

# Comparativa de diferentes métodos de aprendizaje automático para la clasificación de lesiones adrenales indeterminadas diagnosticadas de forma incidental en TC con contraste

## Comparison of different machine learning methods for the classification of indeterminate adrenal lesions incidentally diagnosed in contrast enhanced CT

Daniel Prieto Morán<sup>1,\*</sup>, Miguel Ángel Gómez Bermejo<sup>2</sup>, Elena Canales Lachen<sup>2</sup>, Ernesto García Santana<sup>3</sup>, Raquel García Latorre<sup>2</sup>, Miguel Cámara Gallego<sup>1</sup>, Rafael Colmenares Fernández<sup>1</sup>, Ana Belén Capuz Suárez<sup>1</sup>, María José Béjar Navarro<sup>1</sup>, Juan David García Fuentes<sup>1</sup>, David Sevillano Martínez<sup>1</sup>, Rafael Morís Pablos<sup>1</sup>, Javier Blázquez Sanchez<sup>2</sup>, Feliciano García Vicente<sup>1</sup>

<sup>1</sup> Hospital Universitario Ramón y Cajal, Servicio de Radiofísica, Madrid, España.

<sup>2</sup> Hospital Universitario Ramón y Cajal, Servicio de Radiodiagnóstico, Madrid, España.

<sup>3</sup> Hospital Universitario Insular de Gran Canaria, Servicio de Radiodiagnóstico, Las Palmas de Gran Canaria, España.

Fecha de Recepción: 20/12/2023 - Fecha de Aceptación: 21/05/2024

**Objetivo:** Se realiza una comparativa de diferentes métodos de aprendizaje automático para discriminar incidentalomas suprarrenales mediante estudios de TC.

**Métodos:** Se obtuvieron 62 características radiómicas diferentes a partir de una muestra de 61 incidentalomas indeterminados utilizando el software de licencia libre LIFEx. Se realizaron 19 modelos predictivos empleando además diversos métodos de selección de características para optimizar la detección de lesiones malignas. Para todos ellos se evaluaron cuatro métodos de validación cruzada. El contorneado de los adenomas fue realizado por duplicado por 4 radiólogos.

**Resultados:** Se obtienen los valores del área bajo la curva ROC entre 0.42 (0.09-0.81) y 0.92 (0.63-1.00), y exactitud de los modelos entre 0.63 (0.43-0.79) y 0.94 (0.82-1.00). El modelo de mejor rendimiento fue la regresión logística balanceada entrenado con 14 características con un coeficiente intraclass superior a 0.9, con el que se obtuvo una exactitud de 0.94 (0.74-1.00), un AUC ROC de 0.917 (0.63-1.00), una sensibilidad de 0.92 (0.65-1.00) y especificidad de 1.00 (0.71-1.00)

**Conclusiones:** La evaluación, comparación y validación de diferentes modelos predictivos basados en características radiómicas nos ha permitido obtener un modelo optimizado (regresión logística balanceada con ICC > 0.9) para la detección de tumores adrenales malignos entre los incidentalomas diagnosticados de forma incidental en TC con contraste.

**Palabras clave:** Radiómica, inteligencia artificial, LIFEx, adenoma, incidentalomas, suprarrenal.

**Purpose:** A comparison of different machine learning methods to discriminate adrenal incidentalomas by CT studies was performed.

**Methods:** Sixty-two different features were obtained from a sample of 61 incidentalomas using the free license software LIFEx and 19 radiomic studies were performed with different models and feature selection methods to obtain the most efficient determination of possible malignancy. For all of them, four cross-validation methods were evaluated. Adenoma contouring was performed in duplicate by different radiologists evaluating all models in both groups.

**Results:** ROC AUC between 0.42 (0.09-0.81) and 0.92 (0.63-1.00), and accuracy of the models between 0.63 (0.43-0.79) and 0.94 (0.82-1.00). The best-performing model was the balanced logistic regression applied to the 14 features with

\*Correspondencia: daniel.prieto@salud.madrid.org

<https://doi.org/10.37004/sefm/2024.25.2.001>

an intraclass coefficient greater than 0.9, with which accuracy of 0.94 (0.74-1.00), ROC AUC of 0.917 (0.63-1.00), benign recall of 0.92 (0.65-1.00) and malignant recall of 1.00 (0.71-1.00) were obtained.

**Conclusions:** The evaluation and validation of different models has allowed us to obtain an efficient radiomic model (balanced logistic regression ICC > 0.9) for the discrimination of adrenal incidentalomas.

**Key words:** Radiomics, machine learning, LIFEx, adenoma, incidentaloma, suprarenal.

## 1. Introducción

Los tumores suprarenales son un hallazgo incidental frecuente en las pruebas de imagen y suelen denominarse incidentalomas suprarenales.<sup>1</sup> Su prevalencia en los estudios radiológicos es del 3-7%.<sup>2</sup> En las series clínicas, el 80% son adenomas suprarenales, el 7% feocromocitomas, el 8% carcinomas y el 5% metástasis.<sup>3</sup> Por diversos motivos, es probable que la frecuencia de los adenomas suprarenales esté infraestimada.<sup>4</sup>

A pesar de la elevada frecuencia de tumores benignos, las sociedades médicas recomiendan la realización de estudios clínicos, analíticos y radiológicos de forma sistemática en muchos pacientes para excluir tumores malignos o productores de hormonas, a pesar de ser poco frecuentes. Los incidentalomas suprarenales son motivo de un elevado número de estudios de imagen. Debido a la baja probabilidad de malignidad en la mayoría de los casos, los resultados suelen ser negativos o falsos positivos. Esto se traduce en costes significativos, estrés emocional para el paciente y una dosis de radiación que podría conllevar un riesgo de cáncer mortal similar al riesgo de que el tumor suprarenal sea maligno.<sup>4, 5</sup> Por ello, es necesario mejorar la selección de pacientes para futuros estudios.

La radiómica extrae una gran cantidad de información cuantitativa de las imágenes radiológicas que podría utilizarse para elaborar modelos predictivos capaces de aumentar la precisión diagnóstica en comparación con la inspección visual o las medidas cuantitativas simples.<sup>6, 7</sup>

Para ello, se desarrollan modelos matemáticos predictivos a partir de los valores de las diferentes características radiómicas obtenidas de las imágenes.

La elección del modelo adecuado es a veces una tarea compleja que requiere diferentes métodos y una evaluación detallada de los resultados. Para mejorar el resultado de los modelos y ajustar los datos a la finalidad requerida, la selección de variables desempeña un papel muy importante. El reto es aún mayor cuando la muestra se encuentra desequilibrada, existiendo pocos casos de la patología a detectar.

El objetivo de este trabajo es comparar los resultados obtenidos con diferentes modelos predictivos radiómicos y varios métodos de selección de variables en la clasificación de incidentalomas adrenales en

tomografía computerizada (TC) con contraste, entre benignos y malignos.

## 2. Material y métodos

### 2.1. Datos

Este estudio cuenta con la aprobación del comité de ética en investigación médica local. Todos los procedimientos con pacientes se realizaron siguiendo la declaración de Helsinki de 1964 y sus posteriores modificaciones.

Los pacientes seleccionados corresponden a estudios de TC realizados entre 2012 y 2020 en nuestro centro. Tres radiólogos especialistas con al menos 3 años de experiencia y un residente de tercer año realizaron la selección. Los criterios de inclusión fueron los siguientes:

- Tumor suprarenal mayor de 1 cm diagnosticado histológicamente o con resonancia (RM).
- Detectado por primera vez en TC con contraste en adquirida en fase venosa portal.

Se excluyeron los pacientes que presentaban:

- Hallazgos diagnósticos típicos (grasa macroscópica, ausencia de realce o calcificaciones).
- Neoplasias malignas conocidas en el momento del diagnóstico del incidentaloma.

