

Detección de fallos mediante clasificación multi-etiqueta para flujos de datos

Aurora Esteban, Amelia Zafra, Sebastián Ventura

Dept. de Informática y Análisis Numérico

Instituto Andaluz de Investigación en Data Science and Computational Intelligence (DaSCI), Universidad de Córdoba
Córdoba, España

aestebant@uco.es, azafra@uco.es, sventura@uco.es

Resumen—El Mantenimiento Predictivo se ha establecido como un componente crítico de la Industria 4.0, optimizando la eficiencia operativa y minimizando los costos de mantenimiento gracias a los avances en Aprendizaje Automático a partir de datos sensorizados. Sin embargo, abordar fallos concurrentes en un sistema complejo donde múltiples eventos de mantenimiento pueden ocurrir simultáneamente sigue siendo un desafío en el área. Para abordar este problema, este trabajo propone un ensemble *online* de árboles de Hoeffding multi-etiqueta, un novedoso modelo para el aprendizaje incremental en sistemas industriales que generan grandes cantidades de datos y pueden verse afectados a lo largo del tiempo por diferentes condiciones que afecten a su funcionamiento. La propuesta se valida en un problema de mantenimiento predictivo multi-etiqueta para un sistema naval. Los resultados experimentales revelan el potencial del enfoque *online* en el campo, con mejoras del 30 % respecto a trabajos previos basados en aprendizaje por lotes, y la efectividad de la propuesta para aprender las correlaciones entre fallos con mejoras del 20 % respecto a otros modelos *online*.

Index Terms—mantenimiento predictivo, flujos de datos, aprendizaje multi-etiqueta, árboles de decisión

I. INTRODUCCIÓN

La Industria 4.0 representa una nueva era en los procesos de manufactura e industriales, impulsada por nuevas tecnologías como el Internet de las Cosas, *Internet of Things* (IoT), el análisis de grandes volúmenes de datos y el Aprendizaje Automático (AA), que han resultado en la integración de sistemas ciberfísicos para transformar la industria [1]. En este contexto, el Mantenimiento Predictivo, *Predictive Maintenance* (PdM) es un componente crítico que optimiza la eficiencia operativa y minimiza los costos de mantenimiento [2]. Partiendo de modelos predictivos y tecnologías IoT, el PdM mejora la fiabilidad del equipo mediante la detección temprana de posibles fallos. Este enfoque proactivo minimiza el tiempo de inactividad no planificado causado por mantenimientos correctivos y evita los reemplazos prematuros y tiempos de inactividad innecesarios del mantenimiento preventivo que no tienen en cuenta la condición actual de la maquinaria.

El PdM ha sido un campo de investigación muy activo en los últimos años por su interés para la industria, así como por los avances en AA que permiten modelos cada vez más eficientes en aprender de grandes cantidades de datos. Así, existen múltiples tareas de PdM abordable con AA, como la estimación de la vida útil restante, la predicción de fallos o el diagnósticos de averías [2]. Entre éstas, este trabajo se

centra en la tarea de predecir múltiples fallos, potencialmente correlacionados y que podrían ocurrir concurrentemente dentro de un sistema o un conjunto de equipos interconectados. Este problema se aborda con la Clasificación Multi-Etiqueta, *Multi-Label Classification* (MLC) [3], un paradigma del AA que ha demostrado mejorar el rendimiento en problemas donde los patrones pueden tener más de una clase asociada. La predicción simultánea de fallos es una tarea poco explorada dentro del PdM, principalmente debido a una mayor complejidad en la modelización de interacciones entre múltiples componentes [4], [5]. Las propuestas previas se han centrado en transformaciones a problemas clásicos multi-clase [5], por lo que pierden la capacidad de aprender la posible interacción entre fallos; o en modelos de Aprendizaje Profundo, *Deep Learning* (DL) diseñados para producir múltiples salidas [6], que requieren grandes cantidades de datos y recursos computacionales para aprender. Además, en ambos casos, los modelos son estáticos: una vez entrenados, no se pueden actualizar eficientemente si nuevas condiciones afectan al sistema industrial monitorizado.

Tradicionalmente, los modelos PdM se han entrenado asumiendo que la distribución de datos siempre reflejará el conjunto de observaciones que el modelo tendrá que predecir [2]. Sin embargo, en los entornos industriales actuales cada vez se trabaja con mayores volúmenes de datos y velocidad a la que se generan. En este contexto, es posible que el sistema eventualmente se enfrente a nuevos escenarios no contemplados en el conjunto de entrenamiento inicial, lo que conduce a fluctuaciones en la distribución aprendida de entradas y salidas que pueden variar con el tiempo. Esto hace que el sistema PdM se comporte de manera errática, ya que recibirá información inusual de la entrada, lo que afectará directamente a la precisión de las predicciones. En este contexto, el aprendizaje *online* para flujos de datos tiene gran potencial. En este subcampo del AA se desarrollan algoritmos capaces de aprender incrementalmente conforme nuevos datos de entrenamiento llegan al sistema. A diferencia del aprendizaje por lotes tradicional, el aprendizaje *online* permite actualizaciones instantáneas y eficientes al modelo.

