Universidad de Jaén

Escuela Politécnica
Superior de Jaén

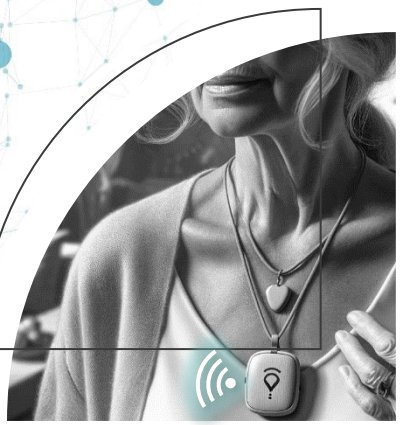
Integración de tecnologías no invasivas para promover la autonomía y mejorar la calidad de vida de personas frágiles

Autor: Aurora Polo Rodríguez

Director de la tesis: Dr. Javier Medina Quero
Departamento: Informática

Fecha: 09/02/2024

Universidad de Jaén



FEBRERO, 2024

INTEGRACIÓN DE TECNOLOGÍAS NO INVASIVAS

para promover la autonomía y mejorar la calidad de vida de personas frágiles

Aurora Polo Rodríguez

Universidad de Jaén

Departamento de Informática

Doctora en Tecnologías de la Información y la Comunicación

Director: Dr. Javier Medina Quero

INTEGRACIÓN DE TECNOLOGÍAS NO INVASIVAS PARA PROMOVER LA AUTONOMÍA Y MEJORAR LA CALIDAD DE VIDA DE PERSONAS FRÁGILES

La presente tesis doctoral nace para fomentar la autonomía y la calidad de vida de las personas frágiles mediante la confección de un sistema de tecnologías no invasivas adaptado al contexto de cada individuo. En base a esta alentadora temática, el estudio se enfoca en los siguientes requisitos de investigación prioritarios relacionados con la sensorización y arquitecturas propuestas: i) baja invasividad, ii) accesibilidad (bajo coste y sencilla adquisición), iii) alta autonomía, iv) escalabilidad, v) inmersividad en el entorno y vi) fácil despliegue; sustentados por la necesidad intrínseca de privacidad y seguridad que demandan los usuarios.

Considerando estos requerimientos y unificando los avances de modelos de cuidados, salud y tecnología que representan nuestras líneas prioritarias, se propone el desarrollo de un sistema distribuido formado por sensores multimodales. Esta infraestructura tecnológica integra sensores de diversa índole, tales como dispositivos de audio, de visión, de localización y ambientales; junto con modelos de reconocimiento de actividades orientados a personas con fragilidad. En este contexto, el reconocimiento de actividades y eventos de la vida cotidiana en entornos inteligentes se adapta tanto a situaciones individuales como a espacios con múltiples ocupantes. Para lograrlo, se emplean modelos basados en conocimiento utilizando Lógica Difusa, y modelos orientados a datos empleando técnicas de Aprendizaje Profundo.

Cada una de las propuestas de sensorización y trazabilidad de usuarios se ha evaluado en casos de estudio que permiten explorar su viabilidad, sensibilidad y precisión en entornos reales.

Palabras clave: Reconocimiento de actividades; personas frágiles; multiocupación; entornos inteligentes; sensores de baja invasividad.

INTEGRATION OF NONINVASIVE TECHNOLOGIES TO ENHANCE AUTONOMY AND IMPROVE THE QUALITY OF LIFE OF FRAIL INDIVIDUALS

This doctoral thesis is developed to promote the autonomy and quality of life of frail individuals through the design of a system of non-invasive technologies tailored to the context of each person. Based on this encouraging topic, the study focuses on the following priority research requirements related to sensorization and proposed architectures: i) low invasiveness, ii) accessibility (affordable cost and easy acquisition), iii) high autonomy, iv) scalability, v) environmental immersion, and vi) easy deployment; supported by the intrinsic need for privacy and security that users demand.

Acknowledging these prerequisites and integrating advancements in care models, health, and technology that represent our priority lines, the development of a distributed system composed of multimodal sensors is proposed. This technological infrastructure integrates a variety of sensors, such as audio devices, vision, location, and environmental sensors; along with activity recognition models aimed at frail individuals. In this context, the identification of activities and events of daily life in smart environments is adaptable to both individual scenarios and multi-occupant settings. To achieve this, knowledge-based models using Fuzzy Logic, and data-oriented models employing Deep Learning techniques are used.

Each proposed sensorization and user traceability approach has been evaluated through case studies that assess their viability, sensitivity, and precision in real-world scenarios.

Keywords: Activity recognition; frail individuals; multi-occupancy; smart environments; low-invasiveness sensors.

A la mujer que me otorgó la vida,
quien celebra cada uno de mis logros
como el triunfo del fruto de sus entrañas.

AGRADECIMIENTOS

Deseo expresar mi profunda gratitud a cada persona que ha sido parte de este largo recorrido, una maratón que parecía interminable. Quiero dar las gracias a mi familia, y en especial a mi madre, por sus palabras llenas de cariño, ternura y apoyo en todo momento.

Mi amor y gratitud a mi compañero de vida, cuya paciencia y dedicación han sido mi motor, manteniendo siempre la fe en mi capacidad para alcanzar mis objetivos.

Pero, en particular, mi sincero agradecimiento a Javier, mi mentor, quien ha estado a mi lado desde el inicio de mi carrera universitaria. Su inmensa confianza en mí ha sido un regalo invaluable, enseñándome que los límites son solo puntos de partida hacia lo extraordinario. Las palabras no bastan para expresar mi agradecimiento por la oportunidad que me ha brindado de crecer tanto personal como profesionalmente, aprendiendo el verdadero significado de amistad y compañerismo.

Finalmente, agradecer el apoyo de las instituaciones que han soportado esta tesis doctoral. Principalmente, el proyecto 857188 (Pharaon Project 'Pilots for Healthy and Active Ageing') del Programa de investigación e innovación Horizonte 2020 de la Unión Europea, junto con el proyecto de investigación DTS21/00047 (entornos autónomos con sensorización de planes de cuidados en personas con discapacidad cognitiva (AMALTEA)) del Instituto Español de Salud Carlos III y las ayudas de la EDUJA para la realización de estancias para la obtención de Mención Internacional.

TABLA DE CONTENIDO

ÍNDICE DE TABLAS

ÍNDICE DE ILUSTRACIONES

ACRÓNIMOS

1D Unidimensional. 73, 88

2D Bidimensional. 53, 54, 59, 65, 73, 83, 85, 98–102, 106, 126, 148

3D Tridimensional. 40, 53, 56, 73, 100–102, 105–107, 147, 148

AOA Ángulo de Llegada (Angle of Arrival, por sus siglas en inglés). 77

AVD Actividades de la Vida Diaria. 6, 8, 11, 24, 25, 34, 39, 40, 47, 48, 75, 83, 111–113, 122

BB Caja delimitadora (Bounding Box, por sus siglas en inglés). 135

BLE Bluetooth de Baja Energía (Bluetooth Low Energy, por sus siglas en inglés). 13, 28, 74, 80–82, 90–95, 144, 148

CNN Red Neuronal Convolucional (Convolutional Neural Network, por sus siglas en inglés). 16, 40, 41, 44–46, 48–51, 53, 54, 57, 61, 63–65, 74–76, 82, 83, 88, 89, 93, 94, 96–98, 145–147

DL Aprendizaje Profundo (Deep Learning, por sus siglas en inglés). 6, 15, 16, 40–42, 48, 49, 52–54, 56, 57, 61, 63–65, 70, 74, 78, 83, 88, 89, 93, 94, 96, 98, 99, 103, 143, 146, 147, 149, 150

EI Entornos Inteligentes. 2, 6–9, 15, 24, 31, 33, 39, 42, 52, 75, 78, 83, 99, 106, 107, 111, 141, 144, 148

FPS Imágenes por Segundo (Frames Per Second, por sus siglas en inglés). 69, 103, 108, 110

FTW Ventanas temporales difusas (Fuzzy Temporal Windows, por sus siglas en inglés). 117, 118, 120, 121, 139

HA Home Assistant. 32, 34, 37–39

IoT Internet de las Cosas (Internet of Things, por sus siglas en inglés). 2, 7, 9, 17, 21, 23, 40, 41, 53–56, 58, 61, 62, 67, 125, 145

LM Espectrograma Log-Mel. 40, 41, 44, 46, 48–50

LSTM Red de Memoria a Corto y Largo Plazo (Long Short-Term Memory, por sus siglas en inglés). 16, 53, 54, 61, 64, 65, 74, 82, 83, 88, 89, 93, 94, 96–98, 146, 147

MAE Error Absoluto Medio (Mean Absolute Error, por sus siglas en inglés). 88, 96, 98

MFCC Coeficientes Cepstrales de Frecuencia Mel (Mel Frequency Cepstral Coefficients, por sus siglas en inglés). 40, 44–46, 48–51, 145

MQTT Message Queuing Telemetry Transport. 30, 36, 39, 42, 43, 90, 93, 106, 125, 126

NFC Comunicación de Campo Cercano (Near Field Communication, por sus siglas en inglés). 43, 85

PIR Sensor Infrarrojo Pasivo (Passive Infrared Sensor, por sus siglas en inglés). 32, 75

POE Energía sobre Ethernet (Power over Ethernet, por sus siglas en inglés). 85

RA Reconocimiento de Actividades Humanas. 3–8, 11, 12, 14–18, 24, 25, 31, 33, 39, 43, 51–53, 72, 82, 112, 116, 117, 141, 143–145, 147–149

RANSAC Consenso de Muestras Aleatorias (RANdom SAmple Consensus, por sus siglas en inglés). 102

ResNet Red Residual (Residual Network, por sus siglas en inglés). 52, 54, 55, 57–60, 67, 82, 99, 103

RF Bosques Aleatorios (Rain Forest, por sus siglas en inglés). 16, 82, 93, 96, 97, 135, 136, 147

RMSE Error Cuadrático Medio Raíz (Root Mean Square Error, por sus siglas en inglés). 81, 146

RSSI Intensidad de la Señal Recibida (Received Signal Strength Indicator, por sus siglas en inglés). 74, 76, 78–81, 83, 84, 86–88, 90, 92–95, 98, 135, 147

RTLS Sistema de Localización en Tiempo Real (Real-Time Location System, por sus siglas en inglés). 12, 76, 78, 125, 144

SVM Máquinas de Soporte Vectorial (Support Vector Machines, por sus siglas en inglés). 16, 40, 82, 93, 96, 97, 135, 136, 147

TDOA Diferencia de Tiempo de Llegada (Time Difference of Arrival, por sus siglas en inglés). 76, 77, 79, 82, 85, 98, 147

TOF Tiempo de Vuelo (Time of Flight, por sus siglas en inglés). 77, 79–81, 147

TTA Tiempo de Llegada Total (Total Time of Arriva, por sus siglas en inglésl). 77, 79

TWR Respuesta de Tiempo de Vuelo (Time-of-Flight Response, por sus siglas en inglés). 81, 82, 85

UWB Tecnología de banda ultra ancha (Ultra-wideband, por sus siglas en inglés). 4, 13, 27, 28, 73, 74, 78, 80–83, 90–95, 111, 112, 122, 123, 126, 135, 144, 147, 148, 150, 151

CAPÍTULO 1

INTRODUCCIÓN

El vehículo para la transformación más vertiginosa de nuestra historia se está generando en tecnología y conocimiento, permitiendo crear y modificar velozmente los modelos de desarrollo sociales [1]. Aunque estos cambios habilitan soluciones hacia una sociedad más solidaria, sana e independiente, se requiere de un esfuerzo multidisciplinar para que los sistemas tradicionales de salud abarquen las necesidades de personas frágiles o dependientes [2].

De forma destacada en este sector, se encuentran las personas mayores. En los países desarrollados, y con especial hincapié en España, el número de personas mayores está manteniendo una tendencia creciente que cambiará el perfil demográfico de nuestra sociedad [3]. Este cambio repercutirá de forma profunda en el ámbito social y sanitario [4] cuyos recursos y medios actuales no son suficientes para abordarlo con los procesos de cuidados actuales. El incremento de esta población generará que los mayores de 85 sean el segmento de edad más numerosos en 2050 [5]. Por otra parte, en España viven más de 250.000 personas con discapacidad intelectual [6] y otros tipos de trastornos, como el espectro autista, se cifra en más de 450.000 personas [7].

La Convención sobre los Derechos de las Personas con Discapacidad dicta en el artículo 19 el derecho a *vivir de forma independiente y ser incluido en la comunidad* [8]. Sin embargo, la estrategia Europea de Discapacidad 2010-2020 de la Comisión Europea ha identificado la

exclusión social como uno de los grandes desafíos a los que se enfrentan las personas con discapacidad en Europa [9].

Con la finalidad de abordar estos dos grandes retos nace la propuesta de tesis doctoral, donde se presenta el desarrollo de un sistema que ayuda a mejorar la vida y la autonomía de este segmento poblacional en su lugar de residencia. La motivación y justificación de esta investigación viene reforzada por el positivo impacto de los Entornos Inteligentes (EI) para monitorizar las actividades diarias de las personas con fragilidad [10], que tiene como pilares fundamentales los campos de Internet de las Cosas (Internet of Things, por sus siglas en inglés) (IoT) e Inteligencia Ambiental y que cada vez está penetrando más en hogares inteligentes y aplicaciones de asistencia sanitaria [11, 12].

Además, esta propuesta se enmarca dentro de las líneas de investigación prioritarias AES establecidas por el Reto en Salud, Cambio Demográfico y Bienestar:

- Los trastornos y tecnologías asociadas al envejecimiento y la discapacidad, así como la rehabilitación y el desarrollo de entornos asistidos y orientados al abordaje de la fragilidad.

- El uso y difusión de las tecnologías habilitadoras como eje vertebrador de un espacio global de e-health en el área de epidemiología, salud pública y servicios de salud, así como en el ámbito de la organización y gestión del Sistema Nacional de Salud.

Así como en las líneas de investigación prioritarias de la convocatoria de la Acción Estratégica en Salud (AES) 2021: investigación e innovación en cuidados de salud, salud mental y tecnologías de la información y comunicación aplicadas a la salud, el fomento de la salud participativa, la atención de la cronicidad y la innovación en cuidados de salud. Por último, también se enmarca en las líneas de actuación preferentes en Tecnologías de la Información y la Comunicación (TIC) del Horizonte 2020 europeo dentro de los retos de la sociedad en el campo de la salud, cambio demográfico y bienestar.

1.1. Hipótesis y objetivos

En la actualidad, las instituciones están fomentando iniciativas para promover que las personas con fragilidad puedan residir de forma segura e independiente en sus hogares o en residencias, disfrutando de una vida digna y enriquecedora.

La hipótesis central de esta tesis doctoral sostiene que *el uso de sensores inmersivos y dispositivos vestibles de mínima invasividad integrados en entornos inteligentes, posibilita la identificación y Reconocimiento de Actividades Humanas (RA) cotidianas. La integración de estas tecnologías permiten una monitorización en tiempo real, ofreciendo así una herramienta valiosa para la optimización de la asistencia y cuidados a individuos con fragilidad garantizando su seguridad y bienestar.*

En base a la hipótesis formulada, los principales objetivos específicos a abordar se definen a continuación:

i) Aumentar la autonomía y seguridad de personas con discapacidad o fragilidad que viven de forma independiente usando dispositivos no invasivos.

ii) Construir un sistema que integre diversos sensores multimodales no invasivos con un esquema reproducible en entornos reales.

iii) Desarrollar procesos de extracción y reconocimiento de patrones mediante Machine Learning sobre sensores no invasivos, de larga autonomía y bajo coste que permitan describir el desarrollo de actividades diarias.

iv) Evaluar, mediante casos de estudios en entornos reales, la capacidad de los dispositivos no invasivos para reconocer actividades cotidianas.

v) Diseminar los resultados de investigación en formatos abiertos que permitan impulsar el avance científico, técnico y la implantación de soluciones derivadas dentro de este campo.

1.2. Estructura

La presente tesis doctoral está estructurada en los siguientes capítulos:

- **Capítulo 2. Estado del arte**: Se analiza el estado actual de la investigación sobre el RA humanas en entornos inteligentes, resaltando la importancia de las tecnologías no invasivas para la monitorización y asistencia de personas con fragilidad. Se discuten los avances en sistemas de RA y se detallan las necesidades específicas de este segmento poblacional.

- **Capítulo 3. Arquitectura y procesamiento de sensores multimodales en entornos inteligentes:** Se propone una arquitectura para el RA en entornos domésticos utilizando sensores multimodales. Se describen las tecnologías empleadas, incluyendo sensores ambientales, de audio y visión, y se presentan casos de estudio para ilustrar su aplicación. Se destaca la importancia de la aproximación multimodal y se introducen metodologías innovadoras para la clasificación de sonidos y el procesamiento de imágenes térmicas.

- **Capítulo 4. Trazabilidad en entornos de multiocupación:** Se examinan tecnologías para la localización y seguimiento en espacios interiores, evaluando sistemas basados en Tecnología de banda ultra ancha (Ultra-wideband, por sus siglas en inglés) (UWB), Bluetooth y sensores de visión. Se discuten los desafíos y soluciones para el seguimiento detallado de interacciones en entornos de multiocupación, destacando el rendimiento de diferentes metodologías y tecnologías en contextos reales.

- **Capítulo 5. RA cotidianas mediante protoformas difusas en entornos de multiocupación:** Se presenta un marco innovador para la discriminación de actividades en entornos de multiocupación, que relaciona espacial y temporalmente a los habitantes con el área donde se ha producido un evento asociado a uno o varios sensores. Se detalla un caso de estudio que demuestra la eficacia del discriminador en la identificación de interacciones específicas y la monitorización de actividades prolongadas. A su vez, se presenta un modelo de protoformas difusas basado en cocimiento para el RA a largo

plazo, relacionadas con las actividades cotidianas de personas frágiles.

- **Capítulo 6. Conclusiones y trabajos futuros:** Finalmente, se sintetizan los principales hallazgos y contribuciones de la tesis, evaluando el impacto de las tecnologías y metodologías desarrolladas en el campo del RA. Se discuten las limitaciones del estudio y se proponen líneas de investigación futuras para continuar avanzando en este ámbito.

CAPÍTULO 2

ESTADO DEL ARTE

En este capítulo, se abordan tres áreas clave que son fundamentales para entender el estado actual y las perspectivas futuras del RA en EI. En primer lugar, se explora cómo la integración de tecnologías avanzadas en entornos cotidianos puede enriquecer nuestra comprensión e identificación de las actividades humanas. Esta revisión abarca desde desarrollos tecnológicos hasta enfoques teóricos y prácticos que han contribuido significativamente a este campo.

La segunda área de enfoque es el procesamiento, modelos y arquitecturas de RA, donde se examinan las técnicas y metodologías utilizadas para procesar y analizar datos de Actividades de la Vida Diaria (AVD). Esta sección incluye una discusión sobre diversos modelos y arquitecturas, desde métodos convencionales hasta técnicas de Aprendizaje Profundo (Deep Learning, por sus siglas en inglés) (DL), destacando sus aplicaciones, ventajas y limitaciones en el contexto de EI.

Finalizando y enfatizando la motivación principal de esta tesis doctoral, se dedica una sección a las aplicaciones en personas frágiles; donde se descubre cómo las tecnologías y metodologías investigadas pueden ser empleadas para enriquecer la vida de individuos vulnerables, tales como los ancianos o personas con discapacidades. Se profundiza en cómo los EI y el RA tienen el potencial de ofrecer asistencia personalizada, mejorar la seguridad y fomentar la autonomía de estas personas, proporcionando una visión completa de su influencia y po-

sibilidades en la sociedad. Esta sección establece una base sólida para nuestra investigación, situándola dentro de un contexto amplio de conocimiento e indagación en el ámbito del RA en EI.

2.1. Entornos inteligentes y reconocimiento de actividades

La disciplina de EI surge con el propósito de respaldar las actividades cotidianas de manera envolvente y omnipresente. Relacionada con este término, surge con anterioridad un concepto revolucionario en Ciencias de la Computación: la Computación Ubicua, definida por Weiser como *"la era de la tecnología serena, cuando la tecnología se desvanece en el trasfondo de nuestras vidas"* [13]. Desde esta perspectiva visionaria cuyos inicios se remontan a la década de 1990 hasta el actual panorama del Internet de las Cosas (Internet of Things, por sus siglas en inglés) (IoT), la investigación científício técnica se ha concentrado en tres grandes ramas durante estos 30 años: (i) la integración de sensores no intrusivos e inmersivos, tanto en nuestros cuerpos como en nuestro entorno, (ii) la comunicación inteligente entre dispositivos de larga duración de batería y bajo consumo, y (iii) la creación de métodos o mecanismos de RA eficientes adaptados al contexto y capaces de proporcionar resultados interpretables derivados del procesamiento y tratamiento de datos de sensores.

Actualmente, las soluciones de última generación destinadas a EI posibilitan la monitorización de las actividades humanas de manera cada vez menos invasiva y se han convertido en un campo prolífero para la comunidad científica [14]. Su implementación ostenta el potencial de optimizar la prestación de servicios de atención médica, a la vez que viabiliza y prolonga el tiempo de autonomía de las personas en sus domicilios [15]. Dentro de este campo, destaca el RA, que se enfoca en la creación de modelos predictivos destinados a la identificación de actividades humanas [16] en el marco de un EI con el objetivo de ofrecer asistencia y supervisión a sus habitatantes.

Los sensores ambientales se relacionan con los sentidos del entorno inteligente, que otorgan información del mundo y del espacio cambiante para que modelos inteligentes procesen

información compleja en tiempo real transformando datos en conocimiento interpretable.

Así, los datos recolectados pueden, posteriormente, ser utilizados para describir las AVD de los individuos mediante la supervisión de la interacción entre el usuario y el entorno [17, 18]. Tal como se introduce previamente, los sensores binarios se propusieron con la finalidad de describir las actividades diarias en espacios interiores [19] con alta adaptación por su naturaleza inmersiva y bajo coste. Estos sensores generaron avances alentadores en conjuntos de datos meticulosamente etiquetados usando enfoques orientados a datos [18]. Dentro de la categoría de sensores binarios, es relevante destacar varios dispositivos específicos que desempeñan funciones distintas en un EI:

- Sensores de presencia o movimiento pasivo: activados por el desplazamiento de objetos, estos sensores registran la presencia o el movimiento de elementos dentro del entorno.
- Sensores de apertura de puertas y ventanas: detectan la separación de componentes físicos que normalmente están unidos, permitiendo identificar la apertura y cierre de puertas y ventanas.
- Botones de activación: diseñados para registrar la activación de interruptores, paneles de luz u otros dispositivos, como la activación de una cisterna.
- Sensores de movimiento o vibración: activados por fuerza física, generalmente a través de un giroscopio, para detectar cambios de posición o movimiento.
- Sensores de fuga de agua, gas o humo: identifican situaciones de riesgo al detectar la presencia de sustancias peligrosas, como agua, gas o humo, cuando su concentración excede cierto umbral.

A pesar de que no se categorizan como sensores binarios, dado que su medida no se expresa como una activación booleana, sino como una medida multivaluada, existen otros tipos de sensores ambientales de bajo coste e invasividad para medir diversos parámetros. Estos sensores resultan particularmente relevantes en el ámbito del RA y los EI:

- Sensores de temperatura y humedad: miden la temperatura y la humedad relativa del

entorno, siendo fundamentales para la regulación del clima en interiores y el seguimiento de condiciones ambientales específicas.

- Sensores de calidad del agua: cruciales para evaluar la calidad del agua potable u otras fuentes de agua, detectando posibles contaminantes o variaciones en los niveles de pH, turbidez y otros parámetros esenciales para la seguridad y la salud.

- Sensores de calidad del aire: elementos esenciales para analizar la composición y contaminación del aire, midiendo la concentración de compuestos como dióxido de carbono, monóxido de carbono, ozono y partículas en suspensión, lo que permite evaluar la calidad del aire en entornos cerrados. [20].

- Sensores de ruido: estos dispositivos miden los niveles de ruido ambiental, siendo esenciales para evaluar la calidad del entorno sonoro, detectar ruidos excesivos o inusuales y aplicaciones que requieren control acústico.

- Sensores de luminosidad: miden la intensidad de la luz en un entorno, lo que es fundamental para regular la iluminación en interiores, así como para detectar cambios en las condiciones de luz que pueden afectar la visibilidad o el confort visual en un espacio determinado.

La principal característica inherente a estos dispositivos radica en su baja intrusividad, pero esto también limita su capacidad para discernir la ocupación múltiple de habitantes en EI, dado que su activación no permite distinguir qué usuario ha interactuado con los objetos o está relacionado con una medida de un sensor ambiental. No obstante, el actual incremento en la oferta y la demanda de dispositivos inteligentes ha propiciado el desarrollo de sensores multimodales, que integran altas prestaciones sensoriales mediante la recolección de imagen, audio y sensores portátiles o vestibles [21] integrados en los usuarios. Además, estos dispositivos IoT están dotados altos rendimientos computacionales en constante crecimiento. Desde una perspectiva de hardware, esta tecnología ha evolucionado hacia una mayor sofisticación, incrementando los recursos computacionales, la duración de batería y los métodos de comunicación, al tiempo que se ha vuelto más asequible propiciando su democratización. Este avance ha allanado el camino de nuevas tendencias que han convergido en los llamados

sensores sintéticos [22], desplegados con la intención de cubrir por completo un espacio específico, permitiendo la implementación de tecnologías de detección de propósito general que supervisen actividades mediante la fusión de información sensorial. Dentro de esta amplia categoría, se destacan los siguientes:

- Sensores de visión: Estos dispositivos tienen la capacidad de adquirir información visual en distintos espectros, incluyendo el visible, térmico o infrarrojo [23]. Los sensores de visión emplean tecnología de cámaras y dispositivos especializados para capturar imágenes o secuencias visuales. La información recopilada abarca datos sobre formas, colores, texturas, movimientos, patrones y estructuras presentes en la escena observada. En particular, los sensores térmicos o infrarrojos permiten detectar la radiación emitida por los cuerpos según el calor que estos generan, y se utilizan en diversas aplicaciones, desde la vigilancia de la seguridad hasta la visión por computadora.

- Sensores de audio: Registran y reconocen información acústica, que comprende formas de onda sonora y, en aplicaciones más avanzadas, puede incluir espectrogramas, representaciones visuales del espectro de frecuencias y sus variaciones temporales [24]. Facilitan el análisis de patrones acústicos, identificando sonidos específicos o eventos de interés, como voz, música, ruidos ambientales o alertas. Son fundamentales para aplicaciones que demandan análisis de audio, tales como la detección de alarmas, control de ruido, asistentes de voz y sistemas de seguridad basados en sonido.

- Sensores vestibles o *wearable*: Diseñados para monitorizar y analizar diversas actividades, movimientos y gestos de las personas [25]. Los datos obtenidos de sensores wearables varían según el tipo de sensor integrado en el dispositivo. Estos datos incluyen información proveniente del acelerómetro (movimiento lineal), giroscopio (orientación), pulsómetro (frecuencia cardíaca), podómetro (número de pasos), sensores de temperatura corporal, calidad del sueño, entre otros. Estos datos permiten un seguimiento detallado de la actividad física, la salud y el bienestar del usuario.

- Sensores de localización: Permiten identificar la ubicación espacial de un objeto o persona [26]. Hacen uso de distintas tecnologías como GPS, RFID, sistemas de posicionamiento

por satélite o redes inalámbricas para determinar la posición de un objeto o individuo. Dependiendo de la tecnología empleada (GPS, RFID, UWB, etc.), se obtienen coordenadas precisas que pueden incluir información sobre latitud, longitud, altitud y, en algunos casos, detalles sobre la velocidad, dirección, orientación y distancia a un objeto de interés. Son cruciales para sistemas de navegación, logística, seguimiento de activos, servicios basados en la ubicación y aplicaciones de realidad aumentada.

En este contexto, el RA se basa en la utilización de dispositivos interconectados con el propósito de monitorizar las actividades cotidianas de las personas y reconocer las acciones que llevan a cabo, empleando datos recolectados a partir de sensores. Estos datos se representan en los modelos en forma de secuencias temporales o flujos de información que capturan valores específicos, a menudo sujetos a cierta incertidumbre, o las transiciones entre diferentes estados [27].

Como fuente de representación de la información, los datos desempeñan un papel esencial en la obtención de conocimiento relacionado con las AVD, lo que plantea importantes cuestiones de privacidad [28, 29]. Por esta razón, los sensores binarios y los dispositivos vestibles o *wearable* se han consolidado como las tecnologías preeminentes en este ámbito, debido a su naturaleza discreta y menos intrusiva [30]. Sin embargo, lograr el RA en el contexto de hogares inteligentes plantea un desafío significativo, debido a la inherente complejidad y variabilidad de las actividades humanas. Estas actividades pueden variar de un día a otro y entre diferentes ocupantes, cada uno de los cuales exhibe patrones de comportamiento y capacidades distintivas [31].

En función del lugar donde se sitúe el dipositivo para llevar a cabo el RA, se ha realizado una división en dos grandes grupos:

- **Sensores desplegados en el entorno:**

  i) **Sensores de visión**, que implica la instalación de cámaras en ubicaciones estratégicas dentro del hogar. Estos dispositivos permiten recopilar información esencial sobre el entorno y las interacciones que allí tienen lugar, incluyendo la estimación de *body landmarks* o puntos de referencia del cuerpo, detección de la postura, identifi-

cación de la persona a través de la forma corporal o el rostro, así como la estimación de actividades basadas en los objetos reconocidos en la escena, entre otros aspectos relevantes. Sin embargo, estas soluciones presentan desafíos, como variaciones significativas en la precisión debido a cambios en la iluminación, además de la inquietud de las personas respecto a la privacidad [32, 33].

ii) **Sensores ambientales y binarios**, que involucran un despliegue algo más complejo, ya que requiere identificar objetos y variables clave a monitorizar. Es necesario estudiar las condiciones del entorno a monitorizar e identificar las actividades a reconocer, dado que cada electrodoméstico u objeto en el entorno requiere estar equipado con sensores. Por ejemplo, estos sensores permiten detectar la interacción con elementos clave en la cocina, como el frigorífico o microondas con sensores de apertura y cierre, o la vitrocerámica con sensores de consumo, temperatura y humedad. La principal ventaja de estos sensores radica en su menor intrusión en comparación con otros tipos, aunque enfrentan desafíos en la interpretación de los datos recopilados para modelar el RA, lo que requiere el uso de técnicas de fusión de datos y conocimiento experto. [34].

iii) **Sensores de audio**, que capturan información sonora del entorno y permiten reconocer eventos específicos o actividades basadas en patrones acústicos. Requieren la instalación de micrófonos en ubicaciones estratégicas, teniendo en cuenta que pueden existir solapamientos de sonidos.

iv) **Sensores de localización**, que utilizan tecnologías conocidas como Sistema de Localización en Tiempo Real (Real-Time Location System, por sus siglas en inglés) (RTLS) para interiores y que requieren de la instalación de diversas balizas para poder trazar la localización del usuario. Estos dispositivos, conocidos como anclas, dependen de que los usuarios porten un tag para su identificación única mediante diferentes técnicas.

v) **Sensores basados en radar**, que emplean ondas electromagnéticas para detectar la presencia y movimiento de objetos o personas en el entorno, siendo especialmente útiles en situaciones donde la visión directa puede ser limitada. La limitación de

estos sensores radica de nuevo en la incapacidad de identificar a la persona.

- **Sensores vestibles o *wearable***: representan la opción más popular gracias a la proliferación de dispositivos como teléfonos inteligentes y relojes inteligentes, los cuales incorporan una variedad de sensores, lo que los hace idóneos para monitorizar la salud y la actividad física de los usuarios, así como para rastrear la posición del usuario en interiores a través de señales como BLE o UWB. Sin embargo, la principal limitación radica en la duración relativamente corta de la batería y la posibilidad de que los sujetos de estudio olviden llevar consigo el dispositivo correspondiente [35].

  i) **Sensores inerciales**, que miden la aceleración, la velocidad angular y, en algunos casos, la orientación del cuerpo, proporcionando información valiosa sobre la actividad física y el movimiento del usuario, por ejemplo, los pasos o la intensidad del ejercicio.

  ii) **Sensores biométricos**, que registran datos fisiológicos como la frecuencia cardíaca, la temperatura corporal, la calidad del sueño y otros parámetros biológicos para evaluar el estado de salud y el nivel de esfuerzo físico.

  iii) **Sensores de localización**, que, en el contexto de dispositivos vestibles, pueden hacer uso de tecnologías específicas como GPS, UWB, Bluetooth de Baja Energía (Bluetooth Low Energy, por sus siglas en inglés) (BLE) o sensores de proximidad para lograr una determinación más precisa de la ubicación del usuario. Estos sensores suelen depender de otros que actúan como anclas instaladas en el entorno, siendo denominados *tags* en este caso.

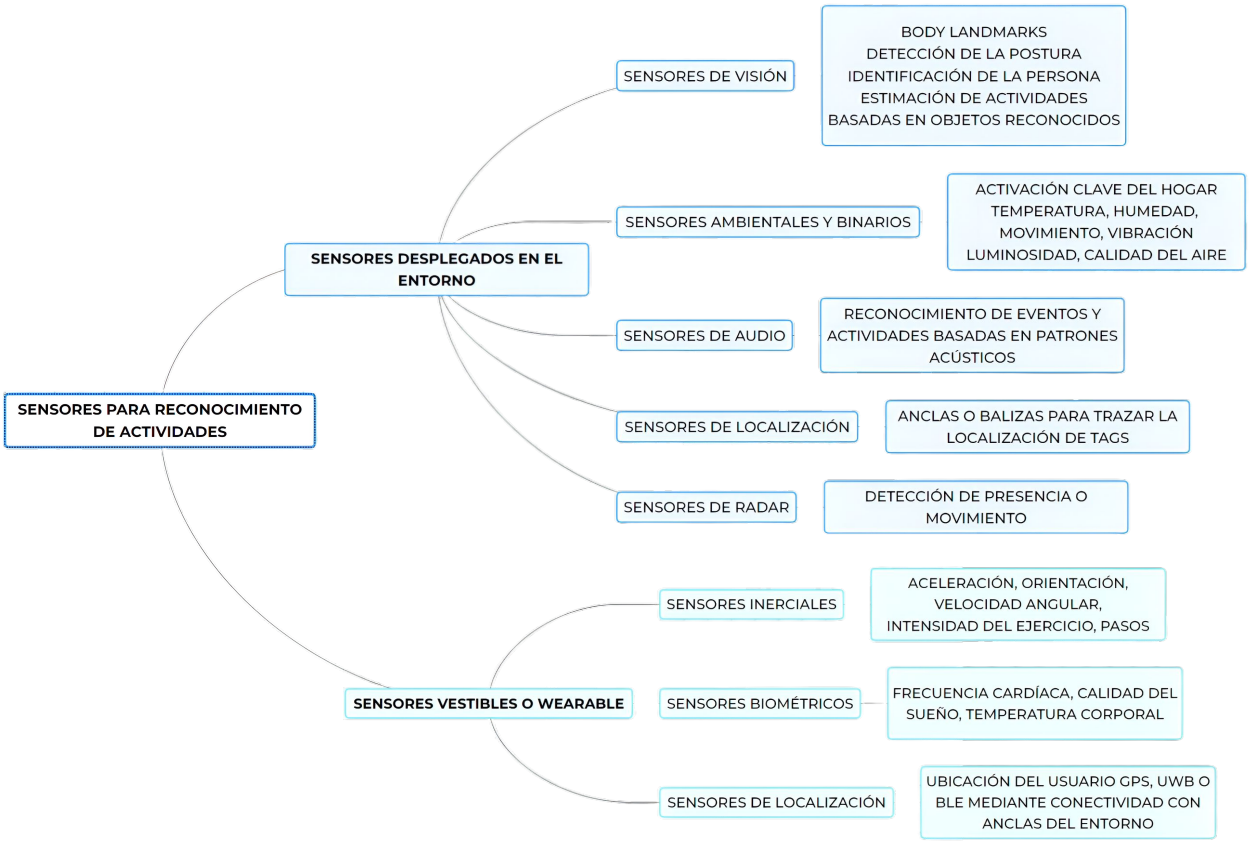


Figura 2.1: Esquema de clasificación de sensores según su ubicación: desplegados en el entorno o vestibles.

Una vez recopilados los datos de los sensores, es factible desarrollar modelos de RA, que abarca métodos basados en datos y métodos basados en conocimiento.

2.2. Procesamiento, modelos y arquitecturas de RA

Según la naturaleza de los modelos utilizados en RA, podemos definir dos categorías principales: enfoques basados en conocimiento [36] y enfoques dirigidos por datos [18].

Los modelos basados en conocimiento utilizan conocimiento explícito de un experto sobre un dominio específico para realizar tareas de razonamiento y toma de decisiones. En ellos, destabamos los modelos basados en reglas y lógica descriptivias derivadas de conocimientos previos, aportando interpretabilidad intrínseca [37]. Su flexibilidad ante cambios imprevistos es limitada, afectando su capacidad de generalización. Por otro lado, los modelos basados en

datos, desarrollados con algoritmos de aprendizaje automático que extraen patrones directamente de los datos, son flexibles y adaptables, aunque suelen tener menor interpretabilidad [38, 39]. En la práctica de entornos reales, ambos enfoques pueden convivir [40].

Uno de los procesos clave en RA que ha generado importantes retos de investigación es la segmentación [41] e identificación de los tamaños de ventana temporal [42] para el reconocimiento y evaluación de patrones de sensores . En cuanto a la metodología de segmentación de datos, se distinguen dos enfoques: la segmentación explícita [43], cuando se tiene conocimiento previo sobre el inicio y el final de las actividades, y la segmentación en tiempo real, necesaria en el contexto de RA para condiciones del mundo real [44], utilizando enfoques de ventanas deslizantes [43].

En el ámbito de los modelos de DL aplicados al RA, existen dos enfoques fundamentales en la estructuración del proceso de aprendizaje: aprendizaje supervisado y aprendizaje no supervisado [45]. El aprendizaje supervisado implica entrenar el modelo con un conjunto de datos que incluye ejemplos de entrada y sus respectivas salidas o etiquetas, y que incluye generalmente las propuestas basadas en datos junto a sus etiquetas [46]. En cambio, el aprendizaje no supervisado trabaja con datos sin etiquetar, buscando descubrir patrones, estructuras o relaciones intrínsecas sin la orientación de salidas previamente conocidas [47]. Hasta ahora, los enfoques de aprendizaje supervisado han predominado en el campo del RA, debido a la naturaleza de los entornos donde se implementan las soluciones. Sin embargo, un desafío importante en EI ha sido desarrollar modelos de RA eficaces y transferibles a diferentes contextos [48, 49]. A pesar de los avances, los modelos diseñados para RA suelen estar estrechamente ligados a implementaciones específicas de sensores que generan los datos de entrenamiento, lo que dificulta su adaptación a diferentes entornos y tipos de actividades, ya que requieren reconfiguración o ajustes para adecuarse a la configuración específica de sensores en el nuevo entorno. Además, en el aprendizaje automático supervisado, la transferencia se ve limitada por la necesidad de disponer de datos etiquetados.

En este contexto, la representación de la información sensorial y los modelos empleados se han visto influenciados por el tipo de sensores involucrados. La elección de los modelos utilizados en este campo depende de la entrada que reciben y su complejidad [46, 50], entre

los que destacan:

- K-NN (K-Vecinos Más Cercanos): Usados para la clasificación de actividades basada en similitudes con instancias previas.
- Máquinas de Soporte Vectorial (Support Vector Machines, por sus siglas en inglés) (SVM): Utilizadas para clasificar patrones de actividad basados en características extraídas de datos, eficaces en tareas como la identificación de posturas o gestos.
- Bosques Aleatorios (Rain Forest, por sus siglas en inglés) (RF): Aplicados en la clasificación de actividades, efectivos con datos temporales o secuenciales y robustos ante el sobreajuste.
- CNN: Efectivas en el procesamiento de datos estructurados como imágenes, utilizadas en la clasificación de actividades humanas en videos o imágenes.
- Red de Memoria a Corto y Largo Plazo (Long Short-Term Memory, por sus siglas en inglés) (LSTM): Ideales para modelar secuencias temporales y aplicadas en el RA con secuencias de eventos.
- Redes Generativas Adversarias (Generative Adversarial Networks, por sus siglas en inglés): Empleadas para generar datos sintéticos realistas, útiles en la ampliación de conjuntos de datos y la mejora de la robustez del modelo.

Especialmente, los modelos de DL han demostrado ser estrategias idóneas en el ámbito del RA en comparación con los algoritmos de aprendizaje automático tradicional, permitiendo descubrir y extraer características relevantes de los datos sensoriales [51, 52].

En el ámbito de la arquitectura de componentes enfocada en el aprendizaje y la comunicación entre dispositivos, los conceptos de computación en el borde *(edge computing)* [53] y computación en la niebla *(fog computing)* [54] han revolucionado la gestión de datos y servicios, trasladándolos a los dispositivos que albergan los sensores. Estos paradigmas se caracterizan por una red de dispositivos que interactúan y colaboran mutuamente para alcanzar objetivos compartidos [55]. La eficacia en la transmisión de datos y en la distribución

de procesos de aprendizaje en tales arquitecturas resulta fundamental para el progreso en el campo del RA.

- **Computación en la niebla [54] (Fog Computing)**. Es un paradigma emergente en el procesamiento de datos que resulta especialmente relevante en el contexto de sensores y RA. En contraste con la computación en la nube, que centraliza el procesamiento de datos en servidores remotos, el fog computing lleva la capacidad de procesamiento y almacenamiento más cerca del borde de la red, es decir, cerca de donde se generan los datos. Esto es particularmente beneficioso para aplicaciones que involucran sensores y RA, donde la rapidez y la eficiencia en el procesamiento son cruciales.

  En aplicaciones de RA, como el seguimiento de la salud, la seguridad o la monitorización de deportes, los sensores generan grandes cantidades de datos en tiempo real. El fog computing permite procesar estos datos localmente o en nodos cercanos, reduciendo la latencia y el ancho de banda necesario para enviar datos a la nube. Esto significa que las decisiones pueden tomarse más rápidamente, lo cual es esencial para respuestas en tiempo real, como alertas médicas de emergencia o ajustes automáticos en dispositivos de asistencia.

  Además, el fog computing facilita la privacidad y la seguridad de los datos, ya que menos información sensible se transmite a través de la red. Esto es particularmente importante en aplicaciones de salud y seguridad, donde la protección de los datos personales es fundamental.

- **Computación en el borde [53] (Edge Computing)**. Es un enfoque en el procesamiento de datos que, al igual que el *fog computing*, se centra en trasladar las capacidades de cómputo más cerca de la fuente de datos, pero con algunas diferencias clave. La principal diferencia entre edge computing y fog computing radica en la ubicación y la proximidad del procesamiento de datos. Mientras que el fog computing implica procesar los datos en nodos o capas intermedias entre los dispositivos de origen y la nube, el edge computing lleva este procesamiento directamente a los dispositivos en el borde de la red, como sensores, cámaras, y otros dispositivos IoT. Esto genera en una reducción

aún mayor de la latencia, ya que los datos no necesitan viajar a otros nodos para su procesamiento.

En aplicaciones de RA, esto significa que los datos generados por los sensores pueden ser procesados inmediatamente en el mismo dispositivo o en un servidor local muy cercano. Por ejemplo, un reloj inteligente que monitoriza la actividad física puede analizar los datos de movimiento directamente en el dispositivo, permitiendo una respuesta casi instantánea y personalizada, o un dispositivo de visión térmica puede procesar directamente el reconocimiento de las personas, sin enviar el flujo de datos en tiempo real a un nodo de procesamiento explícito, con el coste computacional y de transferencia que conlleva.

El edge computing optimiza la eficiencia en el uso del ancho de banda, ya que solo los datos procesados y relevantes pueden necesitar ser enviados a la nube o a otros sistemas para su análisis o almacenamiento a largo plazo. Esto es especialmente útil en entornos con conectividad limitada o costosa.

En cuanto a la privacidad y seguridad, el edge computing ofrece ventajas similares a las del fog computing, ya que menos datos sensibles se transmiten a través de la red. Sin embargo, el procesamiento edge puede plantear desafíos adicionales en términos de gestión y seguridad de los dispositivos, ya que cada uno se convierte en un punto de procesamiento de datos.

La selección y diseño de las etapas de procesamiento, modelos de reconocimiento y arquitecturas en el diseño de sistemas de RA es fundamental para garantizar su eficacia, precisión y eficiencia. Cada una de estas decisiones influye significativamente en cómo el sistema procesará y analizará los datos para identificar y clasificar distintas actividades humanas. Las etapas de procesamiento, que pueden incluir la adquisición de datos, preprocesamiento, segmentación, extracción de características y clasificación, deben ser cuidadosamente seleccionadas y optimizadas para el tipo de actividad que se quiere reconocer y el entorno en el que se opera. Por ejemplo, el preprocesamiento puede requerir filtros específicos para eliminar ruido de los datos de sensores, mientras que la segmentación y la extracción de características deben ser capaces de capturar eficazmente los patrones distintivos de cada actividad [56].

La selección del modelo y la arquitectura del sistema es igualmente crítica. Los modelos pueden variar desde enfoques tradicionales de aprendizaje automático, como máquinas de soporte vectorial o árboles de decisión, hasta métodos más avanzados como redes neuronales profundas y aprendizaje por refuerzo. La elección depende de varios factores, incluyendo la complejidad de las actividades a reconocer, la cantidad y tipo de datos disponibles, y los recursos computacionales a disposición. La arquitectura del sistema, ya sea basada en computación en la nube, edge computing o una combinación de ambos, también juega un papel crucial [57], especialmente en términos de latencia, capacidad de procesamiento en tiempo real y gestión de la privacidad de los datos. El diseño de las arquitecturas determina crucialmiente su despliegue y la escalabilidad y seguridad en entornos reales.

2.3. Aplicaciones en personas frágiles

La fragilidad, definida por la Sociedad Española de Medicina Interna como un *"síndrome biológico de disminución de la reserva funcional y resistencia a los estresores, debido al declive acumulado de múltiples sistemas fisiológicos que originan pérdida de la capacidad homeostática y vulnerabilidad a eventos adversos"*, plantea una serie de desafíos intrínsecos relacionados con el envejecimiento, la disminución progresiva de las capacidades físicas y cognitivas, y la coexistencia de patologías crónicas. A pesar de la importancia de esta problemática, aún persiste un desconocimiento en algunos ámbitos profesionales respecto a su prevalencia y definición. La incidencia de la fragilidad crece exponencialmente con la edad, variando desde un 3,2 % en personas de 65 años, hasta un 16,3 % en mayores de 80 años y un 23,1 % en aquellos de 90 años o más. Este síndrome es particularmente relevante debido a que actúa como un predictor clave de eventos adversos graves en la población anciana, incluyendo mortalidad, institucionalización, caídas, deterioro de la movilidad, incremento en la dependencia para realizar actividades básicas e instrumentales de la vida diaria y hospitalizaciones. Por lo tanto, el síndrome de fragilidad engloba a personas con discapacidades, adultos mayores y aquellos con diversas condiciones de salud que afectan negativamente su calidad de vida [58, 59].

Para proporcionar una atención óptima a este segmento problacional, es esencial descartar y tratar diferentes patologías que pueden producir debilidad progresiva, cansancio, pérdida de peso, menor tolerancia al ejercicio, menor rendimiento y disminución de la velocidad en la marcha, factores que podrían desencadenar una "fragilidad secundaria" como depresión, insuficiencia cardiaca, hipotiroidismo, tumores, entre otros. El tratamiento debe estar enfocado en disminuir diversos marcadores de fragilidad, así como en detectar y prevenir el desarrollo de discapacidad [60]. En este contexto, uno de los tratamientos más frecuentes se enfocan en la pérdida de peso. Los ejercicios de resistencia han demostrado ser eficaces, aumentando 2 a 3 veces la masa corporal magra con mejoras en la fuerza, tolerancia al ejercicio y velocidad de la marcha, beneficio que crece con la adición de suplementos nutricionales. Asimismo, la actividad física, adaptada a las capacidades y limitaciones individuales, juega un rol fundamental en la mitigación del declive funcional asociado con la inactividad física. La implementación de programas de ejercicio supervisado contribuye no solo a la mejora o mantenimiento de la capacidad funcional, sino también a la reducción del riesgo de caídas y al fortalecimiento de la salud mental, factores todos ellos que contribuyen a la preservación de la independencia y a una mejora global de la calidad de vida [61, 62, 63].

Además, la monitorización y atención en tiempo real adquieren una importancia crucial. Los pacientes con fragilidad tienden a tener menor capacidad para tolerar factores estresantes como procedimientos médicos u hospitalización, aumentando el riesgo de discapacidad u otros resultados adversos [58]. La presencia de este síndrome es común en el entorno hospitalario, con una prevalencia del 46,8 % al 57,4 % entre las personas mayores hospitalizadas [59]. Sumado a este hecho, la rehabilitación puede estar comprometida por fluctuaciones en el estado de salud y un alto riesgo de complicaciones médicas. Por lo tanto, la supervisión y vigilancia médica continua son fundamentales para la prevención de estos problemas, mejorando la efectividad de la rehabilitación y el pronóstico funcional de estos pacientes [64]. Paralelamente, la asistencia en las tareas básicas de la vida diaria adquiere una relevancia singular. La progresiva pérdida de fuerza muscular, la disminución de la movilidad y el deterioro en las habilidades motoras finas, que son características de la fragilidad, limitan la capacidad de los individuos para realizar actividades cotidianas, incidiendo directamente en su independencia,

autoestima y bienestar emocional [65]. La necesidad de asistencia en las actividades básicas de la vida diaria se duplica con cada década hasta los 84 años, y se triplica entre los 85 y 95 años. La proporción de personas mayores con dependencia en actividades instrumentales alcanza el 10-20 %, y aumenta hasta el 30 % en mayores de 80 años. La asistencia proporcionada en estas áreas es clave para mantener la dignidad del individuo, asegurando un entorno de vida seguro y confortable [58].

La comunicación y la integración social se erigen como pilares esenciales en el manejo de la fragilidad, dado que el aislamiento social y la soledad son problemáticas prevalentes en este grupo poblacional [66, 67]. Fomentar la interacción social y el apoyo emocional a través de actividades y grupos de soporte es vital para mantener la salud mental, prevenir trastornos como la depresión y promover un sentido de pertenencia y propósito. En este sentido, la estimulación cognitiva es un elemento clave para contrarrestar el declive cognitivo asociado al proceso de envejecimiento y la fragilidad [68, 69]. A través de programas específicos y actividades adaptadas que fomentan la actividad mental, es posible preservar la función cognitiva, mejorando la autoestima y promoviendo la independencia.

La atención a las personas con fragilidad demanda un enfoque holístico y multidisciplinario que abarque todas las dimensiones del bienestar y la salud. En este sentido, el IoT surge como una herramienta innovadora con un enorme potencial para brindar soporte integral a estas personas. La tecnología ofrece soluciones personalizadas que atienden las necesidades específicas de este segmento poblacional, potenciando notablemente su calidad de vida y fomentando su autonomía.

Mediante el uso de dispositivos interconectados, sensores y sistemas avanzados, el IoT se adecúa a las circunstancias personales, ofreciendo un soporte completo y personalizado. Una de las principales ventajas del IoT es su capacidad para proporcionar monitorización y asistencia en tiempo real [70]. Los dispositivos IoT, capaces de recopilar datos de manera continua, facilitan respuestas inmediatas ante emergencias o cambios en el estado de salud, siendo cruciales para personas que requieren atención constante. Existen numerosas propuestas que incorporan esta tecnología para monitorizar signos vitales como el ritmo cardíaco, la presión arterial y los niveles de oxígeno en sangre [25, 71, 12, 72]. En el contexto de pacientes

con Alzheimer, por ejemplo, los dispositivos de localización pueden alertar a los cuidadores si el paciente abandona una zona geográfica segura, previniendo así situaciones de riesgo como el extravío [73, 74, 75, 76, 77]. Estos dispositivos poseen además la capacidad fomentar la continuidad y seguimiento de la actividad física, elemento indispensable en el cuidado de este grupo poblacional [78, 79, 80, 81].

Estas tecnologías no invasivas se integran de forma fluida en la vida cotidiana, abarcando desde la automatización del hogar hasta dispositivos portátiles y aplicaciones móviles. No solo asisten en tareas diarias, sino que también se adaptan a necesidades en evolución, asegurando que el apoyo proporcionado se ajuste a las circunstancias cambiantes del individuo [82, 83]. Los sistemas de automatización del hogar, por ejemplo, facilitan el control de luces, termostatos y electrodomésticos mediante comandos de voz o aplicaciones móviles, lo cual resulta especialmente beneficioso para personas con discapacidad motriz o visual. Adicionalmente, los asistentes virtuales con inteligencia artificial pueden ofrecer recordatorios para la toma de medicamentos, citas médicas y actividades diarias, incrementando la independencia y disminuyendo la carga para los cuidadores [84, 85, 86]. Otras iniciativas han integrado el uso de robots asistentes que apoyan en tareas como la alimentación o la movilización dentro del hogar [87, 88].