Todas las imágenes fueron adquiridas en nuestras instalaciones con diferentes equipos TC: Siemens Somatom Sensation 64, Philips Brilliance 64, Philips Brilliance 16, Philips Incisive y Canon Aquilion One. Los protocolos de adquisición se obtuvieron a 120 kVp y con modulación de corriente activada (100-200 mAs). Para obtener características más robustas, se uniformizó el tamaño del píxel y el grosor del corte,<sup>8</sup> fijando el espesor de corte de la imagen en 1,5 mm y el muestreo en x e y en 1 mm.

Se emplea el software de licencia libre LIFEx 6.3<sup>9</sup> para delimitar los contornos y obtener las características. El software extrae 95 características radiómicas.<sup>10</sup>

Con una primera selección de características se eliminan aquellas que no son resultado numérico, son constantes, matrices, no se recomiendan para estudios de TC o proporcionan parámetros de cálculo o información técnica del estudio, reduciendo finalmente el número de características analizadas a 62 (Tabla 1).

Cada volumen de interés (VOI) fue delimitado por dos radiólogos diferentes de forma independiente, de los cuatro implicados, formando dos grupos separados de datos (A y B). Se analizó la correlación interoperador mediante el coeficiente de correlación intraclase (ICC)<sup>11, 12</sup> para las 62 características radiómicas.

## 2.2. Tratamiento de los datos, modelos empleados y validación

Se emplea el lenguaje de programación Python para la manipulación de datos, estadística, obtención de modelos y validación, utilizando las librerías de aprendizaje automático más comunes (*sklearn, pandas, ...*).

Se calcula la matriz de correlación de Pearson para cada grupo (A y B) de forma separada.

Los VOI se distribuyeron entre los grupos de entrenamiento y validación siguiendo una proporción de 75% y 25% respectivamente, garantizando 12 benignos y 4 malignos en el grupo de validación. Antes de entrenar los modelos, los datos se escalan utilizando el método mínimo-máximo.<sup>13</sup>

Los métodos de aprendizaje automático evaluados han sido: la regresión logística (*logistic regression*) (LR),<sup>14</sup> regresión logística balanceada (*balanced logistic regression*) (BLR),<sup>15</sup> k-vecinos (*k-neighbors*) (kN) ( $n = 3$ ),<sup>16</sup> árbol de decisión (*decision tree*) (DT),<sup>17</sup> clasificador de bosque aleatorio (*random forest*) (RF),<sup>18</sup> clasificador de bosque aleatorio balanceado (*balanced random forest*) (BRF), clasificador de aumento del gradiente (*gradient boost classifier*) (GB),<sup>19</sup> clasificador de aumento de gradiente supervisado (*random under-sampling AdaBoost*) (RUS)<sup>20,21</sup> y clasificador ponderado por varios modelos sencillos (*easy ensemble classifier*) (EE)<sup>22</sup>.

El modelo de regresión logística balanceada se aplicó a las 62 características resultantes. Para mejorar el algoritmo, se realizaron diferentes métodos de selección de variables, entrenando el modelo con las características resultantes. Los distintos métodos utilizados son:

- Eliminación de variables fuertemente correlacionadas (Pearson). Umbral en 0.999 (10 características eliminadas) (BLR 0.999) y 0.998 (12 características eliminadas) (BLR 0.998).
- Selección de variables mediante el algoritmo clasificador de bosque aleatorio con el umbral en la mediana (31 características eliminadas) (BLR

Mediana) y la media (37 características eliminadas) (BLR Media).

- Selección iterativa de variables<sup>23</sup> con el estimador de bosque aleatorio (random forest estimator) (RFE) (42 características eliminadas) (BLR RFE).
- Selección de las cinco variables más importantes para el árbol de decisión (*Discretized\_Histo\_Entropy\_log10, GLZLM\_LZLGE, Conventional\_HUmin, Discretized\_HUQ3, GLZLM\_LZE*) (57 características eliminadas) (BLR DT 5).
- Selección del valor F de ANOVA<sup>24, 25</sup> (42 características eliminadas) (BLR AF).
- Variables con un coeficiente intraclase, ICC > 0.9<sup>11,12</sup> (48 características eliminadas), con el fin de eliminar las incertidumbres introducidas por el radiólogo responsable y el equipo seleccionado (BLR ICC).
- Variables morfológicas (*Shape\_Volume(mL), Shape\_Volume(vx), Shape\_Sphericity, Shape\_Surface(mm2), Shape\_Compacity*) (57 características eliminadas), dado que las características morfológicas, como veremos en la sección de resultados, se encuentran entre las que presentan un valor de ICC más elevado y son, en principio, las que tienen más probabilidades de diferir cuando la región de interés está delimitada por diferentes personas (BLR Mor pH).

El algoritmo de árbol de decisión se modeló con las 62 características y las 5 más importantes (DT 5MI).

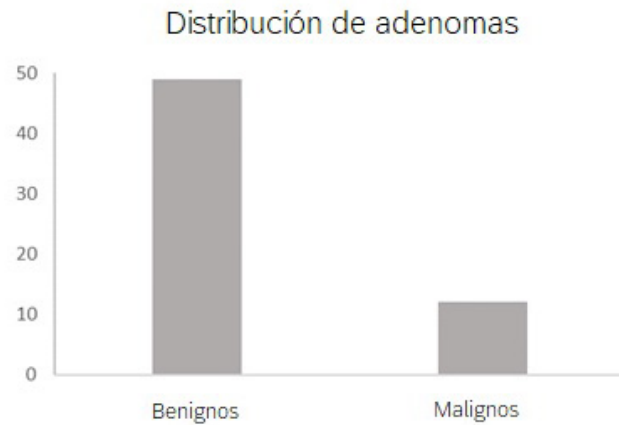
Para cada modelo evaluado, se calcula la matriz de confusión (verdaderos y falso positivos/negativos), se representa la curva ROC y se obtienen los valores de la exactitud ("accuracy"), sensibilidad (tasa de verdaderos positivos, "benign recall"), especificidad (tasa de verdaderos negativos, "specificity") y *f1-score*.

Los métodos de aprendizaje automático se validaron mediante validación cruzada utilizando cuatro métodos, dos generales<sup>26</sup> *k-fold* y *shuffle split*, y sus versiones estratificadas<sup>27,28</sup> para muestras balanceadas, *k-fold* estratificado y *shuffle split* estratificado.<sup>29</sup> En todos los casos, se utilizaron tres subgrupos de cálculo obteniendo la precisión para cada uno de ellos y la desviación estándar de los tres valores.

El intervalo de confianza se obtuvo mediante el método *out-of-bag bootstrapping*,<sup>30</sup> realizando 200 iteraciones. Presentando en este trabajo para cada una de las métricas el intervalo de confianza del 95%.

## 3. Resultados

La selección de pacientes detallada anteriormente dio como resultado un total de 61 VOIs, de las cuales 49 correspondían a tumores benignos y 12 a malignos.



**Fig. 1.** Comparación entre el número de tumores benignos y malignos del estudio.

Dados los dos grupos de 61 regiones de interés, se obtuvieron los ICC del análisis intraclase entre el grupo A y B, para las 62 características obtenidas, tal como se presenta en la Tabla 1.