En este trabajo se abordan conjuntamente los dos desafíos del PdM comentados con un ensemble *online* de árboles de Hoeffding multi-etiqueta, una propuesta para la predicción simultánea de fallos de sistemas industriales multicomponente en entornos dinámicos. Este nuevo enfoque combina el apren-

dizaje *online* y la MLC mediante Árbol de Decisión, *Decision Trees* (DTs) incrementales, una familia de métodos *online* diseñados para actualizar la estructura del DT dinámicamente conforme llegan los datos. Además, el enfoque multi-etiqueta propuesto es capaz de aprender la correlación entre etiquetas incluso en un entorno PdM donde las averías pueden ser infrecuentes y por tanto se tiene un gran desbalanceo de datos. La propuesta es validada con un caso real de PdM [7]: la predicción de fallos en una fragata, concretamente en los componentes de su Sistema de Propulsión Naval (SPN).

El resto del trabajo se organiza de la siguiente manera. La Sección II hace un repaso del trabajo previo en PdM multi-etiqueta y *online*. La Sección III presenta la propuesta de este trabajo. La Sección IV describe el estudio de caso y los resultados obtenidos. Finalmente la Sección V analiza las conclusiones obtenidas.

II. TRABAJO RELACIONADO

Aunque los modelos mayoritarios de PdM están diseñados para detectar fallos excluyentes entre sí, la creciente complejidad de los sistemas ciber-físicos está haciendo crecer la relevancia de los modelos para detectar fallos en sistemas multi-componente mediante MLC [2]. La aproximación más común es la transformación del problema para aplicar modelos de AA tradicional. Destacan las transformaciones basadas en Máquina de Vectores Soporte, *Support Vector Machines* (SVMs) aplicadas a pilas de combustible de óxido sólido [8] o sistemas de propulsión naval [5], los modelos basados en vecinos más cercanos para trabajar con las señales de vibración provenientes de rotores [9], o los modelos basados en DT por sus propiedades interpretativas [10]. La otra gran línea para abordar la detección de fallos simultáneos con MLC son modelos de DL capaces de producir múltiples salidas. En este sentido destacan los modelos basados en redes convolucionales, debido a su habilidad para capturar relaciones espaciales en datos secuenciales mediante la transformación de las series de datos multi-variantes de los sistemas industriales monitorizados a imágenes. Así, se han aplicado a rotores en líneas de producción [6] o a trenes de alta velocidad [11] entre otros.

Todos los trabajos de PdM multi-etiqueta encontrados en la literatura se centran en el AA tradicional que aprende por lotes. Y, en general, pocos trabajos exploran otros esquemas de aprendizaje más apropiados para entornos en los que se generan grandes cantidades de datos y nuevos escenarios pueden surgir. No obstante, algunos trabajos exploran el potencial del aprendizaje *online* para predicción de fallos auto-excluyentes entre sí. En [12] se propone un modelo online basado en la teoría de la resonancia adaptativa aplicado los rodamientos de aerogeneradores. En [13] se sigue otro enfoque *online* basado en *clustering* adaptativo sobre los datos de funcionamiento de las puertas de una flota de trenes. Nuestro trabajo continúa esta prometedora línea abordando la MLC desde una perspectiva *online* para proponer un novedoso sistema PdM para detección de fallos simultáneos que se adapta dinámicamente a las nuevas condiciones que puedan aparecer a lo largo del tiempo.

III. ENSEMBLE ONLINE DE ÁRBOLES DE Hoeffding MULTI-ETIQUETA

En esta sección se presenta la propuesta de ensemble de árboles para la predicción de fallos simultáneos, con la especificación del modelo base en la Sección III-A y las características del ensemble online en la Sección III-B.

III-A. Árbol de Hoeffding Multi-Etiqueta

Los DTs clásicos se basan en el particionamiento recursivo del conjunto de entrenamiento, asumiendo que todas las instancias de entrenamiento están disponibles, para obtener subconjuntos lo más puros posible para un objetivo dado, o conjunto de etiquetas en MLC. Por el contrario, los DTs incrementales aprenden de flujos de datos a medida que los datos etiquetados están disponibles, actualizando la estructura del árbol incrementalmente, es decir, generando nuevas ramas o podando ramas que ya no son útiles. Este paradigma permite que el modelo predictivo produzca diagnósticos en cualquier momento, incluido al comienzo del período de monitoreo o cuando aparecen nuevas fallos. La propuesta de este trabajo sigue los principios de Árbol de Hoeffding, *Hoeffding Trees* (HTs) previos [14], [15] e incorpora adaptaciones específicas para MLC.

