Cross-Domain Drone Audio Classification: A Comparative Domain Transfer Study

Kalpana Algotar, Rita Chattopadhyay
Intel Corp
Chandler, Arizona, USA
{kalpana.a.algotar, rita.chattopadhyay}@intel.com

Srividya Bansal, Ajay Bansal
School of Computing and Augmented Intelligence
Arizona State University, Arizona, USA
{skbansa2, abansa10}@asu.edu

ABSTRACT

Unmanned Aerial Vehicles (UAVs) pose security concerns due to potential malicious use. Audio-based detection systems provide effective solutions, but face significant domain shift challenges in real-world deployment, which compromises their generalization capability across different domains in practical deployments. To tackle the domain shift problem, this study proposes a comprehensive domain adaptation framework evaluating five methodologies for drone detection across-domains: baseline machine learning (ML) such as SVM, baseline deep learning (DL) such as neural networks, enhanced ML with maximum mean discrepancy (MMD), DL integrated with MMD, and Domain advanced neural networks (DANN). The framework employs 234-dimensional feature extraction encompassing MFCCs, mel-spectrograms, chroma features, spectral characteristics, and temporal properties across source domains (FT2D, HPSS, RepetSim) and target domain (Public dataset). The system addresses class imbalance through SMOTE-Tomek resampling and incorporates four-fold augmentation techniques (Gaussian noise, time stretching, pitch shifting, volume variation) to enhance cross-domain transfer. MMD enables explicit domain alignment through multi-scale kernel computation and adaptive instance weighting, while DANN employs adversarial training with gradient reversal for domain-invariant feature learning. This study shows that MMD-enhanced ML outperforms (16-21%) baseline methods by 48-83 percentage points, achieving 99.75% accuracy in five-fold cross-validation. To verify the superiority and stability of our approach, paired t-tests were conducted. The results show significant improvements (p¡0.001) for both MMD+Aug ML and MMD+Aug DL compared with baseline, confirming consistently more efficient performance.

***Index Terms*—drone audio; uav classification; machine learning; deep learning; transfer learning; maximum mean discrepancy.**

I. INTRODUCTION

The rapid proliferation of commercial and recreational drone has introduced globally significant security and surveillance risks. The Federal Aviation Administration (FAA) reports over 1.5 million drone registrations in the United States [16], and the market is expected to expand from \$14 billion (2018) to roughly \$43 billion (2025) [6, 16]. The technology's rapid evolution is fueled by enhanced battery capability, smaller GPS components, and smarter flight controls[5, 16].

The rapid expansion of drone applications across agricultural, emergency response, and recreational domains has generated considerable security threats stemming from their easy accessibility. Unregistered drones violate restricted airspace and may transport dangerous materials or conduct illegal monitoring activities. Significant security incidents include an unauthorized drone incursion at the White House in 2015 [4], disruptions to diplomatic events involving Chancellor Angela Merkel in 2013, and the attempted drone-based assassination of President Nicolás Maduro in 2018 [3]. Traditional UAV detection methodologies encompass radar systems, radio frequency monitoring, computer vision, and acoustic analysis [1], each possessing distinct technical and operational constraints: optical systems fail under adverse weather conditions, radar technology cannot reliably detect small-profile aircraft, Radio Frequency based approaches are ineffective against autonomous flight modes, and acoustic detection suffers from ambient noise interference [2].

This study addresses the critical challenge of cross-domain acoustic UAV detection in real-world surveillance applications. Audio-based detection systems show effectiveness under controlled conditions [1, 2]. These systems experience significant performance degradation across different acoustic domains due to variations in recording setups, background noise levels, and the characteristics of the recording equipment [3, 4].

Most UAV detection studies concentrate on single-domain scenarios [5, 8], leaving a significant gap in understanding cross-domain generalization capabilities. Different recording environments, background noise conditions, and hardware configurations lead to shift in acoustic signatures that create domain shift challenges. Conventional ML approaches cannot adequately address these challenges [9, 10]. This limitation greatly hinders the real-world deployment of acoustic UAV detection systems in diverse operational environments [11].

