Optimización de procesadores CFAR 2D para la mitigación de clutter en mapas rango-doppler mediante entornos de simulación

Dota Fernandez Kevin Joseph y Leppe Bravo Lucca Nicolas

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



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

MSc. Lara Mina Marco Fernando

Director

C.C.: 171514011-5

Departamento de Eléctrica, Electrónica y Telecomunicaciones

Carrera de Telecomunicaciones

Certificación

Certifico que el trabajo de integración curricular: “Optimización de procesadores CFAR 2D para la mitigación de clutter en mapas rango-doppler mediante entornos de simulación” fue realizado por los señores Dota Fernandez, Kevin Joseph y Leppe Bravo Lucca Nicolas, 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, 24 de julio de 2026

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

MSc. Lara Mina Marco Fernando

Departamento de Eléctrica, Electrónica y Telecomunicaciones

Carrera de Telecomunicaciones

Responsabilidad de Autoría

Nosotros, Dota Fernandez, Kevin Joseph y Leppe Bravo Lucca Nicolas, con cédulas de ciudadanía n° 1718820385 y n° 1718828070., declaramos que el contenido, ideas y criterios del trabajo de integración curricular**:** Optimización de procesadores CFAR 2D para la mitigación de clutter en mapas rango-doppler mediante entornos de simulación

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, 24 de julio de 2026

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
KEVIN JOSEPH DOTA FERNANDEZ

Dota Fernandez Kevin Joseph

C.C.: 171882035-8

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
LUCCA NICOLAS LEPPE BRAVO

Leppe Bravo Lucca Nicolas

C.C.: 171882807-0

Departamento de Eléctrica, Electrónica y Telecomunicaciones

Carrera de Telecomunicaciones

Autorización de Publicación

Nosotros, Dota Fernandez, Kevin Joseph y Leppe Bravo Lucca Nicolas, con cédulas de ciudadanía n° 1718820385 y n° 1718828070. autorizamos a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de integración curricular: Optimización de procesadores CFAR 2D para la mitigación de clutter en mapas rango-doppler mediante entornos de simulación en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi/nuestra responsabilidad.

Sangolquí, 24 de julio de 2026

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
KEVIN JOSEPH DOTA FERNANDEZ

Dota Fernandez Kevin Joseph

C.C.: 171882035-8

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
LUCCA NICOLAS LEPPE BRAVO

Leppe Bravo Lucca Nicolas

C.C.: 171882807-0

Dedicatoria

Dedico este trabajo a quienes más aportaron en mi crecimiento como ser humano.

A Enma Córdova, quien fue, es y será mi guía espiritual; a Ángel Fernández y Leopoldo Dota, que son el ejemplo de trabajo duro y de quienes anhelo aprender el valor del esfuerzo y la constancia Un espacio muy especial está reservado para mi compañero más leal en los momentos de soledad y en las largas noches de silencio: mi gato Willy, quien me enseñó que el sacrificio hecho por amor es la recompensa más grande. Por todo esto, este trabajo está dedicado a quienes ahora me acompañan desde la eternidad, para preservar su espíritu.

Dedicatoria

A mis padres y a mi hermano por estar siempre conmigo cuando los necesité, por ser un apoyo incondicional entregándome su confianza, cariño y amor más sincero, siempre creyendo en mis capacidades.

A mi novia, por acompañarme durante los últimos años de mi formación académica, por su comprensión, cariño y apoyo.

A todos ellos, les dedico este logro, porque su presencia y confianza fueron fundamentales para alcanzar esta meta.

Agradecimiento

Quiero expresar mi más profundo agradecimiento a mis padres, José Dota y Mélida Fernández. Sus valores, enseñanzas y la educación moral y espiritual que me han inculcado han sido el pilar fundamental para perseverar y culminar esta etapa de mi vida. A mis hermanos Darwin Gabriela y Bryan, por su aliento constante en el día a día. A mi compañero y amigo Lucca Leppe por todo el tiempo dedicado a este trabajo y por su tiempo de calidad. A mis amigos, Joel Alejandro, Santiago, Eduardo, Sebastian, Francisco, Ariel, Fernando, Johand, Joel, Jhylmar, Edwin y Erick por su compañía incondicional y por brindarme los espacios de descanso necesarios durante este proceso. Agradezco con todo mi corazón y cariño a Melany Villa quien fue mi compañera de clases y de vida durante esta maravillosa etapa, a sus padres Angel y Jaqueline que fueron un apoyo incondicional dentro de mis estudios y su hermana Anahi.

Finalmente, extiendo mi gratitud al Ing. Marco Fernando Lara Mina, MSc., por su invaluable guía metodológica, su tiempo y los conocimientos impartidos, los cuales fueron determinantes para el desarrollo de este trabajo.

Agradecimiento

Agradezco profundamente al Ing. Marco Fernando Lara Mina, por su paciencia, dedicación y apoyo durante este trabajo de investigación, por sus conocimientos y guía que fueron un pilar fundamental para el cumplimiento de nuestro objetivo.

De igual manera, expreso mi gratitud a los docentes de la carrera, quienes contribuyeron a convertir esta etapa académica en una experiencia caracterizada por el respeto, la cercanía y la amistad. Gracias a ellos, nunca me sentí como un desconocido, sino como parte de una comunidad en la que pude disfrutar del tiempo compartido, de sus enseñanzas y de su compañía.

Finalmente expreso mi más profunda gratitud a mi compañero y amigo Kevin Dota, por acompañarme en este viaje final, y a mi amigo Nicolás Aulestia por su aporte y apoyo durante todo el proceso educativo. También agradezco infinitamente a mis padres, a mi hermano y mi novia por siempre preocuparse por mí y por convertirse en mi principal motivación para culminar la carrera.

Índice de contenidos

Certificación 3
Responsabilidad de Autoría 4
Dedicatoria 6
Dedicatoria 7
Agradecimiento 8
Agradecimiento 9
Resumen 16
Abstract 17
Capítulo I 18
Introducción 18
Estado del arte 20
Objetivos 23
Objetivo general: 23
Objetivos específicos: 23
Justificación 23
Capítulo II 26
Marco teórico 26
Radares FMCW 26
Radar Distance2Go 27
Estimación de distancia con FMCW 29
Configuraciones de Chirps y Tramas 30
Frecuencia Doppler en Sistemas FMCW 31
Estimación Doppler con FMCW 32
Micro Doppler 33
Estimación Espectral 33

Tiempo Frecuencia 33
Transformada de Fourier de Tiempo Corto (STFT) 34
Procesamiento Espectral y Extracción del Desplazamiento Doppler 35
Aprendizaje Profundo 36
CNN y su uso en la clasificación de Micro Doppler 37
Arquitectura de Redes: AlexNet, SqueezeNet y Resnet-18 37
Capítulo III 40
Metodología 40
Adquisición de señales 40
Protocolo de medición 41
Diseño del escenario 42
Procesamiento Inicial 43
Calibración Estática y Escalamiento de la Señal 43
Transformada de Fourier para Rango (Range-FFT) y Detección de Objetivos 43
Procesamiento Doppler 44
Normalización, Segmentación Energética y Recorte de Imágenes 44
Exportación del Conjunto de Datos 45
Aumento de Datos 46
Selección de la Técnica de Aumento de Datos 47
Modelo de ruido AWGN 48
Procedimiento de generación del conjunto aumentado 50
Contribución del aumento de datos al entrenamiento de la CNN 50
Entrenamiento de Modelos 50
Entrenamiento mediante AlexNet 51
Adaptación de la arquitectura 52
Preparacion del conjunto de entrenamiento 52
Entrenamiento Mediante SqueezeNet 53
Adaptación de la arquitectura 53

Preparación del conjunto de datos 54
Entrenamiento mediante ResNet-18 54
Adaptación de la arquitectura 55
Preparación del conjunto de Datos 56
Validación cruzada K-Fold 56
Configuración del entrenamiento 57
Selección estadística 59
Estadística descriptiva 59
Prueba Global de Friedman 60
Comparaciones múltiples mediante la prueba de Wilcoxon 60
Correlación por comparaciones múltiples 61
Selección del mejor modelo 61
Despliegue Practico 62
Adquisición de la información radar 62
Procesamiento de la señal 62
Preparación de la imagen para la clasificación 63
Clasificación mediante la CNN seleccionada 63
Presentación del resultado 63
Capítulo IV 65
Análisis y discusión de los resultados 65
Entrenamiento de modelos de Deep Learning 67
Resultados con datos previos al procesamiento 67
Resultados con data augmentation 69
Influencia de la data augmentation 72
Resultados de Friedman globales y de Wilcoxon con datos previos al procesamiento 73
Resultados de Friedman globales y de Wilcoxon con datos con data augmentation 74
Capítulo V 76
Conclusiones 76

Recomendaciones 79

Referencias 80

Índice de tablas

Tabla 2. 1 Especificaciones técnicas y parámetros operativos del radar FMCW Distance2Go 28

Tabla 3. 1 Resumen de los parámetros de optimización y validación para las arquitecturas de aprendizaje profundo 58

Tabla 4. 1 Desempeño general de las CNN 68

Tabla 4. 2 F1-score global por clase 68

Tabla 4.3 Aciertos y errores de AlexNet 69

Tabla 4.4 Desempeño general de las CNN 70

Tabla 4.5 F1-score global por clase 70

Tabla 4.6 Aciertos y errores de ResNet-18 72

Tabla 4.7 F1-score macro obtenido por las arquitecturas con datos aumentados 73

Tabla 4.8 Resultados de Friedman y Wilcoxon, para datos previos al procesamiento, con $\alpha = 0.05$ 74

Tabla 4. 9 Resultados de Friedman y Wilcoxon para los datos con data augmentation, con $\alpha = 0.05$ 75

Índice de figuras

Figura 2. 1 Principio básico de operación del sistema de radar FMCW, (tomado de [8]) 27

Figura 2. 2 configuración de múltiples chirps de FMCW para estimación Doppler (Tomado de [8]) 30

Figura 2. 3 Definición espacial del ángulo α respecto al centro del haz del radar 32

Figura 2. 4 Espectrograma representativo de una señal micro-Doppler adquirida mediante radar FMCW 35

Figura 2. 5 Estructura general de una CNN diseñada con 3 categorías de clasificación 37

Figura 3. 1 Diagrama de bloques de la metodología utilizada 40

Figura 3. 2 Área definida por el campo de visión del radar dentro del entorno de trabajo 41

Figura 3. 3 Aumento de datos donde se selecciona la zona con mayor potencia del espectrograma y añadiendo ruido blanco Gaussiano 47

Figura 3. 4 Diagrama de flujo del despliegue practico realizado 64

Figura 4. 1 Espectrograma de persona 65

Figura 4. 2 Espectrograma de bicicleta 66

Figura 4. 3 Espectrograma de vehículo 67

Resumen

Los sistemas avanzados de asistencia al conductor (ADAS) requieren mecanismos confiables para clasificar objetivos ante condiciones climáticas adversas, donde los sensores ópticos tradicionales, como cámaras, disminuyen notablemente su precisión. Este proyecto desarrolla un sistema automatizado para clasificar con precisión tres tipos de objetivos viales (peatones, bicicletas y vehículos) mediante el análisis espectral de señales capturadas con un radar de onda continua modulada en frecuencia (FMCW) operando a 24 GHz. Las firmas micro-doppler características de cada clase se obtuvieron transformando las series temporales en espectrogramas mediante la Transformada de Fourier de Tiempo Corto (STFT), permitiendo capturar el movimiento dinámico intrínseco de cada blanco. Para mitigar el desbalance, evitar el sobreajuste y robustecer el entrenamiento, se amplió el conjunto de datos aplicando aumento artificial con Ruido Blanco Gaussiano Aditivo (AWGN) bajo tres niveles controlados de relación señal a ruido (SNR). Posteriormente, se entrenaron y evaluaron tres redes neuronales convolucionales preentrenadas (AlexNet, SqueezeNet y ResNet-18) empleando técnicas de aprendizaje por transferencia y un esquema riguroso de validación cruzada para medir su capacidad de generalización. Finalmente, se aplicaron pruebas estadísticas no paramétricas, específicamente los tests de Friedman y Wilcoxon con corrección de Holm para validar con significancia estadística el modelo con el mejor desempeño global. Esta metodología propone una solución robusta y eficiente para fortalecer los sistemas de percepción en futuras tecnologías de conducción asistida.

Palabras clave: Clasificación de objetivos, Radar de onda continua modulada en frecuencia, Firmas micro Doppler, Redes neuronales convolucionales, Sistemas avanzados de asistencia al conductor.

Abstract

Advanced Driver Assistance Systems (ADAS) require reliable target classification mechanisms in adverse weather environments, such as dense fog, rain, or low lighting, where conventional optical sensors like cameras experience substantial degradation in accuracy. This research develops a robust, automated framework to classify three vulnerable and common road targets (pedestrians, bicycles, and vehicles) using spectral representations derived from a 24 GHz Frequency-Modulated Continuous-Wave (FMCW) radar. Distinctive micro-Doppler signatures corresponding to the kinematics and structural motion of each target category were systematically extracted by converting raw radar beat signals into high-resolution time-frequency spectrograms via the Short-Time Fourier Transform (STFT). To alleviate data scarcity, mitigate overfitting, and enhance classification resilience against environmental interference, synthetic data augmentation was implemented using Additive White Gaussian Noise (AWGN) across three controlled Signal-to-Noise Ratio (SNR) profiles. Three prominent pretrained deep convolutional neural network architectures AlexNet, SqueezeNet, and ResNet-18 were fine-tuned using transfer learning and benchmarked via a stratified k-fold cross-validation scheme. Comprehensive performance metrics, including classification accuracy, confusion matrices, and computational efficiency, were evaluated, followed by non-parametric statistical hypothesis testing through the Friedman and Wilcoxon signed-rank tests with Holm corrections to rigorously confirm model superiority. Ultimately, this methodology delivers an accurate, computationally feasible perception solution tailored for integration into next-generation intelligent transportation and autonomous vehicle architectures.

Keywords: Target classification, Frequency modulated continuous wave radar, Micro Doppler signatures, Convolutional neural networks, Advanced driver assistance systems.

Capítulo I

Introducción

En la actualidad, las aplicaciones de radar abarcan un espectro diverso de aplicaciones, que van desde el monitoreo de la salud, la arqueología hasta radares de pronóstico meteorológico (Greco, Li, & Zoubir, 2019). No obstante, una de las áreas con mayor cantidad de desafíos, es el sector automotriz, donde, durante las últimas décadas, la ingeniería se ha centrado con especial énfasis en resolver desafíos inherentes a las aplicaciones automotrices. Debido a este interés sostenido, el desarrollo e implementación de algoritmos avanzados de procesamiento de señales ha experimentado un crecimiento constante (Greco, Li, & Zoubir, 2019).

Los avances mencionados han permitido la consolidación de los Sistemas Avanzados de Asistencia al Conductor (ADAS, del inglés Advanced Driver Assistence System). Sin embargo, a pesar de su adopción masiva, los sistemas ADAS actuales enfrentan desafíos críticos de fiabilidad (Antich, 2023). Uno de los problemas más reportados es el denominado frenado fantasma (*phantom braking*), una desaceleración repentina e inesperada del vehículo sin un peligro aparente. Este fenómeno suele originarse por errores en la calibración de los sensores o por la presencia de objetos en el entorno que generan falsos positivos (Antich, 2023).

Aunado a esto, estudios sobre los Sistemas de Frenado de Emergencia Autónomo (AEB, del inglés Autonomous Emergency Braking) señalan que el rendimiento disminuye significativamente bajo condiciones meteorológicas adversas, como lluvia, niebla y baja iluminación. Es precisamente ante estas limitaciones donde la mejora en el procesamiento de señales de radar y la clasificación inteligente de objetivos se vuelven indispensables para garantizar la seguridad vial (Smethurst, 2023).

En la literatura técnica, el Reconocimiento de Actividades Humanas (HAR, del inglés Human Activity Recognition) se ha abordado tradicionalmente mediante sistemas basados en visión artificial y sistemas basados en sensores. Si bien los sistemas visuales han

demostrado eficacia en condiciones ideales, presentan limitaciones críticas en la vía pública, especialmente ante la oclusión por obstáculos sólidos o la degradación de la imagen por factores climáticos. Por el contrario, los sensores de radar mantienen su operabilidad en entornos donde el clima afecta severamente a los sistemas ópticos, ya sea por lluvia, niebla o condiciones de baja iluminación, lo que garantiza una detección constante (Li, He, & Jing, 2019).

Este proyecto toma como base las estructuras de procesamiento utilizadas en aplicaciones HAR y adapta dichas metodologías para la clasificación de objetivos viales: peatones, bicicletas y vehículos. Como un soporte a la identificación visual, se utiliza la técnica de firmas micro doppler para obtener mapas tiempo-frecuencia. Estos mapas permiten capturar las variaciones de frecuencia características de cada clase de objetivo, lo que proporciona un conjunto de datos robustos que facilitan la distinción entre los tres tipos de entidades viales mencionados mediante el análisis de su respuesta al radar.