En el aspecto cognitivo, la tecnología aporta soluciones innovadoras para mejorar la comunicación y la integración social [89, 90]. Los dispositivos y aplicaciones se pueden personalizar para ajustarse a las necesidades específicas de comunicación, tales como aplicaciones que traducen texto a lenguaje hablado y viceversa. Esto no solo optimiza la comunicación, sino que también permite a estas personas participar de manera más activa en la sociedad [10]. En el contexto de individuos con autismo, se han realizado varios estudios que emplean robots como herramienta para potenciar las habilidades comunicativas [91]. La mayoría de propuestas que involucran a personas pertenecientes al espectro autista se centran en el uso de la tecnología para facilitar interacciones directas, así como para el reconocimiento de gestos o emociones, con el fin de comprender de manera más efectiva la condición del individuo [92, 93]. Cada vez es más común encontrar proyectos enfocados en prevenir el aislamiento social y la soledad no deseada, abordando estos problemas mediante tecnología. Por ejemplo, el proyecto

Pharaon [94] se enfoca en el desarrollo de plataformas abiertas integradas, personalizables e interoperables que integran una variedad de tecnologías avanzadas. Dentro de este proyecto, el piloto implementado en Andalucía establece objetivos específicos como combatir la soledad no deseada y mejorar las habilidades cognitivas y físicas de los usuarios. Para ello, el piloto incorpora el uso de redes sociales y juegos cognitivos, así como dispositivos de medición de actividad física. Estas herramientas contribuyen al desarrollo y fortalecimiento de diversas áreas de la memoria y otras capacidades cognitivas, representando un enfoque innovador para mejorar la calidad de vida de los usuarios [95].

No obstante, la implementación del IoT en el cuidado de personas frágiles presenta desafíos significativos, como la privacidad y seguridad de los datos, y la accesibilidad y asequibilidad de las tecnologías. Es crucial garantizar que los avances sean accesibles para todos, independientemente de su situación económica o ubicación geográfica. Además, para maximizar su potencial, es fundamental la colaboración entre tecnólogos, profesionales de la salud, responsables políticos y, especialmente, las propias personas frágiles y sus comunidades. Esta participación asegurará que las soluciones desarrolladas sean tecnológicamente avanzadas y humanamente sensibles, creando un entorno de cuidado que responda efectivamente a las necesidades de este grupo poblacional.

CAPÍTULO 3

ARQUITECTURA Y PROCESAMIENTO DE SENSORES MULTIMODALES EN ENTORNOS INTELIGENTES

La interacción entre humanos y tecnología ha evolucionado hacia la convergencia de sistemas sensoriales y la inteligencia artificial, abriendo la puerta a EI que responden de manera adaptativa a las necesidades de las personas. En este contexto, la fusión de datos provenientes de diversos sensores se vuelve esencial para comprender de manera holística los entornos y las actividades humanas. En este capítulo, se analiza el tratamiento inteligente de sensores multimodales como columna vertebral de esta tesis doctoral, base a partir de la cual se realiza el reconocimiento de AVD.

El presente capítulo se inicia con una propuesta arquitectura integradora de sensores ambientales, multimodales y de trazabilidad para el RA diarias. Esta arquitectura, concebida bajo un enfoque Edge-Fog, actúa como pilar para la integración permitiendo la eficiente recopilación, procesamiento y distribución de datos provenientes de diversas fuentes. Posteriormente, se detalla en la sección 3.2 la capacidad de sensorización de los sensores ambientales, fundamentales en la creación de un sistema de RA.

Uno de los enfoques innovadores de esta tesis es la integración de dispositivos multimedia, cuyo tratamiento y procesamiento requiere de modelos y diseño adecuados para el RA en EI.

Por una parte, se integran dispositivos de sonido donde las muestras de audio se convierten en una valiosa herramienta para descripción de eventos y AVD. En la sección 3.3 se explora cómo las técnicas de procesamiento de señales desentrañan eventos representativos de los usuarios en la vida cotidiana. Por otra parte, se ha ampliado la integración multimodal con dispositivos de visión. En relación al procesamiento de datos provenientes de sensores de visión, en la sección 3.4 se describen las cámaras térmicas, que habilitan la detección de los puntos de referencia del cuerpo y RA físicas. Este apartado propone una integración que identifica la pose de la persona o las actividades físicas que realiza y salvaguardando la privacidad de los individuos, ofreciendo así una solución integral y respetuosa en el ámbito del RA.

3.1. Arquitectura para la sensorización de hogares inteligentes

En esta sección, se expone la arquitectura diseñada para la integración de sensores multimodales de fuentes heterogéneas en el ámbito de hogares inteligentes. El propósito de esta arquitectura es facilitar una monitorización exhaustiva y precisa de usuarios y eventos domésticos en entornos reales. Los sensores, que constituyen elementos clave en la construcción de este sistema integrado, proporcionan la infraestructura esencial para el desarrollo de una solución tecnológica avanzada, caracterizada por su precisión y fiabilidad. Inicialmente concebido para el RA en entornos de alta privacidad (como actividades higiénicas del cuarto de baño) y la monitorización de usuarios [96, 97], este sistema ha evolucionado progresivamente hacia un sistema de RA en entornos de ocupación múltiple, protegido por derechos de propiedad intelectual bajo el nombre *Mercedes 2.0*. Este sistema integra los componentes necesarios para el seguimiento y la trazabilidad de actividades en espacios interiores. El modelo de discriminación utilizado en este sistema, que es clave para identificar y correlacionar actividades y usuarios, se detalla posteriormente en el capítulo 5.

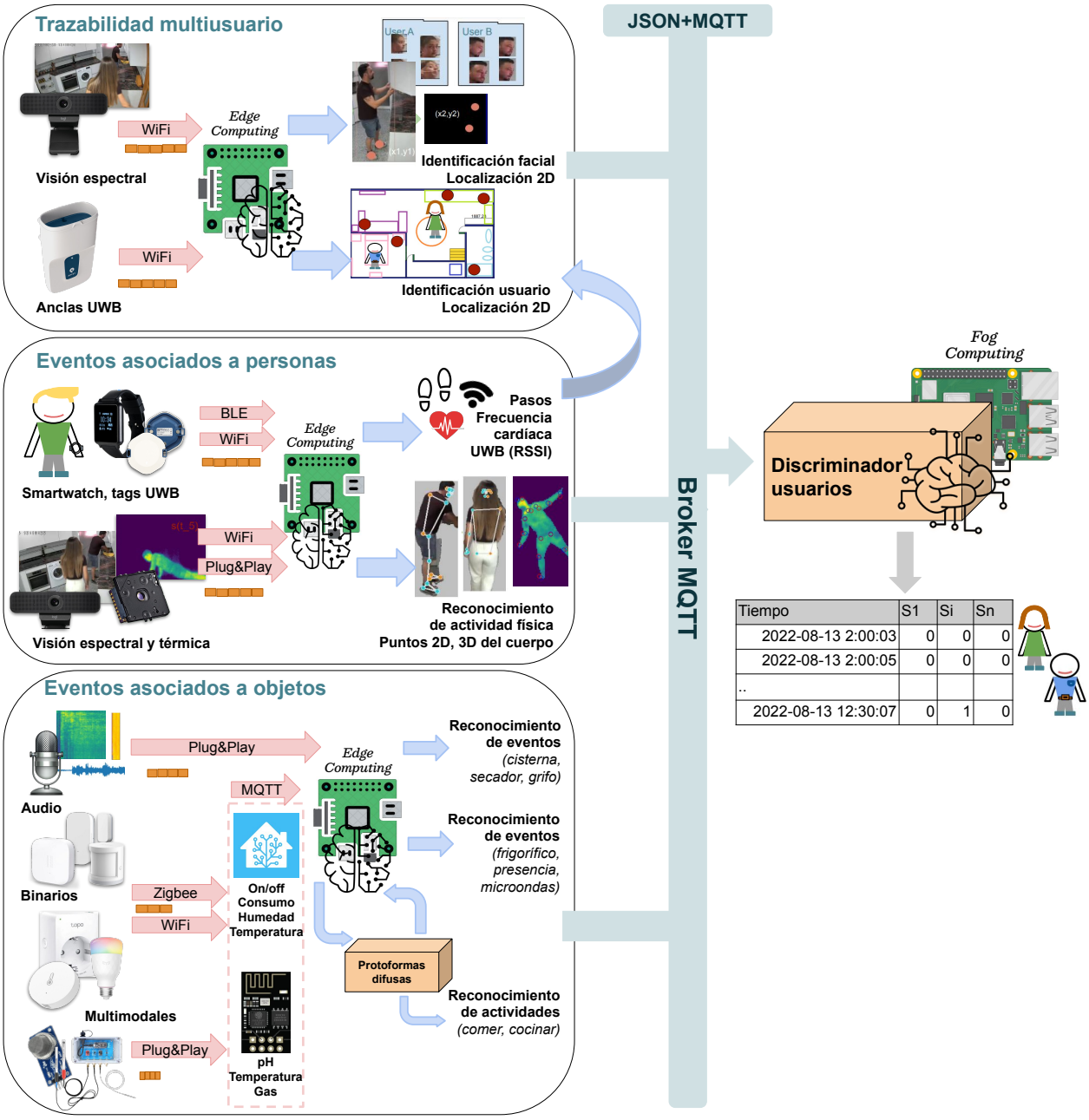


Figura 3.1: Arquitectura general que integra los componentes de la tesis doctoral: módulo de trazabilidad, módulo de eventos asociados a personas y módulo de eventos asociados a objetos

En la Figura 3.1, detallamos la arquitectura propuesta para la integración de los componentes de la tesis doctoral: módulo de trazabilidad, módulo de eventos asociados a personas y módulo de eventos asociados a objetos. Para comprender en profundidad cómo se inte-

gran y comunican los distintos componentes del sistema, es esencial distinguir la función que desempeñan cada uno de ellos:

- **Componente de trazabilidad multiusuario**: Este módulo se encarga de la localización y la identificación unívoca de sujetos dentro de un entorno de multiocupación. Para tal fin, se emplean tecnologías avanzadas, como sensores de visión o sistemas de localización UWB. Los datos procedentes de los sensores se transmiten vía WiFi, siendo recolectados y procesados en un nodo edge. Este enfoque prioriza la privacidad del usuario, distribuyendo únicamente la identificación del usuario y las coordenadas espaciales 2D (x, y). Ambas tecnologías han demostrado ser eficaces en contextos de trazabilidad en entornos de multiocupación, aunque presentan diferentes pros y contras.
  - Sensores de visión espectral: A pesar de que el procesamiento de datos se lleva a cabo en un nodo edge, la captura continua de imágenes puede resultar intrusiva, particularmente en entornos domésticos o clínicos. Para alcanzar una eficacia óptima, es crucial instalar un número significativo de cámaras, optimizando su ángulo de visión y configurando adecuadamente los parámetros para el procesamiento de coordenadas 2D mediante técnicas de homografía. El nodo edge requiere una capacidad de procesamiento considerable, incorporando modelos de reconocimiento facial avanzados como DeepFace. No obstante, el coste asociado a este tipo de instalación es relativamente bajo. En el capítulo 4.3, se detalla el uso del sensor de visión para este propósito.
  - Sensores de localización UWB: destacan por su facilidad de instalación y su enfoque en la privacidad. Las anclas UWB se distribuyen estratégicamente en el entorno, y el usuario porta un sensor vestible o etiqueta identificativa. La principal limitación de esta tecnología es su coste elevado, mitigable mediante la implementación de técnicas como *fingerprinting* que pueden ser computadas en un nodo edge. Esta metodología incrementa la precisión del modelo con un número menor de anclas reduciendo costes mediante un mapeo del entorno que asigna intensidades de señal a localizaciones concretas. En este escenario, las demandas de procesamiento del nodo edge son menores, dado que los modelos procesan intensidades de señal directamen-

te vinculadas a un usuario específico. Es crucial que el usuario porte el tag para garantizar la efectividad del sistema de localización. Una descripción exhaustiva del uso de los sensores de localización para este fin se encuentra en el capítulo 4.2.

- **Componente de detección de datos y eventos asociados a personas**: Este módulo se encarga de medir variables y reconocer actividades que dependen exclusivamente del usuario, sin necesidad de interactuar con el entorno. Los datos capturados por los sensores se transmiten mediante BLE, WiFi o directamente al nodo edge para su procesamiento y conversión en información interpretable, excluyendo datos sensibles. Posteriormente, esta información se envía a través del protocolo de publicación-suscripción. Las tecnologías que integran este módulo son:
  - Relojes inteligentes: Estos dispositivos transmiten información vía BLE, que es recogida por el nodo edge. Los datos incluyen parámetros como pasos dados, frecuencia cardíaca e intensidad del movimiento. El nodo edge procesa esta información para determinar estados de actividad o inactividad basándose en los pasos del usuario, o para reconocer la realización de ejercicio físico a través de la frecuencia cardíaca y el acelerómetro. La identificación del usuario se asocia directamente con la dirección MAC del dispositivo.
  - Tags UWB: Transmiten la información vía WiFi al nodo edge. Este componente se relaciona con el sistema de trazabilidad, ya que envía la intensidad de la señal del dispositivo a las anclas distribuidas en el entorno, proporcionando así una identificación directa. Por tanto, sólo se transmiten los datos de RSSI a cada ancla al bróker. Se aborda en detalle el uso de la tecnología UWB para este propósito en el capítulo 4.2.
  - Sensores de visión espectral: Estos sensores envían los datos a través de WiFi para su recopilación y procesamiento en el nodo edge. Los modelos de procesamiento permiten la detección de puntos 3D del cuerpo mediante modelos avanzados como YOLO integrados en la placa IoT, distribuyendo únicamente dicha información. El capítulo 4.3 proporciona una explicación detallada sobre el uso del sensor de visión espectral para este fin.

  - Sensores de visión térmica: conectados directamente al nodo edge, recoge y procesa la información recibida con alta privacidad. El procesamiento incluye la detección de puntos 2D del cuerpo utilizando modelos de aprendizaje profundo adaptados a imágenes térmicas, como versiones especializadas de YOLO, y la clasificación de ciertas actividades deportivas. Solo los puntos 2D del sujeto y/o la actividad física detectada se distribuyen en tiempo real. El capítulo 3.4 se enfoca en detallar el procesamiento de imágenes térmica para este propósito.

- **Componente de detección de eventos asociados a objetos**: Este módulo se centra en la detección de eventos y actividades que involucran la interacción del usuario con elementos del entorno. La diversidad de sensores integrados, provenientes de fuentes heterogéneas, hace esencial la centralización de datos en el nodo edge para su análisis y procesamiento. Este componente facilita el reconocimiento de interacciones con objetos clave, como electrodomésticos o luces, así como la identificación de actividades que implican la interacción con varios elementos. Dentro de este módulo se incluyen las siguientes tecnologías:

  - Sensores de audio: Capturan pistas de sonido y espectrogramas, conectándose directamente al nodo edge para el procesamiento de datos. El análisis se realiza mediante modelos de aprendizaje profundo enfocados en el reconocimiento de eventos sonoros específicos, como el sonido de una cisterna, un secador o un grifo. Solo el evento identificado se transmite al bróker. Detalles sobre el procesamiento del sensor de audio para este propósito se pueden encontrar en el capítulo 3.3.
  - Sensores ambientales (binarios y multimodales): Conectados mediante ZigBee o Wi-Fi a plataformas como Home Assistant, que actúan como gateways para la integración de diversos sensores comerciales. Estos sensores envían información sobre cambios de estado y mediciones (como la activación de una puerta o el consumo de un electrodoméstico) a través del protocolo de publicación-suscripción. Estas mediciones son luego procesadas en el nodo edge para identificar eventos específicos (como la apertura de un frigorífico) o para reconocer actividades más prolongadas que involucran varios sensores. Para el reconocimiento de estas últimas, se emplean

técnicas como protoformas difusas y reglas IF-THEN. La actividad o evento identificados, junto con el instante o período de tiempo correspondientes, se transmiten posteriormente al bróker y son recogidos por el nodo fog. Para una comprensión detallada de las especificaciones, aplicaciones y plataforma de integración de estos sensores, véase el capítulo 3.2.

- Sensores ad hoc no comerciales: Se conectan a placas IoT de menor capacidad de procesamiento, como la ESP8266, mediante pines. Estos sensores recopilan mediciones diversas, desde la calidad del agua (oxidación, pH) hasta la concentración de diferentes gases específicos. La información recogida se envía al nodo edge a través del protocolo de publicación suscripción, donde se procesa de manera similar a los sensores multimodales, ya sea para eventos puntuales o actividades de mayor duración. El capítulo 3.2 ofrece una visión completa sobre las especificaciones, las aplicaciones y la integración de estos sensores.

Tras la adquisición y procesamiento de datos provenientes de los sensores en el nodo edge, se procede de la publicación de la información reconocida en el entorno. En nuestro caso se ha integrado el protocolo de publicación-suscripción de Message Queuing Telemetry Transport (MQTT) donde uno de los nodos fog actúa como bróker. En este esquema es especialmente relevante el discriminador de usuarios (ver capítulo 5), como un nodo fog suscrito a los eventos del entorno y usuarios que recibe dicha información para su posterior análisis. Este discriminador sintetiza y correlaciona los datos emanados de los módulos de detección de eventos asociados tanto a personas como a objetos, junto con la información procedente del módulo de trazabilidad. El proceso establece una asociación precisa espacio temporal entre la ubicación del usuario y la zona donde se ha generado el evento, logrando así una discriminación inequívoca del individuo que ejecutó la actividad. Este mecanismo permite, por ejemplo, en situaciones de eventos puntuales como la apertura de una puerta, identificar con exactitud qué usuario ha sido el responsable. De manera análoga, en el contexto de eventos de larga duración, como la actividad de cocinar, el sistema es capaz de discernir el grado de participación de cada usuario en la actividad (activo, parcial o inactivo), así como cuantificar el tiempo empleado en la realización de dicha actividad.

La arquitectura propuesta proporciona un esquema eficiente para el manejo y procesamiento de información sensible. En esta estructura, la captura inicial de datos se efectúa a través de diversos sensores, cuyo procesamiento preliminar se lleva a cabo en el nodo edge donde está conectado. Esta etapa inicial de procesamiento es crucial, ya que permite una filtración y condensación de la información, asegurando que solo los datos relevantes sean transmitidos a etapas posteriores, reduciendo el ancho de banda y aportando alto grado de privacidad. Posteriormente, la información procesada se transfiere al nodo fog. Este nodo asume un rol fundamental en la estructura propuesta, actuando como un puente entre el procesamiento realizado en el nodo edge y la posible integración con sistemas de almacenamiento en la nube. Entre los beneficios de esta integración se encuentran la accesibilidad remota y la capacidad de compartir información con terceros, como profesionales de la salud, cuidadores, familiares o cualquier usuario debidamente autorizado. Estos usuarios pueden acceder a los datos mediante aplicaciones móviles o interfaces web específicamente diseñadas para tal propósito.

En conclusión, la arquitectura propuesta garantiza un procesamiento de datos eficiente, seguro y escalable. Facilita la interacción y el acceso remoto a la información de manera controlada y protegida, resultando ideal para aplicaciones en sectores como la salud, el bienestar y la monitorización en entornos domésticos y hospitalarios.

3.2. Sensores ambientales para el reconocimiento de actividades asociadas a objetos

En la vanguardia de la tecnología aplicada a EI, se ha logrado un significativo avance en la capacidad de monitorizar las actividades humanas de manera cada vez menos intrusiva, con el RA emergiendo como un área de investigación fundamental. En este contexto, los sensores multimodales representan una faceta crucial en la transformación de entornos convencionales en ecosistemas interactivos y adaptables. La utilización de sensores ambientales ha sido una elección recurrente para recopilar datos que describan las actividades cotidianas, permitiendo la supervisión de la interacción entre el individuo y su entorno [16]. Los avances tecnológicos

recientes han impulsado mejoras notables tanto en el hardware como en el software, con un énfasis en el aprendizaje automático supervisado [98]. Sus principales ventajas son las siguientes:

i) Facilidad de instalación: Estos sensores suelen adherirse a muebles o electrodomésticos mediante pegatinas o adhesivos, y cuentan con una autonomía de batería propia, lo que prescinde de la necesidad de instalaciones eléctricas.

ii) Tamaño reducido: Dado que deben integrarse como parte de los objetos en el entorno, suelen ser de dimensiones pequeñas.

iii) Coste económico: La activación binaria de estos sensores no requiere componentes electrónicos complejos, lo que los convierte en una opción de bajo coste.

iv) Mínima invasividad: En comparación con dispositivos como cámaras y micrófonos, estos sensores binarios presentan un nivel mínimo de intrusión [14].

3.2.1. Selección de dispositivos ambientales

En esta sección se incluye una descripción de diferentes dispositivos ambientales que pueden ser integrados en el sistema *Mercedes y Mercedes 2.0*, propiedad intelectual precursora de la plataforma descrita en esta tesis. Se ha optado por una selección de dispositivos ambientales compatibles con la plataforma HA (ver subsección 3.2.2), caracterizados por su bajo coste y su capacidad para describir actividades y vincularse con el entorno de las personas. Indicamos que dichos dispositivos son compatibles con la arquitectura y plataforma propuesta en la tesis doctoral y se han integrado para diferentes propósitos en trabajos de investigación relacionados [97, 99, 92].

1. **Detección de presencia Xiaomi Mi/Aqara Motion Sensor.** Este dispositivo posibilita la descripción de la actividad humana sin afectar la privacidad ni la comodidad del usuario, dado que emplea sensores de infrarrojos pasivos Sensor Infrarrojo Pasivo (Passive Infrared Sensor, por sus siglas en inglés) (PIR) para detectar el movimiento.

2. **Apertura/Cierre de puertas Aqara Door/Window Sensor.** Estos sensores desempeñan un papel crucial en el seguimiento de la interacción entre individuos y objetos cotidianos. Permiten la identificación de los objetos de mayor interés y facilitan la comprensión de comportamientos compulsivos, como la repetida verificación de puertas, común en personas dependientes. Generalmente, se instalan en electrodomésticos clave para el RA, como se ejemplifica en el caso de estudio del Capítulo 5 que se centra en la cocina. Asimismo, son fundamentales para determinar los momentos en que una persona entra o sale de su vivienda.

3. **Detección de agua Aqara Water Leak Sensor.** Para una evaluación exhaustiva de la higiene, resulta fundamental tener conocimiento sobre actividades como el uso del inodoro, las duchas o la apertura de grifos. Además, estos dispositivos desempeñan un papel crucial en la detección de posibles fugas de agua.

4. **Bombilla inteligente como dispositivo actuador Xiaomi Yeelight LED Bulb ll.** Tiene la función de alertar a los usuarios sobre posibles riesgos mediante estímulos visuales, así como de establecer automatizaciones para apagar las luces olvidadas al salir de casa de manera autónoma.

5. **Sonido/altavoces e interacción de voz Google Home Mini.** El sonido, como actuador, juega un papel fundamental en un EI. Mantener un ambiente sonoro tranquilo y relajante resulta crucial para el bienestar de las personas dependientes, teniendo un impacto directo en su estado de ánimo. Además, se pueden establecer rutinas o recordatorios, como la toma de medicación, y facilitar la comunicación con un agente conversacional para resolver dudas específicas.

6. **Sensor de temperatura y humedad Aqara.** Este sensor recopila variables ambientales relevantes, como temperatura, humedad y presión atmosférica. Teniendo en cuenta los factores ambientales que se ven afectados de forma inherente, el aumento de la humedad en un baño se produce por una alta concentración de vapor de agua caliente respecto a las condiciones estándar. Este hecho puede ser asociado con alta probabilidad a la actividad de la ducha, ya que provoca la condensación de una gran cantidad de

vapor cálido, que se impregna en paredes y suelos. Del mismo modo, se puede identificar la activación de la vitrocerámica a través del aumento de la temperatura y la humedad del ambiente próximo al sensor. Estos detalles son fundamentales para el reconocimiento de AVD y para la regulación de la temperatura en el interior de la vivienda.

Se incluye una representación visual de los sensores comerciales detallados en el listado previo, mostrada en la Figura 3.2.

Figura 3.2: Sensores y actuadores comerciales integrados en el sistema *Mercedes* a través de HA.

En referencia a sensores no comerciales o sin integración directa a través de HA, se pueden resaltar tres sensores utilizados para la detección de actividades vinculadas a la higiene:

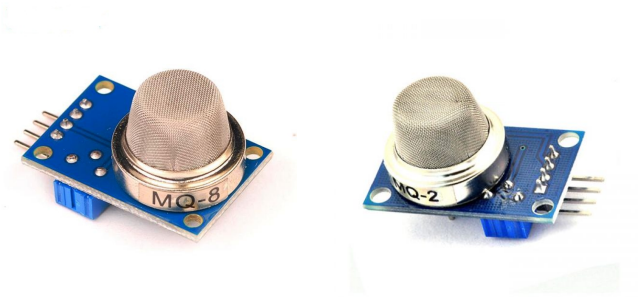
1. **Sensores de gas MQ.** En los gases intestinales humanos, se ha observado que la concentración de hidrógeno (H2) se sitúa en aproximadamente un 20,9 %, mientras que el contenido de metano (CH4) suele rondar un 7,2 %. Normalmente, en el aire, la concentra-

ción de estos gases se encuentra en niveles inferiores al 1 %. No obstante, investigaciones previas han demostrado la factibilidad de emplear la concentración de estos dos compuestos para detectar la actividad de defecación.

a) **Sensor MQ-8.** Un sensor apropiado para detectar concentraciones de hidrógeno en el aire. Este sensor es capaz de detectar concentraciones de gas hidrógeno en un rango de 100 a 10,000 partículas por millón.

b) **Sensor MQ-2.** Este sensor se clasifica como un sensor de gas combustible y tiene la capacidad de detectar gas metano en un rango de 300 a 10,000 partículas por millón.

El material de sensorización utilizado en estos sensores es el SNO_2, cuya conductividad es inferior en un entorno limpio y mayor en presencia de gases específicos. La figura 3.3 muestra una imagen de los mismos.



(a) MQ-8 (b) MQ-2

Figura 3.3: Sensores de detección de gases específicos.

2. **Sensores de calidad del agua.** Este kit de sensores de la empresa Atlas Scientific denominado "Wi-Fi Pool Kit Quality Monitoring"(https://atlas-scientific.com/kits/wi-fi-pool-kit/) incluye sensores de pH, ORP (Oxidation Reduction Potential asociada a la suciedad del agua) y temperatura. El dispositivo opera a 5V con un consumo de corriente de 280 mA a 5V y está clasificado como IP64, garantizando un desempeño robusto contra el polvo y la resistencia al agua en inmersiones. Incluye una placa ESP8266 como controlador de sensores, permitiendo el desarrollo de soluciones embebidas para la monitorización del agua. Esta placa proporciona conectividad WiFi para la transferencia

de datos en tiempo real a través de TCP, HTTP o MQTT. Atlas Scientific proporciona el código y las instrucciones necesarias para iniciar el proyecto desde cero. Además, el dispositivo consta de varios circuitos y sondas:

- Circuito de pH: determina la acidez/alcalinidad del agua con una resolución de 0.01 de acidez.
- Circuito de Potencial de Oxidación-Reducción (ORP): su rango es de +/-1100 mV.
- Circuito de temperatura de resistencia (RTD).
- Sonda de pH, sonda de ORP y sonda de temperatura PT-1000.

En la figura 3.4, describimos el dispositivo Wi-Fi Pool Quality Monitoring y detallamos los componentes que lo integran:

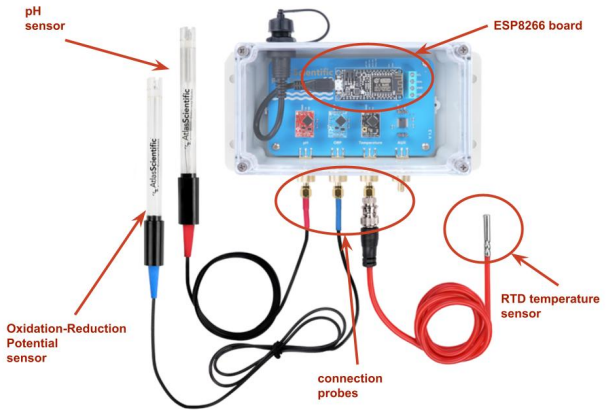


Figura 3.4: Wi-Fi Pool Kit Quality Monitoring. En círculos rojos se detallan los principales componentes: Placa ESP8266, sondas de conexión y sensores RTD, pH y ORP.

La tasa de recogida predeterminada es configurada por defecto, con un tiempo de retardo para la actualización de las mediciones establecido en 1 segundo. Sin embargo, en la práctica, para la lectura, cálculo y estimación de valores posteriores a la calibración de referencia, se obtiene una tasa real de muestreo de una medida de los 3 sensores cada 3 segundos.

3.2.2. Plataforma de integración de sensores multimodales

HA es una plataforma de automatización del hogar de código abierto que se centra en la privacidad y la localidad de los datos, permitiendo a los usuarios mantener el control total sobre su información sin depender de soluciones basadas en la nube. Posibilita la integración y control de una amplia variedad de dispositivos y servicios inteligentes desde una única interfaz. Algunas de sus características clave son:

- Interoperabilidad: Home Assistant es compatible con miles de dispositivos y servicios, lo que permite a los usuarios integrar elementos de diferentes marcas y tecnologías en un sistema unificado.
- Automatización y Escenas: Los usuarios pueden crear automatizaciones y escenas personalizadas que les permiten programar y controlar dispositivos en función de eventos específicos, horarios o condiciones.
- Interfaz de usuario personalizable: Ofrece una interfaz de usuario configurable que se puede adaptar a las necesidades individuales, permitiendo a los usuarios crear paneles de control personalizados.
- Control de voz y asistentes virtuales: Se puede integrar con asistentes de voz como Google Assistant y Amazon Alexa, lo que permite el control de dispositivos mediante comandos de voz.
- Código abierto y flexible: Como una plataforma de código abierto, Home Assistant permite a los desarrolladores contribuir al proyecto y expandir sus capacidades. También es flexible, lo que significa que se puede ejecutar en una variedad de hardware, desde dispositivos de bajo costo como Raspberry Pi hasta servidores más potentes.
- Privacidad y funcionamiento local: A diferencia de muchas otras soluciones de domótica, Home Assistant funciona principalmente de manera local, sin depender de la nube para la mayoría de sus funciones, lo que mejora la privacidad y la velocidad de respuesta.

- Comunidad activa: Tiene una gran comunidad de usuarios y desarrolladores que constantemente contribuyen con nuevas integraciones, soporte y mejoras.

Ejecutada en Python 3, permite automatizar el rastreo, seguimiento y control del comportamiento de los dispositivos del hogar. Utiliza un sistema basado en reglas que cuenta con tres elementos indispensables: Trigger (desencadenante), Condition (condición) y Action (acción). Aunque inicialmente pueda parecer complejo, esta plataforma ofrece múltiples plantillas y ejemplos completamente funcionales que permiten ampliar y personalizar las automatizaciones. Además, brinda la posibilidad de utilizar scripts en Python para usuarios que requieran opciones más avanzadas. La accesibilidad, documentación exhaustiva y su interfaz adaptable basada en Material Design son aspectos destacados de HA. Su interfaz web intuitiva y receptiva a WebSockets posibilita la visualización de datos en tiempo real sin necesidad de recargar la página. Adicionalmente, ofrece una interfaz sencilla para dispositivos Android.

La selección de esta plataforma para la integración de los sensores se ha fundamentado principalmente en la facilidad de instalación independientemente del fabricante y protocolo empleado por el dispositivo. Esto permite una integración heterogénea y transparente para el usuario. En el caso de dispositivos que utilizan el protocolo ZigBee como son los de Aqara y Xiaomi, se ha hecho uso del USB-gateway Conbee II (ver figura 3.5) y el complemento *"deCONZ"* disponible en la plataforma. Cuando se trata de la integración de dispositivos Wi-Fi, el proceso es más directo en comparación con los dispositivos ZigBee. A través de la función de *"discovery"* o descubrimiento, HA escanea los dispositivos presentes en su alcance, permitiendo su control en cuestión de minutos.

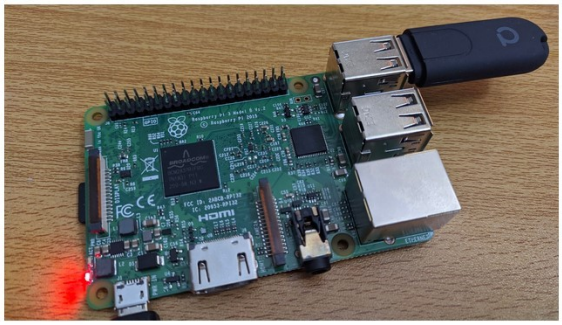


Figura 3.5: Raspberry Pi y Conbee ll para integración de sensores con protocolo ZigBee en HA.

Para la transferencia de datos en el diseño del sistema *Mercedes*, se ha optado por el uso de MQTT, integrado en HA como base. Esta elección se fundamenta en la naturaleza bidireccional del protocolo, permitiendo la publicación de datos para procesamiento externo y la suscripción a temas específicos para la captura de información. Esta versatilidad es estratégica, posibilitando tanto la transmisión de datos desde los sensores a la plataforma como la recepción y registro de información relevante de fuentes o dispositivos externos. La arquitectura liviana y eficiente de MQTT en la transferencia de mensajes es crucial, especialmente en entornos con limitaciones de ancho de banda o en dispositivos con recursos limitados, como los sensores. La estabilidad y fiabilidad en la comunicación, junto con su extensa adopción, lo convierten en una opción idónea para asegurar la conectividad, la eficacia en la transmisión de datos y la integración sin inconvenientes de diversos dispositivos en el marco del sistema de monitorización y RA implementado.

3.3. Procesamiento de sensores de audio para el reconocimiento de actividades asociadas a objetos

La disciplina del RA ha cobrado prominencia como un tema activo de investigación [100], centrándose en la detección de comportamientos humanos en EI [101]. El RA ha encontrado aplicación en hogares inteligentes [102] con el propósito de mejorar la calidad de vida y promover la independencia de las personas en sus hogares [103].

Los enfoques iniciales se basaron principalmente en sensores binarios que describen AVD de manera no invasiva. No obstante, surgió una nueva generación de dispositivos que ofrecen una perspectiva más rica, entre los que destacan: (i) dispositivos portátiles, empleados para analizar actividades y gestos [104]; (ii) dispositivos de localización, que en la actualidad alcanzan una precisión aceptable en contextos interiores [105]; (iii) sensores de visión (visibles o infrarrojos) en secuencias de video e imágenes [106] y (iv) sensores de audio [24] que reconocen eventos basados en información acústica. Este avance fue acompañado por la tendencia a emplear sensores multimodales, lo que permitió la utilización de tecnologías de detección de uso general para monitorizar actividades.

En el campo del reconocimiento de audio, la combinación de CNN [107] con la utilización de espectrogramas para representar el sonido [108] ha mostrado resultados muy positivos. Este enfoque se ha aplicado con éxito en la clasificación de sonidos ambientales [109, 110, 111] y en el análisis de señales musicales [112, 113]. Específicamente, el uso del Espectrograma Log-Mel (LM) y el Coeficientes Cepstrales de Frecuencia Mel (Mel Frequency Cepstral Coefficients, por sus siglas en inglés) (MFCC) han sido destacados como representaciones robustas para la clasificación de sonidos [114, 115].

En cuanto al reconocimiento de sonidos ambientales en interiores, la literatura científica presenta diversas aproximaciones. Beltrán et al. [116] abordaron la identificación de eventos como el rebote de una pelota o el canto de un grillo, utilizando la representación espectral del sonido y características a nivel de trama aprendidas mediante modelos de Markov. En un estudio posterior del mismo autor [117], se logró la clasificación de categorías de sonidos (golpes y lavado de manos) usando representaciones espectrales e histogramas sonoros, empleando SVM en un entorno residencial geriátrico. Otro estudio, realizado por Cruz et al. [24], utilizó micrófonos direccionales espaciales 3D para la captura de audio multidireccional de alta calidad, lo que permitió detectar eventos y localizar sonidos en un entorno específico. Los investigadores calcularon coeficientes cepstrales de frecuencia de Mel como características espaciales, utilizando modelos Gaussianos y modelos de Markov ocultos. Por otra parte, Laput et al. [118] recopilaron un conjunto de 30 eventos para reconocer actividades realizadas en siete espacios diferentes: baño, dormitorio, entrada, cocina, oficina, exterior y taller.

En base a estos trabajos previos basados en sonido, la presente sección describe una novedosa propuesta basada en el **reconocimiento de eventos cotidianos a través de dispositivos de sonido ambiental y la implementación de modelos de DL integrados en placas IoT**. El contenido se sintetiza en los siguientes puntos:

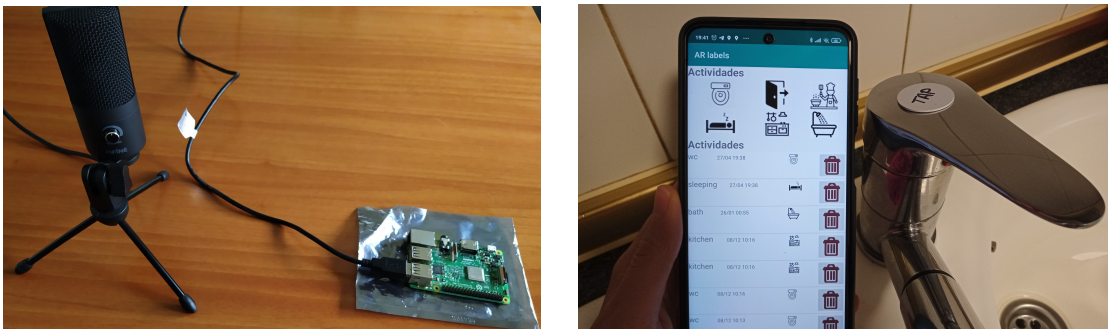
- Compilación de un conjunto de datos de muestras de audio de eventos relacionados con AVD, generados en entornos interiores;
- Integración de una arquitectura edge-fog con placas IoT donde se capturan las muestras de audio para lograr un reconocimiento en tiempo real de eventos acústicos;

- Evaluación del rendimiento de los modelos de DL para el reconocimiento offline y en tiempo real de eventos cotidianos en condiciones naturalistas.

3.3.1. Metodología

En esta sección, se expone la propuesta de dispositivos, arquitectura y métodos para el reconocimiento de sonidos ambientales correspondientes a eventos cotidianos, empleando placas inteligentes y CNN. En primer lugar, se detalle la configuración de la placa IoT y los sensores de audio en un enfoque de arquitectura edge-fog, destinado a la captura y etiquetado de sonidos presentes en el entorno. Posteriormente, se presenta un modelo de DL diseñado para el reconocimiento de los sonidos ambientales relacionados con los eventos diarios, mediante la utilización de un LM y CNN.

En el contexto del IoT y la computación ubicua, la integración de dispositivos en los espacios donde se recopilan los datos se caracteriza por su inmersión y baja invasividad. Para abordar esto, se ha implementado un enfoque de computación edge-fog con el propósito de evitar la transmisión de datos sensibles, y en su lugar, transmitir únicamente los datos procesados. En una primera instancia, se emplean sensores de audio que están conectados a placas inteligentes, con el fin de llevar a cabo la recolección y el reconocimiento de eventos sonoros. La placa inteligente elegida para esta tarea ha sido la Raspberry Pi [119], la cual habilita capacidades de cómputo adecuadas para el aprendizaje automático, incluyendo la implementación de modelos de DL [120]. Los sensores de audio integrados consisten en micrófonos de bajo coste con un conector USB, permitiendo la conectividad plug-and-play con la Raspberry Pi bajo el sistema operativo Raspbian. En la Figura 3.6 se ilustran ambos dispositivos conectados y dispuestos en un cuarto de baño.



(a) (b)

Figura 3.6: (**a**) Raspberry Pi B+ con micrófono USB para la recolección y reconocimiento de eventos de sonido ambiental (**b**) Aplicación móvil y etiqueta NFC para el etiquetado de eventos.

Los objetivos de la integración de sensores de audio en placas inteligentes para el reconocimiento de eventos cotidianos son los siguientes:

i) Recolectar muestras de sonido con el propósito de crear un dataset de entrenamiento.

ii) Entrenar modelos de DL utilizando las muestras de sonido previamente etiquetadas.

iii) Realizar el reconocimiento en tiempo real de eventos de audio, con el fin de evaluar el desempeño de los modelos entrenados.

La aplicación implementada en la Raspberry Pi se ha programado en Python [121], mientras que los modelos de DL se han desarrollado utilizando Keras, una biblioteca de código abierto diseñada para redes neuronales [122]. Los servicios remotos destinados a etiquetar los datos y transmitir las salidas reconocidas de los eventos de audio en tiempo real se han implementado utilizando MQTT, un protocolo de publicación/suscripción utilizado en redes inalámbricas de sensores [123]. Esta estrategia se ha inspirado en los paradigmas de la computación edge-fog [124].

Para llevar a cabo la recolección y etiquetado de muestras de sonido en EI, la Raspberry Pi adquiere muestras de sonido de una duración determinada en tiempo real. Asimismo, la placa Raspberry Pi se suscribe a un topic MQTT donde se publican los inicios y finales de cada evento para etiquetar un determinado evento de sonido específico desde una aplicación móvil creada para ese propósito. Durante el intervalo de tiempo entre el inicio y el fin, la placa

almacena las muestras de sonido, asignando a cada instancia una etiqueta. La aplicación móvil utilizada para el etiquetado de las muestras de sonido se ha desarrollado en Android [125], proporcionando una herramienta para etiquetar eventos en un dispositivo portátil. Con el fin de simplificar la tarea de etiquetado mientras se realizan las actividades diarias, se han colocado etiquetas Comunicación de Campo Cercano (Near Field Communication, por sus siglas en inglés) (NFC) en los objetos y muebles relacionados con los eventos, como puertas o grifos. Estas etiquetas NFC activan automáticamente el etiquetado en la aplicación móvil cuando el usuario las toca con el dispositivo, enviando el inicio y el final de una etiqueta de sonido a través de MQTT. En la Figura 3.6, se ilustran las etiquetas NFC y la aplicación móvil utilizada para el etiquetado de eventos de sonido.

Finalmente, el modelo de reconocimiento de eventos de sonido se entrena con los datos etiquetados y, posteriormente, es capaz de reconocer sonidos en tiempo real. Con este propósito, el modelo de reconocimiento de sonido descrito en la Sección 3.3.1.1 se entrena previamente y se almacena en la Raspberry Pi. El modelo recibe segmentos de muestras de audio provenientes del sensor como entrada y los clasifica de acuerdo con las etiquetas objetivo. La predicción para cada etiqueta objetivo se publica en tiempo real mediante MQTT para que otros dispositivos inteligentes o modelos de RA puedan acceder a ella. La arquitectura se detalla en la Figura 3.7.

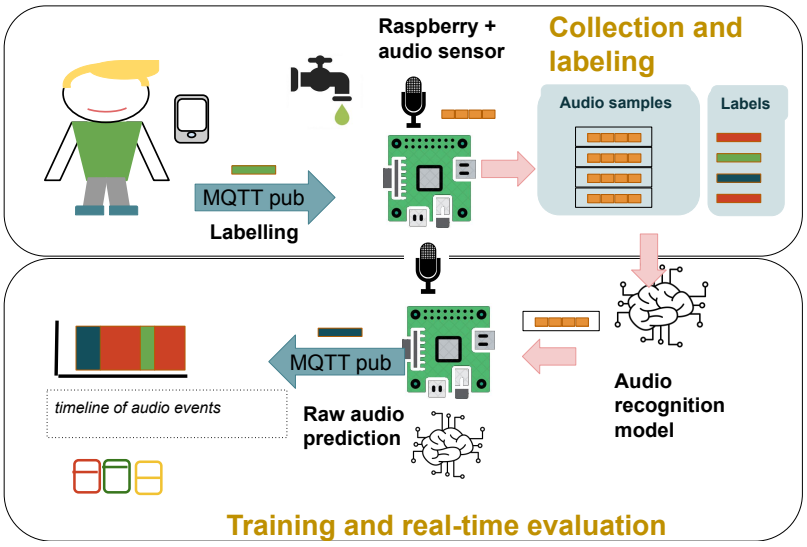


Figura 3.7: Arquitectura de componentes para el reconocimiento de sonidos ambientales de eventos cotidianos.

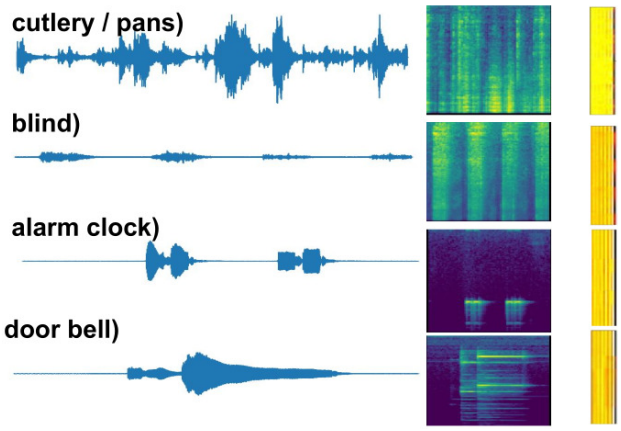
3.3.1.1. Configuración del modelo de DL

En esta sección, describimos un modelo de clasificación diseñado para el reconocimiento de sonidos ambientales en eventos cotidianos, basado en la representación espectral. Inicialmente, y como se mencionó anteriormente, la transformación de las muestras de audio digital unidimensionales en una representación espacial bidimensional mediante características de espectrograma (una ïmagen"del sonido) ha demostrado ser eficaz en la clasificación de audio ambiental [108]. En este estudio, se ha establecido un tamaño de ventana de 3 segundos para segmentar y recolectar muestras de audio ambiental, ya que este intervalo resulta apropiado para el reconocimiento de sonidos [109]. La frecuencia de muestreo del sensor de audio ambiental se ha configurado a 44.1 kHz. Posteriormente, se extraen dos representaciones espectrales de cada muestra de sonido, las cuales son evaluadas como entradas por distintas CNNs:

- El LM se calcula a partir de la representación tiempo-frecuencia de las señales de audio, empleando un espectro de potencia logarítmico en una escala de frecuencia Mel no lineal. Se ha establecido la longitud de la ventana para la transformada de Fourier rápida para generan imágenes de tamaño 128×130.

- Los MFCC escalados en logaritmo, con 13 componentes extraídos directamente de las señales de audio en bruto. Estos coeficientes se calculan utilizando una transformada coseno lineal de un espectro de potencia logarítmico en una escala de frecuencia Mel no lineal [126]. Dado que los MFCC tradicionales suelen emplear entre 8 y 13 coeficientes cepstrales [127], en este estudio se ha optado por 13 características para capturar la información más representativa de las muestras de audio. Con esta configuración, el espectrograma MFCC resultante de las frecuencias positivas produce imágenes de tamaño 13×130.

En la Figura 3.8, se presentan los MFCC de las muestras de audio recogidas, que luego se utilizan como entrada para su clasificación con las etiquetas de sonido correspondientes mediante una CNN.



.

Figura 3.8: Ejemplo de señales de audio en bruto a 44.1 kHz, LM y MFCC de los eventos de audio ambiental: cubiertos, persiana, despertador y timbre.

El enfoque de las CNN se ha destacado por su eficacia tanto en la extracción de características como en la clasificación dentro del campo del reconocimiento de imágenes, según se indica en investigaciones previas [128]. En el ámbito del reconocimiento de audio ambiental, se han propuesto diversos modelos de CNN que incluyen múltiples capas para la extracción de características [114, 109]. Considerando la representación espectral de los sonidos, en este estudio se examinan dos modelos específicos de CNN:

i) Un modelo de CNN que utiliza cinco capas convolucionales para procesar los MFCC. Este modelo incorpora una capa de agrupamiento promedio tras las operaciones convolucionales, ya que las entradas tienen dimensiones reducidas (13×130×1). Un segundo modelo de CNN, que consta de cinco capas convolucionales seguidas de capas de agrupamiento máximo para reducir el tamaño de entrada (128×130×1). Las configuraciones detalladas de este modelo están especificadas en la Tabla 3.1.

Tabla 3.1: Arquitectura del modelo CNN+MFCC y CNN+LM

Arquitectura del modelo CNN+MFCC	
Input	13×130×1
Conv(3×3)	11×128×16
Conv(3×3)	9×126×16
Conv(3×3)	7×124×32
Conv(3×3)	5×122×64
Conv(3×3)	3×120×128
Conv(3×3)	1×118×256
GlobalAvgPool2D	256
Dense	1024
Dense	15
Arquitectura del modelo CNN+LM	
Input	128×130×1
Conv(2×2)	127×129×16
Max-Pool(2×2)	63×64×16
Conv(2×2)	62×63×32
Max-Pool(2×2)	31×31×32
Conv(2×2)	30×30×64
Max-Pool(2×2)	15×15×64
Conv(2×2)	14×14×128
Conv(2×2)	13×13×128
Flatten	21632
Dense	1024
Dense	1024
Dense	15

Los modelos mencionados han sido desarrollados mediante la biblioteca Keras en el lenguaje de programación Python. Esta elección permite la compatibilidad con la Raspberry Pi, facilitando la integración en tiempo real. Dicha integración se ajusta a un enfoque de edge computing, que consiste en la publicación de los eventos detectados sin exponer información sensible proveniente del sensor de audio ubicado en los hogares. De esta manera, se preserva la privacidad del habitante.

3.3.2. Resultados

En esta sección, se presentan los resultados obtenidos en el reconocimiento de eventos de sonido. Se comienza con la descripción de un conjunto de datos compuesto por sonidos ambientales recogidos en un entorno doméstico. Este conjunto se utiliza para evaluar la metodología propuesta, tanto en contextos offline como en tiempo real, a través de varios casos de estudio. Los datos se han recopilado en una vivienda con cuatro habitaciones (sala de estar, dormitorio, cocina y baño), habitada por una persona.

Para comenzar la evaluación, se ha creado un conjunto de datos específico, formado por sonidos ambientales típicos de las AVD. Los eventos/actividades seleccionados para su reconocimiento en los casos de estudio se describen detalladamente en la Tabla 3.2. Para cada categoría, se ha conseguido reunir un conjunto balanceado, compuesto por 100 muestras de sonido de 3 segundos de duración, recogidas en condiciones naturalistas. En cuanto a la categorización de los eventos, se ha desarrollado una aplicación móvil, descrita en la Sección 3.3.1, para determinar el inicio y el final de cada evento.

Tabla 3.2: Eventos de sonido desarrollados en el caso de estudio.

Clase	Descripción
Vaccum cleaner	Muestra de audio de aspiradora
Tank	Muestra de audio de descarga de inodoro
Cutlery + pans	Muestra de audio de cubertería y sartenes
Alarm clock	Muestra de audio del sonido de un despertador
Shower	Muestra de audio de una ducha
Extractor	Muestra de audio de un extractor de cocina
Kitchen tap	Muestra de audio de un grifo de cocina
Bathroom tap	Muestra de audio de un grifo de baño
Printer	Muestra de audio de una impresora en funcionamiento
Microwave	Muestra de audio de un microondas en funcionamiento
Blind	Muestra de audio de una persiana
Door	Muestra de audio de una puerta abriéndose o cerrándose
Phone	Muestra de audio de un teléfono sonando
Doorbell	Muestra de audio de un timbre de una puerta sonando

En la fase inicial de evaluación, se ha empleado una técnica de validación cruzada para probar las capacidades del modelo de reconocimiento de audio en un entorno offline, realizando una segmentación explícita de las muestras de audio con ventanas de tiempo de 3 segundos. Posteriormente, el enfoque se somete a una evaluación en tiempo real en cuatro escenarios distintos. En estos escenarios, se han recopilado muestras de audio mientras el individuo realizaba sus AVD en condiciones naturalistas. Cada caso de estudio cubre un período de 2220 segundos, sumando un total de 760 muestras sujetas a análisis.

El conjunto de datos con las muestras de audio recogidas en este estudio, así como las etiquetas correspondientes a cada escena, están disponibles en el siguiente repositorio: https://github.com/AuroraPR/Ambiental-Sound-Recognition. En este repositorio, también se incluye la implementación de los métodos propuestos, realizada con Python y la biblioteca Keras.

