ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

1922 - ECUADOR

TRABAJO DE INTEGRACIÓN CURRICULAR

Diseño y validación conceptual de un sistema de identificación acústica de vehículos aéreos y terrestres en entornos militares mediante el uso de aprendizaje automático

Andrade Pico Luis Clemente • **Gómez Díaz Melissa Malayca**

Carrera de Telecomunicaciones

Director: Ing. Carrera Erazo Enrique Vinicio, PhD | Julio 2026



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

CONTENIDO

01 Contexto y problema
02 Objetivos
03 Metodología
04 Resultados
05 Conclusiones y recomendaciones

PLANTEAMIENTO DEL PROBLEMA

- → Escenarios beligerantes complejos con empleo de plataformas tripuladas y no tripuladas móviles.
- → Los vehículos y plataformas poseen firmas acústicas particulares y diferenciables, incluso por variaciones en su carga útil.
- → Los sistemas tradicionales basados en sensores ópticos, radar, infrarrojo y radiofrecuencia presentan altos costos y limitaciones operativas.

Problema central

Diseñar y validar conceptualmente un sistema capaz de identificar vehículos aéreos y terrestres en entornos militares, considerando ruido, variabilidad operacional, similitud acústica entre plataformas y condiciones reales de adquisición.

OBJETIVOS

Objetivo general

Diseñar y validar conceptualmente un sistema para la identificación automática de vehículos aéreos y terrestres mediante el análisis de sus señales acústicas, con aplicación directa en sistemas de vigilancia y seguridad en entornos militares.