La clasificación automática de los datos robustos se llevará a cabo mediante Redes Neuronales Convolucionales (CNN, del inglés Convolutional Neural Networks). Las CNN son de las primeras arquitecturas de Aprendizaje Automático (*Machine Learning*) en utilizar múltiples capas jerárquicas (MathWorks) lo que las hace destacar en el estado del arte del Aprendizaje Profundo (*Deep Learning)*. Esta estructura permite extraer automáticamente características desde niveles básicos (bordes y formas en las representaciones tiempo-frecuencia) hasta niveles de abstracción superior, al optimizar el reconocimiento de patrones complejos que los algoritmos tradicionales de procesamiento de señales podrían omitir (MathWorks).

En este contexto, el presente trabajo propone el desarrollo de un sistema de clasificación de objetivos viales haciendo uso del radar de onda continúa modulada en frecuencia (FMCW, del inglés Frecuency Modulated Continuous Wave) para esto se usa un radar Infineon Xensiv Distance2Go, el cual opera en la banda ISM (Industrial, Scientific, Medical) de 24 GHz (Infineon Technologies AG, 2019). La elección de esta tecnología se fundamenta en su capacidad para medir la distancia y velocidad de los objetos mediante la

mezcla de señales transmitidas y reflejadas, lo que genera una señal de frecuencia intermedia (IF, del inglés Intermediate Frequency) o frecuencia de batido (*beat frequency*) proporcional a la distancia del cuerpo detectado. Esta arquitectura integrada permite el acceso a datos brutos de fase y cuadratura, esenciales para el procesamiento avanzado de señales (Mavares F, 2021).

Para lograr la distinción entre peatones, bicicletas y vehículos, se aplicará un análisis espectral basado en la Transformada Rápida de Fourier (FFT, del inglés Fast Fourier Transform) para obtener mapas de rango doppler (a partir de las representaciones tiempo-frecuencia) (Infineon Technologies AG, 2017). A partir de estos, se extraerán las firmas micro doppler, las cuales capturan las variaciones de frecuencia producidas por los movimientos característicos de cada objetivo. Estas firmas servirán como entrada para el entrenamiento de modelos de aprendizaje profundo, con el fin de evaluar su desempeño mediante métricas de exactitud y probabilidad de detección en escenarios viales controlados (Mavares F, 2021) (Infineon Technologies AG, 2017).

Estado del arte

En la actualidad, los sistemas de radar, que en un inicio fueron utilizados principalmente para la detección de misiles y aeronaves, han encontrado nuevas aplicaciones en sistemas ADAS (Greco M. , Li, Long, & Zoubir, 2019). Estos sistemas integran diferentes tipos de sensores, como cámaras, sensores infrarrojos y radares, con el objetivo de mejorar la seguridad vial y prevenir colisiones. Dentro de estos sistemas, el radar presenta varias ventajas frente a otros sensores, ya que es menos sensible a las condiciones de iluminación, condiciones climáticas y posee una mayor capacidad de penetración a través de ciertos obstáculos y permite preservar la privacidad de los individuos, debido a que no captura imágenes directas de personas o vehículos (Diaz, 2026).

Las tecnologías de radar empleadas para la detección y clasificación de objetos se basan en la emisión de pulsos de radio que se reflejan en los objetivos y regresan al receptor del radar. Una de las técnicas más utilizadas para la clasificación de objetos es el

análisis de las firmas micro doppler, las cuales pueden ser procesadas mediante modelos de aprendizaje automático, como redes neuronales, que permiten identificar patrones característicos asociados a diferentes tipos de objetivos en distintos entornos (Sanaullah, Piort, & Michal, 2026).

Estudios actuales convergen en el uso de arquitecturas de aprendizaje profundo alimentadas por señales de radiofrecuencia. Se destacan cuatro elementos principales:

Deep Learning para Clasificación: Se emplean predominantemente CNN para el análisis de imágenes de espectrogramas y Redes Neuronales Recurrentes (RNN, del inglés Recurrent Neural Network) y redes de memoria a corto y largo plazo (LTSM, del inglés Long Short Term Memory) para capturar la dependencia temporal de los movimientos humanos clasificados dentro de 3 categorías, actividades de la vida diaria, movimientos transicionales y movimientos anómalos o de peligro (Ghufron, Fiky Y, & Istiqomah, 2026).

Radar de Banda Ultra Ancha (UWB, del inglés Ultra WideBand): Modelos específicos diseñados para la detección de víctimas en entornos críticos y la discriminación precisa entre seres humanos y animales basándose en patrones biométricos como la respiración (Xin, y otros, 2026)

Espectrogramas y Firmas Micro Doppler: El modelo estándar consiste en transformar la señal cruda en representaciones tiempo-frecuencia haciendo uso de la transformada de Fourier en tiempo corto (STFT, del inglés Short-time Fourier Transform) para extraer patrones de movimiento intrínsecos de objetos que produzcan o sean afectados por vibraciones, rotaciones u oscilaciones mecánicas (Haoran, Tao, Gang, Yifan, & Yu, 2026).

Radares de FMCW: Utilizados ampliamente por su capacidad de medir distancia y velocidad simultáneamente en aplicaciones de sistemas ADAS en la industria automotriz (Greco M. , Li, Long, & Zoubir, 2019).

Robustez Ambiental: Han demostrado que el radar supera a los sensores ópticos en condiciones adversas (niebla, oscuridad, humo) y garantiza la privacidad al no capturar

imágenes identificables dentro de escenarios abiertos con niebla densa, tormentas de lluvia o nieve y entornos con polvo o humo suspendido (Elena, 2025).

Clasificación Multiclase: Los estudios han logrado distinguir con alta precisión entre peatones, ciclistas y vehículos en entornos urbanos complejos (Andras, Jiaao, Julian F, & Dariu M, 2020).

Detección Biométrica: Se ha perfeccionado la capacidad de detectar signos vitales mínimos, lo que permite aplicaciones en el cuidado de la salud (monitoreo de ancianos y pacientes) sin necesidad de contacto físico (Enrique, 2024).

A pesar de los avances, existen brechas técnicas identificadas en la bibliografía:

Generalización de los Modelos: Muchos sistemas de clasificación basados en micro doppler presentan dificultades para generalizar cuando los sujetos realizan movimientos no predefinidos o en ángulos de incidencia oblicuos al radar (Andras, Jiaao, Julian F, & Dariu M, 2020).

Consumo y Costo en ADAS: Aunque se evalúa el uso de radares FMCW para ADAS, todavía falta optimizar los algoritmos para que funcionen en sistemas de conducción, evaluando riesgos, detección de recursos en el control automatizado (Pallabi, Samadendra, Rabindranath, Agbotiname, & Chun-Ta, 2025).

Robustez ante el Ruido y Escasez de Datos: Aunque la literatura reporta altas tasas de precisión en la clasificación multiclase, la mayoría de los modelos asumen condiciones ideales de adquisición o dependen de conjuntos de datos masivos. Las técnicas tradicionales de visión por computador alteran la cinemática de las firmas micro doppler, por lo cual es imperativo explorar estrategias en el dominio del tiempo (como la inyección de ruido AWGN). Esta aproximación es fundamental para entrenar arquitecturas eficientes a través de *Transfer Learning* (Jaya, Sunil, & Gireesh, 2022) y garantizar su robustez frente a distintos niveles de relación señal a ruido (SNR, del inglés Signal Noise Ratio) en escenarios con adquisiciones limitadas. Existe una carencia significativa de metodologías orientadas al aumento de datos físicamente consistentes para radares FMCW (Neda , Passafiume, Sadeghibakhi, Collodi, & Cidronali, 2023).

Objetivos

Objetivo general:

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

Objetivos específicos:

Analizar y definir de las características de radares FMCW de onda milimétrica (mmWave)

Adquirir señales en fase y cuadratura de radar FMCW a 24GHz.

Procesar señales de radar para obtener micro firmas Doppler.

Implementar algoritmos de clasificación de blancos basados en Machine/Deep Learning.

Evaluar el desempeño del sistema en clasificación de peatones, bicicletas y automóviles.

Justificación

En los últimos años se ha experimentado un crecimiento acelerado de aplicaciones para sistemas de radar (Greco, Li, & Zoubir, 2019). Durante décadas se ha tratado de buscar una implementación real del procesamiento de señales en retos asociados a aplicaciones automovilísticas (Greco, Li, & Zoubir, 2019). Actualmente, estos sistemas se utilizan para reconocer el movimiento de personas, especialmente en ambientes donde la privacidad es un aspecto primordial (Elena, 2025). Se utiliza el mismo razonamiento y se implementan radares para la clasificación de objetivos dentro de ambientes donde pueden existir elementos que provocan falsos positivos en la distinción de los objetos claves (Andras, Jiaao, Julian F, & Dariu M, 2020).

La motivación de investigación de este trabajo surge de la evolución de los sistemas de radar como elemento principal en los sistemas ADAS para vehículos, con el fin de desarrollar tareas críticas como la detección pedestre (Pallabi, Samadendra, Rabindranath, Agbotiname, & Chun-Ta, 2025). La importancia radica en que los sistemas actuales ya no

solo deben cumplir tareas tradicionales como medir distancia, velocidad o ángulo ni detección general de objetivos, sino que se necesita reconocer y clasificar objetivos mediante la determinación de firmas doppler y micro firmas doppler (Neda , Passafiume, Sadeghibakhi, Collodi, & Cidronali, 2023). Este proyecto es importante para la seguridad vial, ya que permite que la toma de decisiones esté en manos de los sistemas ADAS.

La revisión de investigaciones previas del uso de sistemas ADAS se trata como una solución a la alta demanda de sistemas de radares en clasificar automáticamente la gran variedad de objetivos potenciales (Pallabi, Samadendra, Rabindranath, Agbotiname, & Chun-Ta, 2025).

En respuesta a estos desafíos, el presente trabajo de titulación propone el desarrollo de un sistema de clasificación multiclase a través de un radar FMCW a 24 GHz (Ghufron, Fiky Y, & Istiqomah, 2026). La investigación aborda de manera directa la clasificación de firmas micro doppler correspondientes a peatones, bicicletas y vehículos (Andras, Jiaao, Julian F, & Dariu M, 2020). Para superar las limitaciones asociadas a la escasez de datos experimentales y asegurar un nivel óptimo de generalización ante entornos reales, la metodología incorpora una etapa de aumento de datos mediante la inyección de ruido blanco gaussiano aditivo (AWGN, del inglés Additive White Gaussian Noise). A partir de estas representaciones, el reconocimiento de los objetivos se ejecuta a través de arquitecturas de aprendizaje profundo preentrenadas (Ghufron, Fiky Y, & Istiqomah, 2026). Este enfoque consolida un clasificador eficiente, robusto y directamente aplicable a la mejora en la toma de decisiones de los sistemas ADAS.

El presente trabajo de integración curricular se estructura en cinco capítulos. El Capítulo I expone la introducción, el estado del arte, los objetivos y la justificación de la investigación, con énfasis en la necesidad de desarrollar sistemas confiables para la clasificación de objetivos viales mediante señales de radar FMCW y técnicas de aprendizaje profundo.

En el Capítulo II se presenta los fundamentos teóricos relacionados con los radares FMCW, el radar XENSIV™ Distance2Go, los efectos Doppler y micro doppler, la estimación

espectral, la Transformada de Fourier de Tiempo Corto, las redes neuronales convolucionales utilizadas en el estudio y a el análisis estadístico a través de pruebas Friedman y Wilcoxon que se van a realizar para determinar la red neuronal con el mejor desempeño.

El Capítulo III describe la metodología aplicada, la cual comprende: la adquisición de señales de peatones, bicicletas y vehículos; el procesamiento de los datos en fase y cuadratura; la generación de representaciones tiempo-frecuencia de firmas micro doppler; el aumento de datos mediante ruido AWGN y recortes en tiempo; el entrenamiento de las arquitecturas AlexNet, SqueezeNet y ResNet-18; la validación cruzada de cinco particiones; la selección estadística del modelo y su despliegue práctico.

El Capítulo IV presenta el análisis y la discusión de los resultados obtenidos durante la evaluación de las redes neuronales. Finalmente, el Capítulo V reúne las conclusiones y recomendaciones derivadas del desarrollo y la evaluación del sistema propuesto.

Capítulo II

Marco teórico

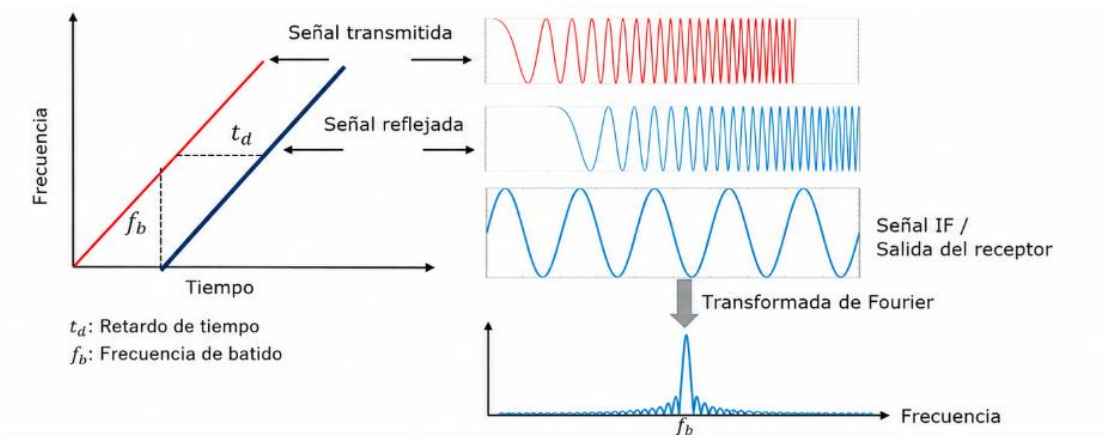
Un radar consiste en un sistema electrónico diseñado para la detección y seguimiento de objetivos (targets) a distancia a través de ondas electromagnéticas (Infineon Technologies AG, 2019). El equipo opera mediante la transmisión de señales de radiofrecuencia hacia el entorno, cuando estas ondas alcanzan un objetivo la energía de la señal se refleja de vuelta hacia la antena receptora (Infineon Technologies AG, 2019). El procesamiento posterior de esta señal reflejada permite extraer información precisa como la distancia, la posición y la velocidad del objetivo independientemente si este permanece fijo o se desplaza (Andras, Jiaao, Julian F, & Dariu M, 2020).

Radares FMCW

Los radares basados en FMCW emiten una señal cuya frecuencia cambia a través del tiempo, mediante un barrido a lo largo del ancho de banda asignado. En el receptor, la señal reflejada en el objetivo se mezcla con la señal original de transmisión. Este proceso produce una señal de baja frecuencia conocida como Frecuencia Intermedia (IF, del inglés Intermediate Frequency) (Jankiraman, 2018).

En la Figura 2.1, se representa la generación del pulso, se observa el cambio de frecuencia que ocurre cuando en el receptor llega la señal reflejada por el objetivo generando una diferencia en frecuencia conocida como frecuencia de batido (f_b).

Figura 2. 1 Principio básico de operación del sistema de radar FMCW, (tomado de (Infineon Technologies AG, 2017))



Radar Distance2Go

El kit Distance2Go es una plataforma de demostración para el chipset de radar BGT24MTR11 de 24 GHz, el cual está basado en Silicio-Germanio (SiGe) de Infineon. Esta placa está diseñada para medir la distancia, la velocidad y la dirección del movimiento (Infineon Technologies AG, 2019). Como se muestra en la Tabla 2.1 tenemos las especificaciones utilizadas del radar mencionado.

Tabla 2. 1

Especificaciones técnicas y parámetros operativos del radar FMCW Distance2Go

Parámetro	Unidad	Valor
Velocidad Mínima	km/h	0.33
Velocidad Máxima	km/h	7
Distancia Mínima	cm	50
Distancia Máxima	m	15
Precisión del alcance con $RCS = 1\ m^2$	cm	± 30
Precisión del alcance con $RCS = 10\ m^2$	cm	± 20
Frecuencia del Transmisor	GHz	24 - 24.25
Potencia Radiada Isotrópica Efectiva (PIRE)	dBm	+21
Frecuencia del receptor	GHz	24 - 24.25
Tipo de antena	Patch	2×4
Horizontal al ancho de haz de -3 dB	Grados	42
Elevación al ancho de haz de -3 dB	Grados	20

El análisis de las especificaciones técnicas del radar FMCW revela las capacidades y restricciones físicas impuestas durante la adquisición de firmas micro doppler. Para las pruebas, el radar FMCW opera en la banda de 24 a 24.25 GHz. Se fijó una Potencia Radiada Isotrópica Efectiva (PIRE) de $+21\ dBm$ para equilibrar la penetración de la señal frente a posibles interferencias del entorno. El alcance operativo es riguroso. El sensor restringe la detección de objetivos a una ventana espacial de 0.5 a 15 metros. Dentro de este margen, el hardware capta velocidades desde los 0.33 km/h hasta de 7 km/h. Estos parámetros determinan una restricción física estricta: el desplazamiento de los vehículos, las bicicletas y los peatones capturados en el conjunto de datos se ejecutó a baja velocidad. Este escenario emula condiciones de tráfico controlado, zonas de aproximación o áreas de alta seguridad peatonal.

La Tabla 2.1 también expone la dependencia directa entre la precisión del radar y la Sección Transversal del Radar (RCS, del inglés Radar Cross Section) del objetivo. El hardware responde distinto según la reflectividad del objetivo. Frente a un RCS de 10 m² (equivalente a un vehículo estándar), la estimación de alcance tiene un error de $\pm 20\ cm$.

