



GOBIERNO
DE COLOMBIA



DNP DEPARTAMENTO
NACIONAL
DE PLANEACIÓN

Departamento Nacional de Planeación

www.dnp.gov.co

— **Análisis de** —
imágenes satelitales
para la identificación
de vías terciarias
proyecto piloto Depto.
Santander

Dirección de Desarrollo Digital
Unidad de científicos de datos

Septiembre, 2018
dnp.gov.co



DNP DEPARTAMENTO
NACIONAL
DE PLANEACIÓN

1. Objetivo
2. Pre-procesamiento
3. Modelos y Resultados



— Análisis de —
**imágenes satelitales
para la identificación
de vías terciarias**
proyecto piloto Depto.
Santander

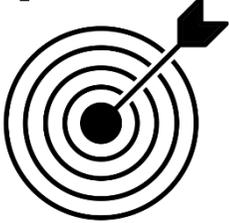
Objetivo

1

— Análisis de —
**imágenes satelitales
para la identificación
de vías terciarias**
proyecto piloto Depto.
Santander

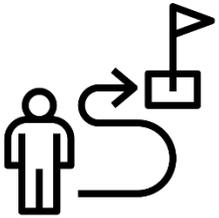
Imágenes satelitales para identificación de vías

Colombia no cuenta con un registro de todas sus vías terciarias y se estima que estas superan los 150.000 kilómetros en todo el territorio nacional



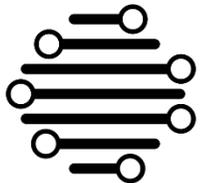
Objetivo

Determinar un intervalo de precisión que identifique si es factible la inversión de un equipo dedicado a obtener el primer archivo cartográfico con la malla vial terciaria de Colombia.



Metodología

Máquinas de Soporte Vectorial
Redes Neuronales Convolucionales



Insumo

S

Imágenes satelitales del departamento de Santander: 146 Imágenes (4304x8817), 43 GB

