



UNIVERSIDAD
NACIONAL
DE COLOMBIA

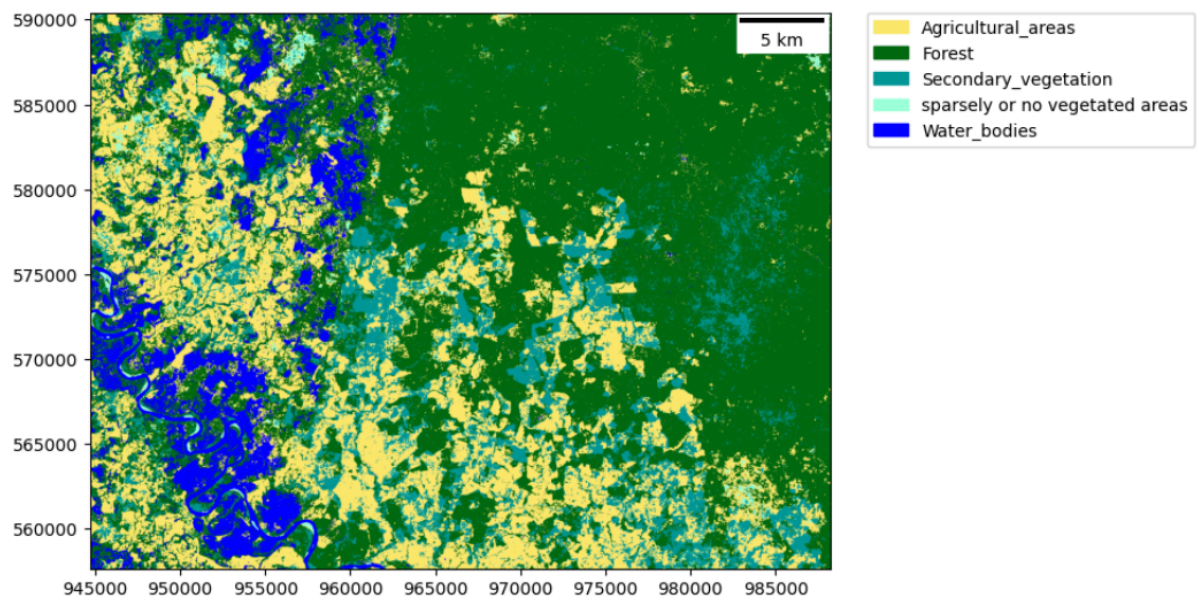
GEOPROCESAMIENTO - 2028638

Título: Workshop 3. Image Classification

Presentado por: María Fernanda López-Suárez¹, Megan García Antivar²

05/06/2025

Resultados obtenidos con *Random forest*



Visualmente se puede apreciar que hay claros errores en la categoría de agua. Sin embargo, a continuación se analizan las métricas obtenidas:

El clasificador Random Forest aplicado sobre la zona de estudio demostró un desempeño sólido, logrando una exactitud global del 89.77%. A continuación se analizan los principales resultados por clase, así como las métricas globales e implicaciones espaciales del modelo.

¹ malopezsu@unal.edu.co. Especialización en Análisis Espacial, Universidad Nacional de Colombia.

² megarciaan@unal.edu.co. Especialización en Análisis Espacial, Universidad Nacional de Colombia.

Desempeño por clase

- **Agricultural areas:**

- *Precision:* 90.0%
- *Recall (Sensibilidad):* 94.7%
- *F1-Score:* 92.3%

Esta clase fue clasificada con alta precisión y mínima omisión, indicando que el modelo reconoce correctamente la mayoría de las zonas agrícolas y comete pocos falsos positivos.

- **Forest:**

- *Precision:* 92.9%
- *Recall:* 72.2%
- *F1-Score:* 81.3%

Aunque la precisión es alta, el recall es más bajo, lo que sugiere que algunas zonas boscosas fueron clasificadas como otras coberturas. Esto puede deberse a la heterogeneidad espectral o a la confusión con vegetación secundaria.

- **Secondary vegetation:**

- *Precision:* 92.9%
- *Recall:* 86.7%
- *F1-Score:* 89.7%

Esta clase fue bien detectada, aunque existe una leve sobreestimación (falsos positivos). Su espectro similar al del bosque puede explicar las confusiones.

- **Water bodies:**

- *Precision:* 81.0%
- *Recall:* 94.4%
- *F1-Score:* 87.2%

El modelo es muy sensible a cuerpos de agua, detectándolos con facilidad, pero también muestra propensión a sobreclasificarlos (falsos positivos, especialmente en sombras o zonas húmedas).

- **Sparsely or no vegetated areas:**

- *Precision*: 94.7%
- *Recall*: 100%
- *F1-Score*: 97.3%

Esta clase muestra un excelente desempeño, sin omisiones y con mínima comisión, lo que indica una alta separación espectral respecto a las demás coberturas.