El árbol base de nuestra propuesta se construye a medida que recibe datos multi-etiqueta de forma continua. Así, el modelo empieza con un único nodo que es alimentado por las instancias que se van sucediendo. Cuando las instancias recibidas son suficientes y suficientemente diversas entre sí, el nodo se divide en dos sub-ramas utilizando como punto discriminatorio el atributo que minimice la entropía multi-etiqueta del conjunto de instancias observadas por el nodo. Para determinar el momento en el que hacer la partición se utiliza el límite de Hoeffding [15] con adaptaciones para considerar la co-ocurrencia entre etiquetas:

$$\epsilon_{spl} = \sqrt{\frac{\log_2(|L|)^2 \ln(1/\delta_{spl})}{2W}} \quad (1)$$

donde $|L|$ es el número de etiquetas observadas en el nodo n , W es la cantidad de instancias recibidas en n y δ_{spl} es un parámetro que determina el error que se permite. Si la diferencia entre ganancias de información G de los dos puntos de corte más prometedoras s_1 y s_2 superan este límite, $G(s_1) - G(s_2) > \epsilon_{spl}$, se producirá la partición del nodo en favor de s_1 , consiguiendo así el crecimiento incremental del DT. A su vez, cada ganancia de información G es una diferencia de entropías antes H_0 y después $H(s)$ de que los datos observados en n se dividan en el punto de corte a evaluar, es decir, una medida de cómo de bien el camino hasta dicho nodo n discrimina entre conjuntos de etiquetas. Por tanto, el cálculo de las entropías debe tener en cuenta la co-ocurrencia entre etiquetas, para lo cual, su distribución se modela siguiendo una distribución de Bernoulli multi-variante [16] que conduce al cálculo de la entropía multi-etiqueta como:

$$H = \sum_{i=1}^{\mathcal{L}} -p(l_i) \log_2(p(l_i)) - (1 - p(l_i)) \log_2(1 - p(l_i)) \quad (2)$$

donde $p(l_i)$ es la probabilidad de tener la i° etiqueta del espacio de etiquetas \mathcal{L} en el nodo n .

El árbol multi-etiqueta utiliza un proceso de clasificación dinámico en las hojas para mantener un buen rendimiento mientras aún no se tiene suficiente evidencia para particionar el nodo. Es decir, la clasificación se produce en primera instancia según el camino que el dato de entrada recorre en el árbol y, una vez que llega a un nodo, se tiene un proceso de clasificación adicional que evoluciona según la complejidad del flujo de datos, pudiendo ir desde predecir un conjunto de etiquetas fijo a entrenar un clasificador multi-etiqueta. Así, se definen cuatro escenarios de clasificación en cada nodo hoja en función de su entropía H_0 y cardinalidad W (número de instancias recibidas): en primer lugar, cuando la entropía es nula ($H_0 = 0$), lo que indica una separación perfecta de instancias con el mismo conjunto de etiquetas. En este caso no es necesario el entrenamiento de clasificadores multi-etiqueta adicionales, ya que la predicción del conjunto de etiquetas mayoritario es precisa y eficiente computacionalmente. En segundo lugar, se contempla un estado intermedio con $H_0 > 0$ y $W < \eta$, donde η es un parámetro que define el umbral de cardinalidad. En este caso se emplea un clasificador de baja cardinalidad $\gamma_{\downarrow C}$, concretamente basado en vecinos más cercanos, para detectar relaciones entre etiquetas en escenarios con pocos datos. En tercer lugar, se contempla un estado ruidoso donde la ganancia de información es baja a pesar de un alto número de instancias ($G < \epsilon_{spl}$ y $W > \eta$), se necesita un clasificador más exigente en datos y computacionalmente costoso $\gamma_{\uparrow C}$, concretamente un ensemble de tipo bagging, para obtener un mejor rendimiento en flujos de datos más grandes manteniendo la eficiencia computacional. Por último, si la entropía aumenta lo suficiente para pasar el límite de Hoeffding ϵ_{spl} , se considera que el nodo está listo para dividirse tal y como se ha descrito anteriormente.

