



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
MAESTRIA EN CIENCIAS MEDICAS, ODONTOLOGICAS Y DE LA SALUD.
CAMPO DISCIPLINARIO: CIENCIAS MEDICAS

ASOCIACIÓN DE LAS MUTACIONES SOMÁTICAS CON EL FENOTIPO METABÓLICO DE LA
IMAGEN MOLECULAR PET/CT CON 18F-FDG EN CÁNCER DE PULMÓN DE CÉLULAS NO
PEQUEÑAS.

TESIS
QUE PARA OPTAR POR EL GRADO DE:
MAESTRIA EN CIENCIAS MEDICAS

PRESENTA:
ROBERTO PEDRERO PIEDRAS

TUTOR:
DR. ODILÓN FÉLIX QUIJANO CASTRO
INSTITUTO MEXICANO DEL SEGURO SOCIAL

MIEMBRO DEL COMITÉ TUTOR:
DR. FRANCISCO OSVALDO GARCÍA PÉREZ
INSTITUTO NACIONAL DE CANCEROLOGIA

CIUDAD UNIVERSITARIA, CD. MX. AGOSTO 2023



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Índice

1. MARCO TEÓRICO.....	4
2. JUSTIFICACIÓN.	8
3. PLANTEAMIENTO DEL PROBLEMA.	9
4. PREGUNTA DE INVESTIGACIÓN.....	10
5. OBJETIVOS.	10
6. HIPÓTESIS DE TRABAJO.	10
7. MATERIAL Y MÉTODOS:.....	10
8. ASPECTOS ÉTICOS.	17
9. ASPECTOS DE BIOSEGURIDAD.....	18
10. RECURSOS, FINANCIAMIENTO Y FACTIBILIDAD.	18
11. ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS.....	19
12. DISCUSIÓN.	35
13. CONCLUSIONES.....	36
14. BIBLIOGRAFÍA.....	36

GLOSARIO DE ABREVIATURAS

^{18}F : Isotopo radiactivo Fluor-18 (^{18}F).

ALK: Gen del receptor de la tirosina quinasa del linfoma anaplásico.

AT: Análisis textural.

CPCNP: Cáncer de pulmón de células no pequeñas.

CT: Tomografía helicoidal computarizada.

EGFR: Receptor del factor de crecimiento epidérmico.

^{18}F -FDG: 2-deoxi-2- ^{18}F fluoro-D-glucosa.

GLCM: Matriz de co-ocurrencia de niveles de gris.

GLZSM: Matriz de tamaño de zona de nivel de gris.

IBSI: Image Biomarker Standardization Initiative

KRAS: Oncogén Kirsten Rat Sarcoma Viral.

NGTDM: Matriz de diferencia de tono de gris vecinal

PACS: Picture Archiving and Communication System (Sistema de Comunicación y Archivado de Imágenes).

PET: Tomografía por emisión de positrones.

SPECT: Tomografía por emisión de fotón único.

SUV: Valor de captación estandarizado.

TKI: Inhibidores de tirosina cinasa.

TNM: Escala de estadificación para el cáncer; T: Tumor, N: Ganglios linfáticos, M: Metástasis.

1. Marco Teórico.

1.1 Análisis radiómico y cáncer de pulmón.

El análisis radiómico ha revolucionado la evaluación cuantitativa de las imágenes médicas, revelando datos cuantitativos dentro de las imágenes más allá de la interpretación visual limitada. Los datos cuantitativos se extraen de imágenes de alta calidad y se convierten en información que se correlaciona con características propias de la entidad patogénica como pueden ser genéticas, variantes patogénicas, histopatológicas, clínicas o de pronóstico relevantes. Los datos extraídos, se pueden emplear para el diagnóstico, estadificación, estratificación de riesgo, pronóstico y evaluación de respuesta a la terapia. Las características derivadas de PET/CT con 18F-FDG por ejemplo, han mostrado un valor particularmente promisorio en el diagnóstico y caracterización de nódulos pulmonares de naturaleza maligna, diferenciación de subtipos de cáncer de pulmón y predicción de la respuesta a diferentes terapias, así como de la supervivencia (1).

Caracterizar la heterogeneidad tumoral (macroscópica, fisiológica, microscópica o genética) es un desafío en oncología, ya que la coexistencia de diferentes clones celulares en un tumor afecta la evolución del cáncer, por lo que en la actualidad es crucial tener en cuenta el estado mutacional del tumor para seleccionar el enfoque terapéutico más adecuado.

El cáncer de pulmón lidera las causas de muerte relacionadas con cáncer en los Estados Unidos, presenta una incidencia estimada para 2022 de más de 236,740 nuevos casos, con un índice de supervivencia a 5 años de 21%, de los cuales, el cáncer de pulmón de células no pequeñas (CPCNP) representa aproximadamente el 80% de todos los casos (2). Los pacientes con CPCNP con mutación del receptor de factor de crecimiento epidérmico (EGFR) son un grupo especial de individuos, ya que la etiología del cáncer, las características clinicopatológicas, el pronóstico y los paradigmas del tratamiento son todos diferentes cuando se comparan con otros pacientes con CPCNP.

1.2 El estado mutacional y la elección del tratamiento.

Los avances en la genómica del cáncer han demostrado que el CPCNP se rige por mutaciones somáticas en oncogenes clave como EGFR, el cual está presente normalmente en la superficie de células epiteliales y es frecuentemente sobreexpresado por una variedad de malignidades, siendo más frecuente en mujeres que en hombres, pacientes no fumadores vs fumadores, adenocarcinomas y etnicidad asiática. La mutación en EGFR, no está asociada con la edad del paciente al diagnóstico, estadio clínico, histología bronquioalveolar o la supervivencia global (3–6).

Las mutaciones de EGFR en CPCNP son múltiples, ocurren en el dominio de tirosina cinasa (TK) entre los exones 18 y 21, se ha descrito que la delección del exón 19 ocurre en hasta 50% y la mutación puntual L858R en el exón 21 en 40% siendo las más frecuentes, en cambio la inserción en el exón 20 ocurre en 3 a 7%. Otras mutaciones de EGFR menos frecuentes, incluyen G718X, S768I, L861Q, etc (7).

Estos descubrimientos han llevado al uso y aprobación regulatoria de inhibidores de la TK específicos de EGFR para el tratamiento de pacientes con mutación de dicho gen, mostrando una gran mayoría de pacientes con mutaciones sensibilizantes de EGFR que responden favorablemente a los inhibidores de TK dirigidos a EGFR de primera y segunda generación cuando se utilizan en el entorno de primera línea(8–10). Por otro lado, se ha encontrado que los pacientes con mutaciones KRAS responden pobremente al uso tanto de inhibidores de la tirosina cinasa como a la quimioterapia convencional basada en cisplatino(11,12). Asimismo, se sabe que los tumores con mutación EGFR en general no responden adecuadamente al tratamiento con inhibidores de puntos de control inmunitarios anti PD-1 o anti PDL-1 cuando se compara con pacientes con expresión EGFR wild-type (13).

Dada la importancia en el manejo del estatus mutacional en la eficacia de las terapias de CPCNP, su identificación es crucial para la selección del tratamiento más efectivo y determinar su pronóstico. El estándar actual de atención utiliza pruebas moleculares para identificar el estado de mutación basado en biopsias de tejido tumoral o resección quirúrgica, sin embargo, pueden estar limitados por procedimientos invasivos, largos tiempos de procesamiento, errores en el muestreo y muestras de tejido que no siempre se encuentran fácilmente disponibles (5,14).

1.3 La biología tumoral por imagen PET/CT con Fluorodesoxiglucosa (FDG).

El abordaje inicial del nódulo pulmonar y/o cáncer de pulmón es multimodal, sin embargo, el método que sobresale entre todos es la imagen híbrida PET/CT o tomografía por emisión de positrones (PET) con ¹⁸F-FDG con adquisición de tomografía computarizada simultánea (CT), puesto que no es invasivo y proporciona información imagenológica del metabolismo glucolítico, así como características morfológicas, siendo parte del abordaje inicial estandarizado para el CPCNP (15,16).

El fenotipo metabólico tumoral obtenido en las imágenes de PET puede ser indicativo de la biología subyacente que es de gran utilidad en la caracterización de los tumores al poder asociarse con las mutaciones somáticas, sin embargo, en la literatura publicada existen resultados inconsistentes (17–22). Estos reportes inconsistentes pueden deberse al hecho de que la simple medición del valor estandarizado de captación (SUV) falla al capturar las relaciones espaciales entre los vóxeles de la imagen, los cuales pueden ser más informativos de la biología de las mutaciones y describir el grado de heterogeneidad tumoral (23,24), dichas relaciones y detalles son susceptibles de ser mejor valorados y entendidos mediante el análisis radiómico. El término heterogeneidad suele tener diferentes significados según la modalidad de la imagen. Cuando se considera el componente PET, se refiere a la distribución espacial de la captación del radiotrazador, que puede reflejar, según el radiotrazador utilizado, la combinación de procesos biológicos subyacentes como el metabolismo de la glucosa u otros sustratos, la hipoxia, la proliferación celular, la vascularización y la necrosis (25). Asimismo, se ha descrito la asociación de características cuantitativas de la heterogeneidad tumoral con fluorodesoxiglucosa (FDG) para diferenciar la expresión de genes entre tejido tumoral y tejido sano, relacionando dichos genes con las vías de proliferación celular u otras vías involucradas con la apoptosis (26).

1.4 El análisis radiómico.

El cáncer es un ecosistema complejo, evolutivo y de múltiples escalas que tiene alta heterogeneidad espacio - temporal. La obtención de imágenes es fundamental para su investigación porque es posible caracterizar de forma no destructiva y longitudinal las variaciones espaciales en el fenotipo y el entorno tumoral, para que la dinámica del sistema a lo largo del tiempo pueda capturarse cuantitativamente. El efecto de un genoma alterado se manifiesta en los genes que se expresan (transcriptoma) que pueden verse afectados por eventos genéticos, como mutaciones o polimorfismos de un solo nucleótido, por eventos epigenéticos, como la metilación del ADN, la metilación de histonas, la acetilación, la dinámica de la cromatina, y la actividad de los micro-ARN (27).

Los múltiples fenotipos y escalas espaciales que interactúan se pueden entender mejor con modelos matemáticos y computacionales apropiados. Estos diferentes niveles de organización pueden evaluarse con imágenes moleculares, funcionales o anatómicas no invasivas. Los tres tipos de imágenes están interrelacionados y cada una tiene características distintivas. Las imágenes moleculares están relegadas a medir los niveles o actividades de macromoléculas específicas o vías metabólicas *in vivo*. Las imágenes funcionales se dedican a medir funciones de órganos específicos, como la perfusión y la densidad celular. En el nivel más alto de organización, las imágenes anatómicas pueden ser muy cuantitativas y pueden identificar la heterogeneidad subyacente en las condiciones microambientales, la expresión génica y los fenotipos metabólicos, a través de la radiómica.

El término radiómica (radiomics), se introdujo por primera vez en 2010, que es la extracción de características cuantitativas de imágenes médicas. Relaciona éstas con patrones de expresión génica utilizando enfoques bioinformáticos sofisticados y a veces se denomina "radiogenómica". La hipótesis central de la radiómica del cáncer es que las características de las imágenes tumorales reflejan patrones de expresión génica subyacentes. Los cambios en la expresión de genes específicos pueden afectar parámetros específicos de imagen, los cuales pueden medirse mediante resonancia magnética o tomografía por emisión de positrones (28).

Durante más de una década se ha estudiado cada vez con más interés y profundidad el análisis de las imágenes médicas mediante procesamiento radiómico, actualmente consolidándose como un campo de estudio con aplicaciones potenciales muy diversas y de gran relevancia, sin embargo, son muchos los desafíos que se han ido superando a través de tiempo y muchos más los que quedan por aclararse.

La radiómica es un campo dedicado a la obtención y análisis automatizado de un gran volumen de datos extraídos de las características cuantitativas de las imágenes médicas, como el PET/CT con 18F-FDG y hacer asociaciones con otros datos; genéticos, patológicos, clínicos o pronósticos, ofreciendo una alternativa de abordaje donde la extracción de datos a partir de imágenes y la consecuente utilización de éstos en algoritmos de caracterización para la creación de modelos de soporte de decisión (predictivos de pronóstico y respuesta al tratamiento) de alta exactitud y eficiencia, pudiendo estudiar el comportamiento de lesiones y mejorar la toma de decisiones clínicas, ya que tiene el potencial de valorar características específicas que son difíciles de identificar por la visión humana además de ser observador independiente. Se ha iniciado el estudio de esta herramienta en pacientes oncológicos, pero es potencialmente aplicable a todas las enfermedades. La radiómica se puede aplicar a la mayoría de las modalidades de imagen como tomografía computarizada (TC), resonancia magnética (RM), ultrasonido e imágenes híbridas (SPECT/CT y PET/CT)(29–31).

En la práctica clínica, la respuesta tumoral a la terapia se mide solo en una dimensión del tumor (criterios de RECIST 1.1). Aunque un cambio en el tamaño tumoral puede indicar respuesta al tratamiento, este normalmente no predice la supervivencia global o supervivencia libre de progresión. Por su parte, el análisis radiómico, es un campo emergente que convierte los datos de una imagen en un espacio de características dimensionales de alta capacidad, utilizando un gran número de algoritmos de caracterización para los datos extraídos automáticamente (32).

La alta heterogeneidad tumoral está frecuentemente asociada con pobre pronóstico debido a que la resistencia a las terapias puede ser desarrollada en subpoblaciones de células tumorales (33,34), de manera que la cuantificación precisa de la heterogeneidad tumoral por las imágenes PET puede proveer información importante para la identificación del estatus de mutación y la medicina de precisión.

La heterogeneidad en los fenotipos tumorales puede ser descrita cuantitativamente a través de las características radiómicas(23,35,36). Algunos autores han investigado la relación entre el patrón metabólico de captación de FDG (2-deoxi-2-[18F]fluoro-D-glucosa) y la apariencia histopatológica intratumoral, observando variaciones en la captación de FDG que correlacionan directamente con la presencia de células tumorales y la ausencia o menor captación se ha correlacionado con la presencia de tejido estromal y necrótico, por lo cual, se sugiere evaluar con precaución el comportamiento tumoral con base en la heterogeneidad tumoral(35). Las características radiómicas emplean modelos matemáticos avanzados para cuantificar las relaciones espaciales entre los vóxeles de la imagen que representa la heterogeneidad tumoral (37).

El análisis de las características radiómicas puede realizarse de forma complementaria al TNM, es bien sabido el uso clínico de dicha clasificación para la estratificación de riesgo y la toma de

decisiones en el tratamiento de los pacientes, sin embargo, el sistema TNM está basado primordialmente en la resecabilidad tumoral, mientras que un gran número de pacientes con CPCNP recibirá tratamiento primario a base de radioterapia sola o combinada con quimioterapia (37).

La utilización de los datos radiómicos comienza con la adquisición de imágenes de alta calidad. A partir de estas imágenes y trazado de una región de interés (ROI) que contiene todo el tumor o subregiones (es decir, segmentos) dentro del tumor. Estos segmentos son delimitados en sus tres dimensiones (3D) y a partir de ellos se extraen características cuantitativas y volúmenes para generar un informe que se coloca en una base de datos junto con otros datos, tales como datos clínicos y genómicos. Estos datos son entonces extraídos para desarrollar modelos de diagnóstico, predictivo o de pronóstico para los resultados de interés.

El número de características de imagen calculadas es típicamente mayor que el número de pacientes estudiados, por lo que la correcta selección de características y la construcción de modelos predictores es crucial para así evitar el sobreajuste de los algoritmos (38–43).

