Passive Defence
Vol. 16, No. 4, Winter 2026, Serial No. 64
ISSN: 2980-8030 , E-ISSN: 2008-6849

# Acoustic Shield: Lightweight Neural Network for Audio-Based Drone Detection and Classification

**Fatemeh Alimadadi**
Correspondence: M.Sc. in Artificial Intelligence & Robotics, Faculty of Computer Engineering, Malek ashtar University, Tehran, Iran. Email Address: afatem041@gmail.com

ARTICLE INFO

***Article history:***

Article Type: Research paper

Received: 26 July 2025

Received in revised form: 20 August 2025

Accepted: 20 Septembe r 2025

Available online: 20 January 2026

***Keywords:***

Real-time Drone Detection

Acoustic Drone Classification

Signal Processing

Lightweight Neural Network

Drone Recognition in Complex Environments

ABSTRACT

The widespread use of UAVs has intensified the need for advanced security measures to prevent unauthorized airspace intrusions and mitigate potential threats. Audio-based drone detection systems, which leverage the unique acoustic signatures of drones, offer a viable solution for remote monitoring and surveillance in sensitive environments such as military zones, secured facilities, and urban areas. In this paper, we propose a powerful framework based on a lightweight deep neural network architecture derived from ConvNeXt for accurate and real-time acoustic drone detection and classification. The proposed model is trained and evaluated on a diverse collection of drone and environmental audio recordings to ensure high performance and generalization across various conditions. Experimental results demonstrate the model's outstanding ability to accurately detect and classify a wide range of drones in acoustically complex environments, while also maintaining low latency suitable for real-time applications. Moreover, the proposed multi-task model outperforms existing methods and proves to be a practical solution for deployment in resource-constrained audio surveillance systems. Despite achieving impressive accuracy about 99/55% in detection and 99/21% in classification, the model contains only 0/62 million trainable parameters (625542), making it highly suitable for integration into low-power, real-time environmental monitoring systems.

**Cite this article:** F. Alimadadi, "Acoustic Shield: Lightweight Neural Network for Audio-Based Drone Detection and Classification," Journal of Passive Defence, vol. 16, no. 4, pp. 121-135, 2026. **DOI: https://doi.org/10.47176/PD.2026.1559**



**Publisher**: Imam Hossein University.

## Introduction

Unmanned aerial vehicles (UAVs) have become an integral component of modern military and security capabilities due to their ability to operate autonomously or remotely, perform missions in both day and night conditions, cover long and short ranges, and function under diverse weather scenarios. These capabilities have simultaneously transformed low-cost and agile UAVs into serious threats to sensitive areas, strategic infrastructures, and urban environments. Consequently, there is a growing demand for robust, efficient, and rapidly deployable counter-UAV technologies capable of timely detection, tracking, and threat mitigation.

Among existing solutions, non-invasive, passive, and low-cost systems that do not rely on electromagnetic emissions have attracted significant attention, particularly in operational environments. Acoustic-based UAV detection systems leverage the distinctive sound signatures generated by UAV motors and propellers, enabling identification even without line-of-sight and under challenging environmental conditions. Recent advances in deep learning and lightweight neural networks have further enabled the design of accurate, low-latency models suitable for deployment on resource-constrained edge devices.

Despite their advantages, acoustic methods face challenges such as high environmental noise, inter-class acoustic similarity, distance- and angle-dependent signal variations, and limited availability of diverse real-world datasets. To address these challenges, this paper proposes a lightweight deep learning model based on a ConvNeXt architecture, designed within a multi-task learning framework to simultaneously detect UAV presence and classify UAV types. The proposed approach emphasizes efficient learning from limited data, real-time processing capability, and robust performance in realistic and noisy environments, making it well-suited for practical military and security surveillance applications.

## Result & Discussion

The proposed acoustic-based UAV detection and classification framework was evaluated using diverse and challenging datasets to assess both accuracy and practical robustness. Experimental results demonstrate that the lightweight ConvNeXt-based multi-task model achieves reliable performance across different acoustic feature representations, while maintaining low computational complexity suitable for real-time deployment. Among the evaluated features, log-Mel spectrograms and MFCCs consistently provided superior discrimination capability compared to raw FFT features, particularly under noisy and long-range recording conditions.

The multi-task learning formulation, which jointly optimizes UAV presence detection and type classification, led to more stable feature representations and improved overall accuracy compared to single-task baselines. Despite the very compact model size (approximately 0.62 million parameters), the proposed architecture preserved strong modeling capacity, outperforming or matching heavier models reported in related studies.

Table (1) reports the UAV type classification accuracy on the reference dataset using an 80% training and 20% evaluation split. This table demonstrates that the proposed lightweight ConvNeXt-based model significantly outperforms the baseline method, particularly when MFCC and Log-Mel features are used. The results confirm that, even with a compact architecture, the proposed model achieves very high classification accuracy, establishing a strong baseline performance under standard training conditions.

Table (1): UAV type classification accuracy

| Model | accuracy |
|---|---|
| Baseline | 93.6 |
| Proposed Model (FFT) | 66 |
| Proposed Model (LogMel) | 96.99 |
| Proposed Model (MFCC) | **99.21** |

Table (2) presents detailed UAV detection results under the same 80–20 split, including Accuracy, Precision, Recall, and F1-score. These metrics highlight the reliability of the proposed model in distinguishing UAV sounds from environmental noise. The consistently high F1-scores, especially for MFCC-based inputs, indicate balanced performance with low false-alarm and miss-detection rates, validating the effectiveness of the multi-task learning framework in the baseline scenario.