Información GPS de algunas vías terciarias en el depto.: 31 archivos shp, 0.5 GB

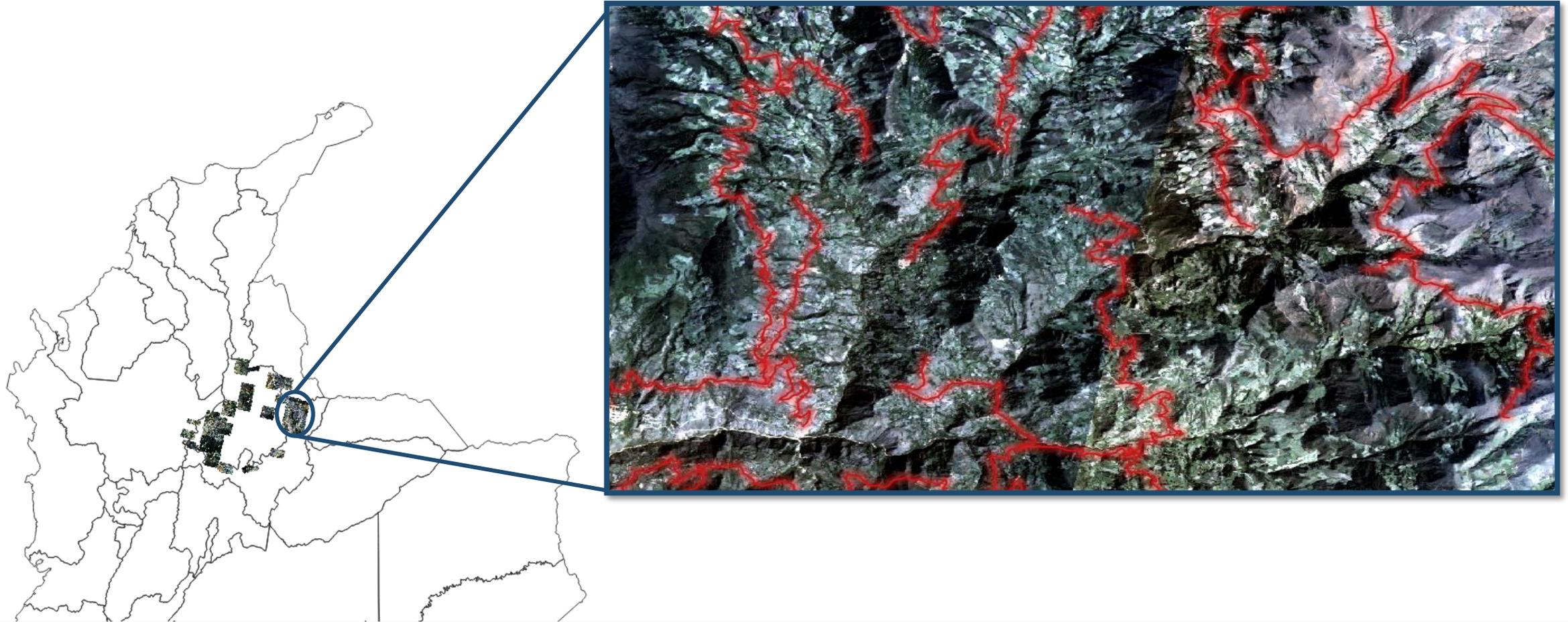
Pre-procesamiento

2

— Análisis de —
**imágenes satelitales
para la identificación
de vías terciarias**
proyecto piloto Depto.
Santander

Etiquetas para cada imagen satelital

Cada imagen tiene un correspondiente archivo .shp que contiene información GPS de algunas vías capturadas manualmente, esta información es esencial para discernir dónde existe o no una vía.



Proceso de análisis de imágenes satelitales

Cada imagen requiere separar sus capas de colores para luego construir matrices. Estas matrices son insumo fundamental para la creación de cualquier algoritmo.



Las imágenes suministradas cuentan con una **cuarta capa:** la capa Alpha o transparencia

Proceso de análisis de imágenes satelitales

Cada imagen se divide en ventanas más pequeñas de dimensiones establecidas bajo criterio personal con el fin de identificar si para cada subdivisión, existe o no, vías terciarias.



Identificación de vías *

No	No	No	No	No	Si	No	No	No	No
No	No	No	No	No	Si	No	No	No	No
No	No	No	No	No	Si	No	No	No	No
No	No	No	No	No	No	Si	Si	Si	No

* Para la clasificación de ventanas individuales, se utilizaron métricas proporcionales sobre el total de píxeles de la imagen y presencia de vía en el píxel central de la ventana

Al tener cada **ventana clasificada** entre si tiene vías o no. Es posible aplicar algoritmos de **aprendizaje supervisado**

Los modelos planteados para este proyecto fueron:

1. Máquinas de Soporte Vectorial (SVM)
2. Redes Neuronales Convolucionales (CNN)

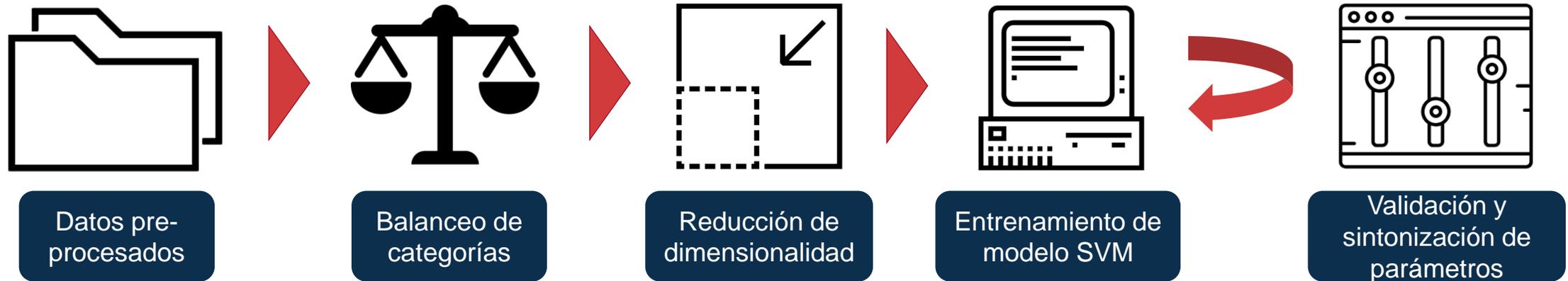
Modelos y Resultados

3

— Análisis de —
**imágenes satelitales
para la identificación
de vías terciarias**
proyecto piloto Depto.
Santander

Modelo de Máquinas de Soporte Vectorial (SVM)

Para el desarrollo del algoritmo de identificación, se utilizó el siguiente flujo de trabajo



Desarrollo de modelo SVM

Balanceo de categorías en los datos

- Solo una proporción muy pequeña de los datos contiene vías terciarias (alrededor de **1%**).
- La mayoría de algoritmos de clasificación (incluyendo SVM) no funcionan bien si son entrenados con datos tan poco balanceados. Para solucionar esto, se tomaron 3 medidas:

1. Trabajar con más imágenes (10).
2. Trabajar solo con una muestra de los registros negativos (ventanas en las que no hay vía).
3. Modificar y aumentar los registros positivos (ventanas que contienen vía).

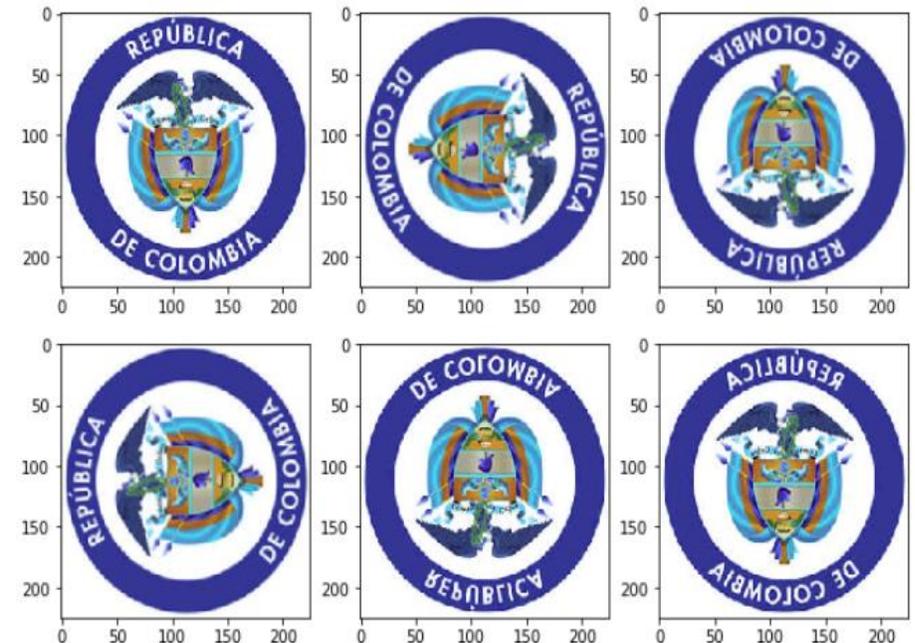
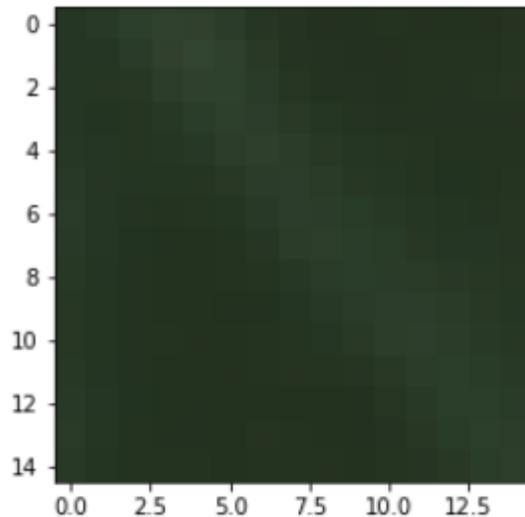


Gráfico: Ejemplo: Modificaciones para obtener 6 imágenes a partir de una sola.

