

TESIS DOCTORAL



UCAM

UNIVERSIDAD CATÓLICA
DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO

Programa de Doctorado en Ciencias de la Salud

IDENTIFICACIÓN DE PATRONES ULTRASONOGRÁFICOS EN EPICONDILITIS MEDIANTE MACHINE LEARNING

Autor/a:

D. Guillermo Droppelmann Díaz

Directores/as:

Dr. D. David Prieto Merino

Dr. D. Felipe Feijoo Palacios

Murcia, Julio de 2024

TESIS DOCTORAL



UCAM

UNIVERSIDAD CATÓLICA
DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO

Programa de Doctorado en Ciencias de la Salud

IDENTIFICACIÓN DE PATRONES ULTRASONOGRÁFICOS EN EPICONDILITIS MEDIANTE MACHINE LEARNING

Autor/a:

D. Guillermo Droppelmann Díaz

Directores/as:

Dr. D. David Prieto Merino

Dr. D. Felipe Feijoo Palacios

Murcia, Julio de 2024

AUTORIZACIÓN DEL DIRECTOR DE LA TESIS PARA SU PRESENTACIÓN

El Dr. D. David Prieto Merino y el Dr. D. Felipe Feijoo Palacios como Directores de la Tesis Doctoral titulada “Identificación de patrones ultrasonográficos en epicondilitis mediante machine learning” realizada por D. Guillermo Droppelmann Díaz en el Programa de Doctorado en Ciencias de la Salud, autoriza su presentación a trámite dado que reúne las condiciones necesarias para su defensa.

Lo que firmo, para dar cumplimiento al Real Decreto 99/2011 de 28 de febrero, en Murcia a 3 de marzo de 2024.



Dr. D. David Prieto Merino



Dr. D. Felipe Feijoo Palacios

RESUMEN

Introducción: La tendinopatía lateral del codo (TLC) es una condición musculoesquelética que afecta significativamente los tendones del origen de los extensores de la muñeca. Actualmente, existe una amplia variedad de opciones terapéuticas, cuya eficacia depende de una correcta asignación diagnóstica. La ecografía sigue siendo una de las herramientas diagnósticas más costo-efectivas, pero su alta dependencia del operador impacta en la precisión de la detección y, por ende, en el éxito de las terapias. La inteligencia artificial (IA), a través de algoritmos de *machine learning* (ML), se ha convertido en un aliado estratégico en la clasificación de imágenes médicas, permitiendo identificar hallazgos y patrones no observables para el ojo humano y mejorando significativamente la precisión diagnóstica.

Material y métodos: Se desarrolló un estudio en múltiples etapas. En primer lugar, se realizó un metaanálisis para identificar vacíos diagnósticos en la TLC. En segundo lugar, se llevó a cabo un extenso estudio descriptivo para caracterizar los hallazgos ultrasonográficos más relevantes en pacientes con esta condición. Posteriormente, se diseñó y validó una prueba clínica para detectar un subtipo de TLC. Finalmente, se entrenaron y validaron modelos de ML para identificar, de manera individual y simultánea, distintos tipos de lesiones degenerativas presentes en estos pacientes.

Resultados: El metaanálisis evidenció una alta capacidad de los modelos de IA para el diagnóstico de patologías de la extremidad superior. El estudio descriptivo permitió identificar patrones degenerativos característicos, destacando la tendinosis, la neovascularización, las calcificaciones, las

entesopatías y las roturas intrasustancia. La nueva prueba clínica demostró una alta precisión diagnóstica en pacientes con rotura intrasustancia. Finalmente, los modelos ML mostraron una alta exactitud en el diagnóstico, destacando el modelo de bosques aleatorios en la clasificación binaria y multietiqueta, particularmente en la detección de roturas intrasustancia.

Conclusiones: Este estudio demostró el potencial del ML para mejorar la precisión diagnóstica de la TLC a partir de imágenes ecográficas. La identificación de patrones ultrasonográficos permite avanzar en una estandarización del diagnóstico, facilitando la asignación individualizada de tratamientos. Es necesario integrar herramientas de IA y ML en la práctica clínica para optimizar la detección temprana y mejorar la efectividad terapéutica en estos pacientes.

ABSTRACT

Introduction: Lateral elbow tendinopathy (LET) is a musculoskeletal condition that significantly affects the tendons at the origin of the wrist extensors. Currently, a wide variety of therapeutic options are available, the efficacy of which depends on accurate diagnostic assessment. Ultrasound remains one of the most cost-effective diagnostic tools; however, its high operator dependence affects detection accuracy and, consequently, the success of therapies. Artificial intelligence (AI), particularly through machine learning (ML) algorithms, has emerged as a strategic ally in medical image classification, enabling the identification of findings and patterns not detectable by the human eye and significantly enhancing diagnostic accuracy.

Material and methods: A multi-stage study was conducted. First, a meta-analysis was performed to identify diagnostic gaps in LET. Second, an extensive descriptive study was carried out to characterize the most relevant ultrasound findings in patients with this condition. Subsequently, a clinical test was designed and validated to detect a subtype of LET. Finally, ML models were trained and validated to individually and simultaneously identify different types of degenerative lesions in these patients.

Results: The meta-analysis demonstrated the high capability of AI models in diagnosing upper limb pathologies. The descriptive study enabled the identification of characteristic degenerative patterns, highlighting tendinosis, neovascularization, calcifications, enthesopathies, and

intrasubstance tears. The new clinical test exhibited high diagnostic accuracy in patients with intrasubstance tears. Finally, the ML models achieved high diagnostic accuracy, with the random forest model performing particularly well in binary and multi-label classification, especially in detecting intrasubstance tears.

Conclusions: This study demonstrated the potential of ML to improve the diagnostic accuracy of LET using ultrasound images. The identification of ultrasound patterns enables progress toward the standardization of diagnosis, facilitating the individualized assignment of treatments. Integrating AI and ML tools into clinical practice is essential to optimizing early detection and enhancing therapeutic effectiveness in these patients.

PALABRAS CLAVE

Diagnóstico por imagen; Inteligencia artificial; Radiología; Traumatología; Visión artificial.

KEYWORDS

Artificial intelligence; Artificial vision; Diagnostic Imaging; Radiology; Traumatology.

ÍNDICE GENERAL

I -	INTRODUCCIÓN.....	25
1.1.	ANTECEDENTES	25
1.2.	ETIOLOGÍA.....	28
1.3.	EPIDEMIOLOGÍA	29
1.4.	PATOFISIOLOGÍA.....	30
1.5.	HISTOPATOLOGÍA	32
1.6.	CUADRO CLÍNICO	33
1.7.	ANAMNESIS.....	34
1.8.	EVALUACIÓN FÍSICA	36
1.9.	DIAGNÓSTICO IMAGENOLÓGICO.....	37
1.10.	MACHINE LEARNING	38
1.10.1.	ML e imágenes.....	40
1.10.2.	Preparación e identificación de objetos	40
1.10.3.	Extracción de características	41
1.10.4.	Reducción de dimensiones	42
1.10.5.	Modelos de ML.....	43
1.10.6.	Regresión lineal	43
1.10.7.	Regresión no lineal	44
1.10.8.	Árboles de decisión, <i>Decision tree learning</i> (DTL).....	45
1.10.9.	Bosques aleatorios, <i>Random Forest</i> , (RF).....	45
1.10.10.	Máquinas de Soporte Vectorial, <i>Support Vector Machine</i> (SVM).....	45
1.10.11.	Redes neuronales, <i>Neural Networks</i> (NN) y Aprendizaje Profundo, <i>Deep Learning</i> , (DL)	46
1.11.	ENTRENAMIENTO Y VALIDACIÓN: APRENDIZAJE	46

ÍNDICE GENERAL	13
1.12. EVALUACIÓN DE RESULTADOS.....	47
II - JUSTIFICACIÓN.....	51
III - OBJETIVOS	57
3.1. OBJETIVO GENERAL.....	57
3.2. OBJETIVOS ESPECÍFICOS.....	57
3.3. OBJETIVOS OPERACIONALES	57
IV - MATERIAL Y MÉTODO.....	61
4.1. ESTUDIO N1	61
4.1.1. Reporte	61
4.1.2. Pregunta de investigación	61
4.1.3. Estrategia de búsqueda y fuentes de información.....	61
4.1.4. Criterios de selección	62
4.1.5. Extracción de datos	62
4.1.6. Consideraciones éticas	63
4.1.7. Evaluación de riesgo de sesgos	63
4.1.8. Análisis estadísticos	63
4.1.9. Análisis univariados	63
4.1.10. Análisis bivariado.....	64
4.1.11. Análisis de heterogeneidad	64
4.1.12. Paquetes y reportes	65
4.2. ESTUDIO N2	66
4.2.1. Consideraciones éticas	66
4.2.2. Diseño del estudio.....	66
4.2.3. Clasificación de los hallazgos ultrasonográficos.....	67
4.2.4. Análisis estadísticos	67
4.3. ESTUDIO N3	69
4.3.1. Consideraciones éticas	69

4.3.2.	Diseño de estudio.....	69
4.3.3.	Participantes	69
4.3.4.	Tamaño de muestra.....	70
4.3.5.	Instrumentación	70
4.3.6.	Maniobra física	71
4.3.7.	Análisis estadísticos	71
4.4.	ESTUDIO N4	73
4.4.1.	Consideraciones éticas	73
4.4.2.	Diseño de estudio.....	73
4.4.3.	Participantes	73
4.4.4.	Instrumentación	74
4.4.5.	Construcción de base de datos	75
4.4.6.	Análisis estadísticos y ML	77
V -	RESULTADOS.....	81
5.1.	ESTUDIO N1	81
5.1.1.	Resultados de la búsqueda	81
5.1.2.	Características de los estudios.....	82
5.1.3.	Riesgo de sesgo.....	84
5.1.4.	Análisis univariados	85
5.1.5.	Análisis bivariados.....	88
5.2.	ESTUDIO N2	91
5.2.1.	Diseño del estudio.....	91
5.2.2.	Relación entre rotura el tendón y hallazgos degenerativos	92
5.3.	ESTUDIO N3	97
5.3.1.	Participantes	97
5.3.2.	Resultados de la prueba.....	99
5.4.	ESTUDIO N4	100

ÍNDICE GENERAL	15
5.4.1. Tendinopatía del extensor común.....	100
5.4.2. Modelos de ML como clasificador binario	100
5.4.3. Modelos de ML como clasificador multietiqueta	105
5.4.4. Rendimiento diagnóstico	108
6. DISCUSIÓN	111
6.1. ESTUDIO N1	111
6.2. ESTUDIO N2	115
6.3. ESTUDIO N3	118
6.4. ESTUDIO N4	121
6.5. Desafíos y oportunidades para investigaciones futuras de modelos de IA en salud musculoesquelética.....	124
7. CONCLUSIONES	137
8. REFERENCIAS BIBLIOGRÁFICAS	140

SIGLAS Y ABREVIATURAS

AC, Accuracy

C, Cálcula

CECA, Comité de Ética Científico Adulto

CNN, Convolutional neural networks

DASH, Disabilities of the Arm, Shoulder, and Hand

DICOM, Digital Imaging and Communication in Medicine

DL, Deep learning

DTL, Decision tree learning

E, Entesopatía

ERCC, Extensor radial corto del carpo

ERLC, Extensor radial largo del carpo

ERN, Entesopatía rotuliana del nódulo

EVA, Escala visual análoga

FN, Falso negativo

FP, Falso positivo

HV, Hipervascular

IA, Inteligencia artificial

IC, Intervalo de confianza

ML, Machine learning

NN, Neural networks

NV, Neovascular

PACS, Sistema de archivo y comunicación de imágenes

PCA, Análisis de componentes principales

PR, Precisión

PRP, Plasma rico en plaquetas

PRTEE, Patient-Rated Tennis Elbow Evaluation

PRWE, Patient-Rated Wrist Evaluation

RIS, Rotura intrasustancia

RF, Random forest

ROC, Característica operativa del receptor

RMN, Resonancia magnética nuclear

ROI, Region of interest

RL, Reinforcement learning

SF-36, Short Form -36

SL, Supervised learning

SN, Sensibilidad

SROC, Característica operativa del receptor resumen

SVM, Support vector machine

SSME, Servicio de Salud Metropolitano Oriente

TLC, Tendinopatía lateral del codo

US, Ultrasonido

USL, Unsupervised learning

VP, Verdadero positivo

VN, Verdadero negativo

ÍNDICE DE FIGURAS, DE TABLAS Y DE ANEXOS

ÍNDICE DE FIGURAS

Figura 1. Flujo de los participantes.....	74
Figura 2. Flujo de selección de datos y sujetos.....	76
Figura 3. Diagrama de flujo PRISMA 2020	82
Figura 4. Evaluación de sesgos de los estudios seleccionados.	84
Figura 5. Evaluación de calidad de los estudios seleccionados.....	85
Figura 6. Representación gráfica del bosque de sensibilidad agrupada.....	86
Figura 7. Representación gráfica del bosque de especificidad agrupada.	87
Figura 8. Representación de las tasas de odds diagnósticas logarítmicas.	88
Figura 9. Representación de bosque de las tasas de odds.....	89
Figura 10. Curva SROC para la precisión de la prueba diagnóstica.	90
Figura 11. Relación largo de la RIS(mm) e hipoecogeneidad.	93
Figura 12. Relación ancho de la RIS(mm) y la neovascularización.	94
Figura 13. Relación profundidad de la RIS(mm) y la hipoecogeneidad.....	94
Figura 14. Diagrama de flujo STARD.	97
Figura 15. Curva ROC para RF y dataset A.	108

ÍNDICE DE TABLAS

Tabla 1. Resumen de las técnicas de extracción de características.	42
Tabla 2. Categorías de Machine Learning y técnicas de análisis.	44
Tabla 3. Matriz de confusión binaria y métricas de evaluación de rendimiento. ..	48
Tabla 4. Características de los estudios incluidos.	83
Tabla 5. Características demográficas y hallazgos ecográficos degenerativos.	92
Tabla 6. Coeficientes estimados para la RIS por dimensión según variables.	95
Tabla 7. Características demográficas y hallazgos.	98
Tabla 8. Mediciones diagnósticas de las pruebas.	99
Tabla 9. Hallazgos ultrasonográficos según sexo.	100
Tabla 10. Rendimiento según clasificadores y dataset	102
Tabla 11. Rendimiento multietiqueta según modelo.	106

I – INTRODUCCIÓN

I - INTRODUCCIÓN

1.1. ANTECEDENTES

El aparato locomotor es el sistema encargado de proteger órganos vitales internos, pero específicamente de liderar la ejecución coordinada y balanceada de cada movimiento que realizan los seres humanos desde los más globales como caminar, hasta los más específicos como tocar un instrumento musical. Su complejidad radica no sólo en el desarrollo de sus extensas unidades especializadas, sino que también mediante su extraordinaria integración guiada por el complejo sistema nervioso, pudiendo responder instantáneamente ante infinitos estímulos extrínsecos e intrínsecos, donde la disposición de sus elementos refleja la evolución de miles de años adquirida por nuestra especie (1) .

Para poder preservar cada una de sus funciones, resulta fundamental mantener una correcta salud musculoesquelética donde debe primar el equilibrio entre ejercicios cardiovasculares, de fuerza, y estiramientos combinados con una balanceada alimentación, estilos de vida y calidad en el descanso que potencien una sinergia permanente entre demanda y reparación (2) (3). En este sentido, cuando alguno de estos elementos se encuentra comprometido, se podría desarrollar una afección musculoesquelética, las cuales se pueden presentar de forma súbita como una fractura, o progresivas en el tiempo como una tendinopatía producto de la acumulación de exigencias físicas en un sistema inestable. A su vez, se destaca que estas problemáticas pueden impactar al sistema de forma sistémica involucrando a más de un segmento corporal como las extremidades o columna vertebral, o de forma más específica como huesos, músculos, articulaciones, ligamentos y tendones (4).

Dentro de las condiciones musculoesqueléticas más frecuentes que se afectan a la extremidad superior se encuentran las tendinopatías. Entre ellas, destaca por su alta prevalencia la afección del tendón del origen de los extensores de muñeca debido a que participa de forma integrada en la mayoría de las actividades que determinan la función y el control del complejo codo-muñeca-mano (5).

Actualmente se acepta con el nombre de epicondilitis, sin embargo, su denominación ha ido variando en el tiempo. Se pueden encontrar diferentes

reportes en la literatura científica haciendo alusión a su nomenclatura asociado específicamente al tipo de población que afecta. Por ejemplo, a principios del siglo XIX fue descrita como patología exclusiva de escritores, vinculándosela fuertemente con este grupo de individuos. También se le ha reportado como una afección en jardineros, trabajadores manuales y en sujetos que realizan actividades deportivas que utilizan raquetas adquiriendo el nombre de codo del tenista, donde todas tiene como factor común ser actividades repetitivas con una demanda moderada de la extremidad superior distal más la realización de una prensión manual controlada (6).

A su vez, debido a su fisiopatogenia, se la denominó inicialmente y por mucho tiempo como tendinitis adquiriendo el rótulo de epicondilitis, término ampliamente aceptado por los clínicos especialistas en esta patología. Sin embargo, diferentes estudios han demostrado que no es precisamente una condición inflamatoria ya que múltiples cortes histológicos en pacientes con esta condición no fueron capaces de confirmar la presencia de células inflamatorias. Por este motivo, diversos autores respaldaban fuertemente la tesis que se trataba de un proceso más bien degenerativo, es decir de un desgaste progresivo e irreversible de los elementos celulares y biológicos presentes en el tendón, impactando severamente en la capacidad funcional de la estructura, adquiriendo el nombre de tendinosis o tendinopatía lateral del codo (TLC) de esta forma se involucra no sólo su estado histológico, sino que también su ubicación en la región del codo (7).

Al igual que en otras tendinopatías, actualmente existen más de cien opciones terapéuticas tanto médico-quirúrgicas, farmacológicas y conservadoras cuyos efectos disímiles no favorecen una estandarización para poder hacer frente a esta condición musculoesquelética. Esto se debe a que a pesar de contar con múltiples alternativas terapéuticas la efectividad reportada en distintos metaanálisis es bastante similar, concluyendo en la gran mayoría de los casos que falta evidencia suficiente para poder respaldar o rechazar sus usos (8). En este escenario, es factible hipotetizar que los esfuerzos por encontrar alternativas de tratamientos han sido suficientes, y que la problemática se debiera centrar con mayor atención en la capacidad diagnóstica, entendiendo que, al obtener una mayor precisión diagnóstica, mejor debiera ser la asignación de una opción terapéutica específica. En este sentido, las opciones diagnósticas disponibles tales como la examinación física, siendo bastante simple incluso para los profesionales no especialistas, no

reconoce una diferenciación de los hallazgos degenerativos de forma individual que podrían estar presentes en un individuo con TLC (9). Para mejorar la precisión en el diagnóstico y pesquisar estos hallazgos se debe complementar la exploración diagnóstica con exámenes imagenológicos como la resonancia magnética nuclear (RMN), pero su alto valor económico, exposición a radiación y la falta de acceso de la población lo posiciona como un limitante (10).

Una opción diagnóstica altamente costo efectiva es la ecografía de partes blandas, resultando en un examen de elección por la mayoría de los clínicos y pacientes. No obstante, no siempre se solicita este examen ya que la sola presencia de dolor en la región lateral del codo asociado a actividades repetitivas resulta suficiente, pero altamente inespecífico para detectar la problemática. Otro aspecto por destacar es que a pesar de que el examen ecográfico es capaz de discriminar cada uno de los hallazgos degenerativos que pudiera estar presentes en un paciente que padece esta condición, éstos han sido descritos sólo conceptualmente y no se han otorgado frecuencias en su presentación, así como tampoco una asociación entre ellos, resultando de vital importancia si se quisiera asignar correctamente un tratamiento específico (11). Otro elemento que resulta clave en el diagnóstico ecográfico es que éste se caracteriza por ser altamente operador dependiente, por lo tanto, el nivel de entrenamiento y experiencia del profesional resultan cruciales, condicionando la precisión de los resultados informados, registrándose altas variabilidades interoperadores, sino que también intraoperadores (12). A su vez, la escasez de radiólogos ecografistas especialistas en patologías musculoesqueléticas en diferentes servicios de salud, dificulta aún más la disponibilidad de acceso de la población a contar con la disponibilidad de este tipo de diagnósticos (13).

Recientemente, la incorporación de diferentes herramientas provenientes desde la inteligencia artificial (IA) para el diagnóstico de imágenes médicas, han permitido contar con un elemento adicional para mejorar la exactitud de los resultados y consecuentemente la toma de decisiones del equipo de salud. Sus ventajas son múltiples, y dependiendo del tipo de algoritmo que se implemente y de la cantidad de imágenes que se analicen sólo se necesitará capacidad computacional para obtener información sobre la patología de interés. Además, una vez entrenado el modelo, no es necesario contar con más profesionales para obtener la información, pudiendo automatizar un proceso en el apoyo diagnóstico (14). No obstante, la mayor parte de la literatura publicada hasta el momento

relacionada con el uso de IA y patologías musculoesqueléticas se relaciona con el uso de imágenes estandarizadas secuencialmente tales como las radiografías, escáner, y RMN, siendo prácticamente inexistentes experiencias con ecografías debido a que el operador determina el corte y serie de imágenes que le permite obtener su impresión diagnóstica. Por otro lado, las publicaciones reportadas para el diagnóstico automatizando se caracterizan por la dicotomización de los resultados o presentan múltiples debilidades (15) (16). Además, entendiendo que en la TLC los hallazgos ecográficos pueden estar presentes más de uno simultáneamente, resulta completamente necesario que el escenario descrito sea particularmente interesante de estudiar en la presente tesis.

1.2. ETIOLOGÍA

La TLC es uno de los desórdenes musculoesqueléticos más frecuentes de la extremidad superior, y al igual que la mayoría de las tendinopatías, describen un espectro de cambios estructurales al interior del tendón, provocando en ocasiones dolor, impotencia funcional y reduciendo drásticamente la función de los tejidos comprometidos y circundantes (17). En particular, se caracteriza por afectar la inserción del grupo de tendones que componen el origen de los extensores de muñeca ubicados en el epicóndilo en la región lateral del codo. Debido a la conformación biomecánica de los músculos extensores que lo componen y más el desbalance anatómico existente producto de una longitud desigual, resulta en un torque ineficiente del complejo de extensores que sumado a demandas repetitivas de la mano y la muñeca pueden desencadenar un cuadro de tendinopatía (18).

Adicionalmente, diversos estudios electromiográficos han demostrado la incapacidad de contracción coordinada de cada uno de los músculos una vez presentada esta afección. En otras palabras, se presenta una elevada actividad eléctrica que en condiciones basales se encuentra muy superior en comparación con otros músculos respondiendo exageradamente a pequeños estímulos facilitando un mecanismo de lesión temprano (19). A su vez, el músculo que más se encuentra comprometido es el extensor radial corto del carpo (ERCC), seguido del extensor radial largo del carpo (ERLC), y en casos más severos pudiendo llegar a afectar el aparato extensor en su totalidad (20). Se destaca que estos músculos presentan capacidad retráctil y gran capacidad de contracción, sin embargo, los tendones

donde finalizan no, ya que se caracterizan por ser más bien rígidos sobre todo desde los 35 a 40 años y dependiendo de las actividades que realice el individuo pudiendo incluso condicionar la aparición de una TLC a temprana edad (21).

1.3. EPIDEMIOLOGÍA

La prevalencia e incidencia de la TLC ha sido estudiada por casi 50 años en diferentes latitudes del mundo y por múltiples autores debido a que constituye una de las condiciones musculoesqueléticas que más se repiten en la extremidad superior. No obstante, no resulta fácil determinar un valor absoluto con precisión acerca de su presentación en la población general debido a que no es una condición exclusiva de un grupo específico, sino que la padecen transversalmente tanto mujeres como hombres y en diferentes espectros de edad por lo que pesquisarla resulta aún más complejo (22).

En general, con respecto a su distribución por sexo, algunos estudios han determinado que existen diferencias menores entre hombres y mujeres. No obstante, otras publicaciones respaldan que las diferencias son prácticamente mínimas y que no determinan en absoluto su presentación a la presencia de algún factor hormonal o anatómico propio del género (23). Esto es respaldado con los datos obtenidos y analizados en la presente tesis donde descriptivamente se presentó la distribución tras analizar más de 4000 pacientes con TLC. Por su parte, la edad resulta un factor de análisis muy interesante, ya que al ser una condición descrita como degenerativa el lector pensaría inmediatamente que se relaciona con una población más envejecida. Sin embargo, su rango de presentación se encuentra fuertemente entre el intervalo que comprende los 35 a 55 años, lo cual no excluye que puedan existir casos aislados en los extremos de menor y mayor edad respectivamente. La edad de inicio tiene directa relación con el detrimento que experimenta el tendón alrededor de esa edad, resultando más susceptible a padecer hallazgos degenerativos e impactar directamente en su función (23) (24).

Dentro de los grupos de sujetos que más han reportado una alta prevalencia de TLC se encuentran aquellos relacionados con el desarrollo de actividades manufactureras, artesanos, o cualquier tipo de oficio que involucre una actividad de alta demanda de las manos y muñecas, las cuales a través de una repetición permanente en periodos de tiempo por sobre las cuatro horas generan una

modificación en el ambiente celular generando una modificación de los tejidos y dando inicio a un proceso degenerativo (25).

Tradicionalmente se ha asociado la TLC a deportistas que practican actividades de alta demanda de la prensión manual como las que se presentan en deportes como el tenis, pádel tenis, ráquetbol, squash, entre otras (26). No obstante, la prevalencia en tenistas de élite que la presentan es bastante menor que lo que se puede encontrar en población (27). Esto en parte puede ser explicado porque el deportista de alto rendimiento mantiene un plan balanceado de actividades que le permiten adaptarse con facilidad a la gran demanda deportiva que realiza. No obstante, esta relación cambia cuando se trata de torneos internacionales de alta competición cuando la exigencia del segmento se ve altamente comprometida (27).

1.4. PATOFISIOLOGÍA

Históricamente se ha expuesto que el mecanismo de lesión de la TLC es una sobrecarga, es decir, una actividad repetitiva cuyo origen puede ser laboral, deportiva o cotidiana como el uso del computador, resultando en la sumatoria de estímulos con cargas pequeñas mantenidas en el tiempo o una mayor demanda de peso en un corto periodo (25). Además, si se considera una incorrecta ejecución o deficiencia energética del sistema demandado resultaría en una aceleración de la aparición de esta condición. Esta concepción se ha mantenido debido a que se ha aceptado ampliamente que la función del tendón es la transmisión de fuerzas dentro del sistema musculoesquelético. Esto no es del todo cierto, ya que, si realmente transmitiera la fuerza, impactaría en los huesos ya sea como micro impactos pudiendo generar una fractura por estrés en el caso de los tendones inmaduros o una fractura por avulsión en un tendón más maduro. Por este motivo, la verdadera función de un tendón es desacelerar los módulos de fuerza que son generados por el músculo, esto se puede realizar debido a que los tendones poseen cierta capacidad limitada de almacenamiento energético, generándose un equilibrio de tensión-estrés (28). Cuando este sistema de balance energético se ve interrumpido, se inician de inmediato procesos celulares reparativos que, si no se permite el tiempo adecuado de recuperación, se comienzan a generar microtraumas acumulativos pudiendo conducir una degeneración progresiva e irreversible (29).

Este mismo conjunto de cargas que son capaces de producir daño en el tendón cuando el ambiente es desequilibrado si se hace de forma correcta, progresiva y balanceada favorecen el proceso reparativo del tejido. Es así como la mayoría de los planes de rehabilitación propuestos se basan específicamente en reconocer aquellos patrones de cargas que conducen a una reparación de los tejidos dañados. El ejemplo más clásico y la propuesta más aceptada es la realizada por Julius Wolff hace más de 100 años atrás la cual fue denominada como mecanotransducción, en donde se proponía que las cargas mecánicas producen en los tejidos una remodelación propia (30). Específicamente, se explicaba cómo el tejido óseo era capaz de modelarse de acuerdo con la duración, magnitud y tasa de carga. En el caso del tendón, este rol es asumido en gran parte por la matriz extracelular, la cual permite la transmisión de señales mecánicas a las células denominadas tenocitos, unidades estructurales del tendón, impulsando las respuestas al ejercicio o en casos de lesión (31).

El uso excesivo de este tejido puede finalizar en una carga mecánica ineficaz alterando las interacciones de la matriz extracelular incitando a una subestimulación mecanobiológica, apoptosis celular y reparación ineficaz (32). Finalmente, cuando la situación es irreversible, se termina degenerando el tendón y subsecuentemente provocando una microrrotura en el tendón, estado que debe ser considerado con particular atención debido a las dificultades que tiene el tejido de revertir el proceso, alterando no sólo su función, sino que impactando en su entorno (33).

En general se acepta la existencia de tres potenciales modelos que explican hipotéticamente la patofisiología del tendón. La primera se centra en el colágeno, el cual al presentar un pequeño desgarro podría explicar la debilidad del tendón debido a que es un elemento de vital importancia en la conformación estructural del tendón. A su vez, la falta de transmisión de carga a través de las fibras de colágeno desgarradas estimularía insuficientemente las células del tendón, contribuyendo a la patología degenerativa. No obstante, ha sido fuertemente cuestionada debido a que *in vivo* no ha podido ser posible demostrar esta condición (34).

En segundo lugar, se ha propuesto un modelo inflamatorio, el cual es altamente complejo ya que no sigue un patrón inflamatorio tradicional. La respuesta de las células tendinosas detecta cambios en la carga y el entorno

bioquímico, lo que desencadena múltiples respuestas y adaptaciones en el tendón. Aunque se han observado células inflamatorias en tendones patológicos, no parece ser una respuesta inflamación normal. Se han detectado aumentos en citocinas inflamatorias en la tendinopatía por uso excesivo, pero la presencia de estas sustancias no respalda necesariamente la afirmación de que la inflamación es el evento principal en la patología del tendón, resultando en una alteración en la síntesis y degradación del tendón (35).

Finalmente, se presenta el modelo de respuesta de células tendinosas, donde el tenocito es el principal responsable de mantener la matriz extracelular en respuesta a su entorno, éstos detectan los cambios en la carga y en el entorno bioquímico y responde activándose, expresando proteoglicanos, así como cambios en los tipos de colágeno. Por lo tanto, esta sinergia de estimulación entre los cambios en la carga del tendón y el entorno bioquímico al ser detectados por las células propias del tendón permitirían una cascada de respuestas (activación celular, expresión de proteoglicanos y cambios en el tipo de colágeno) (36).

1.5. HISTOPATOLOGÍA

Desde su descripción inicial se mantuvo siempre la concepción de que la TLC era una afección musculoesquelética inflamatoria, de ahí su denominación con el uso del sufijo “itis” que denota la existencia de inflamación (epicondilitis). No obstante, se ha demostrado en diferentes estudios histológicos que no existe presencia de células o elementos característicos de una inflamación por lo que esta denominación a pesar de ser ampliamente aceptada por los clínicos debiera de estar cada vez más en desuso (37).