La última característica diferenciadora del árbol multi-etiqueta es la poda de ramas cuando se detecta que éstas ya no representan el estado actual del flujo de datos, o lo que es lo mismo, se ha producido una deriva de concepto. Para tal fin, cada rama del árbol cuenta un detector de cambio de concepto α que monitoriza el rendimiento en el MLC para determinar si el punto de corte que la caracteriza sigue siendo válida para la distribución de datos actual. α se basa en el algoritmo Ventaneo Adaptativo, *Adaptive Windowing* (ADWIN) [17] para detectar cambios en la distribución de datos sobre una ventana de instancias pasadas recibidas en el nodo. Si se detectara una degradación incipiente en el rendimiento, el nodo n inicia un árbol en su segundo plano $DT(n)'$, que continúa creciendo de la misma manera que el árbol principal, alimentado con instancias que llegan al nodo. En el momento en el que se confirma la deriva de concepto, $DT(n)'$ reemplaza al subárbol que existiera en n , actualizando el modelo principal. Para determinar el cambio de concepto, se utiliza nuevamente el límite de Hoeffding, en este caso comparando los errores del árbol principal e y alternativo e' , ponderados con las instancias que cada uno ha observado, W

y W' respectivamente:

$$\epsilon_{alt} = \sqrt{\frac{2e(1-e')(W+W') \ln(2/\delta_{alt})}{W \cdot W'}} \quad (3)$$

Si la diferencia entre errores es mayor que el límite $|e - e'| > \epsilon_{alt}$, significa que se ha confirmado el cambio de concepto, con una confianza de $1 - \delta_{alt}$, por lo que la estructura del árbol base debe actualizarse.

III-B. Ensemble Online

El DT incremental se ha definido en la sección anterior como un árbol de Hoeffding con mejoras específicas para la MLC. Sin embargo, en problemas altamente desbalanceados como la predicción de fallos en PdM, un sólo modelo puede no ser suficiente. Para abordar este problema en un escenario de flujo de datos se debe tener en cuenta que el modelo, además de tener potencia predictiva deben ser eficientes procesando los datos cuando éstos llegan uno a uno y potencialmente sin fin. Para abordar estos retos, en este trabajo se propone la integración de este DT en un ensemble *online* de *bagging* combinado con sub-espacios de características como se muestra en el Algoritmo 1. En este entorno *online*, *bagging* se implementa siguiendo una distribución de *Poisson*(λ) con $\lambda = 1$ para distribuir cada instancia con diferente peso a cada uno de los M modelos que componen el ensemble. Además, para añadir más poder de diversificación al ensemble a la vez que se mantiene controlado el coste computacional, se extiende a MLC el método *Streaming Random Patches* (SRP) [18], que no entrena cada modelo del ensemble con la misma instancia, sino que cada uno verá una representación distinta de la misma que incluya únicamente una porción σ de las características que componen el espacio de entrada, escogidas de forma aleatoria. Así, como se ve en el Algoritmo 1, este ensemble tiene dos parámetros que se fijan siguiendo los valores típicos de la literatura: $M = 10$ y $\sigma = 60\%$ del número de características original.

Algoritmo 1: Algoritmo propuesto

Parámetros:

M : número de modelos base

σ : tamaño del sub-espacio de características

Símbolos:

S : flujo de datos de entrenamiento

DT : cada árbol de Hoeffding multi-etiqueta base

w : peso de la instancia en el entrenamiento

F : espacio de características en la entrada

```

1 for  $S_i \equiv (X_i, Y_i) \in S$  do
2   for  $DT_i, i \in \{1, 2, \dots, M\}$  do
3      $w \leftarrow Poisson(1)$ 
4      $X'_i \leftarrow$  muestreo aleatorio de  $\sigma$  sobre  $F$  para  $X_i$ 
5     Entrena  $DT_i$  con  $(X'_i, Y_i) \times w$ 
6   end for
7 end for
```

