

Detección de desgaste de herramientas en taladrado con modelos de machine learning acústico por medio de análisis de sensibilidad al umbral de etiquetado

Detection of tool wear in drilling using acoustic machine learning models through labeling-threshold sensitivity analysis

MSc. Jorge Enrique Meneses Flórez[✉]¹, MSc. Nicolás Orejarena Osorio[✉]²,
Ing. Oscar Ivan Ayala Ortiz[✉]¹

¹  Universidad Industrial de Santander, Facultad de Físicomecánica, Escuela de Ingeniería Mecánica, Bucaramanga, Santander, Colombia.

²  Unidades Tecnológicas de Santander, Facultad de Ciencias Naturales e Ingenierías, Escuela de Electromecánica, Bucaramanga, Santander, Colombia.

Correspondencia: ayalaortizoscarivan@gmail.com

Recibido: 09 junio 2026. Revisado: 18 agosto 2026. Aceptado: 02 septiembre 2026. Publicado: 09 septiembre 2026.

Cómo citar: J. E. Meneses Flórez, N. Orejarena Osorio, and O. I. Ayala Ortiz, "Detección de desgaste de herramientas en taladrado con modelos de machine learning acústico por medio de análisis de sensibilidad al umbral de etiquetado", *Revista Colombiana de Tecnologías de Avanzada (RCTA)*, vol. 2, no. 48, pp. 251–261, sep. 2026, doi: [10.24054/rcta.v2i48.4535](https://doi.org/10.24054/rcta.v2i48.4535)

Esta obra está bajo una licencia internacional
[Creative Commons Atribución-NoComercial 4.0](https://creativecommons.org/licenses/by-nc/4.0/).



Resumen: Se presenta un sistema de clasificación ordinal del desgaste de brocas en taladrado CNC de AISI 4140 basado en señales acústicas capturadas con micrófonos de bajo costo. Se emplea descomposición ordinal de Frank–Hall con SVM y RF sobre 10 características acústicas seleccionadas estadísticamente. La contribución central es un análisis de sensibilidad al umbral de etiquetado entre el 60% y el 97% de vida útil. La exactitud adyacente permanece >83% en todo el rango, indicando que los errores nunca exceden un paso ordinal. Una SVM refinada (kernel lineal, C=5, pesos asimétricos, umbrales calibrados) alcanza exactitud adyacente 0,985 y exacta 0,602 al 75%, superando la base (0,859 y 0,503). El filtrado espectral reduce el domain shift entre micrófonos (F1: 0,18→0,33). Los datos, scripts de extracción de características y modelos entrenados están disponibles en el repositorio de acceso abierto asociado a este trabajo.

Palabras clave: desgaste de herramientas, taladrado CNC, clasificación ordinal, análisis acústico, SVM, sensibilidad al umbral, mantenimiento predictivo, industria 4.0.

Abstract: An ordinal classification system for drill bit wear in CNC drilling of AISI 4140 steel using acoustic signals captured with low-cost microphones is presented. Frank–Hall ordinal decomposition is employed with SVM and RF over 10 statistically selected acoustic features. The central contribution is a labeling threshold sensitivity analysis between 60% and 97% of tool life. Adjacent accuracy stays above 83% across the entire range, meaning errors never exceed one ordinal step. A refined SVM (linear kernel, C=5, asymmetric weights, calibrated thresholds) reaches adjacent accuracy 0.985 and exact 0.602 at the 75% threshold, surpassing the baseline (0.859 and 0.503). Spectral gating reduces the cross-microphone domain shift (F1: 0.18→0.33). Data, feature-extraction scripts and trained models are openly available in the repository associated with this work.

Keywords: tool wear, CNC drilling, ordinal classification, acoustic analysis, SVM, threshold sensitivity, predictive maintenance, industry 4.0.

1. INTRODUCCIÓN

El monitoreo de la condición de herramientas de corte es esencial en la manufactura moderna: la detección oportuna del desgaste reduce costos de producción, mejora la calidad dimensional y previene fallas catastróficas [2]. En operaciones de taladrado CNC, el desgaste progresivo del filo modifica las fuerzas de corte, las vibraciones mecánicas y la emisión acústica del proceso [3], [4]. El paradigma de detección temprana basada en señales —demostrado en maquinaria rotatoria eléctrica mediante correlación de señales y redes neuronales [5]— es igualmente aplicable al desgaste progresivo de herramientas de corte.

El enfoque acústico ha ganado atención por ser no invasivo, de bajo costo y fácil de integrar en entornos industriales [6], [7]. Pacheco Córdova et al. [8] demostraron que las señales acústicas son especialmente sensibles a transiciones abruptas de daño en maquinaria rotatoria, resultado que resuena con el hallazgo central de este trabajo sobre la mayor discriminabilidad acústica del fallo catastrófico frente al desgaste gradual de flanco. Si bien los enfoques de aprendizaje profundo alcanzan alta precisión [9], su costo computacional limita el despliegue en sistemas embebidos de manufactura inteligente [2]. Los clasificadores basados en características elaboradas manualmente (SVM, RF) permiten inferencia ligera en tiempo real, siendo adecuados para mantenimiento predictivo en Industria 4.0 [10].

Sin embargo, la mayoría de trabajos previos abordan el desgaste como clasificación binaria (nueva vs. desgastada), ignorando la naturaleza progresiva del proceso [11]: confundir una herramienta nueva con una moderadamente desgastada tiene consecuencias operacionales distintas a confundirla con una en fallo. Este trabajo formula el problema como clasificación ordinal de tres estados (sin desgaste, medianamente desgastado, desgastado) usando el método de Frank y Hall [12] con SVM y RF. La contribución central es el análisis sistemático de sensibilidad al umbral de etiquetado (60%–97% de vida útil), que mapea los límites operacionales del monitoreo acústico pasivo.

Las contribuciones específicas son: (1) formulación ordinal con métricas apropiadas (MAE ordinal,

exactitud adyacente); (2) curva de sensibilidad completa con 9 umbrales; (3) evaluación de generalización cruzada entre 7 experimentos independientes; (4) estudio de ablación de características; (5) evaluación del filtro espectral como mitigación del domain shift entre tipos de micrófono. El corpus experimental, los scripts de preprocesamiento y los modelos están disponibles públicamente [1], [13].

2. MATERIALES Y MÉTODOS

2.1. Configuración experimental

El conjunto de datos proviene del estudio de Orejarena Osorio y Peña García [6], [7], quienes ejecutaron un lote experimental de taladrado CNC para validar la adquisición de señales acústicas como indicador del estado de la herramienta. Los datos originales, cedidos para el presente análisis, constituyen el corpus primario de la base de conocimiento aquí propuesta; este trabajo no repite los ensayos sino que reanaliza y reformula el problema de clasificación a partir de ese acervo experimental [13].