**Tabla 1.** Lista de las características empleadas ordenadas por el coeficiente intraclase obtenido.

FEATURES	ICC
CONVENTIONAL_HUQ2	0.97 (0.94-0.98)
CONVENTIONAL_HUQ3	0.96 (0.94-0.98)
CONVENTIONAL_HUQ1	0.96 (0.93-0.97)
GLZLM_ZLNU	0.96 (0.92-0.97)
CONVENTIONAL_HUmean	0.95 (0.93-0.97)
GLRLM_RLNU	0.95 (0.92-0.97)
SHAPE_Volume(mL)	0.95 (0.91-0.97)
SHAPE_Volume(vx)	0.95 (0.91-0.97)
NGLDM_Busyness	0.94 (0.91-0.97)
CONVENTIONAL_HUcalciumAgatstonScore[onlyForCT]	0.94 (0.90-0.96)
GLZLM_GLNU	0.94 (0.89-0.96)
SHAPE_Surface(mm2)[onlyFor3DROI]	0.93 (0.89-0.96)
GLRLM_GLNU	0.93 (0.88-0.96)
SHAPE_Compacity[onlyFor3DROI]	0.92 (0.85-0.95)
GLCM_Entropy_log2[=JointEntropy]	0.81 (0.70-0.88)
GLCM_Entropy_log10	0.81 (0.70-0.88)
CONVENTIONAL_HUmax	0.78 (0.66-0.86)

**Tabla 1 (cont.).** Lista de las características empleadas ordenadas por el coeficiente intraclase obtenido.

FEATURES	ICC
NGLDM_Coarseness	0.77 (0.65-0.86)
GLCM_Correlation	0.65 (0.48-0.77)
GLRLM_SRHGE	0.62 (0.44-0.75)
GLRLM_HGRE	0.60 (0.41-0.74)
CONVENTIONAL_HUstd	0.59 (0.40-0.73)
DISCRETIZED_AUC_CSH	0.58 (0.39-0.73)
DISCRETIZED_HISTO_Entropy_log2	0.54 (0.34-0.70)
DISCRETIZED_HISTO_Entropy_log10	0.54 (0.34-0.70)
GLCM_Dissimilarity	0.48 (0.26-0.65)
DISCRETIZED_HUQ2	0.48 (0.26-0.65)
GLCM_Contrast[=Variance]	0.46 (0.24-0.64)
SHAPE_Sphericity[onlyFor3DROI]	0.44 (0.22-0.62)
GLZLM_LZHGE	0.40 (0.17-0.58)
DISCRETIZED_HISTO_Energy[=Uniformity]	0.39 (0.15-0.58)
GLCM_Homogeneity[=InverseDifference]	0.39 (0.15-0.58)
DISCRETIZED_HUmax	0.37 (0.13-0.56)
GLZLM_HGZE	0.34 (0.10-0.55)
DISCRETIZED_HUQ3	0.30 (0.06-0.52)
GLCM_Energy[=AngularSecondMoment]	0.29 (0.04-0.50)
GLRLM_LRHGE	0.29 (0.04-0.50)
GLZLM_SZE	0.27 (0.02-0.48)
DISCRETIZED_HUSkewness	0.26 (0.02-0.48)
GLZLM_ZP	0.26 (0.01-0.48)
DISCRETIZED_HUmin	0.24 (0.00-0.45)
GLRLM_SRE	0.23 (-0.02-0.45)
GLZLM_SZHGE	0.23 (-0.02-0.45)
GLRLM_RP	0.20 (-0.03-0.45)
NGLDM_Contrast	0.16 (-0.05-0.42)
CONVENTIONAL_HUSkewness	0.16 (-0.10-0.39)
GLRLM_LRE	0.14 (-0.09-0.39)
GLZLM_LGZE	0.12 (-0.11-0.37)
CONVENTIONAL_HUmin	0.11 (-0.14-0.35)
DISCRETIZED_HUQ1	0.09 (-0.16-0.33)

**Tabla 1 (cont.).** Lista de las características empleadas ordenadas por el coeficiente intraclase obtenido.

FEATURES	ICC
GLZLM_SZLGE	0.06 (-0.20-0.30)
DISCRETIZED_HUExcessKurtosis	0.05 (-0.20-0.30)
DISCRETIZED_HUKurtosis	0.05 (-0.20-0.29)
CONVENTIONAL_HUExcessKurtosis	0.05 (-0.20-0.29)
CONVENTIONAL_HUKurtosis	0.05 (-0.20-0.28)
DISCRETIZED_HUstd	0.03 (-0.22-0.28)
GLZLM_LZE	0.02 (-0.24-0.27)
GLRLM_SRLGE	0.02 (-0.24-0.27)
GLRLM_LGRE	0.01 (-0.25-0.26)
DISCRETIZED_HUmean	0.00 (-0.25-0.25)
GLZLM_LZLGE	-0.01 (-0.26-0.24)
GLRLM_LRLGE	-0.04 (-0.28-0.21)

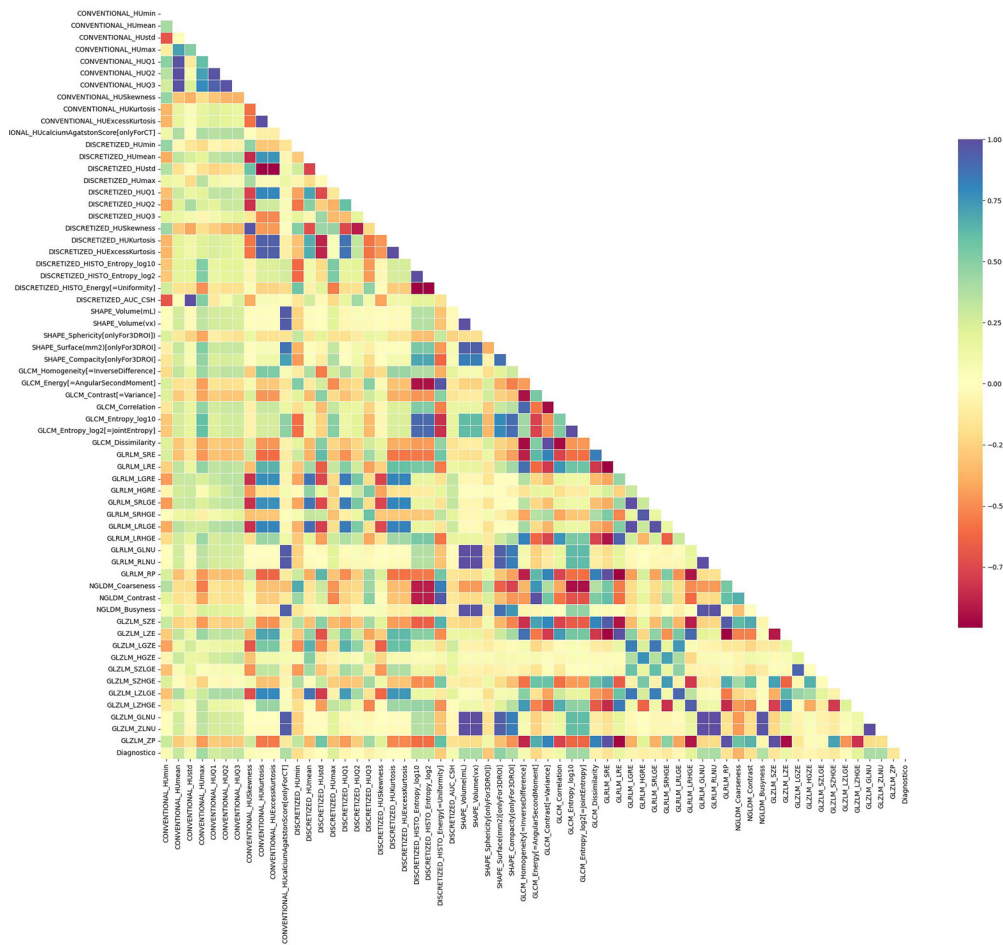
Las matrices de correlación de Pearson (Figs. 2 y 3) muestran que sólo una minoría de variables presentan fuertes correlaciones entre sí.