Table (1): UAV detection accuracy

| Model | accuracy |
|---|---|
| Baseline | 97.7 |
| Proposed Model (FFT) | 94.5 |
| Proposed Model (LogMel) | 97.91 |
| Proposed Model (MFCC) | **99.55** |

Notably, the model exhibited robust generalization to unseen UAV configurations. Evaluation on a modified and previously unseen Race UAV, with altered battery, propellers, and operating voltage, showed high detection reliability, indicating that the learned representations capture intrinsic acoustic characteristics rather than overfitting to specific operating conditions. Furthermore, the inclusion of diverse environmental noises and acoustically similar non-UAV sources, such as aircraft and helicopters, significantly reduced false alarms and improved real-world reliability.

## Conclusion

A comprehensive comparison of acoustic feature representations showed that MFCC and log-Mel spectrograms consistently outperform FFT-based features, particularly in low-data regimes and noisy environments. Importantly, the proposed system exhibited strong generalization capability to previously unseen UAV configurations, confirming its robustness against variations in hardware, operating conditions, and acoustic signatures. The inclusion of acoustically similar non-UAV sources, such as helicopters and fixed-wing aircraft, further validated the model's ability to reduce false alarms in realistic deployment scenarios. From a computational perspective, the measured inference and feature-extraction times confirm that the proposed approach is suitable for real-time operation on embedded and edge devices. Overall, the results indicate that the proposed lightweight ConvNeXt-based model offers a practical and scalable solution for passive UAV surveillance systems in military and security applications.

پدافند غیرعامل

سال پانزدهم، شماره ۴، زمستان ۱۴۰۴، (پیاپی ۶۴): صص ۱۳۵-۱۲۱

شاپای چاپی: ۶۹۴۹-۲۰۰۸ | شاپای الکترونیکی: ۸۰۳۰-۲۹۸۰

# سپر صوتی: شناسایی و طبقه‌بندی صوتی پهپاد با شبکه عصبی سبک وزن

**فاطمه علی‌مددی**

کارشناسی ارشد هوش مصنوعی و رباتیکز، دانشکده کامپیوتر، دانشگاه صنعتی مالک اشتر، تهران، ایران (نویسنده مسئول). رایانامه: afatem041@gmail.com

## چکیده

استفاده گسترده از پهپادها، نیاز به تدابیر امنیتی پیشرفته برای جلوگیری از نفوذ غیرمجاز به حریم هوایی و مقابله با تهدیدات امنیتی را افزایش داده است. سامانه‌های شناسایی صوتی پهپاد، با بهره‌گیری از ویژگی‌های صوتی منحصربه‌فرد این پرنده‌ها، امکان پایش و نظارت از راه دور را در محیط‌های حساس نظیر مناطق نظامی، اماکن حفاظت‌شده، و فضاهای شهری فراهم می‌سازند. در این مقاله، ما چارچوبی قدرتمند مبتنی بر معماری شبکه عصبی عمیق بر اساس شبکه ConvNeXt برای شناسایی و طبقه‌بندی صوتی پهپادها ارائه می‌شود. مدل پیشنهادی با بهره‌گیری از معماری سبک‌وزن، بر روی مجموعه‌ای متنوع از داده‌های صوتی آموزش دیده و ارزیابی شده است تا دقت عملکرد و قابلیت تعمیم آن در شرایط گوناگون تضمین گردد. نتایج آزمایش‌ها نشان‌دهنده عملکرد بسیار بالای مدل پیشنهادی در تشخیص دقیق و طبقه‌بندی انواع مختلف پهپادها در محیط‌های پیچیده صوتی است و نشان می‌دهد که مدل طراحی‌شده علاوه بر دقت بالا، دارای سرعت پردازش مناسب برای استفاده در کاربردهای بلادرنگ نیز می‌باشد. همچنین مدل چند وظیفه‌ای ما نسبت به روش‌های موجود برتری داشته و می‌تواند به‌عنوان یک راهکار عملی در سامانه‌های نظارتی صوتی و سازگار با منابع سخت‌افزاری محدود مورد استفاده قرار گیرد. با وجود عملکرد دقیق که موفق شد دقت شناسایی را تا ۹۹/۵۵ درصد و دقت طبقه‌بندی را تا ۹۹/۲۱ درصد افزایش دهد، حجم مولفه‌های قابل آموزش تنها ۰/۶۲ میلیون (۶۲۵۵۴۲ مولفه) می‌باشد که این مدل را برای پیاده‌سازی در سامانه‌های نظارت محیطی کم‌مصرف مناسب می‌سازد.

## مشخصات مقاله

**تاریخچه مقاله:**

نوع مقاله: علمی پژوهشی

دریافت: ۱۴۰۴/۰۵/۰۴

بازنگری: ۱۴۰۴/۰۵/۲۹

پذیرش: ۱۴۰۴/۰۶/۲۹

ارائه آنلاین: ۱۴۰۴/۱۰/۱۰

**کلیدواژه‌ها:**

شناسایی بلادرنگ پهپاد

طبقه‌بندی صوتی پهپادها

پردازش سیگنال

شبکه عصبی سبک‌وزن

تشخیص پهپاد در محیط‌های پیچیده

**استناد**: علی‌مددی، فاطمه، " سپر صوتی: شناسایی و طبقه‌بندی صوتی پهپاد با شبکه عصبی سبک وزن"، نشریه پدافند غیرعامل، دوره ۱۶، شماره۴، صفحات ۱۳۵-۱۲۱، ۱۴۰۴. DOI: https://doi.org/10.47176/PD.2026.1559



**ناشر**: دانشگاه جام امام حسین(ع).

## ۱- مقدمه

استفاده از پرنده‌های بدون سرنشین امروزه نقش بسیار مهمی در مجموعه قدرت نظامی کشورهای مختلف پیدا کرده و آنچه نظر مدافعان و مسئولان نظامی کشورها را به خود جلب کرده است، توان اجرای عملیات در شب و روز در مناطق دور و نزدیک بر ضد اهداف ساکن و متحرک و در تمام شرایط آب‌وهوایی و امکان پروازهای هدایت‌شونده از دور و یا تمام‌خودکار است [۱]. این قابلیت‌ها، پهپادها را به تهدیداتی جدی برای امنیت مناطق حساس، تأسیسات راهبردی و محیط‌های شهری تبدیل کرده‌اند.