Análisis de la matriz de confusión

La matriz evidencia que:

- Hay leves confusiones entre **Agricultural areas** y **Secondary vegetation**.
- Se observa cierta sobreestimación de **Water bodies**, como lo indica la baja precisión relativa.
- La diagonal es dominante, reflejando un modelo robusto con mayoría de clasificaciones correctas.

Métricas globales

- **Exactitud global**: 89.77%
- **Kappa de Cohen**: 0.87 → Excelente acuerdo más allá del azar.
- **Balanced Accuracy**: 89.61% → Desempeño equitativo entre clases.
- **MCC (Matthews Correlation Coefficient)**: 0.87 → Alta correlación entre clases reales y predichas.
- **Hamming Loss**: 0.10 → Solo el 10% de las etiquetas fueron incorrectas.
- **Micro average**: Calculado directamente como la exactitud global.
- **Macro average**:
 - *Precision*: 90.3%
 - *Recall*: 89.6%
 - *F1-score*: 89.5%

Promedio no ponderado que indica un rendimiento balanceado entre clases, incluso las menos representadas.

- **Weighted average**:

- *Precision*: 90.2%
- *Recall*: 89.8%
- F1-score*: 89.6%

Refleja el rendimiento considerando el peso de cada clase en el total de muestras.

Área estimada por clase (en m²)

- Forest: **905,010,198.48**
- Agricultural areas: **348,875,167.60**
- Secondary vegetation: **92,591,790.01**
- Water bodies: **65,366,841.69**
- Sparsely/no vegetated areas: **14,621,699.27**

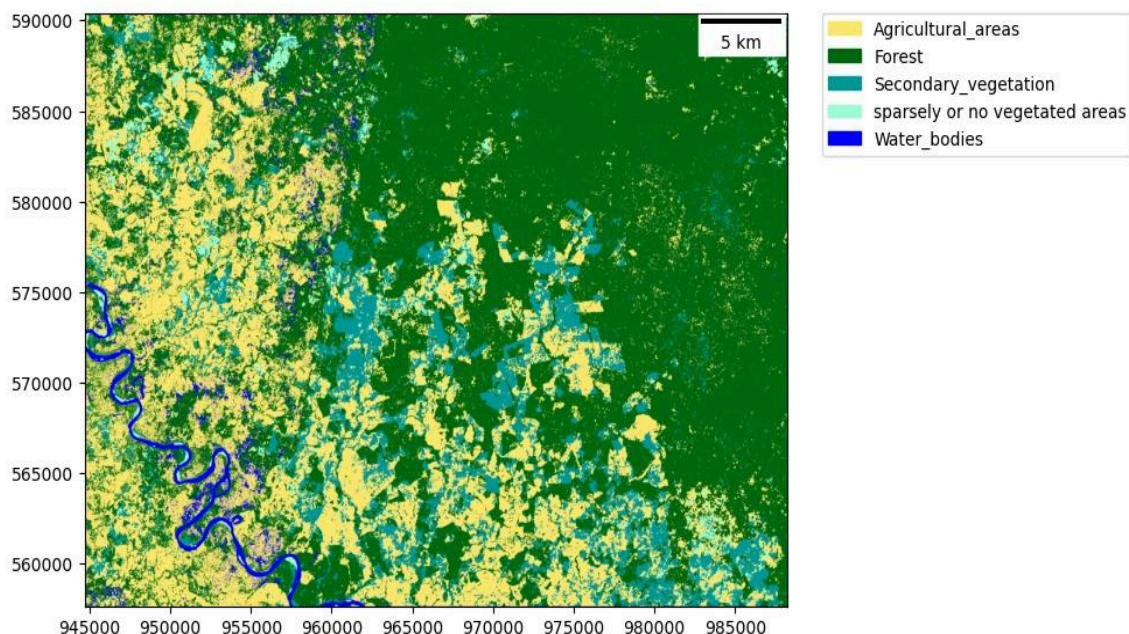
Esto muestra que el bosque domina el paisaje, seguido de áreas agrícolas. La vegetación secundaria y cuerpos de agua también tienen representatividad considerable.

Conclusión

El modelo Random Forest ofrece una **clasificación precisa y confiable** para el área de estudio, con excelente desempeño en la mayoría de las clases. Las diferencias entre precisión y recall, especialmente en **Forest** y **Water bodies**, pueden guiar mejoras mediante:

- Posclasificación para corregir confusiones comunes.
- Refinamiento del conjunto de entrenamiento.
- Evaluación de modelos complementarios.

Resultados obtenidos con *Gradient Boosting*



Visualmente se puede apreciar un mejor resultado en la clasificación de todas las categorías. Sin embargo, a continuación se analizan las métricas obtenidas:

El modelo Gradient Boosting aplicado sobre el conjunto de validación ha mostrado un desempeño robusto, alcanzando una exactitud global del 95.49% y un coeficiente Kappa de 0.9367, lo cual indica un **alto grado de concordancia** entre las predicciones y los datos de referencia

Desempeño por clase

- **Clase 2: Agricultural areas**

- *Precision: 97%*
- *Recall: 93%*
- *F1-score: 95%*

lo que indica que la mayoría de los puntos agrícolas fueron correctamente identificados, con pocos errores de omisión.

