

## Vigilancia acústica inteligente de interiores para el análisis y modelado de escenas para la detección de comportamientos anómalos de personas y máquinas.

Inma Mohino-Herranz <sup>(1)</sup>, Sagrario Alonso-Díaz <sup>(1)</sup>, Roberto Gil-Pita <sup>(2)</sup>, Manuel Rosa-Zurera <sup>(2)</sup>, Joaquín García-Gómez <sup>(2)</sup>, Manuel Utrilla-Manso <sup>(2)</sup>, Fang-Fang Zhu-Zhou <sup>(2)</sup>, Manuel Bernal <sup>(1)</sup> y Robert Benyon <sup>(1)</sup>

<sup>(1)</sup> INTA Centro de Metrología y Calibración. Laboratorio de Factores Humanos.  
Ctra. de Ajalvir, km. 4,500 (28850) Torrejón de Ardoz

<sup>(2)</sup> Universidad de Alcalá. Escuela Politécnica Superior. Departamento de Teoría de la Señal y Comunicaciones. Ctra. Madrid-Barcelona km. 33,600, (28805) Alcalá de Henares

<sup>(1)</sup> TEL: (34) 91 520 63 85, mmohher@inta.es

### RESUMEN

Presentación del proyecto estatal solicitado. En principal objetivo del proyecto es la detección de anomalías para vigilancia inteligente en infraestructuras críticas, debido al comportamiento humano y la funcionalidad de las máquinas. En el caso de las personas, se estudiarán comportamientos anómalos, como pueden ser personas corriendo o con movimientos de flujo inusuales, personas gritando o situaciones violentas. Dichas situaciones se podrían dar en el acceso a estaciones o su interior, centros comerciales, etc. La detección de anomalías en máquinas es una tarea con alta dificultad ya que es complicado obtener una lista completa de anomalías, pero de especial relevancia debido al desarrollo exponencial del IoT (Internet of Things).

Serán utilizados varios sensores de diversa naturaleza, como son sensores acústicos, cámaras, sensores para monitorizar señales fisiológicas, entre otros. Una vez registrada la información, para procesarla de manera conveniente, se utilizarán diversas técnicas.

### 1. INTRODUCCIÓN

El presente trabajo trata sobre la investigación de técnicas para la detección de anomalías en el comportamiento del personal y máquinas en infraestructuras críticas para la seguridad.

La detección de anomalías es un tema de investigación muy estudiado en el que se aplica la tecnología de sensores, el procesamiento de señales y la ciencia de datos [1]. El desarrollo del Internet de las Cosas (IoT), el crecimiento explosivo de los big data y las técnicas de aprendizaje automático, así como las importantes aplicaciones en las que se puede aplicar, han motivado el interés actual por este tema. La detección de anomalías puede aplicarse a temas muy diversos: aplicaciones médicas, deportes, biometría del comportamiento, cuidado de ancianos, robótica, arte y entretenimiento, vigilancia inteligente, y análisis de gestos y posturas, entre otros [2]. En el caso de la detección de anomalías en el comportamiento humano, se debe llevar a cabo la definición del comportamiento anormal de un individuo, que tiene sus raíces en la psicología.

Para poner la investigación propuesta en contexto, es necesario definir los conceptos de anomalía, comportamiento y emoción [3][4].

- Las **Emociones** son reacciones internas de la persona ya sea a eventos, individuos y objetos.
- El **Comportamiento** es la reacción a una emoción interna o evento externo.
- Una **Anomalía** es la diferencia en el comportamiento del individuo con respecto a los demás que le rodean (el individuo puede ser una persona o una máquina en este contexto).