IV. CASO DE ESTUDIO: MANTENIMIENTO PREDICTIVO MULTI-ETIQUETA DE UNA FRAGATA

Este caso de estudio es un conjunto de datos públicos de un simulador numérico validado con datos reales de una fragata [7], específicamente su SPN. La fragata está impulsada por dos hélices de paso controlable que son alimentadas por una turbina de gas a través de una caja de engranajes interconectada. Además, cada eje cuenta con un motor de propulsión eléctrico. Dos embragues entre la caja de engranajes y los dos motores de propulsión eléctricos y otro embrague entre la turbina de gas y la caja de engranajes aseguran la posibilidad de usar dos tipos diferentes de motores principales: propulsión eléctrica o de gas. La energía eléctrica es suministrada por cuatro generadores diésel. El simulador proporciona 25 medidas asociadas con el SPN, que serán las entradas del sistema PdM, todas ellas telemétricas y de tipo numérico continuo. La simulación completa cubre alrededor de 589.000 ciclos de trabajo. Sin embargo, trabajos previos [5], [7] argumentan que mientras la máquina no alcance la velocidad de crucero, las mediciones no son representativas ni útiles. Por lo tanto, para reproducir las condiciones experimentales lo más fielmente posible, este artículo solo considerará registros a velocidad de crucero (15 nudos), lo que deja el flujo de datos en 65.473 ciclos de trabajo. Atendiendo al objetivo de aprendizaje, el conjunto de datos original proporciona el estado de degradación de los diferentes componentes del SPN: turbina de gas (GT), compresor de turbina de gas (GTC), casco (HLL) y hélice (PRP). Los detalles de los modelos de degradación en el tiempo y la velocidad de los componentes se describen en [7], y a grandes rasgos son cuatro variables continuas, kK_t , kM_c , kH y kK_t , que representan el estado de salud de los componentes asociados. A partir de esto, siguiendo el proceso utilizado en [5], [7], se establecen umbrales para estas variables para crear estados binarios (sano o degradado) y generar el esquema de múltiples etiquetas que aprenderá el sistema PdM:

$$\begin{aligned} GT : kM_t &\in \begin{cases} [0,97 - 0,99) - \text{degradado} \\ [0,99 - 1] - \text{sano} \end{cases} \\ GTC : kM_c &\in \begin{cases} [0,95 - 0,98) - \text{degradado} \\ [0,98 - 1] - \text{sano} \end{cases} \\ HLL : kH &\in \begin{cases} [1 - 1,1) - \text{degradado} \\ (1,1 - 1,2] - \text{sano} \end{cases} \\ PRP : kK_t &\in \begin{cases} [0,9 - 0,95) - \text{degradado} \\ [0,95 - 1] - \text{sano} \end{cases} \end{aligned} \quad (4)$$

A partir de estas ecuaciones, el espacio de etiquetas \mathcal{L} se define como:

$$\mathcal{L} = (GT = l_1, GTC = l_2, HLL = l_3, PRP = l_4) \quad (5)$$

$$\forall l_i \in \{0, 1\}.$$

Y tras aplicar las funciones de (4) sobre el flujo de datos compuesto por 65.473 ciclos, se obtiene un dataset multi-label

con 4 etiquetas, cardinalidad de 2,00 y densidad de 0,50. A continuación se definen las condiciones experimentales en las que se evalúa la propuesta en este estudio y los resultados obtenidos.

IV-A. Condiciones de evaluación

Debido a la naturaleza incremental de los flujos de datos, la forma canónica de evaluar el rendimiento al trabajar con ellos sigue un esquema de evaluación seguido de entrenamiento conocido como *prequential evaluation* [19]. En este trabajo, se aplica esta metodología a todos los experimentos estableciendo un factor de olvido de $\alpha = 0,995$.

Para comparar el rendimiento entre algoritmos, se estudian 12 métricas que evalúan la corrección total y parcial de la predicción multi-etiqueta [3], definidas sobre L etiquetas y n instancias. Las métricas basadas en ejemplos evalúan la diferencia entre los conjuntos de etiquetas reales y predichos, promediados sobre n . Dado un conjunto de etiquetas verdaderas $Y_i = y_{i1} \dots y_{iL}$ y uno predicho $Z_i = z_{i1} \dots z_{iL}$, las métricas basadas en ejemplos consideradas son:

$$\begin{aligned} \text{Subset accuracy} &= \frac{1}{n} \sum_{i=0}^n 1|Y_i = Z_i \\ \text{Hamming loss} &= \frac{1}{nL} \sum_{i=0}^n \sum_{l=0}^L 1|y_{il} \neq z_{il} \\ \text{Example-based F1} &= \frac{1}{n} \sum_{i=0}^n \frac{2|Y_i \cap Z_i|}{|Y_i| + |Z_i|} \end{aligned} \quad (6)$$