To address these challenges, domain adaptation techniques have emerged as promising solutions to reduce the mismatch between training data and real-world deployment scenarios [12, 13]. However, current research suffers from limited comprehensive evaluation of domain adaptation strategies for

UAV acoustic detection, specifically in comparing statistical alignment techniques and adversarial training paradigms [14, 15].

This research addresses this gap by providing a comprehensive comparative assessment of domain adaptation techniques for cross-domain acoustic UAV detection. Our contributions include: (1) systematic evaluation of five distinct approaches ranging from traditional baselines to advanced domain adaptation methods, (2) comprehensive feature engineering pipeline optimized for cross-domain acoustic transfer, (3) a comprehensive experimental framework with statistical validation across multiple domains employing FT2D, HPSS, RepetSim [6] as source domains and Public dataset [7, 17] as target domain, and (4) practical insights for deploying acoustic UAV detection technologies in real-world scenarios.

Experimental validation demonstrates that domain adaptation techniques are not only primarily beneficial but essential for practical cross-domain deployment, with statistical analysis demonstrating significant performance gains compared to traditional methods [16]. These findings provide valuable knowledge for developing resilient acoustic surveillance systems capable of operating reliably across a wide range of environmental conditions.

The manuscript structure proceeds as follows: Section II reviews related work; Section III presents the proposed methodology; Section IV describes the experimental design; Section V analyzes performance results; Section VI concludes with key findings and Section VII outlines future research directions.

II. Related Work

Recent developments in Transfer Learning (TL) have tackled various challenges across multiple domains with substantial practical implications.

Yi et al. [8] develop a multitask learning framework for sound-based drone fault classification, achieving improved diagnostic accuracy through simultaneous learning of multiple fault detection tasks and demonstrating effective shared acoustic feature representations to identify mechanical anomalies in UAV systems. The approach utilizes spectral and temporal audio features to distinguish between normal operations and various failure modes including motor degradation, propeller imbalance, and structural vibrations, establishing the foundations for predictive maintenance systems through acoustic signature analysis.

Weiss et al. [9] examine TL methodologies for classification, regression, and clustering tasks, detailing homogeneous and heterogeneous TL while addressing negative transfer where source knowledge degrades target performance.

Sanodiya and Yao [10] propose unsupervised TL combining MMD with relative distance comparisons, eliminating target domain labels while preserving geometric structure.

Singhal et al. [11] conduct a survey of Domain Adaptation methods, classifying discrepancy-based, adversarial, and parameter-based techniques while examining domain shift and negative transfer issues across computer vision and Natural Language Processing (NLP) applications.

Schwendemann et al. [12] propose a Layered MMD framework for bearing fault diagnosis, obtaining 89% accuracy in laboratory-to-industrial transfers via hierarchical adaptation that addresses domain shifts.

Yang et al. [13] introduce discriminative TL that combines MMD with pseudolabel refinement for cross-city driving pattern recognition using confidence-based filtering approaches.

Chattopadhyay et al. [14] introduce multi-source domain adaptation for electromyography signals, enhancing muscle fatigue detection through utilization of electrical activity patterns across multiple subjects.

Yi Qin et al. [15] introduce Deep Joint Distribution Alignment (DJDA) for fault diagnosis, attaining 87% accuracy through alignment of data distributions, fault-specific characteristics, and data-label correlations across different environments.

III. Proposed Methodology and System Architecture

The framework employs systematic domain adaptation to address cross-domain drone detection, benchmarking conventional methods against three advanced adaptation techniques. Knowledge is transferred from multiple source domains (FT2D, HPSS, RepetSim)[6] to the target domain (Public dataset) [7, 17] using comprehensive feature extraction, balanced sampling, and comprehensive cross-validation framework.

Figure 1 illustrates the complete analytical pipeline, including multi-domain acoustic data acquisition, feature extraction, model-specific training, and performance evaluation. The architecture focuses on domain adaptation through statistical alignment and adversarial training techniques.