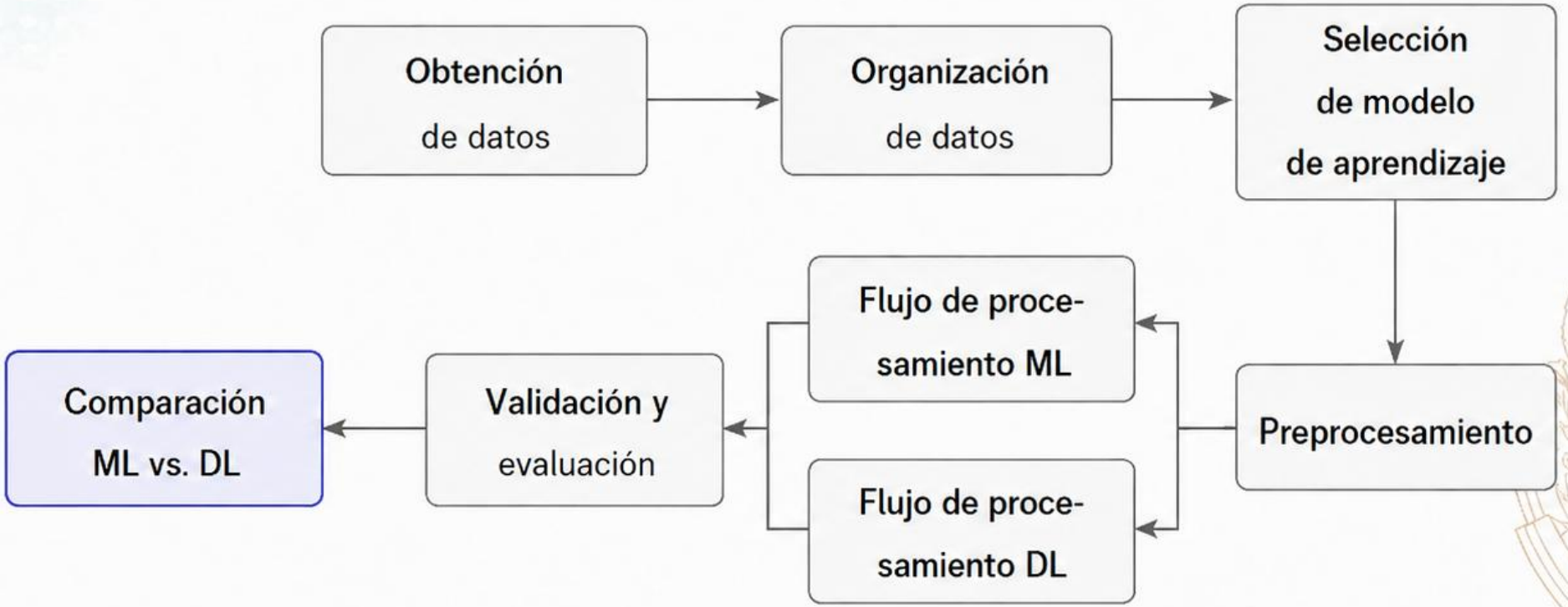
Objetivos específicos

1. Emplear arquitecturas de aprendizaje automático tradicional y de aprendizaje profundo para la clasificación acústica de vehículos aéreos y terrestres.
2. Evaluar si el empleo de una arquitectura jerárquica para clasificación mejora la identificación de los vehículos en distintos niveles de decisión, desde categorías generales hasta modelos específicos.
3. Evaluar y comparar los modelos empleados mediante métricas de clasificación, con el fin de determinar cuál ofrece el desempeño más óptimo en cada nivel jerárquico.
4. Validar la factibilidad conceptual del sistema propuesto de identificación acústica.

ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

METODOLOGÍA

FLUJO METODOLÓGICO



EQUIPO DE ADQUISICIÓN ACÚSTICA



Zoom Handy Recorder H4n

- Grabadora portátil
- Micrófonos condensadores estéreo XY
- Apertura ajustable: 90° / 120°
- Formato: WAV sin compresión
- 24 bits — Resolución: 24 bits
- Frecuencia de muestreo: 96 000 Hz

Captura en condiciones acústicas reales y no simuladas

CONSTRUCCIÓN DEL DATASET

ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

- 1 mes
- 100 % propio
- Condiciones reales









Campañas de campo en unidades militares

REGISTROS ORIGINALES DEL DATASET

Concepto	Valor
Total de clases finales	**22**
Total de registros originales	**157**
Duración total	**05:02:45.89**
Duración total en minutos	**302.76 min**

AUDIOS ORGANIZADOS

SUBCLASE	AUDIOS ORIGINALES	% DEL TOTAL	TIEMPO DE AUDIO ORIGINAL
AIR3S	4	2.55 %	00:02:26.32
DRAGONEYE2	4	2.55 %	00:09:29.63
MATRICE_30	2	1.27 %	00:02:30.61
MATRICE_350	2	1.27 %	00:04:35.77
MATRICE_400	2	1.27 %	00:03:52.78
MAVIC 3T	8	5.10 %	00:07:50.09
MAVIC2_PRO_ENTERPRISE	2	1.27 %	00:03:58.05
MINI4PRO	2	1.27 %	00:02:14.86
AS332B_SUPERPUMA_EUROCOPTER_H215	4	2.55 %	00:07:44.06
ECUREUIL_AS350B_B3_AIRBUS_H125	16	10.19 %	00:15:16.31
FENNEC_AS550-C3E	2	1.27 %	00:08:31.28
H145_COBRA	1	0.64 %	00:01:05.23
AMX 13	6	3.82 %	00:20:40.39
AMX_105mm	6	3.82 %	00:12:21.08
AMX_155mm	12	7.64 %	00:36:30.44
APC	10	6.37 %	00:19:21.46
CASCABEL	10	6.37 %	00:19:42.50
COBRA II	18	11.46 %	00:41:33.88
DAVID	6	3.82 %	00:10:58.19
MBOMBE	24	15.29 %	00:49:06.00
URAL	8	5.10 %	00:09:53.79
URUTÚ	8	5.10 %	00:13:03.19
TOTAL	**157**	**100.00 %**	**05:02:45.89**

SELECCIÓN DE MODELOS

ML MACHINE LEARNING

- Buen comportamiento binomial y multiclase
- Kernel RBF flexible
- Robusto con dataset limitado
- Minimización del riesgo estructural