Para blancos más pequeños como peatones (RCS = 1m²), esta incertidumbre salta a $\pm 30\ cm$.

Todo este proceso de detección se realiza mediante un arreglo de antenas tipo parche de 2×4, el cual irradia un lóbulo con una apertura de 42° en azimut y 20° en elevación en la transmision dependiendo de su polarizacion. Este patrón de radiación resulta ideal para monitorizar el movimiento sobre la superficie terrestre. A través de esta apertura concentrada, se minimiza la captura de reflexiones espurias provenientes de objetos elevados ajenos al área de interés.

Estimación de distancia con FMCW

La señal de IF resultante tiene un comportamiento sinusoidal y su frecuencia corresponde a la frecuencia de batido. El análisis espectral de esta frecuencia permite estimar con exactitud la distancia (R) hacia el objetivo (Infineon Technologies AG, 2017).

La relación matemática se presenta en Ec. 1 es:

$$R = \frac{c \cdot T_c \cdot f_b}{2 \cdot B}, \quad (2.1)$$

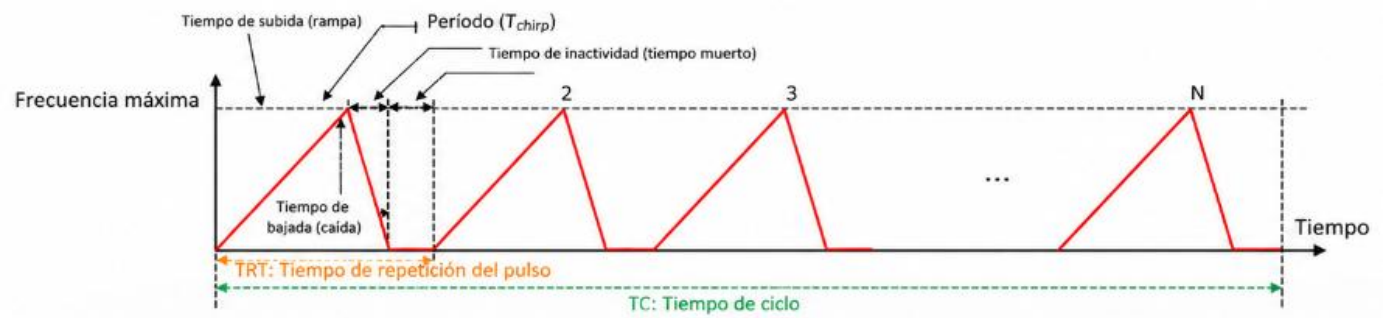
donde: c es la velocidad de la luz en el vacío (3×10^8 m/s), T_c representa el tiempo de duración del pulso ascendente (up-chirp), f_b es la frecuencia de batido y B corresponde al ancho de banda de la señal (Infineon Technologies AG, 2017).

Configuraciones de Chirps y Tramas

El funcionamiento del radar FMCW se basa en la emisión continua de chirps y tramas. Un chirp es una señal caracterizada por un cambio lineal de frecuencia en el tiempo como se muestra en la Figura 2.2, la cual se define por su frecuencia de inicio (f_c), su ancho de banda (B) y su duración (T_c) (Infineon Technologies AG, 2017). Esta variación de frecuencia puede ser ascendente en frecuencia (up-chirp) o descendente en frecuencia (down-chirp) a lo largo del pulso (Infineon Technologies AG, 2017).

En Figura 2.2, vemos el tiempo de ciclo de transmisión de una señal, la cual, inicia desde una frecuencia determinada hasta una frecuencia máxima, esta frecuencia máxima depende de la frecuencia máxima de trabajo del radar, esto se repite N veces, dependiendo del tiempo en que se expone el objetivo al radar.

Figura 2. 2 configuración de múltiples chirps de FMCW para estimación Doppler (Tomado de (Infineon Technologies AG, 2017))



Por su parte, una trama (frame) consiste en una secuencia de múltiples chirps idénticos separados por intervalos de tiempo iguales (Infineon Technologies AG, 2017). Para el cálculo de la velocidad de un objetivo, una técnica simple de comparación de fase resulta insuficiente con un solo pulso; por esta razón, los sistemas de radar requieren la configuración de múltiples chirps consecutivos dentro de una misma trama (Infineon Technologies AG, 2017).

Frecuencia Doppler en Sistemas FMCW

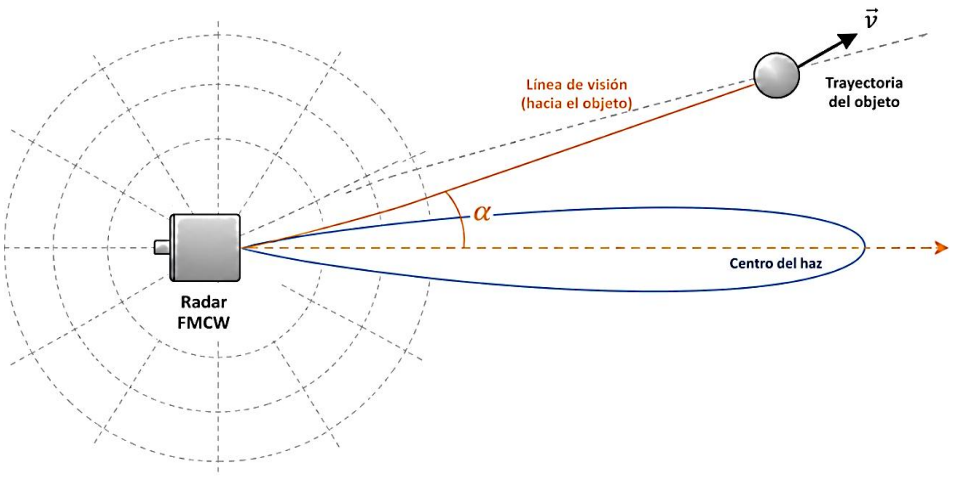
Se define como frecuencia Doppler al cambio aparente en la frecuencia de una onda percibido por un receptor cuando existe un movimiento relativo entre la fuente emisora y dicho receptor (Andras, Jiaao, Julian F, & Dariu M, 2020). El radar Doppler opera bajo el principio de emitir ondas electromagnéticas hacia un objeto en movimiento. Cuando la onda impacta el objetivo, la onda se refleja de regreso hacia el receptor con un desplazamiento en su frecuencia original (Andras, Jiaao, Julian F, & Dariu M, 2020).

Este desplazamiento o frecuencia Doppler (f_d) guarda una relación directa con la velocidad radial del objetivo (v) y se modela mediante las ecuaciones Ec. 2 donde describe la frecuencia Doppler (Infineon Technologies AG, 2017)

$$f_d = \frac{2 \cdot f_{Tx} \cdot v}{c} \cdot cos(\alpha), \qquad 2.2)$$

donde: f_{Tx} representa la frecuencia portadora (en el caso del radar usado en el presente estudio 24×10^9 Hz), c es la velocidad de la luz en el vacío, y α es el ángulo entre el centro del haz de radar y la trayectoria de movimiento del objetivo como se muestra en la Figura 2.3.

Figura 2. 3 Definición espacial del ángulo α respecto al centro del haz del radar



Estimación Doppler con FMCW

En la arquitectura específica de un radar FMCW, el movimiento del objetivo no altera la frecuencia de la señal de batido de forma significativa, pero sí introduce un cambio medible en su fase a través de múltiples chirps (Infineon Technologies AG, 2017). Esta variación de fase (Δ_ω) se asocia a la velocidad radial del blanco en movimiento (v_r) la expresion determinada por Ec. 4 [8]:

$$\Delta_\omega = \frac{4\pi \cdot v_r \cdot T_c}{\lambda}, \quad (2.3)$$

donde: T_c es la duración del chirp y λ es la longitud de onda de la señal (Infineon Technologies AG, 2017). La evaluación matemática de este cambio de fase a lo largo de una secuencia de chirps consecutivos permite al sistema calcular con precisión la velocidad exacta del objetivo (Infineon Technologies AG, 2017).

Micro Doppler

El efecto micro doppler es un fenómeno que describe pequeños cambios de frecuencia causados por el movimiento de las partes móviles de un objetivo, aparte del movimiento relativo del objetivo con respecto al radar.

Dentro de las aplicaciones de los radares FMCW, existe una disciplina especializada enfocada en el análisis de las frecuencias micro doppler. El uso de esta tecnología representa un método avanzado para la clasificación de objetivos a partir de los movimientos secundarios o micro movimientos de sus componentes (Neda , Passafiume, Sadeghibakhi, Collodi, & Cidronali, 2023).

En un esquema FMCW, las partes móviles de un blanco inducen modulaciones periódicas adicionales sobre la señal de retorno principal. Estas firmas micro doppler contienen un volumen significativo de información respecto a las características intrínsecas del objetivo, razón por la cual poseen un uso extendido en el área del reconocimiento automático de blancos y en la monitorización de las actividades físicas humanas (Neda , Passafiume, Sadeghibakhi, Collodi, & Cidronali, 2023).

Estimación Espectral

El análisis espectral comprende todos los métodos de procesamiento que caracteriza el contenido en función de la frecuencia que transporta una señal medida.

Tiempo Frecuencia

La base de la estimación espectral es la Transformada de Fourier, la cual, de manera matemática relaciona una señal representada en el dominio del tiempo con su equivalente representación en el dominio de la frecuencia (Harun & Levent, 2026). Ya que en un escenario real la señal que se procesa posee características ruidosas o aleatorias, el proceso se transforma en una estimación ya que se tiene que evaluar un intervalo finito de la señal. La base fundamental de este análisis se lo conoce como periodo grama, es un método que evalúa la envolvente al cuadrado de los coeficientes de la transformada para estimar la distribución de la potencia de la señal a lo largo de un espectro continuo de frecuencia.

Transformada de Fourier de Tiempo Corto (STFT)

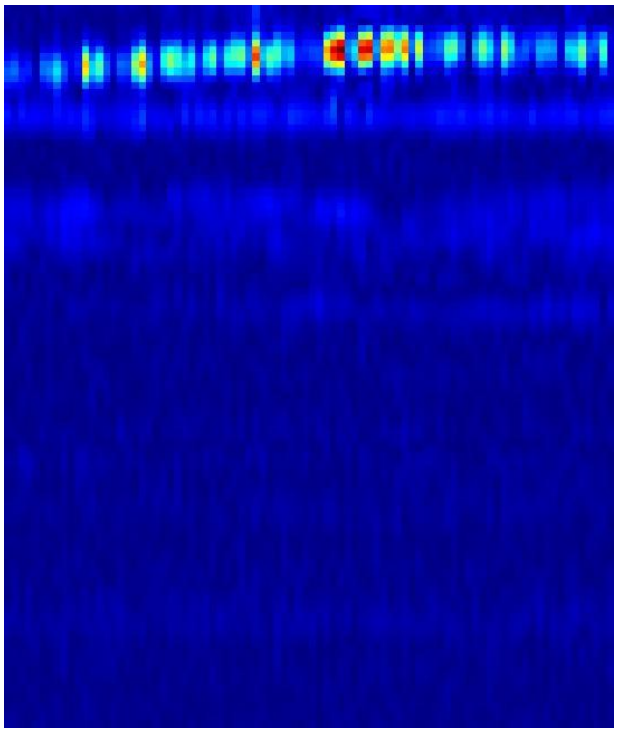
La extracción de firmas micro doppler a partir de señales de radar requiere herramientas matemáticas capaces de analizar variaciones espectrales dinámicas (Haoran, Tao, Gang, Yifan, & Yu, 2026). La Transformada de Fourier tradicional revela las frecuencias presentes en una señal, pero carece de resolución temporal para ubicar en qué instante exacto ocurre cada variación (Harun & Levent, 2026). Para solucionar esta limitación, el procesamiento de señales de radar FMCW emplea la STFT.

Esta técnica matemática segmenta la señal temporal cruda en múltiples ventanas superpuestas de corta duración a través de una función de enventanado (*windowing*) (Wickramarachi, 2003). Posteriormente, aplica la transformada a cada segmento de manera individual. Esta operación genera una representación en dos dimensiones denominada espectrograma, la cual expone la evolución simultánea del tiempo y la frecuencia doppler (Harun & Levent, 2026).

En el contexto del reconocimiento de objetivos, la STFT resulta fundamental. Gracias a esta transformación espacial, los micro movimientos sutiles de las partes móviles de un objetivo como el pedaleo en bicicletas, el balanceo de brazos en peatones o la rotación de neumáticos en vehículos se visualizan como patrones únicos de energía (Harun & Levent, 2026). Esta representación bidimensional transforma un problema complejo de análisis de ondas electromagnéticas en un problema de visión por computador, un formato estructuralmente idóneo para la clasificación autónoma a través de redes neuronales convolucionales (Andras, Jiaao, Julian F, & Dariu M, 2020).

La Figura 2.4 ilustra un espectrograma, el cual constituye la representación visual bidimensional de las firmas extraídas por el radar FMCW. En este formato, el eje horizontal corresponde a la evolución temporal de la captura, el eje vertical indica el desplazamiento de la frecuencia Doppler, y la escala de colores refleja la intensidad o potencia de la señal electromagnética reflejada.

Figura 2. 4 Espectrograma representativo de una señal micro-Doppler adquirida mediante radar FMCW



Procesamiento Espectral y Extracción del Desplazamiento Doppler

En la arquitectura de los sistemas digitales actuales, el procesamiento espectral se ejecuta comúnmente sobre señales de datos con valores complejos. Este tipo de señales es el resultado directo del proceso de demodulación en la recepción de los pulsos (Infineon Technologies AG, 2017). Al aplicar el análisis sobre estos segmentos finitos, cualquier movimiento relativo del objetivo genera un desplazamiento (*Doppler Shift*) que modifica la frecuencia de retorno (Harun & Levent, 2026). En la representación espectral compleja, estos desplazamientos se visualizan de manera integral como frecuencias positivas (movimiento de acercamiento) o negativas (movimiento de alejamiento) como se observa en Ec. 4.

$$fd = \frac{2 \cdot v \cdot f_0}{c}, \qquad (2.4)$$

donde: fd es el desplazamiento de frecuencia Doppler en Hz, $\mathcal{V}$ representa la velocidad relativa del objetivo en m/s y f_0 representa la frecuencia central de la portadora (Infineon Technologies AG, 2017).

Aprendizaje Profundo

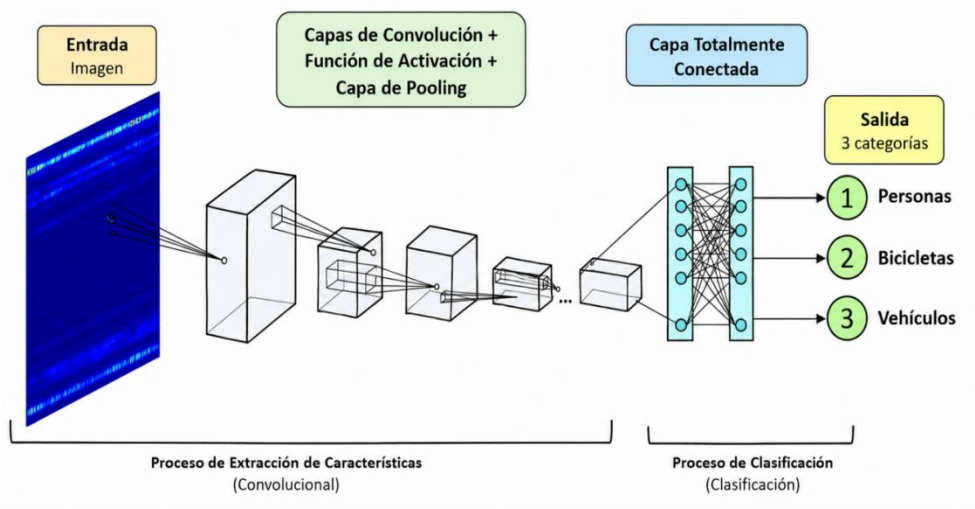
El deep learning constituye un subcampo de la inteligencia artificial que capacita a los sistemas computacionales para el procesamiento y reconocimiento de patrones complejos en imágenes, textos y sonidos a partir de grandes volúmenes de datos (MathWorks). El fin primordial de estos modelos es el suministro de predicciones con un alto grado de exactitud ante escenarios del mundo real (Andras, Jiaao, Julian F, & Dariu M, 2020).

Estas estructuras se fundamentan en arquitecturas de redes neuronales artificiales inspiradas en la corteza cerebral humana (MathWorks). Una red neuronal se compone de unidades de procesamiento denominadas nodos, los cuales se organizan e interconectan a través de una estructura de capas como se muestra en la Figura 2.5. La presencia de múltiples capas ocultas entre la entrada y la salida del sistema es el factor que otorga el término "profundo" a esta tecnología (MathWorks).

La categorización de los modelos de aprendizaje profundo depende estrictamente de la topología de la red implementada; entre las principales configuraciones se encuentran las CNN (Qiwei, Ruixun, & XiuLi, 2019) y las Redes Neuronales Recurrentes (RNN, del inglés Recurrent Neural Networks).

