

# Predicción de zonas susceptible a los procesos de remoción masa aplicando el modelo matemático red neuronal artificial en la localidad de Guatapurí y Chemesquemena, Colombia

## Prediction of areas susceptible to landslide processes by applying the mathematical model artificial neural network in the locality of Guatapuri and Chemesquemena, Colombia

Javier Estrada-Romero <sup>1a</sup>, Fabio Carrillo <sup>1b</sup>, Dino Carmelo Manco-Jaraba <sup>2</sup>, Janer Cantillo <sup>3</sup>

<sup>1</sup> Grupo de investigación Geológico Minero, Fundación Universitaria del Área Andina, Colombia. Orcid: 0009-0008-6493-5231 <sup>a</sup>, 0000-0001-6021-8632 <sup>b</sup>. Correo electrónico: [Ingenierogeologojavier@gmail.com](mailto:Ingenierogeologojavier@gmail.com) <sup>a</sup>, [acarrillo@areandina.edu.co](mailto:acarrillo@areandina.edu.co) <sup>b</sup>.

<sup>2</sup> Grupo de investigación Geológico Minero, Universidad de La Guajira, Colombia. Orcid: 0000-0001-8506-094X. Correo electrónico: [dinomancojaraba@gmail.com](mailto:dinomancojaraba@gmail.com)

<sup>3</sup> Esri Colombia, Colombia. Orcid: 0000-0001-6925-447X. Correo electrónico: [jrcantillo93@gmail.com](mailto:jrcantillo93@gmail.com)

Recibido: 31 marzo, 2024. Aceptado: 9 octubre, 2024. Versión final: 23 noviembre, 2024.

### Resumen

Esta investigación tiene como objetivo predecir zonas susceptibles a procesos de remoción en masa aplicando modelo matemático red neuronal artificial en la localidad de Guatapurí y Chemesquemena. Se delimitaron zonas susceptibles a fenómenos de remoción en masa a partir de la identificación de factores condicionantes ((1) Unidades Geológicas Superficiales, (2) Pendiente del terreno, (3) Cobertura vegetal, (4) Índice de Rugosidad de Terreno, (5) Geomorfología y (6) Acuenca)) y realización de mapas y procesamiento de imágenes satelitales (Landsat) aplicando modelos matemáticos de tipo redes neuronales artificiales. La evaluación de susceptibilidad destacó una distribución desigual en Guatapurí y Chemesquemena. Las zonas de "muy alta" susceptibilidad (43% del área) se caracterizaron por presentar pendientes pronunciadas, patrones distintos de flujo y relieves moderados a muy elevados. En contraste, las zonas catalogadas como de "muy baja" susceptibilidad (34% del área), presentan pendientes suaves a casi planas, con escorrentía lenta y materiales menos propensos a deslizamientos.

**Palabras clave:** Chemesquemena; Guatapurí; Predicción; Procesos de remoción en masa; Redes neuronales artificiales; Susceptibilidad.

ISSN en línea: 2145 – 8456, [CC BY-ND 4.0](https://creativecommons.org/licenses/by-nd/4.0/) 

Como citar: J. Estrada-Romero, F. Carrillo, D. C. Manco-Jaraba, J. Cantillo, "Predicción de zonas susceptible a los procesos de remoción masa aplicando el modelo matemático red neuronal artificial en la localidad de Guatapurí y Chemesquemena, Colombia," *Rev. UIS Ing.*, vol. 23, no. 4, pp. 69-84, 2024, doi: <https://doi.org/10.18273/revuin.v23n4-2024006>

## Abstract

The objective of this research is to predict areas susceptible to landslide processes by applying an artificial neural network mathematical model in the locality of Guatapurí and Chemesquemena. Areas susceptible to mass landslide phenomena were delimited based on the identification of conditioning factors ((1) Surface Geological Units, (2) Terrain Slope, (3) Vegetation Cover, (4) Terrain Roughness Index, (5) Geomorphology and (6) Watershed) and mapping and processing of satellite images (Landsat) by applying mathematical models of artificial neural networks type. The susceptibility assessment highlighted an uneven distribution in Guatapurí and Chemesquemena. The "very high" susceptibility zones (43% of the area) were characterized by steep slopes, distinct flow patterns and moderate to very high relief. In contrast, the zones classified as "very low" susceptibility (34% of the area) have gentle to almost flat slopes, with slow runoff and materials less prone to landslides.

**Keywords:** Artificial neural networks; Chemesquemena; Guatapurí; Mass removal processes; Prediction; Susceptibility.

## 1. Introducción

Los deslizamientos de tierra y movimientos en masa representan un riesgo significativo para las comunidades, desencadenando pérdidas humanas, daños a infraestructura e interrupción de servicios públicos [1]. La falta de control por parte de entes gubernamentales y el rápido crecimiento desmedido e incontrolado por la población urbana en las últimas décadas ha llevado a la ocupación de áreas vulnerables a estos fenómenos [2].

Los mapas de susceptibilidad a movimientos en masa describen la distribución espacial de la probabilidad de que ocurran estos eventos en un área determinada, basándose en factores geológicos y ambientales [3], que pueden ser de naturaleza categórica, como unidades geológicas superficiales (UGS), geomorfología y cobertura vegetal, o de naturaleza continua, como datos raster derivados del modelo digital de elevación (MDE)[2], pendiente, índice de rugosidad del terreno (TRI), y acuena, entre otros.

Los métodos simples para mapear deslizamientos de tierra se basan en inventarios de deslizamientos previos y recopilando información (deslizamientos pasados) proporcionando indicaciones básicas de las áreas propensas. Los métodos más sofisticados emplean enfoques cuantitativos basados en estadísticas y probabilidad [4], [5]. Sin embargo, estos métodos a veces no aprovechan completamente la naturaleza categórica o continua de los factores condicionantes, debido a que requieren de conversión de datos continuos a discretos o viceversa, afectando valores inherentes de las variables. Actualmente, se están implementando el uso de métodos de aprendizaje automático, como máquinas de vectores de soporte [6], Naïve Bayes [7] y árboles de decisión [8]. Cada uno de estos métodos tiene sus propias suposiciones, ventajas y desventajas. Sin embargo, las redes neuronales artificiales (RNA), en particular el algoritmo de backpropagation, se ha destacado como un

enfoque ventajoso para la Landslide susceptibility mass (LSM.). Las RNA no dependen de suposiciones sobre la distribución estadística de los datos y pueden manejar diversos tipos de datos, incluso si son imprecisos, lo que las convierte en un enfoque prometedor para la evaluación de la susceptibilidad a deslizamientos de tierra [9].

Los corregimientos de Guatapurí y Chemesquemena corresponde zonas con características diferenciables, geofomas altamente susceptibles a movimientos en masa debido al relieve escarpado y montañoso, pendientes predominantes, rocas ígneas y metamórficas, relieve colinado y procesos de meteorización de suelos residuales poco cohesivos; permitiendo identificar áreas susceptibles a movimientos en masa. La presente investigación tiene como objetivo predecir zonas susceptibles a los procesos de remoción de remoción en masa aplicando modelo matemático red neuronal artificial en la localidad de Guatapurí y Chemesquemena, Colombia.

## 2. Localización

El área de estudio se encuentra ubicada al norte de los Andes, específicamente en los corregimientos de Guatapurí y Chemesquemena, departamento del Cesar, Colombia. Limitando al norte con el departamento de La Guajira y Magdalena, sur departamento de Santander, Bolívar y Norte de Santander, y este con la República Bolivariana de Venezuela (Figura 1).

## 3. Metodología

Para el desarrollo de esta investigación, inicialmente se realizó una revisión histórica documental en diferentes bases de datos académicas y científicas; posteriormente, fases de campo para el reconocimiento general de la zona, inventario de movimiento en masa según los formatos del [2], toma y descripción macroscópica de 10

muestras de roca según el Manual de Muestreo para exploración [10] y 46 estaciones de control. 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 de Sas-Planet del año 2023. Este proceso involucró caracterización y documentación de 66 eventos de movimientos en masa, de los cuales 11 fueron descargados del catálogo histórico sistema de información de movimientos en masa – SIMMA [11], que registra eventos en la zona de estudio desde el año 2009, los movimientos de masas fueron clasificados según Varnes, (1978): deslizamientos, caídas, flujos de detritos y erosión.