Los elementos celulares que la caracterizan son propios de un estado degenerativo, es decir de una tendinosis, la cual desde el punto de vista histológico se caracteriza por presentar anomalías tempranamente defectos en la celularidad, hiperplasia vascular, disposición del colágeno y aumento en la actividad metabólica con alta presencia de fibroblastos sin diferenciación (38).

Actualmente se aceptan cuatro estadios de un estado degenerativo. En el estado 1, el patrón adquiere una conformación producto de la disposición ondulada del colágeno, existiendo un aumento en la producción de colágeno tipo

3. En el estado 2, hay presencia de tendinosis e hiperplasia angiofibroblástica, es decir, un crecimiento de los vasos sanguíneos con mayor desorganización de las fibras de colágeno, e hiperplasia celular. En el estado 3 de la tendinosis, se presenta la apoptosis o muerte celular programada, provocando el agotamiento de las células tendinosas funcionales y la descomposición del colágeno y la matriz extracelular. Finalmente, el estado 4 se presenta con disrupción estructural grave y falla mecánica. Se destaca que todos estos cambios celulares e histológicos impactan directamente en un mal funcionamiento biomecánico del segmento y de las articulaciones subsecuentes (17).

Se destaca que estos hallazgos degenerativos descritos por estadios y que son visibles al microscopio en la medida que van aumentando son perfectamente reconocibles en una imagen médica como una RMN o ecografía, las cuales serán descritas con mayor detalle en la sección diagnóstico imagenológico (39).

1.6. CUADRO CLÍNICO

Distintos signos y síntomas son patognomónicos de la lesión. Sin embargo, su presentación puede variar dependiendo de las características propias de los sujetos y de acuerdo con la demanda de actividad que realice con el segmento comprometido. En particular, es habitual una presentación de dolor extenso en la cara lateral del codo, caracterizada por una larga data de tiempo, existiendo incluso reportes en la literatura por hasta 24 meses, cuya resolución aparentemente es espontánea. A su vez, los aspectos funcionales parecieran caracterizarse según etapas, presentándose inicialmente una marcada impotencia funcional que en ocasiones es severa, seguida de una adaptación funcional propia de modificaciones histopatológicas con hallazgos ultrasonográficos propios (21).

En general, no existen diferencias significativas entre hombres y mujeres, y el periodo de presentación habitualmente suele ser en la edad media de las personas, vinculado fuertemente a actividades de alta demanda de flexo-extensión del complejo mano muñeca y que se requiera de la realización de una fuerza prensil. Específicamente, el codo se encontrará en una cadena cinética abierta, sin apoyo, sumado a un gesto de prensión de la mano con un movimiento de extensión de falanges, carpo y articulación radioulnar (40). Por este motivo, es altamente

recomendable que uno de los aspectos fundamentales en la evaluación del cuadro clínico sea incluir elementos de gestos deportivos como las actividades laborales que realicen los pacientes para poder determinar con mayor precisión el factor de riesgo. Se estima que cargas menores repetitivas en periodos cercanos a las 4 horas generan un impacto negativo en la capacidad de absorber y desacelerar cargas en el tendón pudiendo iniciar rápidamente un proceso degenerativo (41).

Actualmente, se ha estudiado la relación existente entre las tendinopatías y la presencia de factores intrínsecos como son las enfermedades crónicas no transmisibles debido a que una gran mayoría de los pacientes relata tener presente al menos una. Esto no queda claro si es por la edad de presentación se relaciona con una mayor probabilidad de presentar alguna de ellas o por qué realmente se encuentren vinculadas. Enfermedades tales como la diabetes, hipertensión, hipercolesterolemias, y alteraciones en la tiroides se encuentran entre las más frecuentes. Algunas hipótesis sostienen que los cambios se producen en el ambiente celular del tendón, específicamente en la modificación del pH, alterando fuertemente el intercambio gaseoso de los tenocitos, desencadenando de esta manera una apoptosis celular temprana (42) (43).

También se han descrito antecedentes de patologías musculoesqueléticas como tendinopatías, tenosinovitis, o neuropatías por atrapamiento, seguramente se presenta una predisposición genética en la calidad de los tejidos que determine la aparición de estas y otras enfermedades relacionadas con el aparato locomotor. Adicionalmente, los hábitos y la calidad de vida de los individuos presentan relación con la aparición de la TLC. Por ejemplo, consumo de alcohol, tabaquismo, tratamientos extensos con corticoesteroides orales, inyecciones previas con corticoides, uso de esteroides, terapias hormonales, quimioterapia y radioterapia (44).

1.7. ANAMNESIS

Antes de evaluar el segmento específico, es importante realizar una anamnesis completa para identificar múltiples elementos que ayuden a determinar la causa y el mecanismo de la lesión de la TLC. Por lo general, se recomienda una evaluación completa de la extremidad superior, buscando desbalances, desde el hombro hasta restricciones en los dígitos. Es habitual encontrar alteraciones

capsulares en el hombro y activaciones musculares ineficientes en la región escapulotorácica (45).

El relato del paciente generalmente incluirá actividades que implican el uso repetitivo del antebrazo, como tejer, escribir en el computador, podar o utilizar herramientas que requieran fuerza de agarre o precisión. Esta condición es común en personas que realizan trabajos de escritorio sin soporte adecuado para los antebrazos. También es relevante investigar la participación en actividades deportivas, especialmente deportes que involucren movimientos de agarre, como aquellos que utilizan diversos tipos de raquetas (46).

En relación con este aspecto específico, es importante que el evaluador tenga en cuenta que ciertos elementos relacionados con la raqueta pueden desencadenar la TLC, además de la sobrecarga causada por jugar partidos consecutivos o prolongados. Entre los factores que aumentan la predisposición a desarrollar epicondilitis se encuentran; el uso de raquetas de marcas rígidas, por lo que se sugiere optar por marcas más flexibles. Es fundamental que la raqueta se encuentre equilibrada, es decir, sin una carga excesiva hacia el mango o la cabeza. El peso de la raqueta no debe exceder los 300 gramos y la tensión de las cuerdas debe ser inferior a 24 kilos. Por último, es importante que el mango se adapte correctamente al tamaño del agarre, lo cual se puede determinar dejando aproximadamente un dedo de espacio entre los cuatro dedos y la base de la mano al cerrarla alrededor de la raqueta. Todas estas sugerencias no sólo disminuyen la probabilidad de presentar un evento, sino que prolongan la vida útil de los deportistas (47).

Otro aspecto que debe ser considerado en una anamnesis y como parte de toda evaluación es que aplicar un instrumento que permita evaluar la funcionalidad del paciente. Habitualmente, se utiliza la conocida Escala Visual Análoga (EVA) para tener una aproximación cualitativa de la percepción del dolor que tiene el individuo. Es importante preguntar si este dolor varía durante el día o la noche, durante las actividades o después de éstas, de esta forma, el clínico podrá tener una perspectiva más amplia de la etapa que podría estar la TLC (48) .

A su vez, es de vital importancia contar con la aplicación de un cuestionario, cuyo objetivo sea cuantificar lo más objetivamente posible el estado actual de paciente, así como valorizar las diferentes actividades que realiza en su vida cotidiana. Para este propósito existe una serie de instrumentos que permiten dar respuesta a esta inquietud tales como; el *Short Form*, (SF-36); *Disabilities of the Arm*,

Shoulder and Hand, (DASH); *Quick DASH*, y el *Patient-Rated Tennis Elbow Evaluation*, (PRWE), que son más genéricos. Por este motivo se recomienda el uso de cuestionarios más específicos para evaluar la función del codo tales como el *Patient-Rated Tennis Elbow Evaluation*, (PRTEE) (49).

1.8. EVALUACIÓN FÍSICA

Después de analizar las actividades cotidianas y deportivas del paciente, es recomendable realizar una evaluación exploratoria exhaustiva segmento por segmento de la extremidad superior. Se deben identificar todo tipo de anomalías en la piel, como cicatrices, quemaduras u otros hallazgos que puedan limitar la movilidad de esa área. Además, se recomienda realizar una evaluación del tejido miofascial en busca de restricciones (50).

La zona de mayor dolor será siempre cercana al origen del extensor común de la muñeca, específicamente localizado en la región del epicóndilo lateral del codo y distalmente hacia el antebrazo (51). Sin embargo, se debe identificar cualquier fuente importante de síntomas que no necesariamente provenga del codo. Por este motivo se debe seguir con una evaluación minuciosa de la movilidad pasiva desde el cuello, hombro, codo, muñeca hasta los dedos, teniendo especial cuidado en cada articulación. Se deberá proceder con una evaluación activa no asistida de los segmentos mencionados y luego terminar con una evaluación de fuerza de los segmentos realizados por un terapeuta para determinar la presencia de desbalances o alteraciones en el movimiento (52).

La examinación clínica podrá ser complementada con la realización de diversas pruebas diagnósticas que permitan reconocer con mayor precisión el origen de los síntomas. Estas maniobras físicas se han propuesto para pesquisar la TLC, las cuales permiten discriminar con diferentes grados de sensibilidad si la estructura dañada es el tendón o algún tejido adyacente como causa de los signos y síntomas. Sin embargo, estas pruebas se consideran positivas, es decir detectan la alteración, debido a que provocan una exacerbación de los síntomas del paciente al aplicar estrés sobre la estructura comprometida. Por esta razón, su uso debe realizarse mediante la acción de un profesional con precaución y de manera criteriosa, ya que podrían causar un mayor daño en el tejido afectado (53).

Dentro de las más utilizadas en la práctica clínica y en investigación debido a su validez diagnóstica se encuentran las pruebas de Cozen, Mill y Maudsley (53). En general, todas estas técnicas presentan un resultado positivo cuando se gatilla dolor en la región epicondilar del codo (45). No obstante, desde el punto de vista clínico, esta aproximación suele ser poco precisa ya que no considera los hallazgos que pudieran estar presentes en imágenes, centrándose únicamente en los síntomas que pueden variar entre pacientes, lo que resulta en incertidumbre en cuanto al enfoque terapéutico. Además, existen factores que pueden impactar en la sensibilidad y especificidad de las pruebas debido a la experiencia que posea el evaluador, así como el relato propio del paciente. Por esta razón, a pesar de tener una fuerte sospecha de que el diagnóstico pueda ser una TLC, es altamente recomendable complementar toda exploración física con un examen imagenológico realizado un equipo de especialistas en radiología musculoesquelética. Esto permitirá obtener una evaluación más completa y precisa para designar un tratamiento más preciso para los pacientes que padezcan esta condición (54).

1.9. DIAGNÓSTICO IMAGENOLÓGICO

Múltiples son las opciones disponibles para diagnosticar la TLC mediante el uso de imágenes. La RNM, es sin lugar a duda el referente diagnóstico o también llamado *gold standard*, ya que es el método más efectivo para el diagnóstico de la TLC, es decir, el que posee mayor sensibilidad y especificidad diagnóstica. No obstante, teniendo en cuenta su alto costo, la alta demanda del equipo, la incomodidad durante su realización, el periodo de tiempo extenso en la obtención de los diferentes cortes imagenológicos, contar con la disponibilidad de profesionales expertos a la lectura de estas imágenes, y la exposición a pequeñas dosis de radiación, resulta no ser el método más costo efectivo para el diagnóstico de esta patología ni el de primera elección por parte de los clínicos (55).

En este sentido, el ultrasonido es menos costoso y más accesible económicamente que la RNM, como también la ausencia de contraindicaciones. Diversos estudios han demostrado una mayor sensibilidad y especificidad en el ultrasonido para este tipo de lesiones, por lo que se considera ser la herramienta más útil para determinar el grado de daño del tendón en esta patología (56) (57).

De esta manera, las escalas de clasificación de la epicondilitis mediante ultrasonido pueden ser confiables para evaluar la anormalidad de los tendones afectados (58). Se destaca la propuesta diagnóstica mediante la diferenciación de patrones ultrasonográficos que permite subdividir la del origen de los extensores, en cuatro grandes tipos: Cálctica, C; Entesopatía, E; Rotura intrasustancia, RIS, e Hipervascular, HV, todas presentan como base la disrupción del patrón fibrilar característico de procesos degenerativos. Estos patrones ultrasonográficos se pueden presentar de manera aislada o en forma combinada, por lo que el tratamiento debiera también ser diferenciado según el tipo de lesión (59).

Ahora la pregunta que se puede hacer un clínico es ¿Qué pasa si no tengo acceso a una ecografía?, ¿Es igualmente válida esta diferenciación? La experiencia adquirida a lo largo de estos años en el manejo de esta patología ha permitido establecer algunas relaciones para poder diferenciar los patrones sin el uso de un equipo ultrasonográfico, pero debe quedar claro que es sólo orientativo y no reemplaza el juicio y diagnóstico del médico ni el análisis del radiólogo. Clínicamente se pueden reconocer los patrones mencionados de la siguiente forma: Si es una lesión hipervascular generalmente se identifica al examen físico con la presencia de una zona con aumento de temperatura y un ligero aumento de volumen en relación con la extremidad contralateral. El patrón cálcico generalmente se puede identificar porque el paciente relata presentar una sensación de pellizcamiento al realizar flexo-extensión de la extremidad, las calcificaciones que se encuentran en la proximidad del hueso, es decir que son calcificaciones entesopáticas se reconocen por presentar un dolor insercional. Las epicondilitis con predominio entesopático son las más sencillas de examinar ya que presentan principalmente un dolor en la inserción en el epicóndilo. Finalmente, las lesiones con un patrón de rotura en la mayoría de los casos se destaca la presencia de dolor nocturno donde los sujetos relatan muchas veces despertar por este hallazgo (60).

1.10. MACHINE LEARNING

Un campo que ha liderado avances en la mejora de la precisión de sus procedimientos y los resultados de sus operadores es la Radiología y la Imagenología. Esto se debe a que estas prácticas abarcan un amplio rango de

resultados finales, y los diagnósticos dependen de diversos factores, como el tipo de imagen reportada, la calidad y resolución de la imagen capturada, el nivel de entrenamiento y especialización del médico que realiza el informe y del personal de apoyo en la unidad de análisis, el tiempo disponible para la ejecución e interpretación de la tarea, la gravedad o complejidad de las lesiones informadas, entre otros. Por lo tanto, todos estos elementos podrían influir en la conclusión diagnóstica debido a la generación de una alta variabilidad intra e interobservador (61).

Una estrategia de apoyo diagnóstico que contribuye a mejorar la precisión de los resultados es la implementación de modelos de IA, en particular, del Aprendizaje Automático o *Machine Learning* (ML). Estos modelos, a través de iterativos procesos, son capaces de identificar de manera automatizada aspectos que podrían pasar desapercibidos a los ojos humanos. Actualmente, la relación entre Radiología y ML se estrecha cada vez más, con la mayoría de las publicaciones orientadas tanto al respaldo como al apoyo diagnóstico (62).

La descripción de ML según Arthur Samuel sugiere que se basa en algoritmos que posibilitan a las computadoras aprender a partir de datos, mejorando por sí mismas con la información disponible. Desde una perspectiva más informática o de ingeniería de software, ML se define como una categoría de algoritmo que, mediante la aplicación de un programa, se vuelve más preciso en predecir o inferir resultados específicos sin necesidad de programación directa. En términos estadísticos, ML se conceptualiza como la construcción de modelos a través de algoritmos y análisis estadísticos para predecir eventos de interés. En resumen, ML se posiciona como una disciplina científica que fusiona la estadística tradicional con las ciencias de la computación, facilitando la creación de modelos y la implementación de algoritmos matemáticos alimentados por grandes conjuntos de datos. Esta interacción posibilita que las computadoras logren un aprendizaje automatizado, influyendo positivamente en la eficiencia de los procesos (63).

La esencia del ML radica en identificar objetos mediante un proceso de aprendizaje, donde se emplea una función $f: X \rightarrow Y$. Esta función mapea el dominio de entrada X (datos) al dominio de salida Y (posibles predicciones). La selección de funciones f varía según la clase de algoritmo de aprendizaje utilizado. Esta capacidad resulta especialmente beneficiosa en radiología, ya que permite realizar predicciones y clasificaciones con fines diagnósticos (63).

1.10.1. ML e imágenes

Se destaca que diversas especialidades médicas están utilizando cada vez más estas estrategias. En el ámbito de la radiología, se ha implementado el uso de estas estrategias en diversas modalidades, como rayos X, tomografías computarizadas, resonancias magnéticas, tomografías por emisión de positrones y ultrasonografía, con el objetivo de mejorar y respaldar los resultados obtenidos. En numerosas aplicaciones, los sistemas basados en aprendizaje automático han demostrado un rendimiento excepcional, por lo que contar con el respaldo de estas herramientas en el futuro podría ser esencial para mejorar la toma de decisiones diagnósticas y el seguimiento del estado de salud de los pacientes (63).

Existen varios aspectos que deben considerarse al realizar este tipo de intervenciones. Un elemento crucial para la obtención de datos en el procesamiento de imágenes es la extracción y representación de características, ya que permite obtener la información más relevante contenida en los patrones de interés, favoreciendo la optimización de la tarea de clasificación de patrones. Esta actividad busca extraer la máxima información con la menor dimensión posible, simplificando el procesamiento de variables al analizar solo el conjunto reducido de vectores o características de representación. No obstante, es importante destacar que este proceso requiere etapas tanto previas como posteriores. De acuerdo con diversos autores, la estandarización de un flujo de trabajo adecuado para el procesamiento, análisis y clasificación de imágenes podría comprender las siguientes fases (64).

1.10.2. Preparación e identificación de objetos

Se enfatiza la importancia de mantener una calidad predefinida en la adquisición de imágenes para asegurar análisis posteriores. El formato *Digital Imaging and Communication In Medicine*, (DICOM) es comúnmente utilizado, aunque otros como JPG, PNG, entre otros, son aceptables siempre que cumplan con condiciones análogas en términos de compresión de colores y escala de grises. Una vez establecido este paso, se requiere la selección del plano de examen para la extracción de características, como sagital, coronal, axial, oblicuo, entre otros. En esta etapa, se debe evaluar el tamaño de las imágenes, recomendando trabajar con dimensiones estandarizadas. Este proceso de "limpieza" tiene como objetivo

principal simplificar las variables de entrada a los modelos empleados, asegurando que el entrenamiento alcance la máxima calidad al eliminar posibles ruidos, señales o elementos que podrían afectar la precisión de los resultados (65).

En adición, es fundamental identificar los objetos de estudio, es decir, los elementos que posibilitarán la asociación de la variable de interés con los hallazgos presentes en la imagen, también conocido como etiquetado. Este procedimiento puede llevarse a cabo basándose en información previa existente mediante un proceso supervisado, como un diagnóstico médico que acompaña a la imagen (por ejemplo, una rotura parcial del tendón). También es viable realizar este procedimiento sin información previa, utilizando un enfoque no supervisado, permitiendo que el algoritmo identifique patrones comunes (66).

1.10.3. Extracción de características

Es el elemento fundamental empleado para llevar a cabo análisis. Actualmente, se observa un interés creciente en la extracción e identificación de características lo más fidedignas posible con el propósito de mejorar el rendimiento en las clasificaciones. En el dominio de las imágenes, se emplean diversos métodos para la extracción de características, tales como funciones basadas en regiones de interés, formas, texturas, características de colores, histogramas, momentos de color y vectores de coherencia de color. Se resalta, además, la utilización de grupos de píxeles como un enfoque para obtener texturas, entre otras aplicaciones. No obstante, debido a los extensos intervalos de contraste presentes en diversas imágenes, las técnicas de análisis visual mencionadas podrían resultar insuficientes (67). La Tabla 1, elaborada por Zhang y Sejdić en 2019, proporciona una síntesis de las principales técnicas empleadas para la obtención de características en el ámbito de las imágenes médicas (67).

Tabla 1. Resumen de las técnicas de extracción de características.

Características	Descripción	Ejemplos
Color	Invariante de diferente tamaño y dirección.	Histograma
Forma	Representación binaria de imágenes	Esfericidad
Textura	Uso de la estructura de la imagen, aleatoriedad, linealidad, rugosidad, granulación y homogeneidad.	Características de Haralick, Características de Gabor, Co-ocurrencia, Curvatura, Wavelet.
Local	Uso de la información de la imagen local usando región, objeto de interés, esquinas o bordes.	Patrón binario local, Transformación de características invariantes de escala, Aceleración de funciones robustas.
Otros	Otros métodos de extracción de características de imágenes.	Redes neuronales convolucionales.

1.10.4. Reducción de dimensiones

Aunque constituye una etapa facultativa, se recomienda su implementación debido a que el manejo de datos con elevadas dimensiones suele plantear un desafío considerable. Esta dificultad se origina principalmente en la posibilidad de enfrentar problemas de sobreajuste, ya sea debido a la carencia de conjuntos de datos o a la inclusión de valores de entrada no pertinentes. Asimismo, existe la posibilidad opuesta, en la cual se incorporan datos irrelevantes o redundantes para los análisis. En el contexto de las imágenes médicas, es conocido que existen elementos superfluos que no contribuyen al análisis, como la información del paciente (nombre, edad, lateralidad, centro, fecha) que suele estar presente en la imagen, o incluso la presencia de áreas particularmente oscuras que se repiten sin aportar al patrón de interés (68).

Por consiguiente, la reducción de dimensiones involucra la aplicación de metodologías que posibilitan la transformación de los puntos de datos originales

de alta dimensión en puntos de datos con una dimensión inferior, al tiempo que se conservan las propiedades clave de los puntos de datos originales que son pertinentes para una tarea específica. Entre las técnicas más comúnmente empleadas se encuentran el análisis de componentes principales (PCA), el análisis de componente lineal, los métodos de filtro, los métodos de envoltura y los métodos embebidos (68).

1.10.5. Modelos de ML

Dentro del ámbito de los modelos de ML, se clasifican en diversas categorías y subcategorías (63). Estas incluyen los Modelos de Aprendizaje Supervisado, *Supervised Learning* (SL), los Modelos de Aprendizaje No Supervisado *Unsupervised Learning* (USL), y los Modelos de Aprendizaje Reforzado, *Reinforcement Learning* (RL).

En el caso del SL, el entrenamiento es dirigido, lo que significa que se reconoce el objetivo de interés mediante un proceso de etiquetado, permitiendo que el modelo se entrene con la información proporcionada por el usuario. Por otro lado, en los modelos de USL, los métodos no requieren etiquetado y se utilizan para descubrir patrones entre los datos. En cuanto al RL, el algoritmo de aprendizaje se construye mediante la valoración obtenida por las respuestas existentes, lo que implica una retroalimentación continua de la estructura. Actualmente, se han reportado numerosas subcategorías dentro de ML, sin un esquema claro que indique a qué área pertenece cada una. Esto puede deberse a que, en ocasiones, cada subcategoría puede contribuir al desarrollo de más de una actividad. Por este motivo, se ha propuesto utilizar esquemas generales para las principales categorías con sus respectivas técnicas de análisis (69) (70). Tabla 2.

1.10.6. Regresión lineal

Es uno de los enfoques más comúnmente empleados para anticipar la probabilidad de que ocurra un evento en la investigación clínica. Este método se caracteriza por ser una técnica simple que realiza predicciones de una respuesta cuantitativa, representada por Y , basándose en una única variable predictora, denotada como X . Se parte del supuesto de que existe una relación lineal aproximada entre X e Y (71).

Tabla 2. Categorías de Machine Learning y técnicas de análisis.

Supervisado			No supervisado		Reforzamiento
Regresión	Clasificación	Aprendizaje profundo	Clúster	Reducción de dimensiones	Toma de decisiones
Lineal, No lineal, Modelo lineal generalizado	Árbol de decisión	Redes neuronales convolucionales	Agrupación jerárquica	Análisis lineal discriminatorio	Programación dinámica
Local (LOESS)	K-vecinos más cercanos	Red neuronal recurrente	Agrupamiento K- medias, Fuzzy C- medias, K- medoides	Análisis general discriminatorio	Métodos de Monte Carlo
Mínimos cuadrados ordinarios	Máquinas de soporte vectorial		Mezcla Gaussiana	Análisis de componentes principales	Aprendizaje por diferencia temporal
Métodos de conjunto	Bosques aleatorios		Modelo oculto de Markov		Aprendizaje-Q
Regresión vectorial de soporte, Georradar	Clasificador Bayesiano ingenuo		Redes neuronales		Aprendizaje-R
Redes neuronales	Redes neuronales				

1.10.7. Regresión no lineal

La más utilizada es la regresión logística, la cual considera un conjunto de datos, que poseen una respuesta predeterminada binaria, es decir, que sólo presenta dos categorías de interés (presente/ausente; si/no). En lugar de modelar directamente la respuesta Y , este método considera la probabilidad (Pr) de que Y pertenezca a una categoría en particular (71).

1.10.8. Árboles de decisión, *Decision tree learning* (DTL)

Estos métodos pueden ser empleados en problemas de regresión, pero son especialmente prominentes como técnicas de clasificación en el campo del aprendizaje automático. Se trata de un gráfico acíclico que se utiliza para la toma de decisiones. Un árbol consta de nodos divididos (s), los cuales realizan una partición de los datos x hacia el nodo secundario o hoja izquierdo $cl(s)$ o hacia el nodo secundario o hoja derecho $cr(s)$. Cada hoja representa una clase, pudiendo indicar el valor objetivo más adecuado o la probabilidad de que el objetivo tenga un valor específico (72):

1.10.9. Bosques aleatorios, *Random Forest*, (RF)

Este algoritmo pertenece a la categoría de aprendizaje supervisado y se aplica tanto a problemas de regresión como de clasificación. Al igual que la técnica anterior, enfrenta desafíos de sobreajuste a medida que el árbol crece y los datos se dividen. Los bosques aleatorios abordan esta problemática al construir múltiples árboles de decisión (73).

1.10.10. Máquinas de Soporte Vectorial, *Support Vector Machine* (SVM)

Estos utilizan subconjuntos de datos y características de forma aleatoria. La salida del modelo se obtiene al promediar las predicciones de los árboles individuales, lo que reduce la varianza y previene el sobreajuste (72). Es un algoritmo cuya función permite clasificar de forma binaria con dos clases. Estos modelos construyen hiperplanos en el espacio en el que se encuentran las covariables separando las observaciones según la clase a la que pertenecen. Para esto utiliza funciones de Kernel ($K(x_i, x)$), facilitando relaciones no lineales entre el resultado y las covariables. Como resultado se encuentra una separación con un margen suave dividiendo los puntos pertenecientes a las diferentes clases (74). Donde $x_i = 1, 2, \dots, N$ son vectores de características del conjunto de entrenamiento x , y del indicador de clase correspondiente $y_i \in \{-1, +1\}$.

1.10.11. Redes neuronales, *Neural Networks* (NN) y Aprendizaje Profundo, *Deep Learning*, (DL)

Las redes neuronales son un componente integral del aprendizaje automático. Son algoritmos que se inspiran en las conexiones sinápticas del cerebro, estableciendo conexiones entre capas de entrada y salida a través de nodos. En la actualidad, se han desarrollado capas más profundas para ampliar las redes neuronales con el objetivo de abordar problemas más complejos y facilitar la toma de decisiones entre las capas (75).

Por otro lado, en el ámbito del aprendizaje profundo, se ha observado un aumento significativo en su aplicación para la detección de hallazgos y la clasificación de diversas enfermedades (76). Específicamente, las Redes Neuronales Convolucionales, *convolutional neural networks* (CNN), por sus siglas en inglés) se destacan como los algoritmos más representativos del aprendizaje profundo en el análisis de imágenes. Las arquitecturas de las CNN pueden variar en términos del número y tipo de capas implementadas, adaptándose a las necesidades específicas de la aplicación. Sus beneficios son múltiples, ya que ofrecen predicciones más precisas al mismo tiempo que automatizan la extracción de diferentes niveles de representación de los datos. Esto resulta especialmente útil para aprender representaciones jerárquicas de datos, como en el caso de las imágenes (77)}.

1.11. ENTRENAMIENTO Y VALIDACIÓN: APRENDIZAJE

Para gestionar la complejidad de los modelos, se emplean uno o más parámetros. No obstante, la complejidad excesiva de un modelo puede llevar a un sobreajuste, especialmente si este exhibe un rendimiento destacado en las muestras de entrenamiento, pero un rendimiento deficiente en muestras nuevas. Para equilibrar estos aspectos, se realiza un proceso de validación cruzada, que implica la división de los datos en un subconjunto de entrenamiento y otro de prueba o validación. El subconjunto de entrenamiento se utiliza para modelar diversas configuraciones, que luego se contrastan con el subconjunto de validación, no utilizado previamente, pero cuyos orígenes son conocidos. Una vez optimizados los parámetros del modelo, este se puede emplear para tareas de clasificación de imágenes (78). No obstante, es importante tener en cuenta que, aunque la validación cruzada es un método eficaz para estimar la precisión, una limitación

clave es que cada conjunto de iteraciones de entrenamiento y prueba produce un modelo diferente, por lo que no hay un modelo único que se pueda utilizar al final (61).

Para ilustrar este proceso, imagine que se desea identificar una rotura intrasustancia del tendón del origen de los extensores de muñeca. Se dispone de una base de datos de ecografías de partes blandas pertenecientes a 200 tenistas, divididos en 100 con el diagnóstico de interés y 100 sin la condición. Identificada previamente la región de interés, *region of interest* (ROI), se tienen 100 vectores de entrada en cada grupo. En un escenario de validación cruzada 70/30, se seleccionan 70 imágenes de la ROI del tendón sano y 70 del tendón lesionado, enviándolas al sistema del algoritmo de aprendizaje automático que asignará pesos aleatorios a cada característica. Si la suma es mayor a 0, se clasificará la ROI como una rotura del tendón; de lo contrario, se considerará normal. Luego, se intentará ajustar uno de los pesos para reducir el número de interpretaciones incorrectas. El sistema continuará ajustando hasta que no se observe una mejora adicional en la precisión. A continuación, se tomarán los 30 ejemplos restantes de cada ROI normal y afectado, evaluando la precisión de la predicción. Este proceso se repetirá varias veces para obtener una precisión media del algoritmo en ese conjunto de datos (61).

1.12. EVALUACIÓN DE RESULTADOS

Las medidas de rendimiento de los clasificadores constituyen un aspecto crucial. La mayoría de ellas se fundamenta en una matriz de confusión binaria, con la posibilidad de definir un promedio por clase para cada métrica. La Tabla 3 proporciona una descripción concisa de una matriz de confusión binaria y las métricas correspondientes de rendimiento de clasificación (79) .

Tabla 3. Matriz de confusión binaria y métricas de evaluación de rendimiento.