با افزایش روزافزون استفاده از پهپادهای کم‌هزینه و چابک، نیاز به توسعه‌ی فناوری‌های مقاوم و کارآمد برای مقابله با این پرنده‌ها به منظور شناسایی، ردیابی و خنثی‌سازی تهدیدات احتمالی، بیش از پیش احساس می‌شود. در این میان، توسعه‌ی سامانه‌هایی غیرتهاجمی، غیرفعال (فاقد نشر امواج الکترومغناطیسی) و کم‌هزینه که از قابلیت پیاده‌سازی در محیط‌های عملیاتی برخوردار باشند و در شرایط واقعی با دقت و پایداری بالا عمل کنند (نظیر روش‌های شناسایی بر پایه‌ی امضای صوتی پهپادها)، به یکی از اولویت‌های راهبردی در حوزه‌ی دفاعی و امنیتی بدل شده است.

امضای صوتی خاص هر پهپاد، ناشی از ساختار فیزیکی و ویژگی‌های موتور و پروانه آن، امکان تفکیک و شناسایی این پرنده‌ها را حتی در فواصل نسبتا دور فراهم می‌کند. در این میان، بهره‌گیری از الگوریتم‌های یادگیری عمیق و شبکه‌های عصبی سبک‌وزن می‌تواند زمینه‌ساز طراحی سامانه‌هایی با دقت بالا، زمان کوتاه پاسخ‌دهی و قابلیت استقرار بر روی سخت‌افزارهای محدود را فراهم کند. با توسعه سخت‌افزاری و مکانیکی سامانه‌های ضد هوایی در کنار مدل‌های حاضر،می‌توان به سامانه‌های ضد هوایی کاملا هوشمند که قابلیت تشخیص به موقع و ساقط‌سازی خودکار پهپادها را دارا هستند، دست پیدا کرد [۲]. در روش‌های بینایی ماشین ارائه شده [۲،۳]، همچنان چالش‌های عملکرد محدود در شرایط نوری نامناسب و حساسیت بالا به موانع و پوشش محیط که خط دید را مسدود و منجر به کاهش دقت در فواصل زیاد یا زاویه‌های غیرمستقیم دید را می‌کند، وجود دارد.

با وجود مزایای متعدد روش‌های شناسایی صوتی، این رویکرد با چالش‌هایی اساسی نیز مواجه است. از جمله‌ی این چالش‌ها می‌توان به نویزهای محیطی بالا، شباهت صوتی میان برخی مدل‌های پهپاد، تغییرات صوتی ناشی از فاصله یا زاویه منبع صوت و همچنین نیاز به پردازش بلادرنگ با منابع سخت‌افزاری محدود اشاره کرد. افزون بر این، داده‌های صوتی پهپاد معمولا دارای تنوع محدود بوده و در شرایط واقعی ممکن است دستخوش اعوجاج یا تداخل فرکانسی شوند که دقت سامانه‌های شناسایی را کاهش می‌دهد.

برای غلبه بر این محدودیت‌ها، در این پژوهش یک مدل یادگیری عمیق سبک‌وزن مبتنی بر معماری ConvNeXt [۴] طراحی و پیاده‌سازی شده است که ضمن حفظ دقت بالا، از توانایی پردازش بلادرنگ بر روی پردازنده‌های کم‌مصرف و قابل‌حمل نیز برخوردار است. این مدل با استفاده از یک چارچوب چندوظیفه‌ای قادر است به‌صورت هم‌زمان حضور پهپاد را تشخیص داده و نوع آن را طبقه‌بندی کند.

با انجام آزمایشات، دریافتیم که مدل قادر به آموزش موثر با حجم بسیار محدودی از داده‌ها بوده و در عین حال، در آزمون‌های تجربی بر روی پهپاد دیده‌نشده دقت بسیار بالایی را نشان داده است. با آموزش و ارزیابی بر روی مجموعه‌داده‌های صوتی متنوع با مقایسه چندین روش استخراج ویژگی صوتی، سعی شده است پایداری عملکرد مدل در شرایط واقعی تضمین گردد.

ترکیب معماری سبک‌وزن، یادگیری کارآمد از داده‌های محدود و عملکرد دقیق در محیط‌های چالش‌برانگیز، مدل پیشنهادی را به گزینه‌ای مناسب برای سامانه‌های نظارتی و امنیتی در کاربردهای نظامی تبدیل می‌کند، به‌ویژه در شرایطی که نیاز به استقرار سریع، پردازش محلی و واکنش فوری وجود دارد.

در این بخش، به ذکر مقدمه‌ای بر اهمیت فزاینده‌ی شناسایی پهپادها، چالش‌های موجود در روش‌های صوتی و ضرورت توسعه‌ی سامانه‌های سبک، بلادرنگ و دقیق پرداختیم. همچنین، انگیزه‌ها و اهداف اصلی پژوهش حاضر، در راستای طراحی یک مدل سبک‌وزن مبتنی بر یادگیری عمیق برای شناسایی و طبقه‌بندی صوتی پهپادها، تبیین شد. در ادامه، مروری بر کارهای مرتبط در زمینه‌ی شناسایی و طبقه‌بندی صوتی پهپاد ارائه می‌شود. بخش سوم به تشریح روش پیشنهادی، شامل معماری مدل، رویکرد چندوظیفه‌ای، ویژگی‌های صوتی مورد استفاده، تنظیمات آزمایشی و مجموعه‌داده‌ها اختصاص دارد. در بخش

چهارم، نتایج حاصل از ارزیابی‌های مختلف روی مجموعه‌داده‌ها گزارش و تحلیل می‌گردد. در نهایت، بخش پنجم به جمع‌بندی یافته‌ها و پیشنهاد مسیرهای آتی پژوهش اختصاص یافته است.