A. Signal Preprocessing and Normalization

Acoustic signals are processed with standardized protocols for consistency across datasets. Audio is trimmed to 3 seconds at 22,050 Hz, yielding 66,150-sample input vectors. Files below 0.5 seconds or corrupted are removed. Preprocessing uses StandardScaler for z-score normalization, replaces NaN values with zeros, and checks for data leakage between domains.

B. Feature Engineering

Feature extraction transforms raw acoustic signals into 234-dimensional representations specifically optimized for cross-domain transfer. The framework encompasses six analytical components:

Cepstral Analysis: Mel-Frequency Cepstral Coefficients (MFCC) features are extracted using 40 coefficients with statistical aggregations (mean, standard deviation, maximum, minimum), capturing spectral envelope characteristics across 160 dimensions for robust speech-like pattern recognition.

Spectral Analysis: Mel-spectrogram features from 20 frequency bands with statistical measures (mean, standard deviation) provide 40-dimensional frequency domain representations for capturing energy distribution across frequency ranges.

Harmonic Features: Chroma features are extracted by mapping audio signals into 12 distinct pitch class bins, each corresponding to one of the semitones in an octave. These bins collectively capture the tonal distribution present in the audio signal.

Spectral Characteristics: Spectral centroid, rolloff, and bandwidth measurements with statistical aggregations provide 6 dimensions for effectively characterizing spectral shape and distribution properties.

Temporal Properties: Zero-crossing rate (ZCR) and Root Mean Square (RMS) energy analysis are utilized to generate 4-dimensional time domain features, which capture temporal dynamics and signal energy variations.

The resulting high-dimensional feature space provides a comprehensive acoustic representation which facilitates diverse classification methods and enhances cross-domain adaptability.

C. Class Balancing and Data Augmentation

To address class imbalance and domain gaps, we implement a two-stage enhancement approach:

Synthetic Minority Oversampling Technique (SMOTE)-Tomek Balancing: This combined approach balances class distributions by using SMOTE to generate artificial samples for minority classes using k-nearest-neighbor interpolation and Tomek links to remove ambiguous boundary instances, resulting in improved data quality and classifier performance.

Multi-Technique Augmentation: Expands data diversity using: (1) Gaussian noise addition ($\sigma = 0.01$), (2) time scaling via feature adjustment (0.95-1.05×), (3) pitch adjustment (±2 MFCC bins), and (4) amplitude scaling (0.9-1.1×). This methodology increases the dataset by a factor of three, ensuring that acoustic integrity is preserved for domain transfer applications.

D. Domain Adaptation Methodologies

1) Baseline Methods:

ML Baseline: Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel (C=1.0, gamma='scale') trained directly on source domain data without domain adaptation mechanisms.

DL Baseline: Three-layer neural network architecture (234→256→256→128) with batch normalization, ReLU activation, and dropout regularization (p=0.3-0.4). Training employs AdamW optimizer (lr=0.001, weight_decay=0.01) over 100 epochs with early stopping.

2) MMD-Enhanced Machine Learning: MMD measures the distributional distance between source and target domains by comparing their feature representations in a reproducing kernel Hilbert space. This approach implements adaptive sample weighting through multi-scale MMD computation:

Domain Distance Computation:

$$\text{MMD}^2 = \frac{1}{|\Gamma|} \sum_{\gamma \in \Gamma} \left(\frac{\sum K_{ss}^{(\gamma)}}{n_s^2} + \frac{\sum K_{tt}^{(\gamma)}}{n_t^2} - \frac{2 \sum K_{st}^{(\gamma)}}{n_s \cdot n_t} \right) \quad (1)$$

where n_s and n_t denote the number of source and target samples respectively, $K_{ss}^{(\gamma)}$, $K_{tt}^{(\gamma)}$, and $K_{st}^{(\gamma)}$ represent RBF kernel matrices for source-source, target-target, and source-target pairs, calculated using the bandwidth set $\Gamma = [0.001, 0.01, 0.1, 1.0, 10.0]$.

