

Salamanca, Guanajuato, 19 de enero del 2024.

MTRO. HERIBERTO GUTIÉRREZ MARTIN  
COORDINADOR DE ASUNTOS ESCOLARES  
P R E S E N T E

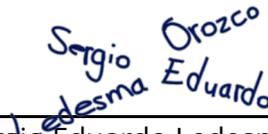
Por medio de la presente, se otorga autorización para proceder a los trámites de impresión, empastado de tesis y titulación al alumno: **Yair Alejandro Andrade Ambriz**, del Programa: **Doctorado en Ingeniería Eléctrica** y cuyo **NUA** es: **801346**, del cual soy director. El título de la tesis es: **"Aplicaciones de la inteligencia artificial en ámbitos de uso real"**.

Hago constar que he revisado dicho trabajo y he tenido comunicación con los sinodales asignados para la revisión de la tesis, por lo que no hay impedimento alguno para fijar la fecha de examen de titulación.

ATENTAMENTE



\_\_\_\_\_  
Dra. Dora Luz Almanza Ojeda  
**SECRETARIO**  
DIRECTOR DE TESIS



\_\_\_\_\_  
Dr. Sergio Eduardo Ledesma Orozco  
DIRECTOR DE TESIS



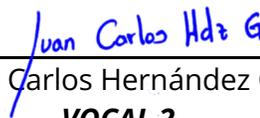
\_\_\_\_\_  
Dr. Juan Manuel Belman Flores  
DIRECTOR DE TESIS



\_\_\_\_\_  
Dr. José Ruíz Pinales  
**PRESIDENTE**



\_\_\_\_\_  
Dr. David Camarena Martínez  
**VOCAL 1**



\_\_\_\_\_  
Dr. Juan Carlos Hernández García  
**VOCAL 2**



\_\_\_\_\_  
Dr. Carlos Andrés Pérez Ramírez  
**VOCAL 3**



# Universidad de Guanajuato

---

---

Campus Irapuato - Salamanca  
División de Ingenierías

*“Aplicaciones de la inteligencia artificial  
en ámbitos de uso real”*

## TESIS PROFESIONAL

QUE PARA OBTENER EL GRADO DE  
DOCTOR EN INGENIERÍA ELÉCTRICA

PRESENTA:

*M. en I. Yair Alejandro Andrade Ambriz*

DIRECTORES:

*Dra. Dora Luz Almanza Ojeda  
Dr. Sergio Eduardo Ledesma Orozco  
Dr. Juan Manuel Belman Flores*

Salamanca, Guanajuato.

Enero, 2024.

# Dedicatoria

Dedico este trabajo y logro a mis padres, Sergio Andrade Cahue y Ma. Guadalupe Ambriz Velazquez, cuya inquebrantable dedicación y apoyo han sido fundamentales para alcanzar esta etapa significativa en mi vida. También, extendiendo esta dedicación a mi hermano Sergio Andrade Ambriz, cuyo invaluable consejo y apoyo han sido una fuente constante de inspiración y fortaleza a lo largo de este camino. Este logro no hubiera sido posible sin su apoyo y aliento constantes.

Finalmente, quiero dedicar este trabajo de manera especial a la memoria y vida de mi hermano Mario Alberto Andrade Ambriz, un hermano extraordinario que siempre creyó en mí y me impulsó a superarme constantemente. Su presencia y aliento han dejado una huella imborrable en mi camino, y esta etapa lleva consigo la influencia positiva de su memoria.

*“La muerte no nos roba a los seres amados. Al contrario, nos los guarda y nos los  
inmortaliza en el recuerdo.”  
François Mauriac*

# Agradecimientos

Quiero expresar mi profundo agradecimiento a mis asesores, el Dr. Sergio Eduardo Ledesma Orozco, la Dra. Dora Luz Almanza Ojeda y el Dr. Juan Manuel Belman Flores, por su total apoyo y consejo invaluable en la realización de este trabajo. Su orientación ha sido fundamental para alcanzar la culminación de esta etapa en mi vida.

# Agradecimientos Institucionales

Al Consejo Nacional de Humanidades Ciencias y Tecnologías de México, bajo la beca otorgada en la convocatoria “Becas Nacionales 2020”, con el número de Currículum Vitae Único: 846226.



A la Dirección de Apoyo a la Investigación y al Posgrado (Universidad de Guanajuato) por los apoyos otorgados en las convocatorias “Convocatoria de apoyo a los posgrados UG 2022” y “Convocatoria de apoyo a los posgrados UG 2023”, con el número de alumno: 801346.



Dirección de Apoyo a la Investigación y al Posgrado

Al Departamento de Ingeniería Electrónica (División de Ingenierías Campus Irapuato - Salamanca, Universidad de Guanajuato), por el apoyo otorgado para la realización de una estancia de investigación en la Universidad de Alberta, Alberta, Canadá, con el número de alumno: 801346.



Campus Irapuato-Salamanca | División de Ingenierías

# Resumen

En el ámbito de la interacción humano - máquina y en diversos campos, la detección y el reconocimiento de las acciones humanas revisten una importancia primordial. En la última década, los avances en el reconocimiento de la actividad humana han aprovechado los dispositivos informáticos de alto rendimiento, utilizando una multitud de sensores para mejorar la calidad y la eficiencia de los resultados. En busca de un enfoque no invasivo, proponemos un novedoso diseño que emplea una red neuronal convolucional temporal. Esta red aprovecha las características espacio - temporales para el análisis y el reconocimiento de las actividades humanas basándose únicamente en entradas de videos cortos. La arquitectura propuesta está basada en una capa convolucional 3D y una capa convolucional de memoria a corto - largo plazo, combinando características de tiempo con características espaciales. Esta innovadora metodología optimiza los recursos computacionales para un entrenamiento/clasificación rápidos, logrando resultados de clasificación en tiempo real. En particular, las simulaciones muestran resultados de clasificación superiores a los de los métodos más avanzados en el estado del arte, incluso a los que requieren datos de más sensores.

Ampliando nuestro enfoque, presentamos un método innovador para estimar la escarcha acumulada sobre la superficie de un evaporador, utilizando señales acústicas, imágenes RGB e inteligencia artificial. Dirigido a la superficie del evaporador, nuestro enfoque emplea diversos valores de frecuencia para generar señales acústicas que interactúan dentro de la cavidad del evaporador. Concurrentemente, se captura una imagen RGB del evaporador donde visualmente se obtiene la acumulación de la escarcha. Mediante un modelo inteligente acoplado a un altavoz activo y un micrófono estándar, se estima la acumulación de escarcha. Los niveles de presión sonora se emplean para reducir la dimensionalidad de los datos, formando el conjunto de datos de entrenamiento y validación. Las redes neuronales artificiales y las redes neuronales probabilísticas sirven de base a dos modelos inteligentes que muestran una alta exactitud en la estimación de cuatro niveles de acumulación de escarcha en un refrigerador doméstico. Las pruebas realizadas en diferentes refrigeradores en condiciones de uso típicas validan su exactitud. Estos enfoques, con una exactitud de clasificación del 94 % (señales acústicas) y 99,75 % (imágenes RGB), destacan como una solución de bajo costo, precisa y fácilmente integrable para evaluar la acumulación de escarcha en la superficie del evaporador, un aspecto fundamental para optimizar la eficiencia de la refrigeración.

# Abstract

In the field of human-machine interaction and in various fields, the detection and recognition of human actions are of paramount importance. In the last decade, advances in human activity recognition have taken advantage of high-performance computing devices, using a multitude of sensors to improve the quality and efficiency of the results. In search of a non-invasive approach, we propose a novel design that employs a temporal convolutional neural network. This network takes advantage of spatio-temporal features for the analysis and recognition of human activities based solely on short video inputs. The proposed architecture is based on a 3D convolutional layer and a short-long term memory convolutional layer, combining time features with spatial features. This innovative methodology optimizes computational resources for fast training/classification, achieving real-time classification results. In particular, simulations show classification results superior to state-of-the-art methods, even those requiring more sensor data.

Extending our approach, we present an innovative method to estimate the frost accumulated on the surface of an evaporator using acoustic signals, RGB imaging and artificial intelligence. Targeting the evaporator surface, our approach employs various frequency values to generate acoustic signals that interact within the evaporator cavity. Concurrently, an RGB image of the evaporator is captured where visually the frost accumulation is obtained. The frost accumulation is estimated using an intelligent model coupled with an active loudspeaker and a standard microphone. Sound pressure levels are used to reduce the dimensionality of the data, forming the training and validation data set. Artificial neural networks and probabilistic neural networks serve as the basis for two intelligent models that show high accuracy in classifying four levels of frost accumulation in a domestic refrigerator. Tests on different refrigerators under typical use conditions validate their accuracy. These approaches, with a classification accuracy of 94 % (acoustic signals) and 99.75 % (RGB images), stand out as a low-cost, accurate and easily integrated solution for assessing frost buildup on the evaporator surface, a key aspect for optimizing refrigeration efficiency.

# Índice general

Dedicatoria	II
Agradecimientos	III
Agradecimientos Institucionales	IV
Resumen	v
Abstract	VI
<b>1. Introducción</b>	<b>1</b>
1.1. Objetivos . . . . .	2
1.2. Objetivos Específicos . . . . .	2
1.3. Justificación . . . . .	3
1.4. Estado del Arte . . . . .	4
1.5. Organización de la Tesis . . . . .	5
<b>2. Aplicaciones de la inteligencia artificial</b>	<b>6</b>
2.1. Inteligencia artificial . . . . .	6
2.2. Aprendizaje máquina . . . . .	7

2.3. Redes neuronales artificiales . . . . .	9
2.3.1. Funciones de activación . . . . .	9
2.4. Red neuronal convolucional . . . . .	11
2.5. Métricas de evaluación . . . . .	12
<b>3. Reconocimiento de actividades</b>	<b>15</b>
3.1. Introducción . . . . .	15
3.2. Reconocimiento de actividades humanas . . . . .	17
3.3. Metodología . . . . .	19
3.3.1. Bases de datos . . . . .	21
3.4. Resultados obtenidos . . . . .	25
3.4.1. Evaluación global de la arquitectura propuesta . . . . .	25
3.4.2. Comparación con trabajos relacionados . . . . .	25
3.4.3. Análisis del tiempo de procesamiento de la arquitectura . . . . .	28
3.5. Discusión . . . . .	29
<b>4. Refrigeración doméstica</b>	<b>31</b>
4.1. Introducción . . . . .	31
4.2. Señales acústicas . . . . .	33
4.3. Información visual . . . . .	37
4.4. Banco experimental . . . . .	38
4.4.1. Refrigerador doméstico A . . . . .	38
4.4.2. Refrigerador doméstico B . . . . .	39
4.4.3. Detección de escarcha en la superficie del evaporador . . . . .	40

<i>ÍNDICE GENERAL</i>	IX
4.5. Resultados . . . . .	42
4.6. Discusión . . . . .	47
<b>5. Conclusiones</b>	<b>49</b>
Índice de figuras	x
Índice de tablas	xii
Bibliografía	xiii

# Capítulo 1

## Introducción

La inteligencia artificial (IA) es un campo de la informática que se enfoca en crear sistemas que tratan de imitar la inteligencia humana. La IA se utiliza principalmente para tareas como el aprendizaje, el razonamiento, percepción, entre otros. Para realizar tareas que normalmente requieren de inteligencia humana, la IA necesita de un algoritmo inteligente. Un algoritmo inteligente es un algoritmo que puede aprender y adaptarse a nuevos datos, a diferencia de un algoritmo tradicional, que siguen un conjunto de reglas predefinidas. Los algoritmos inteligentes se basan en el aprendizaje automático, que es la capacidad de obtener un aprendizaje de forma autónoma sin estar programado específicamente para ello. Además, los algoritmos inteligentes mejoran con el uso, al obtener información recopilada de diferentes fuentes o sensores, logrando ser capaces de realizar tareas como: reconocimiento de patrones, clasificación de información, interacción con humanos o con otras máquinas, entre otros.

La IA ofrece herramientas y técnicas que permiten la mejora y la optimización de procesos en diversas áreas, esto la convierte en una valiosa herramienta para la resolución de problemas y toma de decisiones. Asimismo, permite automatizar tareas repetitivas y complejas, además de optimizar procesos y sistemas. Del mismo modo, puede ser usada para diagnóstico y mantenimiento, así como para la predicción de eventos y tendencias. Por lo tanto, la IA es una disciplina multidisciplinaria que se basa en técnicas de diversas áreas del conocimiento. Esto permite combinar conocimientos de distintos campos para crear nuevas herramientas o aplicaciones.

En este sentido, la presente tesis doctoral plantea el análisis de algoritmos inteligentes para la solución de problemas cotidianos. Específicamente, se plantea el diseño y desarrollo de redes neuronales artificiales, como pueden ser: redes neuronales multicapa, redes probabilísticas, redes convolucionales, entre otras. La tesis doctoral plantea dos casos de estudio enfocados en dos vertientes actuales: la interacción humano – máquina y la refrigeración.

El primer caso de estudio: “Reconocimiento de actividades” examina la interacción humano - máquina. La interacción humano – máquina estudia la asociación de los humanos

con tecnologías computacionales. Se inicia con la detección de actividades humanas, lo cual implica la detección y clasificación de acciones humanas o eventos con base en sensores como cámaras, micrófonos, acelerómetros, entre otros. Esas actividades podrían ser: pararse, caminar, saludar, entre otros. Por consiguiente, se plantea un sistema de detección de actividades humanas basado en redes neuronales convolucionales temporales.

Por otra parte, el segundo caso de estudio propuesto: “Refrigeración doméstica” analiza el crecimiento de la escarcha en la superficie del evaporador. Los refrigeradores son uno de los electrodomésticos más ampliamente manufacturados, principalmente aquellos del tipo “*no-frost*” (sin escarcha). En consecuencia, se plantea el diseño de un sistema de detección de escarcha acumulada sobre la superficie del evaporador, basado en sensores de visión por computadora y señales acústicas.

## 1.1. Objetivos

Analizar diferentes algoritmos inteligentes para la aplicación de la IA en ámbitos de uso real. Examinar modelos inteligentes como pueden ser: redes neuronales artificiales, redes probabilísticas, redes neuronales convolucionales (CNN - *Convolutional Neural Network*), entre otros. Identificar dos casos de estudio de aplicación de la IA: “Reconocimiento de actividades” y “Refrigeración doméstica”. Identificar actividades humanas para la interacción humano – máquina. Desarrollar un modelo inteligente basado en CNN y redes de memoria a largo/corto plazo, que sean capaces de detectar diversas actividades humanas con técnicas no invasivas como una secuencia corta de video. Aplicar la IA en el proceso de deshielo de un refrigerador del tipo “*no-frost*”, a fin de detectar el nivel de acumulación de escarcha en la superficie del evaporador, tomando en cuenta los diversos factores que propician la creación de la escarcha, así como los hábitos de uso más representativos.

## 1.2. Objetivos Específicos

- Recolectar diferentes bases de datos con videos o imágenes de actividades humanas para crear una base de datos completa que permita entrenar y probar modelos basados en redes neuronales artificiales.
- Evaluar el desempeño de bases de datos para el reconocimiento de actividades humanas.
- Obtener una arquitectura convolucional temporal con alta exactitud para el reconocimiento de actividades humanas, usando información de sensores no invasivos.
- Modelado e instrumentación de sensores de audio y de visión para la captura de datos de un evaporador en diferentes refrigeradores domésticos.

- Diseñar e implementar al menos un modelo inteligente para la detección de la escarcha sobre la superficie del evaporador de un refrigerador doméstico.
- Realizar pruebas de desempeño de los modelos inteligentes, validando su exactitud con diferentes refrigeradores domésticos, además de diferentes condiciones de uso.
- Obtener al menos dos publicaciones indexadas en JCR (*Journal Citation Report*).

### 1.3. Justificación

Actualmente la IA es un campo en constante evolución y expansión. Se encuentra en aplicaciones de procesamiento natural de lenguaje usados por asistentes personales, de igual manera, se integra en el desarrollo de vehículos autónomos o sistemas autónomos. Además se utiliza para automatización industrial, optimización en procesos de refrigeración, hasta en el análisis de imágenes médicas para ayuda al diagnosticar enfermedades. La IA se encuentra presente en prácticamente todos los aspectos de la vida cotidiana, y se prevé que continúe desempeñando un papel cada vez más importante en la sociedad conforme pasen los años.

En la actualidad, los sistemas autónomos proporcionan asistencia para realizar actividades en diferentes campos como pueden ser: entretenimiento, educación, domótica, entre otros. Sin embargo, para completar satisfactoriamente una interacción humano – máquina, estos sistemas deben ser capaces de detectar e interpretar ciertos comportamientos humanos. En este sentido, la detección de actividades humanas ha tenido un aumento en su importancia, gracias a una mayor interacción de los humanos en tecnologías computacionales, además de, el aumento del poder computacional del hardware y el desarrollo de modelos inteligentes. El reconocimiento de actividades humanas incrementa la interacción humano - máquina en actividades como detección de caídas, casas inteligentes, detección temprana de enfermedades, entre otros. Esto es gracias al avance de diferentes tipos de sensores, como pueden ser cámaras a color, cámaras de profundidad, sensores portátiles (teléfonos y relojes inteligentes), entre otros. Estos sensores tienen la capacidad de capturar datos de las actividades del usuario, para posteriormente analizar los datos y detectar puntos clave para clasificar la actividad realizada. Por lo tanto, existe la oportunidad de desarrollar un nuevo sistema basado en visión por computadora, sustentado en una red temporal convolucional. Este sistema deberá ser capaz de detectar al menos 12 actividades humanas usando secuencias cortas de video.

Por otra parte, los refrigeradores del tipo “no-frost” son los más usados debido a la facilidad en su uso, ya que, este tipo de refrigeradores elimina de forma automática la formación de escarcha (medio poroso) sobre el evaporador. De esta forma se evita la formación de bloques de hielo sobre el evaporador, que repercuten en el funcionamiento térmico y energético del refrigerador. Para lograr esto, generalmente se utiliza un elemento resistivo el cual al activarse genera calor, y en gran parte por el mecanismo de la radiación funde la escarcha

que se pudo haber acumulado durante el uso habitual del refrigerador, dejando así la superficie limpia del evaporador. Este proceso puede representar hasta un 15% en el uso de la energía anual de la refrigeración doméstica [1]. Finalmente, los hábitos de uso también juegan un papel muy importante en el comportamiento térmico y energético de un refrigerador doméstico [2], estos suelen ser diferentes en cada familia o en cada hogar, además de las condiciones ambientales donde se ubique el refrigerador. En consecuencia, se propone el desarrollo de un sistema de detección de acumulación de escarcha sobre el evaporador de un refrigerador doméstico. Este sistema estará basado en la emisión/detección de señales acústicas, además de un sistema de visión por computadora.

## 1.4. Estado del Arte

La interacción humano – máquina requiere de un sistema capaz de reconocer e inclusive imitar ciertas actividades hechas por humanos. En este sentido, se ha podido proporcionar a robots la capacidad de identificar comandos de voz [3], reconocer acciones humanas [4], o ambos, para habilitar una comunicación/interacción con los humanos. Combinando la información procedente de diversos sensores se pueden identificar muchas actividades humanas con gran precisión [5]. Sin embargo, el manejo de datos de diferentes sensores implica procesos computacionales adicionales para sincronizar y filtrar la información útil. Recientemente, otros enfoques proponen la clasificación de la actividad utilizando CNN [6]. En este contexto, las ventajas de las CNN son dobles ya que pueden operar simultáneamente como un extractor de características y un clasificador [7]. Aprovechando esta característica de las CNN, Neili *et al.* [8] propusieron un enfoque de reconocimiento de la pose humana para estimar la posición conjunta, junto con una máquina de vectores de soporte (SVM - *Support Vector Machine*) para obtener una clasificación final de la pose.

Por otro lado, debido al gran consumo energético de la refrigeración doméstica basada con compresión de vapor, en los últimos años se ha buscado optimizar el funcionamiento de un refrigerador a través de diferentes mecanismos [9]. A todo esto, desde hace varios años ha ido en aumento el uso de gases de efecto invernadero, afectando directamente a la temperatura promedio del ambiente. Muchos de estos gases son producidos debido a la generación de energía eléctrica. En este sentido, el campo de la refrigeración también contribuye de manera indirecta por el tipo de energía de entrada al sistema, además, de forma directa por el tipo de refrigerantes usados, con una alta afectación al calentamiento global. Lo anterior propicia que se investiguen diferentes áreas con el fin de proponer sistemas de refrigeración más eficientes y amigables con el medio ambiente. Por ejemplo, Mina y Gutiérrez [10] analizaron desde el punto de vista teórico la ubicación óptima de un elemento resistivo usando dos criterios de optimización: maximizando el calor suministrado al tubo que recibe la menor cantidad de calor y minimizando la desviación estándar normalizada del calor recibido por los tubos, esto para una configuración (diseño) particular de evaporador. Cabe destacar que el intercambiador de calor o evaporador de un refrigerador doméstico sin escarcha acumulada presenta una mejor transferencia de calor, así como una mejor distribución del flujo de aire

hacia los compartimentos, con lo que se logra tener mejores desempeños energéticos.

