Implementación de algoritmos de clasificación de blancos en radares de onda milimétrica para interiores

Quinatoa Chicaiza Cristhian Alexander & Ruano Ponce Alexander Javier

Departamento de Eléctrica, Electrónica y Telecomunicaciones

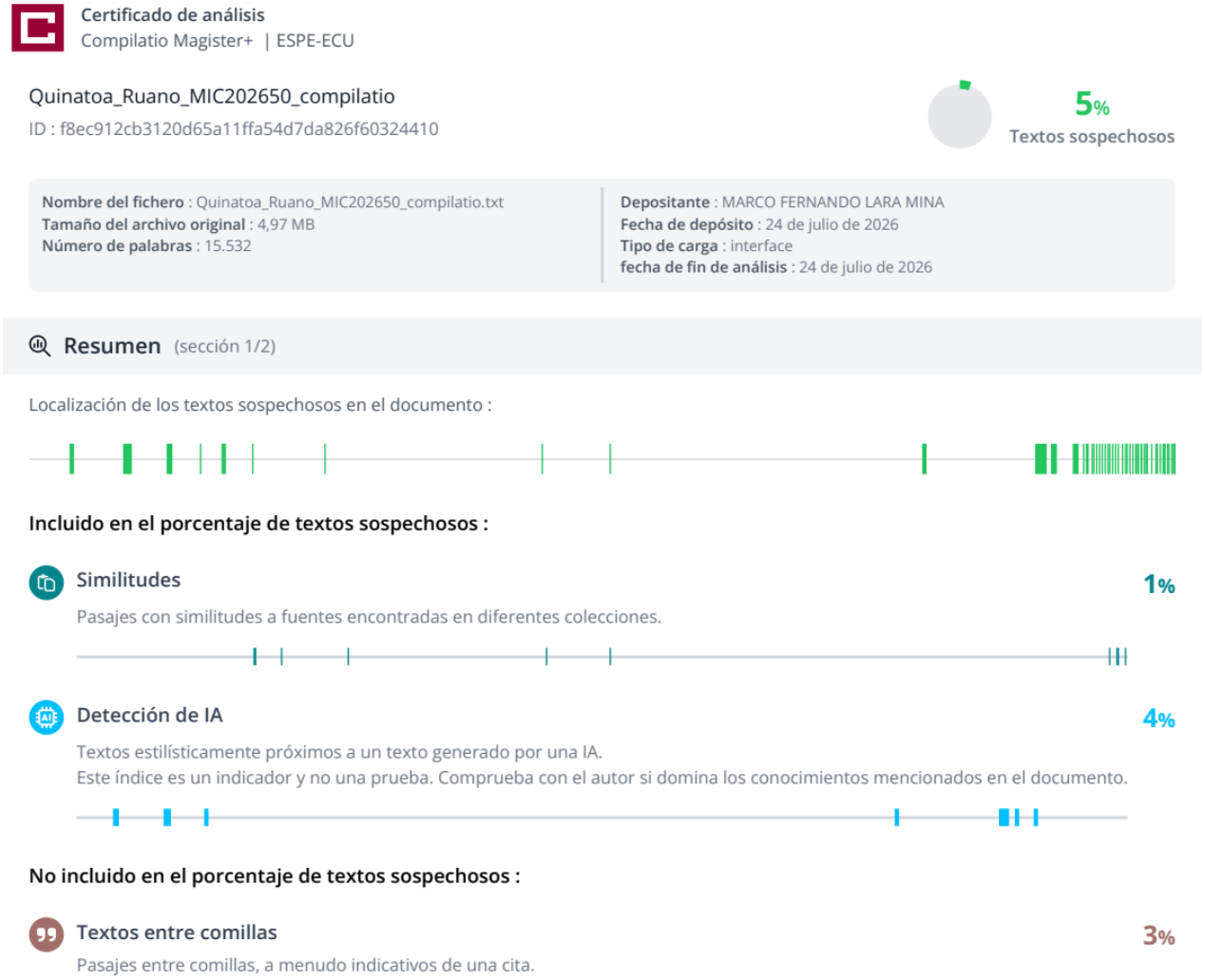
Carrera de Telecomunicaciones

Trabajo de integración curricular, previo a la obtención del título de

Ingeniero en Telecomunicaciones

Msc. Lara Mina Marco Fernando

24 de julio de 2026



Certificado de análisis
Compilatio Magister+ | ESPE-ECU

Quinatoa_Ruano_MIC202650_compilatio
ID : f8ec912cb3120d65a11ffa54d7da826f60324410

5%
Textos sospechosos

Nombre del fichero : Quinatoa_Ruano_MIC202650_compilatio.txt
Tamaño del archivo original : 4,97 MB
Número de palabras : 15.532

Depositante : MARCO FERNANDO LARA MINA
Fecha de depósito : 24 de julio de 2026
Tipo de carga : interface
fecha de fin de análisis : 24 de julio de 2026

Resumen (sección 1/2)

Localización de los textos sospechosos en el documento :

Incluido en el porcentaje de textos sospechosos :

Similitudes 1%
Pasajes con similitudes a fuentes encontradas en diferentes colecciones.

Detección de IA 4%
Textos estilísticamente próximos a un texto generado por una IA.
Este índice es un indicador y no una prueba. Comprueba con el autor si domina los conocimientos mencionados en el documento.

No incluido en el porcentaje de textos sospechosos :

Textos entre comillas 3%
Pasajes entre comillas, a menudo indicativos de una cita.

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
MARCO FERNANDO LARA MINA

Msc. Lara Mina Marco Fernando

Director

Departamento de Eléctrica, Electrónica y Telecomunicaciones

Carrera de Telecomunicaciones

Certificación

Certifico que el trabajo de integración curricular: "**Implementación de algoritmos de clasificación de blancos en radares de onda milimétrica para interiores.**" fue realizado por los señores **Quinatoa Chicaiza Cristhian Alexander y Ruano Ponce Alexander Javier**, el mismo que cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE; habiéndose efectuado una revisión integral de forma y de fondo conforme a los lineamientos y formatos institucionales vigentes. Además, fue revisado y analizada en su totalidad por la herramienta de prevención y/o verificación de similitud de contenidos; razón por la cual me permito acreditar y autorizar para que se lo sustente públicamente.

Sangolquí, viernes 24 de julio del 2026

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
MARCO FERNANDO LARA MINA

Msc. Lara Mina Marco Fernando

Director

C.C.: 1715140115

Departamento de Eléctrica, Electrónica y Telecomunicaciones

Carrera de Telecomunicaciones

Responsabilidad de Autoría

Nosotros, **Quinatoa Chicaiza Cristhian Alexander y Ruano Ponce Alexander Javier**, con cédulas de ciudadanía n° 1723704696 y n° 1727582387, declaramos que el contenido, ideas y criterios del trabajo de integración curricular: "**Implementación de algoritmos de clasificación de blancos en radares de onda milimétrica para interiores"**, es de nuestra autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos, y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, viernes 24 de julio del 2026

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
CRISTHIAN ALEXANDER
QUINATOA CHICAIZA

Quinatoa Chicaiza Cristhian Alexander

C.C.: 1723704696

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
ALEXANDER JAVIER
RUANO PONCE

Ruano Ponce Alexander Javier

C.C.: 1727582387

Departamento de Eléctrica, Electrónica y Telecomunicaciones

Carrera de Telecomunicaciones

Autorización de Publicación

Nosotros, **Quinatoa Chicaiza Cristhian Alexander y Ruano Ponce Alexander Javier**, con cédulas de ciudadanía n° 1723704696 y n° 1727582387, autorizamos a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de integración curricular: "**Implementación de algoritmos de clasificación de blancos en radares de onda milimétrica para interiores**." en el Repositorio Institucional, cuyo contenido, ideas y criterios son de nuestra responsabilidad.

Sangolquí, viernes 24 de julio del 2026

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
CRISTHIAN ALEXANDER QUINATOA CHICAIZA

Quinatoa Chicaiza Cristhian Alexander

C.C.: 1723704696

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
ALEXANDER JAVIER RUANO PONCE

Ruano Ponce Alexander Javier

C.C.: 1727582387

Dedicatoria

Dedico este trabajo, con profundo cariño, a mi madre y a mi abuelita, quienes desde que era pequeño se dedicaron a cuidarme y a sacarme adelante, incluso cuando las circunstancias no fueron las más favorables. Su esfuerzo y entrega son la base de todo lo que he logrado.

Agradezco también a mis tíos, quienes, a pesar de que nunca tuve una figura paterna presente, supieron enseñarme el valor del estudio, del esfuerzo constante y de no conformarme con lo común, guiándome con su ejemplo y sus valores.

A mis mascotas, por haber sido parte fundamental de este proceso de investigación, pues gracias a su compañía nunca me sentí solo durante las desveladas, las largas noches de estudio y los momentos más difíciles de mi vida universitaria. En especial a mi perrito Bolt, quien con su presencia me impulsó a continuar cuando más lo necesité. Aunque en ocasiones sentí que la carrera universitaria no me estaba dando los frutos que merecía, y aunque en algún momento pensé en desistir, no lo hice. Fue precisamente a partir de la partida de mi Bolti que comprendí el verdadero significado de la vida y la importancia de seguir adelante sin rendirse, pues él, a pesar de su edad, jamás dejó de levantarse para recibirme de la mejor manera cada vez que yo llegaba a casa.

Alexander Javier Ruano Ponce.

Agradecimiento

Dedico este trabajo, con profundo cariño, a mis padres, hermanos y a toda mi familia, quienes desde muy temprana edad me enseñaron que todo logro se alcanza con esfuerzo, dedicación y sacrificio; que las cosas no llegan solas ni se obtienen sin trabajo.

Gracias por su apoyo incondicional, tanto emocional como moral, por confiar siempre en mí y acompañarme durante todo este proceso. A pesar de las dificultades, los obstáculos y los momentos de incertidumbre, nunca dejaron de creer en mí ni de brindarme su respaldo. Su paciencia, comprensión y palabras de aliento fueron fundamentales para alcanzar esta meta.

También quiero dedicar este trabajo a mis mascotas, quienes, con su compañía y cariño incondicional, estuvieron presentes en los momentos de mayor estrés y cansancio, brindándome tranquilidad y motivación para seguir adelante.

Este logro también les pertenece a ustedes. Gracias por ser el pilar que me impulsó a nunca rendirme.

Cristhian Alexander Quinatoa Chicaiza.

Agradecimiento

Expreso mi más sincero agradecimiento al Ing. Marco Fernando Lara Mina, director de esta tesis de grado, por su guía académica, disposición y acompañamiento constante a lo largo del desarrollo de este trabajo. Sus conocimientos, experiencia y orientación resultaron fundamentales para la culminación exitosa de esta investigación.

De igual manera, extiendo mi gratitud a la Universidad de las Fuerzas Armadas ESPE y a la carrera de Ingeniería en Telecomunicaciones, por brindarme la formación profesional, los recursos y el entorno académico necesarios para mi crecimiento tanto personal como profesional. A los docentes de la carrera les agradezco de manera especial, pues a lo largo de mi formación compartieron sus conocimientos, experiencias y valores, contribuyendo de forma significativa a mi desarrollo académico y humano, y sentando las bases que me permitieron alcanzar esta meta.

Finalmente, agradezco de manera muy especial a mi compañero de tesis por su dedicación y trabajo en equipo durante todo este proceso. Extiendo también mi más profunda gratitud a mi familia, por su respaldo permanente y su confianza a lo largo de mi formación universitaria, pilares fundamentales para la consecución de este logro.

Alexander Javier Ruano Ponce.

Índice de contenidos

Dedicatoria **6**

Agradecimiento **7**

Agradecimiento **8**

Resumen **14**

Abstract **15**

Capítulo 1 Introducción **16**

Objetivo general: 21

Objetivos específicos: 21

Capítulo 2 Marco Teórico **23**

Fundamentos teóricos de ondas milimétricas 23

Teoría de Radar 25

Efecto Doppler 26

Efecto micro doppler 29

Radar de Onda Continua 31

Radar CW Infineon: 33

Transformada de Fourier de tiempo corto (STFT) 34

Función de ventana y principio de incertidumbre 36

Redes Neuronales Convolucionales 37

Operación de convolución: 37

Arquitectura: 38

AlexNet 40

ResNet50 41

MobileNetV2 42

SqueezeNet 43

Capítulo 3 Metodología **45**

Adquisición de datos 45
Data augmentation 48
Preparación de datasets 48
Transferencia de aprendizaje 49
Entrenamiento y evaluación 51
Configuración experimental 52
Experimentos 1-Alcance de radar: 52
Experimento 2-Pruebas de espectrogramas: 53
Experimento 3-Entrenamiento de CNNs con personas y animales: 55
Experimento 4-Entrenamiento de CNNs con personas, gatos y perros: 56
Experimento 5-Analisis estadístico de CNNs con Wilkinson: 56
Capítulo 4 Resultados 58
Representación Tiempo Frecuencia 58
Entrenamiento Inicial con AlexNet sobre dataset sin smoothing 59
Evaluación del efecto del suavizado (Smoothing) con Alexnet 61
AlexNet con Data Augmentation sin smoothing 62
AlexNet con Data Augmentation con Smothing 62
Comparativa todas las arquitecturas, sin smoothing con data augmentation 63
Comparativa todas las arquitecturas, con Smothing y data augmentation para tres clases. 65
Prueba estadística de Wilcoxon 68
Comparación global del desempeño de las arquitecturas CNN 69
Capítulo 5 Conclusiones 72
Trabajos Futuros: 74
Referencias 75

Índice de tablas

Tabla 4.1 Métricas de clasificación de AlexNet, sin smoothing sin data augmentation 60

Tabla 4.2 Métricas de clasificación con AlexNet, con smoothing sin data augmentation. 61

Tabla 4.3 Métricas de clasificación con AlexNet, sin smoothing y con data augmentation 62

Tabla 4.4 Métricas de clasificación con AlexNet, con smoothing y data augmentation 62

Tabla 4.5 Matrices de confusión consolidadas de todas las arquitecturas, sin smoothing con data augmentation 63

Tabla 4.6 Comparativa de métricas de clasificación con todas las arquitecturas, sin smoothing y con data augmentation 64

Tabla 4.7 Métricas por clase de todas las arquitecturas, 3 clases con smoothing con data augmentation. 65

Tabla 4.8 Comparación del desempeño de las arquitecturas CNN mediante Accuracy, Recall Macro Average y F1-Score Macro Average para la clasificación de tres clases. 66

Tabla 4.9 Matrices de confusión consolidadas de todas las arquitecturas 3 clases con smoothing y con data augmentation. 67

Tabla 4.10 Test de Wilcoxon con respecto a AlexNet, escenario 2 clases ($\alpha = 0,05$) 68

Tabla 4.11 Test de Wilcoxon con respecto a AlexNet, escenario 3 clases ($\alpha = 0,05$) 68

Índice de figuras

Figura 2.1 Ondas electromagnéticas. 23

Figura 2.2 Distancia radar. Tomado de [19] 26

Figura 2.3 Comportamiento del efecto doppler. Tomado de [24] 27

Figura 2.4 Ejemplo del comportamiento del efecto doppler 29

Figura 2.5 Firmas micro doppler de una persona caminando moviendo los brazos. Tomado de [27] 30

Figura 2.6 Diferencia de Fase. Tomado de [25] 32

Figura 2.7 Radar BGT60LTR11AIP 33

Figura 2.8 Procedimiento de la Transformada rápida de Fourier (STFT). Tomado de [29] 35

Figura 2.9 Representación visual de frecuencias mediante un Espectrograma. 36

Figura 2.10 Ejemplo de convolución 2-D sin inversión del núcleo (kernel flipping), Tomado de [32]. 38

Figura 2.11 Arquetipo de una Red Neuronal Convolucional (CNN), Tomado de [33]. 39

Figura 2.12 Arquitetura de Alexnet. Tomado de [35] 41

Figura 2.13 Diagrama de bloque residual (residual block) con la conexión de salto (skip connection). Tomado de [36]. 42

Figura 2.14 Comparación entre (2.14a) el bloque residual clásico y (2.14b) el bloque residual invertido (inverted residual block) de MobileNetV2. Tomado de [37]. 43

Figura 2.15 Diagrama del módulo Fire donde muestra la capa squeeze y la capa expand. Tomado de [38]. 44

Figura 3.1 Diagrama de Bloques de procesos- Metodología 45

Figura 3.2 Trayectoria de objetivos 46

Figura 3.3 Señal cruda recolectada 47

Figura 3.4 Proceso de Data Augmentation 48

Figura 3.5 Smoothing de una muestra original 49

Figura 3.6 Proceso de Transfer Learning..... 50

Figura 3.7 Proceso de 5 K folds..... 52

Figura 3.8 Espectro de Potencia de una muestra del dataset con STFT 53

Figura 3.9 Espectrograma CWT - Transformada Wavelet Continua..... 54

Figura 3.10 Transformada Adaptativa de Superlets ASLT Espectrograma..... 55

Figura 4.1 Representaciones tiempo frecuencia, Figura (4.1a). STFT, Figura (4.1b). Transformada Wavelet Continua (CWT) y en Figura (4.1c). Transformada Adaptativa de Superlets (ASLT)..... 58

Figura 4.2 Matriz de Confusión Global de AlexNet para dos clases, sin smoothing, sin data augmentation..... 60

Figura 4.3 Efecto visual de smoothing sobre un espectrograma..... 61

Figura 4.4 Comparación de las arquitecturas CNN con datos sin smoothing y data augmentation para el escenario de dos clases con evaluación de Accuracy, tiempo de predicción y tamaño del modelo. 70

Figura 4.5 Comparación de las arquitecturas CNN con smoothing y con data augmentation para el escenario de tres clases con evaluación de Accuracy, tiempo de predicción y tamaño del modelo. 70

Resumen

Los radares de onda milimétrica constituyen una tecnología cada vez más requerida dentro del internet de las cosas, el cual atraviesa una etapa de auge y crecimiento acelerado, con presencia en prácticamente todos los ámbitos de la vida cotidiana. Esta expansión exige dispositivos pequeños, compactos, de bajo consumo energético y respetuosos con la privacidad del usuario; características que los radares de onda milimétrica satisfacen, permiten la detección de presencia y movimiento en entornos interiores sin comprometer la privacidad de los usuarios, de modo que posteriormente ejecute acciones acordes a cada situación. Sin embargo, clasificar con precisión humanos y animales en espacios de interiores es un desafío dado que, las trayectorias múltiples y la variabilidad del movimiento biológico generan objetivos falsos que reducen la eficacia del sistema. Por tal motivo, el presente trabajo tuvo como objetivo mejorar el desempeño del radar de onda milimétrica en la discriminación entre humanos y animales mediante la integración de procesamiento digital de imágenes y aprendizaje automatizado profundo. Se empleó una metodología cuantitativa y experimental, mediante un dataset propio de señales recolectadas de personas, perros y gatos en escenarios reales. Para diversificar las muestras y optimizar la calidad visual de los espectrogramas micro doppler se aplicaron técnicas de data augmentation y smoothing respectivamente antes de su ingreso a las redes neuronales. Con estas muestras se entrenaron cuatro arquitecturas de redes neuronales convolucionales, comparadas estadísticamente para identificar el modelo con mejor capacidad predictiva. Los resultados demuestran que la integración de un dataset representativo, procesamiento digital de imágenes y CNNs seleccionadas estadísticamente supera las limitaciones de los métodos tradicionales, donde alcanza una exactitud superior al 95% y con automatización de la clasificación de blancos en interiores.

Palabras clave: micro doppler, procesamiento digital de imágenes, predicción, aprendizaje automático.

Abstract

Millimeter wave radars constitute an increasingly required technology within the Internet of Things, which is undergoing a stage of accelerated growth and expansion, with a presence in virtually every area of daily life. This expansion demands devices that are small, compact, energy efficient, and respectful of user privacy, characteristics that millimeter-wave radars satisfy, enabling the detection of presence and movement in indoor environments without compromising user privacy, so that appropriate actions can subsequently be executed according to each situation. However, accurately classifying humans and animals in indoor spaces is a challenge, given that multiple trajectories and the variability of biological movement generate false targets that reduce the system's effectiveness. For this reason, the present work aimed to improve the performance of millimeter wave radar in discriminating between humans and animals through the integration of digital image processing and deep machine learning. A quantitative and experimental methodology was employed, using a proprietary dataset of signals collected from people, dogs, and cats in real world scenarios. To diversify the samples and optimize the visual quality of the micro doppler spectrograms, data augmentation and smoothing techniques were applied, respectively, prior to their input into the neural networks. With these samples, four convolutional neural network architectures were trained and statistically compared to identify the model with the best predictive capability. The results demonstrate that the integration of a representative dataset, digital image processing, and statistically selected CNNs surpasses the limitations of traditional methods, achieving an accuracy above 95% while automating target classification in indoor environments.

Keywords: micro doppler, digital image processing, prediction, machine learning.

Capítulo 1

Introducción

Durante las últimas décadas, el radar fue una tecnología exclusiva del ámbito militar y aeroespacial, donde se empleaba para detectar objetivos y medir velocidades sin preocuparse por el tamaño, el costo o el consumo energético de los equipos [1].

Con el avance tecnológico fue posible integrar todos los componentes del radar; transmisor, receptor, procesador de señal; en un único chip de pequeñas dimensiones y bajo consumo. Este salto tecnológico abrió la puerta a su adopción masiva en el mundo civil y comercial [2], [3].

