



Departamento Nacional de Planeación

[www.dnp.gov.co](http://www.dnp.gov.co)

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

**Unidad de Científicos de Datos**  
Director de Desarrollo Digital

Julio, 2018  
dnp.gov.co



AGENDA - Julio, 2018

1. Objetivo
2. Pre-procesamiento
3. Modelos y Resultados
4. Siguietes pasos



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



# Objetivo

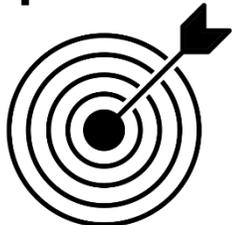
1

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



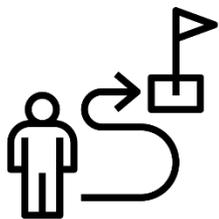
# Imágenes satelitales para identificación de vías

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



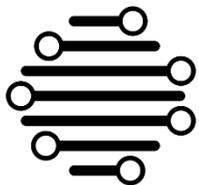
## Objetivo

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



## Insumos

*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

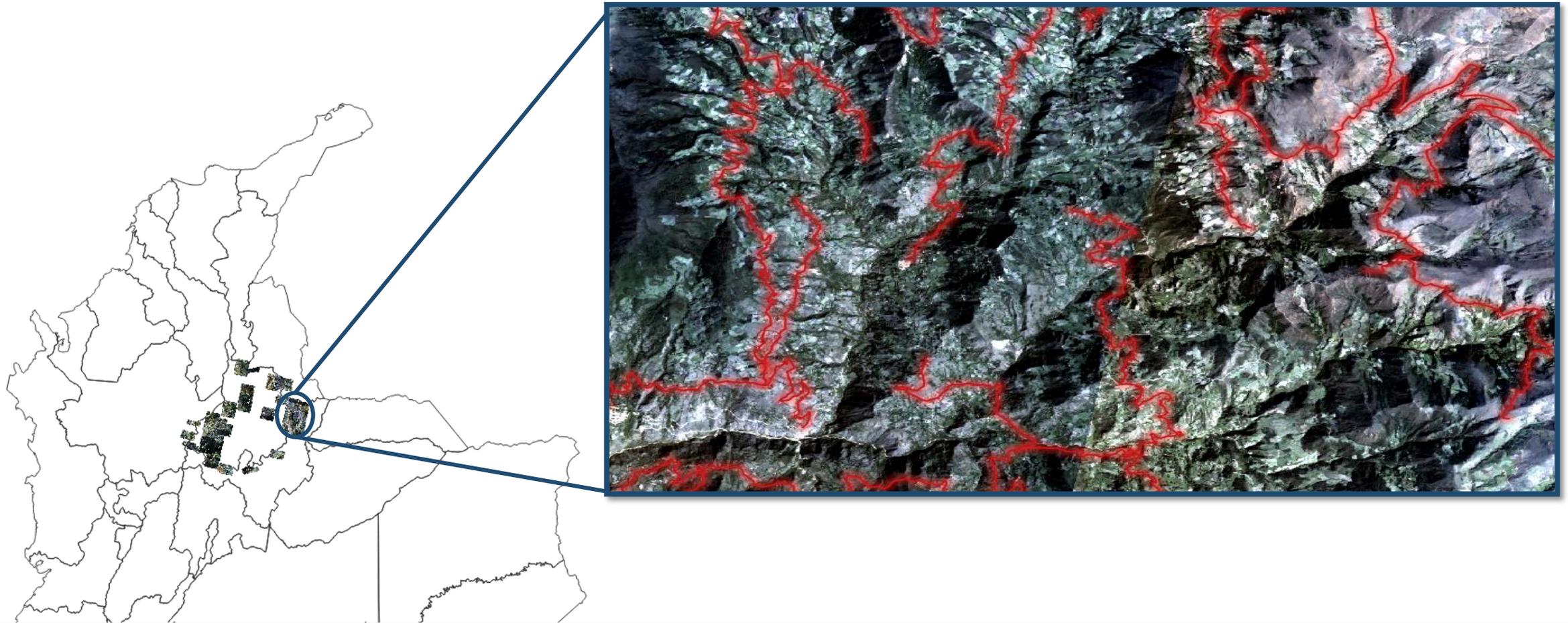
2

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



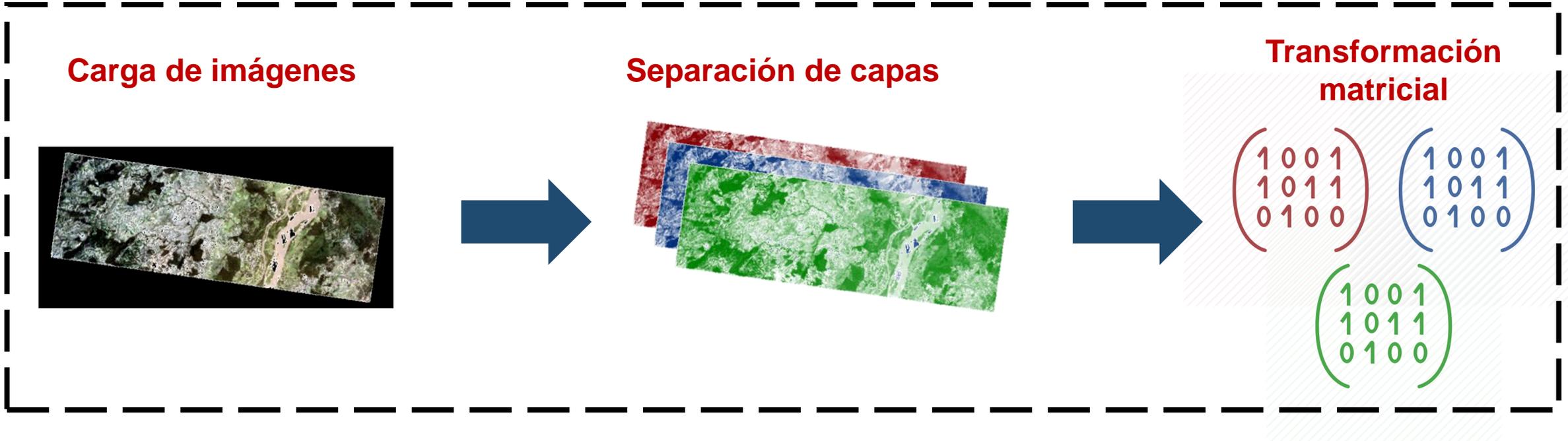
# Etiquetas para cada imagen satelital

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

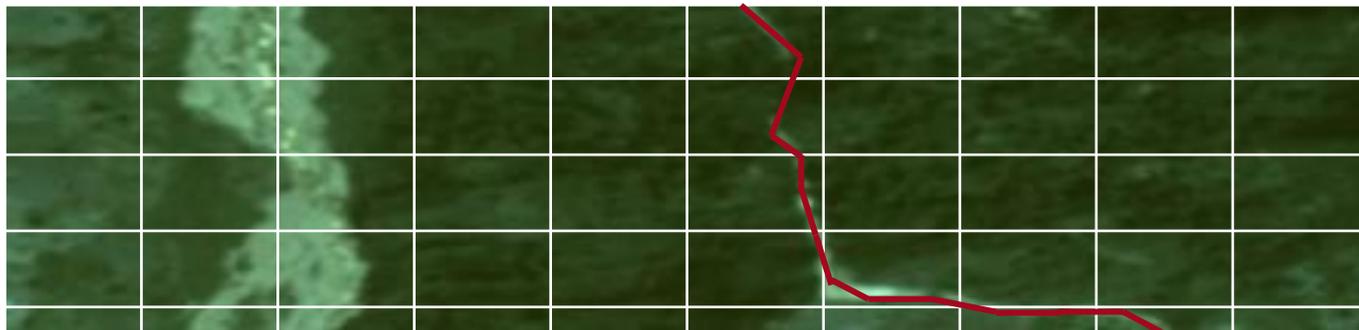
Cada imagen requiere separar sus capas de colores para luego construir matrices. Estas matrices son insumo fundamental para la creación de cualquier algoritmo.



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

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)

