DESDE 2023
https://rictrevista.org

RICT Revista de Investigación Científica, Tecnológica e Innovación

Publicación Semestral RICT Vol. 4 No. 7 (2026) P. 63-71

ISSN: 2992-7315

Análisis comparativo de algoritmos de árboles de decisión para la predicción de tratamientos en lesiones óseas de muñeca
Comparative Analysis of Decision Tree Algorithms for Treatment Prediction in Wrist Bone Injuries

Edgar Corona Organiche[a], Brandon Moya Bucio[b], Abraham Jorge Jiménez Alfaro[c], Griselda Cortés Griselda[d], Norma Karen Valencia Vázquez[e]

[a ,c, d]*Laboratorio Nacional CONAHCYT en Inteligencia Artificial y Ciencia de Datos (LNC-IACD)-ISC, Tecnológico Nacional de México, Tecnológico de Estudios Superiores de Ecatepec, 55210, Ecatepec de Morelos, Estado de México, México.*
[b]*Secretaria Anticorrupcion y Buen Gobierno, Av. de los Insurgentes Sur 1735, Col. Guadalupe Inn, Alcaldía Álvaro Obregón, C.P. 01020, Ciudad de México.*
[e]*Tecnológico Nacional de México, Tecnológico de Estudios Superiores de Chimalhuacán, C. Ópalo 18, Sta Maria Nativitas, 56335 Chimalhuacán, Méx.*

Resumen

Las lesiones óseas de muñeca, en particular las fracturas de radio distal y las artralgias asociadas a trauma, generan alta demanda asistencial y requieren decisiones terapéuticas oportunas. Este estudio adapta y fortalece un modelo comparativo de árboles de decisión para recomendar tratamientos en pacientes adultos con lesiones de muñeca, empleando registros clínicos retrospectivos de 3,360 pacientes atendidos entre 2023 y 2024. La metodología integra el proceso KDD, depuración de expedientes, codificación de variables clínicas, validación cruzada y evaluación con matrices de confusión, métricas de clasificación y curvas ROC. Se compararon los algoritmos ID3 y C4.5/J48 por su interpretabilidad y su utilidad para sistemas de apoyo a la decisión clínica. Los resultados muestran desempeño superior de C4.5, con precisión de 92.50%, tasa de error de 7.50%, F1-score de 91.36% y AUC-ROC de 0.94, frente a 70.00%, 30.00%, 68.29% y 0.88 obtenidos por ID3. El árbol C4.5 identificó la gravedad de la lesión como nodo raíz, seguido por tipo de molestia y edad, generando reglas clínicas comprensibles para recomendar reposo, fisioterapia, medicamentos o cirugía. Se concluye que C4.5 constituye una alternativa robusta e interpretable para apoyar la recomendación terapéutica personalizada.

Palabras clave: minería de datos, árboles de decisión, C4.5, ID3, lesiones de muñeca, sistemas de apoyo a la decisión clínica

Abstract

Wrist bone injuries, particularly distal radius fractures and trauma-associated arthralgias, generate high care demand and require timely therapeutic decisions. This study adapts and strengthens a comparative model of decision trees to recommend treatments in adult patients with wrist injuries, using retrospective clinical records of 3,360 patients treated between 2023 and 2024. The methodology integrates the KDD process, record cleaning, coding of clinical variables, cross-validation, and evaluation with confusion matrices, classification metrics, and ROC curves. ID3 and C4.5/J48 algorithms were compared for their interpretability and usefulness for clinical decision support systems. Results show superior performance of C4.5, with accuracy of 92.50%, error rate of 7.50%, F1-score of 91.36%, and AUC-ROC of 0.94, versus 70.00%, 30.00%, 68.29%, and 0.88 obtained by ID3. The C4.5 tree identified injury severity as the root node, followed by type of complaint and age, generating comprehensible clinical rules to recommend rest, physiotherapy, medication, or surgery. It is concluded that C4.5 constitutes a robust and interpretable alternative to support personalized therapeutic recommendation.

Keywords: data mining, decision trees, C4.5, ID3, wrist injuries, clinical decision support systems

*Autor para la correspondencia: josejuan@dominio1.edu.mx
Correo electrónico: josejuan@dominio1.edu.mx (José-Juan Sotelo-Ibarra), gabysanchez@dominio2.edu.mx (Gabriela Sánchez-Manríquez), Esta área se llena cuando el manuscrito se aceptado con los nombres completos y correos de todos los autores.

Historial del manuscrito: recibido el DD/MM/AAAA, última versión-revisada recibida el DD/MM/AAAA, aceptado el DD/MM/AAAA, en línea (postprint) desde el DD/MM/AAAA, publicado el DD/MM/AAAA. **DOI:** https://doi.org/10.5281/zenodo.23929595

1. Introducción

Las lesiones óseas de muñeca, en particular las fracturas de radio distal y las artralgias asociadas a trauma, representan una de las causas más frecuentes de consulta en servicios de urgencias y ortopedia a nivel mundial (Arroyo-Berezowsky et al., 2021; Nellans et al., 2012). Estas lesiones generan una alta demanda asistencial y requieren decisiones terapéuticas oportunas que consideren variables como el mecanismo de lesión, comorbilidades y contexto de atención (Arroyo-Berezowsky et al., 2021; Nellans et al., 2012).

La selección del tratamiento suele apoyarse en la valoración médica, estudios auxiliares, estabilidad de la lesión, dolor, funcionalidad y presencia de factores de riesgo. No obstante, la variabilidad clínica puede generar diferencias en la recomendación terapéutica cuando no existe un soporte cuantitativo sistematizado. En esta línea, los modelos predictivos permiten integrar datos históricos y variables clínicas para producir estimaciones reproducibles que apoyen la toma de decisiones, sin desplazar el juicio profesional ni los protocolos institucionales (Collins et al., 2015; Steyerberg, 2019).