En la actualidad, estos dispositivos compactos encuentran aplicación en numerosos sectores, tales como, los hogares inteligentes que son uno de los ámbitos de mayor expansión. En este contexto, los radares de ondas milimétricas (mmWave) permiten identificar la presencia de personas, estimar la distancia y velocidad de objetos, así como interpretar gestos para el control de dispositivos domésticos, todo ello sin requerir cámaras ni sensores físicos de contacto [4].

A diferencia de soluciones como las cámaras o los sensores infrarrojos, el radar mmwave funciona de forma ininterrumpida independientemente de las condiciones de luz del entorno, y al no capturar imágenes, garantiza la privacidad de las personas en sus hogares [4]. Estas ventajas lo posicionan como una alternativa idónea para espacios domésticos donde la seguridad, el confort y la discreción son aspectos fundamentales.

Con el avance del Internet de las Cosas (IoT, del inglés Internet of Things), se impulsa la integración de radares en entornos interiores, los cuales demandan cada vez mayor precisión para discriminar seres humanos de animales.

Lo cual resulta fundamental en diversas aplicaciones de respuesta inmediata, como la activación de alertas ante personas no autorizadas, el monitoreo de adultos mayores para detectar caídas o movimientos inusuales [5], y la supervisión de mascotas en vehículos mediante mecanismos de climatización automática. Por ello, una clasificación precisa entre personas y animales es crítica, ya que

un error puede derivar en falsas alarmas, omisión de riesgos reales o consecuencias irreversibles como la muerte de una mascota o un menor de edad por exposición a temperaturas extremas [6], [7].

Para poder distinguir entre objetivos se necesita recabar información mediante la captura de señales de las variaciones dinámicas producidas por los objetivos detectados en el rango del radar, conocidas como firmas de frecuencia micro doppler [8].

Las cuales contienen información detallada sobre los patrones de movimientos únicos entre personas y animales no obstante, para lograr esta clasificación precisa es indispensable el desarrollo e implementación de algoritmos de aprendizaje automáticos debidamente entrenados que sean capaces de interpretar y discriminar dichos patrones de forma eficiente.

Para lograr dicho resultado se debe conocer el comportamiento del efecto micro doppler en interiores. La clasificación entre humanos y animales en entornos dinámicos persiste como un reto técnico aun no resuelto por los métodos tradicionales. Este desafío se intensifica con falsas alarmas en la detección en entornos interiores; debido a la reflexión de las ondas electromagnéticas en techos, paredes, pisos y otros objetos; es decir generan blancos falsos con características similares a las de un objetivo real. En cuanto a la eliminación de trayectorias múltiples, los métodos más comunes pueden dividirse en dos categorías:

- Métodos basados en datos, estos emplean algoritmos de aprendizaje automático para discriminar blancos reales de falsos. Según [9] ha evidenciado en el estudio que los clasificadores de redes neuronales profundas han alcanzado una precisión promedio del 87,36 %. No obstante, la falta de datos representativos y su elevada complejidad computacional restringen su uso práctico; además, la extracción manual de características, como el Análisis de Componentes Principales (PCA, del inglés Principal Component Analysis) o la Descomposición en Valores Singulares (SVD, del inglés Singular Value Decomposition), no logran capturar adecuadamente la variabilidad de los patrones de movimiento en blancos no rígidos [10].

- Métodos basados en modelos optimizan la detección mediante formulaciones matemáticas y procesamiento de señales, técnicas como la entropía ponderada, canales diferenciales temporales, mínimos cuadrados, NLOS (Non Line of Sight) y la combinación MIMO (Multiple Input Multiple Output) junto con la transformada de Hough para generar mapas de rango doppler que suprimen ecos falsos y detectan patrones. Además, se complementan con filtros híbridos de clutter (HCM LMB, del inglés Hybrid Clutter Model based Labeled Multi Bernoulli Filter), localización MAL (Multipath Assisted Localization), y coincidencia de trayectorias, validados en entornos complejos. Su robustez y baja demanda computacional los consolidan como estándar en la supresión de ecos parásitos [9].

En [11] se aborda la discriminación entre humanos y animales mediante radar de banda ultra ancha (UWB, del inglés, Ultra Wideband) en operaciones de búsqueda y rescate post desastre donde se demuestra que es posible distinguir entre ambas especies a partir del coeficiente de correlación de micro vibración corporal (CCMV, del inglés Correlation Coefficient of Body Micro Vibration) con el fin de superar las limitaciones de los métodos convencionales basados en actividad respiratoria, donde se obtuvo un CCMV promedio de 0.980 para objetivos humanos y 0.868 para perros diferencia significativa que permitió la discriminación efectiva entre ambas especies mediante radar UWB.

De igual forma, en Chen et. [12] establecieron el modelo matemático del efecto micro doppler en el cual exponen la Transformada de Fourier de Tiempo Corto (STFT, del inglés Short-Time Fourier Transform) como herramienta principal para el análisis en el dominio tiempo-frecuencia, que es la base para la clasificación de objetivos a partir de sus firmas de movimiento. Sobre todo, demostraron que las firmas micro doppler de la marcha humana permiten clasificar actividades con alta precisión mediante algoritmos de Machine Learning y Depp Learning.

Cabe destacar, que la implementación de los sistemas de radar ha variado históricamente según el tipo de objeto que se desea localizar. Por ejemplo, en aviones, barcos, drones y helicópteros realiza las detecciones de estos objetos a través de las estimaciones de la sección transversal de radar (RCS, del inglés Radar Cross Section). De igual manera, admite el análisis de firmas de modulación y la cinemática

del objetivo; en efecto permite evaluar su trayectoria global. No obstante, en el caso de las personas y animales la dificultad técnica aumenta. Debido a que sus firmas de radar no son estables ni predecibles, sino que dependen de factores biológicos, tal como la complexión física y la cadencia del desplazamiento [10].

Ahora bien, para extraer estas características de los objetivos se requiere de un hardware especializado, como el radar de ondas milimétricas. Este dispositivo de detección avanzada emite y recibe señales de forma autónoma para detectar objetos y medir sus características. Por ejemplo, en [13] se presenta un radar mmWave de 60 GHz, el cual permite detectar movimiento y dirección en un formato extremadamente compacto y de bajo consumo energético.

Así mismo, al tener una longitud de onda reducida de 5mm, maximiza la sensibilidad de las firmas micro doppler [14]. Esto se traduce en una mayor resolución espectral, lo que permite una extracción de una densidad superior de características discriminatorias [15]; por lo mismo, capturan con precisión los matices sutiles del movimiento de las extremidades; a su vez, facilita la extracción de características morfológicas y cinemáticas, estas son necesarias para el entrenamiento de las arquitecturas de redes neuronales convolucionales (CNN, del inglés Convolutional Neuronal Networks).

En cuanto, el auge de los radares en el IoT, se debe a su rol como puente entre el mundo físico y el digital; en contraste con sensores tradicionales, el radar mmWave opera en oscuridad total, atraviesa barreras no metálicos y entrega mediciones en tiempo real sin comprometer la privacidad del usuario, cualidades que lo posicionan como un componente idóneo dentro de ecosistemas de IoT orientados a la detección de presencia y movimiento en interiores.

Estas ventajas , combinadas con la proyección estimada de 75 mil millones de dispositivos conectados para el 2025 [16], han disparado la demanda de sistemas de rastreo de ubicación y movimiento precisos y seguros.

Para sostener estas aplicaciones criticas como: domótica, seguridad y salud digital , se hace indispensable el despliegue masivo de hardware miniaturizado y de bajo consumo energético [17].

Ahora bien, en este trabajo se propone el reconocimiento (detección y clasificación) de personas y animales en espacios interiores. Para esto las CNN surgen como solución para extraer características jerárquicas de los espectrogramas con la información micro doppler de los objetivos; tales como, la captura de patrones espacio temporales complejos y rasgos sutiles que los métodos estadísticos lineales pasan por alto. Esto permite una discriminación fiable en escenarios complejos, y así superar la impredecibilidad propia de los movimientos biológicos.

El uso de CNN permite tener una mejora significativa en el desempeño; a partir de diferentes representaciones en tiempo frecuencia. Así pues, permite extraer características discriminativas de las firmas micro doppler y utilizarlas para entrenar modelos de clasificación automática.

A diferencia del estudio [18], desarrollado en condiciones experimentales ideales y entornos controlados, la presente, investigación aborda la complejidad de escenarios reales mediante la generación de un conjunto de datos (dataset) propio; con la captura de diversos blancos bajo múltiples ángulos de visión. Así mismo, representar con mayor fidelidad el efecto micro doppler en presencia de reflexiones y obstrucciones comunes en interiores, como resultado reducir el sobreajuste y la fortalecer la capacidad de generalización de los algoritmos.

Para validar esta aproximación, se evaluaron estadísticamente cuatro arquitecturas de CNN, con el objetivo de identificar cuál ofrece el mejor desempeño en la discriminación entre humanos y animales.

Este trabajo contribuye al avance de la infraestructura de detección inteligente y sistema de respuesta inmediata; alineándose con el objetivo de fomentar la innovación tecnológica (ODS 9) en el área de telecomunicaciones.

Mas allá de la precisión algorítmica, esta tecnología resuelve desafíos sociales para preservar un bien cada vez más escaso en la actualidad “la privacidad del usuario”. Como se mencionó anteriormente a diferencia de otros sensores (como: cámaras, micrófonos) el radar mmwave no captura imágenes, rostros ni audios, sino solo patrones de movimiento mediante el efecto micro doppler.

Surge como una solución anónima y apta para entornos sensibles como hogares, clínicas o residencias de adultos mayores, en aplicaciones de monitoreo remoto de la salud, en operaciones de

búsqueda y rescate tras desastres naturales, donde optimizar el tiempo de respuesta es vital. Un ejemplo claro es el caso de Japón, donde más del 29% de su población supera los 65 años y el gobierno ha impulsado activamente sistemas de monitoreo domiciliario inteligente para salvaguardar el bienestar de sus adultos mayores, sin embargo, la resistencia cultural a sentirse vigilados representa un obstáculo significativo. En este escenario, el radar mmWave se consolida como una alternativa no invasiva: sin cámaras, sin micrófonos y sin dispositivos adheridos al cuerpo, que permiten monitorizar caídas, localizar individuos en el hogar o identificar irregularidades respiratorias durante el sueño, todo ello para procesar únicamente señales de radiofrecuencia que no revelan identidad ni imagen alguna del individuo [19].

En definitiva, este estudio promueve una transición hacia entornos asistenciales inteligentes alineado con el despliegue masivo de IoT; donde se requiere dispositivos compactos, de bajo consumo energético, capaces de procesar y decidir en tiempo real en entornos donde la seguridad y la privacidad del usuario convergen a través de la tecnología mmWave, sin sacrificar la dignidad e intimidad de las personas en nombre de la eficiencia tecnológica.

Objetivo general:

Implementar algoritmos de clasificación de blancos en radares de onda milimétrica para entornos interiores.

Objetivos específicos:

- Analizar y definir las características de radares doppler de onda milimétrica (mmWave).
- Adquirir señales en fase y cuadratura de radar doppler a 60GHz
- Procesar señales de radar para obtener micro firmas doppler.
- Implementar algoritmos de clasificación de blancosbasados en Machine/Deep Learning.
- Evaluar el desempeño del sistema en clasificación de humanos y animales.

El presente trabajo se estructura en seis capítulos. Los siguientes capítulos son: Capítulo II donde desarrolla el marco teórico, que aborda los fundamentos de las ondas milimétricas, la teoría del radar, el efecto doppler y el efecto micro doppler, el radar de onda continua, la transformada de tiempo corto de

Fourier (STFT) y finalmente las arquitecturas de redes neuronales convolucionales empleadas: AlexNet, ResNet50, MobileNetV2 y SqueezeNet.

El Capítulo 3 presenta la metodología experimental, que comprende el proceso de adquisición de datos con el radar BGT60LTR11AIP, las técnicas de aumento de datos (data augmentation), mejoramiento de espectrogramas micro doppler mediante procesamiento digital de imágenes (smoothing), la preparación de los datasets, la transferencia de aprendizaje aplicada a cada arquitectura y el procedimiento de entrenamiento y evaluación mediante validación cruzada de K folds.

Los Capítulos 4 presenta los resultados obtenidos, donde se compara estadísticamente el desempeño de las cuatro arquitecturas evaluadas y su interpretación de las métricas de clasificación alcanzadas. Finalmente, el Capítulo 5 expone las conclusiones derivadas del estudio, y de esta forma responder los objetivos planteados y señalando posibles líneas de trabajo futuro.

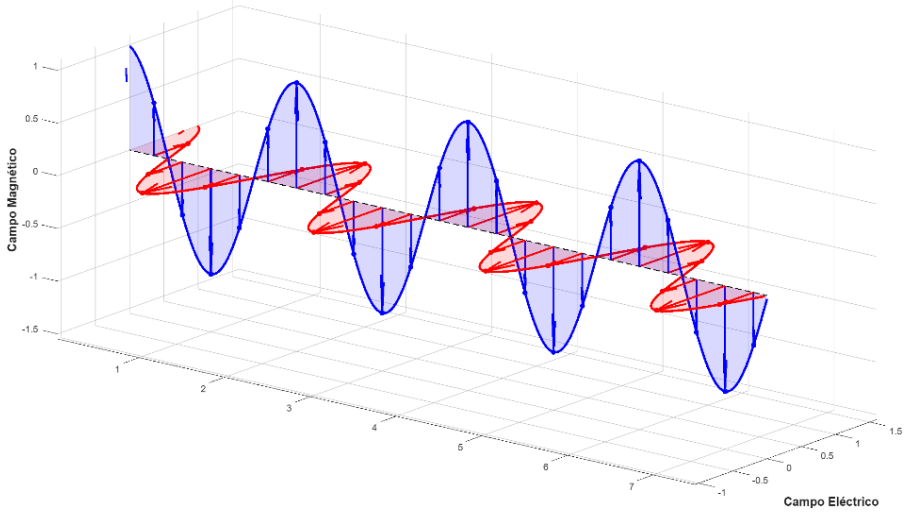
Capítulo 2

Marco Teórico

Fundamentos teóricos de ondas milimétricas

En física, una onda es el proceso por el cual una perturbación viaja desde un punto emisor hasta otro receptor, transporta energía, pero sin desplazar materia. Cada perturbación individual se llama pulso, y cuando varios pulsos se juntan, forman lo que se conoce como tren de ondas. Las ondas electromagnéticas son un caso especial, ya que no necesitan un medio físico para propagarse: su energía se transmite mediante la vibración de los campos eléctrico y magnético [20]. La velocidad con que se propaga la onda electromagnética dependerá de las propiedades eléctricas y magnéticas presentes en el medio.

Figura 2.1
Ondas electromagnéticas.



En la Figura 2.1 se observa la representación esquemática de una onda electromagnética la cual muestra la oscilación perpendicular de los campos eléctrico y magnético respecto a la dirección de propagación.

El espectro electromagnético comprende una extensa gama de frecuencias que va desde las ondas de radio de baja frecuencia hasta los rayos gamma de elevada energía cada una con propiedades físicas y aplicaciones distintas. Dentro del amplio espectro electromagnético, las mmWave constituyen un

subconjunto particular situado en la región de frecuencias más elevadas. Se encuentran entre 30 GHz y 300 GHz, cuyas λ oscilan entre 1 mm y 10 mm, las cuales están designadas como banda de frecuencia extremadamente alta (EHF, del inglés Extremely High Frequency) por la Unión Internacional de Telecomunicaciones (ITU, del inglés International Telecommunications Union).

Gracias a la amplia disponibilidad de ancho de banda en este rango frecuencial, las comunicaciones mmWave se han convertido en una tecnología fundamental para sistemas de quinta generación (5G), redes de alta velocidad, aplicaciones de sensado remoto y radares avanzados, como las aplicadas en este estudio.

Sin embargo, la propagación de las señales en esta banda de frecuencia cuenta con limitaciones. Debida a su reducida longitud de onda, las ondas milimétricas sufren una alta pérdida por trayectoria en el espacio libre (FSPL, del inglés Free Space Path Loss) en comparación con las frecuencias más bajas, según la relación (ver, Ec. 1) [21]:

$$F_p = 32.44 + 20\log_{10}(d) + 20\log_{10}(f), \quad (1)$$

donde,

- d = distancia en kilómetros (Km)
- f = frecuencia en (Mhz)
- F_p = perdidas en espacio libre (dBm)

Además, requieren línea de vista directa (LoS, del inglés Line of Sight) estricta entre el transmisor y el receptor. La presencia de obstáculos físicos, ya sean estructuras edilicias, follaje o el propio cuerpo humano, provoca atenuaciones severas o bloqueos totales de la señal [21]. Por esta razón, el comportamiento de las comunicaciones mmWave difiere considerablemente del de los sistemas tradicionales de microondas, dado que presenta:

Amplio ancho de banda disponible: Las bandas mmWave tiene la disponibilidad de grandes porciones de espectro, lo que permite velocidades de transmisión muy superiores a las obtenidas en bandas tradicionales.

Longitud de onda reducida: Dado que las dimensiones físicas de una antena son directamente proporcionales a su longitud de onda, operar en EHF permite el diseño de arreglos de antenas sumamente compactos. Esta miniaturización facilita la integración del sistema completo de radiofrecuencia y procesamiento de señales en un único circuito integrado, lo que reduce el tamaño, peso, costo y consumo energético del hardware [21].

Alta direccionalidad: Las antenas de onda milimétrica generan haces muy estrechos lo que permite concentrar la energía en una dirección específica. Esto incrementa la ganancia de antena y reduce la interferencia entre usuarios.

Alta resolución espacial: Debido a su pequeña longitud de onda, las señales mmWave permiten una mejor resolución para aplicaciones de radar, detección de objetos y monitoreo de movimientos humanos [22].

Teoría de Radar

El radar es un sistema electrónico que utiliza la reflexión de ondas electromagnéticas para determinar la posición, distancia y velocidad de los objetos. Desarrollado inicialmente con fines militares, su uso se extendió rápidamente a ámbitos como la aviación, la navegación marítima, la meteorología y el control del tráfico aéreo [23]. Con el paso del tiempo y el avance tecnológico en circuitos integrados y sistemas computacionales, esta tecnología ha experimentado una evolución continua hasta consolidarse como un instrumento de alta precisión, capaz de operar con fiabilidad en condiciones adversas donde otros sistemas de detección resultan insuficientes [2].

Existen dos tipos principales de radares: los primarios, que detectan cualquier objeto (embarcaciones, aeronaves, obstáculos naturales) y constituyen la mayoría de los sistemas actuales; y los secundarios, más sofisticados y empleados en control de tráfico aéreo y marítimo. Estos últimos emplean un transponder que responde con información codificada del objetivo al recibir la señal, en ausencia de transponder activo, solo los radares primarios permiten la detección. Independientemente del tipo de radar, todos comparten un mismo principio físico de operación para determinar la posición de los objetivos.

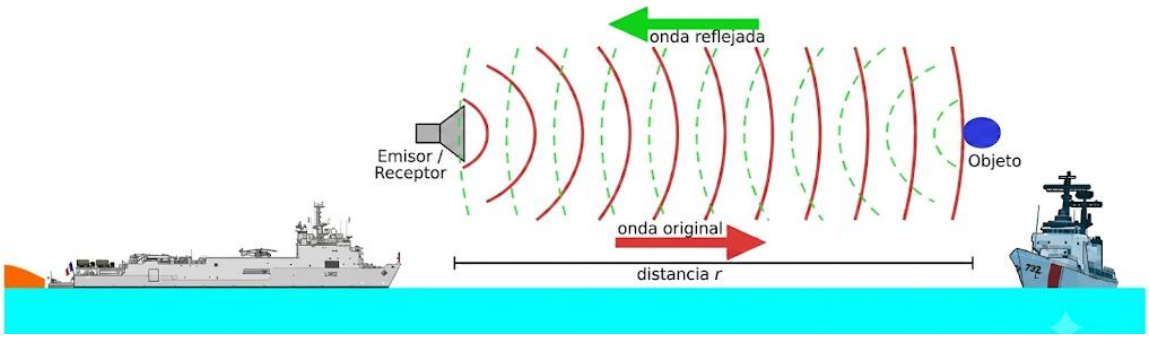
El radar mide el tiempo que tarda un pulso electromagnético en ir hasta el objetivo y regresar. Como la señal recorre dos veces la distancia (ida y vuelta), se divide el resultado entre 2 para obtener la distancia real al objeto (ver. Ec 2) [20]:

$$R = \frac{c * t}{2}, \tag{2}$$

donde,

- C=3x10^8 m/s
- t =tiempo de vuelo de la señal (s).
- R= Rango del objetivo – distancia del objetivo (m)

Figura 2.2
Distancia radar. Tomado de [19]

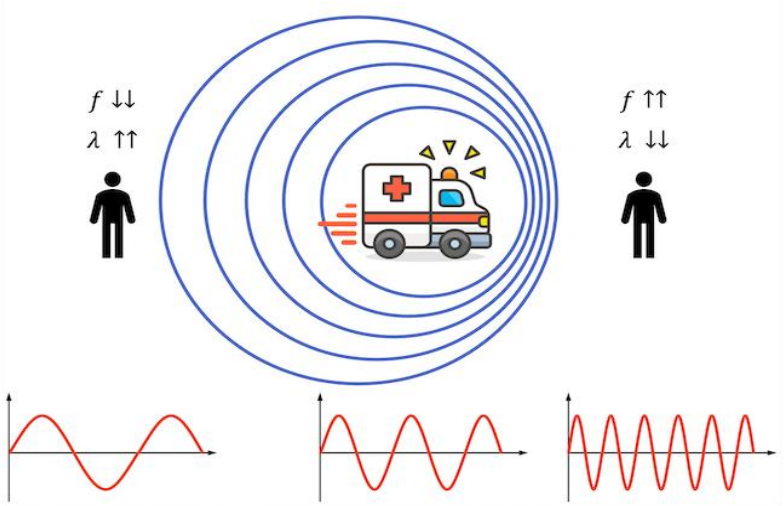


En Figura 2.2 se puede evidenciar cómo funciona la operación del radar para determinar la distancia mediante los ecos de la señal que rebota en el objetivo.