Los mapas fueron digitalizados mediante los softwares ArcGIS 10.8 y QGIS, escala 1:25.000, teniendo en cuenta la resolución espacial del modelo digital de elevación (DEM) de 12,5 metros. Para el cálculo de la susceptibilidad se escogieron 6 variables de acuerdo con

las condiciones geológicas y ambientales con una escala de 1:25000 (X1 unidades geológicas superficiales (UGS) según la guía del SGC [2], X2 pendiente, X3 geomorfología según Carvajal (2012), X4 índice de rugosidad del terreno (TRI), X5 cobertura vegetal según la metodología de CORINE Land Cover [14] y X6 acuenca que condicionan y se relacionan con la ocurrencia de remoción en masa según [2], [15]. Entre las variables categóricas (vectorial) se incluyen X1, X3, X5 Para las variables continuas (raster), se escogieron aquellas derivadas del MDE tomada de ALOS PALSAR, (2024) X2, X4, X6 y resolución 12.5 m. Por razones de eficiencia en la toma de decisiones, se optó por discretizar las variables categóricas en formato vectorial, debido a que las técnicas de aprendizaje asumen que todas las variables de entrada son continuas.

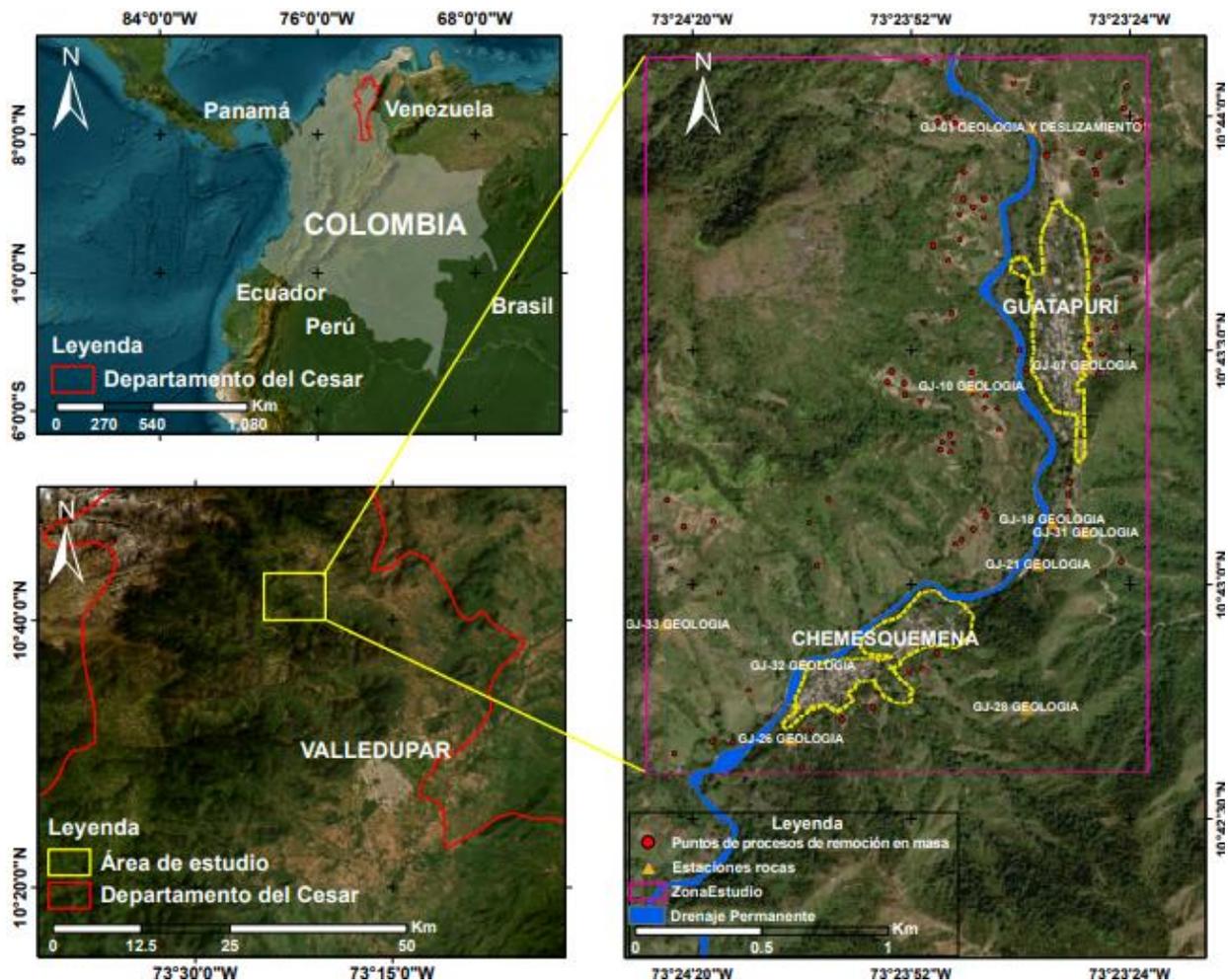


Figura. 1. Localización del área de estudio.

La discretización de parámetros categóricos a numéricos puede lograrse mediante la creación de variables ficticias o codificando y clasificando las clases según el porcentaje relativo del área afectada por deslizamientos de tierra (método de Weight of Evidence-WoE) [16], [17] dando como resultado las nuevas variables causantes continuas (raster) K1(LSI), K2 (pendiente), K3(TRI) y K4 (Acuena).

Seguidamente, se procedió a la unión de todos los puntos de alta susceptibilidad o PR para un total de 102 ocurrencias de mecanismos de procesos de remoción de masa. Para simplificar el análisis, se enfatizó en los procesos de remoción de masa específicos, como deslizamientos de tierra, erosión y flujos de detritos, excluyéndose dos (2) caídas de roca, para un total de 100 puntos de ocurrencias de procesos de remoción de masa. Seguidamente, se determinaron de manera heurística las zonas de baja susceptibilidad relativa. Estas zonas son fundamentales para ubicar aleatoriamente los 100 puntos de no procesos de remoción en masa (no\_PR), siguiendo una relación 1:1, conforme a la metodología propuesta por [18]. Esta relación equilibrada se implementa para garantizar que la red neuronal no muestre preferencia por una clase en particular durante el proceso de entrenamiento, donde los 100 puntos de PR los etiquetamos como "1" (indicando la alta susceptibilidad u ocurrencia de movimientos en masa) y los 100 puntos de baja susceptibilidad que catalogamos como "0" (indicando la ausencia de movimientos en masa). Estos puntos se utilizan como máscara para extraer los valores de los factores condicionantes en formato "dbf". Posteriormente, estos valores se consolidan en una única tabla de datos de entrenamiento. Esta **Tabla 1** resultante sirve como la base de datos principal para el desarrollo de la red neuronal.

Se utilizó la librería de Python (GDAL) [19], para la creación de ráster virtuales permitiendo combinar múltiples mosaicos en un único archivo. Además, se utilizaron 4 variables causantes (K1, K2, K3 y K4), permitiendo consolidar los valores de los píxeles que se emplearon en la predicción de la ocurrencia de procesos de remoción de masa en el área de estudio.

Una vez establecidas las entradas y los puntos de PR y no\_PR, se procedió a determinar los parámetros de entrenamiento de la red. Fueron creados 200 puntos para el entrenamiento y validación de los modelos de Redes Neuronales Artificiales. La creación y entrenamiento de estos modelos se realizó utilizando Python como lenguaje de programación, en el entorno de desarrollo integrado Google Colab Pro, utilizando GPU para mejorar el rendimiento. Los datos de entrenamiento fueron cargados y visualizados en Colab Pro utilizando

la librería Pandas [20]. Una vez definidos los parámetros de entrenamiento se genera el código en Colab y se procede a correr la red, registrándose el número de iteraciones "epoch", "Performance", R2. Seleccionada la red neuronal artificial. Así mismo, la validación de los resultados de los análisis, se realizó la curva de características operativas del receptor (ROC) y se calculó el valor del área bajo la curva (AUC) para cada modelo [21], [22]. La AUC se calcula a partir de la matriz de confusión del modelo para evaluar su capacidad de clasificación.