El objetivo del proyecto es, por lo tanto, la **detección de anomalías para la vigilancia inteligente en infraestructuras críticas**, principalmente debidas al **comportamiento humano y al funcionamiento de las máquinas**. El comportamiento de las personas en los espacios públicos suele seguir ciertos patrones en función del entorno: entrada/salida de un recinto, acceso o interior

de un centro comercial, estaciones, etc. En cada caso, se pueden identificar comportamientos normales (movimientos de personas caminando, grupos de personas hablando, etc.), así como otros comportamientos que podrían considerarse "anómalos" (personas corriendo, flujos de movimiento inusuales, personas gritando o situaciones violentas). La detección de anomalías también puede aplicarse a las máquinas, y este tema ha recibido recientemente una mayor atención debido al desarrollo del Internet de las Cosas (IoT) y de las técnicas de aprendizaje automático [5]. La detección de anomalías es una tarea muy difícil debido a que, normalmente, no se conoce una lista completa de anomalías.

## 2. DESCRIPCIÓN

Muchos algoritmos de detección de anomalías utilizan la información adquirida con diferentes tipos de sensores, principalmente cámaras y sensores acústicos, pero pueden utilizarse otros, como son los sensores inerciales o los sensores incorporados en las prendas textiles (wearable). Aunque los sistemas de visión proporcionan una gran cantidad de información, no siempre pueden utilizarse para detectar comportamientos anómalos, debido a iluminación o visibilidad insuficiente. Además, el uso de cámaras plantea importantes problemas de privacidad, que pueden ser importantes en la Unión Europea, donde debe aplicarse el Reglamento General de Protección de Datos (GDPR). El uso de sensores en textiles inteligentes para monitorizar la actividad humana se ha extendido recientemente, especialmente para aplicaciones sanitarias, militares y de seguridad [6] [7]. Los sensores para detectar señales fisiológicas pueden ser intrusivos y engorrosos, y su uso requiere la colaboración del sujeto, lo que no va a ser posible en la mayoría de las aplicaciones. Sin embargo, pueden utilizarse a nivel de investigación, para ayudar a validar los datos de actividad de las personas que participan en la prueba del sistema.

Por otro lado, sabemos que la mayoría de las actividades humanas cotidianas generan sonidos, y estos sonidos contienen una gran cantidad de información sobre la actividad que se realiza. La monitorización acústica permite detectar, localizar y rastrear eventos y fuentes sonoras, generando un conocimiento exhaustivo de la escena sonora. **Este análisis ayuda no sólo a monitorizar la actividad, sino también a detectar las anomalías que se producen en la escena.** Tanto la detección de eventos sonoros como la clasificación de la escena acústica son técnicas de percepción del entorno o de la situación [8]:

- La detección y el reconocimiento de eventos sonoros abarca técnicas y métodos para detectar y clasificar sonidos específicos generados por eventos de duración relativamente corta [9]. El reconocimiento de eventos sonoros sigue siendo un problema difícil, principalmente porque la señal sonora no está estructurada, no es estacionaria y es diversa [10]. La diversidad de sonidos que pueden asociarse a un mismo tipo de evento dificulta aún más la tarea. Por ejemplo, la caída de un determinado objeto puede producir diferentes sonidos en función del material sobre el que caiga. Para complicar aún más las cosas, pueden producirse múltiples eventos sonoros al mismo tiempo. No obstante, la investigación sobre el reconocimiento de eventos sonoros ha crecido en los últimos años, donde predominan los métodos basados en el aprendizaje profundo.
- El análisis de escenas auditivas consiste en el análisis de mezclas de sonidos para obtener objetos correspondientes a fuentes de sonido individuales [11]. La clasificación de escenas acústicas se utiliza para clasificar lugares, situaciones y entornos mediante el aprendizaje de múltiples eventos sonoros en un espacio específico [12]. En este contexto, Marr [13] postuló que el propósito del análisis de escenas es proporcionar una representación de lo

que está presente en el entorno. Este análisis suele seguir una estrategia ascendente, emulando los procesos que tienen lugar en la mente humana.

Para obtener un conocimiento completo de la escena, deben aplicarse algoritmos de detección y clasificación de eventos sonoros, siendo estas técnicas parte del análisis de la escena. Al mismo tiempo, el conocimiento de la escena puede mejorar el rendimiento de los algoritmos de detección y clasificación de eventos sonoros. Por ello, cuando la persona o el sistema automático adquieren conocimiento del contexto (context-aware sound event recognition, CoSER), la información contextual puede asociarse a los eventos sonoros, mejorando el rendimiento de los sistemas.