El tipo de elemento usado para evitar la acumulación de escarcha define de igual manera el consumo energético, por esta razón, Melo *et al.* [11] realizaron un estudio experimental para comparar diferentes elementos o calentadores eléctricos aplicados a la refrigeración doméstica, como calentadores distribuidos, de tubo de vidrio y calentadores tubulares. Los autores concluyeron que los calentadores estudiados ofrecieron la misma eficiencia al descongelar el intercambiador de calor, pero un menor consumo energético al usar un modo de activación del calentador que redujo la potencia de este conforme pasaba el tiempo, encontrando hasta un 48 % de eficiencia en el descongelamiento. Harrington *et al.* [12] instrumentaron un refrigerador doméstico para analizar el impacto energético al operar el sistema de deshielo, lo cual proporcionó información importante para describir los patrones de eliminación de la escarcha, además, propusieron un método para estimar la energía usada asociada a los sistemas de deshielo automáticos. Li *et al.* [13] propusieron un método para el descongelamiento de un evaporador, basándose en la colocación de diversos elementos resistivos en diferentes posiciones, logrando un 28 % de eficiencia en el descongelamiento.

## 1.5. Organización de la Tesis

Este documento se encuentra dividido en 5 capítulos. El texto descrito anteriormente proporcionó una breve introducción al contenido de la presente tesis doctoral. El capítulo 2 determina el marco teórico necesario y a la vez, establece las bases para realizar este proyecto. En el capítulo 3 se describe el primer caso de estudio: “Reconocimiento de actividades” donde se encuentra la metodología usada para llevar a cabo los objetivos generales y particulares del caso de estudio. El capítulo 4 detalla el segundo caso de estudio: “Refrigeración doméstica”, estableciendo los bancos experimentales usados así como cada una de las pruebas y procedimientos para la obtención de los resultados. En el capítulo 5 se expone la discusión de los casos de estudio, llegando así a las conclusiones finales del proyecto y perspectivas a futuro.

# Capítulo 2

## Aplicaciones de la inteligencia artificial

En este capítulo se describen los principales trabajos relacionados al presente proyecto. Se presenta además, la introducción a las bases teóricas que soportan los casos de estudio propuestos.

### 2.1. Inteligencia artificial

La IA es el estudio de las ciencias computacionales que permiten percibir, razonar y actuar [14]. Implica el aprendizaje interactivo o el desarrollo mediante diferentes algoritmos de optimización. La IA también permite a los sistemas interactuar con su entorno a través de sensores o datos preparados. En consecuencia, estos datos serán procesados y cambiará la forma de actuar de estos sistemas.

Los datos son una representación simbólica de hechos, conceptos o instrucciones. Estos pueden ser almacenados, procesados y comunicados. Cuando estos datos son procesados, se obtiene información, la cual tiene un significado en un contexto específico. El procesamiento de datos es fundamental para el ámbito de la ciencia de datos e inteligencia artificial. Los datos tienen una gran importancia, ya que cuando se procesa o analiza el conjunto de datos correcto es cuando se ha dado pie a los mayores avances en la IA [15].

Los datos se obtienen generalmente mediante sensores o dispositivos de medición, bases de datos, encuestas, entre otros. Los sensores son dispositivos diseñados para captar estímulos en su entorno. Estos capturan diferentes tipos de señales. Una señal es una representación física o abstracta de una magnitud que varía con respecto al tiempo, el espacio o cualquier otra variable independiente [16]. Existen diferentes tipos de señales naturales como pueden

ser señales de voz, de electrocardiogramas, incluso una señal de imagen [17]. Una imagen puede contener un canal, o lo que es lo mismo, una imagen en escala de grises. También hay imágenes con tres canales, o imágenes a color (RGB - *red, green, blue*). Finalmente, existen imágenes con hasta cuatro dimensiones, este último canal agrega profundidad (RGB-D). Las señales pueden transportar datos, los cuales puede ser utilizados en diferentes campos, como puede ser la electrónica, telecomunicaciones, procesamiento de imágenes, entre otros.

Como se observa en la Figura 2.1, existen diferentes ramas de la IA, entre ellas se encuentran los sistemas expertos, el procesamiento natural del lenguaje, aprendizaje máquina, entre otros. El procesamiento del lenguaje natural es una rama de la IA que se ocupa de la acción mutua entre las computadoras y los lenguajes humanos (naturales), en particular, para el análisis de grandes cantidades de datos en lenguaje natural. Ahora bien, un sistema experto trata de imitar el proceso de razonamiento de un humano para resolver un problema específico. Finalmente, el aprendizaje máquina es una rama de la IA que aspira a la mejora de algoritmos basados en el procesamiento de datos. Intenta obtener relaciones o patrones significativos de esos datos procesados [18].

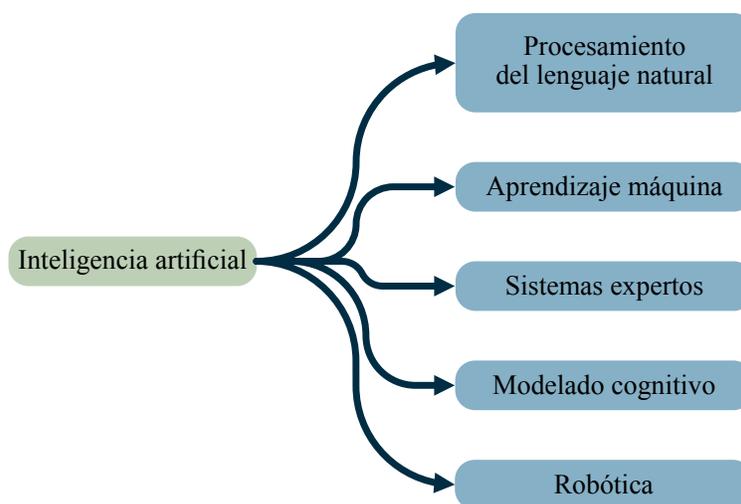


Figura 2.1: Principales ramas de la inteligencia artificial [19].

## 2.2. Aprendizaje máquina

El aprendizaje máquina es un área esencial de la IA. Sus aplicaciones generales van desde la visión por computadora, el análisis de documentos para extracción de texto, hasta en sistemas de recomendaciones, entre otros [20]. Las técnicas de aprendizaje máquina son usadas principalmente cuando un modelo clásico no es suficiente [19, 21], debido a que es demasiado lento o no encuentra una solución. El aprendizaje máquina se puede dividir principalmente en tres categorías: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje reforzado, ver Figura 2.2. El objetivo principal del aprendizaje máquina es agrupar datos que

tengan características afines, sin embargo para lograr esto cada categoría trabaja de diferente forma.

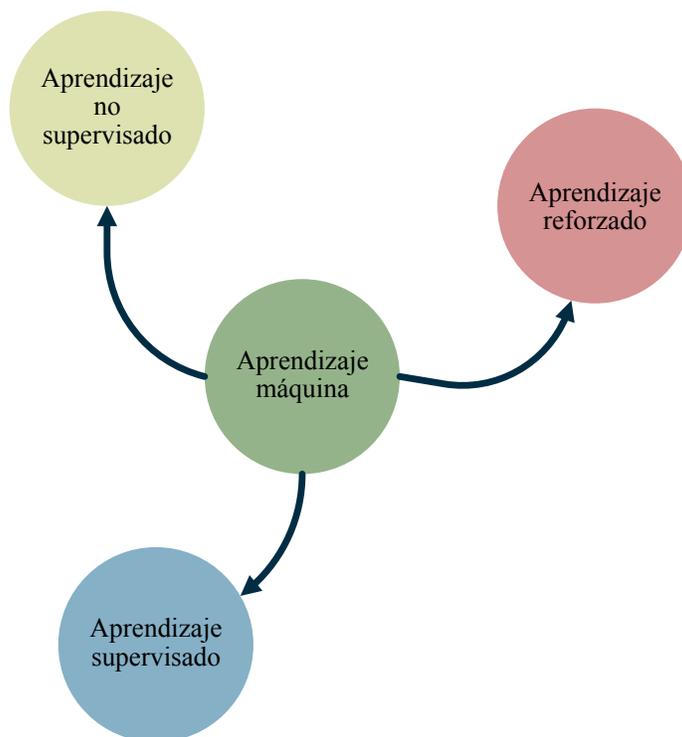


Figura 2.2: Principales técnicas del aprendizaje máquina [22].

En el aprendizaje supervisado existen diferentes enfoques para modelar datos, como las técnicas de aprendizaje automático y los modelos inteligentes. Un algoritmo de aprendizaje automático (supervisado) y/o un modelo inteligente son métodos que adaptan automáticamente sus parámetros para modelar con éxito un conjunto de entradas [23]. Esta adaptación se realiza a través del entrenamiento. Para ello, se requiere un conjunto de datos previamente clasificados por un sistema experto, este conjunto de datos es llamado conjunto de entrenamiento [21], dividido generalmente en conjunto de datos de entrenamiento y de validación. El algoritmo supervisado destaca las características principales de estos datos y agrupa aquellos con características similares. Esta categoría proporciona una retroalimentación indicando si una predicción es correcta o incorrecta, resultando en la clasificación correcta de la información. Al contrario, el aprendizaje no supervisado no tiene una respuesta o retroalimentación, el algoritmo simplemente intenta categorizar el conjunto de datos basado en sus características. Finalmente, el aprendizaje reforzado es similar al aprendizaje supervisado ya que recibe una retroalimentación, pero con una principal diferencia, ya que esta retroalimentación no es necesaria en cada entrada o estado [19].

## 2.3. Redes neuronales artificiales

Una de las principales ramas del aprendizaje supervisado son las redes neuronales artificiales (ANN - *Artificial neural networks*). Una ANN es un método computacional inspirado en procesos biológicos para resolver problemas difíciles para los sistemas de cómputo [24]. Una red neuronal cuenta con una gran ventaja sobre algoritmos tradicionales, ya que se puede adaptar a un amplio conjunto de problemas en los cuales falta un modelo matemático [25]. La unidad básica de una ANN es una neurona. En la Figura 2.3 se observa la estructura interna de una neurona, donde  $X_1, X_2, X_3$  son las entradas, junto a una entrada con valor 1 llamado *Bias*. Los valores  $W_{1X}$  son los pesos de cada entrada y finalmente,  $Z$  la salida.

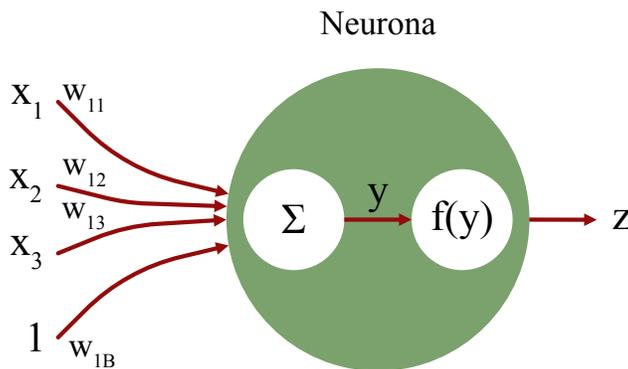


Figura 2.3: Estructura interna de una neurona.

### 2.3.1. Funciones de activación

Una función de activación es usualmente una función no lineal que determina la salida de una neurona [24]. La función de activación determina cuando una neurona está activa o no, dependiente de su salida  $Z$ , donde  $Z \approx 1$  refleja una neurona activa y  $Z \approx 0$  revela una neurona inactiva. La activación/desactivación de neuronas permiten establecer una relación entre los datos de entrada/salida. Esta relación se consigue mediante una función de activación. Existen diferentes funciones de activación, donde las más comunes son:

$$z = f(y) = \text{logsig}(y) = \frac{1}{1 + e^{-y}} \quad (2.1)$$

$$z = f(y) = \text{tanh}(y) = \frac{e^y - e^{-y}}{e^y + e^{-y}} \quad (2.2)$$

y

$$z = f(y) = y^+ = \max(0, y) = \frac{y + |y|}{2} = \begin{cases} y & \text{si } y > 0, \\ 0 & \text{lo contrario.} \end{cases} \quad (2.3)$$

El parámetro  $y$  de las funciones de activación se obtiene mediante:

$$y = \sum_{i=1}^n x_i \cdot w_i + Bias \cdot w_B \quad (2.4)$$

Donde  $x_i$  son los valores de entrada de cada neurona,  $w_i$  los pesos asociados al valor  $x_i$ ,  $Bias = 1$  y  $w_B$  el peso asociado a  $Bias$ , como se observa en la Figura 2.3.

Por lo tanto, al juntar varias neuronas se obtiene una red neuronal artificial. Dentro de estas, los pesos  $W_{1n}$  se utilizan como parámetros internos, para ajustar durante un entrenamiento supervisado. Junto a los pesos, el  $Bias$  permite a la neurona aumentar su flexibilidad de aprendizaje desplazando el valor de  $y$  según sea necesario. Este ajuste fino de parámetros o aprendizaje supervisado tienen como objetivo obtener una relación entre los datos de entrada y salida [22].

Si se observa la Figura 2.4, una ANN cuenta con al menos dos capas. Estas dos capas son una capa de entrada ( $i$ ) y una capa de salida ( $o$ ), donde cada capa tiene un número fijo de neuronas. Opcionalmente, una ANN cuenta con una o más capas ocultas ( $h_1 \dots h_n$ ), dando como resultado una red neuronal artificial multicapa.

Existen diferentes tipos de redes neuronales artificiales, como por ejemplo: las redes neuronales probabilísticas (PNN - *Probabilistic neural network*), donde utilizan una función exponencial definida por [26]:

$$R(x) = e^{(X_i - 1)/\sigma^2} \quad (2.5)$$

Donde  $\sigma$  es un “parámetro de suavizado” [26]. Una PNN puede verse como un método estadístico para analizar y modelar datos de forma similar al análisis realizado con una ANN. En la Figura 2.5 se puede observar el diagrama general de una PNN. En esta se observan las unidades de entrada, de patrón, de suma y finalmente, las unidades de salida. Estas unidades son necesarias para el funcionamiento de una PNN. Donde  $f_{A1}(X)$  y  $f_{B1}(X)$  son funciones de densidad de probabilidad para las categorías  $A$  y  $B$ .

Otro tipo específico de ANN es una red neuronal convolucional (CNN - *Convolutional Neural Network*). Este tipo de arquitectura permite realizar un mejor trabajo en tareas de visión por computador al detectar objetos y patrones, entre otros [27].

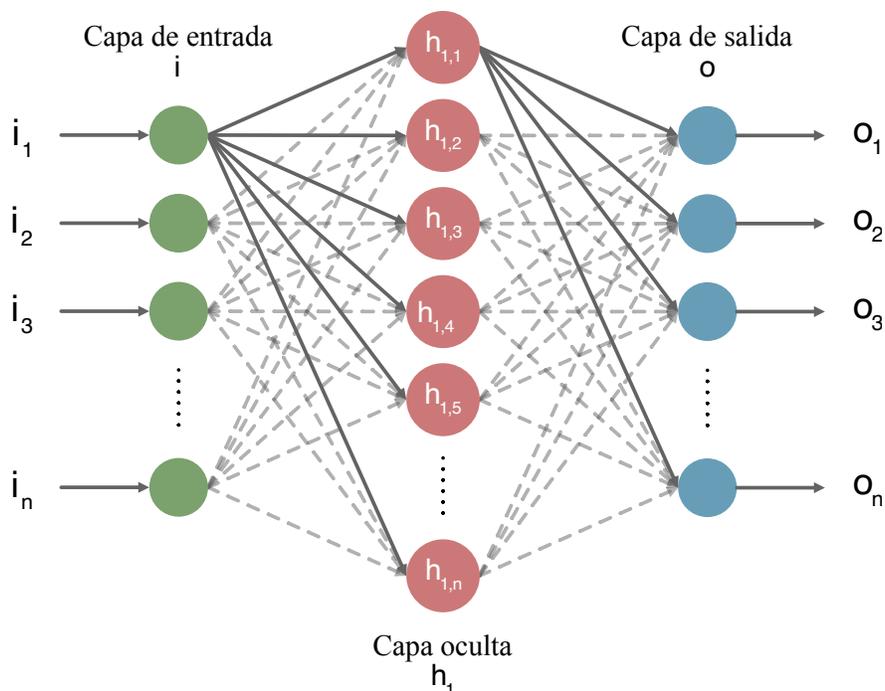


Figura 2.4: Estructura general de una red neuronal artificial.

## 2.4. Red neuronal convolucional

Una CNN es una red neuronal multicapa. En la Figura 2.6 se observa el diagrama general de una red neuronal convolucional.

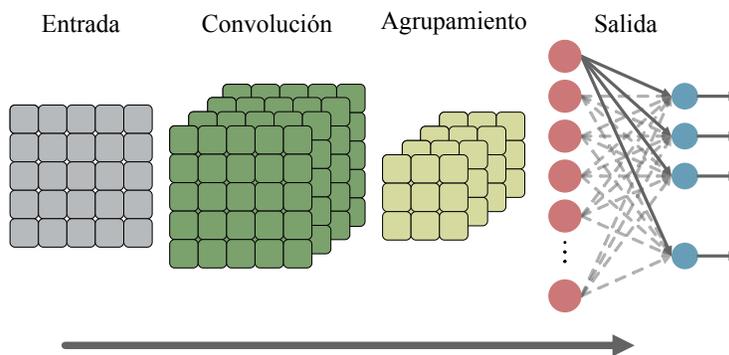


Figura 2.6: Diagrama general de una red neuronal convolucional.

Una red neuronal convolucional se divide en dos partes. La primera sección de una CNN tiene  $N$  capas convulsionales; este tipo de capa utiliza una convulsión, al ser en dos dimensiones, es descrita por:

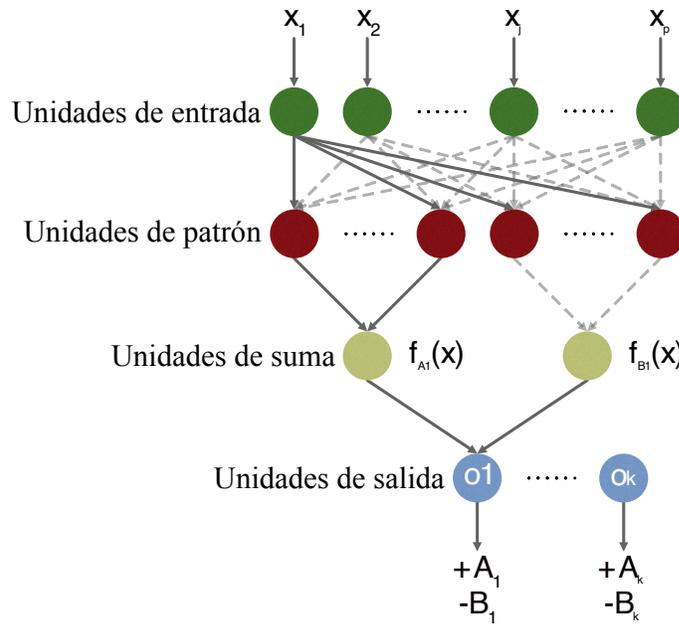


Figura 2.5: Diagrama general de una red neuronal probabilística [26].

$$B_{H,W} = \sum_i \sum_j A_{H+i,W+j} \cdot K_{i,j} \quad (2.6)$$

Donde  $B$  es la matriz resultante,  $A$  una matriz de tamaño  $H \times W$  y  $K$  una matriz convolucional de tamaño  $i \times j$  [28].

La segunda parte de una CNN suelen ser capas totalmente conectadas, o una red neuronal artificial. Estas capas reciben las características extraídas por las capas convolucionales. El entrenamiento del modelo alimenta las funciones de activación, que agrupan características similares para lograr un modelado correcto.

Una CNN requiere del uso de matrices convolucionales. Una matriz convolucional o kernel convolucional es un conjunto de ponderaciones que se aplica a los valores de cada elemento de un grupo de vectores o matrices, con la finalidad de resaltar o extraer un cúmulo de características de los datos originales. En la Figura 2.7 se muestran los dos primeros pasos al aplicar un kernel convolucional a una matriz en dos dimensiones.

## 2.5. Métricas de evaluación

Con el objetivo de medir el rendimiento o comportamiento de un modelo neuronal existen diversas métricas, como pueden ser: precisión, exactitud, F1-Score, entre otros. Estas métricas son basadas en los valores predichos por el modelo, en comparación con la etiqueta real

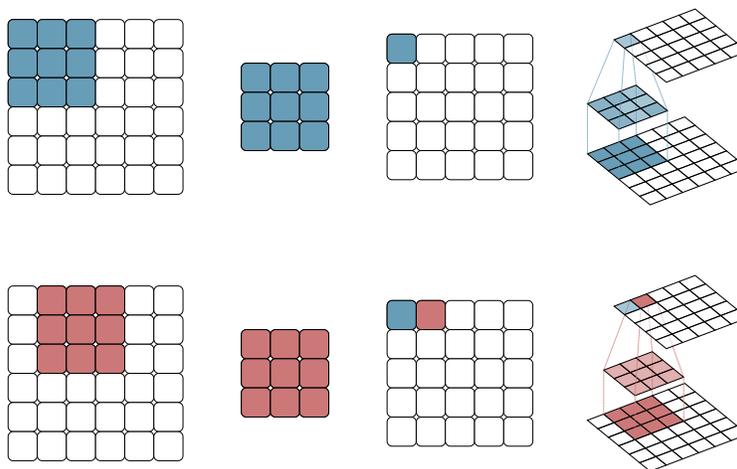


Figura 2.7: Primeros pasos de la aplicación de una matriz convolucional.

de los datos [29]. Por lo tanto, existen verdaderos positivos (TP - *True Positive*), verdaderos negativos (TN - *True Negative*), los cuales son las instancias que el modelo predijo correctamente. Por otra parte, existen falsos positivos (FP - *False Positive*) y falsos negativos (FN - *False Negative*), o aquellas que el modelo falló en su predicción/clasificación.