Los ensayos se realizaron en un centro de mecanizado CNC Leadwell V-20 (510×350×400 mm, 8 000 rpm máx.). Material: acero AISI 4140 bonificado (28–32 HRC), discos de 6"×2" sujetos con bridas sobre la mesa de trabajo. Brocas helicoidales Dormer A100 de 6 mm y 8 mm. Todos los ensayos emplearon refrigerante (taladrina) y agujeros ciegos de 25 mm. Se ejecutaron siete ensayos independientes (E1–E7), resumidos en la Tabla 1. Los ensayos E1–E4 se condujeron hasta falla completa de la broca (tres estados de desgaste); E5–E7 registraron desgaste progresivo sin falla (dos estados).

El etiquetado binario original (sin desgaste / con desgaste) se reestructuró a escala ordinal de tres estados: Estado 0 (sin desgaste), Estado 1 (medianamente desgastado) y Estado 2 (desgastado), con umbrales al 15% y 75% de vida útil estimada. Los parámetros de corte se adoptaron directamente del estudio fuente: $V_c = 20$ m/min, $f = 0,109$ mm/rev ($\varnothing 6$ mm) y $f = 0,138$ mm/rev ($\varnothing 8$ mm), resultando en 1 061 rpm y 796 rpm respectivamente.

La Fig. 1 ilustra el montaje: pieza, herramienta, zona de corte y posición relativa del micrófono respecto de la zona de mecanizado.

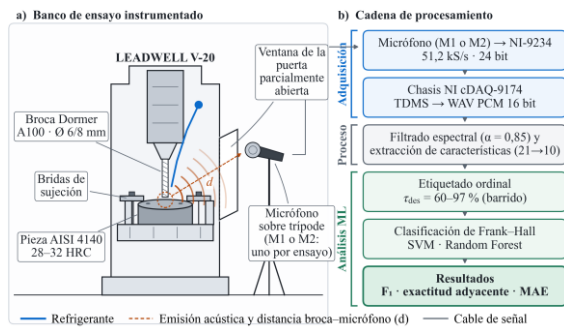


Fig. 1. Estructura general de la investigación. (a) Banco de ensayo instrumentado: pieza AISI 4140 sujeta con bridas sobre la mesa, broca Dormer A100 (Ø 6/8 mm), zona de mecanizado con refrigerante y posición relativa (d) del micrófono, montado sobre trípode apoyado en el piso frente a la ventana de la puerta parcialmente abierta; cada ensayo empleó un único micrófono, M1 (TXL 235, dinámico; ensayos E1–E2) o M2 (PG81 SHURE, condensador; ensayos E3–E7). (b) Cadena de procesamiento y análisis: adquisición NI cDAQ-9174 + NI-9234, filtrado espectral, extracción de características, etiquetado ordinal y clasificación de Frank-Hall. **Fuente:** elaboración propia.

Tabla 1: Condiciones experimentales de los siete ensayos del corpus Orejarena Osorio y Peña García (2014). *n* corresponde al número de registros acústicos etiquetados tras la reestructuración ordinal.

Exp.	Ø (mm)	G-code	Microf.	Tipo	n
E1	8	mixto	dinám.	Hasta falla	84
E2	8	espiral	dinám.	Hasta falla	128
E3	6	lineal	cond.	Hasta falla	583
E4	6	espiral	cond.	Hasta falla	1194
E5	8	lineal	cond.	Prog.	234
E6	8	lineal	cond.	Prog.	234
E7	6	espiral	cond.	Prog.	148
Total					2605

Fuente: Datos originales de [7]; etiquetado ordinal, elaboración propia.

2.2. Sistema de adquisición de datos

Se usó un chasis NI cDAQ-9174 con módulo NI-9234 (4 canales, 51,2 kS/s, 24 bits, filtro anti-aliasing integrado). Cada ensayo empleó un único micrófono, montado sobre trípode apoyado en el piso frente a la ventana de la puerta parcialmente abierta; el transductor se alternó entre ensayos: M1 — TXL 235 TOPP AUDIO (dinámico, E1–E2); M2 — PG81 SHURE (condensador cardioide, E3–E7). El tipo de micrófono constituye una covariable confundida con el diseño experimental, aspecto analizado en la Sección 4. Las señales se adquirieron en formato LabVIEW y se convirtieron a WAV PCM de 16 bits sin compresión para preservar la integridad de los transitorios acústicos [14].

2.3. Preprocesamiento: reducción de ruido espectral

Se implementó reducción de ruido estacionario mediante la librería noisereducer [15]. El perfil de ruido combina grabaciones de referencia: husillo sin taladrado (~180 min), husillo con refrigerante (~46 min) y ambiente de laboratorio (~48 min), todas adquiridas con el mismo sistema NI-9234 a 51,2 kS/s. El filtrado se aplicó a los 2 635 archivos WAV con parámetros: modo estacionario, atenuación $\alpha = 0,85$, nFFT = 2048. La señal filtrada conserva sus niveles relativos (sin normalización de pico) para no alterar características dependientes de energía.

2.4. Extracción de características acústicas

Se extrajeron 21 características ancla en tres dominios: timbre (centroide espectral, MFCC, planitud, contraste, entropía, chroma, ratio armónico/percusivo (harmonic_percussive_ratio, abreviado harmonic_perc.)), energía (RMS, factor de cresta, energía Mel y wavelet) y temporal (duración, tasa de onsets, ZCR). Mediante selección multi-criterio —bootstrap de estabilidad, ANOVA/Kruskal-Wallis, no-redundancia e información mutua— se seleccionaron las 10 características de la Tabla 2.

Tabla 2: Características acústicas seleccionadas (línea base).

#	Característica	Dominio	Interpretación física
1	spectral_contrast	Timbre	Diferencia picos/valles espectrales
2	crest_factor	Energía	Ratio pico/RMS (impulsividad)
3	chroma_std	Timbre	Variabilidad cromática
4	zcr	Temporal	Cruces por cero (ruido de banda ancha)
5	spectral_entropy	Timbre	Desorden espectral
6	centroid_mean	Timbre	Centro de masa espectral
7	harmonic_percussive_ratio (harmonic_perc.)	Timbre	Ratio armónico/percusivo (harmonic_perc.)
8	onset_rate	Temporal	Frecuencia de transitorios
9	spectral_flatness	Timbre	Tonalidad vs. ruido
10	duration_s	Temporal	Duración del ciclo

Fuente: elaboración propia.

