CEA Comité Español de Automática
XLVII JORNADAS DE AUTOMÁTICA CÓRDOBA 2026
UNIVERSIDADE DA CORUÑA

Jornadas de Automática

Sistema Automático de Mantenimiento Predictivo para Herramientas de Mecanizado mediante Entrenamiento Semi-Supervisado

P. Sanchis[a,*], S. García-Nieto[a], I. Cebreiro[b], J. Sanz[b]

[a] *Instituto Universitario de Automática e Informática Industrial (ai2). Universitat Politècnica de València. Camino de Vera, s/n. 46022 Valencia - España*
[b] *Departamento de Nuevas Tecnologías. Planta de Motores. Ford Almussafes. 46440 Valencia - España*

To cite this article: Sanchis. P, García-Nieto. S, Cebreiro. I, Sanz. J. 2026. Automatic Predictive Maintenance System for Machining Tools using Semi-Supervised Learning. Jornadas de Automática, 47. https://doi.org/10.17979/ja-cea.2026.47.13802

Resumen

La monitorización del estado de la herramienta es fundamental para el mantenimiento predictivo en máquinas de mecanizado. Sin embargo, la ausencia de datos históricos etiquetados y el fuerte desbalanceo entre ciclos de funcionamiento normal y fallos dificultan la implementación de modelos supervisados. Este trabajo presenta un sistema inteligente con ajuste semi-supervisado para clasificar y etiquetar ciclos de mecanizado que parte de un conjunto de características estadísticas extraídas de la señal de par motor. El algoritmo reduce la dimensionalidad mediante el Análisis de Componentes Principales (PCA) y emplea Modelos de Mezclas Gaussianas (GMM) para identificar el comportamiento nominal. Posteriormente, aísla anomalías usando un índice de novedad (*Novelty Score*), genera datos sintéticos para balancear estas clases minoritarias y aplica un clasificador bayesiano final. Así, se obtiene un modelo listo para ser utilizado para monitorización en producción o para etiquetar conjuntos de datos en la fase de entrenamiento de modelos. Los resultados demuestran la capacidad del algoritmo de agrupar y clasificar, así como la posibilidad de semi-supervisión que proporciona la modificación del umbral del *Novelty Score*.

Palabras clave: Aprendizaje automático, Monitorización de condición, Detección y diagnóstico de fallos, Métodos bayesianos, Sistemas de mantenimiento inteligente, Métodos estadísticos/análisis de señales para FDI.

Automatic Predictive Maintenance System for Machining Tools using Semi-Supervised Learning

Abstract

Tool condition monitoring is fundamental for predictive maintenance in machining equipment. However, the lack of labeled historical data and the severe imbalance between normal operating cycles and failures hinder the implementation of supervised models. This paper presents an intelligent system with semi-supervised adjustment for classification and labeling of machining cycles based on a set of statistical features extracted from the torque signal. The algorithm reduces dimensionality via Principal Component Analysis (PCA) and employs Gaussian Mixture Models (GMM) to identify nominal behavior. Subsequently, it isolates anomalies using a Novelty Score, generates synthetic data to balance these minority classes, and applies a final Bayesian classifier. This results in a model that is ready for use in production monitoring or for labelling datasets during the model training phase. The results demonstrate the algorithm's capability to cluster and classify, as well as the potential for semi-supervision provided by adjusting the Novelty Score threshold.

Keywords: Machine Learning, Condition Monitoring, Fault detection and diagnosis, Bayesian methods, Intelligent maintenance systems, Statistical methods/signal analysis for FDI.

*Autor para correspondencia: psanjor@etsii.upv.es

Para la elaboración de este artículo, se han empleado herramientas de Inteligencia Artificial Generativa (modelo Gemini 3.1 Pro) con el único propósito de mejorar la legibilidad y redacción del texto.

1. Introducción

Las máquinas de Control Numérico por Computadora (CNC) representan actualmente el núcleo de las operaciones de mecanizado, donde las paradas no planificadas y los fallos en las herramientas generan importantes pérdidas económicas (Thomas and Weiss (2021)). Es por esto que, cada vez más, se busca sustituir el mantenimiento preventivo, el cual se efectúa de forma periódica bajo intervalos de tiempo fijos sin importar el estado de los equipos; por el mantenimiento predictivo, con el cual se utiliza la información del estado actual del equipo para predecir su estado futuro y planificar las acciones correctivas pertinentes (Mallioris et al. (2024)).

Dentro de las estrategias de mantenimiento predictivo enfocadas a máquinas de mecanizado, la monitorización del estado de la herramienta (*Tool Condition Monitoring, TCM*) es una de las áreas de investigación más activas. Tran et al. (2023) destacan su importancia exponiendo que el 20 % del tiempo de parada es debido al desgaste en las herramientas de trabajo, lo que provoca problemas de calidad y un sobrecoste económico. Los sistemas TCM utilizan datos provenientes de sensores como vibraciones, emisiones acústicas o temperatura, los cuales se preprocesan y analizan cada vez más junto a técnicas de *Machine Learning*, que han demostrado su potencial en aplicaciones prácticas (Kuntoğlu et al. (2020)).