## ۲- مروری بر روش‌های پیشین

در این بخش، مطالعات اخیر روش‌های شناسایی و طبقه‌بندی پهپادها و روش‌های مبتنی بر صوت را بررسی می‌کنیم.

### ۲-۱- روش‌های شناسایی و طبقه‌بندی پهپاد

در حوزه‌ی شناسایی و طبقه‌بندی پهپادها، رویکردهای مختلفی با تکیه بر داده‌های حسگری گوناگون توسعه یافته‌اند که می‌توان آن‌ها را در پنج دسته‌ی کلی جای داد: روش‌های مبتنی بر فرکانس رادیویی، صوت، رادار، بینایی و روش‌های چندحسگر [۵]. در روش‌های فرکانس رادیویی، امضای الکترومغناطیسی پهپاد برای شناسایی نوع یا وضعیت پرواز مورد استفاده قرار می‌گیرد. این رویکردها [۶-۸] در محیط‌های کنترل‌شده عملکرد خوبی دارند اما در شرایط واقعی با مشکلاتی چون تراکم طیف، تداخل سیگنال، حملات مخرب و تغییر دامنه مواجه‌اند.

در روش‌های راداری، از امضای میکرو-دوپلر برای تشخیص حرکت‌های مکانیکی استفاده می‌شود که آن‌ها را برای شرایط نوری نامساعد مناسب می‌سازد. با این حال، هزینه‌ی بالا، نیاز به کالیبراسیون دقیق و سخت‌افزار تخصصی، چالش‌هایی جدی به‌شمار می‌روند؛ هرچند در سال‌های اخیر تلاش‌هایی [۹-۱۱] برای به‌کارگیری رادارهای تجاری با هزینه‌ی پایین‌تر در ترکیب با شبکه‌های عصبی، به‌منظور ساده‌سازی پیاده‌سازی و افزایش کارایی انجام شده است.

روش‌های بینایی با بهره‌گیری از پردازش تصاویر به دنبال شناسایی مدل یا رفتار پروازی پهپاد هستند [۱۲،۱۳]. این روش‌ها در شرایط نوری مناسب دقت بالایی دارند، اما در حضور نور ضعیف، آب‌وهوای نامساعد یا ازدحام بصری، عملکرد آن‌ها کاهش می‌یابد. افزون بر این، یکی از چالش‌های مشترک در روش‌های بینایی، ناتوانی در تمایز دقیق میان پهپادهای مختلف به‌دلیل شباهت‌های ظاهری آن‌هاست که دقت طبقه‌بندی را با مشکل مواجه می‌سازد.

در روش‌های صوتی، امضای صوتی پهپاد تحلیل می‌شود [۱۴-۱۶]. این رویکرد به دلیل نیاز سخت‌افزاری کم و سادگی پیاده‌سازی، برای سامانه‌های سبک‌وزن و بلادرنگ مناسب است. با این حال، نویز محیطی و تداخل صوتی از جمله چالش‌های آن است که با استفاده از فیلترهای حذف نویز و آرایه‌های میکروفونی قابل کاهش است.

در نهایت، روش‌های چندحسگر با ترکیب داده‌های حسگرهای مختلف [۱۷،۱۸]، دقت و پایداری سامانه را افزایش می‌دهند. این روش‌ها هرچند از نظر عملکردی قدرتمند هستند، اما به‌دلیل پیچیدگی هم‌زمان‌سازی حسگرها، مصرف بالای منابع و نیاز به طراحی دقیق، چالش‌های عملیاتی مهمی دارند.

روش‌های صوتی در طبقه‌بندی پهپادها به دلیل سادگی سخت‌افزار، هزینه‌ی پایین و قابلیت استفاده در شرایط نوری یا محیطی نامساعد، گزینه‌ای عملی و موثر محسوب می‌شوند. این رویکرد با تحلیل امضای صوتی ناشی از موتور و ملخ‌ها، بدون نیاز به دید مستقیم یا تجهیزات پیچیده، امکان شناسایی پهپاد را فراهم می‌کند.

میکروفون‌های سبک‌وزن و قابل‌حمل، این سامانه‌ها را برای کاربردهای بلادرنگ و مبتنی بر لبه مناسب ساخته‌اند. همچنین، مدل‌سازی دقیق ویژگی‌های صوتی از طریق یادگیری عمیق، دقت بالایی را حتی در حضور نویز محیطی ممکن کرده است.

در مجموع، با وجود چالش‌هایی مانند حساسیت به نویز و کاهش دقت در فواصل زیاد، مزایای فنی و عملی روش‌های صوتی آن‌ها را به گزینه‌ای برتر در بسیاری از سناریوهای واقعی تبدیل کرده است.