Modelo seleccionado: SVM

DL DEEP LEARNING

- Uso de pesos compartidos entre filtros
- Más eficiente con parámetros limitados
- Capacidad de análisis tiempo-frecuencia

Modelo seleccionado: CNN

- **SVM:** usa un conjunto de descriptores diseñado
- **CNN:** aprende características directamente del dato

NO JERÁRQUICO VS JERÁRQUICO

NO JERÁRQUICO

Clasificacćón plana

↓

22 clases finales

- Una sola decisión multiclase
- Mayor complejidad de separación
- Más confusión entre clases similares

JERÁRQUICO

AÉREOS / TERRESTRES

AUV / HELICÓPTEROS | ORUGA / RUEDA

MODELOS ESPECÍFICOS

- Decisiones progresivas más simples
- Nodos especializados por grupo
- Menor confusión acústica
- Mejor ajuste al problema

DECISIÓN ADOPTADA: CLASIFICACIÓN JERÁRQUICA

Permite dividir el problema y especializar la clasificación en grupos acústicamente relacionados.

PREPROCESAMIENTO

SEGMENTADOR

1. Toma los audios originales del dataset
2. Convierte a mono si es necesario
3. Remuestrea a 96,000 Hz
4. Segmenta en ventanas de 4.0 s
5. Salto de 4.0 s (sin solapamiento)
6. Descarta segmentos con RMS < 0.003
7. Guarda segmentos en WAV PCM 16 bits
8. Genera el archivo manifest_segmentos.csv

Autores	Tiempo	Uso
Salamon y Bello (2017)	3 s	entrada CNN
Salamon y Bello (2017)	4 s	sonido ambiental
Piczak (2015)	5 s	sonido ambiental
Seo, Jang e Im (2018)	10 s	drones

AUDIOS SEGMENTADOS

AUV			
SUBCLASE	MUESTRAS	%	TIEMPO
AIR3S	34	0.76 %	00:02:16
DRAGONEYE2	142	3.18 %	00:09:28
MATRICE_30	36	0.81 %	00:02:24
MATRICE_350	68	1.52 %	00:04:32
MATRICE_400	58	1.30 %	00:03:52
MAVIC 3T	114	2.56 %	00:07:36
MAVIC2_PRO_ENTERPRISE	58	1.30 %	00:03:52
MINI4PRO	32	0.72 %	00:02:08

HELICÓPTEROS			
SUBCLASE	MUESTRAS	%	TIEMPO
AS332B_SUPERPUMA_EUROCOPTER_H215	114	2.56 %	00:07:36
ECUREUIL_AS350_B3_AIRBUS_H125	221	4.96 %	00:14:44
FENNEC_AS550-C3E	127	2.85 %	00:08:28
H145_COBRA	16	0.36 %	00:01:04

TERRESTRES			
SUBCLASE	MUESTRAS	%	TIEMPO
AMX 13	308	6.91 %	00:20:32
AMX_105mm	182	4.08 %	00:12:08
AMX_155mm	546	12.24 %	00:36:24
APC	286	6.41 %	00:19:04
CASCABEL	290	6.50 %	00:19:20
COBRA II	610	13.68 %	00:40:40
DAVID	162	3.63 %	00:10:48
MBOMBE	722	16.19 %	00:48:08
URAL	142	3.18 %	00:09:28
URUTÚ	192	4.30 %	00:12:48

TOTAL	4460	100.00 %	04:57:20

FLUJO DE PROCESAMIENTO DE LOS MODELOS

FLUJO DE PROCESAMIENTO DEL MODELO ML

Sistema jerárquico de clasificación

Aéreos

AUV	Helicópteros
AIR3S	AS332B SUPERPUMA
DRAGONEYE2	ECUREUIL AS350
MATRICE 30	FENNEC AS550-C3E
MATRICE 350	H145 COBRA
MATRICE 400	
MAVIC 3T	
MAVIC2 PRO ENTERPRISE	
MINI4 PRO	