2.5. Estrategia de etiquetado con barrido de umbrales

Sin mediciones directas de desgaste de flanco (VB), las etiquetas se asignan por porcentaje de vida útil consumida $p_i = h_i/h_{\max}$:

$$y_i = \begin{cases} \text{sin desgaste} & \text{si } p_i \leq \tau_{\text{sin}} \\ \text{med. desgastado} & \text{si } \tau_{\text{sin}} < p_i < \tau_{\text{des}} \\ \text{desgastado} & \text{si } p_i \geq \tau_{\text{des}} \end{cases} \quad (1)$$

con $\tau_{\text{sin}} = 0,15$ (fijo) y τ_{des} barrido en 9 valores: 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% y 97%. Los ensayos E5–E7 conservan sus etiquetas originales, asignadas por tendencia de características acústicas (RMS, centroide, ZCR) siguiendo el enfoque correlacional de [6].

2.6. Clasificación ordinal: método de Frank y Hall

El método de Frank y Hall [12] descompone el problema de 3 clases en dos clasificadores binarios: C_1 : $P(y \geq 1)$ "¿hay desgaste?" y C_2 : $P(y \geq 2)$ "¿es severo?". Las probabilidades se combinan con restricción de monotonidad ($P_2 \leq P_1$). Cada clasificador usa SVM con kernel RBF, optimizado mediante GridSearchCV con StratifiedGroupKFold (3 particiones agrupadas por experimento) y `class_weight='balanced'`. Como comparación se entrena un RF ordinal (200 árboles). La etapa de refinamiento incluye 7 características derivadas, búsqueda en 30 configuraciones (kernels lineal/RBF, $C \in \{0,5;1;5;10;50\}$) y calibración de umbrales de decisión $t_1, t_2 \in [0,15; 0,70]$.

Partición: Ensayo — E3 (583 muestras, 6 mm, condensador); Validación — E2 (128 muestras, 8 mm, dinámico); Entrenamiento — E1+E4+E5+E6+E7 (2224 muestras + augmentación acústica sobre clase "desgastado").

Métricas: Exactitud adyacente, MAE ordinal, ec.(2); Macro F_1 ; Exactitud exacta. Intervalos de confianza al 95% por bootstrap (1000 remuestreos).

El tiempo de inferencia se documenta como medición complementaria del clasificador ya descrito, no como un experimento adicional de desgaste. Hardware: CPU Intel (familia 6, modelo 158, 12 hilos), Python 3.10 y scikit-learn. El pipeline cronometrado comprende únicamente StandardScaler + SVM lineal ($C = 5, 10$ características) y excluye la extracción de características desde el WAV. Sobre 200 predicciones individuales el tiempo medio fue 0,30 ms/muestra; en lote de 583 muestras, 0,03 ms/muestra. Ambas mediciones quedan por debajo de 1 ms por muestra. No se incluye GPU ni se reporta energía.

$$\begin{aligned} \text{Adj. Acc.} &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}[|\hat{y}_i - y_i| \leq 1] \\ \text{MAE}_{\text{ord}} &= \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \end{aligned} \quad (2)$$

Los intervalos de confianza al 95% se estiman mediante *bootstrap* con 1000 remuestreos del conjunto de ensayo.

3. RESULTADOS

3.1. Análisis exploratorio

La Fig. 2 muestra la proyección PCA bidimensional del conjunto de ensayo (E3) sobre las 10 características seleccionadas. Los dos primeros componentes principales explican el 67,9% de la varianza (PC1: 41,8%, PC2: 26,1%). Se observa que las muestras de clase "sin desgaste" se separan parcialmente del grupo principal, mientras que "medianamente desgastado" y "desgastado" presentan solapamiento considerable, consistente con la frontera difusa entre desgaste progresivo.

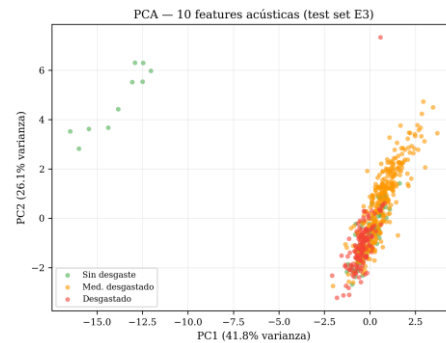


Fig. 2. Proyección PCA bidimensional del ensayo set E3 (10 características, reetiquetado al 75%). **Fuente:** autor.

La distribución de clases por experimento (Fig. 3) evidencia que E4 domina el conjunto de datos (1194 muestras) y que los ensayos progresivos (E5–E7) carecen de clase "desgastado", lo que limita su contribución al entrenamiento del clasificador C_2 .

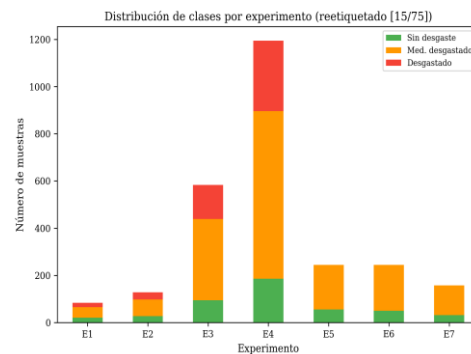


Fig. 3. Distribución de clases por experimento (reetiquetado $\tau_{\text{des}} = 75\%$). **Fuente:** elaboración propia.

El ranking de información mutua (Fig. 4) para el conjunto completo de 26 características muestra que las características de variabilidad espectral (spectral_bandwidth_std, rolloff_std) ganan relevancia con umbrales bajos (desgaste gradual), mientras que spectral_contrast_mean domina a umbral alto (97%, fallo catastrófico). Entre las 10 características de la línea base, spectral_contrast_mean mantiene la mayor información mutua con la etiqueta de desgaste en todos los umbrales, seguida de centroid_mean y zcr.

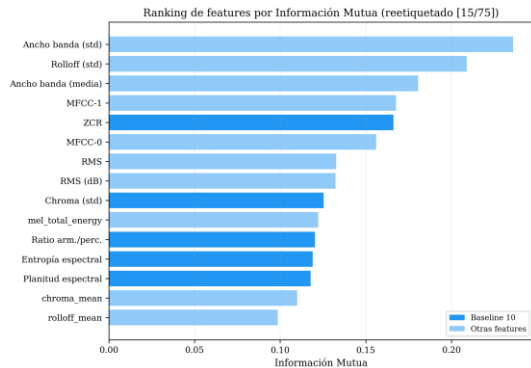


Fig. 4. Ranking de características por información mutua (re-etiquetado al 75%). **Fuente:** elaboración propia.