El análisis entre las variables de entrada permitió garantizar que los factores que pueden afectar el rendimiento de los algoritmos sean considerados y corregidos, analizar características estadísticas del conjunto de datos y estandarizar las variables con el uso de la librería Scikit-learn [23] en Python. El procedimiento se basó en escalar los valores para que tengan las propiedades de distribución normal estándar con  $\mu = 0$  y  $\sigma = 1$  donde  $\mu$  es el promedio y  $\sigma$  la desviación estándar [24]. De esta manera se realizaron análisis estadísticos de estas variables que incluyeron la estandarización, analizar la dispersión y correlación. Se analizó la correlación o multicolinealidad entre las variables de entrada. El modelo de redes neuronales artificiales se implementó utilizando la librería Keras de TensorFlow, configurando los hiperparámetros resultando una arquitectura compuesta por cinco capas ocultas, cada una con diez neuronas; la capa dropout proviene del sobre entrenamiento al desconectar de manera aleatoria unidades de la red neuronal, temporalmente eliminándolas del modelo junto con todas sus conexiones entrantes y salientes [25].

## 4. Resultados

### 4.1. Parámetros de acondicionamiento de deslizamientos

Las UGS son parámetros importantes que permitieron identificar zonas vulnerables a deslizamiento de tierra, de las cuales se identificaron 14 de diferentes orígenes, tipos y característica, destacando: Suelo Residual Cohesivo de la Diorita de la unidad del Batolito de Atanquez, Suelo Residual Cohesivo de la Granodiorita de la unidad del Batolito de Atanquez, Suelo Residual Cohesivo de la Cuarzomonzonita de la unidad del Batolito de Atanquez, Suelo Residual Cohesivo de la Aplita de la unidad del Batolito de Atanquez, Suelo Residual Cohesivo de la Basalto de la unidad de la Riolita de los Goleros (**Figura 2**). Los deslizamientos de tierra se ubicaron principalmente sobre los suelos residuales del batolito de Atanquez y Riolita de Golero.

El mapa geomorfológico (Figura 3) ilustra que la zona está principalmente influenciada y modelada por esfuerzos tectónicos con subunidades de tipo espolón moderado de longitud media, cerro estructural, espolón facetado y ladera estructural, y ambiente denudacional que se manifiesta en subunidades como la Colina residual

disectada. Además, la interacción con subunidades de ambiente fluvial (Barra longitudinal, Cauce aluvial, Montículos y Barra puntual) y ambiente antrópico (subunidades de tipo superficies de explanación) contribuyen a procesos asociados a meteorización, incrementando su propensión a la inestabilidad.

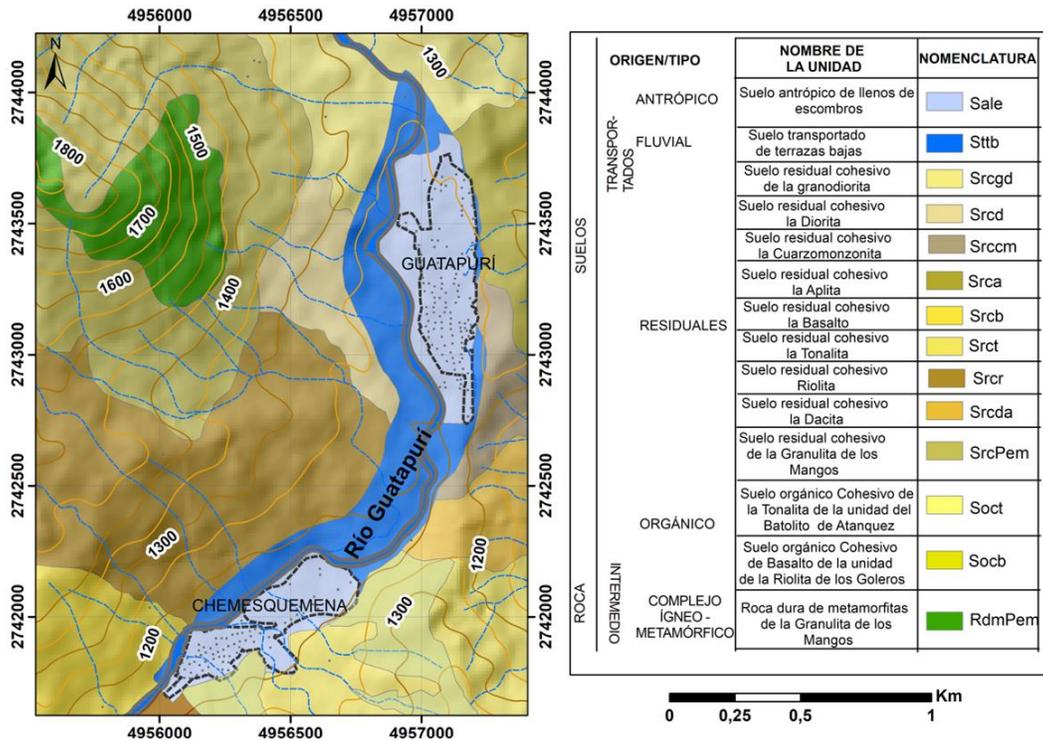


Figura 2. Mapa de unidad geológica superficial.

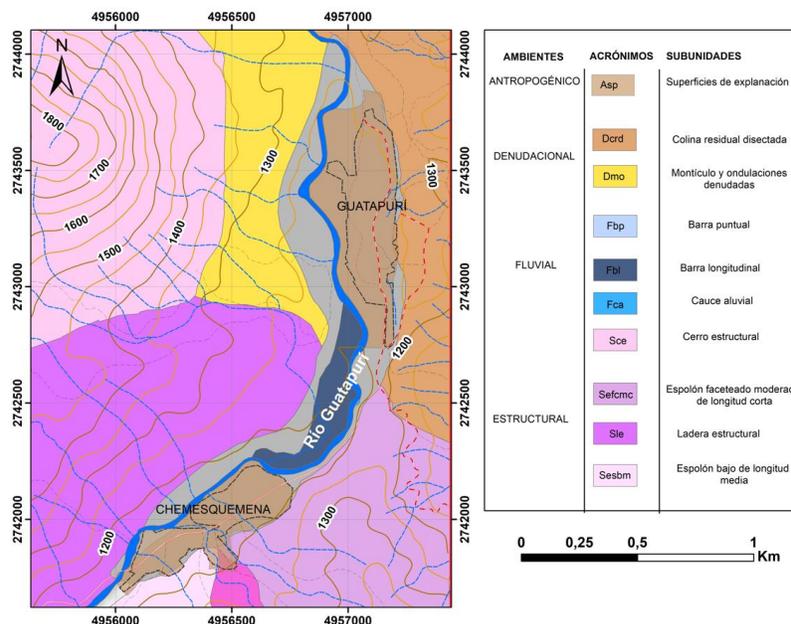


Figura 3. Mapa Geomorfológico.