Adaptive Instance Weighting: Sample importance weights are determined using similarity metrics based on MMD, which emphasize source samples exhibiting higher affinity to the target domain during training:

$$w_i = \text{clip}\left(\frac{\sum_\gamma \beta_\gamma \cdot \text{mean}(K_{st}^{(\gamma)}[i,:])}{|\Gamma|} \cdot |X_{\text{source}}|, 0.1, 10.0 \right) \quad (2)$$

where w_i represents the importance weight for source sample i, $\beta_\gamma = [0.1, 0.2, 0.4, 0.2, 0.1]$ provides multi-scale weighting coefficients corresponding to bandwidth values $\gamma \in [0.001, 0.01, 0.1, 1.0, 10.0]$, $K_{st}^{(\gamma)}[i,:]$ denotes similarity scores between source sample i and all target samples, and $|X_{\text{source}}|$ represents the total source sample count. Aggregating mean similarity enables adaptive weighting, thereby increasing the relevance of source samples that are more similar to the target domain. The application of clipping within the range [0.1, 10.0] is used to maintain stability during training.

Two-Stage Training: The fine-tuning process uses two training phases with different sample weights: first training only on source data using weights w_i, then training on both domains where source weights are decreased (0.3 × w_i) and target weights are increased (3.0×) to focus on target domain performance. The specific multipliers (0.3 and 3.0) represent an empirically effective balance between preserving useful source information and maximizing target-domain adaptation performance while ensuring stability throughout training.

3) MMD-Integrated Deep Learning: A three-phase training strategy combining classification and domain alignment objectives:

Phase 1: Source pre-training for 100 epochs using standard cross-entropy loss to establish baseline feature representations.

Phase 2: Domain adaptation occurs over 50 epochs with composite loss function:

$$\mathcal{L}\text{combined} = \mathcal{L}\text{classification} + \lambda \cdot \text{MMD}(\mathcal{F}_s, \mathcal{F}_t) \quad (3)$$

where $\lambda = 0.3$ balances classification accuracy and domain alignment, $\mathcal{F}_s$ and $\mathcal{F}_t$ represent learned feature representations from source and target domains respectively.

Phase 3: Target fine-tuning is performed across 30 epochs using available labeled target samples to optimize target domain performance.

The progressive reduction in epoch count (100→50→30) reflects the diminishing training requirements as the model progresses from general feature learning to domain-specific adaptation. This schedule also mitigates overfitting risks during later phases when labeled target data is limited.

4) Domain-Adversarial Neural Networks (DANN): Adversarial training utilises shared feature extraction with distinct task and domain classifiers to learn domain-invariant representations:

Architecture: The model consists of a feature extractor (234→256→256→128), a task classifier (128→64→2), and a domain classifier (128→64→32→2). ReLU activations and dropout regularizations are used to promote robust learning.

Progressive Alpha Scheduling:

$$\alpha = \min\left(1, \frac{2}{1+\exp(-5p)} - 1\right), \quad p = \frac{\text{epoch}}{N_{\text{epochs}}} \tag{4}$$

where p represents the training progress ranging from 0 to 1, and α controls the strength of gradient reversal, starting at a minimal value and gradually increasing to its maximum ($\alpha = 1$) for stable adversarial optimization.

Combined Loss Function:

$$\begin{aligned}\mathcal{L}_{\text{DANN}} = \mathcal{L}_{\text{task}} + 0.5 \cdot \text{GRL}_{\alpha}(&\mathcal{L}_{\text{domain\_source}} \\ &+ \mathcal{L}_{\text{domain\_target}})\end{aligned} \tag{5}$$

where GRL_{α} represents gradient reversal with coefficients α, creating adversarial optimization between feature extractor and domain classifier for domain-invariant representations. Training spans 150 epochs for adversarial learning followed by 50 epochs of target fine-tuning.