Para obtener el conocimiento del contexto en la monitorización acústica, es necesario determinar el número de fuentes, su naturaleza y su posición en relación con un sistema de coordenadas. Las propuestas clásicas para la localización de fuentes acústicas pueden dividirse en tres categorías: basadas en la estimación del retardo temporal [14], basadas en la formación de haces [15], y en métodos de estimación espectral de alta resolución [16]. Posteriormente, han aparecido estrategias basadas en la explotación de técnicas de optimización y propiedades matemáticas de las medidas relacionadas para aplicaciones en las que se han perdido datos [17], o para reducir el número de sensores [18], así como el uso del aprendizaje automático [19][20], y el aprendizaje profundo (Deep Learning) [21]. Existen iniciativas prometedoras para obtener directamente la posición geométrica tridimensional de la fuente [22]. Uno de los inconvenientes fundamentales de los sistemas basados en redes neuronales profundas es su dependencia de las condiciones geométricas de la base de datos de entrenamiento, lo que dificulta su despliegue en nuevos entornos geoméricamente diferentes y/o con configuraciones de micrófono distintas.

En la comunidad de investigación que trabaja en la clasificación de escenas acústicas, el DCASE Challenge [23] representa el foro donde se puede encontrar el estado del arte en este campo. En concreto, en DCASE 2022 Challenge [24], la Tarea 2 estudia la detección no supervisada de sonidos anómalos para monitorización del estado de las máquinas aplicando técnicas de generalización de dominios.

Este foro y otros talleres de evaluación se organizan desde el año 2013. En los últimos años se ha producido un auge de las soluciones basadas en el aprendizaje profundo para diversos problemas de clasificación. De hecho, en 2016, los trabajos evaluados que utilizaron redes neuronales representaron algo menos de la mitad, mientras que, en los últimos dos años, las soluciones presentadas para la subtarea de clasificación de escenas acústicas que no incluyeron un componente de aprendizaje profundo, estuvieron por debajo del 5%. Ello da cuenta del incremento de uso de este tipo de técnicas para clasificación de escenas acústicas.

La detección de anomalías también se ha aplicado ampliamente para el mantenimiento preventivo de la maquinaria industrial. Se basa en la información proporcionada por diferentes sensores equipados para medir la temperatura, la presión, la corriente eléctrica, las vibraciones y el sonido, entre otros. Los sensores de sonido tienen muchas ventajas, principalmente por su fácil despliegue, su coste reducido y porque muchos fallos pueden asociarse a una firma distinguible [5]. Todo ello se enmarca en lo que se conoce como Industria 4.0 y Mantenimiento 4.0, que incluye Internet Industrial de las Cosas (IIoT), sensores inalámbricos, computación en la nube, inteligencia artificial y aprendizaje automático. El mantenimiento predictivo [25] es un componente clave del Mantenimiento 4.0, y se basa en la detección de anomalías. Un ejemplo de ello es la monitorización del sonido aplicado al mantenimiento predictivo en procesos de fresado [26] para predecir la vida útil restante de un brazo robótico [27] o para analizar el funcionamiento de los elementos giratorios de una cinta transportadora [27].

Dos cuestiones importantes en cualquier sistema de detección de anomalías son la cantidad de datos utilizados para el entrenamiento y su representatividad. Muchos sistemas de detección de

anomalías se entrenan sólo con datos normales, y se espera que estos sistemas sean capaces de clasificar los datos anómalos como pertenecientes a una clase diferente. Un reto importante desde el punto de vista de los datos de entrenamiento es tener la mejor representación de los datos de forma que ni se generalice con las anomalías, para no generar falsos negativos, ni se sobreajuste a los datos de entrenamiento existentes de forma que reconozca los datos normales que nunca se han visto como anomalías, lo que da lugar a alertar al usuario con demasiada frecuencia [28]. Estos problemas surgen, por ejemplo, cuando el conjunto de datos de entrenamiento está corrompido por el ruido, o cuando hay valores atípicos en el conjunto de datos de entrenamiento. Hay muchos métodos de procesamiento de señales que se han aplicado para mejorar la relación señal-ruido en los sistemas de detección de anomalías. Como ejemplo de esto último, está el caso en el que el sonido está corrompido por el ruido industrial. Entre ellos, podemos mencionar el Análisis de Componentes Independientes (ICA) o la Descomposición Modal Empírica (EMD), entre otros [5].