3.3.2.1. Evaluación del caso de estudio modo offline

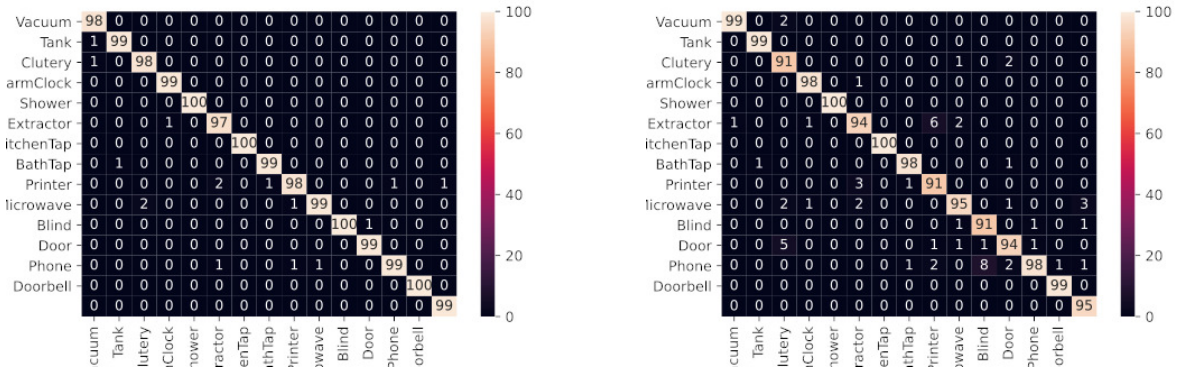
En esta sección, se presentan los resultados derivados de la aplicación de modelos de DL basados en CNN y las representaciones espectrales LM y MFCC, para el reconocimiento de sonidos ambientales. Estos resultados se han obtenido mediante la evaluación de los datos recopilados en el estudio empleando una estrategia de validación cruzada de 10 pliegues (10-fold cross-validation).

- **Conjunto de Datos de Audio Ambiental Ad Hoc:** Este conjunto de datos incorpora muestras de audio capturadas en un entorno doméstico singular, bajo condiciones controladas y etiquetadas mediante segmentación explícita de 3 segundos, como se detalló en la Sección 3.3.1. Cabe destacar que todas las clases descritas en la Tabla 3.2 están representadas en este conjunto de datos.

Para el conjunto de datos mencionado, se muestran las matrices de confusión obtenidas a partir de cada pliegue de la validación cruzada. En la Figura 3.9, se presentan los resultados de desempeño de los modelos de DL en el reconocimiento de sonidos ambientales del conjunto de datos ad hoc. Por otro lado, en la Tabla 3.3, se detallan las métricas de F1-score, precisión

y recall correspondientes a los modelos de DL y el conjunto de datos evaluado.

Es evidente que el conjunto de datos de audio ambiente ad hoc arroja resultados excepcionales en condiciones controladas para ambos modelos de CNN, destacándose especialmente la versión CNN+MFCC.



(a) CNN+MFCC (b) CNN+LM

Figura 3.9: Matrices de confusión en el conjunto de datos de audio ambiente ad hoc.

Tabla 3.3: Métricas de clasificación de la evaluación del caso de estudio en modo offline.

Modelo	Accuracy	Precision	Recall	F1-Score
CNN+MFCC (ad hoc dataset)	0.99	0.99	0.99	0.99
CNN+LM (ad hoc dataset)	0.96	0.96	0.96	0.96

Como se puede inferir, la recolección de un conjunto de datos de audio resulta altamente aconsejable, dadas las notables deficiencias en calidad y cantidad de las muestras extraídas de fuentes heterogéneas. A partir del conjunto de datos de audio creado en el contexto ad hoc, se han efectuado mediciones que abarcan el número de parámetros susceptibles de entrenamiento, el período de tiempo requerido para el proceso de aprendizaje, el conteo de millones de instrucciones (hasta 40 épocas) y el intervalo temporal demandado para la evaluación. Estos registros se han efectuado en una Raspberry Pi 3B que opera a una frecuencia de 400 MHz, y los resultados se presentan de manera concisa en la Tabla 3.4.

Tabla 3.4: Parámetros entrenables, tiempo de aprendizaje, millones de instrucciones (MI) y tiempo de evaluación.

Modelo	Parámetros entrenables	Tiempo aprendizaje	Millones instrucciones (MI)	Tiempo evaluación
CNN+MFCC	1.7 M	96 min	230.4×10^3 MI	2.53 s
CNN+LM	23.3 M	207 min	496.8×10^3 MI	2.81 s

Basándonos en estos resultados, la siguiente sección se adentra en la evaluación en tiempo real, haciendo uso de la configuración más óptima con el conjunto de datos de audio ambiente ad hoc. Dicha evaluación se lleva a cabo mediante el empleo del modelo sustentado en la combinación de CNN+MFCC, el cual asimismo demanda recursos computacionales más reducidos para la fase de aprendizaje y para la evaluación del proceso de reconocimiento de audio.

3.3.2.2. Evaluación del caso de estudio en tiempo real

A continuación, se presentan los resultados derivados de la evaluación en condiciones naturalistas llevada a cabo en cuatro escenarios distintos dentro de un hogar. Esta evaluación ha sido realizada utilizando el modelo CNN+MFCC, el cual ha sido entrenado con el conjunto de datos de audio ambiente ad hoc. Las seis secuencias de actividades comprenden:

- (Escena 1) El residente llega a casa, se dirige a la cocina y comienza a hablar. Luego utiliza los cubiertos, enciende el extractor durante un período prolongado, abre el grifo, enciende el microondas y recibe una llamada telefónica.
- (Escena 2) El residente llega a casa, se dirige a la sala de estar y comienza a hablar. Luego utiliza una aspiradora, abre y cierra las persianas, y recibe una llamada telefónica.
- (Escena 3) El residente llega a casa, se dirige al dormitorio y comienza a hablar. Luego utiliza una aspiradora, suena la alarma durante un tiempo prolongado, imprime algunos documentos y finalmente, abre y cierra las persianas.
- (Escena 4) El residente se dirige al baño y comienza a hablar. Luego abre el grifo, toma una ducha durante un período prolongado, utiliza la aspiradora y, finalmente, tira de la cadena del inodoro.

- (Escena 5) El residente habla en la cocina, luego utiliza una aspiradora en la sala, vuelve a hablar y utiliza cubiertos. Después, abre y cierra las persianas, abre el grifo y, finalmente, utiliza el microondas.

- (Escena 6) El residente se encuentra en el baño utilizando la aspiradora y comienza a hablar. Luego se ducha durante un tiempo prolongado, recibe una llamada telefónica y abre el grifo; finalmente, sale de la habitación cerrando la puerta.

En este contexto, se ha introducido una nueva categoría llamada *inactivo (idle)*, cuyo objetivo es identificar la actividad asociada a la no ocurrencia de eventos de interés. Esta categoría incluye situaciones de silencio y otros sonidos ambientales producidos por el residente. La adición de la etiqueta *inactivo* es crucial para el aprendizaje de la RA en tiempo real [129, 43]. Para realizar una evaluación precisa, se ha implementado la categoría *inactivo* mediante un enfoque de validación cruzada basado en escenarios. En esta metodología, cada escenario se entrena con muestras de audio inactivas recopiladas de otros escenarios, junto con el conjunto de datos offline que contiene los eventos de interés.

En la Tabla 3.5, se detalla el rendimiento del modelo de reconocimiento de sonido ambiental CNN+MFCC, comparando el ground truth contra la clasificación predecida por el modelo mediante F1-score, precisión y recall para cada escena.

Tabla 3.5: Métricas de clasificación de la evaluación del caso de estudio en tiempo real para cada escena.

	Accuracy	**Precision**	**Recall**	**F1-Score**
Escena 1	0.95	0.97	0.95	0.96
Escena 2	0.99	0.99	0.98	0.98
Escena 3	0.97	0.98	0.97	0.98
Escena 4	0.96	0.97	0.96	0.96
Escena 5	0.91	0.93	0.91	0.92
Escena 6	0.92	0.92	0.90	0.91

Como podemos observar, los métodos y dispositivos propuestos para el RA muestran resultados prometedores tanto en el contexto offline como en el reconocimiento en tiempo real de eventos de audio. Un conjunto de datos balanceado, con 100 muestras por categoría,

resulta adecuado para operar bajo condiciones controladas y naturalistas. Sin embargo, para trasladar estos resultados a aplicaciones en entornos reales", sería necesario disponer de un conjunto de datos más amplio y emplear métodos adicionales de preprocesamiento, tales como técnicas de agrupamiento y aumento de datos. La evaluación de eventos de audio en diversos dominios requiere conjuntos de datos extensos y un procesamiento más complejo para gestionar los métodos de adaptación de dominio [130].

3.4. Procesamiento de sensores de visión térmica para el reconocimiento de actividades asociadas a personas

Las corrientes actuales en el ámbito de los EI han incorporado sensores de visión, los cuales han sido propuestos en conjunto con modelos de visión por computador con el propósito de analizar observaciones visuales para el reconocimiento de patrones en los habitantes [32]. De manera más reciente, se han introducido sensores de visión térmica debido a su capacidad para ofrecer un nivel de privacidad que no proporcionan los métodos tradicionales de cámaras, ya que no permiten la identificación de los ocupantes. Este enfoque posee el potencial de ser más aceptado que las cámaras de visión estándar [131] en entornos privados.

En el contexto de la visión por computador, el DL ha obtenido los resultados más destacados [132] en RA. Además, el uso de Red Residual (Residual Network, por sus siglas en inglés) (ResNet) [133, 134] ha sido aplicado con éxito en diversos campos de visión por computador y procesamiento de lenguaje natural, alcanzando un rendimiento líder con arquitecturas más profundas y amplias, y un número reducido de parámetros de configuración basados en la creación de bloques residuales con capas internas predefinidas. En el campo del procesamiento de imágenes, la adopción de sensores de visión térmica ha surgido como un enfoque no invasivo para la monitorización de personas en hogares [135, 136]. A pesar de las ventajas en términos de privacidad que ofrecen [137, 138], la literatura científica se ha enfocado principalmente en sensores de espectro visible, impulsada por avances en análisis de datos multimedia con DL [139]. Sin embargo, es crucial establecer un entorno donde los

usuarios se sientan seguros y puedan desarrollar su rutina normalmente para promover la aceptación de estas tecnologías [140].

Los sensores de visión térmica han demostrado eficacia en clasificación de posturas [137, 141], detección de caídas [18] y clasificación de actividades [36]. A pesar de la prevalencia de sensores de espectro visible, como Kinect [142], existen desafíos en la calidad de etiquetado en conjuntos de datos, afectando el rendimiento de modelos de DL en conjuntos diferentes al original [143]. Esta discrepancia se intensifica al considerar las diferencias entre imágenes de espectro visible e imágenes térmicas.

Las CNN se han mostrado eficaces en este contexto [144], con OpenPose destacando como el primer sistema en tiempo real para detección de puntos clave en el cuerpo humano, incluyendo 135 puntos clave en imágenes individuales, con resultados destacados [145]. El desarrollo de modelos optimizados para IoT, como OpenPose, ha mejorado la estimación de puntos clave del cuerpo y la optimización de recursos computacionales [146, 147].

En cuanto a la monitorización de actividad física, estudios recientes se han enfocado en el reconocimiento de deportes con CNN, como Sozykin et al. [148], quienes asociaron imágenes con deportes específicos usando CNN Tridimensional (3D). Además, sistemas especializados se han aplicado en análisis deportivo en tiempo real, integrando sensores de movimiento y visión [149]. La clasificación de actividades deportivas con sensores de visión térmica es un área en crecimiento, con investigaciones como Nadeem et al. [150] y Gochoo et al. [141], quienes han explorado el RA mediante detección de siluetas y modelos de partes del cuerpo. La combinación de CNN y redes LSTM también ha sido objeto de atención, como en el trabajo de Zhang et al. [149], proponiendo un marco basado en LSTM para integrar datos de movimiento y visión.

Dada la eficacia comprobada de las CNN en la clasificación de actividades físicas, su aplicación ha demostrado ser generalmente la estrategia más efectiva para obtener resultados óptimos [144].

En el contexto de estos trabajos, en particular sobre visión térmica, la presente sección describe dos áreas de investigación que ilustran el potencial de esta tecnología: la estimación en 2D de puntos de referencia corporales y la clasificación de actividades físicas utilizan-

do sensores de visión térmica. Estos estudios demuestran cómo la convergencia de técnicas avanzadas puede brindar soluciones efectivas para desafíos complejos en el análisis visual y la interpretación de actividades humanas.

- **Estimación Bidimensional (2D) de puntos de referencia corporales (comúnmente denominados *body landmarks*) en imágenes de espectro visible RGB a través del etiquetado de imágenes de visión térmica.** En el marco de esta investigación, se ha conseguido realizar una estimación en tiempo real de los puntos de referencia corporales de múltiples individuos en el espectro visible, empleando una conjunción de técnicas de DL y métodos de afinidad [151]. Esta aproximación evita la necesidad de partir desde cero en el proceso de adquisición de conocimiento. Para lograr este cometido, se presentan estrategias de adaptación de dominio que posibilitan la creación de un mapeo entre las distribuciones de datos de origen y destino [152]. De igual manera, se plantea la utilización de una ResNet, cuya implementación en placas de IoT brinda la capacidad de abordar escenarios en tiempo real, teniendo en cuenta factores como oclusiones, identificación parcial de partes corporales y segmentación individual [153, 154].

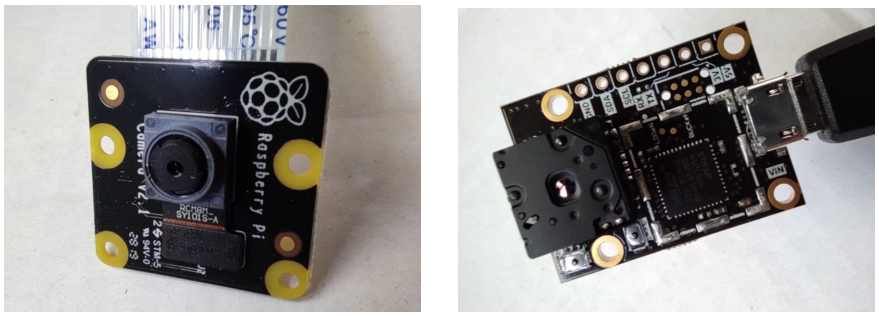
- **Clasificación de cinco actividades físicas utilizando un sensor de visión térmica.** En el marco de este estudio, se ha abordado la tarea de clasificación de cinco actividades físicas empleando un sensor de visión térmica. Con dicho propósito, se ha recopilado un conjunto de datos de sesiones deportivas que comprenden una gama de ejercicios, entre los cuales se incluyen flexiones, abdominales, saltos, sentadillas y planchas. Con miras a optimizar la calidad intrínseca de los datos obtenidos y atenuar la problemática del sobreajuste, se ha implementado un método de aumento de datos personalizado. Por último, se ha evaluado un modelo de DLconfigurado como un híbrido de las capacidades inherentes a las CNN –diseñadas para extraer características de naturaleza espacial– y las LSTM, encargadas de modelar secuencias de imágenes de manera efectiva.

3.4.1. Metodología

3.4.1.1. Estimación en 2D de puntos de referencia corporales

Esta sección detalla la propuesta desarrollada para la estimación de puntos de referencia corporales en dos dimensiones, incluyendo la propuesta de dispositivos, arquitectura y configuración específica de la ResNet empleada. Para llevar a cabo la recopilación de datos y la asociación de imágenes en el espectro visible y el dominio térmico, se ha diseñado un dispositivo IoT compuesto por un sensor de visión térmica y un sensor de espectro visible directamente conectados a una Raspberry Pi 4. La Figura 3.10 muestra los dispositivos descritos a continuación:

- El sensor de espectro visible seleccionado es el **Módulo de Cámara Raspberry Pi NoIR V2**. Se trata de un sensor de bajo coste (aproximadamente 30 €) capaz de capturar imágenes en condiciones de poca luz gracias a que no utiliza filtro infrarrojo (NoIR = No Infrarrojo).
- El sensor de visión térmica seleccionado es el **FLIR Lepton 3.5, integrado con el Módulo Inteligente PureThermal 2 IO**. Se trata de una cámara de infrarrojos de longitud de onda larga que captura datos de temperatura precisos, calibrados y sin contacto en cada píxel de cada imagen, y se encuentra valorado en alrededor de 350 €.



(a) (b)

Figura 3.10: (**a**) Módulo de Cámara Raspberry Pi NoIR V2 (**b**) Módulo Inteligente PureThermal 2 IO + FLIR Lepton 3.5.

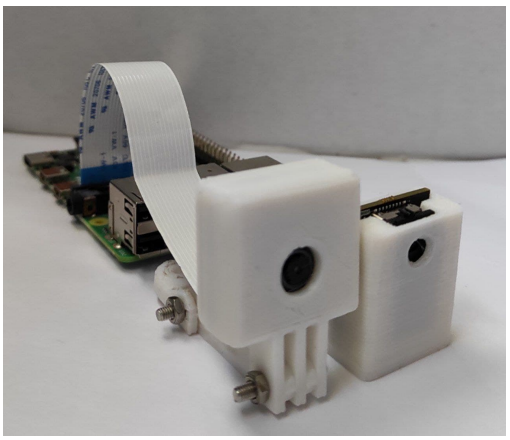


Figura 3.11: Dispositivo IoT con cámara dual (espectro visible y térmico).

Ambas cámaras capturan imágenes de 160 x 120 píxeles a través de una aplicación en Python, lo que permite llevar a cabo la recopilación de datos de manera simultánea en ambos dispositivos. La Figura 3.11 muestra la configuración final de los dispositivos junto con una carcasa diseñada específicamente y fabricada con una impresora 3D. A la izquierda, se encuentra el módulo de cámara Raspberry Pi NoIR V2; a la derecha el módulo PureThermal 2 IO + FLIR Lepton 3.5 y, en la parte trasera, la Raspberry Pi.

3.4.1.1.1. Preprocesamiento y aumentación de datos

La utilización de modelos de DL para llevar a cabo esta estimación de puntos de referencia presenta un desempeño sobresaliente [155]. Sin embargo, uno de los desafíos inherentes a los modelos de DL radica en la necesidad de una amplia cantidad de datos para el proceso de entrenamiento. Esto implica la construcción de un conjunto de datos sustancial en el cual cada imagen térmica deba ser etiquetada por un observador humano. Con el propósito de abordar esta problemática, se presenta una novedosa metodología de autoetiquetado durante la fase de aprendizaje (no en la etapa final de implementación):

- Un dispositivo IoT dotado de sensores de visión térmica y espectro visible recolecta imágenes emparejadas. Sobre los píxeles x_i de los datos térmicos se aplica una función de pertenencia difusa $\mu_{N(x_i)}$, con el propósito de i) normalizar los datos en el intervalo [0,1] y ii) suprimir el ruido de fondo (Figura 3.12). Este paso de preprocesamiento demuestra un rendimiento superior en comparación al uso de imágenes térmicas en bruto en el proceso de entrenamiento de la red. La función de pertenencia se define mediante una forma trapezoidal izquierda, con el fin de normalizar y eliminar los píxeles que componen el fondo de las imágenes térmicas:

$$\mu_{N(x)} = TL(x)[l_1, l_2] = \begin{cases} 0 & x \leq l_1 \\ (x - l_1)/(l_2 - l_1) & l_1 \leq x \leq l_2 \\ 1 & l_2 \leq x \end{cases}$$

- El etiquetado de la imagen de espectro visible se realiza utilizando modelos de DL estables: OpenPose [151].
- El etiquetado de las imágenes de espectro visible se relaciona con las térmicas mediante la homografía H.
- El proceso se repite recopilando muestras en un entorno doméstico de manera automática en tiempo real (con diversas posturas y el usuario).

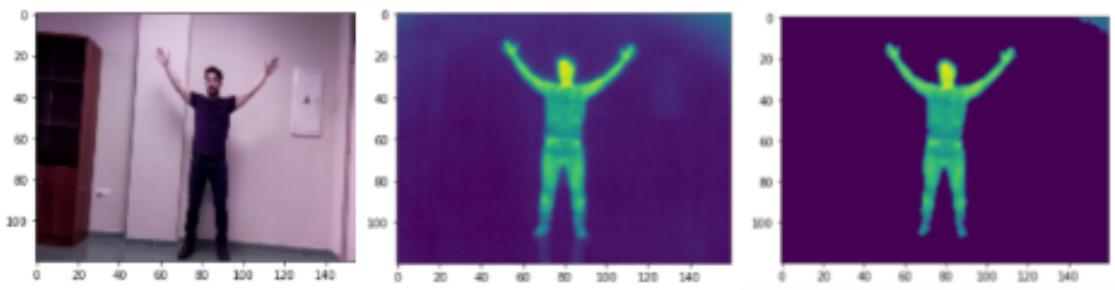


Figura 3.12: Segmentación del fondo de la imagen térmica.

En primer lugar, se utiliza OpenPose para la estimación de puntos de referencia corporales a partir de las imágenes capturadas por el sensor de espectro visible. Dicho enfoque se basa en el modelo MPI, que se simplifica en este caso con cuatro etapas de CNN. Esta configuración más compacta de quince puntos clave, aunque menos precisa en el reconocimiento, se ha seleccionado por generar una mayor velocidad de procesamiento en el tiempo real. El etiquetado automático resultante de la imagen representa el *ground truth* que la ResNet aprenderá para relacionar los datos termales con esos puntos.

Seguidamente, se procede al cálculo de la homografía que relaciona las imágenes térmicas con las del espectro visible, empleando el método de Consenso de Muestras Aleatorias (RANdom SAmple Consensus, por sus siglas en inglés). [156]. En este proceso, se seleccionan diez

puntos clave emparejados en las imágenes térmicas y de espectro visible, los cuales corresponden a ubicaciones coincidentes. La determinación de la homografía entre ambas cámaras se realiza a través de la selección manual de estos puntos relacionados. A continuación, el método RANSAC calcula la homografía que se aplica desde el espectro visible hasta el térmico, logrando una correspondencia precisa de los puntos clave.

Por último, se adopta un método de aumento de datos [157] que aborda la problemática del aprendizaje con grandes volúmenes de datos etiquetados. En este contexto, se implementa un proceso de aumento de datos que transforma cada imagen original para generar múltiples imágenes sintéticas. Tal método involucra escala, volteo horizontal y vertical, así como rotación y traslación de los puntos en la imagen original. Estos procedimientos, realizados con parámetros aleatorios, deben ser aplicados a la homografía final para asegurar la correlación entre las imágenes etiquetadas y las originales. La Figura 3.13 ilustra un ejemplo de aumento de datos para la misma imagen:

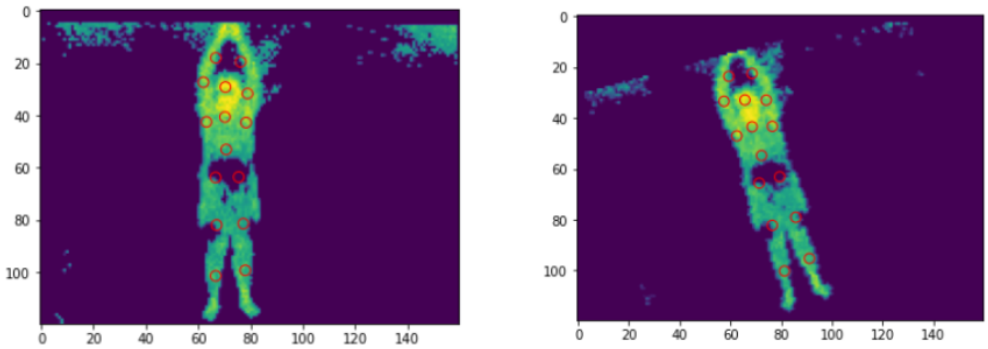


Figura 3.13: Imagen original e imagen sintética generada mediante aumentación de datos, junto con los puntos de referencia relacionados.

3.4.1.1.2. Configuración del modelo de DL

En esta sección, detallamos el modelado de la ResNet que se ha desarrollado y optimizado, destacando los siguientes puntos clave:

i) Se trata de un modelo más ligero diseñado para su implementación en placas IoT con capacidades básicas de cómputo.

ii) Se incorpora una capa de dropout en los bloques residuales para prevenir el sobreajuste.

iii) Se ha desarrollado un regresor progresivo de múltiples capas que va desde la capa flatten hasta 30 valores, los cuales representan las coordenadas x, y de los 15 puntos clave MPI correspondientes a los puntos de referencia del cuerpo.

La estimación de los puntos de referencia se realiza mediante un proceso de regresión (15 puntos con coordenadas x, y) basado en una salida de 30 valores. Para este propósito, se ha propuesto un modelo de ResNet [133]. Una ResNet es un tipo de red neuronal artificial que se basa en estructuras celulares piramidales similares a las presentes en el córtex cerebral. Estas redes se caracterizan por utilizar conexiones de salto o atajos que saltan ciertas capas. Los modelos ResNet convencionales se componen de interrupciones de capas que incorporan normalización de valores y filtrado. En nuestro caso, es fundamental incluir la capa de dropout en los bloques ResNet. La técnica de dropout fue introducida por primera vez en [158] y ha sido adoptada por diversas arquitecturas con muy buenos resultados, como [159], especialmente en modelos amplios con un alto número de parámetros, con el propósito de evitar la coadaptación de características y el sobreajuste.

La creación de las ResNet se realiza mediante bloques que encapsulan funciones de filtro y reducción características de los filtros clásicos. Estas redes han demostrado ser altamente eficaces en tareas de reconocimiento de patrones. Los bloques se componen de 2 a 3 capas, donde la capa final establece conexiones (o atajos) entre las entradas o salidas. En la Figura 3.14, ilustramos el diseño de bloques de esta red.

A partir de estos bloques básicos, se puede diseñar de manera sencilla un modelo de red multicapa. En nuestro caso, hemos definido 10 bloques que configuran una red no profunda, apta para ser implementada de manera eficiente en placas inteligentes. Dentro de este diseño, las entradas son imágenes de dimensiones 120x160 píxeles. Hemos concebido dos modelos, A y B, con la disposición de bloques residuales de la siguiente manera:

- En el modelo A se aplican 2 capas Conv-2D, mientras que en el modelo B se emplean 3 capas Conv-2D.
- El stride determina el número de desplazamientos de píxeles sobre la entrada en Conv2D. Para el modelo A, se fija en stride=1, mientras que en el bloque B se establece en

stride=2. El valor de stride=2 produce un submuestreo del tamaño original de la entrada.

- Los kernels definen las dimensiones del núcleo de convolución. Se establecen en un tamaño de 3.
- La cantidad de filtros establece el número de canales de salida creados después de que el núcleo de convolución actúa sobre la entrada. Estos valores aumentan progresivamente desde 32 hasta duplicarse en tamaño en los diferentes bloques.

La creación de los 10 bloques residuales se intercala entre los modelos A y B en la secuencia $A, B, \cdot, A, B$. Para finalizar, la salida de la última capa, con dimensiones (8, 10, 512), se incorpora en última instancia a través de una capa de normalización mediante una capa promedio (en contraposición a aplicar el máximo, como es habitual en las redes neuronales convencionales). La secuencia de bloques conlleva la generación de 17 capas para la formación de la red definitiva. Con el fin de relacionar la producción abstracta final de la última capa con la salida deseada, se ha establecido una red neuronal conformada por 3 capas densas. Dicha red reduce los patrones abstractos empleando 2048, 512 y 128 neuronas, con el propósito de converger hacia 30 valores de estimación de regresión. Estos últimos valores determinan en última instancia las posiciones de los puntos de referencia (x e y). El modelo completo requiere más de 40 millones de parámetros para la estimación. La ResNet procede con su proceso de aprendizaje de manera iterativa, utilizando el error cuadrático medio para la rectificación y el método Adam como optimizador de tasa de aprendizaje adaptativo [160]. El esquema final de configuración de la red se presenta en la Figura 3.14.

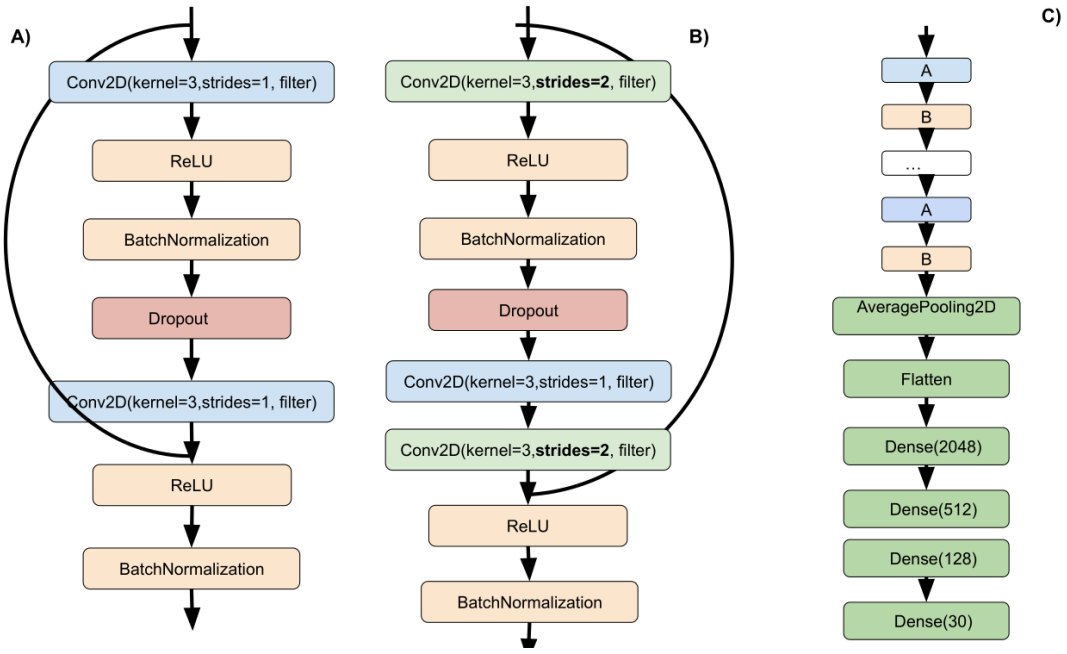


Figura 3.14: Diseño de los bloques residuales para los modelos A y B. El modelo B realiza un muestreo descendente del tamaño de entrada. C) ResNet basada en la configuración de los bloques A y B con capas de optimización finales.

3.4.1.2. Clasificación de actividad física

En esta sección, se presenta un caso de estudio con integración de sensores de visión térmica y modelos de DL secuenciales para la clasificación de actividades físicas. Además de su aplicabilidad en la clasificación de actividades deportivas, esta investigación adquiere una relevancia aún mayor al considerar su potencial en la mejora de la calidad de vida de personas frágiles o dependientes. La monitorización y el análisis preciso de las actividades físicas en este grupo podrían contribuir de manera significativa a la adaptación y personalización de programas de ejercicios terapéuticos, fomentando la autonomía y el bienestar de aquellos con mayores necesidades [161, 162].

Para iniciar esta sección, se describe el dispositivo IoT empleado para la captura de imágenes térmicas. Seguidamente, se detalla la representación de la secuencia de datos recolectados, así como el proceso de preprocesamiento que incluye la técnica de aumento de datos. En tercer lugar, presentamos un modelo de DL que combina CNN y redes LSTM. Esta combinación de enfoques permite abordar de manera eficiente y precisa la tarea de clasificación de actividades

físicas en nuestro estudio.

La recopilación de imágenes del dominio térmico se ha llevado a cabo a través de un dispositivo IoT compuesto por un sensor de visión térmica conectado a una placa de IoT de bajo coste (Raspberry Pi 4). El sensor de visión térmica utilizado es el **FLIR Lepton 3.5, integrado con el módulo PureThermal 2 Smart IO**, presentado en la sección 3.4.1.1.

Como se mencionó anteriormente, este dispositivo destaca por su alta resolución de imagen (160 x 120 píxeles) a pesar de su tamaño mínimo (mucho más pequeño que una moneda), como se muestra en la Figura 3.15 y 3.10 (b). El script desarrollado en Python y ejecutado a través de la Raspberry Pi se conecta directamente con el sensor de visión térmica, solicitando la captura de fotogramas cada 2 segundos. Esta frecuencia de captura permite una recopilación más extensa de datos en condiciones naturalistas.

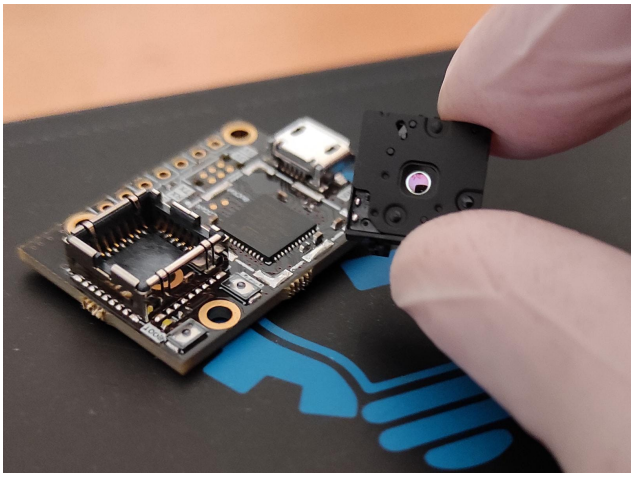


Figura 3.15: Módulo PureThermal 2 Smart I/O + FLIR Lepton 3.5.

3.4.1.2.1. Preprocesamiento y aumentación de datos

Siguiendo una definición formal, un sensor s recopila datos en tiempo real en forma de un par $\overline{s_i} = s_i, t_i$, donde s_i representa una medición específica y t_i es el instante de tiempo. En el caso de un sensor de visión, s_i está compuesto por una matriz de valores de dimensión WxH, donde $s_i[x][y], x \in [0, W], y \in [0, H]$ representa el punto de calor capturado por el sensor térmico en la posición (x, y). Por lo tanto, el flujo de datos de la fuente del sensor s se define como $\overline{S_s} = \overline{s_0}, \ldots, \overline{s_i}$, y un valor dado en un instante de tiempo t_i se define como $S_s(t_i) = s_i$. A partir del flujo de datos $\overline{S_s}$, definimos el tamaño de una secuencia temporal T que segmenta el flujo para obtener una secuencia de valores anteriores $S^(t)$ para cada punto

de tiempo t^*:

$$S^*(t^*) = \{S_s(t^* - T) \rightarrow S_s(t^*T - 1) \rightarrow \ldots \rightarrow S_s(t^*)\}$$

Este $S^(t)$ representa la secuencia de imágenes que es calculada por el modelo de DL propuesto. En la Figura 3.16, presentamos un ejemplo de secuencia de fotogramas para $t^* = 5, T = 5, W = 160, H = 120$ en condiciones reales.

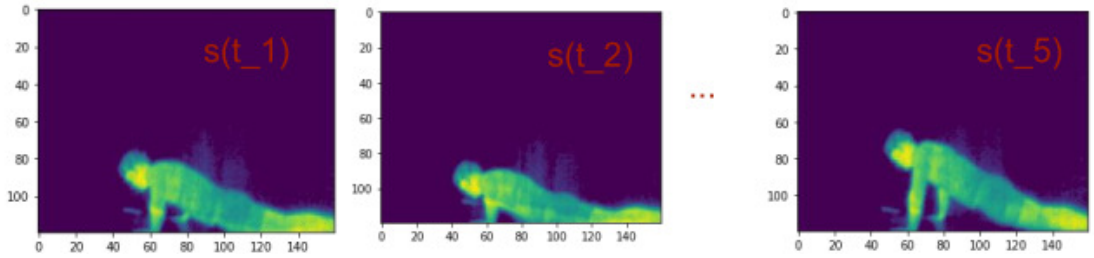


Figura 3.16: Ejemplo de secuencia de fotogramas para $t^* = 5, T = 5, W = 160, H = 120$ en condiciones reales.

El proceso de aprendizaje de las CNN requiere un volumen sustancial de datos [163]. Para abordar esta necesidad, hemos ampliado la cantidad de casos de entrenamiento a partir de un conjunto de datos inicialmente limitado [164], lo cual tiene el efecto de mejorar el rendimiento del modelo y mitigar el sobreajuste [165]. En este contexto, hemos desarrollado una aplicación para realizar un proceso de aumento y expansión (AAE) de los datos de imágenes originales recopilados por ACL. Este proceso involucra diversas transformaciones, como la translación, rotación y escala, con el objetivo de generar un conjunto de datos más amplio y diversificado. Las transformaciones específicas incluyen:

- *Traslación*: La imagen original se desplaza dentro de una ventana de tamaño máximo $[t_x, t_y]^+$ mediante un proceso aleatorio que aplica una transformación de translación $[t_x, t_y]$ con $t_x \in [0, t_x^+]$ y $t_y \in [0, t_y^+]$.

- *Volteo*: La imagen se refleja horizontalmente en un proceso aleatorio que se aplica a un porcentaje de casos F.

- *Rotación*: Se aplica una rotación aleatoria definida por un ángulo máximo de rotación α y una escala aleatoria $s \in [1 - s^+, 1 + s^+]$.

- *Escala*: Se utiliza un factor de escala s^+ para generar una escala aleatoria $s \in [1-s^+, 1+s^+]$.
- *Alteración de píxeles*: Cada píxel $s_i[x][y]$ se modifica mediante un umbral δ utilizando una distribución normal $N(\mu = s_i[x][y], \delta)$.

La Figura 3.17 ilustra un ejemplo de este proceso de aumento de datos con parámetros específicos. Esta metodología de aumento y expansión de datos contribuye a enriquecer y diversificar el conjunto de entrenamiento, lo que a su vez mejora la capacidad del modelo para generalizar y realizar inferencias precisas en una variedad de situaciones.

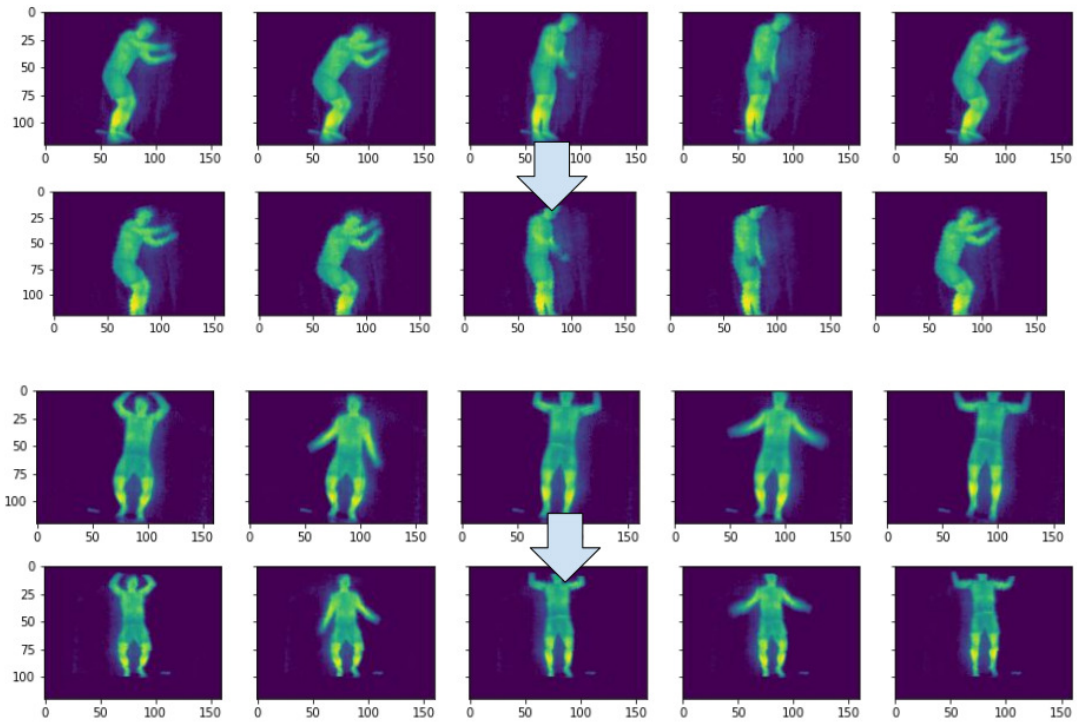


Figura 3.17: Ejemplo de aumento de datos en dos secuencias de imágenes en condiciones reales.

3.4.1.2.2. Configuración del modelo de DL

En esta sección, se presenta en detalle el modelo de DL propuesto para procesar la secuencia de imágenes recopiladas por el sensor térmico. Además, siguiendo lo descrito en el subcapítulo 2.1, hemos optado por una configuración adecuada que combina CNN y LSTM

para la extracción de características espacio-temporales.

En primer lugar, se incorporan las CNN para llevar a cabo la extracción automática de características espaciales a partir de los valores de la imagen térmica $s_i[x][y]$. Se implementa una arquitectura de CNN con 7 capas de convolución 2D. Cada capa de convolución 2D se caracteriza por Conv2D(K, S, ST), donde K denota el número de núcleos (kernels), S representa el tamaño de la convolución y ST define el paso (stride). Se utiliza un valor de paso ST=2, lo que reduce el tamaño de la matriz de entrada a la mitad en la matriz de salida. Para mitigar el problema del desvanecimiento del gradiente y mejorar el rendimiento del aprendizaje, se aplica la función de activación rectificada lineal (ReLU) en las capas intermedias. La Figura 3.18 ilustra el diseño del modelo CNN propuesto.

Posteriormente, se emplea una red LSTM para modelar la secuencia temporal de las características espaciales extraídas por la CNN en cada fotograma. Se integra una estructura LSTM con 2 capas, junto con dos capas de perceptrones multicapa densos que determinan la salida final de la red neuronal. En la Figura 3.18, se describe la modelización temporal del modelo DL y la configuración de la arquitectura CNN+LSTM integrada. Se utiliza la función de pérdida de entropía cruzada y optimizamos los pesos de la Red Neuronal mediante el método de optimización Adam, un algoritmo de descenso de gradiente estocástico ampliamente utilizado en el entrenamiento de modelos de DL.

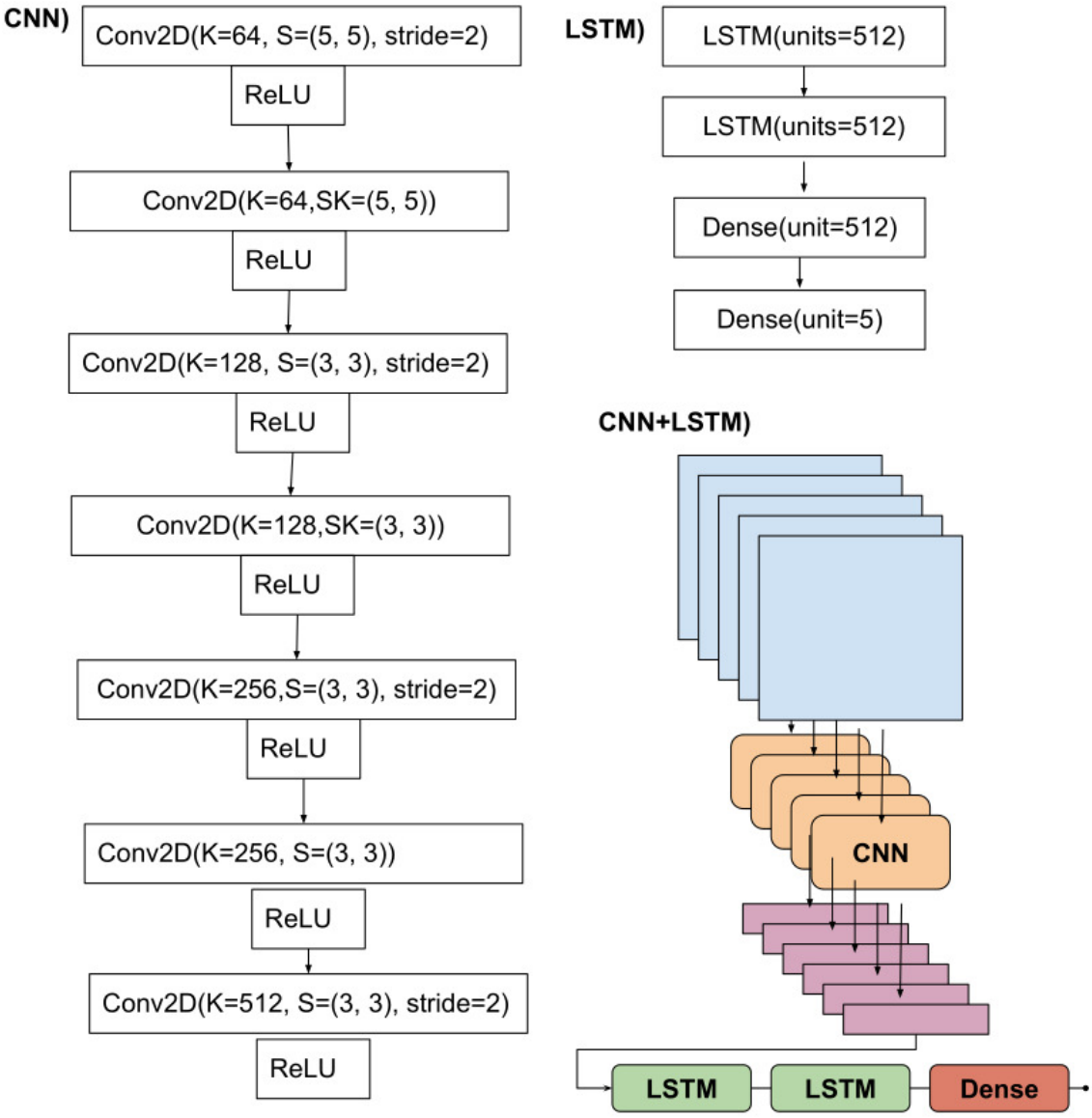


Figura 3.18: (Izquierda) Modelo CNN definido por capas Conv2D y ReLU. (Arriba derecha) Capas LSTM y densas para la configuración de la salida final. (Abajo derecha) Representación del modelo CNN+LSTM.

3.4.2. Resultados

3.4.2.1. Estimación en 2D de puntos de referencia corporales

Esta sección presenta los resultados obtenidos a partir de la evaluación de los puntos de referencia del cuerpo de una persona mediante imágenes térmicas. El caso de estudio

involucra a 4 participantes. En primer lugar, es importante destacar que se ha llevado a cabo una recolección de datos muy ágil. Se han recopilado 120 muestras para cada participante (un total de 480) realizando diferentes poses frontales con cámaras de visión térmica y de espectro visible. Los participantes han realizado ejercicios de tai chi durante 6 minutos mientras el dispositivo IoT recopilaba las imágenes a intervalos de 3 segundos.

Para la eliminación de ruido, normalización y resaltado del cuerpo en las imágenes térmicas, se ha aplicado una función de filtrado mediante una función pertenencia, cuyos valores son $l_1 = 122$ y $l_2 = 255$, respectivamente, a los valores de la matriz térmica. Los datos térmicos se han aumentado (K=10 veces) y procesado con OpenPose para etiquetar automáticamente las imágenes térmicas utilizando el formato de puntos de referencia del cuerpo MPI.

Se ha aplicado una validación cruzada de *leave-one-out* para realizar la evaluación utilizando datos que no han sido observados en el conjunto de entrenamiento. Por lo tanto, el conjunto de datos de entrenamiento está compuesto por 3 participantes (360 imágenes originales), el cual ha sido aumentado (K=10) para obtener 3600 imágenes sintéticas de entrenamiento. El conjunto de datos de prueba está compuesto por el participante excluido (120 imágenes originales).

El modelo de ResNet propuesto para la regresión de puntos de referencia se ha compilado en un lote de 64 imágenes para el aprendizaje, que se cargan en dicho tamaño de bloque y cuyos pesos se optimizan al final del lote. El aprendizaje tuvo lugar durante un total de 50 iteraciones en un ordenador con un procesador Intel i7 de séptima generación y 16 GB de RAM. Se demoró 12 horas y media.

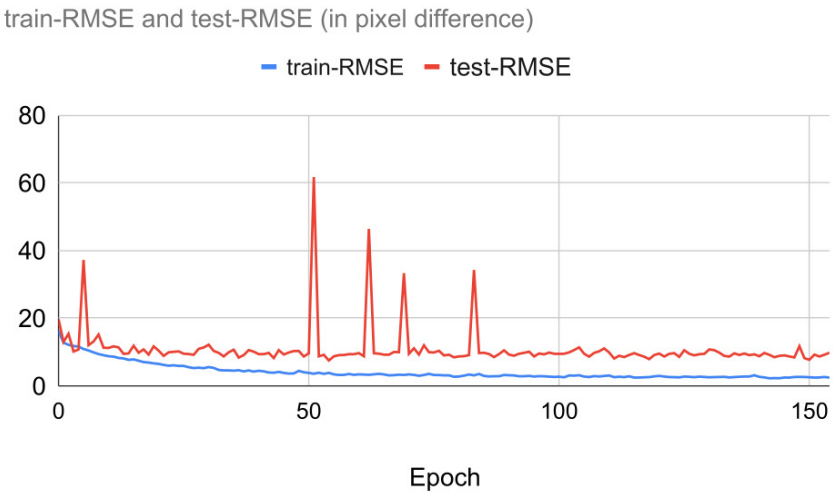


Figura 3.19: Evolución del RMSE en el entrenamiento y las pruebas a lo largo de las épocas de aprendizaje en la ResNet.

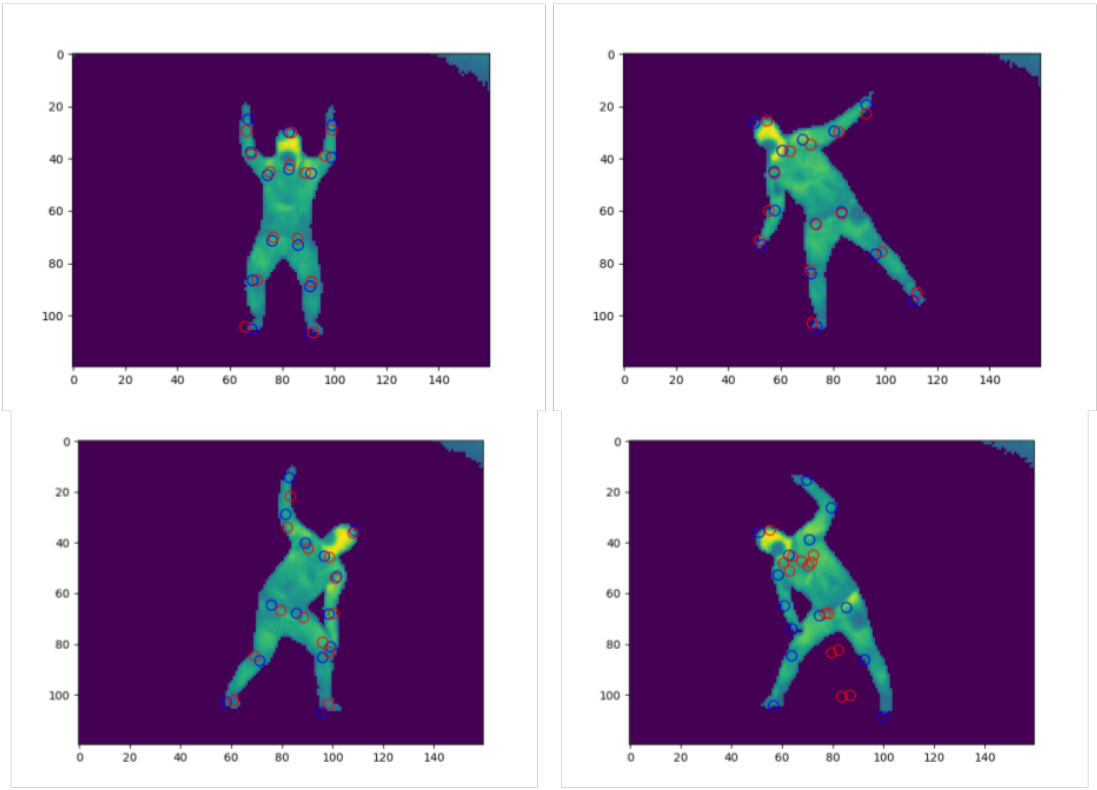


Figura 3.20: Resultados de estimación en las imágenes térmicas.

Es importante destacar nuevamente que las imágenes de prueba no fueron incluidas en el proceso de aprendizaje, y los resultados obtenidos son muy alentadores, especialmente considerando el número limitado de imágenes utilizadas en la fase de recolección de datos. Trabajos futuros podrían implicar la recolección de datos de un mayor número de usuarios, lo que permitiría perfeccionar un modelo de reconocimiento de posturas más preciso.

3.4.2.2. Clasificación de actividades físicas

En esta sección, se presentan los resultados obtenidos en la clasificación de actividades físicas en tiempo real. El caso de estudio incluye tres escenarios donde un participante realiza determinados ejercicios físicos. Específicamente, el participante realiza flexiones, abdominales, saltos, sentadillas y planchas durante un minuto en sesiones de 5 a 6 minutos de duración. El conjunto de datos desarrollado contiene un total de 2089 imágenes (sin el aumento de datos), capturadas a una velocidad de 2 Imágenes por Segundo (Frames Per Second, por sus siglas en inglés) (FPS).

A continuación, el inicio y el final de cada ejercicio es etiquetado por un observador externo. Con fines de aprendizaje, se ha aplicado una validación cruzada dejando fuera una sesión en cada caso, donde una sesión se utiliza para las pruebas y la otra para aprendizaje del modelo.