E. Cross-Validation and Evaluation Framework

The validation protocol employs 5-fold stratified cross-validation to maintain balanced class distributions within each fold. All methods apply uniform random seed initialization to facilitate equitable comparative assessment. Statistical significance of observed performance enhancements is examined using paired t-tests ($\alpha = 0.05$). Performance metrics like accuracy, precision, recall, and F1-score are presented as the mean ± standard deviation across all validation folds:

$$\begin{aligned}\text{Accuracy} &= \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \\ \text{Precision} &= \frac{\text{TP}}{\text{TP}+\text{FP}} \\ \text{Recall} &= \frac{\text{TP}}{\text{TP}+\text{FN}} \\ \text{F1-Score} &= \frac{2\cdot\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}}\end{aligned} \tag{6}$$

Domain Adaptation Framework Architecture

Data Processing Layer

Multi-Domain Audio → Feature Extraction (234D) → SMOTE + Augmentation

Domain Adaptation Layer

Baseline	Adaptation
SVM	MMD-Enhanced ML
Neural Net	MMD-Integrated DL
	Adversarial DANN

Evaluation Layer

5-Fold Cross-Validation → Statistical Analysis → Performance Comparison

Fig. 1. Domain Adaptation Framework Architecture

TABLE I
PERFORMANCE COMPARISON OF DOMAIN ADAPTATION METHODS

Method	Overall				Non-Drone		Drone	
	Acc	Prec	Rec	F1	Prec	Rec	Prec	Rec
Baseline ML	16.62	0.125	0.166	0.143	0.251	0.344	0.000	0.000
Baseline DL	20.94	0.147	0.209	0.173	0.321	0.440	0.000	0.000
ML+MMD	**99.75**	**0.998**	**0.998**	**0.998**	**1.000**	**0.996**	**0.994**	**1.000**
DL+MMD	69.56	0.706	0.696	0.691	0.750	0.585	0.682	0.802
DANN	93.00	0.939	0.930	0.930	**1.000**	0.863	0.878	**1.000**

Note: Values shown without std. dev. for space. DA methods show 48-83pp improvement.

IV. EXPERIMENTAL FRAMEWORK

A. Dataset Configuration and Domain Setup

The multi-domain framework employs three source domains (FT2D, HPSS, RepetSim) with 200 samples per class, totaling 1,200 labeled samples. The target domain integrates ESC-10 environmental sounds (crackling fire, sneezing, rooster crowing, crying baby, chainsaw) with Drone A recordings, creating realistic acoustic interference scenarios with 400 samples (150 drone, 250 non-drone). Only 80 samples (20%) are labeled for training while 320 samples (80%) remain unlabeled for evaluation, simulating practical deployment where source domains provide abundant labeled data but target domains offer limited supervision. This configuration enables robust assessment of cross-domain generalization capabilities.

B. Data Processing Pipeline

Audio preprocessing follows standardized steps outlined in Section III-A, including 3-second normalization, 22,050 Hz sampling, and extraction of 234-dimensional features. Ensure data integrity by handling NaN values, applying z-score normalization, and checking for data leakage via duplicate and similarity analysis. To address class imbalance and domain gaps, SMOTE-Tomek resampling and four augmentation methods — noise injection, time-stretching, pitch-shifting, and volume variation. These methods are carefully selected to maintain acoustic characteristics required for cross-domain transfer.

C. Experimental Protocol

The experimental evaluation assesses five methods outlined in Section III-D4: two baselines (ML and DL) and three domain adaptation techniques (MMD-Enhanced ML, MMD-Integrated DL, and DANN).

1) Model Specifications:

Baselines: SVM with RBF kernel (C=1.0, gamma='scale') and three-layer neural network (234→256→256→128) with batch normalization and dropout rates between 0.3 and 0.4.

Domain Adaptation: MMD-Enhanced ML uses adaptive instance weighting with two-stage training; MMD-Integrated DL employs three-phase training ($\lambda = 0.3$); DANN uses adversarial training with gradient reversal.

2) MMD Implementation: MMD employs RBF kernels with bandwidth set $\Gamma = \{0.001, 0.01, 0.1, 1.0, 10.0\}$ and corresponding weighting coefficients $\beta_{\gamma} = \{0.1, 0.2, 0.4, 0.2, 0.1\}$. Instance weights are constrained to the interval $[0.1, 10.0]$. A two-stage training is adopted, employing source weight reduction by a factor of $(0.3\times)$ and target amplification by a factor of $(3.0\times)$.