Las métricas basadas en etiquetas miden el rendimiento por etiqueta, que puede ser micro o macro promediadas dependiendo de si se utilizan las estadísticas conjuntas para todas las etiquetas, o las medidas por etiqueta se promedian en un único valor. Desde la perspectiva macro se consideran el F1 macro-promediado como la media armónica de la macro-precisión y la macro-recall, definidas como:

$$\begin{aligned} \text{Macro-precisión} &= \frac{1}{L} \sum_{l=0}^L \frac{tp_l}{tp_l + fp_l} \\ \text{Macro-recall} &= \frac{1}{L} \sum_{l=0}^L \frac{tp_l}{tp_l + fn_l} \end{aligned} \quad (7)$$

dada para cada etiqueta l sus verdaderos positivos $tp_l = \sum_{i=0}^n 1|y_{il} = z_{il} = 1$, verdaderos negativos $tn_l = \sum_{i=0}^n 1|y_{il} = z_{il} = 0$, falsos positivos $fp_l = \sum_{i=0}^n 1|y_{il} = 0, z_{il} = 1$, y falsos negativos $fn_l = \sum_{i=0}^n 1|y_{il} = 1, z_{il} = 0$. De igual forma, se define la F1 micro-promediada como la media armónica de la micro-precisión y la micro-recall, definidas como:

$$\begin{aligned} \text{Micro-precisión} &= \frac{\sum_{l=0}^L tp_l}{\sum_{l=0}^L tp_l + \sum_{l=0}^L fp_l} \\ \text{Micro-recall} &= \frac{\sum_{l=0}^L tp_l}{\sum_{l=0}^L tp_l + \sum_{l=0}^L fn_l} \end{aligned} \quad (8)$$

En cuanto los métodos incluidos en el estudio experimental, se incluyen por un lado métodos *online* ya que nuestra

propuesta pertenece a este paradigma de aprendizaje y se debe comparar su rendimiento en el mismo. Por otro lado, el caso de estudio de este trabajo sólo se ha abordado en la literatura previa con modelos de AA tradicional de aprendizaje por lotes u *offline*, por lo que nuestra propuesta también se compara con estos trabajos adaptando la evaluación para conseguir unas condiciones lo más igualitarias posible. La comparativa *online* incluye los métodos del estado del arte directamente relacionados con nuestra propuesta: los DTs multi-etiqueta previos *Multi-Label Hoeffding Tree* (MLHT) [15] e *incremental Structured-Output Prediction Tree* (iSOUPT) [20]. La configuración de sus parámetros corresponde con los parámetros por defecto sugeridos por los autores de las propuestas: *grace period: 200, delta: 1e-5, leaf model: pruned set of Hoeffding tree (grace period: 100, delta: 1e-5)* para MLHT, y *grace period: 200, delta: 1e-5, leaf model: logistic regression* para iSOUPT.

En cuanto a la comparativa con el trabajo previo, dos trabajos [5], [7] han abordado previamente este caso de estudio. Cipollini et al. [7] publicaron el conjunto de datos y compararon el rendimiento de diferentes modelos para predecir la degradación de cada componente de SPN por separado. Este estudio determinó que un modelo de Red Neuronal Profunda, *Deep Neural Network* (DNN) obtiene los mejores resultados, por lo tanto, es el que se utiliza en nuestra comparativa combinado con *Binary Relevance* (BR) [3] para adaptarlo al contexto multi-etiqueta. Por otro lado, Tan et al. [5] abordan directamente el problema como una tarea de MLC que también compara diferentes modelos de AA tradicionales para determinar que la combinación más prometedora es BR con SVM. Los resultados de ambos modelos se han obtenido de las configuraciones proporcionadas por los autores en sus artículos. En todos los casos se tratan de propuestas pertenecientes al paradigma de AA tradicional de aprendizaje por lotes u *offline*, que no pueden ser adaptadas al modelo *online* por la naturaleza de sus algoritmos. Para igualar las condiciones de la evaluación tanto como sea posible en lo que sería un entorno que evoluciona con el tiempo, estas propuestas se han entrenado con el primer 60% del flujo de datos, y testeadas sobre el siguiente 40%.

Los parámetros con los que se ha ejecutado nuestra propuesta han sido obtenidos mediante pruebas preliminares incluyendo la misma porción del flujo de datos utilizada para entrenar las propuestas *offline*. Estos parámetros son $\kappa_{spl} = 65$, $\delta_{spl} = 7,38e - 4$, $\kappa_{alt} = 250$, $\delta_{alt} = 0,019$, $\eta = 640$ y $\lambda = 4$

IV-B. Estudio experimental

La Tabla I muestra los resultados sobre el caso de estudio de nuestra propuesta contra la literatura previa del caso y propuestas *online* previas para MLC tal y como se ha comentado en la Sección IV-A.