Clasificador de clase		
Clase de datos	Clasifica como Positivo	Clasifica como Negativo
Positivo	VP (verdadero positivo)	FN (falso negativo)
Negativo	FP (falso positivo)	VN (verdadero negativo)

Métrica	Objetivo	Fórmula
Accuracy (AC)	Efectividad promedio del clasificador.	$AC = VP + VN / VP + VN + FP + FN$
Precisión (PR)	Acuerdo de clase de las etiquetas de datos con las etiquetas positivas dadas por el clasificador.	$PR = VP / VP + FP$
Sensibilidad (SN)	Efectividad de un clasificador para identificar etiquetas positivas.	$SN = VP / VP + FN$
Especificidad (SP)	Con qué eficacia un clasificador identifica etiquetas negativas.	$SP = VN / VN + FP$
Característica Operativa del Receptor (ROC)	Es un gráfico cuya función muestra el rendimiento de un modelo de clasificación en todos los umbrales de clasificación.	$ROC = \sqrt{SN^2 + SP^2} / \sqrt{2}$
Área Bajo la Curva (AUC)	Es un gráfico que mide toda el área bidimensional por debajo de la curva ROC completa.	$n_e = Z^2_{\alpha/2} * V(AUC) / \epsilon^2$

Abreviaciones: n_e , número de sujetos enfermos; $Z_{1-\alpha/2}$, es un valor normal estandarizado; α , nivel de confianza; $V(AUC)$, varianza anticipada del área bajo la curva ROC; ϵ , precisión absoluta requerida en el otro lado del área bajo la curva.

II – JUSTIFICACIÓN

II - JUSTIFICACIÓN

Las tendinopatías constituyen unas de las condiciones musculoesqueléticas más prevalentes en todo el mundo. A pesar de que existe un consenso mundial de expertos de que debe ser considerada como un verdadero problema de salud pública (80) su incidencia y prevalencia aún es inconclusa y depende de una serie de factores metodológicos y de diseño propiamente tal, dificultando obtener una cifra precisa. Es así como una revisión sistemática logró estimar una alta variabilidad de estos indicadores a través de un análisis de 42 estudios de cohorte en distintas poblaciones en el mundo, concluyendo que su impacto para los pacientes y la sociedad aún se encuentra subestimado, particularmente en poblaciones activas, como atletas y trabajadores (81).

A pesar de que las tasas de esta enfermedad siguen siendo escasamente definidas, la carga de enfermedad por esta condición sigue aumentando anualmente. Un análisis en los Estados Unidos a más de 85.300 pacientes demostró categóricamente que tanto el reembolso total como el reembolso promedio por paciente ha ido en aumento exponencialmente con el tiempo. Evidenciando el incremento de la carga económica para el sistema de atención médica, la calidad de vida de los individuos y la productividad laboral (82). Otros estudios más recientes acerca de la evaluación de morbilidad y costo sanitarios acerca de las tendinopatías han concluido que las cargas del sistema de salud debido a estas patologías están en aumento y que debido a los costos de atención médica realzan la necesidad inmediata de intervenciones multifacéticas y transdisciplinarias para reducir la incidencia y mejorar los resultados clínicos de las lesiones de tendones (83).

La TLC al igual que otras tendinopatías, presenta una gran variedad de alternativas e intervenciones terapéuticas con diferentes efectos. Entre ellos se encuentran las intervenciones médicas, tales como las intervenciones quirúrgicas (84), infiltraciones con corticoides, procedimientos de infiltración con plasma rico en plaqueta o terapias celulares (84). Pero también, se encuentran modalidades no invasivas como elementos provenientes de la fisioterapia, tales como el láser (85), las movilizaciones y ajustes articulares de terapia manual (85), el ejercicio terapéutico (53) con énfasis en la contracción excéntrica (86), y más reciente

propuestas de rehabilitación con realidad virtual (87). No obstante, diferentes revisiones sistemáticas y metaanálisis presentan en común que todas las intervenciones poseen múltiples efectos tanto en el corto, mediano, pero en el largo plazo se tienden a igualar los efectos de las intervenciones. Concluyendo que no existe una única recomendación del mejor tratamiento disponible para esta patología musculoesquelética (88) (89).

Se puede observar, que los esfuerzos se han concentrado mayoritariamente en obtener resoluciones terapéuticas de la lesión, enfocados particularmente en disminuir los síntomas y acortar los periodos de latencia. Sin embargo, escasos esfuerzos han derivado de explorar en profundidad el diagnóstico de la TLC, independiente de que existan maniobras físicas para su detección tales como Maudsley, Mills (90), Cozen (91), y técnicas de imágenes médicas que complementan el diagnóstico como la RM o la ecografía de partes blandas (92). Contar con nuevos aspectos propios del diagnóstico de la TLC y herramientas que permitan mejorar la sensibilidad y la especificidad permitiría recabar más información acerca de la condición degenerativa del tejido. A su vez, entendiendo que las roturas intrasustancias son la condición de culminación del desgaste de los tendones, sería interesante comprender en profundidad cómo es el comportamiento tisular en pacientes que padecen TLC.

En este sentido, poseer una definición mejor aceptada por los profesionales médicos y el público que padece esta condición se hace fundamental en el sistema de salud para mejorar la carga de morbilidad y para otorgar una óptima y oportuna resolución (81). Sin embargo, tal como se ha mencionado, el abordaje de esta patología musculoesquelética depende de múltiples factores y en general su diagnóstico se encuentra condicionada a una estructura bastante rígida en el ejercicio de la profesión (51). En la presente tesis se propone conocer en detalle el estado del arte de la literatura en cuanto a los métodos diagnósticos guiados por modelos de machine learning en la extremidad superior, evidenciando cuales son las propuestas más recientes para el soporte diagnóstico. En segundo lugar, se identificarán las principales condiciones degenerativas que se encuentran presenten en los sujetos que padecen esta condición mediante la evaluación exhaustiva de imágenes ultrasonográficas con más de 4000 ecografías, convirtiéndola en una de las series más extensas reportadas en la actualidad en la literatura científica. En tercer lugar, se propone desarrollar una novel prueba

diagnóstica a través de una maniobra física sumado a algún síntoma característico que permita mejorar la capacidad de detección de las roturas intrasustancias de los sujetos que poseen TLC. Finalmente, se emplearán modelos de aprendizaje automático para mejorar la capacidad diagnóstica de los cambios intratendíneos y la capacidad de identificación de roturas intrasustancias del tendón en imágenes ecográficas, permitiendo una comprensión más profunda de los aspectos degenerativos del tendón en las TLC. Estos avances representan hitos fundamentales para una mejor caracterización y comprensión de la lesión, facilitando así una terapéutica guiada en función de la condición específica de cada paciente.

III – OBJETIVOS

III - OBJETIVOS

La presente tesis considerará cuatro objetivos específicos los cuales se encuentran vinculados con sus respectivos objetivos operacionales. Cada uno de ellos serán desarrollados en secciones independientes en el apartado de material y método, resultados y discusión.

3.1. OBJETIVO GENERAL

Mejorar la precisión diagnóstica en la detección de hallazgos degenerativos y roturas intrasustancia del tendón de pacientes con tendinopatía lateral del codo mediante el desarrollo de una prueba diagnóstica y la aplicación de modelos de machine learning.

3.2. OBJETIVOS ESPECÍFICOS

- I. Determinar el rendimiento diagnóstico de modelos de inteligencia artificial en la detección de patologías musculoesqueléticas de la extremidad superior utilizando diferentes modalidades de imágenes.
- II. Identificar los principales hallazgos degenerativos del tendón y su asociación con roturas intrasustancias según tamaño en pacientes con tendinopatía lateral del codo.
- III. Evaluar la precisión diagnóstica de una prueba de índice para detectar una rotura intrasustancia confirmada por ultrasonido en pacientes con tendinopatía lateral del codo.
- IV. Estimar la precisión diagnóstica simultánea de múltiples hallazgos degenerativos y de roturas intrasustancias en imágenes de ultrasonido de pacientes con diagnóstico de tendinopatía lateral del codo.

3.3. OBJETIVOS OPERACIONALES

- I. Realizar un metaanálisis de estudios diagnósticos para analizar la evidencia disponible de la literatura sobre la detección de patologías

musculoesqueléticas de la extremidad superior según diferentes modalidades de imágenes utilizando modelos de machine learning.

- II. Desarrollar un estudio descriptivo utilizando una extensa fuente de exámenes e imágenes ultrasonográficas de pacientes con diagnóstico de tendinopatía lateral del codo.
- III. Elaborar una nueva prueba diagnóstica para detectar roturas intrasustancias en pacientes que padecen tendinopatía lateral del codo.
- IV. Ejecutar modelos de clasificación binaria y multietiqueta mediante herramientas de machine learning para detectar hallazgos degenerativos y roturas intrasustancias simultáneos en imágenes de ultrasonido de pacientes con diagnóstico de tendinopatía lateral del codo.

IV - MATERIAL Y MÉTODO

IV -MATERIAL Y MÉTODO

La presente tesis se encuentra desarrollada mediante cuatro secciones las que corresponden al cumplimiento en profundidad de los respectivos objetivos específicos propuestos.

4.1. ESTUDIO N1

Para poder llevar a cabo este objetivo se desarrolló un metaanálisis el cual se detalla a continuación (93).

4.1.1. Reporte

La ejecución del presente metaanálisis se realizó considerando las directrices del Informe de Elementos para Revisiones Sistemáticas y Metaanálisis (PRISMA) (94).

4.1.2. Pregunta de investigación

Para la pregunta de investigación se consideraron los criterios del acrónimo PICOT (Participantes, P; Intervenciones, I; Comparación, C; Resultado, O; y Tiempo/Tipo de estudio, T), los cuales son:

- P: Pacientes con diagnóstico de patología musculoesquelética de extremidad superior con cualquier tipo de imagen diagnóstica.
- I: Modelos de IA con fines diagnósticos.
- C: Método diagnóstico tradicional.
- O: Rendimiento diagnóstico de los modelos de IA.
- T: Estudios diagnósticos.

4.1.3. Estrategia de búsqueda y fuentes de información

Se llevó a cabo una exhaustiva revisión bibliográfica en las bases de datos MEDLINE / PubMed, SCOPUS, Cochrane Library, Lilacs y SciELO. Se consideró un intervalo de tiempo de diez años desde agosto de 2023 hacia atrás.

Las palabras clave fueron seleccionadas por especialistas en los temas musculoesqueléticos, radiología e inteligencia artificial respectivamente. Se seleccionaron los términos: *"ligament"*, *"muscle"*, *"tendon"*, *"bone"*, *"convolutional neural network"*, *"deep learning"*, *"machine learning"*, *"artificial intelligence"*, *"accuracy"*, *"specificity"*, *"ROC"*, *"sensitivity"*, *"wrist"*, *"hand"*, *"shoulder"* y *"elbow"*. Se emplearon todas las posibles combinaciones, se evitó el uso de texto libre. Se logró acceder al número total de artículos encontrados.

4.1.4. Criterios de selección

Los criterios de inclusión fueron los siguientes: Artículos que abordaran patologías musculoesqueléticas de la extremidad superior, artículos que emplearan análisis de imágenes biomédicas, artículos que incorporaran técnicas de análisis derivadas de la inteligencia artificial, artículos que proporcionaran valores de verdaderos positivos, falsos positivos, verdaderos negativos, y falsos negativos con el fin de calcular la sensibilidad y la especificidad. En el caso de que un artículo presentara más de un modelo de inteligencia artificial se consideraron las métricas reportadas para el análisis. Adicionalmente se seleccionaron artículos redactados en inglés, español y/o portugués, se consideró un periodo máximo de publicación de 10 años hasta agosto de 2023.

En cuanto a los criterios de exclusión, se excluyeron: 1) artículos de revisión, cartas, informes de congresos, trabajos, artículos cadavéricos y descripciones de técnicas; 2) dispositivos médicos o tecnológicos, sensores, realidad virtual o cualquier tipo de objetos tangibles (*hardware*) o intangibles (*software*) que no hicieran uso de algoritmos de inteligencia artificial.

4.1.5. Extracción de datos

Se examinaron los títulos, resúmenes y textos duplicados. Se incorporaron los artículos que cumplieran con los criterios de inclusión y exclusión propuestos. La información extraída incluyó autores y año de publicación, país de origen, patología musculoesquelética, cantidad de pacientes, cantidad de imágenes, algoritmo diagnóstico, valores de VP, VN, FP, FN.

4.1.6. Consideraciones éticas

Se constató que cada artículo seleccionado adhiera a los principios de la Declaración de Helsinki y que recibiera la aprobación de un comité ético científico. Se aseguró que se contara con un consentimiento informado de todos los participantes involucrados en cada estudio.

4.1.7. Evaluación de riesgo de sesgos

La evaluación de calidad y detección de posibles sesgos en los estudios seleccionados se aseguró mediante el uso del instrumento denominado Evaluación de la Calidad de Estudios de Exactitud Diagnóstica (QUADAS-2) (94). Este asigna tres categorías (bajo, incierto o alto) de sesgo para cada aspecto evaluado.

A) Riesgo de sesgo: Selección de pacientes; Prueba índice; Estándar de referencia; Flujo y sincronización.

B) Preocupaciones de aplicabilidad: Selección de pacientes; Prueba índice; Estándar de referencia. La selección de pacientes se evaluó en términos de cómo se eligieron los participantes para el estudio.

4.1.8. Análisis estadísticos

Se calcularon estadísticas resumidas para evaluar la precisión diagnóstica de las pruebas utilizando las métricas de VP, VN, FP, FN. Se llevaron a cabo análisis univariados y bivariados para cada modelo de IA identificado.

4.1.9. Análisis univariados

El tamaño total del efecto se determinó a partir del número de eventos y el número de muestras en los datos de tipo proporción. Se realizaron cálculos separados de sensibilidad y especificidad para cada modelo, así como un cálculo global.

Se calcularon las razones de verosimilitud positiva y negativa. Se aplicó el método que se basa en la transformación logit y luego en la transformación inversa (la transformación logit y el método de Clopper-Pearson) (95), debido a que produce resultados más estables con datos de tipo proporción.

La estimación del efecto combinado se llevó a cabo mediante el cálculo de una razón de probabilidades de diagnóstico (DOR) y su posterior transformación logarítmica ($\ln DOR$) para comparar y combinar los resultados variados (96). Se utilizaron diagramas de bosque, *forest plots* para presentar de manera gráfica la información obtenida.

4.1.10. Análisis bivariado

Se realizó un análisis de subgrupos, designados como "g", según el tipo de imágenes. En el caso de imágenes por resonancia magnética, el Grupo 0 se consideró como punto de referencia, mientras que a las imágenes que no cumplían con este estándar, como radiografías y ecografías, se les asignó un valor de 1.

Se calcularon las razones de probabilidades de diagnóstico para cada subgrupo, y los resultados fueron presentados gráficamente a través de un diagrama de bosque. Se calcularon medidas resumidas de la precisión diagnóstica utilizando el estimador de la curva del área bajo la curva, y se construyó una Curva de Característica Operativa del Receptor Resumen (SROC) (96).

4.1.11. Análisis de heterogeneidad

Debido a la diversidad entre los estudios se empleó un modelo de efectos aleatorios. La cuantificación de ésta se llevó a cabo mediante el método de inversa de varianza al calcular los pesos de los estudios individuales, y el valor de tau, que refleja la variabilidad entre estudios, se determinó utilizando el estimador de DerSimonian-Lair (97).

La proporción de variabilidad atribuida a la diversidad en lugar de al azar entre los estudios, tanto en el conjunto de datos general como en subgrupos, fue evaluada utilizando el indicador I^2 de Higgins. Los valores entre 0% y 40% indican que la diversidad podría no ser relevante, valores entre 30% y 60% señalan diversidad moderada, valores entre 50% y 90% apuntan a diversidad sustancial, y valores entre 75% y 100% indican diversidad considerable (98).

4.1.12. Paquetes y reportes

Para llevar a cabo el metaanálisis de precisión diagnóstica se emplearon los siguientes paquetes: "mada", "mvtnorm", "ellipse", "mvmeta", "meta", "metafor", "rmeta" dentro del programa estadístico R. Se estableció un nivel de significancia de $p < 0.05$ y se utilizaron intervalos de confianza (IC) del 95%. La presentación de resultados se realizó considerando tres decimales. Todas las análisis estadísticos y gráficos se ejecutaron utilizando el paquete de software R (v 4.1.3) (99).

4.2. ESTUDIO N2

Para poder llevar a cabo este objetivo se desarrolló un diseño con alcance asociativo el cual se detalla a continuación (59).

4.2.1. Consideraciones éticas

Se siguieron los principios de la Declaración de Helsinki y cumplió con la legislación chilena correspondiente. La aprobación del estudio fue otorgada por el Comité de Ética Científico Adulto (CECA) del Servicio de Salud Metropolitano Oriente de la ciudad de Santiago de Chile (SSMO). Dado el carácter retrospectivo de la investigación, el comité ético no solicitó el consentimiento informado de los pacientes. El proyecto recibió la aprobación el 7 de agosto de 2018.

4.2.2. Diseño del estudio

Se realizó un estudio descriptivo retrospectivo y multicéntrico. Se cumplieron con las directrices del Reporte de Estudios de Observación en Epidemiología (STROBE) (100). El estudio comenzó el 14 de enero de 2019 y se seleccionaron pacientes con registros de ecografías de codo en la Clínica MEDS en la ciudad de Santiago, Región Metropolitana, Chile, entre el 1 de enero de 2017 y el 31 de diciembre de 2018. Los pacientes incluidos presentaban sospecha de TLC con cualquier signo o síntoma unilateral o bilateral en este período, incluyendo múltiples episodios de la condición si los había.

El software de imágenes médicas (Carestream RIS, v.11) permitió buscar palabras en toda la base de datos de registros electrónicos de ecografías entre 2017 y 2018. Se incluyeron cambios estructurales en el tendón de pacientes con exámenes que contenían la palabra TLC. Cinco radiólogos musculoesqueléticos con más de diez años de experiencia diagnosticaron los exámenes.

Se utilizó un sistema de ultrasonido Aplio 500 (Toshiba America Medical Systems, Inc., Tustin, CA) con un transductor lineal multifrecuencia de 18 MHz para evaluar la condición de TLC. Se registraron datos demográficos como sexo, edad y tamaño de la lesión, excluyendo del análisis exámenes con tratamientos guiados por ecografía, como inyecciones o cirugías. No se proporcionaron datos sobre actividad física, puntuación de dolor ni tiempo de síntomas.

4.2.3. Clasificación de los hallazgos ultrasonográficos

Se presentó una propuesta de categorización estandarizada para evaluar la presencia de cinco criterios principales de ultrasonido, con subdivisiones basadas en informes previos en la literatura.

Se consideró un tendón como normal aquel sin alteraciones en su estructura. La hipoeocogenicidad, caracterizada por su aspecto redondeado y no asociada a la interrupción del tendón, fue evaluada según la extensión de la estructura tendinosa, asignándole una puntuación ordinal de 0 a 3. En esta estructura, 0 representó un nivel normal de ecogenicidad del tendón, el nivel leve (HE1) abarcó del 1% al 30%, el nivel moderado (HE2) del 31% al 50%, y el nivel grave (HE3) superó el 50%.

La neovascularización se examinó a través del número máximo de píxeles de color que ocupan el origen del tendón, utilizando una puntuación ordinal de 0 a 2, donde 0 indicó ausencia de neovascularización, 1 reflejó neovascularización focal (NV1) con el 30% de los vasos sanguíneos, y 2 se refirió a neovascularización múltiple (NV2) que comprende más del 31% de los vasos sanguíneos.

Las roturas lineales intrasustancias se definieron como focos hipoeoicos lineales asociados con la discontinuidad de las fibras del tendón. Se registró el número de roturas tendinosas como una rotura intrasustancia focal (RIS1) o múltiple (RIS2). El tamaño de la anomalía se midió en milímetros y se evaluó la ubicación de la anomalía en dos planos, asignándola como predominantemente afectando las fibras superficiales, medias o profundas y las fibras anteriores, medias o posteriores.

La calcificación (C) se definió como depósitos de calcio en la sustancia del tendón, y la entesopatía (E) se definió como las patologías que afectan las entesis. Ambos hallazgos se categorizaron como una puntuación binaria indicando su ausencia o presencia.

4.2.4. Análisis estadísticos

La información se presentó mediante frecuencias, medias y desviaciones estándar (DS), según correspondiera. Se empleó la prueba de Chi-cuadrado para comparar datos categóricos y la prueba T para variables continuas. Las

comparaciones univariadas entre el tamaño de la rotura en cada dimensión y los hallazgos degenerativos se llevaron a cabo mediante la prueba de Kruskal-Wallis. Se ajustó un modelo de regresión lineal múltiple separado para analizar la asociación entre el sexo, la edad y los hallazgos degenerativos con la rotura del tendón en cada dimensión (longitud, anchura y profundidad).

Ni las variables de exposición finales ni los datos de resultados presentaron valores faltantes. Se consideró un valor de $p < 0,05$ como estadísticamente significativo, y se utilizaron estadísticas de dos colas. Todos los gráficos y análisis se realizaron utilizando R (The R Foundation for Statistical Computing, v. 3.6.2) y (RStudio, v.4.1.0) (99).

4.3. ESTUDIO N3

Para poder llevar a cabo este objetivo se desarrolló una prueba diagnóstica la cual se detalla a continuación (60).

4.3.1. Consideraciones éticas

Se realizó este estudio conforme a la última edición de la Declaración de Helsinki y la normativa científica vigente en Chile. La aprobación del estudio fue otorgada por el CECA-SSMO de la ciudad de Santiago de Chile. Debido al diseño retrospectivo de la investigación, no se solicitó consentimiento informado por parte del comité ético. La aprobación del proyecto se obtuvo el 23 de noviembre de 2021. La privacidad de los participantes se aseguró, confirmando que los datos fueron anonimizados y manejados de manera confidencial.

4.3.2. Diseño de estudio

El diseño de este estudio corresponde a un análisis retrospectivo y multicéntrico con enfoque diagnóstico. Se redactó conforme a las directrices establecidas en Reporting Diagnostic Accuracy Studies (STARD) (101). Se incluyeron todos los registros de pacientes que se sometieron a un examen de ecografía de codo en la Clínica MEDS en la ciudad de Santiago, Región Metropolitana, Chile, durante el periodo comprendido entre el 1 de enero de 2020 y el 30 de junio de 2020. La ejecución de este estudio comenzó el 15 de diciembre de 2021.

4.3.3. Participantes

Los criterios de selección comprendieron la totalidad de los registros electrónicos de pacientes de ambos sexos, mayores de 18 años, que presentaran indicios o síntomas unilaterales o bilaterales de TLC durante el periodo seleccionado. En cuanto a los criterios de inclusión, se tuvo en cuenta el primer episodio de dolor que hubiera persistido entre una y seis semanas, la sensibilidad al tacto en el epicóndilo lateral y la presencia positiva de maniobras clásicas como las pruebas de Cozen, Mill o Maudsley.

También se consideró el dolor nocturno, y la confirmación de la rotura del tendón se llevó a cabo mediante un examen de ecografía. Aunque esta técnica no alcanza el estándar de oro como la RMN la ecografía se revela como una modalidad de imagen valiosa apta para su empleo como herramienta de detección en la epicondilitis lateral (102).

La rotura intrasustancia se definió como un foco hipoeoico lineal asociado con la discontinuidad de las fibras del tendón (59). En cuanto a los criterios de exclusión, se descartaron tratamientos previos como el plasma rico en plaquetas (PRP), la inyección de corticoesteroides o intervenciones quirúrgicas. También se excluyeron otros diagnósticos para el dolor en el codo lateral, así como fracturas alrededor del codo y condiciones en el hombro o el cuello.

4.3.4. Tamaño de muestra

La estimación del tamaño de la muestra se basó en un mínimo de 28 registros médicos y ecográficos de individuos, se emplearon datos provenientes de dos estudios publicados en 2009 y 2014, respectivamente (103), (104).

4.3.5. Instrumentación

La plataforma de imágenes médicas (Carestream RIS, v.11) facilita la realización de búsquedas de palabras en toda la base de datos electrónica de registros de exámenes de ecografía. Otros colaboradores permitieron la creación de una estructura de base de datos en línea de acuerdo con los criterios de inclusión.

La interpretación de los exámenes estuvo a cargo de cinco radiólogos musculoesqueléticos con más de diez años de experiencia. A pesar de que la ecografía es una modalidad de diagnóstico que depende en gran medida del operador, no se evaluó la concordancia entre los cinco radiólogos musculoesqueléticos. Los profesionales emplearon un sistema de ecografía Aplio 500 (Toshiba America Medical Systems, Inc., Tustin, CA, ca.medical.canon) equipado con un transductor lineal multifrecuencia de 18 MHz para informar sobre el RIS en los exámenes de TLC.

4.3.6. Maniobra física

Se realizaron pruebas índice en todos los pacientes a cargo de dos especialistas en cirugía ortopédica con experiencia en patologías de codo y mano (CG, AV), cada uno con más de 15 años de práctica. No se evaluó la concordancia entre los especialistas en cirugía ortopédica enfocados en codo.

Los pacientes fueron posicionados en una silla con el hombro ligeramente abducido, codo flexionado a 90°, antebrazo pronado y muñeca flexionada de modo que la palma quedara hacia abajo. A continuación, el cirujano ortopédico se situó junto al lado afectado del paciente y resistió el esfuerzo máximo de extensión de la muñeca. La incapacidad para resistir dicho esfuerzo mientras mantenían esta posición se consideró un signo positivo de esta nueva maniobra física. Los resultados clínicos establecieron una asociación con la rotura del tendón extensor cuando los pacientes no pudieron resistir el esfuerzo máximo de extensión dorsal de la muñeca contra resistencia y manifestaron dolor nocturno.

4.3.7. Análisis estadísticos

Los datos fueron calculados mediante la precisión diagnóstica, la sensibilidad y la especificidad. El primer paso consistió en crear una tabla 2×2 con los pacientes divididos por RIS en filas y categorías según la prueba índice, el dolor nocturno y los resultados de la prueba de Cozen en columnas.

El VP se definió como el número de casos correctamente identificados con RIS, mientras que el FP representó los casos incorrectamente identificados con RIS. El VN fue el número de casos correctamente identificados sin RIS, y el FN fue el número de casos incorrectamente identificados sin RIS (105).

La precisión fue la clasificación correcta, es decir, la capacidad de diferenciar correctamente entre pacientes con y sin RIS. Se identificó la proporción de verdaderos positivos y verdaderos negativos en todos los casos evaluados $(VP+VN)/(VP+VN+FP+FN)$. La sensibilidad se expresó en porcentaje y definió la proporción de pacientes VP con RIS en el total de pacientes con la condición del tendón $(VP/VP+FN)$. La especificidad fue una medida de la precisión de la prueba diagnóstica, complementaria a la sensibilidad, y se definió como la proporción de pacientes sin RIS con un resultado de prueba negativo en el total de sujetos sin RIS

(VN/VN+FP) (106). También se calcularon intervalos de confianza del 95% (IC del 95%) para todas las métricas diagnósticas.

No se calcularon los valores predictivos positivos y negativos debido a su dependencia en gran medida de la prevalencia de la enfermedad en la población examinada. Se asumió que todos los pacientes tenían síntomas al ser examinados. La curva ROC no se calculó ya que se presentaron resultados binarios, con pocos umbrales que subestiman significativamente el área real bajo la curva (107).

Ni las variables de exposición finales ni los datos de resultado presentaron valores faltantes. Todos los gráficos y análisis se llevaron a cabo utilizando R (The R Foundation for Statistical Computing, v. 3.6.2) y RStudio, v.4.1.0 (99).

4.4. ESTUDIO N4

Para poder llevar a cabo este objetivo se implementó un modelo de detección diagnóstica el cual se detalla a continuación (108).

4.4.1. Consideraciones éticas

Se realizó esta investigación conforme a la versión más reciente de la Declaración de Helsinki y la legislación científica de Chile. La aprobación del estudio fue otorgada por el CECA-SSMO de la ciudad de Santiago de Chile. Dada la naturaleza del estudio, el Comité de Ética no requirió la obtención de consentimiento informado. El proyecto recibió la aprobación el 7 de agosto de 2018.

4.4.2. Diseño de estudio

Este estudio fue concebido como una investigación retrospectiva y multicéntrica, desarrollada conforme a las directrices del Fortalecimiento del Reporte de Estudios Observacionales en Epidemiología (STROBE) (100). Se tomaron en cuenta todos los registros de pacientes que se sometieron a un examen de ultrasonido de codo en la Clínica MEDS en Santiago, Región Metropolitana, Chile. La iniciativa de investigación se puso en marcha el 1 de marzo de 2019.

4.4.3. Participantes

Solo se consideraron imágenes del tendón común extensor. Se eligieron imágenes de ultrasonido y registros médicos de pacientes con diagnóstico de TLC desde el 1 de enero de 2017 hasta el 30 de diciembre de 2018. Los criterios de inclusión abarcaron: (1) diagnóstico clínico de TLC establecido por ortopedas, médicos especializados en medicina deportiva o cualquier especialista en musculoesquelético, (2) realización del examen de ultrasonografía (US) en el centro médico de interés, (3) informe del examen de US elaborado por un radiólogo musculoesquelético con más de 10 años de experiencia y (4) ausencia de restricciones de raza o edad.

Por otro lado, los criterios de exclusión consistieron en: (1) procedimientos guiados por US, como inyecciones de corticoesteroides, células madre y plasma

rico en plaquetas, (2) cirugía previa de TLC y (3) imágenes duplicadas o no distinguibles, las cuales fueron eliminadas del conjunto de datos. La Figura 1 presenta el flujo de trabajo para la selección de los participantes.

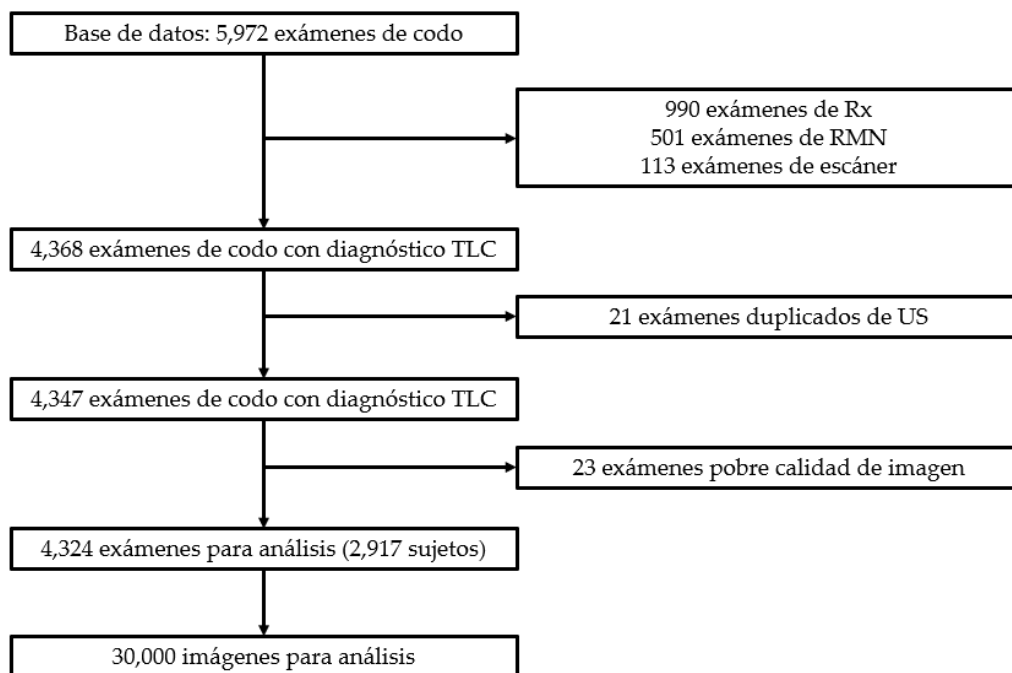


Figura 1. Flujo de los participantes.