### 3. RESULTADOS Y DISCUSIÓN

Los métodos utilizados para la detección de anomalías pueden agruparse en dos grupos principales, modelos *Generativos* o *Discriminativos*:

- Los modelos generativos se utilizan para generar muestras a partir de alguna distribución de datos aprendida  $p(x)$  que, en el caso de la detección de anomalías, es la distribución de los datos normales [29]. Los modelos generativos se entrenan principalmente con datos normales para aprender de su distribución y, una vez entrenados, se utilizan para calcular una función de puntuación de anomalías  $f: \mathcal{X} \rightarrow \mathcal{R}$ , para una entrada  $x \in \mathcal{X}$ , siendo  $\mathcal{X}$  el espacio de datos. Dos ejemplos de modelos generativos son los autocodificadores y los modelos adversariales:
  - *Los autocodificadores* constan de dos estructuras de aprendizaje profundo, un codificador y un decodificador. El codificador se entrena para mapear la entrada a un espacio vectorial de baja dimensión, mientras que el decodificador se entrena para obtener una aproximación del vector de entrada,  $x$ . Ambas partes se entrenan para minimizar el error de reconstrucción utilizando una función de error. Cuando se aplica una anomalía a la entrada, es probable que el autocodificador no produzca una buena reconstrucción, ya que no ha sido entrenado con muestras similares. Los autocodificadores se han aplicado a la detección de anomalías en tiempo real basada, entre otros, en la acústica de los activos [30].
  - *Las redes adversariales generativas (GAN)* también se han utilizado para la detección de anomalías. Las GAN constan de dos adversarios, un generador y un discriminador. El generador está entrenado para crear muestras similares a los datos reales, mientras que el discriminador está entrenado para detectar las muestras falsas entre un conjunto de entradas reales. La puntuación de la anomalía se calcula como una media ponderada del resultado del discriminador y un error de reconstrucción del generador.
- Los modelos discriminantes se entrenan para clasificar entre  $p(x)$  y  $q(x)$ , las distribuciones de los datos normales y la distribución de las anomalías, respectivamente. Actualmente se utilizan redes neuronales profundas entrenadas de forma supervisada. En las GAN se utiliza el modelo discriminativo para el discriminador, por lo que la GAN es un modelo

híbrido. Los modelos discriminativos se han utilizado, por ejemplo, para la detección de anomalías en el funcionamiento de máquinas [31].

Aunque estos modelos ya se han aplicado a la detección de anomalías mediante señales sonoras, la mayoría de las contribuciones se sitúan en el campo del procesamiento de imágenes. Nuestro objetivo es evaluar estos modelos en la detección de anomalías en el comportamiento humano y en el funcionamiento de las máquinas, en combinación con la información contextual obtenida con un sistema computacional de análisis de escenas auditivas, siguiendo en su conjunto el modelo teórico de Endsley de Situational Awareness [32], que se implementa en tres niveles: percepción, comprensión y proyección:

- La **percepción** involucra los procesos de monitoreo, detección y reconocimiento simple, que conducen a una conciencia de múltiples elementos situacionales (fuentes de sonido, eventos, personas, máquinas, etc.), y sus estados actuales (ubicación, condiciones, acciones, etc.).
- La **comprensión** implica el reconocimiento de patrones, la interpretación y la evaluación. Requiere integrar la información obtenida para comprender cómo repercutirá en las metas y objetivos del individuo.
- La **proyección** se consigue extrapolando en el tiempo la información proporcionada por los dos primeros niveles con el objetivo de determinar cómo afectará a futuros estados del entorno operativo.