Los métodos de aprendizaje automático que utilizan algoritmos que tienen en cuenta el desequilibrio de los datos de la muestra (BLR, BRF) mejoran los resultados obtenidos en la matriz de confusión con respecto a sus análogos que no lo tienen en cuenta. Esto es más significativo en el grupo A que en el grupo B.

Para localizar los métodos con un mejor grado de ajuste, representamos gráficamente los valores de exactitud, el área bajo la curva ROC (“AUC ROC”) y los valores de sensibilidad y especificidad de todos los modelos evaluados (Fig. 4).

El clasificador gradient boost classifier, random under-sampling integrated in the learning of AdaBoost y los modelos de árbol de decisión con sólo las cinco características más importantes son los modelos para los que la exactitud y el área bajo la curva ROC obtienen valores más bajos.

Sólo dos de los métodos evaluados presentaron, para cualquiera de los conjuntos de datos, un área bajo la curva ROC inferior a 0.5. Los mejores resultados de



**Fig. 2.** Matriz de correlación para el grupo A.



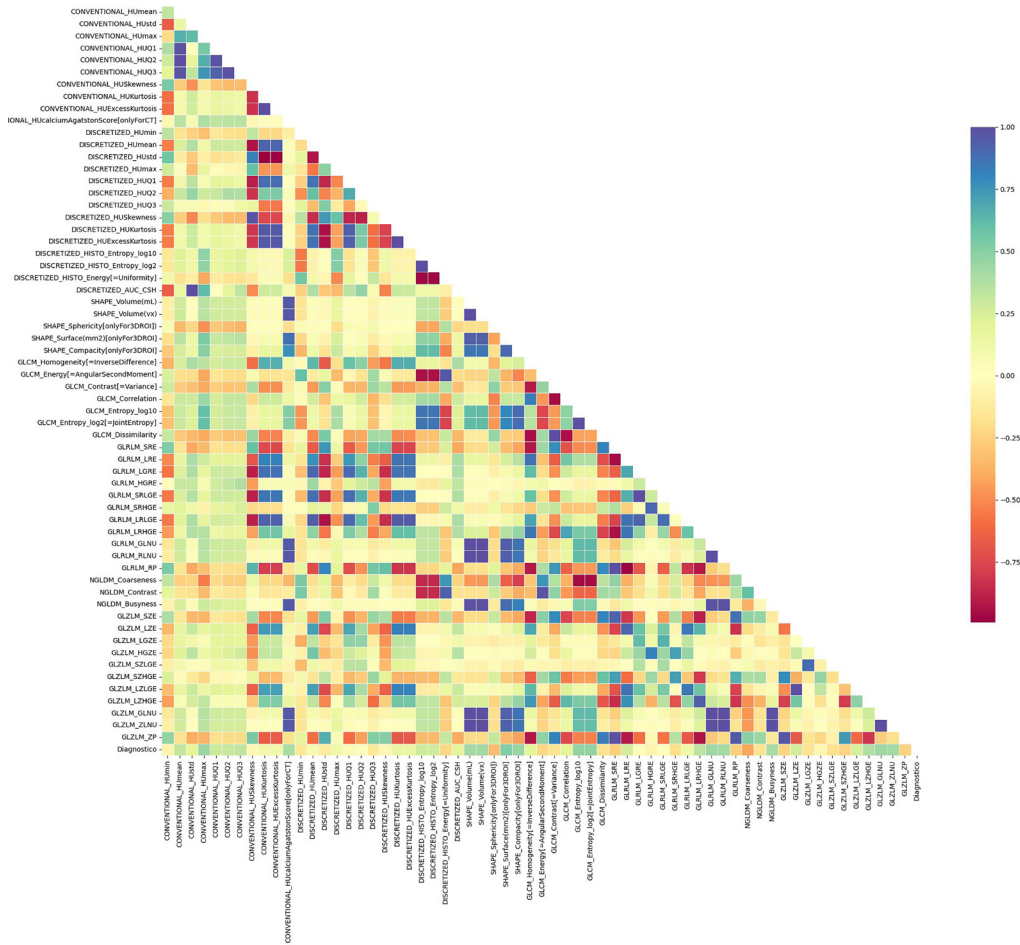


Fig. 3. Matriz de correlación para el grupo B.

la curva ROC fueron los correspondientes a la regresión logística balanceada con selección de variables ANOVA y con selección de variables ICC > 0.9. La implementación de forma conjunta de ambos métodos de selección de variables, (ICC y ANOVA) devuelve exactamente las variables morfológicas, obteniendo por tanto los mismos resultados ya evaluados.

Todos los valores de exactitud fueron superiores al 60%. El peor resultado se obtuvo para la selección de variables que escoge las variables más importantes para el modelo de árbol de decisión. Y el mejor es para la regresión logística balanceada con selección de variables ANOVA, 0.94 (0.82-1.00), y con selección de variables ICC > 0.9, 0.92 (0.74-1.00).

Dada la importancia del valor de la especificidad, en la Fig. 4 comparamos los valores de sensibilidad y especificidad de todos los modelos evaluados, con el fin de visualizar si existe o no una pérdida en el valor de sensibilidad. Obtuvimos una tasa de acierto para los adenomas benignos siempre superior al 50% sin observar una pérdida debido a la mejora en la tasa de acierto para los adenomas malignos.

Dado que nuestro trabajo es una clasificación binaria, el método de regresión logística es adecuado, por ello, estudiamos el impacto de utilizar diferentes métodos de selección de variables. La Fig. 5 muestra la desviación estándar de la exactitud de las tres submuestras ( $k = 3$ ) obtenidas con los cuatro métodos diferentes de validación cruzada. La desviación estándar de la exactitud obtenida en cada uno de los métodos se sitúa entre 0 y 0.24.

A la vista de los resultados, consideramos que los métodos de aprendizaje automático adecuados para nuestra muestra son aquellos que presentan al menos:

- Una especificidad de 0.75 para ambos grupos (detección de al menos el 75%).
- Área bajo la curva ROC > 0.7 para ambos grupos.
- Desviación estándar de la exactitud en los diferentes métodos de validación cruzada < 0.2.

Así, los tres modelos que cumplen estas condiciones (Figs. 4 y 5) son:

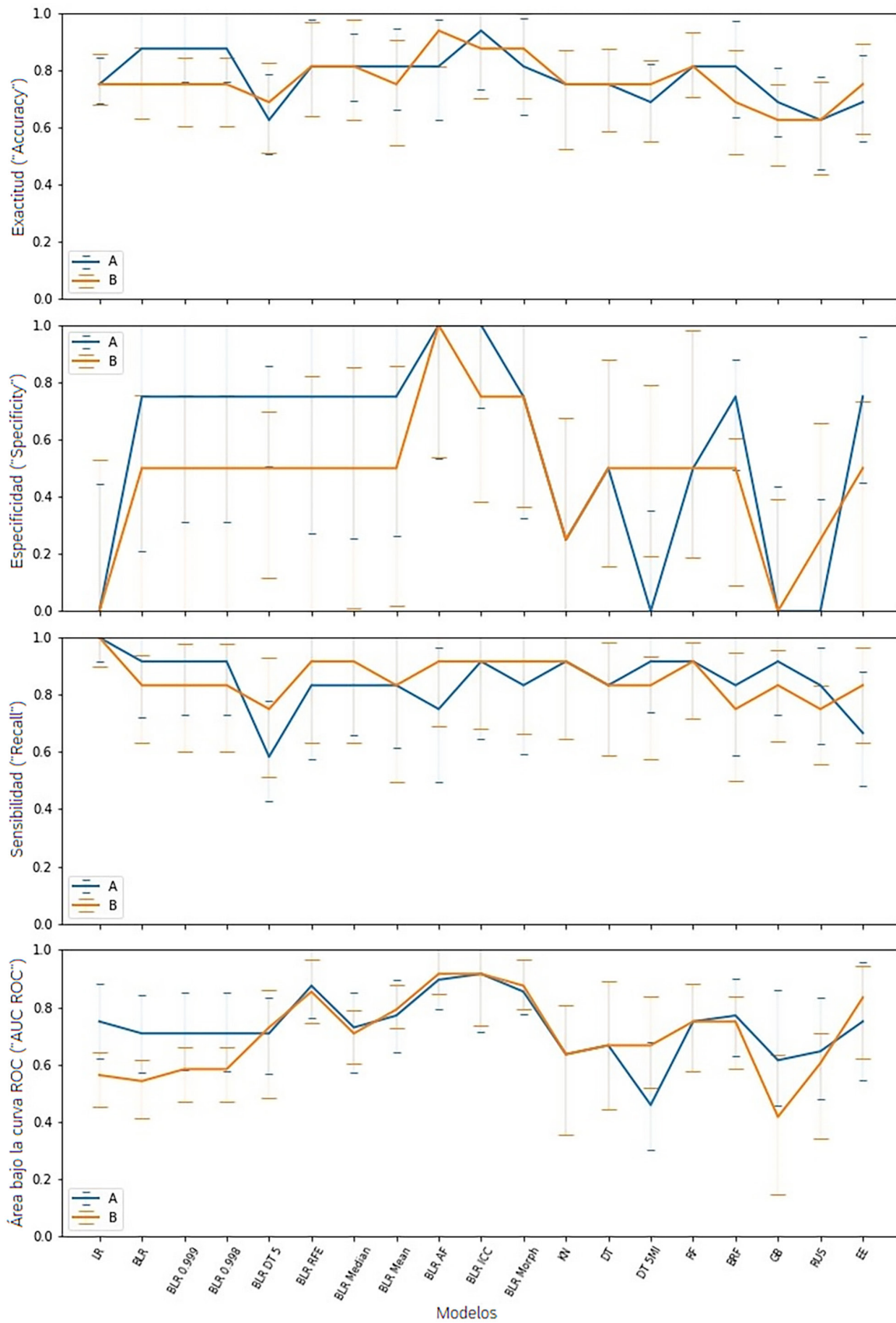
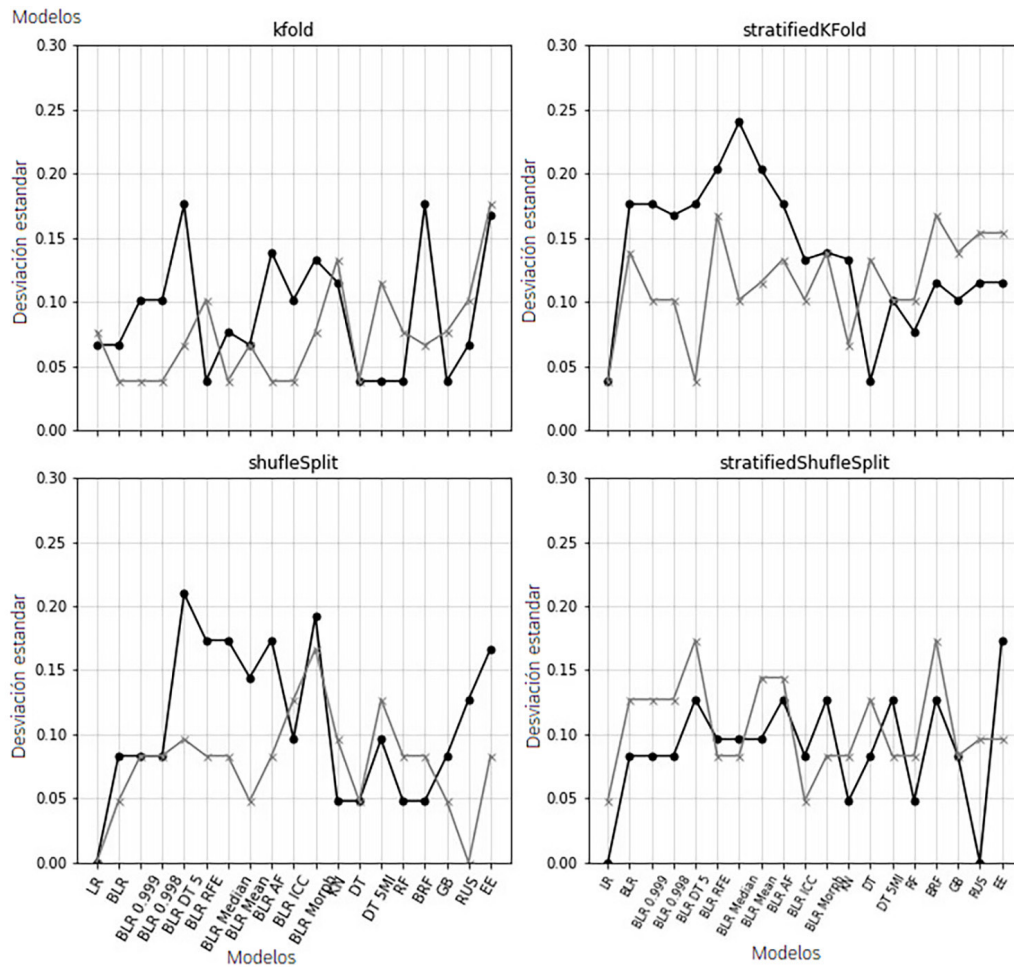


Fig. 4. Exactitud, especificidad, sensibilidad, y AUC ROC para todos los modelos evaluados en ambos grupos (A and B). Intervalos de confianza del 95%.



**Fig. 5.** Desviación estándar de la exactitud en cada grupo para los cuatro modelos de validación cruzada ( $k=3$ ): k-fold, shuffle split, stratified k-fold and stratified shuffle split.

- Regresión logística balanceada. Selección de características ANOVA.
- Regresión logística balanceada. Selección de características ICC > 0.9.
- Regresión logística balanceada. Selección de características sólo morfológicas.

Los resultados de exactitud, área bajo la curva ROC, especificidad y sensibilidad de estos métodos se presentan en la tabla 2.

En el caso de la selección de variables por el método ANOVA, se obtuvo un 100% de predicción de casos malignos para ambos grupos, único modelo en el que se dio esta situación, 1.00 (0.54-1.00) para el grupo A y 1.00 (0.54-1.00) para el B.

Considerando los resultados, las principales diferencias entre los métodos radican en dos puntos:

- en la validación cruzada, que en el caso de la selección de variables por el ICC > 0.9 presenta una des-

viación estándar menor en los valores de exactitud obtenidos en los métodos estratificados

- el área bajo la curva ROC, que presenta el valor más alto para la selección de variables por ICC > 0.9 en ambos grupos

## 4. Discusión

Los estudios radiómicos en adenomas son trabajos novedosos, no encontrándose un gran número de estudios en la literatura,<sup>31</sup> pero pueden ser muy útiles en el desarrollo de la práctica clínica. La gran mayoría de estos autores tienen un número relativamente pequeño de pacientes<sup>32,33</sup> debido a que en pocos casos se confirma el diagnóstico de adenoma mediante RM o estudios histológicos. Es interesante disponer de estos estudios como ocurre en otras patologías.<sup>34,35</sup> Los métodos de aprendizaje automático evaluados en nuestro estudio son ampliamente conocidos<sup>36</sup> y evaluados en estudios similares, tanto en otras pato-



**Tabla 2.** Exactitud, área bajo la curva ROC, especificidad y sensibilidad validados con cuatro métodos de validación cruzada ( $k = 3$ ) para los tres mejores métodos de aprendizaje automático.