Las características radiómicas de primer orden estiman las propiedades de los valores de vóxeles individuales, ignorando la interacción espacial entre ellos (no pueden considerarse características "texturales", porque no diferencian los arreglos y patrones espaciales), mientras que las características de segundo orden y de orden superior estiman las propiedades de dos o más valores de vóxel que ocurren en ubicaciones específicas en relación entre sí. Para estas características de análisis textural de segundo y orden superior, los primeros pasos generalmente consisten en volver a muestrear o interpolar las cuadrículas de vóxeles no cúbicos en vóxeles cúbicos (rara vez se lleva a cabo) y realizar la cuantificación (realizada sistemáticamente, también llamada *discretización*, reducción de *muestreo* o *re-muestreo*) de las intensidades originales en un conjunto discreto de valores. Este número determina el tamaño de las matrices que se construyen y en las que posteriormente se calculan las métricas de análisis textural. Se han propuesto varios métodos para realizar esta cuantificación, como una distribución lineal en un número determinado de contenedores (por ejemplo, 32 o 64), el uso de un algoritmo de agrupamiento (Max-Lloyd) o en contenedores de ancho fijo (por ejemplo, 0,25 SUV o 0,5 SUV). El enfoque y el valor de cuantificación elegidos pueden tener un impacto importante en las métricas de análisis textural resultantes, así como en su relación con el volumen del tumor o el SUV máximo y, por lo tanto, es un factor importante que no debe pasarse por alto (44).

El segundo paso consiste en construir las matrices de textura, de las cuales existen varias (por ejemplo, matriz de co-ocurrencia de niveles de gris, GLCM; matriz de diferencia de tono de gris vecinal, NGTDM; y matriz de tamaño de zona de nivel de gris, GLZSM) y que se pueden construir en diferentes maneras. Por ejemplo, las matrices de coocurrencia cuantifican las relaciones entre pares de vóxeles. Por lo general, se definen de acuerdo con una dirección espacial dada y una distancia dada entre los pares de vóxeles. Para un análisis 3D, a menudo se consideran 13 direcciones y se construye una matriz por dirección. Luego se calcula la métrica de análisis textural (AT) en cada una de estas matrices y se promedian los 13 valores resultantes. Por lo general, la distancia se establece en un vóxel. La modificación de estas opciones (p. ej., usar solo una matriz para todas las direcciones) puede conducir a diferentes distribuciones de características de análisis textural, valor complementario asociado con otras métricas y, como consecuencia, valor clínico general.

En la gran mayoría de los estudios publicados, no se realizó un análisis multivariado que incluyera posibles factores de confusión, ni una corrección para múltiples pruebas, y muy rara vez se utilizó un aprendizaje automático sólido con validación cruzada.

Las características radiómicas extraídas de PET/CT con 18F-FDG pueden distinguir entre lesiones pulmonares primarias y metastásicas. La combinación del análisis radiómico y los métodos de

aprendizaje profundo y automatizado puede mejorar la exactitud en los diagnósticos oncológicos y puede aplicarse a la práctica clínica para promover el desarrollo de la medicina de precisión (45).

Idealmente, estos estudios deberían combinarse en un metaanálisis, pero debido a problemas con la forma en que los resultados se informan con frecuencia, dicho metaanálisis es prácticamente imposible.

La validación adicional de la idoneidad de estos biomarcadores derivados de imágenes debe estar respaldada por pruebas biológicas y estadísticas apropiadas antes de que se investigue su asociación con el resultado del paciente en estudios prospectivos.

El desafío es identificar qué parte de la correlación proviene de la realidad biológica, de las limitaciones de imagen y/o de la definición matemática y algorítmica de las métricas de heterogeneidad. Otro desafío es identificar el nivel de correlación por encima del cual una característica de AT debe excluirse del análisis multivariante posterior, ya que es poco probable que proporcione información complementaria. Esto es menos trivial de lo que puede parecer, ya que el nivel absoluto de correlación varía según los coeficientes elegidos. Los coeficientes de Pearson pueden subestimar significativamente la correlación entre dos métricas cuando la correlación no es lineal. Los coeficientes de Kendall y Spearman proporcionan evaluaciones de correlación de rango, pero los coeficientes de Kendall suelen ser más pequeños que los coeficientes de Spearman. Estos requerirían, por tanto, diferentes escalas para distinguir correlaciones fuertes, moderadas y débiles. El uso de coeficientes de correlación simples para seleccionar características para combinar en un modelo multiparamétrico puede no ser óptimo por lo que se recomienda utilizar técnicas robustas de aprendizaje automático para lograr un mejor análisis de redundancia y selección/combinación de características.

La mayoría de estos estudios recientes han concluido que el análisis textural o radiómico puede proporcionar métricas cuantitativas útiles con respecto al manejo del paciente (subtipo tumoral, agresividad, pronóstico, predicción de respuesta a la terapia y de metástasis a distancia) en diferentes modelos de cáncer (46). Sin embargo, se ha descrito como una necesidad urgente, la estandarización y recomendaciones/directrices para permitir que la realización de estudios de alta calidad y rigor en el diseño, mismas que recientemente han comenzado a darse a conocer por los campos interesados para el desarrollo, integración, evaluación y validación de las nuevas herramientas de aprendizaje automático e inteligencia artificial en el entorno clínico (47–50).

2. Justificación.

El cáncer de pulmón es una de las principales causas de mortalidad relacionada con el cáncer tanto a nivel global como en México, lo cual pone en perspectiva la magnitud del problema que buscamos abordar con nuestro estudio. Dicha enfermedad no sólo tiene un impacto devastador en la salud y la calidad de vida de los pacientes, sino que también supone una carga económica significativa debido a los costos asociados con el diagnóstico, tratamiento y cuidados paliativos(51). La detección temprana y precisa del cáncer de pulmón es de suma importancia para mejorar el pronóstico de los pacientes, y el desarrollo de estrategias de tratamiento personalizadas puede aumentar las posibilidades de éxito terapéutico. Por tanto, nuestro estudio, centrado en la imagen funcional por PET/CT y la extracción de características radiómicas en CPCNP, tiene una trascendencia crucial, ya que puede proporcionar una herramienta diagnóstica adicional de fácil obtención y sin gastos adicionales, que podría ayudar a optimizar la toma de decisiones clínicas, mejorar las capacidades diagnósticas, y potencialmente también las de evaluación de la respuesta a los tratamientos. Además, nuestro estudio incluirá a la población mexicana, un grupo que no ha sido ampliamente estudiado en este contexto, lo cual añade una relevancia adicional a nuestra investigación.

En cáncer de pulmón la radiómica ha sido estudiada en el diagnóstico, evaluación del tratamiento y pronóstico, con múltiples publicaciones al respecto utilizando métodos de imagen morfológicos. Sin embargo, el uso de la imagen funcional por PET/CT y la extracción de características radiómicas en CPCNP, ha demostrado beneficios adicionales, por ejemplo, en la predicción de metástasis a distancia (52); Asimismo, Rios et al. desarrollaron modelos radiómicos para distinguir mutaciones somáticas en un estudio de 763 pacientes con adenocarcinoma de pulmón (53) y existen numerosos estudios que valoran la relación con el estatus mutacional EGFR, su respuesta al tratamiento o la predicción de supervivencia libre de progresión después de la terapia y el pronóstico con resultados prometedores (39–43,54). Sin embargo, el análisis radiómico es un campo aún en exploración, que requiere mayor número de estudios para validar y ampliar los hallazgos obtenidos, máxime en una población como la nuestra que, además difiere sustancialmente en relación con el resto de Latinoamérica y el mundo (55), por lo que un estudio como el propuesto aportaría información que coadyuve a el personal médico a entender mejor la biología subyacente del cáncer, la heterogeneidad tumoral, optimizar la toma de muestras por biopsia para su estudio e influir en la elección dirigida y personalizada de la terapia, así como proponer una herramienta diagnóstica adicional de fácil obtención, que no requiere gastos adicionales y puede mejorar las capacidades diagnósticas y potencialmente también las de evaluación de la respuesta a los tratamientos, tomando en cuenta que la optimización en el abordaje diagnóstico deriva en un tratamiento más oportuno y personalizado que coadyuve a mejorar el pronóstico de estos pacientes. Asimismo, debemos exaltar la muestra poblacional que pretendemos integrar, en la población mexicana que no ha sido estudiada, la cumplimentación con los estándares actuales IBSI, así como el empleo de un modelo estadístico de selección de variables basado en machine learning (regresión logística penalizada) que ayudaría a validar los hallazgos recientes en otras poblaciones y robustecería el empleo de métodos avanzados en el análisis y comprensión de la imagen diagnóstica.

3. Planteamiento del Problema.

El cáncer de pulmón, principal causa de mortalidad por cáncer en ambos sexos, plantea un reto significativo en su manejo debido a su heterogeneidad biológica. Aproximadamente el 80% de estos cánceres son carcinomas de células no pequeñas (CPCNP), cuyo comportamiento y respuesta terapéutica están fuertemente influenciados por mutaciones somáticas, especialmente en los genes EGFR y KRAS. La presencia de mutaciones en EGFR, mutuamente excluyente con KRAS, suele asociarse con una mejor respuesta a terapias dirigidas. Sin embargo, la identificación del estatus mutacional mediante métodos convencionales es costosa, limitada en disponibilidad, invasiva y susceptible a errores de muestreo, lo que puede retrasar, limitar o incluso errar la implementación de las terapias más adecuadas.

Los tumores con mutación EGFR suelen tener un comportamiento biológicamente más indolente y presentar niveles más bajos de metabolismo de la glucosa que los tumores de tipo salvaje, un fenómeno detectable a través de la PET/CT con 18F-FDG. Esta evidencia sugiere la posibilidad de distinguir, mediante análisis fenotípicos, entre tumores mutados y no mutados.

La irrupción de las terapias dirigidas ha supuesto un cambio paradigmático en el manejo del CPCNP, pero aún solo una fracción de los pacientes en todo el mundo tiene acceso a las pruebas de biomarcadores requeridas. Es esencial optimizar tanto la recolección de tejido tumoral para el diagnóstico y las pruebas de biomarcadores como la implementación de métodos emergentes, no invasivos y específicos como el análisis radiómico, que podría facilitar la identificación del estatus mutacional y apoyar en la decisión terapéutica.

En este escenario, se hace patente la necesidad de una estrategia multidisciplinaria para mejorar el acceso y precisión de la oncología personalizada, donde nuestro estudio busca aportar al ofrecer un análisis detallado del potencial del análisis radiómico en la identificación del estatus mutacional EGFR en CPCNP.

4. Pregunta de Investigación.

¿Existe una correlación entre las características radiómicas obtenidas del análisis de imágenes PET/CT con FDG y el estatus mutacional EGFR en pacientes con carcinoma de células no pequeñas (CPCNP)?

5. Objetivos.

5.1 Objetivo General

- Determinar la asociación entre las características radiómicas del PET/CT con FDG en CPCNP y el estado mutacional de EGFR.

5.2 Objetivos Específicos

- Determinar y seleccionar las variables con mejor asociación con la presencia o ausencia de la mutación EGFR.
- Designar los modelos (firmas radiómicas) que mejor se ajusten para la predicción de estatus de la mutación EGFR en CPCNP.
- Contrastar la asociación de los modelos radiómicos con la mutación EGFR según el subtipo mutacional.
- Realizar análisis estratificado para sexo, edad, estadio clínico y tabaquismo.

6. Hipótesis de trabajo.

Las características radiómicas del CPCNP obtenidas de la tomografía por emisión de positrones con fluorodesoxiglucosa (PET/CT – 18F-FDG) están asociadas a la presencia o ausencia de la mutación EGFR.

7. Material y Métodos:

7.1 Diseño del estudio.

7.1.1 Tipo de estudio.

- Por el control de la maniobra experimental por el investigador: Observacional.
- Por la captación de la información: Retrolectivo.
- Por la medición del fenómeno en el tiempo: Transversal.
- Por la direccionalidad causa-efecto: Correlacional.
- Por la búsqueda de inferencia causal: Analítico.
- Por el tipo de muestreo: No probabilístico de casos consecutivos.

7.1.2 Diseño de investigación:

- Estudio transversal analítico.

7.2 Lugares en donde se realizará el protocolo

Pacientes adultos que han acudido al Instituto Nacional de Cancerología (INCan) por ser población no derechohabiente o para realización de estudio diagnóstico PET/CT con FDG como parte del convenio entre IMSS - INCan, los cuales son abordados y tratados en cualquiera de las instituciones involucradas (INCan y/o en Hospital de Oncología de CMN Siglo XXI).

7.3 Duración del estudio.

Se tomarán en cuenta pacientes que cumplan con los criterios de inclusión en el periodo comprendido entre el 01 de enero de 2015 al 31 de diciembre de 2021

7.4 Fuente de información.

Registro en expediente electrónico del Hospital de Oncología CMN Siglo XXI y expediente electrónico del INCan.

7.5 Población de estudio.

Pacientes con diagnóstico histopatológico de cáncer de pulmón de células no pequeñas (CPCNP), así como determinación del estado mutacional de EGFR y estudio PET/CT con 18F-FDG, realizados previo a cualquier tipo de intervención terapéutica.

7.6 Cálculo de tamaño de muestra.

En nuestro estudio, ya conocemos a priori el estado de la mutación de los pacientes y deseamos investigar qué variables independientes pueden influir en la presencia o ausencia de la mutación EGFR en el CPCNP, por tanto, el análisis a realizar por defecto es una regresión logística.

El enfoque más común para determinar el tamaño de la muestra en un análisis de regresión logística es utilizar la regla de eventos por variable (EPV). Según esta regla, se recomienda tener al menos 10 eventos (en este caso, pacientes con la mutación) por cada variable independiente que se planea incluir en el modelo. Dado que el estudio está basado en análisis radiómico y este incluye un aproximado de 143 variables independientes, resulta inviable utilizar inicialmente este enfoque puesto que además existen diversos supuestos que no se cumplen, para resolver este problema, se empleara inicialmente un método avanzado de aprendizaje automático supervisado para selección de variables, conocido como regresión logística penalizada que incluye; Ridge, LASSO (Least Absolute Shrinkage and Selection Operator) y Elastic Nets.

La regresión logística penalizada es un método de regresión que realiza tanto la selección de variables como la regularización para mejorar la precisión y la interpretabilidad del modelo de regresión estadística.

En términos de cálculo del tamaño de la muestra, no hay una regla fija para la regresión penalizada como la regla de eventos por variable (EPV) en la regresión logística estándar. Sin embargo, hay algunas consideraciones generales a tener en cuenta:

- Número de eventos: Al igual que con la regresión logística estándar, se necesitará un número suficiente de eventos (en este caso, pacientes con la mutación) para tener suficiente potencia para detectar las relaciones que se están buscando.
- Número de predictores: Aunque la RLP puede manejar un gran número de predictores, incluso con un tamaño de muestra relativamente pequeño, todavía se necesitará un número suficiente de observaciones para cada predictor que se esté considerando. Si se tiene un número muy grande de predictores en relación con el tamaño de muestra, se puede correr el riesgo de sobreajuste.
- Complejidad del modelo: La RLP tiende a seleccionar modelos más simples, por lo que, si se espera que el modelo final incluya muchas variables, es posible que se necesite un tamaño de muestra más grande.
- Efecto esperado de los predictores: Si se espera que los efectos de los predictores sean pequeños, se necesitará un tamaño de muestra más grande para detectar estos efectos.

Elastic Net, esta técnica combina las propiedades de Ridge y Lasso, ofreciendo una solución más flexible y robusta. Elastic Net utiliza un parámetro de mezcla que permite controlar el grado de regularización y, por lo tanto, puede adaptarse mejor a diferentes

situaciones. La principal ventaja de Elastic Net radica en su capacidad para manejar conjuntos de datos con alta dimensionalidad y correlaciones entre predictores, a la vez que supera algunas limitaciones de Ridge y Lasso.

El cálculo de la muestra que estamos utilizando se basa en un método de validación cruzada. La validación cruzada es un método para evaluar el rendimiento de un modelo estadístico utilizando una muestra limitada de datos. Específicamente, estamos utilizando el método de validación cruzada de K iteraciones (a menudo llamado K-fold), que divide el conjunto de datos en K subconjuntos de aproximadamente el mismo tamaño. El modelo se entrena en K-1 de estos subconjuntos y luego se evalúa en el subconjunto restante. Esto se repite K veces, utilizando cada subconjunto exactamente una vez como conjunto de prueba. El rendimiento del modelo se promedia entonces sobre las K iteraciones para obtener una estimación más robusta del rendimiento del modelo.

Además, estamos variando el tamaño de la muestra para ver cómo afecta al rendimiento del modelo. Para cada tamaño de muestra, tomamos una muestra aleatoria de los datos, y luego realizamos la validación cruzada en esta muestra. Repetimos este proceso para una serie de diferentes tamaños de muestra para ver cómo varía el rendimiento del modelo en función del tamaño de la muestra.