Sin embargo, la implementación de estos modelos en la realidad industrial se enfrenta a un problema fundamental: la necesidad de grandes volúmenes de datos históricos correctamente etiquetados. Este hecho afecta principalmente a los enfoques de aprendizajes supervisados dada la necesidad de etiquetas en los datos. No obstante registrar y clasificar manualmente el estado de la herramienta en cada ciclo de mecanizado supone una tarea inviable y costosa (Dhasaian et al. (2025)). Además, los conjuntos de datos industriales presentan una naturaleza intrínsecamente desbalanceada, tal y como mencionan Fomin et al. (2024), donde los ciclos de mecanizado de funcionamiento normal son la inmensa mayoría, mientras que los eventos anómalos o fallos son eventos atípicos y escasos. Esta falta de etiquetas y de representatividad de todas las fases de desgaste dificultan el entrenamiento de algoritmos robustos.

Para abordar esta problemática, en este trabajo se presenta el desarrollo de un algoritmo de clasificación automático semi-supervisado de ciclos de mecanizado. Aprovechando técnicas de reducción de dimensionalidad y la extracción de métricas estadísticas representativas de la señal de par, el algoritmo propuesto es capaz de clasificar los datos minimizando la intervención humana.

2. Descripción del proceso

Esta investigación se ha llevado a cabo en una planta de fabricación que dispone de diversas máquinas de mecanizado funcionando en un entorno industrial real. Para el desarrollo y posterior experimentación del algoritmo se ha escogido una máquina de tipo transfer, cuyo funcionamiento detalla Belmokhtar et al. (2006) en su trabajo. Dado que esta máquina cuenta con cuatro estaciones y cada una de ellas realiza operaciones diferentes, los datos empleados se corresponden con una de estas estaciones cuyo esquema se plasma en la Figura 1. El proceso de mecanizado que se lleva a cabo en esta estación consiste en el mandrinado de cuatro orificios. Esta estación mecaniza dos piezas al mismo tiempo, con cuatro herramientas diferentes. Los husillos de estas herramientas están conducidos por el mismo motor a pares, es decir, las herramientas que mecanizan los orificios 1A y 2A comparten el mismo motor mediante una correa; ídem para las herramientas de los orificios 1B y 2B.

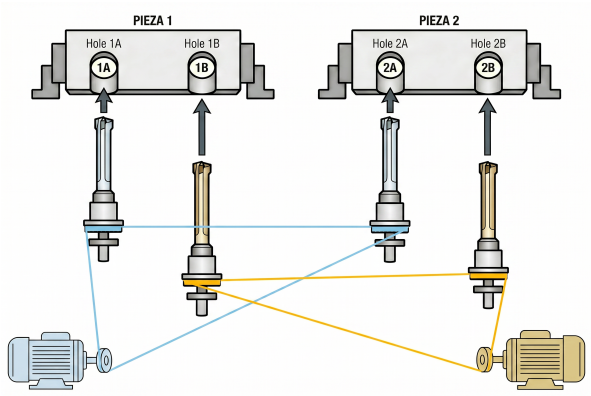

Figura 1: Esquema estación de mecanizado
Fuente: Generado utilizando IA y modificaciones propias

El dato extraído y posteriormente analizado es el par motor del conjunto de husillos 1A y 2A. Este hecho dificulta el análisis de los datos dado que la monitorización se realiza sobre dos herramientas simultáneamente, lo cual debe tenerse en cuenta en la interpretación de los resultados.

En la Figura 2 se muestra la evolución temporal del par motor durante el mecanizado para los orificios 1A y 2A.

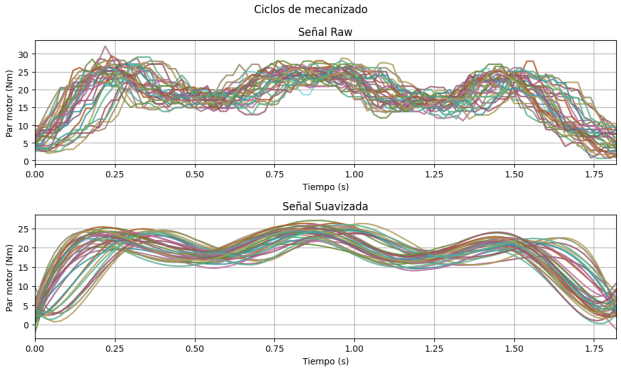

Figura 2: Evolución del par motor durante el mecanizado

3. Análisis previo

El desarrollo metodológico de esta investigación parte de la monitorización continua de la máquina descrita en la Sección 2 durante su funcionamiento habitual. La variable principal objeto de estudio es la señal temporal del par motor, registrada de forma secuencial para cada uno de los sucesivos ciclos de mecanizado. Dado que las señales en bruto adquiridas en entornos industriales suelen presentar ruido y variaciones temporales por diversos factores, se debe aplicar un flujo de preprocesado riguroso antes de la fase de etiquetado.