La exactitud muestra un porcentaje de las predicciones correctas sobre el total de las predicciones [29]. La exactitud se define como:

$$Exactitud = \frac{TP + TN}{TP + TN + FP + FN} * 100 \quad (2.7)$$

La precisión se usa principalmente cuando un conjunto de datos no se encuentra balanceado. Un conjunto de datos no balanceado cuenta con un mayor número de elementos de entrenamiento de alguna clase en específico. Esta situación muestra normalmente un alto valor en la exactitud, valor el cual no sería totalmente correcto. La precisión se representa como:

$$Precisión = \frac{TP}{TP + FP} \quad (2.8)$$

La inteligencia artificial es un conjunto de herramientas o algoritmos que tratan de dotar a una máquina o sistema informático la capacidad de razonar, con el objetivo de resolver problemas que normalmente requieren de inteligencia humana. Estas actividades o problemas incluyen el aprendizaje, percepción visual, reconocimiento de la voz, entre otros. Con el fin de lograr la solución de esos problemas, la IA requiere un conjunto de diferentes áreas o disciplinas. Por tal razón, se requiere la colaboración de diferentes campos de la investigación para poder aplicar la IA de manera efectiva.

Existen diversos campos actuales con un gran interés en el desarrollo de algoritmos inteligentes. Principalmente en el área del reconocimiento de actividades humanas y en el área de la refrigeración doméstica [30, 31, 32, 33]. Existen además, diferentes tecnologías para el desarrollo de algoritmos inteligentes. Entre ellas se encuentran bibliotecas de aprendizaje automático de código abierto como TensorFlow [34], Pytorch [35], entre otros. Este tipo de bibliotecas permiten la construcción y entrenamiento de redes neuronales. Además, permiten la ejecución de las herramientas creadas/entrenadas directamente en el CPU (Unidad central de procesamiento - *Central processing unit*) o GPU (Unidad de procesamiento gráfico - *Graphics processing unit*) del sistema de cómputo.

Debido al creciente interés en el desarrollo de algoritmos inteligentes en las áreas del reconocimiento de actividades humanas, además de, la refrigeración doméstica, en este trabajo doctoral se proponen dos casos de estudio.

El caso de estudio “Reconocimiento de actividades” (Capítulo 3) introduce los conceptos necesarios para el análisis de esta campo, así como también, una breve introducción y estado del arte general actual. Se plantea un modelo inteligente para la clasificación de actividades humanas. Este modelo está basado en las redes neuronales convolucionales, agregando una capa capaz de analizar los cambios en el tiempo. Finalmente, el caso de estudio muestra los resultados obtenidos y perspectivas a futuro.

El caso de estudio “Refrigeración doméstica” (Capítulo 4) describe los bancos y pruebas experimentales necesarios para su realización. Además, se introducen los modelos inteligentes utilizados para la clasificación de la acumulación de la escarcha. Se presenta una introducción de conceptos utilizados, asimismo del estado del arte actual. Por último, se presentan los resultados experimentales y perspectivas a futuro.

# Capítulo 3

## Reconocimiento de actividades

En este capítulo se presenta el caso de estudio: “Reconocimiento de actividades”. Se introducen las actividades humanas y su importancia en la investigación actual. Se propone una metodología de aprendizaje profundo, por lo tanto, se diseña un modelo inteligente convolucional temporal para el análisis de las actividades humanas a partir de secuencias cortas de video. Con el fin de analizar el desempeño del modelo propuesto, se utilizan tres diferentes bases de datos públicas: “*Kinect Activity Recognition*”, “*Cornell Activity*” y “*MSR Daily Activity 3D*”. Estas bases de datos son ampliamente usadas para la clasificación de actividades humanas. Asimismo, contienen diferentes escenas de interiores, así como al menos 10 actividades diferentes a reconocer, entre las cuales incluyen: beber, comer, leer un libro, caminar, sentarse, entre otros. Finalmente, se muestran los resultados obtenidos para el presente caso de estudio.

### 3.1. Introducción

En la actualidad, los sistemas autónomos proporcionan asistencia para realizar actividades en distintos campos, como la medicina, el entretenimiento, la educación o la mejora del hogar, entre otros. Sin embargo, para lograr una interacción satisfactoria con el sistema autónomo, estos sistemas deben ser capaces de identificar e interpretar ciertos comportamientos humanos. En este sentido, el reconocimiento de actividades humanas (HAR - *Human Activity Recognition*) se ha convertido en un tema inmensamente popular con el uso de sensores rápidos y potentes para adquirir los movimientos realizados por los humanos [36, 37]. Entre estos sensores, existe el sistema de Captura de Movimiento (MoCap - *Motion Capture*), que se utiliza para capturar, grabar y transmitir datos de movimiento y/o localización de humanos, animales u otros objetos. MoCap es un sistema muy utilizado en la industria cinematográfica porque ofrece resultados precisos en tiempo real. Del mismo modo, MoCap es capaz de recrear movimientos complejos e interacciones físicas realistas

con gran precisión [38]. Debido a estas importantes características, MoCap es una excelente opción para HAR. Sin embargo, MoCap es un sistema costoso que requiere trajes especiales (con marcadores adheridos) y matrices ópticas. Por otra parte, MoCap debe ser utilizado por personas con una formación especial para obtener un rendimiento óptimo. Finalmente, es importante tener en cuenta que el procesamiento de los datos producidos por MoCap requiere hardware y software especiales, además del elevado tiempo de procesamiento y los requisitos energéticos [39].

A diferencia de los sistemas MoCap, existen otros tipos de sistemas denominados “*markerless*”, los cuales no requieren un traje especial con marcadores. En este tipo de sistemas, los sensores proporcionan información de profundidad para obtener la posición real de un objeto y capturar su movimiento durante una actividad específica. Un ejemplo de sensor “*markerless*” es el Asus Xtion PRO, que es un sensor de cámara de luz estructurada RGB-D de bajo costo [40]. Este capta los movimientos corporales del usuario en tiempo real e incluye herramientas para desarrollar aplicaciones personalizadas. En los últimos años, el sensor Asus Xtion PRO se ha utilizado para la localización de robots, mapeo de obstáculos y la planificación de rutas en entornos 3D en tiempo real [41]. Del mismo modo, este sensor fue la base de un método interactivo para controlar una silla de ruedas mediante el reconocimiento de gestos, lo que facilita al usuario la realización de diversas tareas [42]. Incluso, este tipo de sensor se ha utilizado para la implementación de un sistema de teleoperación bimanual, aprovechando la extracción del esqueleto tridimensional [43]. Por último, el sensor Asus Xtion PRO se ha utilizado para la construcción de un robot de asistencia social para personas mayores [44], donde el reto de reconocer acciones humanas cobra mayor importancia.

Otro tipo de sensores muy utilizados para HAR son el acelerómetro y el giroscopio contenidos en la Unidad de Medición Inercial (IMU - *Inertial Measurement Unit*). En este caso, las señales de aceleración se adquieren utilizando múltiples sensores vestibulares. Estos sensores se fijan al cuerpo humano (normalmente en la muñeca, el tobillo y el pecho) cuando una persona está realizando actividades físicas [5, 45]. Las señales proporcionadas por un teléfono inteligente se han vuelto muy populares para el reconocimiento de la actividad humana [6], debido a que la mayoría de la gente tiene un teléfono inteligente. Un teléfono inteligente dispone de varios sensores, conexión inalámbrica y puede estar situado dentro del bolsillo de la persona que realiza la actividad. En consecuencia, se ha convertido en un dispositivo remoto habitual para adquirir, almacenar y procesar datos.

Por otra parte, las soluciones basadas en la visión computacional para HAR suelen implicar el uso de imágenes RGB-D (imágenes con profundidad). Estas son adquiridas por sensores que contienen una cámara RGB y un proyector de infrarrojos para medir la información de profundidad de la escena. En el mercado, varias marcas ofrecen cámaras para capturar imágenes RGB-D a gran velocidad. Sin embargo, las cámaras de profundidad *Intel RealSense* y *Microsoft Kinect* son muy populares ya que ofrecen una buena relación precio-calidad.

Sin embargo, debido al elevado costo de procesamiento, a la gran cantidad de datos generados por las imágenes en color y profundidad, es importante explorar otras opciones de HAR. Así, en este caso de estudio, presentamos una nueva metodología para el reconocimiento de

actividades humanas en secuencias cortas de video. Este método está basado en un proceso de convolución 3D así como en un proceso convolucional de memoria a largo-corto plazo (LSTM - *Long Short-Term Memory*). La información espacio-temporal del comportamiento humano es resaltada por la arquitectura basada en Redes Neuronales Convolucionales (CNN - *Convolutional Neural Networks*). La estructura propuesta proporciona una gran precisión y requiere pocos recursos de hardware al procesar la información espacio-temporal de las secuencias. La metodología propuesta fue evaluada en los conjuntos de datos de referencia: KARD [46], CAD-60 [47], y el *MSR Daily Activity 3D* [48], los cuales contienen al menos diez actividades diferentes.

## 3.2. Reconocimiento de actividades humanas

La interacción hombre-máquina requiere sistemas capaces de reconocer e incluso imitar ciertas actividades realizadas por los humanos. En este sentido, algunos enfoques ofrecen a los robots la capacidad de identificar comandos de voz [3], reconocer acciones humanas [4], o ambas, para permitir la comunicación e interacción con los humanos. Uno de los retos de HAR es la detección precisa y rápida de la acción realizada. Por lo tanto, los objetivos beneficiosos de los enfoques HAR que superan este reto incluyen la detección de acciones que pueden poner en peligro a las personas mayores para prevenir caídas [49] o ayudar a las personas en la rehabilitación física.

Combinando información procedente de diversos sensores, se pueden identificar muchas actividades humanas con gran precisión [5, 50]. Sin embargo, el manejo de datos procedentes de diferentes sensores implica procesos computacionales adicionales para sincronizar y filtrar la información útil. Las articulaciones esqueléticas se han utilizado ampliamente en HAR para medir con precisión las 25 articulaciones típicas del cuerpo humano. El uso de articulaciones esqueléticas supone un conjunto de datos reducido (en bytes) a pesar del gran número de ejemplos que pueden almacenarse. Las articulaciones esqueléticas proporcionan buenos resultados para un pequeño número de clases [51, 8], sin embargo, a medida que aumenta el número de clases, la calidad de los resultados disminuye [52]. Por otro lado, el reconocimiento de actividad mediante técnicas novedosas se ha utilizado eficazmente en imágenes RGB-D capturadas por el sensor *Microsoft Kinect* consiguiendo buenos resultados. Estas técnicas incluyen *Naive Bayes* [53], *Random Forest* [54] y *Dynamic Bayesian Mixture Models* [51].

Recientemente, otros enfoques proponen la clasificación de las actividades utilizando CNNs [45, 6, 52, 7, 55]. En este contexto, las ventajas de las CNNs son dobles, ya que pueden operar simultáneamente como un extractor de características y un clasificador [7, 56]. Aprovechando esta característica de las CNNs, Neili *et al.* [8] propusieron un enfoque de reconocimiento de la pose humana para estimar la posición conjunta, junto con una Máquina de Vectores Soporte (SVM - *Support Vector Machine*) para obtener una clasificación final de la pose. Tomas y Biswas [56], diseñaron una arquitectura profunda combinada, el modelo CNN aprende la representación del movimiento mientras que los Auto-Encodificadores

Dispersos (SAE - *Sparse AutoEncoder*) modelan los diferentes movimientos de las articulaciones humanas. Después, las puntuaciones de clase de cada red se combinan para obtener una puntuación final a partir de los fotogramas de entrada muestreados. Kim y Reiter [52], utilizaron un modelo CNN temporal para interpretar acciones humanas en 3D. En el mismo contexto, Caetano *et al.* [7] introdujeron una imagen arbórea de articulaciones de referencia para analizar las posiciones de las articulaciones humanas y al mismo tiempo, utilizan esta imagen como entrada para una CNN. Zhang *et al.* [57], determinaron mediante modelado matemático y experimental que los métodos convencionales de aumento de datos, como el recorte, generan ejemplos deficientes para el reconocimiento de actividades. Los autores desarrollaron un método de CNN siamesa basado en parches de movimiento que mejora la información de movimiento para los sistemas de reconocimiento de actividades. Su método determina una región en la que se encuentra la información crítica para el reconocimiento de la actividad humana. No obstante, uno de los inconvenientes del reconocimiento HAR mediante RGB-D es la falta de privacidad en la obtención de la información. Por este motivo, Rajput *et al.* [58] desarrollaron un método de CNN profunda en la nube con la prioridad de preservar la privacidad. En su método, ofuscaron la información mediante el uso de imágenes de la historia del movimiento (MHI - *Motion History Images*) y tres mapas de movimiento en profundidad (DMM - *Depth Motion Maps*). Banerjee *et al.* [59] calculan las posiciones de las articulaciones del esqueleto a partir de imágenes monocromáticas en lugar de imágenes en color para reducir el número de datos a procesar. Este método se complementa con características geométricas extraídas del análisis de la acción humana. Una combinación de modelos difusos y CNN permite una gran precisión en el reconocimiento de la actividad humana.

Recientemente se ha utilizado un nuevo método para HAR que utiliza redes neuronales capaces de detectar características a lo largo del tiempo. Adama *et al.* [60], presentaron un nuevo marco de Segmentación Adaptativa y Aprendizaje Secuencial (ASSL - *Adaptive Segmentation and Sequence Learning*). Este marco procesa la información que se encuentra en las articulaciones 3D del cuerpo humano. El proceso utiliza la energía del movimiento que, a su vez, obtiene acciones clave. Estas acciones clave muestran cambios en la aceleración durante algunas actividades humanas. Este marco utiliza una red neuronal recurrente LSTM.

Todos los enfoques anteriores suelen utilizar imágenes RGB-D y/o un acelerómetro para adquirir los movimientos humanos. Combinando las imágenes y la posición de las articulaciones para la clasificación de las actividades. Una técnica poco explorada para HAR es el uso de sólo información de fotogramas RGB contenidos en videos para realizar la actividad de reconocimiento. El reto de utilizar sólo imágenes en color implica perder la información de profundidad del entorno, lo que reduce la precisión en la detección de actividades. Para superar este problema, en este caso de estudio se propone una nueva metodología basada en modelos inteligentes para obtener características que modelen la posición del cuerpo a lo largo del tiempo. En concreto, se extrae una colección de características del conjunto de fotogramas consecutivos utilizando una combinación de un proceso convolucional 3D y una capa convolucional LSTM.

La capa convolucional LSTM utiliza la información proporcionada por la capa convolu-

cional 3D (que contiene información en el espacio y el tiempo a corto plazo) para aprender características a largo plazo. Las principales aportaciones de este caso de estudio son: 1) La combinación de una capa convolucional 3D y una capa convolucional LSTM para extraer y analizar simultáneamente información espacio-temporal de secuencias de video. 2) Un método que requiere menos recursos computacionales e información que los métodos del estado del arte, obteniendo al mismo tiempo excelentes resultados.

### 3.3. Metodología

Este caso de estudio propone el desarrollo de una red neuronal convolucional temporal para la extracción de características. A continuación, estas características se utilizan para clasificar las actividades humanas.

El método propuesto se muestra en la Figura 3.1. En este caso, la red recibe  $nF$  fotogramas; cada fotograma tiene un tamaño de  $160 \times 120 \times 3$  píxeles donde  $W$  representa la anchura y  $H$  la altura. Estos fotogramas proceden de un video corto que contiene la actividad humana que se desea reconocer.

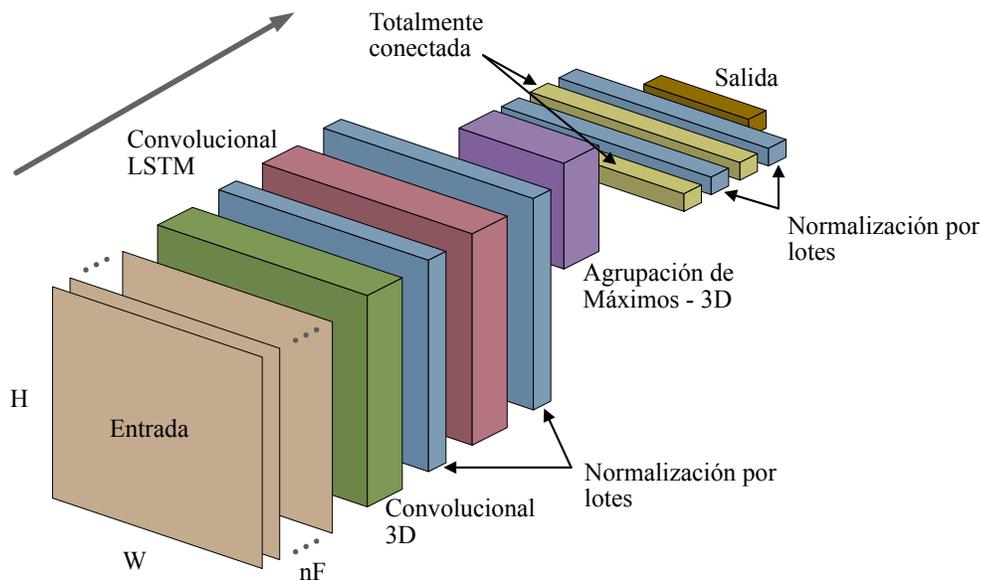


Figura 3.1: Arquitectura propuesta para la red neuronal convolucional temporal.

El enfoque propuesto es capaz de clasificar hasta 2 segundos de video a una velocidad de 30 fotogramas por segundo. En la Figura 3.2 se observan cuatro fotogramas de la actividad “ondulación horizontal del brazo” del conjunto de datos KARD [46]. Este tipo de secuencia es utilizado como entrada del modelo convolucional temporal ( $nF$  fotogramas) con la finalidad de la clasificación de las actividades humanas. La entrada del modelo convolucional temporal es extraída de segmentos cortos del video original, estos segmentos tienen una duración de

dos segundos y están conformados por 30 cuadros o fotogramas por segundo.

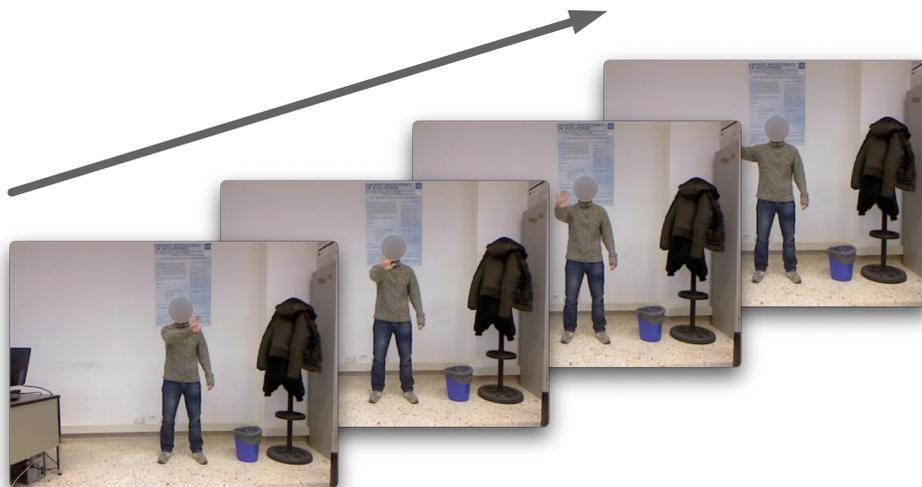
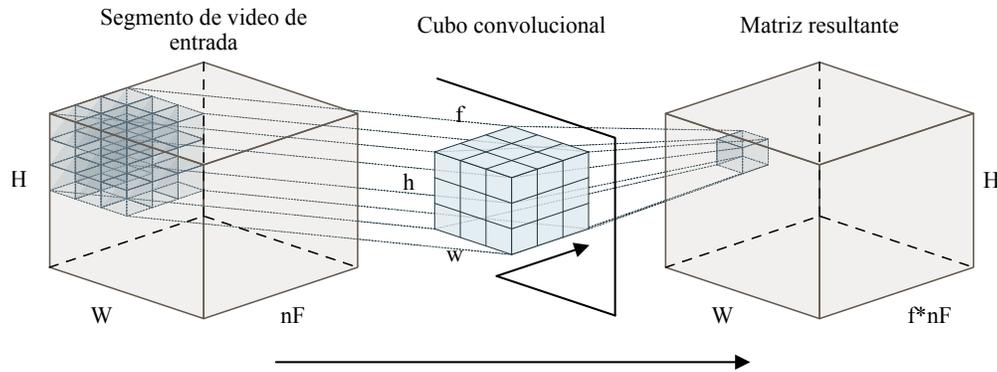


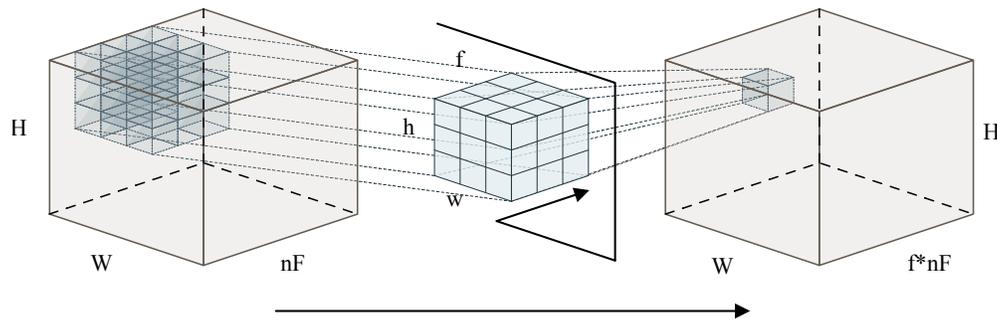
Figura 3.2: Secuencia con cuatro fotogramas de la actividad “ondulación horizontal del brazo” del conjunto de datos KARD [46].