Desarrollo de modelo SVM

Reducción de dimensionalidad

- El algoritmo SVM se vuelve muy costoso computacionalmente a medida que aumenta el número de variables a considerar.
- Se utiliza la técnica de análisis de componentes principales (PCA) para reducir la dimensionalidad (# de variables) de los datos, sin perder mucha información de los mismos.



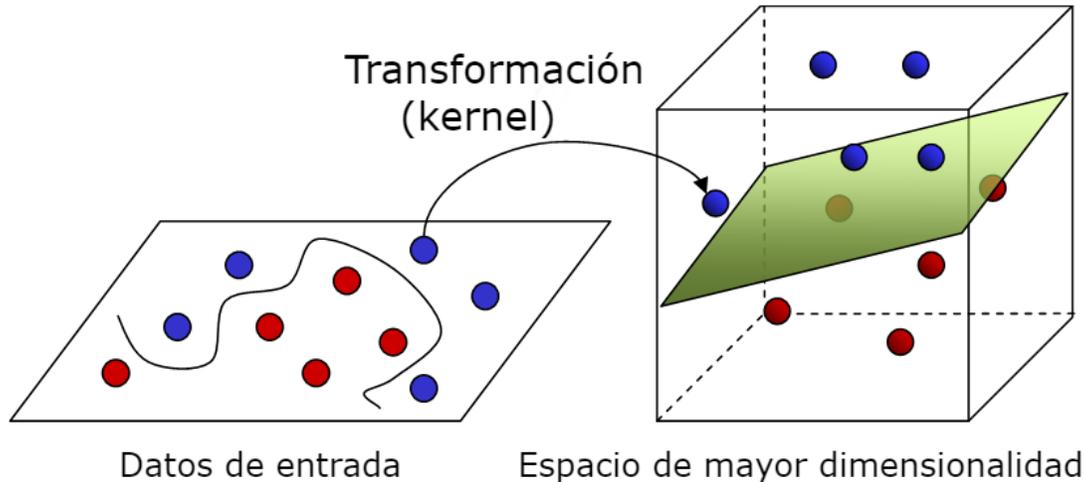
- Ventana de 15 pixeles en el espacio RGB $\rightarrow 15*15*3 =$ **675** variables.
- Utilizando PCA, el 97.1% de la varianza de los datos se puede representar con los primeros **18** componentes principales.