En la Figura 2.5 se describen las capas pertenecientes a una CNN, base principal para clasificación

Figura 2. 5 Estructura general de una CNN diseñada con 3 categorías de clasificación



CNN y su uso en la clasificación de Micro Doppler

Una estructura CNN efectúa operaciones de convolución sobre los datos de entrada para la extracción automática de características jerárquicas mediante el uso de capas convolucionales bidimensionales. Por consiguiente, esta arquitectura demuestra una alta eficiencia en el procesamiento de matrices de datos en dos dimensiones, tales como imágenes digitales (MathWorks).

En el contexto del análisis de señales FMCW, las firmas micro doppler se representan en el dominio del tiempo y la frecuencia mediante la STFT, lo cual genera un espectrograma bidimensional (Harun & Levent, 2026). Las CNN aprovechan su capacidad de análisis espacial para tratar a estos espectrogramas como imágenes. La red extrae patrones visuales únicos presentes en las trazas temporales de la frecuencia, lo que permite la discriminación y clasificación automatizada de diferentes tipos de objetivos en movimiento (Andras, Jiaao, Julian F, & Dariu M, 2020), tales como peatones, bicicletas y vehículos.

Arquitectura de Redes: AlexNet, SqueezeNet y Resnet-18

El diseño de arquitecturas convolucionales evoluciona de manera constante hacia la optimización del rendimiento y el uso eficiente de los recursos computacionales. Tres modelos prominentes con aplicaciones directas en la clasificación de firmas de radar son AlexNet, SqueezeNet y ResNet-18.

La arquitectura de AlexNet se compone estructuralmente de más de 20 capas operacionales secuenciales. Sin embargo, en términos de aprendizaje de características, la red se caracteriza por poseer 8 capas con parámetros entrenables: cinco capas convolucionales encargadas de la extracción de características espaciales, seguidas de tres capas densas (*fully connected*) (Md Zahangir , y otros, 2018). Este modelo introdujo innovaciones críticas en el aprendizaje profundo, tales como el uso de la función de activación por Unidad Lineal Rectificada (ReLU, del inglés *Rectified Linear Unit*) y la técnica de regularización por abandono de nodos (*dropout*) para mitigar el sobreajuste (Md Zahangir , y otros, 2018). Al aplicarse sobre espectrogramas micro-Doppler, AlexNet exhibe una alta capacidad para la extracción de características complejas de la señal, aunque demanda una carga computacional y un almacenamiento de memoria elevados debido a su gran cantidad de parámetros (Md Zahangir , y otros, 2018).

Por otra parte, la arquitectura de SqueezeNet se compone estructuralmente de más de 60 capas operacionales secuenciales. Sin embargo, en términos de aprendizaje de características, la red se caracteriza por poseer 26 capas con parámetros entrenables: veintiséis capas convolucionales encargadas de la extracción de características espaciales, prescindiendo por completo de capas densas (*fully connected*) (Hassanpour & Malek, 2019). Este modelo introdujo innovaciones críticas en el aprendizaje profundo orientadas a la eficiencia, tales como la implementación del módulo *Fire*, el cual se compone de una capa de compresión (*squeeze*) dotada únicamente de filtros de dimensión 1x1, seguida de una capa de expansión (*expand*) que combina filtros de 1x1 y 3x3 (Hassanpour & Malek, 2019). Al aplicarse sobre espectrogramas micro-Doppler generados por sistemas de radar FMCW, SqueezeNet exhibe un rendimiento robusto en la categorización de las firmas espectrales, garantizando una precisión equivalente a la de AlexNet pero reduciendo el número de parámetros de forma notable, lo que facilita su despliegue en plataformas de hardware con recursos limitados (Hassanpour & Malek, 2019).

Finalmente, La arquitectura de ResNet-18 se compone estructuralmente de más de 70 capas operacionales secuenciales. Sin embargo, en términos de aprendizaje de características, la red se caracteriza por poseer 18 capas con parámetros entrenables: diecisiete capas convolucionales encargadas de la extracción de características espaciales, seguidas de una única capa densa (*fully connected*) (Bhogal & Singh , 2025). ResNet-18 resolvió uno de los grandes problemas del aprendizaje profundo mediante las conexiones de salto (skip connections). Este mecanismo inyecta la señal original saltando capas intermedias para evitar que el gradiente se desvanezca a medida que la red se entrena (Bhogal & Singh , 2025). Gracias a sus 18 capas operacionales, la arquitectura presenta una relación excelente entre consumo de recursos y precisión. Por este motivo, se adapta perfectamente a la clasificación de nuestras firmas micro-Doppler. Al procesar los espectrogramas, el modelo distingue rápidamente las dinámicas de los vehículos frente a bicicletas o peatones, asegurando una convergencia veloz (Bhogal & Singh , 2025).

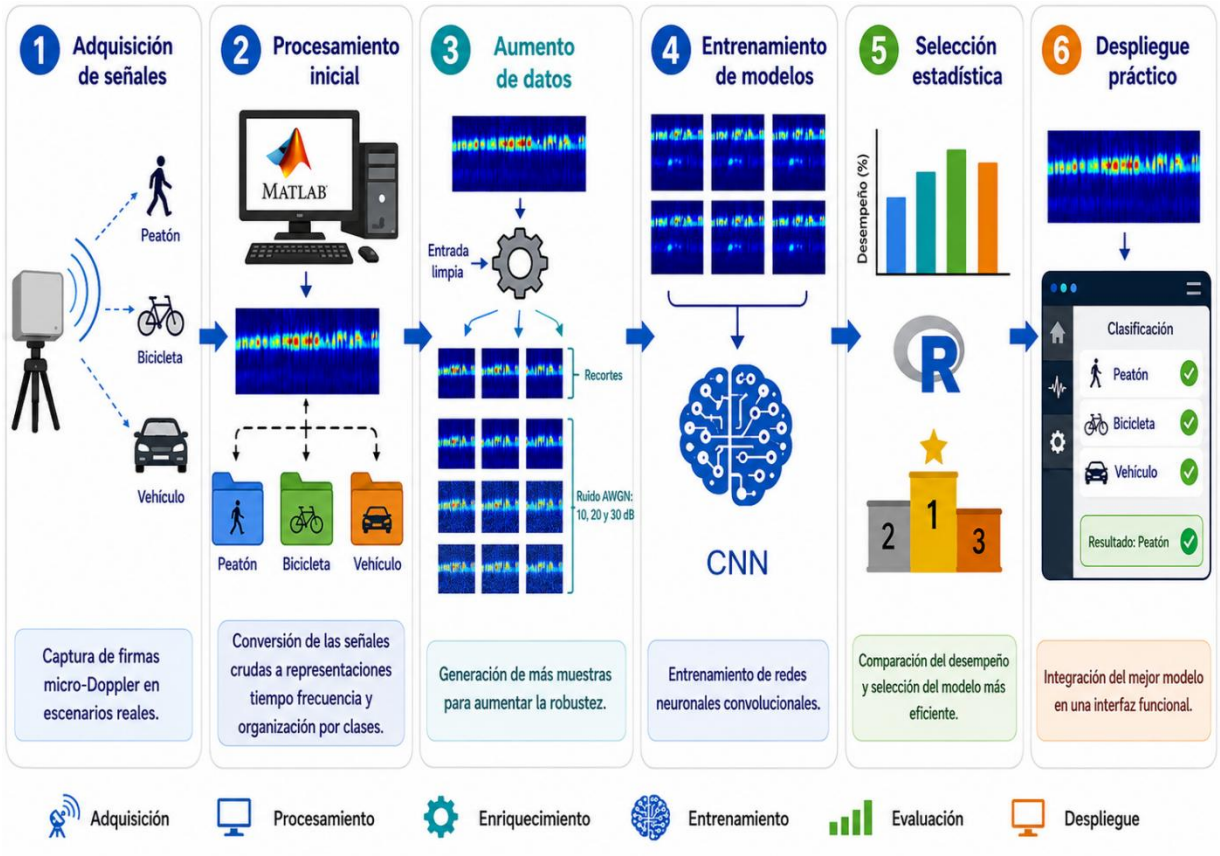
Capítulo III

Metodología

Para cumplir con el objetivo de la clasificación de objetivos en función de las componentes micro doppler, la sección presenta la metodología utilizada para la adquisición de señales, procesamiento inicial para la clasificación de imágenes, aumento de datos, entrenamiento de distintos modelos de deep learning, selección estadística y la integración de una interfaz, siguiendo el diagrama de flujo presentado en la Figura 3.1.

Figura 3. 1

Diagrama de bloques de la metodología utilizada



Adquisición de señales

Para la adquisición de señales se planteó un solo escenario con tres clases diferentes: de señales: personas, vehículos y bicicletas, dentro del cual definió un protocolo

de pruebas, para garantizar la generalización de los datos, manteniendo distancia y tiempos homogéneos.

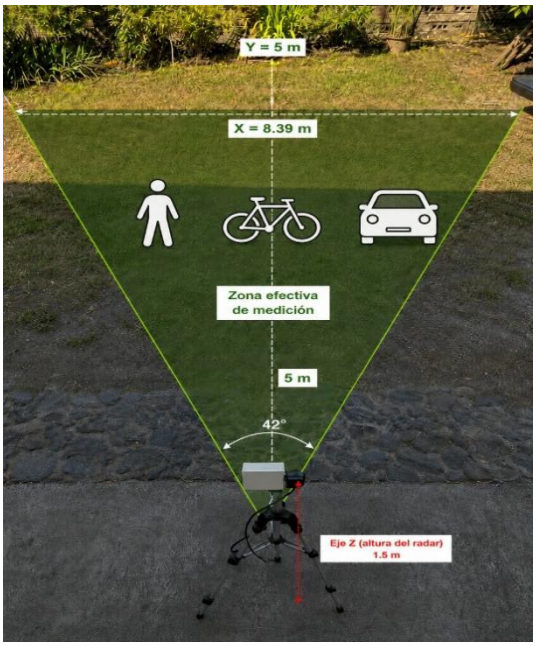
Esta etapa comprende la captura de las firmas micro doppler en escenarios reales mediante el uso del modulador de radar Distance2Go.

Protocolo de medición

El radar se posicionó de manera estática a una altura constante. Los objetivos transitaban a lo largo y ancho de un perímetro delimitado dentro del campo de visión de las antenas como se muestra en la Figura 3.2. Para garantizar un conjunto de datos robusto, se adquirieron 300 señales por cada tipo de objetivo. Las trayectorias incluyeron tanto aproximaciones frontales (orientación radial directa) como cruces transversales al radar, con el fin de capturar la variación de la RCS y el vector de velocidad relativa a lo largo del tiempo.

Figura 3. 2

Área definida por el campo de visión del radar dentro del entorno de trabajo



La inclusión específica de peatones, bicicletas y vehículos en el plan de medición tiene como objetivo principal capturar y analizar la marcada distinción en sus firmas micro doppler. Mientras que el efecto Doppler principal proporciona la velocidad de traslación del

centro de masa del objetivo, la cinemática interna de cada entidad genera modulaciones de frecuencia secundarias únicas:

Peatones (Dinámica articulada): Su huella en el espectrograma es altamente cíclica y compleja. El radar rastrea el avance del torso, así como también el balanceo constante de brazos y piernas. Este movimiento alternado generado por diferentes velocidades causa que existan diferencias en frecuencia lo que provoca una mayor diversidad en firmas micro doppler.

Bicicletas (Dinámica semi-articulada): Arrojan una firma híbrida. Por un lado, el cuadro de la bicicleta produce un desplazamiento Doppler constante. Por el otro, el giro de los pedales y los radios de las ruedas suman componentes rotacionales. El resultado visual es un ensanchamiento espectral marcadamente asimétrico donde existe una mayor reflexión aumentando la intensidad de potencia y con un reducido número de firmas micro doppler.

Vehículos (Cuerpo rígido): Gracias a su alta reflectividad (RCS), actúan prácticamente como un reflector puntual masivo. Su firma Doppler es intensa, sólida y muy concentrada. Aquí las micro-modulaciones son casi inexistentes; el sistema solo llega a captar pequeños ecos aislados de las llantas si el chasis no las cubre por completo.

Al recolectar 300 muestras por objetivo en distintos entornos, se garantiza una base de datos con suficiente varianza estadística para que modelos computacionales avanzados puedan extraer y aprender estas características espacio-temporales únicas para su posterior clasificación.

Diseño del escenario

El diseño del escenario experimental se planificó para capturar a los objetivos bajo distintas condiciones de propagación, asegurando la diversidad necesaria para el posterior análisis y clasificación de las señales.

Las pruebas de recolección de datos se dividieron en dos entornos principales para evaluar la robustez del sistema frente al ecos no deseados (*clutter*):

Entorno de línea de vista directa: Un campo abierto sin presencia de árboles, paredes o estructuras reflectantes cercanas, lo que establece una línea base con un nivel de ruido de fondo mínimo.

Entorno con obstáculos: Zonas con presencia de árboles y vegetación, introduciendo dispersión multietapa (*multipath*) y variaciones en la atenuación de la señal.

Procesamiento Inicial

Esta etapa consiste en la transformación de las señales crudas digitalizadas en matrices bidimensionales de rango versus tiempo (espectrogramas), con el objetivo de estructurar el conjunto de datos para la posterior fase de entrenamiento. La clasificación y organización de las muestras en las categorías de estudio (Automóvil, Persona y Bicicleta) se efectúa de forma sistemática en este punto mediante la modificación manual de las rutas de los directorios de salida en el script de procesamiento.

Calibración Estática y Escalamiento de la Señal

Con el propósito de suprimir el *clutter* estático y los reflejos no deseados del entorno (componente de corriente continua o DC), el programa inicia el proceso de extraer el primer cuadro de datos (*frame*) de la medición y calcula el perfil promedio de sus ráfagas (*chirps*) a través de la segunda dimensión después, este vector de calibración estática se resta de cada uno de los *frames* subsecuentes de la medición. Posterior a esta sustracción, la amplitud de la señal se amplifica mediante un factor de escala de IF equivalente a 52.8. Finalmente, se aplica una sustracción de la media por cada ráfaga empleando una función la cual asegura la remoción de cualquier desfase residual de DC antes de iniciar el análisis espectral.

Transformada de Fourier para Rango (Range-FFT) y Detección de Objetivos

Para la conversión de las señales del dominio del tiempo al dominio del rango por software, el algoritmo implementa los siguientes pasos:

Ventaneo de Rango: Se aplica una función de ventana Blackman de longitud igual al número de muestras por ráfaga, multiplicada por un factor de escala de 2, con el fin de atenuar los lóbulos secundarios del espectro.

El primer paso del procesamiento consiste en aplicar una Transformada Rápida de Fourier (FFT) de 256 puntos a lo largo de las muestras adquiridas. Los resultados de esta operación no se descartan, sino que se van acumulando hasta formar una matriz de datos tridimensional. Una vez estructurada esta información, el sistema pasa a localizar los objetivos. Para lograrlo, el algoritmo evalúa la dimensión de los *chirps* y extrae únicamente su valor absoluto máximo. Es precisamente este perfil de rango depurado el que sirve como entrada para la función detectora de picos. Dicha función está configurada con un umbral estricto de amplitud de rango igual a 80, una restricción de búsqueda de hasta 2 objetivos simultáneos y límites de distancia operativa entre un mínimo de $0.5\ m$ y un máximo de $10.0\ m$.

Procesamiento Doppler

En los casos donde la función de búsqueda confirma la existencia de objetivos, el código procede al aislamiento de sus frecuencias doppler. Para lograrlo, primero se resta la media doppler en los índices de rango detectados con el fin de resaltar los componentes en movimiento. El acondicionamiento de la señal inicia aplicando una ventana de Blackman, cuya longitud se ajusta al número de ráfagas escalado al doble. Una vez filtrados los datos, el algoritmo procesa una STFT de 32 puntos. A este cálculo se le suma un desplazamiento espectral con un objetivo claro: garantizar que la frecuencia Doppler cero quede perfectamente centrada. El último filtro es de magnitud. Cuando el pico espectral logra romper el umbral Doppler de 50, el sistema toma ese índice y lo convierte de inmediato a velocidad lineal (m/s) aplicando la constante espectral correspondiente.

Normalización, Segmentación Energética y Recorte de Imágenes

Una vez procesados todos los cuadros que componen la adquisición completa (cuya duración nominal es de 5.0 segundos), el algoritmo genera dos tipos de representaciones visuales estandarizadas para la base de datos (*dataset)*:

Espectrograma Original con Recorte Espacial (Y): Se genera la matriz de rango versus tiempo a partir del valor absoluto máximo de la Transformada de Fourier acumulada. Con el fin de optimizar el espacio de características y eliminar zonas de largo alcance sin información útil, se aplica un recorte vertical en el eje Y. Este proceso conserva únicamente el primer 40% de las filas de rango.

Espectrograma de Máxima Actividad de 2 Segundos (Recorte X + Y): Con el objetivo de extraer el intervalo temporal con mayor dinamismo del objetivo dentro de los 5 segundos totales, se desarrolla un algoritmo de ventana móvil:

Inicialmente, la matriz original se normaliza a una escala de grises mediante una función que transforma a escala de grises cualquier matriz de imágenes y se suaviza a través de un filtro gaussiano con una desviación estándar de 1, lo cual estabiliza la medición frente a picos de ruido aleatorio. El procesamiento continúa con el uso de una ventana móvil. Esta herramienta evalúa la matriz para calcular la energía total (la suma de los píxeles al cuadrado) en bloques equivalentes a 2.0 segundos esto permite detectar en la imagen los dos segundos que presentan mayor energía mostrando el tiempo donde se capturó al objetivo. En términos prácticos, esto abarca el 40% de las columnas del eje temporal. Con esta métrica, el algoritmo rastrea la zona de mayor actividad: detecta el índice de inicio con la máxima energía acumulada y extrae ese bloque específico directamente de la matriz cruda sin filtrar. Una vez aislado este segmento de 2 segundos, se ejecuta una reducción espacial aplicando un recorte vertical del 40% a lo largo del eje de rango (Y).