El módulo de extracción de características consta de una capa convolucional tridimensional y una capa convolucional bidimensional LSTM, ambas con 48 filtros cada una. Las pruebas computacionales determinaron que 48 filtros ofrecían una buena relación entre el consumo de recursos y la correcta extracción de características, así como un buen rendimiento al combinar estos tipos de capas. Juntas, extraen características temporales a largo y corto plazo y, al mismo tiempo, extraen características espaciales. En consecuencia, esto permite a la arquitectura propuesta generalizar la acción realizada sobre la secuencia de entrada. Una vez extraídas las características temporales-espaciales, son clasificadas por dos capas totalmente conectadas con una función de activación Unidad Lineal Rectificada (ReLU - *Rectified Linear Unit*). Por último, la arquitectura propuesta utiliza una capa totalmente conectada con una función de activación softmax que contiene  $N$  salidas para  $N$  clases.

Se puede utilizar una convolución tridimensional para extraer características espaciales del mismo modo que una convolución bidimensional. Además, una convolución tridimensional puede extraer características temporales a corto plazo utilizando la información contenida en la dimensión adicional. En consecuencia, el cubo convolucional ( $h \times w \times f$ ) se utiliza para procesar espacialmente los fotogramas  $nF$  de entrada, así como para procesar las relaciones temporales que existen entre estos fotogramas. La Figura 3.3 muestra el proceso de una convolución tridimensional. En esta, la primera matriz representa el segmento de video de entrada que es convolucionado con el cubo convolucional con  $f$  filtros, dando lugar a una matriz 3D con la información resultante. Esta matriz tiene información temporal-espacial del video de entrada, es decir, las características temporales-espaciales. Esta información incluye la anchura, la altura y el resultado de los filtros  $f$  obtenidos para todos los  $nF$  fotogramas de entrada ( $W \times H \times f * nF$ ).



(a) Representación de un paso de una convolución.



(b) Siguiendo el proceso de convolución, representación de otro paso de una operación de convolución.

Figura 3.3: Representación de una convolución 3D.

La estructura de una capa LSTM convolucional bidimensional se muestra en la Figura 3.4. Además de utilizar las características visuales de las imágenes, la capa puede almacenar y procesar atributos temporales a largo plazo que se utilizan para crear la secuencia para el desarrollo de la actividad a reconocer. En este caso particular, todas las entradas  $X_1, \dots, X_t$ , salidas de celda  $C_1, \dots, C_t$ , estados ocultos  $H_1, \dots, H_t$  [61] son las matrices resultantes de la capa convolucional 3D. Combinando las características espaciales y temporales, obtenemos una correlación de características de corto-largo plazo que este tipo de capa puede generalizar.

### 3.3.1. Bases de datos

Con la finalidad de evaluar el rendimiento del modelo propuesto se probaron varias bases de datos ampliamente utilizadas para el reconocimiento de las actividades humanas. Los conjuntos de datos son: “*Kinect Activity Recognition: KARD*”, “*Cornell Activity: CAD-60*” y “*MSR Daily Activity 3D*”. Los conjuntos de datos cuentan con 18, 12 y 16 actividades humanas respectivamente. Asimismo, estos conjuntos de datos cuentan con un amplio análisis en el estado del arte, por lo tanto, son una excelente opción para comparar el rendimiento de diferentes clasificadores. Por último, es importante señalar que durante estas pruebas sólo se

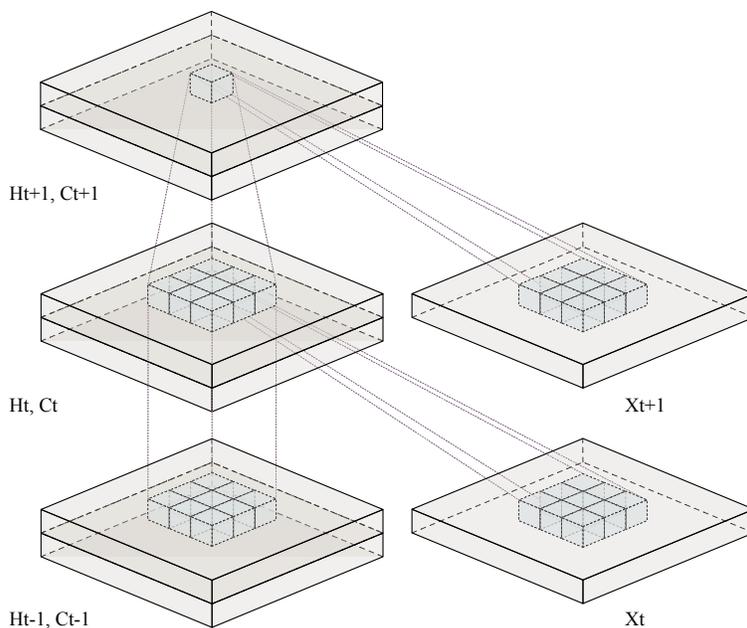


Figura 3.4: Representación de una capa ConvLSTM, con sus entradas, celdas de salida y estados ocultos [61].

utilizaron las secuencias de video (imágenes RGB) de las bases de datos, es decir, se descartó la demás información, como los datos de profundidad.

### Kinect Activity Recognition: KARD

En 2015 [46] se desarrolló el conjunto de datos KARD para el reconocimiento de actividades humanas. Este conjunto de datos consta de 18 actividades divididas en diez gestos (ondulación horizontal del brazo, ondulación alta del brazo, ondulación a dos manos, patada alta, dibujar x, dibujar tic, patada frontal, patada lateral, doblar y aplaudir con la mano) y ocho acciones (agarrar un sombrero, tirar un papel, sostener un paraguas, caminar, llamar por teléfono, beber, sentarse y ponerse de pie). Los gestos y las acciones fueron realizados tres veces por diez sujetos diferentes en escenarios de interior. Este conjunto de datos proporciona el mapa de profundidad, los respectivos videos RGB ( $640 \times 480 \times 3$ ) y las posiciones de las articulaciones. Obsérvese que estas posiciones están disponibles tanto en coordenadas reales como en coordenadas de pantalla.

En la Figura 3.5 se observan fotogramas de dos actividades pertenecientes al conjunto de datos KARD. De igual manera, también se observa el escenario de interior en donde se realizan las actividades.

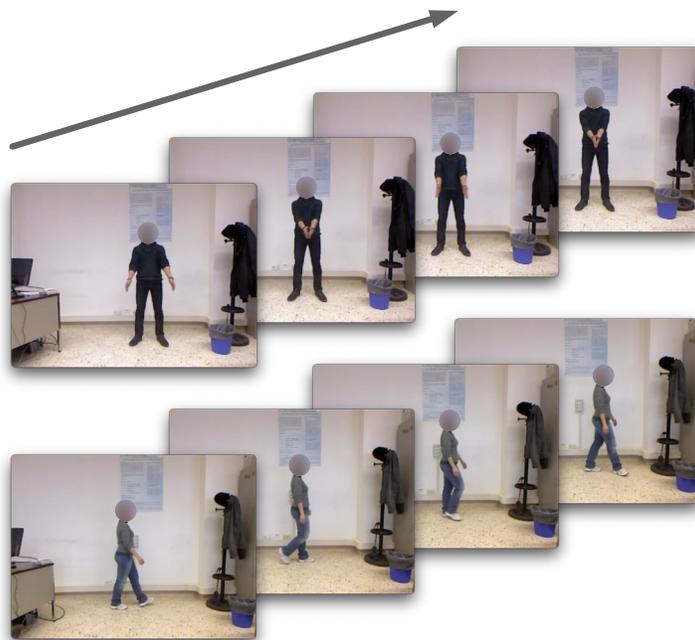


Figura 3.5: Fotogramas de dos actividades diferentes del conjunto de datos KARD [46].

### Cornell Activity: CAD-60

En 2012 [47] se desarrolló el conjunto de datos CAD-60 que contiene secuencias de video RGB-D de humanos realizando actividades comunes. El conjunto de datos contiene 12 actividades diferentes: enjuagarse la boca, cepillarse los dientes, llevar lentillas, hablar por teléfono, beber agua, abrir un envase de pastillas, cocinar-cortar, cocinar-remover, hablar en el sofá, relajarse en el sofá, escribir en una pizarra y trabajar en un ordenador. Estas actividades fueron realizadas por cuatro personas en distintos entornos o escenarios, como la oficina, la cocina, el dormitorio, el baño y el salón. En la Figura 3.6 se observan fotogramas de diferentes actividades del conjunto de datos CAD-60.

### MSR Daily Activity 3D

En 2012 [48] se preparó este conjunto de datos recopilando información de actividades cotidianas comunes realizadas por los seres humanos. Contiene 16 tipos de actividades, entre las que se incluyen: beber, comer, leer un libro, llamar por teléfono móvil, escribir en un papel, utilizar un ordenador portátil, utilizar una aspiradora, animarse, permanecer sentado, lanzar un papel, jugar a un juego, tumbarse en un sofá, caminar, tocar la guitarra, levantarse o sentarse. Estas acciones fueron realizadas por diez personas diferentes mientras estaban sentadas o de pie cerca de un sofá. El conjunto de datos proporciona, además de los datos



Figura 3.6: Fotogramas de diferentes actividades del conjunto de datos CAD-60 [47].

de profundidad, los respectivos datos RGB y las posiciones conjuntas de los actores. En la Figura 3.7 se observan actividades pertenecientes al conjunto de datos *MSR Daily Activity 3D*.

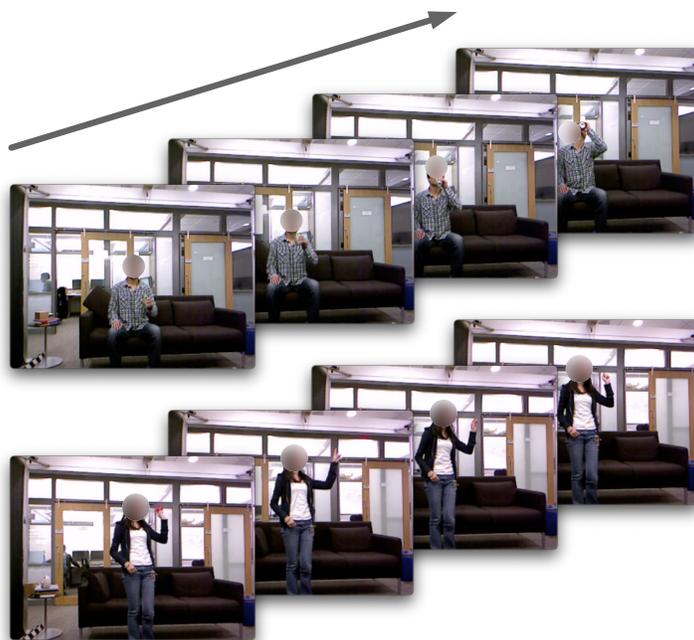


Figura 3.7: Fotogramas de diferentes actividades del conjunto de datos *MSR Daily Activity 3D* [48].

## 3.4. Resultados obtenidos

Esta sección valida el rendimiento de la arquitectura propuesta analizada en la Sección 3.3, incluyendo el tiempo de entrenamiento del modelo desde cero y el tiempo de clasificación de  $nF$  fotogramas de video. Además, el rendimiento del método propuesto se compara con otros trabajos relacionados en el estado del arte.

### 3.4.1. Evaluación global de la arquitectura propuesta

Una vez finalizado el diseño de la arquitectura propuesta y analizados los respectivos conjuntos de datos, se procedió a entrenar el modelo propuesto utilizando cada uno de los conjuntos de datos. A continuación, se analizan los resultados para calcular la matriz de confusión, que permite visualizar rápidamente los resultados de la validación. En otras palabras, la matriz de confusión compara la clase verdadera de los datos con la clase estimada por la red neuronal.

Las Figuras 3.8, 3.9 y 3.10 muestran la matriz de confusión obtenida para el conjunto de datos KARD [46], el conjunto de datos CAD-60 [47] y el conjunto de datos de actividad diaria MSR [48], respectivamente. En la Figura 3.8 se puede ver que se reconocen completamente las 18 clases del conjunto de datos KARD. En la Figura 3.9 se observa que las 12 clases del conjunto de datos CAD-60 también son clasificadas correctamente por la red neuronal. Debido a la gran variabilidad de las actividades, es posible obtener resultados de clasificación con una alta precisión y exactitud con valores hasta del 100 %. En el caso del conjunto de datos *MSR Daily Activity 3D*, se obtuvo una precisión de clasificación del 95.6 %. Como se muestra en la matriz de confusión de la Figura 3.10, actividades como ponerse de pie y levantarse muestran una alta similitud para el reconocimiento de las actividades humanas.

### 3.4.2. Comparación con trabajos relacionados

Otra parte importante de una arquitectura además de su rendimiento computacional, es su capacidad para clasificar correctamente los datos proporcionados. En este sentido, la red neuronal temporal propuesta fue comparada con algunos trabajos relacionados que utilizan secuencias de imágenes para HAR, y así, estimar el rendimiento del método propuesto para clasificar actividades humanas.

La Tabla 3.1 muestra los datos de precisión y sensibilidad del método propuesto y dos métodos comunes en el estado del arte para el conjunto de datos KARD. De esta tabla se desprende que el método propuesto supera a los otros dos métodos obteniendo excelentes resultados.

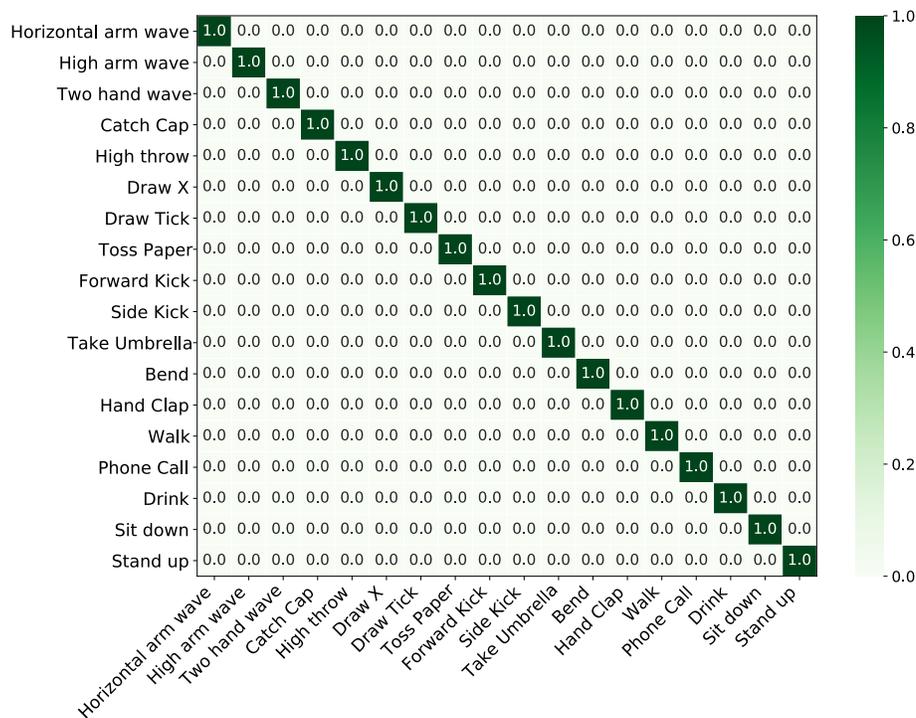


Figura 3.8: Matriz de confusión para el conjunto de datos KARD [46].

Tabla 3.1: Valores de precisión y sensibilidad del estado del arte (%) en el conjunto de datos KARD

	Método	Exactitud (%)	Precisión (%)	Sensibilidad (%)
Gaglio <i>et al.</i> [46]	Km-SVM-HMM	95	84.8	84.5
Cippitelli <i>et al.</i> [62]	Ca-SVM	95	95.1	95
Nuestra propuesta	T-CNN	100	100	100

La Tabla 3.2 muestra los resultados para el conjunto de datos CAD-60 utilizando el método propuesto y otros seis métodos comunes en el estado del arte. A partir de esta tabla, se observa que este método proporciona un excelente rendimiento para este conjunto de datos.

La Tabla 3.3 muestra los resultados para el conjunto de datos MSR Daily Activity 3D utilizando el método propuesto y dos métodos típicos en el estado del arte. Debido a la complejidad del conjunto de datos, ninguno de los métodos obtuvo una exactitud del 100%. En este caso, este método obtuvo la mayor exactitud para este conjunto de datos específico.

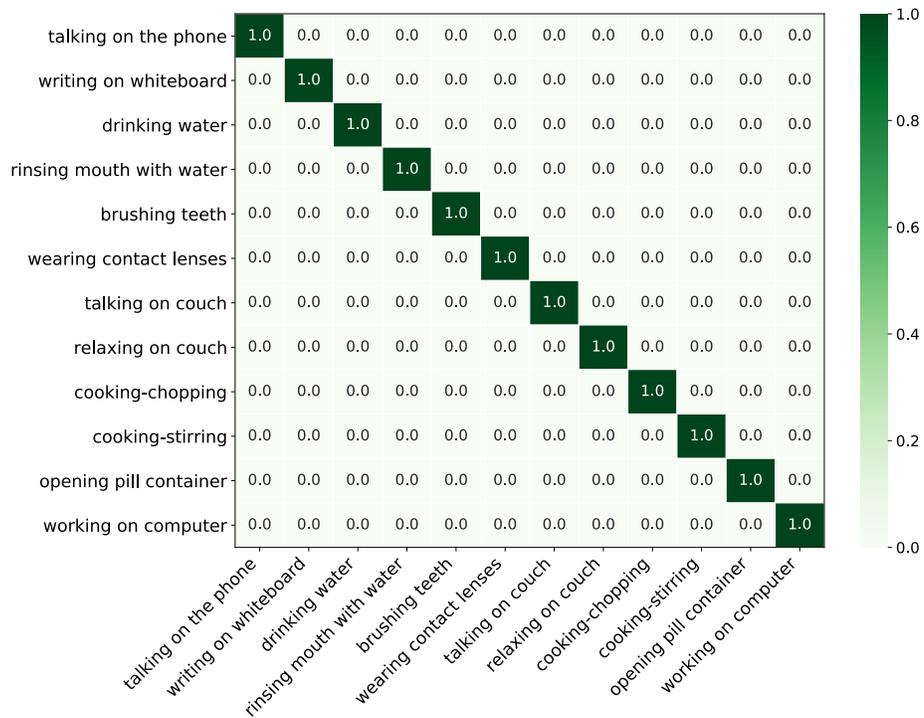


Figura 3.9: Matriz de confusión para el conjunto de datos CAD-60 [47].

Tabla 3.2: Valores de precisión y sensibilidad del estado del arte (%) en el conjunto de datos CAD-60.

	Método	Precisión (%)	Sensibilidad (%)
Yang y Tian [53]	NBNN	71.9	66.6
Gaglio <i>et al.</i> [46]	Km-SVM-HMM	77.3	76.7
Faria <i>et al.</i> [51]	DBMM	91.1	91.9
Cippitelli <i>et al.</i> [62]	Ca-SVM	93.9	93.5
Neili <i>et al.</i> [8]	CNN-SVM	97.08	97.2
Li <i>et al.</i> [50]	ShapeDTW-NBNN	97.8	97.3
Nuestra propuesta	T-CNN	100	100

Tabla 3.3: Valores de exactitud del estado del arte (%) en el conjunto de datos MSR Daily Activity 3D.

	Método	Exactitud (%)
Li <i>et al.</i> [50]	ShapeDTW-NBNN	90.8
Tomas y Biswas [56]	CNN-SAE	91.3
Nuestra propuesta	T-CNN	95.6

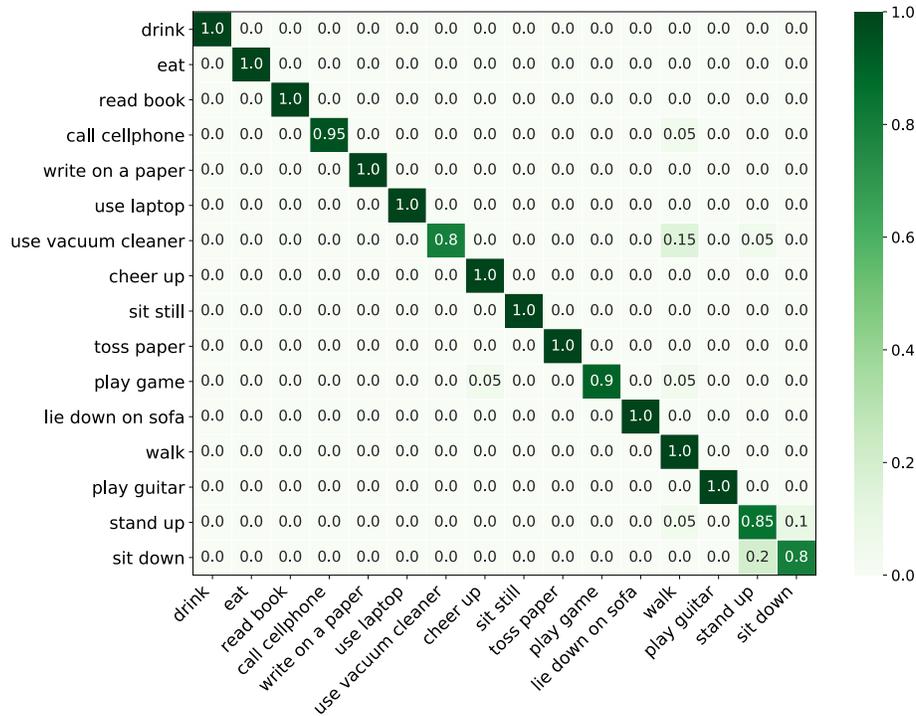


Figura 3.10: Matriz de confusión para el conjunto de datos *MSR Daily Activity 3D* [48].