1.2. Planteamiento del problema

A pesar de los avances en informática médica y sistemas de apoyo a la decisión clínica, persiste una brecha significativa en la implementación de modelos predictivos interpretables que asistan la selección de tratamientos para lesiones óseas de muñeca en entornos clínicos reales. Los sistemas existentes tienden a priorizar la exactitud estadística sobre la interpretabilidad, lo que dificulta su adopción por parte del personal médico, quien requiere comprender la lógica subyacente a cada recomendación para validarla contra su conocimiento clínico y los hallazgos de la exploración física (Caruana et al., 2015; Rajkomar et al., 2019).

Adicionalmente, la literatura reporta que los algoritmos de árboles de decisión como ID3 y C4.5 presentan diferencias sustanciales en su desempeño cuando se aplican a datos clínicos mixtos, con variables continuas y categóricas, como los que caracterizan a las lesiones de muñeca. Sin embargo, no existe consenso sobre cuál algoritmo ofrece el mejor equilibrio entre precisión predictiva, estabilidad y utilidad clínica para esta aplicación específica (Batra & Agrawal, 2018; Quinlan, 1993; Witten et al., 2016). Esta falta de evidencia comparativa limita la capacidad de las instituciones de salud para seleccionar la herramienta más adecuada para sus sistemas de apoyo a la decisión.

1.3. Pregunta de investigación

¿Cuál de los algoritmos de árboles de decisión, ID3 o C4.5, ofrece un mejor desempeño en términos de precisión, estabilidad e interpretabilidad para predecir tratamientos en pacientes adultos con lesiones óseas de muñeca, y qué reglas clínicas comprensibles puede derivarse de dicho modelo?

1.4. Justificación

La relevancia de este estudio radica en tres aspectos fundamentales. Primero, la creciente demanda de sistemas de apoyo a la decisión clínica que sean precisos y a la vez transparentes para el personal médico. Segundo, la necesidad de aprovechar los registros clínicos digitales disponibles en instituciones de salud para generar conocimiento útil que mejore la estandarización de decisiones terapéuticas. Tercero, la oportunidad de contribuir a la literatura sobre la comparación de algoritmos de árboles de decisión en aplicaciones clínicas específicas, donde las diferencias en el desempeño pueden tener implicaciones directas en la calidad de la atención.

La minería de datos y el aprendizaje automático han adquirido relevancia en informática médica porque transforman registros clínicos en patrones útiles para predicción, clasificación y estratificación de riesgo (Fayyad et al., 1996; Sidey-Gibbons & Sidey-Gibbons, 2019). Dentro de estas técnicas, los árboles de decisión destacan por su interpretabilidad: cada predicción se puede explicar mediante rutas, nodos y reglas que conectan variables clínicas con una salida terapéutica. Esta transparencia es especialmente valiosa en salud, donde la confianza del usuario y la auditabilidad del modelo resultan tan importantes como la exactitud estadística (Caruana et al., 2015; Quinlan, 1986).

1.5. Objetivo

El objetivo de este estudio es adaptar y robustecer un análisis comparativo de ID3 y C4.5 para la predicción de tratamientos en lesiones óseas de muñeca.

2. Materiales y Método

2.2. Diseño del estudio y fuente de datos

El estudio se planteó como un análisis retrospectivo de minería de datos, orientado a construir y evaluar modelos de clasificación para recomendar tratamientos en pacientes adultos con lesiones óseas o molestias de muñeca. La fuente de información correspondió a expedientes clínicos digitales de la Clínica Integral Universitaria, con registros comprendidos entre 2023 y 2024.

La población analizada incluyó 3,360 pacientes de 27 a 59 años con diagnóstico de artralgias o fracturas de muñeca. El criterio de inclusión fue la disponibilidad de las variables clínicas mínimas requeridas para el modelado. Los registros con más del 5% de valores faltantes en variables predictoras clave se excluyeron para preservar consistencia y reducir sesgos en la fase de entrenamiento.

2.3. Variables del modelo

La variable objetivo fue el tratamiento recomendado, representado por cuatro clases: reposo, fisioterapia,

medicamentos antiinflamatorios no esteroideos (AINEs) y cirugía. Las variables predictoras integraron características demográficas, clínicas y de valoración médica:

• X1: edad del paciente, variable numérica continua dentro del intervalo 27-59 años.

• X2: género, variable categórica registrada como masculino o femenino.

• X3: enfermedad crónica degenerativa, variable categórica que permite identificar comorbilidades relevantes.

• X4: tipo de molestia, variable categórica asociada con presentación aguda o crónica.

• X5: gravedad de la lesión, variable numérica en escala 0-10.

• X6: método de valoración, variable categórica que resume el mecanismo clínico de evaluación.

• X7: estudios auxiliares, variable categórica vinculada con radiografía, ultrasonido u otros estudios disponibles.

La incorporación conjunta de variables clínicas y demográficas permite que el árbol construya rutas de decisión comprensibles. Este aspecto es fundamental en aplicaciones médicas, donde la explicabilidad facilita la revisión clínica de las reglas y la detección de recomendaciones potencialmente inconsistentes (Caruana et al., 2015; Rajkomar et al., 2019).

2.4. Proceso KDD y preprocesamiento

El flujo metodológico siguió el proceso KDD, que organiza el descubrimiento de conocimiento en selección, preprocesamiento, transformación, minería de datos e interpretación de resultados (Fayyad et al., 1996). En la selección se integraron expedientes con información suficiente; en el preprocesamiento se revisó completitud, coherencia de categorías y consistencia de rangos; y en la transformación se codificaron variables categóricas mediante representaciones binarias o discretas compatibles con los algoritmos implementados.