Se ha definido $T = 5$ (2.5 segundos) como el número de fotogramas que determina la secuencia de entrada para el modelo. El intervalo de tiempo entre la toma de datos se ha establecido en 1 segundo (lo que genera un solapamiento parcial de datos entre secuencias consecutivas). Se destaca nuevamente la rapidez en la recopilación de datos, teniendo que aplicar un método de aumentación de datos (como se describe en la Sección 3.4.1.2).

i) La traslación se establece en $[t_x, t_y]^+ = [-15, 15]$.

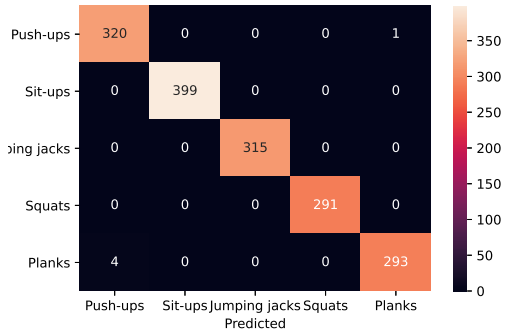
ii) El porcentaje de volteo se establece en $F = 0.5$.

iii) La rotación máxima se establece en $\alpha = 5°$.

iv) La escala de la imagen está en el rango $[1 - s^+, 1 + s^+] = [0.8, 1.2]$.

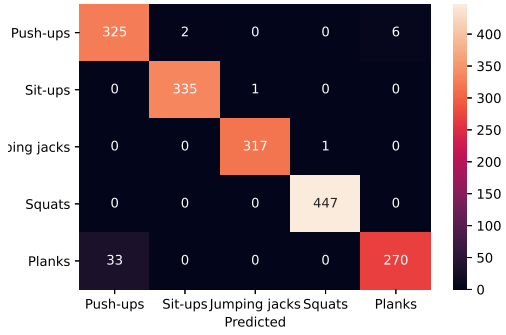
v) La modificación de píxeles se configura con una desviación estándar de $\delta = 0.05$.

El modelo de DL ha sido configurado con un tamaño de lote de 64 y un corto proceso de aprendizaje de 10 épocas. El tiempo de aprendizaje en un sistema con un procesador iCore 17 y 16 GB de RAM fue de 25 minutos. En la Figura 3.21, se presenta la matriz de confusión y en la Tabla 3.6 se detallan el support, el F1-score, la precisión y el recall para cada sesión de prueba.

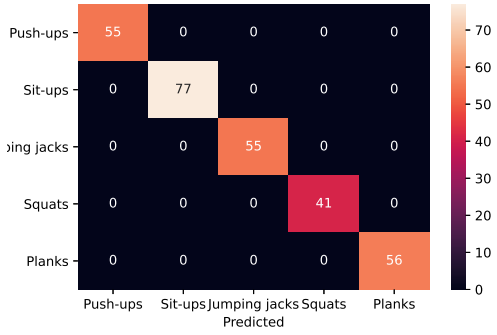


(a) Escena 1 - Entrenamiento

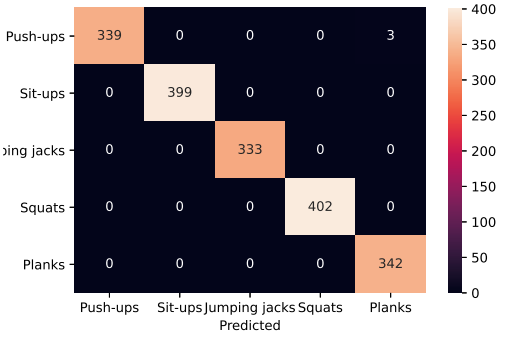
(b) Escena 1 - Evaluación



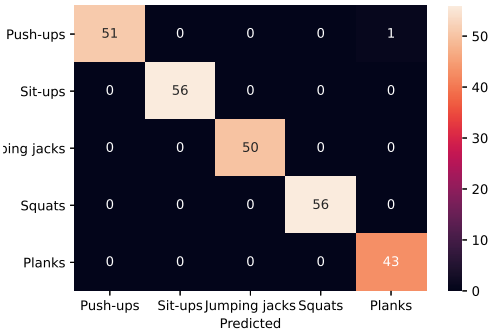
(c) Escena 2 - Entrenamiento



(d) Escena 2 - Evaluación



(e) Escena 3 - Entrenamiento



(f) Escena 3 - Evaluación

Figura 3.21: Matrices de confusión en entrenamiento y evaluación para cada escena.

Tabla 3.6: Métricas de clasificación obtenidas del entrenamiento y validación de las escenas 1, 2 y 3

ESCENA 1

ENTRENAMIENTO

	precision	recall	f1-score	support
0	0.99	1.00	0.99	321
1	1.00	1.00	1.00	399
2	1.00	1.00	1.00	315
3	1.00	1.00	1.00	291
4	1.00	0.99	0.99	297
accuracy			1.00	1623
macro avg	1.00	1.00	1.00	1623
weighted avg	1.00	1.00	1.00	1623

EVALUACIÓN

	precision	recall	f1-score	support
0	0.98	1.00	0.99	59
1	1.00	1.00	1.00	56
2	1.00	1.00	1.00	56
3	1.00	1.00	1.00	93
4	1.00	0.98	0.99	58
accuracy			1.00	322
macro avg	1.00	1.00	1.00	322
weighted avg	1.00	1.00	1.00	322

ESCENA 2

ENTRENAMIENTO

	precision	recall	f1-score	support
0	0.91	0.98	0.94	333
1	0.99	1.00	1.00	336
2	1.00	1.00	1.00	318
3	1.00	1.00	1.00	447
4	0.98	0.89	0.93	303
accuracy			0.98	1737
macro avg	0.97	0.97	0.97	1737
weighted avg	0.98	0.98	0.98	1737

EVALUACIÓN

	precision	recall	f1-score	support
0	1.00	1.00	1.00	55
1	1.00	1.00	1.00	77
2	1.00	1.00	1.00	55
3	1.00	1.00	1.00	41
4	1.00	1.00	1.00	56
accuracy			1.00	284
macro avg	1.00	1.00	1.00	284
weighted avg	1.00	1.00	1.00	284

ESCENA 3

ENTRENAMIENTO

	precision	recall	f1-score	support
0	1.00	0.99	1.00	342
1	1.00	1.00	1.00	399
2	1.00	1.00	1.00	333
3	1.00	1.00	1.00	402
4	0.99	1.00	1.00	342
accuracy			1.00	1818
macro avg	1.00	1.00	1.00	1818
weighted avg	1.00	1.00	1.00	1818

EVALUACIÓN

	precision	recall	f1-score	support
0	1.00	0.98	0.99	52
1	1.00	1.00	1.00	56
2	1.00	1.00	1.00	50
3	1.00	1.00	1.00	56
4	0.98	1.00	0.99	43
accuracy			1.00	257
macro avg	1.00	1.00	1.00	257
weighted avg	1.00	1.00	1.00	257

Como se puede observar en los datos, los resultados son muy alentadores, demostrando una excelente precisión en la clasificación. Se observa que las métricas de entrenamiento presentan un rendimiento ligeramente inferior a las métricas de evaluación, lo cual se atribuiye al impacto de la aumentación de datos en la generación de muestras desafiantes para el proceso de entrenamiento. No obstante, esta diferencia contribuye al desarrollo de un modelo sólido y confiable para propósitos de evaluación.

Tanto los datos recopilados como el código fuente creado están disponibles en GitHub: https://github.com/AuroraPR/Sport-Related-Thermal.

CAPÍTULO 4

TRAZABILIDAD EN ENTORNOS DE MULTIOCUPACIÓN

Este capítulo aborda un valioso desafío dentro del campo del RA: la localización e identificación del usuario. Estos sistemas aportan información clave en el reconocimiento de actividad, especialmente en entornos como industria y hogares inteligentes. Estas tecnologías permiten no solo identificar y seguir el movimiento de personas o activos dentro de un edificio, sino también reconocer y analizar sus actividades.

- Localización: La localización en interiores implica determinar la posición o cómputo del área donde se ubican personas u objetos dentro de un edificio.
- Identificación: La identificación en interiores se centra en reconocer a individuos o etiquetar objetos específicos. Esto se puede lograr mediante tecnologías de corto alcance, reconocimiento facial, o escaneo de códigos QR. En el reconocimiento de actividad, la identificación permite asociar actividades con personas u objetos concretos, lo cual es importante para aplicaciones personalizadas o para mantener un registro detallado de quién realizó qué actividad.
- Trazabilidad: Unificando localización e identificación permite rastrear la historia de localización de personas o objetos en un espacio interior. En el reconocimiento de actividad, la trazabilidad es crucial para comprender cómo se desarrollan las actividades a lo largo

del tiempo y el espacio. La trazabilidad de individuos es clave en contextos de multiocupación, donde existe un entorno compartido; por ejemplo, en un piso habitado por varias personas [166].

Grosso modo, se pueden utilizar múltiples tecnologías para la localización o identificación en espacios interiores [167]: (1) propupestas basadas en radio (UWB, Wi-Fi, Bluetooth, etc.); (2) sistemas ópticos (cámara de video, infrarrojo, etc.); (3) magnético (intensidad magnética); o (4) dispositivos acústicos (ultrasonido).

Dentro del ámbito de posicionamiento en entornos interiores, se distinguen dos enfoques fundamentales: Unidimensional (1D), 2D y 3D [168]. Las propuestas 1D, los iniciales de menor complejidad indican la distancia o presencia a un objeto o sensor, estando relacionados con sensores pasivos de presencia, movimiento como los PIR. Los métodos 2D se fundamentan en señales de corto alcance, como Bluetooth [169], ZigBee [170] o Wi-Fi, para determinar la ubicación. Estos métodos combinan la intensidad de la señal o atributos espacio-temporales con algoritmos de localización. Por otro lado, el enfoque 3D emplea tecnologías como infrarrojo [171], UWB [172], o sensores ultrasónicos [173] para el posicionamiento en tres dimensiones. Además de localizar, estos métodos permiten la identificación de los individuos. Por ejemplo, tecnologías como UWB no solo brindan precisión en el posicionamiento, sino que también posibilitan la identificación única de dispositivos o individuos en entornos con alta densidad de usuarios. Se establece así no solo la ubicación precisa en el espacio, sino también una identificación específica.

De forma paralela, la integración de sensores de visión también facilita la localización y distinción de usuarios. Estos enfoques, al identificar los puntos tridimensionales del cuerpo, comúnmente denominados *body landmarks* [174], posibilitan no solo determinar el número de individuos presentes en un espacio dado, sino también su posición y postura. Además, la identificación facial [175] a través de sensores de visión permite correlacionar las formas corporales con las identidades individuales, lo que contribuye a una identificación más precisa y contextualizada de las personas presentes en el entorno [176].

En este sentido, el presente capítulo presenta una serie de propuestas dirigidas a la identificación y localización de personas en entornos interiores. Los aspectos clave que se abordan

se pueden resumir de la siguiente manera:

- **Evaluación de tecnologías de posicionamiento en interiores, específicamente UWB y BLE, con el objetivo de mejorar la precisión de localización de los habitantes (sección 4.2)**. Para lograrlo, se emplean técnicas de *fingerprint* que permiten optimizar la precisión de la ubicación a partir de mediciones de las señales RSSI mediante datos de un contexto específico. Adicionalmente, se han integrado modelos de DL, en concreto Red Neuronal Convolucional (Convolutional Neural Network, por sus siglas en inglés) (CNN) y LSTM, para predecir la ubicación de los usuarios basándose en la información de las señales Intensidad de la Señal Recibida (Received Signal Strength Indicator, por sus siglas en inglés) (RSSI) recopiladas. Mediante una serie de casos de estudio en entornos residenciales reales, se logra capturar y analizar la actividad cotidiana de los habitantes en situaciones naturales.

- **Metodología para la detección de puntos de referencia tridimensionales en el cuerpo humano, su localización y su identificación en escenarios de multiocupación (sección 4.3)**. Para lograrlo, se emplean modelos de DL avanzados, concretamente Yolo, DeepFace y MediaPipe, con el fin de estimar con eficacia estos puntos y llevar a cabo la identificación correspondiente de los individuos. Posteriormente, se estima la localización bidimensional en el suelo mediante una proyección de homografía. Adicionalmente, se fusiona el seguimiento y reconocimiento facial con trazabilidad mediante un enfoque no supervisado para identificar a los individuos y relacionar los puntos de referencia. Adicionalmente, se presenta un caso de estudio con resultados prometedores para el seguimiento de dos personas en diversos escenarios.

Antes de profundizar en las dos propuestas, a continuación se muestra una descripción de los principales métodos de localización y su evolución tecnológica a lo largo de estos años.

4.1. Evolución de dispositivos y métodos de trazabilidad en interiores

Introducción y primeras aproximaciones con PIR

En EI con multiocupación, la diferenciación de las AVD supone un reto importante [166]. La principal limitación es que la mayoría de los sensores no proporcionan datos individualizados ni identifican al usuario que ha interactuado con los mismos [177, 178]. Por lo tanto, es crucial identificar con precisión la posición y la identidad de cada ocupante para abordar esta problemática [177, 179].

Varios enfoques han sido explorados por la comunidad científica para resolver este desafío, incluyendo el uso de sensores ambientales, dispositivos *wearable* y sistemas de visión. Los sensores PIR, en particular, han sido ampliamente utilizados en residencias para detectar visitas a pacientes [180]. La estrategia principal consiste en analizar los datos de ocupación recabados por estos sensores y aplicar métricas de entropía, como la Entropía Aproximada, la Entropía de Muestra y la Entropía Difusa, para establecer umbrales que indiquen la presencia de personas, basándose en la variabilidad de los patrones de ocupación. Muchas investigaciones han adoptado esta metodología, utilizando sensores PIR para obtener datos precisos sobre la ocupación en interiores [181, 182, 183, 184].

A pesar de los avances en la investigación sobre detección de personas, la tendencia se ha enfocado en hogares unipersonales o en el conteo de individuos en entornos específicos. El desafío clave sigue siendo correlacionar de manera precisa los datos de los sensores ambientales con los usuarios que han generado su activación. Los sensores de visión, aunque más precisos, generan preocupaciones de privacidad [23] y enfrentan dificultades en la identificación precisa de sujetos.

Identificación mediante dispositivos de visión

En investigaciones recientes se ha explorado el uso de cámaras térmicas de baja resolución [185, 79, 186, 187] en combinación con CNN para abordar estos desafíos. En el estudio de [185], se utiliza una cámara de visión térmica y un descomponedor para detectar caídas en presencia

de múltiples ocupantes. La CNN posteriormente determina individualmente si ha ocurrido una caída, superando en rendimiento a propuestas que no consideran la ocupación múltiple. Además, [188] alcanza un seguimiento preciso de las trayectorias de los usuarios mediante el análisis de las diferencias entre imágenes consecutivas, combinando un algoritmo de visión con una CNN de 19 capas. Otros enfoques notables incluyen [189], que utiliza un sensor LiDAR y un método basado en agrupación para separar nubes de puntos correspondientes a cada persona. Este sistema, entrenado con conjuntos de datos en tiempo real y de acceso abierto, demuestra un rendimiento sólido tras aplicar técnicas de adaptación de dominio.

Trazabilidad con RTLS

Los RTLS se han establecido como soluciones avanzadas que complementan los sensores existentes en la identificación de usuarios [190, 191]. Los componentes esenciales de un RTLS incluyen generalmente: i) Etiquetas, que son dispositivos compactos adheridos a objetos o llevados por los usuarios, emitiendo señales que contienen información de identificación única y datos de localización. ii) Anclas, dispositivos de infraestructura estratégicamente distribuidos en el área de cobertura del sistema, diseñados para recibir y procesar las señales de las etiquetas. Los RTLS emplean técnicas como Diferencia de Tiempo de Llegada (Time Difference of Arrival, por sus siglas en inglés) (TDOA) y RSSI para determinar con precisión las posiciones, basándose en las señales recibidas de múltiples anclas [192, 193, 194], donde el uso de varias anclas mejora la exactitud del posicionamiento [195, 196]. Un componente crucial en los RTLS es el middleware, responsable de gestionar el procesamiento de datos, el filtrado, los cálculos de posición y la integración con otros sistemas [197, 198]. El software de monitorización y visualización facilita el seguimiento y presentación en tiempo real de los datos de ubicación mediante interfaces de usuario. Los RTLS se utilizan en sectores como la atención médica, logística, comercio y seguridad, aportando a una mayor eficiencia operativa, mejor seguridad y toma de decisiones basada en datos [199].

Los métodos de localización constituyen técnicas fundamentales en la determinación precisa de la posición de un objeto o dispositivo, basándose en la recepción y análisis de señales provenientes de diversas fuentes, tales como antenas, satélites o balizas. Estas técnicas han encontrado una amplia gama de aplicaciones en sistemas de navegación, rastreo de activos y

sistemas de posicionamiento en tiempo real. A continuación, se definen dichos métodos según la literatura científica [200, 201]:

- Los métodos de **triangulación y trilateración** [202, 203, 204] son enfoques clásicos utilizados en esta disciplina. La triangulación opera mediante el cálculo de las diferencias de distancia o tiempo entre el objeto a localizar y múltiples fuentes emisoras de señales conocidas. En este contexto, la determinación de la posición se logra mediante la evaluación de los ángulos formados por las líneas de visión y las distancias relativas. Por otro lado, la trilateración se apoya en la medición de las diferencias de tiempo o distancia entre el objeto y al menos tres fuentes emisoras, lo que permite calcular la ubicación de manera precisa [205]. En este grupo, destaca la técnica Ángulo de Llegada (Angle of Arrival, por sus siglas en inglés) (AOA) [206], basada en medir el ángulo de llegada de las señales desde múltiples direcciones.

- El método denominado **Tiempo de Llegada Total (Total Time of Arriva, por sus siglas en inglésl) (TTA)** [207], por su parte, se centra en la medición del tiempo que una señal requiere para viajar desde una fuente de emisión conocida hasta el objeto a ser localizado. Comparando los tiempos de llegada de la señal desde diversas fuentes, es posible determinar la distancia entre cada fuente y el objeto, pudiendo este método ser complementado con la trilateración para obtener una ubicación precisa.

- El método **Tiempo de Vuelo (Time of Flight, por sus siglas en inglés) (TOF)** [208], en contraste, mide el tiempo que una señal específica tarda en viajar desde una fuente hasta el objeto y de regreso. Integrando múltiples mediciones temporales de vuelo desde diversas fuentes (anclas), se puede calcular con precisión la distancia entre cada fuente y el objeto. Este método se caracteriza por su enfoque en mediciones temporales altamente precisas y puede emplearse en conjunto con técnicas de trilateración.

- El método **TDOA** [209], por último, se fundamenta en medir las diferencias temporales en las llegadas de señales desde diferentes anclas. Cada ancla sincroniza sus transmisiones con una referencia temporal común, mientras que el objeto a localizar mide las discrepancias temporales en la llegada de la señal desde múltiples anclas. A partir de es-

tas diferencias temporales, se puede calcular la posición relativa del objeto, presentando la posibilidad de combinar este enfoque con la trilateración.

Recientemente, los RTLS, han adquirido relevancia en la comunidad científica e industrial como herramientas para la monitorización precisa del posicionamiento de individuos en espacios interiores [210]. Sus potenciales aplicaciones han potenciado la integración de esta tecnología en dispositivos como teléfonos inteligentes y otros dispositivos que buscan una ubicación de alta precisión [211], especialmente en productos de marcas como Samsung, Google, Apple y Xiaomi, incluyendo Galaxy, Pixel, iPhone, Apple Watch y MIX4, respectivamente. A pesar de los beneficios en términos de precisión, la tecnología UWB plantea problemáticas en materia de seguridad y privacidad debido a su capacidad para rastrear con gran exactitud en términos de distancia a individuos en ambientes interiores.

Señales inalámbricas de corto alcance y aprendizaje *fingerpriting*

Los métodos predominantes para la identificación de personas en EI suelen centrarse en técnicas basadas en la localización. Aunque numerosos estudios han integrado el seguimiento de individuos en contextos específicos, el seguimiento de la localización en interiores continúa siendo un desafío, debido a frecuentes falsas activaciones en los sistemas y dispositivos actuales [178]. La medida de RSSI, que indica la intensidad de la señal de radio recibida, es ampliamente empleada en este campo. Sus ventajas incluyen que no necesita emparejamiento de dispositivos al ser un protocolo básico, es de bajo coste [212], y ofrece un rendimiento aceptable a pesar de la variabilidad de señales entre dispositivos [213]. Una técnica común que usa RSSI es el *fingerprinting* o huella digital, que crea una firma única del entorno mediante un mapeo sistemático y el registro de señales en cada ubicación. Al transformar el espacio en una cuadrícula, se genera una base de datos con la intensidad de señal de cada punto, facilitando la correspondencia de intensidades de señal con ubicaciones específicas [214, 215, 216]. Esta metodología puede combinarse con técnicas de DL para mejorar los modelos de estimación y seguimiento de ubicaciones. Entre las metodologías que utilizan esta técnica, destaca el trabajo de Bai et al. [217], que integra resultados de triangulación a través de RSSI con los obtenidos mediante *fingerprinting*. Estrategias similares se encuentran en los estudios de Xia

et al. [168], que presentan tecnologías de posicionamiento usando huella digital Wi-Fi, y Zhu et al. [218], que investigan la aplicación de aprendizaje automático y algoritmos avanzados en el proceso de *fingerprinting*.

En síntesis, mientras que los métodos de triangulación y trilateración son enfoques tradicionales basados en la medición de ángulos y distancias, los métodos TTA, TOF y TDOA se basan en mediciones temporales y diferencias temporales para estimar la posición. Es importante tener en cuenta que la precisión de estos métodos puede verse afectada por diversos factores, como la calidad de las señales, la interferencia electromagnética, el retraso de la señal y la geometría del entorno [219]. En el marco de esta investigación, se ha optado por abordar el tema de la estimación de la ubicación a través del RSSI, fundamentándonos en las siguientes justificaciones:

i) La simplicidad y el bajo coste del hardware utilizado [220];

ii) Un menor consumo energético al emplear esta técnica, dado que opera en la capa física del sistema de comunicación [221];

iii) En comparación con métodos que se basan en la medición temporal, los cuales exigen componentes adicionales como antenas direccionales o fuentes externas de señal, el enfoque basado en RSSI evita la necesidad de tales recursos extra y minimiza el consumo energético [222];

iv) Las estrategias que dependen de mediciones temporales son diseñadas para situaciones ideales de línea de visión y entornos con mínima obstrucción. En contraste, el enfoque de medición RSSI demuestra resultados más favorables en contextos con trayectorias dinámicas y presencia de obstáculos [219].

No obstante, obtener datos precisos en entornos interiores implica la colocación estratégica de anclas que capturen las señales emitidas por las etiquetas. Estas anclas recopilan las señales de las etiquetas y las transmiten a un servidor central, el cual calcula la ubicación en tiempo real de la entidad en cuestión. Otro punto crucial es el número de anclas utilizadas, ya que se requiere un gran número de ellas en línea de visión para lograr una precisión inferior a un

metro. Para abordar este desafío, la presente propuesta incorpora la técnica de *fingerprinting* con el objetivo de optimizar la predicción de la posición del usuario, como se detalla en las subsecciones siguientes.

Trazabilidad con Bluetooth

En el ámbito de la localización en interiores, la tecnología de red inalámbrica BLE se ha establecido como una de las más empleadas debido a su bajo coste y facilidad de implementación. Comúnmente utilizada en dispositivos vestibles y smartphones, configurados como etiquetas o balizas, BLE capta valores de RSSI. El principal reto de BLE es la variabilidad de la señal [223] y las diferencias entre entornos causadas por la ubicación de muebles y objetos decorativos, lo cual a menudo requiere una alta densidad de balizas para una implementación efectiva [224, 225]. En cuanto a métodos basados en tiempo, aunque el algoritmo de TOF ofrece un rendimiento superior [226], conlleva requisitos más estrictos en cuanto a procesamiento computacional, sincronización y emparejamiento. Además, la alta demanda de comunicación y la necesidad de circuitos de sincronización adecuados incrementan los costos y el mantenimiento [227]. Para la estimación de posiciones, diversas investigaciones han aplicado técnicas como la trilateración y el *fingerprinting* [202], [203] y [204]. Combinar ambas técnicas es común, ya que la trilateración tiene limitaciones con pocas balizas o en presencia de ruido ambiental, mientras que el *fingerprinting* requiere una extensa base de datos etiquetada por entorno [228].

Trazabilidad con UWB

En los últimos años, la adopción de la tecnología UWB ha visto un incremento notable en una variedad de sectores, tanto científicos como industriales. Este auge se atribuye a su mayor rentabilidad y creciente popularidad [167, 191]. Caracterizada por un amplio ancho de banda espectral y pulsos de radio de corta duración, UWB se distingue de las tecnologías inalámbricas convencionales, como Wi-Fi o Bluetooth, por su operación en un extenso rango de frecuencias, desde megahercios hasta gigahercios. Esta particularidad permite una medición precisa del TOF facilitando la localización exacta de dispositivos mediante coordenadas, incluso con precisiones de centímetros [199].

En las últimas propuestas de investigación en este campo, se han explorado métodos de posicionamiento interior utilizando UWB. Xia et al. [229] han implementado DS-TWR y Particle Filter en conjunción con BLE para lograr una precisión de 2-3 cm utilizando el algoritmo EKF. Che et al. [230] han aplicado radar UWB y algoritmo NB para diferenciar entre entornos con línea de visión (LOS) y sin línea de visión (NLOS), empleando 4 anclas y 1 etiqueta. Ambrose et al. [231] combinaron placas DWM3000EVB y nRF52840-DK (UWB+BLE) con trilateración, alcanzando un error de 344 mm, empleando 1 etiqueta y 3 anclas en un espacio reducido. Efendi et al. [232] utilizaron UWB EWINE con aprendizaje basado en árboles de decisión, logrando un 90 % de precisión en la detección de LOS y NLOS en interiores. Volpi et al. [233] emplearon el kit Qorvo MDEK1001 con DWM1001 para trilateración, obteniendo precisión de unos 10 centímetros empleando 4-6 anclas y 1-3 etiquetas en un laboratorio. Gna et al. [234] compararon RSSI y mediciones de TOF con el módulo DW1000, y Kim et al. [235] emplearon Decawave DWM1000 con AI-EKF para condiciones NLOS, logrando un Error Cuadrático Medio Raíz (Root Mean Square Error, por sus siglas en inglés) (RMSE) de 0.53 con 5 anclas y 1 etiqueta. Otros estudios notables incluyen el de Li et al. [236] con un enfoque en el mapeo de rutas que emplea 2 etiquetas y 8 anclas, y Nakamura et al. [237] que usaron antenas multiestáticas y Newton-Raphson con 3 anclas y 1 etiqueta para la localización de destinos. Yin et al. [238] propusieron un sistema de posicionamiento interior con UWB y mediciones RTOF, mientras que Bregar et al. [239] implementaron el módulo DecaWave DW1000 y el algoritmo Respuesta de Tiempo de Vuelo (Time-of-Flight Response, por sus siglas en inglés) (TWR), alcanzando una precisión de 0.59 m.

No obstante, la integración de UWB en entornos domésticos y trazabilidad del usuario presenta desafíos, tales como:

- Reducción del rendimiento en edificios con obstáculos físicos como paredes y muebles [240].
- Aumento en el error y la incertidumbre debido a la necesidad de estabilidad en los sistemas de localización pasiva, con baja fluctuación y señales fuertes [241].
- Influencia de la ubicación donde se sitúan los sensores [242, 240].

UWB y reconocimiento de actividades

Tras la identificación del usuario y la determinación de su ubicación, surge la oportunidad de diseñar sistemas que fusionen de manera eficaz la tecnología UWB con sensores ambientales, con el fin de implementar el RA. Durante los últimos cuatro años, diversos estudios han investigado esta sinergia, evidenciando avances significativos:

Cheng et al. [243] implementaron el módulo DW1000, complementado con algoritmos de trilateración y mínimos cuadrados no lineales, para localización y RA en interiores, tales como sentarse, levantarse, permanecer de pie y caminar. Lograron una precisión en el RA del 87.2 % y 80.2 % usando SVM en diferentes configuraciones. Por otro lado, Matre et al. [244] utilizaron un sistema de radar UWB y modelos de aprendizaje profundo como Stacked LSTM, CNN-LSTM y ResNet para la clasificación de una amplia gama de actividades, incluyendo beber, dormir, vestirse, realizar tareas domésticas, cocinar, entre otras, alcanzando una precisión del 94 % con CNN+LSTM. Pajak et al. [245] demostraron el uso de la tecnología Pozyx con sensores *wearable* UWB y algoritmos TDOA y TWR, en combinación con CNN, para el RA físicas como flexiones, sentadillas y dips, obteniendo una alta precisión tanto en entrenamiento (97.5 %) como en test (94.7 %). Tabbakha et al. [246] propusieron una metodología que integra el seguimiento de ubicaciones en interiores y el RA mediante tecnología BLE y algoritmos de RF y SVM, aunque presentaron limitaciones en la variedad de actividades y su correlación con ubicaciones específicas. Zhan et al. [247] desarrollaron MoSen, un sistema de aprendizaje automático para inferir ubicaciones y reconocer actividades en hogares con múltiples ocupantes, enfrentando ciertas limitaciones en términos de generalización y diferenciación de actividades. Finalmente, Arrotta et al. [248] implementaron un enfoque de razonamiento basado en el conocimiento para el análisis de datos contextuales, proponiendo el uso de estas tecnologías para una asociación confiable de datos, aunque excluyeron actividades que no se ajustaban al contexto especificado.

Estos estudios destacan la viabilidad de combinar UWB y RA en diversos entornos, incluyendo espacios interiores y domésticos. No obstante, también señalan desafíos en cuanto a generalización, precisión en condiciones reales y la vinculación de actividades con ubicaciones específicas para mejorar los resultados. Además, se observa que muchos estudios con

alta precisión no detallan los sensores utilizados, no comparten sus datos con la comunidad científica o no permiten la replicación de sus métodos, limitando la posibilidad de contrastar sus resultados.

4.2. Trazabilidad en interiores basada en UWB y modelos de DL

Esta investigación se centra en la evaluación de tecnologías de posicionamiento en interiores, en particular UWB y Bluetooth, con el fin de mejorar la precisión en la localización de los habitantes. Para aumentar la precisión de la ubicación, se utilizan técnicas de huella digital o *fingerprinting*, que aprovechan las mediciones de RSSI de estas tecnologías. Además, se ha integrado el uso de modelos de DL, específicamente CNN y LSTM, con el objetivo de predecir la ubicación de los usuarios basándose en los datos recopilados de las señales RSSI. A través de dos casos de estudio llevados a cabo en entornos reales, se ha logrado capturar y analizar las AVD de los habitantes en situaciones naturales, demostrando la efectividad de estas tecnologías y metodologías en el reconocimiento preciso de la ubicación en interiores.

4.2.1. Metodología

En esta sección, se expone la arquitectura y métodos empleados para la estimación de la posición en interiores de los usuarios, detallando los dispositivos utilizados y procesamiento de datos. Asimismo, se proporciona una descripción exhaustiva del modelo de DL. En este contexto, la arquitectura se compone de un EI equipado con anclas dispuestas estratégicamente en las paredes y el techo, mientras que los usuarios portan etiquetas que suministran datos de RSSI a dichas anclas.

Los valores de RSSI captados por las anclas se transmiten en tiempo real a un nodoo receptor, el cual procede a segmentar y preprocesar la información utilizando una segmentación de ventana deslizante, ajustando la recepción de entrada de datos para ser procesadada por el modelo de posicionamiento. Este modelo genera, como salida, las coordenadas 2D correspondientes a la posición de cada residente. Para ello, realiza predicciones basadas en el *ground*

truth y en los datos de RSSI provenientes los datos etiquetados previamante. Finalmente, las ubicaciones de los residentes se almacenan en una base de datos, permitiendo el acceso a ellos tiempo real de forma remota. La Figura 4.1 ilustra los componentes de la arquitectura propuesta.

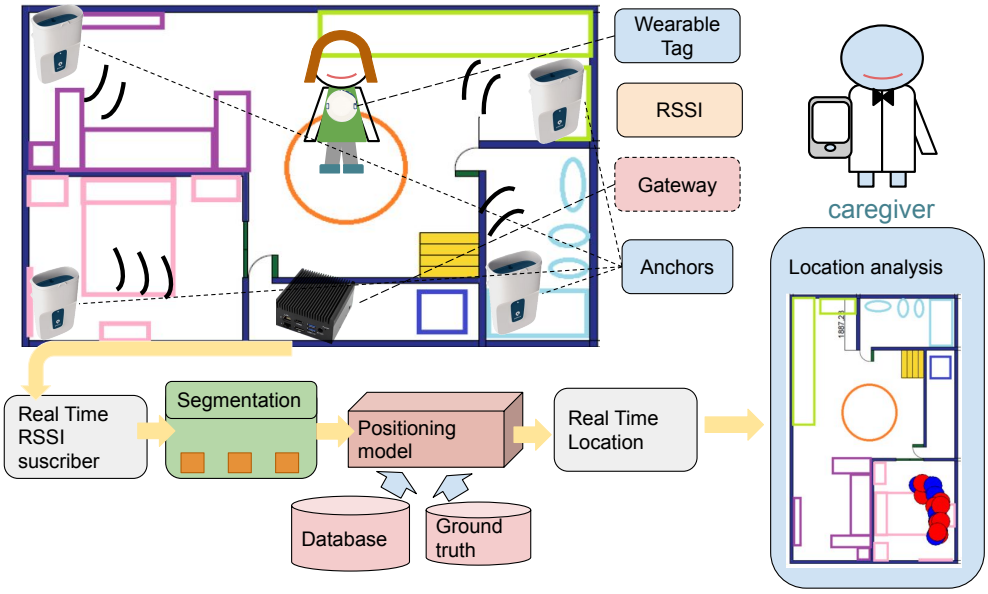


Figura 4.1: Arquitectura de componentes para el sistema de posicionamiento en interiores.

Los dispositivos empleados en esta investigación son parte del kit *Enterprise Kit Lite* de Pozyx [1], que se basa en el principio de trilateración y que puede procesar hasta 100 posiciones por segundo. Como se ha mencionado previamente, la efectividad de este sistema está determinada por el número y la ubicación de los dispositivos que funcionan como anclas. El precio base de este kit es de alrededor de 3.000 euros e incluye un gateway, cuatro anclas y cuatro etiquetas portátiles que pueden ser utilizadas como colgante. La Figura 4.2 presenta uno de estos dispositivos.

[1] https://www.pozyx.io (consultado el 26 de noviembre de 2022)

Figura 4.2: Componentes del sistema Pozyx: anclas, etiquetas y gateway

- **Anclas.** Las anclas constituyen un componente esencial para lograr un posicionamiento preciso. Su función principal es capturar las señales, procesar los datos y transmitirlos a la plataforma central. Para realizar cálculos de posición en 2D, el sistema requiere la presencia de al menos tres anclas visibles entre sí. Estas anclas son compatibles con dos protocolos de posicionamiento: TDOA y TWR. Están disponibles en variantes con clasificación IP20 o IP66/67 y se encuentran diseñados para montaje en paredes y techos mediante soportes VESA. Las anclas pueden ser alimentadas a través de un conector mediante tecnologías Energía sobre Ethernet (Power over Ethernet, por sus siglas en inglés) (POE) (IEEE 802.3af) o POE+ (IEEE 802.3at), permitiendo la conexión en cadena de hasta 4 anclas.

- **Etiquetas portátiles.** Estos dispositivos compactos poseen dimensiones de 66 mm x 65.4 mm x 17 mm (L × A × H) y ofrecen una vida útil de batería de hasta 5 años, con la posibilidad de reemplazar las baterías. Su consumo energético es notablemente bajo, y pueden alcanzar una velocidad máxima de actualización de 10 Hz. Las etiquetas incluyen conectividad NFC para activación, instalación y configuración, además de permitir la modificación de su ID para facilitar su identificación. Están diseñadas para ser compatibles con el sistema de posicionamiento TDOA mencionado anteriormente. Además de su función principal, estas etiquetas incorporan un acelerómetro de 3 ejes como sensor adicional y cuentan con una clasificación IP66/67 para resistencia a condiciones

adversas.

4.2.1.1. Preprocesamiento de datos

Siguiendo una definición formal, una ancla s detecta la presencia en tiempo real de una etiqueta e, manifestándose como un par $\overline{(s,e)_i} = (s_i, t_i)$, en donde $(s,e)i$ denota el valor de RSSI y t_i representa el instante de tiempo correspondiente a la medición. De esta manera, se genera un flujo de datos asociado a un par de ancla y etiqueta (s,e), definido como $\overline{S(s,e)} = \overline{(s,e)_0}, \ldots, \overline{(s,e)i}$, y se obtiene el valor en un instante de tiempo t_i como $S(s,e)(t_i) = s_i$.

Para asegurar una homogeneidad en los datos obtenidos de distintos sensores durante el tiempo y frecuencia de recolección, se definen múltiples ventanas temporales de deslizamiento simétricas que agrupan y resumen los mismos. Estas ventanas están determinadas por el tamaño de ventana W_w que delimita un intervalo temporal $[W_w^-, W_w^+]$. Las ventanas segmentan las muestras del flujo de datos del sensor RSSI $\overline{S_{(s,e)}}$, y los valores $\overline{(s,e)_i}$ dentro de los intercalos de estas ventanas, que son agregados mediante funciones de agregación. Este proceso se describe de la siguiente manera:

$$T_t(S_{(s,e)}, W_w, t^) = \bigcup_{\overline{(s,e)i} \in S(s,e)} s_i, \quad t_i \in [t^- W_w^-, t^* + W_w^+] \tag{4.1}$$

De esta forma, las funciones de agregación $\cup$ se aplican a los datos $(s,e)_i$ que caen dentro de un intervalo de tiempo $W_w = [W_w^-, W_w^+]$ en un instante específico t^*. Específicamente, en este estudio se proponen las siguientes funciones de agregación $\cup = \mu, \max, \min$, que se consideran descriptores fundamentales para flujos de alta velocidad [52]. Estas funciones de agregación ofrecen una representación altamente descriptiva de los datos segmentados por intervalos de tiempo, lo que permite homogeneizar los datos en bruto del sensor que se recopilan a tasas de muestreo intermitentes.

Mediante la aplicación de estas técnicas de agregación de datos, se lleva a cabo la segmentación de la señal a medio plazo mediante varias ventanas temporales deslizantes de corta duración. Esto se logra con ventanas que se definen de la siguiente manera: $W = W_1 = [W_1^-, W_1^+], \ldots, W_i = [W_i^-, W_i^+], W_i^- = W_{i-1}^+$.

En este proceso, se obtiene una secuencia de características a partir de los flujos de datos de los sensores para cada etiqueta e, generando una estructura de datos con dimensiones $|S| \times |\cup| \times |W|$:

$$S_e^*(t^*) \rightarrow \begin{cases} S_e^1(t^*) = T_t(S_{(S1,e)}, [W_1^-, W_1^+], t^*) \ldots, \\ \rightarrow T_t(S_{(S1,e)}, [W_i^-, W_i^+], t^*) \\ \ldots \\ S_e^s(t^*) = T_t(S_{(s,e)}, [W_1^-, W_1^+) \ldots, \\ \rightarrow T_t(S_{(s,e)}, [W_i^-, W_i^+], t^*) \end{cases}$$

En esta etapa, los flujos de sensores $S_e^s(t^*)$ para cada etiqueta e y ancla s: i) se sincronizan en el mismo instante de tiempo t^*, ii) se dividen en segmentos temporales mediante ventanas de deslizamiento homogéneas W, y iii) se representan como vectores de características mediante funciones de agregación $\cup = \{\mu, max, min\}$. La Figura 4.3 proporciona una representación visual de cómo se lleva a cabo la segmentación y agregación en un flujo de sensores RSSI. Se observan las ventanas temporales de deslizamiento junto con las agregaciones de promedio, mínimo y máximo.

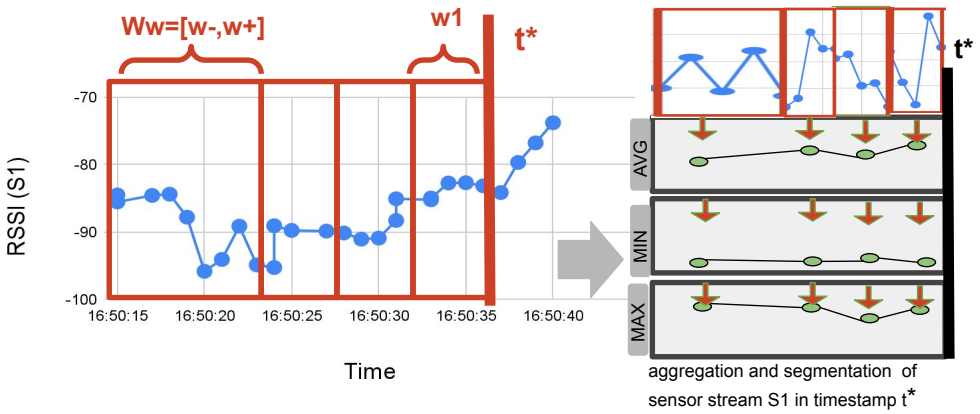


Figura 4.3: Segmentación y agregación mediante promedio, mínimo y máximo de un flujo de sensor de RSSI definido por ventanas temporales de deslizamiento.

4.2.1.2. Configuración del modelo de DL

El modelo de DL, diseñado para predecir la posición de los habitantes en tiempo real, se describe de la siguiente manera: i) para cada etiqueta e y en un instante de tiempo específico t, la entrada $S_e^s(t^)$ se forma utilizando múltiples señales de sensor S, ii) las mediciones de RSSI son recopiladas por cada ancla ambiental. Estas señales se describlen mediante una matriz de tamaño $S \times W$, donde W define el número de pasos en la ventana temporal.

El modelo se compone de tres capas de Convolución 1D (CNN) que actúan como extractores de características espaciales. Estas capas tienen filtros cuyos tamaños son 2, 3 y 3 respectivamente. A continuación, se incorporan dos capas de LSTM con 32 unidades cada una, diseñadas para capturar las relaciones temporales de las características extraídas por la CNN. Para mitigar el sobreajuste y la coadaptación de características, se utilizan capas de Dropout [159].

La elección de una red híbrida CNN-LSTM se basa en su eficacia demostrada en aplicaciones de fingerprinting [249]. Posteriormente, un modelo Perceptrón Multicapa con dos capas de 512 y 256 unidades aprende patrones espaciales y temporales para generar la salida, que representa las coordenadas X e Y de la posición de la persona. Esta salida se produce mediante una función de activación sigmoide. El modelo se entrena utilizando retropropagación y la métrica de pérdida utilizada es el Error Absoluto Medio (Mean Absolute Error, por sus siglas en inglés) (MAE), optimizado mediante el algoritmo Adam. La Figura 4.4 ofrece una representación visual de la disposición de capas en el modelo de DL propuesto.

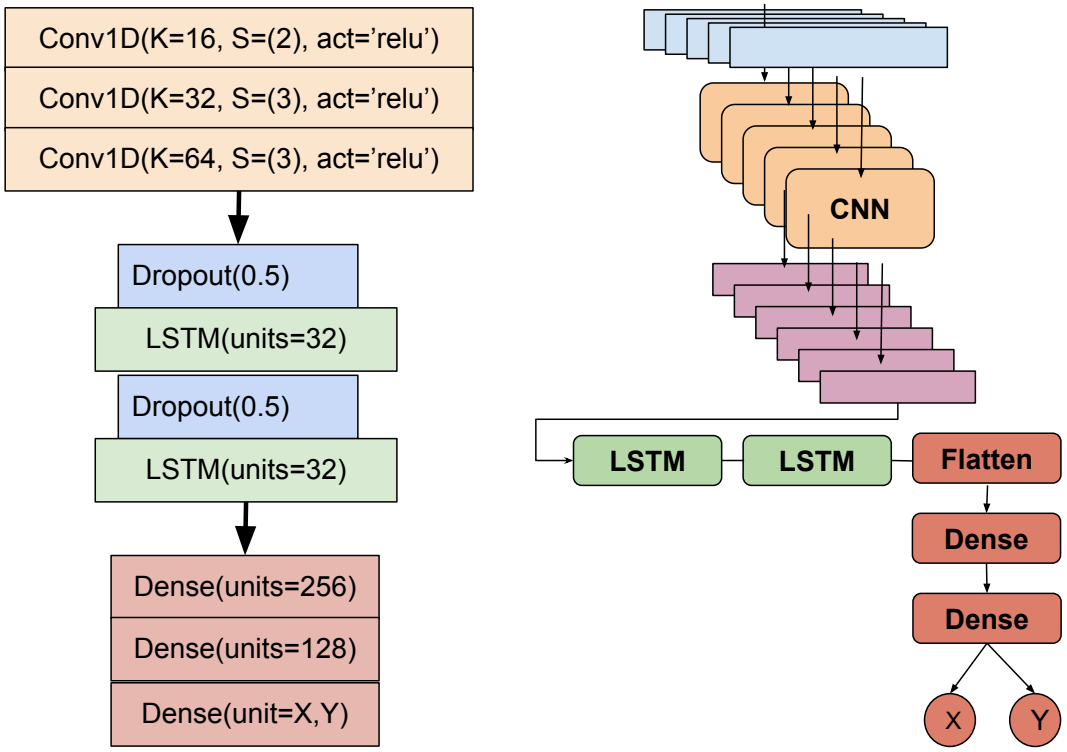


Figura 4.4: Configuración de modelos de DL, incluyendo capas de CNN y LSTM.

En el marco de este estudio, hemos evaluado el desempeño de diferentes arquitecturas de modelos de DL: CNN, LSTM y una combinación de ambas conocida como CNN+LSTM (detallada anteriormente). Entre las distintas arquitecturas evaluadas, hemos observado que la combinación CNN+LSTM produce los resultados más óptimos.

En lo que respecta a la estimación de habitaciones, hemos configurado la salida final del modelo con N salidas que representan las predicciones para cada habitación desde una perspectiva de clasificación. En este contexto, aplicamos la función de activación softmax en la salida y utilizamos la métrica de pérdida de entropía cruzada binaria en el proceso de retropropagación para el aprendizaje.

4.2.2. Resultados

En esta sección, presentamos los casos de estudio desarrollados en la evaluación de aplicabilidad de la tecnología UWB en un entorno real. Los dispositivos empleados son parte del *Enterprise Kit Lite* de Pozyx, detallados anteriormente. En este contexto, cada ancla transmite el RSSI a una etiqueta específica en tiempo real mediante el protocolo MQTT. Incluso si las anclas no están en línea de visión entre sí y no permiten la trilateración para proporcionar coordenadas (x, y), el flujo de datos continúa. La recolección de datos ha sido facilitada por un usuario que llevaba la etiqueta UWB y, en tiempo real, etiquetaba su ubicación en el interior de la vivienda al hacer clic en las coordenadas en una tableta que mostraba un mapa del lugar. Los datos de etiquetados en el mapa por el usuario son utilizados como *ground truth* para el modelo.

Adicionalmente, se ha implementado una configuración similar utilizando BLE en ambos despliegues. Esto nos ha permitido comparar las capacidades de BLE y UWB en condiciones de despliegue reales en entornos domésticos. En ambos casos de estudio, seguimos un enfoque de instalación uniforme de sensores para abarcar todas las áreas de las viviendas, recopilando información de RSSI tanto para UWB (anclas) como para BLE (utilizando una Raspberry Pi configurada como baliza receptora de señales). En el caso de UWB, los habitantes llevaron una etiqueta UWB como un colgante. Para BLE, empleamos la pulsera *Amazfit GTS 2e* con conectividad Bluetooth. Es importante destacar que cualquier pulsera de actividad con soporte BLE y una dirección MAC conocida podría ser integrada.

Como se ha mencionado, este caso de estudio abarca la implementación en dos entornos residenciales con distribuciones y dimensiones distintas. Además, se han llevado a cabo dos configuraciones de dispositivos distintas, las cuales se detallan en las secciones posteriores. Este objetivo proporciona dos contextos diferentes con variaciones en los despliegues y el número de anclas para su evaluación científica. Esta metodología de recopilación de datos en condiciones similares pero con ubicaciones y configuraciones de sensores diferentes ha sido adoptada en trabajos previos relevantes [250, 251, 129].

Piso A (configuración con 4 anclas)

El primer escenario fue desarrollado en un apartamento de 60 m^2. Cuatro anclas UWB y cuatro balizas BLE fueron desplegadas para abarcar las habitaciones principales, que incluyen la sala de estar, la cocina, el baño y el dormitorio. En la Figura 4.5 (a) se puede observar la disposición de las anclas (representados en verde) y las balizas BLE (representadas en azul) en el plano del apartamento A.

La Figura 4.5 (b) ilustra la disposición de los sensores UWB y BLE en el entorno. El mapa utilizado para tomar las muestras presentaba unas dimensiones de 460 x 753 píxeles, correspondiendo a un área en la vivienda de 5800 x 9500 mm. Se registraron ocho trayectos del habitante, cada uno involucrando diferentes rutas que reflejaban actividades cotidianas dentro del entorno. La duración de estos trayectos fue de entre 5 y 10 minutos, y descritos de la siguiente manera: Trayecto 1) 6 minutos y 14 segundos, Trayecto 2) 9 minutos y 32 segundos, Trayecto 3) 6 minutos y 55 segundos, Trayecto 4) 5 minutos, Trayecto 5) 7 minutos y 16 segundos, Trayecto 6) 10 minutos y 50 segundos, Trayecto 7) 7 minutos y 5 segundos, Trayecto 8) 10 minutos y 34 segundos, acumulando un total de 63 minutos y 23 segundos de grabación.

(a) (b)

Figura 4.5: (a) Plano del apartamento A y (b) despliegue del mismo.

Piso B (configuración con 6 anclas)

El despliegue fue desarrollado en un apartamento de 100 m^2. Seis anclas UWB fueron distribuidas en la vivienda de la siguiente manera: dos en la sala de estar, una en el baño, una en la cocina, una en el dormitorio y una en el pasillo central. La cantidad y ubicación de las balizas BLE se mantuvieron iguales al despliegue previamente descrito. La Figura 4.6 ilustra la posición de cada ancla (en verde) y las balizas BLE (en azul) en el plano del apartamento B. El mapa utilizado para la toma de muestras presentaba unas dimensiones de 536 x 621 píxeles, correspondiendo a un área de 9500 x 11100 mm. Se llevaron a cabo diez trayectos, que se describen como sigue: Trayecto 1) 3 minutos y 1 segundo, Trayecto 2) 2 minutos y 6 segundos, Trayecto 3) 2 minutos y 1 segundo, Trayecto 4) 1 minuto y 56 segundos, Trayecto 5) 2 minutos y 23 segundos, Trayecto 6) 2 minutos y 34 segundos, Trayecto 7) 2 minutos y 23 segundos, Trayecto 8) 1 minuto y 51 segundos, Trayecto 9) 2 minutos y 43 segundos, Trayecto 10) 1 minuto y 41 segundos, resultando en un total de 22 minutos y 30 segundos de grabación.

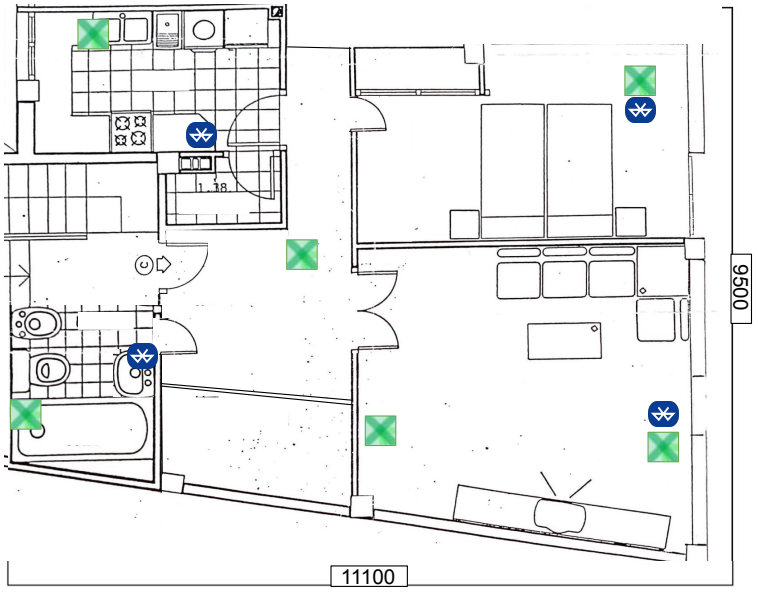


Figura 4.6: Plano del apartamento B.

En el piso A, se recolectaron un total de 16 879 muestras de señales RSSI, mientras que en el piso B se obtuvieron 9 176 muestras. En términos de datos etiquetados, se recopilaron 3 068 muestras para el piso A (aproximadamente una muestra por segundo) y 2 658 para el

piso B (aproximadamente una muestra cada medio segundo). Se estableció una frecuencia de etiquetado muy alta para asegurar un *ground truth* preciso, con valores de ($\mu = 0.98s, \sigma = 0.70s$) para el piso A y ($\mu = 0.41s, \sigma = 0.61s$) para el piso B.