4.4.4. Instrumentación

Se realizó la evaluación de todos los tendones extensores comunes en los pacientes con TLC seleccionados mediante un sistema de ultrasonido Aplio 500 (Toshiba America Medical Systems, Inc., Tustin, CA, EE. UU.) equipado con un transductor lineal multifrecuencia, seleccionando una frecuencia de 18 MHz. Las imágenes se guardaron en formato DICOM y se revisaron utilizando un sistema de archivo y comunicación de imágenes (PACS).

Los pacientes diagnosticados con epicondilitis lateral fueron examinados en posición sentada, con el codo flexionado a 90 grados, la muñeca pronada y el brazo apoyado en una mesa (109).

La ecografía en escala de grises y la ecografía Doppler a color son métodos estándar empleados para evaluar cambios estructurales en el tendón (110). Siguiendo las recomendaciones de la literatura, se seleccionaron cuatro hallazgos degenerativos comúnmente prevalentes en los exámenes de ultrasonido, como la hipoecogenicidad, la neovascularización, la entesopatía y la rotura intrasustancia (59).

Se definió una región hipoecoica focal como aquella que era redondeada y no estaba asociada con la interrupción del tendón. La neovascularización se evaluó mediante la presencia de flujo sanguíneo en el Doppler a color, y la entesopatía se evaluó como anomalías óseas en la inserción del tendón. La rotura intrasustancia lineal se definió como un foco hipoecoico lineal asociado con la discontinuidad de las fibras del tendón (111) (112) (113) (21) (7).

Cada hallazgo se evaluó con una puntuación binaria indicando su presencia o ausencia. Se registró cuando un examen presentaba más de un hallazgo degenerativo.

4.4.5. Construcción de base de datos

Se tomaron en cuenta diversas recomendaciones para el preprocesamiento de datos (imágenes), la detección de objetos y la extracción de características (114)(113) (67).

Se crearon dos conjuntos de datos (A y B) destinados a entrenar y probar modelos. En el paso de preprocesamiento, se consideró la eliminación de elementos que generaban ruido en las imágenes, como iluminación desigual, tamaños diversos o porciones de la imagen sin información. La detección de objetos se centró en una zona de interés específica para el análisis, en este caso, la imagen del tendón común extensor. La extracción de características, fundamental en la construcción de cualquier clasificación de patrones, tiene como objetivo extraer la información relevante que caracteriza a cada clase (115).

De acuerdo con la 7ª Conferencia Internacional de Ingeniería de Sistemas y Tecnología de 2017, el análisis y clasificación de texturas en imágenes médicas de

US pueden hacer uso de la extracción de características y las características de textura para determinar las particularidades del patrón de eco (116). Entre las más utilizadas se encuentran la distribución de intensidades (media y desviación estándar), patrones de coocurrencia de píxeles y granularidad de escalas. Luego, se extrajo el contorno de la forma, donde se cuantificó la textura de los píxeles.

Las imágenes de ultrasonido se etiquetaron manualmente con cuatro salidas de clasificación de hallazgos degenerativos (hipoeocogenicidad, neovascularización, entesopatía y rotura intrasustancia) (117), y se incorporaron datos complementarios del paciente como sexo, edad y lado de la lesión (derecho o izquierdo). El proceso final consistió en una combinación de la información del paciente y el análisis de la imagen.

El conjunto de datos A se utilizó para la predicción de imágenes e incluyó datos extraídos de 95 características morfológicas, formas y variables de textura, donde cada imagen correspondía a un diagnóstico (30,007 filas). El conjunto de datos B se empleó para la predicción de pacientes y abarcó 380 variables de extracción de datos, como mediana, desviación estándar, mínimo y máximo, donde cada examen correspondía a un diagnóstico (4,321 filas). La Figura 2 representa el proceso del flujo de trabajo del estudio.

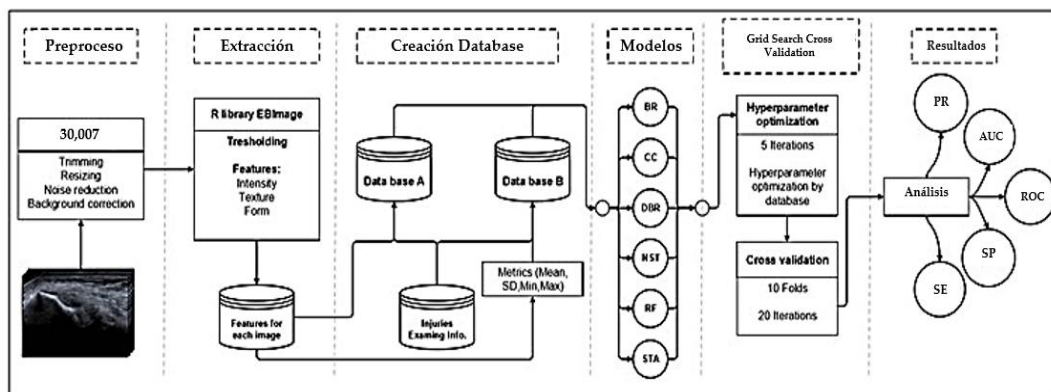


Figura 2. Flujo de selección de datos y sujetos.

4.4.6. Análisis estadísticos y ML

Se optó por el aprendizaje supervisado debido a su prevalencia en las aplicaciones de aprendizaje automático para US. Ambos conjuntos de datos se incorporaron en algoritmos de clasificación binaria y multietiqueta mediante seis métodos de aprendizaje automático: modelo de relevancia binaria, modelo de cadenas de clasificadores, modelo de apilamiento anidado, modelo de relevancia binaria dependiente, generalización por apilamiento y bosque aleatorio.

Todos los modelos se entrenaron para clasificar cuatro hallazgos en el tendón HE, NV, E, RIS en imágenes con diagnóstico de TLC. Inicialmente, cada patrón se identificó de manera individual, y posteriormente, se reconocieron los cuatro hallazgos de manera simultánea. Se utilizaron diversas métricas para evaluar la clasificación de los modelos de aprendizaje automático. Se implementó una validación cruzada repetida en K-fold (KFRCV) con un número de pliegues establecido en diez. Tras este proceso, se obtuvieron los valores de medias e IC.

El análisis de datos se llevó a cabo con R versión 3.6.2 (R Foundation for Statistical Computing). Se emplearon los siguientes paquetes: "EBImage" para la extracción de características, "mlr" para cada algoritmo de aprendizaje automático y "randomForest" para el bosque aleatorio (116) (118) (56). Además, para evaluar la predicción multietiqueta (clasificación), se utilizaron precisión multietiqueta, sensibilidad, especificidad y la curva ROC. También se incluyó un valor predictivo positivo. Se analizaron las diferencias en los hallazgos de ultrasonido entre mujeres y hombres utilizando la prueba t de Student y la prueba de chi-cuadrado. El nivel de significancia se estableció en $p < (0.05)$ y un IC del 95% para todas las métricas.

V – RESULTADOS

V - RESULTADOS

A continuación, se desarrolla la sección de resultados. Para facilitar la lectura, se encontrarán los resultados individualizados por cada uno de los objetivos propuestos.

5.1. ESTUDIO N1

Para poder dar cumplimiento a los resultados de este objetivo se obtuvo la siguiente información del metaanálisis (93).

5.1.1. Resultados de la búsqueda

La Figura 3, presenta los estudios incluidos en este metaanálisis. Después de implementar la estrategia de búsqueda y explorar diversas bases de datos, se logró identificar un total de 3,535 artículos científicos. Tras excluir aquellos que no cumplieran con los criterios de selección, se llevó a cabo una evaluación minuciosa de 60 referencias. Sin embargo, solo 42 artículos fueron considerados aptos. En última instancia, el análisis se ejecutó utilizando únicamente 10 artículos que suministraron toda la información necesaria.

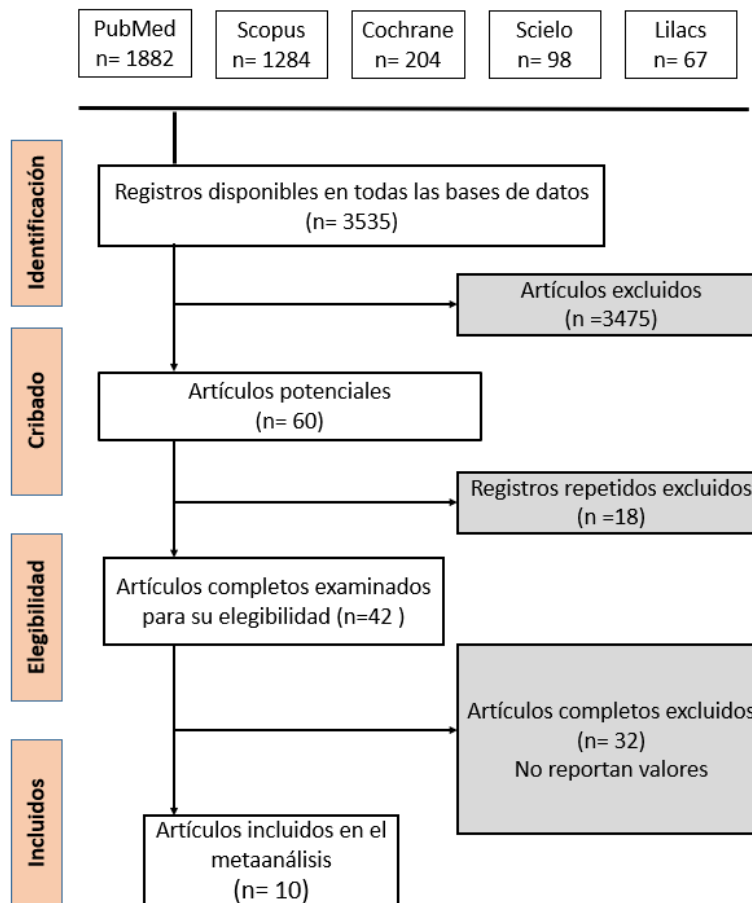


Figura 3. Diagrama de flujo PRISMA 2020

5.1.2. Características de los estudios

Un conjunto total de 4,341 pacientes presentó un total de 31,880 imágenes, que abarcaron radiografías (6 estudios), ecografías (1 estudio) y RMN (3 estudios). Los estudios incluidos provienen de diversas regiones del mundo. Se destaca la presencia de los Estados Unidos y Turquía, ambos contribuyendo con 2 estudios cada uno. Le siguen en la lista Chile, Francia, Corea, Rumania, Suiza y Taiwán, cada uno con la contribución de un artículo. Todos los estudios seleccionados abordaron diagnósticos médicos diversos, centrados exclusivamente en afecciones de la extremidad superior. La muestra incluyó participantes con la condición (N1) y sin la condición de interés (N2) de ambos géneros y de todas las edades.

En la evaluación de la precisión diagnóstica, se emplearon varios modelos de inteligencia artificial, seleccionando únicamente aquellos que proporcionaron todas las métricas relevantes. Entre los modelos destacados se encontraron BoneView, CNN-1, CNN-2, DesNet-161, EfficientNetB3, FSN-8-CNN, Inception-V3, nnU-Net, ResNet-50, ResNet-152, VGG-16 y VGG-19. Los detalles específicos se presentan en la Tabla 4.

Tabla 4. Características de los estudios incluidos.

Referencia	Country	Imagen	Condición	N1	N2	TP	FP	FN	TN	Algoritmo
Cohen et al. 2023	Francia	RX	Fractura muñeca	637	1917	206	41	17	373	Bone View (Gleamer)
Georgeanu et al. 2022	Rumania	RMN	Tumor óseo	23	2106	80	20	4	63	Res-Net 50
Hess et al. 2023	Suiza	RMN	Mango rotador	76	171	2	2	0	59	nnU-Net
Kim et al. 2021	Corea	RX	Fractura ulna radio	2609	9984	271	66	29	624	Des-Net 161
						269	72	31	618	Res-Net 152
Lin et al. 2021	Taiwán	US	Peritendón bíceps	-	3801	72	7	3	18	CNN 2
Ozkaya et al. 2022	Turquía	RX	Fractura escafoides	390	390	28	46	4	12	CNN 1
Saavedra et al. 2023	Chile	RMN	Infiltración grasa manguito rotador	-	606	108	39	6	1512	VGG 19
						105	31	9	1520	Res-Net 50
						99	29	15	1522	Inception V3
Tecle et al. 2020	EE.UU	RX	Osteoporosis	-	265	28	10	6	221	FSN-8-CNN
Üreten et al. 2022	Turquía	RX	Fractura muñeca	-	275	61	2	7	65	VGG 16
Yoon et al. 2021	EE.UU	RX	Fractura escafoides	-	11838	900	469	26	910	Efficient-Net B3

5.1.3. Riesgo de sesgo

Los diez artículos fueron evaluados mediante el uso de la herramienta QUADAS-2. Únicamente en dos de los estudios se observó un riesgo significativo de sesgo en el dominio Flujo y temporización, debido a la falta de detalle en los métodos de selección de pacientes. Adicionalmente, un estudio presentó un riesgo alto de sesgo en los dominios Prueba índice y Estándar de referencia, respectivamente, debido a que no quedó claro si los resultados de la prueba fueron interpretados sin conocimiento de los resultados del estándar de referencia. Es importante resaltar que todos los estudios siguieron estándares de referencia rigurosos, respaldando su clasificación de bajo riesgo de sesgo en los demás dominios evaluados, tal como se refleja en las Figuras 4 y 5.

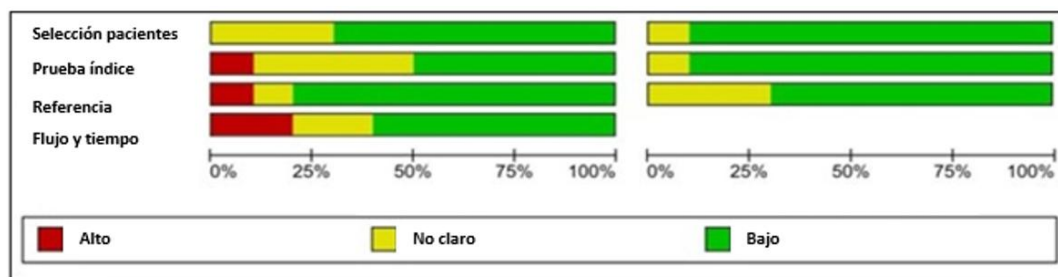


Figura 4. Evaluación de sesgos de los estudios seleccionados.

	Riesgo de sesgo				Preocupación de aplicabilidad		
	Selección pacientes	Prueba índice	Referencia	Flujo y tiempo	Selección pacientes	Prueba índice	Referencia
Cohen et al. (2023)	+	+	+	+	+	+	?
Georgeanu et al. (2022)	?	+	+	-	+	+	+
Hess et al. (2023)	+	+	+	+	?	+	?
Kim et al. (2021)	+	?	-	?	+	+	+
Lin et al. (2020)	+	?	+	+	+	+	?
Ozkaya et al. (2022)	+	?	+	+	+	?	+
Saavedra et al. (2023)	+	+	+	+	+	+	+
Tecle et al. (2020)	?	-	?	-	+	+	+
Üreten et al. (2022)	+	?	+	+	+	+	+
Yoon et al. (2021)	?	+	+	?	+	+	+

- Alto	? No claro	+ Bajo
---	--	---

Figura 5. Evaluación de calidad de los estudios seleccionados.

5.1.4. Análisis univariados

Se evaluaron los análisis individuales de 13 modelos pertenecientes a 10 artículos. En el modelo de efectos aleatorios, la sensibilidad fue de 0.926 (IC del 95%: 0.900; 0.945), indicando una confianza sustancial en la precisión del modelo para detectar correctamente casos positivos. Esto implica que los modelos analizados son capaces de identificar correctamente el 93% de los casos positivos en el conjunto de datos evaluado. El valor (τ^2) de 0.2034 señala una variabilidad significativa en la sensibilidad entre los estudios incluidos en el metaanálisis, independientemente del azar. Con un valor de I^2 del 74% (IC del 95%: 55; 85), se explica que el 74% de la variabilidad en la sensibilidad se atribuye a la heterogeneidad. El valor de H para la sensibilidad fue de 1.96 (IC del 95%: 1.49;

2.58), indicando evidencia de una heterogeneidad ($p < 0,01$) significativa entre los estudios, como se observa en la Figura 6.

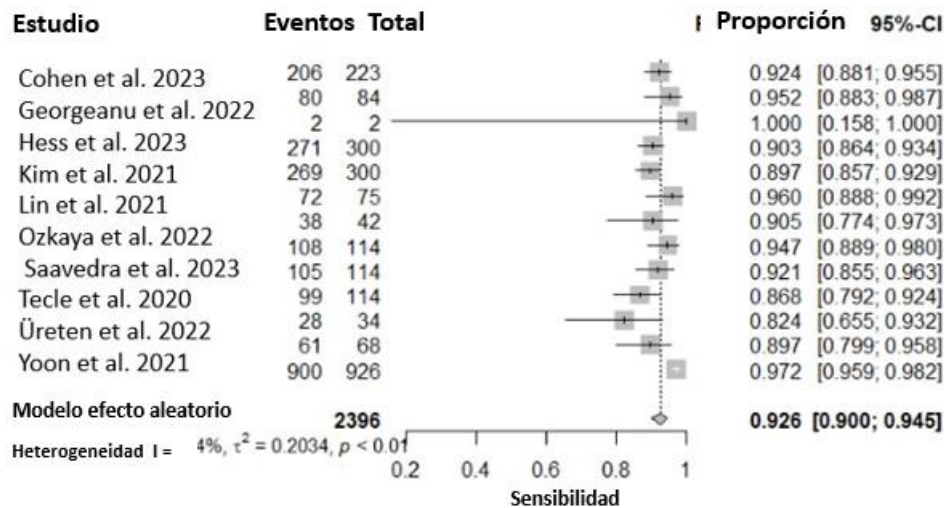


Figura 6. Representación gráfica del bosque de sensibilidad agrupada.

En cuanto a la especificidad reportada en el modelo de efectos aleatorios, fue de 0.908 (IC del 95%: 0.810; 0.958), indicando un alto grado de confianza en la precisión del modelo para detectar correctamente casos negativos. En otras palabras, los modelos tienen la capacidad de identificar correctamente el 90\% de los casos negativos en el conjunto de datos analizado. Con un valor (τ^2) de 2.239, se sugiere que los resultados de los estudios individuales difieren significativamente más allá de lo que se esperaría por azar. Un I^2 del 98.7% (IC del 95%: 98.4; 99) indica que prácticamente toda la variabilidad en los resultados de la especificidad entre los estudios no se puede atribuir simplemente al azar. El valor de H para la especificidad fue de 8.81 (IC del 95%: 7.88; 9.85), señalando evidencia de una heterogeneidad ($p < 0,01$) significativa entre los estudios, como se muestra en la Figura 7.

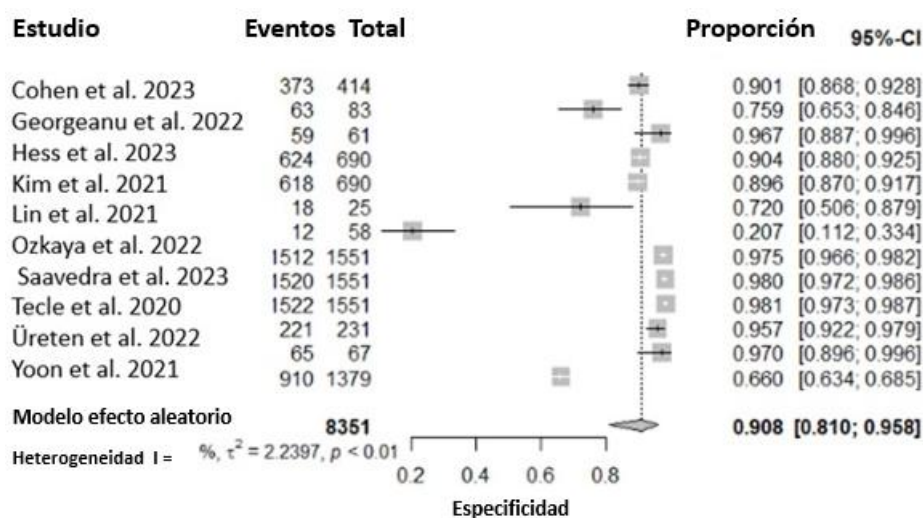


Figura 7. Representación gráfica del bosque de especificidad agrupada.

La Razón de Probabilidad Positiva (RPP) fue de 19.18 (IC del 95%: 8.90; 29.34), lo que significa que la probabilidad de obtener un resultado positivo en individuos con la condición de interés es aproximadamente 19 veces mayor que la probabilidad de obtener un resultado positivo en individuos sin la condición. Por lo tanto, las pruebas realizadas son efectivas para identificar individuos que tienen la condición evaluada.

Adicionalmente, la Razón de Probabilidad Negativa (RPN) obtenida fue de 0.11 (IC del 95%: 0.18; 0.46), indicando que la probabilidad de obtener un resultado negativo en individuos con la condición de interés es aproximadamente 0.11 veces la probabilidad de obtener un resultado negativo en individuos sin la condición. En otras palabras, la prueba tiene una buena capacidad para descartar la presencia de la condición en individuos que realmente no la tienen. Esto destaca la utilidad de la prueba, ya que un bajo valor de RPN indica que la prueba es efectiva para descartar la condición en individuos sin ella.

El estimador de efecto agrupado proporcionado por la Razón de Disparidad Diagnóstica (RDD) fue de 4.68 (IC del 95%: 3.91; 5.45) con un ($p < 0.001$). Por lo tanto, la prueba diagnóstica muestra un poder discriminativo moderadamente alto, ya que la probabilidad de obtener un resultado positivo en individuos mediante estas herramientas es 4.68 veces mayor que la probabilidad de obtener un resultado

positivo en sujetos con el uso de métodos tradicionales de diagnóstico. El lnRDD produjo un valor ligeramente inferior de 4.62 (IC del 95%: 4.02; 5.22) con un ($p < 0.001$), como se muestra en la Figura 8.

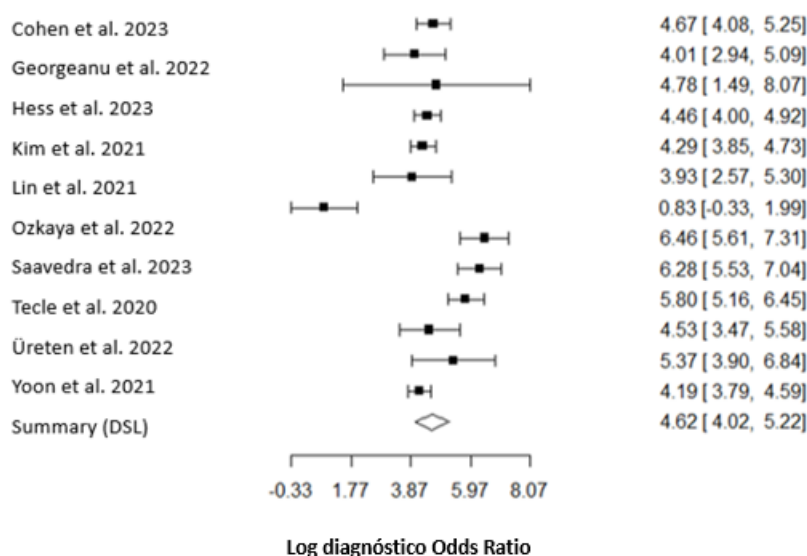


Figura 8. Representación de las tasas de odds diagnósticas logarítmicas.

5.1.5. Análisis bivariados

La estimación de la razón de probabilidades global en el modelo de efectos aleatorios fue de 108.211 (IC del 95%: 50.124; 233.611) con un valor de $p < 0.001$. Este hallazgo sugiere una sólida conexión entre las categorías comparadas en el estudio, respaldada por una significancia estadística sustancial. El parámetro (τ^2) registró 1.689 (IC del 95%: 0.715; 5.089), indicando una variabilidad importante entre los efectos estimados de diversos estudios. Con un I^2 del 87.5%, se revela una heterogeneidad destacada entre los estudios incluidos en el metaanálisis. Este fenómeno implica que los resultados de los estudios individuales difieren considerablemente y no pueden atribuirse simplemente al azar. La estadística de heterogeneidad H fue de 2.83 (IC del 95%: 2.26; 3.55), corroborando la existencia de una cantidad apreciable de variabilidad entre los estudios. La Figura 9 presenta los valores individuales que facilitan el análisis de diferencias sustanciales entre los

grupos (g1: radiografías y ecografías; g0: resonancia magnética), detallando los valores de OR, tau², I² y la estadística Q, respectivamente.

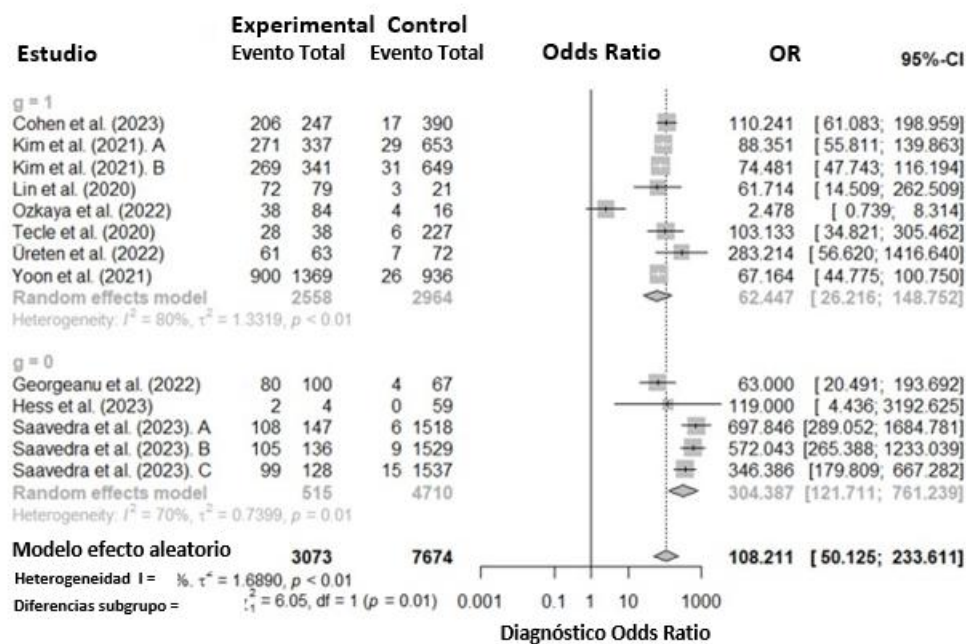


Figura 9. Representación de bosque de las tasas de odds.

La estimación del AUC reportada en este artículo es del 95% (0.913, 0.979), indicando una alta probabilidad de que el modelo realice clasificaciones correctas y, por ende, demostrando un destacado poder discriminativo. En conclusión, la Figura 10 ilustra el rendimiento de diversos modelos de inteligencia artificial en distintos contextos clínicos, ofreciendo una representación visual del desempeño global de los modelos considerados.

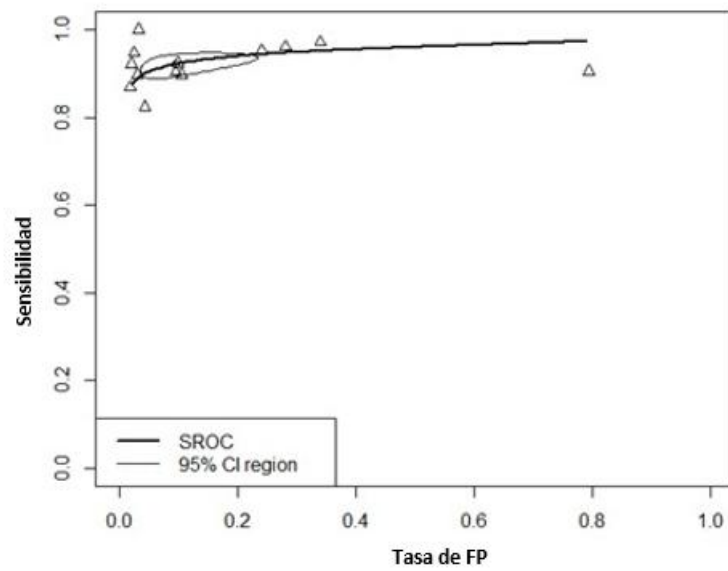


Figura 10. Curva SROC para la precisión de la prueba diagnóstica.

5.2. ESTUDIO N2

Para poder llevar a cabo este objetivo se desarrolló un diseño descriptivo con alcance asociativo el cual se detalla a continuación (59).

5.2.1. Diseño del estudio

Se analizaron un total de 4,335 exámenes de US realizados a 2,926 pacientes con sospecha de TLC. Se excluyeron 11 exámenes por la falta de informes finales. La edad de los pacientes osciló entre 18 y 91 años. Los resultados indican que las mujeres presentan una edad promedio de 47.18 ± 11.00 años ($p < 0.001$), siendo en promedio un año mayor que los hombres. El 70% de los casos involucraron el codo derecho como la zona afectada.

La HE1 fue la anomalía estructural más frecuentemente identificada en los exámenes (53.07%), seguida de entesopatía (35.01%), neovascularización focal (33.55%) y RIS focal (31.91%). Sin embargo, al considerar la clasificación por subtipos, se observó un cambio en la distribución de frecuencias. La mayoría de los hallazgos degenerativos mostraron una frecuencia similar entre mujeres y hombres. Más del 33% de los exámenes con TLC presentaron una RIS en el tendón.

Además, los hombres con RIS focal y múltiple, en las tres dimensiones (longitud, ancho, profundidad), tuvieron un tamaño de desgarró mayor que las mujeres. En general, las variables edad ($p < 0.001$), HE1 ($p < 0.001$) y profundidad RIS1 ($P < 0.01$) mostraron diferencias estadísticamente significativas entre mujeres y hombres. Se proporciona una síntesis en la Tabla 5 para una comprensión más clara.

Tabla 5. Características demográficas y hallazgos ecográficos degenerativos.