Para las variables numéricas se revisaron valores extremos mediante el rango intercuartílico. En lugar de eliminar todos los posibles valores atípicos, se aplicó una estrategia conservadora de winsorización cuando el registro era clínicamente plausible, con el fin de reducir el impacto de extremos sin perder información. Finalmente, el conjunto se aleatorizó y dividió de forma estratificada en entrenamiento y prueba, manteniendo la proporción de la variable objetivo en ambos subconjuntos. El conjunto total de 3,360 pacientes se dividió en 80% para entrenamiento (2,688 casos) y 20% para prueba (672 casos). Las métricas reportadas se calculan sobre el conjunto de datos para prueba, es decir, sobre los 672 casos.

2.5. Algoritmos evaluados

ID3 construye el árbol seleccionando en cada nodo el atributo que maximiza la ganancia de información. Su principal ventaja es la simplicidad, pero presenta limitaciones ante atributos continuos y atributos con muchos valores, ya que puede favorecer particiones altamente específicas. C4.5 corrige parte de este sesgo mediante la ratio de ganancia y añade procedimientos de poda, lo cual reduce el riesgo de sobreajuste y mejora la capacidad de generalización (Quinlan, 1993; Batra & Agrawal, 2017).

La implementación experimental se realizó con Weka, particularmente mediante J48 para C4.5, y se complementó con un prototipo web en ReactJS orientado a capturar datos clínicos y entregar una recomendación terapéutica en tiempo real. Weka es una plataforma ampliamente utilizada para experimentación reproducible en minería de datos y aprendizaje automático (Hall et al., 2009; Witten et al., 2016), véase figura 1.

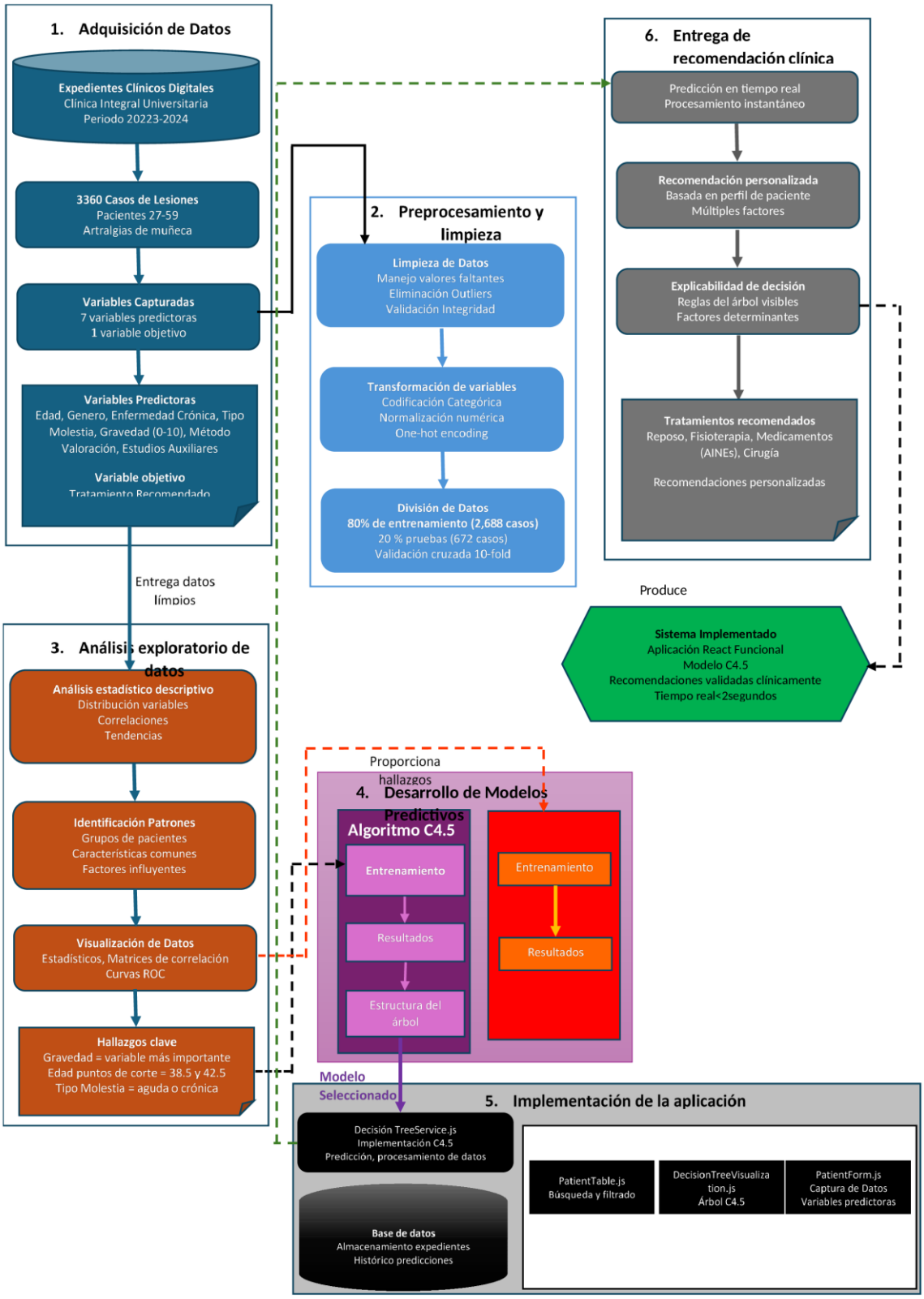


Figura 1. Arquitectura computacional del sistema de predicción: adquisición de datos, limpieza, análisis exploratorio, entrenamiento de modelos, implementación web y recomendación clínica.