Efecto Doppler

El efecto doppler es el cambio aparente en la frecuencia percibida de un movimiento ondulatorio cuando existe desplazamiento relativo entre la fuente emisora y el receptor. Este fenómeno constituye uno de los principios fundamentales para la detección y medición de velocidades mediante sistemas de radar [24].

Figura 2.3
Comportamiento del efecto doppler. Tomado de [24]



En la Figura 2.3 se puede visualizar la variación de la frecuencia y longitud de onda en el efecto doppler debido a el movimiento relativo entre el transmisor y el receptor.

El efecto fue descrito por primera vez para ondas sonoras por el físico austriaco Christian Andreas doppler en 1842 en [20], quien observó que el tono percibido de una fuente sonora se agudizaba al acercarse al observador y se volvía más grave al alejarse. Años más tarde, en 1848, el físico francés Armand Hippolyte Louis Fizeau identificó de manera independiente un fenómeno análogo en ondas electromagnéticas, motivo por el cual este principio también se conoce como efecto doppler-Fizeau [20].
Fórmula matemática General:

Para un caso general de propagación ondulatoria en un medio, la frecuencia percibida por el receptor se expresa como (ver. Ec 3) [20], [25]:

$$f_a = \left(\frac{V + V_r}{V + V_s}\right) * f_0, \tag{3}$$

donde:

- f_a= frecuencia doppler observada (Hz)
- f_0 = frecuencia emitida (Hz)
- V = velocidad de propagación de ondas en el medio

- V_r =velocidad del receptor. Si el receptor se mueve hacia el emisor, V_r es positiva. Pero si el receptor mueve en sentido opuesto, V_r es negativa.
- V_s = velocidad de la fuente. Si la fuente se aleja del receptor, V_s es positiva. Pero si la fuente se acerca al receptor, V_s es negativa.

En un radar monoestático, donde transmisor y receptor comparten ubicación, la señal recorre una distancia de ida y vuelta entre el radar y el objetivo. Dado que las ondas electromagnéticas se propagan a la velocidad de la luz (c), la frecuencia doppler (f_d), definida como la diferencia entre la frecuencia reflejada (f_r) y la frecuencia transmitida (f_t), se expresa como (ver. Ec 4) [20]:

$$f_d = f_r - f_t = \frac{2vf_t}{c-v} \approx \frac{2vf_t}{c}, \tag{4}$$

donde,

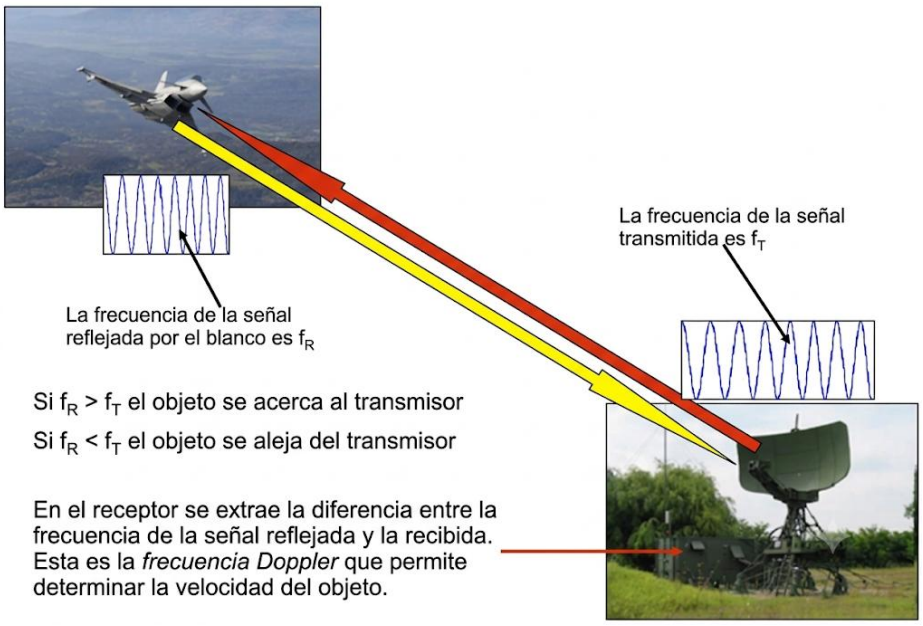
- f_r= frecuencia reflejada por el blanco
- f_t=frecuencia transmitida por el radar.
- f_d=frecuencia doppler
- v=velocidad relativa entre la fuente radar y el blanco.
- c= velocidad de la luz

El coeficiente 2 presente en el numerador de la formula surge precisamente del trayecto de ida y vuelta de la señal, mientras que la aproximación final es válida cuando $V \ll C$, condición que se cumple en la generalidad de las aplicaciones de radar [20], [26].

A partir de esta relación, conocidas la frecuencia transmitida F_t y la frecuencia doppler F_d, la velocidad del objetivo puede determinarse mediante (ver. Ec 5) [20]:

$$V = \frac{c * f_d}{2f_t}, \tag{5}$$

Figura 2.4
Ejemplo del comportamiento del efecto doppler



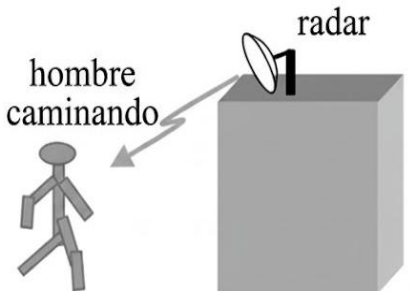
En la Figura 2.4 se puede visualizar el comportamiento del efecto doppler en un caso real en el ámbito militar.

Efecto micro doppler

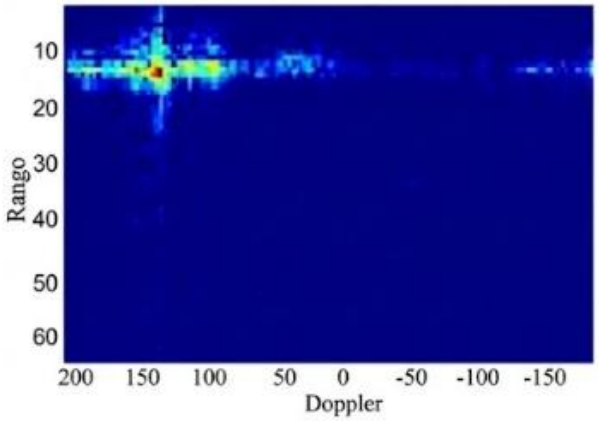
El efecto micro doppler corresponde a las modulaciones adicionales en la frecuencia de la señal reflejada, generadas por movimientos internos del objetivo (vibraciones, oscilaciones o rotaciones) que se superponen al movimiento de traslación. Desde el punto de vista matemático, el desplazamiento doppler total se compone de la suma de dos elementos: la componente doppler traslacional y la componente micro doppler.Estas modulaciones se manifiestan como bandas laterales alrededor del desplazamiento doppler principal y constituyen una firma única del objetivo, ya que reflejan sus características dinámicas y mecánicas. Los micro movimientos pueden provenir de rotores, hélices, vibraciones de motores o el movimiento de extremidades humanas durante la marcha, lo que permite diferenciar entre tipos de vehículos o analizar patrones de movimiento humano [4], [12], [27].

Figura 2.5
Firmas micro doppler de una persona caminando moviendo los brazos. Tomado de [27]

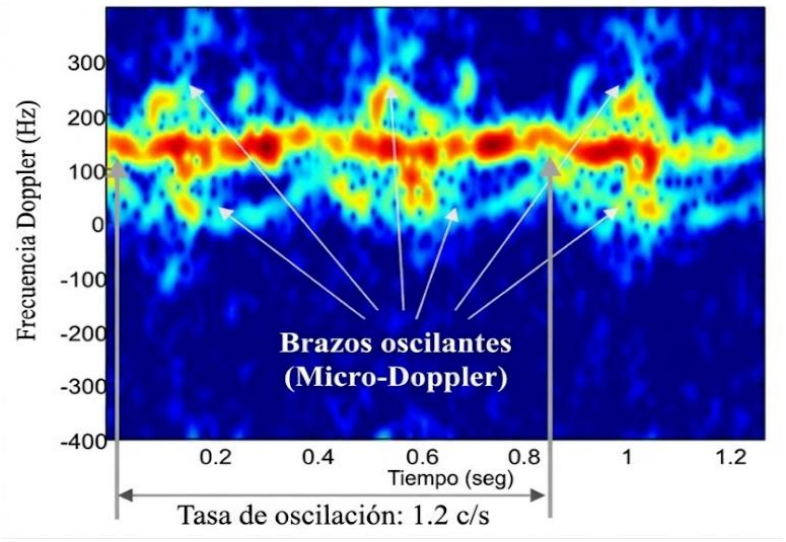
(2.5a)



(2.5b)



(2.5c)



En la Figura 2.5 se puede visualizar el análisis de la firma micro doppler en un objetivo humano en movimiento. En la Figura (2.5a) muestra a una persona caminando frente al sensor de radar. En la Figura (2.5b), muestra la imagen de radar en el dominio tiempo frecuencia, donde se localiza el blanco en la celda de estudio. En la Figura (2.5c), muestra el espectrograma de la señal ampliada, donde muestra la componente doppler central del torso y las bandas laterales oscilantes correspondientes a la firma micro doppler generada por el balanceo de los brazos.

Radar de Onda Continua

Un Radar Doppler o también conocido como radar de onda continua (CW, del inglés Continuous Wave), emite una señal de alta frecuencia con amplitud constante de forma continua, mientras que simultáneamente recibe y procesa de forma permanente la señal reflejada por los objetivos en su campo de detección.

La señal de eco recibida tiene la misma frecuencia de emisión cuando el objetivo se encuentra estático; sin embargo, en presencia de un objetivo con movimiento de acercamiento o alejamiento, la frecuencia de eco experimenta un corrimiento proporcional a su velocidad, este fenómeno se conoce como efecto doppler [26], [28].

Este radar emplea una configuración de antenas separadas, con una designada para transmisión de la señal electromagnética y la otra a la recepción del eco. Esta arquitectura garantiza un alto aislamiento entre el transmisor y el receptor, lo que minimiza la fuga directa de señal (spillover) y el clutter generado por reflexiones de objetos cercanos. Como resultado, se incrementa notablemente la sensibilidad del sistema y se previene la saturación del receptor.

Si bien los radares CW pueden operar con una sola antena mediante el uso de duplexores y técnicas de cancelación, en aplicaciones de baja potencia; ya que el sistema de cancelación no puede compensar completamente las reflexiones producidas por el duplexor y la propia antena, lo que provoca degradación en el receptor. El diseño de doble antena se consolida como la opción preferida en aplicaciones de alta sensibilidad debido a su superior desempeño y eficacia en la mitigación de interferencias internas [26].

Cálculo del alcance del radar (ver. Ec 6), [25], [26]:

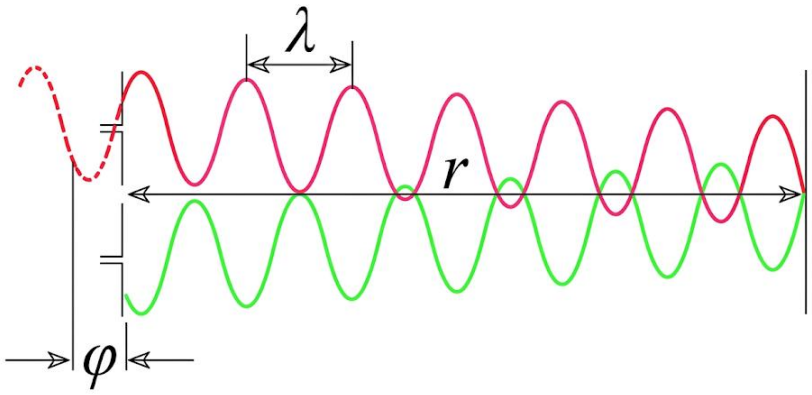
$$R_{max} = \sqrt[4]{\frac{P_s * G^2 * \lambda^2 * \sigma}{P_{Emin} * (4\pi)^3 * L}}, \tag{6}$$

donde,

- P_s= Potencia de Transmisión (w)
- G= Ganancia de la Antena
- σ = Radar Cross Section (m/s^2)
- P_{Emin}=Potencia mínima Detectable por el Receptor
- L=Pérdidas totales del sistema
- λ=Lambda
- R_{max}= Alcance máximo del radar (m)

El radar de onda continua evalúa la diferencia de fase φ entre la señal transmitida y la señal recibida. La magnitud de esta diferencia de fase es la relación entre la distancia recorrida por la onda electromagnética y la longitud de onda de la señal transmitida, multiplicado por la división de grados del círculo completo (2·π).

Figura 2.6
Diferencia de Fase. Tomado de [25]



En la Figura 2.6 se puede visualizar el principio de medición por diferencia de fase en radares de onda continua: el desfase (φ) entre la señal emitida y la reflejada es proporcional a la distancia recorrida por la onda electromagnética, expresado en función de la longitud de onda (λ).

La magnitud de esta diferencia de fase es la relación entre la longitud de onda de la señal transmitida y la distancia recorrida. Si la distancia al reflector no cambia, entonces es constante y se calcula de la siguiente manera (ver. Ec 7) [25], [26]:

$$\varphi = -2\pi\frac{2r}{\lambda} = -4\pi\frac{r(t)}{\lambda}\ , \tag{7}$$

donde,

- φ= Diferencia de fase
- r=distancia del reflector a la antena

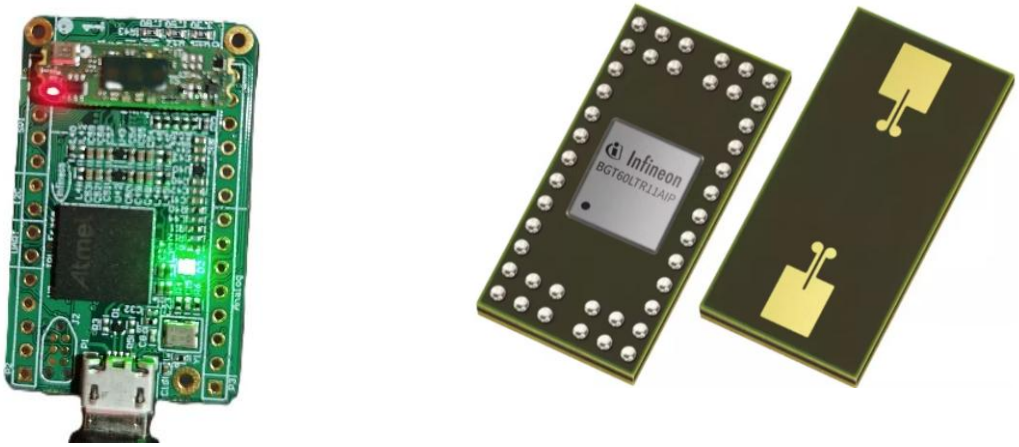
Un cambio constante y dependiente del tiempo en la diferencia de fase entre dos señales sinusoidales durante el periodo de medición corresponde nuevamente a una curva de señal sinusoidal.

Radar CW Infineon:

El sensor BGT60LTR11AIP de Infineon Technologies opera en la banda de 60 GHz catalogada mundialmente como banda ISM (Industrial, Scientific and Medical) de uso libre. Esta asignación espectral combinada con la integración de antenas mediante tecnología AIP (Antennas in Package) ha impulsado el desarrollo de sensores comerciales compactos lo que hace al radar de 60GHz de infineon una plataforma accesible y legalmente viable para proyectos de investigación.

Figura 2.7
Radar BGT60LTR11AIP

(2.7a) (2.7b)



En la Figura (2.7a) se puede observar la PCB que en el cual está integrado el radar BGT60LTR11AIP visualizado en la Figura (2.7b), tomada de [13].

Este dispositivo presenta un tamaño empaquetado de 3,3 × 6,7 × 0,56 mm³, lo que permite su implementación sobre placas de circuito impreso basadas en material estándar FR4, sin requerir sustratos de alta frecuencia costosos. Opera en el rango de 61 GHz a 61,5 GHz con un ancho de banda de 500 MHz, una ganancia de antena de 6 dBi y un campo de visión (HPBW, Half Power Beam Width) de 90°.

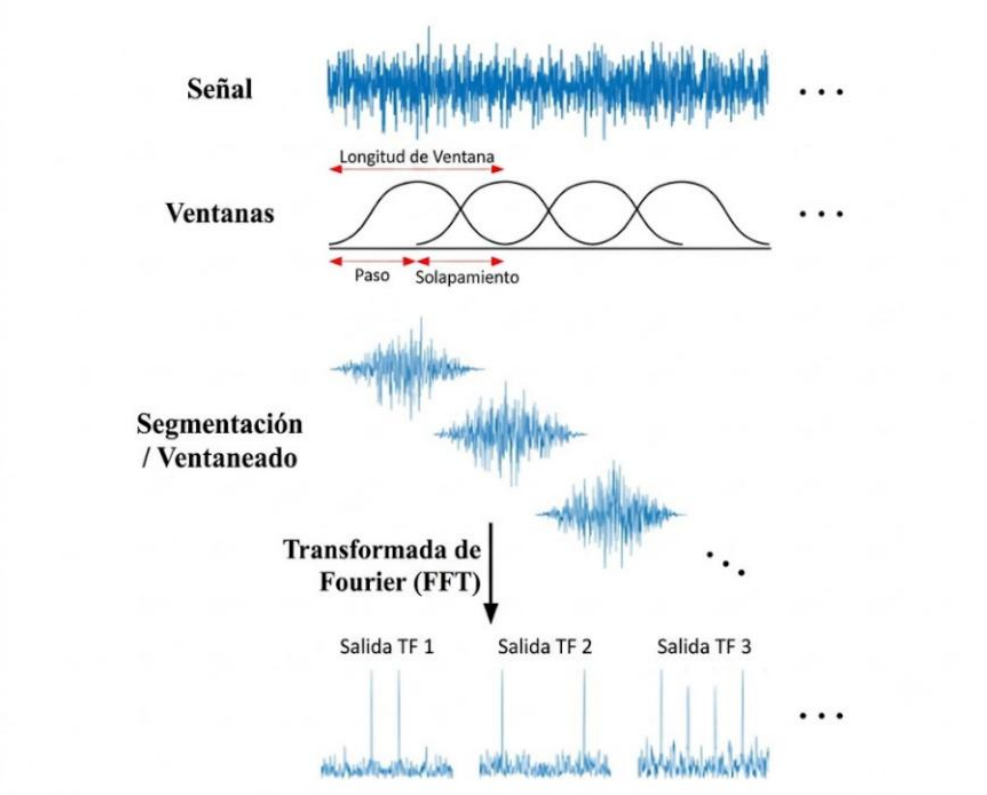
El BGT60LTR11AIP destaca por su flexibilidad operativa al disponer de dos modos de funcionamiento diferenciados. En su modo autónomo utiliza detectores de movimiento y dirección integrados en el hardware lo que permite una detección básica de objetivos humanos a una distancia de hasta 10 m con un consumo energético inferior a 2 mW.

Para aplicaciones de mayor complejidad, como la extracción de firmas micro doppler para la clasificación de objetivos, el radar permite una configuración avanzada mediante la interfaz periférica serial (SPI, Serial Peripheral Interface). Al trabajar en conjunto con una unidad de microcontrolador externa (MCU), se habilitan algoritmos de procesamiento digital de señales y se incrementa el rango máximo de detección para objetivos humanos hasta 14m [13].

Transformada de Fourier de tiempo corto (STFT)

La STFT constituye una herramienta de análisis que combina información temporal y frecuencial simultáneamente, lo cual la hace especialmente adecuada para analizar señales no estacionarias, cuyo contenido espectral cambia conforme transcurre el tiempo.". A diferencia de la Transformada de Fourier clásica, en donde se abarca toda la duración de la señal para el análisis de frecuencia, la STFT analiza la señal a través de ventana temporales finitas y que se encunetar solapadas.

Figura 2.8
Procedimiento de la Transformada rápida de Fourier (STFT). Tomado de [29].



En la Figura 2.8 se muestra el proceso de análisis de la señal mediante ventanas temporales solapadas para obtener la representación en el dominio tiempo - frecuencia.

