mapa de susceptibilidad por RNA.

Fuente. Autores

El análisis del área de movimiento en masa en cada clase, también ilustrado en la Figura 7, revela que 85,76% de los movimientos en masa corresponden a zonas de susceptibilidad muy alta. Las áreas alta susceptibilidad concentran 10,13% de los movimientos en masa, mientras las zonas de susceptibilidad moderada representan 2,5%. Por su parte, las áreas de baja susceptibilidad acumulan 1,33% de los movimientos en masa, y zonas de susceptibilidad muy baja 0,28%.

5 CONCLUSIONES

Los análisis estadísticos revelaron que existía multicolinealidad significativa entre las variables predictoras, lo cual podría afectar negativamente el rendimiento de los Variables como Slope, Slope2sen, LS_Factor, Terrain_Ru y Surface_Ar, que presentaron correlaciones moderadas o fuertes entre sí (correlaciones superiores a 0.5), se agruparon en un solo componente principal (PC1), que capturó el 88% de la varianza acumulada explicada en los datos. Esta reducción de dimensiones mejoró la eficiencia e interpretabilidad del modelo. Segundo PCA a otro conjunto de variables (curvature, plan curve, profile curve, valley depth y multi scale) seleccionando dos componentes principales, PC2 y PC3, que capturaron 91% de la varianza acumulada explicada en los datos.

La zona de estudio mostró alta susceptibilidad a movimientos en masa, con el 50% del área clasificada como de susceptibilidad muy alta y alta. Estas áreas se caracterizan por pendientes pronunciadas y patrones de flujo de agua diversos, lo que aumenta la susceptibilidad. En términos de cobertura del área, la susceptibilidad muy alta cubre aproximadamente 1173 ha, mientras la susceptibilidad alta 716 ha, la susceptibilidad moderada abarca el 15% del área total (557 ha), susceptibilidad baja 16% (624 ha) y muy baja 19% (727 ha). Esta distribución resalta la necesidad de implementar medidas de mitigación específicas en las áreas más vulnerables para prevenir posibles desastres.

El modelo de redes neuronales artificiales (RNA) se diseñó utilizando la librería Keras de TensorFlow y consistió en una arquitectura con cinco capas ocultas, configuradas con 82, 5, 125, 126 y 58 neuronas respectivamente. La

algoritmos computacionales. La aplicación del Análisis de Componentes Principales (PCA) permitió reducir la multicolinealidad al consolidar variables correlacionadas en componentes principales. implementación de una tasa de abandono “dropout rate” de 0.32837 ayudó a prevenir sobreentrenamiento. La función de activación sigmoidea y una tasa de aprendizaje de 0.00012 fue seleccionada para optimizar el proceso de aprendizaje. El modelo fue entrenado durante 500 épocas utilizando el optimizador Adam, logró valor AUC (0.98), indica alta precisión en predicción de áreas de susceptibilidad a movimientos en masa. Esta alta precisión se logró mediante el uso de doce variables de entrada, confirmando que el diseño y configuración del modelo fueron efectivos para capturar relaciones complejas entre las variables predictoras y la susceptibilidad.

6 REFERENCIAS BIBLIOGRÁFICAS

- Aristizábal, E., López, S., Sánchez, O., Vásquez, M., Rincón, F., Ruiz-Vásquez, D., Restrepo, S., & Valencia, J. S. (2019). Evaluación de la amenaza por movimientos en masa detonados por lluvias para una región de los Andes colombianos estimando la probabilidad espacial, temporal, y magnitud. *Revista Boletín de Geología*, 41(3).
<https://doi.org/10.18273/revbol.v41n3-2019004>
- Ayalew, L., Yamagishi, H., Marui, H., & Kanno, T. (2005). Landslides in Sado Island of Japan: Part II. GIS-based susceptibility mapping with comparisons of results from two methods and verifications. *Engineering Geology*, 81(4).
<https://doi.org/10.1016/j.enggeo.2005.08.004>
- Baeza, C., & Corominas, J. (2001). Assessment of shallow landslide susceptibility by means of multivariate