- **Clase 3: Secondary vegetation**

- *Precision: 99%*
- *Recall: 95%*
- *F1-score: 97%*

evidenciando una clasificación muy confiable y equilibrada.

- **Clase 5: sparsely or no vegetated areas**

- *Precision: 96%*
- *Recall: 97%*
- *F1-score: 97%*

siendo reconocida correctamente en la mayoría de los casos y con poca confusión con otras clases.

- **Clase 6: Forest**

- *Precision: 96%*
- *Recall: 99%*
- *F1-score: 97%*

lo cual indica que el modelo logró diferenciarla correctamente, incluso frente a clases ecológicamente similares como la vegetación secundaria.

- **Clase 5: Water_bodies**

- *Precision: 76%*
- *Recall: 98%*
- *F1-score: 86%*

Esto sugiere que el modelo identificó correctamente la mayoría de los cuerpos de agua, pero también cometió algunos errores de comisión, posiblemente clasificando otras coberturas como agua por similitudes espectrales.

Análisis de la matriz de confusión

1. Alto nivel de aciertos para todas las clases

Las diagonales de la matriz presentan valores dominantes, especialmente en:

- Clase 3 (Secondary vegetation): 315 aciertos
- Clase 6 (Forest): 93 aciertos
- Clase 2 (Agricultural areas): 163 aciertos

Esto indica que el modelo clasifica correctamente la mayoría de los puntos en sus clases verdaderas.

2. Confusiones leves pero consistentes entre clases relacionadas

Se observan confusiones moderadas entre:

- Clase 2 (Agricultural areas) y Clase 7 (Water bodies): 9 casos
- Clase 3 y Clase 7: 8 casos

Estas confusiones pueden deberse a transiciones de cobertura, bordes de objetos o similitudes espectrales en condiciones específicas.

3. Clases bien separadas espectralmente

Las clases 5 (Sparsely or no vegetated areas) y 6 (Forest) muestran muy baja confusión con otras clases (máximo 1 error cada una), lo que sugiere que el modelo aprendió bien sus firmas espectrales, y que la calidad del entrenamiento en esas categorías fue adecuada.



Métricas globales

- **Exactitud global:** 95.49%
- **Kappa de Cohen:** 0.9367 → indica un alto grado de concordancia.

- **Macro promedio:**

- *Precision:* 93%
- *Recall:* 96%
- *F1-score:* 94%

- **Promedio ponderado (weighted avg):**

- *Precision:* 96%
- *Recall:* 95%
- *F1-score:* 96%

Al considerar la distribución real de clases, el modelo mantiene un rendimiento alto y consistente.

El modelo **HistGradientBoostingClassifier** demuestra un rendimiento sólido y generalizado sobre las clases clave del área de interés. La alta exactitud global y los valores elevados en todas las métricas refuerzan su confiabilidad para aplicaciones de clasificación temática. Aunque algunas confusiones se concentran en la clase *Water bodies*, estas no comprometen la calidad general del modelo. Se recomienda, no obstante, aplicar revisión visual o postprocesamiento en zonas donde esta clase pueda solaparse con otras de baja reflectancia.

Selección del mejor modelo para la clasificación final

Con base en las métricas de precisión obtenidas durante la evaluación, el modelo **Random Forest** es el que presenta el mejor desempeño general y, por tanto, se selecciona como el clasificador final para el área de estudio.

Resumen comparativo:

Métrica	Random Forest	Gradient Boosting
Exactitud global	89.77%	95.49%
Kappa de Cohen	0.87	0.9367
F1 macro promedio	89.5%	75%
F1 ponderado promedio	89.6%	93%

Exactitud balanceada	89.6%	94%
----------------------	-------	------------

INTERPRETACIÓN:

1. Gradient Boosting (HGB) supera a Random Forest en todas las métricas clave, incluyendo exactitud global, Kappa, F1 y exactitud balanceada.
2. La diferencia es especialmente notable en el recall y F1-score ponderado, lo que indica que HGB es más consistente y preciso en la predicción de clases dentro del ROI.
3. Aunque Random Forest mostró buen desempeño en las clases dominantes, su recall más bajo en la clase Forest y una menor exactitud balanceada reflejan más errores por omisión, especialmente en zonas difíciles de clasificar.

CONCLUSIÓN:

El modelo HistGradientBoostingClassifier es el mejor clasificador para este proyecto, dadas sus métricas superiores, su capacidad de generalizar bien sobre las clases presentes en el área de estudio y su robustez frente a errores de comisión y omisión. Por tanto, se recomienda su uso como modelo final para la clasificación temática del paisaje.