### 3.4.3. Análisis del tiempo de procesamiento de la arquitectura

Uno de los principales objetivos de la arquitectura propuesta es que pueda ser fácilmente entrenada desde cero. Además, el método propuesto proporciona un tiempo de ejecución/clasificación rápido, y en consecuencia, puede ser utilizado en aplicaciones en tiempo real. Las pruebas de rendimiento se realizaron con un CPU Ryzen 7 1700 @ 3 GHz, un GPU NVIDIA GTX 1080 y memoria RAM DDR4 @ 2666 MHz, utilizando Windows 10 PRO (20H2) y CUDA 10.1.

La Tabla 3.4 muestra el tiempo que se tarda en entrenar la arquitectura propuesta para diferentes conjuntos de datos. La diferencia de tiempo se debe al tamaño y configuración de los conjuntos de datos. De la tabla se obtiene que el tiempo que tarda la arquitectura propuesta para clasificar es de  $60 \pm 10$  milisegundos. Clasificando hasta 2 segundos de video continuo, considerando una tasa de 30 fotogramas por segundo.

Tabla 3.4: Tiempo de entrenamiento y clasificación de hasta 2 segundos de video simultáneo.

Conjunto de datos	Entrenamiento (minutos)	Clasificación (milisegundos)
KARD	$70 \pm 5$	$60 \pm 10$
CAD-60	$50 \pm 5$	$60 \pm 10$
MSR Daily Activity 3D	$60 \pm 5$	$60 \pm 10$

A continuación se expondrán las principales aportaciones de este caso de estudio. Los métodos del estado del arte suelen analizar el video inspeccionando la información de cada fotograma, es decir, fotograma a fotograma. En cambio, el método propuesto en este caso de estudio analiza simultáneamente la información de varios fotogramas. Así, la principal contribución de este caso de estudio es que este método es capaz de detectar y medir diferencias y similitudes entre objetos de la escena, a partir de diferentes fotogramas de la secuencia de video.

Adicionalmente, se propone el uso de redes neuronales convolucionales temporales para mejorar la detección de actividades humanas. El uso de este tipo de redes es la segunda aportación de este caso de estudio. La principal ventaja de utilizar una red neuronal convolucional temporal es que este tipo de red puede detectar de forma natural el desplazamiento y el cambio de tamaño/posición de varios objetos en la escena de los fotogramas de video. La red propuesta en este trabajo se diseñó con sólo dos capas para la extracción de características y tres capas para la clasificación. Esto significa que este método requiere menos recursos computacionales que otros métodos del estado del arte, asimismo, permite el entrenamiento/clasificación de hasta dos segundos de video por actividad. Este diseño compacto requiere muy pocas operaciones, lo que permite clasificar rápidamente los movimientos humanos sin sacrificar la exactitud. Por lo tanto, se concluye que el método propuesto es una opción adecuada para el reconocimiento de actividades humanas debido a todas sus ventajas.

### 3.5. Discusión

En los últimos años, la alta disponibilidad de potencia computacional y el constante crecimiento de la interacción hombre-máquina, han promovido el desarrollo de diferentes estrategias para interpretar las acciones realizadas por los humanos a través de una máquina. Este desarrollo se encuentra en constante expansión en diversos campos como la industria 4.0 y la salud. En este contexto, se han desarrollado varios sistemas para extraer información de las muestras de datos recogidas por diferentes tipos de sensores. Estos sistemas se utilizan normalmente para analizar y reconocer actividades humanas.

En este caso de estudio, se propone una red neuronal convolucional temporal para analizar y clasificar actividades humanas. La arquitectura propuesta procesa un video corto de aproximadamente 2 segundos, a una tasa de 30 cuadros por segundo. En consecuencia, este método puede analizar simultáneamente varios fotogramas del video. A diferencia de los métodos más avanzados que analizan fotograma a fotograma, este método tiene la ventaja de poder detectar y medir la diferencia y similitud entre objetos de la escena a partir de fotogramas secuenciales. De este modo, la red neuronal convolucional temporal es capaz de identificar de forma natural el desplazamiento de los objetos o su cambio de tamaño.

Uno de los principales objetivos del diseño de esta arquitectura de red neuronal convolucional temporal es reducir el número de recursos computacionales. Por lo tanto, la red

propuesta tiene un tamaño muy reducido y puede entrenarse rápidamente en comparación con otros modelos del estado del arte. Por último, se concluye que la arquitectura propuesta puede utilizarse en lugares donde exista una restricción de recursos de hardware para analizar y reconocer actividades humanas en tiempo real.

Los resultados experimentales muestran que la arquitectura propuesta tarda poco tiempo en entrenarse cuando se utilizan algunos conjuntos de datos clásicos del estado del arte. Además, las simulaciones por ordenador indican que este método consigue mejores resultados de clasificación que los obtenidos por otros métodos del estado del arte para actividades humanas. Finalmente, este enfoque cobra especial relevancia en la vida cotidiana, donde el reconocimiento de actividades humanas se integra de manera omnipresente. Desde la extracción constante de datos a través de sensores portátiles como teléfonos inteligentes, hasta su aplicación en entornos urbanos inteligentes donde se propone mejorar continuamente la calidad de vida de los habitantes [63, 64].

# Capítulo 4

## Refrigeración doméstica

En este capítulo se presenta el caso de estudio: “Refrigeración doméstica”. Se introducen los bancos experimentales, los cuales son diseñados con el objetivo de construir al menos dos conjuntos de datos y realizar las pruebas experimentales. Estos conjuntos de datos constan de señales acústicas e imágenes RGB. Estos datos son capturados al interior del compartimento del evaporador de un refrigerador doméstico. Al mismo tiempo, se desarrollan pruebas experimentales utilizando técnicas de aprendizaje - máquina usando los conjuntos de datos obtenidos. Las pruebas experimentales tienen la finalidad de clasificar la cantidad de escarcha acumulada sobre la superficie del evaporador en un refrigerador doméstico. Por lo tanto, se proponen metodologías de modelos neuronales artificiales, probabilísticos y convolucionales para la clasificación de la escarcha acumulada. Finalmente, se muestran los resultados y discusiones obtenidos para el presente caso de estudio.

### 4.1. Introducción

Un refrigerador es un electrodoméstico utilizado principalmente para la conservación de alimentos. Basado en la compresión de vapor, un refrigerador doméstico representa uno de los artículos de mayor consumo energético en cualquier hogar. El sector de la refrigeración se estima que consume alrededor del 20 % de la energía generada en todo el mundo [1]. Dentro de este segmento existen los refrigeradores del tipo “*no-frost*” que evitan la acumulación de escarcha incorporando una técnica clásica en su sistema de descongelación. Esta técnica generalmente consiste en un elemento resistivo, el cual permite eliminar la acumulación de la escarcha. Como consecuencia, esto supone un costo energético muy elevado.

En los refrigeradores domésticos, la mayor parte de la humedad se concentra en el compartimento de alimentos frescos, lo que provoca una concentración de escarcha en el congelador como medio poroso, más concretamente sobre la superficie del evaporador, tal y como se

muestra en la Figura 4.1. Esta humedad conduce a la formación y acumulación de escarcha alrededor de los tubos y aletas en todo el evaporador, Figura 4.1b. A medida que aumenta el grosor de la capa de escarcha, se induce una disminución del flujo de aire y de la transferencia térmica, lo que provoca una reducción en la capacidad de enfriamiento del refrigerador y por ende, en su desempeño [65].

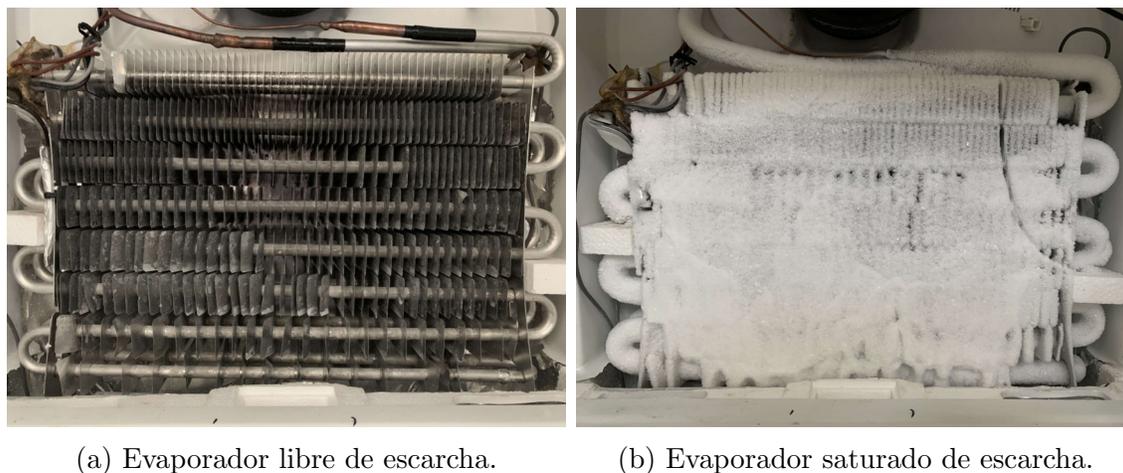


Figura 4.1: Evaporador típico de un refrigerador doméstico.

Con el fin de prevenir la disminución del desempeño del refrigerador, se pueden utilizar varias técnicas de descongelación [66]. Estas técnicas son mecanismos que eliminan la escarcha acumulada. Amer y Wang [67] presentaron un estudio experimental para la descongelación de una placa fría. El estudio utilizó vibración ultrasónica bajo convección natural. Los autores obtuvieron una reducción del espesor de la escarcha del 24%. En los refrigeradores domésticos “no-frost”, el método más utilizado es el uso de un elemento calefactor resistivo. Este elemento calefactor tiene ciertas ventajas, pero también ciertos inconvenientes. La desventaja más importante es el elevado consumo de energía [9]. En resumen, esta técnica común de descongelación influye considerablemente en el rendimiento energético del refrigerador, ya que este método requiere un elemento calefactor para eliminar la escarcha del evaporador.

Debido al elevado consumo energético de la refrigeración doméstica, en los últimos años se ha incrementado la investigación para optimizar globalmente el funcionamiento de un refrigerador a través de diferentes mecanismos [9]. Respecto a estos mecanismos, existen varios estudios para analizar diversas técnicas de desescarche para reducir el consumo energético [66]. Melo *et al.* [68] realizaron un estudio experimental para comparar diferentes calentadores eléctricos, como calentadores distribuidos, calentadores de tubo de vidrio y calentadores tubulares. Li *et al.* [13] desarrollaron un método para descongelar un evaporador basado en la colocación de varios elementos resistivos; los autores del estudio alcanzaron una eficiencia de descongelación del 28%. Harrington *et al.* [12] instrumentaron un refrigerador doméstico para analizar el impacto energético del funcionamiento del sistema de descongelación; propusieron un método para estimar la energía asociada a la descongelación automática.

Además del uso y estudio de los diferentes métodos para reducir la acumulación de escarcha comentados anteriormente, existen numerosos métodos de control y medición en refrigeración doméstica basados en inteligencia artificial. Ledesma y Belman-Flores [69] desarrollaron una herramienta basada en redes neuronales artificiales para construir mapas energéticos de un sistema de compresión de vapor; concluyeron que las redes neuronales podían utilizarse para predecir el coeficiente de rendimiento (COP - *Coefficient Of Performance*) de un sistema de compresión de vapor. Belman-Flores *et al.* [70] diseñaron e implementaron un novedoso controlador basado en lógica difusa para optimizar el consumo energético con relación a los hábitos de uso; lograron una mejora del 3% en el ahorro energético.

Sin embargo, los enfoques más avanzados requieren sensores muy especializados y un trabajo experimental complejo. En consecuencia, existe la oportunidad de explorar alternativas para clasificar la cantidad de escarcha en la superficie del evaporador de un refrigerador doméstico. La novedad del presente caso de estudio es la unión de la visión por computadora, análisis de señales acústicas y la inteligencia artificial, todo ello aplicado a la refrigeración doméstica. Por lo tanto, en este caso de estudio, se propone el desarrollo de un novedoso método inteligente que sea capaz de detectar/clasificar la acumulación de escarcha en la superficie de un evaporador. El método propuesto es fácil de implementar debido a los sensores utilizados. Este método clasifica la acumulación de escarcha en el evaporador utilizando ANN, PNN y CNN. ANN y PNN son técnicas de modelado utilizadas principalmente para simulación, optimización, reconocimiento de patrones, entre otros [71]. Estas técnicas permiten clasificar en tiempo real la acumulación de escarcha en la superficie del evaporador en cuatro niveles (clases) de escarcha diferentes. Estos modelos utilizan una cantidad mínima de recursos computacionales y son fáciles de implementar.

## 4.2. Señales acústicas

El uso de señales acústicas podría reducir la complejidad de un nuevo método de detección de escarcha. Se obtiene así un sistema económico y adaptable de gran precisión para la detección de escarcha. Las señales acústicas o sonido están formadas por variaciones de presión en un medio, como el aire. Estas variaciones se generan por la vibración de algunos objetos o estructuras, como un altavoz electrónico. El sonido se suele representar principalmente por dos propiedades, su amplitud y su frecuencia [72]. La amplitud del sonido indica la magnitud de las variaciones de presión. Cuanto mayor es este valor, mayor es la cantidad de sonido que puede percibirse. Por otro lado, la frecuencia del sonido indica la velocidad de estas variaciones de presión por segundo.

Según el análisis espectral, una señal de audio u onda sonora puede representarse como la suma de varias señales sinusoidales con diferentes amplitudes y frecuencias [17]. Una onda acústica tiene otras propiedades físicas como la absorción, la reflexión y la refracción durante su propagación. Cuando una onda acústica alcanza una superficie rígida, la mayor parte de la señal se refleja, y sólo un pequeño porcentaje de la señal se absorbe. La energía de la

señal absorbida se disipa cuando alcanza el nuevo medio. Este fenómeno físico afecta a la propagación del sonido [73].

Las señales acústicas o las ondas sonoras interactúan en mayor o menor medida dependiendo del tipo de material o medio en el que viajan. Un tipo de material que absorbe las señales acústicas a altas frecuencias es un medio poroso. Cuanto mayor es la frecuencia, mayor es la absorción de la onda sonora [74]. La escarcha acumulada en la superficie del evaporador es un medio poroso; por lo tanto, las ondas sonoras de alta frecuencia serán absorbidas. Esto significa que la energía de las ondas reflejadas será pequeña. En resumen, puede decirse que la cantidad de energía de las ondas reflejadas depende de la onda de frecuencia y de la cantidad de escarcha acumulada.

Por otra parte, la cavidad del evaporador de un refrigerador doméstico representa una caja acústica en la que las ondas sonoras o señales acústicas interactúan con el evaporador y todos los elementos circundantes, como se muestra en la Figura 4.2. Esto significa que el entorno de la cavidad del evaporador afecta a la absorción y reflexión de las ondas sonoras. Cuanto mayor sea la acumulación de escarcha en la superficie del evaporador, mayor será la interacción de las señales acústicas. Así, la propagación de las ondas sonoras en la cavidad del evaporador presenta diferentes fenómenos físicos como absorción, reflexión, entre otros.

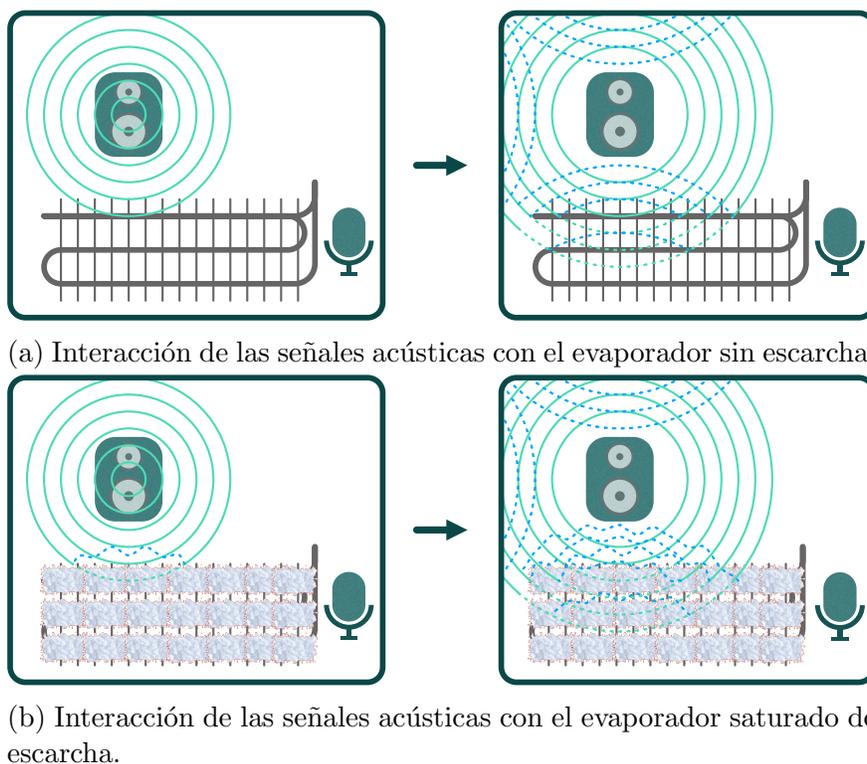


Figura 4.2: Propagación de señales acústicas en el interior del compartimento del evaporador.

Con el propósito de crear un sistema capaz de medir la cantidad de escarcha acumulada en la superficie de un evaporador, se diseñó un algoritmo paralelo para distribuir el flujo de

trabajo de operación de los sensores de forma concurrente. La Figura 4.3 muestra el diagrama general del flujo de trabajo para operar el sistema de generación/adquisición de datos.

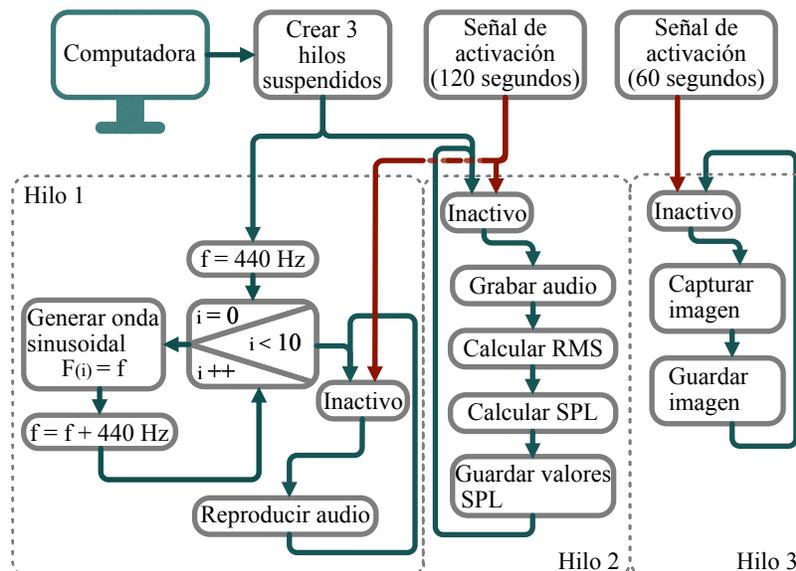


Figura 4.3: Esquema general del sistema de generación/adquisición de datos.

El algoritmo del sistema de adquisición de datos crea tres hilos de ejecución. Estos hilos pueden estar inactivos (suspendidos) o en ejecución. Los hilos están la mayor parte del tiempo inactivos; dos de ellos se despiertan y comienzan la ejecución cada 120 segundos, el hilo restante despierta cada 60 segundos. El hilo de ejecución 1 se encarga de la creación y reproducción de las señales digitales. El hilo de ejecución 2 captura las señales de audio propagadas en el interior de la cavidad del evaporador, tal y como se observa en la Figura 4.2. El hilo de ejecución 3 captura y almacena una imagen RGB de una parte del evaporador, el cual se describirá en la subsección siguiente (ver 4.3).

El sistema de adquisición de datos para las señales acústicas consta de dos partes, una que genera diez señales sinusoidales en diferentes frecuencias y otra que captura la presencia de estas señales en la cavidad del evaporador. El hilo de ejecución 1 crea estas señales digitales utilizando la función seno. Cada una de estas señales tiene una duración de dos segundos; tienen una frecuencia de muestreo de  $44.1 \text{ kHz}$  y una resolución de muestreo de 16 bits. La primera señal se crea con una frecuencia de  $440 \text{ Hz}$ . La segunda señal tiene una frecuencia de  $880 \text{ Hz}$ , que es una frecuencia de  $440 \text{ Hz}$  superior a la frecuencia de la primera señal. Las ocho señales restantes tienen valores de frecuencia con incrementos de  $440 \text{ Hz}$  hasta que la última frecuencia alcanza los  $4400 \text{ Hz}$ .