Para lograr la sincronización entre las mediciones de señales RSSI y los datos etiquetados por el usuario en tiempo real, se emplea el protocolo MQTT como un medio centralizado. Los datos de RSSI y las etiquetas del usuario se distribuyen a través de mensajes de publicación de MQTT en tiempo real. Un nodo suscriptor, cuyo propósito es sincronizar y almacenar los datos, genera una marca de tiempo utilizando el reloj de su computadora al recibir los mensajes de datos de RSSI y etiquetado mediante el protocolo de suscripción.

En la fase de aprendizaje, se establece un intervalo de tiempo de un segundo mediante un enfoque de ventana deslizante que segmenta los valores de localización mediante los datos de entrada de RSSI. La ubicación, salida del modelo, se relaciona en cada intervalo de tiempo con los datos de entrada. La ubicación o posición del usuario se calcula mediante interpolación lineal de las posiciones etiquetadas por el usuario (posiciones previa y posterior a ese instante de tiempo). Se han excluido los datos etiquetados por el usuario que presenten una diferencia de más de 5 segundos entre muestras, considerándolos como intervalos sin etiquetado por parte del usuario durante el proceso. En el caso del piso A, el 2.43 % de las muestras presentaban esta característica y se descartaron, mientras que en el caso del piso B, solo el 0.72 % de las muestras se descartaron por esta razón.

En nuestros resultados comparativos con otros enfoques, hemos incorporado propuestas con modelos relacionados. Estos enfoques incluyen:

i) Una clasificación de áreas o habitaciones en las que ha estado el usuario basada en señales RSSI de UWB y BLE [179].

ii) El uso de SVM como modelo de fingerprint basado en RSSI para la estimación de ubicación [214].

iii) La integración de RF como modelo de fingerprint basado en RSSI para la estimación de la posición [215].

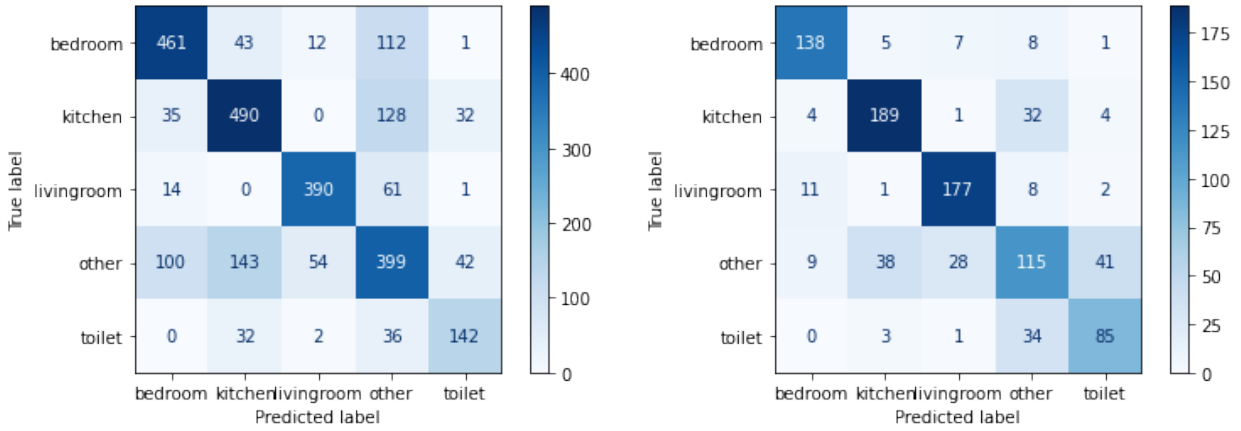
iv) La incorporación de DL (CNN+LSTM) como modelo de fingerprint basado en RSSI

para la estimación de la posición. Cabe mencionar que previamente se habían utilizado LSTM y CNN por separado en [252].

Inicialmente, se evalúan los modelos de DL (CNN, LSTM y CNN+LSTM) para la estimación de la posición de los habitantes. En segundo lugar, se realiza un análisis del impacto en el rendimiento utilizando varias configuraciones de ventanas temporales con distintos tamaños, considerando enfoques tanto pasados como actuales [253].

Siguiendo la metodología de [18], el tamaño acumulativo de la ventana se emplea para determinar la activación de cada sensor dentro de una representación a largo, mediano y corto plazo. Esto permite capturar la activación temporal de sensores en forma binaria. Hemos adoptado tamaños de ventana de 4, 12, 20 y 30 segundos, escalando de 1 a 2 segundos en incrementos sucesivos. Este enfoque se alinea con una metodología de ventana de tiempo múltiple e incremental [18, 253, 254]. Seguidamente, se emplea la métrica de error cuadrado medio y el enfoque de validación cruzada con 10 pliegues para la evaluación de la regresión de la posición. Por último, se efectúa una evaluación específica del modelo CNN+LSTM, que ha demostrado ser la configuración más efectiva en los enfoques de DL, tanto para BLE como para UWB (ancla-etiqueta). En este contexto, se ha abordado un problema de clasificación para estimar la habitación donde se encuentra el habitante [179].

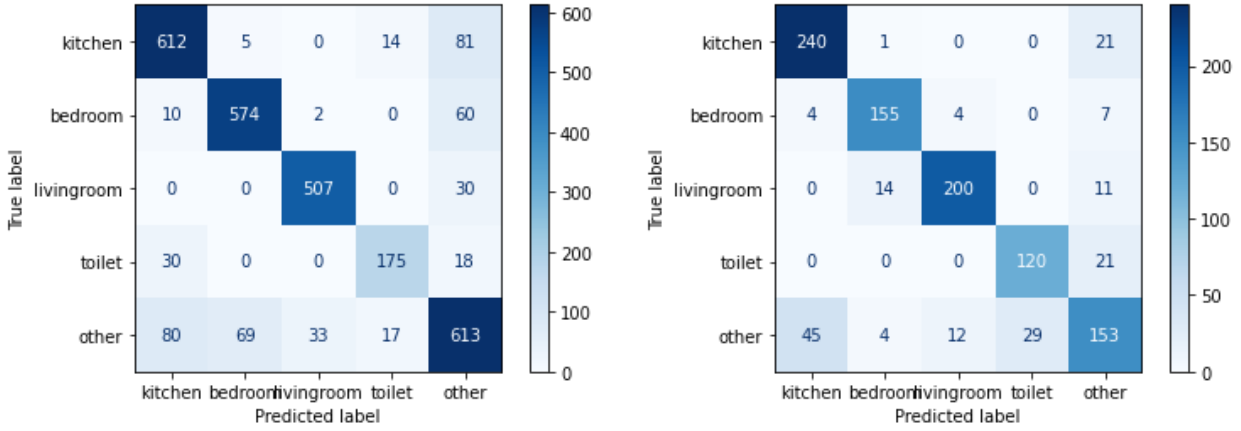
Se excluyeron las muestras de datos que carecían de valores RSSI válidos, dado que no proporcionaban información sobre la ubicación del usuario. Esto podría deberse a la indisponibilidad del dispositivo o a valores infinitos que indican una distancia demasiado grande. Cabe destacar que esta situación solo se presentó con los valores RSSI de BLE. En el caso de UWB, los datos de RSSI siempre estuvieron disponibles desde las anclas. En el contexto del piso A, el algoritmo que se basó en los datos de RSSI de BLE logró predecir con éxito la habitación donde se ubicaba el usuario con una tasa de predicción del 0.70. Por otro lado, en el piso B, esta misma aproximación alcanzó una precisión del 0.75 (consulte la Figura 4.7 para más detalles).



(a) (b)

Figura 4.7: Precisión de la poisición de los habitantes en las habitaciones a partir de los datos de RSSI de BLE en los pisos A y B.

Respecto a los datos de RSSI de UWB, se obtuvieron resultados más precisos. Específicamente, en el piso A, se logró una tasa de predicción del 0.85 en la clasificación de habitaciones, mientras que en el piso B se alcanzó una precisión del 0.83. Estas cifras representan una mejora significativa en comparación con las predicciones basadas en RSSI de BLE (consulte la Figura 4.8 para más detalles).



(a) (b)

Figura 4.8: Precisión de la poisición de los habitantes en las habitaciones a partir de los datos de RSSI de UWB en los pisos A y B.

Adicionalmente, después de realizar las predicciones, se calculó el error absoluto acumulado tanto para la variable x como para la variable y. Se realizaron pruebas utilizando diferentes

métodos de DL, incluyendo CNN, LSTM, y una combinación de ambos, LSTM+CNN. Los cálculos se llevaron a cabo utilizando los datos recopilados de ambos pisos, y los resultados fueron consistentes en ambos casos. En ambas situaciones, el método más efectivo resultó ser LSTM+CNN, seguido por CNN. Las Tablas 4.1 y 4.2 presentan una comparación exhaustiva de los resultados obtenidos. En la columna *WINDOWING*, se indica si se emplearon ventanas temporales pasadas y futuras (PAST+FUTURE) o solamente pasadas (ONLY PAST). En *MODEL* se especifica si se utilizó CNN, LSTM o una combinación LSTM+CNN. La columna *W. SIZE* muestra el tamaño de la ventana, mientras que MAE proporciona el promedio del MAE de las coordenadas X e Y (en metros). Cabe destacar que el error MAE se calcula comparando los datos estimados por los modelos con el *ground truth* proporcionado por el usuario en situaciones de alta frecuencia y tiempo real.

Tabla 4.1: Comparación de los resultados obtenidos con los diferentes métodos en el piso A.

WINDOWING	MODEL	W. SIZE	MAE
PAST+FUTURE	CNN		0.57
PAST+FUTURE	LSTM		1.80
PAST+FUTURE	LSTM+CNN		0.64
PAST+FUTURE	SVM	4	0.93
PAST+FUTURE	RF		0.70
ONLYPAST	SVM		0.98
ONLYPAST	RF		0.80
PAST+FUTURE	CNN		0.40
PAST+FUTURE	LSTM		1.33
PAST+FUTURE	LSTM+CNN		0.32
PAST+FUTURE	SVM	12	0.71
PAST+FUTURE	RF		0.55
ONLYPAST	LSTM+CNN		0.33
ONLYPAST	SVM		0.74
ONLYPAST	RF		0.55
PAST+FUTURE	CNN		0.33
PAST+FUTURE	LSTM		1.40
PAST+FUTURE	LSTM+CNN		0.23
PAST+FUTURE	SVM	20	0.61
PAST+FUTURE	RF		0.53
ONLYPAST	LSTM+CNN		0.23
ONLYPAST	SVM		0.68
ONLYPAST	RF		0.53
PAST+FUTURE	LSTM+CNN		0.17
PAST+FUTURE	SVM		0.58
PAST+FUTURE	RF		0.48
ONLYPAST	LSTM+CNN	30	0.18
ONLYPAST	CNN		0.28
ONLYPAST	SVM		0.64
ONLYPAST	RF		0.51

Tabla 4.2: Comparación de los resultados obtenidos con los diferentes métodos en el piso B.

WINDOWING	MODEL	W. SIZE	MAE
PAST+FUTURE	CNN		0.74
PAST+FUTURE	LSTM		2.10
PAST+FUTURE	LSTM+CNN		0.66
PAST+FUTURE	SVM	4	0.72
PAST+FUTURE	RF		0.56
ONLYPAST	SVM		0.79
ONLYPAST	RF		0.53
PAST+FUTURE	CNN		0.55
PAST+FUTURE	LSTM		1.24
PAST+FUTURE	LSTM+CNN		0.37
ONLYPAST	LSTM+CNN	12	0.35
PAST+FUTURE	SVM		0.79
PAST+FUTURE	RF		0.41
ONLYPAST	SVM		0.66
ONLYPAST	RF		0.47
PAST+FUTURE	CNN		0.50
PAST+FUTURE	LSTM		0.61
PAST+FUTURE	LSTM+CNN		0.32
ONLYPAST	LSTM+CNN	20	0.30
PAST+FUTURE	SVM		1.08
PAST+FUTURE	RF		0.43
ONLYPAST	SVM		0.73
ONLYPAST	RF		0.49
PAST+FUTURE	LSTM+CNN		0.25
ONLYPAST	LSTM+CNN		0.25
ONLYPAST	CNN		0.44
PAST+FUTURE	SVM	30	1.47
PAST+FUTURE	RF		0.43
ONLYPAST	SVM		0.89
ONLYPAST	RF		0.41

También se realizó una exploración del impacto del tamaño de las ventanas temporales en los resultados. Se observó una relación inversamente proporcional entre el tamaño de la ventana y el error. A medida que el tamaño de la ventana aumenta, el error disminuye; sin embargo, se identificó un fenómeno de rendimiento decreciente: conforme el tamaño de la ventana aumenta, la disminución del error es menos pronunciada y el tiempo de ejecución se extiende. Asimismo, se realizó una comparación del error promedio en las coordenadas X e Y utilizando dos configuraciones: i) ventanas temporales que abarcan tanto el pasado como el futuro, y ii) enfoque centrado solo en ventanas pasadas. Los resultados revelaron que no existen diferencias significativas al utilizar ventanas futuras, ya que las diferencias no superaron en promedio un punto, y en ocasiones el enfoque de ventanas pasadas resultó ser incluso más efectivo.

En la Tabla 4.3, se ha realizado una comparación exhaustiva de los resultados de la estimación de ubicación 2D utilizando el enfoque CNN+LSTM en relación con la trilateración basada en TDOA obtenida de Pozyx. Junto con el MAE en las coordenadas X e Y, también se ha incluido el número de marcas temporales en las que los modelos no pudieron proporcionar una estimación de posición. Es importante resaltar el sólido desempeño del enfoque basado en RSSI y DL, así como su robustez en la estimación de ubicación en áreas con cobertura deficiente, donde la trilateración no es capaz de calcular la ubicación. Además, se ha grabado un video que dinámicamente muestra la ubicación predicha por el sistema en comparación con la ubicación etiquetada por el usuario en ambos pisos. En el video, los círculos rojos representan la ubicación predicha y los círculos azules representan la ubicación etiquetada por el usuario. La velocidad de reproducción del video para el piso A es tres veces más rápida que para el piso B debido a las diferencias en la duración de los conjuntos de datos: aproximadamente 60 minutos para el piso A y alrededor de 20 minutos para el piso B. El video está disponible en el siguiente enlace de YouTube: https://youtu.be/FUluz8Dz3ns

Tabla 4.3: Comparación de los resultados obtenidos con UWB mediante trilateración + TDOA frente a LSTM+CNN basado en fingerprint + datos de RSSI.

ALGORITHM	**FLAT**	**MAE (x)**	**MAE (y)**	**LOST EST.**
LSTM+CNN	A	0.14	0.23	0
TDOA	A	0.46	0.64	4.0 %
LSTM+CNN	B	0.20	0.30	0
TDOA	B	0.75	1.41	18.6 %

4.3. Trazabilidad tridimensional mediante dispositivos de visión en entornos de multiocupación

La integración de puntos de referencia tridimensionales anatómicos, la estimación precisa de la ubicación espacial y la identificación del usuario ha cobrado primordial relevancia en la comunidad investigadora [174]. En esta sección, se introduce una propuesta destinada a facilitar el cálculo de puntos de referencia tridimensionales del cuerpo humano en contextos caracterizados por la presencia de múltiples ocupantes. Esto viabiliza la representación no

invasiva de personas a partir de la información recogida por sensores de visión. Este enfoque fusiona las capacidades inherentes de los sensores de visión con las aplicaciones fundamentadas en técnicas de DL, con el propósito de construir representaciones virtuales. En este sentido, aprovechando el poder de procesamiento que ofrecen los sensores de visión, es factible calcular los puntos de referencia tridimensionales del cuerpo [255], los cuales desempeñan un papel esencial en la comprensión del movimiento humano y las interacciones en el EI [256].

Con el propósito de alcanzar una estimación precisa con localización bidimensional en el suelo (2D) en EI, se introduce una estrategia de proyección fundamentada en el concepto de homografía. Dicha técnica se respalda en la estimación de la posición de los pies como punto de partida para situar al usuario en el espacio del mundo real [257]. Además, también se incorpora un método de seguimiento no supervisado basado en características de DL [258, 259]. Finalmente, la identificación facial [175, 260] establece una relación entre las formas corporales y la identidad del individuo, contribuyendo a la mejora de la representación del usuario en el entorno virtual en el cual se proyectan sus posturas.

En efecto, la virtualización de información visual se sitúa como un campo de investigación con perspectivas prometedoras [261], donde la morfología del cuerpo humano da lugar a avatares que emulan la apariencia humana y retienen sus atributos intrínsecos y coherencia anatómica [262]. Asimismo, este concepto guarda cierta relación con la creación de gemelos digitales, que establece escenarios enriquecedores en el contexto de la industria 4.0 [263], así como con el reconocimiento de actividad humana, donde se generan representaciones virtuales de individuos para reflejar sus comportamientos, movimientos y acciones en el mundo real [264].

En lo que respecta a los métodos y aplicaciones propuestos, la esencia la metodología propuesta se apoya en un conjunto de herramientas de visión altamente eficientes, entre las que se incluyen Yolo [265], YoloFace [266], ResNet, DeepFace [267] y MediaPipe [268], desempeñando un rol crucial en la estimación de los puntos de referencia tridimensionales del cuerpo.

4.3.1. Metodología

Esta sección presenta los materiales y métodos empleados para la estimación de puntos de referencia 3D del cuerpo, así como la estimación de la posición e identificación de los individuos haciendo uso de sensores de visión. En primer lugar, se abordan en detalle el tipo de sensor empleado, su disposición y configuración. Posteriormente, se detallan las estrategias utilizadas para la segmentación y estimación de los puntos de referencia 3D, el cálculo de la posición en 2D y la identificación del individuo en situaciones de ocupación múltiple. Por último, se expone la herramienta diseñada para representar los datos obtenidos en un entorno virtual.

El enfoque adoptado tiene como objetivo principal la utilización de sensores de espectro visible como cámaras IP o sensores de visión integrados para adquirir datos en tiempo real. Con el propósito de obtener un amplio campo de visión, los dispositivos se ubican en las esquinas de la habitación. Esta posición de las cámaras, en contraposición a enfoques que las sitúan en posiciones cenitales o en el techo, resulta esencial para posibilitar la identificación facial y el seguimiento simultáneo del cuerpo y los pies de los habitantes.

La metodología propuesta ha sido concebida para la utilización de un único sensor de visión. Durante su instalación, se requiere la siguiente configuración:

- **Coordenadas del suelo en el mundo real:** Es necesaria la determinación de la ubicación de marcadores o mediciones para facilitar la obtención de la ubicación en 2D del usuario en términos de coordenadas del mundo real. Para lograrlo de manera automática, se precisan múltiples puntos que conecten las imágenes de origen con las correspondientes coordenadas en el mundo real.

- **Base de datos facial:** Cada residente ha de transitar por el entorno durante un período breve para establecer una base de datos facial destinada a la identificación del usuario en las escenas. Este enfoque realiza la recopilación automática de imágenes faciales tanto frontales como laterales.

Una vez que la cámara se ha configurado con esta configuración mínima, se exponen los

métodos de procesamiento visual utilizados en las siguientes subsecciones.

4.3.1.1. Preprocesamiento de datos

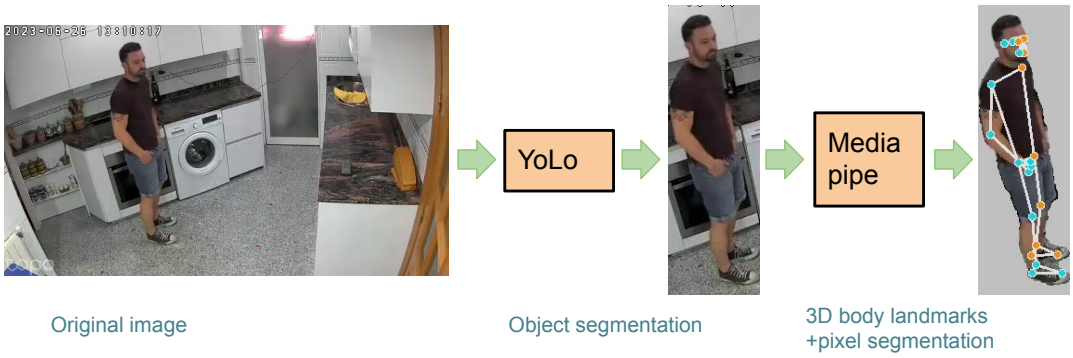
En esta sección, se detalla la metodología empleada para la segmentación del cuerpo y la obtención de puntos de referencia 3D en el marco de este enfoque. El proceso se divide en dos etapas principales:

Primero, se realiza la detección de objetos en la escena utilizando la herramienta *You Only Look Once* (YoLo) en su versión 7 [269, 270]. Esta versión de YoLo ha destacado por su alto rendimiento en una variedad de dominios, ya que permite la detección de objetos y personas en la imagen mediante el uso de cajas delimitadoras. Si bien este estudio se centra en la detección de habitantes, futuros trabajos planean expandirse para incluir otros objetos clave en el entorno virtual.

Una vez que se ha detectado la caja delimitadora (bounding box) de cada persona, la segunda etapa implica la utilización de Mediapipe para llevar a cabo la estimación de la postura de los sujetos [271, 272]. Concretamente, se calculan los puntos de referencia corporales en 3D para cada caja delimitadora identificada. Los puntos de referencia corporales en 3D y la segmentación de píxeles proporcionada por Mediapipe desempeñan un papel crucial en este enfoque por las siguientes razones:

- Proporcionan una representación descriptiva y no invasiva de la figura de los habitantes en el entorno virtual, salvaguardando así su privacidad.
- Permiten el cálculo de la ubicación 2D de los habitantes en el suelo mediante el uso de una homografía que mapea la imagen a las coordenadas del suelo en un contexto de visión monocular.
- Facilitan la eliminación de los píxeles correspondientes al fondo en la imagen segmentada del cuerpo, lo que contribuye a reducir el ruido en el seguimiento de las personas, evitando interferencias basadas en la ropa y características del cuerpo.

En la Figura 4.9, se presenta un ejemplo que ilustra una imagen con un sujeto, las cajas delimitadoras y los puntos de referencia corporales en 3D.



.

Figura 4.9: Ejemplo ilustrativo de una imagen que contiene a un individuo junto con cajas delimitadoras proporcionadas por el modelo YoLo, así como puntos de referencia 3D del cuerpo y la segmentación de píxeles realizada por Mediapipe.

Los resultados generados al realizar la segmentación del cuerpo y la estimación de puntos de referencia corporales en Tridimensional (3D) desempeñan un papel fundamental, ya que constituyen las principales entradas para los subsiguientes componentes encargados de la localización 2D y el seguimiento de múltiples ocupantes utilizando datos provenientes de sensores de visión.

Con respecto al proceso de cálculo de la ubicación 2D de los habitantes en el entorno, se propone una solución basada en visión monocular para cada sensor de visión. La razón detrás de esta elección radica en la intención de desarrollar un enfoque en el cual los sensores de visión operen de manera autónoma en una primera etapa, y luego compartan información para generar una representación global de la escena con coordenadas en el mundo real.

Se parte de la premisa de que los habitantes no están en un estado de levitación. Por lo tanto, se puede establecer una relación directa entre la superficie del suelo vista desde una perspectiva cenital en metros y los puntos en píxeles de la imagen, que corresponden a la ubicación de los pies de los habitantes. Con base en esta relación, se vinculan puntos en la imagen original (x_1, y_1) con coordenadas en el mundo real (x_2, y_2). La homografía $(x_2, y_2) = H \cdot (x_1, y_1)$ se calcula utilizando el método de RANSAC [273]. La Figura 4.10 presenta un ejemplo que ilustra el proceso de localización en 2D de habitantes en el mundo real, basado en la estimación de la posición de los pies.

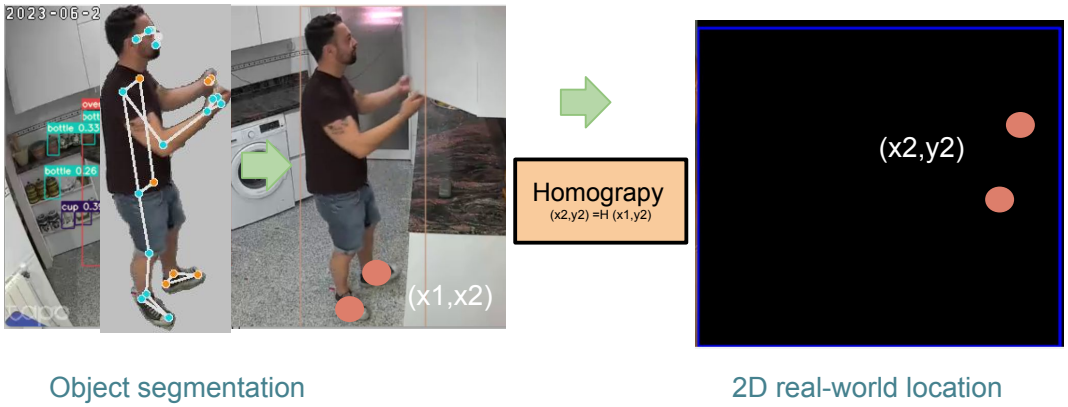


Figura 4.10: Ubicación en el mundo real en 2D de los habitantes basada en la estimación de los pies y la homografía a partir del sensor de visión monocular.

4.3.1.2. Configuración del modelo de DL

Esta sección describe el seguimiento de múltiples habitantes en un entorno con ocupación múltiple. Un desafío clave en este proceso es establecer relaciones entre las secuencias de imágenes para cada persona en intervalos de tiempo específicos. Es relevante señalar que la frecuencia de captura de imágenes es relativamente baja, aproximadamente 1 FPS, debido a la carga computacional del enfoque. Durante este período, las características físicas y las relaciones entre los habitantes cambian rápidamente en términos de orientación, forma y colores. Por lo tanto, hemos descartado métodos basados en desplazamiento (SHIFT) [274].

En este enfoque, se presenta una solución innovadora de seguimiento no supervisado basada en DL. Se ha utilizado el modelo preentrenado ResNet50 [275] para extraer características relevantes [276] que permiten evaluar la similitud entre candidatos. La última capa abstracta del modelo ResNet contiene características de baja y alta abstracción relacionadas con la rotación y la traslación de cada habitante. Usando la similitud coseno, se calcula una matriz de similitud entre los habitantes anteriores j y los candidatos en la imagen actual i: $S[i][j] = |F_i - F_j|$, donde $S[i][j]$ refleja una puntuación de similitud.

Luego, se aplica un enfoque de búsqueda voraz para encontrar el valor mínimo y sus respectivas posiciones en la matriz S^*, es decir, $S^* = min(S), i^*, j^*$. Si el valor de S^* supera un umbral predefinido, se considera que se trata de un nuevo habitante debido a las diferencias

significativas; de lo contrario, se relaciona al candidato potencial i^* con el habitante anterior j^*. En los siguientes pasos: i) se vacía la fila y la columna correspondientes a i^*, j^*, y ii) se actualizan las características de los habitantes F_k, repitiendo el proceso. Es importante destacar que un habitante puede generar varios candidatos durante el tiempo cuando su forma cambia abruptamente, como cuando se sienta después de caminar. Por lo tanto, los candidatos finales están agrupados en función de las formas de los habitantes y se les traza posteriormente como una sola persona usando el reconocimiento facial, un aspecto que se detalla a continuación.

La Figura 4.11 proporciona un ejemplo ilustrativo del seguimiento de múltiples habitantes en imágenes mediante este enfoque.

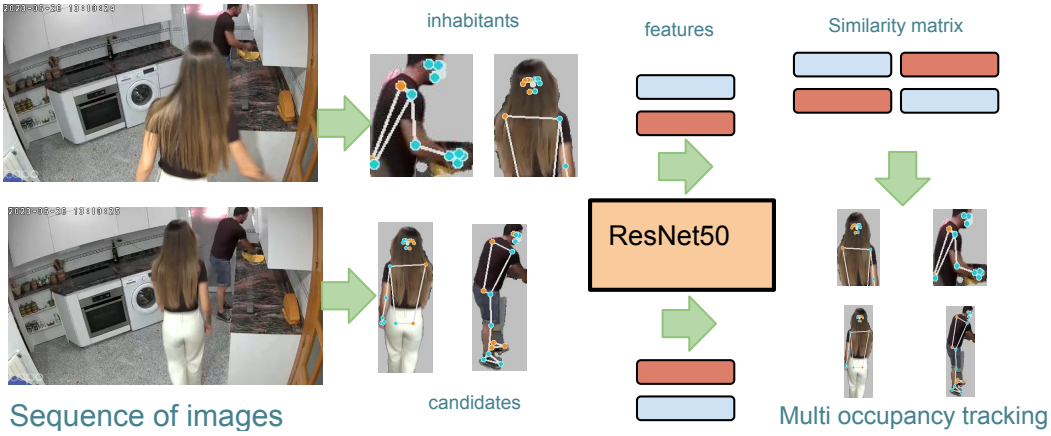


Figura 4.11: Ejemplo de seguimiento de múltiples habitantes mediante imágenes bajo un enfoque no supervisado.

Para abordar el desafío de la identificación facial de personas en entornos con ocupación múltiple, se han integrado las técnicas DeepFace utilizando VGG-Face [267, 277] y YoloFace [266]. Este proceso consta de tres etapas, que se detallan en la Figura 4.12:

i) En primer lugar, como se introdujo al inicio de esta sección, la configuración de este sistema requiere una fase inicial de identificación facial. Para lograr esto, se llevaron a cabo sesiones de identificación facial que tomaron algunos minutos para generar una base de datos facial individualizada para cada habitante antes de evaluar el contexto de multiocupación.

ii) Posteriormente, se aplica YoloFace a las cajas delimitadoras de cada habitante calculadas previamente por YoLo-v7. Cabe destacar que YoloFace está entrenado específicamente para detectar rostros en posiciones frontales y laterales, lo que resulta beneficioso para nuestro enfoque al reducir los falsos positivos en la identificación facial. Gracias al seguimiento de multiocupación, podemos rastrear a los habitantes en la escena y aplicar la identificación facial solo en ciertos puntos de su trayectoria.

iii) Cuando YoloFace detecta un rostro frontal o lateral dentro de la caja delimitadora de la imagen, DeepFace busca en la base de datos facial de los usuarios y proporciona un valor de similitud de coseno para cada imagen. Luego, se aplica un algoritmo de vecino más cercano (K=3) y establecemos un umbral para la similitud de coseno promedio con el fin de lograr un rendimiento sólido en la identificación facial.

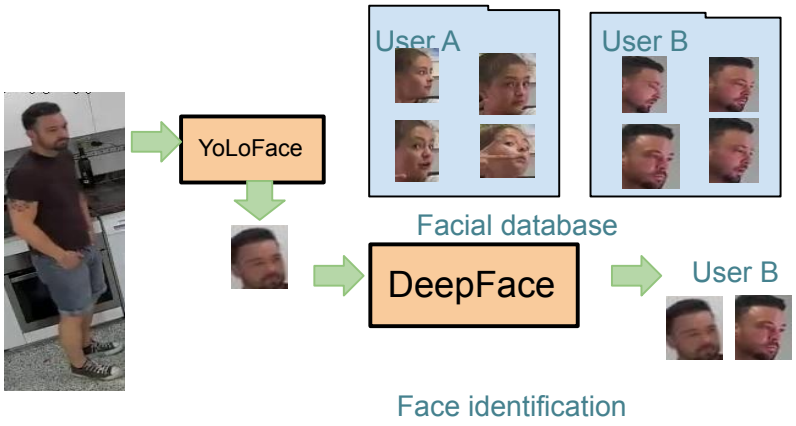


Figura 4.12: Componentes para la identificación de la persona a través de imágenes faciales.

4.3.1.3. Representación de datos en un entorno virtual

En esta sección, se explica el proceso para fusionar la información extraída por los sensores de visión y representarla en un entorno virtual, con el objetivo de preservar la privacidad mientras se realiza una síntesis basada en las imágenes capturadas. Para lograr esto, se ha desarrollado una aplicación utilizando Unity. Unity es una plataforma de desarrollo ampliamente conocida y un motor de juegos que facilita a los creadores diseñar, animar y programar avatares 3D de manera relativamente sencilla.

La aplicación resume la información en una síntesis no invasiva utilizando avatares. Los datos pueden ser cargados desde un archivo de formato o a través de MQTT en tiempo real. La información de los diferentes componentes descritos se muestra en los avatares de la siguiente manera:

- Las ubicaciones en 2D de los habitantes se traducen en coordenadas del mundo real sobre el suelo y se reflejan en la posición del avatar.
- La identificación facial se traduce en un cambio de color y la apariencia de la cabeza del avatar, que varía según el reconocimiento facial realizado.
- Los puntos de referencia corporales en 3D, que describen la postura de los habitantes identificados y rastreados, se utilizan para ajustar la pose del avatar en tiempo real.

Cada avatar desarrollado en Unity representa a un habitante y está vinculado a un "gemelo digital". Estos gemelos digitales proporcionan una representación virtual de una persona del mundo real en el EI. Los gemelos digitales se utilizan principalmente para el análisis, monitoreo y optimización de sistemas del mundo real, así como para el seguimiento de la actividad. En futuros trabajos, los movimientos de los avatares 3D generados por Unity podrán aprenderse a partir de contextos del mundo real, lo que permitirá proporcionar datos sintéticos que reproduzcan el comportamiento de procesos relacionados con el reconocimiento de la actividad. La Figura 4.13 muestra una imagen que compara la representación en Unity con una imagen del mundo real.

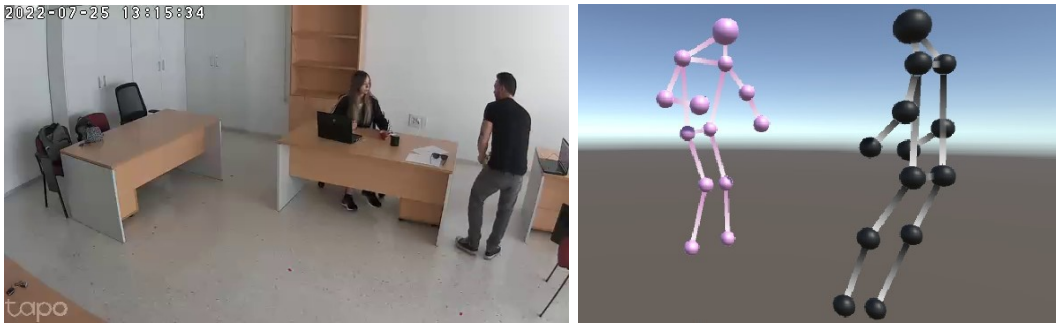


Figura 4.13: Ejemplo de representación de Unity, contexto de oficina (a la izquierda representación real, a la derecha virtual).

4.3.2. Resultados

En esta sección, se aborda el caso de estudio desarrollado en este trabajo para evaluar el desempeño del enfoque en la generación de síntesis no invasivas, que incluye la representación de puntos de referencia 3D del cuerpo, ubicaciones y la identificación a partir de sensores de visión en un EI.

Para llevar a cabo este estudio, se han seleccionado escenarios de la vida real y se ha utilizado una cámara para la captura de imágenes. Se ha posicionado una cámara IP en una esquina de una habitación y dos individuos han participado llevando a cabo trayectos en cinco escenarios diferentes. Estos individuos, un hombre y una mujer, han participado en cinco situaciones distintas que ocurrieron en dos contextos diferentes (cuatro en una sala de estar y una en una oficina), que se describen a continuación:

- Escena 1: En esta escena, la mujer está leyendo en el sofá, y el hombre entra en la habitación para ver la televisión junto a ella.
- Escena 2: En esta escena, el hombre utiliza el teléfono y luego le entrega el teléfono a la otra persona para finalizar la llamada.
- Escena 3: Aquí, el hombre está trabajando en su ordenador cuando la mujer se acerca para hacerle una consulta y conversar.
- Escena 4: En esta escena, los dos protagonistas se sientan en el sofá, toman café y mantienen una conversación.
- Escena 5: La última escena comienza con la mujer y el hombre trabajando en dos ordenadores portátiles y luego él le prepara una bebida.

Este caso de estudio se ha llevado a cabo para evaluar este enfoque en situaciones de la vida real y cómo puede capturar de manera efectiva la pose corporal, las ubicaciones y la identificación de los ocupantes en un EI, proporcionando información importante sobre la eficacia y aplicabilidad de la presente metodología.

En la figura 4.14, se muestra un ejemplo de imágenes recogidas en los dos contextos.

Figura 4.14: Ejemplo de habitantes en dos contextos diferentes (oficina y salón).

Los datos recopilados y el código de los modelos desarrollados están disponibles en: https://github.com/AuroraPR/Real2Virtual3D-UCAmI.

Respecto a los resultados, en primer lugar, se ha evaluado el desempeño del Seguimiento no Supervisado para multiocupación y la eficacia de Yolo en la detección de cuerpos. En la Tabla 4.4, se presentan los resultados, que incluyen el número de fotogramas en los que Yolo detecta a un habitante (formas corporales detectadas), la cantidad de grupos de formas corporales calculados por el seguimiento facial y corporal, o Facial Body Tracking (FBT), en escenarios de ocupación múltiple, así como el accuracy y el error en la separación de los habitantes por parte de FBT (donde el error se refiere al número de formas corporales que se asignan incorrectamente a otros usuarios no relacionados).

Es relevante destacar el sólido rendimiento tanto en la detección de caras como de formas corporales en los fotogramas, gracias a Deep Face y Yolo. Asimismo, se subraya la importancia de FBT para discriminar grupos de candidatos y vincular la identificación facial con precisión, lo que conlleva una mejora significativa en el rendimiento de estas herramientas en comparación con su uso de forma independiente. En cuanto a las métricas de falso positivo o error:

- **Error de FBT:** Esto refleja el número de segundos (FPS=1) en los que un habitante se rastrea erróneamente como otro.
- **No detección de FBT:** Representa la cantidad de grupos de formas corporales que no están vinculados con la identificación facial, lo que significa que se detecta la presencia

de otra persona pero no se sabe quién es.

Tabla 4.4: Formas corporales detectadas, número de clusters, error y precisión para discriminar clusters de candidatos de FBT.

Cara y seguimiento / escenas	**Hombre**	**Mujer**	**Ninguno**	**Total**
Cara identificada(1)	46	74	14	134
FBT(1)	51	79	4	134
Cara identificada(2)	18	14	34	66
FBT(2)	33	32	1	66
Cara identificada(3)	67	16	28	111
FBT(3)	75	34	2	111
Cara identificada(4)	69	33	36	138
FBT(4)	90	48	0	138
Cara identificada(5)	23	88	65	176
FBT(5)	40	136	0	176

	Formas corporales	**N. clústers**	**Error FBT**	**Accuracy FBT**
Escena-1	134	5	7	0,95
Escena-2	66	3	3	0,95
Escena-3	111	3	3	0,97
Escena-4	138	3	5	0,96
Escena-5	176	5	0	1,00

En segundo lugar, se ha realizado una evaluación de la identificación facial proporcionada por YoLo Face y Deep Face en las cinco escenas y dos contextos mencionados. En este proceso, se ha identificado que tres rostros no estaban relacionados con personas reales, sino con otros objetos. Además, solo se produjeron tres falsos positivos en la identificación de usuarios (recall: 1.0, precisión: 0.97, puntuación F1: 0.98). Los resultados detallados de esta evaluación se presentan en la matriz de confusión que se muestra en la Tabla 4.5.

Tabla 4.5: Matriz de confusión del reconocimiento facial de Deep Face en las cinco escenas.

predicho\real	**Mujer**	**Hombre**	**Ninguna**
Mujer	180	0	0
Hombre	3	225	3
Ninguno	0	0	0

Los resultados de esta sección presentan principalmente limitaciones en los siguientes aspectos:

- La evaluación se ha llevado a cabo en un caso de estudio simple y de corta duración que involucra dos habitaciones y dos habitantes, lo que limita la generalización de nuestro enfoque a entornos más complejos con condiciones variables, oclusiones y múltiples ocupantes involucrados en diversas actividades.

- Se requiere abordar la falta de adaptación de nuestro sistema a diferentes tipos de cuerpo, variaciones en la vestimenta y diferencias culturales que pueden afectar la efectividad de la estimación de puntos de referencia corporales y la identificación.

- La complejidad computacional de las herramientas de visión utilizadas podría limitar el rendimiento en tiempo real, especialmente en situaciones con múltiples ocupantes; aunque el caso de estudio proporciona información sobre la velocidad de procesamiento y los requisitos de hardware para una implementación con un FPS de 1, se necesita trabajar en la eficiencia computacional para permitir una implementación en tiempo real más ágil y precisa en futuros desarrollos.

CAPÍTULO 5

RECONOCIMIENTO DE ACTIVIDADES COTIDIANAS MEDIANTE PROFORMAS DIFUSAS EN ENTORNOS DE MULTIOCUPACIÓN

La trazabilidad de las AVD desempeña un papel de importancia crítica en diversos ámbitos, incluyendo el sector de la atención médica, el desarrollo de casas inteligentes y la vida asistida [278]. De una forma general, según la norma *ISO 8402*, la trazabilidad es la *aptitud para rastrear la historia, la aplicación o la localización de una entidad mediante indicaciones registradas*.

Con el avance de tecnologías basadas en sensores binarios y de visión, así como desarrollos en UWB, el seguimiento de la ubicación en interiores ha experimentado un notable progreso, como se ha introducido previamente en el capítulo 4. Esto incluye la implementación de técnicas de aprendizaje automático y la fusión de datos de sensores [279], además de mejoras en el rendimiento para mitigar errores e interferencias de señal [280, 281]. En este sentido, se propone el empleo de la lógica difusa como un paradigma influyente para describir las actividades en EI basado conocimiento [282]. La lógica difusa proporciona capacidades de fusión de datos de sensores [283] y la habilidad de extraer información sobre las actividades diarias a partir de secuencias de datos provenientes de sensores [124].

En el ámbito del RA, los sensores binarios han sido ampliamente utilizados debido a su coste reducido y su capacidad de preservar la privacidad al detectar la presencia o ausencia de una persona en un área específica. A pesar de sus ventajas, se enfrentan a limitaciones en la trazabilidad de las AVD en entornos con múltiples ocupantes, dado que la mayoría de los sensores ambientales no pueden diferenciar entre individuos ni identificar a la persona específica que realiza la actividad [177, 30, 284].

Por otro lado, uno de los puntos clave de la presente propuesta es el modelado de actividades basado en conocimiento mediante lógica difusa y el uso de protoformas para describir conocimiento. Desde los inicios en el campo de RA [283] hasta investigaciones más recientes [52], la lógica difusa ha demostrado ser una metodología efectiva para representar datos procedentes de sensores. Su relevancia se ha incrementado en el contexto de arquitecturas distribuidas de borde y niebla, particularmente en la integración y agregación de datos heterogéneos de sensores [285]. La aplicación de lógica difusa para representar características temporales ha mostrado resultados positivos en varios contextos [18, 253], proporcionando una interpretación comprensible de los datos de bajo nivel obtenidos de los sensores [286] y mejorando la precisión en situaciones de incertidumbre o imprecisión de los datos sensoriales [287]. Zadeh introdujo las protoformas y la lógica difusa como modelos útiles para el razonamiento [288] y la sumarización de datos bajo condiciones de incertidumbre [289, 290]. El uso de protoformas [291] y reglas difusas para la inferencia de conocimiento ha demostrado ser efectivo, ofreciendo representaciones adecuadas y aplicables en diversos contextos [292].

En base a estos dispositivos y modelos basados en conocimiento, el presente capítulo introduce un sistema basado en conocimiento mediante lógica difusa para la identificación de actividades humanas en entornos con múltiples ocupantes. Dicho sistema de basa en la interacción cercana (conocida como *nearby interaction*) haciendo uso de UWB:

- Se emplea una metodología de lógica difusa para modelar de manera precisa la ubicación espacial de áreas de interés, permitiendo así la discriminación de interacciones de eventos de corta duración entre usuarios.

- Se utilizan protoformas lingüísticas junto con reglas difusas para caracterizar eventos de

largo plazo y extraer información relacionada con actividades humanas en contextos de múltiples ocupantes.

- Para respaldar la efectividad de este enfoque, se ha recogido y etiquetado meticulosamente un conjunto de datos multimodal que ilustra su aplicabilidad. Los resultados obtenidos demuestran un rendimiento alentador, con una precisión del 90 % en la capacidad de distinguir entre múltiples ocupantes.

5.1. Metodología

En esta sección, se presenta una descripción formal del modelo difuso propuesto para la extracción de actividades humanas basado en la interacción cercana de la activación de sensores y la proximidad del usuario. El enfoque de esta investogación se centra principalmente en la descripción y discriminación de dos procesos relacionados con las AVD:

i) **Eventos basados en sensores:** Estos eventos se vinculan con la activación a corto plazo de sensores binarios.

ii) **Actividades:** Estas actividades están relacionadas con eventos a largo plazo derivados de la activación de sensores y la posición del usuario.

La Figura 5.1 expone la arquitectura de los componentes que integran la presente propuesta. En una primera etapa, se recopilan las activaciones generadas por los sensores binarios, así como la localización de los usuarios, recolectandos las secuencias de datos generadas por los sensores. Posteriormente, se efectúa la discriminación de las activaciones de los sensores binarios de manera individualizada para cada usuario. Por último, se introducen los eventos de larga duración, previamente definidos mediante protoformas, los cuales representan los sucesos de mayor extensión temporal, a partir de la información de activación suministrada por los sensores.

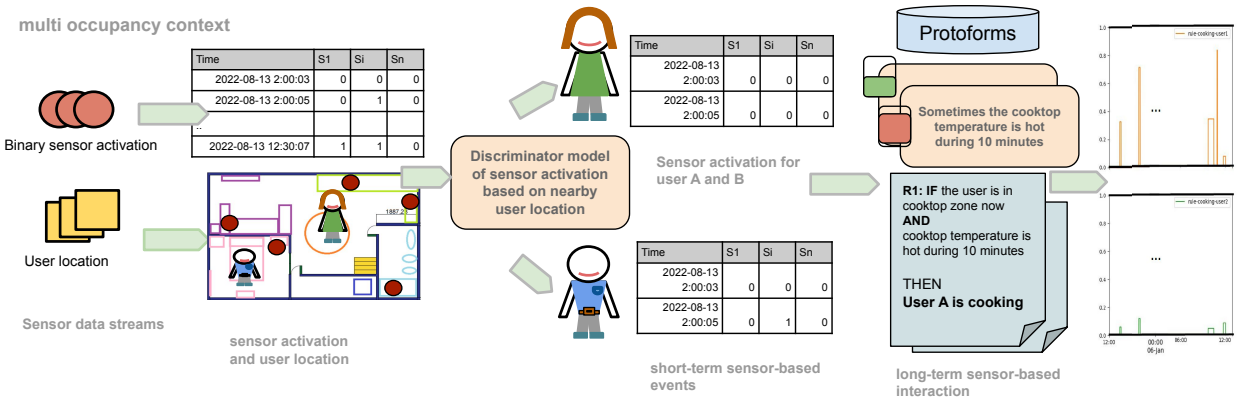


Figura 5.1: Arquitectura de componentes que configuran la propuesta. Inicialmente, el sistema recopila las activaciones de sensores y los datos de ubicación de los usuarios de las secuencias de datos de sensores. Posteriormente, estos datos se procesan para calcular tanto patrones de activación de sensores a corto plazo como eventos a largo plazo para cada usuario.

En las siguientes subsecciones, se describe el proceso de modelización de flujos de datos difusos generados a partir de la interacción cercana del usuario con los sensores del entorno. Posteriormente, se presenta el proceso de discriminación de eventos basados en sensores tanto a corto como a largo plazo, derivados de la mencionada interacción cercana del usuario.

5.1.1. Modelo de discriminación de activación

5.1.1.1. Modelado de flujos de datos difusos a partir de la interacción cercana del usuario en una región de interés

En primer lugar, se establece la definición de un área de posición u_i para un usuario u en cada instante de tiempo t_i mediante una representación geométrica, como un punto, un círculo, una elipse o un cuadro delimitador (rectángulo). Cada uno de estos espacios geométricos introduce una medida aproximada con cierto grado de incertidumbre e imprecisión. Los flujos de datos se caracterizan en instantes de tiempo t_i, los cuales se distribuyen en un intervalo de tiempo especificado, denotado como $[t_0, t_N]$. El intervalo de tiempo se divide en incrementos uniformes de tamaño Δ. Por lo tanto, para un usuario particular, representado como u, su trayectoria se sigue a través de un flujo de posición que puede expresarse como un conjunto ordenado de posiciones $\vec{u} = u_0, u_1, \ldots, u_i$, donde cada u_i corresponde a la posición

en el instante t_i.

En segundo lugar, se define una región de interés, denotada como r, la cual se describe mediante una o varias áreas de posición de interacción $L(r) = \{l_0^r, \ldots, l_i^r\}$. Cada área, representada como l_i^r, se caracteriza mediante un grado de interacción $\overline{l_i^r}$, permitiendo así la flexibilidad en la descripción de múltiples áreas con distintos niveles de interacción. El grado de interacción entre la región de interés r y la ubicación del usuario u_i en el instante t_i se calcula a través de la agregación de las intersecciones de las áreas de posición $\bigcup_{l_j^r}^{L(r)}$, las cuales se ponderan por sus respectivos grados de interacción $\overline{l_j^r}$ de la siguiente manera:

$$L(r)_i^u = L(r) \cap u_i = \bigcup_{l_j^r}^{l_j^s \in L(r)} (\overline{l_j^s} \otimes l_j^s \cap u_i) \tag{5.1}$$

La interacción se define mediante un grado $L(r)_i^u \in [0, 1]$ cuya semántica se relaciona con la intersección espacial entre la ubicación del usuario u en la región de interés r en el instante de tiempo t_i. De esta manera, a partir del flujo de ubicación de un usuario y una región de interés r, somos capaces de calcular un flujo de datos difuso a lo largo del tiempo $\overline{L(r)^u} = L(r)_0^u, \ldots, L(r)_i^u$.

5.1.1.2. Modelado de flujo de datos difusos de sensores ambientales

Un flujo de datos $\vec{s}$ de un sensor s, consiste en un conjunto de mediciones $\vec{s} = s_0, \ldots, s_i$. Cada medición s_i se recopila en un instante de tiempo t_i, el cual está definido en el intervalo de tiempo de interés $[t_0, t_N]$, cuyo incremento de tiempo es Δ.

Bajo un enfoque lingüístico [124], el flujo de datos del sensor $\vec{s}$ se describe empleando términos difusos v para describir las mediciones del sensor de una manera más interpretable. Cada término v está caracterizado por una función de membresía $\overline{v(s_i)}$, la cual define un grado de pertenencia $\overline{s^v}_i \in [0, 1]$ para cada medición s_i. Por ejemplo, los sensores binarios se relacionan directamente con los términos *activo* e *inactivo*, mientras que los sensores de temperatura utilizan términos como *bajo*, *medio* y *alto*. De este modo, a partir de un flujo de datos del sensor $\vec{s}$ y términos lingüísticos v, obtenemos un flujo de datos difuso $\overline{s_v} = \overline{s_{v0}}, \ldots, \overline{s_{vi}}$.

5.1.2. Modelo de reconocimiento de reglas en multiocupación

5.1.2.1. Discriminación de eventos a corto plazo basados en sensores a partir de la interacción cercana del usuario

En esta sección, se detallan los métodos empleados para la discriminación de eventos a corto plazo en el RA basado en la interacción cercana del usuario. Los eventos a corto plazo se vinculan principalmente con la activación de sensores binarios situados en el entorno, encargados de recopilar información sobre la interacción del usuario con electrodomésticos o la apertura de puertas.

Un evento a corto plazo se registra en un instante de tiempo específico t_i cuando la activación de un sensor dado s se verifica con un grado de activación $\overline{s_{Activo_i}} > 0$.

Con el propósito de contextualizar el evento en términos espaciales, se establece una zona de interacción del sensor $L(s)$ para cada sensor binario s de interés. Se calcula el grado de interacción $\overline{s^u_{Activoi}} = L(s) \bigcap u_i$ en el instante t_i, generando así un flujo de activación del sensor para cada usuario u y sensor s. Este flujo, representado por $\overline{sActivo^u} = \overline{s^u_{Activo0}}, \ldots, \overline{sActivo_i{}^u}$, se utiliza para discriminar de manera individual las actividades de cada usuario.

Por lo tanto, el grado de interacción $L(s)i^u > 0$ de un usuario u en una región $L(s)$ cuando se registra una activación del sensor $\overline{sActivo_i} > 0$, determina la relación espacio-temporal del usuario según la siguiente regla:

$$R : \text{IF } L(s)^u_i \text{ and } \overline{s_{Activo_i}} > 0 \text{ THEN } \overline{s^u_{Activo_i}}$$

No obstante, es posible que múltiples usuarios interactúen en el área de un sensor cuando dicho sensor se encuentra activo en el momento t_i. Para abordar esta situación, se aplica una interacción exclusiva de un solo usuario, lo que significa que solamente un usuario puede haber interactuado con el sensor en ese instante. En este contexto, la interacción se relaciona de manera única con un mayor grado de usuario, el cual es definido mediante el operador *sup* y se representa como el valor máximo en este caso. Se han descrito otras políticas de interacción múltiple en investigaciones anteriores [225], y se abordarán en la próxima sección,

específicamente en relación a eventos a largo plazo basados en sensores.