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

# Modelos y Resultados

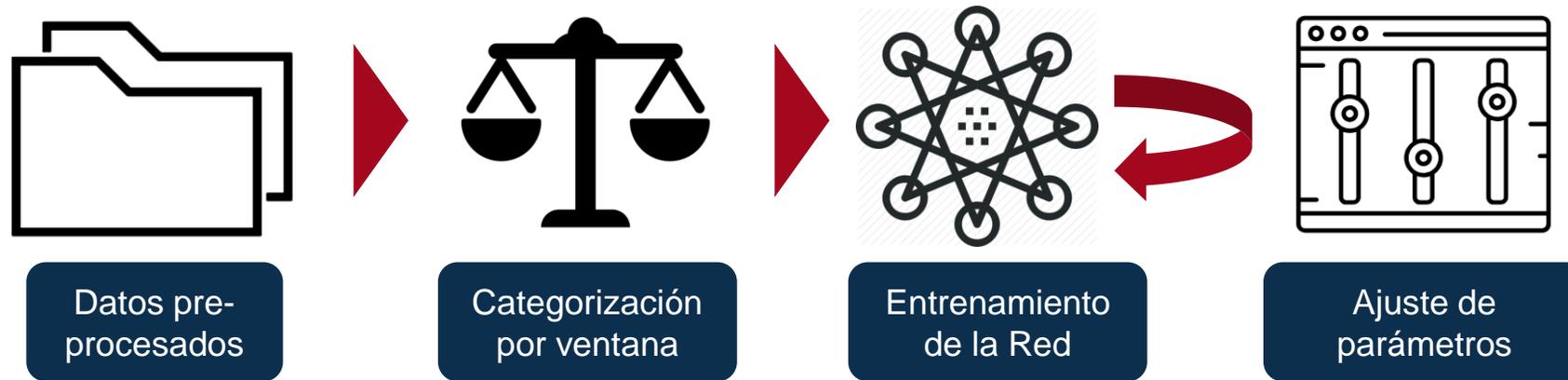
2

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



# Modelo de Redes Neuronales Convolutacional (CNN)

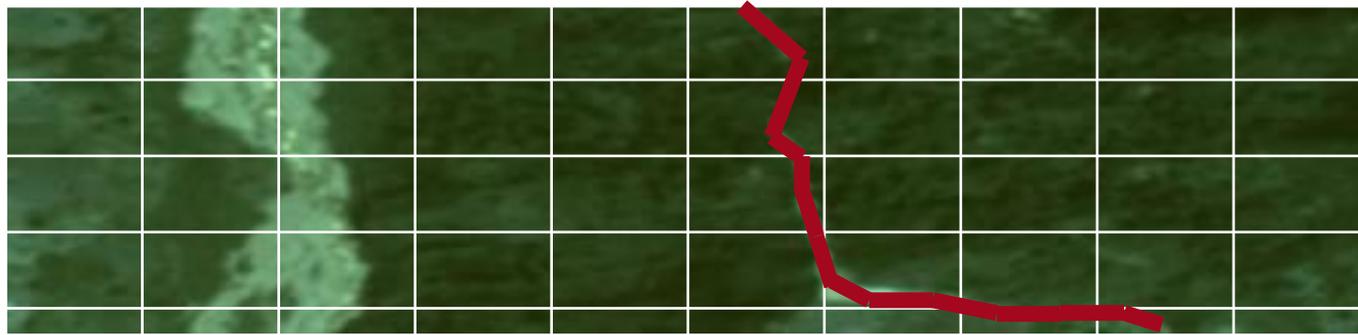
Para el desarrollo del algoritmo de clasificación, se utilizó el siguiente flujo de trabajo



Pregunta de clasificación: **Un segmento de la imagen, ¿contiene o no vías terciarias?**

# Red Neuronal Convolutiva (CNN)

Son redes comúnmente utilizadas para visión artificial en la clasificación y segmentación de imágenes



Algunos píxeles **contienen**  
etiqueta de vía:

**1**

Los píxeles **no contienen**  
etiqueta de vía:

**0**



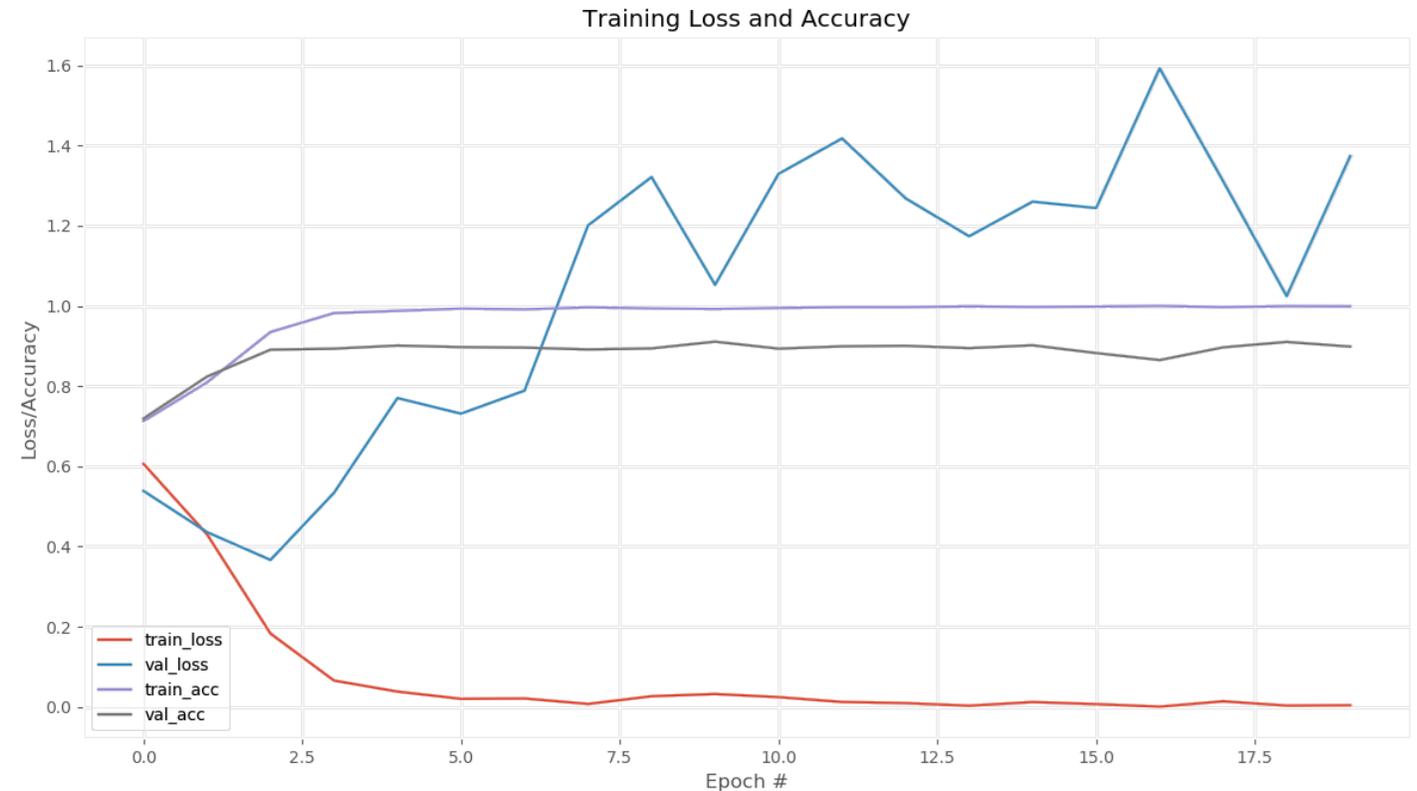
# Resultados de la red

Los resultados obtenidos para la red optimizada tanto con las imágenes a color, como las imágenes en escala de grises fueron:

Evaluación sobre 265 ventanas nuevas.