Desarrollo de modelo SVM

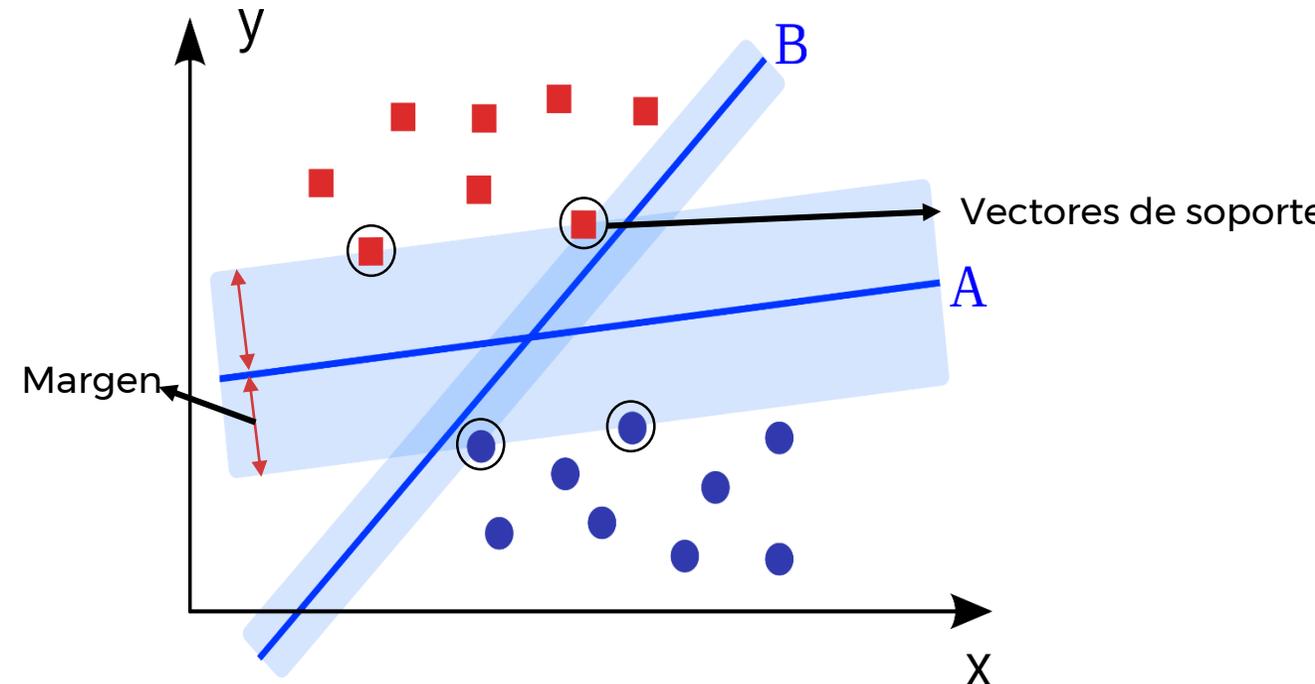
Máquinas de soporte vectorial (SVM)

- Algoritmo de aprendizaje de máquina que sirve para realizar clasificación binaria.
- Involucra un problema de optimización que se puede resumir en dos pasos:

1. Llevar los datos a un espacio dimensional en el que sean linealmente separables.



2. Escoger hiperplano de separación que maximice el margen entre las dos categorías.



Desarrollo de modelo SVM

Sintonización de parámetros

El desarrollo del modelo contiene varios parámetros que pueden ser variados para mejorar el desempeño del mismo.

- Parámetros de los datos de entrada:
 - Tamaño de la ventana → afecta número de datos en cada registro.
 - Escala de color a utilizar (grises, RGB, RGBA) → afecta número y tipo de datos.
- Parámetros del modelo:
 - C: Penalidad que el algoritmo aplica al error.
 - Kernel: Tipo de transformación que aplica el algoritmo a los datos de entrada.
 - Gamma: Parámetro del kernel.

Resultados

Desempeño del modelo SVM

	Datos de entrenamiento			Datos de prueba		
Precisión	86,7%			86.5%		
Matriz de confusión	Real / Predicción	No vía	Vía	Real / Predicción	No vía	Vía
	No vía	34747	1382	No vía	11537	454
	Vía	6779	18445	Vía	2301	6159