2.6. Métricas de evaluación

Los modelos se evaluaron con matrices de confusión, precisión global, tasa de error, sensibilidad, especificidad, F1-score y área bajo la curva ROC. La matriz de confusión permite identificar qué clases se confunden y evaluar la importancia clínica del error. La curva ROC y el AUC resumen la capacidad discriminativa del modelo en distintos umbrales de decisión, por lo que son útiles para comparar clasificadores en escenarios médicos y multiclase (Fawcett, 2006; Powers, 2011).

Además de las métricas, se revisó la interpretabilidad de las rutas del árbol. En modelos clínicos, un resultado estadísticamente alto no es suficiente si la regla no se puede explicar, auditar o contrastar con el conocimiento médico; por ello, la selección del modelo final consideró desempeño, sencillez estructural y congruencia clínica (Collins et al., 2015; Steyerberg, 2019).

3. Resultados

La comparación entre ID3 y C4.5 evidenció diferencias relevantes en la clasificación de tratamientos. El algoritmo C4.5 alcanzó mejor desempeño global y mostró errores clínicamente menos críticos, concentrados principalmente entre tratamientos conservadores. ID3, aunque mantuvo un desempeño aceptable, presentó mayor número de confusiones y menor estabilidad entre clases, véase figura 2a y 2b, tablas 1, 1a,1b,

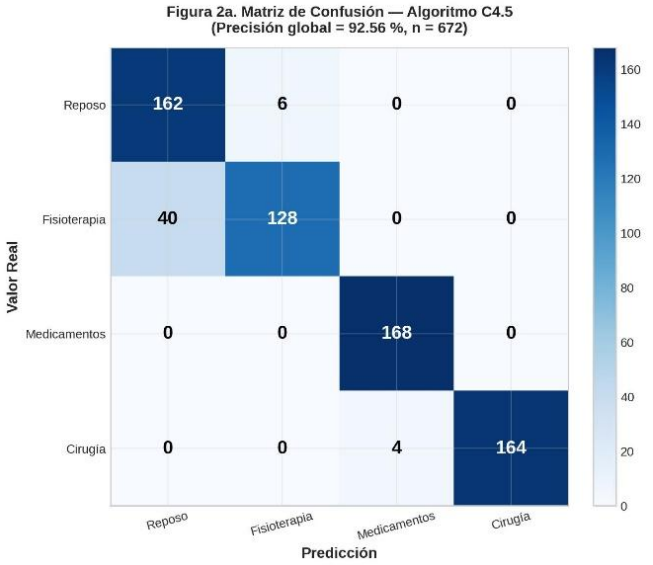


Figura 2a. Matriz de confusión del algoritmo C4.5, con precisión global reportada de 92.56%.

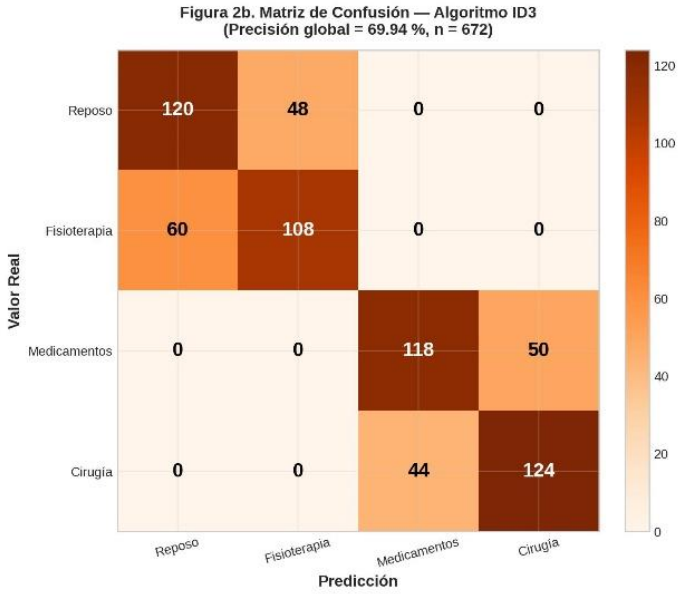


Figura 2b. Matriz de confusión del algoritmo ID3, con precisión global reportada de 69.94%.

La tabla 1, resume los resultados que se presentan en ambas matrices de confusión, permitiendo contrastar y evaluar el desempeño de ambos algoritmos.

Tabla 1. Comparación del desempeño de los algoritmos C4.5 e ID3

Valor Real /	Reposo		Fisioterapia		Medicamentos		Cirugía	
Predicción	C4.5	ID3	C4.5	ID3	C4.5	ID3	C4.5	ID3
Reposo	162	120	6	48	0	0	0	0
Fisioterapia	40	60	128	108	0	0	0	0
Medicamentos	0	0	0	0	168	118	0	50
Cirugía	0	0	0	0	4	44	164	124

La tabla 2, muestra el rendimiento de ambos modelos en las diferentes métricas. En ella se observa que el algoritmo C4.5 superó consistentemente a ID3 en todas las métricas evaluadas. La curva ROC para C4.5 se ubicó más cerca de la esquina superior izquierda, indicando una mayor capacidad para discriminar entre las diferentes clases de tratamiento.

Tabla 2. Comparación del Rendimiento de los Modelos ID3 y C4.5

Métrica	Algoritmo ID3	Algoritmo C4.5
Precisión (Accuracy)	70.00%	92.50%
Tasa de error	30.00%	7.50%
AUC-ROC	0.88	0.94
F1-score	68.29%	91.36%
Sensibilidad (Recall)	69.94%	92.56%
Especificidad	89.98%	97.52%

El análisis de las curvas ROC proporciona una evaluación integral de la capacidad discriminativa de ambos algoritmos a través de todos los umbrales de clasificación posibles. Como se muestra en la Figura 3, presenta las curvas ROC a) para los algoritmos C4.5 e b) ID3 de manera comparativa, permitiendo

visualizar su desempeño en la clasificación multiclase de tratamientos.