### ۲-۲- مروری بر روش‌های شناسایی و طبقه‌بندی صوتی پهپاد

با رشد فزاینده‌ی تهدیدات هوایی ناشی از پهپادها در ماموریت‌های نظامی و امنیتی، توسعه‌ی سامانه‌های شناسایی و طبقه‌بندی موثر، به‌ویژه در محیط‌های باز، مرزی و پرتداخل، از اهمیت راهبردی برخوردار شده است. در این راستا، روش‌های شناسایی صوتی به‌عنوان گزینه‌ای عملیاتی و مقرون‌به‌صرفه مطرح شده‌اند. این سامانه‌ها بدون نیاز به دید مستقیم عمل می‌کنند، در تمام شرایط آب‌وهوایی قابل استفاده‌اند و امکان استقرار سریع بر روی تجهیزات سبک‌وزن و قابل‌حمل را فراهم می‌سازند. با وجود این چالش‌ها، توازن میان کارایی، سادگی استقرار و هزینه‌ی پایین، سامانه‌های صوتی را به گزینه‌ای جذاب برای تشخیص سریع پهپاد در بسیاری از سناریوهای نظامی و امنیتی تبدیل کرده است [۱۴].

فرآیند شناسایی صوتی معمولا شامل چند مرحله کلیدی است: ۱) استخراج ویژگی‌های صوتی، ۲) پیش‌پردازش داده‌ها برای کاهش نویز و استانداردسازی، ۳) شناسایی پهپاد (تشخیص دودویی: پهپاد / محیط) و طبقه‌بندی نوع پهپاد در صورت شناسایی موفقیت‌آمیز.

دو مورد از چالش‌های اصلی روش‌های مبتنی بر صوت در تشخیص و طبقه‌بندی پهپاد، تاثیر نویز پس‌زمینه بر عملکرد راه‌حل مبتنی بر صوت و همچنین، در دسترس بودن مقادیر زیادی از داده‌های صوتی متنوع پهپاد است [۱۵].

راه حل‌هایی برای غلبه بر تاثیر نویز پس‌زمینه در تشخیص پهپا ارائه شده است. یکی از این راهکارها، افزودن نمونه‌های دارای نویزهای مختلف محیطی به مجموعه داده‌های آموزشی است تا مدل بتواند شرایط دنیای واقعی را بهتر شبیه‌سازی کرده و در مواجهه با نویزهای متنوع عملکرد مقاوم‌تری از خود نشان دهد. مجموعه داده‌های ارائه شده [۱۵،۱۶،۱۹،۲۰،۲۱] با جمع‌آوری مجموعه داده‌های جداگانه نویزی، افزودن نویز به نمونه‌های صوتی پهپاد و یا ضبط داده‌های صوتی پهپاد در محیط‌های واقعی و پر سر و صدا، تلاش کرده‌اند شرایط متنوع دنیای واقعی را شبیه‌سازی کنند تا مدل‌ها، عملکردی مقاوم‌تر و قابل اتکا در مواجهه با نویز پس‌زمینه و فواصل مختلف پهپاد داشته باشند.

انتخاب ویژگی‌های صوتی مناسب، نقش بسیار مهمی در بهبود دقت و مقاومت مدل در برابر نویزهای محیطی دارد. استخراج ویژگی‌های مبتنی بر فرکانس صدا به دلیل این که اجزای مکانیکی پهپاد، مانند پروانه‌ها و پره‌ها، امضای فرکانسی مشخص و منحصربه‌فردی ایجاد می‌کنند، می‌توانند به تشخیص دقیق‌تر کمک کند. برای مثال، مدل [۱۶] با مقایسه ویژگی‌های صوتی متنوع اعم از تبدیل فوریه سریع (FFT[1])، MFCC[2]، GTCC[3] و الگوریتم پرونی[4]، به کمک طبقه‌بندی کننده ماشین‌های بردار پشتیبان (SVM[5])، الگوریتم نزدیک‌ترین همسایه (KNN[6]) و شبکه‌های عصبی با الگوریتم پس‌انتشار خطا (BPNN[7]) به ارزیابی می‌پردازد. در [۲۲]، برای شناسایی پهپاد برای استقرار در دستگاه‌های تعبیه شده با منابع محدود، از طیف مل[8] به‌عنوان ویژگی صوتی استفاده شده و یک مدل سبک وزن بر پایه MobileNetV3 بهبودیافته طراحی شده است. این مدل با ادغام ویژگی چندمقیاسی، ماژول توجه چندمقیاسی کارآمد و بازطراحی بلوک مدل پایه با مکانیزم توجه، ضمن بهبود دقت در شرایط پیچیده، پارامترها و پیچیدگی محاسباتی را کاهش داده و سرعت عملکرد را افزایش می‌دهد.

در [۲۳]، پس از استخراج MFCC، با مدل‌های شبکه‌های عصبی عمیق، شبکه عصبی کانولوشن، مدل‌سازی وابستگی‌های بلندمدت در توالی (LSTM[9])، ترکیب کانولوشن و LSTM برای بهره‌گیری همزمان از ویژگی‌های محلی و بلندمدت و همچنین رمزگذارهای ترنسفورمر، به مقایسه می‌پردازد. در [۲۴]، پس از جمع‌آوری داده‌های صوتی از ۱۵ مدل پهپاد، ویژگی‌های MFCC استخراج شده و برای طبقه‌بندی پهپادها به یک شبکه عصبی کانولوشن داده می‌شود.

در [۲۰] نیز، پس از جمع‌آوری داده‌های صوتی از ۲۴ مدل پهپاد، به بررسی توانایی مدل در تشخیص پهپادهای ثبت‌شده و تشخیص حمله (نفوذ پهپادهای ناشناس) می‌پردازد. به این صورت که با پهپادهای ثبت‌شده (یعنی مدل با این‌ها آموزش دیده یا باید آن‌ها را بشناسد)، پهپادهای پس‌زمینه و پهپادهای مهاجم، ارزیابی عملکرد مدل احراز هویت صوتی پهپادها را ارزیابی می‌کند. این مقاله از ویژگی‌های صوتی MFCC و انواع الگوریتم‌های طبقه‌بندی استفاده می‌کند.