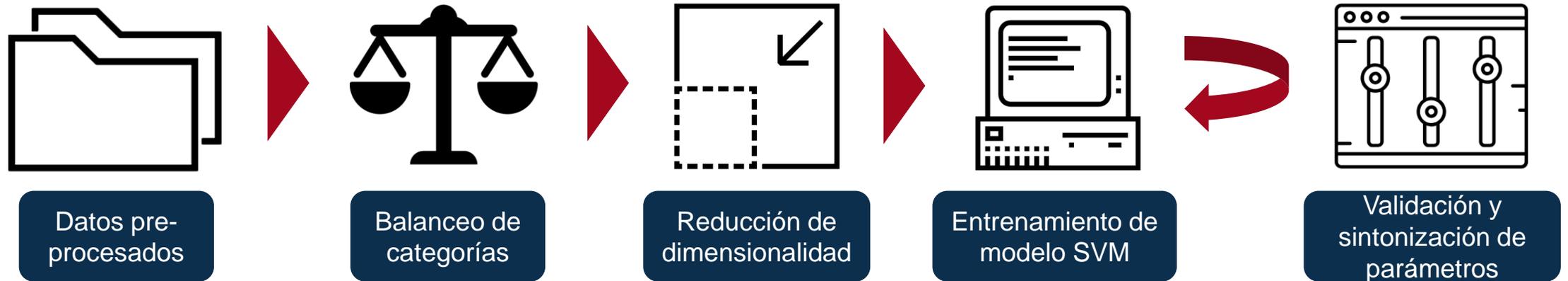
	No vías	Vías
No vías	116	74
Vías	47	28

**Porcentaje de Precisión: 54%**



# Modelo de Máquinas de Soporte Vectorial (SVM)

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

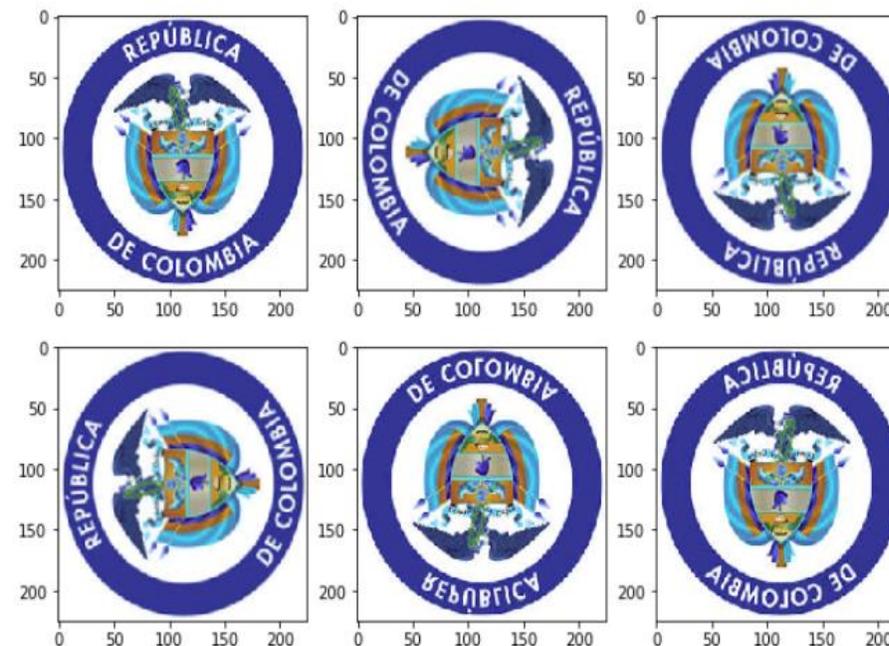


# Desarrollo de modelo SVM

## Balancedo 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).

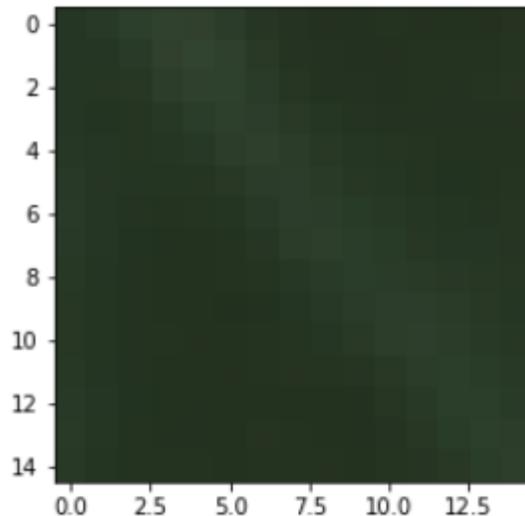


**Gráfico:** Ejemplo: Modificaciones para obtener 6 imágenes a partir de una sola.