Este flujo está compuesto por varias fases entre las que destacan, en primer lugar, un suavizado de la señal mediante un filtro de Savitzky-Golay (Schafer (2011)), para atenuar el ruido de alta frecuencia inherente a la adquisición de datos y a las vibraciones mecánicas de la máquina, preservando al mismo tiempo la dinámica fundamental del proceso de corte. En la Figura 2 se observa la comparación entre la señal en crudo y la señal filtrada por este método. También se realiza un alineamiento de las señales en el eje temporal para garantizar que las diferentes fases de los ciclos de mecanizado estén sincronizadas entre todos los registros, permitiendo una comparación coherente.

3.1. Extracción de características (Feature Engineering)

Con el objetivo de caracterizar el comportamiento de la herramienta en cada ciclo se ha extraído un conjunto de métricas estadísticas a partir de la señal temporal de par (Wang et al. (2007)). Estas métricas se pueden agrupar conceptualmente tal y como se muestra en la Tabla 1.

Métrica	**Fórmula**
Tendencia central y dispersión	
Media	$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$
Mediana	$\text{Med}(x)$
Desviación típica	$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$
Varianza	σ^2
IQR	$Q_3 - Q_1$
Coef. de variación	$\text{CV} = \sigma/\bar{x}$
Energía y amplitud	
Máximo	$x_{\text{máx}} = \text{máx}_i\, x_i$
Mínimo	$x_{\text{mín}} = \text{mín}_i\, x_i$
Sumatorio	$S = \sum_{i=1}^{N} x_i$
Valor eficaz	$\text{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$
Factor de cresta	$C_f = x_{\text{máx}}/\text{RMS}$
Forma de la distribución	
Skewness	$\gamma_1 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i-\bar{x}}{\sigma}\right)^3$
Curtosis	$\gamma_2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i-\bar{x}}{\sigma}\right)^4 - 3$
Dinámica temporal	
Posición del pico	$i^* = \arg\text{máx}_i\, x_i$
Posición relativa del pico	$p_{\text{pico}} = i^*/N$
Índice de máxima pendiente	$i_\nabla = \arg\text{máx}_i\, \nabla x_i; \nabla x_i = x_{i+1} - x_i$
Tasa de cruces por la media	$\text{TC} = \frac{1}{N}\,\#\{i : \text{sgn}(x_{i+1} - \bar{x}) \neq \text{sgn}(x_i - \bar{x})\}$
Históricas	
Energía acumulada	$E_c^{\text{acum}} = \sum_{k=1}^{c} S_k$
Delta de la media	$\Delta\bar{x}_c = \bar{x}_c - \bar{x}_{c-1}$

Tabla 1: Características estadísticas extraídas de la señal temporal de par.

En las fórmulas de la Tabla 1, x_i denota el valor i-ésimo de la señal temporal de par dentro de un ciclo, siendo N el número total de muestras del ciclo. La función $\text{sgn}(\cdot)$ indica el signo de su argumento y $\#\{\cdot\}$ denota el cardinal (número de elementos) del conjunto. Por último, el subíndice c identifica el ciclo actual y k recorre los ciclos previos.

3.2. Reducción de dimensionalidad

El proceso de extracción de características descrito genera un espacio de datos de alta dimensionalidad, donde cada ciclo de mecanizado queda representado por un conjunto de 19 variables. Si bien es cierto que se dispone de información de calidad, puede haber cierta redundancia y el manejo de múltiples variables correlacionadas dificulta la interpretabilidad e imposibilita la visualización de los datos.

Dado que el objetivo de este trabajo es proponer un enfoque de clasificación semi-supervisado, resulta indispensable integrar la figura del experto humano en el bucle de decisión (*Human-in-the-loop*) tal y como introducen Chai and Li (2020). Para que un operario o ingeniero pueda visualizar el espacio de datos, verificar las agrupaciones del algoritmo y realizar ajustes manuales si fuera necesario, el espacio de representación debe ser comprensible.

Por tanto, se ha aplicado el Análisis de Componentes Principales (PCA) sobre el conjunto de características extraídas. Esta técnica consiste en una transformación lineal que permite proyectar los datos originales en un nuevo espacio de menor dimensión maximizando la varianza retenida (Sama et al. (2010)). En este trabajo, el tamaño del problema se ha reducido de las 19 variables originales a únicamente 3 componentes principales, lo que ha posibilitado la representación de todos los ciclos de mecanizado en un gráfico tridimensional y facilitado la identificación visual de clústeres de funcionamiento normal y valores atípicos.

4. Algoritmo implementado

El algoritmo implementado se estructura en una serie de bloques cuyo objetivo es identificar el comportamiento nominal de la máquina y aislar aquellos comportamientos anómalos que se alejen de este estado. Para abordar la falta de etiquetado previo y el fuerte desbalanceo natural de los datos industriales, el algoritmo se ha dividido en las siguiente fases, que agrupan los bloques representados en el flujograma de la Figura 3.

4.1. *Agrupamiento global en el espacio latente*

Partiendo del espacio tridimensional generado por el Análisis de Componentes Principales en el *Bloque 0*, el algoritmo comienza ajustando un Modelo de Mezclas Gaussianas (GMM) en el *Bloque 1* sobre la totalidad del conjunto de datos para capturar la distribución subyacente de los mismos. Se busca el número óptimo de componentes Gaussianas minimizando el Criterio de Información de Akaike (AIC), el cual busca un compromiso entre la bondad de ajuste del modelo y su complejidad. Se formula como $AIC = 2k - 2ln(\widehat{L})$, donde el término $-2ln(\widehat{L})$ representa la bondad de ajuste y k es el número total de parámetros libres estimados en el modelo, que varía según el número de componentes seleccionado bajo este criterio.