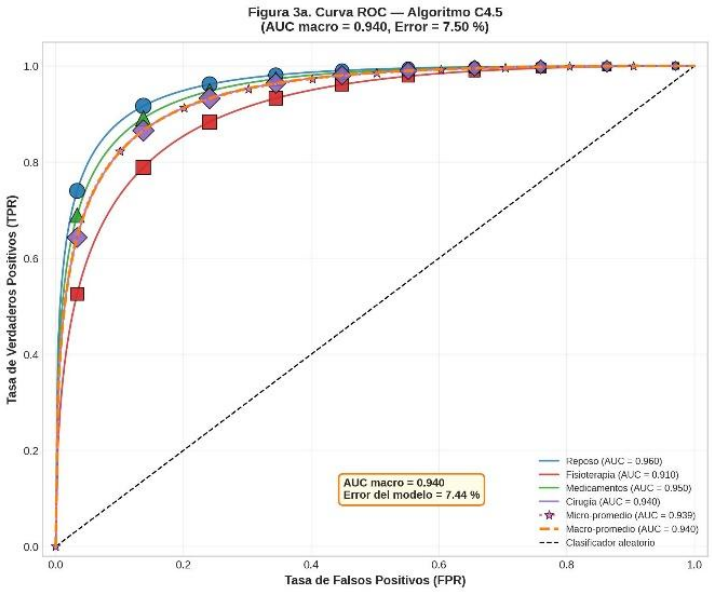


Figura 3a. Curva ROC del algoritmo C4.5, AUC macro = 0.940.

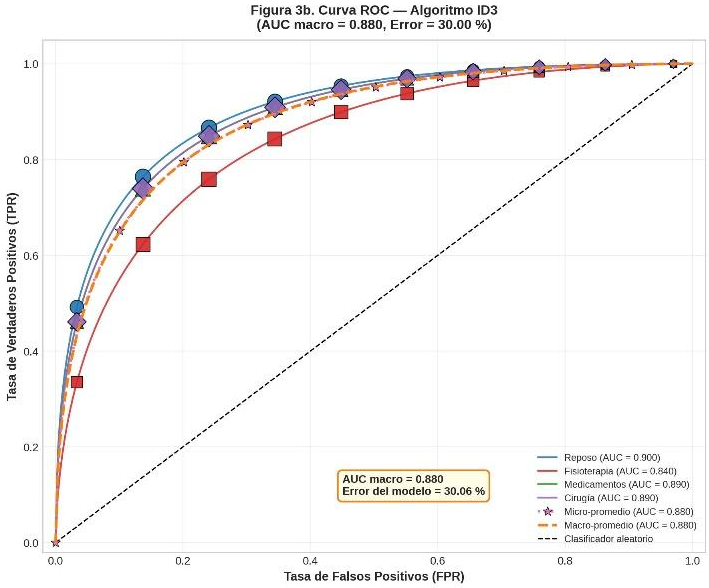


Figura 3b. Curva ROC del algoritmo ID3, AUC macro = 0.880.

Finalmente el árbol de decisión C4.5 simplificado obtenido a partir de las reglas clínicas del modelo, ver figura 4.

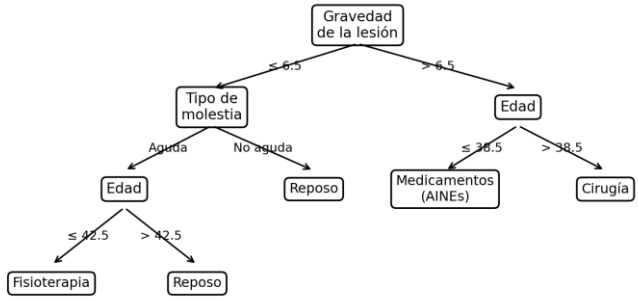


Figura 4. Árbol de decisión C4.5 simplificado obtenido a partir de las reglas clínicas del modelo.

3.2. Interpretación de matrices de confusión

La matriz de C4.5 muestra una clasificación casi perfecta en las clases medicamentos y cirugía, aspecto clínicamente relevante porque confundir una intervención farmacológica con una quirúrgica puede producir subtratamiento o sobretratamiento. Los errores observados se concentran en la relación reposo-fisioterapia, es decir, en alternativas conservadoras con menor riesgo relativo, lo que sugiere un patrón de error clínicamente tolerable bajo supervisión médica.

En ID3, las confusiones se distribuyen de forma más amplia. La presencia de errores entre medicamentos y cirugía indica menor capacidad para separar escenarios de mayor gravedad. Esta diferencia se explica por la estructura del algoritmo: ID3 selecciona atributos por ganancia de información, mientras que C4.5 utiliza ratio de ganancia y puntos de corte para variables continuas, reduciendo sesgos y mejorando la representación de umbrales clínicos (Quinlan, 1993; Witten et al., 2016).

3.3. Análisis ROC y desempeño comparativo

El AUC macro de C4.5 fue 0.94, superior al 0.88 de ID3. Esta diferencia de 0.06 puntos es importante porque resume una mejor discriminación entre clases terapéuticas. Una curva más próxima a la esquina superior izquierda implica mayor tasa de verdaderos positivos con baja tasa de falsos positivos, condición deseable en sistemas de apoyo a la decisión clínica (Fawcett, 2006).