# 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.



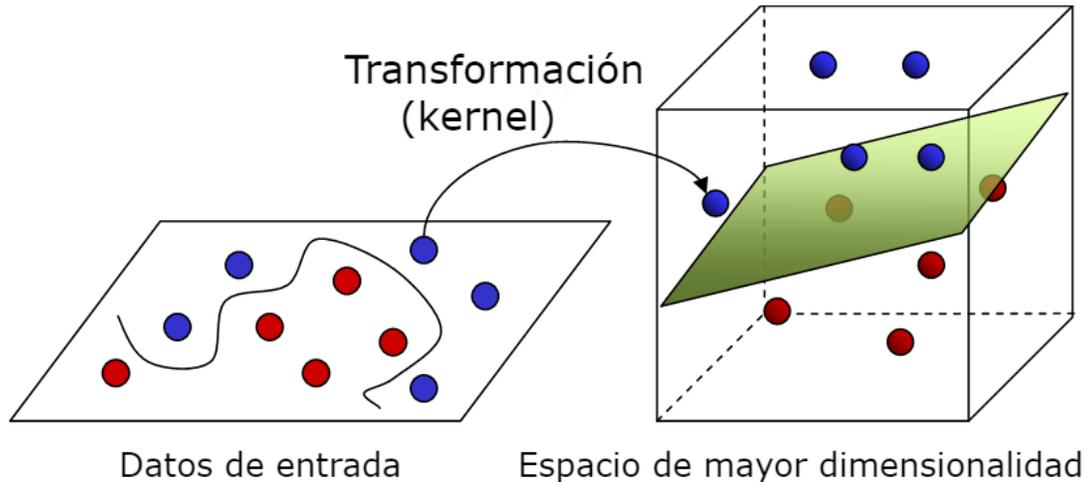
- Ventana de 15 pixeles en el espacio RGB  $\rightarrow 15*15*3 = 675$  variables.
- Utilizando PCA, el 97.1% de la varianza de los datos se puede representar con los primeros **18** componentes principales.

# Desarrollo de modelo SVM

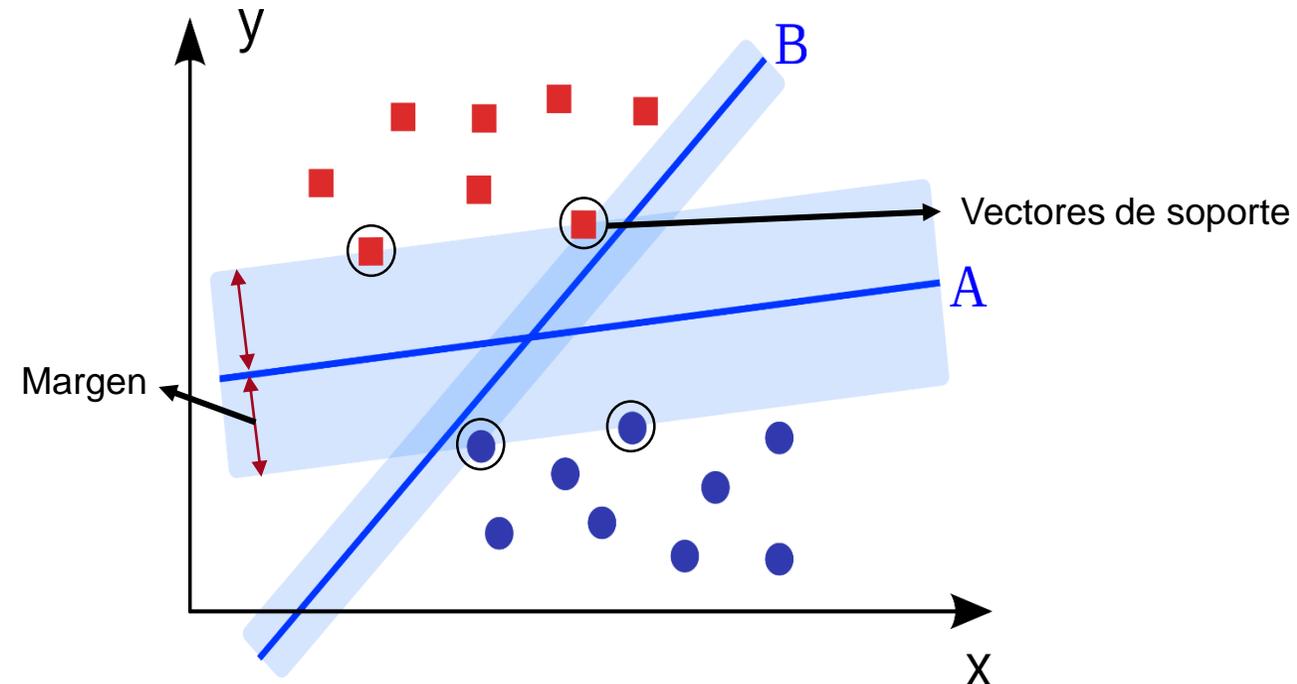
## Máquinas de soporte vectorial (SVM)

- 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.

