

Variabilidad interobservador en la detección y segmentación del adenocarcinoma ductal pancreático en TC

- Estandarización mediante consenso e inteligencia artificial -

María Montserrat Duh, Juan Moreno-Vedia, Elisabet Palomera Fanegas, Mateu Serra Prat,
Javier García López, Meritxell Riera-Marín, Anton Aubanell, Júlia Rodríguez-Comas.

Hospital Universitari de Mataró, Hospital Universitari de Bellvitge, Sycai Technologies SL

Introducción – la variabilidad interobservador

Reto:

Variabilidad anatómica y morfológica del PDAC dificulta su interpretación en TC

Problema:

Discrepancias interobservador en detección y delimitación del tumor

Objetivo:

Cuantificar la variabilidad entre radiólogos expertos en la práctica real



Demografía:

Cohorte

120 estudios TC con PDAC
confirmado
50% mujeres · 50% hombres ·
Mediana edad: **72 años**



Variable	Valor
Sexo, n (%)	
Mujer	60 (50.0%)
Hombre	60 (50.0%)
Edad	
Mediana	72 (63-77)
Rango	41-89
Distribución, n (%)	
40-60 años	22 (18.3%)
60-69 años	27 (22.5%)
70-79 años	46 (38.3%)
≥ 80 años	25 (20.8%)

Demografía:

Cohorte

120 estudios TC con PDAC
confirmado
50% mujeres · 50% hombres ·
Mediana edad: **72 años**



Variable	Valor
Fabricante TC, n (%)	
Philips	52 (43.3%)
Siemens	34 (28.3%)
Toshiba	26 (21.7%)
Otros (GE / Canon)	8 (6.6%)

Materiales y métodos:

Lectores

4 radiólogos independientes: **1, 3, 7 y 22 años** de experiencia

Análisis no enmascarado
(con diagnóstico conocido)

Referencia estándar

5ª etiqueta consensuada
mediante plataforma **Panorama**

Variables y estadística



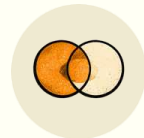
Índice Kappa (κ)

Mide el acuerdo entre observadores más allá del azar. Rango: -1 a 1. Valores > 0.6 indican concordancia sustancial.



Coeficiente de Correlación Intraclass (CCI)

Evalúa la consistencia de mediciones cuantitativas entre múltiples observadores. Rango: 0 a 1. > 0.75 indica excelente acuerdo.



Dice Similarity Coefficient (DCS)

Mide el solapamiento espacial entre segmentaciones de diferentes observadores. Rango: 0 a 1. Valores > 0.7 indican excelente concordancia en la delimitación del tumor.

Resultados:

Detección por radiólogo:



Sensibilidad de detección

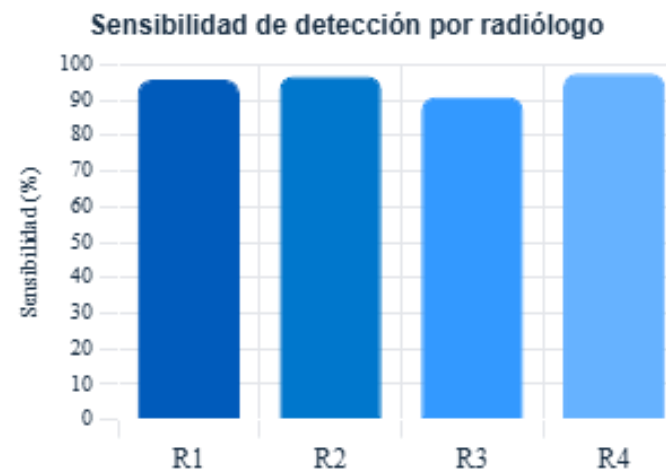
91–97%



Tasa de falsos negativos

3–8.5%

Alta capacidad individual
para **detectar** PDAC
(diagnóstico no
enmascarado, con
diagnóstico conocido
previamente)



Concordancia global:

⚠️ Concordancia global
baja: $\kappa = -0,05$ a $0,38$

A pesar de la alta
sensibilidad individual, el
acuerdo entre lectores es
limitado



Resultados:

Localización anatómica

Concordancia **moderada-alta**:

$\kappa = 0.31$ a 0.51 ($p < 0.001$)



Tamaño de la lesión

Concordancia **moderada-alta**:

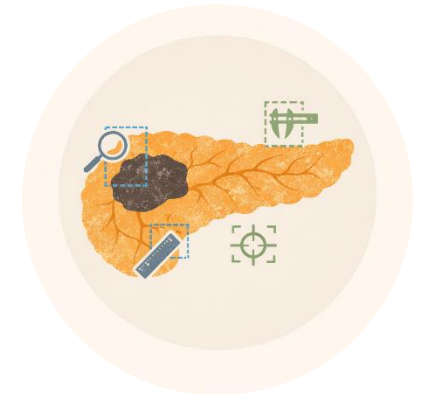
CCI = 0.65 a 0.83

Volumen de la lesión

Mayor variabilidad
interobservador

CCI = **0.49** a 0.82

La segmentación volumétrica es el paso más **variable y crítico** del proceso



Resultados:

»» Alta capacidad de **detección**

Los radiólogos detectan el tumor con una excelente sensibilidad individual, si bien este análisis se realizó en condiciones no enmascaradas, con conocimiento previo del diagnóstico.

»» **Variabilidad** en delimitación

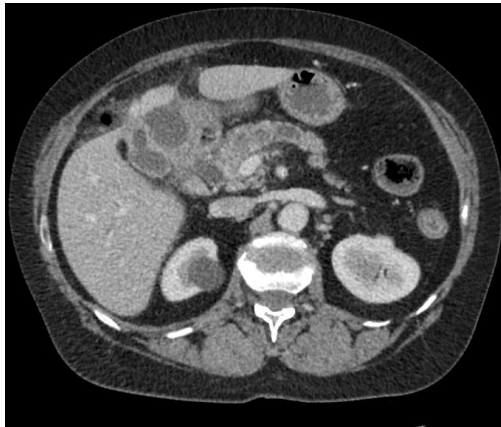
La **segmentación** y **cuantificación** presentan discrepancias considerables entre lectores

Necesidad de estandarización

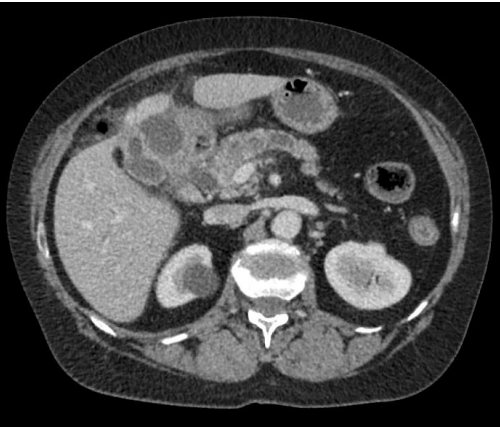
Optimizar protocolos y validar **herramientas automáticas** es prioritario para la práctica clínica



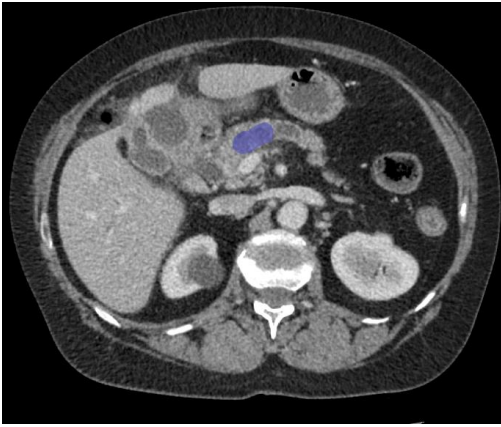
Segmentaciones y variabilidad entre radiólogos:



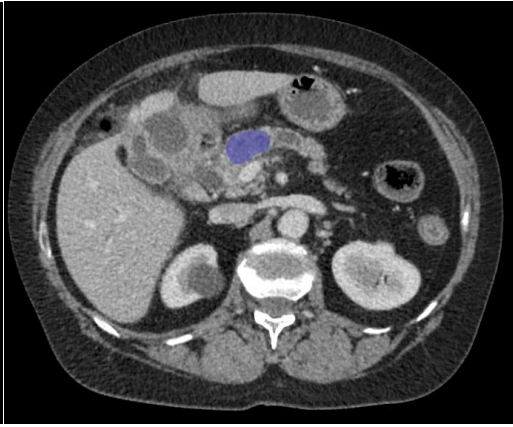
R1 – Slice 117 X



R2 – Slice 117 X



R3 – Slice 117 ✓

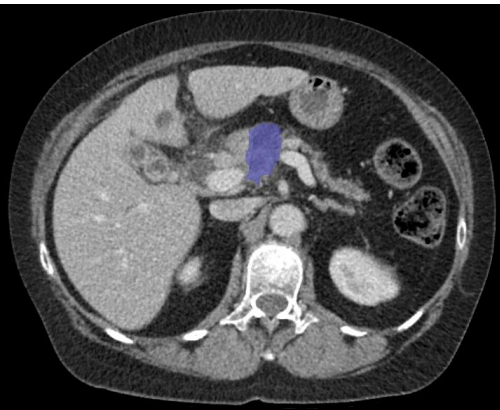


R4 – Slice 117 ✓

*PDAC segmentado en Slices diferentes:

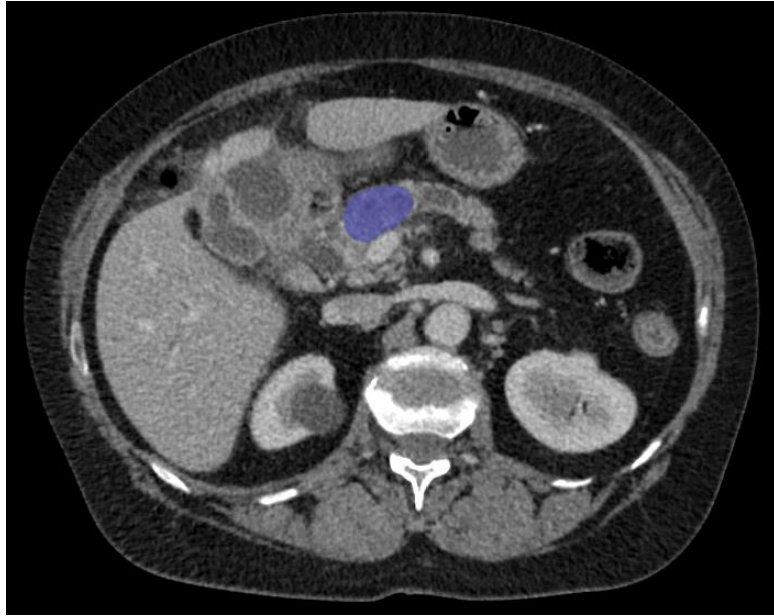


R1 – Slice 121



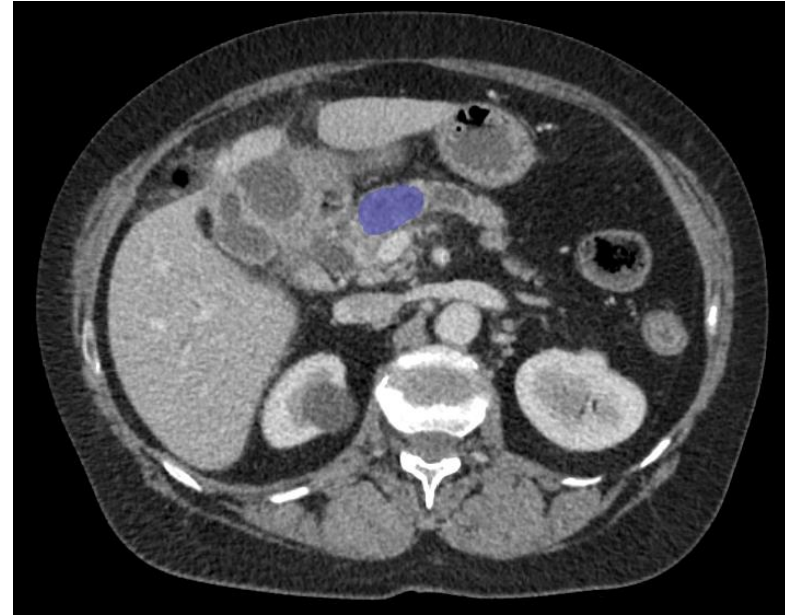
R2 – Slice 121

Importancia de la estandarización:



5ª etiqueta consensuada mediante plataforma **Panorama**

Slice 117 ✓



Etiqueta estandarizada – método
STAPLE (etiquetas de los 4 radiólogos +
etiqueta Panorama) Slice 117 ✓