4.2. *Identificación y refinamiento del comportamiento normal*

Una vez obtenidas las distribuciones globales, el algoritmo debe discernir cuales de ellas representa el funcionamiento normal de la máquina. Esta identificación se realiza en el *Bloque 2* en base a un índice calculado como $S_K = \frac{\pi_K}{Tr(\Sigma_k)}$,

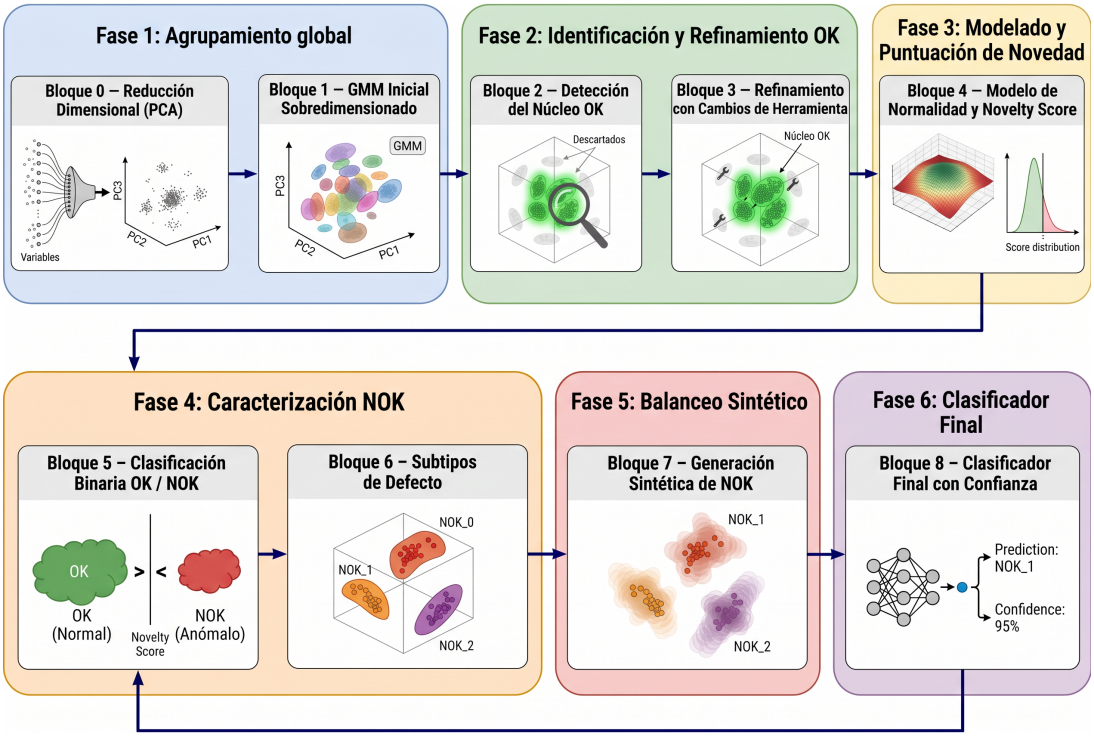

Figura 3: Flujograma por bloques del algoritmo

Fuente: Generado utilizando IA

donde K es el número de componentes o clusters seleccionados en la fase anterior, π_K es la proporción de datos sobre el total que agrupa cada componente y $Tr(\Sigma_k)$ es la traza de la matriz de covarianzas. Este índice responde a la asunción de que los clústeres que forman el núcleo de funcionamiento normal son aquellos de mayor tamaño y menor dispersión. A continuación, en el *Bloque 3* se integra información contextual del proceso para dotar de robustez al sistema. En particular, se utilizan los ciclos ejecutados inmediatamente después de un cambio de herramienta documentado para ajustar el núcleo de normalidad y garantizar que incluya los ciclos con herramientas nuevas o en buen estado.

4.3. *Modelado de normalidad*

Con el núcleo de ciclos normales identificado, en el *Bloque 4* se entrena un segundo modelo GMM exclusivo para describir cómo es el funcionamiento normal de la máquina. A continuación, se evalúa cada ciclo histórico frente a este modelo calculando un parámetro al que se ha llamado *Novelty Score*, que es la log-verosimilitud negativa ($NS = -lnP(x_i|modeloOK)$). Para hacer este parámetro interpretable, se aplica una estandarización robusta basada en la mediana y el rango intercuartílico de los ciclos normales. Finalmente, se emplea una transformación logarítmica simétrica ($LogNS = sign(NS_{norm}) \cdot ln(1 + |NS_{norm}|)$) para comprimir el rango de los valores atípicos y estabilizar el espacio de decisión. De esta forma, cuanto más cercano a cero sea el parámetro $LogNS$ más se asemeja un ciclo al funcionamiento normal, y cuanto mayor sea más anómalo es.