Terrestres

Oruga	Rueda
AMX 13	CASCABEL
AMX 105mm	COBRA II
AMX 155mm	DAVID
APC	MBOMBE
	URAL
	URUTÚ

FLUJO DE PROCESAMIENTO DEL MODELO ML

Sistema jerárquico de clasificación

Nodo 1	**Aéreos**	**Terrestres**
Nodo 2	**AUV** / **Helicópteros**	**Oruga** / **Rueda**
Nodo 3	AIR3S DRAGONEYE2 MATRICE 30 MATRICE 350 MATRICE 400 MAVIC 3T MAVIC2 PRO ENTERPRISE MINI4 PRO	AS332B SUPERPUMA ECUREUIL AS350 FENNEC AS550-C3E H145 COBRA
		AMX 13 AMX 105mm AMX 155mm APC
		CASCABEL COBRA II DAVID MBOMBE URAL URUTÚ

FLUJO DE PROCESAMIENTO DEL MODELO ML

Aumento de datos y balanceo jerárquico

1 Balanceo jerárquico

Desde las clases inferiores hasta las clases superiores.

2 Separación por fold

Train	Test
70 %	30 %

Estratificado

La aumentación se aplica solo en train.

3 Balanceo local

Se realiza según el nodo que se está entrenando.

Ejemplo (AUV)

AIR3S	=	20 segmentos train
DRAGONEYE2	=	80 segmentos train
MATRICE_30	=	25 segmentos train
MINI4PRO	=	18 segmentos train

Objetivo: igualar a la clase mayor

target = 80

Entonces:

AIR3S necesita 60 aumentados
MATRICE_30 necesita 55 aumentados
MINI4PRO necesita 62 aumentados
DRAGONEYE2 necesita 0

FLUJO DE PROCESAMIENTO DEL MODELO ML

Aumento de datos y balanceo jerárquico

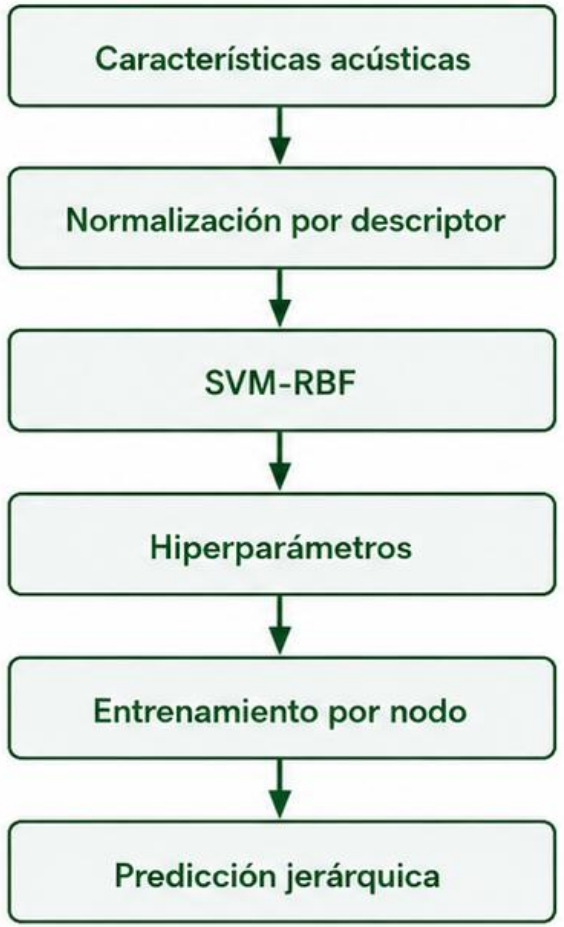
Técnica	Transformación aplicada	Proporción
Desplazamiento circular	Reubicación circular; hasta 10 %; misma duración.	40 %
Inversión de polaridad	Inversión de signo; misma duración; misma energía.	25 %
Variación de ganancia	Variación aleatoria ±3 dB; simula nivel de captura.	25 %
Estiramiento temporal leve	Compresión/expansión leve; factor 0.95–1.05; reajuste final.	10 %