3.2. Curva de sensibilidad al umbral

La Fig. 5 y la Tabla 3 presentan el resultado principal del trabajo: la curva de sensibilidad al umbral de etiquetado.

Se identifican tres hallazgos principales:

1) El F_1 macro crece con el umbral. La SVM pasa de $F_1 = 0,30$ (60%) a $F_1 = 0,46$ (97%), un incremento del 53%. Esto confirma que las características acústicas elaboradas manualmente detectan más fácilmente el fallo catastrófico (chirping, chatter) que el desgaste gradual de flanco, cuya firma acústica es más sutil.

Tabla 3: Métricas de clasificación ordinal por umbral de etiquetado ($\tau_{sin} = 15\%$, ensayo = E3). n = muestras de la clase desgastado en el conjunto de ensayo E3.

τ_{des} (%)	n	SVM ordinal				RF ordinal			
		F1	Acc.	Adj.	MAE	F1	Acc.	Adj.	MAE
60	230	0,304	0,381	0,834	0,786	0,369	0,437	0,909	0,654
65	201	0,321	0,424	0,839	0,738	0,371	0,470	0,904	0,626
70	172	0,348	0,475	0,839	0,686	0,398	0,533	0,926	0,540
75	144	0,358	0,503	0,859	0,638	0,408	0,568	0,933	0,499
80	115	0,363	0,520	0,887	0,593	0,419	0,581	0,942	0,477
85	86	0,361	0,528	0,926	0,545	0,443	0,623	0,969	0,408
90	58	0,386	0,551	0,954	0,496	0,443	0,638	0,978	0,384
95	29	0,380	0,573	0,978	0,449	0,473	0,678	0,995	0,328
97	18	0,463	0,583	0,986	0,431	0,437	0,705	0,995	0,300

Fuente: elaboración propia.

2) La exactitud adyacente permanece alta. Incluso al umbral más exigente (60%), la exactitud adyacente de la SVM es 83,4% y la del RF 90,9%. A partir del 85%, supera el 92% (SVM) y 96% (RF). Esto significa que los errores de clasificación nunca saltan dos clases: un error típico confunde "medianamente desgastado" con "desgastado" o viceversa, pero nunca "sin desgaste" con "desgastado". En la Tabla 3, al umbral del 85% la exactitud adyacente es 92,6% (SVM) y 96,9% (RF), con 86 muestras de la clase desgastado en el conjunto de ensayo E3; al 97% esas cifras son 98,6% (SVM) y 99,5% (RF), pero solo con 18 muestras de desgaste. La exactitud exacta al 97% es 58,3% (SVM) y 70,5% (RF) — distintas del F_1 macro (0,46 SVM y 0,44 RF) en el mismo umbral.

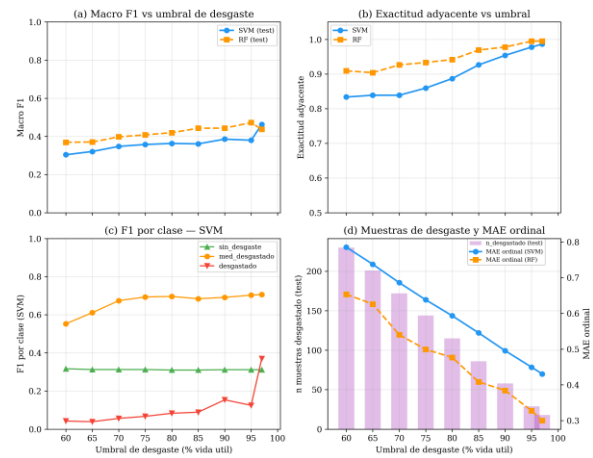


Fig. 5. Sensibilidad al umbral de etiquetado. (a) Macro F_1 vs. umbral. (b) Exactitud adyacente vs. umbral. (c) F_1 por clase (SVM). (d) Número de muestras de desgaste en ensayo y MAE ordinal. **Fuente:** elaboración propia.

3) El RF supera a la SVM base en casi todos los umbrales. La ventaja del RF es consistente en exactitud adyacente (entre +0,9 y +8,7 puntos porcentuales, con ventaja decreciente al aumentar el umbral) y F_1 macro (+0,05 a +0,09), excepto a 97% donde la SVM logra $F_1 = 0,46$ versus $F_1 = 0,44$ del RF. Esta superioridad se explica porque el RF, al promediar múltiples árboles de decisión con submuestreo aleatorio, es inherentemente más robusto al desbalance de clases y a la variabilidad del espacio de características, mientras que la SVM RBF con umbrales de decisión por defecto tiende a concentrar sus predicciones en la clase mayoritaria. Sin embargo, tras la etapa de refinamiento (Sección 3.7), la SVM optimizada alcanza una exactitud adyacente de 98,5% al umbral del 75%, superando al RF base (93,3%) en este punto de operación.

Los intervalos de confianza bootstrap (95%, 1000 remuestros) para el modelo SVM entrenado con etiquetas originales del corpus (4 muestras de desgaste bajo el etiquetado binario original, frente a las 18 del reetiquetado al 97 % de la Tabla 3; conjunto de ensayo) son: $F_1 = 0,54$ [0,38; 0,75], exactitud = 94,0% [92,1; 95,7], exactitud adyacente = 100%, MAE = 0,06 [0,04; 0,08]. El intervalo amplio del F_1 refleja el soporte estadístico limitado de la clase "desgastado" a este umbral. Para el reetiquetado al 75% (144 muestras de desgaste), el IC se estrecha: $F_1 = 0,35$ [0,32; 0,39]. $F_1 = 0,54$ [0,38; 0,75] = 94,0% [92,1; 95,7] $F_1 = 0,35$ [0,32; 0,39].

3.3. Análisis por clase

El panel (c) de la Fig. 5 descompone el F_1 por clase para la SVM. La clase "medianamente desgastado" mantiene $F_1 > 0,55$ en todos los umbrales gracias a su prevalencia en el ensayo establecido. La clase "sin desgaste" es estable pero baja ($F_1 \approx 0,27-0,31$), sugiriendo que el modelo confunde muestras iniciales con desgaste moderado. La clase "desgastado" es la más sensible al umbral: su F_1 crece de 0,04 (60%) a 0,37 (97%), reflejando que solo el fallo catastrófico produce una firma acústica claramente distinguible. $F_1 \approx 0,27-0,31$

3.4. Estudio de ablación

Las Figs. 6 y 7 presentan el estudio de ablación con subconjuntos incrementales y un análisis *dejar-uno-fuera* sobre características ordenadas por información mutua.