4.4. *Caracterización de estados anómalos*

En el *Bloque 5* se realiza una clasificación binaria OK/-NOK, de forma que aquellos ciclos cuyo $LogNS$ supera un umbral definido por el usuario, son clasificados temporalmente como anómalos (NOK). Posteriormente, en el *Bloque 6* el algoritmo aísla todos los ciclos identificados como NOK y ajusta sobre ellos un nuevo modelo GMM para detectar subgrupos de modos de fallo. Cada uno de estos sub-clústeres es modelado de forma independiente para permitir la posterior asignación manual de etiquetas semánticas.

4.5. *Balanceo de clases mediante generación sintética*

Uno de los mayores retos en la monitorización de herramientas es la escasez de ejemplos de fallo. Para mitigar esto y permitir el entrenamiento de modelos predictivos robustos, el algoritmo incorpora una fase de aumento de datos (*Data Augmentation*). En el *Bloque 7* se generan muestras sintéticas a partir de las distribuciones gaussianas ajustadas para cada clase anómala descubierta en la fase anterior. Para garantizar la validez física de estos datos artificiales, se emplea la técnica de muestreo por rechazo (*Rejection Sampling*). Una muestra sintética $\tilde{x}$ generada por la clase anómala j solo es aceptada si su probabilidad a posteriori de pertenecer a dicha clase supera un nivel de confianza mínimo C_{min} frente al resto de modelos, incluyendo el modelo OK. Matemáticamente, la muestra se acepta si: $P(NOK_j|\tilde{x}) \geq C_{min}$ y $argmax_c P(c|\tilde{x}) = NOK_j$, donde c es el número resultante de clústeres OK y NOK.

4.6. *Clasificador Bayesiano Final*

En la útlima fase del algoritmo (*Bloque 8*) se consolida un conjunto de entrenamiento balanceado que contiene datos reales y sintéticos. A diferencia de los enfoques discriminativos tradicionales, el algoritmo utiliza el teorema de Bayes cruzado con múltiples GMM (Bruneau et al. (2010)), generando hiperplanos de decisión no lineales. La expresión matemática

del teorema de Bayes aplicada a este problema se muestra en la Ecuación 1.

$$P(C_k|x) = \frac{P(x|C_k) \cdot P(C_k)}{\sum_j P(x|C_j) \cdot P(C_j)} \quad (1)$$

Donde x representa cada uno de los puntos del conjunto de entrenamiento y C_k es cada uno de los clústeres resultantes. Para resolver esta ecuación garantizando estabilidad computacional se aplica el logaritmo natural, resultando en la Ecuación 2.

$$lnP(C_k|x) = lnP(x|C_k) + lnP(C_k)) - ln(\sum_j P(x|C_j) \cdot P(C_j)) \quad (2)$$

Partiendo de las ecuaciones 1 y 2, el algoritmo desarrolla el numerador y denominador por separado:

1. Para cada estado C_K, calcula su probabilidad a priori $P(C_k) = \frac{N_k}{\sum_j Nj}$, siendo N_k el número de muestras para cada una de las clases. Esto representa la esperanza base de que ocurra dicho estado antes de hacer cualquier evaluación a los datos del ciclo.
2. Entrena un modelo GMM para cada uno de los clusters previamente detectados y evalúa cada punto, obteniendo su log-verosimilitud $lnP(x|C_k)$.
3. Para que las probabilidades de todos los estados sumen exactamente 100 %, se debe dividir el numerador de cada estado entre la suma de los numeradores de todos los estados. No obstante, al estar en formato logarítmico no se pueden sumar directamente. Se calcula la exponencial de cada uno, se suman y se calcula el logaritmo ($ln(\sum_j P(x|C_j) \cdot P(C_j)) = ln(\sum_j exp(lnP(x|C_j) + lnP(C_j))))$. Hacer directamente este cálculo causaría *Underflow*. Para evitarlo, se emplea la función *LogSumExp* que es computacionalmente estable aislando el valor máximo del exponente antes de realizar la suma.
4. Una vez hecho el cálculo de la Ecuación 2 se aplica la exponencial para recuperar el valor de la Ecuación 1 que representa una probabilidad entre 0 y 1 interpretable.

El resultado final es una matriz de probabilidades de pertenencia de cada ciclo de mecanizado a cada uno de los clusters. El algoritmo asigna a cada ciclo la etiqueta final del cluster con mayor probabilidad.

5. Resultados de Validación

En la fase de validación del algoritmo se parte de un conjunto de datos históricos previamente etiquetados por un equipo de ingenieros de proceso. Las etiquetas asignadas a los ciclos de mecanizado, y que representan el estado de la herramienta en cada uno de ellos son: *OK* para la herramienta en buen estado, *SCRAP* para herramienta rota o ausente y *CRIT* para herramienta con desgaste crítico.

La validación del algoritmo se ha realizado mediante diferentes experimentos modificando el umbral de *LogNS* que controla la separación entre los grupos OK y NOK, como se aprecia en la Figura 4. Este parámetro es el que el operario o ingeniero supervisa y customiza para ajustar el modelo de clasificación. Los resultados obtenidos con el modelo entrenado en cada experimento se ha comparado con la realizada por los expertos.