Matemáticamente, para una señal continua $x(t)$, la STFT se define la (ver Ec 8). [30]:

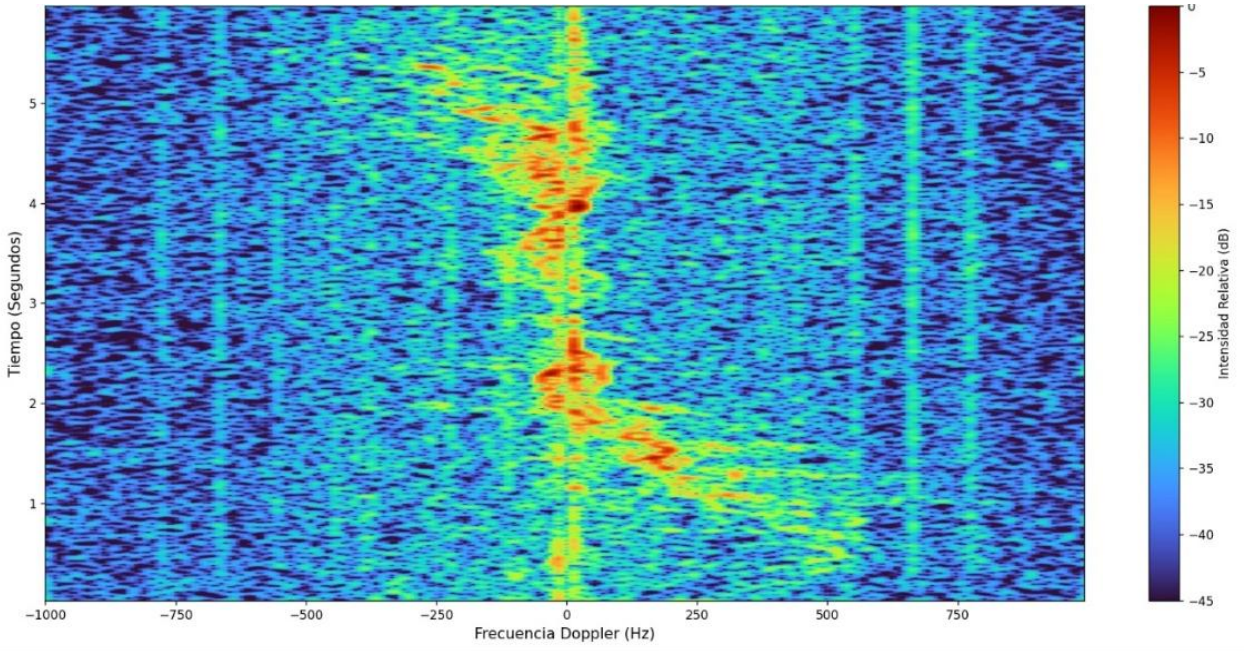
$$STFT\{x(t)\}(\tau,\omega) = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-j\omega t}\,, \tag{8}$$

donde $w(t-\tau)$ es la función de ventana centrada en el instante τ, y w es la frecuencia angular. En tiempo discreto, la STFT se calcula al multiplicar la señal $x[n]$ por una función de ventana $w(n-n_k)$ centrada en el instante de trama n_k seguido de la transformada de Fourier, lo que produce la representación tiempo-frecuencia $X[n_k, f_i]$.

El cuadrado del módulo de la STFT da lugar al espectrograma, que constituye una representación visual de la densidad espectral de potencia de la señal en función del tiempo (ver Ec.9) [30].

$$S(\tau, \omega) = |STFT\{x(t)\}(\tau, \omega)|^2 , \quad (9)$$

Figura 2.9
Representación visual de frecuencias mediante un Espectrograma.



En la Figura 2.9 muestra representación de la densidad espectral de potencia en función del tiempo mediante un espectrograma, obtenido como el cuadrado del módulo de la STFT.

Función de ventana y principio de incertidumbre

La elección de una ventana es fundamental en la STFT, Harris, en su trabajo seminal de 1978, realizo un análisis de las propiedades de diversas funciones de ventana para el análisis armónico con la Transformada de Fourier Discreta, lo que permite establecer las métricas de evaluación que permiten comparar su comportamiento espectral. Las ventanas que más se utilizadas son la rectangular, la de Hann, la de Hamming y la de Blackman-Harris, cada una distintos compromisos en la resolución en frecuencia y la atenuación de lóbulos laterales *(spectral leakage)* [31].

La STFT enfrenta una limitación denominada principio de incertidumbre tiempo-frecuencia: el tamaño de la ventana de análisis determina simultáneamente qué tan bien se puede localizar un evento en el tiempo y qué tan bien se puede identificar su contenido frecuencial. Una ventana corta permite detectar con precisión cuándo ocurre un evento, pero a costa de una pobre resolución espectral; por el contrario, una ventana larga revela con mayor detalle qué frecuencias están presentes, pero difumina la información temporal, esta elección es fundamental para el análisis de una señal y para la obtención de información crítica.

Redes Neuronales Convolucionales

Las Redes Neuronales Convolucionales constituyen una clase especializada de redes neuronales profundas diseñadas para procesar datos con una topología de cuadrícula conocida (grid-like topology). Como establecen Goodfellow, Bengio y Courville, las CNN son capaces de procesar datos de series temporales, que pueden concebirse como una cuadrícula unidimensional (1-D) que toma muestras a intervalos de tiempo regulares, así como datos de imagen, que pueden interpretarse como una cuadrícula bidimensional (2-D) de píxeles [32].

Operación de convolución:

La convolución es una operación sobre dos funciones de argumento real. En el contexto de las CNN, el primer argumento se denomina entrada (input), el segundo se denomina núcleo (kernel), y el resultado se denomina mapa de características (feature map).

Para el caso discreto unidimensional (1-D), la convolución se define (ver Ec.10) [32].

$$s(t) = (x * w)(t) = \sum_{a=-\infty}^{\infty} x(a)w(t-a) \ , \tag{10}$$

Donde x es la señal de entrada y w es el núcleo (kernel). Esta ecuacion expresa que la salida $s(t)$ en cada instante t es un promedio ponderado de los valores pasados de x, donde los pesos están definidos por w.

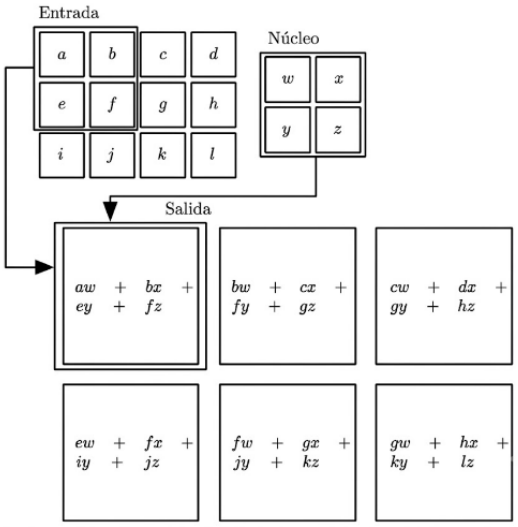
Para el caso bidimensional (2-D), que es el más relevante en procesamiento de imágenes, si se utiliza una imagen Icomo entrada y un núcleo bidimensional K, la convolución discreta se expresa como (ver Ec. 11) [32]:

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(m,n)K(i-m, j-n) \ , \tag{11}$$

Arquitectura:

La etapa de convolución (convolution stage) aplica en paralelo múltiples núcleos (kernels) sobre la entrada para producir un conjunto de mapas de características (feature maps). Cada núcleo se desliza sobre la entrada que multiplica y suma los valores de cada región local, lo que es un único número que indica qué tan presente está una característica, borde, textura o patrón, en esa posición espacial [31].

Figura 2.10
Ejemplo de convolución 2-D sin inversión del núcleo (kernel flipping), Tomado de [32].



En la Figura 2.10 se muestra la operación de convolución 2-D sin inversión del núcleo (kernel flipping), donde cada posición del kernel genera un único valor que indica la presencia de una característica espacial en la entrada.

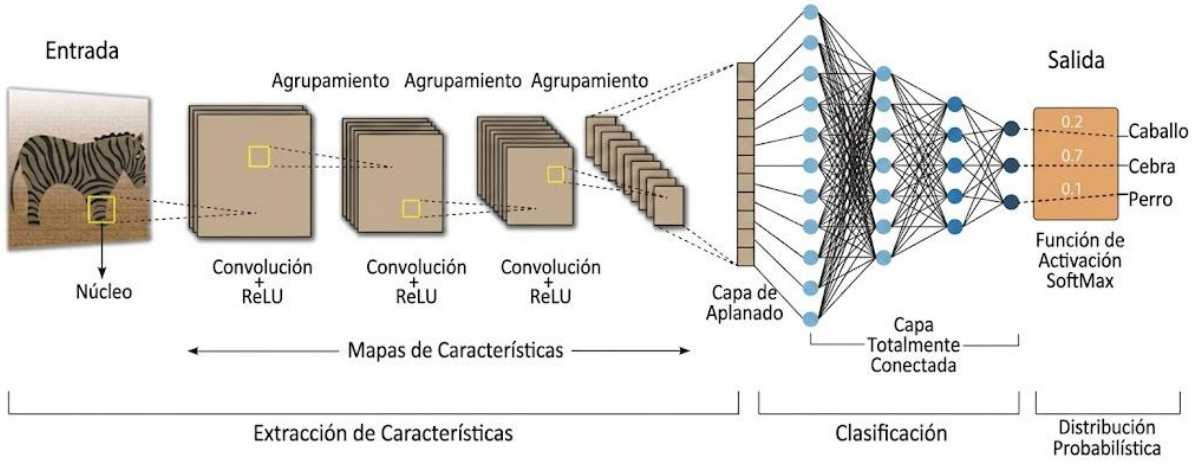
La etapa detectora (detector stage) pasa cada activación por una función de activación no lineal, habitualmente la Unidad Lineal Rectificada (ReLU, Rectified Linear Unit), que deja pasar los valores positivos y convierte a cero los negativos. Esto introduce la no-linealidad necesaria para que la red

aprenda relaciones complejas entre características y evita el problema del desvanecimiento del gradiente (vanishing gradient problem) que presentan otras funciones de activación [31].

Finalmente, la etapa de agrupamiento (pooling stage) reemplaza la salida en cada posición por un estadístico resumen de su vecindad. El más utilizado es el agrupamiento máximo (max pooling), que reporta el valor máximo dentro de una región rectangular, lo que responde a la pregunta ¿está presente esta característica en esta región? sin importar en qué píxel exacto aparece. Esto introduce invarianza aproximada ante pequeñas traslaciones de la entrada y reduce la dimensión espacial de la representación, lo que permite reducir el costo computacional de las capas siguientes [31].

El agrupamiento (pooling) puede interpretarse como la imposición de una distribución a priori (prior) infinitamente fuerte que establece que la función aprendida debe ser invariante a pequeñas traslaciones [31]. Cuando esta suposición es correcta, mejora considerablemente la eficiencia estadística de la red.

Figura 2.11
Arquetipo de una Red Neuronal Convolucional (CNN), Tomado de [33].



En la Figura 2.11 se puede observar el Arquetipo de una CNN, que muestra las etapas de convolución, activación no lineal (ReLU) y agrupamiento (pooling), seguidas por las capas totalmente conectadas para la clasificación.

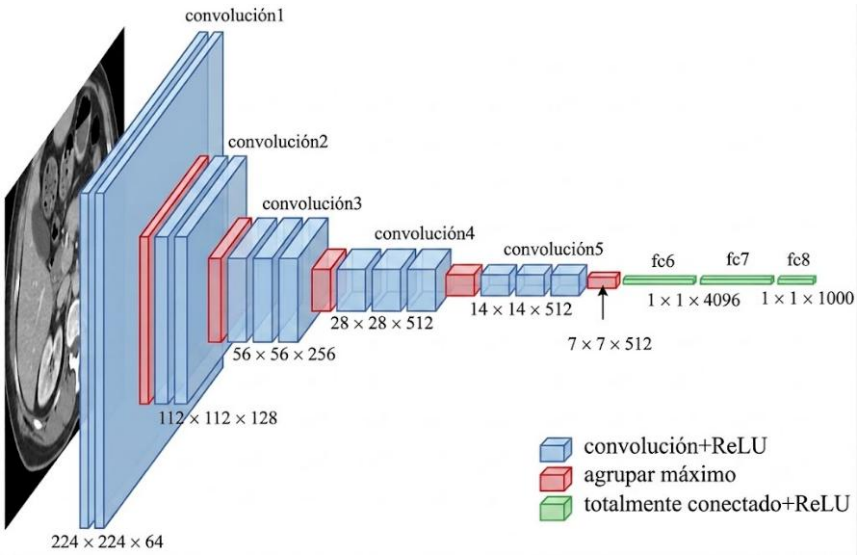
AlexNet

La arquitectura AlexNet, desarrollada por Krizhevsky, Sutskever y Hinton, marcó un punto de inflexión en el campo del reconocimiento visual al ganar el desafío ILSVRC (ImageNet Large Scale Visual Recognition Challenge) en 2012, y así reducir drásticamente la tasa de error del 26.2% al 15.3% [34].

Esta red neuronal convolucional (CNN) está compuesta por un total de ocho capas con pesos entrenables. Su distribución interna consta de cinco capas convolucionales enfocadas en la extracción secuencial de características espaciales. Las capas convolucionales primera, segunda y quinta están seguidas directamente por operaciones de agrupamiento máximo (max pooling) con filtros de 3x3 y un paso (stride) de 2 para reducir la dimensionalidad. La red finaliza su procesamiento con tres capas totalmente conectadas (fully connected layers), de las cuales las dos primeras poseen 4,095 neuronas cada una, mientras que la última se conecta a una función softmax de 1,000 clases.

En total, el modelo concentra aproximadamente 60 millones de parámetros y 650,000 neuronas [34]. Este rendimiento histórico fue posible gracias a innovaciones críticas como la introducción masiva de la función de activación ReLU (Rectified Linear Unit), el uso de la técnica de abandono (dropout) como regularización para prevenir el sobreajuste, y el entrenamiento en paralelo sobre múltiples unidades de procesamiento gráfico (GPU) [34].

Figura 2.12
Arquitetura de Alexnet. Tomado de [35].



En la Figura 2.12 se presenta la Arquitectura de AlexNet, compuesta por cinco capas convolucionales y tres capas totalmente conectadas, con aproximadamente 60 millones de parámetros entrenables.

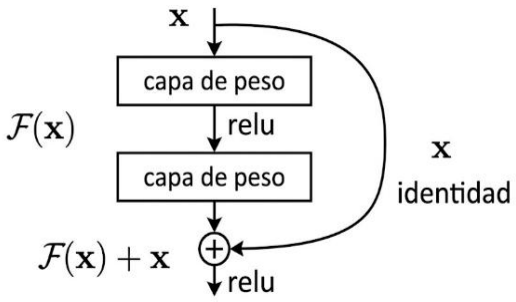
ResNet50

ResNet50 es una variante estructural de la familia Residual Network propuesta por He, Zhang, Ren y Sun, cuya contribución principal es la introducción de las conexiones residuales (skip connections o shortcut connections) para resolver el problema de degradación del gradiente (vanishing gradient problem) en redes muy profundas [25]. El número "50" hace referencia al total de capas con pesos entrenables, las cuales se organizan de forma jerárquica en cinco etapas principales. La primera etapa contiene una capa convolucional con filtros de 7x7 seguida de un max pooling de 3x3, ambas con un paso de 2. El cuerpo principal de la red está conformado por las cuatro etapas siguientes, las cuales se componen de bloques de cuello de botella (bottleneck blocks) que integran internamente capas convolucionales de tamaños 1x1, 3x3 y 1x1 [8]. La operación residual que define a esta arquitectura se expresa matemáticamente como (ver Ec.12) [36].

$$y = \mathcal{F}(x, \{W_i\}) + x \ , \tag{12}$$

donde $\mathcal{F}$ representa la transformación aprendida y x es la identidad de la entrada que se propaga directamente. ResNet50 cuenta con aproximadamente 25.5 millones de parámetros y obtuvo el primer lugar en el ILSVRC (ImageNet Large Scale Visual Recognition Challenge) 2015 con una tasa de error del 3.57% [36].

Figura 2.13
Diagrama de bloque residual (residual block) con la conexión de salto (skip connection). Tomado de [36].



En la Figura 2.13 se muestra el Diagrama del bloque residual (residual block) con su conexión de salto (skip connection), en el que la transformación aprendida F(x, W) se suma directamente a la identidad de la entrada x para mitigar el problema de degradación del gradiente en redes profundas.

MobileNetV2

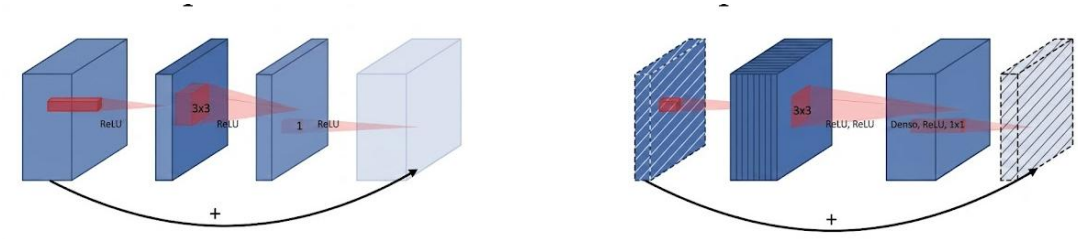
Diseñada bajo una estricta filosofía de optimización, MobileNetV2 es una arquitectura CNN creada para operar eficientemente en dispositivos móviles y sistemas embebidos (embedded systems), propuesta por Sandler et al. [37]. Su diseño se fundamenta en la estructura residual invertida (inverted residual structure), una innovación donde la entrada y la salida del bloque residual son capas de cuello de botella delgadas, y de esta forma operar de manera opuesta a los modelos residuales tradicionales que utilizan representaciones expandidas en sus extremos [37]. La arquitectura inicia con una capa convolucional estándar, seguida de una secuencia de 19 módulos de cuello de botella residual.

Para lograr su ligereza, emplea convoluciones separables en profundidad (depthwise separable convolutions) que actúan como filtros eficientes en la capa de expansión intermedia. Adicionalmente, el modelo implementa cuellos de botella lineales, el las activaciones no lineales en las capas estrechas para mantener la capacidad representacional del sistema [37]. Toda la arquitectura utiliza la función de

activación acotada ReLU6 (Rectified Linear Unit 6), así una red extremadamente compacta que oscila entre 3.4 y 4.2 millones de parámetros sin comprometer la precisión.

Figura 2.14
Comparación entre (2.14a) el bloque residual clásico y (2.14b) el bloque residual invertido (inverted residual block) de MobileNetV2. Tomado de [37].

(2.14a) (2.14b)



En la Figura (2.14a) se puede observar el diagrama comparativo entre el bloque residual clásico y en la Figura (2.14b) el bloque residual invertido (inverted residual block) de MobileNetV2, donde la entrada y salida son representaciones compactas (bottleneck) y la expansión de dimensionalidad ocurre únicamente en la capa intermedia.

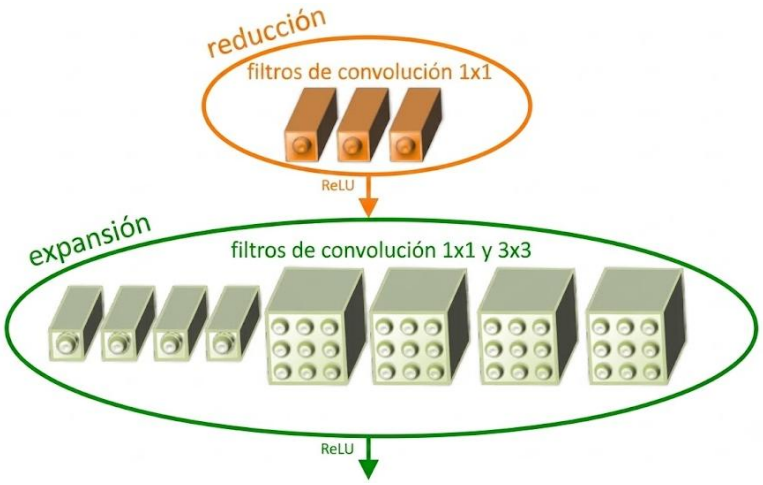
SqueezeNet

SqueezeNet es una arquitectura CNN propuesta por Iandola et al. con el objetivo primordial de lograr una precisión equivalente a la de AlexNet en la base de datos ImageNet, pero con el uso de 50 veces menos parámetros [38]. La motivación principal de este enfoque radica en que los modelos más pequeños ofrecen ventajas sustanciales: requieren una menor comunicación entre servidores durante el entrenamiento distribuido (distributed training), exigen un menor ancho de banda para su despliegue en la nube, y garantizan una mayor viabilidad para su implementación en hardware con memoria limitada, como los dispositivos FPGA (Field-Programmable Gate Arrays) [38].

La característica arquitectónica central de esta red es el módulo Fire (Fire module). Este componente está estructurado en dos fases: una capa squeeze provista de filtros convolucionales de 1x1 que comprime la dimensionalidad del tensor de entrada, seguida inmediatamente por una capa expand que aplica en paralelo filtros de 1x1 y 3x3 para recuperar y expandir la representación espacial. Este diseño estratégico permite reducir drásticamente el número de parámetros sin sacrificar la capacidad de

extracción de características [38]. Al complementar esta estructura con la eliminación total de las capas totalmente conectadas, SqueezeNet base opera con tan solo 1.24 millones de parámetros. Mediante la aplicación de técnicas avanzadas de compresión de modelos (deep compression), esta arquitectura puede ser comprimida a menos de 0.5 MB, volviéndose 510 veces más pequeña que su contraparte AlexNet [38].

Figura 2.15
Diagrama del módulo Fire donde muestra la capa squeeze y la capa expand. Tomado de [38].