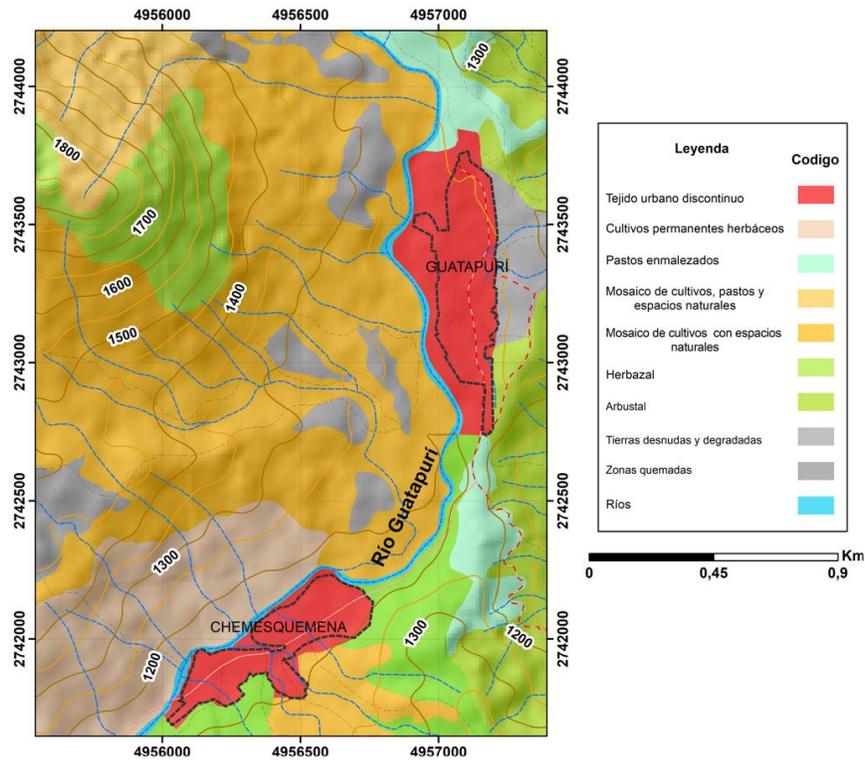


Figura 4. Mapa de cobertura vegetal.

El mapa de cobertura Vegetal (Figura 4), indica que las zonas de estudio con nulos o bajos MM están relacionadas con áreas cercanas a ríos, mosaicos de cultivos, pastizales y espacios naturales, así como zonas de herbazal abierto rocoso y deforestación, abarcando 0.79 km<sup>2</sup> que es el 14% del área. La presencia de MM es moderada en áreas con tejido urbano discontinuo, cultivos permanentes herbáceos, bosques fragmentados con pastizales y cultivos, arbustales densos, y mosaicos de cultivos con espacios naturales, representando alrededor 2.26 km<sup>2</sup> que es el 39% del área de estudio. Por último, las áreas con altos MM, en terrenos desnudos y degradados, en mosaicos de pastizales con espacios naturales, abarcando el 2.26 km<sup>2</sup> que representa el 41% del área total de trabajo.

El parámetro condicionante “pendiente” presenta valores que oscilan de 0° a 64°, predominando altas pendientes aledañas a los centros poblados que están ubicado sobre unas llanuras de baja pendiente (Figura 5a). El quinto y sexto parámetros “TRI” presenta valores entre 0,11 y 0,88 (Figura 5b) y acuencia valores de 10.103 y 20.304 (Figura 5c).

#### 4.2. Preprocesamientos de los datos

Para caracterizar la distribución y variabilidad de los conjuntos de datos de entrada, se calcularon estadísticas básicas como histogramas y se ajustaron distribuciones gaussianas a cada variable. La Figura 6 muestra estos análisis, donde se presentan los histogramas y sus respectivas curvas gaussianas para las variables K1, K2, K3 y K4. La variable K1 (Figura 6a) presenta una distribución dispersa con un sesgo hacia la derecha, lo que sugiere que los valores están más concentrados en los extremos negativos. Esta asimetría podría estar vinculada a la presencia de terrenos inestables o con alta variabilidad geológica, ya que muestra una concentración de valores en el lado negativo del histograma. La distribución observada no sigue un patrón normal, y su comportamiento indica la posible influencia de características físicas específicas del terreno.

Por otro lado, la K2 (Figura 6b) también muestra una asimetría pronunciada hacia la derecha, con una mayor concentración de datos en los valores más bajos del rango. Esto puede estar relacionado con la existencia de áreas de pendiente baja o coberturas de terreno que no favorecen el desarrollo de movimientos en masa. La larga cola hacia la derecha indica que, si bien la mayoría de las áreas analizadas parecen ser estables, algunas zonas

presentan valores más altos de susceptibilidad. En cuanto a la K3 (Figura 6c), se puede observar una distribución cercana a la normal, mostrando un comportamiento más

simétrico, con picos bien definidos y colas menos pronunciadas en comparación con las otras variables.

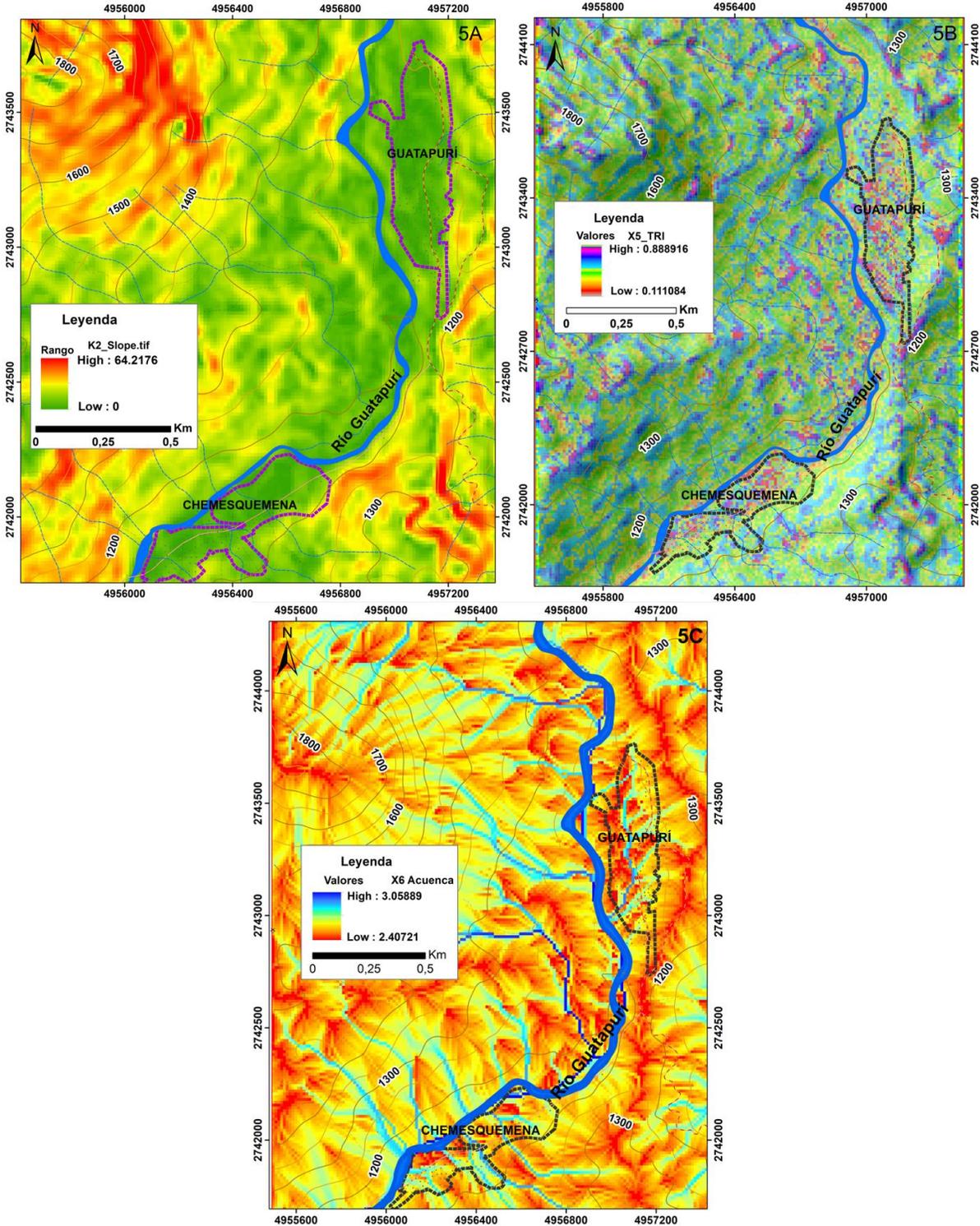


Figura 5. Mapas de derivados del DEM. a) Pendiente, b) TRI, c) Acuenca.