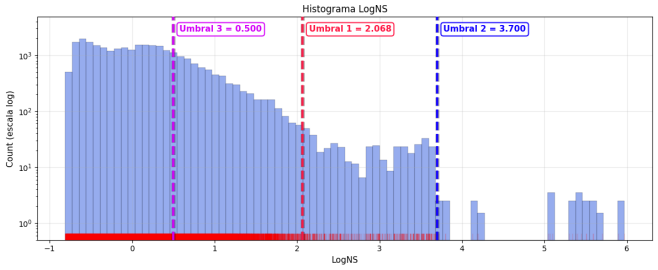

Figura 4: Histograma LogNS y umbrales experimentados

5.1. Experimento 1

En el primer experimento se ha ejecutado el algoritmo con un umbral de *LogNS* en el percentil 95 de su distribución. El resultado de la clasificación se observa en la Figura 5.

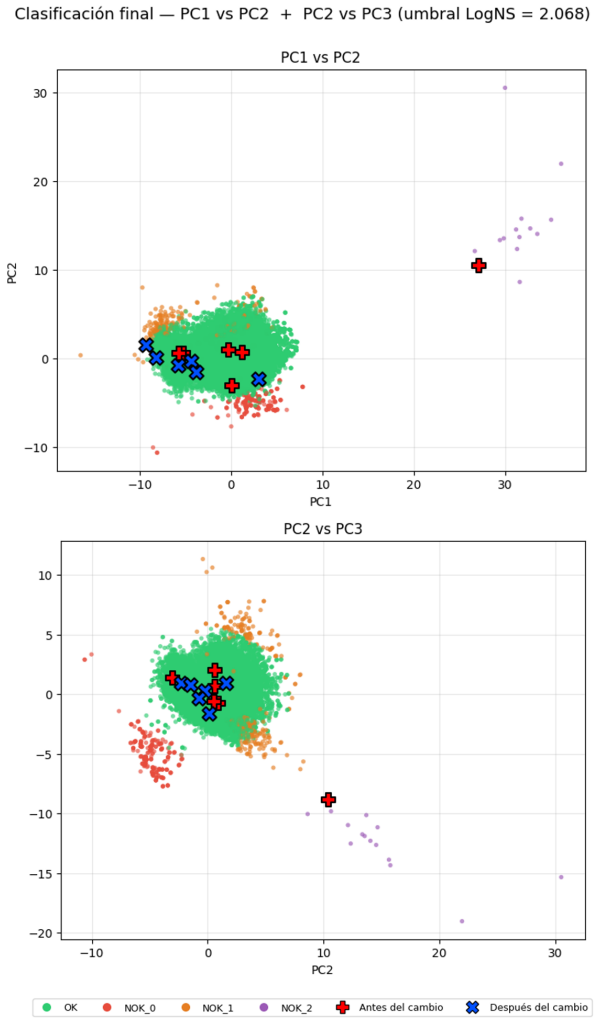

Figura 5: Gráficos PCA experimento 1

Referencia					
	OK	28565	209	252	0
	CRIT	0	0	0	15
	SCRAP	0	3	4	0
		OK	NOK_0	NOK_1	NOK_2
		Predicción			

Tabla 2: Resultados experimento 1

5.2. *Experimento 2*

En el segundo experimento se ha aumentado el umbral de *LogNS* a 3.7, obteniéndose los resultados de la Figura 6. Como se observa en la Tabla 3, el algoritmo ha sido más restrictivo en este caso, clasificando una mayor cantidad de puntos como OK.

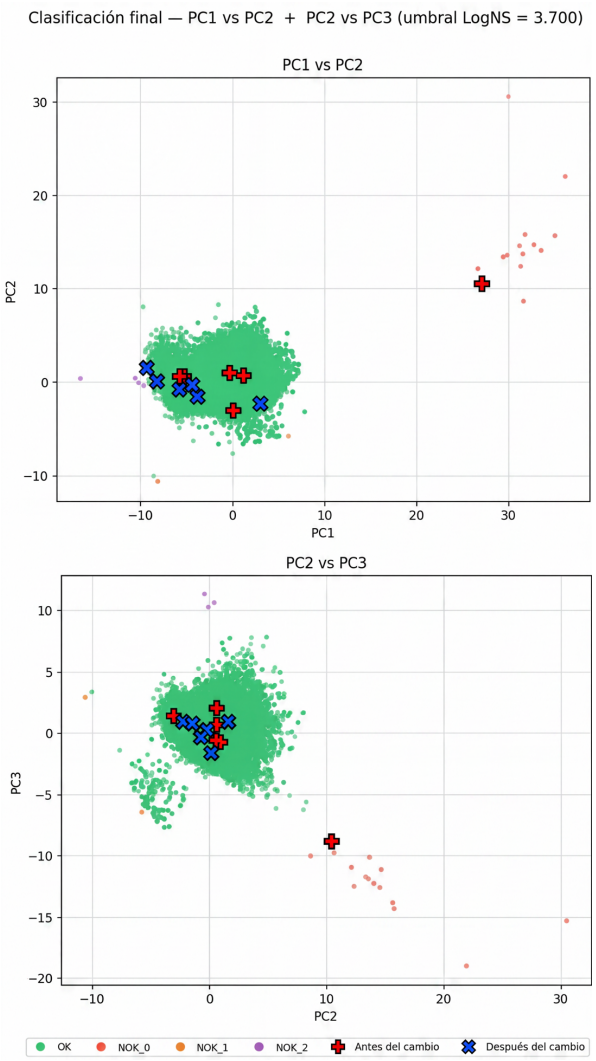

Figura 6: Gráficos PCA experimento 2

Referencia				
	OK	29018	0	8
	CRIT	0	15	0
	SCRAP	0	0	7
		OK	NOK_0	NOK_1
		Predicción		

