

Evaluación de la capacidad predictiva de variables de texturas de imagen de resonancia magnética en cáncer de endometrio

Javier García Gilabert¹ & Pablo Gil Chong¹ & Aida Villalba Ortiz¹ & Jose Miguel Carot Sierra² & Alejandro Rodríguez Ortega³ & Luis Martí Bonmatí⁴

¹Escola Tècnica Superior de Informàtica - Universitat Politècnica de València

²Departamento de Estadística e Investigación Operativa Aplicadas y Calidad- Universitat Politècnica de València

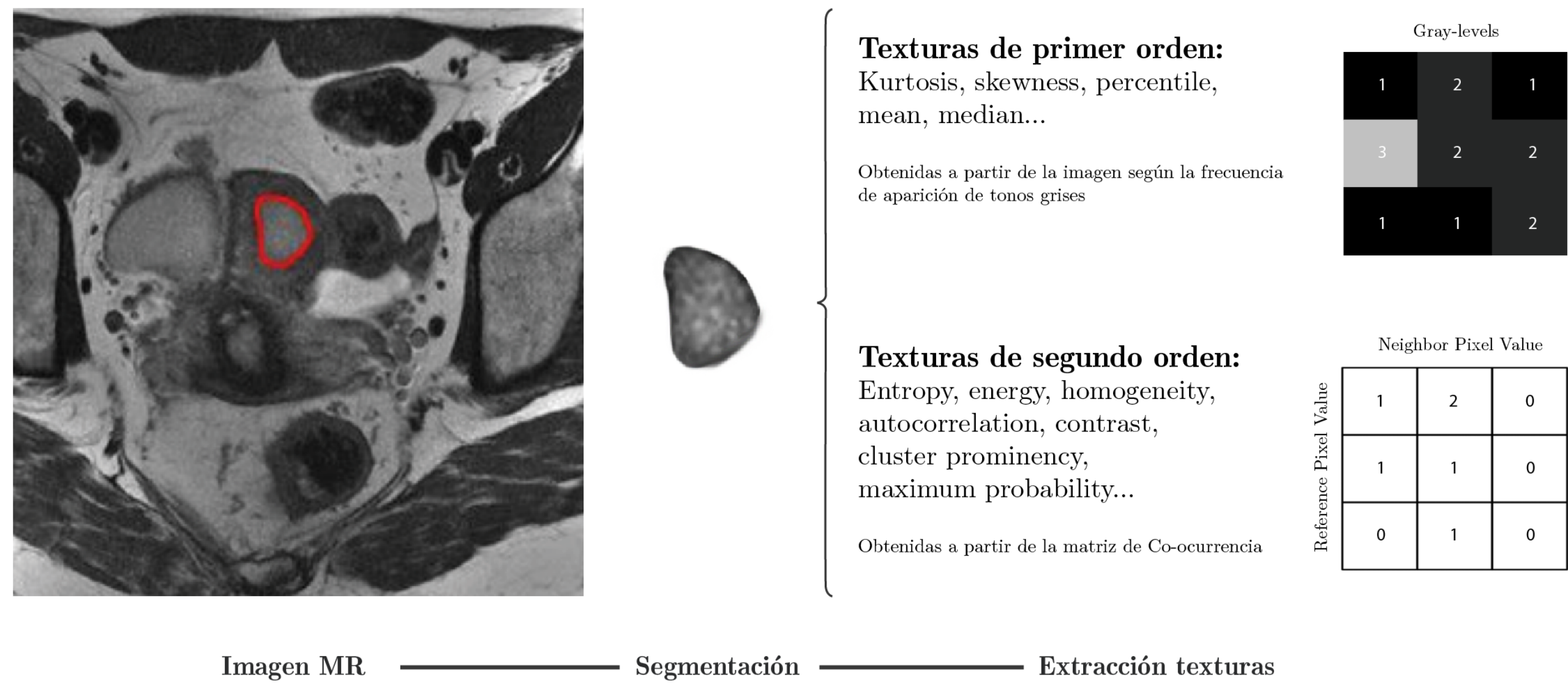
³Grupo de Investigación Biomédica en Imagen (GIBI230). Instituto de Investigación Sanitaria La Fe (IISLAFE)