Estos resultados muestran que la propuesta de ensemble online de árboles de Hoeffding multi-etiqueta obtiene mejores resultados que el trabajo previo en árboles online para MLC, MLHT y iSOUPT, debido a la diversificación que introduce

Tabla I
LA PROPUESTA DE ENSEMBLE ONLINE DE ÁRBOLES DE Hoeffding MULTI-ETIQUETA CONTRA TRABAJOS PREVIOS DEL CASO DE ESTUDIO Y DE ÁRBOLES ONLINE MULTI-ETIQUETA

Model	Subset acc	Hamm. loss	Ex F1	Micro F1	Micro prec	Micro rec	Macro F1	Macro prec	Macro rec
BR+DNN [7]	-	0.256	-	0.764	0.792	0.741	0.679	0.643	0.741
BR+SVM [5]	0.081	0.443	0.677	0.651	0.612	0.697	0.572	0.462	0.750
MLHT [15]	0.227	0.300	0.566	0.574	0.961	0.409	0.558	0.946	0.441
iSOUPT [20]	0.865	0.035	0.948	0.960	0.960	0.960	0.958	0.958	0.958
Nuestra propuesta	0.943	0.013	0.978	0.982	0.977	0.986	0.981	0.976	0.985

en el proceso de entrenamiento al entrenar los árboles base en diferentes subespacios de características, lo cual es más beneficioso en un problema con relativa alta dimensionalidad en el espacio de entrada como éste, con 25 variables. Así, esta propuesta obtiene los mejores resultados tanto en las métricas de precisión global como en la precisión de etiquetas infrecuentes. Específicamente, es aproximadamente un 3% mejor que iSOUPT, el segundo mejor modelo, debido a que también aprende las relaciones entre etiquetas como nuestra propuesta. Los mejores resultados de nuestra propuesta se explican por la alta dimensionalidad del espacio de entrada, así como por la densidad de etiquetas relativamente alta que está relacionada con una alta interacción entre fallos en un sistema complejo como el SPN de una fragata. La diversidad introducida por el ensemble SRP y la adaptación a la deriva de concepto de nuestra propuesta mejoran la detección de correlaciones entre fallos en el problema estudiado.

Atendiendo a los resultados de los métodos previos que han abordado este problema, mostrados también en la Tabla I, se observa igualmente una mejora significativa. Los trabajos previos se enmarcan en el AA tradicional por lotes, en contraste con el modelo *online* propuesto que aprende incrementalmente. Los resultados obtenidos a partir de estos estudios muestran que el enfoque de aprendizaje por lotes no es competitivo contra el modelo *online*, con resultados aproximadamente un 40% peores en todas las métricas. Aunque estos modelos se ajustan mejor al conjunto de entrenamiento porque pueden dar varias pasadas por los datos, dependen de que el conjunto de test se ajuste a la misma distribución observada durante el entrenamiento. Sin embargo, en este problema se trabaja con un sistema en movimiento que atraviesa diferentes condiciones ambientales que afectan las mediciones recopiladas, por lo que hay cambios en la distribución de los datos y condiciones no observadas durante el tramo inicial de operación utilizado para el entrenamiento. También se observa que la DNN tiene un mejor rendimiento que la SVM debido al problema añadido de la alta dimensionalidad de los datos de entrada, lo que afecta las posibilidades de construir los planos de separación de clases de SVM. En consecuencia, un modelo que aprende de forma *online* a partir de grandes cantidades de datos, como el ensemble de árboles de Hoeffding multi-etiqueta propuesto, obtiene resultados significativamente mejores.

V. CONCLUSIONES

En este artículo se ha presentado un novedoso modelo de ensemble *online* de DTs para la detección de fallos en flujos de datos multi-etiqueta. La propuesta se ha validado en un problema público de PdM con múltiples componentes y fallos concurrentes, demostrando su efectividad tanto en el estado del arte del aprendizaje *online* en flujos de datos, como frente al trabajo previo que ha abordado este problema anteriormente mediante AA por lotes tradicional. El uso de un algoritmo base basado en un árbol de Hoeffding para MLC ha permitido aprender la correlación entre fallos y adaptar el modelo a los cambios de concepto, lo que ha contribuido a resultados altamente competitivos. Por otro lado, el enfoque del ensemble basado en *bagging online* con sub-espacios de características ha demostrado ser muy beneficioso a la hora de aprender eficientemente de espacios de entrada con un gran número de características.

Este trabajo ha abierto la puerta a aplicar aprendizaje *online* en entornos de PdM. Como trabajo futuro se plantea el desarrollo de técnicas de explicabilidad avanzadas que ofrezcan más información acerca de las causas detrás de los fallos predichos.

AGRADECIMIENTOS

Este trabajo está financiado en parte por el proyecto PID2020-115832GB-I00 del Ministerio de Ciencia e Innovación de España y el Fondo Europeo de Desarrollo Regional, en parte por el proyecto ProyExcel-0069 de la Consejería de Universidad, Investigación e Innovación de la Junta de Andalucía, y en parte por la beca predoctoral FPU19/03924 del Ministerio de Universidades de España.