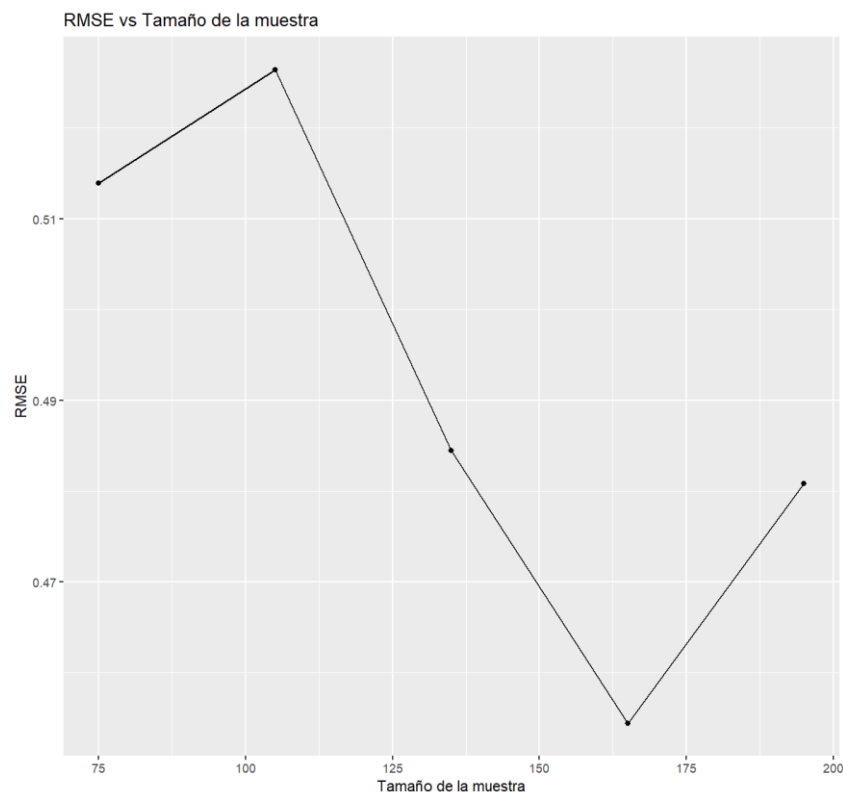


Gráfico 1

La gráfica que obtenemos muestra el error cuadrático medio de la raíz (RMSE) en función del tamaño de la muestra. El RMSE es una medida común del error de un modelo: un RMSE más pequeño indica un modelo más preciso. En la gráfica, cada punto representa un tamaño de muestra distinto, y la línea traza el RMSE para cada tamaño de muestra.

7.7 Criterios de inclusión, exclusión, eliminación.

7.7.1 Criterios de inclusión

- Diagnóstico histopatológico de cáncer de pulmón de células no pequeñas, obtenido por biopsia de tejido o líquida utilizando técnicas estándar de la institución, realizado en un periodo no mayor a 1 mes antes o después de la realización del estudio PET/CT.
- Diagnóstico histopatológico de cáncer de pulmón de células no pequeñas, obtenido por resección quirúrgica del tumor primario mediante técnicas estándar de la institución, posterior a la realización del estudio PET/CT.
- Análisis del perfil mutacional del EGFR llevado a cabo mediante métodos estándar (ej. secuenciación de nueva generación, PCR en tiempo real, etc.), y reportado en el expediente electrónico.
- PET/CT realizado con 18F-FDG antes de haber sido sometido a tratamiento (Farmacológico, Radioterapéutico o Quirúrgico).
- PET/CT realizado en un periodo comprendido entre el 01 de enero de 2015 al 31 de diciembre de 2021.
- Los pacientes se incluirán de forma indistinta en cuanto a su sexo, edad, raza, lugar de origen, etapa clínica y/o factores de riesgo conocidos, sin embargo, estos datos serán recolectados y se incluirán las medidas pertinentes para la estratificación de factores de confusión.

7.7.2 Criterios de exclusión global:

- Pacientes que hayan sido sometidos a resección quirúrgica del tumor primario previo a la realización del estudio PET/CT y aquellos en los que la lesión primaria no sea valorable por dicho método.

7.7.3 Criterios de eliminación:

- Información del paciente o del estudio PET/CT con 18F-FDG incompleta o que no puedan ser recuperados en las bases de datos (PACS) por ningún medio disponible en el Instituto (grabaciones, libretas del área de radiodiagnóstico, informes específicos etc.), en caso de que la recuperación sea parcial se incorporará a la base de datos y se especificará en el análisis.

7.8 Variables incluidas:

Tabla 1

VARIABLES DEMOGRAFICAS				
Definición	Descripción Conceptual	Descripción operacional	Tipo	Escala de medición y codificación
SEXO	Conjunto de características genéticas, biológicas y sociales que determinan si se trata de una mujer u hombre	Se registrará como 1 (Femenino) y 2 (Masculino).	Cualitativa nominal dicotómica	Femenino= 1 Masculino= 2
EDAD	Edad del paciente al momento del abordaje diagnóstico	Se registrará la edad al momento de la realización del estudio PET/CT inicial.	Cuantitativa ordinal	Valor numérico de años cumplidos

Tabla 2

VARIABLE DEPENDIENTE				
Definición	Descripción Conceptual	Descripción operacional	Tipo	Escala de medición y codificación

Estado mutacional de EGFR	Mutación en el gen EGFR	Se recabará resultado del expediente clínico electrónico, independientemente del tipo de muestra y los métodos utilizados (Sanger, PCR, SNG etc.).	Cualitativa nominal dicotómica.	No mutado= 0 Mutado= 1
---------------------------	-------------------------	--	---------------------------------	---------------------------

Tabla 3

VARIABLES INDEPENDIENTES				
Definición	Descripción Conceptual	Descripción Operacional	Tipo	Escala de medición y codificación
Patrón histológico de CPCNP	Descripción de un tumor según cuán anormales se ven las células y los tejidos al microscopio y cuán rápido se podrían multiplicar y diseminar las células cancerosas.	Se registrarán de acuerdo a la clasificación de la OMS, con base en lo enunciado en el reporte histopatológico en el expediente electrónico.	Cualitativa nominal	1= Adenocarcinoma 2=Carcinoma de células escamosas 3=Carcinoma de células grandes 4=Carcinoma sarcomatoide 5=Carcinoma adenoescamoso
SUV max-bw	Valor Máximo De Captación Estándar Por Peso Corporal (Maximum Standardized Uptake Value Body Weight)	Valor cuantitativo calculado mediante software LIFEx 7.3.	Cuantitativa continua	Valor numérico calculado
SUV mean-bw	Valor Promedio De Captación Estándar (Standardized Uptake Value Mean - Body Weight)	Valor cuantitativo calculado mediante software LIFEx 7.3.	Cuantitativa continua	Valor numérico calculado
SUV peak-bw	Valor Pico De Captación Estándar (Standardized Uptake Value Peak- Body Weight)	Valor cuantitativo calculado mediante software LIFEx 7.3.	Cuantitativa continua	Valor numérico calculado
TLG (SUV-bw x cm3) Umbral (H o 40%)	Glicolisis total de la lesion (Total Lesion Glycolysis (Maximum Standardized Uptake Value - Body Weight X Cubic Centimeter))	Valor cuantitativo calculado mediante software LIFEx 7.3.	Cuantitativa continua	Valor numérico calculado
MTV (cm3) Umbral %	Volumen Metabólico Tumoral (Metabolic Tumor Volume)	Valor cuantitativo calculado mediante software LIFEx 7.3.	Cuantitativa continua	Valor numérico calculado
Indices de Histograma	Características de normalidad	Valor cuantitativo calculado mediante software LIFEx 7.3.	Cuantitativa continua	Valor numérico calculado
Indice de forma	Características morfológicas	Valor cuantitativo calculado mediante software LIFEx 7.3.	Cuantitativa continua	Valor numérico calculado
Grey Level Co-occurrence Matrix (GLCM)	Características de textura – Matriz de co-ocurrencia de nivel de gris	Valor cuantitativo calculado mediante software LIFEx 7.3.	Cuantitativa continua	Valor numérico calculado
Grey-Level Run Length Matrix (GLRLM)	Características de intensidad – Matriz de longitud de ejecución de nivel de gris	Valor cuantitativo calculado mediante software LIFEx 7.3.	Cuantitativa continua	Valor numérico calculado

Neighborhood Grey-Level Different Matrix (NGLDM)	Características de intensidad – Matriz de diferencia de nivel de gris vecino.	Valor cuantitativo calculado mediante software LIFE x 7.3.	Cuantitativa continua	Valor numérico calculado
Grey-Level Zone Length Matrix (GLZLM)	Características de definición – Matriz de longitud de zona de nivel de gris.	Valor cuantitativo calculado mediante software LIFE x 7.3.	Cuantitativa continua	Valor numérico calculado

Tabla 4

VARIABLES CONFUSORAS				
Definición	Descripción Conceptual	Descripción operacional	Tipo	Escala de medición y codificación
Tabaquismo	Adicción al consumo del tabaco.	Se recabará del interrogatorio en la historia clínica en expediente electrónico.	Cualitativa nominal dicotómica	Negado= 0 Positivo= 1
Índice tabáquico.	Número de referencia que refleja el promedio de consumo de tabaco en un individuo.	número de cigarrillos consumidos al día por los años durante los cuales se ha fumado y a este resultado se lo divide por 20. (Número de cigarrillos al día) X (años por los que se fumó) dividido entre 20. Se recabará del expediente clínico electrónico.	Cuantitativa Ordinal	Valor numérico calculado
SEXO	Conjunto de características genéticas, biológicas y sociales que determinan si se trata de una mujer u hombre	Se recabará del expediente clínico electrónico.	Cuantitativa ordinal.	Femenino= 1 Masculino= 2
EDAD	Valor numérico de años cumplidos al momento de la realización del estudio PET/CT	Se recabará del expediente clínico electrónico	Cualitativa nominal	Valor numérico de años cumplidos
Subtipo mutacional EGFR	Mutaciones puntuales, sin sentido, inserciones o deleciones presentes en los dominios de los exones del gen que codifica para EGFR.	Se recabará resultado del expediente clínico electrónico, independientemente del tipo de muestra y los métodos utilizados (Sanger, PCR, SNG etc.).	Cualitativa nominal	Delección en exón 19= 1 Sustitución L858R en exón 21= 2 Sustitución L861Q en exón 21= 3 G719X en exón 18= 4 E709X en exón 18=5 Inserción en exón 20= 6 Sustitución S768I en exón 20= 7 T790M en el exón 20= 8
Estadio Clínico	Agrupación de categorías que refleja un pronóstico con base en características clínicas	Se recabará del expediente clínico electrónico.	Cualitativa nominal	IA1 IA2 IB IIA

	derivadas de la extensión del enfermedad.			IIB IIIA IIIB IIIC IVA IVB
Otras variantes patogénicas	Cambio en un gen que impide su funcionamiento correcto (mutación).	Se recabará resultado del expediente clínico electrónico, independientemente del tipo de muestra y los métodos utilizados (Sanger, PCR, SNG etc.).	Cualitativa nominal	Ninguna= 0 ALK= 1 ROS1= 2 ERBB2= 3 RET= 4 BRAF V600E= 5 KRAS= 6 NRAS= 7 TP53= 8

7.9 Procedimientos.

7.9.1 Descripción del estudio:

La adquisición de imágenes siguió las pautas de informes de la Iniciativa de estandarización de biomarcadores de imágenes (IBSI) (48). Todos los pacientes ayunaron durante 6 h antes de la exploración PET/CT. Controlamos los niveles de azúcar en sangre del paciente por debajo de 150 mg/dl y se inyectó por vía intravenosa una dosis de ≈ 5 MBq/kg de ^{18}F -FDG. Todos los pacientes inscritos recibieron exploraciones PET/CT con ^{18}F -FDG en el mismo escáner (Siemens mCT). Los parámetros de la tomografía computarizada fueron 120 kV, 40 mAs, espesor de corte de 5,0 mm y matrices de 512×512 . Los parámetros de exploración PET fueron 60 ± 5 min después de la administración del marcador y 2 min por posición en la cama. Utilizamos los datos de TC adquiridos para realizar la corrección de atenuación en todas las imágenes de PET. Se empleará el software LIFEx versión 7.3 (<http://www.lifexsoft.org>) (56) para analizar las características de textura de las lesiones pulmonares en imágenes PET y CT simultáneamente dentro del mismo volumen de interés (VOI). Se aplicó un umbral de valor máximo de captación estandarizado (SUV_{max}) del 40% para definir (semi)automáticamente el VOI de la lesión objetivo en las imágenes PET. El software discretizará automáticamente la intensidad de las imágenes PET y CT. Para las imágenes de PET, el software reducirá la escala continua de los datos de PET a 64 bits, y los límites de cambio de escala de intensidad variaron de 0 a 20. Para las imágenes de CT, el software redujo la escala continua de los datos de CT a 400 bits, y los límites de cambio de escala de intensidad estuvieron entre - 1000 y 3000 UH. Se calcularán características radiómicas de primer orden, segundo orden y orden superior, proporcionando 48 características para imágenes PET y 41 características para imágenes CT.

Las variables independientes serán determinadas por el investigador mediante el software de acceso libre LIFEx en su versión 7.3, mismo que se considera una herramienta ampliamente validada para análisis radiómico y dicho análisis se realizara sobre el estudio PET/CT con ^{18}F -FDG para la determinación de características radiómicas de primer orden, así como la relación de los parámetros metabólicos con su entorno (características radiómicas de segundo orden). El análisis se realizará de manera cegada sin conocer el estado mutacional de los individuos analizados y se realiza de manera semiautomatizada lo cual reduce considerablemente la posibilidad de sesgos de medición.

7.9.2 Propuesta de análisis estadístico:

1. Preprocesamiento de datos: Iniciaré el análisis con el preprocesamiento de los datos. Esto incluirá la limpieza de los datos, la normalización de las imágenes, la segmentación de las regiones de interés y la extracción de características radiómicas.
2. Análisis descriptivo: Luego, realizaré un análisis descriptivo de los datos. Esto implicará calcular medidas de tendencia central y dispersión para las variables continuas, y frecuencias para las variables categóricas. También generaré gráficos para visualizar los datos.
3. Selección de características con regresión logística penalizada (RLP): A continuación, aplicaré la regresión logística penalizada a los datos para seleccionar un subconjunto de características radiómicas que sean relevantes para predecir la variable de resultado. Para evitar el sobreajuste, dividiré los datos en un conjunto de entrenamiento y un conjunto de validación. También implementaré la validación cruzada durante esta etapa para obtener una evaluación más robusta del rendimiento del modelo.
4. Modelado con regresión logística convencional: Tomaré las características que fueron seleccionadas por RLP y construiré un modelo de regresión logística convencional con ellas en el conjunto de entrenamiento. Esto me permitirá obtener estimaciones de los coeficientes para estas características sin la penalización de la regresión logística.
5. Evaluación del modelo: Evaluaré el rendimiento del modelo en el conjunto de validación utilizando medidas como la exactitud, la sensibilidad, la especificidad, el área bajo la curva ROC, etc.
6. Creación de la firma radiómica: Las características seleccionadas y sus coeficientes en el modelo de regresión logística convencional se utilizarán para crear una "firma radiómica". Esta firma será una ecuación o algoritmo que combina las características de una manera específica para predecir el resultado de interés.
7. Interpretación y comunicación de los resultados: Finalmente, interpretaré los resultados del análisis en el contexto de la literatura existente y las preguntas de investigación de mi estudio. Comunicaré mis hallazgos de manera clara y concisa, asegurándome de discutir cualquier limitación de mi estudio y sugerencias para futuras investigaciones.
8. Para el análisis estadístico se usará el software RStudio y SPSS V 27.0.

8. Aspectos Éticos.

- Riesgos del estudio: Se considero que corresponde con una investigación sin riesgo donde se emplean técnicas y métodos de investigación documental retrospectivos y no se realiza ninguna intervención o modificación intencionada en las variables fisiológicas, psicológicas y sociales de los individuos, de acuerdo con el artículo 17, título segundo, capítulo I, fracción I del reglamento de la ley general de salud en materia de investigación para la salud. No implica el involucro de población vulnerable.
- Apego a normas éticas: Este estudio se realiza de acuerdo con los lineamientos éticos de la 64° asamblea general de la declaración de Helsinki de la asociación médica mundial y a lo establecido en el capítulo 102 del reglamento de la ley general de salud en materia de investigación para la salud.