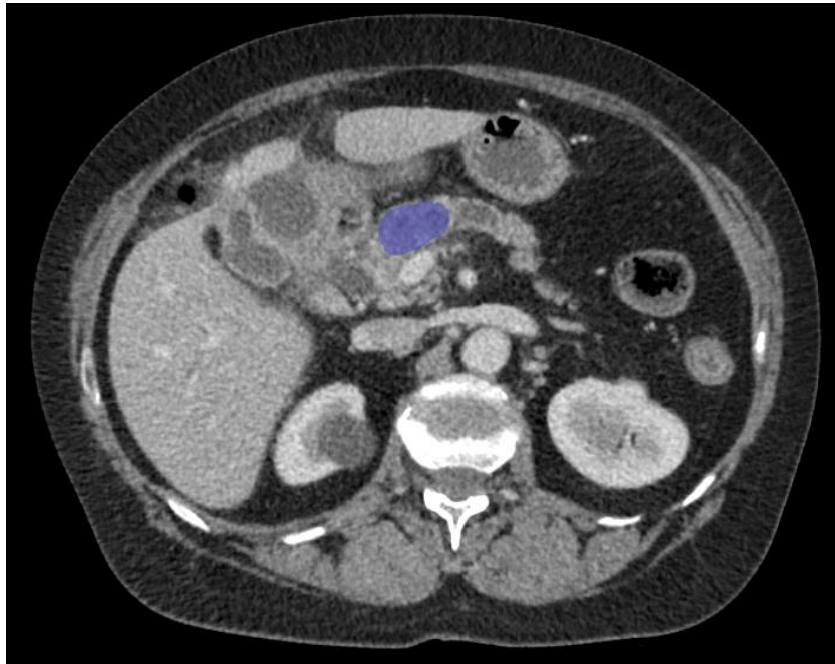
Importancia de la estandarización:

- Se calculó el **DCS** entre las etiquetas de los radiólogos y con la etiqueta radiológica consensuada de referencia (Panorama)
- Se generó la etiqueta **STAPLE**:

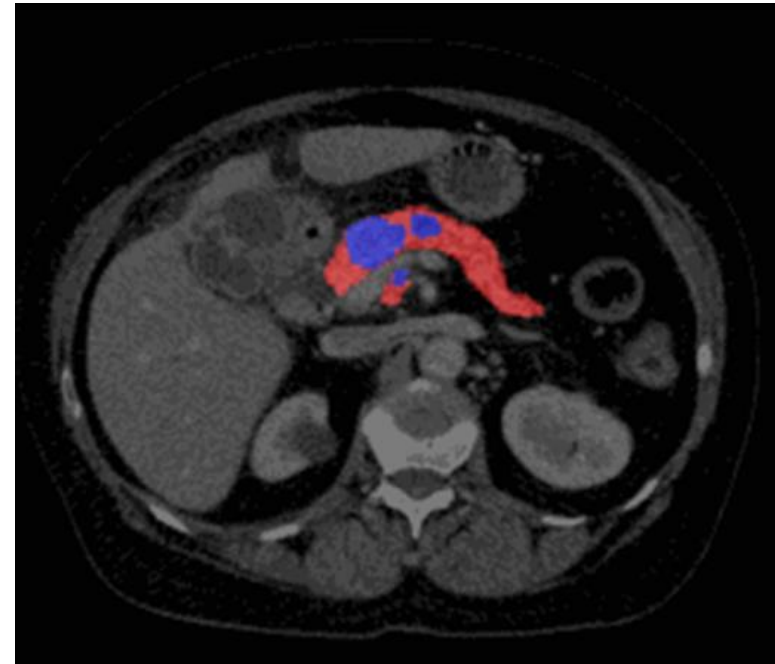
	R1	R2	R3	R4	Panorama
R1		0.81	0.49	0.54	0.58
R2	0.81		0.51	0.57	0.59
R3	0.49	0.51		0.74	0.64
R4	0.54	0.57	0.74		0.71
Panorama	0.58	0.59	0.64	0.71	
»» STAPLE	0.69	0.71	0.74	0.84	0.86

STAPLE (Simultaneous Truth and Performance Level Estimation) se define como una estimación de consenso derivada de la agregación de todas las segmentaciones disponibles, integrando la información de los distintos lectores en una única referencia. No obstante, la fiabilidad de esta aproximación depende del número y la variabilidad de los anotadores, por lo que la incorporación de un mayor número de lectores permitiría obtener un consenso más robusto y representativo.

Predicción con SYCAI MEDICAL:



STAPLE



Se está analizado un modelo de IA para evaluar la concordancia en términos de DSC con STAPLE, con el objetivo de explorar su potencial como herramienta de estandarización.

**DCS =
0.80**

Conclusiones

- »» Existe una **variabilidad** considerable en la delimitación y caracterización del PDAC
- »» Importancia crítica en la **estandarización** e interpretación de las lesiones PDAC
- »» Una herramienta de **IA** automatizada puede contribuir a la estandarización del proceso de segmentación.