Modelo	Grupo	Exactitud ("Accuracy")	Área bajo la curva ROC ("AUC ROC")	Especificidad ("Specificity")	Sensibilidad ("Recall or sensitivity")	Validación cruzada ( $k=3$ ) (valor medio / Desviación estándar)			
						k-F	S.S	S.k-F	S.S.S
BLR ICC	A	0.94 (0.74-1.00)	0.92 (0.63-1.00)	1.00 (0.71-1.00)	0.92 (0.65-1.00)	0.76 / 0.10	0.72 / 0.10	0.67 / 0.13	0.67 / 0.08
	B	0.88 (0.71-1.00)	0.92 (0.55-1.00)	0.75 (0.39-1.00)	0.92 (0.68-1.00)	0.78 / 0.04	0.72 / 0.13	0.69 / 0.10	0.69 / 0.04
BLR-AF	A	0.81 (0.62-0.97)	0.90 (0.44-1.00)	1.00 (0.54-1.00)	0.75 (0.50-0.96)	0.69 / 0.14	0.72 / 0.17	0.67 / 0.18	0.69 / 0.13
	B	0.94 (0.82-1.00)	0.92 (0.46-1.00)	1.00 (0.54-1.00)	0.92 (0.69-1.00)	0.69 / 0.04	0.67 / 0.08	0.67 / 0.13	0.75 / 0.14
BLR-Morph	A	0.81 (0.64-0.98)	0.85 (0.43-1.00)	0.75 (0.33-1.00)	0.83 (0.59-1.00)	0.73 / 0.13	0.72 / 0.10	0.76 / 0.14	0.81 / 0.13
	B	0.88 (0.71-1.00)	0.88 (0.50-1.00)	0.75 (0.38-1.00)	0.92 (0.67-1.00)	0.76 / 0.08	0.75 / 0.17	0.76 / 0.14	0.83 / 0.08

logías,<sup>37-39</sup> en otras técnicas de imagen,<sup>38,40</sup> como en estudios similares.<sup>41</sup>

No se han estudiado otras posibles modalidades de imagen, dado que se busca la utilidad en estos casos incidentales encontrados en los estudios CT.

Las características morfológicas (4) muestran valores de ICC superiores a 0.9, lo que demuestra un buen grado de concordancia entre los *VOI* definidos por los radiólogos.

Dada la matriz de confusión, consideramos que el mejor resultado es siempre aquel en el que se detectan mejor los casos malignos, dada la importancia crucial de identificar con fiabilidad y actuar con rapidez en los casos de sospecha de malignidad.

Dado que nuestra muestra está desequilibrada, debido a los pocos casos malignos (Fig. 1), el uso de un modelo balanceado está claramente justificado.

Los métodos de aprendizaje automático RBC y GBC no son recomendables para la clasificación de adenomas tal y como se han desarrollado en este trabajo, ya que los resultados obtenidos de exactitud, especificidad y área bajo la curva ROC se encuentran entre los valores más bajos obtenidos en la comparativa.

En el caso de la selección de características fuertemente correlacionadas se recurre a valores altos de correlación (0.999 y 0.998) con el fin de enfatizar la pequeña repercusión de su omisión en nuestro estudio, respecto a su inclusión en el método de regresión logística balanceada, a pesar de su clara correlación.

La combinación de los métodos de selección de variables ANOVA y ICC selecciona las variables que hemos catalogado como morfológicas, lo que permite pensar que estas características son, de las seleccionadas, las que mejor explican la varianza de la variable dependiente.

En el caso de la validación *k-fold* las submuestras analizadas pueden no tener casos malignos o tener un número significativamente elevado de casos malignos, lo que mejorará la exactitud en ambas situaciones. Esto también afecta a la desviación estándar observada en la Fig. 5, donde la mayoría de los modelos presentan desviaciones mayores para las validaciones cruzadas con algoritmos estratificados. A pesar de presentar valores ligeramente superiores en algunos métodos, estas discrepancias entre los valores de exactitud de cada submuestra son pequeñas, por lo que todos los modelos evaluados pueden considerarse aceptables.

El resultado obtenido en el método de regresión logística balanceada con la selección de las variables más importantes para el modelo de árbol de decisión es inferior al resto (Fig. 4), ya que el grado de importancia evaluado para este modelo no lo es necesariamente en el caso de la regresión logística balanceada.

Así, podemos considerar este modelo, BLR ICC > 0.9, como el de mejor capacidad para clasificar correctamente los tumores suprarrenales en nuestro estudio. Observamos que esto se debe a que hemos utilizado las variables con mayor correlación entre los diferentes estudios al ser realizados en diferentes equipos (5 TC) y contorneados por diferentes especialistas, adaptándose de forma más eficiente a nuestra institución al estar evaluando lesiones encontradas incidentalmente.

Los métodos evaluados son dependientes de la muestra utilizada limitando su extrapolación a otras muestras de estudio.

El método de aprendizaje automático finalmente seleccionado será el obtenido para el grupo A, ya que se obtiene un 100% de predicción de lesiones malignas.

Los resultados obtenidos son consistentes incluso con los presentados en estudios realizados con un menor sesgo debido al uso de una única unidad de TC<sup>32,42,43</sup>. Como en nuestro estudio, otros autores<sup>32, 41,42</sup> obtienen mejores resultados con el empleo de la regresión logística como método de aprendizaje automático.

## 5. Conclusiones

En un estudio radiómico, la evaluación de diferentes algoritmos será decisiva para encontrar un método de aprendizaje automático satisfactorio. Para ello, evaluamos los modelos en términos de exactitud, área bajo la curva ROC, especificidad y sensibilidad, seguido de diferentes métodos de validación cruzada que nos permitieron evaluar la muestra frente a ellos.

En nuestro estudio, el método que mejor clasifica los tumores suprarrenales es el modelo de regresión logística balanceada utilizando selección de variables por ICC > 0.9. Se obtuvieron unos resultados de especificidad de 1.00 (0.71-1.00) (100% de detección).

La obtención de un método de aprendizaje automático tan eficiente nos permite adecuar las pruebas de los pacientes en función de los tiempos de espera, dando prioridad en la lista a aquellos pacientes con sospecha de malignidad.

## 6. Conflictos

Los autores declaran no tener ningún conflicto de intereses.