TABLE II
STATISTICAL SIGNIFICANCE ANALYSIS OF METHOD COMPARISONS

Comparison	Balance Data (YES)	Balance Data (NO)	Diff.	t-stat	p-value
SMOTE Effect on ML	16.63±0.78	16.81±0.78	-0.19	-1.00	0.374
SMOTE Effect on DL	20.94±2.86	22.44±1.47	-1.50	-1.67	0.170
MMD+Aug Effect on ML	99.75±0.13	16.63±0.78	+83.13	204.01	<0.001
MMD+Aug Effect on DL	69.56±8.13	20.94±2.86	+48.63	16.83	<0.001
ML vs DL (MMD+Aug)	99.75±0.13	69.56±8.13	+30.19	7.42	0.002
DANN vs MMD-DL	93.00±1.77	69.56±8.13	+23.44	6.97	0.002
DANN vs Baseline ML	93.00±1.77	16.81±0.78	+76.19	122.82	<0.001
DANN vs Baseline DL	93.00±1.77	22.44±1.47	+70.56	162.11	<0.001

TABLE III
EXPERIMENTAL CONFIGURATION SUMMARY

Dataset & Processing	
Source Domains	FT2D, HPSS, RepetSim [6]
Target Domain	Public (ESC-10 + Drone A) [7, 17]
Total Samples	1600 (1200 source + 400 target)
Feature Dimensions	234 (MFCC, spectral, temporal)
Audio Duration	3.0 seconds, 22,050 Hz
Methods Compared	
Baseline	SVM, Neural Network
Domain Adaptation	MMD-ML, MMD-DL, DANN
Enhancement	SMOTE + 4× Augmentation
Validation	
Cross-Validation	5-Fold Stratified
Metrics	Accuracy, Precision, Recall, F1
Statistical Test	Paired t-test (= 0.05)

3) Training Configuration: Training follows method-specific protocols from Section III-D4: DL Baseline uses AdamW optimizer (learning rate=0.001) for 100 epochs; MMD-Integrated DL uses three-phase training comprising 100, 50, and 30 epochs; DANN uses adversarial training for 150 epochs followed by a fine-tuning of 50 epochs. For all DL methods, a batch size of 32 is consistently applied.

4) Controlled Experimental Conditions: Experiments were performed using Python 3.13.3 and standard ML libraries on identical hardware. Reproducibility was ensured through systematic seeding (42 + fold index) within a 5-fold cross-validation setup. Each method was executed under identical configurations to facilitate fair comparison.

5) Evaluation and Validation Framework: The evaluation framework assesses target domain generalization, cross-validation stability, computational efficiency, and statistical significance using 5-fold stratified cross-validation with paired t-test ($\alpha = 0.05$). Each fold maintains class proportions within both source and target domains to prevent imbalance-related bias. Performance metrics include accuracy, precision, recall, and F1-score, with emphasis on F1-score for balanced assessment of both precision and recall. Results are reported as mean ± standard deviation across folds with stability analysis.

V. EXPERIMENTAL OUTCOMES AND ANALYSIS

Table I presents comprehensive performance metrics for all evaluated methods, revealing stark contrasts between baseline and domain adaptation approaches.

A. Performance Analysis

Baseline methods demonstrate catastrophic cross-domain failure, achieving only 16.62% (ML) and 20.94% (DL) accuracy. Most critically, both baseline methods achieve 0.000 precision and recall for drone detection, meaning they would miss every drone intrusion in real-world deployment with environmental interference (ESC-10 sounds: fire, sneezing, chainsaw, etc.).

Domain adaptation methods yield transformative improvements: MMD-enhanced ML achieves exceptional 99.75% accuracy (83.13 pp improvement), DANN reaches 93.00% accuracy, and MMD-integrated DL attains 69.56% accuracy. For drone detection specifically, MMD-enhanced ML achieves 99.4% precision with 100% recall, while DANN demonstrates 87.8% precision with 100% recall, ensuring no missed threats.