Variables	Mujeres (n = 1717)	Hombres (n = 2607)	p- value	Total (n = 4324)
Edad	47.18 ±11.00	45.99 ±11.03	< .001 ^b	46.46±11.03
Lado derecho de la lesión	1179 (68.66)	1790 (68.66)	0.98	2969 (68.66)
HE0	516 (30.05)	877 (33.64)	0.014	1393 (32.21)
HE1	961 (55.96)	1334 (51.16)	< .001 ^b	2295 (53.07)
HE2	222 (12.92)	373 (14.30)	0.19	595 (13.76)
HE3	18 (1.04)	23 (0.88)	0.58	41 (0.94)
NV0	1081 (62.95)	1608 (61.68)	0.41	2689 (62.18)
NV1	558 (32.49)	893 (34.25)	0.22	1.451 (33.55)
NV2	78 (4.54)	106 (4.06)	0.45	184 (4.25)
RIS0	1135 (66.10)	1727 (66.24)	0.94	2862 (66.18)
RIS1	552 (32.14)	828 (31.76)	0.79	1380 (31.91)
Largo, mm	3.84 ± 1.72	4.78 ± 3.24	0.91	4.44 ± 2.81
Ancho, mm	1.54 ± 0.86	2.35 ± 1.95	0.88	2.04 ± 1.66
Profundidad, mm	1.66 ± 0.83	3.02 ± 3.84	< .01 ^b	2.83 ± 3.58
RIS2	30 (1.74)	52 (1.99)	0.55	82 (1.89)
Largo, mm	3.30 ± 1.58	3.39 ± 1.87	0.20	3.36 ± 1.76
Ancho, mm	2.28 ± 1.46	2.36 ± 1.63	0.13	2.33 ± 1.56
Profundidad, mm	2.10 ± 1.13	2.52 ± 1.28	0.57	2.36 ± 1.24
C0	1604 (93.41)	2452 (94.05)	0.43	4056 (93.80)
C	113 (6.58)	155 (5.94)	0.39	268 (6.19)
E0	1118 (65.11)	1692 (64.90)	0.91	2810 (64.98)
E	599 (34.88)	915 (35.09)	0.87	1514 (35.01)

Los valores se presentan como media ± desviación estándar o No. (%). HE0, no hipoecogeneidad; HE1, hipoecogeneidad leve; HE2, hipoecogeneidad moderada; HE3, hipoecogeneidad severa; NV0, sin neovascularización; NV1, neovascularización focal; NV2, neovascularización múltiple; RIS0, sin intrasustancia; RIS1, rotura intrasustancia; RIS2, rotura intrasustancia múltiple; C0, sin calcificación; C, calcificación; E0, sin entesopatía; E, entesopatía. b Significación estadística entre los grupos, $p < 0.05$.

5.2.2. Relación entre rotura el tendón y hallazgos degenerativos

La clasificación de hipoecogenicidad está condicionada por las dimensiones de la RIS. La longitud, en relación con las medidas de ancho y profundidad, muestra la variabilidad más notable entre los niveles de hipoecogenicidad. Además, se registra un incremento en la longitud mediana del desgarro cuando el tendón exhibe niveles más elevados de hipoecogenicidad (HE3).

En particular, la Figura 11 indica que la longitud del desgarro en casos de hipoecogenicidad HE3, comparada con HE2, presenta una diferencia significativa

$p = 0.035$. Sin embargo, al considerar el ancho del desgarro del tendón y diversos niveles de hipoecogenicidad, la mayor disparidad $p < 0.001$ se evidenció entre HE1 y HE2 (no representada en la figura).

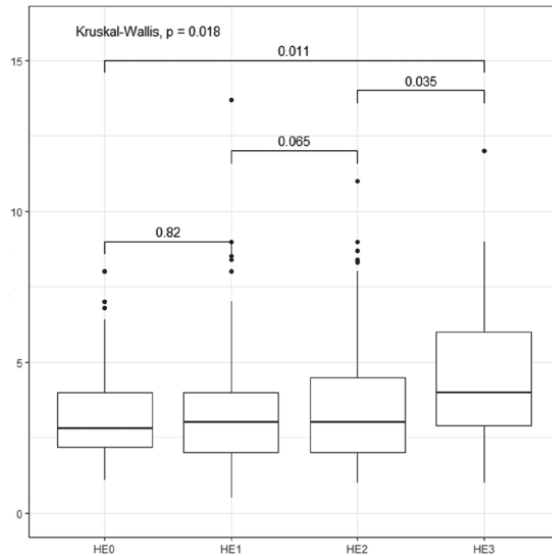


Figura 11. Relación largo de la RIS(mm) e hipoecogeneidad.

De manera similar, en la Figura 12, se presenta la relación correspondiente a la dimensión de ancho del desgarro del tendón en relación con la presencia de dos niveles de NV1 y NV2. Nuevamente, se observó que valores superiores en las dimensiones de ancho del desgarro estuvieron vinculados a la presencia del nivel de neovascularización más elevado $p < 0.005$. Sin embargo, solo las transiciones de NV0 a NV1 demostraron una diferencia significativa $p < 0.004$. Un escenario similar se presentó con la profundidad del desgarro. En este caso, no se identificaron diferencias estadísticamente significativas entre los niveles de neovascularización. Esto podría atribuirse a la baja frecuencia de casos en la dimensión profunda (no representada en la figura).

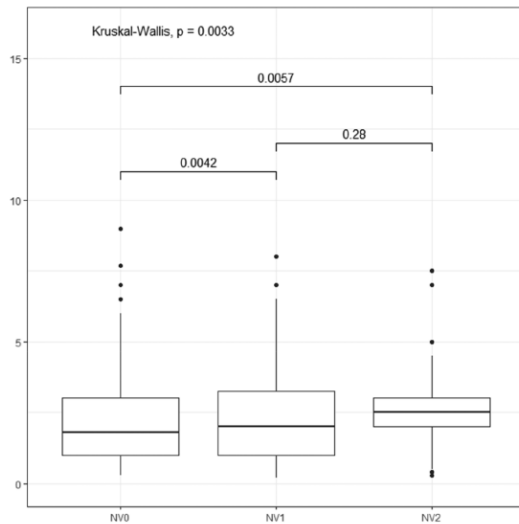


Figura 12. Relación ancho de la RIS(mm) y la neovascularización.

Finalmente, en la Figura 13, se evidencia que la dimensión de profundidad exhibió la diferencia estadística más notable $p < 0.007$ entre HE0 y HE1, sin observarse diferencias sustanciales entre los niveles de hipoecogenicidad HE1 y HE2. Asimismo, la comparación entre HE2 y HE3 muestra significancia estadística $p < 0.025$. No obstante, la cantidad de casos con desgarró en la dimensión de profundidad y HE3 es limitada, por lo que podría no ser representativa.

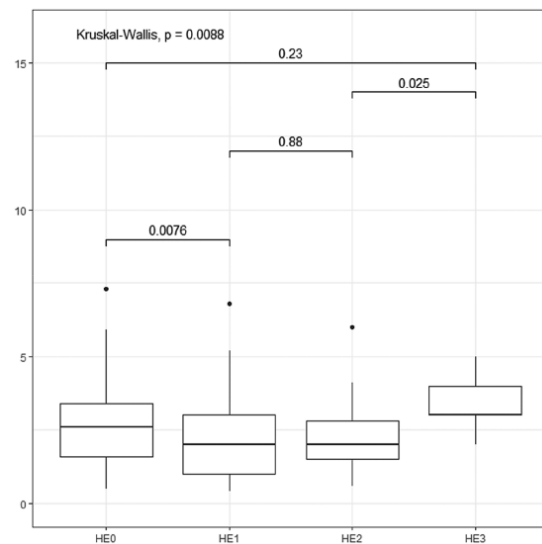


Figura 13. Relación profundidad de la RIS(mm) y la hipoecogenicidad.

Posteriormente, se investigó si las tres dimensiones del desgarro del tendón mostraban diferencias estadísticas en presencia de otras manifestaciones degenerativas. Se llevaron a cabo modelos de regresión lineal para cada dimensión del desgarro, según se detalla en la Tabla 6.

La hipoecogenicidad influyó en todas las dimensiones del desgarro del tendón, con un aumento promedio de H0 a H3 de 1.37 (IC del 95%, 0.57, 2.17]; $p < 0.001$) en longitud, de 1.10 (IC del 95%, 0.33, 1.87]; $p < 0.01$) en ancho y de 1.64 (IC del 95%, 0.40, 2.88]; $p < 0.01$ en profundidad. En la dimensión de longitud, la neovascularización presentó un aumento promedio de 0.42 (IC del 95%, 0.19, 0.65]; $p < 0.001$), y la neovascularización múltiple de 0.71 (IC del 95%, 0.27, 1.15]; $p < 0.01$). La entesopatía redujo el ancho promedio en -0.37 (IC del 95%, -0.62, -0.13]; $p < 0.01$) e incrementó el ancho promedio en 0.57 (IC del 95%, 0.19, 0.95]; $p < 0.01$), respectivamente.

Tabla 6. Coeficientes estimados para la RIS por dimensión según variables.

	Largo n=1107		Ancho n=705		Profundidad n= 226	
Edad						
1 - 40*	93 (8.40)		65 (9.21)		18 (7.96)	
41 - 60	897 (81.02)	0.34 (-0.02, 0.71)	574 (81.41)	0.52 (0.13, 0.91)	189 (83.62)	-0.12 (-0.71, 0.46)
61 +	117	0.88 (0.40, 1.35) ^b	66	0.61 (0.09, 1.14) ^b	19 (8.40)	0.05 (-0.73, 0.83)
Sexo						
Mujeres*	435 (39.30)		279 (39.58)		86 (38.05)	
Hombres	672 (60.70)	0.13 (-0.07, 0.34)	426 (60.42)	0.10 (-0.12, 0.33)	140 (61.95)	0.32 (0.00, 0.65) ^b
Hallazgos degenerativos						
HE0*	101 (9.12)		82 (11.65)		50 (22.13)	
HE1	679 (61.34)	-0.14 (-0.53, 0.24)	394 (55.88)	-0.17 (-0.56, 0.22)	113 (50)	-0.15 (-0.62, 0.31)
HE2	304 (27.46)	0.06 (-0.35, 0.48)	210 (29.78)	0.55 (0.12, 0.98) ^b	58 (25.66)	-0.02 (-0.59, 0.54)
HE3	23 (2.07)	1.37 (0.57, 2.17) ^b	19 (2.69)	1.10 (0.33, 1.87) ^b	5 (2.21)	1.64 (0.40, 2.88) ^b
NV0*	340 (30.71)		230 (32.64)		90 (39.83)	
NV1	693 (62.60)	0.42 (0.19, 0.65) ^b	432 (61.27)	0.22 (-0.02, 0.48)	123 (54.42)	-0.07 (-0.42, 0.28)
NV2	74 (6.68)	0.71 (0.27, 1.15) ^b	43 (6.09)	0.34 (-0.15, 0.84)	13 (5.75)	-0.64 (-1.38, 0.08)
C0*	1019 (92.06)		665 (94.33)		217 (96.02)	
C	88 (7.94)	-0.21 (-0.60, 0.16)	40 (5.67)	-0.22 (-0.71, 0.26)	9 (3.98)	-0.71 (-1.56, 0.14)
E0*	652 (58.90)		380 (53.91)		112 (49.56)	
E	455 (41.10)	-0.20 (-0.43, 0.01)	325 (46.09)	-0.37 (-0.62, -0.13) ^b	114 (50.44)	0.57 (0.19, 0.95) ^b

HE0, no hipoecogeneidad; HE1, hipoecogeneidad leve; HE2, hipoecogeneidad moderada; HE3, hipoecogeneidad severa; NV0, sin neovascularización; NV1, neovascularización focal; NV2, neovascularización múltiple; RIS0, sin intrasustancia; RIS1, rotura intrasustancia; RIS2, rotura intrasustancia múltiple; C0, sin calcificación; C, calcificación; E0, sin entesopatía; E, entesopatía. b Significación estadística entre los grupos, $p < 0.05$.

Para casi todas las dimensiones, la neovascularización y los hallazgos de hipoecogenicidad moderada y severa presentaron coeficientes positivos. A mayor presencia de estos hallazgos, mayor fue el tamaño del desgarro. No obstante, la presencia de calcificaciones y entesopatía se asoció principalmente con una reducción en el tamaño del desgarro.

Se observó que el tamaño del desgarro del tendón, en las tres dimensiones, aumentó con la edad. Pacientes mayores de 60 años presentaron un desgarro del tendón de mayor longitud 0.88 ([IC del 95%, 0.40, 1.35]; $p < 0.001$) y un desgarro del tendón de mayor ancho 0.61 ([IC del 95%, 0.09, 1.14]; $p < 0.05$), ambos con valores de p significativos. Aunque los pacientes varones mostraron un tamaño de desgarro mayor en todas las dimensiones en comparación con las mujeres, esto solo fue ligeramente significativo en la profundidad del desgarro 0.32 ([IC del 95%, 0.00, 0.65]; $p < 0.05$). Es importante destacar que los pacientes en los grupos de edad de 0 a 20 años y mayores de 81 años no presentaron desgarro del tendón.

Finalmente, se evaluó si las interacciones entre las variables de exposición podrían mejorar los modelos en alguna de las tres dimensiones propuestas del desgarro del tendón. Sin embargo, ninguna de las interacciones entre los diferentes tipos de manifestaciones degenerativas mostró significancia estadística.

5.3. ESTUDIO N3

Para poder llevar a cabo este objetivo se desarrolló una prueba diagnóstica la cual se detalla a continuación (60).

5.3.1. Participantes

Este estudio involucró a cuarenta y un pacientes en total. No obstante, la información completa para el análisis estadístico se obtuvo de treinta y nueve pacientes (39 codos), como se ilustra en la Figura 14.

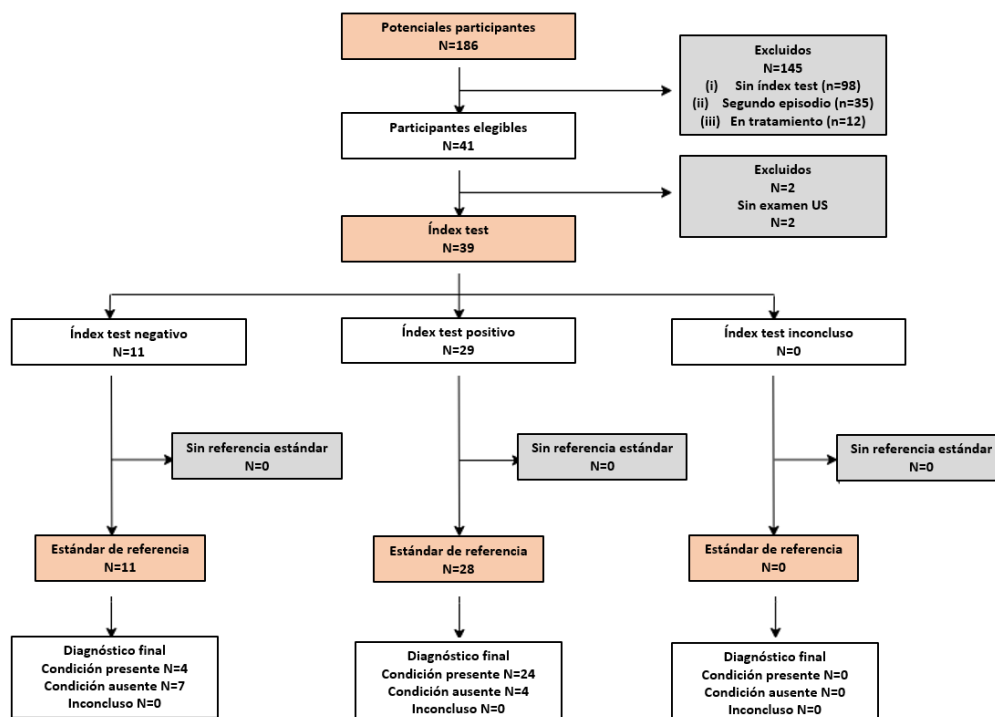


Figura 14. Diagrama de flujo STARD.

La composición de la muestra incluyó a 25 mujeres (64%) y 14 hombres (36%). La edad promedio fue de 47.7 años, con un rango de 34 a 69 años. De todos los codos, el 95% presentaba afectación en el lado dominante, y el 82% se encontraba

en el lado derecho. Por su parte, 23 pacientes (59%) reportaron experimentar dolor nocturno. Asimismo, el 85% de los pacientes mostró un resultado positivo en la Prueba de Cozen. Un total de 28 pacientes (72%) presentaban una rotura intrasustancia, y otros 28 también mostraron un resultado positivo en la prueba índice. La Tabla 7, proporciona información detallada sobre las características demográficas y clínicas iniciales de los participantes.

Tabla 7. Características demográficas y hallazgos.

ID	Sexo	Edad	Dominancia	Lesión	RIS	Nocturno	Cozen	Prueba Índice
1	F	49	1	1	1	1	1	1
2	F	54	1	1	0	0	1	0
3	F	42	1	1	1	1	1	1
4	F	49	1	1	1	1	1	1
5	M	41	1	0	1	1	1	0
6	F	49	1	1	0	0	1	0
7	M	43	1	0	1	0	0	1
8	M	41	1	1	1	0	1	0
9	M	69	1	1	1	0	1	1
10	F	47	1	1	1	1	1	1
11	M	48	1	1	1	0	0	0
12	F	44	1	1	1	0	1	1
13	M	43	1	1	0	1	1	1
14	F	45	1	0	1	1	1	1
15	F	48	1	1	1	0	1	1
16	F	36	1	1	0	0	1	0
17	F	48	1	1	1	1	1	1
18	F	38	1	1	0	0	1	0
19	F	50	1	1	1	1	1	1
20	F	55	1	1	1	1	1	1
21	F	51	1	1	1	1	1	1
22	F	56	1	1	1	1	0	1
23	M	41	1	1	1	1	0	1
24	M	34	1	1	1	1	1	1
25	M	56	1	1	1	0	1	0
26	M	44	1	1	1	1	1	1
27	F	44	1	1	1	1	1	1
28	F	44	1	1	0	0	1	0
29	F	45	1	1	0	0	1	1

ID	Sexo	Edad	Dominancia	Lesión	RIS	Nocturno	Cozen	Prueba Índice
30	M	40	1	1	1	1	0	1
31	F	48	1	0	0	0	1	0
32	M	67	1	0	1	1	1	1
33	M	57	1	1	1	1	1	1
34	F	55	1	1	1	1	1	1
35	F	34	1	1	1	0	1	1
36	M	43	0	0	0	0	1	0
37	F	48	1	1	1	1	1	1
38	F	59	0	1	0	1	1	1
39	F	57	1	0	0	1	0	1

ID, identificador; RIS, rotura intrasustancia, F, femenino; M, masculino; 0, ausente; 1, presente.

5.3.2. Resultados de la prueba

En términos generales, la eficacia de la prueba índice (0.79; IC del 95%, 0.64-0.91) superó la presencia de dolor nocturno (0.72; IC del 95%, 0.55-0.85) y el signo positivo en la prueba de Cozen (0.61; IC del 95%, 0.45-0.77). La sensibilidad o tasa de verdaderos positivos de la prueba índice (0.86; IC del 95%, 0.67-0.96) fue cercana a la del dolor nocturno (0.87; IC del 95%, 0.66-0.97) pero superior a la de la prueba de Cozen (0.70; IC del 95%, 0.51-0.84). Por último, la especificidad o tasa de verdaderos negativos de la prueba índice (0.64; IC del 95%, 0.31-0.89) fue significativamente mayor que el informe de dolor nocturno (0.5; IC del 95%, 0.25-0.75) y la prueba de Cozen (0.17; IC del 95%, 0.00-0.64). La Tabla 8, presenta la exactitud y precisión diagnóstica (con intervalos de confianza del 95%). No se registraron eventos adversos durante la realización de la prueba índice o el estándar de referencia.

Tabla 8. Mediciones diagnósticas de las pruebas.

Examinación	Precisión [95% CI]	Sensibilidad (TPR) [95% CI]	Especificidad (TNR) [95% CI]
Prueba Índice	0.79 (0.64, 0.91)	0.86 (0.67, 0.96)	0.64 (0.31, 0.89)
Presencia nocturna	0.72 (0.55, 0.85)	0.87 (0.66, 0.97)	0.5 (0.25, 0.75)
Prueba de Cozen	0.61 (0.45, 0.77)	0.70 (0.51, 0.84)	0.17 (0.00, 0.64)

5.4. ESTUDIO N4

Para poder llevar a cabo este objetivo se implementó un modelo de detección diagnóstica el cual se detalla a continuación (108).

5.4.1. Tendinopatía del extensor común

En el análisis de datos de este estudio se incorporaron un total de 30,007 imágenes de ultrasonido, con un promedio de 6.9 imágenes en 4,324 exámenes, junto con los expedientes médicos de 2,917 pacientes diagnosticados con TLC. La edad de los pacientes osciló entre 7 y 91 años. Las mujeres, con una media de 47.18 ± 11.00 años, resultaron ser mayores que los hombres ($p < 0.001$), y también exhibieron diferencias estadísticas en la detección de hipoecogenicidad en comparación con los hombres ($p = 0.01$). En todos los exámenes, se observó al menos un hallazgo degenerativo. Las características del ultrasonido se detallan en la Tabla 9.

Tabla 9. Hallazgos ultrasonográficos según sexo.

Hallazgos degenerativos	Femenino (N=1717) Media $\pm$ SD	Masculino (N=2607) Media $\pm$ SD	p-value	Total (N=4324) Media $\pm$ SD
Edad	47,18 $\pm$ 11.00	45,99 $\pm$ 11.03	< .001 ^a	46,46 $\pm$ 11,03
Lado derecho	1179 (68.88)	1790 (68.66)	0.98	2969 (68,66)
HE	1201 (69.94)	1730 (66.35)	.0119 ^b	2931 (67,75)
NV	636 (37.04)	999 (38.31)	0.4093	1635 (37,79)
E	599 (34.88)	915 (35.09)	0.9411	1514 (35,00)
RIS	582 (33.89)	880 (33.75)	0.9521	1462 (33,80)

HE, hipoecogenicidad; NV, neovascularización; E, entesopatía; RIS, rotura intrasustancia; a, p-value < .001; b, p-value < .01.

5.4.2. Modelos de ML como clasificador binario

La Tabla 10, muestra el rendimiento de la clasificación binaria (AUC, sensibilidad y especificidad) para ambos conjuntos de datos (A y B) en cada uno de

los seis algoritmos de aprendizaje automático. Se evaluaron los principales hallazgos degenerativos en TLC, como la hipoecogenicidad, neovascularización, entesopatía y la rotura intrasustancia. Al analizar AUC, sensibilidad y especificidad, se observó variabilidad entre la mayoría de los modelos. Los resultados se describieron en su mayoría con un rango mínimo del 95% de IC, lo que evidencia un rendimiento sólido para todos los modelos. Es relevante destacar que el modelo de RF obtuvo los resultados más destacados.

A su vez, en la Tabla 10, en el conjunto de datos A, el modelo de RF registró los valores promedio más elevados en AUC, sensibilidad y también especificidad para cada hallazgo degenerativo. En particular, el AUC y la sensibilidad destacaron por su rendimiento óptimo en la RIS, con 0.991 [IC del 95%, 0.99 - 0.99] y 0.775 [IC del 95%, 0.77 - 0.77], respectivamente. Por otro lado, la especificidad alcanzó valores superiores en la hipoecogenicidad, marcando 0.821 [IC del 95%, 0.82 - 0.82].

Se observó una situación similar en el conjunto de datos B, donde los valores fueron ligeramente menores para los mismos hallazgos y modelos. Nuevamente, el modelo de RF demostró un rendimiento superior para todas las métricas y características degenerativas. En la Tabla 10, se destacaron los valores más altos de AUC y sensibilidad para el desgarro intrasustancia, alcanzando 0.937 [IC del 95%, 0.93 - 0.94] y 0.82 [IC del 95%, 0.82 - 0.82], respectivamente. La hipoecogenicidad también exhibió una especificidad superior en comparación con otros hallazgos degenerativos, registrando 0.763 [IC del 95%, 0.72 - 0.72].

Tabla 10. Rendimiento según clasificadores y dataset A y B.

Medida A	Modelo	HE [95% CI]		NV [95% CI]		E [95% CI]		RIS [95% CI]	
AUC	BR	0.8064	(0.81,0.81)	0.9016	(0.9008,0.9024)	0.7482	(0.7471,0.7493)	0.9637	(0.9632,0.9643)
	CC	0.8108	(0.81,0.81)	0.8976	(0.8964,0.8988)	0.6954	(0.6893,0.7015)	0.9618	(0.9604,0.9632)
	DBR	0.8048	(0.8,0.81)	0.8921	(0.8912,0.893)	0.6488	(0.6473,0.6504)	0.9566	(0.9542,0.9589)
	NST	0.8064	(0.81,0.81)	0.9014	(0.9006,0.9022)	0.7463	(0.7451,0.7475)	0.9637	(0.9632,0.9643)
	RF	0.9286	(0.93,0.93)	0.9743	(0.9739,0.9746)	0.8993	(0.8987,0.9)	0.9911	(0.9909,0.9913)
	STA	0.8064	(0.81,0.81)	0.8477	(0.8467,0.8487)	0.688	(0.6866,0.6894)	0.9353	(0.9346,0.936)
SN	BR	0.577	(0.58,0.58)	0.7043	(0.7037,0.7048)	0.6568	(0.656,0.6576)	0.7601	(0.7595,0.7607)
	CC	0.5781	(0.58,0.58)	0.7022	(0.7015,0.703)	0.6234	(0.6196,0.6273)	0.759	(0.7581,0.76)
	DBR	0.5766	(0.58,0.58)	0.6994	(0.6988,0.7)	0.594	(0.593,0.5951)	0.7561	(0.7547,0.7575)
	NST	0.577	(0.58,0.58)	0.7041	(0.7036,0.7047)	0.6556	(0.6547,0.6564)	0.7601	(0.7595,0.7607)
	RF	0.6077	(0.61,0.61)	0.7412	(0.7407,0.7418)	0.7522	(0.7516,0.7529)	0.7755	(0.7749,0.776)
	STA	0.577	(0.58,0.58)	0.6768	(0.6763,0.6774)	0.6187	(0.6178,0.6197)	0.7442	(0.7435,0.7448)
SP	BR	0.7293	(0.73,0.73)	0.6972	(0.6966,0.6979)	0.5913	(0.5908,0.5918)	0.7034	(0.7029,0.704)
	CC	0.7326	(0.73,0.73)	0.6952	(0.6945,0.696)	0.5719	(0.5696,0.5742)	0.7026	(0.7018,0.7035)
	DBR	0.7282	(0.73,0.73)	0.6926	(0.6919,0.6932)	0.5548	(0.5542,0.5553)	0.7003	(0.6991,0.7015)
	NST	0.7293	(0.73,0.73)	0.6971	(0.6965,0.6977)	0.5906	(0.5901,0.5911)	0.7034	(0.7029,0.704)
	RF	0.8208	(0.82,0.82)	0.7329	(0.7323,0.7335)	0.6469	(0.6464,0.6475)	0.7155	(0.7149,0.716)
	STA	0.7293	(0.73,0.73)	0.6708	(0.6701,0.6715)	0.5692	(0.5686,0.5697)	0.691	(0.6904,0.6916)

Medida B	Modelo	HE [95% CI]		NV [95% CI]		E [95% CI]		RIS [95% CI]	
AUC	BR	0.8308	(0.83,0.83)	0.9252	(0.9232,0.9273)	0.7811	(0.778,0.7843)	0.9605	(0.9573,0.9636)
	CC	0.8305	(0.83,0.83)	0.9067	(0.9019,0.9115)	0.7228	(0.7142,0.7314)	0.9644	(0.9618,0.9669)
	DBR	0.7888	(0.79,0.79)	0.846	(0.842,0.85)	0.6477	(0.6434,0.652)	0.9656	(0.9633,0.9678)
	NST	0.8308	(0.83,0.83)	0.9269	(0.9253,0.9286)	0.781	(0.7779,0.7842)	0.9605	(0.9573,0.9636)
	RF	0.8887	(0.89,0.89)	0.9654	(0.9644,0.9664)	0.8517	(0.8491,0.8542)	0.9865	(0.9857,0.9872)
	STA	0.8299	(0.83,0.83)	0.8701	(0.8668,0.8734)	0.7222	(0.7176,0.7269)	0.9378	(0.9353,0.9403)
SN	BR	0.6062	(0.61,0.61)	0.7641	(0.7623,0.7659)	0.6821	(0.6798,0.6844)	0.8044	(0.8019,0.8069)
	CC	0.6061	(0.61,0.61)	0.7525	(0.7494,0.7555)	0.6444	(0.6388,0.6501)	0.807	(0.8048,0.8093)
	DBR	0.5927	(0.59,0.59)	0.7149	(0.7122,0.7177)	0.5957	(0.5928,0.5986)	0.8078	(0.8058,0.8099)
	NST	0.6062	(0.61,0.61)	0.7651	(0.7636,0.7667)	0.6821	(0.6798,0.6844)	0.8044	(0.8019,0.8069)
	RF	0.6247	(0.62,0.63)	0.789	(0.7876,0.7905)	0.7279	(0.7257,0.73)	0.8216	(0.8201,0.8232)
	STA	0.6059	(0.6,0.61)	0.7298	(0.7275,0.7321)	0.6441	(0.6409,0.6473)	0.7895	(0.7874,0.7916)
SP	BR	0.7238	(0.72,0.73)	0.6602	(0.6587,0.6616)	0.5983	(0.597,0.5996)	0.6549	(0.6532,0.6567)
	CC	0.7236	(0.72,0.73)	0.6533	(0.651,0.6555)	0.5779	(0.5749,0.5809)	0.6562	(0.6547,0.6578)
	DBR	0.6955	(0.69,0.7)	0.6303	(0.6285,0.632)	0.5516	(0.5501,0.5532)	0.6567	(0.6551,0.6582)
	NST	0.7238	(0.72,0.73)	0.6608	(0.6594,0.6622)	0.5983	(0.597,0.5996)	0.6549	(0.6532,0.6567)
	RF	0.763	(0.76,0.76)	0.6753	(0.6739,0.6767)	0.623	(0.6218,0.6242)	0.6637	(0.6622,0.6652)
	STA	0.7233	(0.72,0.73)	0.6394	(0.6377,0.6411)	0.5776	(0.576,0.5793)	0.6473	(0.6458,0.6489)

AUC, Area Under the Curve; SN; sensibilidad, SP; especificidad; HE, hipoecogeneidad; NV, neovascularización; RIS, rotura intrasustancia; E, entesopatía; BR, binary relevance model; CC, classifier chains model; NST, nested stacking model; DBR, dependent binary relevance model; STA, staking generalization; RF, random forest.

5.4.3. Modelos de ML como clasificador multietiqueta

En la sección de resultados previa, los modelos de aprendizaje automático evaluaron la clasificación binaria de cada hallazgo degenerativo. Ahora, estos métodos emplearon un clasificador multietiqueta para identificar simultáneamente los cuatro tipos de hallazgos en el tendón en ambos conjuntos de datos. En este escenario, el diagnóstico mostró diferentes niveles de precisión en todos los modelos de aprendizaje automático. Cuando el diagnóstico se basó en la combinación de hallazgos degenerativos, el algoritmo de bosques aleatorios nuevamente exhibió el mejor rendimiento entre los modelos seleccionados.