Se obtienen 44100 muestras por segundo para cada señal. Por lo tanto, hay una gran cantidad de datos que necesitan ser procesados. Sin embargo, para los modelos inteligentes, es importante reducir la cantidad de datos mediante una técnica llamada preprocesamiento. El objetivo principal del preprocesamiento es reducir la dimensionalidad de los datos. En este caso de estudio, para reducir la dimensionalidad de los datos se propone el cálculo del nivel de presión sonora (SPL). El SPL mide la intensidad acústica como:

$$SPL = 20 \log \left( \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i)^2}}{P_{ref}} \right) \quad (4.1)$$

donde  $P_{ref} = 20 \mu Pa$  [75] y  $x_i$  son los datos capturados de la señal acústica.

Por esta razón, el hilo de ejecución 2 además de capturar las ondas previamente emitidas, también realiza el preprocesamiento, es decir, la reducción de la cantidad de datos.

Una vez creados los hilos de ejecución se establece un flujo de trabajo, tal y como se muestra en la Figura 4.4. El hilo de ejecución 1 genera los valores digitales de las señales sinusoidales y envía estos valores digitales al dispositivo conversor digital-analógico. A continuación, el dispositivo conversor digital-analógico transforma las señales digitales en señales analógicas. Enseguida, un altavoz activo reproduce estas señales analógicas. Estas señales acústicas se propagan dentro del compartimento del evaporador. Esta propagación permite que estas señales acústicas interactúen con el evaporador y otros elementos del interior de la cavidad, como la escarcha acumulada.

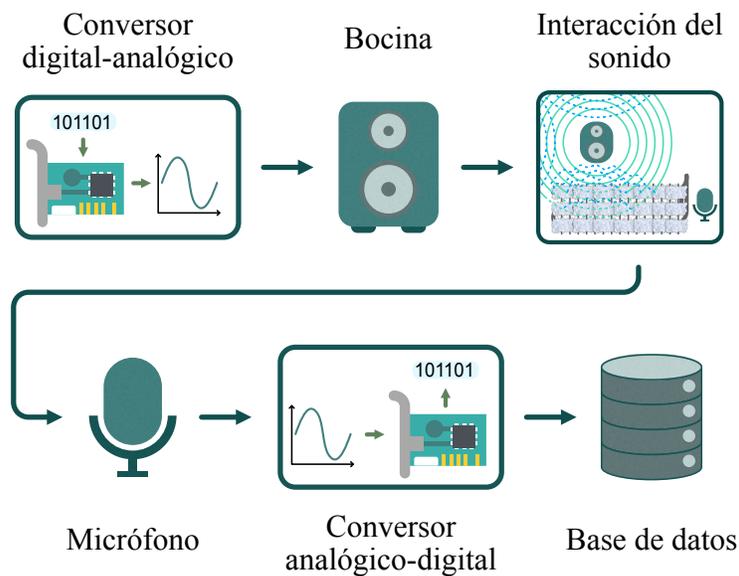


Figura 4.4: Diagrama del proceso de emisión/captura de señales acústicas.

Mientras la adquisición de datos está activa, hay tres hilos de ejecución corriendo simultáneamente. Como se ha indicado anteriormente, el hilo de ejecución 1 se encarga de generar y reproducir señales acústicas de diferentes frecuencias. Por otro lado, el hilo de ejecución 2 se utiliza para la captura de muestras analógicas procedentes de un micrófono situado en la cavidad del evaporador. Este hilo utiliza un dispositivo conversor analógico-digital para convertir las señales analógicas en señales digitales. Por lo tanto, este mismo hilo realiza el preprocesado de las señales digitales para obtener los valores SPL. Finalmente, estos valores SPL se almacenan para crear una base de datos.

### 4.3. Información visual

Una imagen puede definirse como una función en dos dimensiones  $f(x, y)$ , donde  $x$  e  $y$  son coordenadas espaciales, y la amplitud de  $f$  en cualquier par de coordenadas se denomina intensidad del nivel de gris en ese punto [76]. Juntando tres funciones  $f(x, y)$  se crea una imagen RGB, donde cada canal representa la intensidad del color. Una imagen también tiene un tamaño, en anchura y altura, que representa el total de píxeles, donde cada píxel representa las intensidades de color en cada coordenada  $x$  e  $y$ .

Una imagen comprende una colección de píxeles y canales que representan el mundo real en forma digital. La imagen contiene información que puede ser procesada para extraer características clave para identificar personas y patrones, entre otros [77]. Como se muestra en la Figura 4.5, una imagen puede capturar la información del mundo real; en este caso, representa fácilmente la cantidad de escarcha en la superficie de un evaporador doméstico. Esta representación de la cantidad de escarcha contiene las características necesarias para su posterior clasificación.

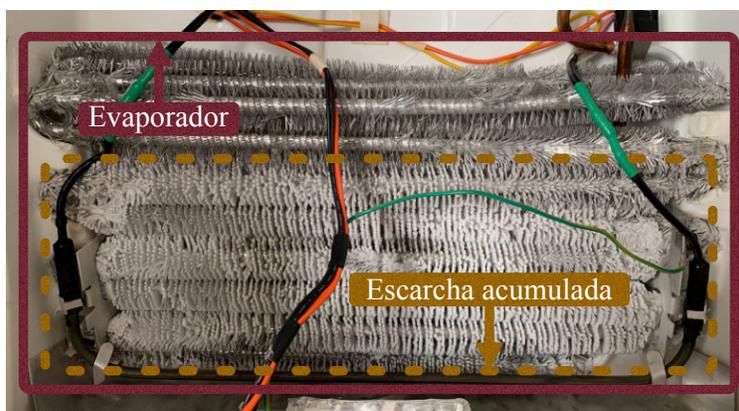


Figura 4.5: Evaporador de un refrigerador doméstico.

Un sistema de adquisición de datos permite obtener información precisa y almacenarla en equipos digitales. Cada sistema se adapta a unas necesidades concretas. Por lo tanto, se desarrolla un sistema de adquisición de datos basado en visión por computadora para lograr una clasificación de la acumulación de escarcha de alto nivel, como se muestra en la Figura 4.6.

El hilo de ejecución 3 en la Figura 4.3, controla el sistema de adquisición el cual se encarga de hacer funcionar una cámara RGB con un LED incorporado. Esta cámara captura imágenes con un tamaño de  $640 \times 480$  píxeles. A través de muchas pruebas experimentales se localizó el lugar con mayor acumulación de escarcha la superficie de los evaporadores. Por esta razón, la cámara se colocó en ese lugar en ambos refrigeradores para que se tomara una imagen cada minuto.

El sistema de adquisición de datos se utiliza para crear una base de datos de imágenes.

Las imágenes RGB contienen los datos necesarios para clasificar la cantidad de escarcha en cuatro niveles diferentes. Estas imágenes capturan una porción del evaporador donde se encuentra la mayor acumulación de escarcha en su superficie.

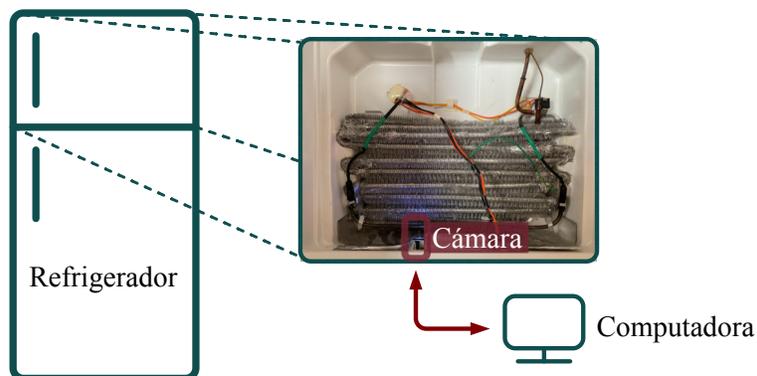


Figura 4.6: Sistema de adquisición de datos usando visión por computadora.

## 4.4. Banco experimental

Se utilizan dos refrigeradores diferentes tipo “no-frost” como bancos experimentales. Con ello se pretende comprobar la solidez de la metodología propuesta detectando la escarcha acumulada en varios diseños de evaporadores. La adquisición de datos y las pruebas experimentales se realizan con un ordenador personal. Todas las pruebas se realizaron en condiciones reales de hábitos de uso dentro de un hogar. La recolección de datos fue continua, lo que implica variaciones en los hábitos de uso y en la temperatura ambiente a lo largo del día, imitando condiciones de uso real.

### 4.4.1. Refrigerador doméstico A

La primera instalación experimental utiliza un refrigerador doméstico “no-frost”. Este refrigerador tiene un congelador en la parte superior y un compartimento de alimentos refrigerados en la parte inferior. La transferencia de calor de este refrigerador se mantiene principalmente por convección forzada a través de un ventilador situado delante del evaporador dentro del congelador. Las dimensiones externas de este refrigerador son de  $1.76m \times 0.65m \times 0.6m$  (*altura*  $\times$  *anchura*  $\times$  *profundidad*), produciendo una capacidad volumétrica de  $0.396 m^3$ . Las dimensiones del evaporador de tubos aletados son de  $0.15m \times 0.5m$  (*altura*  $\times$  *anchura*). Finalmente, este refrigerador funciona con el refrigerante R134a.

Se realizan dos pruebas experimentales utilizando el refrigerador doméstico A. La primera prueba experimental consiste en el procesamiento de señales acústicas. Para esto, se utiliza una disposición de un altavoz activo y un micrófono, como se muestra en la Figura 4.7.

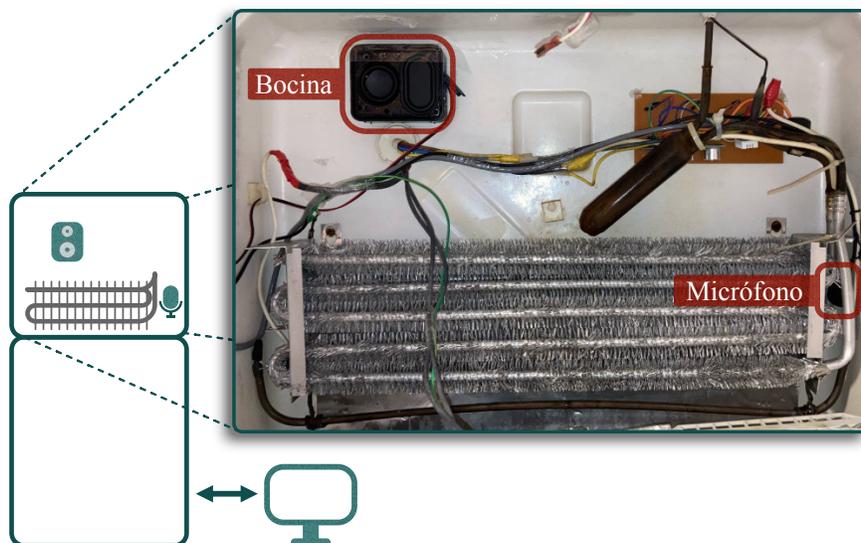


Figura 4.7: Arreglo de sensores para la prueba experimental que utiliza señales acústicas en el refrigerador A.

La Tabla 4.1 muestra las características del altavoz activo y el micrófono utilizados para las pruebas experimentales, como la respuesta en frecuencia, la potencia cuadrática media (RMS - *Root Mean Square*) y la relación señal/ruido (SNR - *Signal to Noise Ratio*).

Tabla 4.1: Características del altavoz activo y micrófono.

Sensor	Respuesta en frecuencia	RMS	SNR
Altavoz	180 $Hz$ – 20 $kHz$	3.1 $W$	80 $dB$
Micrófono	100 $Hz$ – 16 $kHz$	-	-

La segunda prueba experimental consiste en el procesamiento de imágenes para la clasificación del espesor de la escarcha en la superficie del evaporador, utilizando una cámara RGB. Esta cámara cuenta con una resolución de  $640 \times 480$  píxeles (72 píxeles por pulgada) y un LED incorporado como se muestra en la Figura 4.8.

#### 4.4.2. Refrigerador doméstico B

La segunda instalación experimental consiste en un refrigerador doméstico del tipo “no-frost”. Este refrigerador tiene la misma forma que el visto en “Refrigerador doméstico A”. Sin embargo, el tamaño y el diseño del evaporador difieren. La operación de transferencia de calor se mantiene principalmente por convección forzada a través de un ventilador en el compartimento del evaporador. Las dimensiones exteriores del refrigerador son de  $1.68m \times 0.57m \times 0.6m$  (*altura*  $\times$  *anchura*  $\times$  *profundidad*), produciendo una capacidad volumétrica

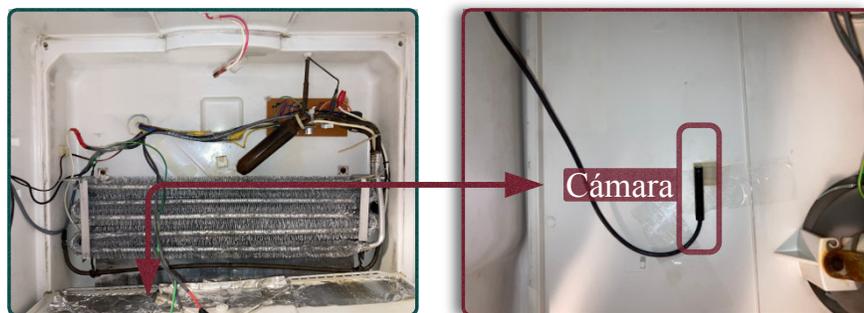


Figura 4.8: Arreglo de sensores para la prueba experimental que utiliza imágenes RGB en el refrigerador A.

de  $0.283 \text{ m}^3$ . Las dimensiones del evaporador de tubos con aletas son 0.5 m de ancho, 0.44 m en la parte inferior y 0.25 m de alto. Además, este refrigerador funciona con el refrigerante R134a.

Se utiliza una cámara RGB para la clasificación del espesor de la escarcha en la superficie del evaporador, como se muestra en la Figura 4.9.

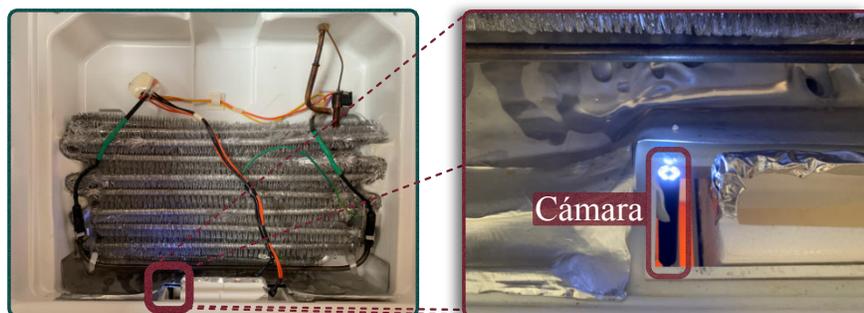


Figura 4.9: Arreglo de sensores para la prueba experimental en el refrigerador B.

#### 4.4.3. Detección de escarcha en la superficie del evaporador

Este caso de estudio propone el uso de diferentes técnicas de inteligencia artificial para la detección de la escarcha acumulada sobre la superficie de un evaporador doméstico. Se propone el uso de redes neuronales artificiales, específicamente el uso de ANN y PNN para el análisis de las señales acústicas. Además, se propone el uso de CNN para el procesamiento de las bases de datos de imágenes RGB.

Se ha diseñado y desarrollado una arquitectura de ANN utilizando el software Neural Lab [24]. La Figura 4.10 muestra la arquitectura de la red neuronal propuesta. Tiene diez entradas, una para cada valor de SPL en cada frecuencia. Una capa oculta con 27 neuronas. Finalmente, la red tiene una capa de salida con cuatro neuronas, una para cada nivel de

acumulación de escarcha. Dado que la red neuronal está configurada como un clasificador, cada salida ( $o_1, o_2, o_3, o_4$ ) corresponde a una de las cuatro clases (“Sin escarcha”, “Poca escarcha”, “Escarcha media” y “Escarcha alta”). Por último, esta red neuronal se entrenó para estimar la escarcha acumulada en el evaporador.

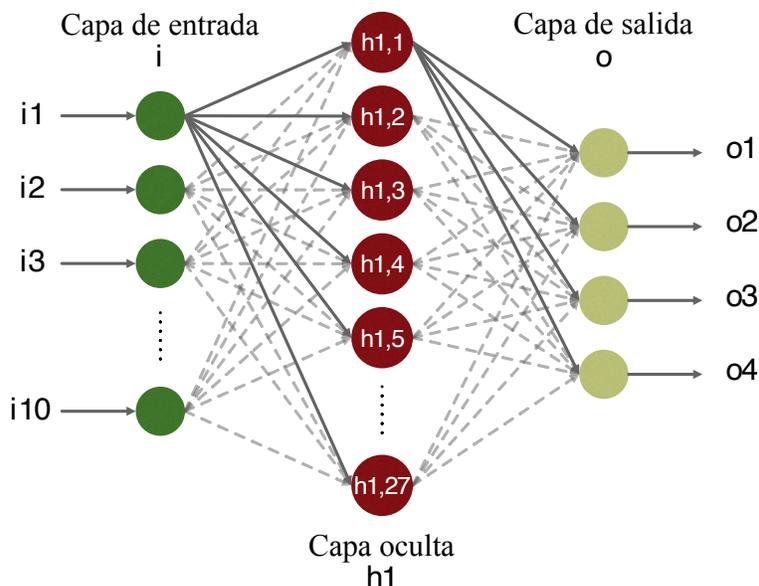


Figura 4.10: Diagrama de la ANN propuesta.

Las simulaciones por ordenador para la PNN también se realizaron utilizando el software Neural Lab aplicando la plantilla PNN estándar del software, véase [24].

Finalmente, se diseña e implementa un modelo inteligente basado en CNN como se muestra en la Figura 4.11 para el procesamiento de la base de datos de imágenes RGB. Se diseñó un modelo inteligente pequeño y fácil de entrenar sin sacrificar la clasificación, fácilmente adaptable a cualquier evaporador. Para determinar el tamaño correcto de las imágenes para el modelo inteligente se realizaron simulaciones con tamaños de  $320 \times 240$  píxeles,  $160 \times 120$  píxeles y  $80 \times 60$  píxeles. El tamaño tuvo en cuenta el mantener la relación aspecto de la imagen original. Las pruebas indicaron el mismo rendimiento para las tres resoluciones; por este motivo, se decidió elegir el tamaño más pequeño, consiguiendo un modelo que necesita menos recursos informáticos para su funcionamiento.

La CNN propuesta recibe una imagen de tamaño  $80 (W) \times 60 (H)$  píxeles. Tiene una capa convolucional en dos dimensiones con 64 filtros de  $3 \times 3$ . Luego se añade una capa de normalización, seguida de una capa de agrupación máxima en dos dimensiones con un tamaño de  $3 \times 3$ . La capa anterior tiene por objeto reducir la dimensionalidad de los datos. Después, se añade una capa de normalización junto con una capa totalmente conectada con 156 neuronas. Por último, se añade una capa totalmente conectada con cuatro neuronas encargadas de realizar la clasificación de la información de entrada. Se realizaron múltiples simulaciones computacionales para determinar el número óptimo de filtros y neuronas para reducir el tamaño del modelo sin sacrificar la clasificación.

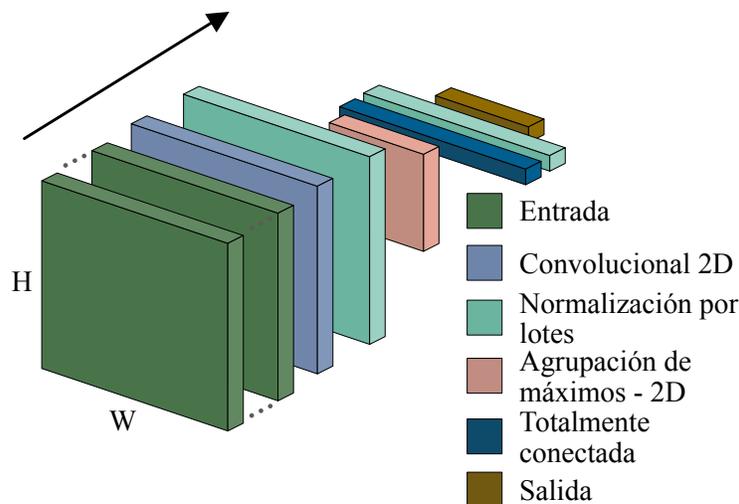


Figura 4.11: Diagrama de la red neuronal convolucional propuesta.