5.1.2.2. Discriminación de eventos a largo plazo basados en sensores a partir de la interacción cercana del usuario

Mediante la aplicación de la lógica difusa, presentamos un modelo de extracción de información de alto nivel relacionada con eventos de larga duración en el contexto del RA, a partir de la activación de sensores, la localización y las interacciones del usuario.

Inicialmente, detallamos el uso de protoformas, las cuales representan un enfoque altamente interpretable, de gran riqueza y expresividad, diseñado para modelar el conocimiento experto con el fin de extraer información de los flujos de datos. A continuación, se proponen las siguientes protoformas específicamente adaptadas para el análisis de flujos de datos provenientes de sensores, las cuales han sido exitosamente incorporadas en investigaciones previas [285, 291]. La estructura de dichas protoformas se describe de la siguiente manera:

$$Q \quad V \quad T$$

En la notación utilizada, se emplean los identificadores Q, V, y T para representar los siguientes términos lingüísticos:

- V hace referencia a la variable que determina el flujo de datos difusos provenientes de los sensores binarios o la posición del usuario.
- T se encarga de definir una Ventanas temporales difusas (Fuzzy Temporal Windows, por sus siglas en inglés) (FTW), denotada como $\overline{T}$, en la cual se agregan los flujos de datos difusos s_v en el dominio temporal $s_v \bigcap T$.
- Q corresponde a un cuantificador que tiene la función de filtrar y transformar el grado de agregación $\overline{Q}(s_v \bigcap T)$.

En este contexto, una protoforma $P = QVT$ se utiliza para definir un nuevo flujo de datos difusos, cuyo grado $\overline{P_i} = \overline{Q}(\overline{V_i} \bigcap \overline{T})$ para cada instante de tiempo t_i, se relaciona con la relevancia temporal del flujo de datos de origen en la FTW transformada por el

cuantificador difuso. En la sección 5.1.2.3 se proporciona una descripción detallada del proceso de agregación de ventanas temporales y cuantificadores a flujos de datos difusos. Un ejemplo concreto de protoforma podría ser "la mayoría del tiempo el horno está caliente durante más de 20 minutos", donde se asignan los siguientes valores a los identificadores: "Q=la mayoría del tiempo", "V=el horno está caliente", y "T=durante más de 20 minutos", teniendo en cuenta que "s=hornoz "v=está caliente".

Para la composición de actividades de alto nivel, que implica la interrelación de protoformas y la discriminación del usuario que lleva a cabo la actividad, se emplea un enfoque basado en reglas con una estructura de tipo IF-THEN de carácter ad hoc. En primer lugar, la unión de las protoformas se lleva a cabo utilizando la t-norma .ANDçon el propósito de modelar los antecedentes, los cuales establecen los intervalos de tiempo en los cuales el grado de verdad de las protoformas es válido. Posteriormente, el consecuente de la regla calcula el grado de interacción para cada usuario dentro de estos intervalos de tiempo, con el fin de determinar quién es el responsable de llevar a cabo la actividad en cuestión.

Una regla R se define de la forma:

$$R : \text{IF } A^1 \ldots \text{ and } A^j \text{ in } T_R \text{ THEN } Q_R(\text{usuario u R })$$

Donde el grado de pertenencia de antecedente $A^1 \ldots$ y A^j en T_R se calcula de la siguiente manera:

- Cada antecedente A^j se corresponde con una protoforma previamente definida, y se utiliza el operador .AND"para calcular una unificación de antecedentes, generando un flujo de datos difusos $\overline{R_i} = \overline{A_i^1} \cap \overline{A_j^1}$ para la regla R en cada marca de tiempo t_i.
- Una ventana temporal difusa T_R se encarga de concatenar el grado del antecedente $\overline{R_i}$ a lo largo del tiempo. Para mantener la interpretabilidad, es esencial reducir el número de intervalos de tiempo, agrupando aquellos que se encuentran en proximidad. La FTW $\overline{T_R}$ se encarga de definir esta proximidad temporal entre los intervalos de unión. Los detalles sobre el método de concatenación de un flujo de datos difusos mediante una FTW se encuentran descritos en la sección 5.1.2.3.

- Los antecedentes proporcionan intervalos de tiempo $\Delta_R = (t_0^-, t_0^+), \ldots, (t_i^-, t_i^+)$, los cuales están definidos por el punto inicial t_0^- y el punto final de tiempo t_0^+. Estos intervalos se calculan en función de un grado $\alpha - cut > 0$ que identifica las marcas de tiempo t_i pertenecientes a los intervalos de tiempo (t_i^-, t_i^+) donde $\overline{R_i} > 0 \forall t_i, t_i \in (t_i^-, t_i^+)$.

Una vez que se determinan los intervalos de tiempo (t_i^-, t_i^+) de activación de los antecedentes, se computa el grado de verdad del consecuete. Para ello, el consecuente procede a calcular el grado de interacción de cada usuario con el fin de determinar quién está llevando a cabo la regla R. El cálculo del grado de interacción de un usuario u con la regla R se lleva a cabo de la siguiente manera:

- Cada regla R se encuentra asociada a una región de interés denominada $L(R)$.
- En el contexto de eventos de larga duración, se permite la posibilidad de múltiples interacciones de usuarios, en las cuales varios usuarios pueden haber participado en la ejecución del evento. En esta situación, el grado de interacción de un usuario u en el intervalo de tiempo (t_i^-, t_i^+) se calcula agregamdp el grado de interacción del usuario con la región de interés $\overline{R_i} \cap u_i$, basado en las marcas de tiempo t_i correspondientes al intervalo de tiempo en cuestión:

$$R_{(t_i^-, t_i^+)} \cap u = \bigcup_{t_i}^{t_i \in (t_i^-, t_i^+)} L(s)_i^u \otimes \overline{R_i}) \tag{5.2}$$

- Es importante destacar que en algunos casos, varios usuarios pueden estar vinculados al sensor de la regla R durante el mismo intervalo de tiempo (t_i^-, t_i^+). Esto se refleja cuando se cumple la condición $R_{(t_i^-, t_i^+)} \cap u_A > 0, R_{(t_i^-, t_i^+)} > 0 \cap u_B, u_A \neq u_B$, lo que indica que los usuarios u_A y u_B están involucrados en la ejecución de la regla R en ese intervalo de tiempo específico.
- Finalmente, se aplica un cuantificador Q_R al grado de interacción del usuario, denotado como $\overline{Q_R}(R_{(t_i^-, t_i^+)} \cap u)$, con el propósito de determinar un valor mínimo y ajustar el grado de interacción final en función de criterios expertos.

5.1.2.3. Proceso de agregación de ventanas temporales y cuantificadores a flujos de datos difusos

Se ha incorporado una sección para proporcionar una comprensión completa de los principios matemáticos y métodos empleados de este capítulo. Esta sección cumple la función de ser un recurso complementario, donde se presenta una compilación de ecuaciones y expresiones que desempeñan un papel esencial en el respaldo del marco teórico de la presente investigación.

- **La operación $V \cup T$ denota la agregación de una ventana temporal difusa T y un flujo de datos difusos $V = s_v$.** La función de pertenencia de la FTW se define temporalmente en función de la distancia $\Delta t_i^* = t^* - t_i, t^* > t_i$ desde un tiempo actual t^* hasta otras marcas de tiempo del flujo de datos t_i. Para cada marca de tiempo t^*, se lleva a cabo la agregación de los grados de los términos en relación con el grado temporal difuso, utilizando los operadores t-norma y co-norma:

$$V \cup T(t^*) = \bigcup_{(v_i,t_i)}^{V} \overline{v_i} \cap \overline{T}(\Delta t_i^*) \in [0,1] \tag{5.3}$$

  El promedio ponderado para multimodalidades se define como funciones de agregación utilizadas para la agregación de FTW en flujos de datos provenientes de sensores, como se describe en [285].

$$V_r \cup T_k(t^*) = \frac{1}{\sum \overline{T}(\Delta t_i^*)} \sum_{(v_i,t_i)}^{V} \overline{v_i} \times \overline{T}(\Delta t_i^*) \in [0,1] \tag{5.4}$$

- **Cuantificador Q.** Un cuantificador aplica una transformación utilizando una función de pertenencia $\overline{Q} : [0,1] \to [0,1]$ con el propósito de modificar y modelar el grado de origen $\overline{Q}(x)$. Este proceso implica la determinación de un valor mínimo denominado $\alpha - cut$ y la posterior rectificación del grado deseado.

- **La operación** $V \cap T$ se refiere a la concatenación de un flujo de datos difusos V

utilizando una FTW T. Este proceso implica la reducción de los intervalos de tiempo más cercanos (t_i^-, t_i^+) en el flujo de datos difusos antes de aplicar un α-cut para obtenerlos. Los principales resultados del método son los siguientes:

i) Se realiza la concatenación de los grados internos de los intervalos de tiempo más cercanos, es decir, aquellos con $v_i > 0$.

ii) No se extiende el punto inicial t_0^- y el punto final de tiempo t_0^+ en el proceso de concatenación.

Para evaluar el punto ii), es necesario que la FTW $\overline{T}$ esté definida por una función de pertenencia con una forma izquierda o derecha. En este contexto, se asume que la FTW T_t corresponde a una función left-shoulder, y $T_t^(-1)$ es la función right-shoulder simétrica. Se tratan de funciones trapezoidales que incluyen solo la parte derecha o izquierda del trapecio. La concatenación se define mediante la unión de las ventanas temporales $(V \cup T) \cap (V \cup T_t^(-1))$, donde la operación de unión $(\cup)$ se calcula como el promedio ponderado, tal como se describió previamente, y la función de corte (α-cut) se establece como el valor mínimo $cap = min$.

- Las funciones mencionadas, $T[l_1, l_2, l_3, l_4](x)$**,**$L[l_1, l_2](x)$**,**$R[l_1, l_2](x)$, corresponden a funciones trapezoidales, left y right shoulder, respectivamente. La función $T[a, b, c, d](x)$ es una función de pertenencia trapezoidal ampliamente reconocida, la cual se define mediante un límite inferior l_1, un límite superior l_4, un límite de soporte inferior l_2, y un límite de soporte superior l_3, como se describe en la Ecuación (5.5). Las funciones trapezoidales left y right shoulder, denotadas como L y R respectivamente, se definen de la siguiente manera: $L[l_1, l_2](x) = T[l_1, l_2, l_2, l_2](x)$ y $R[l_1, l_2](x) = T[l_1, l_1, l_1, l_2](x)$.

$$T(x)[l_1, l_2, l_3, l_4] = \begin{cases} 0 & x \leq l_1 \\ (x - l_1)/(l_2 - l_1) & l_1 < x < l_2 \\ 1 & l_2 \leq x \leq l_3 \\ (l_4 - x)/(l_4 - l_3) & l_3 < x < l_4 \\ 0 & l_4 \leq x \end{cases} \tag{5.5}$$

5.2. Caso de estudio experimental

En esta sección, se proporciona una descripción detallada del caso de estudio y la configuración experimental desarrollados en el marco de este capítulo, con el propósito de evaluar y presentar la aplicación de la metodología en un contexto de la vida real. El caso de estudio se ha llevado a cabo en una cocina donde dos habitantes (un hombre de 71 años y una mujer de 70 años) realizan AVD en un entorno cotidiano. La interacción de estos dos adultos con los electrodomésticos y muebles de la cocina, que incluyen cubertería, nevera, microondas y utensilios de cocina, se ha monitorizado utilizando sensores UWB y sensores ambientales. La elección de la cocina como espacio de estudio se basa en su relevancia y complejidad, ya que es un lugar donde los residentes realizan actividades normalmente con múltiples ocupantes.

El caso de estudio se ha llevado a cabo en una cocina con un área de 29 m^2. Para cubrir este espacio, se han desplegado seis anclas UWB, cuatro sensores de apertura/cierre, dos sensores de presencia y dos cámaras IP, como se muestra en la Figura 5.2. Se han recopilado datos con el propósito de generar dos conjuntos de datos:

- **Conjunto de datos de configuración:** Este conjunto de datos se ha utilizado para definir las áreas de interacción y el comportamiento en función de los datos observados. Ha sido recopilado de forma independiente al conjunto de datos de evaluación y abarca un período de dos días, desde el 20/12/2022 a las 15:00 a. m. hasta el 23/12/2022 a las 15:00 a. m. Incluye registros de actividades como desayuno, cocina, almuerzo y cena. Este conjunto de datos contiene información detallada sobre 220 interacciones de los

sensores de apertura/cierre, cuyos detalles se presentan en la tabla 5.1.

- **Conjunto de datos de evaluación:** Este conjunto se ha empleado para evaluar los métodos propuestos, utilizando las áreas de interacción de sensores definidas en el conjunto de datos de configuración. A diferencia del conjunto de datos de configuración, este conjunto no ha sido observado por expertos y se ha recopilado con fines de evaluación. Cubre un período de dos días, desde el 04/01/2023 a las 12:00 a. m. hasta el 06/01/2023 a las 12:00 a. m., e incluye actividades de desayuno, cocina, almuerzo y cena. Contiene información detallada sobre 238 interacciones de los sensores de apertura/cierre, que se describen en la tabla 5.2. Además, en cuanto a los datos de ubicación UWB, se recolectaron 114.559 muestras del usuario 1 y 110.254 del usuario 2.

Las cámaras IP han registrado imágenes pixeladas a intervalos de 3 segundos, con el propósito de etiquetar los datos recopilados por los sensores desde la perspectiva de un observador externo de forma no invasiva. Las etiquetas obtenidas a través de las imágenes han permitido discriminar la activación del usuario en las interacciones registradas, tal como se describe en detalle en las Tablas 5.2 y 5.1.

Tabla 5.1: Discriminación espacio-temporal de la activación de sensores para los usuarios 1 y 2 (conjunto de datos de configuración).

Habitante	**Frigorífico**	**Cubiertos**	**Microondas**	**Utensilios**
Usuario 1	49	54	6	5
Usuario 2	37	50	7	10
Nadie	0	2	0	0
Total	86	106	13	15

Tabla 5.2: Discriminación espacio-temporal de la activación de sensores para los usuarios 1 y 2 (conjunto de datos de evaluación).

Habitante	**Frigorífico**	**Cubiertos**	**Microondas**	**Utensilios**
Usuario 1	66	41	21	10
Usuario 2	36	50	9	7
Nadie	1	1	0	1
Total	98	92	30	18

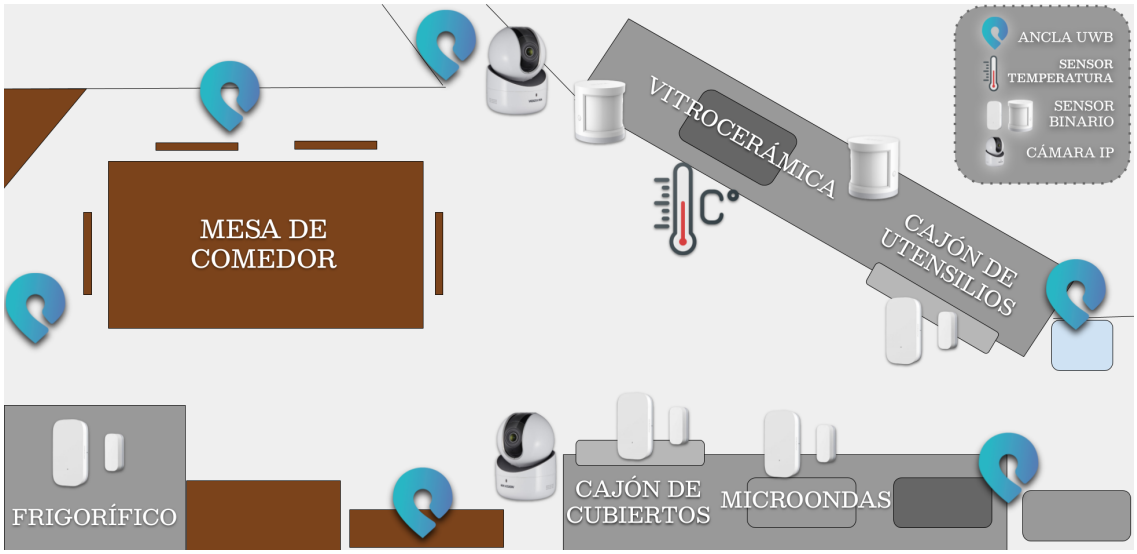


Figura 5.2: Plano de la cocina.

Figura 5.3: Implementación en una cocina que incluye cámaras, sensores ambientales y UWB.

5.2.1. Despliegue del sistema

5.2.1.1. Sensores ambientales

Para la monitorización de la interacción de los habitantes con los electrodomésticos y muebles en la cocina, se ha integrado un conjunto de sensores ambientales. Concretamente, se emplearon sensores de apertura/cierre y temperatura Xiaomi Aqara (https://www.aqara.com/), los cuales se integraron a través de la plataforma Home Assistant (https://www.home-assistant.io/). La recopilación de datos en tiempo real se ha efectuado mediante el protocolo MQTT, un protocolo de comunicación ligero y eficiente diseñado para la transmisión de datos en tiempo real entre dispositivos de IoT. Los sensores se colocaron estratégicamente en electrodomésticos y muebles clave, como se muestra en la ubicación de los sensores en la Figura 5.2.

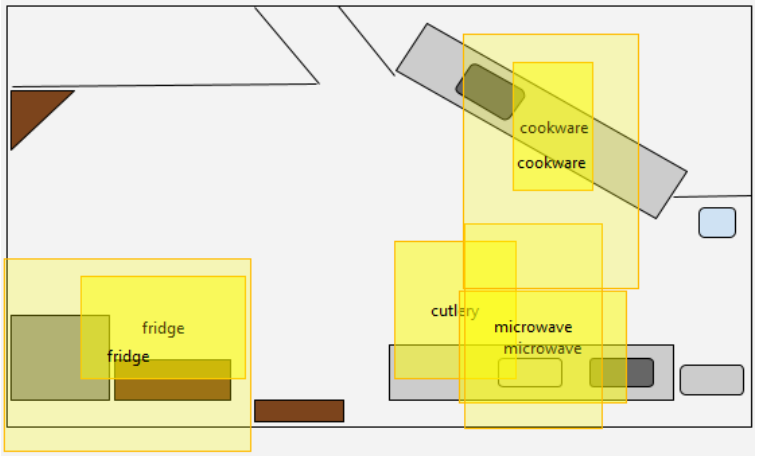


Figura 5.4: Áreas de interacción relacionadas con reglas difusas y localización.

5.2.1.2. RTLS para interacción cercana basado en UWB

Un RTLS es una tecnología que se utiliza para rastrear y determinar la ubicación en tiempo real de objetos o personas dentro de un área específica. Este sistema ofrece información precisa y continua sobre la posición.

En el marco de esta investigación, se ha implementado una solución de seguimiento de

ubicación para monitorizar la actividad de los residentes en el entorno de una cocina. Para ello, se ha utilizado la tecnología UWB proporcionada por https://www.pozyx.io. En dicho entorno, se dispusieron estratégicamente cuatro anclas en las paredes de la cocina para habilitar la localización bidimensional (coordenadas x, y) de los habitantes. Estos usuarios portaban etiquetas UWB a modo de colgante, tal como se muestra en la Figura 5.3. Esta plataforma ofrece una estimación en tiempo real de la ubicación mediante el uso del protocolo MQTT y trilateración.

5.2.2. Definición de la interacción cercana del usuario en la región de interés

Siguiendo la descripción detallada en la sección 5.1, en esta sección se expone la implementación de los métodos utilizados para el sistema de localización en tiempo real UWB y los sensores ambientales desplegados.

- **Definición de la ubicación del usuario:** La ubicación de un usuario se establece mediante una caja delimitadora, denotada como $u_i = (x_i^-, y_i^-), (x_i^+, y_i^+)$, en un instante de tiempo t_i. Para calcular esta caja delimitadora que define la ubicación del usuario, se emplean operaciones de agregación mínimo-máximo sobre las coordenadas 2D estimadas (x_i, y_i) por el sistema de localización en el intervalo de tiempo t_i para cada usuario u. Así, se define $u_i = (min(x_0, \ldots, x_i), min(y_0, \ldots, y_i)), (max(x_0, \ldots, x_i), max(y_0, \ldots, y_i))$.

- **Grado de interacción:** El grado de interacción $(r)_i^u i$ en un momento específico t_i entre las cajas delimitadoras que representan las áreas de ubicación l_j^r y las cajas delimitadoras que definen la ubicación del usuario u_i se fundamenta en la métrica de intersección sobre unión $J(A, B) = \frac{A \cap B}{A \cup B}$. En este contexto, se adapta dicha métrica para otorgar peso exclusivamente a la intersección con el área de ubicación del usuario, es decir, $J(l_j^r, u_i) = \frac{l_j^r \bigcap u_i}{u_r}$:

$$L(r)_i^u i = \sum_{l_j^r}^{l_j^r \in L(r)} \overline{l_j^r} \cdot \frac{l_j^s \bigcap u_i}{u_i} \tag{5.6}$$

5.2.3. Discriminación de eventos a corto plazo en la cocina a partir de la interacción cercana del usuario

Los métodos descritos en la sección 5.1.2.1 de este caso de estudio posibilitan la discriminación de eventos a corto plazo basados en sensores, como la apertura de puertas. En primera instancia, se define la interacción cercana del usuario mediante áreas de interacción de sensores $L(s)$ para cada sensor de apertura de puertas s. La Figura 5.4 proporciona una descripción detallada de las áreas de interacción de sensores, las cuales se han definido de acuerdo a criterios establecidos por expertos humanos y a partir de los datos observables contenidos en el conjunto de datos de configuración.

Basado en estas áreas de interacción, se aplica la siguiente regla difusa para calcular el grado de interacción $L(s)i^u > 0$ para cada usuario u en una región $L(s)$ en el momento en que se produce una activación del sensor $\overline{sActivo_i} > 0$:

$$R : \text{IF } L(s)_i^u \text{ and } \overline{s_{Activo_i}} > 0 \text{ THEN } sup(\overline{s^u_{Activo_i}})$$

En este caso de estudio, se consideran dos usuarios designados como $u = u_1, u_2$. En este contexto, u_1 representa al hombre, mientras que u_2 corresponde a la mujer. Los sensores involucrados son $s =$ cubiertos, frigorífico, microondas, utensilios de cocina.

5.2.4. Discriminación de eventos a largo plazo en la cocina a partir de la interacción cercana del usuario

En esta sección, se detalla el método propuesto para describir actividades a largo plazo en la cocina basadas en la interacción cercana del usuario. Se ha analizado únicamente el conjunto de datos de configuración para definir reglas, protoformas y áreas de interacción según criterios expertos que se basan en datos de ubicación, activación de sensores e imágenes pixeladas de cámaras. En este caso de estudio, se identifican las siguientes actividades a largo plazo: *cocinar* y *sentarse a comer*. La regla de *cocinar* se activa cuando la temperatura en la cocina alcanza niveles elevados, el sensor de presencia en la cocina se encuentra activo, y

el usuario se halla en la zona de la cocina. La regla de *sentarse a comer* se activa cuando el usuario permanece en la zona de la mesa durante un período aproximado de 20 minutos.

Para modelar estas reglas, se han definido directamente las protoformas y las funciones de pertenencia especificadas en la Tabla 5.3. Es importante mencionar que las funciones de pertenencia están definidas utilizando funciones left-shoulder, right-shoulder y trapezoidales, explicadas en detalle en la sección 5.1.2.3. La región de interés para cada regla y la representación de las funciones de pertenencia se muestran en las Figuras 5.5 y 5.6, respectivamente.

Tabla 5.3: Protoformas y reglas para actividades a largo plazo en la cocina.

Id	Protoforma	Q V T		
P1	A veces la temperatura de la vitrocerámica es alta durante 10 minutos	Q = A veces	V ='la temperatura de la vitrocerámica es alta'	T='al menos 10 minutos'
P2	A veces la presencia en la cocina está activa durante 10 minutos	Q = A veces	V ='la presencia en la cocina está activa'	T='al menos 10 minutos'
P3	El usuario u está en la zona de la cocina ahora	$Q = \emptyset$	V ='el usuario u está en la zona de la cocina'	T='ahora'
P4	Principalmente el usuario u está en la zona de la mesa durante aproximadamente 20 minutos	Q = A veces	V ='el usuario u está en la zona de la cocina'	T='ahora'

ID	Regla	Antecedentes
cocinar	El usuario u está cocinando	P1 Y P2 Y P3 durante unos 5 minutos
sentarse a comer	El usuario u está sentado en la mesa	P4 durante unos 5 minutos

Término	Función de pertenencia
A veces	$\overline{Q} = L[0, 0.5](x)$
Ninguno	$\overline{Q} = L[0, 1](x) = x$
Principalmente	$\overline{Q} = L[0.25, 0.75](x) = x$
Temperatura alta	$\overline{v} = R[21°, 25°](x)$
Durante 10 minutos	$\overline{T} = T[-5m, -3m, 3m, 5m](t)$
Ahora	$\overline{T} = T[-15s, 0, 0, 15s](x)$
Alrededor de 20 minutos	$\overline{T} = T[-10m, -8m, 8m, 10m](t)$

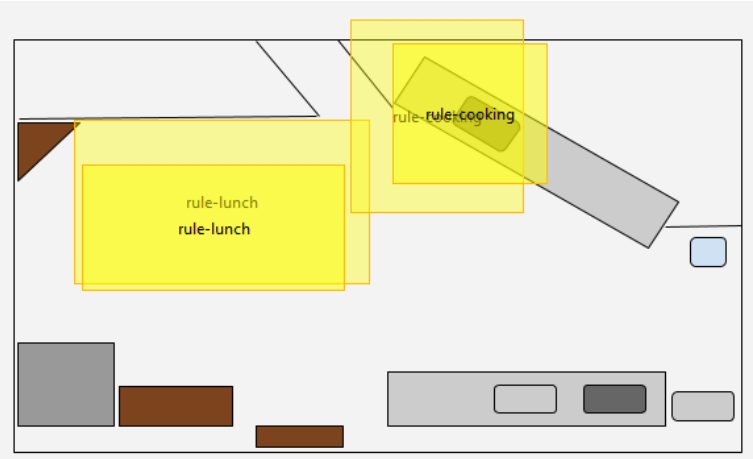


Figura 5.5: Región de interés definida para cada regla.

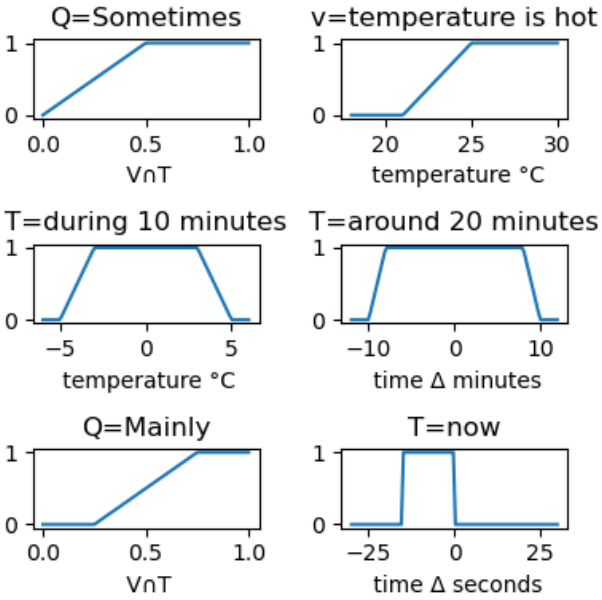


Figura 5.6: Representación de funciones de pertenencia para los términos lingüísticos.

5.3. Resultados

La presente sección expone los resultados y la discusión correspondiente respecto a los datos recopilados en el caso de estudio desarrollado, basado en la interacción cercana del usuario en las actividades cotidianas de la cocina. Se aborda tanto la discriminación de las

actividades a corto plazo basadas en sensores como las actividades a largo plazo basadas en sensores.

El etiquetado de los datos se ha llevado a cabo utilizando una cámara ubicada en la cocina, cuyas imágenes fueron procesadas por un observador externo. Para preservar la privacidad de los usuarios, se ha reducido la resolución de las imágenes capturadas y se han pixelado, siguiendo un método que ha demostrado ser adecuado en la literatura científica [293, 294, 295]. Este proceso se ha efectuado para distinguir de manera efectiva al usuario, la actividad que realiza y su ubicación. La Figura 5.7 presenta un ejemplo de los fotogramas obtenidos durante el caso de estudio. Tanto las actividades a corto plazo como las actividades a largo plazo basadas en sensores han sido etiquetadas por un observador, que identificó al usuario responsable de la actividad y el momento en que esta se produjo. Por último, los datos generados por el modelo difuso propuesto se comparan y evalúan junto con las actividades etiquetadas por el observador. Cabe destacar que el intervalo de tiempo se ha establecido en $\Delta = 15s$, lo que influye en la agregación de posiciones dentro de las cajas delimitadoras y en la granularidad de la activación de los sensores binarios.

Figura 5.7: Ejemplo de una imagen capturada con la cámara y el etiquetado dado por el observador.

Los datos recopilados, así como la implementación del método y los resultados obtenidos en este capítulo, están disponibles para su acceso en el siguiente enlace: https://github.com/AuroraPR/AMALTEA-IoT.

5.3.1. Discriminación de usuarios en eventos basados en sensores de apertura/cierre

En lo que respecta a la discriminación de las actividades basadas en sensores de apertura/-cierre, es importante destacar que se establecieron dos conjuntos de datos para este propósito: i) un conjunto de datos de configuración destinado a definir áreas de interacción y reglas utilizando criterios de expertos, y ii) un conjunto de datos de evaluación que comprende datos no observables, como se detalló al inicio de la sección 5.2.

En relación al conjunto de datos de configuración, la Tabla 5.5 presenta los resultados de precisión, recall y f-score para los usuarios 1 y 2 en el desarrollo de actividades de apertura/-cierre relacionadas con el frigorífico, los cubiertos, el microondas y los utensilios de cocina. En promedio, se logró un rendimiento global de 0.96, con una dispersión reducida tanto en precisión como en recall.

Tabla 5.4: Precisión, recall y f-score de los usuarios 1 y 2 para las actividades de apertura/cierre (conjunto de datos de configuración).

	usuario	**precision**	**recall**	**f1-score**
frigorífico	1	0.96	0.96	0.96
	2	0.97	0.95	0.96
cubiertos	1	0.95	1.00	0.97
	2	1.00	0.94	0.97
microondas	1	1.00	1.00	1.00
	2	1.00	1.00	1.00
utensilios	1	0.83	1.00	0.91
	2	1.00	0.90	0.95
total	1	0.95	0.98	0.97
	2	0.99	0.94	0.97
	avg	0.97	0.96	0.96

En segundo lugar, en el conjunto de datos de evaluación, se presentan en la Tabla 5.5 los

resultados de precisión, recall y F1-score para los usuarios 1 y 2 en el desarrollo de actividades de apertura/cierre relacionadas con el frigorífico, los cubiertos, el microondas y los utensilios de cocina. El rendimiento global es de 0.9 en promedio, con una dispersión reducida tanto en precisión como en recall. Las diferencias en las métricas de discriminación son similares para los usuarios 1 y 2 (0.92 y 0.89, respectivamente). Se observa un rendimiento alentador en datos no observables respecto a la configuración de inicio definida según criterios de expertos.

En la Tabla 5.8, se proporciona una matriz de confusión detallada para cada elemento clave evaluado, lo que permite identificar diferencias significativas en el uso diario, pero no variaciones sustanciales en el rendimiento de discriminación de usuarios del modelo propuesto. Además, la Figura 5.9 ofrece una representación gráfica de la activación de eventos de apertura/cierre y la discriminación desarrollada para cada usuario, ilustrando los resultados del modelo propuesto descrito en esta sección.

Tabla 5.5: Precisión, recall y f-score de los usuarios 1 y 2 para las actividades de apertura/cierre (conjunto de datos de evaluación).

	usuario	**precision**	**recall**	**f1-score**
frigorífico	1	0.94	0.94	0.94
	2	0.91	0.86	0.89
cubiertos	1	0.92	0.83	0.87
	2	0.89	0.82	0.85
microondas	1	0.88	1.00	0.93
	2	1.00	0.67	0.80
utensilios	1	0.91	1.00	0.95
	2	1.00	0.86	0.92
total	1	0.92	0.94	0.93
	2	0.89	0.86	0.88
	avg	0.92	0.88	0.90

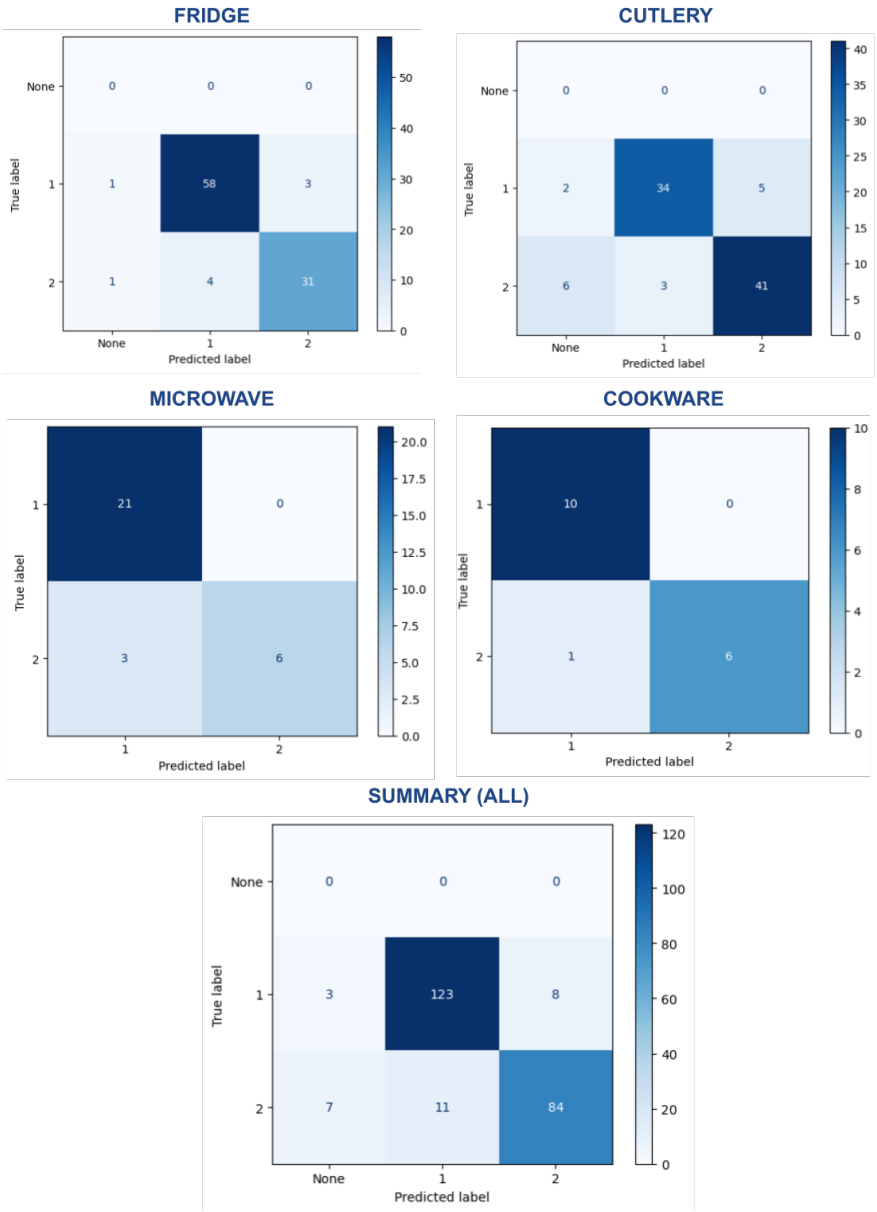


Figura 5.8: Matriz de confusión de los sensores binarios en el conjunto de datos de evaluación.

En la Tabla 5.2, se proporcionan detalles sobre algunos casos especiales detectados durante el caso de estudio, que están relacionados con la fila "ninguno" que merecen ser mencionados: a) En el caso de los cubiertos, se detectó una falsa activación del sensor; b) En el frigorífico, se identificó una falsa activación del sensor; c) Se observaron 4 activaciones de sensores en las que el habitante no llevaba la etiqueta cuando se produjeron; d) En relación a los utensilios de cocina, se registró una falsa activación del sensor. Cabe destacar que existieron situaciones

en los que ambos habitantes podrían haber desarrollado la activación debido a la proximidad de ambos al realizar la actividad. Un ejemplo de esto se puede observar en la Figura 5.7, específicamente en el cuadro B de la misma, que corresponde a la marca de tiempo 2023-01-06 11:52:30. La Figura 5.9 muestra una representación gráfica de los resultados para cada activación del sensor.

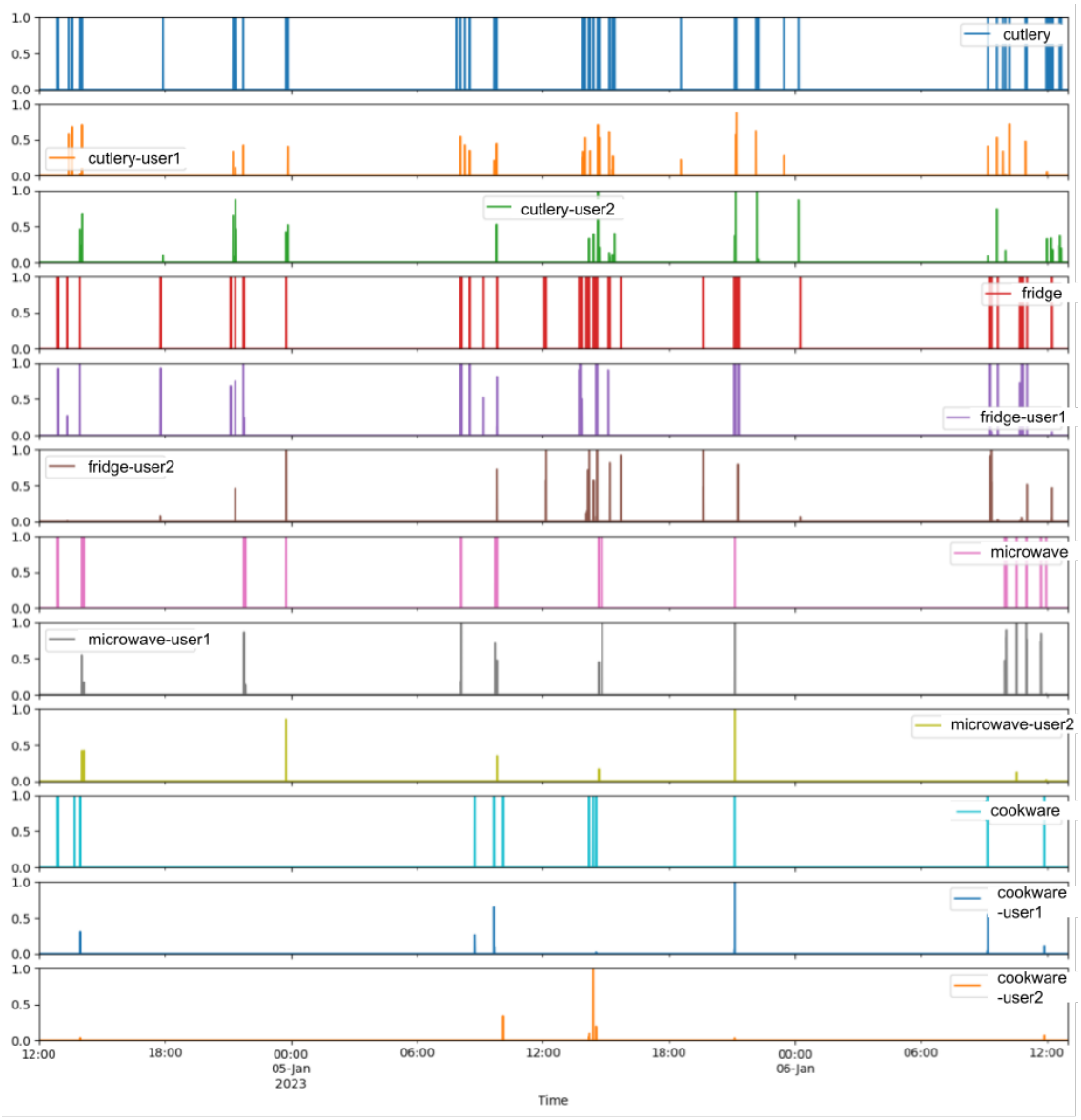


Figura 5.9: Por cada activación del sensor, se tiene el dato en bruto, así como el grado de interacción calculado para el usuario 1 y el usuario 2 mediante el método de discriminación espacial-temporal.

Además de nuestra propuesta, se han evaluado dos modelos diferentes para discriminar la interacción de eventos a corto plazo:

- **Distancia mínima:** Este enfoque se centra en la identificación de usuarios utilizando el método del vecino más cercano, que calcula la distancia entre el Usuario 1 y el Usuario 2 en relación con el sensor activado. La distancia mínima entre los usuarios y los electrodomésticos se calcula mediante la distancia euclidiana. En este caso, no requiere una fase de aprendizaje, por lo que el rendimiento se calcula de manera directa con el conjunto de datos de evaluación, obteniendo una precisión del 0,86.

- **Clasificación de cajas delimitadoras BB:** Este enfoque basado en datos utiliza la señal del sistema de posicionamiento para discriminar la actividad del usuario. Se basa en la propuesta de los autores [246], donde el RSSI se computa mediante SVM y RF para discriminar la interacción del usuario. En el contexto de este trabajo, las cajas delimitadoras con la ubicación del usuario que se obtienen a partir del sistema UWB componen la entrada del modelo. Evaluamos dos configuraciones:

  - En primer lugar, entrenamos y evaluamos al modelo en el conjunto de datos de evaluación, con una división de datos del 20 % para prueba y 80 % para entrenamiento. Los resultados de SVM y RF obtuvieron una precisión del 0,82 y del 0,87, respectivamente. Sin embargo, cabe destacar que existe un ligero sobreajuste al evaluar este conjunto de datos debido a su contexto específico y su tamaño limitado (238 eventos).

  - En segundo lugar, entrenamos el modelo utilizando el conjunto de datos de configuración y luego evaluamos con el conjunto de datos de prueba. Los resultados estuvieron por debajo del rendimiento de nuestro sistema, con SVM logrando una precisión máxima del 0,57 y RF un 0,49.

En las Tablas 5.6 y 5.7, presentamos un resumen del rendimiento de varios métodos explorados en la literatura aplicados a nuestro problema.

Tabla 5.6: Precision, recall y f1-score de BB+SVM, BB+RF y distancia mínima (conjunto de datos de prueba)

		BB+SVM			BB+RF			min-distance		
	usuario	**precision**	**recall**	**f1-score**	**precision**	**recall**	**f1-score**	**precision**	**recall**	**f1-score**
cubiertos	1	1,00	0,33	0,50	0,67	0,86	0,75	0,88	0,68	0,77
	2	0,67	1,00	0,80	0,90	0,75	0,82	0,66	0,87	0,75
frigorífico	1	0,91	1,00	0,95	0,88	1,00	0,94	0,94	0,92	0,93
	2	1,00	0,78	0,88	1,00	0,60	0,75	0,86	0,89	0,87
microondas	1	0,83	0,62	0,71	1,00	1,00	1,00	0,95	0,87	0,91
	2	0,00	0,00	0,00	1,00	1,00	1,00	0,67	0,86	0,75
utensilios	1	1,00	0,60	0,75	0,75	1,00	0,86	1,00	0,83	0,91
	2	0,33	1,00	0,50	0,00	0,00	0,00	0,71	1,00	0,83
total	1	0,92	0,72	0,80	0,83	0,97	0,89	0,93	0,82	0,87
	2	0,65	0,89	0,75	0,93	0,68	0,79	0,74	0,88	0,80
	avg	0,82	0,78	0,78	0,87	0,86	0,85	0,86	0,84	0,85

Tabla 5.7: Precision, recall y f1-score de BB+SV y BB+RF (conjunto de datos de configuración para entrenamiento y conjunto de datos de prueba para evaluación)

		BB+SVM			BB+RF		
	usuario	**precision**	**recall**	**f1-score**	**precision**	**recall**	**f1-score**
cubiertos	1	0,06	0,02	0,03	0,27	0,34	0,30
	2	0,45	0,66	0,54	0,33	0,26	0,29
frigorífico	1	0,73	0,84	0,78	0,77	0,58	0,66
	2	0,63	0,47	0,54	0,49	0,69	0,57
microondas	1	0,75	1,00	0,86	0,00	0,00	0,00
	2	1,00	0,22	0,36	0,74	0,81	0,77
utensilios	1	0,00	0,00	0,00	0,00	0,00	0,00
	2	0,41	1,00	0,58	0,50	0,60	0,55
total	1	0,63	0,55	0,59	0,55	0,54	0,55
	2	0,50	0,58	0,53	0,42	0,40	0,41
	avg	0,57	0,56	0,57	0,49	0,48	0,49

La diferencia en el rendimiento entre los enfoques de BB+SVM y BB+RF en las Tablas 5.6 y 5.7 radica en la variación de las actividades y el contexto entre los usuarios en las escenas que conforman los dos conjuntos de datos. La definición de regiones de interés permite modelar de manera sencilla la interacción y la discriminación de usuarios de manera más flexible y adaptable en comparación con los modelos basados en datos, que están diseñados para el contexto específico de datos del dominio donde se ha entrenado.

5.3.2. Discriminación de usuarios en eventos a largo plazo basados en sensores binarios e interacción cercana

Esta sección refleja los resultados de minería de actividades humanas a partir de sensores usando lógica difusa. Se han modelado las reglas de *cocinar* y *sentarse a comer*, que se definieron en la sección 5.1.2.2. En las Tablas 5.8 y 5.9, se detallan los intervalos de tiempo obtenidos por los modelos en el conjunto de datos de evaluación, que describen, para cada usuario, el grado calculado en función de la interacción cercana en la región de interés correspondiente.

Primero, en relación a la actividad de *cocinar*, se requiere la activación simultánea del sensor de presencia y una alta temperatura en la vitrocerámica durante cierto tiempo. En este escenario, los usuarios 1 y 2 pueden interactuar simultáneamente en el evento, lo cual es etiquetado por un observador externo a partir de imágenes de las cámaras. De este modo, se determinan los minutos activos.

En segundo lugar, la actividad de *sentarse a comer* se relaciona con una presencia a largo plazo en la zona de la mesa, que los usuarios 1 y 2 utilizan para almorzar después de cocinar. Para presentar de manera comprensible el resultado de la activación del usuario para cada regla, el grado calculado para cada usuario y la regla se han categorizado (mediante defuzzificación) en los siguientes términos de actividad: "ninguna"(grado == 0), "parcial"(grado <1/3) y .ªctiva"(grado >=1/3).

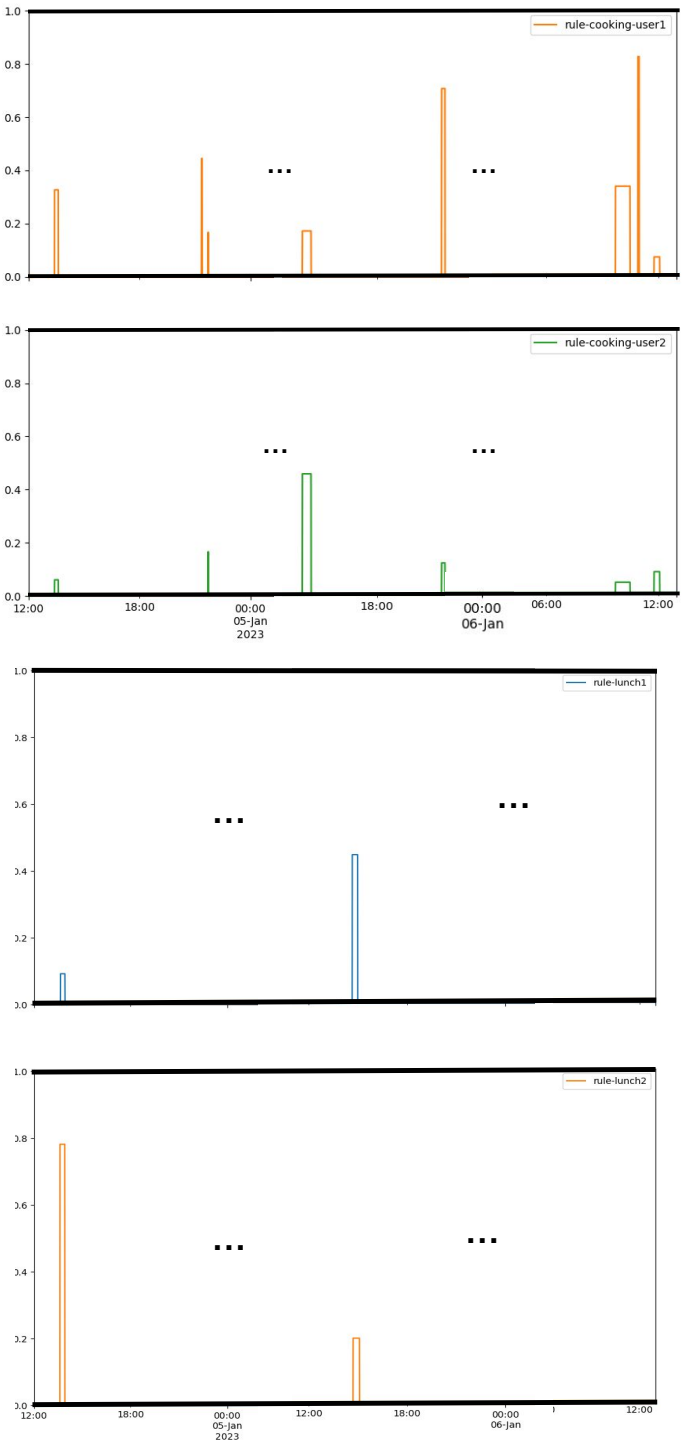


Figura 5.10: Grado de interacción del usuario 1 y el usuario 2, calculado a partir del método de discriminación espacio-temporal para cada intervalo (cocinar y sentarse a comer).

Tabla 5.8: Resultados del reconocimiento de eventos a largo plazo basados en sensores y en las reglas para la actividad de cocinar.

COCINAR							
GROUND TRUTH				MODELO			
start time	duration (min)	min u1	min u2	grado u1	grado u2	u1	u2
2023-01-04 13:24:15	12	8	4	0,842	0,158	Activo	Parcial
2023-01-04 21:21:30	2	2	0	1	0	Activo	Ninguno
2023-01-04 21:42:30	1,25	0,75	1	0,5	0,5	Activo	Activo
2023-01-05 14:16:15	26	3	10	0,273	0,727	Parcial	Activo
2023-01-05 21:14:15	10	10	7	0,853	0,147	Activo	Parcial
2023-01-06 9:42:15	46,75	25	2	0,887	0,113	Activo	Parcial
2023-01-06 10:54:30	4	4	0	1	0	Activo	Ninguno
2023-01-06 11:46:30	18	18	11	0,455	0,545	Activo	Activo

Tabla 5.9: Resultados del reconocimiento de eventos a largo plazo basados en sensores y en las reglas para la actividad de sentarse a comer.

SENTARSE A COMER							
GROUND TRUTH				MODELO			
start time	duration (min)	min u1	min u2	grado u1	grado u2	u1	u2
2023-01-04 13:37:15	19,25	16,5	18	0,857	0,935	Activo	Activo
2023-01-04 14:43:30	21,75	19,75	23,25	0,908	1,069	Activo	Activo

Los resultados muestran que la regla de *sentarse a comer* se ha estimado con una alta precisión en relación al *ground truth*. Los intervalos de tiempo y la presencia están muy bien alineados, con una cobertura del 94 %. El grado de activación es estable, preciso y se calcula mediante la concatenación y agregación utilizando el modelo de FTW y los criterios

de expertos.

En el caso de la regla de *cocinar*, la situación es más compleja debido a la activación del sensor de presencia, que incluye un tiempo de activación impreciso. Esto significa que los usuarios pueden interactuar de manera parcial, cocinar o no cocinar, e interactuar o no en la región de interés relacionada mientras la regla está activa. En este escenario, la desfuzzificación en términos lingüísticos juega un papel crucial de interpretabilidad. Esta desfuzzificación resalta la representación lingüística basada en la incertidumbre y la imprecisión, lo que proporciona una descripción detallada de la activación de cada usuario en relación con la regla.

