

MAPEO DE ÁREAS INUNDABLES MEDIANTE EL USO DE MACHINE LEARNING EN UNA CUENCA DE LLANURA DEL NORESTE DE LA PROVINCIA DE BUENOS AIRES

Guido Borzi^{1,2}, Carolina Pedraza De Marchi^{1,2}, Francisco Aldo Cellone^{1,3}, Eleonora Carol^{1,2}, Orlando Mauricio Quiroz Londoño^{1,4}

1-Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET

2-Centro de Investigaciones Geológicas, CONICET-UNLP, Argentina

3-CIM, Centro de Investigaciones del Medio Ambiente, CONICET-UNLP, ARGENTINA

4-IGCyC, Instituto de Geología de Costas y del Cuaternario "Dr. Enrique Jorge Schnack", Facultad de Ciencias Exactas y Naturales, UNMdP - CICPBA .Instituto de Geología de Costas y del Cuaternario, CONICET - UNMdP, ARGENTINA
gborzi@fcnym.unlp.edu.ar

Introducción

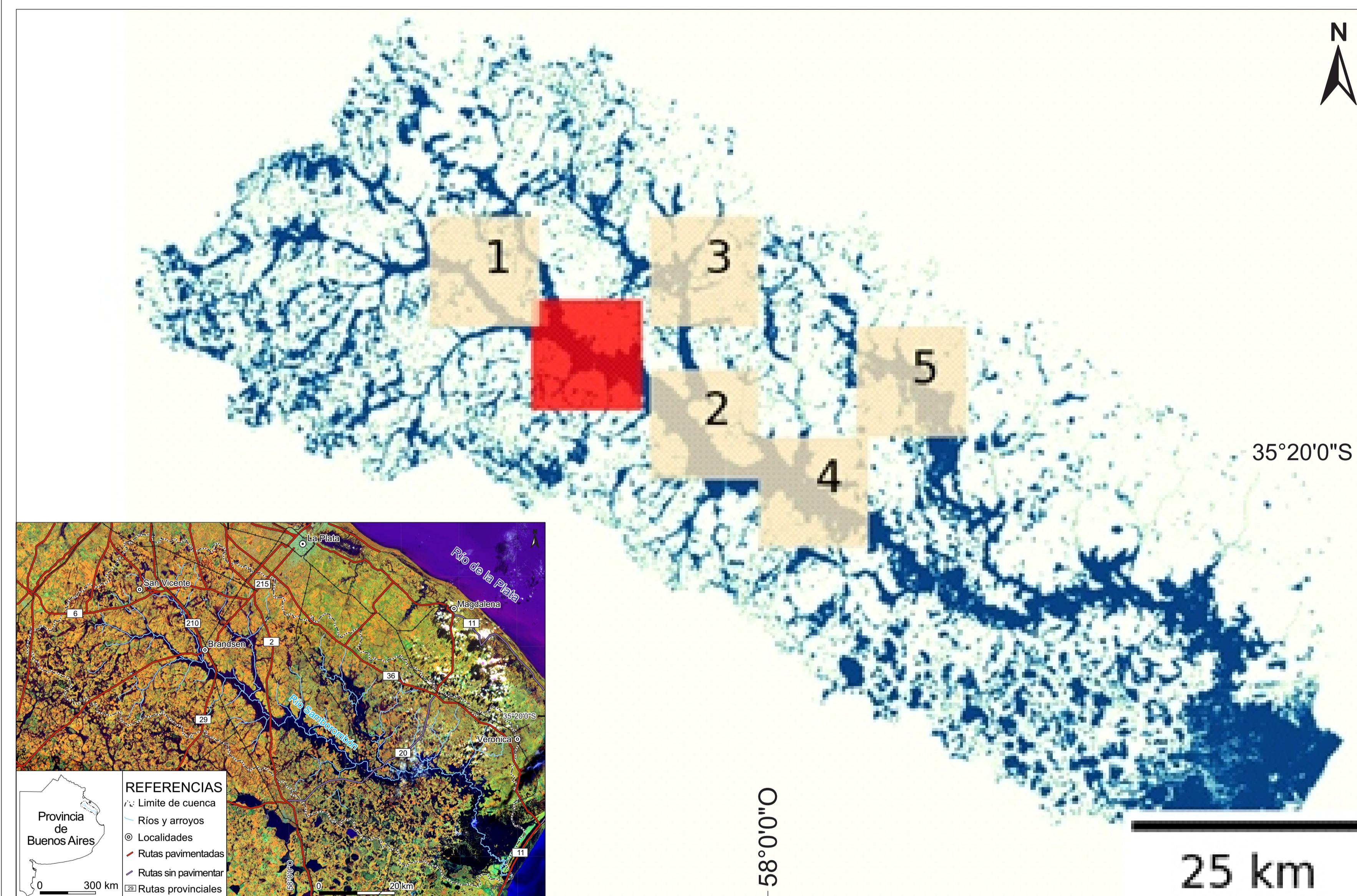
Las inundaciones son uno de los fenómenos naturales que afectan de manera recurrente las actividades humanas. La expansión y el crecimiento urbano actual con planificación que no contempla la zonificación de áreas inundables en su planificación conduce a agravar esta situación. La cuenca del río Samborombón, localizada en el noreste de Buenos Aires, se encuentra en este contexto, por lo que identificar áreas con peligro de inundación actualmente es una tarea prioritaria en sectores ya urbanizados. Esta tarea puede demandar mucho tiempo, por lo que desarrollar una metodología que identifique áreas de riesgo de manera rápida y eficiente sería de gran utilidad para la gestión.