- Este protocolo de investigación aportara a la sociedad y/o confirmara nuevos conocimientos emergentes en relación con el uso del análisis radiómico y su potencial papel en la predicción del estado mutacional de diversas variantes patogénicas que tienen implicación directa en el pronóstico y el tratamiento empleado. Acorde con lo anterior y dado el diseño retrospectivo de la investigación, los beneficios son muchos y los riesgos son inexistentes.
- Para garantizar la confidencialidad de la información, se asignará un número de folio para la identificación de los sujetos sin recolección o utilización de datos privados asimismo los datos DICOM serán anonimizados y foliados para evitar el riesgo de comprometer los datos.
- La selección de los participantes será mediante muestreo no probabilístico de casos consecutivos que acuden a solicitar atención y son sometidos al abordaje convencional de métodos diagnósticos, de manera que todos los pacientes tienen la misma probabilidad de ser incluidos en el protocolo de investigación.
- Los beneficios identificados en este protocolo serán otorgados a la población general susceptible de beneficiarse de estos métodos de forma prospectiva, incrementando el acervo científico y asentando las bases científicas que soporten algoritmos de decisión médica especializada, que mejoren el pronóstico de los pacientes con un abordaje expedito y altamente específico, maximizando el beneficio de los métodos de abordaje, sin incrementar los costos para el paciente o el sistema de salud.

9. Aspectos de Bioseguridad.

- Los estudios PET/CT fueron realizados en el servicio de Medicina Nuclear del Instituto Nacional de Cancerología en el periodo comprendido entre el 01 de enero de 2015 al 31 de diciembre de 2021.
- El tipo de material radiactivo que fue utilizado para la realización del PET/CT, fue:
 - o 18F-FDG (fluorodesoxiglucosa), el cual es un análogo estructural de la glucosa, radiomarcado con Flúor 18, el cual es un isotopo radiactivo del Flúor, con tipo de decaimiento por emisión de positrones con vida media física de 110 minutos.
- El personal técnico del instituto cuenta con capacitación y reconocimiento como personal ocupacionalmente expuesto (POE), por parte de la Comisión Nacional de Seguridad Nuclear y Salvaguardas (CNSNS), los cuales cuentan con registro dosimétrico de exposición mensual y acumulado anual, según consta en los registros de la CNSNS.
- Número de la licencia expedida por la Comisión Nacional de Seguridad Nuclear y Salvaguardas: 2014-2016 (A00.200/0762/2014), 2016-2018 (A00.200/0864/2016) y 2018-actual (A00.200/0838/2018)
- Numero de licencia COFEPRIS para el uso de rayos X (TC): COFEPRIS-09-004 (Numero: 173300904X0005).

La normatividad vigente aplicable para la adquisición de estudios PET/CT y el uso de fuentes radiactivas abiertas, se enuncia a continuación:

- Reglamento General de Seguridad Radiológica (1988)
- Norma Oficial Mexicana NOM-028-NUCL-2009, Manejo de desechos radiactivos en instalaciones radiactivas que utilizan fuentes abiertas
- Norma Oficial Mexicana NOM-004-NUCL-1994, Clasificación de los desechos radiactivos
- Norma Oficial Mexicana NOM-012-STPS-2012, Condiciones de seguridad y salud en los centros de trabajo donde se manejen fuentes de radiación ionizante
- Lineamientos generales para la utilización de "Sustancias Químicas Peligrosas y Tóxicas e Isótopos Radiactivos"

10. Recursos, financiamiento y factibilidad.

El investigador responsable y asociados cuentan con capacitación altamente especializada y participarán directamente en la recolección de los datos, el análisis radiómico y estadístico, los

cuales se obtendrán de estudios diagnósticos que ya han sido obtenidos y/o realizados en los sujetos participantes.

Este protocolo de investigación tiene un diseño observacional, retrospectivo que no requiere de apoyo financiero y no implica uso de recursos adicionales, además de estar apegado a las normas vigentes en el territorio nacional, depende de la realización de pruebas diagnósticas realizadas al abordaje de pacientes con CPCNP, mismas que forman parte del protocolo convencional en los institutos involucrados, la afluencia de pacientes a estos hospitales permite contar con una muestra suficiente, los criterios de inclusión permiten contemplar pacientes de ambos institutos, sin embargo, se tiene presente que la mayoría de la población participante proviene del INCan, debido a que en este instituto se cuenta con los equipo PET/CT y se realizan pruebas mutacionales de forma ordinaria, sin embargo, existe convenio vigente entre ambas instituciones para la colaboración con estudios PET/CT, de acuerdo con la evaluación por el equipo de investigación es factible su realización y culminación, dado su carácter retrospectivo y la disponibilidad inmediata de la información necesaria.

11. Análisis e Interpretación de los Resultados.

Análisis estadístico

Se trata de un diseño de investigación transversal analítico retrolectivo en el que se incluyeron inicialmente 577 pacientes con diagnóstico de CPCNP, de los cuales se pudieron recuperar los datos e información completa de los estudios PET/CT con FDG, estudio mutacional y se ejecutó con éxito el análisis radiómico de 196 pacientes, obteniendo la información de 138 variables independientes de las cuales de las cuales 4 variables fueron clínicas (Sexo, Edad, Etapa Clínica y Tabaquismo) y 134 variables radiómicas (numéricas), así como 1 variable dependiente binomial (EGFR).

Análisis descriptivo:

Para el análisis descriptivo, se calculó la frecuencia y el porcentaje para las variables categóricas, y la media, mediana, desviación estándar, y rango para las variables continuas.

Sexo: El 41.8% de los participantes fueron hombres y el 58.2% fueron mujeres.

Edad: La edad media de los participantes fue de 63 años con una desviación estándar de 14.51. La edad mínima fue de 27 años y la máxima de 94 años.

Tabla 5

Características	EGFR		Total (%)
	Mutado	No mutado	
Edad	88	108	196
Sexo			
- Femenino	57	57	114 (58.2)
- Masculino	31	51	82 (41.8)
Tabaquismo			
- Fumadores	32	58	
- No fumadores	56	50	
Histología			
- Adenocarcinoma	86	101	
- Carcinoma de células escamosas	2	7	
Variante patogénica			
- Deleción del exón 19	61		

- Mutación puntual exón 21 L858R	25		
- Mutación puntual exón 18, G719X	1		
- Mutación puntual exón 21, L861Q	1		
- ALK		5	
<i>Etapa clínica</i>			
I	2	5	7
II	5		5
III	3	6	9
IV	78	97	175

Análisis univariado

En este análisis, buscamos entender la asociación entre diferentes características de pacientes con la presencia de la mutación EGFR. Las variables consideradas fueron el sexo, la etapa clínica y la edad.

Nuestro conjunto de datos consta de pacientes con información sobre su sexo, etapa clínica (EC), edad y presencia de la mutación EGFR. Las variables categóricas fueron recodificadas para facilitar su interpretación.

Las pruebas estadísticas empleadas incluyen:

- La prueba de chi-cuadrado y la prueba de Fisher para evaluar la asociación entre variables categóricas.
- La prueba de Lilliefors para verificar la normalidad de la variable de edad.
- La prueba de Levene para comprobar la homogeneidad de las varianzas.
- La prueba de Mann-Whitney para comparar medianas, utilizada debido a la no normalidad de los datos de edad.

- Prueba de chi cuadrado para Sexo y EGFR

Tabla 6

	No mutado	Mutado
Femenino	57	57
Masculino	51	31

La tabla de contingencia muestra que, de las 114 mujeres, 57 tienen la mutación EGFR, y de los 82 hombres, 31 tienen la mutación.

Pearson's Chi-squared test with Yates' continuity correction

```
data: data$Sexo and data$EGFR
X-squared = 2.3953, df = 1, p-value = 0.1217
```

Los resultados de la prueba de chi-cuadrado indican que no hay una asociación significativa entre el Sexo y la mutación EGFR (X-squared = 2.3953, df = 1, p-value = 0.1217). Por lo tanto, el sexo del paciente no parece influir en la probabilidad de tener la mutación EGFR.

- Prueba de Fisher para Etapa Clínica (EC) y EGFR

Tabla 7

Estadio Clínico	EGFR	
	No mutado	Mutado
I	5	2
II	0	5
III	6	3
IV	97	78

La tabla de contingencia muestra que de las 7 personas en la etapa I, 2 tienen la mutación EGFR, de las 5 en la etapa II, todas tienen la mutación, de las 9 en la etapa III, 3 tienen la mutación y de las 175 en la etapa IV, 78 tienen la mutación.

Fisher's Exact Test for Count Data

```
data: data$EC and data$EGFR
p-value = 0.05557
alternative hypothesis: two.sided
```

La prueba de Fisher indica que no hay una asociación significativa entre la EC y la mutación EGFR (p-value = 0.05557). Aunque el p-valor está cerca del umbral común de 0.05, generalmente aceptado para indicar significancia, aún es superior, por lo que no podemos rechazar la hipótesis nula.

Prueba de Lilliefors para Edad

Lilliefors (Kolmogorov-Smirnov) normality test

```
data: data$Edad
D = 0.094716, p-value = 0.0002012
```

La prueba de Lilliefors (Kolmogorov-Smirnov) sugiere que la variable Edad no sigue una distribución normal (D = 0.094716, p-value = 0.0002012), lo que es importante tener en cuenta si vamos a realizar análisis que asumen la normalidad de los datos.

- Prueba de homogeneidad de las varianzas para Edad y EGFR

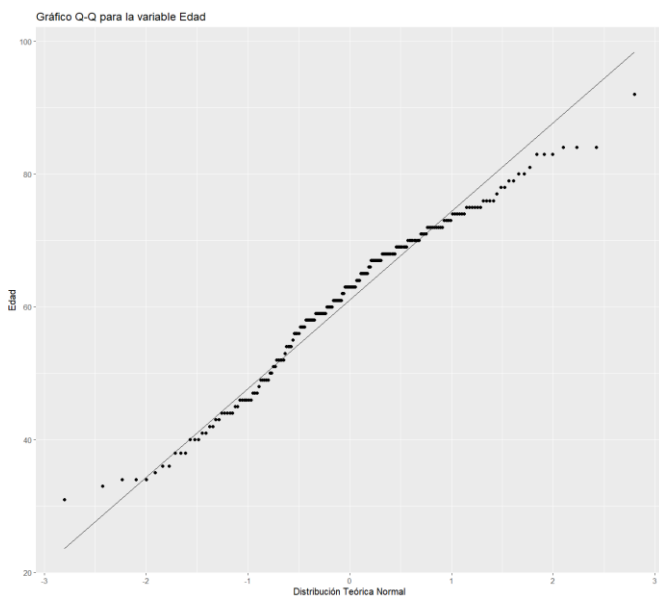


Gráfico 2

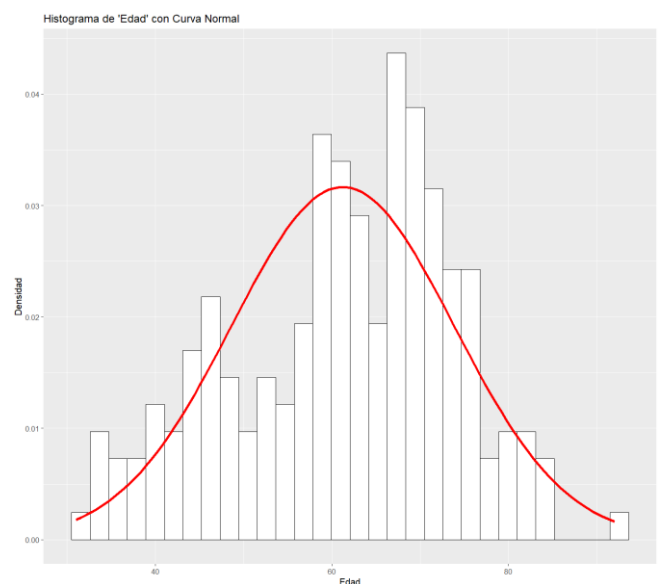


Gráfico 3

Los gráficos Q-Q y de histograma también indican que la Edad no sigue una distribución normal, confirmando los resultados de la prueba de Lilliefors.

```
Levene's Test for Homogeneity of Variance (center = median)
  Df F value Pr(>F)
group 1 0.2153 0.6431
    194
```

La prueba de Levene indica que las varianzas de las edades entre los grupos EGFR son iguales (F value = 0.2153, p-value = 0.6431). Esto significa que podemos proceder con un análisis de varianza (ANOVA) o similar si es necesario, aunque con la precaución de que la distribución de los datos de edad no es normal.

Los resultados de nuestro análisis univariado indican que, en nuestra muestra de datos, no hay una asociación significativa entre la presencia de la mutación EGFR y el sexo, la etapa clínica o la edad del paciente. Es importante destacar que estos resultados son específicos de nuestra muestra de datos y podrían no generalizarse a todas las poblaciones de pacientes. Se requieren estudios más amplios para confirmar estos hallazgos.

Análisis multivariado

Se procedió al análisis del conjunto de datos obtenidos, utilizando modelos de regresión regularizada; Ridge, LASSO y Elastic Net, de las cuales se obtuvo el área bajo la curva y el error mínimo residual estandarizado (RMSE), de los cuales el modelo Elastic Net se ajustó mejor a los datos y tuvo el valor más bajo de RMSE, por lo que se ajustó el modelo, mismo que es una técnica de regresión penalizada que combina la regresión de Ridge y Lasso para optimizar la selección de variables y la regularización. Posterior a la ejecución del modelo de regresión regularizada se seleccionan las variables más importantes para la predicción de la variable dependiente (EGFR) según su coeficiente (21 variables seleccionadas) y se utilizan dichas variables para la creación de un modelo de regresión logística convencional, ajustando el modelo una vez más de acuerdo a las variables estadísticamente significativas (3 variables) y finalmente se realiza la conformación de una firma radiómica y su respectiva evaluación de rendimiento. A continuación, se enuncian los pasos realizados y se reportan los resultados obtenidos:

1. Se cargaron las bibliotecas necesarias para el análisis, incluyendo `glmnet` para el modelo Elastic Net, `caret` para la manipulación y división de los datos, `readxl` para leer datos de Excel y `pROC` para el cálculo de la curva ROC.