CAPÍTULO 6

CONCLUSIONES Y TRABAJOS FUTUROS

En un mundo donde la tecnología avanza a pasos agigantados, ha surgido de forma prioritaria la necesidad de desarrollar soluciones innovadoras que mejoren la calidad de vida de las personas, especialmente aquellas en situaciones de fragilidad. Esta tesis doctoral se centra en la integración de tecnologías avanzadas y no invasivas, que brindan nuevas perspectivas en el ámbito del RA. Esto permite abordar el reconocimiento de eventos más específicos, como los eventos sonoros y la localización precisa en interiores, lo que resulta esencial para superar el desafío de la multiocupación en espacios compartidos. De esta forma, se logra no solo alcanzar el RA antes imperceptibles en los EI, sino también modernizar los esquemas clásicos de sensores y modelos existentes, adaptándolos a las necesidades actuales y emergentes.

El RA y la integración de tecnología se ha convertido en un área de estudio con un impacto significativo en la autonomía y bienestar de individuos con el síndrome de fragilidad, un grupo particularmente vulnerable que a menudo enfrenta desafíos únicos en su vida diaria. El síndrome de fragilidad, caracterizado por una disminución en la resistencia y la capacidad para afrontar factores de estrés, es especialmente prevalente en la población mayor. Estas personas pueden experimentar una variedad de dificultades, incluyendo una movilidad reducida, una mayor susceptibilidad a caídas y enfermedades, y un deterioro cognitivo. Es fundamental ofrecerles una atención de calidad para evitar que su patología derive en fragilidad secundaria,

como depresión o insuficiencia cardíaca, y centrarse en reducir los indicadores de fragilidad para prevenir incapacidades. Las necesidades que posee este segmento poblacional son muy variadas, aunque destacan fundamentalmente el seguimiento de su salud y estado físico, la atención temprana y la comunicación e integración social.

El seguimiento y asistencia son elementos fundamentales para personas con fragilidad, quienes presentan una menor resistencia a los estresores y un mayor riesgo de desarrollar discapacidades. Es clave vigilar sus signos vitales, nivel de actividad física y su habilidad para realizar actividades cotidianas, ya que estos factores son indicadores objetivos de su estado general de salud. Las tecnologías de supervisión y monitorización, como los sensores de localización, audio, visión o binarios, pueden ser particularmente útiles, ya que pueden instalarse en el hogar para reconocer la actividades diarias y detectar cambios en los patrones de movilidad o comportamiento que podrían indicar un deterioro de salud. Por ejemplo, un descenso o desviación en la actividad diaria general o un cambio en los patrones de sueño podría ser un signo temprano de una complicación médica o un inicio de demencia, permitiendo intervenciones oportunas y tempranas. También sería posible la detección de situaciones de riesgo, como cambios significativos en el horario de realización de cierta actividad cotidiana, salidas a deshoras, inactividad inusual o posibles caídas y alertar a los cuidadores o al personal médico, lo que permite una respuesta rápida en caso de una emergencia médica. Esto no solo mejora la seguridad del individuo, sino que también proporciona tranquilidad a los familiares y cuidadores.

De forma específica, se ha demostrado que la realización regular de actividad física es necesaria para fortalecer la capacidad motora y mejorar la salud mental, contribuyendo así a una independencia más notable. Esta actividad física está intrínsecamente vinculada con la rehabilitación, un proceso frecuentemente pautado a estas personas, que puede verse influenciado por cambios en su estado de salud y complicaciones médicas. Esto resalta la necesidad de una supervisión médica continua y atenta, pudiendo realizarse a través de la incorporación de tecnologías en los programas de rehabilitación y ejercicio. Por ejemplo, la utilización de cámaras para monitorizar y asegurar la correcta ejecución de los ejercicios puede ayudar a diseñar rutinas personalizadas enfocadas en fortalecer la fuerza y mejorar el equilibrio. Esto,

a su vez, contribuye a minimizar el riesgo de caídas y otras complicaciones asociadas a la fragilidad.

Por último, la comunicación y la integración social desempeñan un papel importante en el manejo de la fragilidad, dado que el aislamiento y la soledad son comunes entre las personas afectadas por este síndrome. Es fundamental fomentar la interacción social, brindar apoyo emocional y ofrecer estimulación cognitiva para preservar una salud mental robusta, prevenir trastornos como la depresión y promover la autonomía. En este contexto, la tecnología emerge como un recurso valioso para facilitar la expresión, proporcionar compañía y motivación, por ejemplo, mediante el uso de robots que interactúen con ellos, establezcan conexiones con sus cuidadores o familiares, o simplemente les asistan en la expresión de sus emociones y pensamientos. Además, en ambientes donde conviven con otras personas, dispositivos de localización pueden ser útiles para monitorizar la interacción social identificando así posibles situaciones de aislamiento y ayudando a mitigarlas.

Por consiguiente, resulta primordial diseñar soluciones que brinden asistencia efectiva y considerada en entornos cotidianos, tales como residencias de ancianos o viviendas familiares, donde es común la coexistencia de varias personas. En estos contextos, donde se enfrenta el desafío de la multiocupación, es imperativo diferenciar entre los distintos ocupantes y sus respectivas actividades. Esto demanda un alto grado de precisión y adaptabilidad en las tecnologías de RA. Para ello, se deben implementar sistemas avanzados que no solo reconozcan la presencia de las personas, sino que también identifiquen sus patrones de movimiento, rutinas y necesidades individuales. De esta manera, se puede garantizar una asistencia personalizada y eficiente, mejorando la calidad de vida y la seguridad de cada ocupante, al tiempo que se respeta su privacidad e independencia.

Esta tesis doctoral aborda los mecanismos de reconocimiento y la integración de sensorización, allanando el camino para que fabricantes y desarrolladores puedan crear soluciones completas y específicas para cada entorno. Para afrontar estos desafíos, se proponen y evalúan una variedad de propuestas de innovación tecnológica, incluyendo el uso de sensores de diversa índole, sistemas de localización en interiores, modelos de DL y protoformas difusas para el RA. De forma destacada, en la presente tesis doctoral, se han abordado los siguientes

puntos:

- Se han incluido técnicas de procesamiento de dispositivos multimedia: i) sonido para la detección y clasificación de eventos sonoros domésticos, y ii) procesamiento de imágenes térmicas, útiles para detectar la postura y los puntos de referencia corporales en entornos con múltiples ocupantes, así como para reconocer la actividad física.

- Se han evaluado diferentes propuestas sobre tecnologías y modelos para la localización en interiores, incluyendo RTLS con tecnología BLE y UWB, así como el uso de cámaras para la identificación facial y la ubicación de la persona.

- Se ha incluido y evaluado una innovadora metodología de RA que vincula la activación o los eventos de un determinado sensor con la localización del usuario que lo ha activado. Esta metodología permite adaptar los sistemas de RA a las características y patrones individuales de cada usuario, ofreciendo una solución personalizada y precisa independientemente del entorno.

La evaluación de estas tecnologías en casos de estudio reales se ha definido como un requisito fundamental de la tesis doctoral, ya que permite mostrar su efectividad y adaptabilidad en entornos de vida cotidianos. A través de estas investigaciones, se ha demostrado, mediante evaluación de diferentes casos de estudio, cómo las tecnologías propuestas pueden mejorar la seguridad y establecer mecanismos de reconocimiento en personas con síndrome de fragilidad, al tiempo que se respeta su privacidad y dignidad. Estos avances representan un paso significativo hacia la creación de entornos más seguros y adaptativos para las poblaciones vulnerables.

Profundizando en el contenido de esta tesis doctoral, en el capítulo 2 se investiga en detalle el rol esencial que juega el RA humanas en EI. Se ha destacado la aplicación de los sistemas de RA, que no solo facilitan la observación y análisis de comportamientos diarios, sino también proporcionan respuestas rápidas y efectivas en situaciones críticas, incrementando la seguridad y asistencia a personas frágiles. El contexto de este segmento poblacional se detalla en la última sección del capítulo, presentando sus necesidades y resaltando la importancia de

ofrecer sistemas avanzados diseñados según las necesidades de cada individuo. Este capítulo ofrece una visión y motivación hacia la relevancia de soluciones para personas frágiles basadas en tecnología no invasiva.

En el capítulo 3, se ha propuesto una arquitectura y procesamiento de sensores multimodales y casos de estudio para reconocimiento de actividad en entornos domésticos. Se ha realizado además una combinación de distintas tecnologías —como sensores ambientales, de audio y visuales— para mostrar su aplicación con una visión más completa y precisa de las interacciones humanas en un espacio determinado. Esta aproximación multimodal es crucial para el desarrollo de sistemas que se ajusten con precisión a los comportamientos de usuarios que, otros sensores, como los binarios, no permiten recoger. Para comprender de forma completa la construcción del ecosistema que conforman las diversas invertigaciones, se describe la plataforma y la arquitectura subyacente que sirve como base para la integración de estas tecnologías, permitiendo la recolección y procesamiento eficiente de datos de múltiples fuentes, destacando la facilidad de instalación de estos sensores y su bajo coste. Adicionalmente, en este capítulo se introducen dos sensores multimodales, de audio y visión térmica, como dispositivos emergentes que abren nuevas posibilidades en el campo del RA.

En el caso del reconocimiento de audio, se ha desarrollado una metodología de computación Edge-Fog para integrar esta función en dispositivos IoT, que además se utilizan para la recolección de datos (ver sección 3.1). Con este fin, se ha creado y etiquetado un conjunto de datos de eventos sonoros domésticos, como el sonido de una aspiradora, un microondas o una ducha. Para la clasificación de estos sonidos, se ha propuesto un enfoque basado en la representación espectral para la extracción de características, utilizando un modelo de CNN. Adicionalmente, se ha implementado un procesamiento difuso de secuencias de eventos sonoros, empleando restricciones temporales definidas por protoformas, con el objetivo de filtrar y minimizar los falsos positivos en la identificación de eventos en tiempo real. Los resultados de clasificación del conjunto de datos han alcanzado un F1-score de 0.99 con el modelo de CNN+MFCC. Para evaluar la propuesta en tiempo real bajo condiciones naturales, se ha llevado a cabo un caso de estudio en un entorno doméstico comprendido por seis escenas, obteniendo un F1-score de 0.95 en este caso.

Por otra parte, el procesamiento de imágenes térmicas también se introduce como protagonista en este capítulo con un doble propósito: i) la identificación de puntos de referencia corporales (*body landmarks*), y ii) la clasificación de actividades físicas.

En el primer punto, sobre la estimación de puntos de referencia corporales, se utiliza una red neuronal residual para identificar puntos de referencia corporales a partir de imágenes obtenidas mediante un sensor de visión térmica de baja resolución y coste (ver sección 3.4.1.1). Para abordar las dificultades en la captura y el etiquetado de datos, que suelen ser obstáculos en el aprendizaje de modelos de DL, se ha implementado un método de autoetiquetado utilizando cámaras duales, tanto de espectro visible como térmicas, creando un dispositivo que además permite realizar el procesamiento en Edge Computing. Este proceso incluye el uso del modelo OpenPose para reconocer puntos clave y generar el ground truth a partir de las imágenes de espectro visible. De esta manera, la red neuronal residual aprende a interpretar imágenes térmicas basándose en el conocimiento previo adquirido por OpenPose, preentrenado con imágenes de espectro visible. Se ha realizado un caso de estudio con cuatro individuos en los que se han obtenido muy buenos resultados, obteniendo un bajo error (RMSE de 2.46 en el conjunto de prueba y un RMSE de 9.80 en el conjunto de entrenamiento) en comparación con la anotación de espectro visible. No obstante, ciertas posturas complicadas, especialmente aquellas con acortamiento, requieren un entrenamiento más exhaustivo y una ampliación del conjunto de datos, incluyendo un mayor equilibrio en la variedad de posturas.

En el segundo punto, se analiza la monitorización de actividades físicas mediante una clasificación de cinco actividades deportivas detectadas mediante un sensor de visión térmica (ver sección 3.4.1.2). Se ha recopilado un conjunto de datos que comprende tres sesiones de ejercicio (5-6 minutos de duración) de un individuo realizando actividades como flexiones, abdominales, saltos *jumping jack*, sentadillas y ejercicios de plancha. Para mejorar la calidad del conjunto de datos y minimizar el riesgo de sobreajuste del modelo, se ha implementado un método de aumento de datos que incluye rotación, giro horizontal y escala. Posteriormente, se ha utilizado un modelo de DL para analizar secuencias de imágenes del usuario y determinar la actividad realizada. Para ello, se ha empleado una CNN para extraer características significativas del espacio espacial, y una red LSTM para modelar la secuencia de imágenes

y realizar la clasificación final. Los resultados obtenidos (F1-score de 0.99) demuestran un rendimiento destacado y una rápida capacidad de aprendizaje del modelo.

Asimismo, la trazabilidad en entornos de multiocupación ha supuesto un aspecto crucial en esta investigación, analizando cómo el seguimiento detallado de las interacciones y movimientos de los individuos en estos espacios compartidos es esencial para identificar las actividades que se están llevando a cabo en dicho entorno. En el capítulo 4, se han examinado diversas tecnologías para la localización y seguimiento en espacios interiores, incluyendo sistemas basados en radiofrecuencia como UWB y Bluetooth, así como sensores de visión.

En primer lugar, se ha evaluado la capacidad de UWB para seguir las trayectorias de los individuos en interiores utilizando modelos de DL 4.2. La integración de esta tecnología para determinar la ubicación de los usuarios mediante métodos de TOF o TDOA ofrece un rendimiento elevado, sin embargo, este método requiere un número considerable de balizas o anclas y, en entornos reales, las señales UWB pueden verse afectadas por paredes, muebles y otros obstáculos. Por esta razón, se ha utilizado una técnica denominada *fingerprint* para mejorar la correlación entre una medida específica, en este caso, la intensidad de la señal RSSI, con una ubicación en el entorno. Esta tecnología ha sido desplegada y evaluada en dos apartamentos con diferentes configuraciones (60 y 100 m^2) realizando actividades cotidianas. Se ha realizado una comparación entre el desempeño de modelos CNN, LSTM y CNN+LSTM y las capacidades de modelos más tradicionales como RF o SVM, incluyendo también la eficacia de Bluetooth en la clasificación por habitaciones. Los resultados obtenidos son prometedores, con estimaciones aproximadas a los 50 cm o en términos de error absoluto medio con el modelo de CNN+LSTM, lo cual facilita la implementación de esta metodología en el RA humanas en entornos domésticos.

En segundo lugar, se ha evaluado el uso de sensores de visión para reconocer puntos de referencia corporales en 3D, así como para identificar al sujeto 4.3. Se ha utilizado un enfoque que procesa conjuntamente una variedad de herramientas de DL, como YoLo, MediaPipe y DeepFace, incluyendo un sistema de seguimiento facial y corporal. El modelo propuesto en esta investigación es capaz de procesar, identificar, rastrear y obtener puntos de referencia corporales en 3D en situaciones de multiocupación, además de llevar a cabo una proyección de

homografía para estimar la ubicación 2D sobre el plano del suelo, permitiendo así la fusión de datos procedentes de diferentes cámaras. Posteriormente, se incorpora la identificación facial y el seguimiento no supervisado para determinar la identidad de los ocupantes dentro del EI y correlacionar los puntos de referencia corporales con cada sujeto. A través de un caso de estudio que comprende cinco escenas distribuidas en dos estancias, se ha evaluado el rendimiento de la metodología propuesta, obteniendo un F1-score de 0.98 a la hora de identificar al usuario.

Comparando las tecnologías evaluadas en términos de rendimiento, se ha destacado la superioridad de UWB sobre BLE en precisión, debido a la variabilidad y la incertidumbre asociadas a la señal de BLE al localizar usuarios. Sin embargo, el despliegue de sistemas de radiofrecuencia actualmente implica un coste mayor, dado que todavía están en fase de crecimiento y su adopción es más común en entornos industriales que entre usuarios generales, además de que la mayoría ofrece solo posicionamiento en 2D. Por otro lado, el seguimiento mediante cámaras muestra resultados notablemente eficaces y es más económico, permitiendo conocer la posición 3D del usuario en el espacio, e incluso su postura. No obstante, el seguimiento completo en un entorno requiere la instalación de numerosos dispositivos, lo que plantea preocupaciones de privacidad debido a la grabación continua de imágenes. Una solución a esta problemática es la creación de un entorno virtual a partir de los datos recogidos y procesados, lo que podría mitigar los problemas de privacidad al no utilizar imágenes directas de los individuos.

Un aspecto particularmente innovador de esta investigación ha sido la propuesta de un marco de discriminación en entornos de multiocupación que puede aplicarse al contexto de RA y EI de forma general. Dicho marco, basado en conocimiento y desarrollado a lo largo del capítulo 5, permite realizar una configuración rápida y sencilla modelando las zonas donde se produce un evento.

Gracias a los sistemas de trazabilidad, estudiados en profundidad en el capítulo 4, el discriminador establece una relación espacio-temporal entre los potenciales habitantes del entorno. La propuesta se ha modelado formalmente y de forma independiente del contexto, de modo que pueda ser reproducida y desplegada en cualquier entorno. Para evaluar la propuesta, se

ha realizado un caso de estudio donde se concreta un sistema de trazabilidad y generación de eventos en un entorno doméstico real. Para ello, se ha generado un conjunto de datos con cada evento perfectamente etiquetado. Gracias a una grabación mediante cámaras que establecen el ground truth, se ha etiquetado cada evento y comprobado que el modelo del discriminador permite determinar qúe usuario interactuó con los objetos en un porcentaje mayor al 0.90 de f1-score. De forma adicional, se ha extendido el modelo para tratar eventos o actividades más duraderas (no eventos puntuales). Para este propósito, se han definido protoformas linguísticas que minan y extraen información de sensores usando reglas difusas. Dichas protoformas definen mediante términos y ventanas temporales acciones como cocinar o estar sentado durante largo rato en la mesa, y permiten fusionar información de diversos sensores mediante reglas IF-THEN. De forma paralela, el discriminador ha permitido reconocer qué usuario era el que generaba esa regla mediante asociación espacio-temporal. El reconocimiento y discriminación de dos actividades seleccionadas (cocinar y comer), se ha situado en una tasa de cobertura del 94 %.

En conclusión, esta última investigación se ha distinguido por su capacidad para manejar eficazmente la incertidumbre y la variabilidad propias de las acciones humanas en contextos de multiocupación. Mediante la integración de diversas propuestas y tecnologías, se ha conseguido determinar con alta precisión la naturaleza de la actividad realizada, la identidad del sujeto y la duración o el momento en que se lleva a cabo. Se ha abordado el desafío de identificar áreas específicas de interacción y modelar flujos de datos difusos, basándose en la interacción cercana entre usuarios y sensores. Este enfoque facilita la obtención del nivel de interacción entre usuarios y sensores, proporcionando una fuente de datos valiosa para el campo del RA. Los resultados han demostrado que la combinación de tecnologías de localización, sensores ambientales y modelos de DL puede profundizar y mejorar nuestro entendimiento de las actividades humanas en entornos complejos, abriendo nuevas vías en el campo del RA y la inteligencia ambiental.

Estos resultados han sido posibles gracias a una diversa integración de conocimientos, métodos y hallazgos de investigaciones realizadas a lo largo de la tesis. Este proceso acumulativo ha resultado en un sistema completo y escalable a una amplia gama de sensores, incluyen-

do dispositivos ambientales, vestibles y de localización precisa como el UWB, manteniendo siempre la privacidad del individuo. Su capacidad de adaptación a distintos entornos y su potencial de escalabilidad son puntos fuertes de este sistema.

La relevancia de esta tesis doctoral radica en su contribución de soluciones concretas y efectivas a los retos cotidianos de personas en situaciones de fragilidad. Los hallazgos de esta investigación no solo avanzan el conocimiento teórico en el campo de las tecnologías asistivas y el cuidado de la salud, sino que también tienen aplicaciones prácticas directas, brindando herramientas que pueden mejorar significativamente la vida de estos individuos. Sin embargo, aunque los resultados obtenidos son prometedores, es crucial reconocer las limitaciones de la investigación actual. Estas limitaciones subrayan la importancia de continuar explorando y desarrollando estas tecnologías en entornos reales para validar su eficacia y adaptabilidad, ofreciendo un sólido punto de partida para futuros trabajos en este campo. La continuación de esta investigación debería centrarse en varias áreas clave para profundizar y expandir los conocimientos y aplicaciones obtenidos hasta ahora, y que se detallan a continuación.

Una de las áreas de enfoque para futuros estudios sería la ampliación y diversificación del conjunto de datos utilizados, recopilando información de un mayor número de sujetos en diferentes entornos y situaciones, priorizando las condiciones naturales de vida cotidiana. Esto no solo enriquecería los modelos con una variedad de patrones de comportamiento, sino que también mejoraría su generalización y robustez. Paralelamente, la mejora y refinamiento de los modelos de DL, a través de la exploración de nuevas arquitecturas de redes neuronales y técnicas de aprendizaje avanzadas, podría llevar a avances significativos en precisión y eficiencia. Además, la realización de pruebas y validaciones adicionales en entornos reales es esencial para evaluar la efectividad y adaptabilidad de las soluciones propuestas, identificando áreas de mejora y ajustes necesarios.

En cuanto a aplicaciones específicas, la tecnología UWB presenta un amplio espectro de posibilidades. La creación de ambientes inteligentes con robots de asistencia y telepresencia, son otras áreas donde UWB puede jugar un papel crucial. Por otra parte, el análisis de la interacción social y actividades de grupo mediante UWB podría ofrecer indicadores valiosos para el bienestar emocional y mental de personas en situaciones de fragilidad. En este con-

texto, la integración de la tecnología UWB con dispositivos wearables y plataformas abiertas como Home Assistant constituye un avance significativo en el campo de los sistemas integrados, ofreciendo un potencial considerable para el desarrollo de soluciones personalizadas y de alta versatilidad. Actualmente, el coste asociado a la tecnología UWB es relativamente alto, limitando su presencia a dispositivos específicos, predominantemente aquellos fabricados por Apple. Sin embargo, se anticipa que en un futuro próximo, dispositivos como el *Pixel Watch* de Google, que operan con sistemas abiertos y programables, integrarán esta tecnología, posibilitando el desarrollo de aplicaciones contextualizadas, tanto para entornos domésticos como hospitalarios, a un coste significativamente reducido. Se espera que Google lance una red similar a la *Find My* de Apple, facilitando la localización de dispositivos de bajo costo, como los airtags, que no exceden los 30 euros. Esto permitiría democratizar la localización precisa en tiempo real facilitando la monitorización de la salud y el control de dispositivos inteligentes en el hogar.

Adicionalmente, la combinación de tecnologías radar para la detección de presencia con UWB abre nuevas posibilidades, particularmente en su aplicación en residencias de ancianos y otros entornos de asistencia. Esta integración resulta especialmente relevante en contextos donde las visitas de familiares son intermitentes, permitiendo la identificación de estas visitas sin la necesidad de que ellos lleven un dispositivo, ya que solo sería necesario para los pacientes y el personal profesional. Estas tecnologías podrían operar conjuntamente para proporcionar una monitorización detallada y continua, elevando la seguridad y el bienestar de los ocupantes.

La posibilidad de desplegar estos sistemas integrados en una diversidad de entornos, que incluyen hogares, hospitales y residencias de ancianos, destaca la flexibilidad y capacidad de adaptación de estas soluciones. Estos sistemas se presentan como opciones ideales para una amplia gama de escenarios y necesidades, abriendo un abanico de oportunidades para mejorar la calidad de vida y la asistencia en diversos contextos. Su implementación podría marcar un hito importante en la evolución de la tecnología asistiva y de cuidado, destacando la importancia de seguir explorando y desarrollando estas tecnologías para maximizar su impacto y eficacia.

En resumen, la investigación presentada en esta tesis doctoral marca un paso importante

hacia un futuro donde la tecnología asistiva y el cuidado de la salud no solo mejoren la calidad de vida, sino que también empoderen a los individuos para vivir de manera más independiente y con dignidad. La colaboración interdisciplinaria y la adaptación continua de la tecnologías en entornos reales serán claves para alcanzar este valioso objetivo.

Conclusions

In a world where technology is rapidly evolving, the need to develop innovative solutions that improve the quality of life of people, particularly those in fragile conditions, has become a priority. This doctoral thesis focuses on the integration of advanced and non-invasive technologies, which offer new perspectives in the field of Activity Recognition (AR). This allows for addressing more specific event recognition, such as sound events and precise indoor localization, which is essential for overcoming the challenge of multi-occupancy in shared spaces. Thus, it is possible not only to achieve previously imperceptible AR in Intelligent Environments (IEs) but also to modernize the classical schemes of sensors and existing models, adapting them to current and emerging needs.

AR and the integration of technology have become areas of study with a significant impact on the autonomy and well-being of frail individuals, a particularly vulnerable group that often faces unique challenges in their daily lives. Frailty syndrome, characterized by a decrease in resilience and the ability to cope with stressors, is especially prevalent in the elderly population. These individuals may experience a variety of difficulties, including reduced mobility, increased susceptibility to falls and diseases, and cognitive decline. It is crucial to provide quality care to prevent its pathology from leading to secondary frailty, such as depression or heart failure, and to focus on reducing frailty indicators to prevent disabilities. The needs of this group of people are extremely varied and diverse, including physical and health monitoring, early care, and social communication.

Monitoring and assistance are fundamental elements for frail individuals, who have less resistance to stressors and a higher risk of developing disabilities. It is key to monitor its vital signs, level of physical activity, and ability to perform daily activities, as these factors are objective indicators of its general health status. Monitoring and surveillance technologies, such as localization, audio, vision, or binary sensors, can be particularly useful, as they can be installed at home to recognize daily activities and detect changes in mobility or behavior

patterns that could indicate deterioration in health. For example, a decrease or deviation in general daily activity or a change in sleep patterns could be an early sign of a medical complication or the onset of dementia, allowing timely and early interventions. It would also be possible to detect risk situations, such as significant changes in the timing of a certain daily activity, outings at unusual hours, unusual inactivity or potential falls, and alert caregivers or medical staff, allowing for a rapid response in case of a medical emergency. This not only improves the individual's safety, but also provides peace of mind to family members and caregivers.

Specifically, regular physical activity has been shown to be necessary to strengthen motor capacity and improve mental health, thus contributing to greater independence. This physical activity is intrinsically related to rehabilitation, a process that is often prescribed to these individuals, which can be influenced by changes in their health status and medical complications. This highlights the need for continuous and attentive medical supervision, which can be achieved by incorporating technologies into rehabilitation and exercise programs. For example, the use of cameras to monitor and ensure the correct execution of exercises can help design personalized routines focused on strengthening strength and improving balance. This, in turn, contributes to minimizing the risk of falls and other complications associated with frailty.

Lastly, communication and social integration play an important role in the treatment of frailty, as isolation and loneliness are common among people affected by this syndrome. Encourage social interaction, provide emotional support, and offer cognitive stimulation to maintain robust mental health, prevent disorders such as depression, and improve autonomy. In this context, technology emerges as a valuable resource to facilitate expression, provide companionship, and motivation, for example, through the use of robots that interact with them, establish connections with their caregivers or family members, or assist them in expressing their emotions and thoughts. Additionally, in environments where they live with other people, localization devices can be useful for monitoring social interaction, thus identifying possible situations of isolation and helping to mitigate them.

Therefore, it is essential to design solutions that provide effective and considerate assistan-

ce in everyday environments, such as nursing homes or family homes, where the coexistence of several individuals is common. In these contexts, where the challenge of multi-occupancy is faced, it is imperative to differentiate between the different occupants and their respective activities. This demands a high degree of precision and adaptability in AR technologies. To achieve this, we need advanced systems that can not only detect people's presence but also identify their movement patterns, routines, and individual needs. In this way, personalized and efficient assistance can be guaranteed, improving the quality of life and safety of each occupant, while respecting their privacy and independence.

This doctoral thesis addresses recognition mechanisms and sensor integration, paving the way for manufacturers and developers to create complete and specific solutions for each environment. To address these challenges, a variety of technological innovation proposals are proposed and evaluated, including the use of various sensors, indoor localization systems, deep learning (DL) models, and fuzzy protoforms for AR. In particular, in this doctoral thesis, the following points have been addressed:

- Techniques for processing multimedia devices have been included: i) sound for the detection and classification of domestic sound events, and ii) thermal image processing, useful for detecting posture and body landmarks in environments with multiple occupants, as well as for recognizing physical activity.
- Different proposals have been evaluated on technologies and models for indoor localization, including real-time location systems (RTLSs) with Bluetooth Low Energy (BLE) and Ultra-Wideband (UWB) technology, as well as the use of cameras for facial identification and person localization.
- An innovative AR methodology has been included and evaluated that matches the activation or events of a particular sensor with the location of the user who activated it. This methodology allows for the adaptation of AR systems to the characteristics and individual patterns of each user, offering a personalized and precise solution regardless of the environment.

Evaluation of these technologies in real case studies has been defined as a fundamental

requirement of the doctoral thesis as it allows demonstrating their effectiveness and adaptability in everyday living environments. Through these investigations, it has been shown, by evaluating different case studies, how the proposed technologies can improve safety and establish recognition mechanisms in frail individuals while respecting their privacy and dignity. These advances represent a significant step towards creating safer and more adaptive environments for vulnerable populations.

In delving into the content of each chapter of this doctoral thesis, the following conclusions can be drawn:

Chapter 2 thoroughly examines the significant role AR plays in helping frail individuals. The thesis highlights how AR systems facilitate the observation and analysis of daily behaviors, and provide quick and effective responses in critical situations, which increases safety and support for these individuals. The last section of the chapter provides context for this population segment, presenting its needs, and emphasizing the importance of offering advanced systems designed according to the requirements of each individual. Overall, this chapter provides vision and motivation for the relevance of technology-based solutions for frail individuals.

Chapter 3 proposes an architecture and processing approach for multimodal sensors. It also presents case studies for activity recognition in domestic environments. The approach combines different technologies such as environmental, audio, and visual sensors to provide a complete and precise vision of human interactions in a specific space. This multimodal approach is critical for developing systems that accurately adjust to user behaviors, which other sensors, such as binary ones, cannot capture. The platform and underlying architecture provide a foundation for integrating these technologies, allowing efficient collection and processing of data from multiple sources while focusing on the ease of installation and low cost of these sensors. Additionally, this chapter introduces two emerging multimodal sensors, audio and thermal vision, that offer new possibilities in the field of AR.

An Edge-Fog computing methodology has been developed for audio recognition, enabling the integration of this function into IoT devices used for data collection (see Section 3.1). A data set of domestic sound events has been created and labeled, including sounds such as a

vacuum cleaner, a microwave, or a shower. To classify these sounds, a spectral representation-based approach has been proposed for feature extraction using a convolutional neural network (CNN) model. To reduce false positives in real-time event identification, fuzzy processing of sound event sequences has been implemented employing temporal constraints defined by protoforms. The CNN + Mel frequency spectral coefficients (MFCC) model has achieved an F1 score of 0.99 for the classification of the data set. To evaluate the proposal under natural conditions, a case study was conducted in a domestic environment consisting of six scenes, achieving an F1 score of 0.95.

In this chapter, thermal image processing is introduced with two purposes: to identify body landmarks and classify physical activities.

To estimate body landmarks, a residual neural network is used to identify them from images captured using a low-resolution, low-cost thermal vision sensor. Dual cameras, one for the visible spectrum and one for thermal images, are used to create a device that can also process data through Edge Computing (see Section 3.4.1.1). An auto-labeling method has been implemented to address difficulties in data capture and labeling. The OpenPose model is used to recognize key points and generate ground truth from visible-spectrum images. The residual neural network interprets thermal images using the prior knowledge acquired by OpenPose. The model achieved good results in a case study with four individuals, with a low error rate compared to visible-spectrum annotation. However, more training and an expanded dataset are required for certain complicated postures.

To monitor physical activities, five sports activities are classified using a thermal vision sensor (see Section 3.4.1.2). A data set is collected that includes three exercise sessions of an individual performing activities such as push-ups, crunches, jump jacks, squats, and plank exercises. To improve the quality of the dataset, a data augmentation method is implemented, including rotation, horizontal flip, and scale. A DL model is used to analyze sequences of user images and determine the activity performed. A CNN extracts significant features from spatial space, and an LSTM network models the sequence of images and performs the final classification. The model shows excellent performance and rapid learning capacity, achieving an F1 score of 0.99.

Furthermore, traceability in multi-occupancy environments has been a crucial aspect of this research; analyzing how detailed tracking of individual interactions and movements in these shared spaces is essential for identifying the activities being carried out in such an environment. In Chapter 4, various technologies for localization and tracking in indoor spaces have been examined, including radio frequency-based systems such as UWB and Bluetooth, as well as vision sensors.

First, the ability of UWB to follow the trajectories of individuals indoors has been evaluated using DL models 4.2. The integration of this technology to determine user location through methods such as Time of Flight (ToF) or Time Difference of Arrival (TDoA) offers high performance; however, this method requires a considerable number of beacons or anchors, and in real environments, UWB signals can be affected by walls, furniture, and other obstacles. For this reason, a technique called fingerprinting has been used to improve the correlation between a specific measurement, in this case the Received Signal Strength Indicator (RSSI) intensity, and a location in the environment. This technology has been implemented and evaluated in two apartments with different configurations (60 and 100 m^2) performing daily activities. A comparison has been made between the performance of the CNN, LSTM and CNN + LSTM models and the capabilities of more traditional models such as Random Forest (RF) or Support Vector Machine (SVM), including Bluetooth's effectiveness in room classification. The results obtained are promising, with estimates close to 50 cm or in terms of mean absolute error with the CNN+LSTM model, which facilitates the implementation of this methodology in human AR in domestic environments.

Second, the use of vision sensors to recognize body landmarks in 3D and identify the subject has been evaluated 4.3. An approach that processes various DL tools, such as YoLo, MediaPipe, and DeepFace, including a facial and body tracking system has been used. The model proposed in this research is capable of processing, identifying, tracking, and obtaining body landmarks in 3D in multi-occupancy situations, in addition to carrying out a homography projection to estimate the 2D location on the floor plan, thus allowing the fusion of data from different cameras. Subsequently, unsupervised facial identification and tracking are incorporated to determine the identity of the occupants within the IE and correlate the

body landmarks with each subject. Through a case study comprising five scenes distributed in two rooms, the performance of the proposed methodology has been evaluated, obtaining an F1 score of 0.98 when identifying the user.

When comparing the performance of the UWB and BLE technologies, it was found that the UWB has a higher precision due to the variability and uncertainty associated with the BLE signal when locating users. However, radio frequency systems such as UWB are currently more expensive because they are still in the growth phase and are mostly used in industrial environments. Moreover, most radio frequency systems offer only 2D positioning. On the other hand, camera-based tracking is highly effective and economical. It not only allows the user's 3D position in space to be determined but also their posture. However, complete tracking in an environment requires the installation of numerous devices, which raises privacy concerns due to the continuous recording of images. One solution to this privacy concern is the creation of a virtual environment from the collected and processed data, which can mitigate privacy issues by not using direct images of individuals.

One innovative aspect of this research is the proposal of a discrimination framework that can be applied to multi-occupancy environments, in the context of AR and IEs. This framework, developed throughout Chapter 5, is based on knowledge and enables easy configuration by modeling the zones where an event occurs.

Due to the traceability systems studied in depth in Chapter 4, the discriminator establishes a spatial-temporal relationship between the potential inhabitants of the environment. This proposal has been formally modeled in a context-independent manner so that it can be reproduced and implemented in any environment. To evaluate the proposal, a case study was conducted in a real domestic environment where a traceability and event generation system was specified and a data set was generated with each event appropriately labeled. Camera recordings were used to establish the ground truth, and the discriminator model was able to determine which user interacted with the objects with an accuracy of over 0.99 of the F1 score. In addition, the model was extended to deal with long-term events (not punctual events), and linguistic protoforms were defined to extract information from sensors using fuzzy rules. For this purpose, linguistic protoforms have been defined to extract information

from sensors using fuzzy rules. These protoforms define through terms and temporal windows actions such as cooking or sitting for a long time at the table and allow the fusion of information from various sensors using IF-THEN rules. In parallel, the discriminator allowed to recognize which user generated each rule through spatial-temporal association. Recognition and discrimination of two activities (cooking and eating) had a coverage rate of 94

This research has successfully addressed the challenge of handling uncertainty and variability in human actions in multi-occupancy contexts. By integrating different technologies and methodologies, such as environmental sensors, wearable devices, and ultra-wideband (UWB) localization, research has been able to accurately determine the nature of activities, the subject's identity, and the time or point at which these activities occur. The research has successfully identified specific areas of interaction and modeled fuzzy data flows based on close interaction between users and sensors. This approach greatly enhances the interaction between users and sensors, providing valuable data for the field of AR. The combination of localization technologies, environmental sensors, and DL models has resulted in a comprehensive and scalable system that can accommodate a wide range of sensors while maintaining individual privacy. The system's adaptability to different environments and its potential for scalability are its major strengths.

The relevance of this research lies in its contribution to provide practical and effective solutions to the daily challenges faced by frail individuals. The findings of this research not only advance theoretical knowledge in the field of assistive technologies and healthcare but also have direct practical applications, providing tools that can significantly improve the lives of these individuals. However, it is important to acknowledge the limitations of current research and the need to continue exploring and developing these technologies in real-world environments to validate their effectiveness and adaptability, providing a solid foundation for future work in this field.

ANEXOS

Anexo A. Publicaciones

Este apéndice incluye un compendio de las publicaciones generadas a partir de los hallazgos expuestos en esta tesis doctoral, abarcando artículos en revistas y congresos internacionales.

A.1. Revistas internacionales

α) Polo-Rodriguez, A., Vilchez Chiachio, J. M., Paggetti, C., & Medina-Quero, J. (2021). Ambient sound recognition of daily events by means of convolutional neural networks and fuzzy temporal restrictions. *Applied Sciences, 11(15), 6978.* [115]

β) Polo-Rodríguez, A., Dionisio, P., Agnoloni, F., Gómez, A. P., Paggetti, C., López, L. G., ... & Medina-Quero, J. (2022). Challenges of ubiquitous and wearable solutions to address active ageing in the Andalusian community. *Journal of Universal Computer Science, 28(11), 1221.* [95]

γ) Lupion, M., Polo-Rodriguez, A., Medina-Quero, J., Sanjuan, J. F., & Ortigosa, P. M. (2022). On the limits of Conditional Generative Adversarial Neural Networks to reconstruct the identification of inhabitants from IoT low-resolution thermal sensors. *Expert Systems with Applications, 203, 117356.* [176]

δ) Polo-Rodríguez, A., Cavallo, F., Nugent, C., & Medina-Quero, J. (2024). Human activity mining in multi-occupancy contexts based on nearby interaction under a fuzzy approach.

Internet of Things, 25, 101018. [284]

ϵ) Lupión, M., Polo-Rodríguez, A., Medina-Quero, J., Sanjuan, J. F., & Ortigosa, P. M. (2024). 3D Human Pose Estimation from multi-view thermal vision sensors. *Information Fusion, 104, 102154.* [136]

A.2. Congresos internacionales

i) Polo-Rodriguez, A., Montoro-Lendinez, A., Espinilla, M., & Medina-Quero, J. (2022, May). Classifying Sport-Related Human Activity from Thermal Vision Sensors Using CNN and LSTM. In International Conference on Image Analysis and Processing (pp. 38-48). Cham: Springer International Publishing. [79]

ii) Polo-Rodríguez, A., Lupión, M., Ortigosa, P. M., & Medina-Quero, J. (2022, June). Estimating Frontal Body Landmarks from Thermal Sensors Using Residual Neural Networks. In *International Work-Conference on Bioinformatics and Biomedical Engineering (pp. 330-342).* Cham: Springer International Publishing. [153]

iii) Lupión, M., Polo-Rodríguez, A., Ortigosa, P. M., & Medina-Quero, J. (2022, November). ThermalYOLO: A Person Detection Neural Network in Thermal Images for Smart Environments. In *International Conference on Ubiquitous Computing and Ambient Intelligence (pp. 772-783).* Cham: Springer International Publishing. [154]

iv) Polo-Rodríguez, A., Romero-Sanchez, J., Fernández-García, E., Paloma-Castro, O., Porcel-Gálvez, A. M., & Medina-Quero, J. (2023). Review on Internet of Things for Innovation in Nursing Process-A PubMed-Based Search. In *International Conference on Ubiquitous Computing and Ambient Intelligence (pp. 57-70).* Springer, Cham. [12]

v) Polo-Rodriguez, A., Diaz-Jimenez, D., Carvajal, M. A., Baños, O., & Medina-Quero, J. (2023, November). Detection of Sets and Repetitions in Strength Exercises Using IMU-Based Wristband Wearables. In *International Conference on Ubiquitous Computing and Ambient Intelligence (pp. 71-80).* Cham: Springer Nature Switzerland. [78]

vi) Burns, M., Nugent, C., McClean, S., Quero, J. M., & Polo-Rodríguez, A. (2023, November). A Deep Learning and Probabilistic Approach to Recognising Activities of Daily Living with Privacy Preserving Thermal Sensors. In *International Conference on Ubiquitous Computing and Ambient Intelligence (pp. 155-166).* Cham: Springer Nature Switzerland. [29]

vii) Polo-Rodriguez, A., Burns, M., Nugent, C., Florez-Revuelta, F., & Medina-Quero, J. (2023, November). Non-invasive Synthesis from Vision Sensors for the Generation of 3D Body Landmarks, Locations and Identification in Smart Environments. In *International Conference on Ubiquitous Computing and Ambient Intelligence (pp. 57-68)*. Cham: Springer Nature Switzerland. [260]

BIBLIOGRAFÍA

[1] C. Jacob, S. Bourke, and S. Heuss, "From testers to cocreators—the value of and approaches to successful patient engagement in the development of ehealth solutions: Qualitative expert interview study," *JMIR Human Factors*, vol. 9, no. 4, p. e41481, 2022.

[2] M. Li, A. L. Porter, and A. Suominen, "Insights into relationships between disruptive technology/innovation and emerging technology: A bibliometric perspective," *Technological Forecasting and Social Change*, vol. 129, pp. 285–296, 2018.

[3] J. I. Conde-Ruiz and C. I. González, "El proceso de envejecimiento en españa," *Estudios sobre la economía Española*, 2021.

[4] J. P. Díaz and A. A. García, "Retos sanitarios de los cambios demográficos," *Medicina clínica*, vol. 146, no. 12, pp. 536–538, 2016.

[5] J. M. Domènech, "El envejecimiento de la población española y su impacto macroeconómico," *Papeles de Economía Española*, vol. 161, pp. 100–241, 2019.

[6] B. E. d. D. de Personas, "con valoración del grado de discapacidad," *Imserso. es*, 2018.

[7] O. E. de la Discapacidad, "Plan de acción de la estrategia española sobre discapacidad 2014-2020. informe sobre evaluación final del plan de acción 2014-2020 de la estrategia española sobre la discapacidad," *Observatorio Estatal de la Discapacidad*, 2022.

[8] M. T. Fernández, "La discapacidad mental o psicosocial y la convención sobre los derechos de las personas con discapacidad," *Comisión de Derechos Humanos del Distrito Federal*, 2017.

[9] C. O'Mahony and S. Quinlivan, "The eu disability strategy and the future of eu disability policy," in *Research handbook on EU disability law*, pp. 12–28, Edward Elgar Publishing, 2020.

[10] M. L. Fotteler, V. Mühlbauer, S. Brefka, S. Mayer, B. Kohn, F. Holl, W. Swoboda, P. Gaugisch, B. Risch, M. Denkinger, *et al.*, "The effectiveness of assistive technologies for older adults and the influence of frailty: systematic literature review of randomized controlled trials," *JMIR aging*, vol. 5, no. 2, p. e31916, 2022.

[11] Á. V. Espinosa, J. L. L. López, F. M. Mata, and M. E. E. Estevez, "Application of iot in healthcare: Keys to implementation of the sustainable development goals," *Sensors*, vol. 21, no. 7, p. 2330, 2021.

[12] A. Polo-Rodríguez, J. Romero-Sanchez, E. Fernández-García, O. Paloma-Castro, A.-M. Porcel-Gálvez, and J. Medina-Quero, "Review on internet of things for innovation in nursing process-a pubmed-based search," in *International Conference*

on Ubiquitous Computing and Ambient Intelligence, pp. 57–70, Springer, 2023.

[13] M. Weiser, "The computer for the twenty-first century scientific american," *September Elsevier Ltd*, 1991.

[14] A. Lentzas and D. Vrakas, "Non-intrusive human activity recognition and abnormal behavior detection on elderly people: A review," *Artificial Intelligence Review*, vol. 53, no. 3, pp. 1975–2021, 2020.

[15] H. Mshali, T. Lemlouma, M. Moloney, and D. Magoni, "A survey on health monitoring systems for health smart homes," *International Journal of Industrial Ergonomics*, vol. 66, pp. 26–56, 2018.

[16] E. De-La-Hoz-Franco, P. Ariza-Colpas, J. M. Quero, and M. Espinilla, "Sensor-based datasets for human activity recognition–a systematic review of literature," *IEEE Access*, vol. 6, pp. 59192–59210, 2018.

[17] M. Javaid, A. Haleem, S. Rab, R. P. Singh, and R. Suman, "Sensors for daily life: A review," *Sensors International*, vol. 2, p. 100121, 2021.

[18] J. Medina-Quero, S. Zhang, C. Nugent, and M. Espinilla, "Ensemble classifier of long short-term memory with fuzzy temporal windows on binary sensors for activity recognition," *Expert Systems with Applications*, vol. 114, pp. 441–453, 2018.

[19] T. Van Kasteren, G. Englebienne, and B. J. Kröse, "An activity monitoring system for elderly care using generative and discriminative models," *Personal and ubiquitous computing*, vol. 14, no. 6, pp. 489–498, 2010.

[20] J. M. Quero, M. A. L. Medina, A. S. Hidalgo, and M. Espinilla, "Predicting the urgency demand of copd patients from environmental sensors within smart cities with high-environmental sensitivity," *IEEE Access*, vol. 6, pp. 25081–25089, 2018.

[21] B. Fu, N. Damer, F. Kirchbuchner, and A. Kuijper, "Sensing technology for human activity recognition: A comprehensive survey," *Ieee Access*, vol. 8, pp. 83791–83820, 2020.

[22] G. Laput, Y. Zhang, and C. Harrison, "Synthetic sensors: Towards general-purpose sensing," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 3986–3999, 2017.

[23] L. M. Dang, K. Min, H. Wang, M. J. Piran, C. H. Lee, and H. Moon, "Sensor-based and vision-based human activity recognition: A comprehensive survey," *Pattern Recognition*, vol. 108, p. 107561, 2020.

[24] D. Cruz-Sandoval, J. Beltran-Marquez, M. Garcia-Constantino, L. A. Gonzalez-Jasso, J. Favela, I. H. Lopez-Nava, I. Cleland, A. Ennis, N. Hernandez-Cruz, J. Rafferty, *et al.*, "Semi-automated data labeling for activity recognition in pervasive healthcare," *Sensors*, vol. 19, no. 14, p. 3035, 2019.

[25] J. Heikenfeld, A. Jajack, J. Rogers, P. Gutruf, L. Tian, T. Pan, R. Li, M. Khine, J. Kim, and J. Wang, "Wearable sensors: modalities, challenges, and prospects," *Lab on a Chip*, vol. 18, no. 2, pp. 217–248, 2018.

[26] T. Kim Geok, K. Zar Aung, M. Sandar Aung, M. Thu Soe, A. Abdaziz, C. Pao Liew, F. Hossain, C. P. Tso, and W. H. Yong, "Review of indoor positioning: Radio wave technology," *Applied Sciences*, vol. 11, no. 1, p. 279, 2020.

[27] S. K. Yadav, K. Tiwari, H. M. Pandey, and S. A. Akbar, "A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions," *Knowledge-Based Systems*, vol. 223, p. 106970, 2021.

[28] Y. Yang, L. Wu, G. Yin, L. Li, and H. Zhao, "A survey on security and privacy issues in internet-of-things," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1250–1258, 2017.

[29] M. Burns, C. Nugent, S. McClean, J. M. Quero, and A. Polo-Rodríguez, "A deep learning and probabilistic approach to recognising activities of daily living with privacy preserving thermal sensors," in *International Conference on Ubiquitous*

Computing and Ambient Intelligence, pp. 155–166, Springer, 2023.

[30] B. Abade, D. Perez Abreu, and M. Curado, "A non-intrusive approach for indoor occupancy detection in smart environments," *Sensors*, vol. 18, no. 11, p. 3953, 2018.

[31] D. Bouchabou, S. M. Nguyen, C. Lohr, B. LeDuc, and I. Kanellos, "A survey of human activity recognition in smart homes based on iot sensors algorithms: Taxonomies, challenges, and opportunities with deep learning," *Sensors*, vol. 21, no. 18, p. 6037, 2021.

[32] S. Zhang, Z. Wei, J. Nie, L. Huang, S. Wang, and Z. Li, "A review on human activity recognition using vision-based method," *Journal of healthcare engineering*, vol. 2017, 2017.

[33] M. H. Siddiqi and A. Alsirhani, "An efficient feature selection method for video-based activity recognition systems," *Mathematical Problems in Engineering*, vol. 2022, 2022.

[34] J. Maitre, K. Bouchard, C. Bertuglia, and S. Gaboury, "Recognizing activities of daily living from uwb radars and deep learning," *Expert Systems with Applications*, vol. 164, p. 113994, 2021.

[35] F. Serpush, M. B. Menhaj, B. Masoumi, and B. Karasfi, "Wearable sensor-based human activity recognition in the smart healthcare system," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

[36] M. López-Medina, M. Espinilla, I. Cleland, C. Nugent, and J. Medina, "Fuzzy cloud-fog computing approach application for human activity recognition in smart homes," *Journal of Intelligent & Fuzzy Systems*, vol. 38, no. 1, pp. 709–721, 2020.

[37] L. Chen, C. D. Nugent, and H. Wang, "A knowledge-driven approach to activity recognition in smart homes," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 6, pp. 961–974, 2011.

[38] D. Dubois, P. Hájek, and H. Prade, "Knowledge-driven versus data-driven logics," *Journal of logic, Language and information*, vol. 9, pp. 65–89, 2000.

[39] H. F. Nweke, Y. W. Teh, M. A. Al-Garadi, and U. R. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," *Expert Systems with Applications*, vol. 105, pp. 233–261, 2018.

[40] A. S. A. Sukor, A. Zakaria, N. A. Rahim, L. M. Kamarudin, R. Setchi, and H. Nishizaki, "A hybrid approach of knowledge-driven and data-driven reasoning for activity recognition in smart homes," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 5, pp. 4177–4188, 2019.

[41] O. Banos, J. M. Galvez, M. Damas, A. Guillen, L. J. Herrera, H. Pomares, and I. Rojas, "Evaluating the effects of signal segmentation on activity recognition.," in *IWBBIO*, pp. 759–765, 2014.

[42] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, and I. Rojas, "Window size impact in human activity recognition," *Sensors*, vol. 14, no. 4, pp. 6474–6499, 2014.

[43] N. C. Krishnan and D. J. Cook, "Activity recognition on streaming sensor data," *Pervasive and mobile computing*, vol. 10, pp. 138–154, 2014.

[44] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 790–808, 2012.

[45] M. W. Berry, A. Mohamed, and B. W. Yap, *Supervised and unsupervised learning for data science*. Springer, 2019.

[46] Z. Hussain, M. Sheng, and W. E. Zhang, "Different approaches for human activity recognition: A survey," *arXiv preprint arXiv:1906.05074*, 2019.

[47] P. Ariza Colpas, E. Vicario, E. De-La-Hoz-Franco, M. Pineres-Melo, A. Oviedo-Carrascal, and F. Patara, "Unsupervised

human activity recognition using the clustering approach: A review," *Sensors*, vol. 20, no. 9, p. 2702, 2020.

[48] G. Wilson and D. J. Cook, "Multi-purposing domain adaptation discriminators for pseudo labeling confidence," *arXiv preprint arXiv:1907.07802*, 2019.

[49] Y. Chen, J. Wang, M. Huang, and H. Yu, "Cross-position activity recognition with stratified transfer learning," *Pervasive and Mobile Computing*, vol. 57, pp. 1–13, 2019.

[50] X. Li, J. Luo, and R. Younes, "Activitygan: Generative adversarial networks for data augmentation in sensor-based human activity recognition," in *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, pp. 249–254, 2020.

[51] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognition Letters*, vol. 119, pp. 3–11, 2019.

[52] M. A. Lopez Medina, M. Espinilla, C. Paggeti, and J. Medina Quero, "Activity recognition for iot devices using fuzzy spatio-temporal features as environmental sensor fusion," *Sensors*, vol. 19, no. 16, p. 3512, 2019.

[53] W. Shi and S. Dustdar, "The promise of edge computing," *Computer*, vol. 49, no. 5, pp. 78–81, 2016.

[54] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the internet of things," in *Proceedings of the first edition of the MCC workshop on Mobile cloud computing*, pp. 13–16, 2012.

[55] G. Kortuem, F. Kawsar, V. Sundramoorthy, and D. Fitton, "Smart objects as building blocks for the internet of things," *IEEE Internet Computing*, vol. 14, no. 1, pp. 44–51, 2009.