La Tabla 11, revela que los bosques aleatorios en el conjunto de datos A lograron el valor de precisión multietiqueta más alto de 0.772 [IC del 95%, 0.771, 0.773]. De manera similar, para la condición representada en el conjunto de datos B, estos resultados indican que el modelo se desempeña bien en entornos de prueba sin presentar problemas de sobreajuste, algo común en los métodos de aprendizaje profundo o redes neuronales. El valor de precisión multietiqueta fue de 0.723 [IC del 95%, 0.721, 0.726]. Además, se observan puntuaciones elevadas de macro y micro-AUC en los modelos de bosques aleatorios en ambos conjuntos de datos. Estos resultados podrían explicarse por el equilibrio entre sensibilidad y especificidad que demuestran los modelos de bosques aleatorios. Especialmente, la micro-AUC observada en el conjunto de datos A de 0.962 [IC del 95%, 0.962 – 0.963] y 0.942 [IC del 95%, 0.941 – 0.943] en los resultados del conjunto de datos B resulta crucial, ya que agrega las contribuciones de todas las clases para calcular la métrica promedio.

Tabla 11. Rendimiento multietiqueta según modelo

Dataset	Modelo	Macro AUC		Micro AUC		SN		SP		Precisión	
A	BR	0.8549	(0.8544,0.8555)	0.9110	(0.9106,0.9113)	0.7007	(0.7003,0.701)	0.7103	(0.71,0.7105)	0.6835	(0.6825,0.6846)
	CC	0.8414	(0.8398,0.8429)	0.8916	(0.8895,0.8936)	0.6912	(0.6902,0.6922)	0.7004	(0.6993,0.7014)	0.6912	(0.6894,0.6929)
	DBR	0.8256	(0.8246,0.8266)	0.8655	(0.8648,0.8662)	0.6785	(0.6781,0.6788)	0.6870	(0.6866,0.6874)	0.6972	(0.6963,0.698)
	NST	0.8544	(0.8539,0.8549)	0.9108	(0.9104,0.9111)	0.7006	(0.7002,0.7009)	0.7102	(0.7099,0.7104)	0.6834	(0.6824,0.6845)
	RF	0.9482	(0.9479,0.9485)	0.9627	(0.9625,0.963)	0.7259	(0.7256,0.7262)	0.7368	(0.7365,0.7371)	0.7724	(0.7715,0.7733)
	STA	0.8193	(0.8188,0.8199)	0.8977	(0.8973,0.8981)	0.6942	(0.6938,0.6945)	0.7035	(0.7032,0.7038)	0.6835	(0.6825,0.6846)
B	BR	0.8742	(0.8727,0.8756)	0.9183	(0.9173,0.9194)	0.7358	(0.7348,0.7368)	0.6823	(0.6816,0.6829)	0.6651	(0.6623,0.668)
	CC	0.8558	(0.8536,0.858)	0.8999	(0.8978,0.902)	0.7255	(0.7241,0.7269)	0.6742	(0.6732,0.6753)	0.6765	(0.6739,0.6792)
	DBR	0.8117	(0.8096,0.8139)	0.8477	(0.8456,0.8497)	0.6960	(0.6946,0.6973)	0.6515	(0.6504,0.6525)	0.6588	(0.6561,0.6614)
	NST	0.8746	(0.8731,0.876)	0.9187	(0.9177,0.9197)	0.7360	(0.7351,0.737)	0.6824	(0.6817,0.6831)	0.6662	(0.6633,0.6691)
	RF	0.9227	(0.9218,0.9237)	0.9423	(0.9415,0.9431)	0.7493	(0.7484,0.7503)	0.6927	(0.6921,0.6933)	0.7238	(0.7213,0.7263)
	STA	0.8399	(0.8381,0.8417)	0.8982	(0.897,0.8995)	0.7245	(0.7235,0.7255)	0.6735	(0.6728,0.6742)	0.6631	(0.6603,0.6659)

AUC, Area Under the Curve; SE; sensibilidad, SP; especificidad; HE, hipoecogeneidad; NV, neovascularización; RIS, rotura intrasustancia; E, entesopatía; BR, binary relevance model; CC, classifier chains model; NST, nested stacking model; DBR, dependent binary relevance model; STA, staking generalization; RF, random forest.

5.4.4. Rendimiento diagnóstico

La Figura 15 ilustra el conjunto de datos A, y los resultados evidencian la relación entre la sensibilidad y la 1-especificidad para cada hallazgo degenerativo mediante el modelo de RF. En este gráfico, se observa una mayor capacidad discriminante en la detección del diagnóstico, ya que la mayoría de las líneas tienden a situarse más cerca de la esquina superior izquierda en el espacio ROC (característica operativa del receptor). La rotura intrasustancia destaca por su capacidad discriminante significativa en comparación con otras lesiones en el tendón. No obstante, el hallazgo de entesopatía exhibe la capacidad discriminante más reducida en este modelo.

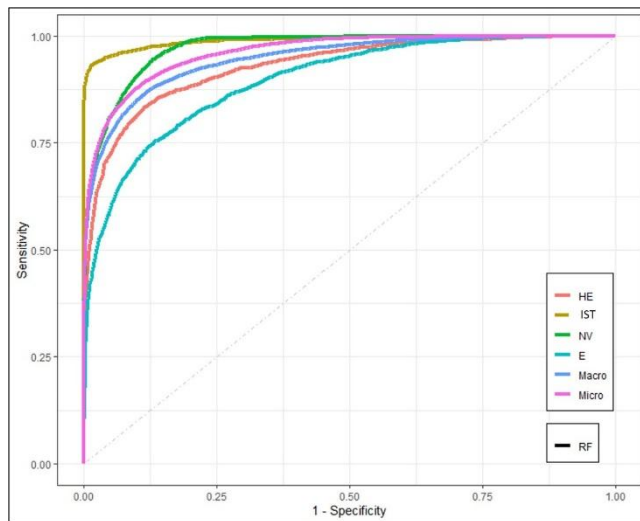


Figura 15. Curva ROC para RF y dataset A.

VI – DISCUSIÓN

6. DISCUSIÓN

6.1. ESTUDIO N1

Los primeros registros de procesamiento de imágenes utilizando técnicas de programación datan de la década de 1960 (119). No obstante, el uso de aprendizaje automático para elementos diagnósticos, especialmente en análisis supervisados donde los usuarios definen parámetros y características basadas en criterios de expertos, se hizo más frecuente aproximadamente veinte años después. Entre los algoritmos más empleados se incluían el análisis de componentes principales, las máquinas de vectores de soporte y las primeras redes neuronales convolucionales (120).

El metaanálisis desarrollado demostró que la utilización de diversos modelos de aprendizaje automático presentó una alta precisión diagnóstica para detectar patologías musculoesqueléticas en la extremidad superior a través de diversas modalidades de imagen. La sensibilidad combinada de todos los artículos analizados en esta revisión fue del 93%, y la especificidad combinada fue del 91%, indicando su capacidad para identificar correctamente los casos positivos y excluir los negativos. Por su parte, el análisis de subgrupos reveló odds ratios elevados, sugiriendo que la aplicación de estas estrategias para el diagnóstico de patologías musculoesqueléticas en varias situaciones de imágenes médicas demuestra una alta precisión diagnóstica.

Es notable que dentro del Grupo 0, un único artículo presenta tres modelos diferentes para el diagnóstico, lo que justifica una mayor atención y la necesidad de más investigación en esta área. A pesar del aumento en los últimos 5 años de publicaciones científicas que utilizan diversas herramientas de aprendizaje automático, tanto para el desarrollo de la investigación como para la aplicabilidad de modelos que permiten la identificación de elementos diagnósticos en imágenes biomédicas (121).

Adicionalmente, es relevante señalar que aún no se ha generalizado el uso de estas técnicas en el campo musculoesquelético, quedando rezagado en comparación con disciplinas como oncología, neumología y la medicina

cardiovascular (122). Por otro lado, se destaca que se publicaron más de 3500 artículos relacionados con el área de interés de este metaanálisis. Sin embargo, la falta de elementos estandarizados en la presentación de resultados de estudios que utilizan herramientas de aprendizaje automático para diagnósticos por imagen dificulta el análisis de este tipo de estudios. Por lo tanto, surge la necesidad de directrices que orienten la estructura metodológica para informar futuras publicaciones que utilicen elementos de aprendizaje automático para el apoyo diagnóstico en imágenes biomédicas.

Recientemente se ha publicado un marco guía para informar estadísticas en estudios biomédicos sobre aprendizaje automático e inteligencia artificial, lo que contribuirá eventualmente a mejorar la estandarización de estos informes en el campo musculoesquelético. En este contexto, resulta especialmente importante que los clínicos adopten estándares para la evaluación de calidad al implementar la inteligencia artificial en estudios diagnósticos. Un informe reciente destacó una adopción incompleta de herramientas en el diagnóstico (123).

Estudios de precisión basados en el uso de algoritmos señalan inconsistencias en todos los ámbitos de evaluación de calidad, convirtiéndose en obstáculos para la implementación clínica (124). Simultáneamente, se han propuesto iniciativas globales para el desarrollo de informes estandarizados para el uso de la inteligencia artificial en la atención médica, subrayando la necesidad de avanzar en estos temas multidisciplinarios (125). Dada la considerable diversidad en los informes de conjuntos de datos existentes, resulta imperativo establecer estándares para su presentación en el ámbito de la salud, especialmente en aplicaciones que dependen del uso de la inteligencia artificial (126).

Una fortaleza significativa de este estudio radica en evidenciar un alto nivel de precisión en la detección y clasificación de condiciones musculoesqueléticas a partir de diversas fuentes de imágenes. Considerando la abundancia de publicaciones encontradas, se sugiere que esta colaboración entre humanos y máquinas para mejorar la precisión diagnóstica en imágenes seguirá evolucionando con el tiempo (127). Además, las métricas diagnósticas elevadas identificadas en esta propuesta sugieren firmemente que el aprendizaje automático es una herramienta eficaz para identificar patrones complejos de lesiones que a veces pueden pasar desapercibidos para ojos expertos. Esto significa que existirá un impacto positivo en la salud pública al proporcionar más elementos para el

diagnóstico y, en última instancia, mejorar la toma de decisiones, especialmente en escenarios con alta incertidumbre diagnóstica, como las patologías musculoesqueléticas (126).

Hay algunas limitaciones en este estudio. En primer lugar, se destaca la alta heterogeneidad entre los estudios evaluados, atribuible no solo al hecho de que los artículos seleccionados no siguieron una estructura metodológica uniforme para la publicación, sino también a la diversidad en el origen de las publicaciones. Cada región puede tener estrategias específicas en el desarrollo de sus avances científicos en este campo. Por esta razón, para minimizar esta alta variabilidad, se optó por realizar los análisis utilizando un modelo de efectos aleatorios (128).

En segundo lugar, los algoritmos de inteligencia artificial analizados fueron diversos, lo que podría conducir a mayores variaciones en los resultados (62). Además, se incluyeron diversas patologías, cada una con criterios de diagnóstico marcadamente diferentes, lo que podría introducir errores en los conjuntos de datos. A su vez, los métodos de referencia no siempre fueron el estándar de oro ideal, que suele ser la intervención quirúrgica. La resonancia magnética se consideró un método de validación importante para los diagnósticos.

Por esta razón, analizar múltiples modelos de inteligencia artificial en patologías musculoesqueléticas tiene varias implicaciones (129). Por ejemplo, lograr una mayor precisión diagnóstica es posible mediante el uso exclusivo de redes neuronales artificiales en lugar de modelos de regresión logística tradicionales (130). La revolución en el diagnóstico musculoesquelético, facilitada por la implementación de algoritmos complejos, sentará las bases para una transición gradual hacia la medicina de precisión. Este enfoque transformador convertirá diversas manifestaciones clínicas en diagnósticos personalizados (129). Además, será esencial contar con conjuntos de datos representativos que abarquen todas las posibles condiciones musculoesqueléticas bajo estudio para garantizar análisis precisos y mitigar sesgos diagnósticos. Por lo tanto, es crucial tener en cuenta las etapas de recopilación de datos, preprocesamiento, desarrollo de modelos, validación e implementación integral (131).

Los profesionales clínicos especializados en el diagnóstico musculoesquelético deben recibir capacitación para comprender cómo los modelos toman decisiones. Esta comprensión es crucial para su colaboración efectiva dentro de equipos interdisciplinarios (132). Por lo tanto, para promover eficazmente la

investigación continua, un plan integral debería enfatizar el desarrollo de modelos específicos para abordar nuevas patologías o mejorar las existentes. Este enfoque impulsará la innovación sostenida en la intersección de la inteligencia artificial y el campo musculoesquelético (133).

En tercer lugar, y finalmente, la estrategia de búsqueda y los criterios de selección de información fueron bastante estrictos para permitir un análisis sistemático, lo que limitó la adquisición de más artículos que presentaran lesiones musculoesqueléticas que no cumplieran con nuestros parámetros de búsqueda. Por ejemplo, solo se consideraron artículos que informaban métricas diagnósticas como VP, FP, FN, VN, lo que redujo significativamente el número de artículos finales para el análisis. Además, considerar los modelos o algoritmos de inteligencia artificial informados como unidades de análisis en lugar de los propios artículos podría introducir un sesgo de selección. Por ende, a pesar de los resultados alentadores de este metaanálisis, que afirman que las herramientas derivadas de la inteligencia artificial para el apoyo diagnóstico en imágenes biomédicas son positivas, deben ser consideradas e interpretadas con moderación por diferentes expertos en el campo.

Múltiples desafíos surgen en la implementación de modelos de inteligencia artificial para el diagnóstico de patologías musculoesqueléticas. Estos desafíos incluyen la complejidad de los datos clínicos asociados con estas condiciones, la variabilidad inter e intraoperador en la interpretación del mismo resultado de imagen, la integración efectiva de modelos en la práctica clínica y la necesidad de validar externamente los resultados del modelo. Sin embargo, se proponen algunas soluciones estratégicas, como la integración avanzada de técnicas de procesamiento y reconocimiento de datos con una mayor cantidad y calidad de información, proporcionando herramientas de presentación más accesibles generadas por los modelos, facilitando la interoperabilidad con sistemas de registros médicos electrónicos y, en última instancia, realizando un mayor número de estudios en diversos entornos clínicos musculoesqueléticos para mejorar la generalización de los resultados.

6.2. ESTUDIO N2

La presencia de tendinopatías es un problema frecuente tanto para la población general como para los atletas. Es importante destacar que la TLC es el trastorno musculoesquelético más diagnosticado en el codo y aún representa un problema no resuelto para la salud pública (80).

Varios estudios han abordado los cambios estructurales que se encuentran exclusivamente en la TLC, como el engrosamiento del tejido, áreas hipoeoicas, zonas de neovascularización, interrupción fibrilar y anormalidades óseas (7). Como se demostró en esta investigación, es poco frecuente que los hallazgos degenerativos estén presentes de manera aislada e independiente. Este estudio permitió determinar la presencia de cuatro grandes grupos de hallazgos degenerativos y la presencia de lesiones graduales, que pueden indicar la existencia de desgarros dentro del tendón.

En una muestra amplia de ecografías obtenidas durante dos años, este análisis indica que la edad y los hallazgos degenerativos del tendón fueron los factores más críticos asociados con un aumento en el desgarro del tendón en cualquier dimensión. En particular, se encontró una relación positiva entre la presencia de niveles de hipoeogenicidad y la longitud del desgarro, así como entre los niveles de neovascularización y el ancho del desgarro (59).

Este resultado sugiere que, si el objetivo es reducir la magnitud y la variabilidad del desgarro del tendón en pacientes con TLC, la hipoeogenicidad y la neovascularización deberían influir en la importancia de los hallazgos degenerativos, de modo que las ecografías preventivas podrían ser útiles para evitar posibles desgarros del tendón (134). Además, esta información puede utilizarse para evaluar la gravedad de la tendinopatía en el codo lateral. Por ejemplo, los investigadores concluyeron que el tamaño de la rotura intrasustancia en una ecografía podría determinar qué pacientes responden al tratamiento no operatorio y la correlación con la discapacidad funcional (135). Sin embargo, la descripción del desgarro del tendón no incluyó la magnitud de la dimensión del desgarro, así como los hallazgos de hipoeogenicidad y neovascularización. Por lo tanto, es esencial contar con esta información en futuros estudios.

Por otro lado, existen varias explicaciones posibles para la asociación positiva de la hipoeogenicidad y la neovascularización con la dimensión del desgarro del

tendón. En primer lugar, el ERCC es el tendón central involucrado en un proceso degenerativo de TLC, y en este estudio solo se consideró aquellas descripciones exhaustivas y el análisis de los hallazgos degenerativos en el interior de este tendón. No obstante, la mayoría de las publicaciones existentes describen la relación entre la TLC y diferentes aspectos anatómicos u otros tejidos adyacentes como nervios, ligamentos, músculos, estructuras óseas o más tendones (136) e incluso otras condiciones de salud que no son únicas para la alteración del origen de los extensores de la muñeca.

En segundo lugar, algunos investigadores han propuesto que el modelo de patogénesis del tendón es continuo, argumentando que hay una transición de 3 estados desde una etapa inicial caracterizada por tendinopatía reactiva, una segunda etapa disruptiva y, finalmente, una tercera etapa degenerativa (137). Esta última etapa también presenta graduación, especialmente con algunos hallazgos degenerativos, ya que pueden subdividirse según su grado de compromiso. Además, la existencia de las asociaciones demostradas podría respaldar futuros modelos predictivos que determinen si la presencia de un hallazgo anticipa la aparición de otro.

Recientemente, un estudio mostró que la TLC afecta principalmente a personas de mediana edad (7) y que en el sexo femenino es mayor la probabilidad de padecerla que en el masculino (138). El presente estudio demuestra que la edad está fuertemente asociada con el tamaño del desgarro del tendón, por lo que cuanto mayor es el paciente, mayor es la magnitud del desgarro independientemente de la dimensión evaluada. Los resultados confirman que los hombres desarrollan TLC casi un año antes que las mujeres (59). Es decir, la importancia clínica radica en identificar las características estructurales de los tendones mediante ecografía.

Algunos autores incluso proponen la realización de exámenes de US incluso en sujetos que no presentan síntomas, pero podrían tener hallazgos degenerativos (139). Como se pudo observar en el presente estudio, los hallazgos degenerativos podrían presentar graduaciones antes del desgarro del tendón, por lo que la identificación temprana de estos hallazgos permitiría el monitoreo de futuros desgarros del tendón y podría mejorar la capacidad de resolución terapéutica. Es importante contar con futuras publicaciones que asignen un factor temporal a los hallazgos degenerativos para comprender la evolución natural de dicho desgarro del tendón (59).

La principal limitación del estudio fue el diseño retrospectivo, que a menudo se considera inferior a los ensayos clínicos prospectivos, aleatorios y controlados (140). Además, no se pudo controlar el procedimiento e imagen, y no se consideró la confiabilidad inter e intraevaluador, perdiendo validez externa.

Otras limitaciones incluyeron la evaluación solo del tendón ERCC, dificultando la generalización de los resultados. A pesar de esto, este estudio se posiciona como uno con el mayor número de exámenes analizados hasta la fecha. Una limitación adicional fue que las imágenes corresponden a pacientes con diferentes tratamientos conservadores, puntajes de dolor, tiempos de evolución, niveles de actividad física, comorbilidades y manifestaciones clínicas. Este estudio se centró únicamente en las características de los exámenes de ultrasonido, sin incorporar estos elementos. Futuros estudios deberían considerar un diseño de cohorte para evaluar estas consideraciones.

6.3. ESTUDIO N3

El propósito principal de este estudio fue evaluar la exactitud diagnóstica de una prueba índice para detectar una rotura intrasustancia del tendón confirmado por ecografía en pacientes TLC. El objetivo secundario fue sugerir que el dolor nocturno también presenta valores similares a la prueba índice. Dos cirujanos ortopédicos llevaron a cabo la evaluación. Se aplicó una nueva maniobra física única debido a la experiencia clínica diagnóstica de los profesionales, sumado con registros médicos electrónicos y ecografías para confirmar una TLC con la presencia de un RIS.

Los resultados de este estudio indicaron que casi el 90% de los pacientes fueron diagnosticados correctamente cuando la prueba índice fue positiva y se confirmó la presencia de RIS mediante ecografía. Este resultado podría explicarse potencialmente por la aguda percepción clínica y la experiencia de los profesionales especialistas. Además, los datos mostraron una fuerte evidencia en la proporción de sujetos correctamente clasificados entre todos los sujetos, también conocida como precisión diagnóstica (eficacia), con un valor cercano al 80%, clasificado como muy bueno.

También hay cierta evidencia de que la prueba índice tenía una capacidad particular para detectar a los pacientes que no tenían la enfermedad cuando la prueba fue negativa, conocida como especificidad. Este valor fue más bajo que los otros dos, y esto se podría explicar porque la especificidad es la probabilidad de que la prueba clasifique correctamente a individuos sanos o la probabilidad de que los individuos sanos obtengan un resultado negativo. En este caso, todos los pacientes ya habían sido diagnosticados previamente con TLC.

Como se mencionó, el diagnóstico de TLC se confirma mediante diferentes estrategias clínicas e imagenológicas. Uno de los más utilizados en la práctica clínica es la reproducción de maniobras físicas provocativas para esta condición musculoesquelética, como la prueba de Cozen, la prueba de Mill, la prueba de Maudsley y la prueba de Polk (141). Aunque estas pruebas han sido ampliamente descritas en la literatura, numerosas pruebas clínicas para la TLC no acompañan la condición en cuanto a precisión diagnóstica (142). De hecho, la capacidad diagnóstica de estas pruebas clínicas de provocación para confirmar la TLC está insuficientemente investigada (143).

A pesar de esto, la prueba de Cozen es la única recomendada internacionalmente por la *Health Safety Executive* del Reino Unido para diagnosticar la TLC (144). Por esta razón, los autores exploraron el rendimiento de la prueba de Cozen en la misma muestra de pacientes. Sorprendentemente, se encontró que los valores más bajos de sensibilidad, especificidad y precisión. Esto podría sugerir que la prueba de Cozen no es la mejor maniobra física en pacientes con sospecha de RIS.

Otros autores, han estudiado la reproducción del dolor durante las actividades diarias para el diagnóstico de TLC. Desde la década de 1970, levantar una silla con la mano pronada y una prueba modificada de recogida de silla es una de las pruebas más utilizadas para diagnosticar TLC (145). Investigadores determinaron recientemente que la posición del miembro superior afecta la fuerza de agarre sin dolor en TLC (146). Sin embargo, la mayoría de estas pruebas de diagnóstico tienen en común que no discriminan según el tipo de hallazgo ecográfico, como un RIS. Solo realizan el diagnóstico al confirmar la presencia o ausencia de la condición de TLC. Por esta razón, detectar la presencia o ausencia de la condición no es suficiente. Detectar la RIS es crucial para determinar los mejores resultados terapéuticos en pacientes con TLC. El desgarro del tendón es la fase final del proceso degenerativo del tejido, y el número de hallazgos ecográficos podría determinar la gravedad de la tendinopatía (59).

Por otro lado, el dolor intenso puede causar trastornos del sueño (7). Según la experiencia registrada en este artículo, el dolor nocturno es típico en pacientes con desgarros del tendón. La sección de resultados demostró una mayor sensibilidad con este signo clínico, pero una especificidad moderada. Hasta la fecha, este es el único estudio en la literatura reciente que intenta identificar una maniobra física y un signo nocturno de forma independiente con la presencia de una RIS del tendón diagnosticado por ecografía musculoesquelética en pacientes con TLC.

Este estudio tiene algunas limitaciones. En primer lugar, fue un diseño retrospectivo, que a menudo se considera inferior a los ensayos clínicos prospectivos, aleatorios y controlados (140). No se consideraron la confiabilidad inter e intraevaluador entre radiólogos y cirujanos ortopédicos, perdiendo validez externa.

Otras limitaciones incluyen la confirmación del diagnóstico de RIS solo mediante ecografía, sin incluir otros hallazgos imagenológicos, dificultando la generalización de los resultados. Una limitación adicional fue que este estudio se centró únicamente en la presencia de RIS y el signo positivo de la prueba índice y el dolor nocturno, pero no se incluyeron puntuaciones de dolor, tiempos de evolución, niveles de actividad física, comorbilidades y otras manifestaciones clínicas. Finalmente, los resultados binarios considerados no permitieron análisis estadísticos más complejos.

Los estudios futuros requerirán reproducibilidad y la implementación de la nueva prueba. Se debe considerar un diseño de cohorte con un tamaño de muestra más grande, confiabilidad inter e intraevaluador entre profesionales y variables continuas para el análisis.

6.4. ESTUDIO N4

Este estudio representa una de las primeras contribuciones en introducir modelos de clasificación multietiqueta mediante algoritmos de aprendizaje automático para identificar hallazgos degenerativos y RIS en imágenes de ultrasonido con diagnóstico de TLC. En este análisis retrospectivo, se tomó en cuenta específicamente una de las series más extensas de imágenes de US del tendón ERCC, y la herramienta de diagnóstico de TLC basada en aprendizaje automático se entrenó con el conjunto de datos más amplio hasta la fecha.

Los resultados más notables de este estudio se lograron mediante la incorporación de varios modelos de aprendizaje automático basados en el conocimiento diagnóstico. Se observaron resultados sobresalientes y valores más altos para todos los hallazgos degenerativos en el rendimiento de la clasificación binaria. Además, cuando el diagnóstico de US se fundamentó en la combinación de hallazgos degenerativos mediante un clasificador multietiqueta, se obtuvieron valores de precisión sólidos.

Los resultados obtenidos indican que el algoritmo de RF exhibió el mejor rendimiento diagnóstico, tanto en modelos binarios como multietiqueta. Estos hallazgos respaldan la idea de que la implementación de herramientas derivadas de la IA puede ser beneficiosa para respaldar la imagenología de las tendinopatías. La colaboración entre el radiólogo y el algoritmo podría mejorar la precisión de los resultados, especialmente en instituciones sin radiólogos especializados en el área musculoesquelética.

Tradicionalmente, el US se ha reconocido como una herramienta rentable para detectar patrones anormales en las estructuras de los tendones. Además, hay evidencia que respalda su utilidad en la detección de TLC. Un metaanálisis publicado en 2014 determinó que la precisión del diagnóstico parece depender en gran medida de variables como la experiencia del operador, el equipo y la etapa de la patología. Sin embargo, el ultrasonido muestra sensibilidad y especificidad variables (sensibilidad: 64–100%; especificidad: 36–100%), lo que disminuye la precisión del diagnóstico clínico (56). Otro artículo publicado en el mismo año informó específicamente sobre la sensibilidad y especificidad para cada hallazgo anormal de ultrasonido utilizando el método de detección tradicional. La hipoecogenicidad presentó la mejor combinación de sensibilidad y especificidad

diagnósticas. Tiene una sensibilidad moderada (0.64 [IC del 95%, 0.56, 0.72]) y una especificidad alta (0.82 [IC del 95%, 0.72, 0.90]). Además, la neovascularización (especificidad: 1.00 [IC del 95%, 0.97, 1.00]), calcificaciones (especificidad: 0.97 [IC del 95%, 0.94, 0.99]) e irregularidades corticales (especificidad: 0.96 [IC del 95%, 0.88, 0.99]) muestran una fuerte especificidad para la TLC crónica (57).

Los resultados obtenidos, especialmente para la detección de RIS mediante el algoritmo de clasificación binaria en ambos conjuntos de datos, exhibieron un rendimiento superior a los métodos de diagnóstico tradicionales de US. En cuanto a la precisión multietiqueta, el rendimiento para ambos indicadores fue menor que los resultados del método binario. Esta situación podría explicarse porque resulta difícil encontrar una función que minimice el error para más clases, es decir, aumenta la variabilidad de la variable de respuesta.

Por ejemplo, en la clasificación binaria, la entesopatía mostró el rendimiento más bajo entre los seis clasificadores de aprendizaje automático. Específicamente, en el modelo de relevancia binaria dependiente del conjunto de datos B, nuestro análisis reveló que la AUC fue de 0.647 [IC del 95%, 0.64, 0.65]. Este resultado es bastante similar a otros informes que indican una sensibilidad del 0.65 y una especificidad del 0.86 para este hallazgo (112). No obstante, los resultados en la clasificación binaria para la detección de lesiones por una RIS mediante algoritmos de bosque aleatorio. El rendimiento mostró un AUC casi de 1.0 (0.99) [IC del 95%, 0.99, 0.99], en contraste con los métodos de diagnóstico tradicionales de ultrasonido para detectar desgarros en el tendón extensor común en el lateral, que mostraron rendimientos más bajos en sensibilidad, especificidad y precisión del 64.52, 85.19 y 72.73%, respectivamente (102).

Una de las fortalezas de esta investigación radica en la aplicación de modelos de aprendizaje automático mediante la detección multietiqueta para hallazgos de lesiones en tendones. Hasta la fecha, se habían publicado pocas experiencias en el área musculoesquelética que utilizaran inteligencia artificial para la detección de patrones en tendones. Algunas experiencias previas han empleado la Detección y Clasificación Automáticas de la Región de Interés de Imágenes de Ultrasonido del Tendón de Aquiles (147), y modelos de aprendizaje profundo para el seguimiento automático de la unión músculo-tendón o incluso la medición de la atrofia muscular (76).

Otras disciplinas también han utilizado otras técnicas de clasificación, como redes neuronales o redes neuronales convolucionales de aprendizaje profundo para la detección de imágenes, demostrando resultados excelentes. Sin embargo, las CNN y DL tienen algunas desventajas que se deben analizar al desarrollar modelos predictivos. En primer lugar, se ha demostrado que DL requiere grandes conjuntos de datos para obtener un mejor rendimiento. Para abordar esto, es comúnmente utilizado el aprendizaje por transferencia. No obstante, las arquitecturas de DL también deben ser reentrenadas y los parámetros del modelo optimizados, prestando atención a posibles patrones de sobreajuste. En segundo lugar, las arquitecturas de DL dependen de un alto rendimiento computacional, y lleva más tiempo probar los resultados. En este sentido, son más complejas de implementar, especialmente en un entorno clínico con una alta demanda de atención, por lo que mejorar la velocidad de diagnóstico sin comprometer la precisión del diagnóstico es crucial tanto para los pacientes como para el sistema de salud. Por lo tanto, los algoritmos de aprendizaje automático son ventajosos cuando la velocidad es un factor crítico. En este caso, los tiempos de ejecución del método propuesto fueron muy bajos, lo que permite su implementación fácil en un entorno hospitalario y su reentrenamiento con nuevos datos generados diariamente. Finalmente, el modelo de clasificación multietiqueta difiere de otros algoritmos más comúnmente utilizados en el diagnóstico por imagen debido a la simplicidad de su implementación.