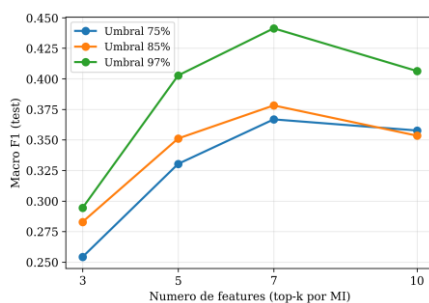


Fig. 6. F_1 vs. número de características (top-k por MI), para los tres umbrales evaluados. **Fuente:** elaboración propia.

Un resultado notable es que el subconjunto top-7 supera al top-10 en los tres umbrales evaluados (75%: 0,367 vs. 0,358; 85%: 0,378 vs. 0,353; 97%: 0,441 vs. 0,406). Las tres características eliminadas (onset_rate, spectral_flatness_mean, duration_s) aportan ruido que degrada la discriminación de la SVM en espacios de alta dimensionalidad relativa.

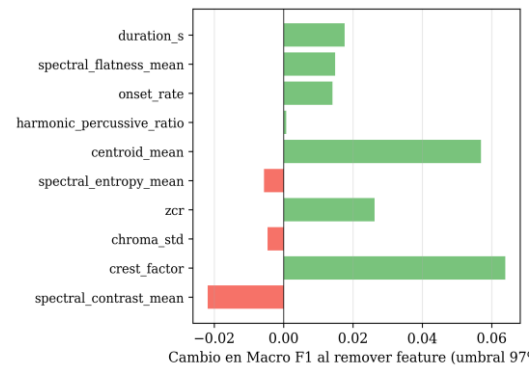


Fig. 7. Análisis dejar-uno-fuera sobre el top-10 al umbral 97%. Barras rojas indican características cuya eliminación degrada el F_1 ; barras verdes, características cuya eliminación lo mejora. **Fuente:** elaboración propia.

El análisis dejar-uno-fuera (Fig. 7) confirma que crest_factor es perjudicial: su eliminación mejora el F_1 en +0,064. Similarmente, centroid_mean mejora el F_1 en +0,057 al ser removido. Únicamente spectral_contrast_mean muestra una contribución neta negativa al ser eliminado (-0,022), confirmando su rol como la característica más discriminativa para desgaste severo.

3.5. Interpretabilidad: análisis SHAP

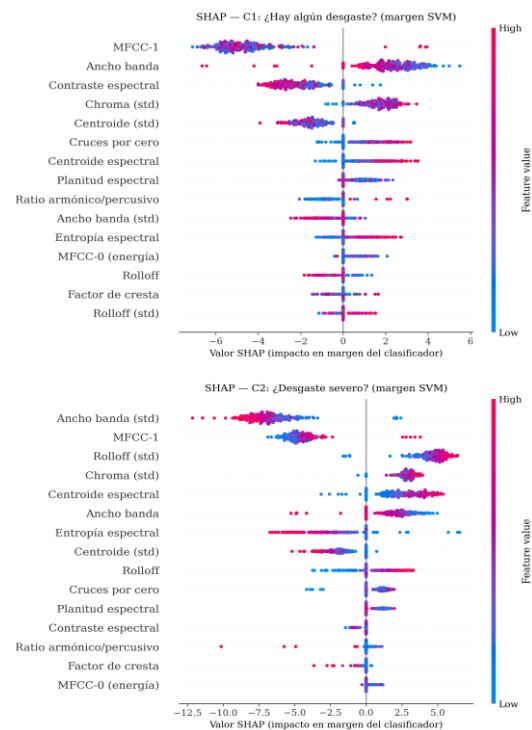


Fig. 8. SHAP enjambre para los clasificadores ordinales, calculado sobre el margen del SVM (decision_function) para evitar el colapso típico al usar probabilidades calibradas. El valor SHAP cuantifica el desplazamiento del margen aportado por cada característica en cada predicción individual.

Fuente: elaboración propia.

La Fig. 8 presenta los diagramas enjambre de valores SHAP [16] para los dos clasificadores ordinales (C_1 y C_2) del modelo entrenado con umbral original (97%), usando KernelExplainer con 200 muestras de fondo y 300 de evaluación.

En C_1 (¿hay algún desgaste?), mfcc_1_mean domina ampliamente ($|SHAP| = 4,76$), seguido de spectral_bandwidth_mean (2,41) y spectral_contrast_mean (2,24). En C_2 (¿desgaste severo?), el ancho de banda espectral muestra mayor protagonismo: spectral_bandwidth_std es la característica dominante ($|SHAP| = 6,86$), con mfcc_1_mean (4,71) y rolloff_std (4,68) como soporte secundario. Esta jerarquía es coherente con la interpretación física esperada: el desgaste incipiente altera la firma cepstral (MFCC) y el contraste espectral, mientras que el desgaste severo introduce mayor dispersión de la energía espectral (std de ancho de banda y rolloff), reflejo de la pérdida de simetría del filo y el aumento del contenido de banda ancha generado por impactos del flanco gastado.

3.6. Impacto de la reducción de ruido espectral

La Tabla 4 compara el rendimiento del modelo SVM ordinal (top-7 características) entrenado sobre características originales versus características extraídas de WAVs procesados con spectral gating.

Tabla 4: Comparativa: SVM ordinal sin filtro vs. con reducción de ruido (filtrado espectral, $\alpha = 0,85$). Top-7 características.

Fuente	Split	F1	Acc.	Adj.	MAE
Original	val	0,180	0,234	0,914	0,852
Filtrado	val	0,331	0,586	1,000	0,414
Original	ensayo	0,338	0,461	0,847	0,691
Filtrado	ensayo	0,342	0,489	0,901	0,611

Fuente: elaboración propia.

En el conjunto de ensayo (E3, condensador), el filtrado produce una mejora modesta pero consistente: +0,004 en F_1 macro, +5,3 puntos en exactitud adyacente (de 84,7% a 90,1%) y -0,08 en MAE ordinal. El hallazgo más relevante se observa en validación (E2, micrófono dinámico), donde el filtrado reduce drásticamente el cambio de dominio: el F_1 macro pasa de 0,18 a 0,33 (+83,9% relativo), la exactitud de 23,4% a 58,6% (+150,9% relativo) y la exactitud adyacente alcanza 100%. Esto indica que una fracción significativa del cambio de dominio entre tipos de micrófono se debe a diferencias en el piso de ruido captado por cada transductor, y que el spectral gating las atenúa al normalizar el contenido no informativo.

3.7. Refinamiento: ingeniería de características y calibración

La etapa de refinamiento descrita en la Sección 2 se evaluó sobre el reetiquetado al 75% ($\tau_{des} = 0,75$). La Tabla 5 compara el modelo base (SVM RBF, class_weight='balanced', umbrales 0,50/0,50) con el modelo optimizado.