- statistical techniques. *Earth Surface Processes and Landforms*, 26(12), 1251–1263.
<https://doi.org/10.1002/esp.263>
- Beguiría, S. (2006). Validation and evaluation of predictive models in hazard assessment and risk management. *Natural Hazards*, 37(3).
<https://doi.org/10.1007/s11069-005-5182-6>
- Cantillo Romero, J. R., Estrada Romero, J. J., & Henríquez Miranda, C. (2023). APLICACIÓN DE ALGORITMOS DE APRENDIZAJE AUTOMÁTICO EN GEOCIENCIA: REVISIÓN INTEGRAL Y DESAFÍO FUTURO. *REVISTA AMBIENTAL AGUA, AIRE Y SUELO*, 14(2 SE-Artículos), 9–18.
<https://doi.org/10.24054/raaas.v14i2.2783>
- Conforti, M., Pascale, S., Robustelli, G., & Sdao, F. (2014). Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy). *Catena*, 113, 236–250.
<https://doi.org/10.1016/j.catena.2013.08.006>
- Corominas, J., van Westen, C., Frattini, P., Cascini, L., Malet, J. P., Fotopoulou, S., Catani, F., Van Den Eeckhaut, M., Mavrouli, O., Agliardi, F., Pitilakis, K., Winter, M. G., Pastor, M., Ferlisi, S., Tofani, V., Hervás, J., & Smith, J. T. (2014). Recommendations for the quantitative analysis of landslide risk. *Bulletin of Engineering Geology and the Environment*, 73(2), 209–263.
<https://doi.org/10.1007/s10064-013-0538-8>
- Corporación OSSO, LA RED, & UNDRR. (2024). *Desinventar*.
<https://db.desinventar.org/>
- Correa-Muñoz, N. A., & Higidio-Castro, J. F. (2017). Determination of landslide susceptibility in linear infrastructure. Case: Aqueduct network in Palacé, Popayan (Colombia). *Ingeniería e Investigación*, 37(2), 17–24.
<https://doi.org/10.15446/ing.investig.v37n2.59654>
- Dai, F. C., Lee, C. F., & Ngai, Y. Y. (2002). Landslide risk assessment and management: An overview. *Engineering Geology*, 64(1), 65–87.
[https://doi.org/10.1016/S0013-7952\(01\)00093-X](https://doi.org/10.1016/S0013-7952(01)00093-X)
- ESRI. (2018). *ArcGis Desktop Help in ArcMap utility*. <https://www.esri.com/en-us/arcgis/products/arcgis-desktop/resources>
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8).
<https://doi.org/10.1016/j.patrec.2005.10.010>
- Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., & Savage, W. Z. (2008). Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. *Engineering Geology*, 102(3–4), 99–111.
<https://doi.org/10.1016/j.enggeo.2008.03.014>
- Frattini, P., Crosta, G., & Carrara, A. (2010). Techniques for evaluating the performance of landslide susceptibility models. *Engineering Geology*, 111(1–4), 62–72.
<https://doi.org/10.1016/j.enggeo.2009.12.004>
- Goyes-Peñafiel, P., & Hernandez-Rojas, A. (2021). Landslide susceptibility index based on the integration of logistic regression and weights of evidence: A case study in Popayan, Colombia. *Engineering Geology*, 280.
<https://doi.org/10.1016/j.enggeo.2020.105958>
- Gwelo, A. S. (2019). PRINCIPAL COMPONENTS TO OVERCOME MULTICOLLINEARITY PROBLEM. *Oradea Journal of Business and Economics*, 4(1).
<https://doi.org/10.47535/1991ojbe062>
- Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A. R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., & Kingsbury, B. (2012). Deep neural

- networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6), 82–97. <https://doi.org/10.1109/MSP.2012.2205597>
- Hong, H., Liu, J., Bui, D. T., Pradhan, B., Acharya, T. D., Pham, B. T., Zhu, A. X., Chen, W., & Ahmad, B. Bin. (2018). Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). *Catena*, 163, 399–413. <https://doi.org/10.1016/j.catena.2018.01.005>
- Hosmer, D. W., & Lemeshow, S. (2002). Applied Logistic Regression. In *Applied Logistic Regression* (Vol. 354). <https://doi.org/10.1002/0471722146>
- Huang, D., Jiang, F., Li, K., Tong, G., & Zhou, G. (2022). Scaled PCA: A New Approach to Dimension Reduction. *Management Science*, 68(3). <https://doi.org/10.1287/mnsc.2021.4020>
- Kalantar, B., Ueda, N., Lay, U. S., Al-Najjar, H. A. H., & Halin, A. A. (2019). Conditioning Factors Determination for Landslide Susceptibility Mapping Using Support Vector Machine Learning. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 9626–9629. <https://doi.org/10.1109/IGARSS.2019.898340>
- Lee, S., Ryu, J. H., Lee, M. J., & Won, J. S. (2003). Use of an artificial neural network for analysis of the susceptibility to landslides at Boun, Korea. *Environmental Geology*, 44(7), 820–833. <https://doi.org/10.1007/s00254-003-0825-y>
- Lei, T. C., Wan, S., Chou, T. Y., & Pai, H. C. (2011). The knowledge expression on debris flow potential analysis through PCA + LDA and rough sets theory: A case study of Chen-Yu-Lan watershed, Nantou, Taiwan. *Environmental Earth Sciences*, 63(5), 981–997. <https://doi.org/10.1007/s12665-010-0775-0>
- Mantilla, A., & Goyes-Peñafiel, P. (2023). *Predicción de la ocurrencia de depósitos minerales tipo pórfido usando técnicas de aprendizaje automático*.
- Mindrila, D., & Balentyne, P. (2017). *Scatterplots and Correlation*. http://www.westga.edu/academics/research/vrc/assets/docs/scatterplots_and_correlation_notes.pdf
- Nandi, A., & Shakoor, A. (2006). *Preparation of a landslide susceptibility map of Summit County, Ohio, USA, using numerical models*. 660, 1–11. <https://www.researchgate.net/publication/n/228392404>
- Ordoñez, J. (2019). Movimiento en masa por lluvias intensas en el Perú. *Repositorio Institucional - SENAMHI*, 1–103. <http://repositorio.senamhi.gob.pe/handle/20.500.12542/291#.XuuzTXE3a2o.mendeley>
- Park, S., Choi, C., Kim, B., & Kim, J. (2013). Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. *Environmental Earth Sciences*, 68(5), 1443–1464. <https://doi.org/10.1007/s12665-012-1842-5>
- Remondo, J., González-Díez, A., Díaz de Terán, J. R., & Cendrero, A. (2003). Landslide susceptibility models utilising spatial data analysis techniques. A case study from the lower Deba Valley, Guipúzcoa (Spain). *Natural Hazards*, 30(3). <https://doi.org/10.1023/B:NHAZ.0000007202.12543.3a>
- Rodríguez, E., Sandoval, J., Chaparro, J., Trejos, G., Medina, E., Ramírez, K. C., Castro, E., Castro, J. A., & Ruiz, G. L. (2017). Guía metodológica para la zonificación de amenaza por movimientos en masa escala 1:

- 25.000. In *Guía metodológica para la zonificación de amenaza por movimientos en masa escala 1: 25.000*. Servicio Geológico Colombiano.
<https://doi.org/10.32685/9789585978225>
- Soeters, R., & Van Westen, C. J. (1996). Slope instability recognition, analysis, and zonation. *Special Report - National Research Council, Transportation Research Board*, 247, 129–177.
https://ezproxy2.utwente.nl/login?url=https://webapps.itc.utwente.nl/library/1996/general/vanwesten_slo.pdf
- Tekin, S., & Çan, T. (2022). Slide type landslide susceptibility assessment of the Büyük Menderes watershed using artificial neural network method. *Environmental Science and Pollution Research*, 29(31), 47174–47188.
<https://doi.org/10.1007/s11356-022-19248-1>
- Wang, Y., Fang, Z., & Hong, H. (2019). Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. *Science of the Total Environment*, 666.
<https://doi.org/10.1016/j.scitotenv.2019.02.263>
- Warmerdam, F., Rouault, E., & Others. (2023, January 8). *gdalwarp — GDAL documentation*.
<https://doi.org/https://doi.org/10.5281/zenodo.5884351>
- Yesilnacar, E., & Topal, T. (2005). Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). *Engineering Geology*, 79(3–4), 251–266.
<https://doi.org/10.1016/J.ENGCEO.2005.02.002>
- Yi, Y., Zhang, Z., Zhang, W., & Xu, C. (2019). Comparison of Different Machine Learning Models for Landslide Susceptibility Mapping. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 9318–9321.
<https://doi.org/10.1109/IGARSS.2019.8898208>
- Yilmaz, I. (2009). A case study from Koyulhisar (Sivas-Turkey) for landslide susceptibility mapping by artificial neural networks. *Bulletin of Engineering Geology and the Environment*, 68(3).
<https://doi.org/10.1007/s10064-009-0185-2>
- Zhang, H., Zhang, G., & Jia, Q. (2019). Integration of Analytical Hierarchy Process and Landslide Susceptibility Index Based Landslide Susceptibility Assessment of the Pearl River Delta Area, China. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(11), 4239–4251.
<https://doi.org/10.1109/JSTARS.2019.2938554>