Tabla 3: Resultados experimento 2

5.3. *Experimento 3*

En el último experimento se ha reproducido el caso contrario al experimento 2, es decir, se ha reducido el umbral de *LogNS* a 0.5 para poder apreciar su influencia en el desempeño del algoritmo, obteniéndose los resultados de la Figura 7. Como se observa en la Tabla 4, al haber establecido un umbral menos restrictivo, el algoritmo ha reducido el tamaño del núcleo OK y ha clasificado más puntos como anómalos.

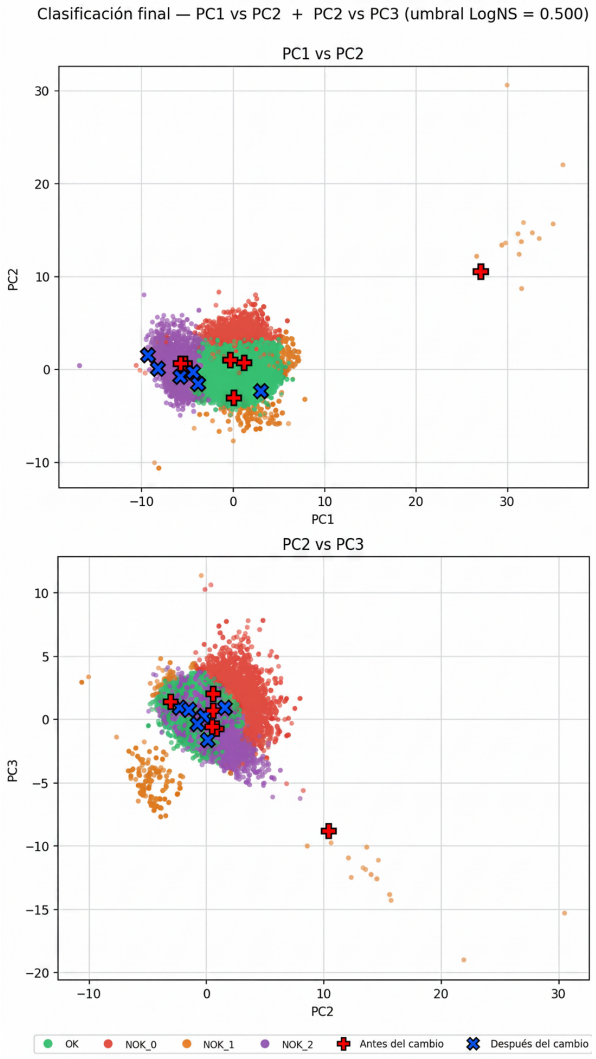

Figura 7: Gráficos PCA experimento 3

Referencia					
	OK	22877	2887	573	2689
	CRIT	0	0	15	0
	SCRAP	0	2	4	1
		OK	NOK_0	NOK_1	NOK_2
		Predicción			

Tabla 4: Resultados experimento 3

Los experimentos de validación realizados ponen de manifiesto que el umbral de *LogNS* es el factor determinante de ajuste del comportamiento del algoritmo. A partir de los gráficos de las Figuras 5 a 7 el operario o ingeniero debe decidir el valor del umbral dado que en una nueva implementación no dispondrá de las Tablas 2 a 4 comparativas. A la vista de

los resultados, destacando la Figura 5, cabe la posibilidad de la existencia de otros modos de fallo no contemplados por los expertos de planta en sus etiquetas que sí han sido detectados por el algoritmo.

6. Conclusiones

En este trabajo se ha presentado y validado un algoritmo de clasificación semi-supervisado diseñado para superar las limitaciones clásicas del mantenimiento predictivo en entornos industriales reales: la ausencia de datos históricos etiquetados, así como el fuerte desbalanceo entre ciclos de funcionamiento normal y anómalo. Con este, se obtiene un modelo listo para clasificación en producción o etiquetado en la fase de modelado.

Con la metodología propuesta, se han utilizado técnicas de reducción de dimensionalidad (PCA) que han permitido visualizar los datos en un espacio tridimensional y se ha modelado de forma precisa la topología de estos mediante Modelos de Mezclas Gaussianas (GMM). La introducción de una métrica de novedad estandarizada y en escala logarítmica, a la que se ha denominado *LogNS*, ha resultado ser un mecanismo robusto para separar el núcleo de funcionamiento normal de los diferentes modos de fallo.

Los resultados observados en los diferentes experimentos han validado que la sensibilidad del algoritmo depende en gran medida de la correcta calibración de *LogNS*. La posibilidad de poder modificar este umbral junto con el hecho de proyectar los resultados en un espacio tridimensional interpretable y proporcionar un nivel de confianza probabilístico para cada predicción facilita la integración del factor humano en el bucle de decisión contribuyendo a la semi-supervisión del algoritmo.