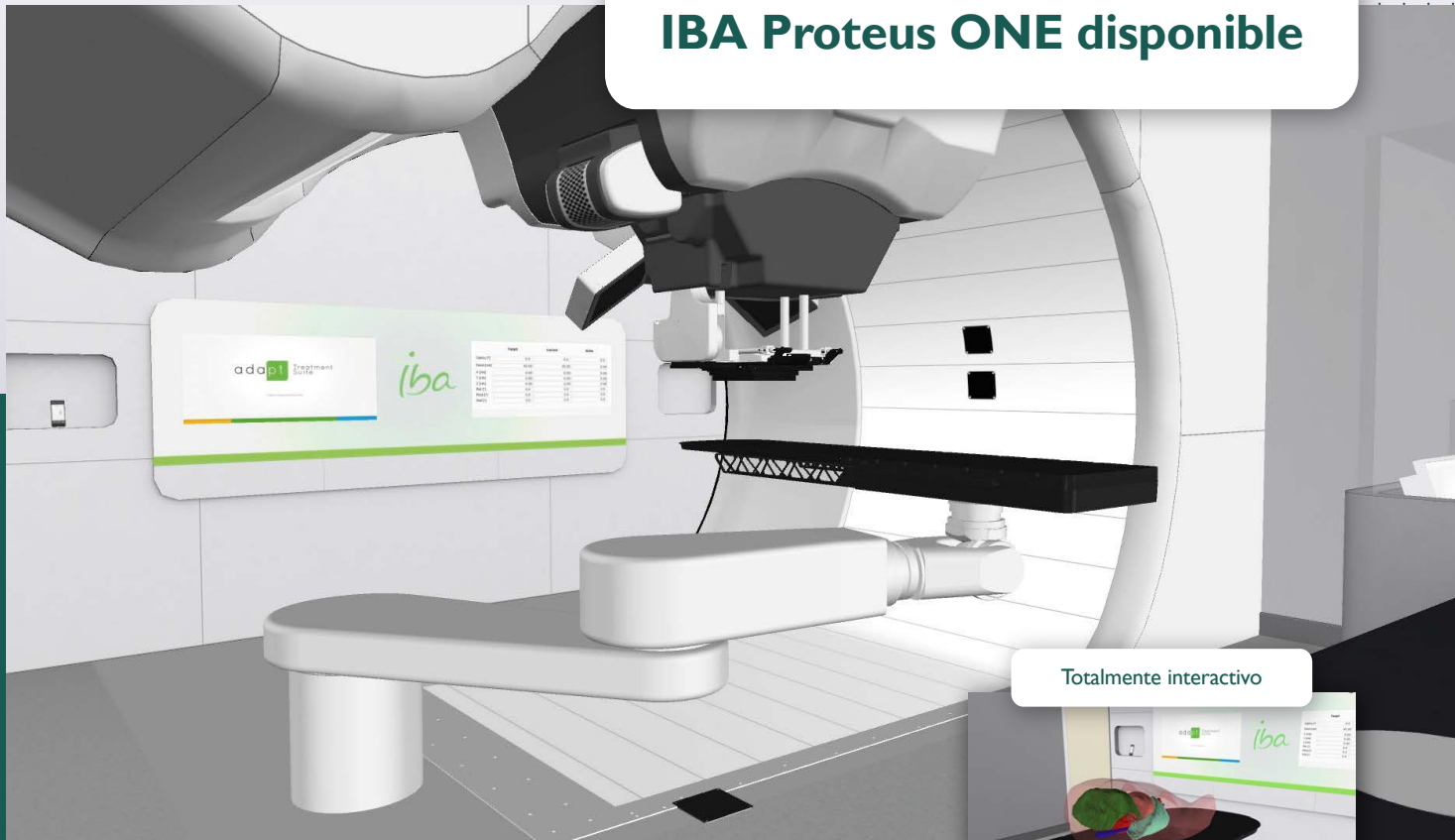
## References

- Saruta T, Suzuki H, Shibata H. Adrenal incidentaloma. *Nippon Naibunpi Gakkai zasshi* 1993; 69(5): 509–519. [https://doi.org/10.1507/endocrine.1927.69.5\\_509](https://doi.org/10.1507/endocrine.1927.69.5_509)
- Mayo-Smith WW, Song JH, Boland GL, et al. Management of Incidental Adrenal Masses: A White Paper of the ACR Incidental Findings Committee. *J. Am. Coll. Radiol.* 2017; 14(8): 1038–1044. <http://dx.doi.org/10.1016/j.jacr.2017.05.001>
- Terzolo M, Stigliano A, Chiodini I, et al. AME position statement on adrenal incidentaloma. *Eur. J. Endocrinol.* 2011; 164(6): 851–870. <https://doi.org/10.1530/EJE-10-1147>
- Cawood TJ, Hunt PJ, O'Shea D, Cole D, Soule S. Recommended evaluation of adrenal incidentalomas is costly, has high false-positive rates and confers a risk of fatal cancer that is similar to the risk of the adrenal lesion becoming malignant; time for a rethink? *Eur. J. Endocrinol.* 2009; 161(4): 513–527. <https://doi.org/10.1530/EJE-09-0234>
- Welch HG, Black WC. Overdiagnosis in cancer. *J. Natl. Cancer Inst.* 2010; 102(9): 605–613. <https://doi.org/10.1093/jnci/djq099>
- Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. *Radiology* 2016; 278(2): 563–577. <https://doi.org/10.1148/radiol.2015151169>
- Larue RTHM, Defraene G, De Ruyscher D, Lambin P, Van Elmpt W. Quantitative radiomics studies for tissue characterization: A review of technology and methodological procedures. *Br. J. Radiol.* 2017; 90(1070). <https://doi.org/10.1259/bjr.20160665>
- Ligero M, Jordi-Ollero O, Bernatowicz K, et al. Minimizing acquisition-related radiomics variability by image resampling and batch effect correction to allow for large-scale data analysis. *Eur. Radiol.* 2021; 31(3): 1460–1470. <https://doi.org/10.1007/s00330-020-07174-0>
- Nioche C, Orhac F, Boughdad S, et al. Lifex: A freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res.* 2018; 78(16): 4786–4789. <https://doi.org/10.1158/0008-5472.CAN-18-0125>
- Nioche C, Orhac F, Buvat I. Texture — User Guide Local Image Features Extraction. <https://www.lifexsoft.org/images/phocagallery/documentation/ProtocolTexture/UserGuide/TextureUserGuide.pdf> (2023).
- Carrasco JL, Jover L. Estimating the Generalized Concordance Correlation Coefficient through Variance Components. *Biometrics* 2003; 59(4): 849–858. <https://doi.org/10.1111/j.0006-341X.2003.00099.x>
- Lahey MA, Downey RG, Saal FE. Intraclass correlations: There's more there than meets the eye. *Psychol. Bull.* 1983; 93(3): 586–595. <https://doi.org/10.1037/0033-2909.93.3.586>
- Raju VNG, Lakshmi KP, Jain VM, Kalidindi A, Padma V. Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification. *Proc. 3rd Int. Conf. Smart Syst. Inven. Technol. ICSSIT 2020 2020; (Icassit): 729–735.* <https://doi.org/10.1109/ICSSIT48917.2020.9214160>
- Peeken JC, Shouman MA, Kroenke M, et al. A CT-based radiomics model to detect prostate cancer lymph node metastases in PSMA radioguided surgery patients. *Eur. J. Nucl. Med. Mol. Imaging* 2020; 47(13): 2968–2977. <https://doi.org/10.1007/s00259-020-04864-1>
- Zhang H, Li Z, Shahriar H, Tao L, Bhattacharya P, Qian Y. Improving prediction accuracy for logistic regression on imbalanced datasets. *Proc. - Int. Comput. Softw. Appl. Conf.* 2019; 1: 918–919. <https://doi.org/10.1109/COMPSAC.2019.00140>
- Zhou CY, Chen YQ. Improving nearest neighbor classification with cam weighted distance. *Pattern Recognit.* 2006; 39(4): 635–645. <https://doi.org/10.1016/j.patcog.2005.09.004>
- Quinlan JR. Induction of decision trees. *Mach. Learn.* 1986; 1(1): 81–106. <https://doi.org/10.1007/bf00116251>
- Jin Z, Shang J, Zhu Q, Ling C, Xie W, Qiang B. RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 2020; 12343 LNCS: 503–515. [https://doi.org/10.1007/978-3-030-62008-0\\_35](https://doi.org/10.1007/978-3-030-62008-0_35)
- de Jesus FM, Yin Y, Mantzorou-Kyriaki E, et al. Machine learning in the differentiation of follicular lymphoma from diffuse large B-cell lymphoma with radiomic [18F]FDG PET/CT features. *Eur. J. Nucl. Med. Mol. Imaging* 2022; 49(5): 1535–1543. <https://doi.org/10.1007/s00259-021-05626-3>
- Masters T. Multilayer Feedforward Networks. *Pract. Neural Netw. Recipes C++* 1993; 77–116.