[56] F. Cruciani, A. Vafeiadis, C. Nugent, I. Cleland, P. McCullagh, K. Votis, D. Giakoumis, D. Tzovaras, L. Chen, and R. Hamzaoui, "Comparing cnn and human crafted features for human activity recognition," in *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pp. 960–967, IEEE, 2019.

[57] M. Laroui, B. Nour, H. Moungla, M. A. Cherif, H. Afifi, and M. Guizani, "Edge and fog computing for iot: A survey on current research activities & future directions," *Computer Communications*, vol. 180, pp. 210–231, 2021.

[58] S. E. de Medicina Interna (SEMI, S. E. de Medicina Familiar y Comunitaria, *et al.*, "Mejoras en la atención a personas con enfermedades crónicas en españa. declaración de posición común de la sociedad española de medicina interna y la sociedad española de medicina familiar y comunitaria," *Atención Primaria*, vol. 52, no. 1, p. 1, 2020.

[59] H. Y. Kaçmaz, A. Döner, H. Kahraman, and S. Akin, "Prevalencia y factores asociados a la fragilidad en pacientes mayores hospitalizados," *Revista Clínica Española*, 2023.

[60] L. R. Mañas, "Fragilidad vs comorbilidad: concepto, marcadores y capacidad predictora," *Envejecer con Salud*, p. 11, 2022.

[61] M. Plaza-Carmona, C. Requena-Hernández, and S. Jiménez-Mola, "El ejercicio físico multicomponente como herramienta de mejora de la fragilidad en personas mayores," *Gerokomos*, vol. 33, no. 1, pp. 16–20, 2022.

[62] D. S. Vásquez Andi, "Efectos del ejercicio físico en el adulto mayor con síndrome de fragilidad.," B.S. thesis, Universidad Nocional de Chimborazo, 2022.

[63] L. Moreno Macaya, "El ejercicio físico como intervención principal en el abordaje del síndrome de fragilidad y del riesgo de caídas en el anciano," Master's thesis, Universidad Pública de Navarra, 2020.

[64] A. Capdevila-Reniu, T. Casanova, E. Sopena, and J. M. Cancio, "Seguimiento telemático a los pacientes con fracturas por fragilidad," *Revista Española de Geriatría y Gerontología*, vol. 55, no. 6, p. 375, 2020.

[65] J. Á. Gregori, M. E. G. Cuevas, E. O. Muñoz, E. María, C. García, M. V. Á. Azofeifa, D. L. A. Gorno, and X. S. Calderón, "Envejecimiento, fragilidad y dependencia," *Salux: revista de ciencias y humanidades*, vol. 6, no. 9, pp. 27–35, 2020.

[66] A. C. C. Vicedomini[1], D. L. Waitzberg[1], N. C. Lopes, N. Magalhães, W. Jacob, A. Busse, D. Ferdinando, T. P. Alvez, R. M. R. Pereira, R. S. Torrinhas, *et al.*, "Impact of social isolation on the fragility and quality of life of elderly," in *ESPEN Virtual Congress*, 2021.

[67] X. Covbasa and G. Şoric, "Social factors of fragility in elderly people," in *Cercetarea în biomedicină și sănătate: calitate, excelență și performanță*, pp. 120–120, 2021.

[68] J. Apóstolo, E. Bobrowicz-Campos, I. Gil, R. Silva, P. Costa, F. Couto, D. Cardoso, A. Barata, and M. Almeida, "Cognitive stimulation in older adults: an innovative good practice supporting successful aging and self-care," *Translational Medicine@ UniSa*, vol. 19, p. 90, 2019.

[69] R. M. Cafferata, B. Hicks, and C. C. von Bastian, "Effectiveness of cognitive stimulation for dementia: A systematic review and meta-analysis.," *Psychological bulletin*, vol. 147, no. 5, p. 455, 2021.

[70] S. L. Ullo and G. R. Sinha, "Advances in smart environment monitoring systems using iot and sensors," *Sensors*, vol. 20, no. 11, p. 3113, 2020.

[71] S. D. Mamdiwar, Z. Shakruwala, U. Chadha, K. Srinivasan, and C.-Y. Chang, "Recent advances on iot-assisted wearable sensor systems for healthcare monitoring," *Biosensors*, vol. 11, no. 10, p. 372, 2021.

[72] N. S. M. Hadis, M. N. Amirnazarullah, M. M. Jafri, and S. Abdullah, "Iot based patient monitoring system using sensors to detect, analyse and monitor two primary vital signs," in *Journal of Physics: Conference Series*, vol. 1535, p. 012004, IOP Publishing, 2020.

[73] R. Chokri, W. Hanini, W. B. Daoud, S. A. Chelloug, and A. M. Makhlouf, "Secure iot assistant-based system for alzheimer's disease," *IEEE Access*, vol. 10, pp. 44305–44314, 2022.

[74] W. Salehi, G. Gupta, S. Bhatia, D. Koundal, A. Mashat, A. Belay, *et al.*, "Iot-based wearable devices for patients suffering from alzheimer disease," *Contrast Media & Molecular Imaging*, vol. 2022, 2022.

[75] R. J. Oskouei, Z. MousaviLou, Z. Bakhtiari, and K. B. Jalbani, "Iot-based healthcare support system for alzheimer's patients," *Wireless Communications and Mobile Computing*, vol. 2020, pp. 1–15, 2020.

[76] H. E. Adardour, M. Hadjila, S. Irid, T. Baouch, and S. Belkhiter, "Outdoor alzheimer's patients tracking using an iot system and a kalman filter estimator," *Wireless Personal Communications*, vol. 116, no. 1, pp. 249–265, 2021.

[77] Z. Munadhil, S. K. Gharghan, A. H. Mutlag, A. Al-Naji, and J. Chahl, "Neural network-based alzheimer's patient localization for wireless sensor network in an indoor environment," *IEEE Access*, vol. 8, pp. 150527–150538, 2020.

[78] A. Polo-Rodriguez, D. Diaz-Jimenez, M. A. Carvajal, O. Baños, and J. Medina-Quero, "Detection of sets and repetitions in strength exercises using imu-based wristband wearables," in *International Conference on Ubiquitous Computing and Ambient Intelligence*, pp. 71–80, Springer, 2023.

[79] A. Polo-Rodriguez, A. Montoro-Lendinez, M. Espinilla, and J. Medina-Quero, "Classifying sport-related human activity from thermal vision sensors using cnn and lstm," in *International Conference on Image Analysis and Processing*, pp. 38–48, Springer, 2022.

[80] M. Trombini, F. Ferraro, M. Morando, G. Regesta, and S. Dellepiane, "A solution for the remote care of frail elderly individuals via exergames," *Sensors*, vol. 21, no. 8, p. 2719, 2021.

[81] J. Calvillo-Arbizu, D. Naranjo-Hernández, G. Barbarov-Rostán, A. Talaminos-Barroso, L. M. Roa-Romero, and J. Reina-Tosina, "A sensor-based mhealth platform for remote monitoring and intervention of frailty patients at home," *International journal of environmental research and public health*, vol. 18, no. 21, p. 11730, 2021.

[82] J. Y. Baek, S. H. Na, H. Lee, H.-W. Jung, E. Lee, M.-W. Jo, Y. R. Park, and I.-Y. Jang, "Implementation of an integrated home internet of things system for vulnerable older adults using a frailty-centered approach," *Scientific Reports*, vol. 12, no. 1, p. 1922, 2022.

[83] F. M. García-Moreno, E. Rodríguez-García, M. J. Rodríguez-Fórtiz, J. L. Garrido, M. Bermúdez-Edo, C. Villaverde-Gutiérrez, and J. M. Pérez-Mármol, "Designing a smart mobile health system for ecological frailty assessment in elderly," *Multidisciplinary Digital Publishing Institute Proceedings*, vol. 31, no. 1, p. 41, 2019.

[84] H. Isyanto, A. S. Arifin, and M. Suryanegara, "Design and implementation of iot-based smart home voice commands for disabled people using google assistant," in *2020 International Conference on Smart Technology and Applications (ICoSTA)*, pp. 1–6, IEEE, 2020.

[85] S. Shalini, T. Levins, E. L. Robinson, K. Lane, G. Park, and M. Skubic, "Development and comparison of customized voice-assistant systems for independent living older adults," in *Human Aspects of IT for the Aged Population. Social Media, Games and Assistive Environments: 5th International Conference, ITAP 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26-31, 2019, Proceedings, Part II 21*, pp. 464–479, Springer, 2019.

[86] A. V. L. Volochtchuk, H. Leite, and A. D. Vieira, "Voice assistant technology applied to populations with developmental and physical disabilities," *Behaviour & Information Technology*, pp. 1–23, 2023.

[87] N. Z. Azlan and N. S. Lukman, "Assist as needed control strategy for upper limb rehabilitation robot in eating activity," *IIUM Engineering Journal*, vol. 22, no. 1, pp. 298–322, 2021.

[88] A. Abou Allaban, M. Wang, and T. Padır, "A systematic review of robotics research in support of in-home care for older adults," *Information*, vol. 11, no. 2, p. 75, 2020.

[89] F. Ibarra, M. Baez, L. Cernuzzi, and F. Casati, "A systematic review on technology-supported interventions to improve old-age social wellbeing: loneliness, social isolation, and connectedness," *Journal of healthcare engineering*, vol. 2020, 2020.

[90] H. K. Choi and S. H. Lee, "Trends and effectiveness of ict interventions for the elderly to reduce loneliness: a systematic review," in *Healthcare*, vol. 9, p. 293, MDPI, 2021.

[91] M. A. Saleh, F. A. Hanapiah, and H. Hashim, "Robot applications for autism: a comprehensive review," *Disability and Rehabilitation: Assistive Technology*, vol. 16, no. 6, pp. 580–602, 2021.

[92] A. Polo-Rodríguez, S. Rotbei, S. Amador, O. Baños, D. Gil, and J. Medina, "Smart architectures for evaluating the autonomy and behaviors of people with autism spectrum disorder in smart homes," in *Neural Engineering Techniques for Autism Spectrum Disorder*, pp. 55–76, Elsevier, 2021.

[93] M. Abdel Hameed, M. Hassaballah, M. E. Hosney, and A. Alqahtani, "An ai-enabled internet of things based autism care system for improving cognitive ability of children with autism spectrum disorders," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

[94] H. Posadas and E. Villar, "Eu fp7-288307 pharaon project," *IEEE Explore*, 2020.

[95] A. Polo-Rodríguez, P. Dionisio, F. Agnoloni, A. P. Gómez, C. Paggetti, L. G. López, A. C. Lendínez, M. Espinilla-Estévez, and J. Medina-Quero, "Challenges of ubiquitous and wearable solutions to address active ageing in the andalusian community," *Journal of Universal Computer Science*, vol. 28, no. 11, p. 1221, 2022.

[96] A. Polo Rodríguez and J. Medina Quero, "Monitorización remota de personas dependientes en el hogar usando dispositivos

wearables y binarios," Master's thesis, Jaén: Universidad de Jaén, 2020.

[97] A. Polo Rodríguez and J. Medina Quero, "Higia-sistema de reconocimiento de actividades higiénicas con sensores multimodales," Master's thesis, Jaén: Universidad de Jaén, 2022.

[98] L. Chen and C. D. Nugent, "Sensor-based activity recognition review," in *Human Activity Recognition and Behaviour Analysis*, pp. 23–47, Springer, 2019.

[99] A. Polo-Rodriguez, F. Cruciani, C. Nugent, and J. Medina-Quero, "Recognition of hygiene activities by means of multimodal sensors," in *Research and Innovation Forum 2021: Managing Continuity, Innovation, and Change in the Post-Covid World: Technology, Politics and Society*, pp. 89–98, Springer, 2021.

[100] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 790–808, 2012.

[101] M. Espinilla, L. Martínez, J. Medina, and C. Nugent, "The experience of developing the ujami smart lab," *Ieee Access*, vol. 6, pp. 34631–34642, 2018.

[102] J. Bravo, L. Fuentes, and D. L. de Ipina, "Theme issue:"ubiquitous computing and ambient intelligence"," 2011.

[103] P. Rashidi and A. Mihailidis, "A survey on ambient-assisted living tools for older adults," *IEEE journal of biomedical and health informatics*, vol. 17, no. 3, pp. 579–590, 2012.

[104] F. J. Ordóñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, 2016.

[105] A. R. J. Ruiz and F. S. Granja, "Comparing ubisense, bespoon, and decawave uwb location systems: Indoor performance analysis," *IEEE Transactions on instrumentation and Measurement*, vol. 66, no. 8, pp. 2106–2117, 2017.

[106] X. Xu, J. Tang, X. Zhang, X. Liu, H. Zhang, and Y. Qiu, "Exploring techniques for vision based human activity recognition: Methods, systems, and evaluation," *sensors*, vol. 13, no. 2, pp. 1635–1650, 2013.

[107] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, pp. 1–6, Ieee, 2017.

[108] L. Wyse, "Audio spectrogram representations for processing with convolutional neural networks," *arXiv preprint arXiv:1706.09559*, 2017.

[109] J. Salamon and J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279–283, 2017.

[110] J. Kim, "Urban sound tagging using multi-channel audio feature with convolutional neural networks," *Detection and Classification of Acoustic Scenes and Events 2019*, 2019.

[111] M. Lasseck, "Acoustic bird detection with deep convolutional neural networks," in *Proceedings of the Detection and Classification of Acoustic Scenes and Events 2018 Workshop (DCASE2018)*, pp. 143–147, 2018.

[112] K. Choi, G. Fazekas, and M. Sandler, "Automatic tagging using deep convolutional neural networks," *arXiv preprint arXiv:1606.00298*, 2016.

[113] J. Pons, O. Slizovskaia, R. Gong, E. Gómez, and X. Serra, "Timbre analysis of music audio signals with convolutional neural networks," in *2017 25th European Signal Processing Conference (EUSIPCO)*, pp. 2744–2748, IEEE, 2017.

[114] Y. Su, K. Zhang, J. Wang, and K. Madani, "Environment sound classification using a two-stream cnn based on decision-level fusion," *Sensors*, vol. 19, no. 7, p. 1733, 2019.

[115] A. Polo-Rodriguez, J. M. Vilchez Chiachio, C. Paggetti, and J. Medina-Quero, "Ambient sound recognition of daily events

by means of convolutional neural networks and fuzzy temporal restrictions," *Applied Sciences*, vol. 11, no. 15, p. 6978, 2021.

[116] J. Beltrán, E. Chávez, and J. Favela, "Scalable identification of mixed environmental sounds, recorded from heterogeneous sources," *Pattern Recognition Letters*, vol. 68, pp. 153–160, 2015.

[117] J. Beltrán, R. Navarro, E. Chávez, J. Favela, V. Soto-Mendoza, and C. Ibarra, "Recognition of audible disruptive behavior from people with dementia," *Personal and Ubiquitous Computing*, vol. 23, no. 1, pp. 145–157, 2019.

[118] G. Laput, K. Ahuja, M. Goel, and C. Harrison, "Ubicoustics: Plug-and-play acoustic activity recognition," in *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, pp. 213–224, 2018.

[119] E. Upton and G. Halfacree, *Raspberry Pi user guide*. John Wiley & Sons, 2014.

[120] A. Monteiro, M. de Oliveira, R. de Oliveira, and T. Da Silva, "Embedded application of convolutional neural networks on raspberry pi for shm," *Electronics Letters*, vol. 54, no. 11, pp. 680–682, 2018.

[121] S. Monk, *Programming the Raspberry Pi: getting started with Python*. McGraw-Hill Education, 2016.

[122] A. Gulli and S. Pal, *Deep learning with Keras*. Packt Publishing Ltd, 2017.

[123] U. Hunkeler, H. L. Truong, and A. Stanford-Clark, "Mqtt-s—a publish/subscribe protocol for wireless sensor networks," in *2008 3rd International Conference on Communication Systems Software and Middleware and Workshops (COMSWARE'08)*, pp. 791–798, IEEE, 2008.

[124] J. Medina, M. Espinilla, D. Zafra, L. Martínez, and C. Nugent, "Fuzzy fog computing: A linguistic approach for knowledge inference in wearable devices," in *International conference on ubiquitous computing and ambient intelligence*, pp. 473–485, Springer, 2017.

[125] I. F. Darwin, *Android Cookbook: Problems and Solutions for Android Developers*. ."O'Reilly Media, Inc.", 2017.

[126] B. Logan *et al.*, "Mel frequency cepstral coefficients for music modeling.," in *Ismir*, vol. 270, pp. 1–11, Citeseer, 2000.

[127] K. S. Rao and A. K. Vuppala, *Speech processing in mobile environments*. Springer, 2014.

[128] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Flexible, high performance convolutional neural networks for image classification," in *Twenty-second international joint conference on artificial intelligence*, 2011.

[129] F. Ordóñez, P. De Toledo, A. Sanchis, *et al.*, "Activity recognition using hybrid generative/discriminative models on home environments using binary sensors," *Sensors*, vol. 13, no. 5, pp. 5460–5477, 2013.

[130] A. Polo-Rodriguez, F. Cruciani, C. D. Nugent, and J. Medina, "Domain adaptation of binary sensors in smart environments through activity alignment," *IEEE Access*, vol. 8, pp. 228804–228817, 2020.

[131] A. Sixsmith and N. Johnson, "A smart sensor to detect the falls of the elderly," *IEEE Pervasive computing*, vol. 3, no. 2, pp. 42–47, 2004.

[132] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Computational intelligence and neuroscience*, vol. 2018, 2018.

[133] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[134] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1492–1500, 2017.

[135] E. Griffiths, S. Assana, and K. Whitehouse, "Privacy-preserving image processing with binocular thermal cameras,"

Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 1, no. 4, pp. 1–25, 2018.

[136] M. Lupión, A. Polo-Rodríguez, J. Medina-Quero, J. F. Sanjuan, and P. M. Ortigosa, "3d human pose estimation from multi-view thermal vision sensors," *Information Fusion*, vol. 104, p. 102154, 2024.

[137] M. Gochoo, T.-H. Tan, T. Batjargal, O. Seredin, and S.-C. Huang, "Device-free non-privacy invasive indoor human posture recognition using low-resolution infrared sensor-based wireless sensor networks and dcnn," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2311–2316, IEEE, 2018.

[138] J. M. Medina-Quero, M. Burns, M. A. Razzaq, C. Nugent, and M. Espinilla, "Detection of falls from non-invasive thermal vision sensors using convolutional neural networks," in *Multidisciplinary Digital Publishing Institute Proceedings*, vol. 2, p. 1236, 2018.

[139] S. Hiriyannaiah, B. Akanksh, A. Koushik, G. Siddesh, and K. Srinivasa, "Deep learning for multimedia data in iot," in *Multimedia Big Data Computing for IoT Applications*, pp. 101–129, Springer, 2020.

[140] X. Kong, Z. Meng, L. Meng, and H. Tomiyama, "A privacy protected fall detection iot system for elderly persons using depth camera," in *2018 International Conference on Advanced Mechatronic Systems (ICAMechS)*, pp. 31–35, IEEE, 2018.

[141] M. Gochoo, T.-H. Tan, S.-C. Huang, T. Batjargal, J.-W. Hsieh, F. S. Alnajjar, and Y.-F. Chen, "Novel iot-based privacy-preserving yoga posture recognition system using low-resolution infrared sensors and deep learning," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 7192–7200, 2019.

[142] H. J. Jara-Quito, L. F. Guerrero-Vasquez, K. A. Parra-Luzuriaga, M. V. Ojeda-Sanchez, and J. F. Bravo-Torres, "Avatar: Human-computer interface for interaction with children using a live animation process based in facial and body landmarks recognition," in *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pp. 715–720, IEEE, 2021.

[143] S. Jin, L. Xu, J. Xu, C. Wang, W. Liu, C. Qian, W. Ouyang, and P. Luo, "Whole-body human pose estimation in the wild," in *European Conference on Computer Vision*, pp. 196–214, Springer, 2020.

[144] A. Martínez-González, M. Villamizar, O. Canévet, and J.-M. Odobez, "Efficient convolutional neural networks for depth-based multi-person pose estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 11, pp. 4207–4221, 2019.

[145] H. Badave and M. Kuber, "Evaluation of person recognition accuracy based on openpose parameters," in *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 635–640, IEEE, 2021.

[146] G. Hidalgo, Y. Raaj, H. Idrees, D. Xiang, H. Joo, T. Simon, and Y. Sheikh, "Single-network whole-body pose estimation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6982–6991, 2019.

[147] D. Osokin, "Real-time 2d multi-person pose estimation on cpu: Lightweight openpose," *arXiv preprint arXiv:1811.12004*, 2018.

[148] K. Sozykin, S. Protasov, A. Khan, R. Hussain, and J. Lee, "Multi-label class-imbalanced action recognition in hockey videos via 3d convolutional neural networks," in *2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, pp. 146–151, IEEE, 2018.

[149] C. Zhang, F. Yang, G. Li, Q. Zhai, Y. Jiang, and D. Xuan, "Mv-sports: a motion and vision sensor integration-based sports analysis system," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, pp. 1070–1078, IEEE, 2018.

[150] A. Nadeem, A. Jalal, and K. Kim, "Automatic human posture estimation for sport activity recognition with robust body parts detection and entropy markov model," *Multimedia Tools and Applications*, vol. 80, no. 14, pp. 21465–21498, 2021.

[151] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "Openpose: realtime multi-person 2d pose estimation using part

affinity fields," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 1, pp. 172–186, 2019.

[152] N. Patricia and B. Caputo, "Learning to learn, from transfer learning to domain adaptation: A unifying perspective," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1442–1449, 2014.

[153] A. Polo-Rodríguez, M. Lupión, P. M. Ortigosa, and J. Medina-Quero, "Estimating frontal body landmarks from thermal sensors using residual neural networks," in *International Work-Conference on Bioinformatics and Biomedical Engineering*, pp. 330–342, Springer, 2022.

[154] M. Lupión, A. Polo-Rodríguez, P. M. Ortigosa, and J. Medina-Quero, "Thermalyolo: A person detection neural network in thermal images for smart environments," in *International Conference on Ubiquitous Computing and Ambient Intelligence*, pp. 772–783, Springer, 2022.

[155] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7291–7299, 2017.

[156] X. Li, Y. Liu, Y. Wang, and D. Yan, "Computing homography with ransac algorithm: a novel method of registration," in *Electronic Imaging and Multimedia Technology IV*, vol. 5637, pp. 109–112, International Society for Optics and Photonics, 2005.

[157] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.

[158] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[159] S. Zagoruyko and N. Komodakis, "Wide residual networks," *arXiv preprint arXiv:1605.07146*, 2016.

[160] I. K. M. Jais, A. R. Ismail, and S. Q. Nisa, "Adam optimization algorithm for wide and deep neural network," *Knowledge Engineering and Data Science*, vol. 2, no. 1, pp. 41–46, 2019.

[161] J. Thomas, K. Thirlaway, N. Bowes, and R. Meyers, "Effects of combining physical activity with psychotherapy on mental health and well-being: A systematic review," *Journal of Affective Disorders*, vol. 265, pp. 475–485, 2020.

[162] Z. Zhang and W. Chen, "A systematic review of measures for psychological well-being in physical activity studies and identification of critical issues," *Journal of affective disorders*, vol. 256, pp. 473–485, 2019.

[163] T. Yamashita, T. Watasue, Y. Yamauchi, and H. Fujiyoshi, "Improving quality of training samples through exhaustless generation and effective selection for deep convolutional neural networks.," in *VISAPP (2)*, pp. 228–235, 2015.

[164] D. C. Cireşan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, "High-performance neural networks for visual object classification," *arXiv preprint arXiv:1102.0183*, 2011.

[165] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[166] Q. Li, R. Gravina, Y. Li, S. H. Alsamhi, F. Sun, and G. Fortino, "Multi-user activity recognition: Challenges and opportunities," *Information Fusion*, vol. 63, pp. 121–135, 2020.

[167] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2568–2599, 2019.

[168] S. Xia, Y. Liu, G. Yuan, M. Zhu, and Z. Wang, "Indoor fingerprint positioning based on wi-fi: An overview," *ISPRS International Journal of Geo-Information*, vol. 6, no. 5, p. 135, 2017.

[169] H. Li, "Low-cost 3d bluetooth indoor positioning with least square," *Wireless personal communications*, vol. 78, no. 2,

pp. 1331–1344, 2014.

[170] M. Sugano, T. Kawazoe, Y. Ohta, and M. Murata, "Indoor localization system using rssi measurement of wireless sensor network based on zigbee standard.," *Wireless and optical communications*, vol. 538, pp. 1–6, 2006.

[171] C. Lee, Y. Chang, G. Park, J. Ryu, S.-G. Jeong, S. Park, J. W. Park, H. C. Lee, K.-s. Hong, and M. H. Lee, "Indoor positioning system based on incident angles of infrared emitters," in *30th Annual Conference of IEEE Industrial Electronics Society, 2004. IECON 2004*, vol. 3, pp. 2218–2222, IEEE, 2004.

[172] T. Gigl, G. J. Janssen, V. Dizdarevic, K. Witrisal, and Z. Irahhauten, "Analysis of a uwb indoor positioning system based on received signal strength," in *2007 4th Workshop on Positioning, Navigation and Communication*, pp. 97–101, IEEE, 2007.

[173] M. Hazas and A. Hopper, "Broadband ultrasonic location systems for improved indoor positioning," *IEEE Transactions on mobile Computing*, vol. 5, no. 5, pp. 536–547, 2006.

[174] P.-L. Liu and C.-C. Chang, "Simple method integrating openpose and rgb-d camera for identifying 3d body landmark locations in various postures," *International Journal of Industrial Ergonomics*, vol. 91, p. 103354, 2022.

[175] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep face recognition: A survey," in *2018 31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI)*, pp. 471–478, IEEE, 2018.

[176] M. Lupion, A. Polo-Rodriguez, J. Medina-Quero, J. F. Sanjuan, and P. M. Ortigosa, "On the limits of conditional generative adversarial neural networks to reconstruct the identification of inhabitants from iot low-resolution thermal sensors," *Expert Systems with Applications*, vol. 203, p. 117356, 2022.

[177] S. Hayward, K. van Lopik, C. Hinde, and A. West, "A survey of indoor location technologies, techniques and applications in industry," *Internet of Things*, vol. 1, p. 100608, 2022.

[178] N. E. ElHady and J. Provost, "A systematic survey on sensor failure detection and fault-tolerance in ambient assisted living," *Sensors*, vol. 18, no. 7, p. 1991, 2018.

[179] A.-P. Albín-Rodríguez, Y.-M. De-La-Fuente-Robles, J.-L. López-Ruiz, Á. Verdejo-Espinosa, and M. Espinilla Estévez, "Ujami location: A fuzzy indoor location system for the elderly," *International Journal of Environmental Research and Public Health*, vol. 18, no. 16, p. 8326, 2021.

[180] A. Howedi, A. Lotfi, and A. Pourabdollah, "Exploring entropy measurements to identify multi-occupancy in activities of daily living," *Entropy*, vol. 21, no. 4, p. 416, 2019.

[181] A. Howedi, A. Lotfi, and A. Pourabdollah, "Employing entropy measures to identify visitors in multi-occupancy environments," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14, 2020.

[182] D. S. Khan, J. Kolarik, C. A. Hviid, and P. Weitzmann, "Method for long-term mapping of occupancy patterns in open-plan and single office spaces by using passive-infrared (pir) sensors mounted below desks," *Energy and Buildings*, vol. 230, p. 110534, 2021.

[183] R. Krishnamurthy, "Determining occupancy of a multi-occupancy space," Mar. 30 2021. US Patent 10,963,683.

[184] A. Howedi, A. Lotfi, and A. Pourabdollah, "Distinguishing activities of daily living in a multi-occupancy environment," *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, 2019.

[185] C. Zhong, W. W. Ng, S. Zhang, C. D. Nugent, C. Shewell, and J. Medina-Quero, "Multi-occupancy fall detection using non-invasive thermal vision sensor," *IEEE Sensors Journal*, vol. 21, no. 4, pp. 5377–5388, 2020.

[186] S. A. Manssor, Z. Ren, R. Huang, and S. Sun, "Human activity recognition in thermal infrared imaging based on deep recurrent neural networks," in *2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 1–7, IEEE, 2021.

[187] S. Zhu, "Privacy-preserving building occupancy estimation via low-resolution infrared thermal cameras," 2021.

[188] M. A. Razzaq, J. M. Quero, I. Cleland, C. Nugent, U. Akhtar, H. S. M. Bilal, U. U. Rehman, and S. Lee, "umodt: an unobtrusive multi-occupant detection and tracking using robust kalman filter for real-time activity recognition," *Multimedia Systems*, vol. 26, no. 5, pp. 553–569, 2020.

[189] M. A. U. Alam, F. Mazzoni, M. M. Rahman, and J. Widberg, "Lamar: Lidar based multi-inhabitant activity recognition," in *MobiQuitous 2020-17th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, pp. 1–9, 2020.

[190] C. Laoudias, A. J. C. Moreira, S. Kim, S. Lee, L. Wirola, and C. Fischione, "A survey of enabling technologies for network localization, tracking, and navigation," *IEEE Communications Surveys & Tutorials*, vol. 20, pp. 3607–3644, 2018.

[191] G. A. Oguntala, R. A. Abd-Alhameed, S. M. R. Jones, J. M. Noras, M. N. Patwary, and J. Rodriguez, "Indoor location identification technologies for real-time iot-based applications: An inclusive survey," *Comput. Sci. Rev.*, vol. 30, pp. 55–79, 2018.

[192] S. Hara and D. Anzai, "Experimental performance comparison of rssi- and tdoa-based location estimation methods," *VTC Spring 2008 - IEEE Vehicular Technology Conference*, pp. 2651–2655, 2008.

[193] M. Karmy, S. ElSayed, and A.-E. H. Zekry, "Performance enhancement of an indoor localization system based on visible light communication using rssi / tdoa hybrid technique," *J. Commun.*, vol. 15, pp. 379–389, 2020.

[194] E. H. Yoshitome, J. V. R. da Cruz, M. E. P. Monteiro, and J. L. Rebelatto, "Lora-aided outdoor localization system: Rssi or tdoa?," *Internet Technology Letters*, vol. 5, 2021.

[195] P. S. Farahsari, A. Farahzadi, J. Rezazadeh, and A. Bagheri, "A survey on indoor positioning systems for iot-based applications," *IEEE Internet of Things Journal*, vol. 9, no. 10, pp. 7680–7699, 2022.

[196] F. Zhang, L. Yang, Y. Liu, Y. Ding, S.-H. Yang, and H. Li, "Design and implementation of real-time localization system (rtls) based on uwb and tdoa algorithm," *Sensors*, vol. 22, no. 12, p. 4353, 2022.

[197] Vikash, L. Mishra, and S. Varma, "Middleware technologies for smart wireless sensor networks towards internet of things: A comparative review," *Wireless Personal Communications*, vol. 116, pp. 1539 – 1574, 2020.

[198] R. Medeiros, S. Fernandes, and P. G. G. Queiroz, "Middleware for the internet of things: a systematic literature review," *J. Univers. Comput. Sci.*, vol. 28, pp. 54–79, 2022.

[199] S. Campaña Bastidas, M. Espinilla, and J. Medina Quero, "Review of ultra wide band (uwb) for indoor positioning with application to the elderly," *ScholarSpace*, vol. 1, 2022.

[200] F. Seco, A. R. Jiménez, C. Prieto, J. Roa, and K. Koutsou, "A survey of mathematical methods for indoor localization," in *2009 IEEE International Symposium on Intelligent Signal Processing*, pp. 9–14, IEEE, 2009.

[201] H. Obeidat, W. Shuaieb, O. Obeidat, and R. Abd-Alhameed, "A review of indoor localization techniques and wireless technologies," *Wireless Personal Communications*, vol. 119, pp. 289–327, 2021.

[202] G. Retscher, "Fusion of location fingerprinting and trilateration based on the example of differential wi-fi positioning.," *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 4, 2017.

[203] T. Kluge, C. Groba, and T. Springer, "Trilateration, fingerprinting, and centroid: taking indoor positioning with blue-

tooth le to the wild," in *2020 IEEE 21st International Symposium on.ª World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*, pp. 264–272, IEEE, 2020.

[204] M. Abd Elgwad and B. I. Sheta, "Wi-fi based indoor localization using trilateration and fingerprinting methods," in *IOP Conference Series: Materials Science and Engineering*, vol. 610, p. 012072, IOP Publishing, 2019.

[205] J. Uren, W. Price, J. Uren, and W. Price, "Triangulation and trilateration," *Surveying for engineers*, pp. 163–187, 1985.

[206] J. Xu, M. Ma, and C. L. Law, "Aoa cooperative position localization," in *IEEE GLOBECOM 2008-2008 IEEE Global Telecommunications Conference*, pp. 1–5, IEEE, 2008.

[207] X. Li, Z. D. Deng, L. T. Rauchenstein, and T. J. Carlson, "Contributed review: Source-localization algorithms and applications using time of arrival and time difference of arrival measurements," *Review of Scientific Instruments*, vol. 87, no. 4, 2016.

[208] S. Lanzisera, D. T. Lin, and K. S. Pister, "Rf time of flight ranging for wireless sensor network localization," in *2006 international workshop on intelligent solutions in embedded systems*, pp. 1–12, IEEE, 2006.

[209] R. Yamasaki, A. Ogino, T. Tamaki, T. Uta, N. Matsuzawa, and T. Kato, "Tdoa location system for ieee 802.11 b wlan," in *IEEE Wireless Communications and Networking Conference, 2005*, vol. 4, pp. 2338–2343, IEEE, 2005.

[210] A. Poulose, O. S. Eyobu, M. Kim, and D. S. Han, "Localization error analysis of indoor positioning system based on uwb measurements," in *2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN)*, pp. 84–88, IEEE, 2019.

[211] A. Ahmed, "Privacy issues of mobile phone companies' usage of ultra-wideband (uwb) technology: Analysing the use of uwb in mobile phones from a multi-actor perspective, magnifying privacy concerns and formulating guidelines," *Sensors, MDPI*, vol. 1, 2021.

[212] M. I. M. Ismail, R. A. Dzyauddin, S. Samsul, N. A. Azmi, Y. Yamada, M. F. M. Yakub, and N. A. B. A. Salleh, "An rssi-based wireless sensor node localisation using trilateration and multilateration methods for outdoor environment," *arXiv preprint arXiv:1912.07801*, vol. 1, 2019.

[213] P.-C. Liang and P. Krause, "Smartphone-based real-time indoor location tracking with 1-m precision," *IEEE journal of biomedical and health informatics*, vol. 20, no. 3, pp. 756–762, 2015.

[214] H. Sallouha, A. Chiumento, and S. Pollin, "Localization in long-range ultra narrow band iot networks using rssi," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6, IEEE, 2017.

[215] X. Luo, Q. Guan, H. Tan, L. Gao, Z. Wang, and X. Luo, "Simultaneous indoor tracking and activity recognition using pyroelectric infrared sensors," *Sensors*, vol. 17, no. 8, p. 1738, 2017.

[216] V.-L. Dao and S. M. Salman, "Deep neural network for indoor positioning based on channel impulse response," in *27th IEEE International Conference on Emerging Technologies and Factory Automation, ETFA 2022, Stuttgart, Germany, 6-9 September 2022*, vol. 2022, Institute of Electrical and Electronics Engineers Inc., 2022.

[217] L. Bai, F. Ciravegna, R. Bond, and M. Mulvenna, "A low cost indoor positioning system using bluetooth low energy," *Ieee Access*, vol. 8, pp. 136858–136871, 2020.

[218] X. Zhu, W. Qu, T. Qiu, L. Zhao, M. Atiquzzaman, and D. O. Wu, "Indoor intelligent fingerprint-based localization: Principles, approaches and challenges," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2634–2657, 2020.

[219] S. Hara and D. Anzai, "Experimental performance comparison of rssi-and tdoa-based location estimation methods," in *VTC Spring 2008-IEEE Vehicular Technology Conference*, pp. 2651–2655, IEEE, 2008.

[220] N. A. Azmi, S. Samsul, Y. Yamada, M. F. M. Yakub, M. I. M. Ismail, and R. A. Dziyauddin, "A survey of localization using rssi and tdoa techniques in wireless sensor network: System architecture," in *2018 2nd International Conference on Telematics and Future Generation Networks (TAFGEN)*, pp. 131–136, IEEE, 2018.

[221] S. Jondhale, R. Deshpande, S. Walke, and A. Jondhale, "Issues and challenges in rssi based target localization and tracking in wireless sensor networks," in *2016 international conference on automatic control and dynamic optimization techniques (ICACDOT)*, pp. 594–598, IEEE, 2016.

[222] Z. Yang, Z. Zhou, and Y. Liu, "From rssi to csi: Indoor localization via channel response," *ACM Computing Surveys (CSUR)*, vol. 46, no. 2, pp. 1–32, 2013.

[223] A. Strzoda, K. Grochla, and K. Połys, "Variability of ble advertisement packets received signal strength and delivery probability in the presence of interferences," in *Proceedings of the 12th ACM International Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications*, pp. 37–44, 2022.

[224] L. Flueratoru, V. Shubina, D. Niculescu, and E. S. Lohan, "On the high fluctuations of received signal strength measurements with ble signals for contact tracing and proximity detection," *IEEE Sensors Journal*, vol. 22, no. 6, pp. 5086–5100, 2021.

[225] A. Polo-Rodriguez and J. Medina-Quero, "Discriminating sensor activation in activity recognition within multi-occupancy environments based on nearby interaction," *arXiv preprint arXiv:2211.10355*, vol. 1, 2022.

[226] M. Laaraiedh, L. Yu, S. Avrillon, and B. Uguen, "Comparison of hybrid localization schemes using rssi, toa, and tdoa," in *17th European Wireless 2011-Sustainable Wireless Technologies*, pp. 1–5, VDE, 2011.

[227] N. Singh, S. Choe, and R. Punmiya, "Machine learning based indoor localization using wi-fi rssi fingerprints: an overview," *IEEE Access*, vol. 9, 2021.

[228] T. Wei and S. Bell, "Indoor localization method comparison: Fingerprinting and trilateration algorithm," *University of Saskatchewan. Accessed March*, vol. 24, p. 2015, 2011.

[229] J. Xia, S. Li, Y. Wang, and B. Jiang, "Research on uwb/ble-based fusion indoor positioning algorithm and system application," *2021 International Symposium on Computer Technology and Information Science (ISCTIS)*, pp. 50–54, 2021.

[230] F. Che, A. Ahmed, Q. Z. Ahmed, S. A. R. Zaidi, and M. Z. Shakir, "Machine learning based approach for indoor localization using ultra-wide bandwidth (uwb) system for industrial internet of things (iiot)," *2020 International Conference on UK-China Emerging Technologies (UCET)*, pp. 1–4, 2020.

[231] A. S. C. Ambrose, C. Savur, and F. Sahin, "Low cost real time location tracking with ultra-wideband," *2022 17th Annual System of Systems Engineering Conference (SOSE)*, pp. 445–450, 2022.

[232] A. M. Efendi, I. G. D. Nugraha, H. Han, D. Choi, S. M. S. Seo, and J. Kim, "A decision tree-based nlos detection method for the uwb indoor location tracking accuracy improvement," *International Journal of Communication Systems*, vol. 32, 2019.

[233] A. Volpi, L. Tebaldi, G. Matrella, R. Montanari, and E. Bottani, "Low-cost uwb based real-time locating system: Development, lab test, industrial implementation and economic assessment," *Sensors (Basel, Switzerland)*, vol. 23, 2023.

[234] O. Gnas, "Precise indoor location system using ultra-wideband technology," *PRZEGLAD ELEKTROTECHNICZNY*, 2023.

[235] D.-H. Kim, A. Farhad, and J.-Y. Pyun, "Uwb positioning system based on lstm classification with mitigated nlos effects," *IEEE Internet of Things Journal*, vol. 10, pp. 1822–1835, 2023.

[236] C. Li, Z. Li, H. Shen, and X. Gao, "Application of uwb indoor positioning system in different types of space," *Academic Journal of Engineering and Technology Science*, 2021.

[237] R. Nakamura and H. Hadama, "Target localization using multi-static uwb sensor for indoor monitoring system," *2017 IEEE Topical Conference on Wireless Sensors and Sensor Networks (WiSNet)*, pp. 37–40, 2017.

[238] Z. Yin, X. Jiang, Z. Yang, N. Zhao, and Y. Chen, "Wub-ip: A high-precision uwb positioning scheme for indoor multiuser applications," *IEEE Systems Journal*, vol. 13, pp. 279–288, 2019.

[239] K. Bregar, A. Hrovat, M. Mohori, and T. Javornik, "Self-calibrated uwb based device-free indoor localization and activity detection approach," *2020 European Conference on Networks and Communications (EuCNC)*, pp. 176–181, 2020.

[240] T. Otim, A. Bahillo, L. E. Díez, P. Lopez-Iturri, and F. Falcone, "Towards sub-meter level uwb indoor localization using body wearable sensors," *IEEE Access*, vol. 8, pp. 178886–178899, 2020.

[241] R. Zetik, G. Shen, and R. S. Thomä, "Evaluation of requirements for uwb localization systems in home-entertainment applications," in *2010 International Conference on Indoor Positioning and Indoor Navigation*, pp. 1–8, IEEE, 2010.

[242] T. Otim, A. Bahillo, L. E. Díez, P. Lopez-Iturri, and F. Falcone, "Impact of body wearable sensor positions on uwb ranging," *IEEE Sensors Journal*, vol. 19, no. 23, pp. 11449–11457, 2019.

[243] L. Cheng, A. Zhao, K. Wang, H. Li, Y. Wang, and R. Chang, "Activity recognition and localization based on uwb indoor positioning system and machine learning," *2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pp. 0528–0533, 2020.

[244] J. Maître, K. Bouchard, C. Bertuglia, and S. Gaboury, "Recognizing activities of daily living from uwb radars and deep learning," *Expert Syst. Appl.*, vol. 164, p. 113994, 2021.

[245] I. Pajak, P. Krutz, J. Patalas-Maliszewska, M. Rehm, G. Pajak, H. Schlegel, and M. Dix, "Sports activity recognition with uwb and inertial sensors using deep learning approach," *2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1–8, 2022.

[246] N. E. Tabbakha, C. P. Ooi, W. H. Tan, and Y.-F. Tan, "A wearable device for machine learning based elderly's activity tracking and indoor location system," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 2, pp. 927–939, 2021.

[247] Y. Zhan and H. Haddadi, "Mosen: Activity modelling in multiple-occupancy smart homes," *ArXiv*, vol. abs/2101.00235, 2021.

[248] L. Arrotta, C. Bettini, and G. Civitarese, "Micar: multi-inhabitant context-aware activity recognition in home environments," *Distributed and Parallel Databases*, pp. 1 – 32, 2022.

[249] M. T. Hoang, B. Yuen, K. Ren, X. Dong, T. Lu, R. Westendorp, and K. Reddy, "A cnn-lstm quantifier for single access point csi indoor localization," *arXiv preprint arXiv:2005.06394*, 2020.

[250] G. Singla, D. J. Cook, and M. Schmitter-Edgecombe, "Tracking activities in complex settings using smart environment technologies," *International journal of biosciences, psychiatry, and technology (IJBSPT)*, vol. 1, no. 1, p. 25, 2009.

[251] D. J. Cook and M. Schmitter-Edgecombe, "Assessing the quality of activities in a smart environment," *Methods of information in medicine*, vol. 48, no. 5, p. 480, 2009.

[252] Y.-M. Lu, J.-P. Sheu, and Y.-C. Kuo, "Deep learning for ultra-wideband indoor positioning," in *2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1260–1266, IEEE, 2021.

[253] R. A. Hamad, A. S. Hidalgo, M.-R. Bouguelia, M. E. Estevez, and J. Medina-Quero, "Efficient activity recognition in smart

homes using delayed fuzzy temporal windows on binary sensors," *IEEE journal of biomedical and health informatics*, vol. 24, no. 2, pp. 387–395, 2019.

[254] J. M. Quero, C. Orr, S. Zang, C. Nugent, A. Salguero, and M. Espinilla, "Real-time recognition of interleaved activities based on ensemble classifier of long short-term memory with fuzzy temporal windows," in *Multidisciplinary digital publishing institute proceedings*, vol. 2, p. 1225, 2018.

[255] M. Zago, M. Luzzago, T. Marangoni, M. De Cecco, M. Tarabini, and M. Galli, "3d tracking of human motion using visual skeletonization and stereoscopic vision," *Frontiers in bioengineering and biotechnology*, vol. 8, p. 181, 2020.

[256] L. Song, G. Yu, J. Yuan, and Z. Liu, "Human pose estimation and its application to action recognition: A survey," *Journal of Visual Communication and Image Representation*, vol. 76, p. 103055, 2021.

[257] J. Castro, M. Delgado, J. Medina, and M. Ruiz-Lozano, "An expert fuzzy system for predicting object collisions. its application for avoiding pedestrian accidents," *Expert Systems with Applications*, vol. 38, no. 1, pp. 486–494, 2011.

[258] S. Cheng, J.-X. Sun, Y.-G. Cao, L.-R. Zhao, *et al.*, "Target tracking based on incremental deep learning," *optics journal*, 2015.

[259] T. Jiang, Q. Zhang, J. Yuan, C. Wang, and C. Li, "Multi-type object tracking based on residual neural network model," *Symmetry*, vol. 14, no. 8, p. 1689, 2022.

[260] A. Polo-Rodriguez, M. Burns, C. Nugent, F. Florez-Revuelta, and J. Medina-Quero, "Non-invasive synthesis from vision sensors for the generation of 3d body landmarks, locations and identification in smart environments," in *International Conference on Ubiquitous Computing and Ambient Intelligence*, pp. 57–68, Springer, 2023.

[261] A. Pereira, P. Carvalho, N. Pereira, P. Viana, and L. Côrte-Real, "From a visual scene to a virtual representation: A cross-domain review," *IEEE Access*, 2023.

[262] Y. Tian, H. Zhang, Y. Liu, and L. Wang, "Recovering 3d human mesh from monocular images: A survey," *arXiv preprint arXiv:2203.01923*, 2022.

[263] N. Nikolakis, K. Alexopoulos, E. Xanthakis, and G. Chryssolouris, "The digital twin implementation for linking the virtual representation of human-based production tasks to their physical counterpart in the factory-floor," *International Journal of Computer Integrated Manufacturing*, vol. 32, no. 1, pp. 1–12, 2019.

[264] V. Gopinath, A. Srija, and C. N. Sravanthi, "Re-design of smart homes with digital twins," in *Journal of Physics: Conference Series*, vol. 1228, p. 012031, IOP Publishing, 2019.

[265] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of yolo algorithm developments," *Procedia Computer Science*, vol. 199, pp. 1066–1073, 2022.

[266] W. Chen, H. Huang, S. Peng, C. Zhou, and C. Zhang, "Yolo-face: a real-time face detector," *The Visual Computer*, vol. 37, pp. 805–813, 2021.

[267] S. Serengil, "Deepface, 2020," *URL https://github. com/serengil/deepface*, 2020.

[268] A. K. Singh, V. A. Kumbhare, and K. Arthi, "Real-time human pose detection and recognition using mediapipe," in *International Conference on Soft Computing and Signal Processing*, pp. 145–154, Springer, 2021.

[269] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7464–7475, 2023.

[270] F. Yang, X. Zhang, and B. Liu, "Video object tracking based on yolov7 and deepsort," *arXiv preprint arXiv:2207.12202*,

2022.

[271] J.-W. Kim, J.-Y. Choi, E.-J. Ha, and J.-H. Choi, "Human pose estimation using mediapipe pose and optimization method based on a humanoid model," *Applied Sciences*, vol. 13, no. 4, p. 2700, 2023.

[272] Y. Lin, X. Jiao, and L. Zhao, "Detection of 3d human posture based on improved mediapipe," *Journal of Computer and Communications*, vol. 11, no. 2, pp. 102–121, 2023.

[273] O. Chum, J. Matas, and J. Kittler, "Locally optimized ransac," in *Pattern Recognition: 25th DAGM Symposium, Magdeburg, Germany, September 10-12, 2003. Proceedings 25*, pp. 236–243, Springer, 2003.

[274] C. Zhu, "Video object tracking using sift and mean shift," *Computer Vision and Image Understanding*, 2011.

[275] S. Jian, H. Kaiming, R. Shaoqing, and Z. Xiangyu, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision & Pattern Recognition*, pp. 770–778, 2016.

[276] E. Saraee, M. Jalal, and M. Betke, "Visual complexity analysis using deep intermediate-layer features," *Computer Vision and Image Understanding*, vol. 195, p. 102949, 2020.

[277] O. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *BMVC 2015-Proceedings of the British Machine Vision Conference 2015*, British Machine Vision Association, 2015.

[278] C. Debes, A. Merentitis, S. Sukhanov, M. Niessen, N. Frangiadakis, and A. Bauer, "Monitoring activities of daily living in smart homes: Understanding human behavior," *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 81–94, 2016.

[279] M. Lupión, J. Medina-Quero, J. F. Sanjuan, and P. M. Ortigosa, "Dolars, a distributed on-line activity recognition system by means of heterogeneous sensors in real-life deployments—a case study in the smart lab of the university of almería," *Sensors*, vol. 21, no. 2, p. 405, 2021.

[280] M. Elsanhoury, P. Mäkelä, J. Koljonen, P. Välisuo, A. Shamsuzzoha, T. Mantere, M. Elmusrati, and H. Kuusniemi, "Precision positioning for smart logistics using ultra-wideband technology-based indoor navigation: A review," *IEEE Access*, 2022.

[281] A. Alarifi, A. Al-Salman, M. Alsaleh, A. Alnafessah, S. Al-Hadhrami, M. A. Al-Ammar, and H. S. Al-Khalifa, "Ultra wideband indoor positioning technologies: Analysis and recent advances," *Sensors*, vol. 16, no. 5, p. 707, 2016.

[282] J. A. Iglesias, P. Angelov, A. Ledezma, and A. Sanchis, "Human activity recognition based on evolving fuzzy systems," *International journal of neural systems*, vol. 20, no. 05, pp. 355–364, 2010.

[283] J.-M. Le Yaouanc and J.-P. Poli, "A fuzzy spatio-temporal-based approach for activity recognition," in *International Conference on Conceptual Modeling*, pp. 314–323, Springer, 2012.

[284] A. Polo-Rodríguez, F. Cavallo, C. Nugent, and J. Medina-Quero, "Human activity mining in multi-occupancy contexts based on nearby interaction under a fuzzy approach," *Internet of Things*, vol. 25, p. 101018, 2024.

[285] J. Medina-Quero, L. Martinez, and M. Espinilla, "Subscribing to fuzzy temporal aggregation of heterogeneous sensor streams in real-time distributed environments," *International Journal of Communication Systems*, vol. 30, no. 5, p. e3238, 2017.

[286] C. Martínez-Cruz, J. Medina-Quero, J. M. Serrano, and S. Gramajo, "Monwatch: A fuzzy application to monitor the user behavior using wearable trackers," in *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1–8, IEEE, 2020.

[287] M. K. Al-Sharman, B. J. Emran, M. A. Jaradat, H. Najjaran, R. Al-Husari, and Y. Zweiri, "Precision landing using an adaptive fuzzy multi-sensor data fusion architecture," *Applied soft computing*, vol. 69, pp. 149–164, 2018.

[288] L. A. Zadeh, "Generalized theory of uncertainty: principal concepts and ideas," in *Fundamental Uncertainty*, pp. 104–150, Springer, 2011.

[289] L. A. Zadeh, "A prototype-centered approach to adding deduction capability to search engines-the concept of protoform," in *2002 Annual Meeting of the North American Fuzzy Information Processing Society Proceedings. NAFIPS-FLINT 2002 (Cat. No. 02TH8622)*, pp. 523–525, IEEE, 2002.

[290] J. Kacprzyk and S. Zadrożny, "Linguistic database summaries and their protoforms: towards natural language based knowledge discovery tools," *Information Sciences*, vol. 173, no. 4, pp. 281–304, 2005.

[291] M. D. Peláez-Aguilera, M. Espinilla, M. R. Fernández Olmo, and J. Medina, "Fuzzy linguistic protoforms to summarize heart rate streams of patients with ischemic heart disease," *Complexity*, vol. 2019, 2019.

[292] M. A. A. Akhoundi and E. Valavi, "Multi-sensor fuzzy data fusion using sensors with different characteristics," *arXiv preprint arXiv:1010.6096*, 2010.

[293] L. Fan, "Image pixelization with differential privacy," in *Data and Applications Security and Privacy XXXII: 32nd Annual IFIP WG 11.3 Conference, DBSec 2018, Bergamo, Italy, July 16–18, 2018, Proceedings 32*, pp. 148–162, Springer, 2018.

[294] L. Rakhmawati *et al.*, "Image privacy protection techniques: A survey," in *TENCON 2018-2018 IEEE Region 10 Conference*, pp. 0076–0080, IEEE, 2018.

[295] P. Korshunov, C. Araimo, F. De Simone, C. Velardo, J.-L. Dugelay, and T. Ebrahimi, "Subjective study of privacy filters in video surveillance," in *2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*, pp. 378–382, Ieee, 2012.