Este estudio también presenta algunas limitaciones. En primer lugar, las imágenes analizadas provienen de la misma institución, y los pacientes compartían condiciones socioeconómicas similares. En segundo lugar, incluimos todas las imágenes estáticas de US del tendón extensor común por paciente, sin considerar la visualización en tiempo real ni otras estructuras o tejidos. En tercer lugar, incluimos tendones con un diagnóstico definitivo de TLC, y no se comparó la variabilidad inter e intraobservador entre radiólogos. En cuarto lugar, se consideró todas las imágenes sin una región de interés, siguiendo la práctica común en la mayoría de las publicaciones. En un futuro cercano, esto podría ser una ventaja potencial. Finalmente, no se repitió el diagnóstico de US para reducir el sesgo retrospectivo. No obstante, el médico radiólogo contaba con más de 10 años de experiencia en disfunciones musculoesqueléticas.

6.5. DESAFÍOS Y OPORTUNIDADES PARA INVESTIGACIONES FUTURAS DE MODELOS DE IA EN SALUD MUSCULOESQUELÉTICA

En la siguiente sección ampliamos la discusión a diversas oportunidades para el desarrollo de nuevas investigaciones, fundamentadas en las limitaciones y alcances de la presente tesis. La identificación de espacios de mejora permite establecer líneas futuras de exploración que podrán complementar y expandir los hallazgos ya obtenidos. En particular, la aplicación de metodologías alternativas podría contribuir en la optimización de la precisión y validez de los resultados en múltiples escenarios hospitalarios.

Validación externa

El autor reconoce la importancia y el valor indiscutible que representa el proceso de evaluación de los modelos implementados, cuyo propósito es garantizar la aplicabilidad clínica de los resultados obtenidos en diferentes escenarios clínicos. En este sentido, esta tesis incorporó exitosamente dos estrategias de validación trascendentales. En primer lugar, se ejecutó un riguroso proceso de validación interna, asegurando que los modelos implementados aprendieran correctamente sin sobreajustarse. Luego, mediante la aplicación de una estrategia de validación cruzada se evaluó el óptimo desempeño de los modelos utilizados en distintos conjuntos de datos, asegurando así su robustez.

Por otro lado, se reconoce que no se implementó directamente un proceso de validación externa, a pesar de que constituye un paso esencial en la generalización de los resultados (148). Su ejecución no fue posible debido a diversos desafíos metodológicos y operacionales que van más allá del alcance del presente estudio. La validación de nuevos sets de imágenes independientes, requieren una serie de recursos humanos y de infraestructura tecnológica significativos como el desarrollo de nuevos protocolos de investigación para ser aprobados por la autoridad sanitaria a través del comité de ética científico, una inversión adicional de más de 1500 horas-hombre en la descarga de las imágenes similar a la que fue utilizada en esta tesis, así como la disponibilidad de radiólogos subespecialistas que distingan el correcto etiquetado de los diagnósticos obtenidos, y por supuesto, nuevos

recursos computacionales para la realización del procesamiento de los conjuntos de datos.

Todos los aspectos mencionados anteriormente, constituyen un proyecto independiente de investigación en sí mismo, se valora profundamente la importancia de esta etapa y se considera que su implementación constituye una extensión natural y necesaria de esta investigación en un futuro cercano para consolidar los resultados obtenidos, así como validar la información en diferentes escenarios clínicos similares al presentado.

Calidad de las imágenes ultrasonográficas

Otro aspecto fundamental a considerar es la calidad de las imágenes ultrasonográficas obtenidas. En este sentido, se destaca que todas las ecografías fueron capturadas entre 2017 y 2018 utilizando un ecógrafo de alta resolución, el Aplio 500 US System de Toshiba America Medical Systems, actualmente Canon Medical Systems Corporation. Este equipo es reconocido como una herramienta de alta gama en el procesamiento de imágenes para estudios radiográficos musculoesqueléticos. Entre sus principales ventajas, el Aplio 500 destaca por su capacidad para reducir el ruido en las imágenes, mejorando la nitidez y definición de las estructuras anatómicas. Asimismo, su avanzada tecnología Doppler permite una detección precisa del flujo sanguíneo, optimizando el análisis vascular (59).

Si bien existen equipos con resoluciones y capacidades de análisis superiores, como el Aplio i800 o el Aplio i900, el uso del Aplio 500 garantiza un alto nivel de precisión en los modelos. Además, dado que muchos centros clínicos aún cuentan con equipos similares, los resultados expuestos son comparables y replicables en otros entornos. En caso de utilizar ecógrafos más avanzados, la calidad de las imágenes podría incluso potenciar aún más la solidez de los hallazgos alcanzados.

En la actualidad, no existe un protocolo estandarizado para la obtención de imágenes ecográficas, ya que su adquisición depende directamente del profesional que la realiza. Esto significa que la calidad y características de las imágenes pueden variar según su nivel de formación, experiencia, destreza técnica y carga de trabajo, lo que genera múltiples enfoques para el análisis de una misma estructura (149).

Sin embargo, lejos de representar una limitación, esta diversidad constituye una fortaleza del estudio, ya que permite explorar nuevas estrategias diagnósticas para la detección de patrones ultrasonográficos. Además, otorga un valor adicional al trabajo, considerando que la mayoría de los estudios en radiología musculoesquelética se basan en imágenes obtenidas de manera estandarizada, sin la intervención humana, como ocurre en la radiografía o la RMN (150).

En este contexto, el principal desafío de esta tesis radicó en la estandarización de una región de interés en imágenes ecográficas, garantizando la detección precisa de las zonas que contenían los hallazgos y permitiendo capturar con fidelidad la naturaleza del problema, lo que convirtió este proceso en una iniciativa pionera dentro de la disciplina debido a su complejidad. Para abordar este reto, se implementaron diversas estrategias, entre ellas el preprocesamiento de imágenes para eliminar ruido asociado a iluminación, homogeneización de tamaños y posiciones, así como la normalización de datos para asegurar su uniformidad (151). La zona de interés se centró en la región anatómica del tendón extensor común, utilizando como referencia anatómica el epicóndilo, en lugar de considerar lesiones específicas. Se aplicó un proceso de extracción de características basado en el análisis de texturas, parámetros, intensidades, distribuciones y relaciones entre píxeles (152). Además, se llevó a cabo un extenso etiquetado manual de imágenes para la clasificación de hallazgos degenerativos y se integraron datos clínicos y de imagen en dos conjuntos diferenciados: el conjunto A, basado en predicciones a partir de imágenes, y el conjunto B, centrado en predicciones a partir de los exámenes.

Este aspecto sigue representando un gran desafío en el procesamiento automático de imágenes musculoesqueléticas, lo que motivó al autor de esta tesis en conjunto con un equipo de trabajo a seguir desarrollando nuevas iniciativas de investigación. Estas culminaron exitosamente en publicaciones adicionales donde se logró identificar regiones de interés en lesiones musculoesqueléticas, segmentándolas y clasificándolas automáticamente en distintos tipos de imágenes (153, 154). Estos avances científicos han servido como referencia para continuar construyendo evidencia para posteriores análisis ecográficos en el codo.

Explicabilidad de los modelos

A diferencia de las redes neuronales profundas, cuya interpretación suele ser más compleja, en esta tesis se emplearon estructuras de ML como binary relevance model, classifier chains model, nested stacking model, dependent binary relevance model, stacking generalization y RF, los cuales presentan un mayor nivel de interpretabilidad (155–158). Sobresale el desempeño obtenido en los resultados por RF el cual permite evaluar la importancia de cada variable en la toma de decisiones a través de métricas como la disminución de impureza (índice Gini). De hecho, al ser una sumatoria de árboles de decisión su explicabilidad posee la capacidad de ser trazable, entregando una mayor transparencia en su comportamiento evitando el efecto de caja negra. De manera similar, enfoques como classifier chains y nested stacking ofrecen una clasificación jerárquica que, aunque introduce cierta dependencia entre variables, sigue siendo más comprensible que los modelos basados en aprendizaje profundo.

Para abordar la problemática de la interpretabilidad en ML, recientemente se han desarrollado técnicas como SHapley Additive Explanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME) y Partial Dependence Plots (PDPs), que permiten analizar en detalle la influencia de cada variable en la predicción final (159). Al implementar estas herramientas, es posible visualizar la contribución individual de cada característica al resultado del modelo, facilitando su interpretación en el ámbito clínico y mejorando su aceptación por parte de los especialistas. A pesar de que se reconoce su utilidad, estos instrumentos no fueron implementados porque quedaban fuera de los alcances de los objetivos desarrollados a lo largo de esta tesis.

En los últimos años, la comunidad clínica y científica ha incorporado técnicas avanzadas, como el aprendizaje profundo con redes neuronales, para el análisis de imágenes médicas para patologías musculoesqueléticas (150). No obstante, esta tesis optó por el uso de modelos más simples, priorizando la interpretabilidad y la eficiencia computacional. Esta elección no solo facilita su comprensión, sino que también mejora su replicabilidad en entornos con recursos limitados, favoreciendo su aplicación en la práctica médica. Futuras investigaciones podrían enfocarse en mejorar la transparencia de estos algoritmos para fortalecer la confianza en su uso. Aunque pueden alcanzar una alta precisión

diagnóstica, su menor grado de explicabilidad puede generar dudas entre los especialistas, limitando su adopción.

Es importante recordar que el diagnóstico no depende exclusivamente de datos e imágenes, sino que requiere la evaluación clínica del paciente con una participación activa del equipo de salud (160). Por este motivo, a pesar de las estrategias para reducir el sesgo en los modelos de ML, este no puede eliminarse por completo. Por ello, estas herramientas deben ser creadas, validadas y supervisadas en conjunto por profesionales clínicos, utilizándose como apoyo en la toma de decisiones. La interacción entre especialistas y tecnología debe ser una colaboración estratégica, donde el modelo aporte su capacidad analítica sin reemplazar el juicio humano.

Se destaca sobresalientemente que esta iniciativa fue implementada desde su origen como una colaboración en conjunto con un equipo multidisciplinario. Participaron diversos actores del equipo de salud musculoesquelética, pensando en los desafíos que impone una transformación digital en salud. Se incluyeron radiólogos, traumatólogos, especialistas musculoesqueléticos, ingenieros y técnicos informáticos, permitiendo abordar de manera integral la implementación de estas nuevas tecnologías. A su vez, para lograr una confianza de los especialistas en el rendimiento de los modelos, se debe involucrar desde la fase inicial del desarrollo hasta la fase final de implementación. Esto permitiría no sólo comprender el funcionamiento y las limitaciones de las arquitecturas seleccionadas, sino que también contar con el apoyo con sus experiencias y entrenamiento para lograr un óptimo desempeño. Esta integración estratégica facilita la confianza en la tecnología, promoviendo una mayor aceptación y uso efectivo en la práctica profesional.

Finalmente, para mejorar la confianza de los miembros del equipo en estas nuevas herramientas diagnosticas resulta fundamental que los clínicos logren entender por qué el algoritmo entrega determinados resultados. En otras palabras, si el modelo reconoce la presencia de un hallazgo degenerativo y este es confirmado por el especialista, se convertirá en un testigo del valor de la herramienta generando credibilidad. Una estrategia podría ser la generación de explicabilidad visual, es decir, crear un mapeo visual de las regiones más susceptibles a presentar hallazgos degenerativos con el propósito que el especialista este monitorizando en tiempo real los alcances del instrumento aportando mayor confianza en el uso de

este. No obstante, el estudio presentado (108), no se centró en la explicabilidad visual, sino más bien en el uso de características predefinidas (textura, píxeles, formas), proporcionando un mayor nivel de transparencia metodológica.

Comparación con otros métodos diagnósticos

Tradicionalmente la TLC se diagnostica mediante un exhaustivo examen físico y una completa anamnesis por parte de un especialista en el área musculoesquelética. El uso de imágenes para su detección suele reservarse para aquellos casos más severos y sospechosos de alteraciones concomitantes. Recomendaciones recientes (59) sugieren solicitar siempre una ecografía cuando se sospeche de este diagnóstico médico a pesar de ser claramente identificable con el examen físico. Esto se explica debido a la multiplicidad de hallazgos ecográficos específicos existentes, permitiendo prescribir dirigidamente los recursos terapéuticos según el tipo de hallazgo predominante, optimizando los recursos y adelantando los tiempos de alta.

En términos generales, la detección ecográfica realizada por expertos radiólogos en la detección de la TLC es variada. La sensibilidad oscila entre el 72% y el 88% y la especificidad entre el 36% y el 48.5% (161). Otras publicaciones más recientes, han destacado el buen balance entre estas dos métricas en este instrumento llegando al 73%, reafirmando que no es necesario el uso de otras estrategias diagnósticas complementarias (162). Esto convierte el uso de la ecografía en un método fiable debido a su alta precisión en la detección de hallazgos patológicos y por su sobresaliente costo efectividad, disponibilidad inmediata y diagnóstico en tiempo real (163).

Por su parte, la RMN sigue siendo el gold standard para el diagnóstico de esta condición. Otras experiencias han comparado su rendimiento con la ecografía señalando que la sensibilidad es aún mayor para la detección de la TLC variando entre el 64% y el 82% para la ecografía y entre el 90% y el 100% en el caso de la RMN. La especificidad ha logrado mejores resultados alcanzado un 67% a un 100% para la ecografía y entre el 83% y el 100% para la RMN (164). Adicionalmente, su alto costo, baja disponibilidad, dificultad de acceso, exposición a radiación y escasez de especialistas refuerzan que el diagnóstico sea realizado mediante una

ecografía. Esto implica que el uso de la resonancia sólo se reserva para casos donde la ecografía no es concluyente.

En el caso del uso de radiografías para la identificación de epicondilitis no es utilizado, debido a que este instrumento no está diseñado para la detección de tejido blando como tendones o músculos, sino más bien para evaluar disposiciones articulares y/o estructuras óseas como fracturas o calcificaciones, restringiendo su uso para esta condición. Complementariamente se ha reportado su uso en epicondilitis para la detección de hallazgos crónicos compatibles con alteraciones estructurales óseas, no reportando indicadores de sensibilidad y especificidad (165).

En el caso del diagnóstico apoyado por algoritmos de IA, hasta la fecha y según el conocimiento del autor no existen reportes científicos que evalúen las capacidades diagnósticas de estos modelos en imágenes de RMN y radiografías en la TLC. En el caso de la RMN, seguramente por lo anteriormente expuesto y en el caso de la radiografía probablemente porque la señal de la patología en este instrumento es sutil y de baja sensibilidad. Se destaca que la única contribución científica hasta el momento ha sido el artículo científico que constituye parte de esta tesis con métricas diagnósticas variables dependiendo específicamente del tipo de hallazgo identificado y el tipo de modelo implementado. De igual forma, resultan ser mayores que la identificación realizada por los expertos (108).

A su vez, los resultados obtenidos en esta tesis impulsaron la oportunidad de avanzar en nuevos proyectos acerca del impacto de la IA en la detección de alteraciones específicas en el tendón, utilizando como criterio de inclusión algoritmos más complejos para la identificación de patrones y computacionalmente más demandantes como el aprendizaje profundo, incluyendo cualquier tipo de alteración del tendón e imagen médica. Se logró exitosamente una reciente publicación de un robusto metaanálisis el cual consideró seis artículos que aportaron con información acerca de 11 algoritmos para evaluar el desempeño diagnóstico de afecciones que involucran tendones utilizando RMN y ecografía dado el bajo desempeño de las radiografías. El modelo de efectos aleatorios ejecutado obtuvo una sensibilidad de 0,910 (IC del 95%: 0,865; 0,940), lo que demostró un alto nivel de precisión en la identificación correcta de casos positivos. En otras palabras, esto significa que los algoritmos evaluados son capaces de identificar efectivamente casos positivos en el 91% de los casos. Por su parte, se

logró encontrar un valor de especificidad de 0,954 (IC del 95%: 0,909; 0,977), lo que indicó una vez más una alta precisión en la identificación correcta de casos negativos. En otras palabras, los algoritmos pueden identificar correctamente casos negativos en el 95% de los casos o tener un error de estimación de solo el 5% (150).

Al comparar los resultados de los modelos implementados, se destaca que aún existen escasos estudios que utilicen tanto modelos de ML como otras arquitecturas más complejas en la detección de hallazgos degenerativos de tendón analizando imágenes ecográficas. Se debe comprender que estos elementos son un complemento importante para el desempeño diagnóstico de los especialistas. Por este motivo, radiólogos expertos podrán identificar con facilidad hallazgos degenerativos como hipoecogenicidad, roturas intrasustancias y calcificaciones en pacientes con TLC, pero aún existen desafíos en la estandarización de la variabilidad interobservador. En este contexto, los modelos de la IA deberán de seguir utilizándose como apoyo al diagnóstico, buscando igualar o superar el desempeño de radiólogos experimentados al reducir la subjetividad.

Cuantificación del impacto clínico

La evaluación del impacto clínico de los modelos de IA en salud musculoesquelética constituye un aspecto fundamental para determinar la efectividad y los alcances reales que tendría su integración en la práctica profesional. Validar las métricas diagnósticas obtenidas en condiciones de demanda hospitalaria en un escenario cotidiano nos entregaría las dimensiones y utilidades de su uso. La incorporación como parte del equipo de salud con sus ventajas y desventajas aumentaría la comprensión de su impacto en cada uno de los procesos asistenciales, en la toma de decisiones y finalmente en la experiencia del paciente. Su integración debiera ser gradual con el propósito de identificar fácilmente desafíos operativos y oportunidades de mejora en la optimización de procesos, asegurando que su implementación sea un aporte tangible para actores dentro del ecosistema de salud (166).

En este sentido, la validación del impacto que poseen estos modelos es un objetivo estratégico que debe ser implementado en la práctica asistencial. Propuestas de estudios comparativos entre especialistas y diseños metodológicos podrían ser la implementación de ensayos clínicos controlados, estudios

observacionales, y metodologías diagnósticas prospectivas podrían ser algunos de los diseños preliminares factibles para establecer de forma inmediata su verdadero aporte en el ámbito de la salud musculoesquelética.

Se destaca que la mayoría de los estudios analizados que utilizan algoritmos para la identificación de lesiones musculoesqueléticas en la extremidad superior como en el reconocimiento de tendinopatías se han validado de forma retrospectiva y en entornos controlados (93,150). Por lo que aún sigue siendo un desafío para la disciplina la evaluación del impacto en tiempo real el desempeño de radiólogos y algoritmos en la identificación de TLC.

Desde el punto de vista operacional, algunas estrategias simples a implementar podrían considerar la integración de los modelos de IA en el sistema de almacenamiento, gestión y transmisión de imágenes médicas digitales. Esto podría quedar resuelto explorando formas de aplicación en el sistema PACS, el cual es ampliamente utilizado en entornos médicos que requieren acceso, procesamiento e interpretación de imágenes. Otra medida es que los modelos participen automáticamente en el análisis de las ecografías de codo que obtienen los especialistas, resaltando las zonas de mayor probabilidad de pesquisar un hallazgo ultrasonográfico, sirviendo como de orientación en tiempo real en las regiones de interés. De esta forma, al orientar y guiar acerca al profesional al diagnóstico, pero dejando la decisión final al especialista, mejorando la confianza en su uso cotidiano.

Recientemente, se ha reportado que la colaboración binomio humano-IA en el mejoramiento del diagnóstico es una estrategia válida para mejorar los resultados que cuando se analizan por separado. Se debe comprender que este enfoque híbrido aprovecha las fortalezas complementarias de la experiencia clínica humana y la alta capacidad de procesamiento de información que posee la IA, reduciendo la variabilidad clínica, obteniendo resultados en segundos y por ende mejorando la performance en diagnósticos complejos (167).

Otras ventajas que podrían ser evaluadas dentro del contexto de la evaluación del impacto clínico en la incorporación de estas tecnologías se enmarcan en la velocidad de análisis de detección, pudiendo descongestionar listas de esperas en servicios mayormente demandados por una atención de salud, administración estratégica y uso eficiente de los equipos médicos mediante el aumento de la disponibilidad de los mismos, disminución de la carga laboral o

demanda asistencial de los profesionales debido al uso de sistemas automatizados de detección diagnóstica (168).

Finalmente, a pesar de reconocer la importancia que significa la cuantificación del impacto clínico, esta actividad queda fuera del alcance de los objetivos planteados en esta tesis. Esto se fundamenta debido a una serie de limitaciones técnicas y operativas que no forman parte de lo anteriormente desarrollado. Entre los principales obstáculos para llevar a cabo un nuevo estudio se encuentran requerimientos técnicos, económicos y humanos. Se destaca que se necesitaría contar con nuevas imágenes ecográficas con una resolución destacada y una calidad suficiente para poder hacer nuevos análisis. Además de necesitar un nuevo dataset extenso cuyo diagnóstico se encuentre validado por especialistas. Por otro lado, al implementar los modelos presentados se requeriría nuevamente necesidades de servidores, cómputo y almacenamiento. Dado lo anterior, el estudio requeriría recursos significativamente superiores a los disponibles teniendo que concursar a fuentes de financiamiento externo para ejecutar cada una de las etapas necesarias. Adicionalmente, se requiere un tiempo protegido exclusivo para el desarrollo en extenso del protocolo de investigación para ser presentado ante las instancias regulatorias correspondiente según la normativa chilena actual de investigación, más un tiempo estimado aproximando de 2 a 3 años para ejecutar exitosamente este estudio de gran escala.

VII – CONCLUSIONES

7. CONCLUSIONES

Los modelos de IA evaluados en el metaanálisis permitieron demostrar una alta precisión diagnóstica para la detección de patologías musculoesqueléticas de la extremidad superior. Estos hallazgos respaldan el potencial de ML como una herramienta de apoyo para profesionales en el área de imágenes. Para evaluar su aplicabilidad y eficacia en la práctica clínica, futuras investigaciones deberán validar estos modelos en estudios prospectivos y en entornos clínicos en tiempo real.

Se identificaron los principales tipos de hallazgos degenerativos en pacientes portadores de TLC. Se confirmó la alta prevalencia de patrones hipoecogénicos o tendinosis, calcificaciones, entesopatías, neovascularización y RIS. Adicionalmente, se determinó la asociación existente entre lesiones hipoecogénicas y neovascularización con la magnitud de las RIS. Estos resultados destacan la importancia de los hallazgos degenerativos como indicadores clave en la progresión de la patología, lo que sugiere su potencial utilidad en la estratificación del riesgo clínico y en la toma de decisiones terapéuticas.

La prueba índice y la presencia de dolor nocturno demostraron alta sensibilidad y precisión diagnóstica en pacientes con TLC que presentaban un hallazgo de RIS. A pesar de que estos hallazgos requieren validación adicional prospectiva para consolidar su aplicabilidad clínica, esta contribución proporciona herramientas adicionales específicas en la detección de RIS en estos sujetos, permitiendo a los profesionales del área musculoesquelética incorporar nuevas estrategias en la evaluación y diagnóstico avanzado de esta condición.

Los modelos de ML resultaron ser una alternativa efectiva para mejorar la sensibilidad y especificidad en la clasificación de hallazgos degenerativos y RIS en pacientes con TLC que cuentan con una imagen ecográfica. En particular, los rendimientos alcanzados fueron sobresalientes tanto en hallazgos individuales como múltiples, confirmando que los algoritmos de ML pueden ser utilizados como complemento para la toma de decisiones diagnósticas en este tipo de individuos.

VIII - REFERENCIAS BIBLIOGRÁFICAS

8. REFERENCIAS BIBLIOGRÁFICAS

1. Sarkodie-Gyan T, Yu H. The Human Locomotor System: Physiological and Technological Foundations. Cham; 2023. 316 p.
2. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. CMAJ Can Med Assoc J J Assoc Medicale Can. 14 de marzo de 2006;174(6):801-9.
3. Hart PD, Buck DJ. The effect of resistance training on health-related quality of life in older adults: Systematic review and meta-analysis. Health Promot Perspect. 2019;9(1):1-12.
4. Williams A, Kamper SJ, Wiggers JH, O'Brien KM, Lee H, Wolfenden L, et al. Musculoskeletal conditions may increase the risk of chronic disease: a systematic review and meta-analysis of cohort studies. BMC Med. 25 de septiembre de 2018;16(1):167.
5. Millar NL, Silbernagel KG, Thorborg K, Kirwan PD, Galatz LM, Abrams GD, et al. Tendinopathy. Nat Rev Dis Primer. 7 de enero de 2021;7(1):1.
6. Karabinov V, Georgiev GP. Lateral epicondylitis: New trends and challenges in treatment. World J Orthop. 18 de abril de 2022;13(4):354-64.
7. Vaquero-Picado A, Barco R, Antuña SA. Lateral epicondylitis of the elbow. EFORT Open Rev. noviembre de 2016;1(11):391-7.
8. Ma KL, Wang HQ. Management of Lateral Epicondylitis: A Narrative Literature Review. Pain Res Manag. 2020;2020:6965381.
9. Di Filippo L, Vincenzi S, Pennella D, Maselli F. Treatment, Diagnostic Criteria and Variability of Terminology for Lateral Elbow Pain: Findings from an Overview of Systematic Reviews. Healthc Basel Switz. 14 de junio de 2022;10(6):1095.
10. Coale M, Schiffman B, Iannuzzi N, Huang J. Magnetic resonance imaging for elbow pathology: overused by both orthopedic surgeons and primary care providers. JSES Int. noviembre de 2022;6(6):1062-6.
11. Radunovic G, Vlad V, Micu MC, Nestorova R, Petranova T, Porta F, et al. Ultrasound assessment of the elbow. Med Ultrason. junio de 2012;14(2):141-6.

12. Del Baño-Aledo ME, Martínez-Payá JJ, Ríos-Díaz J, Mejías-Suárez S, Serrano-Carmona S, de Groot-Ferrando A. Ultrasound measures of tendon thickness: Intra-rater, Inter-rater and Inter-machine reliability. *Muscles Ligaments Tendons J.* 2017;7(1):192-9.
13. Moriarity AK, Brown ML, Schultz LR. Work and retirement preferences of practicing radiologists as a predictor of workforce needs. *Acad Radiol.* agosto de 2014;21(8):1067-71.
14. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer.* agosto de 2018;18(8):500-10.
15. Dicle O. Artificial intelligence in diagnostic ultrasonography. *Diagn Interv Radiol Ank Turk.* 31 de enero de 2023;29(1):40-5.
16. Park SH, Han K. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction. *Radiology.* marzo de 2018;286(3):800-9.
17. Bhabra G, Wang A, Ebert JR, Edwards P, Zheng M, Zheng MH. Lateral Elbow Tendinopathy: Development of a Pathophysiology-Based Treatment Algorithm. *Orthop J Sports Med.* noviembre de 2016;4(11):2325967116670635.
18. Chourasia AO, Buhr KA, Rabago DP, Kijowski R, Lee KS, Ryan MP, et al. Relationships between biomechanics, tendon pathology, and function in individuals with lateral epicondylitis. *J Orthop Sports Phys Ther.* junio de 2013;43(6):368-78.
19. Heales LJ, Bergin MJG, Vicenzino B, Hodges PW. Forearm Muscle Activity in Lateral Epicondylalgia: A Systematic Review with Quantitative Analysis. *Sports Med Auckl NZ.* diciembre de 2016;46(12):1833-45.
20. Nayak SR, Ramanathan L, Krishnamurthy A, Prabhu LV, Madhyastha S, Potu BK, et al. Extensor carpi radialis brevis origin, nerve supply and its role in lateral epicondylitis. *Surg Radiol Anat SRA.* marzo de 2010;32(3):207-11.
21. Coombes BK, Bisset L, Vicenzino B. Management of Lateral Elbow Tendinopathy: One Size Does Not Fit All. *J Orthop Sports Phys Ther.* noviembre de 2015;45(11):938-49.

22. Sanders TL, Maradit Kremers H, Bryan AJ, Ransom JE, Smith J, Morrey BF. The epidemiology and health care burden of tennis elbow: a population-based study. *Am J Sports Med.* mayo de 2015;43(5):1066-71.
23. Droppelmann G, Cárcamo M, Opazo A, Rosales J, Greene C, González F, et al. PATRONES ULTRASONOGRÁFICOS Y PREDICCIONES SEGÚN EDAD EN EPICONDILITIS. *Rev Obs Deporte.* 2017;123-35.
24. Park HB, Gwark JY, Im JH, Na JB. Factors Associated With Lateral Epicondylitis of the Elbow. *Orthop J Sports Med.* mayo de 2021;9(5):23259671211007734.
25. Curti S, Mattioli S, Bonfiglioli R, Farioli A, Violante FS. Elbow tendinopathy and occupational biomechanical overload: A systematic review with best-evidence synthesis. *J Occup Health.* enero de 2021;63(1):e12186.
26. Pitts G, Uhl TL, Day JM. Grip strength is more sensitive to changes in elbow position than isolated wrist extension strength in patients with lateral elbow tendinopathy. *J Hand Ther Off J Am Soc Hand Ther.* 2021;34(3):509-11.
27. Musa RM, Hassan I, Abdullah MR, Latiff Azmi MN, Abdul Majeed APP, Abu Osman NA. A longitudinal analysis of injury characteristics among elite and amateur tennis players at different tournaments from electronic newspaper reports. *Front Public Health.* 2022;10:835119.
28. Thorpe CT, Screen HRC. Tendon Structure and Composition. *Adv Exp Med Biol.* 2016;920:3-10.
29. Meunier M. Lateral Epicondylitis/Extensor Tendon Injury. *Clin Sports Med.* julio de 2020;39(3):657-60.
30. Turner CH, Pavalko FM. Mechanotransduction and functional response of the skeleton to physical stress: the mechanisms and mechanics of bone adaptation. *J Orthop Sci Off J Jpn Orthop Assoc.* 1998;3(6):346-55.
31. Burger EH, Klein-Nulend J. Mechanotransduction in bone--role of the lacuno-canalicular network. *FASEB J Off Publ Fed Am Soc Exp Biol.* 1999;13 Suppl:S101-112.
32. Jadaun PK, Zhang S, Koedam M, Demmers J, Chatterjee S, van Leeuwen JP, et al. Inhibition of hypoxia-induced Mucin 1 alters the proteomic composition of human osteoblast-produced extracellular matrix, leading to

reduced osteogenic and angiogenic potential. *J Cell Physiol.* febrero de 2022;237(2):1440-54.

33. Rineer CA, Ruch DS. Elbow tendinopathy and tendon ruptures: epicondylitis, biceps and triceps ruptures. *J Hand Surg.* marzo de 2009;34(3):566-76.

34. Maffulli N, Longo UG, Spiezia F, Denaro V. The Pathophysiology of Tendon Injury. En: Margheritini F, Rossi R, editores. *Orthopedic Sports Medicine: Principles and Practice* [Internet]. Milano: Springer Milan; 2011 [citado 18 de noviembre de 2024]. p. 21-30. Disponible en: https://doi.org/10.1007/978-88-470-1702-3_2

35. Chisari E, Rehak L, Khan WS, Maffulli N. Tendon healing in presence of chronic low-level inflammation: a systematic review. *Br Med Bull.* 11 de diciembre de 2019;132(1):97-116.

36. Cook JL, Feller JA, Bonar SF, Khan KM. Abnormal tenocyte morphology is more prevalent than collagen disruption in asymptomatic athletes' patellar tendons. *J Orthop Res Off Publ Orthop Res Soc.* marzo de 2004;22(2):334-8.