El F1-score de C4.5 (93.82%) también confirma el balance entre precisión y sensibilidad. En problemas clínicos multiclase, esta métrica es útil porque penaliza simultáneamente falsos positivos y falsos negativos, evitando interpretar como robusto un modelo que solo acierta por predominio de una clase. La superioridad simultánea de precisión, F1-score, recall, especificidad y AUC respalda la selección de C4.5 como modelo preferente.

3.4. Modelo matemático del árbol C4.5

El árbol C4.5 se construyó con base en teoría de la información. Sea S el conjunto de entrenamiento, m el número de clases terapéuticas y p_i la proporción de instancias pertenecientes a la clase i. La entropía mide la incertidumbre del conjunto:

$$Entropía\ (S) = -\sum_{i=1}^{m} p_i \log_2 p_i \qquad (1)$$

En la ecuación (1), H(S) representa la impureza del conjunto; p_i es la frecuencia relativa de la clase i; y m = 4 corresponde a reposo, fisioterapia, medicamentos y cirugía. Si todas las instancias pertenecen a una sola clase, la entropía tiende a cero; si las clases están equilibradas, la incertidumbre aumenta.

Para cada atributo A, la ganancia de información compara la entropía original con la entropía ponderada de los subconjuntos generados al dividir por los valores del atributo:

$$Ganancia\ (S,A) = Entropía\ (S) - \sum_{v \epsilon Valores(A)} \frac{|S_v|}{|S|} Entropía\ (S_v) \qquad (2)$$

En la ecuación (2), Values(A) es el conjunto de valores posibles del atributo A; S_v es el subconjunto de registros cuyo atributo A toma el valor v; |S_v| es el número de registros de dicho subconjunto; y |S| es el total de registros. La ganancia aumenta cuando una partición reduce la mezcla de clases.

C4.5 evita favorecer atributos con muchos valores mediante el cálculo de la información de partición:

$$SplitInfo(S,A) = -\sum_{v \epsilon Valores(A)} \frac{|S_v|}{|S|} \log_2 \frac{|S_v|}{|S|} \qquad (3)$$

En la ecuación (3), SplitInfo(S,A) mide qué tan dispersa queda la partición producida por A. Cuando un atributo genera demasiadas divisiones pequeñas, este término aumenta y modera la selección del atributo. La ratio de ganancia se define como:

$$Ratio_{Ganancia}(S,A) = \frac{Ganancia\ (S,A)}{SplitInfo(S,A)} \qquad (4)$$

En la ecuación (4), C4.5 selecciona el atributo con mayor GainRatio, siempre que la partición sea informativa. Esta operación explica por qué la gravedad de la lesión aparece como nodo raíz: reduce de manera significativa la incertidumbre sobre el tratamiento recomendado. Para atributos numéricos como edad y gravedad, C4.5 evalúa puntos de corte candidatos θ y selecciona el umbral que maximiza la ratio de ganancia:

$$\theta^* = \arg\max_\{\theta \in T_A\}\ \text{GainRatio}(S, A \leq \theta) \qquad (5)$$

En la ecuación (5), T_A representa el conjunto de umbrales candidatos para el atributo numérico A. El resultado del modelo identificó umbrales clínicamente interpretables, como gravedad ≤ 6.5, edad ≤ 42.5 y edad ≤ 38.5.

3.5. Métricas de evaluación

A partir de la matriz de confusión, se calcularon las métricas estándar de rendimiento. Si TP representa verdaderos positivos, TN verdaderos negativos, FP falsos positivos y FN falsos negativos, la precisión global se expresa como:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \qquad (6)$$

La precisión por clase indica la proporción de predicciones positivas que fueron correctas:

$$\text{Precision} = TP / (TP + FP) \qquad (7)$$

La sensibilidad o recall mide la capacidad del modelo para detectar correctamente los casos reales de una clase:

$$\text{Recall} = TP / (TP + FN) \qquad (8)$$

La especificidad resume la capacidad de excluir correctamente los casos que no pertenecen a una clase:

$$\text{Especificidad} = TN / (TN + FP) \qquad (9)$$

El F1-score integra precisión y sensibilidad en una media armónica, útil cuando se busca equilibrio entre falsos positivos y falsos negativos:

$$F1 = 2 \cdot (\text{Precision} \cdot \text{Recall}) / (\text{Precision} + \text{Recall}) \qquad (10)$$

Para el análisis ROC se calcularon la tasa de verdaderos positivos y la tasa de falsos positivos:

$$TPR = TP / (TP + FN), \quad FPR = FP / (FP + TN) \qquad (11)$$

El AUC se obtiene como el área bajo la curva TPR-FPR; valores cercanos a 1 indican alta capacidad discriminativa. En este estudio, C4.5 obtuvo mejor AUC que ID3, por lo que separó con mayor consistencia las clases terapéuticas.

3.6. Reglas clínicas derivadas del árbol

El árbol de decisión C4.5 produjo reglas compactas y clínicamente interpretables. En forma lógica, la recomendación puede expresarse como:

$$\text{Fisioterapia} \Leftarrow (\text{Gravedad} \leq 6.5) \wedge (\text{Molestia} = \text{Aguda}) \wedge (\text{Edad} \leq 42.5) \qquad (12)$$

$$\text{Reposo} \Leftarrow (\text{Gravedad} \leq 6.5) \wedge [(\text{Molestia} \neq \text{Aguda}) \vee (\text{Edad} > 42.5)] \qquad (13)$$

$$\text{Medicamentos} \Leftarrow (\text{Gravedad} > 6.5) \wedge (\text{Edad} \leq 38.5) \quad (14)$$

$$\text{Cirugía} \Leftarrow (\text{Gravedad} > 6.5) \wedge (\text{Edad} > 38.5) \quad (15)$$

Las ecuaciones (12)-(15) muestran cómo se traduce el árbol en reglas de decisión. La variable gravedad opera como criterio inicial de separación entre lesiones de baja/moderada y alta severidad. En lesiones de baja gravedad, el tipo de molestia y la edad orientan hacia fisioterapia o reposo; en lesiones de alta gravedad, la edad ajusta la recomendación entre manejo farmacológico y cirugía. Estas reglas no sustituyen la valoración clínica, pero permiten transparentar la lógica del sistema y documentar por qué se emite una recomendación determinada.