Resultados

Prueba del modelo sobre una imagen

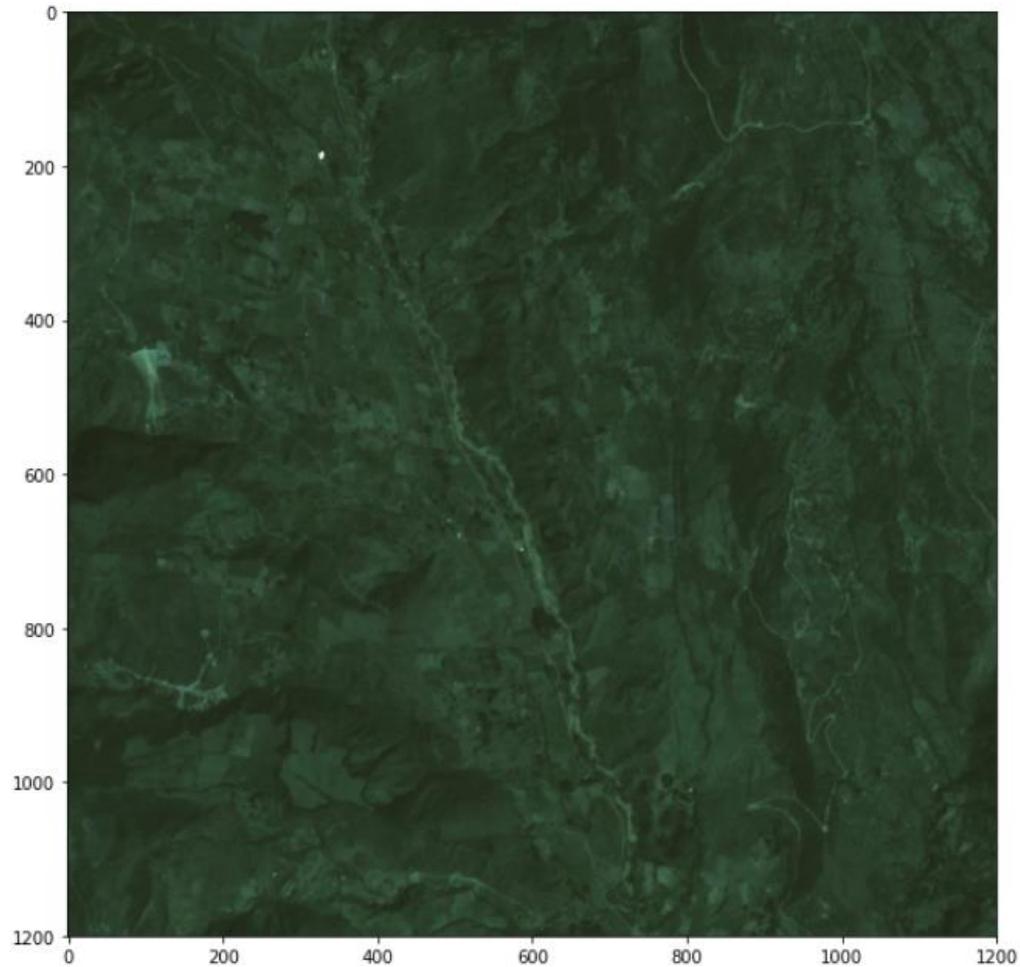


Gráfico: Imagen satelital - Cerrito, Santander.