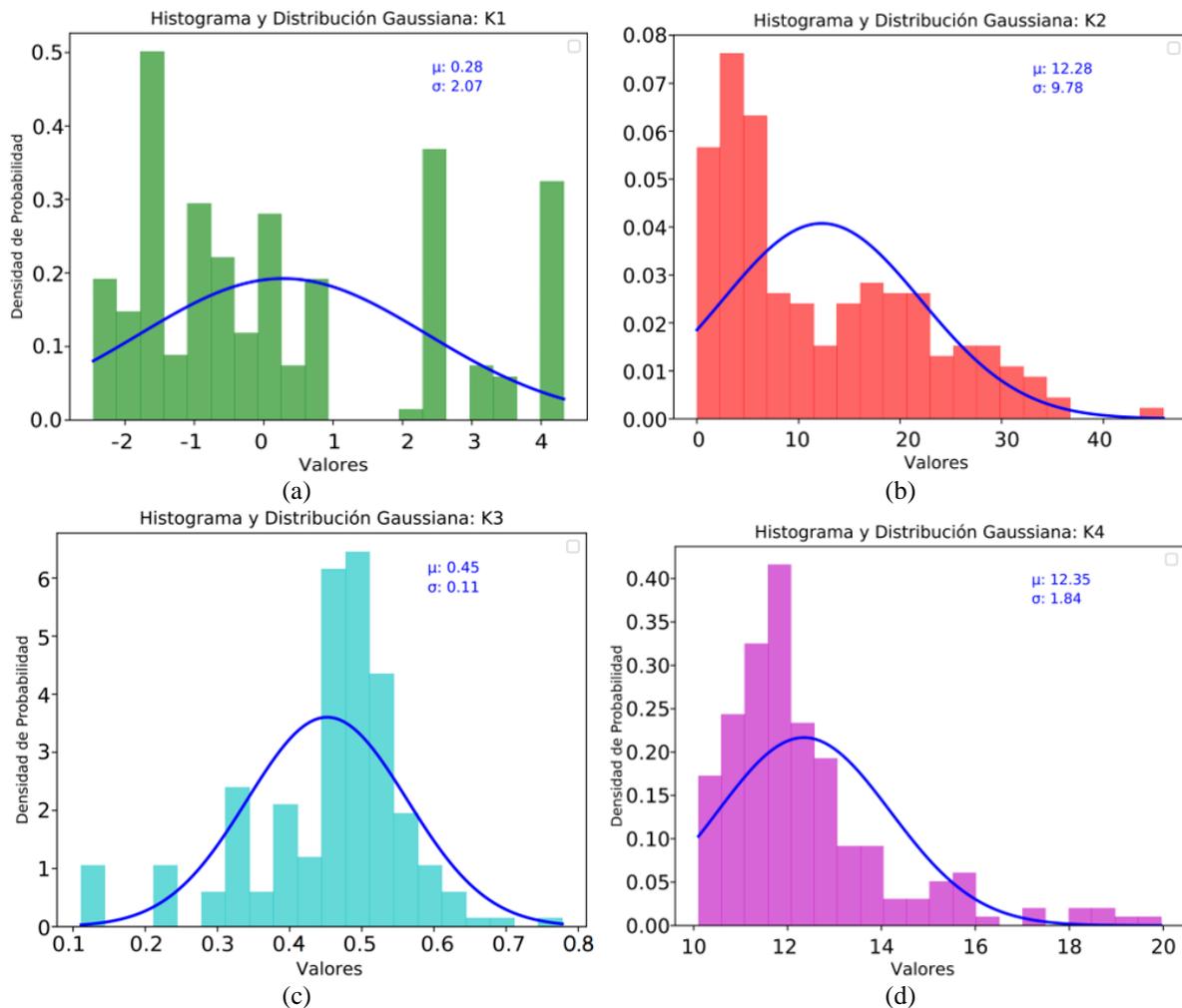


Figura 6. Histograma y distribución Gaussiana: a) K1, b) K2, c) K3, d) K4.

Este patrón sugiere que los valores de esta variable están distribuidos de manera más equilibrada, lo cual podría representar una característica del terreno que afecta de manera uniforme, como la rugosidad o el drenaje superficial.

Finalmente, la variable K4 (Figura 6d) sigue una tendencia similar a K1 y K2, con un claro sesgo hacia la derecha. Esto indica que la mayoría de los datos están concentrados en valores bajos, probablemente representando áreas con baja susceptibilidad a movimientos en masa, aunque existen algunos valores extremos que reflejan zonas con mayor probabilidad de inestabilidad.

Las variables analizadas muestran distribuciones asimétricas y sesgadas hacia la derecha, lo que sugiere que los datos no siguen una distribución normal. La variable K3 es la única que presenta un comportamiento más cercano a la normalidad, mientras que las demás

variables exhiben mayor dispersión y sesgo. Este comportamiento resalta la importancia de utilizar técnicas más robustas en el análisis, capaces de manejar distribuciones asimétricas y ser insensibles al ruido o errores presentes en los datos (Figura 6).

### 4.3. Análisis entre las variables de entrada

La multicolinealidad existe cuando se encuentran correlaciones significativas entre las variables predictoras [26] como se cita en [27], es decir cuando los datos están relacionados linealmente entre sí y por lo tanto tienen coeficientes de correlación altos (Tabla 1).

Para determinar si no existía la correlación entre las variables, se analizó la matriz de correlación con las cuatro (4) variables predictoras K1, K2, K3 y K4, donde se evidencia baja correlación entre las variables, coeficientes de correlación de Pearson con valores absolutos menores de 0.5 (Figura 7), lo cual se considera

correlación moderada a débil según la **Tabla 1**, es decir, se removió la multicolinealidad por ende no existe independencia condicional y evita los problemas de colinealidad entre las variables.

Tabla 1. Interpretación de los valores del coeficiente de correlación de Pearson

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

Fuente: [28].

La **Figura 8** presenta matriz de dispersión para las variables K1, K2, K3 y K4. Se observa amplia dispersión de los puntos de datos a lo largo de los ejes, indicando falta de patrón definido o agrupamiento entre las variables.

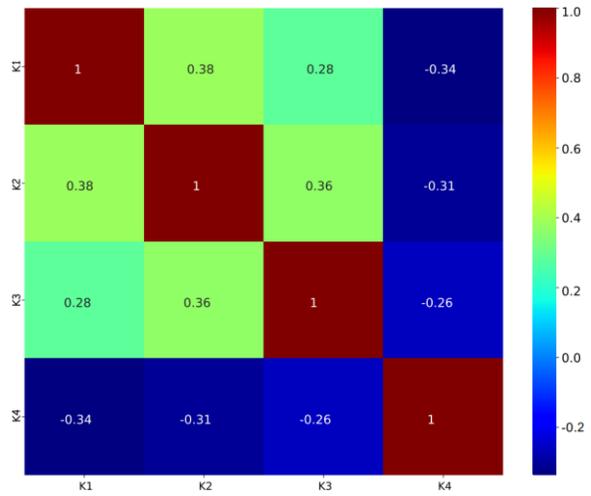


Figura 7. Matriz de correlación de Pearson.

Esta dispersión considerable sugiere una baja correlación entre las variables, respaldada por los valores consistentemente bajos del coeficiente de correlación de Pearson, todos por debajo de 0.57. Esta falta de relación fuerte entre las variables puede simplificar la tarea del algoritmo al separar los datos mediante una superficie de decisión [27].