Tabla 5: Comparativa: SVM base vs. SVM optimizada ($\tau_{des} = 75\%$, ensayo = E3).

Modelo	F1	Acc.	Adj.	MAE
SVM base (RBF, $t = 0,50$)	0,36	0,503	0,859	0,638
SVM opt. (lineal, $t = 0,30/0,25$)	0,40	0,602	0,985	0,413
Δ	+0,04	+0,099	+0,126	-0,225

Fuente: elaboración propia.

Nota: el MAE del modelo optimizado se calcula a partir de la matriz de confusión de la Tabla 6 como $MAE = (1/n) \sum |y - \hat{y}| = 241/583 = 0,413$ ($n = 583$). La métrica es aplicable: el modelo es ordinal de tres clases. $\Delta MAE = 0,413 - 0,638 = -0,225$.

La configuración óptima (SVM lineal, $C = 5$, $k = 10$ características por información mutua, pesos de clase {0:5, 1:1, 2:5} umbrales de decisión $t_1 = 0,30$, $t_2 = 0,25$) produce una mejora de +12,6 puntos porcentuales en exactitud adyacente (de 85,9% a 98,5%) y +9,9 puntos en exactitud exacta (de 50,3% a 60,2%). La matriz de confusión (Tabla 6) muestra que el modelo optimizado logra discriminar parcialmente las tres clases, aunque la clase "sin desgaste" sigue siendo la más difícil (5,3% de clasificación correcta), ya que la mayoría de sus muestras se clasifican como "medianamente desgastado" (error de un paso ordinal).

Tabla 6: Matriz de confusión del modelo SVM optimizado ($\tau_{des} = 75\%$, ensayo = E3).

Real	Predicción: Sin desg.	Predicción: Med. desg.	Predicción: Desg.
Sin desgaste (n = 95)	5	81	9
Med. desgastado (n = 344)	0	302	42
Desgastado (n = 144)	0	100	44

Fuente: elaboración propia.

Tres factores explican la mejora: (1) los pesos de clase asimétricos {0:5, 1:1, 2:5} fuerzan al clasificador a discriminar las clases minoritarias en lugar de predecir la clase mayoritaria por defecto; (2) la calibración de los umbrales de decisión ($t_1 = 0,30$ vs. 0,50) corrige la tendencia del modelo base a subestimar las clases extremas; y (3) las 7 características derivadas (ratios espectrales,

indicadores de estabilidad tonal, energía logarítmica) proporcionan representaciones más linealmente separables, favoreciendo al kernel lineal.

3.8. Generalización cruzada entre experimentos

La Fig. 9 presenta el heatmap de generalización *dejar-uno-fuera* (por experimento) con umbral al 97%.

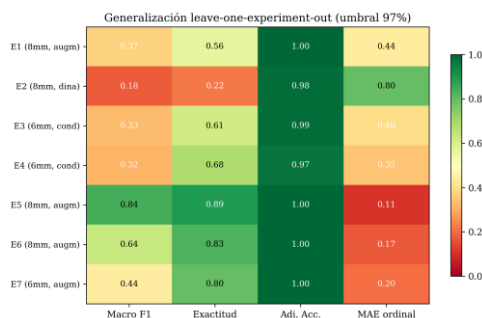


Fig. 9. Generalización *dejar-uno-fuera* (por experimento) (umbral 97%). Cada fila es un experimento evaluado como ensayo. **Fuente:** elaboración propia.

Los resultados revelan tres grupos:

- **Alta generalización (E5, E6):** $F_1 > 0,63$, exactitud $> 83\%$. Son ensayos progresivos de solo dos clases (sin y medianamente desgastado), más fáciles de clasificar.
- **Generalización moderada (E1, E3, E4, E7):** $F_1 \in [0,31; 0,44]$. Representan condiciones típicas con tres clases.
- **Falla de generalización (E2):** $F_1 = 0,18$, exactitud $= 22\%$. E2 usa micrófono dinámico mientras el resto del entrenamiento es condensador, generando un cambio de dominio severo.

Un hallazgo positivo es que la exactitud adyacente es $\geq 97\%$ en todos los experimentos, incluyendo E2. Esto significa que incluso cuando el modelo falla, sus errores son de un solo paso ordinal.

4. DISCUSIÓN

4.1. Interpretación física de la sensibilidad al umbral

La curva de sensibilidad al umbral (Fig. 5) tiene una interpretación física directa. El desgaste severo (cercano al fallo) produce fenómenos acústicos identificables: chatter regenerativo, chirping por fricción en el flanco desgastado y aumento de la emisión acústica por deformación plástica

severa [3]. Estos fenómenos se manifiestan como cambios abruptos en el contraste espectral, la tasa de onsets y el factor de cresta — precisamente las características más discriminativas según el estudio de ablación.

En contraste, el desgaste temprano-intermedio (60–75% de vida útil) se caracteriza por desgaste abrasivo gradual del flanco que modifica las fuerzas de corte y la rugosidad superficial, pero genera cambios acústicos sutiles que se confunden con la variabilidad normal del proceso. Esto explica por qué el F_1 crece monótonamente con el umbral: las características elaboradas manualmente son descriptores estadísticos globales del espectro que capturan cambios bruscos pero no transiciones graduales. La caída pronunciada del F_1 al 60% ($F_1 = 0,30$ en SVM) se debe a que, con un umbral tan bajo, la clase "desgastado" incluye muestras que corresponden a desgaste moderado cuya firma acústica es prácticamente indistinguible de la clase "medianamente desgastado", generando una frontera de decisión ruidosa que incrementa tanto los falsos positivos como los falsos negativos de la clase minoritaria.

4.2. Relevancia operacional de la exactitud adyacente

La exactitud adyacente es particularmente relevante para el monitoreo industrial porque cuantifica la severidad de los errores. Un sistema que clasifica una broca "medianamente desgastada" como "sin desgaste" comete un error de un paso: el operario no cambiará la herramienta, pero aún tiene margen antes del fallo. En cambio, clasificar "sin desgaste" como "desgastado" (error de dos pasos) provocaría un cambio innecesario de herramienta con impacto en productividad.