Chu, H.-C., Zhang, Y.-L., & Chiang, H.-C. (2023)

FLUJO DE PROCESAMIENTO DEL MODELO ML

Normalización por características

1. **Características acústicas**
   Se emplean los descriptores extraídos de cada segmento de audio.
2. **Normalización por descriptor**
   Cada descriptor se escala de forma independiente usando solo estadísticas del conjunto de entrenamiento.
3. **SVM-RBF**
   El clasificador modela fronteras de decisión no lineales entre clases acústicas.
4. **Hiperparámetros**
   Se ajustan los hiperparámetros óptimos para cada nodo jerárquico.
5. **Entrenamiento por nodo**
   Se entrena un SVM local con las clases hijas en competencia.
6. **Predicción jerárquica**
   La decisión avanza por la jerarquía hasta llegar a la clase final.

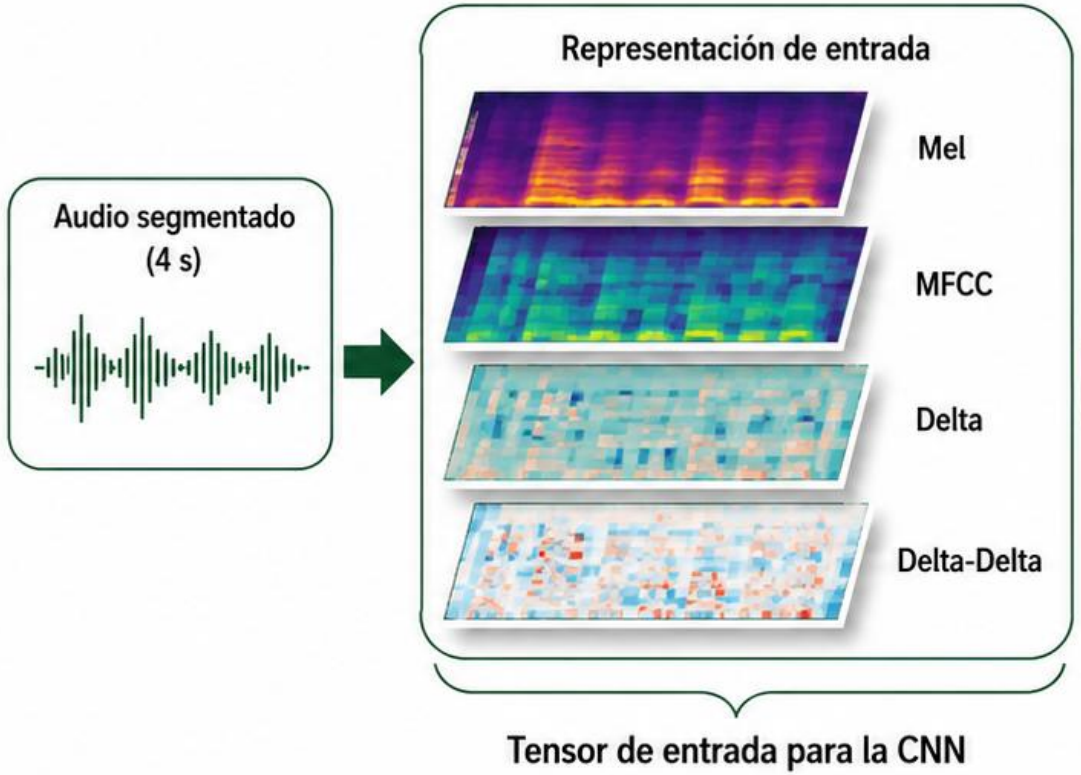


características acústicas → normalización por descriptor → SVM-RBF → hiperparámetros → entrenamiento por nodo → predicción jerárquica

FLUJO DE PROCESAMIENTO DEL MODELO DL

Entrada y representación acústica

1. Mismo preprocesamiento del modelo ML.
2. Ingreso: audios segmentados de 4 s.
3. Cada segmento se transforma en una representación de 4 canales.
4. Canales: Mel, MFCC, Delta y Delta-Delta.