37. Walz DM, Newman JS, Konin GP, Ross G. Epicondylitis: pathogenesis, imaging, and treatment. *Radiogr Rev Publ Radiol Soc N Am Inc.* enero de 2010;30(1):167-84.

38. Dakin SG, Buckley CD, Al-Mossawi MH, Hedley R, Martinez FO, Wheway K, et al. Persistent stromal fibroblast activation is present in chronic tendinopathy. *Arthritis Res Ther.* 25 de enero de 2017;19(1):16.

39. Chang A, Miller TT. Imaging of tendons. *Sports Health.* julio de 2009;1(4):293-300.

40. Hill CE, Heales LJ, Stanton R, Kean CO. Pain-free grip strength in individuals with lateral elbow tendinopathy: Between- and within-session reliability of one versus three trials. *Physiother Theory Pract.* mayo de 2023;39(5):1007-15.

41. Pederzini LA, Prandini M, Di Palma F, Sante ED. Sport-Related Elbow Problems. En: Doral MN, Karlsson J, editores. *Sports Injuries: Prevention, Diagnosis, Treatment and Rehabilitation* [Internet]. Berlin, Heidelberg: Springer; 2015 [citado 18 de noviembre de 2024]. p. 627-39. Disponible en: https://doi.org/10.1007/978-3-642-36569-0_40

42. Lui PPY. Tendinopathy in diabetes mellitus patients-Epidemiology, pathogenesis, and management. *Scand J Med Sci Sports*. agosto de 2017;27(8):776-87.
43. Oliva F, Piccirilli E, Berardi AC, Frizziero A, Tarantino U, Maffulli N. Hormones and tendinopathies: the current evidence. *Br Med Bull*. marzo de 2016;117(1):39-58.
44. Abdul Rahman H, Shah A, Abu Bakar A, Shahar H, Mohd Nazan AIN, Nazan M. A Review on the Factors Contributing to Non-communicable Disease (NCD) Among Low-wage Workers. 1 de octubre de 2021;17:2636-9346.
45. Laratta J, Caldwell JM, Lombardi J, Levine W, Ahmad C. Evaluation of common elbow pathologies: a focus on physical examination. *Phys Sportsmed*. mayo de 2017;45(2):184-90.
46. Pitzer ME, Seidenberg PH, Bader DA. Elbow tendinopathy. *Med Clin North Am*. julio de 2014;98(4):833-49, xiii.
47. Mohandhas BR, Makaram N, Drew TS, Wang W, Arnold GP, Abboud RJ. Racquet string tension directly affects force experienced at the elbow: implications for the development of lateral epicondylitis in tennis players. *Shoulder Elb*. julio de 2016;8(3):184-91.
48. Pustovoit B, Pashkevych S, Duhina L. PHYSICAL THERAPY FOR CHRONIC LATERAL ELBOW TENDOPATHY (TENNIS ELBOW). *Slobozhanskyi Her Sci Sport*. 26 de octubre de 2020;8(5):80-91.
49. Farzad M, MacDermid JC, Shafiee E, Beygi AS, Vafaei A, Varahra A, et al. Clinimetric testing of the Persian version of the Patient-Rated Tennis Elbow Evaluation (PRTEE) and the Disabilities of the Arm, Shoulder, and Hand (DASH) questionnaires in patients with lateral elbow tendinopathy. *Disabil Rehabil*. junio de 2022;44(12):2902-7.
50. Marinelli A, Buondonno C, Al Zoubi A, Guerra E. Evaluation of Common Tendinopathies of the Elbow. En: Lane JG, Gobbi A, Espregueira-Mendes J, Kaleka CC, Adachi N, editores. *The Art of the Musculoskeletal Physical Exam* [Internet]. Cham: Springer International Publishing; 2023 [citado 18 de noviembre de 2024]. p. 159-69. Disponible en: https://doi.org/10.1007/978-3-031-24404-9_19

51. Speers CJ, Bhogal GS, Collins R. Lateral elbow tendinosis: a review of diagnosis and management in general practice. *Br J Gen Pract.* noviembre de 2018;68(676):548.
52. Lucado AM, Dale RB, Kolber MJ, Day JM. ANALYSIS OF RANGE OF MOTION IN FEMALE RECREATIONAL TENNIS PLAYERS WITH AND WITHOUT LATERAL ELBOW TENDINOPATHY. *Int J Sports Phys Ther.* agosto de 2020;15(4):526-36.
53. Karanasios S, Korakakis V, Whiteley R, Vasilogeorgis I, Woodbridge S, Gioftsos G. Exercise interventions in lateral elbow tendinopathy have better outcomes than passive interventions, but the effects are small: a systematic review and meta-analysis of 2123 subjects in 30 trials. *Br J Sports Med.* mayo de 2021;55(9):477-85.
54. Gustas CN, Lee KS. Multimodality Imaging of the Painful Elbow: Current Imaging Concepts and Image-Guided Treatments for the Injured Thrower's Elbow. *Radiol Clin North Am.* septiembre de 2016;54(5):817-39.
55. Tuntiyatorn P, Taweesakulvashra R, Kanchanathepsak T, Rojpitipongsakorn C, Tawonsawatruk T. Validation of a novel magnetic resonance imaging classification and recommended treatment for lateral elbow tendinopathy. *BMC Musculoskelet Disord.* 22 de agosto de 2022;23(1):803.
56. Latham SK, Smith TO. The diagnostic test accuracy of ultrasound for the detection of lateral epicondylitis: a systematic review and meta-analysis. *Orthop Traumatol Surg Res OTSR.* mayo de 2014;100(3):281-6.
57. Dones VC, Grimmer K, Thoires K, Suarez CG, Luker J. The diagnostic validity of musculoskeletal ultrasound in lateral epicondylalgia: a systematic review. *BMC Med Imaging.* 3 de marzo de 2014;14:10.
58. Lee KS, Rosas HG, Craig JG. Musculoskeletal ultrasound: elbow imaging and procedures. *Semin Musculoskelet Radiol.* septiembre de 2010;14(4):449-60.
59. Droppelmann G, Feijoo F, Greene C, Tello M, Rosales J, Yáñez R, et al. Ultrasound findings in lateral elbow tendinopathy: A retrospective analysis of radiological tendon features [Internet]. F1000Research; 2022 [citado 18 de noviembre de 2024]. Disponible en: <https://f1000research.com/articles/11-44>

60. Greene C, Droppelmann G, García N, Jorquera C, Verdugo A. A New Test for the Advanced Diagnosis of Lateral Elbow Tendinopathy with Concomitant Intrastance Tear: Failure to Resist Extension Effort (the Free Test). *Orthop Res Rev.* 2022;14:495-503.
61. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. *Radiogr Rev Publ Radiol Soc N Am Inc.* 2017;37(2):505-15.
62. Nichols JA, Herbert Chan HW, Baker MAB. Machine learning: applications of artificial intelligence to imaging and diagnosis. *Biophys Rev.* febrero de 2019;11(1):111-8.
63. Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Panykh OS, et al. Current Applications and Future Impact of Machine Learning in Radiology. *Radiology.* agosto de 2018;288(2):318-28.
64. Suhaidi M, Kadir RA, Tiun S. A REVIEW OF FEATURE EXTRACTION METHODS ON MACHINE LEARNING. *J Inf Technol Manag JISTM.* 1 de septiembre de 2021;6(22):51-9.
65. Abdallah ZS, Du L, Webb GI. Data Preparation. En: Sammut C, Webb GI, editores. *Encyclopedia of Machine Learning and Data Mining* [Internet]. Boston, MA: Springer US; 2016 [citado 18 de noviembre de 2024]. p. 1-11. Disponible en: https://doi.org/10.1007/978-1-4899-7502-7_62-1
66. Jiang H, Nachum O. Identifying and Correcting Label Bias in Machine Learning. En: *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics* [Internet]. PMLR; 2020 [citado 18 de noviembre de 2024]. p. 702-12. Disponible en: <https://proceedings.mlr.press/v108/jiang20a.html>
67. Zhang Z, Sejdić E. Radiological images and machine learning: Trends, perspectives, and prospects. *Comput Biol Med.* mayo de 2019;108:354-70.
68. Maleki F, Ovens K, Najafian K, Forghani B, Reinhold C, Forghani R. Overview of Machine Learning Part 1: Fundamentals and Classic Approaches. *Neuroimaging Clin N Am.* noviembre de 2020;30(4):e17-32.
69. Sultan K, Ali H, Zhang Z. Big Data Perspective and Challenges in Next Generation Networks. *Future Internet.* julio de 2018;10(7):56.

70. Moubayed A, Injadat M, Nassif AB, Lutfiyya H, Shami A. E-Learning: Challenges and Research Opportunities Using Machine Learning & Data Analytics. *IEEE Access*. 2018;6:39117-38.
71. Jarantow SW, Pisors ED, Chiu ML. Introduction to the Use of Linear and Nonlinear Regression Analysis in Quantitative Biological Assays. *Curr Protoc*. junio de 2023;3(6):e801.
72. Reinders C, Ackermann H, Yang MY, Rosenhahn B. Learning convolutional neural networks for object detection with very little training data. En: *Multimodal Scene Understanding: Algorithms, Applications and Deep Learning* [Internet]. Elsevier; 2019 [citado 18 de noviembre de 2024]. p. 65-100. Disponible en: <https://research.utwente.nl/en/publications/learning-convolutional-neural-networks-for-object-detection-with->
73. Zhang L, Huettmann F, Zhang X, Liu S, Sun P, Yu Z, et al. The use of classification and regression algorithms using the random forests method with presence-only data to model species' distribution. *MethodsX*. 1 de enero de 2019;6:2281-92.
74. Cortes C, Vapnik V. Support-Vector Networks. *Mach Learn*. 1995;20(3):273-97.
75. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to Machine Learning, Neural Networks, and Deep Learning. *Transl Vis Sci Technol*. 27 de febrero de 2020;9(2):14.
76. Kim JY, Ro K, You S, Nam BR, Yook S, Park HS, et al. Development of an automatic muscle atrophy measuring algorithm to calculate the ratio of supraspinatus in supraspinous fossa using deep learning. *Comput Methods Programs Biomed*. diciembre de 2019;182:105063.
77. Bazgir O, Zhang R, Dhruva SR, Rahman R, Ghosh S, Pal R. Representation of features as images with neighborhood dependencies for compatibility with convolutional neural networks. *Nat Commun*. 1 de septiembre de 2020;11(1):4391.
78. Xu Y, Goodacre R. On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. *J Anal Test*. 2018;2(3):249-62.

79. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*. 2 de enero de 2020;21(1):6.
80. Vicenzino B, de Vos RJ, Alfredson H, Bahr R, Cook JL, Coombes BK, et al. ICON 2019-International Scientific Tendinopathy Symposium Consensus: There are nine core health-related domains for tendinopathy (CORE DOMAINS): Delphi study of healthcare professionals and patients. *Br J Sports Med*. abril de 2020;54(8):444-51.
81. Hopkins C, Fu SC, Chua E, Hu X, Rolf C, Mattila VM, et al. Critical review on the socio-economic impact of tendinopathy. *Asia-Pac J Sports Med Arthrosc Rehabil Technol*. abril de 2016;4:9-20.
82. Degen RM, Conti MS, Camp CL, Altchek DW, Dines JS, Werner BC. Epidemiology and Disease Burden of Lateral Epicondylitis in the USA: Analysis of 85,318 Patients. *HSS J Musculoskelet J Hosp Spec Surg*. febrero de 2018;14(1):9-14.
83. Clark ST, Zhu M, Gamble GD, Naot D, Paine SJ, Dalbeth N, et al. Epidemiology of tendon and ligament injuries in Aotearoa/New Zealand between 2010 and 2016. *Inj Epidemiol*. 10 de febrero de 2020;7(1):5.
84. Tavassoli M, Jokar R, Zamani M, Khafri S, Esmailnejad-Ganji SM. Clinical efficacy of local injection therapies for lateral epicondylitis: A systematic review and network meta-analysis. *Casp J Intern Med*. 2022;13(2):311-25.
85. Bjordal JM, Lopes-Martins RA, Joensen J, Couppe C, Ljunggren AE, Stergioulas A, et al. A systematic review with procedural assessments and meta-analysis of Low Level Laser Therapy in lateral elbow tendinopathy (tennis elbow). *BMC Musculoskelet Disord*. 29 de mayo de 2008;9(1):75.
86. Chen Z, Baker NA. Effectiveness of eccentric strengthening in the treatment of lateral elbow tendinopathy: A systematic review with meta-analysis. *J Hand Ther Off J Am Soc Hand Ther*. 2021;34(1):18-28.
87. Janela D, Costa F, Molinos M, Moulder RG, Lains J, Bento V, et al. Digital Rehabilitation for Elbow Pain Musculoskeletal Conditions: A Prospective Longitudinal Cohort Study. *Int J Environ Res Public Health*. 27 de julio de 2022;19(15):9198.
88. Droppelmann G, Torrez A, Garcia N, Greene C. Low-load exercise program and platelet-rich plasma for intrasubstance tear in lateral elbow

tendinopathy: protocol for a double-blinded randomized clinical trial. *Princ Pract Clin Res* [Internet]. 2022 [citado 18 de noviembre de 2024];8(4). Disponible en: <https://journal.ppcr.org/index.php/ppcrjournal/article/view/210>

89. Sims SEG, Miller K, Elfar JC, Hammert WC. Non-surgical treatment of lateral epicondylitis: a systematic review of randomized controlled trials. *Hand N Y N*. diciembre de 2014;9(4):419-46.

90. Geoffroy P, Yaffe MJ, Rohan I. Diagnosing and treating lateral epicondylitis. *Can Fam Physician Med Fam Can*. enero de 1994;40:73-8.

91. Soares MM, Souza PC, Ribeiro AP. Differences in Clinical Tests for Assessing Lateral Epicondylitis Elbow in Adults Concerning Their Physical Activity Level: Test Reliability, Accuracy of Ultrasound Imaging, and Relationship with Energy Expenditure. *Int J Environ Res Public Health*. 18 de enero de 2023;20(3):1794.

92. Miller TT, Shapiro MA, Schultz E, Kalish PE. Comparison of sonography and MRI for diagnosing epicondylitis. *J Clin Ultrasound JCU*. mayo de 2002;30(4):193-202.

93. Droppelmann G, Rodríguez C, Jorquera C, Feijoo F. Artificial intelligence in diagnosing upper limb musculoskeletal disorders: a systematic review and meta-analysis of diagnostic tests. *EFORT Open Rev*. 4 de abril de 2024;9(4):241-51.

94. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 18 de octubre de 2011;155(8):529-36.

95. Puza B, O'Neill T. Generalised Clopper-Pearson confidence intervals for the binomial proportion. *J Stat Comput Simul*. junio de 2006;76(6):489-508.

96. Walter SD. Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. *Stat Med*. 15 de mayo de 2002;21(9):1237-56.

97. Int'Hout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med Res Methodol*. 18 de febrero de 2014;14(1):25.

98. Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychol Methods*. junio de 2006;11(2):193-206.
99. Chambers J. *Software for Data Analysis: Programming with R*. Springer; 2008. 515 p.
100. Cuschieri S. The STROBE guidelines. *Saudi J Anaesth*. abril de 2019;13(Suppl 1):S31-4.
101. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. 14 de noviembre de 2016;6(11):e012799.
102. Bachta A, Rowicki K, Kisiel B, Żabicka M, Elert-Kopeć S, Płomiński J, et al. Ultrasonography versus magnetic resonance imaging in detecting and grading common extensor tendon tear in chronic lateral epicondylitis. *PloS One*. 2017;12(7):e0181828.
103. Tarhan S, Ünlü Z, Ovalı GY, Pabuşçu Y. Value of Ultrasonography on Diagnosis and Assessment of Pain and Grip Strength in Patients with Lateral Epicondylitis. *Arch Rheumatol*. 2009;24(3):123-30.
104. Saroja G, Leo A, venkata sai P. DIAGNOSTIC ACCURACY OF PROVOCATIVE TESTS IN LATERAL EPICONDYLITIS. *Int J Physiother Res*. 11 de diciembre de 2014;2:815-23.
105. Baratloo A, Hosseini M, Negida A, El Ashal G. Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity. *Emerg Tehran Iran*. 2015;3(2):48-9.
106. Šimundić AM. Measures of Diagnostic Accuracy: Basic Definitions. *EJIFCC*. enero de 2009;19(4):203-11.
107. Muschelli J. ROC and AUC with a Binary Predictor: a Potentially Misleading Metric. *J Classif*. octubre de 2020;37(3):696-708.
108. Droppelmann G, Tello M, García N, Greene C, Jorquera C, Feijoo F. Lateral elbow tendinopathy and artificial intelligence: Binary and multilabel findings detection using machine learning algorithms. *Front Med [Internet]*. 23 de septiembre de 2022 [citado 18 de noviembre de 2024];9. Disponible en: <https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2022.945698/full>

109. Draghi F, Danesino GM, de Gautard R, Bianchi S. Ultrasound of the elbow: Examination techniques and US appearance of the normal and pathologic joint. *J Ultrasound*. junio de 2007;10(2):76-84.
110. Palaniswamy V, Ng SK, Manickaraj N, Ryan M, Yelland M, Rabago D, et al. Relationship between ultrasound detected tendon abnormalities, and sensory and clinical characteristics in people with chronic lateral epicondylalgia. *PloS One*. 2018;13(10):e0205171.
111. Connell D, Burke F, Coombes P, McNealy S, Freeman D, Pryde D, et al. Sonographic Examination of Lateral Epicondylitis. *Am J Roentgenol*. marzo de 2001;176(3):777-82.
112. Levin D, Nazarian LN, Miller TT, O'Kane PL, Feld RI, Parker L, et al. Lateral epicondylitis of the elbow: US findings. *Radiology*. octubre de 2005;237(1):230-4.
113. Bianchi S, Martinoli C. *Ultrasound of the Musculoskeletal System*. Springer Science & Business Media; 2007. 974 p.
114. Sommer C, Gerlich DW. Machine learning in cell biology - teaching computers to recognize phenotypes. *J Cell Sci*. 15 de diciembre de 2013;126(Pt 24):5529-39.
115. Sklyar O, Huber W. Image analysis for microscopy screens. *R News*. 1 de diciembre de 2006;6(5):12-6.
116. Nugroho HA, Rahmawaty M, Triyani Y, Ardiyanto I, Choridah L, Indrastuti R. Texture analysis and classification in ultrasound medical images for determining echo pattern characteristics. En: 2017 7th IEEE International Conference on System Engineering and Technology (ICSET) [Internet]. 2017 [citado 18 de noviembre de 2024]. p. 23-6. Disponible en: <https://ieeexplore.ieee.org/document/8123414>
117. Brattain LJ, Telfer BA, Dhyani M, Grajo JR, Samir AE. Machine learning for medical ultrasound: status, methods, and future opportunities. *Abdom Radiol N Y*. abril de 2018;43(4):786-99.
118. Sorower MS. A Literature Survey on Algorithms for Multi-label Learning. En 2010 [citado 18 de noviembre de 2024]. Disponible en: <https://www.semanticscholar.org/paper/A-Literature-Survey-on-Algorithms-for-Multi-label-Sorower/6b5691db1e3a79af5e3c136d2dd322016a687a0b>

119. Shapiro LG. Computer vision: the last 50 years. *Int J Parallel Emergent Distrib Syst.* 3 de marzo de 2020;35(2):112-7.
120. Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput Sci.* 2021;2(3):160.
121. Karpov OE, Pitsik EN, Kurkin SA, Maksimenko VA, Gusev AV, Shusharina NN, et al. Analysis of Publication Activity and Research Trends in the Field of AI Medical Applications: Network Approach. *Int J Environ Res Public Health.* enero de 2023;20(7):5335.
122. Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ.* 2019;7:e7702.
123. Crossnohere NL, Elsaid M, Paskett J, Bose-Brill S, Bridges JFP. Guidelines for Artificial Intelligence in Medicine: Literature Review and Content Analysis of Frameworks. *J Med Internet Res.* 25 de agosto de 2022;24(8):e36823.
124. Jayakumar S, Sounderajah V, Normahani P, Harling L, Markar SR, Ashrafian H, et al. Quality assessment standards in artificial intelligence diagnostic accuracy systematic reviews: a meta-research study. *NPJ Digit Med.* 27 de enero de 2022;5(1):11.
125. Daneshjou R, Barata C, Betz-Stablein B, Celebi ME, Codella N, Combalia M, et al. Checklist for Evaluation of Image-Based Artificial Intelligence Reports in Dermatology: CLEAR Derm Consensus Guidelines From the International Skin Imaging Collaboration Artificial Intelligence Working Group. *JAMA Dermatol.* 1 de enero de 2022;158(1):90-6.
126. Sivanesan U, Wu K, McInnes MDF, Dhindsa K, Salehi F, van der Pol CB. Checklist for Artificial Intelligence in Medical Imaging Reporting Adherence in Peer-Reviewed and Preprint Manuscripts With the Highest Altmetric Attention Scores: A Meta-Research Study. *Can Assoc Radiol J J Assoc Can Radiol.* mayo de 2023;74(2):334-42.
127. Najjar R. Redefining Radiology: A Review of Artificial Intelligence Integration in Medical Imaging. *Diagnostics.* enero de 2023;13(17):2760.
128. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods.* abril de 2010;1(2):97-111.

129. Gyftopoulos S, Lin D, Knoll F, Doshi AM, Rodrigues TC, Recht MP. Artificial Intelligence in Musculoskeletal Imaging: Current Status and Future Directions. *AJR Am J Roentgenol.* septiembre de 2019;213(3):506-13.
130. Hassanipour S, Ghaem H, Arab-Zozani M, Seif M, Fararouei M, Abdzadeh E, et al. Comparison of artificial neural network and logistic regression models for prediction of outcomes in trauma patients: A systematic review and meta-analysis. *Injury.* febrero de 2019;50(2):244-50.
131. de Hond AAH, Leeuwenberg AM, Hooft L, Kant IMJ, Nijman SWJ, van Os HJA, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *NPJ Digit Med.* 10 de enero de 2022;5(1):2.
132. Kusters R, Misevic D, Berry H, Cully A, Le Cunff Y, Dandoy L, et al. Interdisciplinary Research in Artificial Intelligence: Challenges and Opportunities. *Front Big Data [Internet].* 23 de noviembre de 2020 [citado 18 de noviembre de 2024];3. Disponible en: <https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2020.577974/full>
133. Hirschmann A, Cyriac J, Stieltjes B, Kober T, Richiardi J, Omoumi P. Artificial Intelligence in Musculoskeletal Imaging: Review of Current Literature, Challenges, and Trends. *Semin Musculoskelet Radiol.* junio de 2019;23(3):304-11.
134. McAuliffe S, McCreesh K, Culloty F, Purtill H, O'Sullivan K. Can ultrasound imaging predict the development of Achilles and patellar tendinopathy? A systematic review and meta-analysis. *Br J Sports Med.* diciembre de 2016;50(24):1516-23.
135. Clarke AW, Ahmad M, Curtis M, Connell DA. Lateral elbow tendinopathy: correlation of ultrasound findings with pain and functional disability. *Am J Sports Med.* junio de 2010;38(6):1209-14.
136. Obuchowicz R, Bonczar M. Ultrasonographic Differentiation of Lateral Elbow Pain. *Ultrasound Int Open.* mayo de 2016;2(2):E38-46.
137. Fu SC, Rolf C, Cheuk YC, Lui PP, Chan KM. Deciphering the pathogenesis of tendinopathy: a three-stages process. *Sports Med Arthrosc Rehabil Ther Technol SMARTT.* 13 de diciembre de 2010;2:30.

138. Sayampanathan AA, Basha M, Mitra AK. Risk factors of lateral epicondylitis: A meta-analysis. *Surg J R Coll Surg Edinb Irel.* abril de 2020;18(2):122-8.
139. Krogh TP, Fredberg U, Ammitzbøl C, Ellingsen T. Ultrasonographic Characteristics of the Common Extensor Tendon of the Elbow in Asymptomatic Individuals: Thickness, Color Doppler Activity, and Bony Spurs. *Orthop J Sports Med.* mayo de 2017;5(5):2325967117704186.
140. Abbott KV, Barton FB, Terhorst L, Shembel A. Retrospective Studies: A Fresh Look. *Am J Speech Lang Pathol.* 1 de mayo de 2016;25(2):157-63.
141. MacDermid JC, Michlovitz SL. Examination of the elbow: linking diagnosis, prognosis, and outcomes as a framework for maximizing therapy interventions. *J Hand Ther Off J Am Soc Hand Ther.* 2006;19(2):82-97.
142. Zwerus EL, Somford MP, Maissan F, Heisen J, Eygendaal D, van den Bekerom MP. Physical examination of the elbow, what is the evidence? A systematic literature review. *Br J Sports Med.* octubre de 2018;52(19):1253-60.
143. Wright JG. Evidence-Based Orthopaedics E-Book: The Best Answers to Clinical Questions. Elsevier Health Sciences; 2008. 727 p.
144. Palmer K, Walker-Bone K, Linaker C, Reading I, Kellingray S, Coggon D, et al. The southampton examination schedule for the diagnosis of musculoskeletal disorders of the upper limb. *Ann Rheum Dis.* enero de 2000;59(1):5-11.
145. Paoloni JA, Appleyard RC, Murrell GAC. The Orthopaedic Research Institute-Tennis Elbow Testing System: A modified chair pick-up test-interrater and intrarater reliability testing and validity for monitoring lateral epicondylosis. *J Shoulder Elbow Surg.* 2004;13(1):72-7.
146. Cooke N, Obst S, Vicenzino B, Hodges PW, Heales LJ. Upper limb position affects pain-free grip strength in individuals with lateral elbow tendinopathy. *Physiother Res Int J Res Clin Phys Ther.* julio de 2021;26(3):e1906.
147. Benrabha J, Meziane F. Automatic ROI detection and classification of the achilles tendon ultrasound images. En: *Proceedings of the 1st International Conference on Internet of Things and Machine Learning [Internet].* New York, NY, USA: Association for Computing Machinery; 2017 [citado 18 de noviembre de 2024]. p. 1-7. (IML '17). Disponible en: <https://doi.org/10.1145/3109761.3158381>

148. Yu AC, Mohajer B, Eng J. External Validation of Deep Learning Algorithms for Radiologic Diagnosis: A Systematic Review. *Radiol Artif Intell* [Internet]. 4 de mayo de 2022 [citado 10 de marzo de 2025]; Disponible en: <https://pubs.rsna.org/doi/10.1148/ryai.210064>
149. Park SH. Artificial intelligence for ultrasonography: unique opportunities and challenges. *Ultrasonography*. enero de 2021;40(1):3-6.
150. Droppelmann G, Rodríguez C, Smague D, Jorquera C, Feijoo F. Deep learning models for tendinopathy detection: a systematic review and meta-analysis of diagnostic tests. 1 de octubre de 2024 [citado 28 de noviembre de 2024]; Disponible en: <https://eor.bioscientifica.com/view/journals/eor/9/10/EOR-24-0016.xml>
151. Chaki J, Dey N. *A Beginner's Guide to Image Preprocessing Techniques*. Boca Raton: CRC Press; 2018. 114 p.
152. Armi L, Fekri-Ershad S. Texture image analysis and texture classification methods - A review [Internet]. arXiv; 2019 [citado 10 de marzo de 2025]. Disponible en: <http://arxiv.org/abs/1904.06554>
153. Saavedra JP, Droppelmann G, García N, Jorquera C, Feijoo F. High-accuracy detection of supraspinatus fatty infiltration in shoulder MRI using convolutional neural network algorithms. *Front Med* [Internet]. 25 de mayo de 2023 [citado 27 de noviembre de 2024];10. Disponible en: <https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2023.1070499/full>
154. Saavedra JP, Droppelmann G, Jorquera C, Feijoo F. Automated segmentation and classification of supraspinatus fatty infiltration in shoulder magnetic resonance image using a convolutional neural network. *Front Med* [Internet]. 3 de septiembre de 2024 [citado 28 de noviembre de 2024];11. Disponible en: <https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2024.1416169/full>
155. Montañas E, Senge R, Barranquero J, Ramón Quevedo J, José del Coz J, Hüllermeier E. Dependent binary relevance models for multi-label classification. *Pattern Recognit*. 1 de marzo de 2014;47(3):1494-508.

156. Mishra NK, Himthani PK, Singh PK. StaC: Stacked chaining for multi-label classification. *Expert Syst Appl.* 1 de junio de 2023;219:119699.
157. Touw DJW, Velden M van de. Classifier Chain Networks for Multi-Label Classification [Internet]. *arXiv*; 2024 [citado 10 de marzo de 2025]. Disponible en: <http://arxiv.org/abs/2411.02638>
158. Shaik AB, Srinivasan S. A Brief Survey on Random Forest Ensembles in Classification Model. En: Bhattacharyya S, Hassanien AE, Gupta D, Khanna A, Pan I, editores. *International Conference on Innovative Computing and Communications*. Singapore: Springer; 2019. p. 253-60.
159. Islam MdM, Rifat HR, Shahid MdSB, Akhter A, Uddin MA, Uddin KMM. Explainable Machine Learning for Efficient Diabetes Prediction Using Hyperparameter Tuning, SHAP Analysis, Partial Dependency, and LIME. *Eng Rep.* 2025;7(1):e13080.
160. Lambert SI, Madi M, Sopka S, Lenes A, Stange H, Buszello CP, et al. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. *Npj Digit Med.* 10 de junio de 2023;6(1):1-14.
161. Levin D, Nazarian LN, Miller TT, O’Kane PL, Feld RI, Parker L, et al. Lateral epicondylitis of the elbow: US findings. *Radiology.* octubre de 2005;237(1):230-4.
162. Bachta A, Rowicki K, Kisiel B, Żabicka M, Elert-Kopec S, Płomiński J, et al. Ultrasonography versus magnetic resonance imaging in detecting and grading common extensor tendon tear in chronic lateral epicondylitis. *PloS One.* 2017;12(7):e0181828.
163. Bureau NJ, Ziegler D. Economics of Musculoskeletal Ultrasound. *Curr Radiol Rep.* 2016;4:44.
164. Miller TT, Shapiro MA, Schultz E, Kalish PE. Comparison of sonography and MRI for diagnosing epicondylitis. *J Clin Ultrasound JCU.* mayo de 2002;30(4):193-202.
165. Shillito M, Soong M, Martin N. Radiographic and Clinical Analysis of Lateral Epicondylitis. *J Hand Surg.* 1 de junio de 2017;42(6):436-42.
166. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* 29 de octubre de 2019;17(1):195.

167. Zöller N, Berger J, Lin I, Fu N, Komarneni J, Barabucci G, et al. Human-AI collectives produce the most accurate differential diagnoses [Internet]. arXiv; 2024 [citado 10 de marzo de 2025]. Disponible en: <http://arxiv.org/abs/2406.14981>

168. Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ. 22 de septiembre de 2023;23(1):689.