## 4.5. Resultados

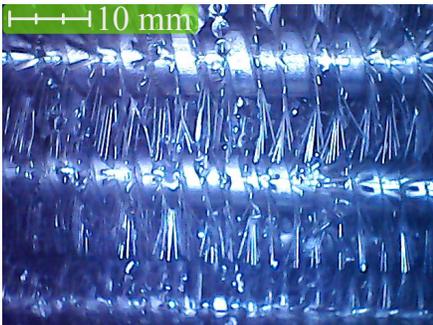
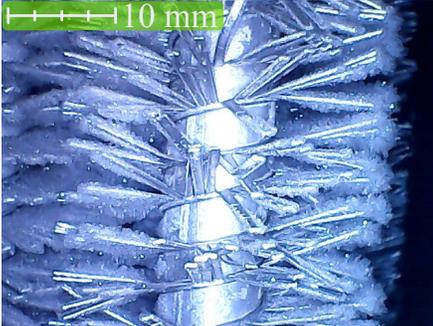
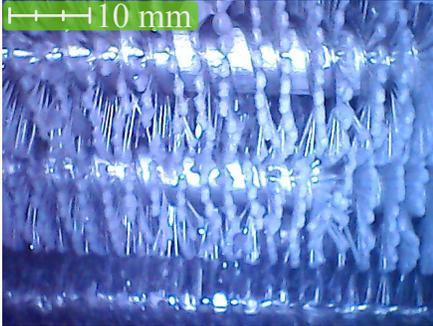
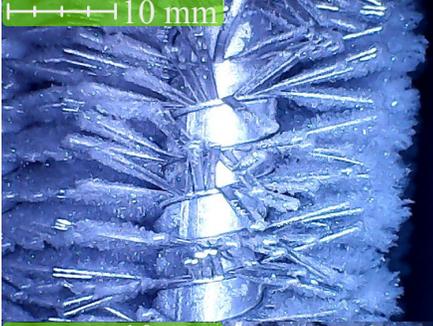
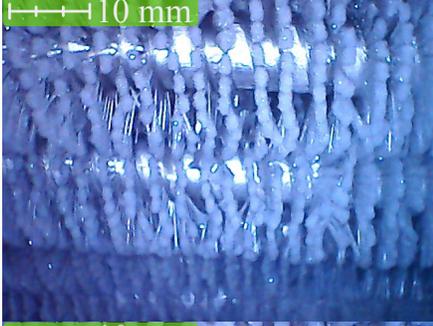
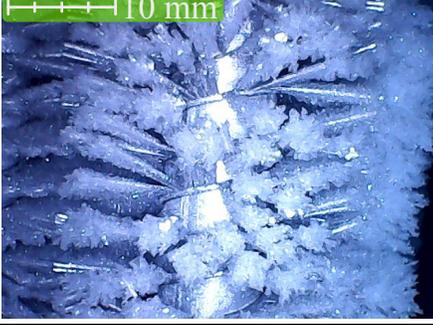
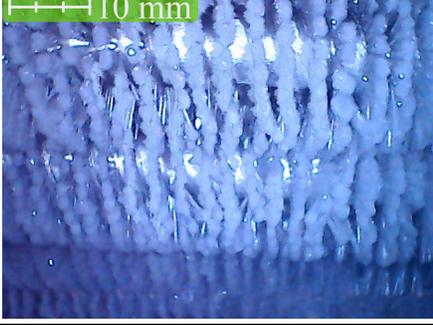
En esta sección se presentan los resultados de los entrenamientos de los modelos propuestos para validar su rendimiento. Previamente, es necesario decidir el número de clases de acumulación de escarcha en la superficie del evaporador. Se realizan varios experimentos utilizando los refrigeradores descritos anteriormente y un software de imagen como dispositivo de medición.

La Tabla 4.2 muestra las clases propuestas en este caso de estudio y una imagen descriptiva (que muestra sus características) para cada clase. Las clases resultantes son: “Sin escarcha”, “Poca escarcha”, “Escarcha media” y “Escarcha alta”. En la clase “Sin escarcha”, como se muestra en la tabla, el evaporador no contiene escarcha acumulada. En la clase “Poca escarcha”, el espesor de la escarcha oscila entre 2.5 mm y 5 mm. En la Tabla 4.2, la clase “Escarcha media” muestra las características del espesor de la escarcha entre 5 mm y 7.5 mm. Por último, la clase “Escarcha alta” se utiliza para la acumulación de escarcha superior a 7.5 mm.

Una vez definidas las cuatro clases para la acumulación de escarcha, utilizando el sistema de generación/adquisición de datos (ver 4.2, 4.3) se crean bases de datos con imágenes y valores SPL. Este sistema de generación/adquisición captura una imagen cada minuto durante 48 horas continuas, además de cada dos minutos la reproducción/captura de las señales acústicas, así como su preprocesamiento o reducción de datos usando SPL. Se realizaron varias pruebas de 48 horas continuas. La reducción de datos ofrece ventajas, como la inmunidad al ruido y a los cambios bruscos en los datos. Se realizaron varias pruebas para la captura de datos, incluyendo aquellas con hábitos de uso reales.

El entrenamiento de una red neuronal proporciona una técnica para extraer características

Tabla 4.2: Niveles de escarcha acumulada en la superficie de un evaporador doméstico.

Nivel de escarcha	Refrigerador A	Refrigerador B
“Sin escarcha” escarcha < 2.5 mm		
“Poca escarcha” 2.5 mm < escarcha < 5 mm		
“Escarcha media” 5 mm < escarcha < 7.5 mm		
“Escarcha alta” escarcha > 7.5 mm		

esenciales de los datos. Estas características crean una relación entre la entrada y la salida de la red neuronal a través de una o varias funciones de activación. Una red neuronal utiliza dos conjuntos de datos, uno de entrenamiento y otro de validación. El conjunto de datos de entrenamiento se utiliza para estimar los valores de peso que conectan las neuronas de la red

neuronal. Así, el conjunto de entrenamiento proporciona muestras para ilustrar la relación entre la entrada y la salida de la red neuronal. Por último, el conjunto de validación se utiliza para medir el rendimiento de la red con un conjunto de datos que no se ha visto antes. Antes del entrenamiento de la red, los datos capturados por el sistema de adquisición se dividieron aleatoriamente. El conjunto de entrenamiento consta de datos conformados por imágenes RGB y señales acústicas. El 80 % de los datos capturados se utilizó para crear el conjunto de entrenamiento y el 20 % restante se utilizó para crear el conjunto de validación.

Una matriz de confusión es una de las mejores maneras de verificar la eficacia de un modelo inteligente [78], específicamente una red neuronal artificial. La matriz de confusión representa una comparación entre los valores reales de las clases frente a los valores predichos por el clasificador [79]. La matriz de confusión es una matriz de identidad cuando un modelo inteligente es perfecto. Así, la diagonal principal de la matriz de confusión tiende a ser 1, y los demás valores tienden a ser 0. Esto significa que los valores predichos por el modelo son precisamente equivalentes a la clase real de los datos analizados. Por último, una matriz de confusión o matriz de error es el punto de partida para obtener datos estadísticos analíticos como la exactitud global, la precisión y la sensibilidad, entre otros [78].

La Figura 4.12 muestra la matriz de confusión obtenida durante la validación para la clasificación de la acumulación de escarcha en el evaporador utilizando el modelo ANN y usando la base de datos de señales acústicas. Como se puede observar en la Figura 4.12, se obtiene una exactitud global del 94 % para los cuatro niveles diferentes de escarcha.

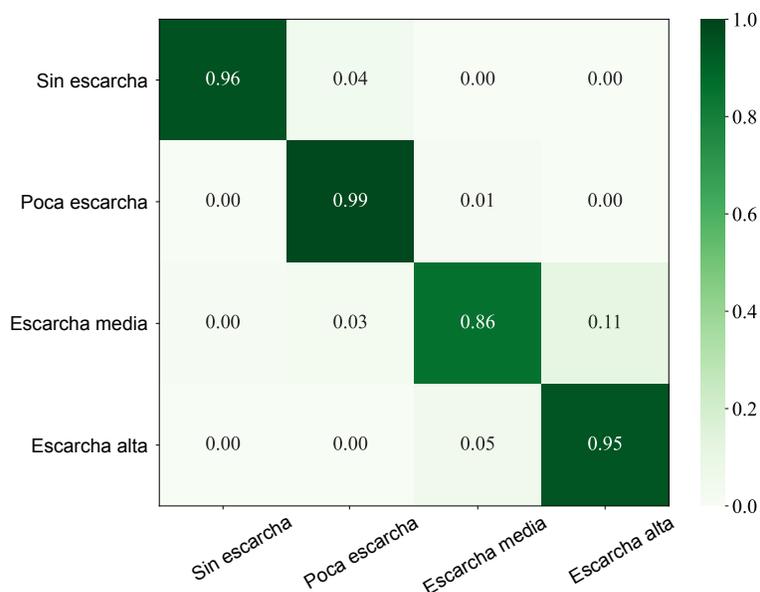


Figura 4.12: Matriz de confusión del modelo ANN en el refrigerador doméstico A, utilizando la base de datos de señales acústicas.

La Figura 4.13 muestra la matriz de confusión resultante obtenida durante la validación para la clasificación de la acumulación de escarcha en el evaporador utilizando el modelo

PNN y usando la base de datos de señales acústicas; se obtiene una exactitud global del 94 % para los cuatro niveles diferentes de acumulación de escarcha.

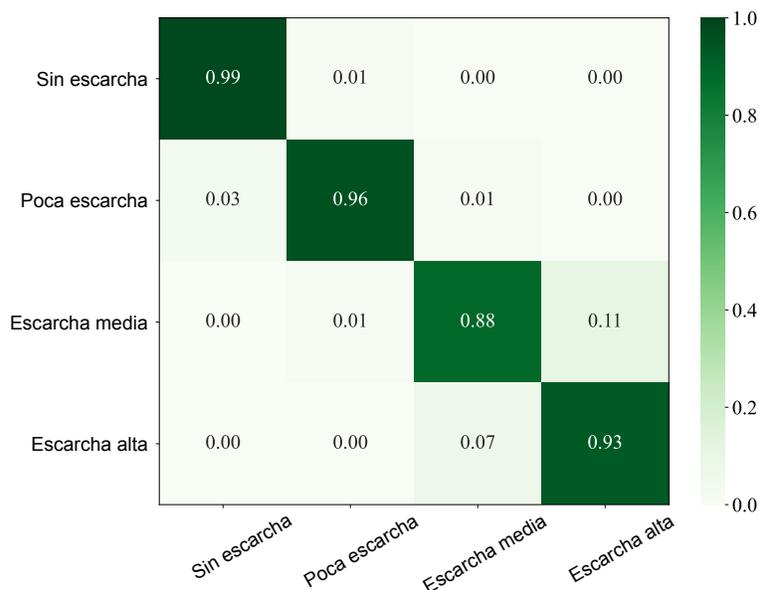


Figura 4.13: Matriz de confusión del modelo PNN en el refrigerador doméstico A, utilizando la base de datos de señales acústicas.

La exactitud global de ambos modelos inteligentes proporcionan una buena clasificación de la escarcha acumulada en la superficie del evaporador de un refrigerador doméstico.

En la Tabla 4.3 se muestran las métricas como precisión, sensibilidad, F-score, exactitud global, para ambas configuraciones experimentales (refrigerador doméstico A y B) usando la base de datos de imágenes RGB. Estas métricas indican el comportamiento del modelo inteligente a la hora de clasificar la cantidad de escarcha acumulada. Obsérvese que el modelo muestra una tasa mínima de falsos positivos y falsos negativos entre las clases.

Tabla 4.3: Métricas de los modelos CNN para los refrigeradores A y B usando la base de datos de imágenes RGB.

Refrigerador	Precisión (%)	Sensibilidad (%)	F-score (%)	Exactitud global (%)
Refrigerador A	99.75	99.75	99.74	99.75
Refrigerador B	99.75	99.75	99.74	99.75

La Figura 4.14 muestra la matriz de confusión obtenida durante la validación para el refrigerador doméstico A, utilizando el modelo CNN y la base de datos de imágenes RGB. Como se observa en su diagonal principal, se obtiene una exactitud global del 99.75 % para cuatro niveles diferentes de escarcha acumulada.

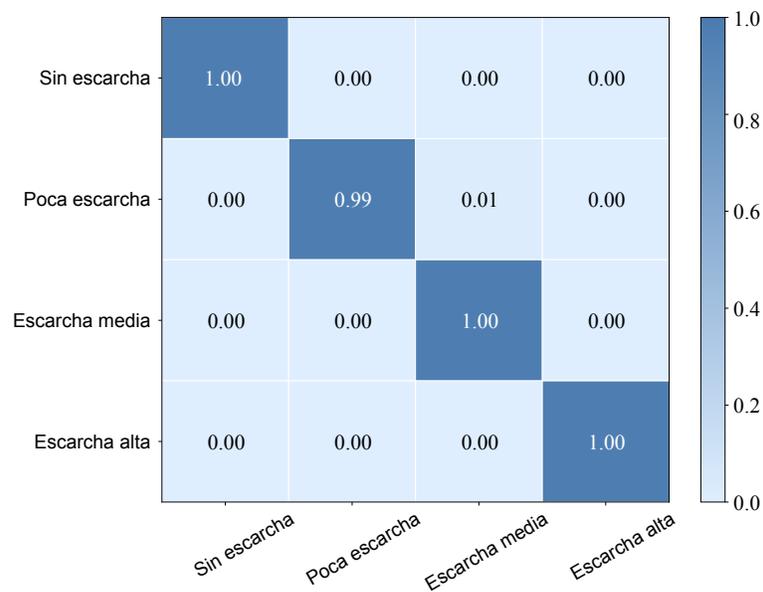


Figura 4.14: Matriz de confusión para el modelo CNN utilizando el refrigerador doméstico A y la base de datos de imágenes RGB.

Del mismo modo, en la Figura 4.15 se observa la diagonal principal de la matriz de confusión de validación utilizando el refrigerador doméstico B, usando de igual forma el modelo CNN y la base de datos de imágenes RGB. Se obtiene una exactitud global del 99.75 %.

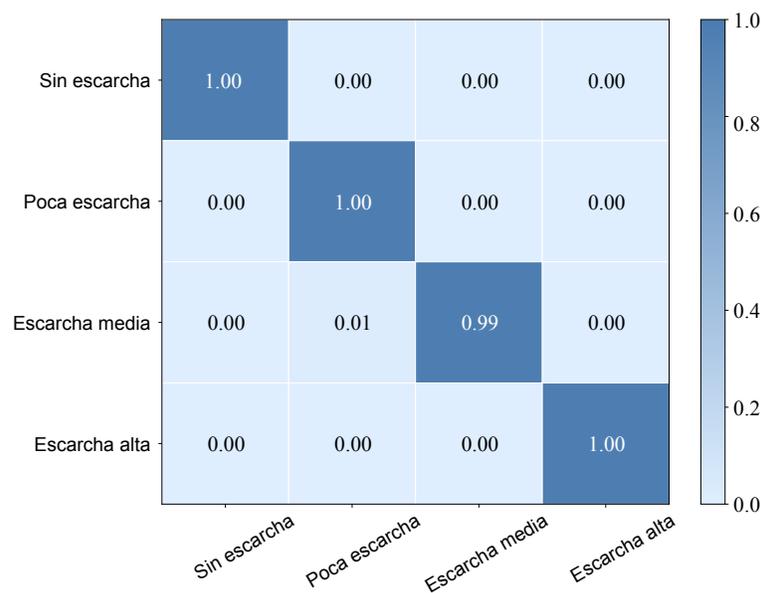


Figura 4.15: Matriz de confusión para el modelo CNN utilizando el refrigerador doméstico B y la base de datos de imágenes RGB.

El rendimiento de la exactitud global del modelo propuesto en ambas configuraciones de

refrigerador para la base de datos de imágenes RBB, muestra una clasificación adecuada de la escarcha acumulada en la superficie del evaporador de un refrigerador doméstico. Utilizando dos diseños con diferentes modelos de evaporador, se demuestra la robustez del modelo CNN propuesto.

## 4.6. Discusión

En este caso de estudio se propone una metodología novedosa para clasificar la acumulación de escarcha sobre la superficie de un evaporador en un refrigerador doméstico. Esta clasificación se obtiene utilizando inteligencia artificial y un conjunto de datos conformado por señales acústicas, junto a datos de imágenes RGB. El objetivo principal es desarrollar un sistema de bajo costo capaz de evaluar la cantidad de escarcha en la superficie del evaporador en un refrigerador doméstico. Otro de los principales objetivos del diseño de los modelos inteligentes, es su adaptación a los hábitos de uso, así como su adaptabilidad a las condiciones ambientales de su ubicación.

En primer lugar, el sistema consta de un sistema de generación/adquisición de datos. Este sistema utiliza hilos de ejecución para manejar simultáneamente un altavoz activo, un micrófono ordinario y una cámara RGB con LED incorporado. El hilo de ejecución del altavoz activo se encarga inicialmente de crear un conjunto de diez ondas sinusoidales. Una vez creado el conjunto de ondas, este hilo de ejecución es encargado de reproducir las señales sinusoidales cada 120 segundos. Esta propagación permite a las ondas sonoras interactuar con su entorno; esta interacción favorece que su energía cambie, es decir, que pueda ser absorbida y/o reflejada. Por lo tanto, el hilo de ejecución del micrófono capta las ondas acústicas que interactúan con su entorno de forma simultánea al otro hilo de ejecución del altavoz, el cual reproduce las ondas. Una vez finalizada la captura de las señales acústicas, el sistema calcula el nivel de presión sonora. Este preprocesamiento permitió reducir la dimensionalidad de los datos. Finalmente, el tercer hilo de ejecución captura y almacena una imagen RGB cada 60 segundos. Se crearon dos conjuntos de datos con los valores de SPL y dos conjuntos de datos con imágenes RGB. Estos conjuntos de datos se han utilizado para el entrenamiento y la validación de los modelos inteligentes.

Se crearon tres modelos inteligentes. Estos se basaron en redes neuronales artificiales, redes neuronales probabilísticas y redes neuronales convolucionales. Uno de los principales objetivos del diseño de estos modelos inteligentes es reducir el consumo de recursos computacionales, de este modo, se obtuvieron modelos de tamaño reducido. Estos pudieron ser entrenados rápidamente y por lo tanto, estimaron instantáneamente la cantidad de escarcha acumulada sin sacrificar la precisión; se obtuvo una exactitud global del 94 % en los modelos inteligentes basados en ANN y PNN. Se obtuvo una exactitud global del 99.75 % para el modelo basado en CNN en ambos refrigeradores experimentales.

Finalmente, se puede concluir que los modelos propuestos pueden ser utilizados en lugares

donde exista una restricción de recursos computacionales y, al mismo tiempo, tener un alto rendimiento en la estimación de la escarcha acumulada. Los trabajos futuros incluyen la provisión de un prototipo funcional embebido de nuestro enfoque que pueda ser fácilmente añadido a los refrigeradores domésticos. Además de un posible control del retiro de escarcha con una retroalimentación a partir de este sistema de generación/adquisición de datos, con el fin de reducir el consumo energético global del refrigerador doméstico.

# Capítulo 5

## Conclusiones

En el primer caso de estudio se propuso una red neuronal convolucional temporal como una solución eficiente para analizar y clasificar actividades humanas, con un énfasis particular en la interacción entre objetos en un video de corta duración. La arquitectura desarrollada permite procesar simultáneamente varios fotogramas, destacando la capacidad para identificar cambios naturales en el movimiento y tamaño de objetos en una escena. La principal ventaja radica en su capacidad para detectar y medir la diferencia y similitud entre objetos en fotogramas secuenciales. Además, se destaca la importancia de reducir el consumo de recursos computacionales mediante la creación de una red de tamaño reducido, lo que facilita su entrenamiento rápido en comparación con otras redes avanzadas. Este caso de estudio permite obtener un modelo inteligente el cual se desempeña con una mejor precisión que los modelos existentes del estado del arte.

A partir del caso de estudio “Reconocimiento de actividades” se obtuvo la siguiente publicación:

- *Yair A. Andrade-Ambroz, Sergio Ledesma, Mario-Alberto Ibarra-Manzano, Marvella I. Oros-Flores, Dora-Luz Almanza-Ojeda, Human activity recognition using temporal convolutional neural network architecture, Expert Systems with Applications, Volume 191, 2022, 116287, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2021.116287>, IF: 8.5 [80].*

En el segundo caso, se presentó una metodología novedosa que utiliza inteligencia artificial y señales acústicas para estimar la acumulación de escarcha en evaporadores de refrigeradores domésticos. El sistema propuesto incorpora un enfoque de bajo costo al utilizar un conjunto de señales acústicas y datos de imágenes RGB. El diseño incluye un sistema de generación/adquisición de datos que emplea hilos de ejecución para un altavoz activo, un micrófono ordinario y una cámara RGB. Se crearon tres modelos inteligentes basados en diferentes tipos de redes neuronales, todos diseñados con el objetivo de reducir el consumo

de recursos computacionales y lograr una rápida capacidad de entrenamiento y clasificación. Los resultados experimentales demuestran la efectividad de los modelos, con exactitudes globales del 94 % y 99.75 % para modelos basados en ANN/PNN y CNN, respectivamente.

Con los resultados obtenidos del caso de estudio “Refrigeración doméstica” se lograron las siguientes publicaciones:

- *Yair A. Andrade-Ambriz, Sergio Ledesma, J.M. Belman-Flores, I. Carvajal-Mariscal, Dora-Luz Almanza-Ojeda, Frost thickness estimation in a domestic refrigerator using acoustic signals and artificial intelligence, Expert Systems with Applications, Volume 201, 2022, 117071, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2022.117071>, IF: 8.5 [81].*
- *Yair A. Andrade-Ambriz, Sergio Ledesma, Dora-Luz Almanza-Ojeda, J.M. Belman-Flores, Accurate classification of frost thickness using visual information in a domestic refrigerator, International Journal of Refrigeration, Volume 145, 2023, Pages 256-263, ISSN 0140-7007, <https://doi.org/10.1016/j.ijrefrig.2022.08.019>, IF: 3.9 [82].*

En conjunto, ambas propuestas comparten la preocupación por la eficiencia en el uso de recursos computacionales, logrando modelos y sistemas de bajo costo que ofrecen resultados precisos en sus respectivos contextos de aplicación: el análisis de actividades humanas y la estimación de escarcha en refrigeradores domésticos. Ambos trabajos demuestran la viabilidad de soluciones tecnológicas novedosas con aplicaciones prácticas en entornos con restricciones de recursos, además de ofrecer perspectivas a futuro con la creciente preocupación en la reducción del consumo energético, así como una mejora continua en la calidad de vida.