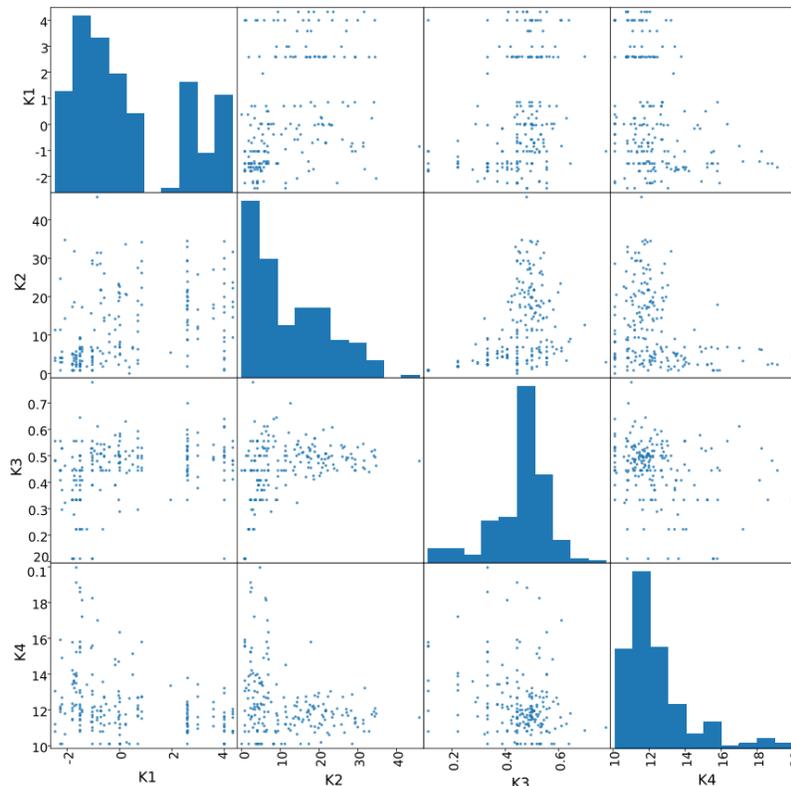


Figura 8. Matriz de dispersión para las variables K1, K2, K3, K4, K5. En la diagonal se muestra el histograma para cada variable.

#### 4.4. Diseño de la Red Neuronal Artificial

##### 4.4.1 Redes Neuronales Artificiales (RNA)

Esta investigación utilizó la función sigmoidea (logística), la cual generó valores entre cero y uno para cada una de las entradas, y se fijaron las tasas de aprendizaje en 0,001, un parámetro crucial que determina el tamaño del paso en cada iteración. La elección de la tasa de aprendizaje es vital: si es muy pequeña, se necesitan más épocas ya que los cambios en las actualizaciones son mínimos; si es demasiado grande, el modelo converge rápidamente, pero corre el riesgo de oscilaciones [29]. Cabe resaltar que no existe una regla general para seleccionar el ritmo de aprendizaje adecuado, sino que se determina experimentalmente para cada problema específico [26].

El proceso de entrenamiento de redes neuronales implicó el uso de algoritmos como “optimizador Adam en conjunto con la backpropagation”, para resolver problemas de optimización en grandes conjuntos de datos, requiriéndose 1000 épocas para marcar el final de la fase de aprendizaje de la red neuronal.

##### 4.4.2 Evaluación de precisión

La validación es fundamental para conocer el valor predictivo de los modelos [30]. Para la validación de los resultados de nuestro análisis, se realizó la curva de características operativas del receptor (ROC) y se calculó el valor del área bajo la curva (AUC) para cada modelo. Este método ha sido ampliamente utilizado para proporcionar predicciones del rendimiento y comparar

los diferentes modelos [31], [32]. El análisis ROC se considera un método poderoso para la validación del desempeño de los modelos LS [33], [34].

En la **Figura 9** la sensibilidad del modelo se muestra en función de la especificidad. La curva ROC correspondiente indica la capacidad del modelo para discriminar correctamente entre observaciones positivas y negativas en el espacio de validación [35]. Una alta sensibilidad indica una gran cantidad de predicciones correctas, mientras que una alta especificidad indica una pequeña cantidad de falsos positivos. El valor del área bajo la curva (AUC) varía de 0,5 a 1,0. El modelo ideal presenta valor AUC cercano a 1,0 (ajuste perfecto), mientras que un valor cercano a 0,5 indica inexactitud en el modelo (ajuste aleatorio) [36], [37].

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. Sin embargo, debido a la escasez de datos disponibles sobre deslizamientos de tierra en el área de estudio, resultó inviable dividir el conjunto de datos de deslizamientos en un conjunto separado de entrenamiento para desarrollar los modelos y otro de predicción para su validación. Como alternativa, se optó por seleccionar aleatoriamente un porcentaje (25%) 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.

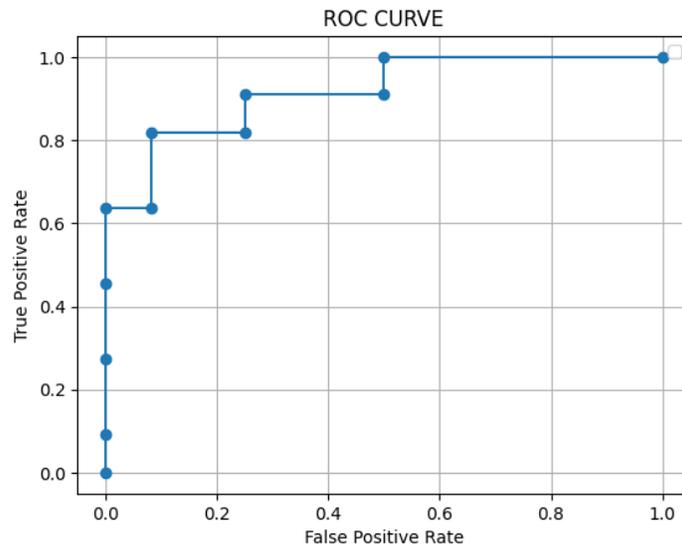


Figura 9. Evaluación de la capacidad de predicción: ROC CURVE.

La evaluación del modelo a partir de los valores del área bajo la curva y características operativas del receptor presentó valor de  $AUC = 0,975$ ; indicando, que la estimación de la curva ROC de la RNA esboza los resultados de la precisión y evaluación de la función de pérdida durante el entrenamiento y validación de las redes neuronales. Presenta un buen desempeño el algoritmo de entrenamiento usando las 4 variables de entrada, llegando hasta una precisión cercana al 100% para los datos de validación.

#### 4.4.3 Evaluación de la susceptibilidad

La susceptibilidad muy alta en la zona de estudio cubre aproximadamente el 43% del área, principalmente en las partes medias y altas, se caracterizan por presentar pendientes pronunciadas y diferentes patrones de flujo de agua, lo que las hace propensas a movimientos en masa y morfológicamente relieve moderadamente a muy altas (Figura 10).

La susceptibilidad alta en la zona de estudio abarca aproximadamente un 6% del área total y se caracteriza por presentar pendientes muy abruptas y diferentes patrones de flujo de agua asociadas a materiales con erosión moderada y relieves relativamente elevados,

geformas denudacionales, suelos residuales relacionados principalmente con coberturas vegetales de bosque de galería y mosaicos de pastos con espacios naturales (Figura 10).

La susceptibilidad moderada en la zona de estudio abarca alrededor del 6% del área total y comparte similitudes con la susceptibilidad alta. Se caracteriza por presentar pendientes inclinadas a moderadamente abruptas, flujo de agua y áreas de escorrentía lenta., materiales relativamente blandos, alta erosión y relieves de 200 a 400 m.s.n.m. Morfológicamente están vinculadas a geformas denudacionales, suelos residuales y cobertura de pastos enmalezados, pastos limpios y tejido urbano (Figura 10).

La susceptibilidad baja en la zona de estudio abarca aproximadamente el 10% del área total. Se caracteriza por presentar pendientes inclinadas a moderadamente inclinadas y áreas de escorrentía lenta asociadas con materiales blandos y relieves de menos de 200 m.s.n.m.

