

Data From The SKY

Proyecto Integrador

Johan Bejarano
Diego Arboleda
Jerónimo Gil

Universidad Nacional de Colombia sede Manizales

23 de julio de 2025

- La infraestructura urbana sufre deterioros invisibles a simple vista.
- La inspección manual de techos es costosa, riesgosa y poco escalable.
- Surge la necesidad de una herramienta automática.
- El presente proyecto aborda esta problemática desde el enfoque de visión computacional y analítica de datos.

Objetivo General

Diseñar un sistema automatizado para detectar y clasificar posibles daños en techos a partir de imágenes aéreas, empleando modelos de segmentación semántica y detección de objetos, complementado con análisis exploratorio y agrupamiento basado en métricas.

Estructura del Proyecto

- 1 Segmentación semántica con SegFormer.
- 2 Detección de objetos con YOLOv8.
- 3 Inferencia y generación de métricas por imagen.
- 4 Análisis exploratorio (EDA) y clustering.
- 5 Preparación de datos para visualización en dashboard interactivo.

- **YOLOv8:** arquitectura de detección de objetos altamente eficiente y precisa.
- **SegFormer:** modelo transformer de segmentación semántica adaptable y liviano.
- **Aplicaciones:** inspección de infraestructura, agricultura de precisión, monitoreo ambiental.

Metodología - Inferencia y Métricas (Notebook 2)

- Se procesaron **481 imágenes del conjunto** train del dataset personalizado.
- Se aplicó un modelo YOLOv8 previamente entrenado.
- Se generaron imágenes renderizadas con predicciones.
- Se extrajeron métricas por imagen: cantidad de daños, densidad, confianza media y máxima.

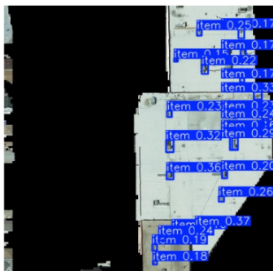


Figura: predicción YOLOv8 con daño detectado

Metodología - EDA y Clustering (Notebook 4)

- Consolidación y validación de 481 archivos *_metrics.csv.
- Visualización de distribuciones y relaciones entre métricas.
- Clustering con KMeans ($k=3$) usando variables estandarizadas.
- Visualización 2D con PCA para explorar patrones entre grupos.

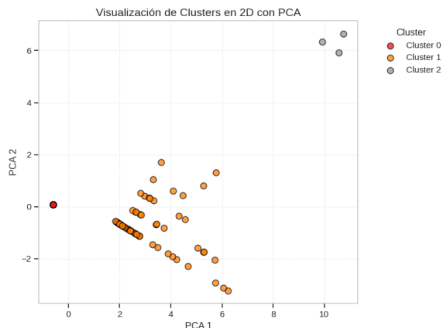


Figura: gráfico PCA de clusters

Preparación de Datos para Dashboard

- Exportación de:
 - `df_metrics_clustered.csv` / `.pkl`
 - `cluster_summary.csv` / `.json`
 - `cluster_representatives.json`
 - `metadata.json` y `log.txt`
- Archivos organizados en carpeta `dashboard_data/` y comprimida como `metrics_data.zip`
- Subida a Hugging Face:
<https://huggingface.co/jobejaranom/yolo-roof-damage>

- Estandarización de nombres de imágenes entre métricas y renderizados.
- Control de errores y robustez del sistema ante archivos faltantes.
- Optimización del pipeline de inferencia masiva.
- Integración modular y automatizada para futuros dashboards.

Impacto y Viabilidad del Proyecto

- Potencial uso en planes de mantenimiento predictivo para infraestructura urbana.
- Reducción de tiempos, costos y riesgos en inspecciones manuales.
- Fácil integración con plataformas de visualización (e.g., Dash, Gradio).
- Modularidad para ampliación a nuevas ciudades o clases de daño.

- Se construyó un sistema robusto de análisis e inferencia sobre techos urbanos.
- Se logró caracterizar el conjunto de imágenes mediante clustering no supervisado.
- Se dejaron todos los insumos listos para el despliegue en un dashboard profesional.

Siguientes Pasos

- Desarrollo del dashboard interactivo con filtros por cluster y métricas.
- Inclusión de segmentación semántica con SegFormer como módulo complementario.
- Posible integración de un sistema de alerta automática en función de severidad.

Gracias por su atención