La conciencia situacional se utiliza para la toma de decisiones y la realización de acciones, en la Figura 1 se muestra el esquema utilizado, basado en el Modelo de conciencia situacional de Endsley. También está influida por variables externas, recogidas del entorno, relacionadas con factores individuales (factores humanos, objetivos, formación, ...) y con factores de la tarea y del entorno (peligros, diseño del sistema, complejidad, etc.).

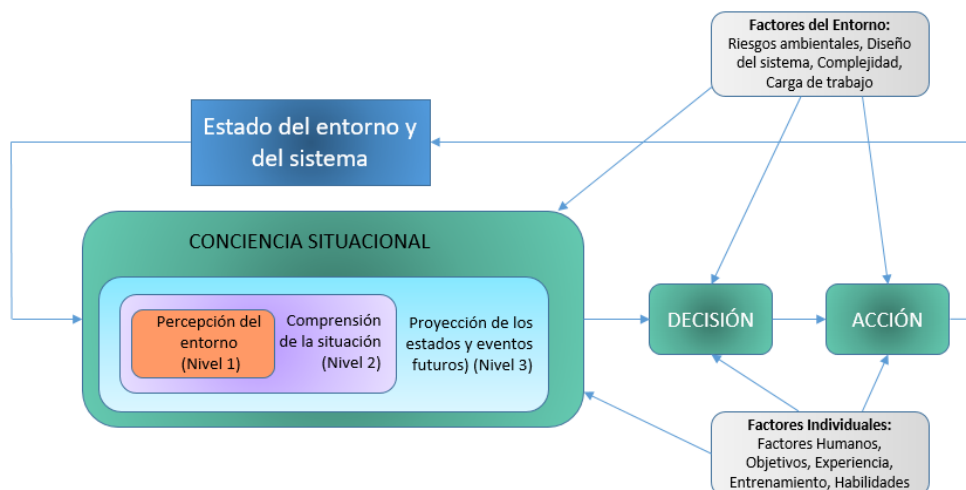


Figura 1. Esquema utilizado en el proyecto basado en el Modelo de conciencia situacional de Endsley

**En nuestro enfoque**, la percepción se realizará principalmente con arrays acústicos (diseñados con arquitecturas novedosas para trabajar con señales multibanda y para minimizar los errores de localización debidos a errores en la estimación de la dirección de llegada de las fuentes sonoras),

pero complementados con sistemas de visión que aporten información adicional de la escena, y, cuando se trate de análisis del comportamiento humano, con un sistema generado en un proyecto anterior, que aporta información de señales biológicas que se utilizan para inferir el estado de estrés físico y mental. Por lo que sabemos, este enfoque aún no se ha aplicado en este campo. Sólo existen algunas propuestas centradas en la descripción textual y abstracta de entornos a partir de información acústica [33][34], que se basan en el conocimiento adquirido mediante la detección y clasificación de eventos y escenas. Este enfoque ha encontrado aplicación en "Smart-cities" [35] y "Smart-home" [36].

La **hipótesis** de la investigación es que la combinación del análisis acústico de la escena (basado en la detección y clasificación de eventos sonoros, y el posicionamiento de la fuente sonora) para generar conciencia situacional, y algoritmos de Deep-Learning para la detección de anomalías, puede ser útil para detectar comportamientos anómalos de humanos y máquinas en aplicaciones de seguridad.

#### 4. CONCLUSIONES

La combinación del análisis acústico de escenas –con el objeto de crear conciencia situacional- y el uso de algoritmos de Deep-Learning para la detección de anomalías, aporta como principal novedad una combinación de métodos de análisis de escenas. El uso de algoritmos de detección de anomalías basados principalmente en algoritmos de aprendizaje profundo, mejorarán los métodos actualmente definidos para detectar comportamientos anómalos de personas en la escena, así como el posible funcionamiento anómalo de máquinas en la industria. Los métodos basados en el aprendizaje profundo tendrán en cuenta 1) la dependencia de la cantidad de datos para el entrenamiento, 2) cómo se comportará el sistema cuando se utilicen redes pre-entrenadas y, 3) cómo se comportará el sistema cuando se utilicen redes entrenadas con una cantidad corta de datos disponibles (aprendizaje de transferencia [37]). Además, para crear la comprensión de la escena, se explotarán los conjuntos de sensores y las redes de sensores. Finalmente, cuando se trate de análisis del comportamiento humano, se estudiará la utilización de la información de señales biológicas en tiempo real para inferir el estado de estrés físico y mental en el contexto de detección de anomalías en aplicaciones de seguridad.