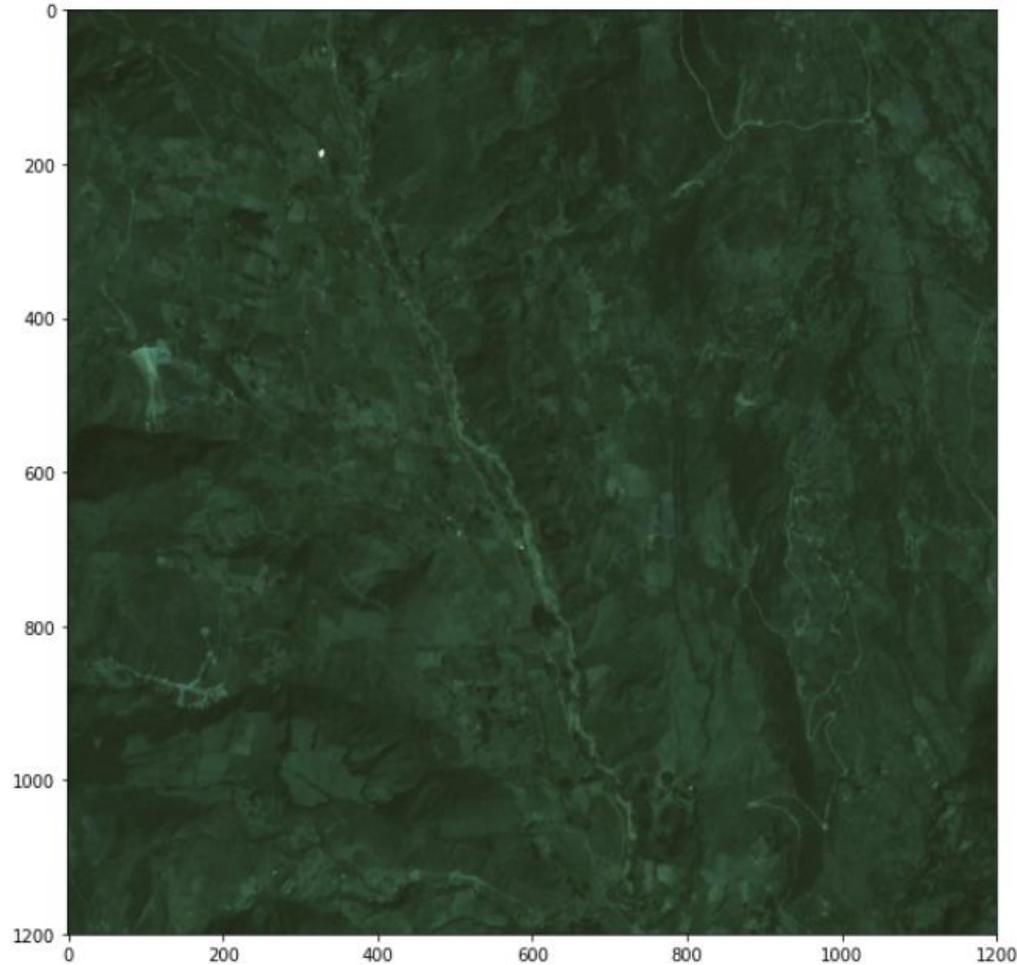
# Resultados

## Desempeño del modelo SVM

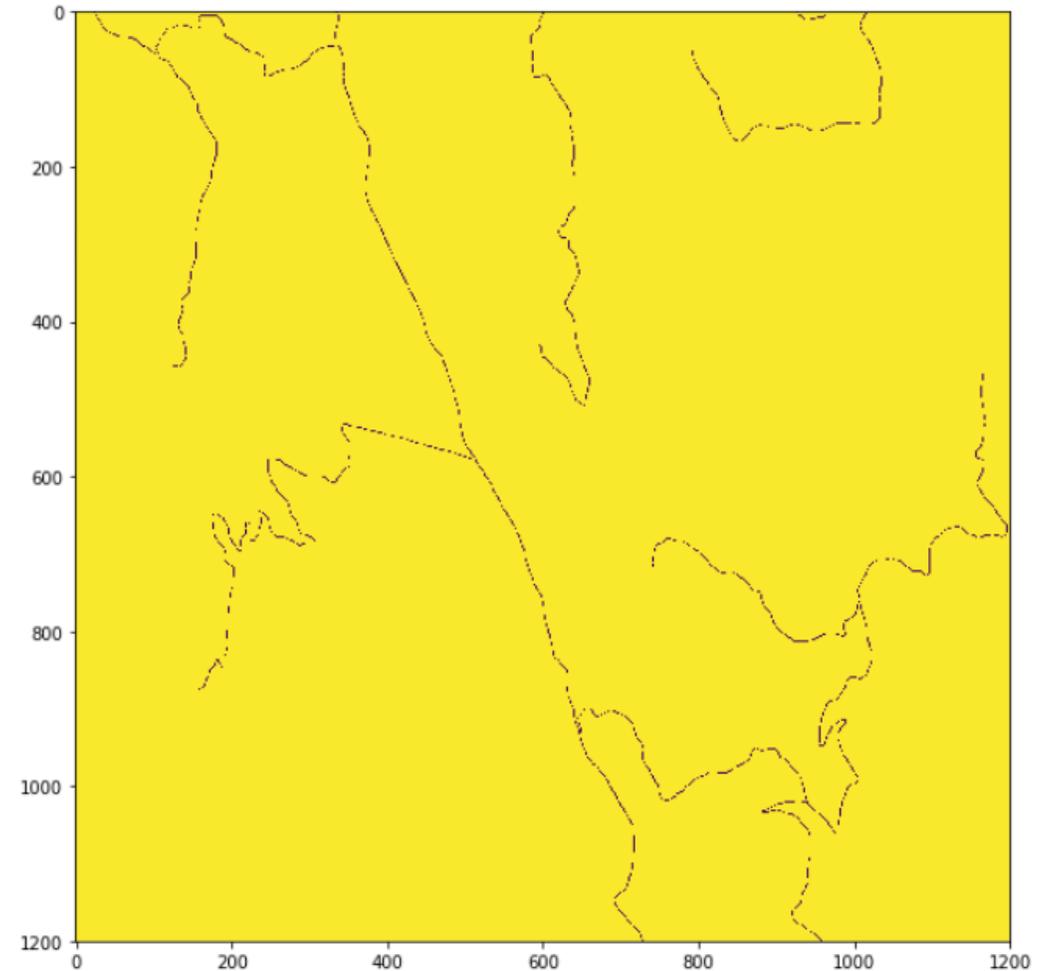
	Datos de entrenamiento			Datos de prueba		
Precisión	86,7%			86.5%		
Matriz de confusión	Real / Predicción	No vía	Vía	Real / Predicción	No vía	Vía
	No vía	34747	1382	No vía	11537	454
	Vía	6779	18445	Vía	2301	6159