En la Figura 2.15 se puede visualizar el diagrama del módulo Fire de SqueezeNet, que muestra la capa squeeze con filtros 1×1 para comprimir la dimensionalidad y la capa expand con filtros 1×1 y 3×3 en paralelo para recuperar la representación espacial, de esta forma se reduce drásticamente el número de parámetros.

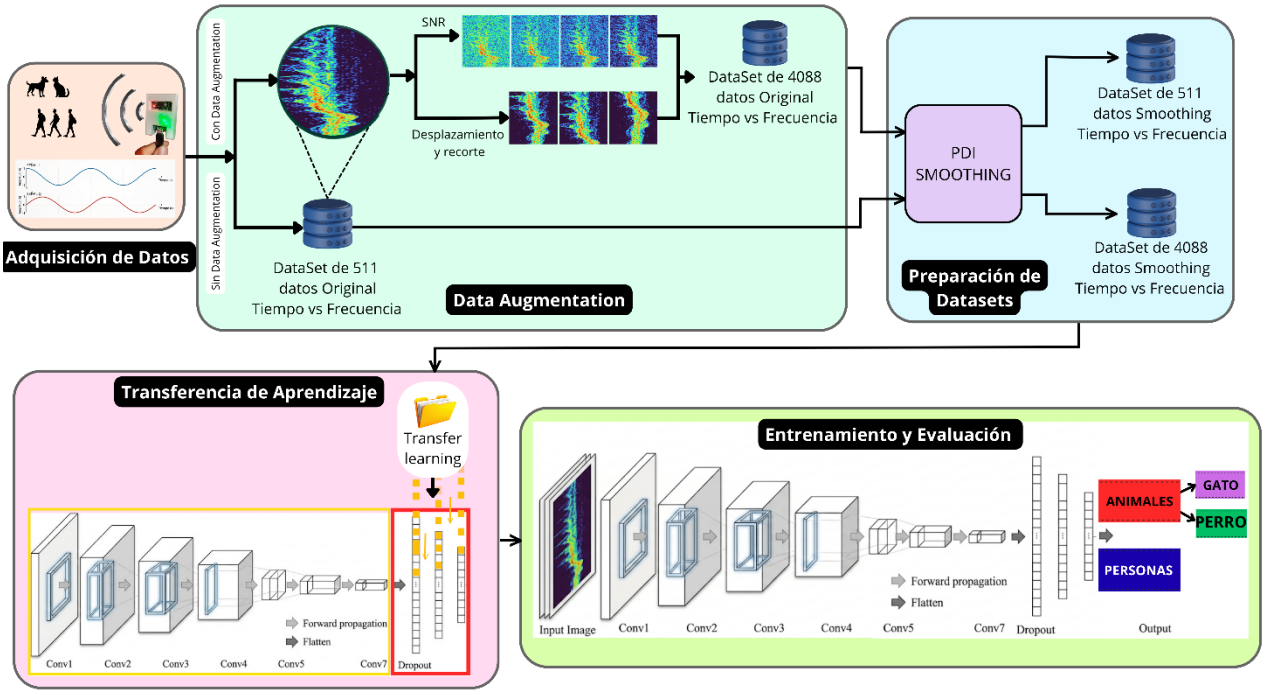
Capítulo 3

Metodología

Para cumplir con el objetivo general de desarrollas un modelo de aprendizaje de máquina capaz de distinguir con agilidad entre personas y animales en entornos de interiores, esta sección presenta la metodología propuesta para la generación del modelo que tenga la capacidad de reconocer con eficacia entre animales y personas en entonos de interiores.

La metodología se organiza en las siguientes etapas: adquisición de datos, data augmentation, preparación de datasets, transferencia de aprendizaje, entrenamiento y evaluación, como se puede observar en la Figura 3.1.

Figura 3.1
Diagrama de Bloques de procesos- Metodología



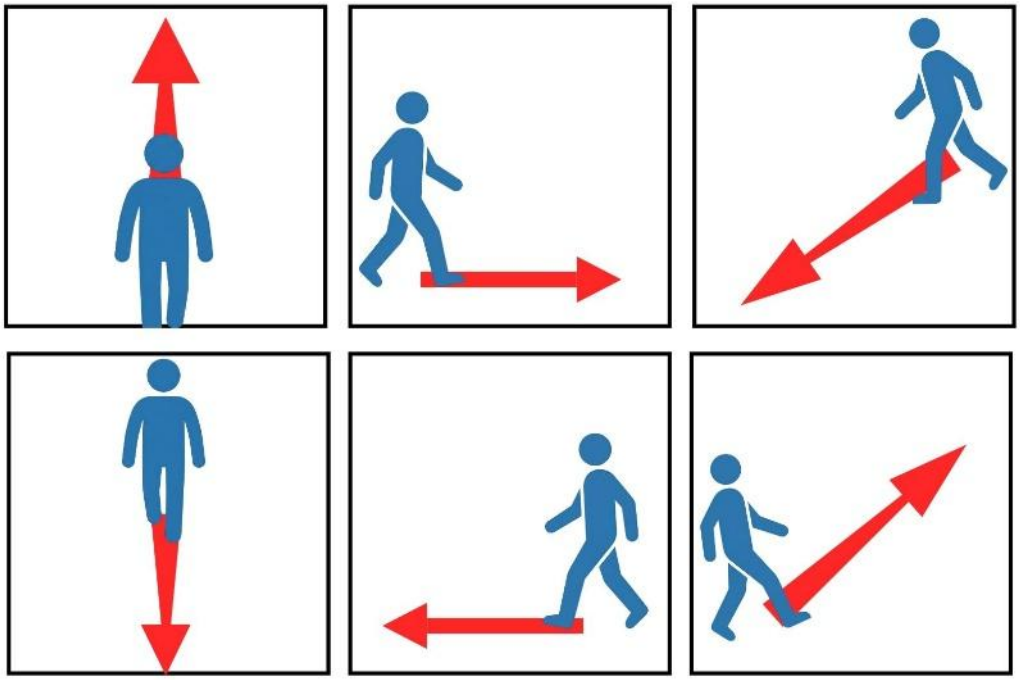
Adquisición de datos

La adquisición de datos se llevó a cabo en diversos entornos de interiores reales, esta metodología descarta los escenarios ideales con el fin de registrar los fenómenos de propagación típicos en los hogares, como por ejemplo el clutter estático y los efectos de multitrayectoria. Dentro de estos espacios los sujetos

ejecutaron trayectorias predefinidas que se mantuvieron estrictamente dentro del rango efectivo de detección del radar.

Se realizo la recolección de datos mediante el radar BGT60LTR11AIP de diferentes objetivos de animales domésticos más comunes como diversas razas y tamaños de perros y gatos, así como de distintas personas como niño, adultos, abuelos, todas estas muestras recabadas desde distintos puntos de visión, como se muestra en la Figura 3.2.

Figura 3.2
Trayectoria de objetivos



Cabe señalar que, debido a la dificultad inherente que se presentó al momento de tratar con animales domésticos en entornos de interiores, no fue posible replicar dichas trayectorias con exactitud; no obstante, se procuró que los recorridos realizados fueran lo más similares posible a las rutas establecidas, para preservar la intención de diversidad angular en la recolección de datos.

Se capturaron las firmas micro doppler de los objetivos de estudio los cuales comprenden dos categorías: personas y animales domésticos. En el caso de las personas se incluye sujetos de distintas franjas generacionales, como niños, adultos y adultos mayores. Para los animales domésticos se seleccionaron perros y gatos de diversas razas y tamaños para garantizar la diversidad de muestras, todas

las muestras fueron recabadas desde múltiples ángulos de observación con el fin de incorporar variabilidad en la geometría de adquisición es decir, en la orientación y posición relativa de los objetivos respecto al radar, lo que contribuye a reducir el sobreajuste y mejorar la capacidad de generalización del algoritmo entrenado.

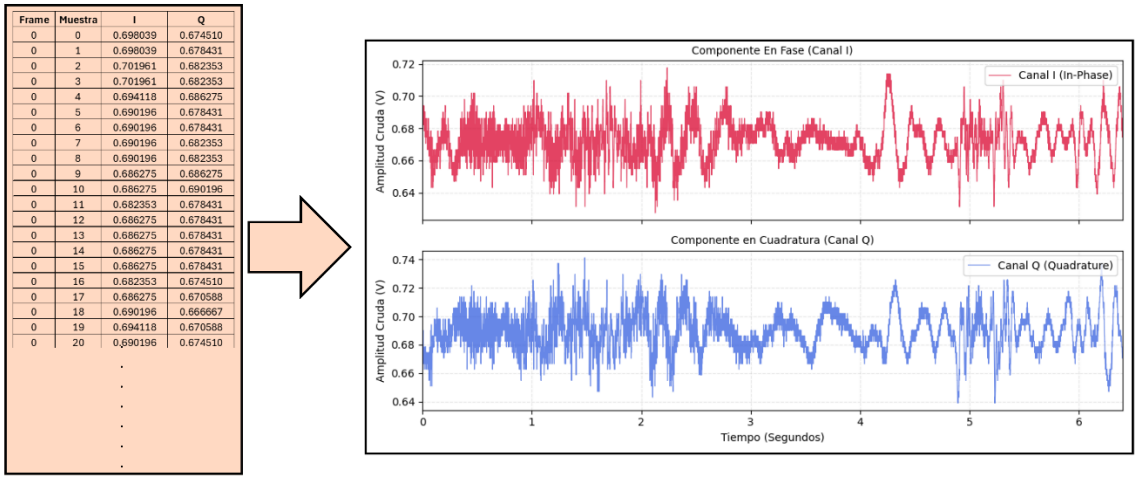
La señal crudos recabada por el radar se encuentran en fase y cuadratura. Cada registro tuvo una duración de aproximadamente 6 segundos para garantizar la captura completa de la dinámica de movimientos del individuo. Este intervalo corresponde a 50 frames, obtenido mediante la expresión (ver Ec.13):

$$T_g(s) = \frac{N_s}{F_s}\ x\ N_f\ , \tag{13}$$

donde,

- $N_s = \ Número\ de\ muestras\ por\ frame$: 256 muestras
- $F_s = \ Frecuencia\ de\ muestreo\ (Hz)$: 2000 Hz
- $N_f = Número\ de\ frames$: 50

Figura 3.3
Señal cruda recolectada



Frame	Muestra	I	Q
0	0	0.698039	0.674510
0	1	0.698039	0.678431
0	2	0.701961	0.682353
0	3	0.701961	0.682353
0	4	0.694118	0.686275
0	5	0.690196	0.678431
0	6	0.690196	0.678431
0	7	0.690196	0.682353
0	8	0.690196	0.682353
0	9	0.686275	0.686275
0	10	0.686275	0.690196
0	11	0.682353	0.678431
0	12	0.686275	0.678431
0	13	0.686275	0.678431
0	14	0.686275	0.678431
0	15	0.686275	0.678431
0	16	0.682353	0.674510
0	17	0.686275	0.670588
0	18	0.690196	0.666667
0	19	0.694118	0.670588
0	20	0.690196	0.674510

En la Figura 3.3 se presenta el formato de fase y cuadratura en el que se recolectan las muestras.

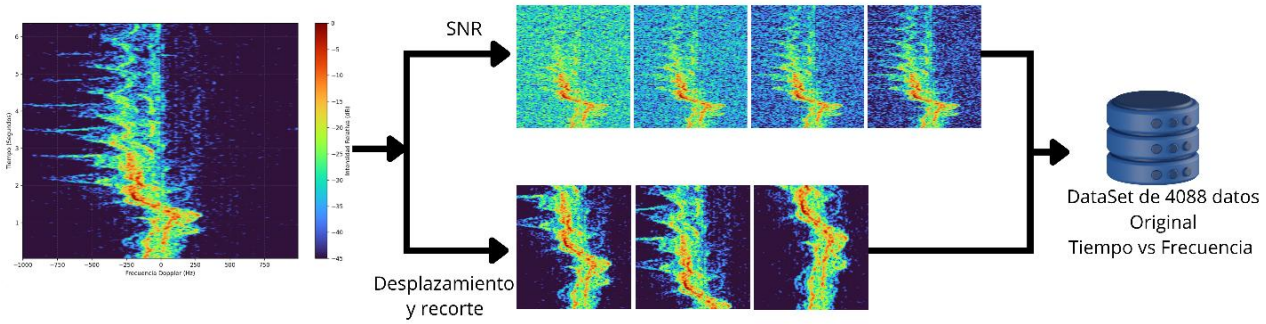
Data augmentation

Una vez recabadas las 511 muestras de en crudo, se procedió a expandir el conjunto de datos mediante las dos técnicas de data augmentation, esta estrategia busca incrementar artificialmente el volumen de muestras, lo cual es fundamental para mejorar la capacidad de generalización de las CNN, mitigar el sobreajuste (overfitting) y optimizar su precisión predictiva. El primer método consistió en degradar la señal original, donde se agregó diferentes niveles de señal a ruido (SNR, del inglés Signal to Noise Ratio) específicamente de 2, 6,10 y 14 dBm para generar cuatro variantes sintéticas.

El segundo método se basó en un recorte dinámico basado en localizar el segmento de máxima activad, es decir, máxima concentración de energía en la señal, se aisló en una ventana de 3 segundos y se generó variaciones mediante desplazamiento temporales (time shifting) de 1 segundo hacia adelante y hacia atrás, de esta forma se consiguió tres variantes sintéticas.

Después de obtener 4088 muestras, se procede a generar a obtener el dataset de imágenes 2D de representación tiempo versus frecuencia mediante el uso de la STFT.

Figura 3.4
Proceso de Data Augmentation



En la Figura 3.4 se observa el proceso que se siguió para aumentar los datos del dataset mediante incremento de SNR y mediante el uso de localización de mayor concentración de energía combinado con corte de 3 segundos y desplazamiento de 1 segundo hacia adelante y 1 segundo hacia atrás.

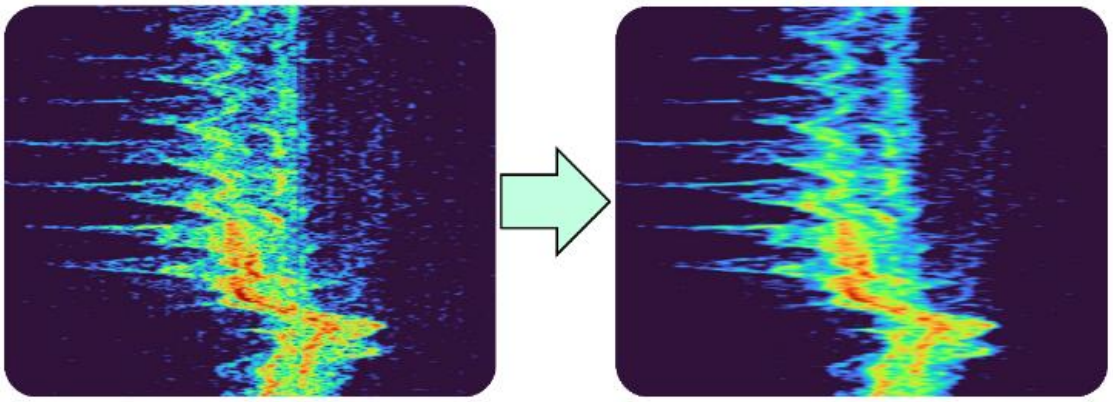
Preparación de datasets

Una vez generadas las representaciones tiempo versus frecuencia mediante la STFT, los espectrogramas resultantes se trataron como imágenes 2D para aplicar técnicas de procesamiento digital

de imágenes (PDI). De manera específica, se incorporó un filtro de suavizado espacial (smoothing) dentro del proceso.

Esta etapa es fundamental ya que la STFT debido a su propia naturaleza matemática genera varianza espectral y ruido de fondo capaces de fragmentar visualmente las firmas micro doppler. La aplicación de smoothing permite atenuar estas fluctuaciones de alta frecuencia y se suprime el ruido disperso y de esta forma mejorar la continuidad de las crestas de frecuencia (que representan el movimiento de las extremidades). Para el posterior entrenamiento de las CNNs, este preprocesamiento es crucial, ya que evita que el modelo sufra de sobreajuste (overfitting) al aprender patrones de ruido, obligándolo a centrarse exclusivamente en la topología y la forma real de la firma dinámica del objetivo.

Figura 3.5
Smoothing de una muestra original



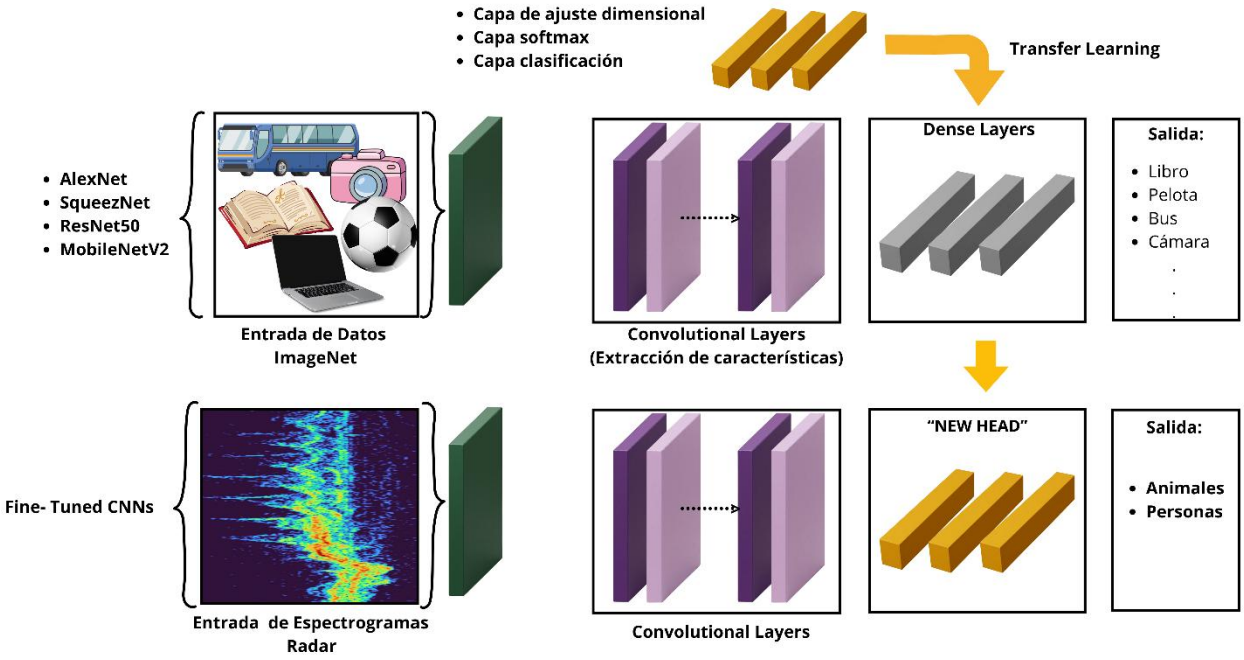
La Figura 3.5 muestra un ejemplo en al cual se verifica como una señal original al aplicarle smoothing permite atenuar clutter disperso y resaltar la señal original.

Transferencia de aprendizaje

Dado que el entrenamiento de CNNs desde cero, demanda un alto costo computacional y base de datos masivas, su implementación resulta inviable en escenarios con recurso limitados. Para mitigar esta limitación, se implementó la técnica de aprendizaje por transferencia (transfer learning), esta estrategia reutiliza los pesos y características extraídas por modelos preentrenados en aplicaciones de gran escala.

Esto permite realizar una adaptación directa de forma eficiente y con menor costo computacional, para la tarea de clasificación de objetivos.

Figura 3.6
Proceso de Transfer Learning



En la Figura 3.6 se visualiza como es el proceso de transfer learning con el uso CNNs preentrenadas para el propósito de este estudio, cabe destacar que el tamaño del bloque verde de entrada corresponde a una dimensión de 227x227x3 píxeles para AlexNet y SqueezeNet, y de 224x224x3 píxeles para ResNet50 y MobileNetV2.

Para comprender el proceso de adaptación, una CNN rentrenada se divide conceptualmente en dos bloques funcionales. El bloque extractor de características (capas convolucionales y pooling), actúa como filtros matemáticos que permite detecta patrones visuales jerárquicos, desde bordes hasta geometrías complejas. Este bloque se mantuvo intacto, pues su capacidad de abstracción es aplicable al análisis de espectrogramas micro doppler.

El bloque de clasificación (capas fully connected y Softmax), originalmente diseñado para discriminar las 1.000 clases o superiores, fue reemplazado por un nuevo conjunto de 3 capas, adaptada a

las dos categorías del estudio: personas y animales, con el fin de permitir la transferencia eficiente del conocimiento preentrenado en la aplicación requerida para este trabajo.

La adaptación de las CNNs (AlexNet, Res Net50, SqueezeNet y MobilNetV2) consistió en sustituir las tres últimas capas originales dedicadas a la clasificación por una nueva estructura de tres capas orientada a las dos clases de este dataset.