Metodología

El uso de técnicas de inteligencia artificial, como el aprendizaje automático (Machine Learning), se destaca por procesar grandes volúmenes de datos de manera más rápida y precisa. La utilización de esta herramienta con un reducido número de capas derivadas de un sistema de información geográfica, proporciona resultados confiables en la predicción de áreas en riesgo de inundación. Para aplicar esta metodología se utilizaron las capas de pendiente, permeabilidad y espesor de zona no saturada y se tuvo en cuenta las subcuencas, donde aquellas que colectan más agua fueron consideradas con mayor potencial de inundación. Los algoritmos utilizados para aprendizaje automático fueron Random forest (RF), Regresión Logística (LR) y Vecino natural (KNN), entrenando los modelos en un área y observando los resultados de las predicciones de los modelos en 5 áreas diferentes.

Resultados

En líneas generales, los resultados obtenidos evidencian una gran correspondencia con el área inundada utilizada para calibración, la cual pertenece a una inundación registrada por el satélite Landsat para mayo del año 2000.



Una de las métricas seleccionadas para evaluar los modelos fue la matriz de confusión, donde se resume el número de predicciones correctas para cada algoritmo utilizado. Los campos de la misma son, Verdadero Positivo (TP) cuando el resultado del modelo predice correctamente la clase positiva (área inundada), Verdadero Negativo (TN) cuando el resultado del modelo predice correctamente la clase negativa (área no inundada), Falso Positivo (FP), donde el resultado del modelo predice incorrectamente la clase positiva y Falso Negativo (FN), cuando el resultado del modelo predice incorrectamente la clase negativa. A continuación se pueden observar las matrices de confusión de prueba obtenidas para cada modelo.