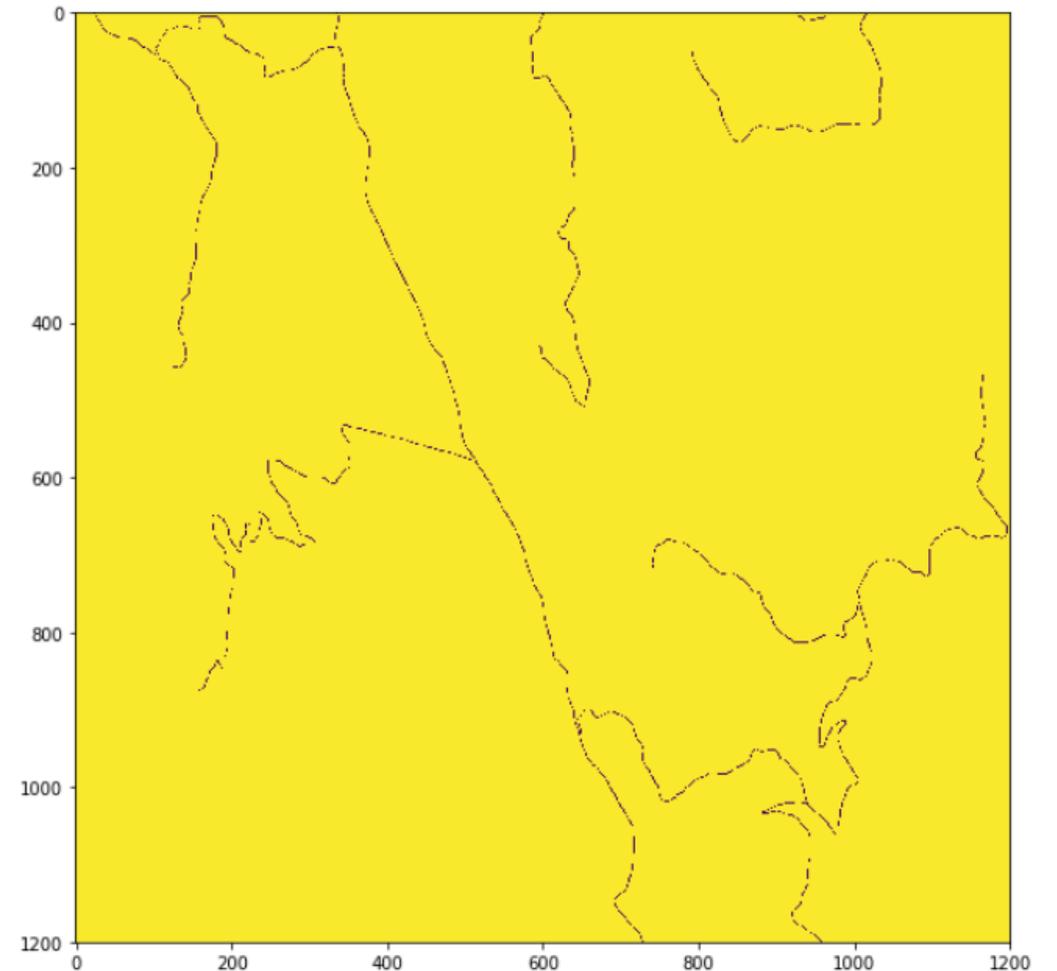
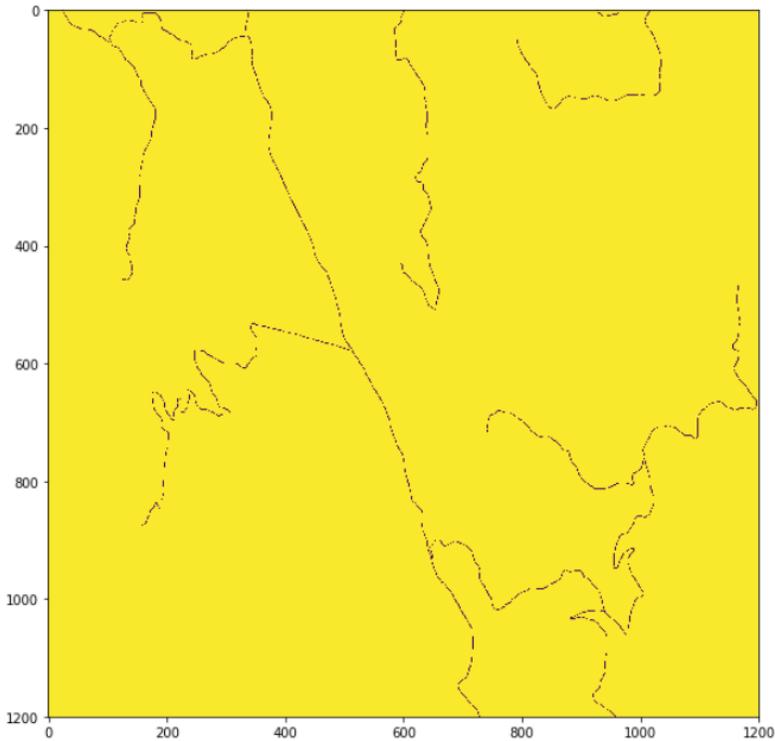


Gráfico: Shapefile de vías terciarias de la misma región.