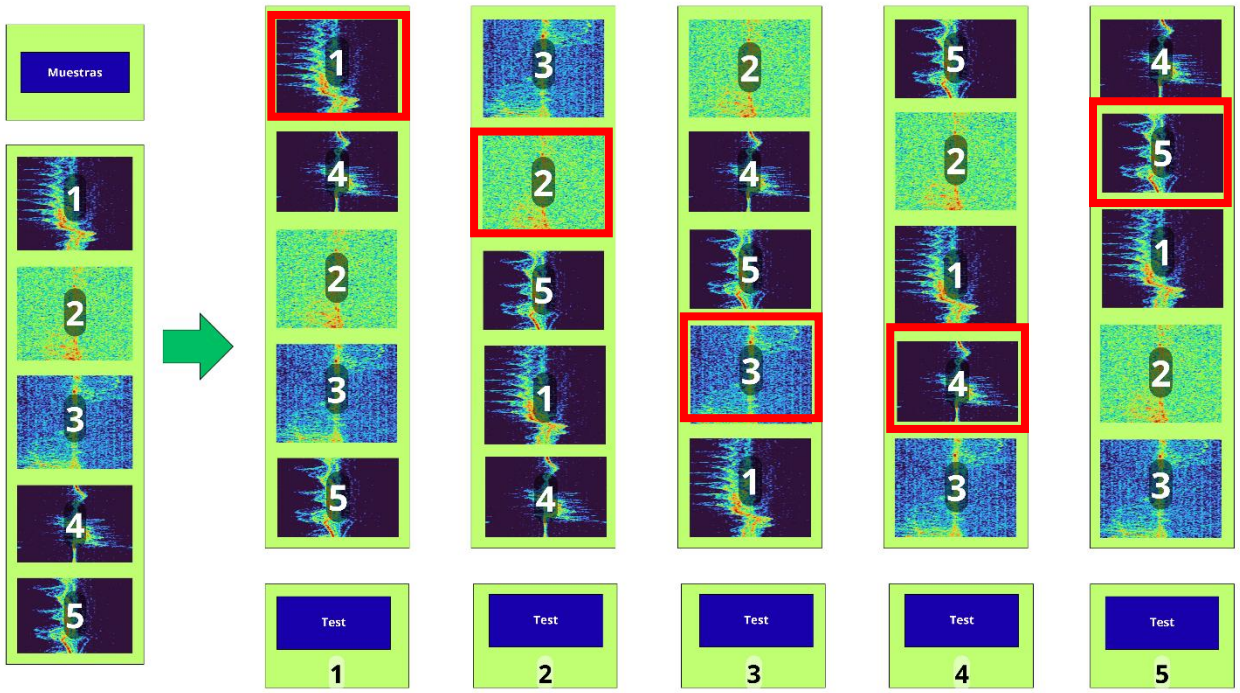
La primera capa es de ajuste dimensional, se reemplaza la salida de 1000 clases, con el uso de capas densamente conectadas (fully connected) para AlexNet, Res Net50 y MobileNetV2 y una convolución 2D de 1X1 para SquezeNet. La segunda es la capa de Softmax, se entrega para transformar las puntuaciones en una distribución de probabilidad normalizada, y la tercera capa de salida permite calcular la perdida de entropía cruzada, es decir mede qué tan lejos está la predicción que hace tu red neuronal de la realidad y determinar la clase con mayor probabilidad.

Entrenamiento y evaluación

Una vez completado el proceso de transferencia de aprendizaje en las CNNs seleccionadas, se procedió con la fase de entrenamiento adaptativo, con el uso de un dataset propio. Durante esta etapa, uno de los desafíos más críticos en el entrenamiento de CNNs es el sobreajuste, fenómeno que se presenta cuando el modelo asimila en exceso las particularidades de los datos de entrenamiento, como por ejemplo el ruido y las particularidades específicas de ese conjunto, pero no logra generalizar a nuevos datos no vistos.

Para garantizar una evaluación estadísticamente robusta del rendimiento de cada arquitectura, se implementó la técnica de validación cruzada de K pliegues (K Fold Cross Validation) con K=5.

Figura 3.7
Proceso de 5 K folds



En la Figura 3.7 se muestra el proceso que consistió en 5 iteraciones independientes por arquitectura, donde en cada ciclo un pliegue diferente se reservó para validación y los restantes para entrenamiento. Durante las 15 épocas, se monitoreó el desempeño cada 10 iteraciones para evaluar la generalización del modelo. Al finalizar, las predicciones de los 5 K folds se consolidaron en una matriz de confusión global, con el fin de calcular métricas agregadas: accuracy promedio, recall y F1-score. Este enfoque eliminó la dependencia de particiones sesgadas y proporcionó una estimación objetiva del rendimiento esperado de cada CNN en entornos reales de radar.

Configuración experimental

Experimentos 1-Alcance de radar:

Previo a la recolección de los datos, se realizó una estimación del alcance real que del radar BGT60LTR11AIP; donde se utilizó una esfera metálica de gran tamaño recubierta de aluminio, material que favorece una alta reflexión de la señal en su superficie, de esta forma se facilita la identificación de las señales captadas por el sensor. Una persona se desplaza con la esfera con trayectoria de frente,

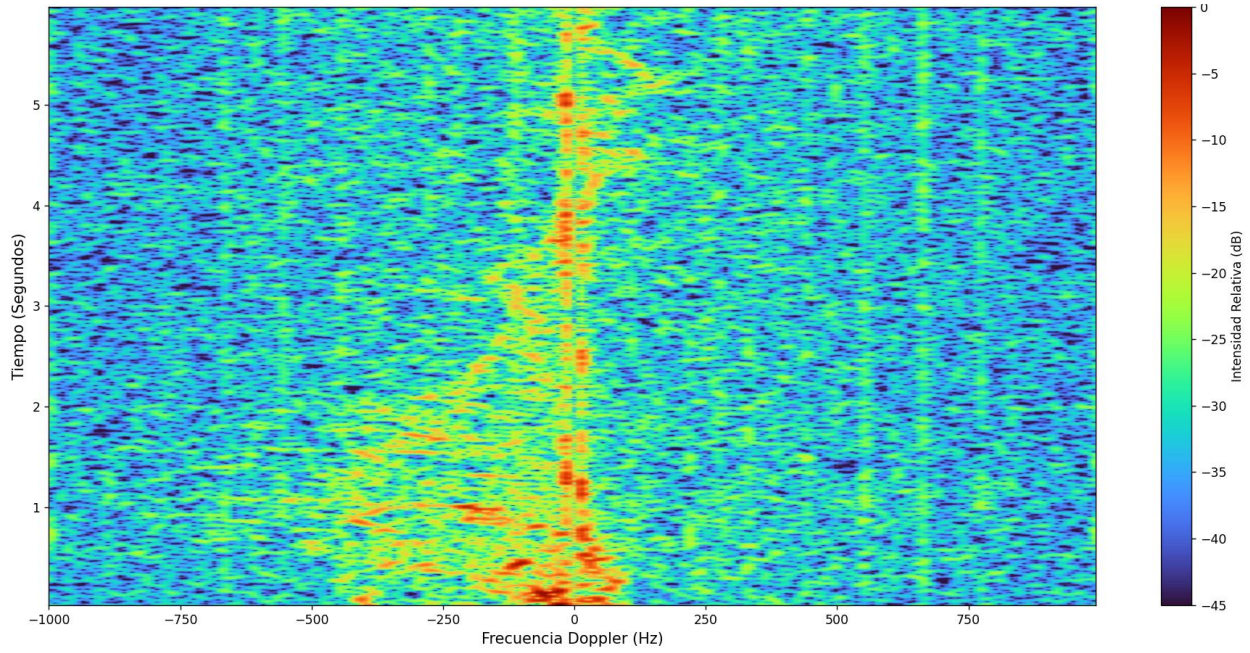
diagonal, horizontal. A partir de este procedimiento, se determinó que la distancia máxima de detección confiable es de 10 metros.

Experimento 2-Pruebas de espectrogramas:

Previo a la definición del flujo de procesamiento, se evaluaron comparativamente tres representaciones tiempo - frecuencia ampliamente utilizadas en el análisis de firmas micro doppler: la STFT, la Transformada Wavelet Continua (CWT, del inglés Continuous Wavelet Transform) y la Transformada Adaptativa de Superlets (ASLT, del inglés Adaptive Superlets). El objetivo de esta evaluación fue determinar cuál representación preserva con mayor fidelidad las características discriminativas de las firmas micro doppler de los objetivos de estudio, criterio fundamental para el posterior entrenamiento de las arquitecturas CNN.

Transformada de Fourier de tiempo corto (STFT):

Figura 3.8
Espectro de Potencia de una muestra del dataset con STFT

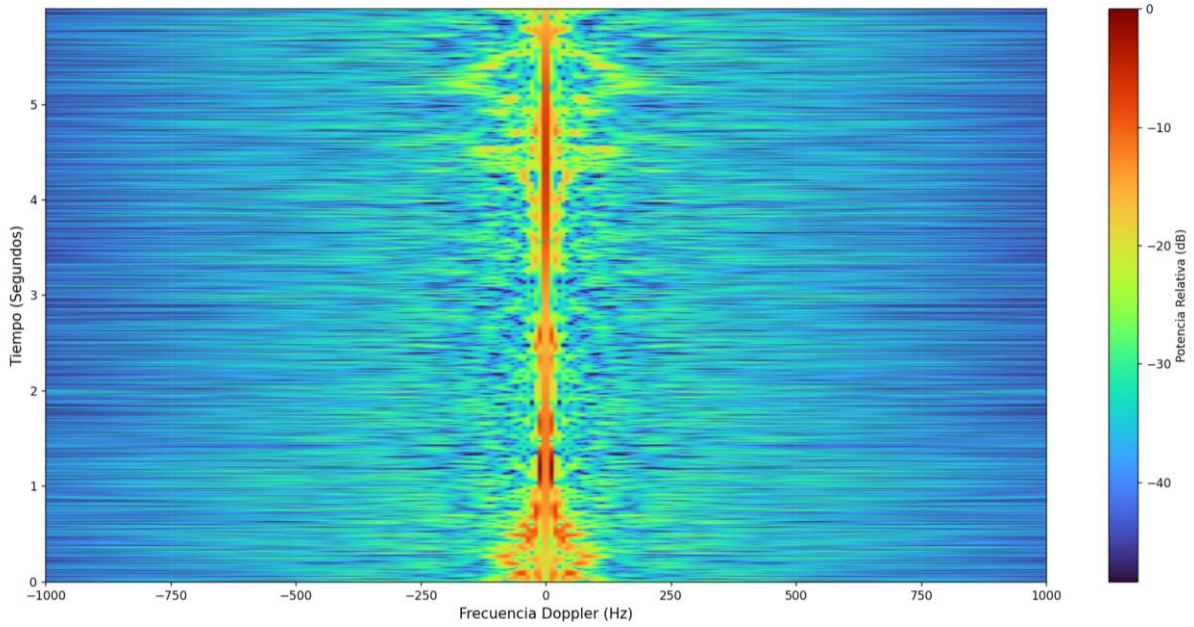


La Figura 3.8 se presenta el espectrograma obtenido mediante STFT sobre una muestra representativa del dataset, en la cual se observar acumulación de máxima energía que sigue un patrón el cual es la componente doppler central correspondiente al desplazamiento del torso del individuo de

estudio, así como las bandas laterales simétricas generadas por el movimiento oscilatorio de las extremidades, y por otro lado zonas donde la energía es muy baja que indica es el ruido de fondo. Esta separación visual entre la firma del objetivo de estudio y el ruido de fondo permite identificar con la estructura dinámica del objetivo, característica esencial para que las CNNs puedan extraer rasgos discriminativos de forma eficiente.

Transformada Wavelet Continua (CWT):

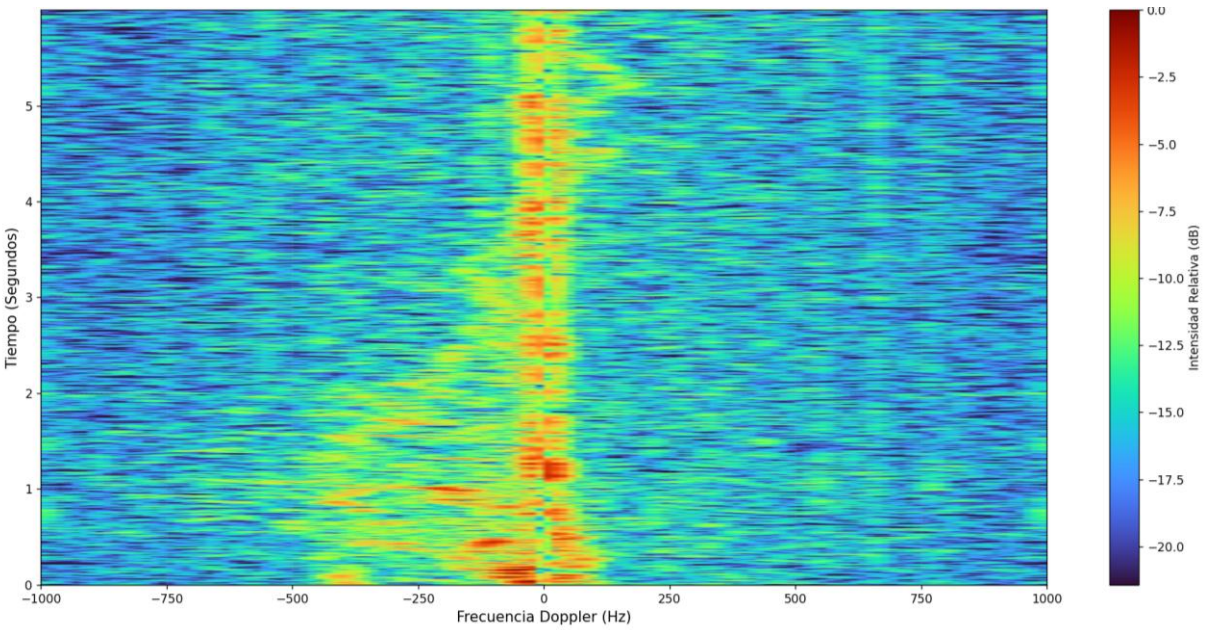
Figura 3.9
Espectrograma CWT - Transformada Wavelet Continua



La Figura 3.9 muestra la representación obtenida mediante CWT. Si bien la firma micro doppler del objetivo permanece visible en torno a la frecuencia central, el espectrograma presenta un fondo densamente poblado de franjas horizontales de alta energía que se estiran y se extienden a lo largo de todo el eje temporal, visualmente con la firma del objetivo y dificulta la separación entre señal útil, lo que dificulta la capacidad de generalización de los clasificadores entrenados sobre estas imágenes.

Transformada Adaptativa de Superlets (ASLT):

Figura 3.10
Transformada Adaptativa de Superlets ASLT Espectrograma



La Figura 3.10 corresponde a la representación generada mediante ASLT. En comparación con la CWT, el fondo presenta menor densidad de componentes espectrales, donde presenta la firma micro doppler que aparece ensanchada y con menor resolución frecuencial, lo que diluye los componentes individuales de las extremidades y reduce la separabilidad entre las crestas de energía. Esta pérdida de definición en la estructura que presenta mayor acumulación de energía, compromete la extracción de características por parte de las capas convolucionales.

Experimento 3-Entrenamiento de CNNs con personas y animales:

Con base en el dataset y el flujo de preprocesamiento detallados en la Figura 3.1 se diseñó un protocolo experimental para evaluar cuatro arquitecturas de redes neuronales convolucionales (AlexNet, SqueezeNet, MobileNetV2 y ResNet50) en una tarea de clasificación binaria de personas y animales en interiores. El estudio se dividió en dos etapas, primero se realizó un análisis comparativo para seleccionar el preprocesamiento óptimo, con la evaluación de tres configuraciones: datos originales, aplicación de smoothing mediante PDI, y data augmentation, lo que permitió expandir el conjunto inicial de 511 a 4088

muestras. El análisis empírico reveló que smoothing no aportaba mejoras discriminativas para estas dos clases por el contrario, la combinación de los datos originales sin smoothing con data augmentation maximizó la precisión del modelo.

En consecuencia, se descartó el procesamiento adicional por PDI y en una segunda etapa, se entrenaron las cuatro arquitecturas bajo esta configuración óptima. Para garantizar la robustez y fiabilidad de las métricas, todo el proceso se ejecutó mediante validación cruzada de 5 k folds.

Experimento 4-Entrenamiento de CNNs con personas, gatos y perros:

A partir de los resultados del escenario binario, se extendió la evaluación a un problema de clasificación de tres clases: persona, perro y gato. Cabe destacar que, si bien el dataset original sin smoothing fue recolectado con un número equilibrado de muestras para estas tres categorías, el experimento principal las simplificó a dos (personas y animales). Al revertir esta fusión, el presente escenario introduce una complejidad adicional, ya que las firmas micro doppler de perros y gatos presentan una alta similitud morfológica. Para abordar esta clasificación se compararon dos configuraciones de preprocesamiento: datos originales sin smoothing con data augmentation, y smoothing con data augmentation. A diferencia del escenario binario, los resultados empíricos demostraron que la aplicación de smoothing junto con data augmentation arrojó la mayor precisión, ya que el filtrado visual facilitó la discriminación de las sutiles diferencias entre las especies animales. Las cuatro arquitecturas CNN fueron evaluadas bajo esta configuración óptima, junto con la validación cruzada de 5 k folds para garantizar la robustez estadística.

Experimento 5-Analisis estadístico de CNNs con Wilkinson:

Con el fin de determinar si las diferencias de desempeño observadas entre las arquitecturas son estadísticamente significativas y no producto de la variabilidad aleatoria entre folds, se aplicó el test de Wilcoxon. Dado que esta prueba no paramétrica trabaja directamente sobre las diferencias pareadas entre los resultados de cada fold sin asumir una distribución normal de los datos resulta idónea para comparar

clasificadores evaluados bajo validación cruzada. AlexNet se estableció como modelo de referencia al haber obtenido el mejor desempeño en ambos escenarios efectuándose comparaciones independientes para los escenarios de dos y tres clases con un nivel de significancia $\alpha = 0,05$.

Capítulo 4

Resultados

En esta sección se presenta los resultados obtenidos a partir de las diferentes pruebas realizadas durante el desarrollo de la investigación. En primer lugar, se evalúan distintas técnicas de representación tiempo frecuencia con el fin de determinar cuál de ellas preserva de mejor manera las firmas micro doppler. Posteriormente, se analiza el impacto de las técnicas de data augmentation y con smoothing en el proceso de entrenamiento de las arquitecturas CNN. Finalmente, se compara el desempeño de las diferentes redes convolucionales mediante diversas métricas de evaluación y se selecciona la arquitectura con mejor rendimiento a partir de un análisis estadístico.

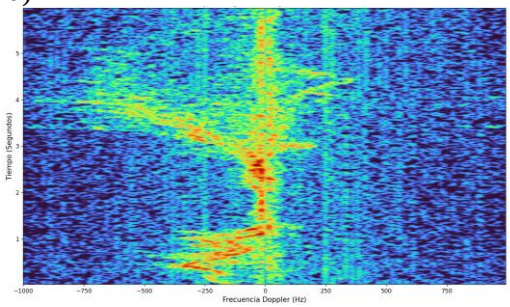
Representación Tiempo Frecuencia

Se compararon las tres técnicas de representación tiempo frecuencia, con el objetivo de identificar la representación que mejor conserva las firmas micro doppler, para generar el conjunto de imágenes, empleado para el entrenamiento de las arquitecturas CNN.

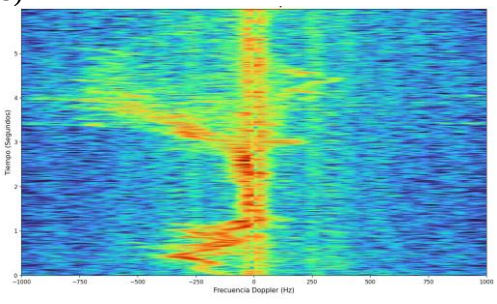
Figura 4.1

Representaciones tiempo frecuencia, Figura (4.1a). STFT, Figura (4.1b). Transformada Wavelet Continua (CWT) y en Figura (4.1c). Transformada Adaptativa de Superlets (ASLT)

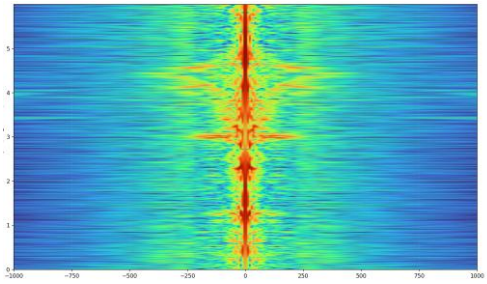
(4.1a)



(4.1b)



(4.1c)



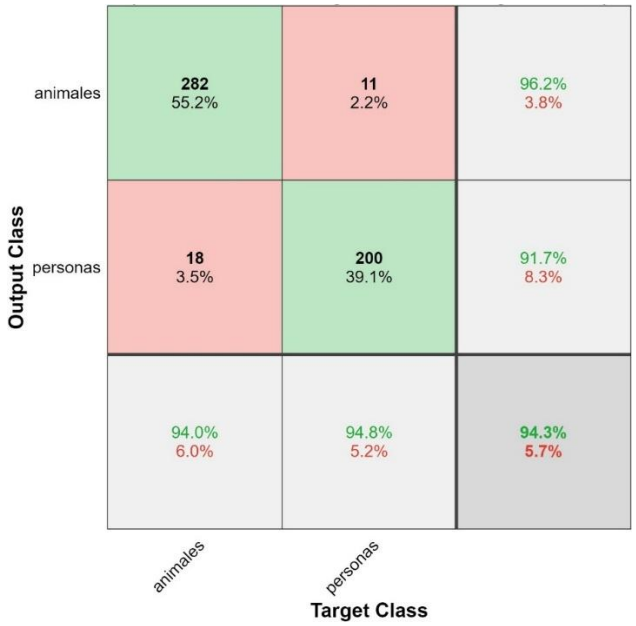
La Figura (4.1a) presenta el espectrograma generado mediante la STFT, donde se observa una representación conjunta en tiempo y frecuencia de las firmas micro doppler. La Figura (4.1b) muestra el resultado obtenido mediante la Transformada Wavelet Continua, la cual ofrece una representación multiresolución de la señal. Finalmente, la Figura (4.1c) corresponde a la Transformada Adaptativa de Superlets que proporciona una mayor resolución espectral, sin embargo, visualmente presenta una menor diferenciación de algunas características relevantes del movimiento.