2. Los datos se cargaron desde un archivo Excel y se graficó el mapa de correlaciones

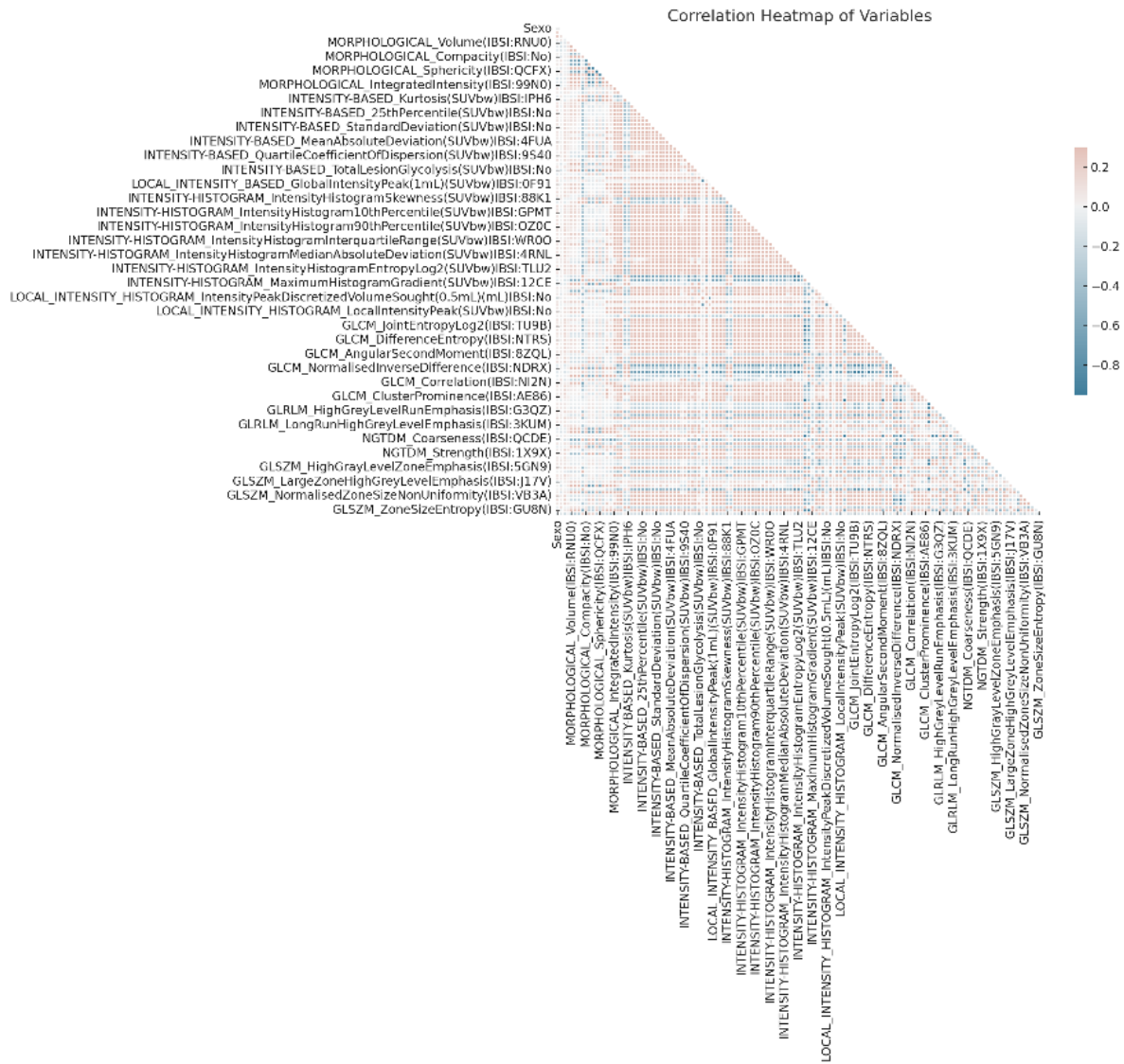


Gráfico 4. Mapa de calor de las correlaciones entre las variables. Las áreas en rojo indican correlaciones positivas fuertes, mientras que las áreas en azul indican correlaciones negativas fuertes. Muchas de las características parecen tener correlaciones bajas a moderadas entre sí. Sin embargo, hay algunas características que tienen correlaciones fuertes. Estas correlaciones fuertes pueden ser un indicio de multicolinealidad, que deberíamos tener en cuenta al ajustar los modelos de regresión y a su vez justifica el uso de la regresión logística regularizada (elastic net), la cual puede trabajar con este tipo de datos, regularizarlos, ajustarlos y penalizarlos para la selección de variables.

3. Se verificó que la variable de respuesta (EGFR) es binaria, con 108 observaciones en la clase 0 (no mutado) y 88 en la clase 1 (mutado).
4. Se estableció una semilla para garantizar la reproducibilidad del análisis (123).
5. Los datos se dividieron en un conjunto de entrenamiento y un conjunto de prueba usando una proporción del 70% para el entrenamiento.
6. Los datos se prepararon y ajustaron para el modelo.
7. Se entrenó el modelo Elastic Net con validación cruzada.

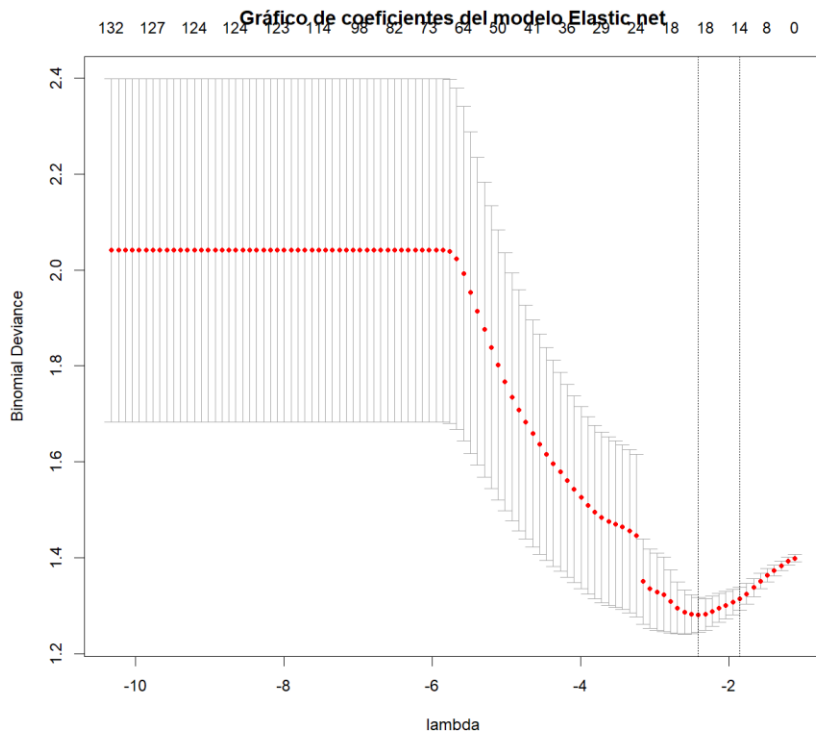


Gráfico 5

8. Se obtuvo el valor óptimo de lambda, que controla el grado de regularización en el modelo Elastic Net.

```
> mejor.lambda
[1] 0.107666
> log(mejor.lambda)
[1] -2.228722
```

9. Se obtuvieron los coeficientes del modelo para este valor óptimo de lambda y se imprimieron.

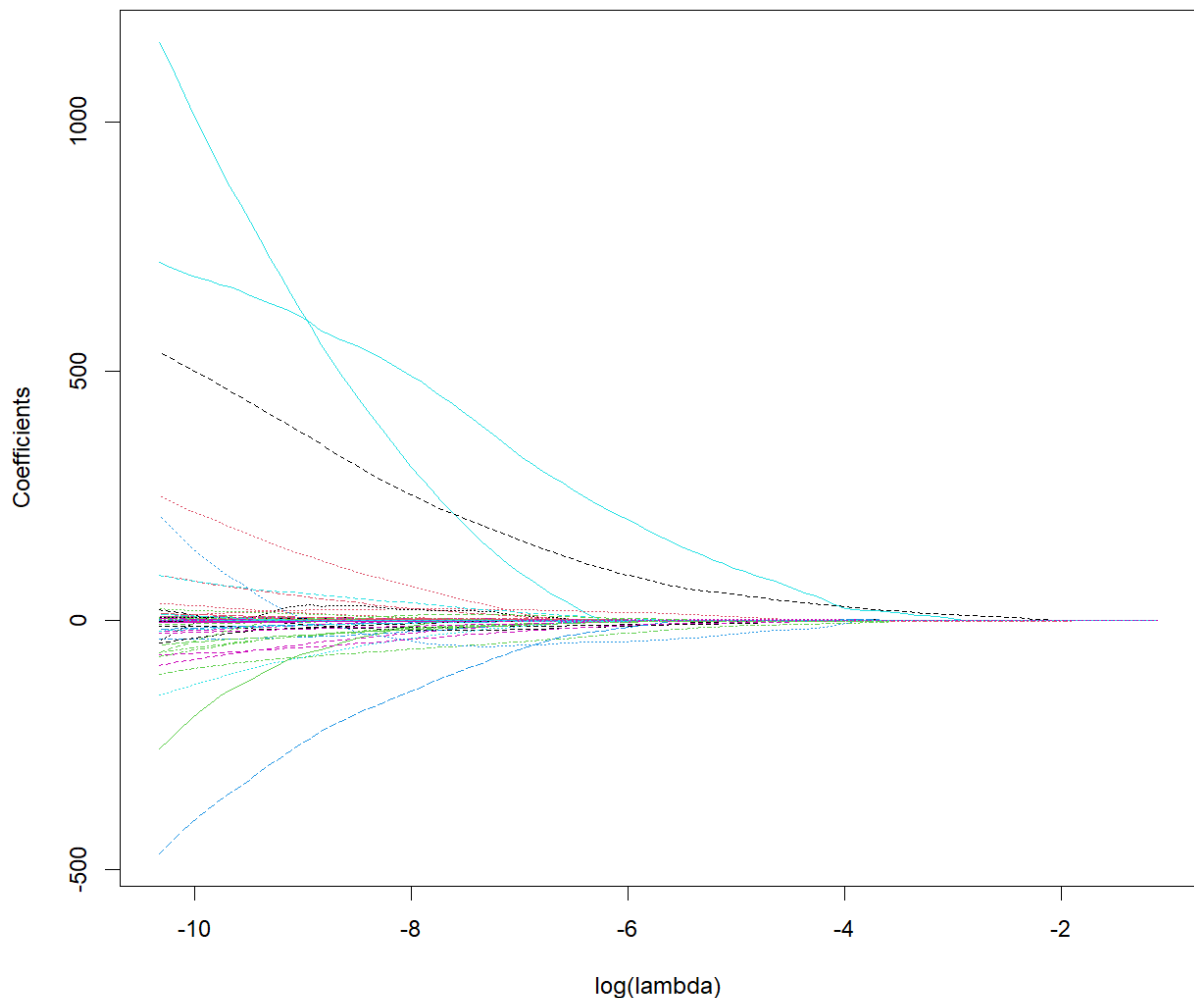


Gráfico 6. Representación visual de la trayectoria de los coeficientes a lo largo de los valores logarítmicos de la penalización lambda en un modelo de red elástica.

En el gráfico, cada línea representa un predictor diferente en tu modelo, y el eje X muestra el logaritmo de lambda. A medida que te mueves hacia la derecha en el eje X (es decir, a medida que lambda aumenta), la penalización por los coeficientes también aumenta y los coeficientes se reducen hacia cero. Esto se conoce como regularización y es una técnica útil para evitar el sobreajuste y para manejar la multicolinealidad en el análisis de regresión.

La regularización es especialmente útil cuando tienes un conjunto de datos de alta dimensión, donde hay más predictores que observaciones, o cuando varios predictores están altamente correlacionados.

En este gráfico, puedes ver cuándo los coeficientes de cada predictor se reducen a cero a medida que aumenta la penalización (moviéndote hacia la derecha en el gráfico). Los predictores cuyos coeficientes se reducen a cero más rápidamente son menos robustos, ya que son más sensibles a la penalización. Por otro lado, los predictores cuyos coeficientes se reducen a cero más lentamente son los que el modelo considera más importantes para predecir la variable de respuesta.

Entonces, la interpretación esencial de este gráfico es que te permite ver cómo cambian los coeficientes del modelo a medida que cambia la fuerza de la regularización, lo cual es un aspecto fundamental de la red elástica y otros modelos de regresión regularizados.

10. Se realizó una validación cruzada para el modelo Elastic Net.

11. Se graficaron los resultados de la validación cruzada y se graficó el mejor valor de lambda.

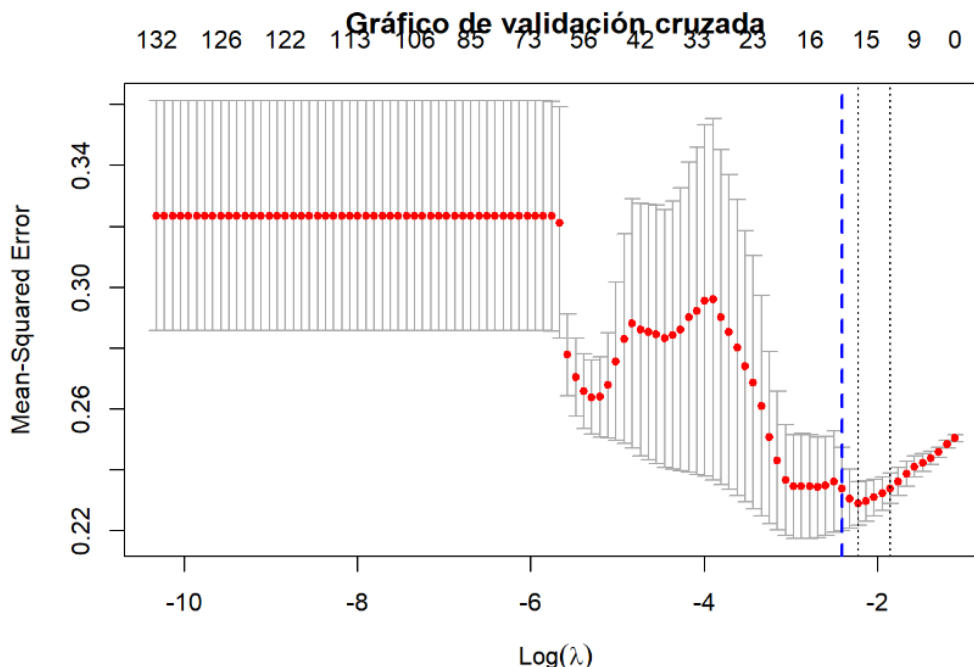


Gráfico 7. El gráfico muestra los resultados de un procedimiento de validación cruzada, que se utiliza para determinar el parámetro de penalización lambda que resulta en el modelo con el mejor rendimiento predictivo, la línea azul intermitente representa el valor del mejor lambda..

En el gráfico, cada punto representa la tasa de error (definida por un criterio específico, como el error cuadrático medio) para un valor específico de lambda, y la línea conecta estos puntos. Usualmente, el eje X representa el logaritmo de lambda, y el eje Y representa el error de validación cruzada.

El valor de lambda que minimiza este error de validación cruzada (es decir, el punto más bajo en el gráfico) es a menudo un buen equilibrio entre sesgo y varianza, y generalmente proporciona un modelo que es ni demasiado complejo (sobreajuste) ni demasiado simple (subajuste).

Se obtiene el "mejor lambda" y luego se calcula el logaritmo de este valor.

En resumen, este gráfico y el valor de "mejor lambda" permiten visualizar y cuantificar cómo varía el rendimiento del modelo a medida que cambia la fuerza de la regularización, y permite elegir un valor que equilibre de manera efectiva el sesgo y la varianza.

12. Se predijeron las probabilidades de la clase positiva usando este mejor valor de lambda y se convirtieron estas probabilidades en predicciones de clase.

13. Se calculó una matriz de confusión para evaluar el desempeño del modelo, con una precisión del 68.84%.

Confusion Matrix and Statistics

Prediction	Reference	
	0	1
0	49	20
1	23	46

Accuracy : 0.6884
 95% CI : (0.6041, 0.7645)
 No Information Rate : 0.5217
 P-Value [Acc > NIR] : 5.106e-05
 Kappa : 0.3768
 McNemar's Test P-Value : 0.7604
 Sensitivity : 0.6970
 Specificity : 0.6806
 Pos Pred Value : 0.6667
 Neg Pred Value : 0.7101
 Prevalence : 0.4783
 Detection Rate : 0.3333
 Detection Prevalence : 0.5000
 Balanced Accuracy : 0.6888
 'Positive' Class : 1

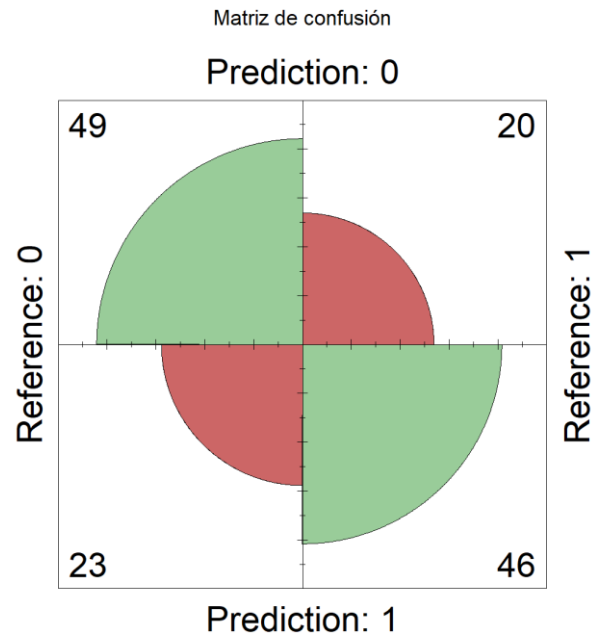


Gráfico 8.