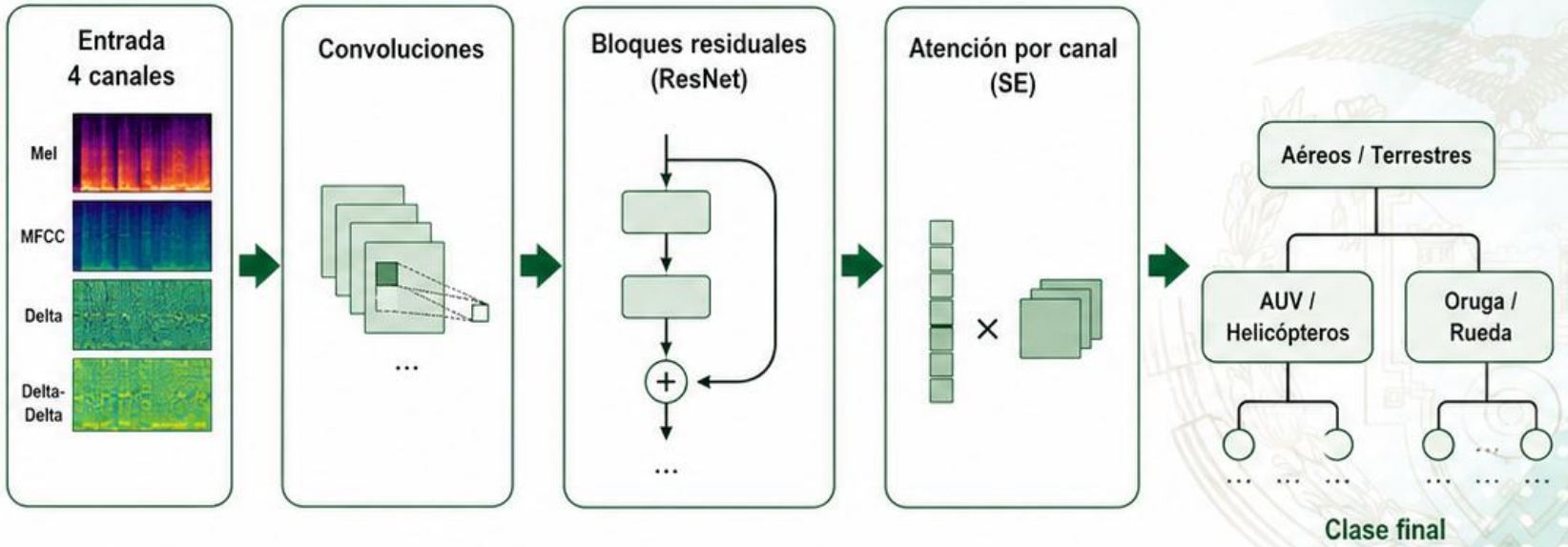


La CNN aprende directamente sobre estas representaciones acústicas.

FLUJO DE PROCESAMIENTO DEL MODELO DL

Aprendizaje y clasificación jerárquica

1. **Convoluciones**: detectan patrones locales del sonido.
2. **Bloques residuales**: facilitan el aprendizaje profundo.
3. **Atención por canal**: pondera Mel, MFCC, Delta y Delta-Delta.
4. **Salida jerárquica**: una CNN por nodo del árbol.



La **CNN** aprende automáticamente qué patrones observar; el **SVM** requiere características calculadas previamente.

ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

RESULTADOS ALCANZADOS

RESULTADOS ALCANZADOS

ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

SVM

Nodo	Clases evaluadas	N.º muestras	Accuracy	Balanced Acc.
N1 – Raíz	AEREOS vs TERRESTRES	4,460	99.64 %	99.63 %
N2 – Aéreos	AUV vs HELICOPTEROS	1,020	95.10 %	94.77 %
N2 – Terrestres	ORUGA vs RUEDA	3,440	96.42 %	96.76 %
N3 – AUV	8 modelos de dron	542	86.72 %	86.03 %
N3 – Helicópteros	3 modelos defendibles	462	74.03 %	67.56 %
N3 – Oruga	4 modelos de oruga	1,322	99.39 %	99.63 %
N3 – Rueda	6 modelos de rueda	2,118	86.83 %	85.97 %

CNN

Nodo	Clases evaluadas	N.º muestras	Accuracy	Balanced Acc.
N1 – Raíz	AEREOS vs TERRESTRES	1,079	99.17 %	98.98 %
N2 – Aéreos	AUV vs HELICOPTEROS	401	95.51 %	95.41 %
N2 – Terrestres	ORUGA vs RUEDA	678	99.41 %	99.36 %
N3 – AUV	8 modelos de dron	205	100.00 %	100.00 %
N3 – Helicópteros	4 modelos de helicóptero	196	42.86 %	55.02 %
N3 – Oruga	4 modelos de oruga	243	88.48 %	93.64 %
N3 – Rueda	6 modelos de rueda	435	90.34 %	93.33 %

RESULTADOS ALCANZADOS

SVM

Nodo	Clases evaluadas	N.º muestras	Accuracy	Balanced Acc.
N1 – Raíz	AEREOS vs TERRESTRES	4,460	99.64 %	99.63 %
N2 – Aéreos	AUV vs HELICOPTEROS	1,020	95.10 %	94.77 %
N2 – Terrestres	ORUGA vs RUEDA	3,440	96.42 %	96.76 %
N3 – AUV	8 modelos de dron	542	86.72 %	86.03 %
N3 – Helicópteros	3 modelos defendibles	462	74.03 %	67.56 %
N3 – Oruga	4 modelos de oruga	1,322	99.39 %	99.63 %
N3 – Rueda	6 modelos de rueda	2,118	86.83 %	85.97 %

CNN

Nodo	Clases evaluadas	N.º muestras	Accuracy	Balanced Acc.
N1 – Raíz	AEREOS vs TERRESTRES	1,079	99.17 %	98.98 %
N2 – Aéreos	AUV vs HELICOPTEROS	401	95.51 %	95.41 %
N2 – Terrestres	ORUGA vs RUEDA	678	99.41 %	99.36 %
N3 – AUV	8 modelos de dron	205	100.00 %	100.00 %
N3 – Helicópteros	4 modelos de helicóptero	196	42.86 %	55.02 %
N3 – Oruga	4 modelos de oruga	243	88.48 %	93.64 %
N3 – Rueda	6 modelos de rueda	435	90.34 %	93.33 %

CONCLUSIONES

1. El empleo de representación jerárquica mostró eficiencia y capacidad de escalamiento.

2. Las condiciones de obtención de las muestras repercuten directamente en la calidad de clasificación, dependiendo del modelo.

3. A pesar de tener muestras limitadas en cantidad y calidad variable, en la mayoría de nodos se logró una clasificación aceptable.

4. Aun aprovechando las particularidades de cada modelo y obteniendo una clasificación aceptable en la mayoría de nodos, existen salvedades en las que es indispensable contar con un mínimo de audios para su correcta ejecución (≥ 2).

Conclusión general: la solución es factible como vigilancia pasiva complementaria, sin atribuirle mayor madurez que la demostrada experimentalmente.