Exportación del Conjunto de Datos

Ambas representaciones espectrales (el registro de 5 segundos con recorte en Y, y el segmento de 2 segundos de máxima energía con recorte en X e Y) se grafican de forma invisible para el usuario a través de una figura desactivada. Las imágenes se configuran bajo el mapa de colores *jet*, con la supresión absoluta de ejes, marcos o márgenes blancos (*axis image off* y posición del eje fijada en $[0\ 0\ 1\ 1]$). Finalmente, las matrices gráficas limpias se exportan en formato PNG hacia las carpetas correspondientes de la clase seleccionada manualmente.

Aumento de Datos

El entrenamiento de modelos de aprendizaje profundo requiere un conjunto de datos suficientemente amplio y diverso para que la red neuronal sea capaz de aprender patrones representativos de cada clase y generalizar adecuadamente frente a nuevos datos.

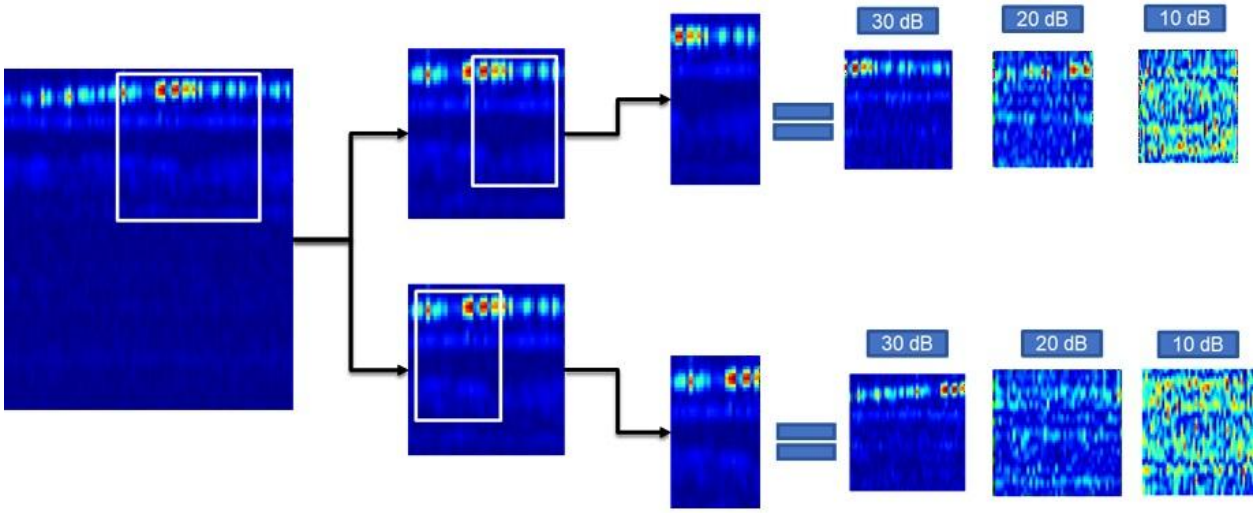
En aplicaciones de radar, la adquisición de un gran número de mediciones puede resultar limitada debido al tiempo requerido para realizar las mediciones de cada uno de los objetivos, la disponibilidad del equipo y las condiciones de experimentación.

En este trabajo, el conjunto de datos (300 señales por objetivo) era limitado para el entrenamiento de redes neuronales convolucionales (CNN) con arquitecturas profundas, tales como AlexNet, SqueezeNet o ResNet-18, se implementó una etapa de aumento de datos (llegando hasta 1600 espectrogramas por objetivo). Esta estrategia tuvo el propósito de incrementar la cantidad de ejemplos disponibles sin la necesidad de ejecutar nuevas adquisiciones experimentales.

El objetivo principal de esta etapa fue generar nuevas muestras que conservaran las características fundamentales de las firmas micro-doppler y, al mismo tiempo, añadieran variaciones como se muestra en la Figura 3.3. Estas técnicas permiten mejorar la capacidad de generalización del modelo y disminuir el riesgo de sobreajuste durante la fase de entrenamiento, un enfoque ampliamente validado en la clasificación de señales de radar mediante *deep learning*.

Figura 3. 3

Aumento de datos donde se selecciona la zona con mayor potencia del espectrograma y añadiendo ruido blanco Gaussiano



Selección de la Técnica de Aumento de Datos

En los problemas tradicionales de clasificación de imágenes dentro del campo de la visión por computadora, las técnicas de aumento de datos consisten habitualmente en transformaciones geométricas tales como rotaciones, escalamientos, traslaciones o reflexiones. Sin embargo, estas operaciones modifican la representación espacial de la matriz y alteran de forma crítica la estructura física de las firmas micro Doppler (Gurbuz & Amin, 2018)**.** Debido a que un espectrograma es una representación tiempo-frecuencia donde los ejes poseen una correspondencia física estricta el eje horizontal representa la evolución temporal y el eje vertical el desplazamiento Doppler, cualquier transformación geométrica directa sobre la imagen destruiría la relación cinemática del objetivo.

Por este motivo, se optó por ejecutar el aumento de datos directamente en el dominio del tiempo sobre la señal digitalizada recibida, de manera previa a la generación de las representaciones de tiempo-frecuencia. Este enfoque asegura que las muestras sintéticas generadas preserven un comportamiento físicamente consistente con las condiciones reales de adquisición del sistema de radar.

Para cumplir con este propósito, se implementó la adición de Ruido Blanco Gaussiano Aditivo (AWGN, del inglés Additive White Gaussian Noise), una técnica fundamental en el procesamiento digital de señales para modelar el ruido térmico e inherente a los sistemas electrónicos y de comunicaciones (Proakis & Manlakis, 2007). Esta metodología permite simular diversos escenarios de Relación Señal a Ruido (SNR, del inglés Signal-to-Noise Ratio), lo cual diversifica el conjunto de datos y mejora la robustez del modelo ante entornos ruidosos, sin desvirtuar las características esenciales de las firmas micro Doppler de los diferentes objetivos.

Modelo de ruido AWGN

El ruido incorporado durante el proceso de aumento de datos corresponde a un modelo de ruido blanco gaussiano aditivo (AWGN), el cual se caracteriza por presentar una distribución normal de media cero y una densidad espectral de potencia constante en todas las frecuencias (Haykin, 2014).

Matemáticamente, la señal con ruido puede expresarse como en Ec. 5:

$$\chi_{AWGN}(n) = \chi(n) + \omega(n), \quad (3.1)$$

donde $\chi(n)$ representa la señal original adquirida por el radar, $\omega(n)$ corresponde a la componente de ruido blanco gaussiano, $\chi_{AWGN}(n)$ representa la nueva señal sintética utilizada para el entrenamiento.

La potencia del ruido se controla mediante la relación señal-ruido definida como la proporción entre la potencia de la señal y la potencia del ruido

$$SNR = \frac{P_{Señal}}{P_{ruido}}, \quad (3.2)$$

O, expresada en decibelios (dB)

$$SNR_{dB} = 10log_{10}\left(\frac{P_{Señal}}{P_{ruido}}\right), \quad (3.3)$$

El empleo de distintos valores de SNR permite la generación señales con diferentes niveles de degradación. Esta estrategia hace posible modelar escenarios de adquisición en los cuales las condiciones del entorno afectan la calidad de la información recibida, lo que a su vez incrementa la robustez de la red neuronal frente a interferencias del mundo real.

Para evitar que el modelo se sobreajuste a escenarios ideales, el conjunto de entrenamiento se diversificó inyectando tres relaciones de señal a ruido (SNR). Se seleccionó un margen de 10 dB para simular entornos de alta interferencia, 20 dB como un punto intermedio, y 30 dB para representar capturas con una señal mayoritariamente libre de ruido. Al exponer a la red neuronal a esta mezcla, el algoritmo se ve obligado a aislar la firma micro-Doppler sin depender de un espectrograma perfecto. Entrenar bajo estos diferentes grados de contaminación electromagnética es lo que realmente garantiza la robustez de las arquitecturas CNN al enfrentarse a las condiciones reales de operación (Seyfioğlu, Özbayoğlu, & Gürbüz, 2018).

Procedimiento de generación del conjunto aumentado

El procedimiento de aumento de datos se implementó en el entorno de desarrollo MATLAB y consistió en la generación automática de nuevas versiones para cada captura original. Inicialmente, se cargaron los archivos en formato MAT correspondientes a las mediciones del radar. Cada archivo contiene una variable, la cual almacena la secuencia completa de tramas (*frames*) adquiridas durante la medición.

Posteriormente, para cada uno de los niveles de SNR establecidos, se calculó la potencia del ruido correspondiente y se generó una matriz de ruido AWGN con distribución gaussiana. Este ruido se sumó de forma individual a cada trama contenida en el archivo. Mediante este procedimiento, se obtuvo una nueva versión de la señal con un nivel de degradación estrictamente controlado.

Para organizar el almacenamiento, los datos sintéticos se separaron en carpetas exclusivas según su nivel de SNR. Esta decisión logística aceleró enormemente el procesamiento computacional necesario para generar los espectrogramas.

Contribución del aumento de datos al entrenamiento de la CNN

La etapa de aumento de datos es el pilar que sostiene el entrenamiento del modelo. Su función es clara: multiplicar la diversidad de las muestras sin deformar la física real del radar. Al someter a la red neuronal a ruido artificial controlado, se emulan las caídas de señal provocadas por el entorno o la propia electrónica del hardware. Esto obliga al algoritmo a rastrear la firma micro-Doppler a pesar de las perturbaciones, blindando al clasificador contra el sobreajuste (*overfitting*) y validando la viabilidad de la red profunda (Shorten & Khoshgoftaar , 2019)**.** De este modo, se contribuye a la obtención de un sistema de clasificación más robusto, capaz de exhibir un mejor desempeño operacional al ser evaluado con conjuntos de datos completamente nuevos y externos al entrenamiento (Goodfellow, Bengio, & Courville, 2016)**.**

Entrenamiento de Modelos

El objetivo central de esta fase consistió en entrenar modelos de aprendizaje profundo utilizando las representaciones tiempo-frecuencia. La meta era lograr que el

algoritmo discriminara automáticamente las firmas correspondientes a peatones, bicicletas y vehículos.

Para optimizar los recursos, se descartó la construcción de una red desde cero. En su lugar, se implementó aprendizaje por transferencia (*Transfer Learning*). Esta técnica es comúnmente usada cuando no se tiene una gran diversidad de datos en un *dataset*; al heredar pesos previamente entrenados, el modelo acelera su convergencia y reduce drásticamente los tiempos de computación.

Bajo este enfoque, se pusieron a prueba tres arquitecturas específicas:

AlexNet: Sirvió como línea base del experimento. Al tener una estructura clásica y poco profunda, permitió establecer el rendimiento mínimo del sistema.

SqueezeNet: Su elección responde a criterios de hardware. Gracias a su bajo número de parámetros, resulta la candidata ideal para futuros despliegues del proyecto en sistemas embebidos.

ResNet-18: Se integró para evaluar el impacto de sus conexiones residuales, las cuales evitan que el gradiente se desvanezca.

Para garantizar un contraste riguroso, las tres redes operaron bajo restricciones idénticas. Se congelaron los mismos parámetros experimentales para evaluar posteriormente su desempeño frente a frente mediante métricas estandarizadas.

.

Entrenamiento mediante AlexNet

AlexNet es una red neuronal convolucional propuesta en 2012, considerada una de las arquitecturas fundamentales que impulsó el desarrollo del aprendizaje profundo en tareas de clasificación de imágenes (Wang & Rawat, 2017)**.** Su estructura consta de cinco capas convolucionales seguidas de tres capas completamente conectadas. Esta configuración topológica permite aprender características jerárquicas, desde patrones visuales simples en las primeras etapas hasta representaciones abstractas de mayor nivel en las capas más profundas.

Debido a que el modelo base fue entrenado originalmente con el conjunto de datos masivo ImageNet, en este trabajo se empleó la estrategia de aprendizaje por transferencia (*Transfer Learning*). Mediante este enfoque, se extraen las características previamente asimiladas por la red y se adaptan únicamente las últimas capas de clasificación. De este modo, la arquitectura queda optimizada para discriminar las firmas micro-Doppler correspondientes a personas, bicicletas y vehículos, sin la necesidad de un entrenamiento exhaustivo desde cero.

Adaptación de la arquitectura

La arquitectura original de AlexNet fue diseñada para discriminar entre 1000 categorías diferentes. Sin embargo, el problema abordado en esta investigación contempla únicamente tres clases: peatón, bicicleta y vehículo. Por esta razón, las tres últimas capas de la topología original se reemplazaron por nuevas estructuras adaptadas a la dimensionalidad del conjunto de datos empleado.

Las capas sustituidas corresponden a:

Capa completamente conectada (*Fully Connected Layer*).

Capa de activación (*Softmax Layer*).

Capa de salida para clasificación (*Classification Output Layer*).

La nueva capa completamente conectada se configuró con tres neuronas de salida, una por cada categoría del problema. Posteriormente, las activaciones resultantes se normalizaron mediante una función Softmax y, finalmente, las predicciones probabilísticas se obtuvieron a través de una capa de clasificación estándar.

Preparacion del conjunto de entrenamiento

Las representaciones gráficas de las firmas micro Doppler se organizaron en directorios independientes en función de la clase a la que pertenecían (peatones, bicicletas y vehículos). El entorno MATLAB identificó automáticamente las etiquetas de cada captura a partir de la jerarquía de las carpetas mediante la creación de un objeto. Esta funcionalidad facilitó la gestión eficiente del conjunto de datos durante la fase de entrenamiento.

Debido a que la arquitectura AlexNet exige que los datos de entrada posean dimensiones estrictas de 227 × 227 × 3 píxeles, todas las imágenes se redimensionaron de forma automática a través de una función capas de redimensionar cualquier imagen que tenga de entrada. Este procedimiento garantizó el ajuste espacial adecuado de las matrices, al mismo tiempo que preservó la integridad de la información espectral necesaria para el correcto aprendizaje de la red neuronal.

Entrenamiento Mediante SqueezeNet

SqueezeNet es una red neuronal convolucional propuesta en 2016 con el objetivo de obtener un desempeño comparable al de arquitecturas más profundas, a través de una reducción significativa en el número de parámetros del modelo. Esta reducción en memoria y tiempo de cómputo es el mayor atractivo del modelo. En la práctica, SqueezeNet logra operar en hardware restringido. El funcionamiento de estos bloques es muy particular. En lugar de procesar todo de golpe, la información ingresa primero a una capa de compresión (*Squeeze Layer*) equipada exclusivamente con filtros de 1×1 para alivianar la carga de parámetros. Una vez comprimidos, los datos pasan a la etapa de expansión (*Expand Layer*). Es precisamente esta eficiencia interna la que la convierte en la opción ideal para sistemas embebidos.

Esta topología permite disminuir la cantidad de filtros convolucionales sin afectar de manera drástica la capacidad de extracción de características del modelo (Lin, Chen, & Yan, 2013)**.**

La arquitectura base fue entrenada mediante el conjunto de datos masivo ImageNet, el cual está compuesto por más de un millón de imágenes distribuidas en 1000 categorías. En esta investigación se implementó la técnica de aprendizaje por transferencia (*Transfer Learning*), la cual aprovecha los pesos sinápticos previamente ajustados y adapta de forma exclusiva las capas finales al problema específico de clasificación de firmas micro-Doppler.

Adaptación de la arquitectura

La versión original de SqueezeNet finaliza con una capa convolucional, cuya salida corresponde a las 1000 categorías de ImageNet. Debido a que el problema abordado en

esta investigación considera únicamente tres clases (peatones, bicicletas y vehículos), fue necesario modificar la arquitectura mediante la sustitución de las capas finales.

En particular, la capa convolucional de salida original de SqueezeNet se sustituyó por una nueva capa convolucional de tamaño 1 × 1, configurada con tres filtros de salida, correspondientes al número de clases del conjunto de datos. Posteriormente, la capa de clasificación original se reemplazó por una nueva estructura adaptada a las etiquetas establecidas para el entrenamiento.

Esta modificación permitió conservar la robusta capacidad de extracción de características asimilada durante el entrenamiento base de SqueezeNet. De este modo, se adaptó de forma exclusiva la etapa algorítmica encargada de ejecutar la clasificación final, preservando la eficiencia paramétrica del modelo.

Preparación del conjunto de datos

El conjunto de imágenes se organizó mediante la creación de un objeto en el entorno de MATLAB. Esta herramienta computacional hizo posible asociar automáticamente las etiquetas de cada muestra a partir de la jerarquía de carpetas correspondiente a cada clase (peatones, bicicletas y vehículos).