14. Se realizó una serie de comprobaciones para garantizar que los datos son adecuados para el cálculo de la curva ROC.

15. Se graficó la curva ROC, que es una representación gráfica del rendimiento del clasificador.

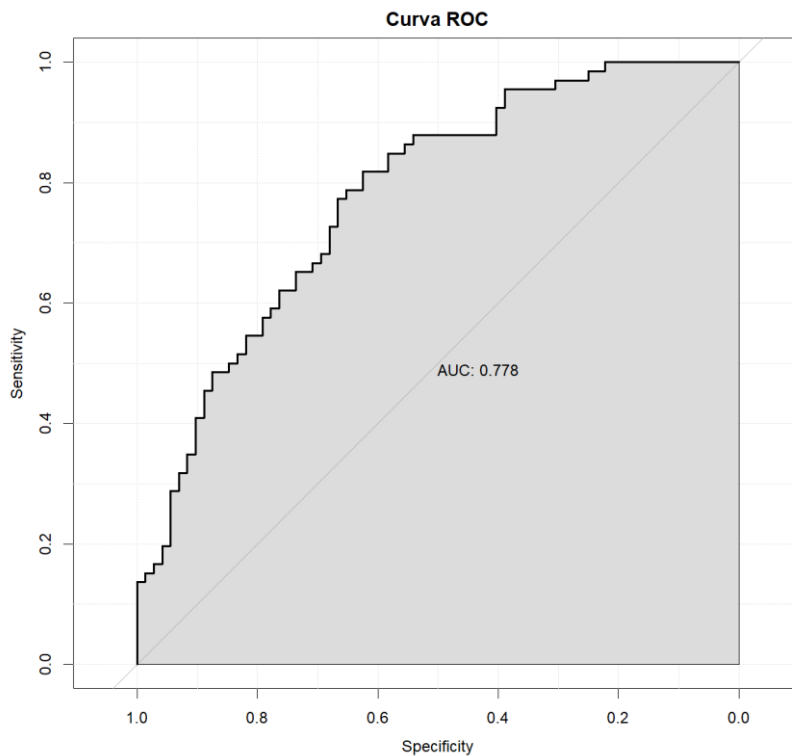


Gráfico 9

16. Se obtuvieron los coeficientes de las variables con importancia en la clasificación (coeficientes no nulos) y se imprimieron. Estas son las variables que el modelo identificó como significativas para la predicción de la clase positiva de la variable EGFR.

Tabla 8

ID	Variable	Coefficiente
1	(Intercept)	1.1573170
2	Edad	-0.0047710
3	Tabaquismo	-0.4602699
4	MORPHOLOGICAL_Volume(IBSI:RNU0)	-0.0000010
5	MORPHOLOGICAL_ApproximateVolume(IBSI:YEKZ)	-0.0000011
6	MORPHOLOGICAL_SurfaceArea(IBSI:C0JK)	-0.0000070
7	MORPHOLOGICAL_SphericalDisproportion(IBSI:KRCK)	-0.0484106
8	MORPHOLOGICAL_Sphericity(IBSI:QCFX)	0.0054944
9	MORPHOLOGICAL_Asphericity(IBSI:25C7)	-0.0476901
10	MORPHOLOGICAL_Maximum3DDiameter(IBSI:L0JK)	-0.0027241
11	MORPHOLOGICAL_IntegratedIntensity(IBSI:99N0)	0.0000000
12	INTENSITY-BASED_MaximumGreyLevel(SUVbw)IBSI:84IY	-0.0005064
13	INTENSITY-BASED_TotalLesionGlycolysis(SUVbw)IBSI:No	-0.0000106
14	LOCAL_INTENSITY_BASED_IntensityPeakDiscretizedVolumeSought(0.5mL)(mL)IBSI:No	0.4094848
15	LOCAL_INTENSITY_HISTOGRAM_IntensityPeakDiscretizedVolumeSought(0.5mL)(mL)IBSI:No	0.4053978
16	GLCM_DifferenceAverage(IBSI:TF7R)	-0.0007852
17	GLCM_Dissimilarity(IBSI:8S9J)	-0.0006805
18	NGTDM_Complexity(IBSI:HDEZ)	-0.0000272
19	GLSZM_SmallZoneLowGreyLevelEmphasis(IBSI:5RAI)	6.2405510
20	GLSZM_GreyLevelNonUniformity(IBSI:JNSA)	-0.0001640
21	GLSZM_NormalisedZoneSizeNonUniformity(IBSI:VB3A)	-1.4255470

17. Se muestra el mapa de correlaciones de las variables seleccionadas a partir del modelo de regresión regularizada elastic net, el cual consta de 21 variables.

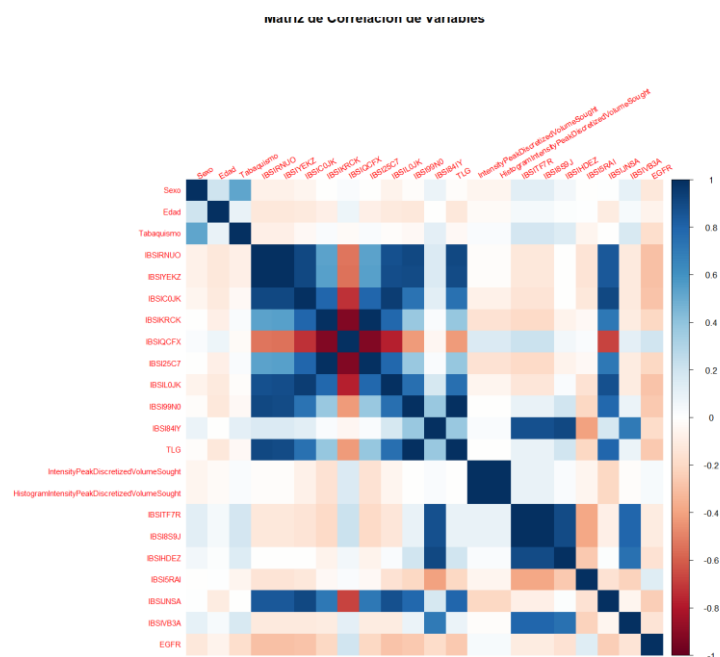


Gráfico 10

18. Una vez realizada la selección de las variables mediante método de regresión penalizada se procedió a realizar un modelo de regresión logística convencional con las variables seleccionadas.
19. Con los datos ya cargados, se ajusta un modelo de regresión logística. Este modelo busca predecir la variable objetivo EGFR (en función de todas las demás variables del conjunto de datos) y estima la probabilidad de que una observación pertenezca a una de dos categorías posibles.
20. Finalmente, se produce un resumen del modelo ajustado.

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.897e+00	8.388e+00	0.584	0.5593
Sexo	-6.102e-02	4.190e-01	-0.146	0.8842
Edad	-1.242e-02	1.236e-02	-1.004	0.3152
Tabaquismo	-8.024e-01	3.999e-01	-2.006	0.0448 *
IBSIRNUO	-7.036e-04	4.194e-04	-1.678	0.0934 .
IBSIYEKZ	6.891e-04	4.145e-04	1.663	0.0964 .
IBSIC0JK	-2.883e-04	1.369e-04	-2.105	0.0353 *
IBSIKRCK	-8.567e-01	2.861e+00	-0.299	0.7646
IBSIQCFX	-3.172e+00	5.793e+00	-0.548	0.5840
IBSI25C7	NA	NA	NA	NA
IBSIL0JK	1.854e-02	1.831e-02	1.013	0.3112
IBSI99N0	8.107e-05	7.051e-05	1.150	0.2503
IBSI84IY	-1.804e-02	9.693e-02	-0.186	0.8523
TLG	-7.923e-02	6.996e-02	-1.133	0.2574
IntensityPeakDiscretizedVolumeSought	2.165e+00	2.557e+00	0.847	0.3971
HistogramIntensityPeakDiscretizedVolumeSought	NA	NA	NA	NA
IBSITF7R	1.100e-01	2.560e-01	0.430	0.6674
IBSI8S9J	NA	NA	NA	NA
IBSIHDEZ	1.924e-05	2.553e-04	0.075	0.9399
IBSI5RAI	1.455e+01	2.914e+01	0.499	0.6176
IBSIJNSA	9.945e-02	4.071e-02	2.443	0.0146 *
IBSIVB3A	-4.951e+00	3.644e+00	-1.359	0.1742

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 269.67 on 195 degrees of freedom
 Residual deviance: 217.68 on 177 degrees of freedom
 AIC: 255.68

Number of Fisher Scoring iterations: 6

21. Este resumen proporciona información sobre el rendimiento del modelo, tales como los coeficientes estimados para cada variable predictora, el grado de significancia de cada coeficiente y el intervalo de confianza asociado, mostrando 3 variables con datos estadísticamente significativos:

- Tabaquismo: Esta variable ha dado un estimado (Estimate) de -0.8024. Esto sugiere que, manteniendo constantes todas las demás variables, un incremento de una unidad en el "Tabaquismo" disminuiría la log-odds de la variable dependiente EGFR en 0.8024. El error estándar (Std. Error) de este estimado es 0.3999, y el valor z asociado es -2.006. El valor p (Pr(>|z|)) asociado es 0.0448, que es menor que 0.05, por lo que podemos decir que la variable "Tabaquismo" es estadísticamente significativa en el nivel de confianza del 95%.
- IBSI:C0JK: Esta variable tiene un estimado de -0.0002883. Este es un coeficiente muy pequeño, lo cual sugiere que el efecto de esta variable en la variable dependiente es relativamente pequeño. El error estándar de este coeficiente es 0.0001369, y el valor z asociado es -2.105. El valor p asociado es 0.0353, que es menor que 0.05, por lo que podemos decir que esta variable es también estadísticamente significativa en el nivel de confianza del 95%.
- IBSI:JNSA: El estimado para esta variable es de 0.09945, lo cual indica que un incremento de una unidad en "IBSIJNSA" incrementaría la log-odds de la variable dependiente en 0.09945, manteniendo constantes todas las demás variables. El error estándar de este coeficiente es 0.04071, y el valor z es 2.443. El valor p

asociado es 0.0146, que es menor que 0.05, indicando que esta variable es estadísticamente significativa en el nivel de confianza del 95%.

- d. Los resultados también incluyen el deviance nulo (Null deviance) y el deviance residual (Residual deviance), que son medidas de la bondad de ajuste del modelo. El deviance nulo es el valor del deviance para un modelo que solo incluye la intersección (269.67), mientras que el deviance residual es el valor del deviance para el modelo que se ajustó (217.68). La disminución en el deviance sugiere que el modelo que se ajustó se ajusta mejor a los datos que el modelo nulo.
- e. El criterio de información de Akaike (AIC) es una medida de la calidad relativa de un modelo estadístico. En este caso, el valor del AIC es 255.68.

22. Finalmente, con base en los datos obtenidos con la regresión logística posterior a la regresión penalizada Elastic net, procedemos a generar la firma radiómica y evaluar su rendimiento.

23. La firma radiómica es una ecuación que combina los coeficientes obtenidos de la regresión logística para cada característica radiómica seleccionada, cada uno de los cuales es ponderado por el valor de esa característica radiómica para el individuo en particular. La puntuación de la firma radiómica para un individuo sería la suma de cada coeficiente multiplicado por el valor de la característica correspondiente para ese individuo.

$$\text{FIRMA RADIÓMICA} = (-0.8024 * \text{Tabaquismo}) + (-0.0002883 * \text{IBSIC0JK}) + (0.09945 * \text{IBSIJNSA})$$

Donde Tabaquismo, IBSIC0JK e IBSIJNSA son las características de un individuo y -0.8024, -0.0002883 y 0.09945 son los coeficientes obtenidos de la regresión logística, respectivamente. IBSIC0JK= MORPHOLOGICAL_SurfaceArea(IBSI:C0JK), IBSIJNSA= GLSZM_GreyLevelNonUniformity(IBSI:JNSA).

Las puntuaciones de la firma radiómica pueden ser interpretadas como las 'puntuaciones de riesgo' para el resultado que se está modelando (en este caso, la variable EGFR). Una puntuación más alta significa un mayor riesgo relativo o mayor probabilidad de tener el resultado, según lo modelado por la regresión logística.

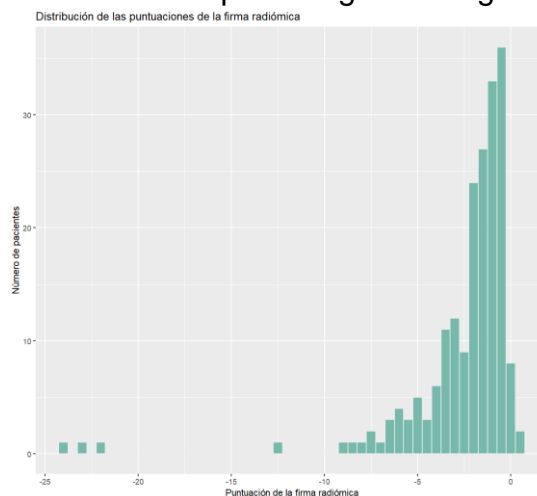


Gráfico 11. Las puntuaciones de la firma radiómica reflejan la combinación de características radiómicas que mejor predicen el resultado de interés, según nuestro modelo. En el eje X (horizontal), la "Puntuación de la firma radiómica". Cada punto en este eje representa un valor específico de la firma radiómica que se ha calculado para los pacientes en tu estudio. En el eje Y (vertical), tienes el "Número de pacientes". Cada punto en este eje indica cuántos pacientes en tu estudio obtuvieron la puntuación correspondiente en el eje X.

24. Se muestra cómo estas puntuaciones se distribuyen entre los individuos del estudio y cómo difieren según el estado del resultado.

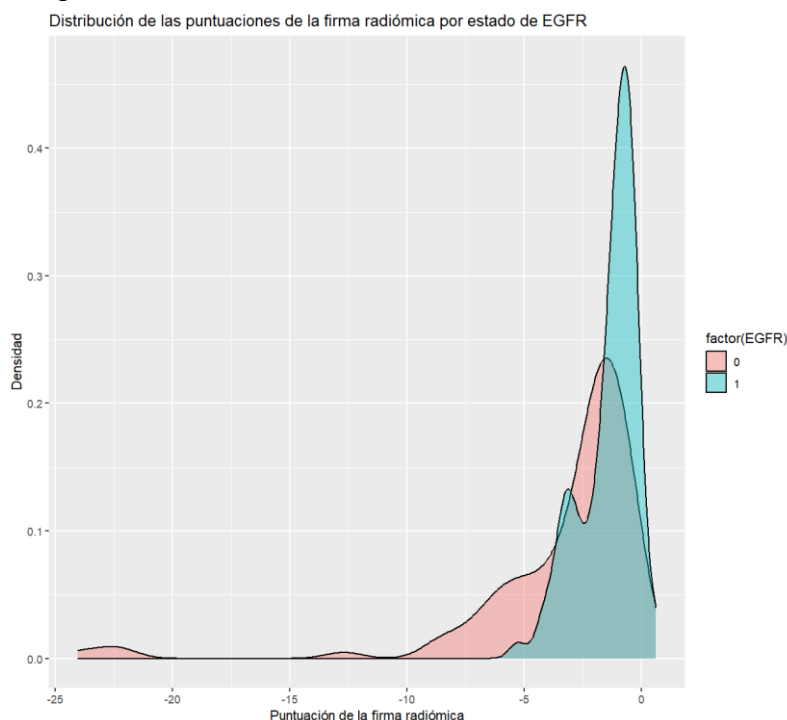


Gráfico 12. Gráfico de densidad para visualizar cómo se distribuyen las puntuaciones de la firma radiómica entre los diferentes estados de la variable dependiente (EGFR).

25. Usamos estas puntuaciones para calcular la probabilidad de tener el resultado (mediante la función logit inversa o sigmoide) lo cual significa que se puede usar la firma radiómica para clasificar a los individuos en categorías de riesgo, lo que puede ser útil para la toma de decisiones clínicas o la estratificación de pacientes en investigaciones futuras.