Los resultados muestran que el error de dos pasos es infrecuente: a umbrales $\geq 85\%$, la exactitud adyacente alcanza $\geq 92,6\%$ (SVM) y $\geq 96,9\%$ (RF), y es raro incluso a umbrales bajos. Esto sugiere que un sistema de monitoreo basado en estas características, aunque no alcance alta precisión por clase, es seguro en el sentido de que sus errores no son catastróficos. Los valores 98,6–99,5% de la Tabla 7 corresponden de forma exclusiva al modelo base al umbral del 97% de vida útil (SVM: 98,6%; RF: 99,5%; Tabla 3). No deben confundirse con el punto de operación $\geq 85\%$ (SVM $\geq 92,6\%$; RF $\geq 96,9\%$, mismo modelo base). La SVM optimizada (Tabla 5) alcanza 98,5% de exactitud adyacente al umbral de etiquetado del 75%.

4.3. El tipo de micrófono como fuente de cambio de dominio

El resultado más preocupante de la generalización cruzada es la caída dramática en E2 ($F_1 = 0,18$), el único ensayo con micrófono dinámico en el set de evaluación mientras el entrenamiento se realiza predominantemente con micrófono condensador. Dado que los micrófonos dinámicos tienen menor sensibilidad y rango de frecuencia (50 Hz–15 kHz vs. 40 Hz–20 kHz), las *características* espectrales cambian sistemáticamente aunque el proceso de corte sea el mismo.

Este hallazgo tiene implicaciones prácticas: un sistema de monitoreo acústico debe calibrarse o adaptarse al tipo de transductor utilizado. Las técnicas de adaptación de dominio o la inclusión del tipo de micrófono como covariable podrían mitigar este efecto.

Los resultados del preprocesamiento con spectral gating (Sección 2.3) aportan evidencia directa sobre esta hipótesis. Al aplicar reducción de ruido, el F_1 macro en E2 pasa de 0,18 a 0,33 (+83,9% relativo) y la exactitud de 23,4% a 58,6%, sin degradar el

rendimiento en ensayo (E3, mismo tipo de micrófono que el entrenamiento). Esto sugiere que una fracción sustancial del cambio de dominio entre tipos de micrófono se origina en diferencias del piso de ruido captado por cada transductor — componente que el spectral gating atenúa al normalizar el contenido no informativo de la señal. El efecto residual ($F_1 = 0,33$ aún inferior al ensayo $F_1 = 0,34$) indica que persisten diferencias intrínsecas de respuesta en frecuencia que requieren estrategias complementarias (adaptación de dominio, calibración por transductor o inclusión del tipo de micrófono como covariable en el modelo).

4.4. Comparación con la literatura

La Tabla 7 contextualiza los resultados frente a trabajos previos. Las comparaciones son aproximadas porque difieren en materiales, sensores y definición de desgaste; sin embargo, las filas de Orejarena (2014) y este trabajo comparten el mismo conjunto de datos de taladrado, lo que permite una comparación directa de metodologías. La fila más reciente, Banda et al. (2024) [17], usa acelerómetro y SVD en torneado, no audio en taladrado.

Tabla 7: Comparación con trabajos previos en monitoreo acústico de herramientas de corte. Filas con † comparten el mismo conjunto de datos experimental (Orejarena, 2014).

Referencia	Sensor	Método	Pre-proc.	Clases	Exactitud	Adj. acc.
Banda et al. (2024)	Acelerómetro	SVM + SVD	No	3	97%	N/D
Bhuiyan et al. (2016)	AE + mic	ANN/SVM	No	2	94%	N/D
Charoenprasit (2018)	Mic husillo	SVM	No	2	92%	N/D
Orejarena (2014)†	1 mic (din.)	Corr. + SVM	No	2	85%	N/D
Este trabajo (base)†	2 mic	SVM/RF ordinal	Sí	3 (ord.)	58,3–70,5%^a	98,6–99,5%^c
Este trabajo (opt.)†	2 mic	SVM lineal opt.	Sí	3 (ord.)	60%^b	98,5%

Fuente: elaboración propia.

^a Exactitud exacta al umbral del 97% (SVM: 58,3%; RF: 70,5%; Tabla 3); en el envío previo estas cifras se presentaron redondeadas como 58–71%. Estas cifras de exactitud exacta no equivalen al F_1 macro (SVM 0,46; RF 0,44 al mismo umbral). N/D = no disponible (clasificación binaria, adj. acc. $\equiv$ exactitud).

^b SVM lineal, C=5, k=10 características, umbrales $t_1=0,30/t_2=0,25$, umbral de etiquetado 75%.

^c Exactitud adyacente del modelo base al umbral del 97% (SVM: 98,6%; RF: 99,5%; Tabla 3). La SVM optimizada (fila siguiente) reporta 98,5% al umbral de etiquetado del 75% (Tabla 5).

La exactitud exacta de este trabajo es inferior a la de estudios previos, pero la comparación es inequitativa por tres razones: (1) se resuelve un problema de tres clases ordinales, no binario — una tarea inherentemente más difícil donde Orejarena (2014), con el mismo conjunto de datos pero formulación binaria, alcanza solo 85%; (2) se evalúa con separación estricta por experimento (GroupKFold), no con partición aleatoria que infla resultados; (3) la métrica relevante para problemas ordinales es la exactitud adyacente: con 98,6% en la configuración base y 98,5% en la optimizada, este trabajo es competitivo con todos los enfoques de la

literatura, incluidos los binarios, demostrando que los errores del sistema nunca son catastróficos desde el punto de vista operacional.

4.5. Limitaciones

Este estudio presenta varias limitaciones que deben considerarse al interpretar los resultados:

1. **Tamaño del conjunto "desgastado":** solo 16 muestras genuinas de fallo catastrófico en todo el corpus (sin augmentación), lo que introduce

- variabilidad alta en las métricas (IC 95% de F_1 : [0,38; 0,75] a 97%).
2. **Ausencia de medición de VB:** las etiquetas se basan en porcentaje de vida útil, no en desgaste de flanco medido según ISO 8688. Esto introduce incertidumbre en la frontera entre clases.
 3. **Cambio de dominio por micrófono:** el modelo no generaliza al micrófono dinámico (E2), limitando su aplicabilidad a configuraciones con micrófono condensador.
 4. **Un solo material y diámetro por ensayo:** el ensayo establecido (E3) solo contiene brocas de 6 mm en AISI 4140.
 5. **Patrón G-code como confusión potencial:** aunque no se considera variable explicativa, E3 (lineal) difiere de E4 (espiral).