B. Statistical Significance and Method Comparison

Paired t-tests confirm all domain adaptation improvements are highly significant ($p < 0.001$). MMD-enhanced ML shows the strongest effect (t-statistic = 204.01), with DANN also demonstrating substantial improvements over both baselines (t-statistics: 122.82 and 162.11). MMD-enhanced ML significantly outperforms other adaptation methods: 30.19 pp better than MMD-DL (t = 7.42, $p = 0.002$) and superior to DANN in overall accuracy while matching its perfect drone recall.

The exceptional performance of MMD-enhanced ML stems from multi-scale MMD alignment, adaptive instance weighting (0.1× to 10.0×), and SVM stability. MMD-integrated DL struggles with competing training objectives, while DANN achieves strong but variable results due to adversarial training complexity.

C. Computational Efficiency and Deployment Readiness

MMD-enhanced ML requires only 2.3 minutes per fold versus 8.7 minutes (DANN) and 6.1 minutes (MMD-DL), providing 3.8× computational advantage. Combined with 99.75% accuracy, 100% drone recall, and minimal hyperparameter sensitivity, MMD-enhanced ML represents the optimal solution for security-critical applications.

The complete baseline failure (0% drone detection) validates that domain adaptation is essential, not optional, for real-world acoustic surveillance. MMD-enhanced ML's near-perfect performance demonstrates reliable cross-domain detection is achievable for operational deployment across diverse environmental conditions.

To facilitate reproducibility and further research in the field, we have released our source code at https://shorturl.at/jn8dK and made the datasets available through [6, 7, 17].

VI. CONCLUSIONS

This comprehensive study evaluates domain adaptation techniques for cross-domain drone detection, demonstrating that MMD-enhanced ML achieves superior performance with 99.75% accuracy, exceptional stability, and computational efficiency compared to deep learning approaches. The methodology substantiates that explicit statistical alignment via MMD outperforms adversarial techniques such as DANN (93.00%) in acoustic surveillance environments, challenging conventional assumptions about deep learning superiority in domain adaptation tasks.

Performance gains of 48.62 to 83.13 percentage points over baseline models confirm that domain adaptation is essential for operational systems. Critically, baseline methods achieve 0% drone detection with environmental sounds, representing complete system failure that would compromise security operations, while MMD-enhanced ML maintains near-perfect performance (99.4% drone precision, 100% drone recall, 100% non-drone precision), ensuring no missed threats while minimizing false alarms.

For practical deployment, MMD-enhanced ML provides the optimal solution with near-perfect accuracy, high reliability (±0.12% variation), and 3.8× faster computation than competing methods. The comprehensive 234-dimensional feature engineering pipeline effectively bridges acoustic domain gaps across diverse recording environments, with statistical significance ($p < 0.001$) confirming robust generalization capabilities. These findings establish the practical viability of traditional ML enhanced with explicit domain alignment for critical security applications, providing a strong foundation for reliable cross-domain acoustic surveillance systems capable of real-world deployment.

VII. FUTURE WORK

Further research should address unsupervised domain adaptation scenarios in which no supervision from the target domain is available, as well as investigate multi-source domain adaptation methods that utilize information from multiple source domains concurrently. Moreover, expanding the current binary detection framework to enable multi-class drone classification has the potential to advance real-world security applications by facilitating the identification of specific drone types and corresponding threat levels.

The results suggest several areas for future research, including developing systems that automatically adapt to changing sound environments and improving feature extraction for real-time detection. Testing the MMD-enhanced method on different audio tasks and acoustic settings would help prove its usefulness beyond just drone detection.

Incorporating edge computing technology and streamlined model designs would enable deployment on resource-limited devices within distributed monitoring networks. In addition, developing combined approaches that use MMD's consistent performance together with deep learning's advanced feature learning could result in more durable acoustic detection systems that work across various domains.