# Índice de figuras

2.1. Principales ramas de la inteligencia artificial [19]. . . . .	7
2.2. Principales técnicas del aprendizaje máquina [22]. . . . .	8
2.3. Estructura interna de una neurona. . . . .	9
2.4. Estructura general de una red neuronal artificial. . . . .	11
2.6. Diagrama general de una red neuronal convolucional. . . . .	11
2.5. Diagrama general de una red neuronal probabilística [26]. . . . .	12
2.7. Primeros pasos de la aplicación de una matriz convolucional. . . . .	13
3.1. Arquitectura propuesta para la red neuronal convolucional temporal. . . . .	19
3.2. Secuencia con cuatro fotogramas de la actividad “ondulación horizontal del brazo” del conjunto de datos KARD [46]. . . . .	20
3.3. Representación de una convolución 3D. . . . .	21
3.4. Representación de una capa ConvLSTM, con sus entradas, celdas de salida y estados ocultos [61]. . . . .	22
3.5. Fotogramas de dos actividades diferentes del conjunto de datos KARD [46]. . . . .	23
3.6. Fotogramas de diferentes actividades del conjunto de datos CAD-60 [47]. . . . .	24
3.7. Fotogramas de diferentes actividades del conjunto de datos <i>MSR Daily Activity 3D</i> [48]. . . . .	24
3.8. Matriz de confusión para el conjunto de datos KARD [46]. . . . .	26

3.9. Matriz de confusión para el conjunto de datos CAD-60 [47]. . . . .	27
3.10. Matriz de confusión para el conjunto de datos <i>MSR Daily Activity 3D</i> [48]. .	28
4.1. Evaporador típico de un refrigerador doméstico. . . . .	32
4.2. Propagación de señales acústicas en el interior del compartimento del evaporador. . . . .	34
4.3. Esquema general del sistema de generación/adquisición de datos. . . . .	35
4.4. Diagrama del proceso de emisión/captura de señales acústicas. . . . .	36
4.5. Evaporador de un refrigerador doméstico. . . . .	37
4.6. Sistema de adquisición de datos usando visión por computadora. . . . .	38
4.7. Arreglo de sensores para la prueba experimental que utiliza señales acústicas en el refrigerador A. . . . .	39
4.8. Arreglo de sensores para la prueba experimental que utiliza imágenes RGB en el refrigerador A. . . . .	40
4.9. Arreglo de sensores para la prueba experimental en el refrigerador B. . . . .	40
4.10. Diagrama de la ANN propuesta. . . . .	41
4.11. Diagrama de la red neuronal convolucional propuesta. . . . .	42
4.12. Matriz de confusión del modelo ANN en el refrigerador doméstico A, utilizando la base de datos de señales acústicas. . . . .	44
4.13. Matriz de confusión del modelo PNN en el refrigerador doméstico A, utilizando la base de datos de señales acústicas. . . . .	45
4.14. Matriz de confusión para el modelo CNN utilizando el refrigerador doméstico A y la base de datos de imágenes RGB. . . . .	46
4.15. Matriz de confusión para el modelo CNN utilizando el refrigerador doméstico B y la base de datos de imágenes RGB. . . . .	46

# Índice de tablas

3.1. Valores de precisión y sensibilidad del estado del arte (%) en el conjunto de datos KARD . . . . .	26
3.2. Valores de precisión y sensibilidad del estado del arte (%) en el conjunto de datos CAD-60. . . . .	27
3.3. Valores de exactitud del estado del arte (%) en el conjunto de datos MSR Daily Activity 3D. . . . .	27
3.4. Tiempo de entrenamiento y clasificación de hasta 2 segundos de video simultáneo. . . . .	28
4.1. Características del altavoz activo y micrófono. . . . .	39
4.2. Niveles de escarcha acumulada en la superficie de un evaporador doméstico. .	43
4.3. Métricas de los modelos CNN para los refrigeradores A y B usando la base de datos de imágenes RGB. . . . .	45

# Bibliografía

- [1] Dupont J. L., Domanski P., Lebrun P., and Ziegler F. 38th note on refrigeration technologies: The role of refrigeration in the global economy (2019), 2019.
- [2] Juan M. Belman-Flores, Diana Pardo-Cely, Francisco Elizalde-Blancas, Armando Gallegos-Muñoz, Vicente Pérez-García, and Miguel A. Gómez-Martínez. Perspectives on consumer habits with domestic refrigerators and its consequences for energy consumption: Case of study in guanajuato, mexico. *Energies*, 12(5), 2019.
- [3] Ing-Jr Ding and Jia-Yi Shi. Kinect microphone array-based speech and speaker recognition for the exhibition control of humanoid robots. *Computers & Electrical Engineering*, 62:719–729, 2017.
- [4] Ing-Jr Ding and Che-Wei Chang. An eigenspace-based method with a user adaptation scheme for human gesture recognition by using kinect 3d data. *Applied Mathematical Modelling*, 39(19):5769–5777, 2015.
- [5] Manuel Gil-Martín, Rubén San-Segundo, Fernando Fernández-Martínez, and Javier Ferreiros-López. Improving physical activity recognition using a new deep learning architecture and post-processing techniques. *Engineering Applications of Artificial Intelligence*, 92:103679, 2020.
- [6] I. K. Ihianle, A. O. Nwajana, S. H. Ebenuwa, R. I. Otuka, K. Owa, and M. O. Orisatoki. A deep learning approach for human activities recognition from multimodal sensing devices. *IEEE Access*, 8:179028–179038, 2020.
- [7] C. Caetano, F. Brémond, and W. R. Schwartz. Skeleton image representation for 3d action recognition based on tree structure and reference joints. In *2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 16–23, Oct 2019.
- [8] Sameh Neili, Sami Gazzah, Mounim A. El Yacoubi, and Najoua Essoukri Ben Amara. Human posture recognition approach based on convnets and svm classifier. In *2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pages 1–6, 2017.
- [9] J.M. Belman-Flores, J.M. Barroso-Maldonado, A.P. Rodríguez-Muñoz, and G. Camacho-Vázquez. Enhancements in domestic refrigeration, approaching a

- sustainable refrigerator – a review. *Renewable and Sustainable Energy Reviews*, 51:955–968, 2015.
- [10] Ehab Mina and Cesar Gutierrez. Heater location through radiative heat optimization in a finned tube evaporator, 2004.
- [11] Claudio Melo, Fernando Testoni Knabben, and Paula do Valle Pereira. An experimental study on defrosted heaters applied to household refrigerators, 2012.
- [12] Lloyd Harrington, Lu Aye, and Bob Fuller. Energy impacts of defrosting in household refrigerators: Lessons from field and laboratory measurements. *International Journal of Refrigeration*, 86:480–494, 2018.
- [13] Zhiqiang Li, Dan Zhao, Guoliang Ding, Tao Ren, Songtao Miao, Xiaolei Han, and Toshinori Noda. Improving defrosting performance by controlling frost distribution to match defrosting heat distribution in frost-free household refrigerators. *International Journal of Refrigeration*, 77:136–148, 2017.
- [14] Patrick Henry Winston. *Artificial Intelligence*. Addison-Wesley, 1992.
- [15] Daochen Zha, Zaid Pervaiz Bhat, Kwei-Herng Lai, Fan Yang, Zhimeng Jiang, Shaochen Zhong, and Xia Hu. Data-centric artificial intelligence: A survey, 2023.
- [16] Pragnan Chakravorty. What is a signal? [lecture notes]. *IEEE Signal Processing Magazine*, 35(5):175–177, 2018.
- [17] John G. Proakis and Dimitris G. Manolakis. *Tratamiento digital de señales*. Prentice Hall, 2007.
- [18] Christian Janiesch, Patrick Zschech, and Kai Heinrich. Machine learning and deep learning. *Electronic Markets*, 31(3):685–695, September 2021.
- [19] Felix Weber. *Artificial Intelligence*, pages 33–64. Springer Fachmedien Wiesbaden, Wiesbaden, 2023.
- [20] Pramila P. Shinde and Seema Shah. A review of machine learning and deep learning applications. In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, pages 1–6, 2018.
- [21] Sergio Ledesma, Mario-Alberto Ibarra-Manzano, Dora-Luz Almanza-Ojeda, Juan Gabriel Avina-Cervantes, and Eduardo Cabal-Yepez. On removing conflicts for machine learning. *Expert Systems with Applications*, 206:117835, 2022.
- [22] Raffaele Pugliese, Stefano Regondi, and Riccardo Marini. Machine learning-based approach: global trends, research directions, and regulatory standpoints. *Data Science and Management*, 4:19–29, 2021.
- [23] Issam El Naqa and Martin J. Murphy. *What Is Machine Learning?*, pages 3–11. Springer International Publishing, Cham, 2015.

- [24] Sergio Ledesma, Mario-Alberto Ibarra-Manzano, Ma-Guadalupe Garcia-Hernandez, and Dora-Luz Almanza-Ojeda. Neural lab a simulator for artificial neural networks. In *2017 Computing Conference*, pages 716–721, 2017.
- [25] Mark W. Craven and Jude W. Shavlik. Learning symbolic rules using artificial neural networks. In *International Conference on Machine Learning*, 1993.
- [26] Donald F. Specht. Probabilistic neural networks. *Neural Networks*, 3(1):109–118, 1990.
- [27] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, 9(4):611–629, Aug 2018.
- [28] Ashkan Tousimojarad, Wim Vanderbauwhede, and W. Paul Cockshott. 2d image convolution using three parallel programming models on the xeon phi, nov 2017.
- [29] P. Celard, E. L. Iglesias, J. M. Sorribes-Fdez, R. Romero, A. Seara Vieira, and L. Borraro. A survey on deep learning applied to medical images: from simple artificial neural networks to generative models. *Neural Computing and Applications*, 35(3):2291–2323, Jan 2023.
- [30] Sravan Kumar Challa, Akhilesh Kumar, Vijay Bhaskar Semwal, and Nidhi Dua. An optimized deep learning model for human activity recognition using inertial measurement units. *Expert Systems*, 40(10):e13457, 2023.
- [31] Hao Wu, Zhichao Zhang, Xiaoyong Li, Kai Shang, Yongming Han, Zhiqiang Geng, and Tingrui Pan. A novel pedal musculoskeletal response based on differential spatio-temporal lstm for human activity recognition. *Knowledge-Based Systems*, 261:110187, 2023.
- [32] Guoqiang Liu, Tong Xiong, Tengfei Sun, Guixiang He, and Gang Yan. Frosting and defrosting characteristics of household refrigerators and freezers: Recent progress and perspectives. *Energy and Buildings*, 303:113755, 2024.
- [33] M.A. Moradkhani, Runmiao GAO, S.H. Hosseini, and Mengjie SONG. Estimation of frost layer characteristics during natural convection on vertical and inverted cold surfaces based on smart approaches. *Applied Thermal Engineering*, 239:122090, 2024.
- [34] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

- [35] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library, 2019.
- [36] Djamila Romaiassa Beddiar, Brahim Nini, Mohammad Sabokrou, and Abdenour Hadid. Vision-based human activity recognition: a survey. *Multimedia Tools and Applications*, 79(41):30509–30555, Nov 2020.
- [37] Gulshan Saleem, Usama Ijaz Bajwa, and Rana Hammad Raza. Toward human activity recognition: a survey. *Neural Computing and Applications*, 35(5):4145–4182, Feb 2023.
- [38] Xiaoyu Zhu, Po-Yao Huang, Junwei Liang, Celso M. de Melo, and Alexander G. Hauptmann. Stmt: A spatial-temporal mesh transformer for mocap-based action recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1526–1536, June 2023.
- [39] John F. Stumpf. Motion capture system, December 2 2010. US Patent 20100304931A1.
- [40] Joana Dias, Pedro Simões, Nuno Soares, Carlos M. Costa, Marcelo R. Petry, Germano Veiga, and Luís F. Rocha. Comparison of 3d sensors for automating bolt-tightening operations in the automotive industry. *Sensors*, 23(9), 2023.
- [41] Daniel Maier, Armin Hornung, and Maren Bennewitz. Real-time navigation in 3d environments based on depth camera data. In *2012 12th IEEE-RAS International Conference on Humanoid Robots*, pages 692–697, 2012.
- [42] N. Kawarazaki, D. Stefanov, and A. I. B. Diaz. Toward gesture controlled wheelchair: A proof of concept study. In *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*, pages 1–6, 2013.
- [43] Emily-Jane Rolley-Parnell, Dimitrios Kanoulas, Arturo Laurenzi, Brian Delhaisse, Leonel Rozo, Darwin G. Caldwell, and Nikos G. Tsagarakis. Bi-manual articulated robot teleoperation using an external rgb-d range sensor. In *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pages 298–304, 2018.
- [44] Markus Bajones, David Fischinger, Astrid Weiss, Daniel Wolf, Markus Vincze, Paloma de la Puente, Tobias Koertner, Markus Weninger, Konstantinos Papoutsakis, Damien Michel, Ammar Qammaz, Pashalis Paderleris, Michalis Foukarakis, Ilia Adami, Danai Ioannidi, Asterios Leonidis, Margherita Antona, Antonis Argyros, Peter Mayer, and Susanne Frennert. Hobbit: Providing fall detection and prevention for the elderly in the real world. *Journal of Robotics*, 2018:1–20, 06 2018.
- [45] Shaohua Wan, Lianyong Qi, Xiaolong Xu, Chao Tong, and Zonghua Gu. Deep learning models for real-time human activity recognition with smartphones. *Mobile Networks and Applications*, 25:743–755, 2019.

- [46] S. Gaglio, G. Lo. Re, and M. Morana. Human activity recognition process using 3-d posture data. *IEEE Transactions on Human-Machine Systems*, 45(5):586–597, 2015.
- [47] J. Soung, C. Ponce, B. Selman, and A. Saxena. Unstructured human activity detection from rgbd images. In *InRobotics and Automation (ICRA)*, pages 842–849, 05 2012.
- [48] J. Wang, Z. Liu, Y. Wu, and J. Yuan. Mining actionlet ensemble for action recognition with depth cameras. In *IEEE Conferenfe on Computer Vision and Pattern Recognition*, pages 1290–1297, 2012.
- [49] Martha Magali Flores-Barranco, Mario-Alberto Ibarra-Mazano, and Irene Cheng. Accidental fall detection based on skeleton joint correlation and activity boundary. In George Bebis, Richard Boyle, Bahram Parvin, Darko Koracin, Ioannis Pavlidis, Rogerio Feris, Tim McGraw, Mark Elendt, Regis Kopper, Eric Ragan, Zhao Ye, and Gunther Weber, editors, *Advances in Visual Computing*, pages 489–498, Cham, 2015. Springer International Publishing.
- [50] Qiming Li, Wenxiong Lin, and Jun Li. Human activity recognition using dynamic representation and matching of skeleton feature sequences from rgb-d images. *Signal Processing: Image Communication*, 68:265–272, 2018.
- [51] Diego R. Faria, Cristiano Premebida, and Urbano Nunes. A probabilistic approach for human everyday activities recognition using body motion from rgb-d images. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 732–737, 2014.
- [52] Tae Soo Kim and Austin Reiter. Interpretable 3d human action analysis with temporal convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1623–1631, 2017.
- [53] Xiaodong Yang and YingLi Tian. Effective 3d action recognition using eigenjoints. *Journal of Visual Communication and Image Representation*, 25(1):2–11, 2014. Visual Understanding and Applications with RGB-D Cameras.
- [54] Jörg Krüger and The Duy Nguyen. Automated vision-based live ergonomics analysis in assembly operations. *CIRP Annals*, 64(1):9–12, 2015.
- [55] Thien Huynh-The, Cam-Hao Hua, and Dong-Seong Kim. Learning action images using deep convolutional neural networks for 3d action recognition. In *2019 IEEE Sensors Applications Symposium*, pages 1–6, 2019.
- [56] A. Tomas and K. K. Biswas. Human activity recognition using combined deep architectures. In *2017 IEEE 2nd International Conference on Signal and Image Processing*, pages 41–45, Aug 2017.
- [57] Yujia Zhang, Lai Man Po, Mengyang Liu, Yasar Abbas Ur Rehman, Weifeng Ou, and Yuzhi Zhao. Data-level information enhancement: Motion-patch-based siamese convolutional neural networks for human activity recognition in videos. *Expert Systems with Applications*, 147:113203, 2020.

- [58] Amitesh Singh Rajput, Balasubramanian Raman, and Javed Imran. Privacy-preserving human action recognition as a remote cloud service using rgb-d sensors and deep cnn. *Expert Systems with Applications*, 152:113349, 2020.
- [59] Avinandan Banerjee, Pawan Kumar Singh, and Ram Sarkar. Fuzzy integral-based cnn classifier fusion for 3d skeleton action recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(6):2206–2216, 2021.
- [60] David Ada Adama, Ahmad Lotfi, and Robert Ranson. Adaptive segmentation and sequence learning of human activities from skeleton data. *Expert Systems with Applications*, 164:113836, 2021.
- [61] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai kin Wong, and Wang chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting, 2015.
- [62] Enea Cippitelli, Samuele Gasparrini, E. Gambi, and Susanna Spinsante. A human activity recognition system using skeleton data from rgb-d sensors. *Computational Intelligence and Neuroscience*, 2016:1–14, 03 2016.
- [63] Syed Hammad Hussain Shah, Anniken Susanne T. Karlsen, Mads Solberg, and Ibrahim A. Hameed. An efficient and lightweight multiperson activity recognition framework for robot-assisted healthcare applications. *Expert Systems with Applications*, 241:122482, 2024.
- [64] Yanjinkham Myagmar-Ochir and Wooseong Kim. A survey of video surveillance systems in smart city. *Electronics*, 12(17), 2023.
- [65] Chili Wu Yiqiang Jiang Mengjie Song, Jiankai Dong and Minglu Qu. Improving the frosting and defrosting performance of air source heat pump units: review and outlook. *HKIE Transactions*, 24(2):88–98, 2017.
- [66] Mohammed Amer and Chi-Chuan Wang. Review of defrosting methods. *Renewable and Sustainable Energy Reviews*, 73:53–74, 2017.
- [67] Mohammed Amer and Chi-Chuan Wang. Experimental investigation on defrosting of a cold flat plate via ultrasonic vibration under natural convection. *Applied Thermal Engineering*, 179:115729, 2020.
- [68] Cláudio Melo, Fernando T. Knabben, and Paula V. Pereira. An experimental study on defrost heaters applied to frost-free household refrigerators. *Applied Thermal Engineering*, 51(1):239–245, 2013.
- [69] Sergio Ledesma and J.M. Belman-Flores. Application of artificial neural networks for generation of energetic maps of a variable speed compression system working with r1234yf. *Applied Thermal Engineering*, 69(1):105–112, 2014.

- [70] J.M. Belman-Flores, Sergio Ledesma, D.A. Rodríguez-Valderrama, and D. Hernández-Fusilier. Energy optimization of a domestic refrigerator controlled by a fuzzy logic system using the status of the door. *International Journal of Refrigeration*, 104:1–8, 2019.
- [71] J.M. Ledesma S. Rangel-Hernández V.H. Cabal-Yépez E. Barroso-Maldonado, J.M. Belman-Flores. Predicting the energy performance of a reciprocating compressor using artificial neural networks and probabilistic neural networks. *Revista Mexicana de Ingeniería Química*, 2017.
- [72] Robert Pasnau. What is sound?. *The Philosophical Quarterly*, 49(196):309–324, 01 2003.
- [73] Rodrigo Avilés-Lopez and Rocío Perera-Martin. *Manual de acústica ambiental y arquitectónica*. Paraninfo, 2017.
- [74] Bao-Lian Su, Clément Sanchez, and Xiao-Yu Yang. *Insights into Hierarchically Structured Porous Materials: From Nanoscience to Catalysis, Separation, Optics, Energy, and Life Science*. John Wiley & Sons, Ltd, 2011.
- [75] Robert Burkard. Sound pressure level measurement and spectral analysis of brief acoustic transients. *Electroencephalography and Clinical Neurophysiology*, 57(1):83–91, 1984.
- [76] Rafael C. Gonzalez and Richard E. Woods. *Digital image processing*. 2 edition, 2002.
- [77] Tianmei Guo, Jiwen Dong, Henjian Li, and Yunxing Gao. Simple convolutional neural network on image classification. In *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, pages 721–724, 2017.
- [78] Russell G. Congalton. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1):35–46, 1991.
- [79] Ivo Düntsch and Günther Gediga. Confusion matrices and rough set data analysis. *Journal of Physics: Conference Series*, 1229(1):012055, may 2019.
- [80] Yair A. Andrade-Ambriz, Sergio Ledesma, Mario-Alberto Ibarra-Manzano, Marvella I. Oros-Flores, and Dora-Luz Almanza-Ojeda. Human activity recognition using temporal convolutional neural network architecture. *Expert Systems with Applications*, 191:116287, 2022.
- [81] Yair A. Andrade-Ambriz, Sergio Ledesma, J.M. Belman-Flores, I. Carvajal-Mariscal, and Dora-Luz Almanza-Ojeda. Frost thickness estimation in a domestic refrigerator using acoustic signals and artificial intelligence. *Expert Systems with Applications*, 201:117071, 2022.
- [82] Yair A. Andrade-Ambriz, Sergio Ledesma, Dora-Luz Almanza-Ojeda, and J.M. Belman-Flores. Accurate classification of frost thickness using visual information in a domestic refrigerator. *International Journal of Refrigeration*, 145:256–263, 2023.