3.7. Validación del modelo

La validación cruzada de 10 pliegues estima el rendimiento medio al entrenar y evaluar el modelo en particiones alternadas del conjunto de datos. Si Acc_k es la precisión obtenida en el pliegue k, el rendimiento promedio se define como:

$$Accurancy_{CV} = \frac{1}{10}\sum_{k=1}^{10} Accurancy_k \quad (16)$$

La ecuación (16) reduce la dependencia de una sola división entrenamiento-prueba y proporciona una estimación más estable. Aun así, para un uso clínico real se requiere validación externa con datos de otras instituciones y evaluación prospectiva, de acuerdo con recomendaciones de reporte y validación de modelos predictivos clínicos (Collins et al., 2015; Steyerberg, 2019), véase figura 5 y figura 6.

Registros del Modelo Predictivo

Total de registros entre 27 y 59 años: **3360**

Tratamiento	idregistro	Edad	Género	Familiar	Tipo de Enfermedad crónica degenerativa	Observación	Dieta	Vacunación
Generar Tratamiento	1	49	Masculino	Madre	Diabetes	Controlado	Vegetariana	Influenza, He
Generar Tratamiento	3	30	Masculino	Hermano	Ninguna	-	Alta en grasa	COVID-19, H
Generar Tratamiento	5	41	Masculino	Hermano	Asma	Controlado	Vegetariana	Tétanos, COV
Generar Tratamiento	6	39	Masculino	Padre	Diabetes	Controlado	Vegetariana	COVID-19, H
Generar Tratamiento	7	27	Femenino	Hermano	Hipertensión	Controlado	Baja en carbohidratos	Hepatitis B, Ir
Generar Tratamiento	8	54	Masculino	Hermano	Diabetes	Controlado	Omnívora	Tétanos, COV
Generar Tratamiento	10	35	Masculino	Ninguno	Asma	Controlado	Vegetariana	Influenza, CC
Generar Tratamiento	11	54	Femenino	Madre	Asma	Controlado	Vegetariana	Tétanos, Hep
Generar Tratamiento	12	30	Masculino	Padre	Asma	Controlado	Omnívora	Hepatitis B, C
Generar Tratamiento	13	27	Masculino	Hermano	Hipertensión	Controlado	Baja en carbohidratos	Hepatitis B, C
Generar Tratamiento	16	54	Femenino	Madre	Ninguna	-	Alta en grasa	Tétanos, COV
Generar Tratamiento	17	31	Masculino	Ninguno	Hipertensión	Controlado	Baja en carbohidratos	Influenza, He
Generar Tratamiento	18	41	Masculino	Hermano	Hipertensión	Controlado	Vegetariana	Influenza, He
Generar Tratamiento	19	56	Masculino	Padre	Ninguna	-	Alta en grasa	Tétanos, Influ

Figura 5. Prototipo web en ReactJS para consulta de registros y generación de recomendaciones.



Figura 6. Salida del sistema con tratamiento sugerido y recomendaciones complementarias.

4. Discusión

Los resultados demuestran que C4.5 supera a ID3 en la predicción de tratamientos para lesiones de muñeca. La diferencia no se limita a métricas globales: también se observa en la naturaleza de los errores. Mientras C4.5 concentra sus confusiones en alternativas conservadoras, ID3 produce confusiones más sensibles entre medicamentos y cirugía. Esta distinción es relevante porque los errores no tienen el mismo peso clínico; confundir reposo y fisioterapia puede corregirse con seguimiento, pero confundir cirugía y tratamiento farmacológico puede modificar de manera sustancial el curso terapéutico.

La selección de la gravedad como nodo raíz es congruente con el razonamiento clínico: la severidad de una lesión ósea condiciona la urgencia, el nivel de intervención y la necesidad de estudios complementarios. La edad y el tipo de molestia funcionan como moduladores de la recomendación, lo que permite generar rutas sencillas y comprensibles. Esta estructura favorece la adopción del modelo porque el profesional puede verificar si la regla es coherente con la exploración física y los hallazgos auxiliares.

La superioridad de C4.5 puede explicarse por tres razones técnicas. Primero, el algoritmo maneja variables continuas mediante umbrales, lo que se adapta mejor a variables como edad y gravedad. Segundo, la ratio de ganancia reduce el sesgo hacia atributos con muchos valores. Tercero, la poda disminuye la complejidad del árbol y favorece la generalización. Estas propiedades coinciden con la literatura que describe a C4.5/J48 como una extensión más robusta que ID3 para datos mixtos y problemas aplicados (Quinlan, 1993; Witten et al., 2016).

Desde el punto de vista de implementación, el prototipo web facilita la interacción entre el modelo y el usuario clínico: el médico puede introducir variables del paciente y obtener una recomendación inmediata con una explicación basada en reglas. No obstante, el sistema debe considerarse una herramienta de apoyo. La decisión final requiere historia clínica, exploración física, estudios de imagen, preferencias del paciente, contraindicaciones y disponibilidad institucional.