RECOMENDACIONES
y trabajos futuros

1. Realizar un modelo híbrido de clasificación, aprovechando la independencia de nodos, a fin de mejorar la clasificación.
   Zahid et al. (2015), Shekar y Hansen (2022)

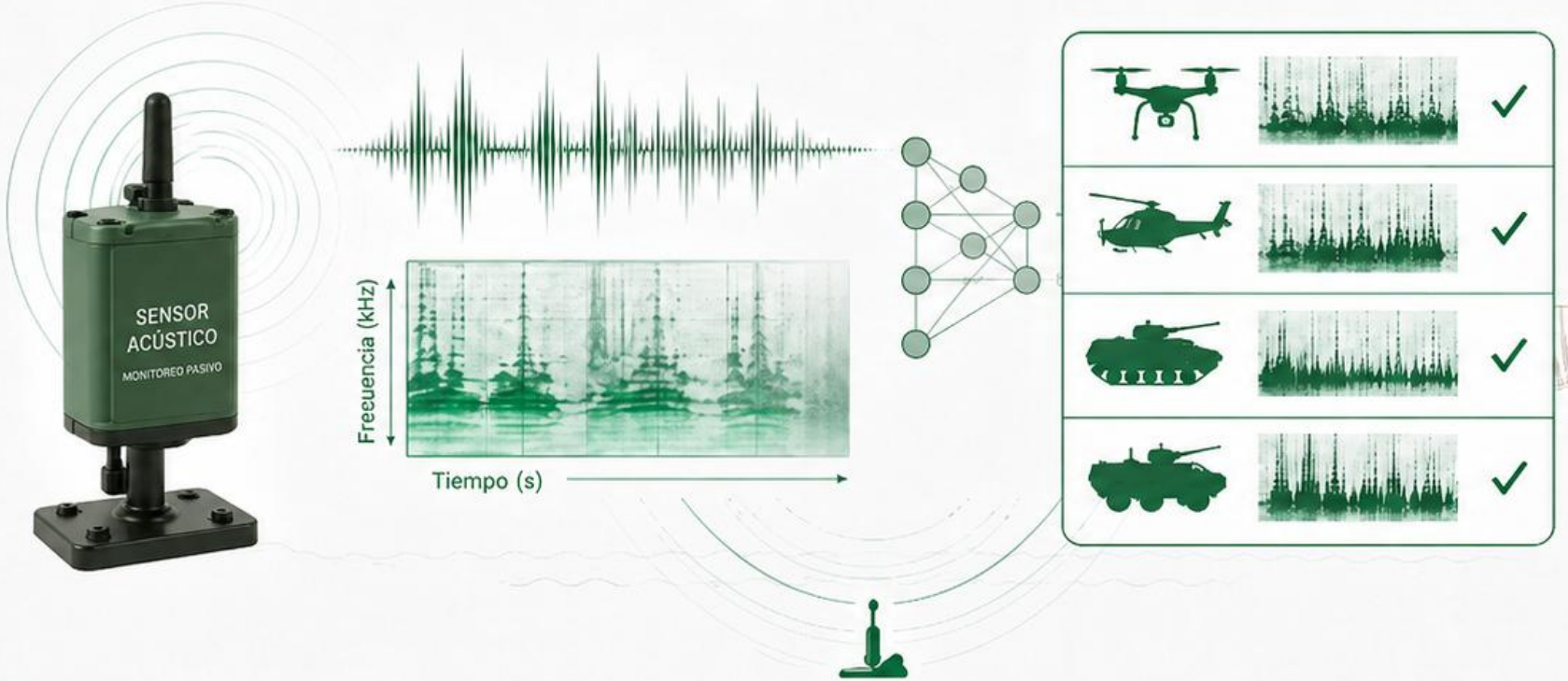
2. Emplear el sistema de representación jerárquica con umbrales por nivel para optimizar la toma de decisiones y aprovechar la fiabilidad de los nodos superiores.

3. Realizar pruebas de embebido para determinar el desempeño de hardware y el costo computacional, y establecer la viabilidad económica de su implementación.

4. En escenarios de sistemas, puede ser un punto de partida para implementar algoritmos TDoA (Time Difference of Arrival), obteniendo más información para la toma de decisiones.

ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

GRACIAS POR SU ATENCIÓN



SECCIÓN DE PREGUNTAS