در کاربردهای عملی شناسایی و طبقه‌بندی صوتی پهپاد، دقت تشخیص و زمان پردازش در شرایط واقعی محیطی از اهمیت بالایی برخوردار است. در چنین سناریوهایی، دستیابی به یک توازن میان سرعت و دقت مدل ضروری است؛ چرا که در سامانه‌های بلادرنگ و منابع محدود (مانند سامانه‌های تعبیه‌شده)، تاخیر پردازش می‌تواند منجر به کاهش کارایی عملیاتی شود. از سوی دیگر، با توجه به محدودیت‌های موجود در داده‌های صوتی پهپادها و در دسترس نبودن حجم بالایی از

[1] Fast Fourier Transform
[2] Mel-Frequency Cepstral Coefficients
[3] Gammatone Cepstral Coefficients
[4] Prony Algorithm
[5] Support Vector Machines
[6] K-Nearest Neighbor
[7] Back Propagation Neural Networks

[8] Mel spectrum
[9] LongShort Term Memory
[10] Log-Mel Spectrogram

نمونه‌های متنوع، انتخاب ویژگی‌های مناسب از سیگنال صوتی نقشی کلیدی در بهبود عملکرد سامانه دارد. به همین دلیل، در این پژوهش سه روش پرکاربرد برای استخراج ویژگی‌های صوتی، شامل FFT، طیف‌نگاشت لگاریتمی مل[1] و MFCC مورد بررسی و مقایسه قرار گرفته‌اند تا تأثیر آن‌ها بر دقت طبقه‌بندی و کارایی محاسباتی مدل مشخص شود.

## ۳- روش پیشنهادی

در این بخش، روش پیشنهادی خود برای شناسایی و طبقه‌بندی پهپاد مبتنی بر صوت را به همراه مشخصات مجموعه داده‌های استفاده شده مورد بحث قرار می‌دهیم. در بخش ۳-۱، روش‌های استخراج ویژگی مورد استفاده را بررسی می‌کنیم. در بخش ۳-۲، مدل شبکه عصبی سبک‌وزن خود را شرح می‌دهیم. در بخش ۳-۳، توضیحاتی در مورد مجموعه داده‌ها، سناریوهای پرواز پهپاد و شرایط محیطی ارائه می‌دهیم.

### ۳-۱- پردازش و استخراج ویژگی از سیگنال صوتی

در سامانه‌های هوشمند شناسایی و طبقه‌بندی صوت، انتخاب ویژگی‌های مناسب از سیگنال صوتی خام نقش بسیار مهمی در عملکرد نهایی مدل یادگیری دارد. از آنجا که سیگنال‌های صوتی طبیعتی غیرایستا دارند، استخراج ویژگی‌های موثر نیازمند تجزیه و تحلیل آن‌ها در حوزه‌ی زمان-فرکانس است. در این پژوهش، با هدف دستیابی به تعادل میان دقت بالا و پردازش سریع در شرایط محیطی واقعی، از روش‌های استاندارد و بهینه‌ی استخراج ویژگی مانند FFT، طیف‌نگاشت لگاریتمی مل و MFCC بهره گرفته شده است. در ادامه، مراحل گام‌به‌گام استخراج ویژگی به تفصیل توضیح داده می‌شود. در شکل (۱) نیز روند کلی را نشان داده‌ایم.

فرض می‌کنیم سیگنال صوتی به‌صورت گسسته‌شده با نرخ نمونه‌برداری $f_s$ برحسب هرتز داده شده باشد. این سیگنال به‌صورت یک دنباله‌ی گسسته در فرمول (۱) نمایش داده شده است که $N$ تعداد کل نمونه‌های سیگنال است:

$$x[n], n = 0,1, \dots, N-1 \quad (۱)$$

به‌منظور بررسی تغییرات زمانی ویژگی‌ها، سیگنال به قطعات زمانی کوتاه موسوم به فریم تقسیم می‌شود. اگر طول فریم برابر با L و گام بین فریم‌ها برابر با H باشد، فریم‌ها طبق فرمول (۲) تعریف می‌شوند که m شماره‌ی فریم است:

$$x_n[m] = x[m + nH], 0 \le m < L \quad (۲)$$

برای کاهش اثرات برش ناگهانی در ابتدا و انتهای فریم‌ها، یک تابع پنجره (برای مثال پنجره همینگ) مانند فرمول (۳) به هر فریم اعمال می‌شود:

$$w[m] = 0.54 - 6.64 \cos\left(\frac{2\pi m}{L-1}\right) \quad (۳)$$

و فریم نهایی طبق فرمول (۴) حاصل می‌شود:

$$x_w[n, m] = x_n[m].w[m] \quad (۴)$$

همانطور که در فرمول (۵) نمایش داده شده است، برای هر فریم، تبدیل فوریه سریع انجام می‌شود تا به حوزه‌ی فرکانس منتقل شود:

$$X_n[k] = \sum_{m=0}^{L-1} x_w[n, m].e^{-j2\pi km/L} \quad (۵)$$

با تکرار محاسبه FFTبرای همه‌ی فریم‌ها، به طیف‌نگاشت کوتاه‌مدت زمان-فرکانس (STFT[2]) طبق فرمول (۶) می‌رسیم:

$$P_n[k] = \frac{1}{L}|X_n[k]|^2 \quad (۶)$$

در این مرحله، طیف فرکانسی حاصل از STFT با مجموعه‌ای از فیلترهای مثلثی بر پایه مقیاس مل نمونه‌برداری می‌شود. این مقیاس با الهام از درک شنوایی انسان طراحی شده و رابطه‌ی تبدیل فرکانس هرتز به مل به‌صورت فرمول (۷) است:

$$f_{mel} = 2595.\log_{10}(\frac{1+f}{700}) \quad (۷)$$

بانک فیلتر شامل M فیلتر مثلثی است که روی بازه‌ی فرکانسی مشخصی پخش شده‌اند. خروجی هر فیلتر طبق فرمول (۸) محاسبه می‌شود که $H_m[k]$ تابع فیلتر مل در فیلتر m است:

$$E_m[n] = \sum_{k=0}^{N-1} P_n[k].H_m[k] \quad (۸)$$

برای شبیه‌سازی مقیاس ادراکی صدا توسط گوش انسان، از لگاریتم انرژی هر فیلتر استفاده می‌شود که در آن مقدار بسیار کوچکی برای جلوگیری از لگاریتم صفر است (مثلا *1e-10*):