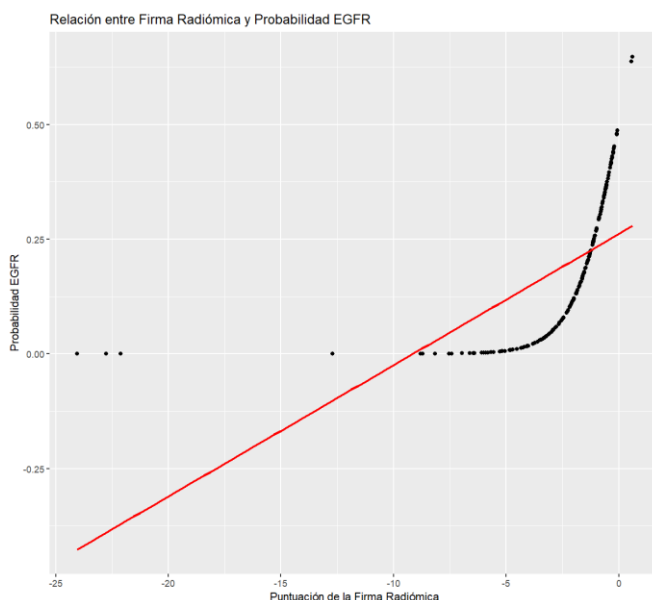


Gráfico 13. Este gráfico presenta la relación entre las puntuaciones de la firma radiómica (eje X) y la probabilidad calculada de que EGFR sea positivo (eje Y), para cada observación en el conjunto de datos. Los puntos individuales en el gráfico representan observaciones individuales del conjunto de datos, cada uno con una puntuación de firma radiómica y una probabilidad correspondiente de que EGFR sea positivo. La línea roja en el gráfico es una línea de ajuste lineal, que se calcula utilizando un modelo de regresión lineal (lm). Esta línea de ajuste intenta describir la tendencia general en los datos, muestra cómo cambia la probabilidad calculada de EGFR en promedio a medida que cambia la puntuación de la firma radiómica.

Es importante tener en cuenta que debido a que estamos trabajando con una probabilidad y una puntuación de firma radiómica, no se espera que esta relación sea perfectamente lineal. Sin embargo, la línea de ajuste puede ayudarnos a visualizar la tendencia general en los datos. La línea de ajuste tiene una pendiente positiva, esto indicaría que a medida que la puntuación de la firma radiómica aumenta, también lo hace la probabilidad de que EGFR sea positivo. Sin embargo, el grado en que esto es cierto (es decir, qué tan fuerte es la relación) y si la relación es estadísticamente significativa, son preguntas que requerirían un análisis más formal.

26. Se calculo la curva ROC y el área bajo la curva (AUC) para evaluar el rendimiento de la firma radiómica

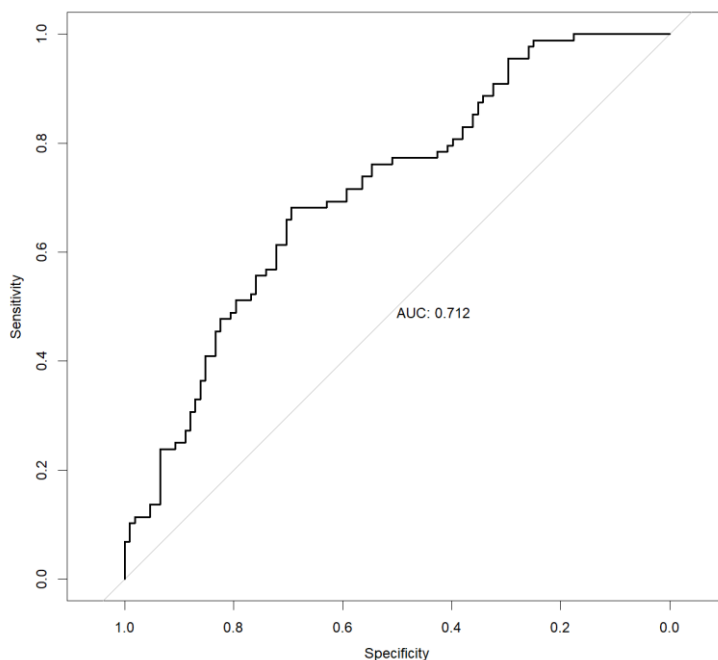


Gráfico 14

27. Determinación de umbral óptimo de clasificación

Buscamos el umbral óptimo, el cual, es un valor de probabilidad predicha que se utiliza para clasificar las observaciones en las dos clases posibles, en este caso EGFR 0 o 1.

La firma radiómica se ha convertido en una probabilidad utilizando la función logística inversa o función sigmoïdal. Por lo tanto, la salida es una probabilidad que varía entre 0 y 1. Al generar la curva ROC, observamos cómo varía el rendimiento del modelo (en términos de sensibilidad y especificidad) para diferentes umbrales de probabilidad.

El umbral óptimo, en este caso, es el valor de probabilidad que maximiza la suma de la sensibilidad y la especificidad y fue igual a 0.24087138321101.

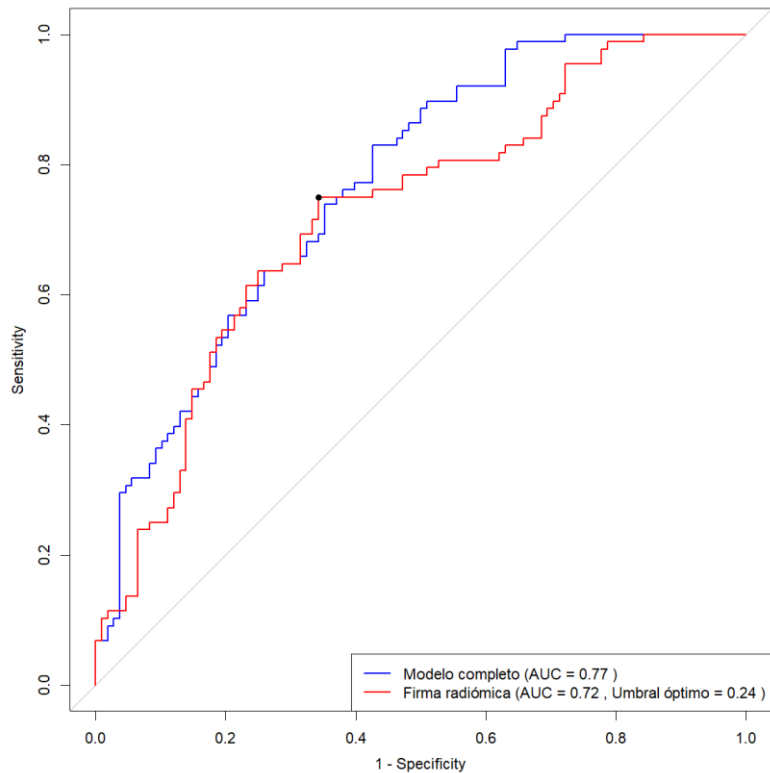


Gráfico 15

Por lo tanto, cuando hablamos de este "umbral óptimo", nos estamos refiriendo a un valor de probabilidad que estamos utilizando para tomar una decisión sobre si predecir EGFR 0 o EGFR 1. Con el valor de umbral óptimo, valoramos gráficamente el rendimiento de la predicción del estado mutacional de EGFR según el puntaje de la firma radiómica en términos probabilísticos.

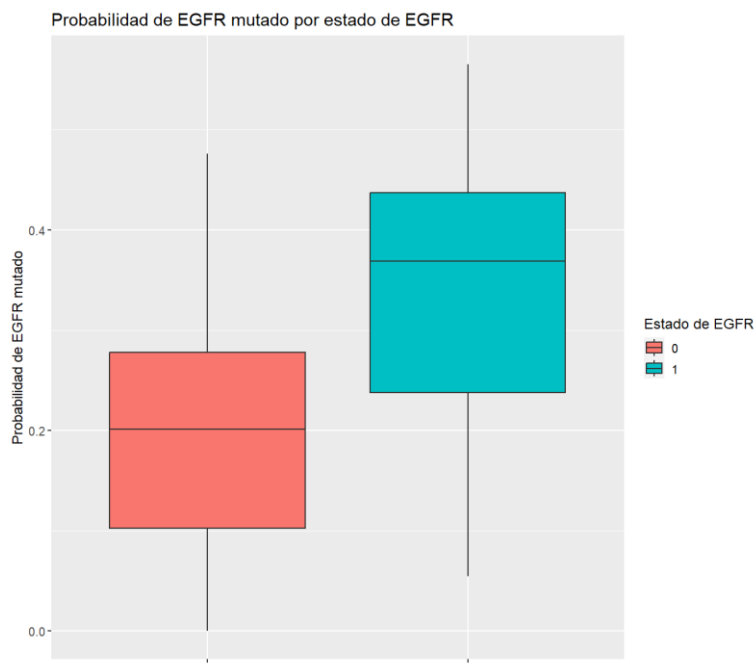


Gráfico 16. Este boxplot muestra la distribución de las probabilidades predichas de tener EGFR mutado (en el eje y) para cada estado real de EGFR (en el eje x, con 0 indicando no mutado y 1 indicando mutado). Las cajas representan el rango intercuartílico (es decir, el 50% central de las observaciones), con la línea dentro de la caja que indica la mediana de la distribución. Los bigotes se extienden desde la caja hasta el valor más extremo dentro de 1.5 veces el rango intercuartílico.

La caja para EGFR=1 (mutado) se encuentra más alta en el eje que la caja para EGFR=0 (no mutado), esto sugeriría que el modelo está prediciendo correctamente una mayor probabilidad de

tener EGFR mutado para los pacientes que realmente tienen EGFR mutado en comparación con aquellos que no lo tienen. En contraste, cualquier superposición entre las cajas indicaría una cierta cantidad de error en las predicciones del modelo.

28. Ejemplo de uso de la firma radiómica en un paciente hipotético:

Supongamos que tenemos un paciente llamado Juan, es fumador (asignamos 1 a Tabaquismo), su valor por análisis radiómico de las variables: IBSI:C0JK es 2000 y para IBSI:JNSA es 0.75.

Aplicando los datos de Juan en la ecuación obtenemos:

$$\text{Firma radiómica de Juan} = (-0.8024 * 1) + (-0.0002883 * 2000) + (0.09945 * 0.75)$$

Posteriormente, para transformar la puntuación de la firma radiómica en una probabilidad de tener el resultado (en este caso, EGFR), aplicamos la función logit inversa (o función sigmoïdal) a la puntuación de la firma radiómica:

$$\text{Probabilidad_EGFR} = \exp(\text{firma_radiómica}) / (1 + \exp(\text{firma_radiómica}))$$

Sustituyendo firma_radiómica con la puntuación que calculamos para Juan, obtenemos una probabilidad que representa el riesgo estimado de Juan de tener EGFR.

Si el objetivo es clasificar a los pacientes en base a su riesgo, podríamos establecer un umbral de probabilidad para determinar si un paciente se considera de "alto riesgo" o "bajo riesgo".

Primero calculamos la firma radiómica para Juan:

$$\text{Firma radiómica de Juan} = (-0.8024 * 1) + (-0.0002883 * 2000) + (0.09945 * 0.75) = -0.8024 - 0.5766 + 0.0745875 = -1.3044125$$

Ahora, vamos a aplicar esta puntuación en la función logit inversa para obtener la probabilidad de EGFR para Juan:

$$\text{Probabilidad_EGFR} = \exp(-1.3044125) / (1 + \exp(-1.3044125))$$

Calculamos las funciones exponenciales, obtenemos:

$$\text{Probabilidad_EGFR} = 0.271 / (1 + 0.271) = 0.271 / 1.271 = 0.213$$

Por lo tanto, la probabilidad estimada de que Juan tenga EGFR mutado, de acuerdo con el modelo y su firma radiómica, es de aproximadamente 0.213 o 21.3%.

Si establecemos un valor umbral de 0.24 (24%) para alto riesgo, dado que este valor es menor que el umbral, clasificaríamos a Juan como un paciente de "bajo riesgo" para la presencia de mutación EGFR.

12. Discusión.

En este estudio, utilizamos técnicas de análisis radiómico, regresión logística regularizada y regresión logística convencional para identificar un conjunto de características radiómicas capaces de predecir el estatus mutacional de EGFR. Estas características incluyen el tabaquismo, MORPHOLOGICAL_SurfaceArea (IBSI:C0JK) y GLSZM_GreyLevelNonUniformity (IBSI:JNSA). Estas características luego se utilizaron para construir una firma radiómica que predice EGFR con un área bajo la curva (AUC) del 72.4% y se determinó su valor umbral óptimo que es 0.24, dicha ecuación o firma radiómica es una propuesta de gran valor por su fácil aplicabilidad en un entorno clínico real.

Los resultados de nuestro estudio coinciden y en algunos casos, expanden los hallazgos previos en la literatura científica de estudios con intenciones similares pero con grandes diferencias en el tipo de análisis radiómico, el tipo de análisis estadístico y por tanto la potencial implementación clínica que en su mayoría es limitada o nula debido a la complejidad de los modelos resultantes y la baja reproducibilidad (57–65). Por ejemplo, confirmamos la asociación inversa entre el tabaquismo y la presencia de la mutación EGFR, un hallazgo que ha sido descrito en varios estudios (66,67). Asimismo, nuestra investigación muestra una fuerte correlación positiva entre GLSZM_GreyLevelNonUniformity (IBSI:JNSA) y la presencia de la mutación EGFR, en línea con lo que previamente se ha observado (60,61). Además, descubrimos una fuerte relación inversa entre MORPHOLOGICAL_SurfaceArea (IBSI:C0JK) y la presencia de la mutación, lo que no ha sido previamente reportado.

A pesar de las fortalezas de nuestro estudio, también reconocemos sus limitaciones. Por ejemplo, nuestra muestra está compuesta mayoritariamente por pacientes en etapas avanzadas de la enfermedad, lo cual refleja la población atendida en hospitales de tercer nivel pero limita la generalización de los resultados a poblaciones en etapas tempranas. Además, en nuestro análisis, los fumadores fueron categorizados de acuerdo con su historia, sin distinguir entre fumadores actuales y aquellos que han dejado el hábito, lo que podría haber impactado en los resultados. Adicionalmente, se excluyeron varias variables radiómicas que no pudieron ser estandarizadas debido a los tipos de valores proporcionados por el software de análisis. Este factor podría haber limitado el potencial de nuestro modelo al omitir variables que podrían haber mejorado su capacidad predictiva.

En términos de la implementación de Machine Learning, es importante considerar el riesgo de sobreajuste, es decir, el modelo se ajusta tan bien a los datos de entrenamiento que pierde capacidad para generalizar en nuevos datos. Aunque nuestro estudio ha utilizado técnicas de regresión regularizada para minimizar este riesgo, la validación en muestras independientes es crucial para evaluar la capacidad de generalización del modelo (68).

A pesar de estas limitaciones, este estudio tiene varias fortalezas. Utilizamos la regresión logística regularizada Elastic net, que combina las fortalezas de la regresión LASSO y Ridge para seleccionar características, permitiendo el manejo de un gran número de características y evitando el sobreajuste. Validamos nuestra firma radiómica en un conjunto de datos independiente, lo que fortalece nuestra confianza en su capacidad para generalizarse a nuevos datos.

Finalmente, la interpretabilidad de los modelos de Machine Learning puede ser un desafío, especialmente en modelos complejos, lo que puede dificultar su adopción en la práctica clínica(69).

Sin embargo, la firma radiómica que desarrollamos está compuesta solo por tres variables (dos radiómicas y una clínica), y proporciona un AUC de 0.712 [95% CI: 0.6526-0.7952 (DeLong)]. Esto sugiere que la firma podría ser fácilmente implementada en la práctica clínica, contribuyendo de forma confiable en la toma de decisiones clínicas que optimicen el abordaje terapéutico.

13. Conclusiones.

Nuestros hallazgos sugieren que el análisis radiómico de PET/CT con FDG y la construcción de una firma radiómica mediante machine learning, ofrecen un método no invasivo, ágil, barato y potencialmente útil en la toma de decisiones clínicas. Sin embargo, aunque este estudio proporciona un punto de partida prometedor, es necesario realizar estudios adicionales para validar y refinar la ecuación de lo que hemos llamado firma radiómica en un estudio multicéntrico, diferentes poblaciones de pacientes y entornos clínicos que permitan la generalización de los resultados.