Posteriormente, todas las matrices visuales se redimensionaron a 227 × 227 × 3 píxeles el formato espacial de entrada requerido de manera estricta por la arquitectura SqueezeNet a través de la misma función utilizada en AlexNet para el redimensionamiento. Este procedimiento garantizó la compatibilidad estructural entre el conjunto de datos empíricos y la red neuronal, al mismo tiempo que preservó la integridad de la información espectral contenida en las representaciones de tiempo-frecuencia (Chen & Wang, 2014).

Entrenamiento mediante ResNet-18

ResNet-18 es una red neuronal convolucional profunda desarrollada como parte de la familia *Residual Networks* (ResNet). Su principal aporte radica en la incorporación de conexiones residuales (*Skip Connections*). Estas conexiones habilitan un flujo directo de

información entre capas no adyacentes, una estrategia arquitectónica que mitiga el problema de la degradación del aprendizaje inherente a los modelos de gran profundidad.

La estructura comprende dieciocho capas parametrizables distribuidas en bloques residuales, donde cada bloque modela exclusivamente la función residual con respecto a su entrada. Este mecanismo algorítmico optimiza la retro propagación del gradiente durante la fase de entrenamiento. Como resultado, se asegura una convergencia más estable y se incrementa el rendimiento general en tareas complejas de clasificación visual (Goodfellow, Bengio, & Courville, 2016)**.**

De manera homóloga a los modelos AlexNet y SqueezeNet, la arquitectura base de ResNet-18 se entrenó con el masivo conjunto de datos ImageNet. Por consiguiente, en esta investigación se implementó el aprendizaje por transferencia (*Transfer Learning*). A través de esta técnica, se conservaron los pesos sinápticos asimilados previamente y se modificaron de forma exclusiva las capas terminales encargadas de la clasificación final de las firmas micro-Doppler.

Adaptación de la arquitectura

La arquitectura original de ResNet-18 finaliza con una capa totalmente conectada denominada *fully connected* de 1000 capas, cuya salida corresponde a las 1000 categorías presentes en ImageNet. Debido a que el problema planteado en esta investigación considera únicamente tres clases (peatón, bicicleta y vehículo), fue necesario modificar la etapa final de clasificación.

Para ello, la capa *fully connected* de 1000 capas fue reemplazada por una nueva capa totalmente conectada con tres neuronas de salida, correspondientes al número de clases del conjunto de datos. Posteriormente, la capa de clasificación original fue sustituida por una nueva capa adaptada a las etiquetas utilizadas durante el entrenamiento.

Mediante esta modificación fue posible conservar la capacidad de extracción de características aprendida por ResNet-18 durante su entrenamiento sobre ImageNet, ajustando únicamente la etapa responsable de asignar la clase correspondiente a cada representación tiempo-frecuencia.

Preparación del conjunto de Datos

Las representaciones de tiempo-frecuencia correspondientes a las firmas micro Doppler se organizaron mediante la misma función utilizada en las CNN anteriores, del entorno Matlab. Esta herramienta computacional hizo posible la asignación automática de las etiquetas a partir de la jerarquía de carpetas del conjunto de datos.

Posteriormente, todas las imágenes se redimensionaron a 224 × 224 × 3 píxeles el formato espacial de entrada requerido de manera estricta por la arquitectura ResNet-18 a través de la función de redimensionamiento ya utilizada. Este procedimiento aseguró la compatibilidad técnica entre las muestras empíricas y la red neuronal, preservando la información espectral de las representaciones.

Validación cruzada K-Fold

Con el objetivo de obtener una estimación rigurosa e imparcial del desempeño de las arquitecturas propuestas, se implementó una estrategia de validación cruzada de cinco particiones (*5- Fold Cross-Validation*). Esta técnica estandarizada se aplicó para todos los modelos evaluados en el estudio.

Para llevar a cabo este proceso, el conjunto de datos se dividió aleatoriamente en cinco subconjuntos iguales en tamaño. Durante cada iteración, cuatro de estos subconjuntos se emplearon para el entrenamiento de la red neuronal, mientras que el subconjunto restante se destinó a la validación. Este procedimiento sistemático se repitió en cinco ocasiones diferentes. Como resultado, la totalidad de las muestras participaron tanto en la fase de entrenamiento como en la fase de evaluación.

Finalmente, el desempeño global de cada modelo clasificador se cuantificó a través del cálculo de la media aritmética de las métricas obtenidas en las cinco iteraciones. Esta metodología metodológica minimiza el sesgo asociado a la selección aleatoria de un único conjunto de prueba y proporciona una estimación de generalización sumamente confiable frente a datos no observados (Hastie, Tibshirani, & Friedman, 2009)

Configuración del entrenamiento

El entrenamiento algorítmico se llevó a cabo mediante el optimizador de Descenso de Gradiente Estocástico con Momento (SGDM, por sus siglas en inglés: *Stochastic Gradient Descent with Momentum*). Este método numérico es ampliamente empleado en la optimización de redes neuronales profundas debido a su comprobada estabilidad y a la aceleración en la rapidez de convergencia (Sutskever, Martens, Dahl, & Hinton, 2013).

En cuanto a la configuración de los hiperparámetros computacionales, el tamaño del lote (*mini-batch size*) se fijó en 10 muestras por iteración. Paralelamente, el proceso de entrenamiento se desarrolló durante un total de seis épocas, aplicando una tasa de aprendizaje inicial (*learning rate*) de 5×10^{-5} como lo se observa en la Tabla 3.1.

Tabla 3. 1

Resumen de los parámetros de optimización y validación para las arquitecturas de aprendizaje profundo

Parámetro	Valor
Optimizador	SGDM
Mini batch Size	10
Épocas	6
Learning Rate	5×10^{-5}
Validación	K-Fold (5)
GPU	VRAM 10 GB RAM 16 GB

La Tabla 3.1 resume la configuración computacional establecida para la fase de entrenamiento de las redes neuronales convolucionales aplicadas a la clasificación de firmas del radar FMCW.

Con el objetivo de garantizar una convergencia estable, la optimización matemática se ejecutó a través del algoritmo de Descenso de Gradiente Estocástico con Momento (SGDM). La tasa de aprendizaje (*learning rate*) se fijó en un valor de 5×10^{-5}. Esta decisión metodológica resulta fundamental en esquemas de aprendizaje por transferencia (*Transfer Learning*), ya que un parámetro reducido protege los pesos sinápticos previamente ajustados e impide la degradación abrupta del conocimiento inicial del modelo.

El tamaño del lote (*mini-batch size*) se delimitó a 10 muestras por iteración. Esta dimensión optimiza el consumo de memoria y aporta un nivel adecuado de variabilidad estocástica al cálculo del gradiente. Paralelamente, el proceso completo consta de 6 épocas de exposición a los datos, un ciclo suficiente para alcanzar la adaptación de las capas finales de clasificación.

Finalmente, el entorno de MATLAB delegó toda la carga tensorial a una unidad de procesamiento gráfico (GPU). Esta aceleración por *hardware* constituye una medida indispensable para gestionar las intensivas iteraciones de cálculo y reducir de forma drástica los tiempos computacionales totales del proyecto.

Selección estadística

Una vez entrenadas las arquitecturas AlexNet, SqueezeNet y ResNet-18, fue necesario determinar si las diferencias observadas en su desempeño correspondían únicamente a variaciones propias del proceso de entrenamiento o si representaban diferencias estadísticamente significativas. Para ello, se realizó un análisis estadístico utilizando como métrica de comparación el F1-score macro obtenido en cada una de las cinco particiones de la validación cruzada (*5-Fold Cross Validation*).

La selección del F1-score macro como indicador principal se fundamenta en que esta métrica considera simultáneamente la precisión (*Precision*) y la sensibilidad (*Recall*) de cada clase, asignando el mismo peso a todas las categorías independientemente del número de muestras disponibles. De esta manera, proporciona una evaluación más equilibrada del desempeño de los clasificadores cuando existen múltiples clases.

Todo el análisis estadístico fue implementado en el software “R”, empleando pruebas no paramétricas adecuadas para datos pareados y un número reducido de observaciones.

Estadística descriptiva

Como primera etapa del análisis, se calculó las medidas descriptivas para cada arquitectura de red neuronal. En particular, se determinó el valor promedio y la desviación estándar del *F1-score* macro a partir de los cinco *folds* generados durante la validación cruzada.

Estas métricas facilitan una primera comparación del comportamiento de cada modelo, lo que permite evaluar tanto su desempeño promedio como la variabilidad de los resultados entre las diferentes particiones del conjunto de datos. Finalmente, estos hallazgos se organizan en una tabla resumen que sirve de base para la posterior comparación cuantitativa entre las arquitecturas.

Prueba Global de Friedman

Con el propósito de determinar si existen diferencias estadísticamente significativas entre las tres arquitecturas evaluadas, se aplica la prueba no paramétrica de Friedman (Demšar, 2006).

Esta prueba constituye una alternativa al análisis de varianza para medidas repetidas (ANOVA de medidas repetidas (Demšar, 2006)) cuando no puede garantizarse el cumplimiento del supuesto de normalidad. Además, resulta adecuada cuando las observaciones provienen de un mismo conjunto experimental evaluado bajo diferentes condiciones, como ocurre en este trabajo, donde los tres modelos se entrenaron y evaluaron a partir de exactamente las mismas particiones generadas mediante la validación cruzada de cinco *folds*.

Las hipótesis consideradas fueron:

H_0:Las tres arquitecturas presentan un desempeño estadísticamente equivalente.

H_1:Al menos una arquitectura presenta un desempeño significativamente diferente.

Donde H_0 es la hipótesis utilizada para Friedman y H_1 es la hipótesis utilizada para las pruebas de Wilcoxon.

El criterio de decisión se estableció con un nivel de significancia de $\alpha = 0.05$ (Demšar, 2006). De esta forma, cuando el valor p obtenido fue inferior al nivel de significancia, se rechaza la hipótesis nula y se concluye que existen diferencias estadísticamente significativas entre las arquitecturas evaluadas (Demšar, 2006).

Comparaciones múltiples mediante la prueba de Wilcoxon

En caso de obtener diferencias significativas mediante la prueba de Friedman, es necesario identificar específicamente entre qué arquitecturas existen dichas discrepancias. Para ello, se realizan comparaciones por pares mediante la prueba de rangos con signo de Wilcoxon para muestras pareadas (Japkowicz & Shah, 2011). Esta prueba es adecuada debido a que los resultados de cada arquitectura provienen de las mismas particiones de validación cruzada, lo cual implica que las observaciones no son independientes entre sí.

Se efectuaron las siguientes comparaciones por pares:

- AlexNet vs. SqueezeNet
- AlexNet vs. ResNet-18
- SqueezeNet vs. ResNet-18

Para cada comparación se calculó el correspondiente valor *p*, lo que permitió establecer si las diferencias observadas entre los modelos eran estadísticamente significativas. Adicionalmente, para controlar la inflación del error Tipo I debido a las comparaciones múltiples, se aplicó la corrección de Holm sobre los niveles de significancia (Demšar, 2006).

Correlación por comparaciones múltiples

Debido a que la realización de múltiples pruebas estadísticas incrementa la probabilidad de obtener resultados significativos por azar, se aplicó una corrección de los valores *p* mediante el procedimiento de Holm (Demšar, 2006).

Este método controla la tasa de error por familia (FWER, del inglés *Family-Wise Error Rate*) y constituye una alternativa menos conservadora que la corrección de Bonferroni (Holm, 1979), lo cual permite mantener un adecuado equilibrio entre el control del error tipo I y la potencia estadística. Finalmente, los valores *p* ajustados se establecieron como el criterio definitivo para determinar la existencia de diferencias significativas entre las arquitecturas evaluadas (Holm, 1979).

Selección del mejor modelo

Una vez finalizado el análisis inferencial, se identifica la arquitectura que alcanzó el mayor valor promedio de *F1-score* macro durante la validación cruzada. La selección de este modelo definitivo integró de forma simultánea dos criterios fundamentales:

- El desempeño métrico promedio obtenido a través del *F1-score* macro.
- La evidencia de significancia estadística proporcionada por los resultados de las pruebas de Friedman y Wilcoxon.

De esta manera, el modelo seleccionado se estableció como la arquitectura que demostró el mejor rendimiento global, cuyo desempeño se respalda por el análisis estadístico realizado.

Despliegue Practico

Una vez concluido el entrenamiento y la evaluación de las arquitecturas de redes neuronales convolucionales, se implementó un sistema capaz de realizar la clasificación automática de objetivos a partir de las señales adquiridas por el radar FMCW Distance2Go. Esta etapa integra todos los procedimientos desarrollados durante la investigación, desde la adquisición de la información radar hasta la identificación de la clase correspondiente al objetivo detectado.

El sistema fue desarrollado en MATLAB, permitiendo la comunicación con el radar, el procesamiento de las señales recibidas, la generación de las representaciones tiempo-frecuencia y la clasificación mediante la arquitectura CNN seleccionada.

Adquisición de la información radar

El proceso inicia con la adquisición de los datos provenientes del radar FMCW Distance2Go de Infineon. Durante la captura, el sensor registra la información correspondiente a las señales reflejadas por los diferentes objetivos presentes en el escenario, almacenando los datos en una secuencia de *frames* que contienen la información necesaria para realizar el procesamiento posterior.

Cada *frame* representa una medición instantánea del entorno y constituye la unidad básica empleada para construir la representación micro doppler utilizada durante la clasificación (Infineon Technologies AG, 2017).

Procesamiento de la señal

Una vez adquiridas las señales radar, cada uno de los *frames* es sometido a las etapas de procesamiento implementadas en MATLAB. En esta fase se realizan las operaciones necesarias para obtener la representación tiempo-frecuencia de la señal, permitiendo resaltar las variaciones micro doppler producidas por el movimiento de los diferentes objetivos.

El resultado de este procesamiento corresponde a una imagen que representa la firma micro doppler característica del objeto observado, la cual constituye la entrada del sistema de clasificación basado en redes neuronales convolucionales.

Preparación de la imagen para la clasificación

Preparación de la imagen para la clasificación

Antes de ser procesada por la red neuronal, la representación tiempo-frecuencia es acondicionada para cumplir con los requisitos de entrada de la arquitectura seleccionada. Este procedimiento incluye el redimensionamiento automático de la imagen a las dimensiones requeridas por la red, manteniendo las características relevantes de la firma micro Doppler.

Posteriormente, la imagen es convertida al formato adecuado para ser utilizada como entrada durante la etapa de entenamiento de la CNN.

Clasificación mediante la CNN seleccionada

Como resultado del proceso de evaluación estadística desarrollado en la etapa anterior, se eligió la arquitectura con mejor desempeño para la implementación final del sistema.

Durante la etapa de entrenamiento, la imagen correspondiente a la firma micro doppler es suministrada a la red neuronal previamente entrenada. A partir de las características aprendidas durante el proceso de entrenamiento, la CNN extrae automáticamente patrones espaciales presentes en la representación tiempo-frecuencia y determina la clase con mayor probabilidad de elegir la clase correcta.

El sistema fue configurado para realizar la clasificación entre tres categorías:

- Persona.
- Bicicleta.
- Vehículo.

Presentación del resultado

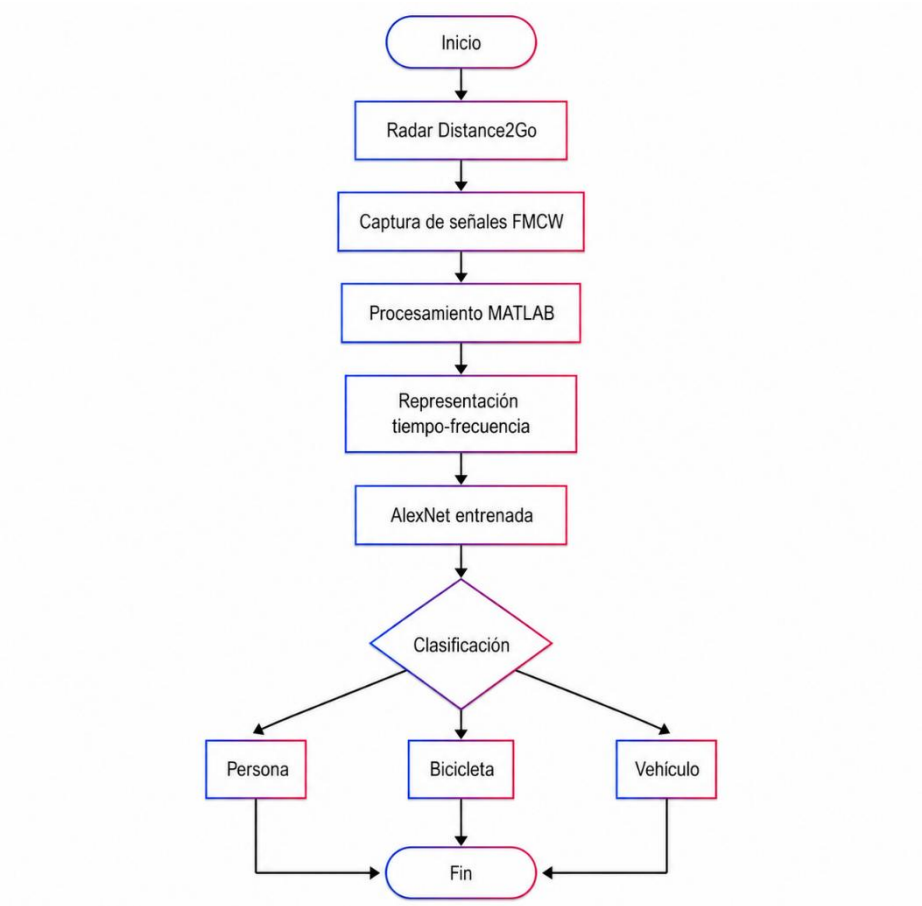
Finalmente, el sistema implementado en MATLAB muestra la categoría asignada por el clasificador junto al porcentaje de exactitud.