⁴Departamento de Radiología. Hospital Universitario y Politécnico La Fe, Valencia

Introducción

El cáncer de endometrio es el tipo de cáncer ginecológico más común. La estadi-ficación de este tumor se hace quirúrgicamente, mediante técnicas que necesitan sedación. Es por ello, que existe la necesidad de usar herramientas de diagnóstico no invasivas, que ayuden a determinar la estadificación del tumor y ayudar a la planificación de la cirugía de los tumores endometriales. Se han empleado técni-cas de imagen por resonancia magnética (MRI) para evaluar el grado histológico de los tumores endometriales, a través de la extracción de diferentes biomarcado-res de estas imágenes.

Texturas



Conjunto de datos

El conjunto de datos consta de **144** pacientes con **4255** variables de texturas junto al grado histológico de cada tumor por cada observación.

Objetivo: Predecir el grado histológico de un tumor usando varia-bles de texturas

Selección de variables

La complejidad del problema de clasificación se ve afectada por la existencia de variables de texturas redundantes. Las técnicas de selección de variables se pueden clasificar en técnicas de tipo *wrapper* o de tipo *filter*. Los de tipo *filter* asignan un valor de importancia a cada variable x_j . Algoritmos canónicos de este tipo son; Fisher Score, coeficiente de correlación de Pearson, el estadístico χ^2 y la varianza.

Spectral Feature Selection

Dado un grafo $\mathbb{G}(V, E)$ donde V es el conjunto de nodos y E el conjunto de enlaces. El i -ésimo vértice v_i se corresponderá con la observación x_i y habrá un enlace entre dos pares de nodos (v_i, v_j) si x_i y x_j pertenecen a la misma clase y son vecinos. El peso de cada enlace será w_{ij} . Una variable x_j preservará la es-structura del grafo si asigna valores similares a observaciones que se encuentran conectadas en el grafo. Como se muestra en el ejemplo de la figura 1, la variable x_j asigna valores salvaguardando la estructura del grafo mientras que la variable x_l no lo hace.