A partir de la evaluación visual de las tres representaciones tiempo frecuencia mostradas en la Figura 4.1, se determinó que la STFT proporciona una representación más adecuada de las firmas micro doppler, al conservar una mayor cantidad de características discriminativas de los objetivos analizados. Por esta razón, la STFT fue seleccionada como el método de representación para generar el dataset de imágenes empleado en el entrenamiento y evaluación de las cuatro arquitecturas CNN desarrolladas en este trabajo.

Entrenamiento Inicial con AlexNet sobre dataset sin smoothing

Una vez seleccionada la STFT como método de representación tiempo frecuencia, se procedió a realizar un entrenamiento inicial con la arquitectura AlexNet sobre el dataset sin smoothing, sin aplicación de data augmentation. El objetivo de esta etapa fue obtener una línea base de desempeño que permita evaluar, en etapas posteriores, el impacto de las técnicas de preprocesamiento implementadas.

Figura 4.2
Matriz de Confusión Global de AlexNet para dos clases, sin smoothing, sin data augmentation



La matriz de confusión global, presentada en la Figura 4.2 muestra que de 300 muestras de animales 282 fueron correctamente clasificadas, y de 211 muestras de personas 200 fueron correctamente identificadas.

El modelo fue evaluado mediante validación cruzada de 5 folds. Los resultados se resumen en la Tabla 4.1

Tabla 4.1
Métricas de clasificación de AlexNet, sin smoothing sin data augmentation

Arquitectura	**Recall** (σ)	**F1-Score** (σ)	**Accuracy global** (σ)
Alexnet	94,4%$\pm$ 2.68%	94,19%$\pm$ 3,28%	94.32% $\pm$ 3,28%

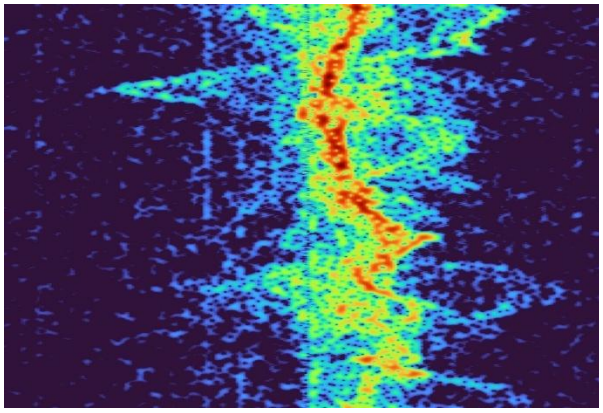
Los resultados obtenidos se resumen en la Tabla 4.1. La exactitud global promedio alcanzada fue de 94,32% $\pm$ 3,28%, mientras que el recall fue superior en personas (94,4%) respecto a animales (94,00%) indicando que el modelo detecta un poco mejor a las personas que a los animales.

Evaluación del efecto del suavizado (Smoothing) con Alexnet

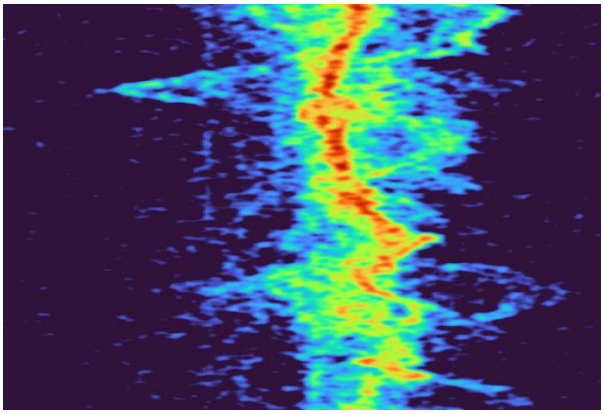
A partir del resultado de línea base obtenido con el dataset original, se procedió a evaluar el impacto de smoothing sobre el desempeño de AlexNet, con las mismas condiciones de entrenamiento: sin data augmentation y con validación cruzada de 5 folds. Las Figuras 4.3a y 4.3b ilustran el efecto visual de smoothing sobre un espectrograma representativo del dataset, donde se aprecia la reducción del ruido disperso y la mejora en la continuidad de la firma micro doppler.

Figura 4.3
Efecto visual de smoothing sobre un espectrograma

(4.3a) Espectrograma sin smoothing



(4.3b) Espectrograma con smoothing



Los resultados de esta evaluación se presentan en la Tabla 2. La exactitud global promedio fue de 94,13% ± 3,11%, con un F1-score de 94,91% para animales y 93,05% para personas.

Tabla 4.2
Métricas de clasificación con AlexNet, con smoothing sin data augmentation.

Arquitectura	**Recall (σ)**	**F1-Score (σ)**	**Accuracy global (σ)**
Alexnet	94,29%± 2,98%	93,99%± 3,16%	94,13% ± 3,11%

La Tabla 4.2 muestra el efecto de aplicar la técnica de smoothing sobre los espectrogramas con l de AlexNet. La exactitud promedio obtenida fue de 94,13% ± 3,11% valor muy similar al alcanzado con el dataset sin smoothing. De igual forma los valores de recall se mantuvieron estables en ambas clases, lo

cual muestra que el suavizado de las imágenes no produjo una mejora significativa en la capacidad del modelo para detectar correctamente las muestras reales. Del mismo modo el F1-score presentó variaciones mínimas respecto al experimento anterior o que indica que el smoothing por sí solo no incrementa de forma considerable el desempeño del clasificador.

AlexNet con Data Augmentation sin smoothing

Tras aplicar las técnicas de data augmentation descritas en la metodología, el dataset sin smoothing se expandió a 4088 muestras, con las cuales se procedió al entrenamiento de AlexNet bajo las mismas condiciones de validación cruzada de 5 folds.

Tabla 4.3
Métricas de clasificación con AlexNet, sin smoothing y con data augmentation

Arquitectura	Recall(σ)	F1-Score (σ)	Accuracy global(σ)
Alexnet	**98,60%± 0,38%**	**98,63%± 0,39%**	**98,68% ± 0,38%**

La incorporación de data augmetation permitió aumentar la variabilidad del conjunto de entrenamiento, de esta forma se mejoró la capacidad de generalización del modelo. Como resultado, se obtuvo una mejora significativa en todas las métricas de clasificación, con una exactitud superior al 98,68%.

AlexNet con Data Augmentation con Smothing

De igual forma, el dataset con smoothing expandido a 4088 muestras fue evaluado bajo las mismas condiciones.

Tabla 4.4
Métricas de clasificación con AlexNet, con smoothing y data augmentation

Arquitectura	Recall (σ)	F1-Score (σ)	Accuracy global (σ)
Alexnet	98,27%± 0,54%	98,31%± 0,50%	98,36% ± 0,48%

La incorporación de data augmentation tuvo una mejora sustancial en el desempeño de AlexNet respecto a las configuraciones sin aumento de datos. La exactitud global pasó de 94,32% a 98,68% con sin smoothing, y de 94,13% a 98,36% con smoothing, lo que representa una ganancia de aproximadamente 4 puntos porcentuales en ambos casos. También la desviación estándar entre folds se redujo de $\pm$ 3,28% a $\pm$ 0,38% en dataset sin smoothing, de esta forma se evidencia una mayor estabilidad del modelo ante distintas particiones de datos.

En cuanto al efecto del smoothing, el dataset sin smoothing con data augmentation superó al smoothing en todas las métricas, con una accuracy de 98.68% frente a 98.36%. Este comportamiento sugiere que, al disponer de un volumen de datos suficientemente amplio, el modelo logra abstraer por sí mismo los patrones relevantes de la firma micro doppler sin requerir preprocesamiento adicional. Con base en estos resultados, se determinó que el dataset sin smoothing con data augmentation constituye la configuración de mayor desempeño para AlexNet, y bajo esta misma condición se procedió a evaluar las arquitecturas restantes: SqueezeNet, MobileNetV2 y ResNet50.

Comparativa todas las arquitecturas, sin smoothing con data augmentation

Tabla 4.5
Matrices de confusión consolidadas de todas las arquitecturas, sin smoothing con data augmentation

Arquitectura	VP Animales	FP/Personas	FP/Animales	VP Personas	Accuracy
AlexNet	**2377 (99,0%)**	**23 (1,0%)**	**31 (1,8%)**	**1657 (98,2%)**	**98,68%**
ResNet50	2376 (99,0%)	24 (1,0%)	43 (2,5%)	1645 (97,5%)	98,36%
SqueezeNet	2353 (98,0%)	47 (2,0%)	38 (2,3%)	1650 (97,7%)	97,92%
MobileNetV2	2306 (96,1%)	94 (3,9%)	39 (2,3%)	1649 (97,7%)	96,75%

VP = Verdaderos Positivos

FP=Falsos Positivos.

FP→Personas = animales clasificados como personas.

FP→Animales = personas clasificadas como animales.

El análisis de las matrices de confusión se puede observar el error que existe entre arquitecturas al clasificar entre animales y personas. AlexNet y ResNet50 presentaron el menor número de falsos positivos en la clase animales, con apenas 23 y 24 muestras mal clasificadas respectivamente, lo que se traduce en un recall del 99,0% para ambas. SqueezeNet incrementó este error a 47 muestras, mientras que MobileNetV2 registró el valor más elevado con 94 muestras de animales clasificadas como personas, esto explica su menor exactitud global de 96,75%.

En la clase personas, el patrón se invierte: ResNet50 presentó el mayor número de confusiones 43 muestras, lo que supera a AlexNet 31, SqueezeNet 38 y MobileNetV2 39. Esto indica que ResNet50, si bien es más conservadora al clasificar animales, tiende a asignar con menor seguridad las muestras de personas, comportamiento que se refleja en su menor recall para esa clase 97,45%. AlexNet logró el mejor equilibrio entre ambas clases, con la reducción de los errores de forma homogénea y con la mayor exactitud global con 98,68%, lo que la posiciona como la arquitectura de mayor desempeño para este escenario de clasificación binaria de dos clases.

Tabla 4.6

Comparativa de métricas de clasificación con todas las arquitecturas, sin smoothing y con data augmentation

Arquitectura	Recall (σ)	F1-Score (σ)	Accuracy global (σ)
AlexNet	**98,60%± 0,38%**	**98,63%± 0,39%**	**98,68% ± 0.38%**
ResNet50	98,23%± 0,30%	98,31%± 0,25%	98,36% ± 0,24%
SqueezeNet	97,89%± 0,30%	97,86%± 0,46%	97,92% ± 0,45
MobileNetV2	96,89%± 0,60%	96,66%± 0,70%	96,75% ± 0,69%

AlexNet obtuvo la mayor exactitud global con 98,68%, seguida de ResNet50 con 98,36%, SqueezeNet con 97,92% y MobileNetV2 con 96,75%. ResNet50 presentó la menor desviación entre folds $\pm$ 0,24%, esto refleja el comportamiento más estable gracias a sus conexiones residuales. MobileNetV2, si bien obtuvo el desempeño más bajo, sigue colocándose de forma competitiva para implementaciones en hardware embebido de bajo consumo. En conjunto, todas las arquitecturas superaron el 96% de exactitud, obtenido en condiciones experimentales más controladas.

La expansión del dataset de 511 a 4088 muestras representó el factor de mayor influencia, esto produce una ganancia de aproximadamente 4 puntos porcentuales en exactitud junto con la reducción de la desviación estándar de $\pm$ 3,28% a $\pm$ 0,38%, lo que evidencia una mayor estabilidad y capacidad de generalización del modelo.

En cuanto al recall de AlexNet y ResNet50 alcanzaron los valores más altos, indicando que ambas arquitecturas detectaron correctamente casi la totalidad de las muestras reales. Asimismo, AlexNet registró el mayor F1-score, lo que evidencia el mejor equilibrio entre precisión y sensibilidad, consolidándose como la arquitectura con mejor desempeño para el problema de clasificación propuesto.

Comparativa todas las arquitecturas, con Smothing y data augmentation para tres clases.

Con el fin de extender el sistema a un escenario más exigente, se evaluaron las cuatro arquitecturas en la clasificación de tres clases: gato, perro y persona, haciendo uso de smoothing y con data augmentation, configuración que mostró mejor rendimiento frente al dataset sin smoothing en este escenario.

Tabla 4.7

Métricas por clase de todas las arquitecturas, 3 clases con smoothing con data augmentation.

Arquitectura	Clase	Recall	F1-Score
AlexNet	**Gato**	**95,70%**	**92,48%**
	Perro	**95,66%**	**96,70%**
	Persona	**98,10%**	**97,90%**
ResNet50	**Gato**	89,96%	90,05%

	Perro	96,23%	96,36%
	Persona	98,10%	97,93%
SqueezeNet	**Gato**	89,14%	87,88%
	Perro	94,40%	95,25%
	Persona	97,69%	97,11%
MobileNetV2	**Gato**	88,53%	90,19%
	Perro	93,88%	94,57%
	Persona	97,28%	95,97%

Tabla 4.8

Comparación del desempeño de las arquitecturas CNN mediante Accuracy, Recall Macro Average y F1-Score Macro Average para la clasificación de tres clases.

Arquitectura	Recall Macro (σ)	F1-Score Macro (σ)	Accuracy global (σ)
AlexNet	**96,49%± 1,03%**	**95,69%± 1,02%**	**96,67% ± 0,74%**
ResNet50	94,76%± 1,17%	94,77%± 0,75%	96,26% ± 0,42%
SqueezeNet	93,74%± 2,15%	93,40%± 1,75%	95,13% ± 1,00%
MobileNetV2	93,23%± 0,23%	93,58%± 0,47%	94,64% ± 0,51%

La Tabla 4.8. presenta la comparación del desempeño de las cuatro arquitecturas CNN evaluadas en el escenario de clasificación de tres clases (gato, perro y persona) con el uso de las métricas Accuracy, Recall Macro Average y F1-Score Macro Average. Los resultados muestran que AlexNet alcanzó el mejor desempeño general con una exactitud del 96,67% un Recall Macro Average de 96,49% y un F1-Score Macro Average de 95,69% lo que demuestra una elevada capacidad para identificar correctamente

las tres clases y poder mantener un equilibrio entre precisión y sensibilidad. En segundo lugar, ResNet50 se posicionó con una exactitud de 96,26% seguido por SqueezeNet con 95,13% y MobileNetV2 con 94,64%. Si bien todas las arquitecturas presentaron un desempeño superior al 94% de exactitud, AlexNet evidenció una mayor capacidad de generalización al obtener los mayores valores en las métricas macro, posicionándose como la arquitectura con mejor rendimiento para la clasificación de firmas micro doppler en el escenario de tres clases.

Tabla 4.9
Matrices de confusión consolidadas de todas las arquitecturas 3 clases con smoothing y con data augmentation.

Arquitectura	VP Gato	VP Perro	VP Personas	Accuracy
AlexNet	**467 (95,7%)**	**1829 (95,7%)**	**1656 (98,1%)**	**96,67%**
ResNet50	439 (90,0%)	1840 (96,2%)	1656 (98,1%)	96,26%
SqueezeNet	435 (89,1%)	1805 (94,4%)	1649 (97,7%)	95,13%
MobileNetV2	432 (88,5%)	1795 (93,9%)	1642 (97,3%)	94,64%

La incorporación de una tercera clase introduce una complejidad adicional, dado que las firmas micro doppler de perros y gatos son estructuralmente más similares entre sí que respecto a las personas. La presencia de clutter disperso introduce ambigüedad visual que dificulta la separación de sus patrones, y al eliminarlo mediante el uso de smoothimg, las CNNs disponen de rasgos más definidos para distinguir entre ambas especies animales. Este efecto se manifiesta de forma consistente en todas las arquitecturas: la clase gato registró las métricas más bajas en todos los modelos, con F1-scores que oscilan entre 87,88% en SqueezeNet y un 92,48% en AlexNet, mientras que las clases perro y persona mantuvieron valores superiores al 94% en la mayoría de los casos.

AlexNet obtuvo nuevamente la mayor exactitud global con 96,67%, seguida de ResNet50 con 96,26%, SqueezeNet con 95,13% y MobileNetV2 con 94,64%. A diferencia del escenario de dos clases, en este caso el uso de smoothing superó al dataset sin smoothing, lo que sugiere que el preprocesamiento espacial aporta mayor beneficio cuando la tarea de discriminación es más fina, lo que facilita la distinción

de los patrones sutiles entre especies animales de tamaño y comportamiento similares. ResNet50 mostró nuevamente la menor desviación estándar entre folds ±0,42%, lo que confirma su estabilidad en distintas particiones del dataset independientemente del número de clases.

Prueba estadística de Wilcoxon

Se aplicó el test de rangos con signo de Wilcoxon con el uso de AlexNet como modelo de referencia, bajo un nivel de significancia $\boldsymbol{\alpha = 0,05}$.

Tabla 4.10
Test de Wilcoxon con respecto a AlexNet, escenario 2 clases (α = 0,05)

Modelo Comparado	Métrica	P-valor (bilateral)	P-valor (unilateral)
ResNet50	Accuracy	0,2500	0,1250
	F1- Score	0,3125	0,1563
SqueezeNet	Accuracy	0,0625	0,0313
	F1- Score	0,0625	0,0313
MobileNetV2	Accuracy	0,0625	0,0313
	F1- Score	0,0625	0,0313

En el escenario de dos clases: los resultados del test unilateral (p=0,0313) indican que AlexNet superó de forma estadísticamente significativa a SqueezeNet y MobileNetV2 tanto en accuracy como en F1- Score. La comparación con ResNet50 no alcanzó significancia (p=0,1250), lo que sugiere que ambas arquitecturas operan a un nivel de desempeño equivalente en este escenario.

Tabla 4.11
Test de Wilcoxon con respecto a AlexNet, escenario 3 clases (α = 0,05)

Modelo Comparado	Métrica	P-valor (bilateral)	P-valor (unilateral)
ResNet50	Accuracy	0,4375	0,2188
	F1-Macro	0,3125	0,1563
SqueezeNet	Accuracy	0,1875	0,0938
	F1-Macro	0,1875	0,0938

MobileNetV2	Accuracy	0,0625	0,0313
	F1-Macro	0,0625	0,0313

En el escenario de tres clases, ninguna comparación alcanzó significancia estadística, con p-valores bilaterales que oscilan entre 0,0625 y 0,4375 lo que indica que al incrementar la complejidad del problema las diferencias de desempeño entre arquitecturas se reducen y ningún modelo destaca sobre el otro por lo que no hay diferencias significativas entre los modelos con el aumento de complejidad al incluir más clases.

Al tomar en conjunto los resultados de clasificación y la evidencia estadística aportada por el test de Wilcoxon, se concluye que AlexNet con dataset sin smoothing y data augmentation constituye la configuración de mayor desempeño para el escenario de dos clases, con una exactitud global de 98**,**68% con superioridad estadísticamente demostrable frente a SqueezeNet y MobileNetV2. Para el escenario de tres clases, si bien AlexNet mantiene la mayor exactitud 96**,**67%, la ausencia de diferencias significativas entre modelos sugiere que cualquiera de las arquitecturas evaluadas representa una solución válida, es importante considerar los recursos computacionales disponible en el hardware de despliegue al momento de tomar una decisión final.

Comparación global del desempeño de las arquitecturas CNN

Con el fin de comparar el desempeño de las arquitecturas CNN evaluadas se generó un diagrama de burbujas para los escenarios de clasificación de dos y tres clases. Esta representación gráfica permite analizar simultáneamente la exactitud promedio, el tiempo de predicción por imagen y el tamaño del modelo en memoria. De esta forma, además de comparar la capacidad de clasificación de cada arquitectura, también es posible evaluar su eficiencia computacional y observar el equilibrio entre precisión, velocidad de procesamiento y consumo de memoria.

Figura 4.4

Comparación de las arquitecturas CNN con datos sin smoothing y data augmentation para el escenario de dos clases con evaluación de Accuracy, tiempo de predicción y tamaño del modelo.