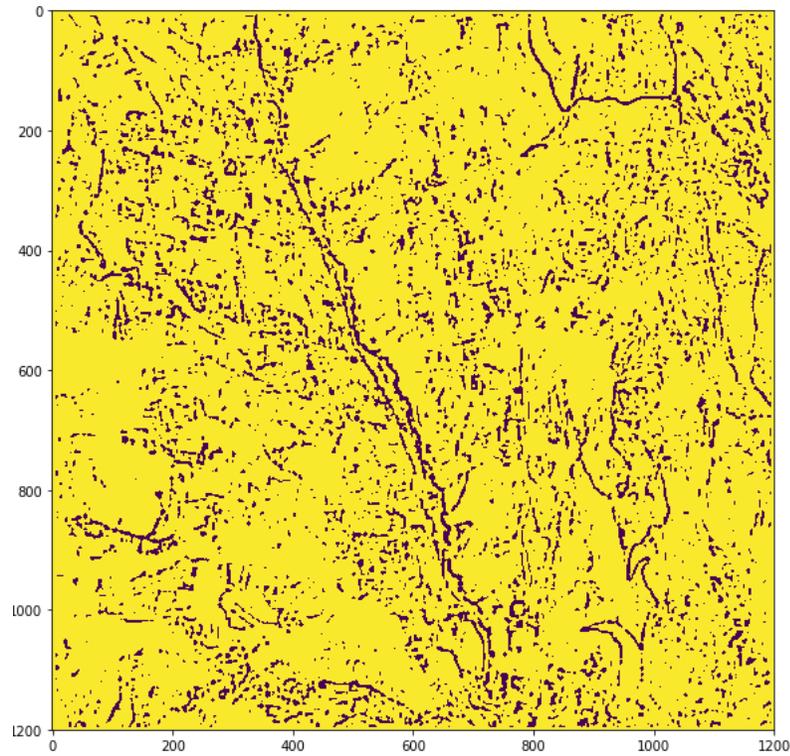
Resultados

Prueba del modelo sobre una imagen

a: Shapefile objetivo.



b: Predicción con umbral =



c: Predicción con umbral =

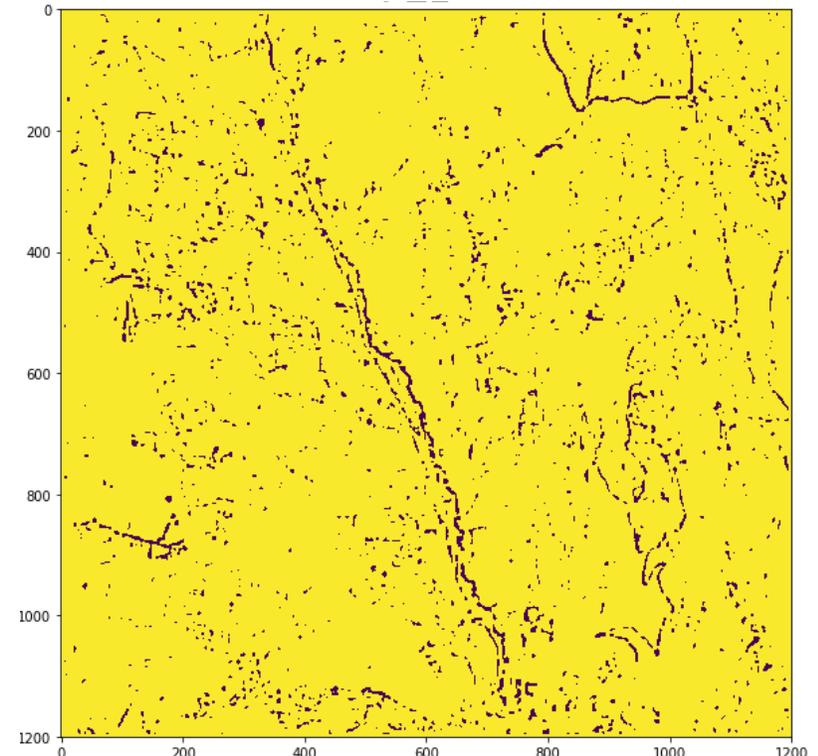


Gráfico: Shapefile de vías terciarias, región de Cerrito, Santander.