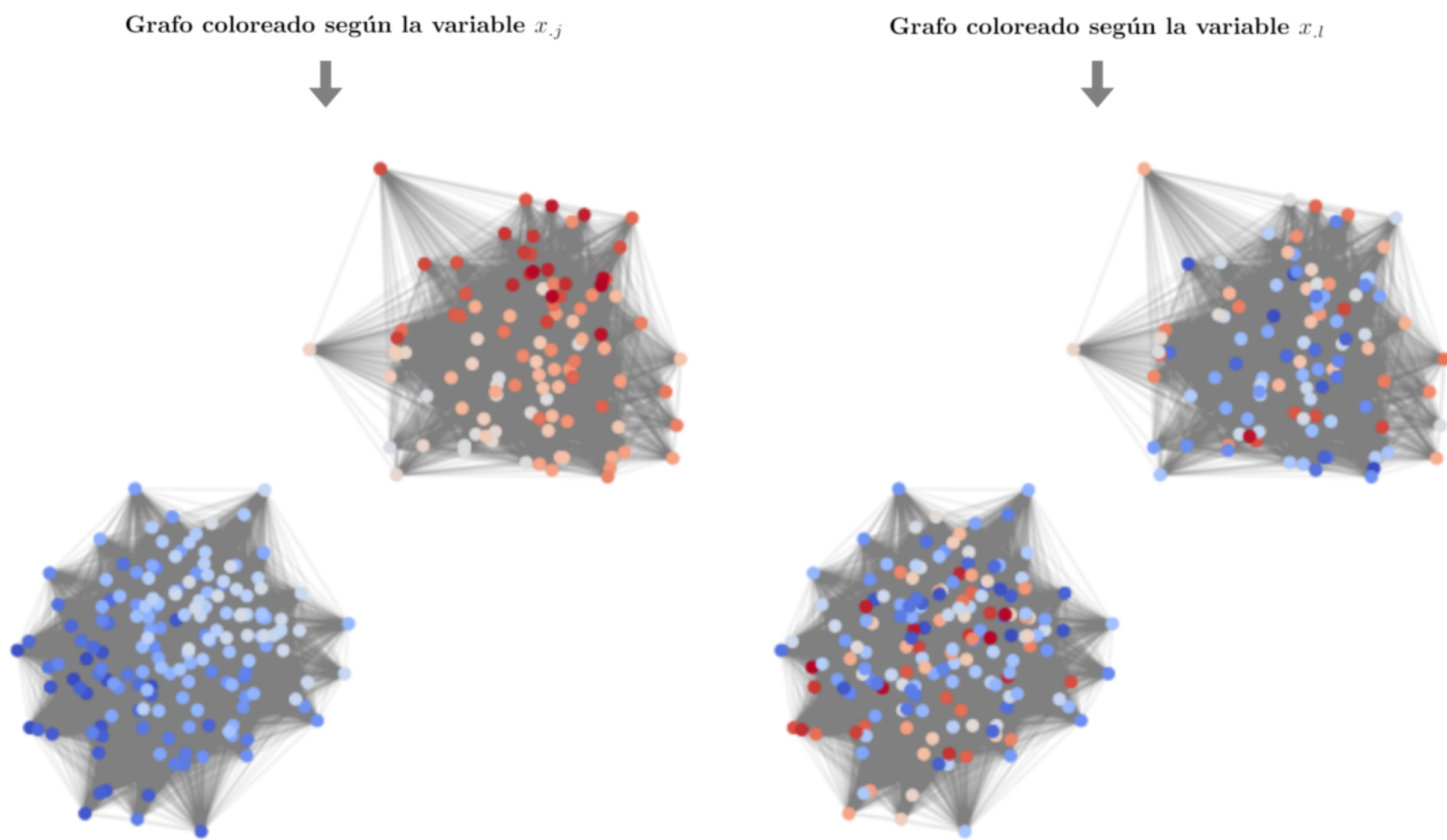


Figura 1: Estructura del grafo coloreada según x_l y x_j

El algoritmo Laplacian Score [1] asignará un valor a cada variable según la ca-pacidad de esta para preservar la geometría del grafo. Fisher Score es un caso especial de Laplacian Score con una estructura determinada del grafo subyacen-te. Los grafos usados para los algoritmos de Laplacian Score han sido;

- Grafo de los ϵ -más cercanos con dos componentes conexas usando un kernel simple para ponderar los enlaces:

$$w_{ij} = \begin{cases} 1 & \text{si } x_i \text{ está conectado con } x_j \\ 0 & \text{en otro caso} \end{cases}$$

- Grafo de los ϵ -más cercanos con dos componentes conexas usando un kernel de tipo multiquadric para ponderar los enlaces:

$$w_{ij} = \begin{cases} \sqrt{\|x - y\|^2 + c^2} & \text{si } x_i \text{ está conectado con } x_j \\ 0 & \text{en otro caso} \end{cases}$$

Resultados

Subconjunto de variables: El subconjunto óptimo de variables se ha escogido mediante validación cruzada de forma que las variables introducidas ordenadas según su valor de Fisher o Laplacian Score maximizasen el área bajo la curva del modelo.

Optimización de parámetros: El número de componentes PCA así como los hiperparámetros de los modelos se obtuvieron mediante validación cruzada.