14. Bibliografía.

1. Manafi-Farid R, Karamzade-Ziarati N, Vali R, Mottaghy FM, Beheshti M. 2-[18F]FDG PET/CT radiomics in lung cancer: An overview of the technical aspect and its emerging role in management of the disease. *Methods*. 2021 Apr 1;188:84–97.
2. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin*. 2022 Jan 12;72(1):7–33.
3. Pao W, Girard N. New driver mutations in non-small-cell lung cancer. *Lancet Oncol*. 2011 Feb;12(2):175–80.
4. Pao W, Hutchinson KE. Chipping away at the lung cancer genome. *Nat Med*. 2012 Mar 6;18(3):349–51.
5. Lindeman NI, Cagle PT, Beasley MB, Chitale DA, Dacic S, Giaccone G, et al. Molecular Testing Guideline for Selection of Lung Cancer Patients for EGFR and ALK Tyrosine Kinase Inhibitors Guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. *The Journal of Molecular Diagnostics*. 2013;15(4):415–53.
6. Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, Wistuba II, et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst*. 2005;97(5):339–46.
7. Hsu WH, Yang JCH, Mok TS, Loong HH. Overview of current systemic management of EGFR-mutant NSCLC. *Annals of Oncology*. 2018;29(Supplement 1):i3–9.
8. Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, et al. EGFR Mutations in Lung Cancer: Correlation with Clinical Response to Gefitinib Therapy. *Science* (1979). 2004;304(5676).
9. Maemondo M, Inoue A, Kobayashi K, Sugawara S, Oizumi S, Isobe H, et al. Gefitinib or Chemotherapy for Non-Small-Cell Lung Cancer with Mutated EGFR. *N Engl J Med*. 2010;25(24).
10. Zhou C, Wu YL, Chen G, Feng J, Liu XQ, Wang C, et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *Lancet Oncol*. 2011 Aug;12(8):735–42.
11. Loriot Y, Mordant P, Deutsch E, Olausson KA, Soria JC. Are RAS mutations predictive markers of resistance to standard chemotherapy? *Nat Rev Clin Oncol*. 2009 Sep 14;6(9):528–34.

12. Pao W, Wang TY, Riely GJ, Miller VA, Pan Q, Ladanyi M, et al. KRAS Mutations and Primary Resistance of Lung Adenocarcinomas to Gefitinib or Erlotinib. *PLoS Med*. 2005;2(1):e17.
13. Lee CK, Man J, Lord S, Links M, GebSKI V, Mok T, et al. Checkpoint Inhibitors in Metastatic EGFR-Mutated Non-Small Cell Lung Cancer—A Meta-Analysis. *Journal of Thoracic Oncology*. 2017;12(2):403–7.
14. Pao W, Ladanyi M. Epidermal Growth Factor Receptor Mutation Testing in Lung Cancer: Searching for the Ideal Method. *Clinical Cancer Research*. 2007;13(17):4954–5.
15. Weber WA, Petersen V, Schmidt B, Tyndale-Hines L, Link T, Peschel C, et al. Positron Emission Tomography in Non-Small-Cell Lung Cancer: Prediction of Response to Chemotherapy by Quantitative Assessment of Glucose Use. *Journal of Clinical Oncology*. 2003 Jul 15;21(14):2651–7.
16. National Comprehensive Cancer Network. NCCN Guidelines Version 3.2023 Non-Small Cell Lung Cancer [Internet]. 2023 Apr [cited 2023 Jul 20]. Available from: <https://www.nccn.org/guidelines/guidelines-detail?category=1&id=1450>
17. Caicedo C, Garcia-Velloso MJ, Lozano MD, Labiano T, Vigil Diaz C, Lopez-Picazo JM, et al. Role of [18F]FDG PET in prediction of KRAS and EGFR mutation status in patients with advanced non-small-cell lung cancer. *Eur J Nucl Med Mol Imaging*. 2014 Nov 3;41(11):2058–65.
18. Na I II, Byun BH, Kim KM, Cheon GJ, Choe DH, Koh JS, et al. 18F-FDG uptake and EGFR mutations in patients with non-small cell lung cancer: A single-institution retrospective analysis. *Lung Cancer*. 2010 Jan;67(1):76–80.
19. Huang CT, Yen RF, Cheng MF, Hsu YC, Wei PF, Tsai YJ, et al. Correlation of F-18 fluorodeoxyglucose-positron emission tomography maximal standardized uptake value and EGFR mutations in advanced lung adenocarcinoma. *Medical Oncology*. 2010 Mar 7;27(1):9–15.
20. Kawada K, Nakamoto Y, Kawada M, Hida K, Matsumoto T, Murakami T, et al. Relationship between 18 F-Fluorodeoxyglucose Accumulation and KRAS / BRAF Mutations in Colorectal Cancer. *Clinical Cancer Research*. 2012 Mar 15;18(6):1696–703.
21. Yip SSF, Kim J, Coroller TP, Parmar C, Velazquez ER, Huynh E, et al. Associations Between Somatic Mutations and Metabolic Imaging Phenotypes in Non-Small Cell Lung Cancer. *Journal of Nuclear Medicine*. 2017 Apr;58(4):569–76.
22. Lv Z, Fan J, Xu J, Wu F, Huang Q, Guo M, et al. Value of 18F–FDG PET/CT for predicting EGFR mutations and positive ALK expression in patients with non-small cell lung cancer: a retrospective analysis of 849 Chinese patients. *Eur J Nucl Med Mol Imaging* [Internet]. 2018 May 21 [cited 2022 Dec 28];45(5):735–50. Available from: <http://link.springer.com/10.1007/s00259-017-3885-z>
23. Cook GJR, Siddique M, Taylor BP, Yip C, Chicklore S, Goh V. Radiomics in PET: principles and applications. *Clin Transl Imaging*. 2014 Jun 3;2(3):269–76.
24. Rahim MK, Kim SE, So H, Kim HJ, Cheon GJ, Lee ES, et al. Recent Trends in PET Image Interpretations Using Volumetric and Texture-based Quantification Methods in Nuclear Oncology. *Nucl Med Mol Imaging*. 2014 Mar 22;48(1):1–15.
25. Hatt M, Tixier F, Visvikis D, Cheze Le Rest C. Radiomics in PET/CT: More Than Meets the Eye? *Journal of Nuclear Medicine*. 2017 Mar;58(3):365–6.
26. Buvat I, Orhac F, Soussan M. Tumor Texture Analysis in PET: Where Do We Stand? *Journal of Nuclear Medicine*. 2015;56(11).
27. Hanahan D. Hallmarks of Cancer: New Dimensions. *Cancer Discov*. 2022;12(1):31–46.
28. Gillies RJ, Anderson AR, Gatenby RA, Morse DL. The biology underlying molecular imaging in oncology: from genome to anatome and back again.

29. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016 Feb;278(2):563–77.
30. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, et al. The applications of radiomics in precision diagnosis and treatment of oncology: Opportunities and challenges. *Theranostics*. 2019;9(5):1303–22.
31. Bailey DL, Willowson KP. Quantitative SPECT/CT: SPECT joins PET as a quantitative imaging modality.
32. Lohith G. Radiomic Imaging Analysis-To Renovate RECIST Criteria? *Cancer Ther Oncol Int J*. 2017 Jul 18;6(1):1–2.
33. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*. 2013 Sep 19;501(7467):338–45.
34. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer*. 2012;12:323–34.
35. Cook GJR, Goh V. What can artificial intelligence teach us about the molecular mechanisms underlying disease? *Eur J Nucl Med Mol Imaging*. 2019 Dec 12;46(13):2715–21.
36. Yip SSF, Aerts HJWL. Applications and limitations of radiomics. *Phys Med Biol*. 2016 Jul 7;61(13):R150–66.
37. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014 Sep 22;5(1):4006.
38. 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. Vol. 90, *British Journal of Radiology*. 2017.
39. Thuillier P, Liberini V, Rampado O, Gallio E, De Santi B, Ceci F, et al. Diagnostic value of conventional pet parameters and radiomic features extracted from 18f-fdg-pet/ct for histologic subtype classification and characterization of lung neuroendocrine neoplasms. *Biomedicines*. 2021;9(3):1–14.
40. Cucchiara F, Del Re M, Valleggi S, Romei C, Petrini I, Lucchesi M, et al. Integrating Liquid Biopsy and Radiomics to Monitor Clonal Heterogeneity of EGFR-Positive Non-Small Cell Lung Cancer. *Front Oncol*. 2020 Dec 16;10(December):1–8.
41. Park C, Na KJ, Choi H, Ock CY, Ha S, Kim M, et al. Tumor immune profiles noninvasively estimated by FDG PET with deep learning correlate with immunotherapy response in lung adenocarcinoma. *Theranostics*. 2020;10(23):10838–48.
42. Shao X, Niu R, Shao X, Jiang Z, Wang Y. Value of 18F-FDG PET/CT-based radiomics model to distinguish the growth patterns of early invasive lung adenocarcinoma manifesting as ground-glass opacity nodules. *EJNMMI Res*. 2020 Dec 13;10(1):80.
43. Hu Y, Zhao X, Zhang J, Han J, Dai M. Value of 18F-FDG PET/CT radiomic features to distinguish solitary lung adenocarcinoma from tuberculosis. *Eur J Nucl Med Mol Imaging*. 2021 Jan 25;48(1):231–40.
44. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present... any future? *Eur J Nucl Med Mol Imaging*. 2017 Jan 6;44(1):151–65.
45. Zhou Y, Ma XL, Zhang T, Wang J, Zhang T, Tian R. Use of radiomics based on 18 F-FDG PET/CT and machine learning methods to aid clinical decision-making in the classification of solitary pulmonary lesions: an innovative approach.
46. Hatt M, Cheze C, Rest L, Antonorsi N, Tixier F, Tankyevych O, et al. Radiomics in PET/CT: Current Status and Future AI-Based Evolutions. 2021;

47. Hatt M, Krizsan AK, Rahmim A, Bradshaw TJ, Costa PF, Forgacs A, et al. Joint EANM/SNMMI Guideline on Radiomics in Nuclear Medicine.
48. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology*. 2020 May 21;295(2):328–38.
49. Delso G, Cirillo D, Kaggie JD, Valencia A, Metser U, Veit-Haibach P. How to Design AI-Driven Clinical Trials in Nuclear Medicine. 2021;
50. Bradshaw TJ, Boellaard R, Dutta J, Jha AK, Jacobs P, Li Q, et al. Nuclear Medicine and Artificial Intelligence: Best Practices for Algorithm Development. *THE JOURNAL OF NUCLEAR MEDICINE*. 2022;63(4):500–10.
51. Barrios CH, Werutsky G, Mohar A, Ferrigno AS, Müller BG, Bychkovsky BL, et al. Cancer control in Latin America and the Caribbean: recent advances and opportunities to move forward [Internet]. Vol. 22, Series Lancet Oncol. 2021. Available from: www.thelancet.com/oncology
52. Wu J, Aguilera T, Shultz D, Gudur M, Rubin DL, Loo BW, et al. Early-Stage Non–Small Cell Lung Cancer: Quantitative Imaging Characteristics of 18 F Fluorodeoxyglucose PET/CT Allow Prediction of Distant Metastasis. *Radiology*. 2016 Oct;281(1):270–8.
53. Rios Velazquez E, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfield O, et al. Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer. *Cancer Res*. 2017 Jul 15;77(14):3922–30.
54. Mostafa R, Kandeel A, Abdelkareem M, Nardo L, Abdelhafez Y. Association of robust radiomic features from staging 18F-FDG PET/CT in lung cancer with EGFR expression and overall survival. *Journal of Nuclear Medicine*. 2021;62(supplement 1).
55. Arrieta O, Cardona AF, Martín C, Más-López L, Corrales-Rodríguez L, Bramuglia G, et al. Updated Frequency of EGFR and KRAS Mutations in NonSmall-Cell Lung Cancer in Latin America: The Latin-American Consortium for the Investigation of Lung Cancer (CLICaP). *Journal of Thoracic Oncology*. 2015 May;10(5):838–43.
56. LIFEx - LIFEx [Internet]. [cited 2022 May 25]. Available from: <https://www.lifexsoft.org/>
57. Lv Z, Fan J, Xu J, Wu F, Huang Q, Guo M, et al. Value of 18F–FDG PET/CT for predicting EGFR mutations and positive ALK expression in patients with non-small cell lung cancer: a retrospective analysis of 849 Chinese patients. *Eur J Nucl Med Mol Imaging* [Internet]. 2018 May 21;45(5):735–50. Available from: <http://link.springer.com/10.1007/s00259-017-3885-z>
58. Cho A, Hur J, Moon YW, Hong SR, Suh YJ, Kim YJ, et al. Correlation between EGFR gene mutation, cytologic tumor markers, 18F-FDG uptake in non-small cell lung cancer. *BMC Cancer* [Internet]. 2016 Dec 16;16(1):224. Available from: <http://bmccancer.biomedcentral.com/articles/10.1186/s12885-016-2251-z>
59. Zheng L, Xie H, Luo X, Yang Y, Zhang Y, Li Y, et al. Radiomic Signatures for Predicting EGFR Mutation Status in Lung Cancer Brain Metastases. *Front Oncol* [Internet]. 2022 Jul 14;12. Available from: <https://www.frontiersin.org/articles/10.3389/fonc.2022.931812/full>
60. Huang W, Wang J, Wang H, Zhang Y, Zhao F, Li K, et al. PET/CT Based EGFR Mutation Status Classification of NSCLC Using Deep Learning Features and Radiomics Features. *Front Pharmacol* [Internet]. 2022 Apr 27;13. Available from: <https://www.frontiersin.org/articles/10.3389/fphar.2022.898529/full>
61. Wang S, Shi J, Ye Z, Dong D, Yu D, Zhou M, et al. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. *European Respiratory Journal* [Internet]. 2019 Mar 1;53(3):1800986. Available from: <http://erj.ersjournals.com/lookup/doi/10.1183/13993003.00986-2018>

62. Jiang M, Zhang Y, Xu J, Ji M, Guo Y, Guo Y, et al. Assessing EGFR gene mutation status in non-small cell lung cancer with imaging features from PET/CT. *Nucl Med Commun* [Internet]. 2019 Aug;40(8):842–9. Available from: <https://journals.lww.com/00006231-201908000-00010>
63. Yang C, Chen W, Gong G, Li Z, Qiu Q, Yin Y. Application of CT radiomics features to predict the EGFR mutation status and therapeutic sensitivity to TKIs of advanced lung adenocarcinoma. *Transl Cancer Res* [Internet]. 2020 Nov 1;9(11):6683–90. Available from: <http://tcr.amegroups.com/article/view/46111/html>
64. Liu Q, Sun D, Li N, Kim J, Feng D, Huang G, et al. Predicting EGFR mutation subtypes in lung adenocarcinoma using 18F-FDG PET/CT radiomic features. *Transl Lung Cancer Res* [Internet]. 2020 Jun 1;9(3):549–62. Available from: <http://tlcr.amegroups.com/article/view/39275/html>
65. Choi YJ, Cho BC, Jeong YH, Seo HJ, Kim HJ, Cho A, et al. Correlation Between 18F-Fluorodeoxyglucose Uptake and Epidermal Growth Factor Receptor Mutations in Advanced Lung Cancer. *Nucl Med Mol Imaging* [Internet]. 2012 Sep 7;46(3):169–75. Available from: <http://link.springer.com/10.1007/s13139-012-0142-z>
66. Midha A, Dearden S, McCormack R. EGFR mutation incidence in non-small-cell lung cancer of adenocarcinoma histology: a systematic review and global map by ethnicity (mutMapII). *Am J Cancer Res* [Internet]. 2015;5(9):2892–911. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26609494>
67. Chapman AM, Sun KY, Ruestow P, Cowan DM, Madl AK. Lung cancer mutation profile of EGFR, ALK, and KRAS: Meta-analysis and comparison of never and ever smokers. *Lung Cancer* [Internet]. 2016 Dec 1;102:122–34. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0169500216305098>
68. Bickel P, Diggle P, Fienberg S, Gather U, Olkin I, Zeger S. *The Elements of Statistical Learning* [Internet]. New York, NY: Springer New York; 2009 [cited 2023 Jul 18]. (Springer Series in Statistics). Available from: <http://link.springer.com/10.1007/b94608>
69. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* [Internet]. 2019 Jan 7;25(1):44–56. Available from: <https://www.nature.com/articles/s41591-018-0300-7>