# Resultados

## Prueba del modelo sobre una imagen



**Gráfico:** Imagen satelital - Cerrito, Santander.

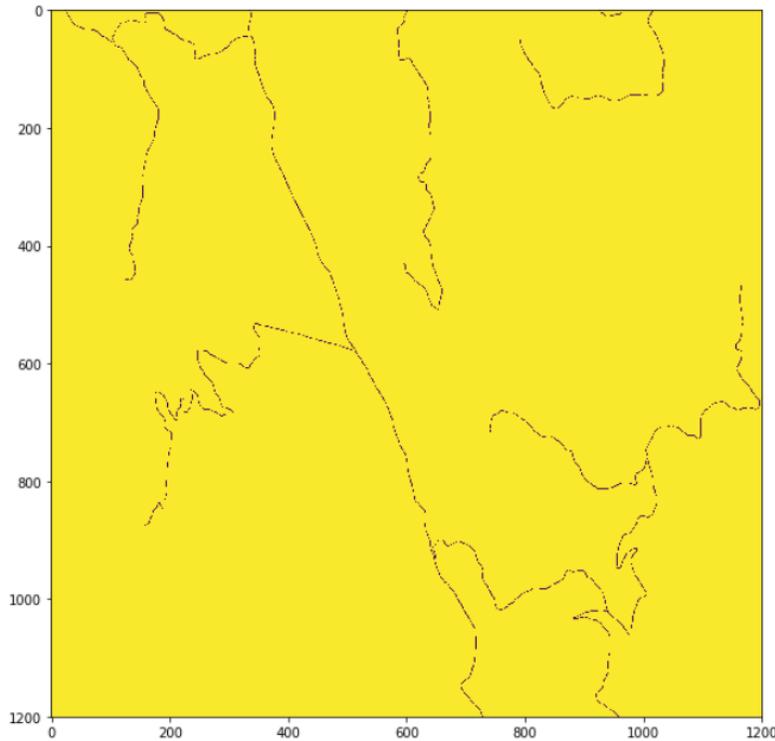


**Gráfico:** Shapefile de vías terciarias de la misma región.

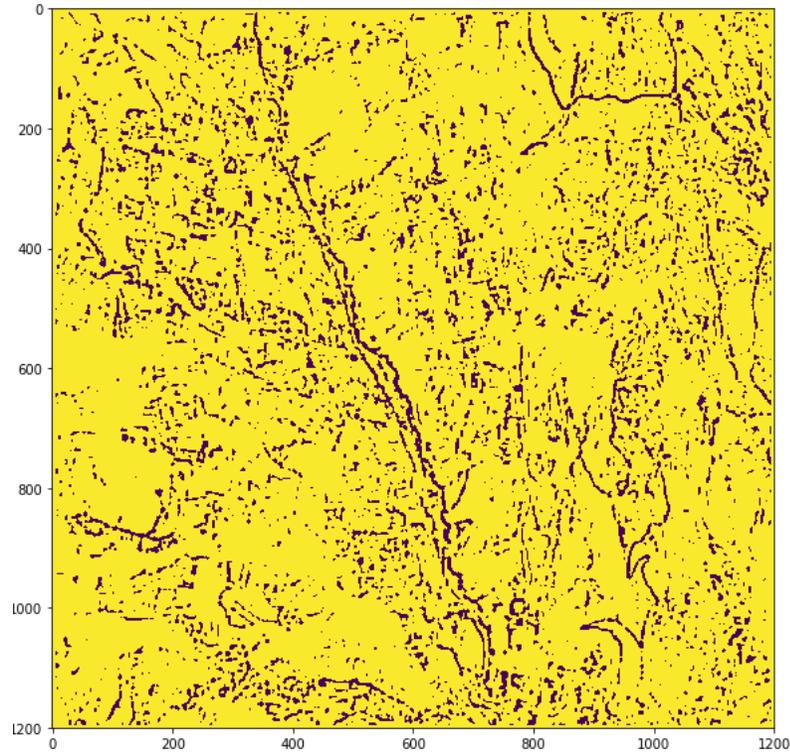
# Resultados

## Prueba del modelo sobre una imagen

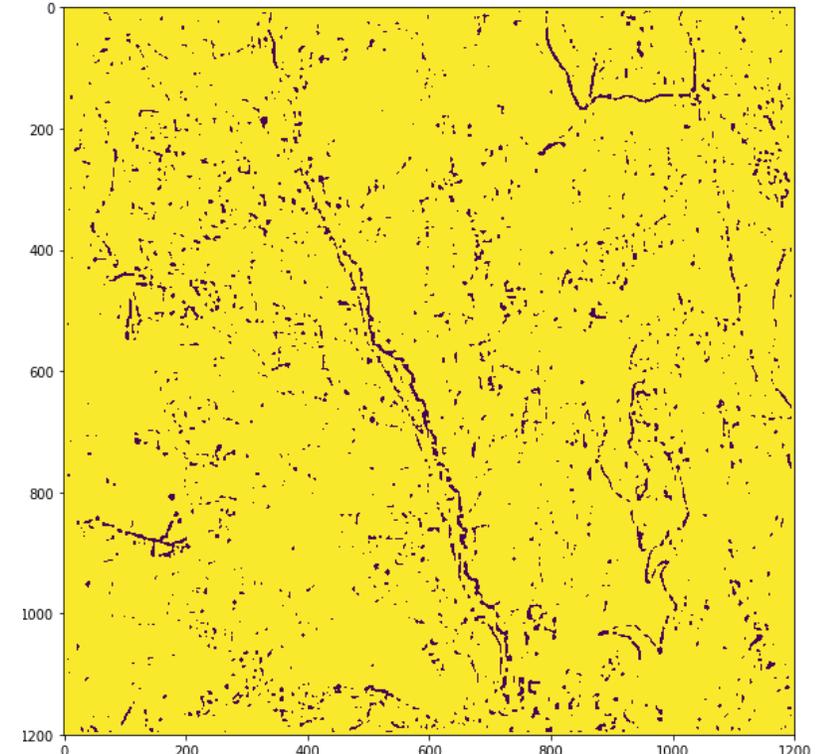
a: Shapefile objetivo.



b: Predicción con umbral = 0.5.



c: Predicción con umbral = 0.75.



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

# Siguientes pasos

4

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



# Conclusiones

Los resultados obtenidos a partir de los modelos desarrollados muestran un buen potencial del uso de estas técnicas para alcanzar el objetivo de identificar vías terciarias.

El desempeño de los modelos puede ser mejorado aún. Algunos factores que pueden mejorar el desempeño obtenido son:

- Utilizar más imágenes para entrenar los modelos.
- Hacer un mejor pre-procesamiento de las imágenes antes de entrenar los modelos.
- Hacer un mejor procesamiento de las imágenes resultantes, con el objetivo de dejar y unir las vías identificadas (grupos de puntos contiguos) y borrar el ruido (puntos dispersos en la imagen).
- Considerar el uso de otros modelos, haciendo énfasis en técnicas de aprendizaje profundo y *boosting*.

# Potenciales desafíos

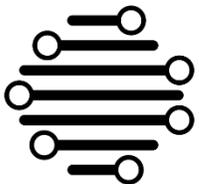
Para llevar el proyecto a una escala nacional, se evidencia la necesidad de contar con:



- Personal capacitado en análisis y manipulación de imágenes georreferenciadas.
- Personal con conocimientos en modelos de aprendizaje de máquina enfocado en visión artificial
- Personal especializado en procesamiento de imágenes



- Alta capacidad de procesamiento en CPU para ejecutar algoritmos sobre las imágenes nuevas
- Alta capacidad de procesamiento en GPU para profundizar en el desarrollo de algoritmos con redes neuronales convolucionales



- Aumentar la calidad de captura de las imágenes satelitales
- Aumentar el número de archivos SHP para otras partes del país que contengan características geográficas propias de cada región