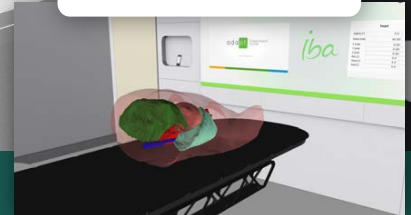
21. Seiffert C, Khoshgoftaar TM, Van Hulse J, Napolitano A. RUSBoost: A hybrid approach to alleviating class imbalance. *IEEE Trans. Syst. Man, Cybern. Part A Systems Humans* 2010; 40(1): 185–197. <https://doi.org/10.1109/TSMCA.2009.2029559>
22. Martínez-García JM, Suárez-Araujo CP, Báez PG. SNEOM: A sanger network based extended over-sampling method. Application to imbalanced biomedical datasets. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 2012; 7666 LNCS(PART 4): 584–592. [https://doi.org/10.1007/978-3-642-34478-7\\_71](https://doi.org/10.1007/978-3-642-34478-7_71)
23. Sharma N V., Yadav NS. An optimal intrusion detection system using recursive feature elimination and ensemble of classifiers. *Microprocess. Microsyst.* 2021; 85(June): 104293. <https://doi.org/10.1016/j.micpro.2021.104293>
24. Hasan KA, Hasan MAM. Prediction of Clinical Risk Factors of Diabetes Using Multiple Machine Learning Techniques Resolving Class Imbalance. *ICCIT 2020 - 23rd Int. Conf. Comput. Inf. Technol. Proc.* 2020; (December). <https://doi.org/10.1109/ICCIT51783.2020.9392694>
25. Ding H, Feng PM, Chen W, Lin H. Identification of bacteriophage virion proteins by the ANOVA feature selection and analysis. *Mol. Biosyst.* 2014; 10(8): 2229–2235. <https://doi.org/10.1039/c4mb00316k>
26. Berrar D. Cross-validation. *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.* 2018; 1–3(January 2018): 542–545. <https://doi.org/10.1016/B978-0-12-809633-8.20349-X>
27. Risk C, James PMA. Optimal Cross-Validation Strategies for Selection of Spatial Interpolation Models for the Canadian Forest Fire Weather Index System. *Earth Sp. Sci.* 2022; 9(2): 1–17. <https://doi.org/10.1029/2021EA002019>
28. Liu X, Lu J, Zhang G, et al. A machine learning approach yields a multiparameter prognostic marker in liver cancer. *Cancer Immunol. Res.* 2021; 9(3): 337–347. <https://doi.org/10.1158/2326-6066.CIR-20-0616>
29. Harrigan MP, Sultan MM, Hernández CX, et al. MSMBUILDER: Statistical Models for Biomolecular Dynamics. *Biophys. J.* 2017; 112(1): 10–15. <https://doi.org/10.1016/j.bpj.2016.10.042>
30. Raschka S. Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. 2018; <http://arxiv.org/abs/1811.12808>
31. Crimi F, Quaia E, Cabrelle G, et al. Diagnostic Accuracy of CT Texture Analysis in Adrenal Masses: A Systematic Review. *Int. J. Mol. Sci.* 2022; 23(2). <https://doi.org/10.3390/ijms23020637>
32. Elmohr MM, Fuentes D, Habra MA, et al. Machine learning-based texture analysis for differentiation of large adrenal cortical tumours on CT. *Clin. Radiol.* 2019; 74(10): 818.e1-818.e7. <http://dx.doi.org/10.1016/j.crad.2019.06.021>
33. Lisa M. Ho, Ehsan Samei, Maciej A. Mazurowski, Yuese Zheng, Brian C. Allen, Rendon C. Nelson and DM. Can Texture Analysis Be Used to Distinguish Benign From Malignant Adrenal Nodules on Unenhanced CT, Contrast-Enhanced CT, or In-Phase and Opposed-Phase MRI? *Am. J. Roentgenol.* 2019; 212:3(March): 554–561. <https://doi.org/10.2214/AJR.18.20097>
34. Sun PAN, Wang D, Mok VCT, Shi LIN. Comparison of Feature Selection Methods and Machine Learning Classifiers for Radiomics Analysis in Glioma Grading. 2019; 7.
35. Ni M, Wang L, Yu H, et al. Radiomics Approaches for Predicting Liver Fibrosis With Nonenhanced T1-Weighted Imaging: Comparison of Different Radiomics Models. 2021; <https://doi.org/10.1002/jmri.27391>
36. Prinzi F, Currier T, Gaglio S, Vitabile S. Shallow and deep learning classifiers in medical image analysis. *Eur. Radiol. Exp.* 2024; 2. <https://doi.org/10.1186/s41747-024-00428-2>
37. Naseri H, Skamene S, Tolba M, et al. Radiomics - based machine learning models to distinguish between metastatic and healthy bone using lesion - center - based geometric regions of interest. *Sci. Rep.* 2022; 1–13. <https://doi.org/10.1038/s41598-022-13379-8>
38. Decoux A, Duron L, Habert P, et al. Comparative performances of machine learning algorithms in radiomics and impacting factors. *Sci. Rep.* 2023; 1–10. <https://doi.org/10.1038/s41598-023-39738-7>
39. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine Learning methods for Quantitative Radiomic Biomarkers. *Nat. Publ. Gr. n.d.*; 1–11. <https://doi.org/10.1038/srep13087>
40. Shao S, Zheng N, Mao N, et al. A triple-classification radiomics model for the differentiation of pleomorphic adenoma, Warthin tumour, and malignant salivary gland tumours on the basis of diffusion-weighted imaging. *Clin. Radiol.* 2021; 76(6): 472.e11-472.e18. <http://dx.doi.org/10.1016/j.crad.2020.10.019>
41. Qi S, Zuo Y, Chang R, Huang K, Liu J, Zhang Z. Using CT radiomic features based on machine learning models to subtype adrenal adenoma. 2023; 1–12. <https://doi.org/10.1186/s12885-023-10562-6>
42. Zheng Y, Liu X, Zhong Y, Lv F, Yang H. A Preliminary Study for Distinguish Hormone-Secreting Functional Adrenocortical Adenoma Subtypes Using Multiparametric CT Radiomics-Based Machine Learning Model and Nomogram. *Front. Oncol.* 2020; 10(September): 1–11. <https://doi.org/10.3389/foonc.2020.570502>
43. Winkelmann MT, Gassenmaier S, Walter SS, et al. Differentiation of adrenal adenomas from adrenal metastases in single-phased staging dual-energy CT and radiomics. *Diagnostic Interv. Radiol.* 2022; 28(3): 208–216. <https://doi.org/10.5152/dir.2022.21691>

# Simulador de protonterapia VERT para formación

**IBA Proteus ONE disponible**



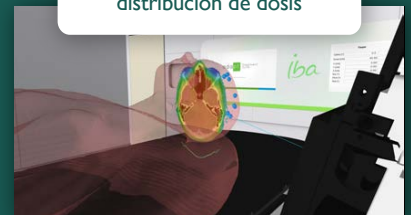
Totalmente interactivo



Modelos de administración de dosis



Comparación de distribución de dosis



Mejora el plan de estudios actual con tecnología punta.



Introduce las últimas técnicas clínicas sin necesidad de prácticas en clínica.



Compara los planes de tratamiento con protones y fotones.



Garantiza que los estudiantes reciban enseñanzas y competencias más actualizadas.



Póngase en la vanguardia de la enseñanza de radioterapia.

**VERTUAL**

Distribuido en España por:

**APLICACIONES TECNOLÓGICAS** | **RADIATION PROTECTION & MEDICAL**

Tel: (+34) 91 352 54 54 | [atfísica@atfísica.com](mailto:atfísica@atfísica.com) | [www.atfísica.com](http://www.atfísica.com)