Esta implementación permite realizar el proceso completo de identificación de objetivos de manera automática, integrando la adquisición de datos, el procesamiento de señales y la clasificación mediante aprendizaje profundo dentro de una única aplicación.

Todo el proceso realizado en le despliegue practico se puede resumir en diagrama de flujo presentado en la Figura 3.4

Figura 3. 4

Diagrama de flujo del despliegue practico realizado



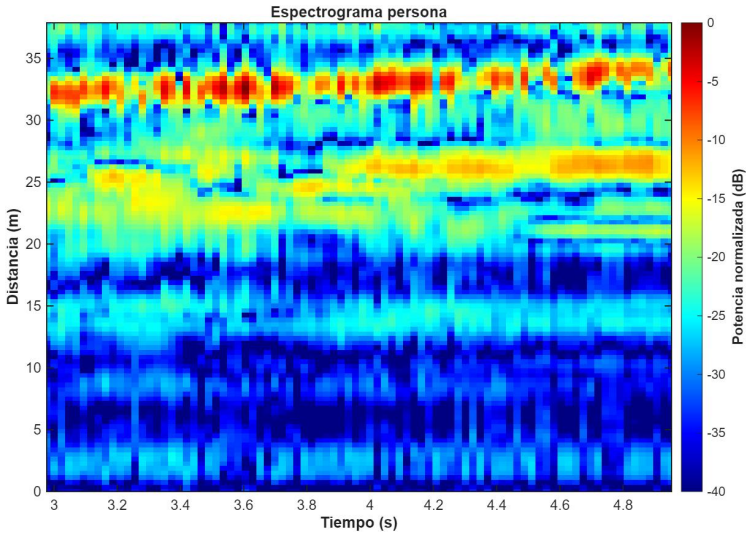
Capítulo IV

Análisis y discusión de los resultados

En el presente capítulo se exponen y evalúan los resultados obtenidos tras la implementación del sistema de clasificación multiclase de objetivos viales mediante el uso del radar FMCW. A partir de las representaciones tiempo-frecuencia de las firmas micro doppler como se muestra en las figuras 4.1, 4.2 y 4.3, mostrando los espectrogramas de personas bicicletas y vehículos. Posteriormente se aplicarán técnicas de aumento de datos (AWGN y tiempo), se analiza el desempeño cuantitativo de las arquitecturas de aprendizaje profundo seleccionadas: AlexNet, SqueezeNet y ResNet-18. La estructura que utilizamos para realizar el análisis empezó con la evaluación de las métricas obtenidas de cada una de las arquitecturas entrenadas, precisión, exactitud y F1 score, calculadas después de cada uno de los 5 folds realizados. Posteriormente, se analizaron las matrices de confusión para poder definir aciertos y errores que tuvo cada CNN en la clasificación de las tres clases evaluadas. Finalmente, se utilizaron métricas estadísticas, cola media y la desviación estándar (DE) para seleccionar el modelo que tuvo el mejor desempeño en la clasificación y utilizarlo en el despliegue práctico.

Figura 4. 1

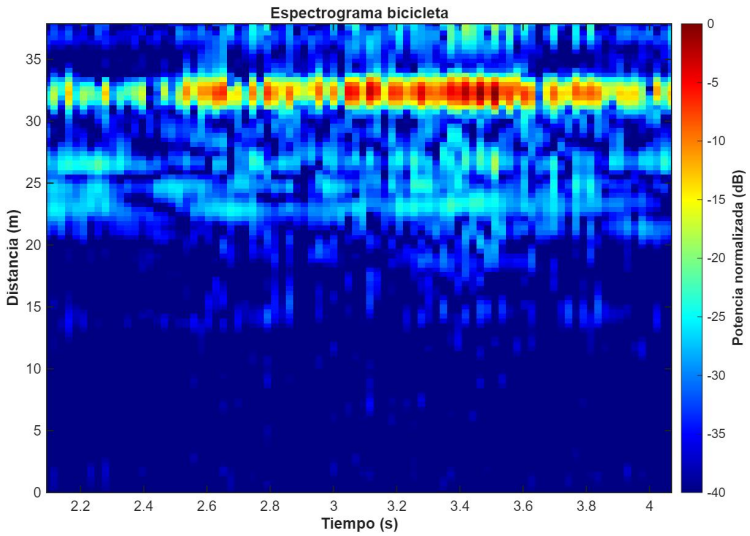
Espectrograma de persona



En la Figura 4.1 se observa el espectrograma correspondiente a personas, la cual muestra una gran cantidad de variaciones en frecuencia debido al movimiento de oscilación de brazos y piernas.

Figura 4. 2

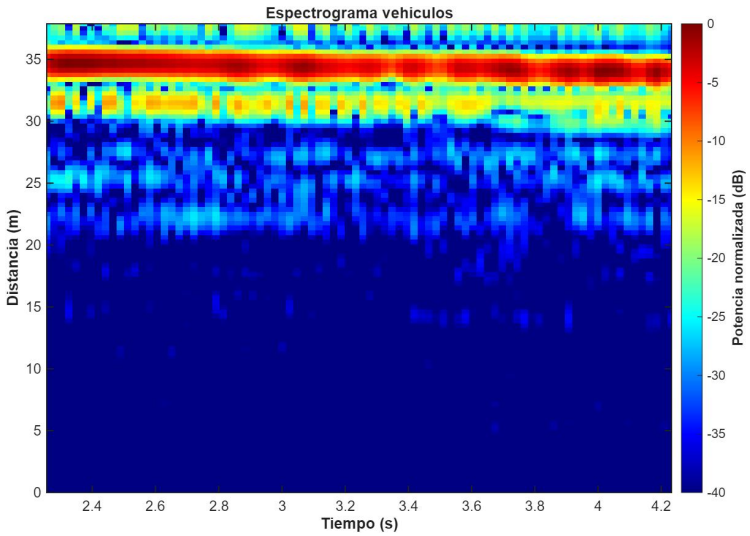
Espectrograma de bicicleta



La Figura 4.2 presenta el espectrograma perteneciente a bicicletas, esta muestra una intensidad más alta en energía, manteniendo un nivel reducido de variaciones en frecuencia o micro doppler, debido a los movimientos que existen en el pedaleo de la persona.

Figura 4. 3

Espectrograma de vehículo



Finalmente, la Figura 4.3 muestra el espectrograma de vehículos, el cual se caracteriza por tener un bloque solido de energía, en este caso se presenta una mina cantidad de firmas micro doppler debido a que el vehículo actúa como un reflector.

Entrenamiento de modelos de Deep Learning

Resultados con datos previos al procesamiento

En la Tabla 4.1 se ve que AlexNet presentó el mejor desempeño general con los datos previos al procesamiento. Esta arquitectura alcanzó los mayores valores promedio en las cuatro métricas evaluadas, con una exactitud de 91,17 % ± 1,77 % y un F1-score macro de 91,14 % ± 1,78 %. Además, registró 537 clasificaciones correctas de un total de 589 muestras, equivalente a una tasa de error de apenas 8,83 %.

ResNet-18 ocupó la segunda posición, con un F1-score macro promedio de 86,35 %, mientras que SqueezeNet presentó el menor desempeño, con 83,16 % ± 5,80 %. También se observa que SqueezeNet tuvo la mayor desviación estándar, lo que indica que su rendimiento varió considerablemente entre los folds. Por el contrario, AlexNet mostró un comportamiento más uniforme y estable durante la validación cruzada.

Tabla 4. 1

Desempeño general de las CNN

Arquitectura	Exactitud media ± DE	Precisión macro ± DE	Recall macro ± DE	F1 macro ± DE	Aciertos/total
AlexNet	91,17 ± 1,77 %	91,63 ± 1,48 %	91,11 ± 1,84 %	91,14 ± 1,78 %	537/589
SqueezeNet	83,19 ± 5,75 %	84,55 ± 4,43 %	83,27 ± 5,62 %	83,16 ± 5,80 %	490/589
ResNet-18	86,42 ± 2,88 %	87,05 ± 2,65 %	86,34 ± 2,91 %	86,35 ± 2,93 %	509/589

Los datos expuestos en la Tabla 4.2 ratifican lo hallado en la Tabla 4.1. AlexNet alcanzó el mayor F1-score en las tres clases evaluadas. Su mejor resultado correspondió a bicicletas, con 92,68 %, seguido de vehículos con 90,40 % y personas con 90,32 %. La diferencia reducida entre las tres categorías evidencia que la arquitectura mantuvo un comportamiento equilibrado, sin depender excesivamente de una clase específica.

ResNet-18 también presentó un desempeño aceptable, especialmente para bicicletas, pero mostró una disminución en la clasificación de personas. SqueezeNet obtuvo su resultado más bajo para vehículos, con un F1-score de 79,69 %, lo que indica una mayor dificultad para equilibrar la precisión y el recall en esa categoría.

Tabla 4. 2

F1-score global por clase

Arquitectura	Personas	Bicicletas	Vehículos	F1 macro global
AlexNet	90,32 %	92,68 %	90,40 %	91,14 %
SqueezeNet	84,88 %	84,95 %	79,69 %	83,17 %
ResNet-18	84,07 %	89,16 %	85,78 %	86,34 %

En la tabla 4.3 se puede ver como AlexNet identificó correctamente 190 de las 200 muestras reales de bicicletas, registrando únicamente 10 omisiones. El recall de 95,00 % indica que la arquitectura reconoció correctamente casi la totalidad de las bicicletas evaluadas.

En la clase vehículos se obtuvieron 179 aciertos, 16 falsas alarmas y 22 omisiones. Aunque esta clase presentó una cantidad ligeramente mayor de muestras no reconocidas, mantuvo una precisión de 91,79 % y un F1-score de 90,40 %.

Para personas se registraron 168 aciertos, 16 falsas alarmas y 20 omisiones. El equilibrio entre precisión y recall produjo un F1-score de 90,32 %. En general, AlexNet mostró resultados consistentes en las tres clases, sin presentar una categoría con un deterioro marcado.

Tabla 4.3

Aciertos y errores de AlexNet

Clase	Muestras reales	VP (aciertos)	FP (falsas alarmas)	FN (omisiones)	Precisión	Recall	F1-score
Bicicletas	200	190	20	10	90,48 %	95,00 %	92,68 %
Vehículos	201	179	16	22	91,79 %	89,05 %	90,40 %
Personas	188	168	16	20	91,30 %	89,36 %	90,32 %

Resultados con data augmentation

En la tabla 4.4 se muestra como con la incorporación de data augmentation ResNet-18 presenta el mejor desempeño general, esta arquitectura alcanzó una exactitud media de 84,23 % ± 1,32 % y un F1- score macro de 84,13 % ± 1,44 %, evidenciando una mejora estadística frente a AlexNet y SqueezeNet.

ResNet-18 obtuvo el mayor número de aciertos, con 3969 sobre un total de 4712 muestras. Este resultado representa 123 aciertos adicionales frente a AlexNet y 254 frente a SqueezeNet.

SqueezeNet mostró la menor DE del valor de F1-score, lo que indica poca variación entre los 5 *folds* realizados. Sin embargo, esta estabilidad se produjo alrededor de un desempeño inferior, ya que su F1-score macro promedio fue de 78,76 %. Por tanto, una desviación estándar baja no significa necesariamente que la arquitectura haya obtenido el mejor resultado, sino que mantuvo valores similares durante las cinco particiones.

Tabla 4.4

Desempeño general de las CNN

Arquitectura	Exactitud media ± DE	Precisión macro ± DE	Recall macro ± DE	F1 macro ± DE	Aciertos/total
AlexNet	81,62 ± 1,97 %	82,22 ± 2,18 %	81,69 ± 2,03 %	81,61 ± 1,99 %	3846/4712
SqueezeNet	78,84 ± 0,94 %	80,15 ± 1,22 %	78,87 ± 0,97 %	78,76 ± 0,88 %	3715/4712
ResNet-18	84,23 ± 1,32 %	84,59 ± 1,35 %	84,17 ± 1,35 %	84,13 ± 1,44 %	3846/4712

En la tabla 4.5 para las tres arquitecturas entrenadas ResNet-18 es la que presenta un F1 macro global más alto. Su mayor desempeño corresponde a la clase de vehículos, con 87,89 %, seguido de bicicletas con 83,40 % y personas con 81,24 %. La clase vehículos fue la mejor clasificada por las tres arquitecturas, mientras que las clases de bicicletas y personas presentaron valores reducidos. Este comportamiento indica que las características micro doppler de los vehículos fueron identificadas con mayor consistencia. En comparación las bicicletas y personas al tener movimientos similares presentan firmas micro doppler parecidas lo que vuelve más complicada su clasificación en el conjunto con data augmentation.

AlexNet ocupó la segunda posición en todas las clases, mientras que SqueezeNet obtuvo los valores más bajos. La diferencia observada entre vehículos y las otras dos categorías podría relacionarse con una mayor similitud entre determinadas características espectrales de bicicletas y personas.

Tabla 4.5

F1-score global por clase

Arquitectura	Personas	Bicicletas	Vehículos	F1 macro global
AlexNet	79,85 %	79,26 %	85,84 %	81,65 %
SqueezeNet	76,96 %	76,88 %	82,74 %	78,86 %
ResNet-18	81,24 %	83,40 %	87,89 %	84,17 %

La Tabla 4.6 detalla que ResNet-18 tuvo 1344 aciertos de las 1600 muestras reales de bicicletas. De igual forma se registraron 279 falsas alarmas y 256 omisiones, produciendo un F1-score de 83,40 %.

La clase vehículos presentó el mejor comportamiento, con 1404 aciertos, 183 falsas alarmas y 204 omisiones. Mostró una precisión de 88,47 % lo que indica que la mayoría de las muestras clasificadas como vehículos pertenecían a esta categoría.

La clase personas presentó el desempeño más bajo de ResNet-18, con 1221 aciertos, 281 falsas alarmas y 283 omisiones. La similitud entre la precisión y el recall produjo un F1-score equilibrado de 81,24 % que es inferior comparado con el F1-socre de las demás clases.

Tabla 4.6

Aciertos y errores de ResNet-18

Clase	Muestras reales	VP (aciertos)	FP (falsas alarmas)	FN (omisiones)	Precisión	Recall	F1-score
Bicicletas	1600	1344	279	256	82.81 %	84,00 %	83,40 %
Vehículos	1608	1404	183	204	88.47 %	87,31 %	87,89 %
Personas	1504	1221	281	283	81.29 %	81,18 %	81,24 %

Influencia de la data augmentation

Los resultados que se ven en la tabla 4.7, muestran que la aplicación de la data augmentation modificó el comportamiento relativo de las arquitecturas. Antes del procesamiento, AlexNet ocupó la primera posición con un F1-score macro de 91,14 %. Sin embargo, en el conjunto con data augmentation su desempeño disminuyó hasta 81,61 %, lo que representó una reducción de 9,54 puntos porcentuales y la ubicó en la segunda posición.

SqueezeNet en ambos escenarios se mantuvo en la tercera posición, demostrando que fue la arquitectura con el menor desempeño en los resultados de la clasificación.

ResNet-18 presentó la menor reducción de variación en puntos porcentuales entre los dos escenarios. Su F1-score pasó de 86,35 % a 84,13 %, con una diferencia de solo 2,21 puntos. Estos resultados permitieron que la arquitectura pasara de la segunda posición en el conjunto de datos sin aumentar a la primera posición en el escenario con data augmentation , antes del procesamiento, a la primera posición en el escenario con data augmentation.

Estos resultados demuestran que el aumento de la cantidad de las muestras no produjo una mejora del F1-score promedio. Sin embargo, permitió observar que las arquitecturas respondieron de forma diferente ante las nuevas condiciones. ResNet-18 conservó una mayor proporción de su desempeño inicial y obtuvo los mejores resultados con data augmentation, mientras que AlexNet presentó una reducción más marcada.

Debe expresarse que este resultado sugiere una mayor estabilidad de ResNet-18 ante la variabilidad incorporada, pero no demuestra por sí solo que el data augmentation haya causado una reducción del desempeño, debido a que los escenarios contienen cantidades y características diferentes de muestras.

Tabla 4.7

F1-score macro obtenido por las arquitecturas con datos aumentados

Arquitectura	Datos previos al procesamiento	Datos con data augmentation	Variación	Posición antes del procesamiento	Posición con data augmentation
AlexNet	91,14 %	81,61 %	−9,54 puntos porcentuales	1.ª	2.ª
SqueezeNet	83,16 %	78,76 %	−4,40 puntos porcentuales	3.ª	3.ª
ResNet-18	86,35 %	84,13 %	−2,21 puntos porcentuales	2.ª	1.ª

Resultados de Friedman globales y de Wilcoxon con datos previos al procesamiento

Con el propósito de determinar si las diferencias observadas entre AlexNet, SqueezeNet y ResNet-18 correspondían a un comportamiento sistemático, se aplicó inicialmente la prueba no paramétrica de Friedman sobre los valores de F1-score macro obtenidos en los cinco folds. La hipótesis nula estableció que las tres arquitecturas presentaban rangos equivalentes de desempeño. Como se presenta en la tabla 4.7 el resultado obtenido fue $p = 0{,}07427$, valor superior al nivel de significancia de 0,05, por lo que no se rechazó la hipótesis nula. En consecuencia, no se encontraron diferencias globales estadísticamente significativas entre las CNN.