ACKNOWLEDGEMENTS

The authors acknowledge the use of public UAV, background noise and ESC-10 datasets. Chat-GPT was employed for English language editing. The authors have reviewed and approved all generated content, taking full responsibility for the final text.

REFERENCES

[1] Xiufang Shi, Chaoqun Yang, Weige Xie, Chao Liang, Zhiguo Shi, and Jiming Chen. Anti-drone system with multiple surveillance technologies: Architecture, implementation, and challenges. IEEE Communications Magazine, 56(4):68 – 74, 2018.

[2] Rahman MH, Sejan MAS, Aziz MA, Tabassum R, Baik J-I, Song H-K. A Comprehensive Survey of Unmanned Aerial Vehicles Detection and Classification Using Machine Learning Approach: Challenges, Solutions, and Future Directions. Remote Sensing. 2024; 16(5):879. https://doi.org/10.3390/rs16050879.

[3] Wang, Y., Fagiani, F. E., Ho, K. E., & Matson, E. T. (2022, February). A Feature Engineering Focused System for Acoustic UAV Payload Detection. In ICAART (3) (pp. 470-475).

[4] Utebayeva, D.; Ilipbayeva, L.; Matson, E.T. Practical Study of Recurrent Neural Networks for Efficient Real-Time Drone Sound Detection: A Review. Drones 2023, 7, 26. https://doi.org/10.3390/drones7010026.

[5] Miah, A. (2020). Drones: The brilliant, the bad and the beautiful. Emerald Group Publishing.

[6] Algotar, K., Chattopadhyay, R., Bansal, S., & Bansal, A. (2024, December). Drone sound classification using chromagram features: A novel approach. In 2024 Eighth IEEE International Conference on Robotic Computing (IRC) (pp. 34-40). IEEE.

[7] Karol J Piczak. Esc: Dataset for environmental sound classification. In Proceedings of the 23rd ACM international conference on Multimedia, pages 1015–1018, 2015.

[8] Yi W, Choi JW, Lee JW. Sound-based drone fault classification using multitask learning. arXiv preprint arXiv:2304.11708. 2023.

[9] Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. Journal of Big Data, 3(9).

[10] Sanodiya, R.K. and Yao, L. (2020). Unsupervised transfer learning via relative distance comparisons. IEEE Access, 8, pp.110290-110305.

[11] Singhal, P., Walambe, R., Ramanna, S., and Kotecha, K. (2023). Domain Adaptation: Challenges, Methods, Datasets, and Applications. IEEE Access, 11, pp. 6973-7020.

[12] Schwendemann, S., Amjad, Z., & Sikora, A. (2021). Bearing fault diagnosis with intermediate domain based layered maximum mean discrepancy: A new transfer learning approach. Engineering Applications of Artificial Intelligence, 105, 104415.

[13] Yang, L., Li, M., Shen, C., Hu, Q., Wen, J. and Xu, S., 2020. Discriminative transfer learning for driving pattern recognition in unlabeled scenes.IEEE Transactions on Cybernetics, 52(3), pp.1429-1442.

[14] Chattopadhyay, R., Sun, Q., Fan, W., Davidson, I., Panchanathan, S., & Ye, J. (2012). Multisource domain adaptation and its application to early detection of fatigue. ACM Transactions on Knowledge Discovery from Data (TKDD), 6(4), 1-26.

[15] Qin, Y., Qian, Q., Luo, J., & Pu, H. (2022). Deep joint distribution alignment: A novel enhanced-domain adaptation mechanism for fault transfer diagnosis. IEEE Transactions on Cybernetics, 53(5), 3128-3138.

[16] S. Salman, J. Mir, M. T. Farooq, A. N. Malik and R. Haleemdeen, ”Machine Learning Inspired Efficient Audio Drone Detection using Acoustic Features,” 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST), Islamabad, Pakistan, 2021, pp. 335-339, doi: 10.1109/IBCAST51254.2021.9393232.

[17] https://github.com/agrija9/Environmental-Sound-Classification-ESC-using-neural-networks-and-other-classifiers