Selección	Modelo	AUC	Accuracy
Laplacian Score Simple Kernel	SVM	$0,71 \pm 0,024$	$0,69 \pm 0,024$
Laplacian Score Simple Kernel	PLS-DA	$0,73 \pm 0,017$	$0,67 \pm 0,037$
Laplacian Score Simple Kernel	PCA + LDA	$0,74 \pm 0,017$	$0,72 \pm 0,013$
Laplacian Score Multiquadric Kernel	SVM	$0,70 \pm 0,023$	$0,71 \pm 0,016$
Laplacian Score Multiquadric Kernel	PLS-DA	$0,72 \pm 0,016$	$0,72 \pm 0,033$
Laplacian Score Multiquadric Kernel	PCA + LDA	$0,74 \pm 0,017$	$0,74 \pm 0,012$
Fisher Score	SVM	$0,69 \pm 0,028$	$0,69 \pm 0,023$
Fisher Score	PLS-DA	$0,74 \pm 0,020$	$0,66 \pm 0,033$
Fisher Score	PCA + LDA	$0,74 \pm 0,017$	$0,72 \pm 0,017$

Cuadro 1: Tabla comparativa resultados

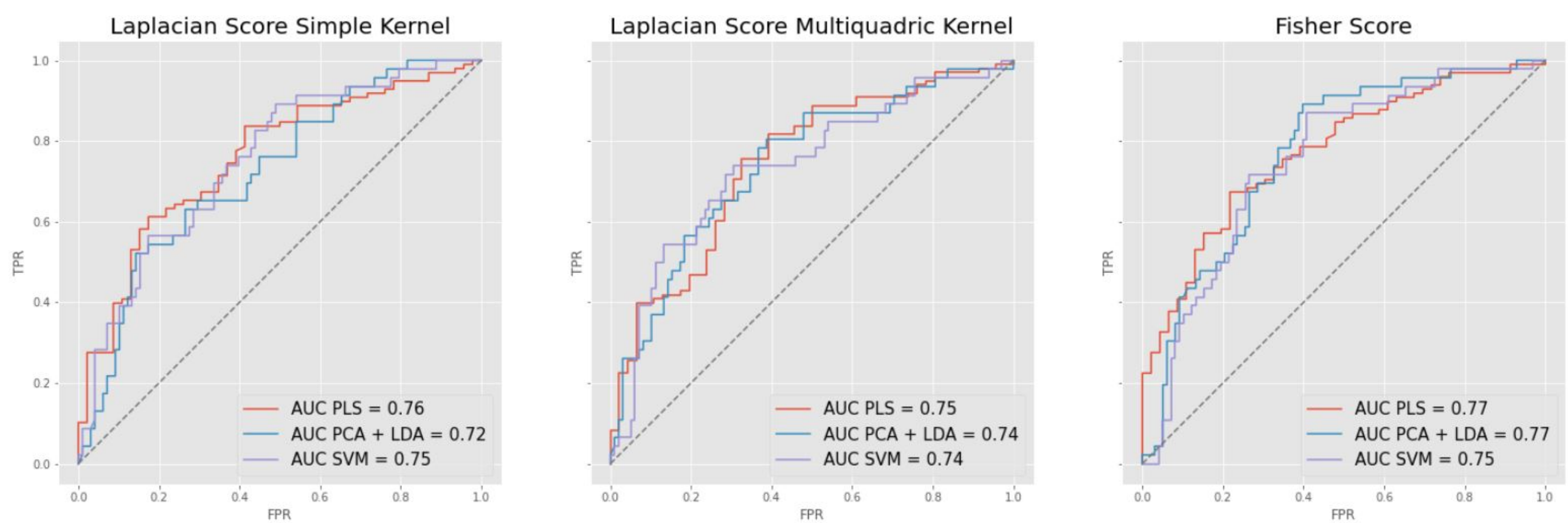


Figura 2: Comparativa curvas ROC

Conclusiones

- Los mejores resultados se obtienen con una selección de variables mediante La-placian Score con un Kernel de tipo Multiquadric y una combinación de PCA más LDA.
- Las variables de texturas podrían ser capaces de predecir con éxito el grado his-tológico sin necesidad de usar variables obtenidas mediante técnicas invasivas.

Referencias

[1] Xiaofei He, Deng Cai, and Partha Niyogi. Laplacian score for feature selec-tion. volume Vol. 18, 01 2005.