#### 5. REFERENCIAS

- [1] V. Chandola, A. Banerjee, V. Kumar, "Anomaly detection: A survey", ACM Comput. Surv., Vol. 41, pp. 1-58, 2009.
- [2] S. Kumar Yadav, K. Tiwari, H. Mohan Pandey, S. Ali Akbar, "A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions", Knowledge-Based Systems, 223-106970, 2021.
- [3] M. W. Baig, E. I. Barakova, I. Marcenaro, M. Rauterber, and C.S. Regazzoni, "Crowd Emotion Detection Using Dynamic Probabilistic Models", From Animals to Animals 13 (A.P. del Pobil et al. Eds.) Springer Berlin Heidelberg, pp. 328-337, 2014.
- [4] G. Gliniskite, E. Lavendelis, M. Pudane, "Affective State Based Anomaly Detection in Crowd", Applied Computer Systems, Vol. 24, No. 2, pp. 134-140, 2019.
- [5] Yuki Tagawa, Rytis Maskeliūnas, Robertas Damaševičius, "Acoustic Anomaly Detection of Mechanical Failures in Noisy Real-Life Factory Environments", Electronics, 10, 2329, 2021.

- [6] C. Zhu, W. Sheng, and M. Liu, "Wearable sensor-based behavioral anomaly detection in smart assisted living systems," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 4, pp. 1225–1234, 2015.
- [7] S.C. Mukhopadhyay, "Wearable sensors for human activity monitoring: A review," *IEEE Sensors Journal.*, vol. 15, no. 3, pp. 1321–1330, 2015.
- [8] M. Jung, S. Chi, "Human activity classification based on sound recognition and residual convolutional neural networks", *Automation in Construction*, Vol. 114, paper 103177, 2020.
- [9] O. Gencoglu, T. Virtanen, H. Huttunen, "Recognition of acoustic events using deep neural networks", *22nd European Signal Processing Conference*, pp. 506-510, 2014.
- [10] H.M. Do, K.C. Welch, W. Sheng, "SoHAM: A Sound-Based Human Activity Monitoring Framework for Home Service Robots", *IEEE Trans. Aut. Sc. and Eng. (early access)*, pp. 1-15, 2021.
- [11] A. S. Bregman, "Auditory scene analysis: the perceptual organization of sound." Cambridge, MA: MIT Press, 1990.
- [12] M. Valenti, S. Squartini, A. Diment, G. Parascandolo, T. Virtanen, "A convolutional neural network approach for acoustic scene classification", *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 1547-1554, Anchorage, AK (USA), 2017.
- [13] D. Marr, "Vision: a computational investigation into the human representation and processing of visual information". Freeman. 1982.
- [14] C. Knapp, G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoustics, Speech and Signal Processing*, Vol. 24, No. 4, pp. 320-327, 1976.
- [15] J.C. Chen, et al., "Source localization and beamforming". *IEEE Signal Processing Magazine*, Vol. 19, No. 2, pp. 30-39, 2002.
- [16] G. N. Kuznetsov, et al., "Spectrogram and localization of a sound source in shallow water". *Acoustical Physics*, Vol. 63, No.4, pp. 449-461, 2017.
- [17] J. Velasco, et al., "TDOA Matrices: Algebraic Properties and Their Application to Robust Denoising with Missing Data". *IEEE Trans. Signal Process*, Vol. 64, pp. 5242-5254, 2016.
- [18] S. Salari, et al., "TDOA Estimation with Compressive Sensing Measurements and Hadamard Matrix". *IEEE Trans. Aerosp. Electron. Syst.* pp. 3137-3142, 2018.
- [19] S. Adavanne, et al., "Sound event localization and detection of overlapping sources using convolutional recurrent neural networks". *IEEE J. Sel. Top. Sig. Process.*, Vol.13, No.1, pp. 34-48, 2018.
- [20] Y. Sun, et al., "Indoor Sound Source Localization with Probabilistic Neural Network". *IEEE Trans. Ind. Electron*, Vol. 65, pp. 6403-6413, 2018.
- [21] N. Ma, et al., "Exploiting Deep Neural Networks and Head Movements for Robust Binaural Localization of Multiple Sources in Reverberant Environments". *IEEE/ACM Trans. Audio Speech Lang. Process*, Vol. 25, pp. 2444-2453, 2017.
- [22] J. Vera-Diaz, et al., "Towards End-to-End Acoustic Localization Using Deep Learning: From Audio Signals to Source Position Coordinates". *Sensors*, Vol. 18, No. 10, 2018.
- [23] I. Martín-Morato, T. Heittola, A. Mesaros, T. Virtanen, "Low-complexity acoustic scene classification for multi-device audio: analysis of DCASE 2021 challenge systems", *Detection and Classification of Acoustic Scenes and Events 2021* (online: <https://arxiv.org/pdf/2105.13675.pdf>).
- [24] Dohi, Kota and Nishida, Tomoya and Purohit, Harsh and Tanabe, Ryo and Endo, Takashi and Yamamoto, Masaaki and Nikaido, Yuki and Kawaguchi, Yohei. "MIMII DG:

Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection for Domain Generalization Task". Detection and Classification of Acoustic Scenes and Events 2022. En arXiv e-prints: 2205.13879.

[25] F. Calabrese, A. Regattieri, M. Bortolini, M. Gamberi, E. Pilati, "Predictive maintenance: A novel framework for a data-driven semi-supervised, and partially online prognostic health management application in industries", Applied Science, Vol. 11, paper 3380, 2021.

[26] C.Y. Peng, U. Raihany, S. W. Kuo, Y.Z. Chen, "Sound detection monitoring tool in CNC milling sounds by K-means clustering algorithm", Sensors, Vol. 21, paper 4288, 2021.

[27] D. Kim, S. Lee, "An applicable predictive maintenance framework for the absence of run to failure data", Applied Science, Vol. 11, paper 5180, 2021.

[28] W.J. Lee, K. Helwani, S. Tenneti, A. Krishnaswamy, "Robust audio anomaly detection", In RobustML Workshop at ICLR, pp. 1-13, 2021.

[29] V. Skvára, T. Pevny, V. Smile, "Are generative deep models for novelty detection truly better?", arXiv preprint arXiv:1807.05027, 2018.

[30] M. Meire, P. Karsmakers, "Comparison of deep autoencoder architectures for real-time acoustic based anomaly detection in assets", 10th IEEE Int. Conf. on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, pp. 786-790, 18-21 September, 2019.

[31] H. Ahn, I. Yeo, "Deep-Learning-based approach to anomaly detection techniques for large acoustic data in machine operation", Sensors, 21, 5446, 2021.

[32] M.R. Endsley, "Toward a theory of situational awareness in dynamic systems", Human Factors, Vol. 37, No. 1, pp. 32-64, 1995.

[33] T. Lu, et al., "Context-based environmental audio event recognition for scene understanding". Multimedia Systems, Vol. 21, No. 5, pp. 507-524, 2015

[34] T. Virtanen, et al., "Computational analysis of sound scenes and events". Springer. 2018.

[35] J. Ye, et al. "Urban sound event classification based on local and global features aggregation". Applied Acoustics, Vol. 117, pp. 246-256, 2017.

[36] M.A. Sehili, et al., "Sound environment analysis in smart home". Int. Joint Conference Ambient Intelligence, pp. 208-223, 2012.

[37] K. Bhat, M. Bhandari, C. Oh, S. Kim, J. Yoo. Transfer Learning Based Automatic Model Creation Tool For Resource Constrained Devices. arXiv preprint arXiv:2012.10056, 2020.

## 6. AGRADECIMIENTOS

Este trabajo ha sido financiado por el Ministerio de Ciencia e Innovación, con el proyecto de referencia PID2021-129043OB-I00.