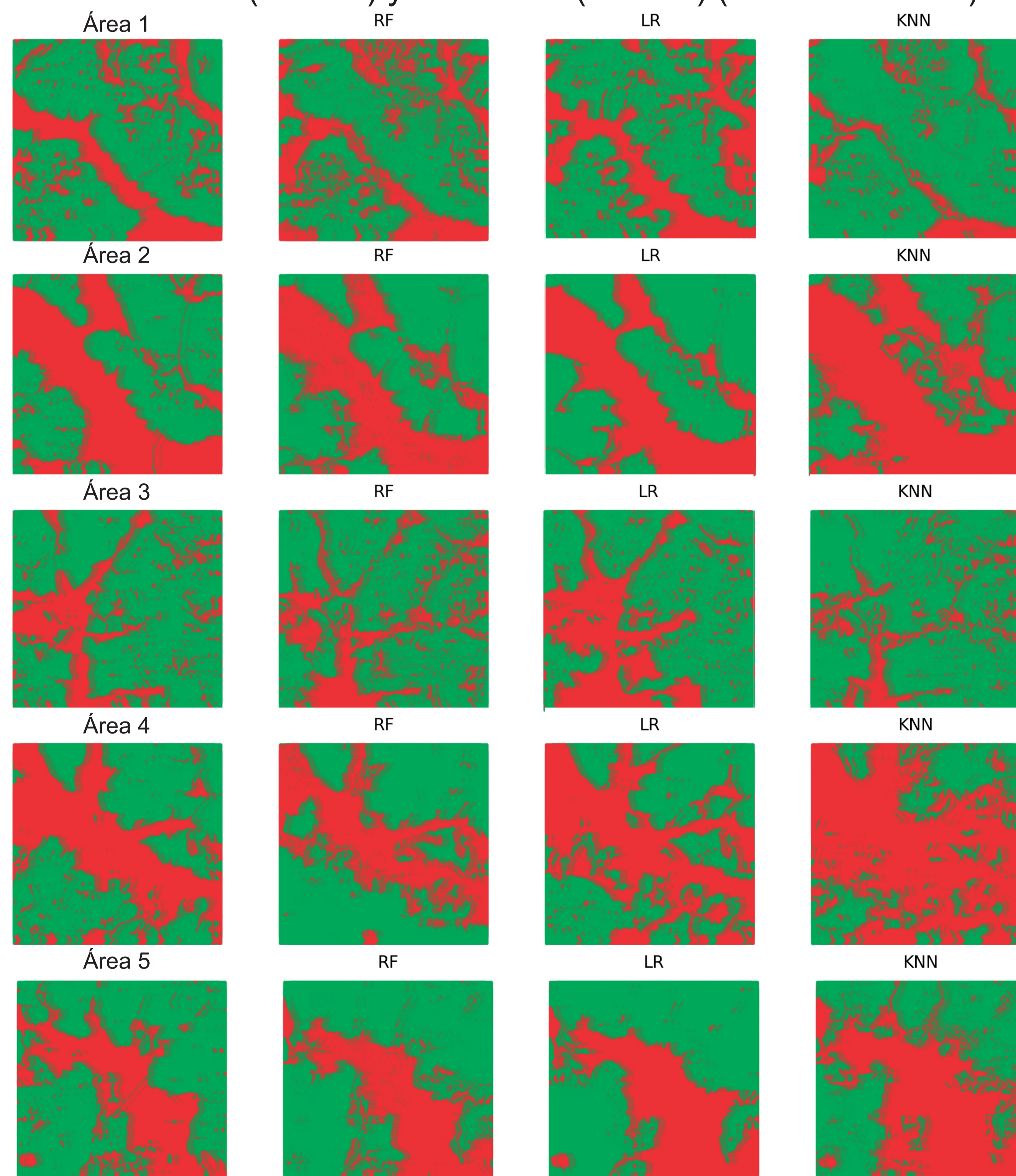
		Agua	Sin agua		
Valores de predicción	Agua	PV	PF	Agua	
	Sin agua	NF	NV		
				Valores reales	

Matriz confusión RF	Matriz confusión LR	Matriz confusión KNN												
<table border="1"> <tr><td>38270</td><td>2175</td></tr> <tr><td>3149</td><td>20406</td></tr> </table>	38270	2175	3149	20406	<table border="1"> <tr><td>38038</td><td>2407</td></tr> <tr><td>5305</td><td>18250</td></tr> </table>	38038	2407	5305	18250	<table border="1"> <tr><td>38118</td><td>2327</td></tr> <tr><td>3187</td><td>20368</td></tr> </table>	38118	2327	3187	20368
38270	2175													
3149	20406													
38038	2407													
5305	18250													
38118	2327													
3187	20368													

De los modelos utilizados, se observa que el modelo RF presenta la mejor correspondencia en la totalidad de las áreas de validación, como también lo muestra la tabla con los resultados de precisión y exactitud de los datos de entrenamiento y prueba. Donde precisión es el valor predictivo positivo o proporción de instancias relevantes entre las instancias recuperadas y exactitud es la proporción de predicciones que el modelo clasificó correctamente.

Datos	Modelo	Presición	Exactitud
Entrenamiento	RF	99.9	99.9
Prueba	RF	87.84	87.82
Entrenamiento	LR	87.84	87.82
Prueba	LR	87.98	87.88
Entrenamiento	KNN	93.78	93.79
Prueba	KNN	91.35	91.38

Precision = TP/(TP+FP) y Exactitud=(TP+TN)/(TP+TN+FP+FN).



Conclusiones

La aplicación de estas técnicas en una cuenca de llanura, como la estudiada, muestra una alta correspondencia con un escenario real. Estos resultados sugieren que el uso de algoritmos de aprendizaje automático para identificar inundaciones en una cuenca de llanura, incluso con un número limitado de capas de datos, produce predicciones confiables. Una posible utilidad de estos modelos es la identificación de zonas inundables en otras cuencas de llanura similares que enfrentan limitaciones en las mediciones utilizadas para el mapeo convencional.