7. Trabajos futuros

Como líneas de trabajo futuro, se plantean tres ejes principales: en primer lugar, la introducción en el algoritmo de un índice que indique el porcentaje de vida restante de la herramienta; en segundo lugar, la evolución del modelo mediante algoritmos avanzados de *Deep Learning*, utilizando redes neuronales *Long Short-Term Memory* (LSTM) para el reconocimiento de patrones; y, finalmente, la exploración de herramientas basadas en *Large Language Models* (LLM) para el análisis de grandes volúmenes de datos históricos por etiquetar, con el objetivo de mitigar el desbalanceo inherente a los datos adquiridos y optimizar la rigurosidad de las etapas de preprocesamiento.

Agradecimientos

Los autores agradecen a la Fundación para el Desarrollo y la Innovación (FDI) y a la Universitat Politècnica de València la financiación proporcionada a través de la Ayuda UPV Predoctoral Subprograma 2 (20251065).

Referencias

Belmokhtar, S., Bratcu, A., Dolgui, A., 2006. Modular machining line design and reconfiguration: Some optimization methods. In: Kordic, V., Lazinica, A., Merdan, M. (Eds.), Manufacturing the Future. IntechOpen, London, Ch. 5.
URL: https://doi.org/10.5772/5047
DOI: 10.5772/5047

Bruneau, P., Gelgon, M., Picarougne, F., 2010. Parsimonious reduction of gaussian mixture models with a variational-bayes approach. Pattern Recognition 43 (3), 850–858.
URL: https://www.sciencedirect.com/science/article/pii/S0031320309003021
DOI: https://doi.org/10.1016/j.patcog.2009.08.006

Chai, C., Li, G., 2020. Human-in-the-loop techniques in machine learning. IEEE Data Eng. Bull. 43, 37–52.
URL: https://api.semanticscholar.org/CorpusID:240291940

Dhasaian, M. R. S., Parnitha, Prasana, Naini, 11 2025. Unsupervised and semi supervised machine learning frameworks for multi class tool wear recognition. American Journal of Management and IOT Medical Computing 4, 22–28.
URL: https://ajmimc.com/journal/index.php/ajmimc/article/view/129
DOI: 10.64751/AJMIMC.2025.V4.N4(1).PP22-28

Fomin, D., Lalanda, P., Morand, D., 2024. Automating the production of machine learning models for imbalanced datasets - application to industry 4.0. 2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events, PerCom Workshops 2024, 39–44.
URL: https://ieeexplore.ieee.org/document/10503057
DOI: 10.1109/PERCOMWORKSHOPS59983.2024.10503057

Kuntoğlu, M., Aslan, A., Pimenov, D. Y., Üsame Ali Usca, Salur, E., Gupta, M. K., Mikolajczyk, T., Giasin, K., Kapłonek, W., Sharma, S., 12 2020. A review of indirect tool condition monitoring systems and decision-making methods in turning: Critical analysis and trends. Sensors 2021 21, 108.
URL: https://www.mdpi.com/1424-8220/21/1/108/htmhttps://www.mdpi.com/1424-8220/21/1/108
DOI: 10.3390/S21010108

Mallioris, P., Aivazidou, E., Bechtsis, D., 6 2024. Predictive maintenance in industry 4.0: A systematic multi-sector mapping. CIRP Journal of Manufacturing Science and Technology 50, 80–103.
URL: https://www.sciencedirect.com/science/article/pii/S1755581724000221#ab0010
DOI: 10.1016/J.CIRPJ.2024.02.003

Sama, A., Pardo-Ayala, D. E., Cabestany, J., Rodríguez-Molinero, A., 2010. Time series analysis of inertial-body signals for the extraction of dynamic properties from human gait. In: The 2010 International Joint Conference on Neural Networks (IJCNN). pp. 1–5.
DOI: 10.1109/IJCNN.2010.5596663

Schafer, R., Jul. 2011. What Is a Savitzky-Golay Filter? [Lecture Notes]. IEEE Signal Processing Magazine 28 (4), 111–117.
DOI: 10.1109/MSP.2011.941097

Thomas, D., Weiss, B., 12 2021. Maintenance costs and advanced maintenance techniques in manufacturing machinery: Survey and analysis. International journal of prognostics and health management 12.
URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC9890517/
DOI: 10.36001/IJPHM.2021.V12I1.2883

Tran, M. Q., Doan, H. P., Vu, V. Q., Vu, L. T., 2 2023. Machine learning and iot-based approach for tool condition monitoring: A review and future prospects. Measurement 207, 112351.
URL: https://www.sciencedirect.com/science/article/pii/S0263224122015470#s0005
DOI: 10.1016/J.MEASUREMENT.2022.112351

Wang, X., Wirth, A., Wang, L., 10 2007. Structure-based statistical features and multivariate time series clustering. Proceedings - IEEE International Conference on Data Mining, ICDM, 351–360.
URL: https://ieeexplore.ieee.org/document/4470259
DOI: 10.1109/ICDM.2007.103