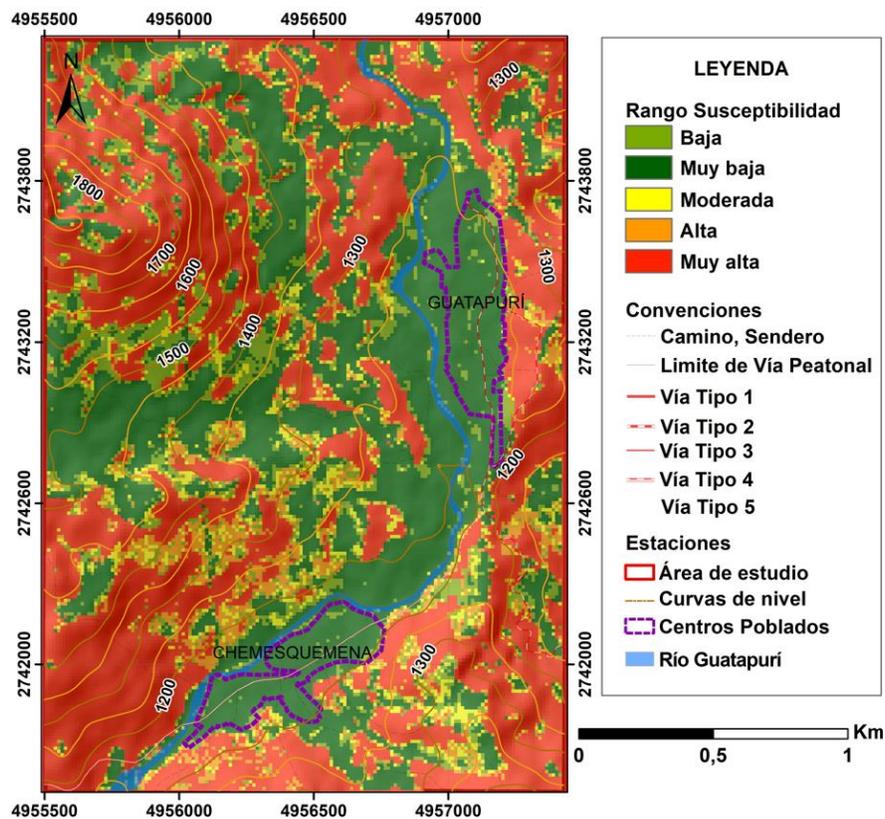


Figura 10. Mapa de susceptibilidad a procesos de remoción en masa a escala 1:25.000.

Morfológicamente, se relacionan con diversas unidades geomorfológicas y litologías, incluidos los depósitos Cuaternarios y algunas rocas de Granulitas de los Mangos, Batolito de Atanquez, Riolita del Golero, suelos transportados y orgánicos desarrollados en ambientes fluviales y denudacionales, cultivos y pastos, con presencia ocasional de herbazales (Figura 10).

La susceptibilidad muy baja ocupa alrededor del 34% del área total de la zona de estudio y se concentran principalmente en los centros poblados de Guatapurí y Chemesquemena. Estas áreas se caracterizan por presentar pendientes suavemente inclinadas a casi planas, asociadas con zonas de escorrentía lenta. Morfológicamente, presentan materiales muy blandos y relieves relativamente bajos. Están vinculadas a ambientes fluviales, antropogénicos y denudacionales, con predominio de suelos aluviales y residuales, mosaicos de cultivos, pastos y espacios naturales (Figura 10).

La distribución de las áreas en las clases de susceptibilidad (Figura 10), indica que la zona presenta principalmente susceptibilidades muy altas con área aproximada de 2.3 Km<sup>2</sup> (43% del área de estudio), susceptibilidad muy baja con un área de 1.87 Km<sup>2</sup> (34% del área de estudio), bajas con un área de 0.55 Km<sup>2</sup> (10.2% del área de estudio), moderadas con un área de 0,3 Km<sup>2</sup> (6 % del área de estudio) y altas con un área de 0.3 Km<sup>2</sup> (6 % del área de estudio).

El área de movimiento en masa en cada clase (Figura 7), indica que el 85,76% de los movimientos en masa corresponden a susceptibilidad muy alta, 10,13% susceptibilidad alta, 2,5% moderada, 1,33% baja y muy baja 0, 28%.

## 5. Conclusiones

La prueba de validación del modelo reveló un alto valor en el área bajo la curva (AUC=0,975), lo que demuestra una capacidad notable para distinguir entre áreas con diferentes niveles de susceptibilidad. La precisión se evidenció en las características geológicas, geomorfológicas y climáticas, asociadas a niveles variables de susceptibilidad a remoción en masa. Este enfoque integral no solo permitió predecir áreas de riesgo, sino también comprender las diversas causas subyacentes que contribuyen a la susceptibilidad en dichas áreas.

Durante el preprocesamiento de los datos, se observó una baja correlación entre las variables predictoras K1, K2, K3 y K4, con coeficientes de correlación de Pearson inferiores a 0.5, lo que indica una correlación moderada

a débil. Esta baja correlación eliminó la multicolinealidad, lo que facilitó el proceso de entrenamiento del modelo al reducir los problemas de colinealidad entre las variables. Además, la dispersión de los puntos de datos a lo largo de los ejes en la matriz de dispersión confirmó la falta de un patrón definido o agrupamiento entre las variables, lo que simplifica la tarea del algoritmo al separar los datos mediante una superficie de decisión. Este análisis reforzó la elección de técnicas robustas, como las redes neuronales, que pueden manejar datos con relaciones complejas y distribuciones no normales.

La investigación utilizó la función sigmoidea (logística) para generar valores entre cero y uno en las entradas, lo que, junto con una tasa de aprendizaje fijada en 0.001, permitió un balance adecuado entre la velocidad de convergencia y la estabilidad del modelo. Se seleccionó el optimizador Adam en conjunto con el algoritmo de backpropagation para resolver problemas de optimización en grandes volúmenes de datos, lo que requirió 1000 épocas para completar el proceso de aprendizaje de la red neuronal.

La evaluación de la susceptibilidad reveló una distribución desigual en las zonas de Guatapurí y Chemesquemena. Las áreas con "muy alta" susceptibilidad, que representan el 43% del área total, se caracterizan por tener pendientes pronunciadas, flujos de agua complejos y relieves elevados, lo que las hace más propensas a movimientos en masa. En contraste, las áreas de "muy baja" susceptibilidad, que cubren el 34% del territorio, presentan pendientes suaves y materiales menos susceptibles a deslizamientos. Esta distribución evidencia la necesidad de enfocar los esfuerzos de mitigación en las zonas de mayor susceptibilidad, que abarcan el 77% del territorio, para reducir la vulnerabilidad ante desastres.

Finalmente, se observó que el 96% de los deslizamientos se concentran en las áreas catalogadas como de alta y muy alta susceptibilidad. Este hallazgo subraya la importancia de implementar estrategias de gestión del riesgo, tales como la planificación del uso del suelo, la construcción de infraestructura resiliente y la educación comunitaria, con el objetivo de mitigar el impacto de estos eventos en las zonas más vulnerables. El uso de redes neuronales permitió abordar la complejidad de las variables geoespaciales, mejorando significativamente la identificación de zonas susceptibles y proporcionando una herramienta clave para la gestión del riesgo de deslizamientos.

## Financiación

No aplica

## Contribución de los autores

J. Estrada-Romero: Preparación de datos, Análisis formal, Investigación, Redacción – borrador original. F. Carrillo: Conceptualización, Preparación de datos, Análisis formal, Investigación, Redacción – revisión y edición. D. C. Manco-Jaraba: Análisis formal, Investigación, Redacción – borrador original, Visualización. J. Cantillo. Análisis formal, Investigación, Redacción – borrador original, Visualización.

Todos los autores han leído y están de acuerdo con la versión publicada del artículo.

## Conflicto de interés

Declaramos que no existe ningún conflicto de intereses en relación con este artículo. Los autores no tienen intereses financieros, personales, ni profesionales que puedan influir en los resultados o en la interpretación de los hallazgos encontrados.

## Declaración de la Junta de Revisión Institucional

No aplica.

## Declaración de consentimiento informado

No aplica.