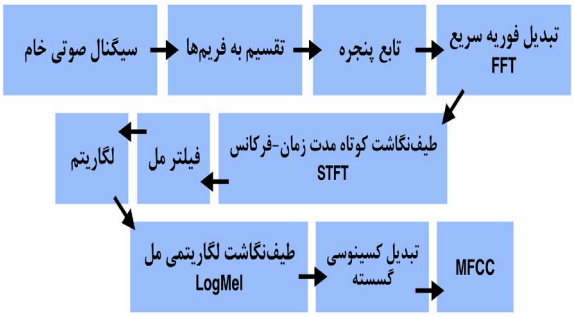
$$LogMel(m) = \log(E_m + \epsilon) \quad (۹)$$

برای کاهش همبستگی بین ویژگی‌ها و فشرده‌سازی اطلاعات، تبدیل کسینوسی گسسته نوع دوم روی بردار لگاریتم انرژی‌ها اعمال می‌شود که همانطور که در فرمول (۱۰) نمایش داده شده؛ C تعداد ضرایب MFCC مورد نظر است و این ضرایب، ویژگی‌های MFCC نام دارند:

[1] Short-Time Fourier Transform

$$MFCC_c[n] = \sum_{m=1}^{M} LogMel(m).\cos\left[\frac{\pi c}{M}(m-0.5)\right], c = 1,2,\dots,C \quad (۱۰)$$

در این پژوهش، از طیف متنوعی از روش‌های استخراج ویژگی شامل FFT، طیف لگاریتمی مل و MFCC به‌عنوان بازنمایی‌های موثر صوتی استفاده شده است. هدف از به‌کارگیری این ویژگی‌ها، تحلیل و مقایسه عملکرد آن‌ها در شناسایی صوتی پهپاد در شرایط محیطی مختلف و با در نظر گرفتن محدودیت‌های داده‌ای می‌باشد.




**شکل (۱):** روند کلی استخراج ویژگی از سیگنال صوتی

## ۳-۲- مدل شبکه عصبی سبک‌وزن

در این پژوهش، یک شبکه عصبی سبک‌وزن مبتنی بر معماری ConvNeXt طراحی و پیاده‌سازی شده است که به‌صورت چندوظیفه‌ای، همزمان وظایف تشخیص حضور پهپاد و طبقه‌بندی نوع آن را انجام می‌دهد. ساختار کلی مدل شامل سه بخش اصلی است: بلوک‌های ConvNeXt برای استخراج ویژگی که در شکل (۲) نشان داده شده است؛ و دو سر مجزا برای تشخیص و طبقه‌بندی.

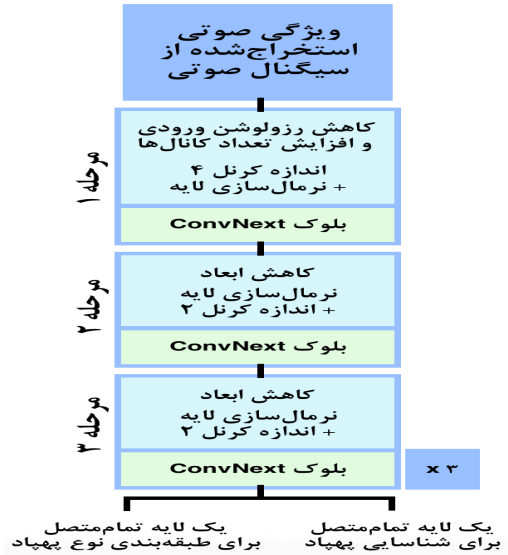

**شکل (۲):** بلوک ConvNeXt

در ستون فقرات مدل[1] برای استخراج ویژگی، از نسخه سبک‌شده‌ای از ConvNeXt استفاده شده که شامل یک بلوک اولیه با یک کانولوشن با کرنل ۴×۴ که رزولوشن ورودی را کاهش داده و تعداد کانال‌ها را افزایش می‌دهد. سپس سه مرحله پی‌درپی که در هر مرحله ابتدا کاهش ابعاد با کانولوشن ۲×۲ انجام شده و سپس چندین بلوک ConvNeXt پشته می‌شوند.

هر بلوک ConvNeXt شامل یک کانولوشن عمقی با کرنل ۷×۷ برای مدل‌سازی وابستگی‌های فضایی، نرمال‌سازی لایه‌ای روی ابعاد کانال، دو لایه تمام‌متصل به‌عنوان پرسپترون چند لایه با نسبت افزایش ویژگی و فعال‌ساز GELU برای ایجاد غیرخطی بودن می‌باشد که با یک مسیر میان‌بر جمع می‌شود.

پس از استخراج ویژگی، دو سر وظیفه‌ای تعریف می‌شوند: سر تشخیص که با یک لایه تمام‌متصل و خروجی تک‌بعدی، وظیفه‌ی تشخیص حضور یا عدم حضور پهپاد را به‌صورت دودویی انجام می‌دهد. سر طبقه‌بندی که با خروجی چندبعدی متناسب با تعداد کلاس‌های پهپاد، نوع پهپاد را طبقه‌بندی می‌کند.

