

Multi-scale Remote Sensing for Tree and Forest Species Identification in Ecuadorian Andes using Machine Learning and Deep Learning Methods

Doctorando: Diego Pacheco Prado

Director: Luis Ángel Ruiz Fernández

RESUMEN

La cordillera de los Andes que atraviesa Ecuador contribuye a la presencia de una enorme variedad de microclimas, que junto a factores como altitud y humedad del suelo influyen fuertemente en la distribución y crecimiento de distintas especies forestales, haciendo del país uno de los más megadiversos del mundo. En el cantón Cuenca (área de estudio), ubicado en la zona centro/sur del Ecuador, se pueden encontrar ecosistemas como páramos, bosques montanos, humedales, lagunas, valles interandinos. De estos, los bosques montanos andinos tienen una alta prioridad de conservación dadas sus funciones de regulación del flujo de agua, secuestro de carbono, regeneración de suelos y retención de nutrientes y sedimentos.

Sin embargo, la deforestación y la expansión agrícola son una amenaza latente para estos ecosistemas. Esta problemática también se manifiesta dentro del área urbana, donde la expansión urbana y el mantenimiento de las áreas verdes públicas han desplazado a las especies nativas de la región. Por ello, la identificación forestal se convierte en una tarea clave para la conservación y el manejo sostenible de los ecosistemas nativos.

En este contexto, la identificación de especies forestales se abordó en tres escalas, definidas según la extensión espacial y el sensor usado para la toma de datos. La primera escala, denominada regional, evaluó los datos satelitales en el cantón Cuenca (Ecuador). La segunda escala denominada local, evaluó los datos fotogramétricos obtenidos con dron a nivel de árbol. En la última escala denominada a corta distancia, se evaluaron fotografías tomadas con teléfonos inteligentes a nivel de árbol. Las dos últimas escalas fueron evaluadas dentro del área urbana del cantón.

En la escala regional (satelital), se consiguió la identificación de rodales del género *Polylepis*, con una fiabilidad global de 0,91 y un coeficiente Kappa de 0,80. Para ello se utilizaron las variables topográficas (elevación, pendiente y orientación), junto con variables espectrales (bandas e índices) obtenidas de Sentinel-2. Esto se realizó en la plataforma Google Earth Engine con el algoritmo Random Forest. Cabe destacar que, en varios mapas oficiales, estos bosques se encontraban generalizados como bosques nativos o eran absorbidos por la clase mayoritaria que rodeaba a estos rodales que es el Páramo.

En la escala local (dron) se identificaron con una fiabilidad global de 0,70 y Kappa de 0,65 las especies *Populus alba* L., y *Melaleuca armillaris* (Sol. ex Gaertn.) Sm, entre ocho especies evaluadas. Para ello se utilizó la información RGB de las nubes de puntos fotogramétricas de dron, complementadas con la caracterización vertical del árbol basada en la coordenada Z (normalizada) de la nube de puntos, a través de los valores de Z máximo y los percentiles 10, 50 y 90; además de la densidad de la nube de puntos por árbol.

Finalmente, en la escala a corta distancia (smartphone), las especies *Yucca gigantea* Lem., *Salix humboldtiana* y *Melaleuca armillaris* (Sol. ex Gaertn.) Sm. tuvieron una alta tasa de acierto entre 14 especies evaluadas. En esta escala se registró una fiabilidad global de 0,83 y kappa de 0,81.

Para la identificación, en la escala regional y local la selección de las características más relevantes se realizó evaluando múltiples variables obtenidas de una revisión bibliográfica. De todas estas se usó la técnica *Recursive Feature Elimination* (RFE) a través de *Random Forest*, para eliminar las variables de menor importancia. Con ello se buscó reducir al mínimo la cantidad de variables a usar en la identificación, y a su vez mantener el nivel de fiabilidad. Finalmente, en la escala a corta distancia, al usar un modelo de *Deep Learning*, donde las características se presentan como una “caja negra”, no se pudieron determinar las variables usadas para la identificación. En su lugar se utilizaron los mapas de activación para determinar qué regiones de la fotografía fueron las más relevantes en la asignación de la especie forestal.

Cada nivel evaluado tuvo su complejidad, principalmente en acceso a datos en campo, necesarios para el entrenamiento de los modelos de *Machine* y *Deep Learning*, los cuales serán analizados en este trabajo.