## Referencias

- [1] V. Ortiz-Maestre, C. Polo-Mendoza, D. Giraless-Puerta, D. Manco-Jaraba, “Análisis de susceptibilidad por movimientos en masa implementando el método Mora-Vahrson modificado para el corregimiento de Chemesquemena (Cesar, Colombia),” *Tecnura*, vol. 27, no. 77, pp. 1–21, 2022, doi: <https://doi.org/10.14483/22487638.19951>
- [2] E. Rodríguez et al., “Guía metodológica para la zonificación de amenaza por movimientos en masa escala 1: 25.000,” Servicio Geológico Colombiano, 2017. doi: <https://doi.org/10.32685/9789585978225>
- [3] F. C. Dai, C. F. Lee, and Y. Y. Ngai, “Landslide risk assessment and management: An overview,” *Eng. Geol.*, vol. 64, no. 1, pp. 65–87, 2002, doi: [https://doi.org/10.1016/S0013-7952\(01\)00093-X](https://doi.org/10.1016/S0013-7952(01)00093-X)

- [4] S. Park, C. Choi, B. Kim, J. Kim, “Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea,” *Environ. Earth Sci.*, vol. 68, no. 5, pp. 1443–1464, 2013, doi: <https://doi.org/10.1007/s12665-012-1842-5>
- [5] H. Zhang, G. Zhang, Q. Jia, “Integration of Analytical Hierarchy Process and Landslide Susceptibility Index Based Landslide Susceptibility Assessment of the Pearl River Delta Area, China,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 12, no. 11, pp. 4239–4251, 2019, doi: <https://doi.org/10.1109/JSTARS.2019.2938554>
- [6] B. Kalantar, N. Ueda, U. S. Lay, H. A. H. Al-Najjar, and A. A. Halin, “Conditioning Factors Determination for Landslide Susceptibility Mapping Using Support Vector Machine Learning,” in *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2019, pp. 9626–9629, doi: <https://doi.org/10.1109/IGARSS.2019.8898340>
- [7] Y. Yi, Z. Zhang, W. Zhang, and C. Xu, “Comparison of Different Machine Learning Models for Landslide Susceptibility Mapping,” in *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2019, pp. 9318–9321, doi: <https://doi.org/10.1109/IGARSS.2019.8898208>
- [8] H. Hong et al., “Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China),” *Catena*, vol. 163, pp. 399–413, 2018, doi: <https://doi.org/10.1016/j.catena.2018.01.005>
- [9] B. Pradhan, S. Lee, “Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models,” *Environ. Earth Sci.*, vol. 60, no. 5, pp. 1037–1054, 2010, doi: <https://doi.org/10.1007/s12665-009-0245-8>
- [10] A. Lambert, *Manual de Muestreo para Exploración, Minería Subterránea Y Rajo Abierto*. 2006.
- [11] Servicio Geológico Colombiano, “Sistema de Información de Movimientos en Masa.” [Online]. Available: <https://simma.sgc.gov.co/#/public/results/>
- [12] D. Varnes, “Slope Movement Types and Processes,” 1978.

- [13] J. H. Carvajal, *Propuesta de estandarización de la cartografía geomorfológica en Colombia*. Imprenta Nacional de Colombia 2012.
- [14] Ideam, Igac, Cormagdalena, “Metodología CORINE Land Cover Adaptada para Colombia escala 1:100.000,” 2016.
- [15] Ministerio de Vivienda, “Decreto Numero 1077 de 2015 ‘Por medio del cual se expide el Decreto Único Reglamentario del Sector Vivienda, Ciudad y Territorio,’” Decreto, vol. 2015, pp. 1–829, 2015.
- [16] Pamela, I. A. Sadisun, Y. Arifianti, “Weights of Evidence Method for Landslide Susceptibility Mapping in Takengon, Central Aceh, Indonesia,” in *IOP Conference Series: Earth and Environmental Science*, 2018, doi: <https://doi.org/10.1088/1755-1315/118/1/012037>
- [17] C. V. Patriche, R. Pirnau, A. Grozavu, and B. Rosca, “A Comparative Analysis of Binary Logistic Regression and Analytical Hierarchy Process for Landslide Susceptibility Assessment in the Dobrovăț River Basin, Romania,” *Pedosphere*, vol. 26, no. 3, 2016, doi: [https://doi.org/10.1016/S1002-0160\(15\)60047-9](https://doi.org/10.1016/S1002-0160(15)60047-9)
- [18] S. Lee, J. H. Ryu, M. J. Lee, and J. S. Won, “Use of an artificial neural network for analysis of the susceptibility to landslides at Boun, Korea,” *Environ. Geol.*, vol. 44, no. 7, pp. 820–833, 2003, doi: <https://doi.org/10.1007/s00254-003-0825-y>
- [19] E. Rouault, “GDAL”. Zenodo, nov. 06, 2024, doi: <https://doi.org/10.5281/zenodo.14046734>
- [20] W. McKinney, “Pandas documentation.” 2023.
- [21] K. H. Zou, A. J. O’Malley, L. Mauri, “Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models,” *Circulation*, vol. 115, pp. 654–657, 2007.
- [22] J. G. Kim, S. H. Shin, H. Kang, “A Case Study on the Use of ROC Curve and AUC in the Evaluation of Discriminant Model,” *J. Korean data Anal. Soc.*, vol. 20, pp. 609–619, 2018.
- [23] F. Pedregosa, “SciKit-Learn: Machine Learning in Python.” *Journal of Machine Learning Research*, vol. 12, 2012.
- [24] A. Burkov, *The hundred-page machine learning book-annotated*. 2019.
- [25] E. Yesilnacar, T. Topal, “Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey),” *Eng. Geol.*, vol. 79, no. 3–4, 2005, doi: <https://doi.org/10.1016/j.enggeo.2005.02.002>
- [26] P. Goyes-Peñañiel, A. Hernandez-Rojas, “Landslide susceptibility index based on the integration of logistic regression and weights of evidence: A case study in Popayan, Colombia,” *Eng. Geol.*, vol. 280, 2021, doi: <https://doi.org/10.1016/j.enggeo.2020.105958>
- [27] A. Mantilla, P. Goyes-Peñañiel, “Predicción de la ocurrencia de depósitos minerales tipo pórfido usando técnicas de aprendizaje automático,” proyecto de grado, Universidad Industrial de Santander, 2023.
- [28] D. Mindrila and P. Balentyne, “Scatterplots and Correlation,” 2017. [Online]. Available: <https://www.studocu.com/es-mx/document/universidad-univer/economy/scatterplots-and-correlation-notes/14672739>
- [29] J. Remondo, A. González-Díez, J. R. Díaz de Terán, A. Cendrero, “Landslide susceptibility models utilising spatial data analysis techniques. A case study from the lower Deba Valley, Guipúzcoa (Spain),” *Nat. Hazards*, vol. 30, pp. 267–279, 2003, doi: <https://doi.org/10.1016/10.1023/B:NHAZ.0000007202.12543.3a>
- [30] S. Beguería, “Validation and evaluation of predictive models in hazard assessment and risk management,” *Nat. Hazards*, vol. 37, pp. 315–329, 2006, doi: <https://doi.org/10.1007/s11069-005-5182-6>
- [31] D. W. Hosmer, S. Lemeshow, *Applied Logistic Regression*. John Wiley & Sons, 2002, doi: <https://doi.org/10.1002/0471722146>
- [32] P. Frattini, G. Crosta, A. Carrara, “Techniques for evaluating the performance of landslide susceptibility models,” *Eng. Geol.*, vol. 111, pp. 62–72, 2010, doi: <https://doi.org/10.1016/j.enggeo.2009.12.004>
- [33] A. Nandi and A. Shakoor, “Preparation of a landslide susceptibility map of Summit County, Ohio , USA , using numerical models,” *IAEG2006*, no. 660, pp. 1–11, 2006.

[34] L. Montrasio, R. Valentino, G. L. Losi, “Towards a real-time susceptibility assessment of rainfall-induced shallow landslides on a regional scale,” *Nat. Hazards Earth Syst. Sci.*, vol. 11, no. 7, pp. 1927–1947, 2011, doi: <https://doi.org/10.5194/nhess-11-1927-2011>

[35] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognit. Lett.*, vol. 27, no. 8, 2006, doi: <https://doi.org/10.1016/j.patrec.2005.10.010>

[36] I. Yilmaz, “A case study from Koyulhisar (Sivas-Turkey) for landslide susceptibility mapping by artificial neural networks,” *Bull. Eng. Geol. Environ.*, vol. 68, no. 3, 2009, doi: <https://doi.org/10.1007/s10064-009-0185-2>

[37] I. Yilmaz, “Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat-Turkey),” *Comput. Geosci.*, vol. 35, no. 6, 2009, doi: <https://doi.org/10.1016/j.cageo.2008.08.007>