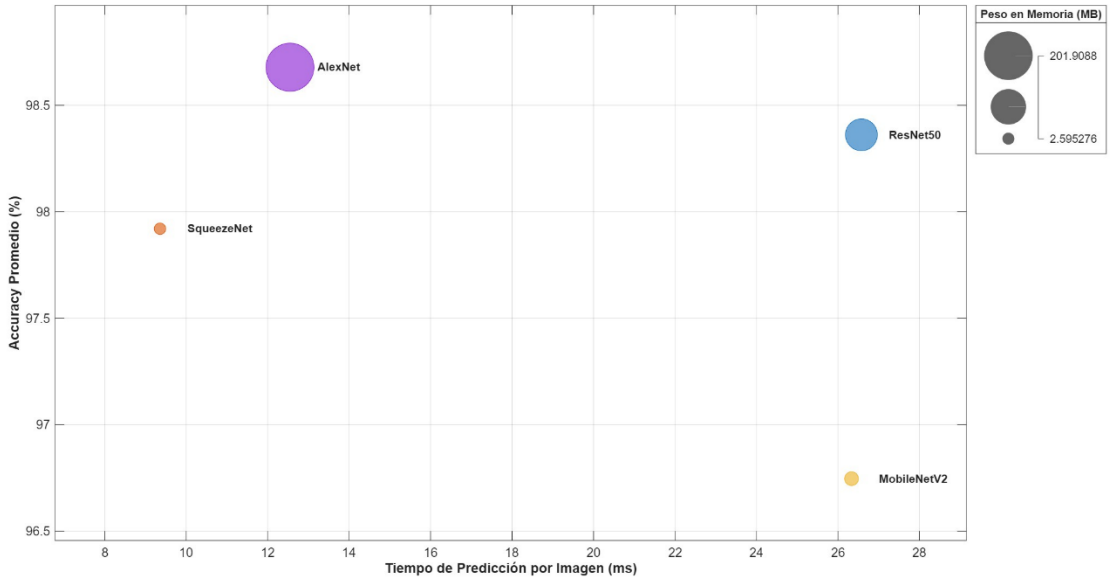
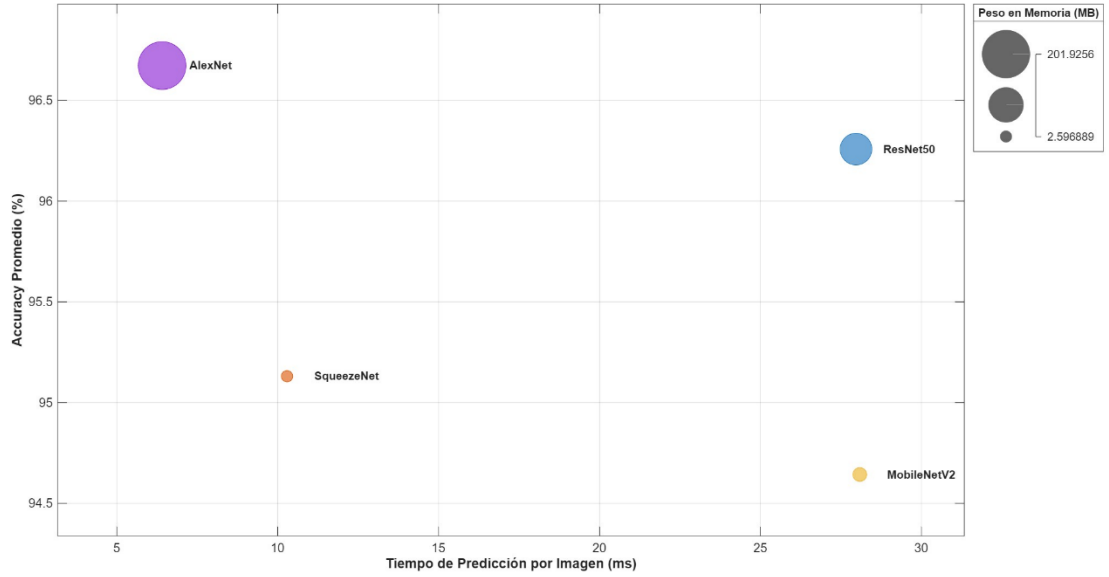


Figura 4.5

Comparación de las arquitecturas CNN con smoothing y con data augmentation para el escenario de tres clases con evaluación de Accuracy, tiempo de predicción y tamaño del modelo.



Las Figuras 4.4 y 4.5 comparan el desempeño de las arquitecturas CNN en los escenarios de clasificación de dos y tres clases considerando la exactitud promedio, el tiempo de predicción por imagen y el peso del modelo en memoria. El número de clases en todas las arquitecturas, contribuyó en una ligera disminución en la precisión. AlexNet obtuvo la mayor precisión en ambos escenarios, alcanzando un 98,68% en dos clases y un 96,67% en tres clases, con tiempos de predicción de 9,31 ms y 8,75 ms respectivamente. Sin embargo, también presentó el mayor tamaño del modelo, cercano a 202 MB, ResNet50 alcanzó precisiones de 98,36% y 96,26% con tiempos de 24,99 ms y 23,62 ms respectivamente con un tamaño intermedio de aproximadamente 83,76 MB. Por su parte, SqueezeNet obtuvo precisiones de 97,92% y 95,13% con tiempos de predicción de 10,60 ms y 10,33 ms siendo la arquitectura más liviana, con un tamaño cercano a 2,60 MB. Finamente con MobileNetV2 se alcanzó exactitudes de 96,75% y 94,64%, con tiempos de 24,11 ms y 26,48 ms respectivamente y un tamaño aproximado de 8,17 MB. Los resultados muestran que al aumentar la complejidad del problema reduce ligeramente el rendimiento de todas las arquitecturas. A pesar del aumento de la complejidad AlexNet mantuvo el mejor equilibrio entre precisión y velocidad mientras que SqueezeNet se destacó por su bajo consumo de memoria.

Cabe destacar que los tiempos de predicción presentados en los diagramas de burbujas son dependientes del entorno de hardware empleados para su evaluación por lo que pueden variar al ejecutarse en un hardware diferente. En esta investigación, las mediciones se realizaron utilizando un computador con procesador AMD Ryzen 5 7235HS de 3,20 GHz, 16GB de memoria RAM y sistema operativo Windows 11 Home de 64 bits. De igual forma, el entrenamiento y la evaluación de las arquitecturas CNN se ejecutaron utilizando una tarjeta gráfica NVIDIA GeForce RTX 3050 Laptop GPU la cual permitió acelerar el procesamiento mediante CUDA. Estas especificaciones se consideran como la base para la comparación del tiempo de predicción entre las diferentes arquitecturas evaluadas.

Capítulo 5

Conclusiones

La adquisición de datos en campo permitió consolidar un conjunto de datos robusto y representativo de las firmas micro doppler generadas por el movimiento de animales y personas, esta etapa de adquisición directa validó la viabilidad del uso de sistemas de radar para la discriminación de objetivos biológicos en entornos de interiores reales, superando las restricciones de los escenarios puramente simulados o controlados.

A partir de la evaluación visual y un análisis cualitativo de las firmas micro Doppler se observó que la STFT proporciona la representación de tiempo versus frecuencia más adecuada en comparación con la Transformada Wavelet Continua y la transformada adaptativa de Superlets, la STFT demostró una mayor capacidad para conservar los rasgos morfológicos y las características discriminativas de los objetivos analizados, estableciéndose como el método óptimo para la generación del conjunto de datos de imágenes destinado al entrenamiento y evaluación de las CNNs.

El análisis experimental demostró que el impacto de la técnica smoothing está condicionado a la complejidad de la tarea de clasificación. En el escenario de dos clases, cualquiera de los cuatro clasificadores no experimentó variaciones significativas con este preprocesamiento, logrando una exactitud equivalente al usar el dataset sin smoothing (94,32% ± 3,28%) o con smoothing (94,13% ±3,11%). No obstante, al momento de realizar un clasificador de tres clases (perro, gato y persona), el dataset con smoothing superó al original. Esto evidencia que la reducción del clutter disperso y la delimitación de los contornos espectrales aporta un beneficio crítico únicamente ante tareas de discriminación fina, donde los patrones de movimiento de especies biológicamente similares (caninos y felinos) tienden a solaparse.

La implementación de técnicas de data augmentation constituyó el factor de mayor influencia en la optimización del rendimiento y estabilidad de los modelos estudiados. Al aumentar el conjunto de datos de 511 a 4088 muestras la arquitectura AlexNet incrementó su exactitud global en aproximadamente 4%, alcanzando un 98,68% en el escenario de datos sin smoothing. De igual manera, la reducción de la

desviación estándar entre folds de validación cruzada (de ±3,28% a ±0,38%) ratifica que la diversificación sintética de las muestras mitiga el sobreajuste (overfitting), robustece la estabilidad y eleva la capacidad de generalización ante variaciones en la firma de radar.

En el escenario de dos clases, AlexNet se posicionó como el modelo de mayor desempeño general con una exactitud del 98,68%, logrando un balance óptimo entre sensibilidad y precisión reflejado en su F1-Score (98,63%). La aplicación del test de rangos con signo de Wilcoxon (α= 0,05) confirmó que la superioridad de AlexNet es estadísticamente significativa frente a SqueezeNet y MobileNetV2 (p = 0,0313), mientras que opera a un nivel estadísticamente equivalente en comparación con ResNet50 (p = 0,1250). Al introducir la tercera clase, el rendimiento global decreció de forma homogénea debido a la complejidad del problema, y aunque AlexNet retuvo la exactitud más alta (96,67%), el test estadístico no arrojó diferencias significativas entre los modelos ($p \geq 0,0625$), lo que demuestra que, al aumentar la dificultad del problema, la diferencia de rendimiento entre las redes se reduce y todas tienden a responder con un nivel de efectividad muy similar.

A través del análisis de las métricas como precisión, velocidad y tamaño, se concluye que existe un compromiso entre los recursos requeridos y la efectividad del clasificador:

- AlexNet resulta ser la alternativa más sólida para el sistema al lograr una exactitud global en ambos escenarios (98,68% en dos clases y 96,67% en tres clases), con un menor tiempo de inferencia por imagen (~9 ms en dos clases y 3 clases) requiriendo a cambio el mayor peso en memoria (hasta ~ 201 MB).
- ResNet50 destaca por su alta exactitud (98,36% en dos clases) y su robustez frente a las variaciones en los folds de validación. Sin embargo, esta precisión impacta negativamente en el tiempo de inferencia por imagen siendo estos mayores a comparación de AlexNet (~43 ms en dos clases y ~21.6 ms en tres clases), on un peso en memoria de ~ 83,76 MB.
- SqueezeNet se consolida como la arquitectura óptima para entornos con severas restricciones de hardware de bajo consumo o almacenamiento, manteniendo exactitudes destacables (superiores al 95%) con la huella de memoria más reducida de todo el grupo (~2,60 MB en

ambos escenarios). Además, ofrece un excelente tiempo de procesamiento (~15 ms en dos clases y ~11.2 ms en tres clases).

- MobileNetV2, a pesar de estar diseñada para la eficiencia, presentó el desempeño más bajo en exactitud en ambos escenarios (96.75% en dos clases y 94.64% en tres clases) junto con los tiempos de predicción más elevados de la comparativa (~49 ms en dos clases y ~22.6 ms en tres clases), superando incluso a ResNet50. No obstante, su reducido peso en memoria de ~8,17 MB la mantiene como una alternativa competitiva y viable frente a modelos masivos cuando el objetivo primordial es el despliegue en sistemas embebidos con recursos físicos limitados.

Trabajos Futuros:

Como trabajos futuros se recomienda explorar la implementación del sistema en operación de tiempo real, migrando el procesamiento desde un entorno de cómputo de escritorio hacia una tarjeta de desarrollo embebida con capacidad de procesamiento suficiente para ejecutar la inferencia de los modelos CNN entrenados. Esto implicaría optimizar las arquitecturas seleccionada de manera que puedan ejecutarse eficientemente en hardware de recursos limitados sin comprometer significativamente la exactitud alcanzada. De igual forma, se propone incorporar un módulo de comunicación mediante el protocolo MQTT que permita el envío de los resultados de clasificación en tiempo real hacia una plataforma de monitoreo o un sistema IoT centralizado para poder integrar el sistema desarrollado dentro de aplicaciones prácticas de hogares o edificios inteligentes.

Referencias

[1] «Evolución de la tecnología de radar: avances y futuro - Plaza Chimalhuacan». Accedido: 29 de mayo de 2026. [En línea]. Disponible en: https://plazachimalhuacan.mx/evolucion-de-la-tecnologia-de-radar-avances-y-futuro/

[2] M. S. Greco, J. Li, T. Long, y A. Zoubir, «Advances in Radar Systems for Modern Civilian and Commercial Applications: Part 1 [From the Guest Editors]», *IEEE Signal Process. Mag.*, vol. 36, n.º 4, pp. 13-15, jul. 2019, doi: 10.1109/MSP.2019.2911108.

[3] «imec magazine March 2018 - Ultra-small, economical and cheap radar made possible thanks to chip technology», imec. Accedido: 12 de julio de 2026. [En línea]. Disponible en: https://www.imec-int.com/en/imec-magazine/imec-magazine-march-2018/ultra-small-economical-and-cheap-radar-made-possible-thanks-to-chip-technology

[4] A. Hanif, M. Muaz, A. Hasan, y M. Adeel, «Micro-Doppler Based Target Recognition With Radars: A Review», *IEEE Sens. J.*, vol. 22, n.º 4, pp. 2948-2961, feb. 2022, doi: 10.1109/JSEN.2022.3141213.

[5] «Sensor de radar de onda milimétrica : Usos clave en industrias», MinewTienda. Accedido: 29 de mayo de 2026. [En línea]. Disponible en: https://www.minewstore.com/millimeter-wave-radar-sensor-advancements-and-applications-in-modern-industries

[6] P. O, «Así funciona el radar antiolvido de niños y mascotas», El Norte de Castilla. Accedido: 29 de mayo de 2026. [En línea]. Disponible en: https://www.elnortedecastilla.es/sociedad/radar-anti-olvido-20221201192523-nt.html

[7] linpowave, «Radar de ondas milimétricas en cabina: mejora la seguridad y una cabina inteligente», Ningbo Linpowave. Accedido: 29 de mayo de 2026. [En línea]. Disponible en: https://linpowave.com/es/blog/in-cabin-millimeter-wave-radar

[8] V. Kozlov *et al.*, «Micro-Doppler radar deception and camouflage of airborne targets», *Sci. Rep.*, vol. 16, n.º 1, p. 847, dic. 2025, doi: 10.1038/s41598-025-30442-2.

[9] Z. Xing, P. Chen, J. Wang, Y. Bai, J. Song, y L. Tian, «Millimeter-Wave Radar Detection and Localization of a Human in Indoor Complex Environments», *Remote Sens.*, vol. 16, n.º 14, p. 2572, jul. 2024, doi: 10.3390/rs16142572.

[10] D. Buchman, M. Drozdov, T. Krilavičius, R. Maskeliūnas, y R. Damaševičius, «Pedestrian and Animal Recognition Using Doppler Radar Signature and Deep Learning», *Sensors*, vol. 22, n.º 9, p. 3456, may 2022, doi: 10.3390/s22093456.

[11] X. Yu, T. J. Jiao, H. Lv, Y. Zhang, Z. Li, y J. Q. Wang, «A new use of UWB radar to detecting victims and discriminating humans from animals», en *2016 16th International Conference on Ground Penetrating Radar (GPR)*, jun. 2016, pp. 1-5. doi: 10.1109/ICGPR.2016.7572661.

[12] V. C. Chen, Fayin Li, Shen-Shyang Ho, y H. Wechsler, «Micro-doppler effect in radar: phenomenon, model, and simulation study», *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, n.º 1, pp. 2-21, ene. 2006, doi: 10.1109/TAES.2006.1603402.

[13] «BGT60LTR11AIP - 60 GHz radar sensors for IoT | Infineon Technologies». Accedido: 11 de mayo de 2026. [En línea]. Disponible en: https://www.infineon.com/part/BGT60LTR11AIP

[14] T.-Y. Huang, J. Lin, y L. Hayward, «Non-invasive measurement of laboratory rat's cardiorespiratory movement using a 60-GHz radar and nonlinear Doppler phase modulation», en *2015 IEEE MTT-S 2015 International Microwave Workshop Series on RF and Wireless Technologies for Biomedical and Healthcare Applications (IMWS-BIO)*, Taipei, Taiwan: IEEE, sep. 2015, pp. 83-84. doi: 10.1109/IMWS-BIO.2015.7303788.

[15] W. Chen *et al.*, «Environment-aware Multi-person Tracking in Indoor Environments with MmWave Radars», *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 7, n.º 3, pp. 1-29, sep. 2023, doi: 10.1145/3610902.

[16] A. B. R. C. D. Mattos, G. Brante, G. L. Moritz, y R. D. Souza, «Human and Small Animal Detection Using Multiple Millimeter-Wave Radars and Data Fusion: Enabling Safe Applications», *Sensors*, vol. 24, n.º 6, p. 1901, mar. 2024, doi: 10.3390/s24061901.

[17] Y. Ying, «Integration of Radar Technology in Smart homes: Applications, Technical Models, and Performance Evaluation», *Sci. Technol. Eng. Chem. Environ. Prot.*, vol. 1, n.º 10, dic. 2024, doi: 10.61173/h19azr74.
[18] Z. Xia, G. Ding, H. Wang, y F. Xu, «Person Identification With Millimeter-Wave Radar in Realistic Smart Home Scenarios», *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1-5, 2022, doi: 10.1109/LGRS.2021.3117001.
[19] J. Nino, «Japón y su innovador modelo de atención domiciliaria inteligente», TITI. Accedido: 3 de junio de 2026. [En línea]. Disponible en: https://titi.social/japon-y-su-innovador-modelo-de-atencion-domiciliaria-inteligente/
[20] J. L. Alonso Cerpa, «Introducción a los sistemas de radar y ARPA», 2015, Accedido: 18 de junio de 2026. [En línea]. Disponible en: https://riull.ull.es/xmlui/handle/915/6842
[21] X. Wang *et al.*, «Millimeter Wave Communication: A Comprehensive Survey», *IEEE Commun. Surv. Tutor.*, vol. 20, n.º 3, pp. 1616-1653, 2018, doi: 10.1109/COMST.2018.2844322.
[22] «What are Millimeter Waves? - everything RF». Accedido: 18 de junio de 2026. [En línea]. Disponible en: https://www.everythingrf.com/community/what-are-millimeter-waves
[23] M. I. Skolnik, «AN INTRODUCTION TO RADAR».
[24] M. G. Juárez, «▷ Efecto Doppler: qué es, ejemplos, fórmula y aplicaciones», Ingenierizando. Accedido: 19 de junio de 2026. [En línea]. Disponible en: https://www.ingenierizando.com/optica/efecto-doppler/
[25] «Efecto Doppler | Fisicalab». Accedido: 19 de junio de 2026. [En línea]. Disponible en: https://www.fisicalab.com/apartado/efecto-doppler
[26] «Chapter 14: CW and FM Radar | GlobalSpec». Accedido: 19 de junio de 2026. [En línea]. Disponible en: https://www.globalspec.com/reference/75045/203279/chapter-14-cw-and-fm-radar
[27] V. C. Chen, «Micro-Doppler effect of micromotion dynamics: a review», presentado en AeroSense 2003, A. J. Bell, M. V. Wickerhauser, y H. H. Szu, Eds., Orlando, FL, abr. 2003, p. 240. doi: 10.1117/12.488855.
[28] D.-I. (FH) C. Wolff, «Radartutorial». Accedido: 19 de junio de 2026. [En línea]. Disponible en: https://www.radartutorial.eu/02.basics/Continuous%20Wave%20Radar.en.html
[29] H. Jeon, Y. Jung, S. Lee, y Y. Jung, «Area-Efficient Short-Time Fourier Transform Processor for Time–Frequency Analysis of Non-Stationary Signals», *Appl. Sci.*, vol. 10, n.º 20, p. 7208, oct. 2020, doi: 10.3390/app10207208.
[30] A. V. Oppenheim, R. W. Schafer, y J. R. Buck, *Discrete-time signal processing*, 2nd ed. Upper Saddle River, N.J: Prentice Hall, 1999.
[31] F. J. Harris, «On the use of windows for harmonic analysis with the discrete Fourier transform», *Proc. IEEE*, vol. 66, n.º 1, pp. 51-83, 1978, doi: 10.1109/PROC.1978.10837.
[32] I. Goodfellow, Y. Bengio, y A. Courville, *Deep learning*. en Adaptive computation and machine learning. Cambridge, Massachusetts: The MIT Press, 2016.
[33] A. Ali, S. Z. Kazmi, G. Shahzadi, M. Rashid, y M. U. Ahsan, «Demystifying CNN with Mathematical Insights: A Prelude with Application to an AI-based Sustainable Solution for Diabetic Retinopathy Diagnosis», 11 de septiembre de 2023, *In Review*. doi: 10.21203/rs.3.rs-3338196/v1.
[34] A. Krizhevsky, I. Sutskever, y G. E. Hinton, «ImageNet Classification with Deep Convolutional Neural Networks», en *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2012. Accedido: 7 de junio de 2026. [En línea]. Disponible en: https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html
[35] D. A. Pérez-Aguilar, R. H. Risco-Ramos, y L. Casaverde-Pacherrez, «Transfer learning en la clasificación binaria de imágenes térmicas», *Ingenius*, n.º 26, pp. 71-86, jun. 2021, doi: 10.17163/ings.n26.2021.07.
[36] K. He, X. Zhang, S. Ren, y J. Sun, «Deep Residual Learning for Image Recognition», 10 de diciembre de 2015, *arXiv*: arXiv:1512.03385. doi: 10.48550/arXiv.1512.03385.

[37] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, y L.-C. Chen, «MobileNetV2: Inverted Residuals and Linear Bottlenecks», 21 de marzo de 2019, *arXiv*: arXiv:1801.04381. doi: 10.48550/arXiv.1801.04381.
[38] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, y K. Keutzer, «SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size», 4 de noviembre de 2016, *arXiv*: arXiv:1602.07360. doi: 10.48550/arXiv.1602.07360.