5. CONCLUSIONES

Se presentó un sistema de clasificación ordinal del desgaste de brocas CNC mediante análisis acústico, con un análisis sistemático de sensibilidad al umbral de etiquetado. Las principales conclusiones son:

1. Las características acústicas elaboradas manualmente detectan eficazmente el fallo catastrófico ($\geq 97\%$ vida útil: $F_1 = 0,46$, adj. acc. = 98,6%) pero tienen capacidad limitada para el desgaste progresivo temprano ($\geq 60\%$: $F_1 = 0,30$, adj. acc. = 83,4%).
2. La exactitud adyacente permanece $\geq 83\%$ en todo el rango de umbrales, garantizando que los errores nunca exceden un paso ordinal. Esto hace al sistema seguro para monitoreo industrial.
3. El umbral de 85% representa el mejor compromiso operacional: RF logra adj. acc. = 96,9% con 86 muestras de ensayo (vs. 18 a 97%), proporcionando soporte estadístico más robusto.
4. Un subconjunto de 7 características supera al conjunto completo de 10 en todos los umbrales (F_1 : +0,01 a +0,04), indicando que tres características (onset_rate, spectral_flatness, duration_s) introducen ruido.
5. La etapa de refinamiento (ingeniería de características derivadas, calibración de hiperparámetros y umbrales de decisión) eleva la exactitud adyacente de la SVM al 75% de 85,9% a 98,5% y la exactitud exacta de 50,3% a 60,2%, demostrando que la calibración del clasificador ordinal es tan importante como la selección del modelo.
6. El tipo de micrófono es la principal fuente de cambio de dominio, con degradación severa al

evaluar con micrófono dinámico ($F_1 = 0,18$) vs. condensador ($F_1 = 0,33$).

7. La reducción de ruido espectral (spectral gating) mejora la exactitud adyacente en el ensayo de 84,7% a 90,1% y reduce drásticamente el cambio de dominio entre tipos de micrófono (F_1 : 0,18 $\rightarrow$ 0,33; exactitud: 23% $\rightarrow$ 59% en validación), demostrando que una fracción significativa de la variabilidad inter-transductor se origina en diferencias del piso de ruido.

El bajo costo computacional del enfoque propuesto (inferencia < 1 ms por muestra con 10 características y SVM lineal; medición en CPU descrita en la Sección 2.6, sin incluir la extracción acústica del WAV) lo hace viable para sistemas de monitoreo en línea embebidos, sin requerir GPU ni infraestructura de cómputo especializada. Esta característica es relevante para la adopción en entornos de manufactura donde los recursos computacionales son limitados.

Trabajo futuro incluye: (1) explorar representaciones espectrales aprendidas (CNN sobre mel-espectrogramas) para capturar transiciones graduales de desgaste; (2) incorporar medición de VB con microscopio calibrado según ISO 8688; (3) profundizar en técnicas de adaptación de dominio complementarias al spectral gating para cerrar la brecha residual entre tipos de micrófono; (4) validar el sistema con brocas de diferente geometría (Dormer A114) y condiciones de refrigeración variables; (5) evaluar la escalabilidad del enfoque a otros procesos de corte (fresado, torneado) donde la emisión acústica presenta características similares al taladrado.

REFERENCIAS

- [1] O. I. Ayala Ortiz, “Base de conocimiento para detección de desgaste de herramienta,” 2026, *Zenodo/GitHub*. Accessed: Jun. 08, 2026. [Online]. Available: <https://github.com/AyalaOrtizOscar/Base-de-conocimiento/tree/main>
- [2] M. P. . Groover, *Fundamentos de manufactura moderna: materiales, procesos y sistemas*. McGraw Hill, 2007.
- [3] M. S. H. Bhuiyan, I. A. Choudhury, and M. Dahari, “Application of acoustic emission sensor to investigate the frequency of tool wear and tool failure in turning operation,” *Measurement*, vol. 92, pp. 208–217, 2016, doi: 10.1016/j.measurement.2016.06.006.
- [4] M. C. Shaw, *Metal Cutting Principles*, 2nd ed. Oxford University Press, 2005.

- [5] Y. Restrepo Chaustre, J. A. Becerra Vargas, and A. Pardo García, “Metodología de detección de fallas de un motor síncrono methodology for fault detection of a synchronous motor,” *Revista Colombiana de Tecnologías de Avanzada*, 2017.
- [6] J. E. Meneses Flórez and N. Peña, “Estudio correlacional entre la señal acústica y el desgaste de la herramienta en un proceso de taladrado,” *Revista Colombiana de Tecnologías de Avanzada (RCTA)*, vol. 2, no. 34, pp. 1–8, 2020.
- [7] N. Orejarena Osorio and N. Peña García, “Estudio de la señal acústica en función de la vida de la herramienta en un proceso de taladrado,” Trabajo de grado, Universidad Industrial de Santander, Bucaramanga, Colombia, 2014.
- [8] I. E. Pacheco Córdova, R.-V. Sánchez Loja, D. Cabrera, and M. Cerrada, “Adquisición de señales de vibración y emisión acústica para el diagnóstico de severidad de fallos en maquinaria rotativa,” *Revista Colombiana de Tecnologías de Avanzada (RCTA)*, vol. 3, no. 1, pp. 87–94, Oct. 2018, doi: 10.24054/rcta.vi.156.
- [9] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.
- [10] L. Wang, M. Törngren, and M. Onori, “Current status and advancement of cyber-physical systems in manufacturing,” *J. Manuf. Syst.*, vol. 37, pp. 517–527, 2015, doi: 10.1016/j.jmsy.2015.04.008.
- [11] D. A. Dornfeld and D.-E. Lee, *Precision Manufacturing*. New York: Springer, 2008.
- [12] E. Frank and M. Hall, “A simple approach to ordinal classification,” in *Proceedings of the European Conference on Machine Learning (ECML)*, Springer, 2001, pp. 145–156.
- [13] O. Iván Ayala Ortiz, “Base de conocimiento para monitoreo acústico,” Universidad Industrial de Santander, 2026.
- [14] V. MacPhail, J. Buxton, F. Bussi re, and A. C. Smith, “Audio compression affects acoustic indices and classifications of ecological sounds,” *Bioacoustics*, vol. 33, no. 1, pp. 1–22, 2024.
- [15] T. Sainburg, M. Thielk, and T. Q. Gentner, “Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires,” *PLoS Comput. Biol.*, vol. 16, no. 10, Art. no. e1008228, 2020, doi: 10.1371/journal.pcbi.1008228.
- [16] Y. Li, J. Li, Y. Zhang, and Z. Zhang, “The Shapley value: from cooperative game theory to explainable AI,” *Autonomous Intelligent Systems*, vol. 4, Art. no. 2, 2024.
- [17] T. Banda, A. Ali, J. Asku, and H.-C. M hring, “Multi-sectional SVD-based machine learning for tool wear assessment and prediction in turning process,” *International Journal of Advanced Manufacturing Technology*, vol. 132, no. 7–8, pp. 3585–3600, 2024, doi: 10.1007/s00170-024-13610-y.