La principal limitación del estudio es su carácter retrospectivo y unicéntrico. Aunque el tamaño de la base de datos es suficiente para un análisis comparativo inicial, la validación externa es indispensable para confirmar generalización. También se recomienda evaluar calibración, análisis de subgrupos, estabilidad temporal del modelo y posibles sesgos por género, edad o comorbilidades. Una versión futura debería incorporar trazabilidad de datos, consentimiento y gobernanza clínica para su integración en sistemas hospitalarios.

5. Conclusiones

El desarrollo matemático explicita las fórmulas de entropía, ganancia, ratio de ganancia, selección de umbrales, métricas de clasificación, ROC/AUC, validación cruzada y reglas finales del árbol.

El algoritmo C4.5 mostró desempeño superior frente a ID3 en todas las métricas reportadas: precisión, tasa de error, AUC-ROC, F1-score, sensibilidad y especificidad. Además, produjo un árbol interpretable con reglas clínicas simples basadas en gravedad de la lesión, tipo de molestia y edad.

El modelo puede apoyar la recomendación terapéutica personalizada para lesiones de muñeca y contribuir a la estandarización de decisiones, pero su uso debe limitarse a un sistema de apoyo. Antes de su adopción clínica se recomienda validación externa, evaluación prospectiva, análisis de seguridad, revisión ética y supervisión por especialistas en ortopedia y rehabilitación.

En términos de transferencia tecnológica, el prototipo en ReactJS muestra una ruta viable para integrar modelos de aprendizaje automático en interfaces clínicas. Su valor principal reside en combinar exactitud predictiva con explicabilidad, permitiendo que la recomendación no sea una caja negra, sino una secuencia de criterios verificables por el profesional de salud.

6. Agradecimientos

Los autores expresan su agradecimiento a la Clínica Integral Universitaria por la colaboración en el desarrollo de la presente investigación, por las facilidades otorgadas para el acceso a la información necesaria y por su disposición para apoyar actividades de análisis.

Asimismo, se reconoce de manera especial al personal técnico y administrativo de la institución participante que contribuyó con la provisión, validación y contextualización de los datos utilizados en el estudio, cuya colaboración fue fundamental para la correcta comprensión de los procesos y para la interpretación de los resultados obtenidos.

Finalmente, los autores reconocen a todas las personas e instancias que, de manera directa o indirecta, participaron en el desarrollo de esta investigación, reiterando que el tratamiento de la información se realizó bajo principios de confidencialidad, ética y uso responsable de los datos.

7. Referencias

Arroyo-Berezowsky, C., Quintero-Fábregas, J., & García-López, A. (2021). Epidemiology of hand and wrist injuries treated in a reference specialty center over a year. Acta Ortopédica Mexicana, 35(5), 435-441.

Batra, M., & Agrawal, R. (2017). Comparative analysis of decision tree algorithms. En N. R. Shetty, L. C. Prasad, & N. H. Prasad (Eds.), Advances in Intelligent Systems and Computing (Vol. 517, pp. 31-36). Springer. https://doi.org/10.1007/978-981-10-3174-8_4

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. Wadsworth.

Bruder, A. M., Shields, N., Dodd, K. J., & Taylor, N. F. (2017). Prescribed exercise programs may not be effective in reducing impairments and improving activity during upper limb fracture rehabilitation: A systematic review. Journal of Physiotherapy, 63(4), 205-220. https://doi.org/10.1016/j.jphys.2017.08.009

Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. En Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1721-1730). Association for Computing Machinery. https://doi.org/10.1145/2783258.2788613

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. Annals of Internal Medicine, 162(1), 55-63. https://doi.org/10.7326/M14-0697

Domínguez-Gasca, L. G., & Villaseñor, S. L. O. (2017). Frecuencia y tipos de fracturas clasificadas por la Asociación para el Estudio de la Osteosíntesis en el Hospital General de León durante un año. Acta Médica Grupo Ángeles, 15(4), 275-286.

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861-874. https://doi.org/10.1016/j.patrec.2005.10.010

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI Magazine, 17(3), 37-54. https://doi.org/10.1609/aimag.v17i3.1230

González, C. E. M., Rodríguez, M. B., & Martínez, F. M. (2016). The wrist joint complex: Anatomical, physiological and biomechanical aspects, characteristics, classification, and treatment of distal radius fractures. Acta Ortopédica Mexicana, 30(3), 141-147.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. ACM SIGKDD Explorations Newsletter, 11(1), 10-18. https://doi.org/10.1145/1656274.1656278

Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3.ª ed.). Morgan Kaufmann.

Lira, F. G. (2011). Aspectos epidemiológicos y mecanismos de lesión. Revista de Ortopedia y Traumatología, 55(1), 13-20.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. En Advances in Neural Information Processing Systems, 30, 4765-4774.

Nellans, K. W., Kowalski, E., & Chung, K. C. (2012). The epidemiology of distal radius fractures. Hand Clinics, 28(2), 113-125. https://doi.org/10.1016/j.hcl.2012.02.001

Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. Journal of Machine Learning Technologies, 2(1), 37-63.

Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1, 81-106. https://doi.org/10.1007/BF00116251

Quinlan, J. R. (1993). C4.5: Programs for machine learning. Morgan Kaufmann.

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. The New England Journal of Medicine, 380(14), 1347-1358. https://doi.org/10.1056/NEJMra1814259

Sidey-Gibbons, J. A. M., & Sidey-Gibbons, C. J. (2019). Machine learning in medicine: A practical introduction. BMC Medical Research Methodology, 19, 64. https://doi.org/10.1186/s12874-019-0681-4

Steyerberg, E. W. (2019). Clinical prediction models: A practical approach to development, validation, and updating (2.ª ed.). Springer. https://doi.org/10.1007/978-3-030-16399-0

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data mining: Practical machine learning tools and techniques (4.ª ed.). Morgan Kaufmann.