REFERENCIAS

- [1] Z. Jan, F. Ahamed, W. Mayer, N. Patel, G. Grossmann, M. Stumptner, and A. Kuusk, "Artificial intelligence for industry 4.0: Systematic review of applications, challenges, and opportunities," *Expert Systems with Applications*, vol. 216, pp. 1–21, 2023.
- [2] A. Esteban, A. Zafra, and S. Ventura, "Data mining in predictive maintenance systems: A taxonomy and systematic review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 12, no. 5, pp. 1–45, 2022.
- [3] E. Gibaja and S. Ventura, "Multi-label learning: A review of the state of the art and ongoing research," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 4, no. 6, pp. 411–444, 2014.
- [4] S. Matzka, "Explainable Artificial Intelligence for Predictive Maintenance Applications," in *3rd the International Conference on Artificial Intelligence for Industries*, 2020, pp. 69–74.
- [5] Y. Tan, J. Zhang, H. Tian, D. Jiang, L. Guo, G. Wang, and Y. Lin, "Multi-label classification for simultaneous fault diagnosis of marine machinery: A comparative study," *Ocean Engineering*, vol. 239, pp. 1–11, 2021.
- [6] N. A. Sonkul, G. S. Dhage, and N. S. Vyas, "Single and Multi-label Fault Classification in rotors from unprocessed multi-sensor data through deep and parallel CNN architectures," *Expert Systems with Applications*, vol. 185, no. March, pp. 1–14, 2021.
- [7] F. Cipollini, L. Oneto, A. Coraddu, A. J. Murphy, and D. Anguita, "Condition-based maintenance of naval propulsion systems: Data analysis with minimal feedback," *Reliability Engineering and System Safety*, vol. 177, pp. 12–23, 2018.
- [8] S. Li, H. Cao, and Y. Yang, "Data-driven simultaneous fault diagnosis for solid oxide fuel cell system using multi-label pattern identification," *Journal of Power Sources*, vol. 378, no. 1, pp. 646–659, 2018.
- [9] D. Liu, X. Zhang, Z. Zhang, and H. Jiang, "A Hybrid Feature Selection and Multi-Label Driven Intelligent Fault Diagnosis Method for Gearbox," *Sensors*, vol. 23, no. 10, pp. 1–25, 2023.
- [10] N. Mylonas, I. Mollas, N. Bassiliades, and G. Tsoumakas, "Local Multi-Label Explanations for Random Forest," in *the 2022 Joint Conference on Machine Learning and Knowledge Discovery in Databases*, 2022, pp. 369–384.
- [11] Y. Wu, W. Jin, Y. Li, and D. Wang, "A novel method for simultaneous-fault diagnosis based on between-class learning," *Measurement: Journal of the International Measurement Confederation*, vol. 172, pp. 1–13, 2021.
- [12] J. Ben, S. Harrath, E. Bechhoefer, and M. Benbouzid, "Online automatic diagnosis of wind turbine bearings progressive degradations under real experimental conditions based on unsupervised machine learning," *Applied Acoustics*, vol. 132, pp. 167–181, 2018.
- [13] M. H. Le-Nguyen, F. Turgis, P. E. Fayemi, and A. Bifet, "Continuous Health Monitoring of Machinery using Online Clustering on Unlabeled Data Streams," in *2022 IEEE International Conference on Big Data*. IEEE, 2022, pp. 1866–1873.
- [14] P. Domingos and G. Hulten, "Mining High-Speed Data Streams," in *6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2000, pp. 71–80.
- [15] J. Read, A. Bifet, G. Holmes, and B. Pfahringer, "Scalable and efficient multi-label classification for evolving data streams," *Machine Learning*, vol. 88, no. 1-2, pp. 243–272, 2012.
- [16] B. Dai, S. Ding, and G. Wahba, "Multivariate Bernoulli distribution," *Bernoulli*, vol. 19, no. 4, pp. 1465–1483, 2013.
- [17] A. Bifet and R. Gavaldà, "Adaptive learning from evolving data streams," in *8th International Symposium on Intelligent Data Analysis*, 2009, pp. 249–260.
- [18] H. M. Gomes, J. Read, and A. Bifet, "Streaming random patches for evolving data stream classification," in *the 2019 IEEE International Conference on Data Mining*, 2019, pp. 240–249.
- [19] J. Gama, P. P. Rodrigues, and R. Sebastião, "Evaluating algorithms that learn from data streams," in *2009 ACM Symposium on Applied Computing*, 2009, pp. 1496–1500.
- [20] A. Osojnik, P. Panov, and S. Džeroski, "Multi-label classification via multi-target regression on data streams," *Machine Learning*, vol. 106, no. 6, pp. 745–770, 2017.