Como análisis complementario, se realizaron comparaciones pareadas mediante la prueba de Wilcoxon de rangos con signo. Para cada contraste, la hipótesis nula planteó que la mediana de las diferencias del F1-score macro entre las dos arquitecturas era igual a cero. Los valores ppp obtenidos fueron 0,0591 para AlexNet frente a SqueezeNet, 0,1056 para AlexNet frente a ResNet-18 y 0,5896 para SqueezeNet frente a ResNet-18. Debido a la realización de múltiples comparaciones, se aplicó la corrección de Holm para controlar el

error de tipo I. Después del ajuste, los valores p fueron 0,1772, 0,2113 y 0,5896, respectivamente. Todos superaron el umbral de 0,05, por lo que ninguna diferencia por pares resultó estadísticamente significativa. Por tanto, las diferencias observadas entre las arquitecturas fueron únicamente descriptivas y no permitieron demostrar la superioridad estadística de una CNN sobre las demás.

Tabla 4.8

Resultados de Friedman y Wilcoxon, para datos previos al procesamiento, con $\alpha = 0.05$

Comparación	p de Friedman	p de Wilcoxon	p ajustado por Holm	Resultado
Friedman para hipótesis global	0.07427	-	-	No significativo
AlexNet vs. SqueezeNet	-	0,0591	0,1772	No significativo
AlexNet vs. ResNet-18	-	0,1056	0,2113	No significativo
SqueezeNet vs. ResNet-18	-	0,5896	0,5896	No significativo

Resultados de Friedman globales y de Wilcoxon con datos con data augmentation

Con el propósito de determinar si las diferencias observadas entre las arquitecturas correspondían a un comportamiento sistemático, se aplicó la prueba no paramétrica de Friedman sobre los valores de F1- score macro obtenidos en los cinco folds. La hipótesis nula estableció que AlexNet, SqueezeNet y ResNet-18 presentaban rangos equivalentes de desempeño. El resultado obtenido fue $p = 0{,}04076$ inferior al nivel de significancia de 0,05, por lo que se rechazó la hipótesis nula y se concluyó que existieron diferencias globales estadísticamente significativas entre las CNN.

Debido a que Friedman no permite identificar directamente entre qué arquitecturas se encuentran las diferencias, se realizaron comparaciones pareadas mediante la prueba de Wilcoxon de rangos con signo. Para cada contraste, la hipótesis nula estableció que la mediana de las diferencias del F1-score macro entre las dos arquitecturas era igual a cero. Los valores p obtenidos fueron 0,1056 para AlexNet frente a SqueezeNet, 0,1056 para AlexNet frente a ResNet-18 y 0,0591 para SqueezeNet frente a ResNet 18.

Con el fin de controlar el error de tipo I asociado con las múltiples comparaciones, se aplicó la corrección de Holm. Después del ajuste, los valores fueron 0,2113, 0,2113 y 0,1772, respectivamente. Todos superaron el umbral de 0,05, por lo que ninguna

comparación por pares resultó estadísticamente significativa. En consecuencia, aunque ResNet-18 alcanzó el mayor F1-score macro promedio, con 84,13%, su ventaja debe interpretarse como una superioridad numérica y no como una superioridad estadística individual frente a las demás arquitecturas.

Tabla 4. 9

Resultados de Friedman y Wilcoxon para los datos con data augmentation, con $\alpha = 0.05$

Comparación	p de Friedman	p de Wilcoxon	p ajustado por Holm	Resultado
Friedman para hipótesis global	0,04076	-	-	Significativo
AlexNet vs. SqueezeNet	-	0,1056	0,2113	No significativo
AlexNet vs. ResNet-18	-	0,1056	0,2113	No significativo
SqueezeNet vs. ResNet-18	-	0,0591	0.1772	No significativo

Capítulo V

Conclusiones

Se desarrolló una metodología integral para la adquisición, procesamiento y clasificación de objetivos mediante un radar FMCW de 24 GHz, utilizando el dispositivo XENSIV™ Distance2Go. La obtención de señales en fase y cuadratura permitió obtener información acerca del movimiento de personas, bicicletas y vehículos en un escenario controlado, mostrando la capacidad que tiene el radar para generar información diferenciable para cada una de las tres categorías evaluadas. El proceso que se aplicó permitió transformar las señales adquiridas de fase y cuadratura en representaciones tiempo-frecuencia para su posterior análisis mediante técnicas de deep learning.

La STFT permitió obtener representaciones de las firmas Doppler y micro doppler características de cada tipo de objetivo, realizando los respectivos espectrogramas. Estas representaciones permitieron el uso efectivo de arquitecturas de redes neuronales convolucionales para realizar una clasificación correcta de las 3 clases presentadas, personas, bicicletas y vehículos.

La inserción de ruido AWGN con niveles de SNR de 10dB, 20dB y 30 dB permitió un aumento eficaz de datos, al replicar condiciones en entornos adversos que se caracterizan por presentar mucha interferencia, este enfoque mejoró la variabilidad de la base de datos, lo que proporcionó a las arquitecturas de CNN una mayor capacidad para reconocer y clasificar señales bajo condiciones desfavorables lo que generó un proceso de clasificación más robusto No obstante, los resultados también evidenciaron que con este aumento en el volumen de datos de manera sintética o una mayor variabilidad no garantizan por sí solos un incremento significativo de las métricas de clasificación, por lo que la influencia del aumento de datos depende de la arquitectura empleada y de las características del conjunto de entrenamiento.

En la evaluación realizada con los datos aumentados, ResNet-18 presentó el mejor desempeño numérico entre las tres arquitecturas analizadas. Alcanzó como se indica en la

tabla 4.4 una exactitud media de 84,23 %, una precisión macro de 84,59 %, un recall macro de 84,17 % y un F1-score macro medio de 84,13 % durante la validación cruzada de cinco folds. Además, obtuvo 3969 clasificaciones correctas de 4712 muestras evaluadas. La prueba de Friedman obtuvo un valor de $p = 0{,}041$, lo que evidenció diferencias globales estadísticamente significativas entre las tres arquitecturas. Las comparaciones pareadas de Wilcoxon, aplicando el ajuste de Holm, arrojaron un resultado claro: estadísticamente, no existen diferencias significativas aisladas entre los modelos. Frente a este empate técnico, la evaluación tuvo que basarse en el rendimiento empírico puro. Al no poder probar una superioridad matemática estricta de ResNet-18 sobre AlexNet o SqueezeNet de forma individual, nos apoyamos en métricas descriptivas globales. Fueron los valores crudos de precisión y exactitud los que finalmente posicionaron a ResNet-18 como la arquitectura más eficiente sobre el conjunto de datos aumentado.

El análisis de desempeño por cada clase presenta que las arquitecturas tuvieron un mejor resultado al clasificar vehículos, en la tabla 4.5 se evidencia que los porcentajes de F1-score fueron los mejores en las tres redes, cada una tuvo valores porcentuales de 85,84 % con AlexNet, 82,74 % con SqueezeNet y 87,89 % con ResNet-18. Por el contrario, las redes presentaron más dificultades al momento de clasificar personas y bicicletas donde la matriz de confusión mostró una mayor cantidad en clasificaciones cruzadas entre ambas, esto debido a la similitud que existe en las firmas micro doppler de los movimientos de peatones y bicicletas.

Los resultados obtenidos con los datos previos al procesamiento mostraron un comportamiento diferente al observado con el conjunto de data augmentation. En este caso, AlexNet alcanzó el mayor desempeño con un puntaje de F1 -score macro de 91,14 %, seguida de ResNet-18 con 86,35 % y SqueezeNet con 83,16 % como se indica en la tabla 4.1. Sin embargo, el valor de p obtenido en la prueba global de Friedman fue de 0,07427, superior al nivel de significancia de 0,05, lo que indica que no se rechaza la hipótesis nula la hipótesis nula donde el comportamiento de las tres arquitecturas es similar, por lo tanto, no existen diferencias estadísticamente significativas entre las CNN. Asimismo, las

comparaciones pareadas de Wilcoxon, ajustadas mediante el método de Holm, tampoco mostraron diferencias estadísticamente significativas entre los modelos. Al no existir diferencias en los procesos estadísticos se utilizó la métrica de F1- score macro para determinar que Alexnet fue la que tuvo un mejor desempeño en la clasificación de estos datos.

Finalmente al comparar los dos conjuntos de datos experimentales se evidencia que la selección de una arquitectura no se debe basar únicamente por sus valores numéricos puntuales (exactitud, precisión, recall, F1-score), las pruebas estadísticas deben complementarse con los valores numéricos puntuales en un análisis de rendimiento para tener un valor confiable en la decisión de cual CNN resultó mejor frente a otra. Si bien no se encontró diferencias estadísticas significativas en ninguna de las pruebas pareadas de Wilcoxon, el uso de estos valores numéricos permitió escoger a Resnet-18 como la red con el mejor desempeño en el caso más desafiante, el de datos aumentados, y a AlexNet como la mejor en el caso con menos datos.

Recomendaciones

Se recomienda verificar los códigos de adquisición de datos del radar para que no existan conflictos con versiones modernas de MATLAB y tampoco se tengan mediciones erróneas.

Se recomienda ampliar la base de datos, mediante nuevas adquisiciones reales y distintas técnicas para la creación de señales sintéticas. Para encontrar una mejora en la capacidad de generalización de los modelos y hacerlos más robustos frente a situaciones adversas.

Se recomienda realizar adquisiciones de datos en nuevos escenarios más diversos, donde se puedan incluir más objetivos dentro de la zona efectiva de medición del radar para que permita evaluar con mayor rigor el comportamiento del sistema en situaciones que se asemejen a un caso real.

Se recomienda analizar a profundidad las clases que tienen mayores similitudes en firmas micro doppler, personas y bicicletas, se podrían encontrar nuevas estrategias de procesamiento de las señales para poder tener espectrogramas distintos y mejora la clasificación de clases en las CNN.

Se recomienda hacer el estudio con diferentes tipos de CNN para tener una mayor diversidad de arquitecturas que permita una selección de la red más adecuada para planteamientos en casos reales.

Referencias

Wang, Z., & Rawat, W. (2017). Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation, 29*(9), 2352-2449.

Andras, P., Jiaao, D., Julian F, K., & Dariu M, G. (2020). CNN Based Road User Detection Using the 3D Radar Cube. *IEEE Robotics and Automation Letters, 5*(2), 1263-1270.

Antich, M. (12 de Abril de 2023). *automotive FLEET*. Recuperado el 06 de abril de 2026, de https://www.automotive-fleet.com/articles/safety-systems-may-cause-phantom-braking

Bhogal, R., & Singh , A. (2025). A Comprehensive Review of ResNet-18: Architecture and Applications. *Advances in Data-Driven Computing and Intelligent Systems, 1333*, 595-607.

Chen, V., & Wang, J. (2014). *Radar and Sonar Imaging and Processing.* CRC Press.

Demšar, J. (2006). Statistical Comparisons of Classifiers. *Journal of Machine Learning Research 7, 1*(30).

Diaz, J. E. (2026). Vehículos autónomos: desafíos y ventajas en su implementación. *Doctrina Distral, 1*(6), 38-62.

Elena, L. C. (2025). *Algoritmo de estimación odométrica basado en radar para vehículos terrestres en condiciones climáticas adversas.* Sevilla: Universidad de Sevilla.

Enrique, T. V. (2024). *Implementación de RADAR de 120 GHz para la detección del ritmo cardiaco en personas.* Comillas: Comillas Universidad Pontifica.

Ghufron, A., Fiky Y, S., & Istiqomah. (2026). Robust Real-Time Human Activity Recognition via 2D Tensor CNN With Change Point Detection and Confidence-Based Prediction Filtering. *IEEE Access, 14*, 80802-80816.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.* MIT Press.

Greco, M., Li, J., & Zoubir, A. (2019). Advances in Radar Systems for Modern Civilian and Commercial Applications: Part 1 [From the Guest Editors],. *IEEE Signal Processing Magazine, 36*(5), 16-18.

Greco, M., Li, J., Long, T., & Zoubir, A. (2019). Advances in radar systems for modern civilian and commercial applications: Part 2. *IEEE Signal Processing Magazine, 36*(5), 16-18.

Gurbuz, S., & Amin, M. (2018). Radar-based human-motion recognition with deep learning: Promising applications for radar systems. *IEEE Signal Processing Magazine, 36*(4), 16-28.

Haoran, D., Tao, S., Gang, Y., Yifan, S., & Yu, C. (2026). *Variational time-frequency mode tracking for micro-Doppler signature extraction.* Beijing: ScienceDirect.

Harun, S., & Levent, S. (2026). Comparative Analysis of Time-Frequency Transformations and Deep Learning Architectures in the Classification of FMCW Radar Micro-Doppler Signatures. *GU J Sci, 13*(2), 692-709.

Hassanpour, M., & Malek, H. (2019). Document Image Classification using SqueezeNet Convolutional Neural Network. *2019 5th Iranian Conference on Signal Processing and Intelligent Systems .* Shahrood.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer.

Haykin, S. (2014). *Communication Systems.* John Wiley & Sons.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics, 6*(2), 65-70.

Infineon Technologies AG. (2017). *DEMO DISTANCE2GOL Software Guide Xensiv 24 GHz radar.* Neubiberg: Infineon Technologies AG.

Infineon Technologies AG. (2019). *Distance2Go XENSIV™ 24 GHz radar demo kit with BGT24MTR11 and XMC4200 32-bit ARM Cortex-M4 MCU for ranging, movement and presence detection (Application Note AN543).* Múnich: Infineon Technologies AG.

Jankiraman, M. (2018). *FMCW.* Norwood: Artech House.

Japkowicz, N., & Shah, M. (2011). *Evaluating Learning Algorithms: A Classification Perspective.* Cambridge: Cambridge University Press.

Jaya, G., Sunil, P., & Gireesh, K. (2022). Deep Learning (CNN) and Transfer Learning: A Review. *Journal of Physics: Conference Series, 2273*, 1742-6596.

Li, X., He, Y., & Jing, X. (2019). A Survey of Deep Learning-Based Human Activity Recognition in Radar. *Remote Sensing, 11*(9), 22.

Lin, M., Chen, Q., & Yan, S. (2013). Network In Network. *Neural and Evolutionary Computing*.

MathWorks. (s.f.). *MATLAB Documentation.* Recuperado el 06 de abril de 2026, de https://la.mathworks.com/help/deeplearning/ug/introduction-to-convolutional-neural-networks.html

Mavares F, P. (2021). Evaluation of a fmcw radar as a teaching tool in the automotive and telecommunication engineering careers. *INGENIUS*(25), 11.

Md Zahangir , A., Taha, T., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M., . . . Asari, V. (2018). *The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches.* arXiv.

Neda , R., Passafiume, M., Sadeghibakhi, M., Collodi, G., & Cidronali, A. (2023). Model-Based Data Augmentation Applied to Deep Learning Networks for Classification of Micro-Doppler Signatures Using FMCW Radar. *IEEE Transactions on Microwave Theory and Techniques, 71*(5), 2222-2236.

Pallabi, B., Samadendra, N., Rabindranath, B., Agbotiname, L., & Chun-Ta, L. (2025). *Advanced Signal Processing and Modeling Techniques for Automotive Radar: Challenges and Innovations in ADAS Applications.* Sikkim: Tech Science Press.

Proakis, J., & Manlakis, D. (2007). *Digital Signal Processing.* Pearson Prentice Hall.

Qiwei, Y., Ruixun, Z., & XiuLi, S. (2019). CNN and RNN mixed model for image classification. *2018 International Joint Conference on Metallurgical and Materials Engineering (JCMME 2018).* Tianjin.

Sanaullah, S., Piort, K., & Michal, S. (2026). Classification of Low-Observable Aerial Objects Using Micro-Doppler Signatures and Deep Learning. *2026 27th International Radar Symposium (IRS).* Krakow.

Seyfioğlu, M., Özbayoğlu, A., & Gürbüz, S. (2018). Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities. *IEEE Transactions on Geoscience and Remote Sensing, 54*(4), 1709-1723.

Shorten, C., & Khoshgoftaar , T. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data, 6*(60), 1-48.

Smethurst, S. (27 de junio de 2023). *brightmile*. Recuperado el 06 de abril de 2026, de https://www-brightmile-io.translate.goog/blog/the-dangers-of-advanced-driver-assistance-systems-adas?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=tc

Sutskever, I., Martens, J., Dahl, G., & Hinton, G. (2013). On the importance of initialization and momentum in deep learning. *Proceedings of the 30th International Conference on Machine Learning.*

Wickramarachi, P. (2003). *Effects of Windowing on the Spectral Content of a Signa.* San Jose, California: Data Physics Corporation.

Xin, L., Lei, H., Zhen, Y., Jun Bo, Z., Si Yang, Q., & Qi, C. (2026). *Research on UWB Radar Human Posture Recognition in Complex Disaster Environments Based on Active Sensing Closed-loop Feedback.* Huludao: SSRN.