مدل پیشنهادی دارای تعداد پارامتر پایین (۶۲۵،۵۴۲ پارامتر) و ساختار فشرده است که آن را مناسب برای پیاده‌سازی در سامانه‌های تعبیه‌شده می‌سازد. همچنین، استفاده از بلاک‌های ConvNeXt باعث حفظ قدرت مدل‌سازی در عین کاهش پیچیدگی محاسباتی شده است. طراحی چندوظیفه‌ای باعث یادگیری بهتر نمایش‌های مشترک و افزایش دقت کلی مدل شده است. در شکل (۳) ساختار کلی مدل شبکه عصبی عمیق را نمایش داده‌ایم.




**شکل (۳):** ساختار شبکه عصبی سبک‌وزن

[1] Backbone

مدل پیشنهادی ما یک نوع سبک از ConvNeXt با پارامترهای بسیار کمتر از انواع استاندارد این مدل است. مدل پیشنهادی این مقاله با الهام از معماری ConvNeXt طراحی شده است، اما با اصلاحات عمده‌ای در عمق، ابعاد کانال‌ها و تعداد بلوک‌ها، به شکل قابل توجهی سبک‌سازی شده است. این مدل تنها از سه مرحله با عمق‌های [۱, ۱, ۳] و کانال‌های [۴۸, ۹۶, ۱۹۲] استفاده می‌کند و در مقایسه با نسخه‌های اصلی ConvNeXt که شامل چهار مرحله با کانال‌های بیشتر (مانند [۹۶, ۱۹۲, ۳۸۴, ۷۶۸]) و عمق‌های زیاد (مثلا [۳, ۳, ۹, ۳] در نسخه Tiny) هستند، ساختار بسیار فشرده‌تری دارد. به‌طور خاص، مدل ما شامل تنها ۶۲۵۵۴۲ پارامتر (۰/۶۲ میلیون) است، در حالی‌که نسخه ConvNeXt-Tiny دارای حدود ۲۸/۶ میلیون پارامتر است، نسخه ConvNeXt-Small حدود ۵۰/۲ میلیون پارامتر دارد و نسخه‌های بزرگ‌تر مانند ConvNeXt-Base و ConvNeXt-Large به‌ترتیب دارای ۸۸/۶ میلیون و ۱۹۷ میلیون پارامتر هستند. این کاهش چشمگیر در اندازه مدل، حاصل به‌کارگیری مراحل کمتر (سه مرحله به‌جای چهار مرحله)، عمق پایین‌تر در هر مرحله، استفاده از تعداد کمتر فیلترها در هر مرحله و استفاده از سرهای ساده (یک لایه خطی برای شناسایی و طبقه‌بندی)، می‌باشد. در نتیجه، مدل طراحی‌شده در این تحقیق، بیش از ۴۵ برابر سبک‌تر از ConvNeXt-Tiny است و از این رو، برای کاربردهای بلادرنگ، سامانه‌های تعبیه شده و دستگاه‌هایی با منابع پردازشی محدود مانند Raspberry Pi بسیار مناسب و قابل استقرار است.

## ۳-۳- مجموعه داده‌های آموزش و ارزیابی مدل

یکی از چالش‌های اصلی در حوزه شناسایی و طبقه‌بندی صوتی پهپادها، دسترسی محدود به داده‌های واقعی و متنوع است. به‌دلیل ملاحظات امنیتی، سختی ضبط صدای پهپاد در شرایط کنترل‌شده و تفاوت‌های فیزیکی و صوتی میان انواع پهپادها، بسیاری از مجموعه‌داده‌های عمومی موجود یا محدود به پهپادهای خاص هستند یا در شرایط غیرواقعی (مثلا بدون نویز محیطی یا از فاصله نزدیک) جمع‌آوری شده‌اند. تعدادی از مجموعه داده‌ها نیز در دسترس عموم قرار نمی‌گیرند و اجازه انتشار ندارند [۲۵-۲۷].

در این پژوهش، تلاش شده است تا مجموعه‌داده‌ای انتخاب و پردازش شود که هم شامل نویزهای محیطی متنوع (نظیر باد، خودرو، پرندگان، و سایر صداهای پس‌زمینه) باشد و هم پهپادها در فواصل نسبتا دور از میکروفون ضبط شده باشند. این امر به‌صورت هدفمند انجام شده است تا شرایط دشوار و واقعی‌تری برای شناسایی و طبقه‌بندی مدل فراهم گردد و مدل نهایی در محیط‌های عملیاتی واقعی نیز عملکرد قابل قبولی داشته باشد.

همچنین سعی شده است از پهپادهای متنوع از نظر نوع، سخت‌افزار و اندازه استفاده شود تا مدل نهایی نسبت به تغییرات طیفی، دامنه فرکانسی و الگوهای صوتی گوناگون مقاوم و تعمیم‌پذیر باشد. این موضوع به‌ویژه در کاربردهای میدانی که با انواع مختلف پهپاد سروکار دارند، بسیار حائز اهمیت است.

در این پژوهش، یکی از منابع داده صوتی مورد استفاده، مجموعه داده منتشر شده در مقاله [۱۶] است. در فرآیند جمع‌آوری داده، صداهای سه نوع پهپاد شامل DJI Mavic، Syma X5، و یک پهپاد دست‌ساز (Race) ضبط شده‌اند. عملیات ضبط در محیط بیرونی و در حضور نویزهای معمول زندگی روزمره با نرخ نمونه‌برداری ۴۴/۱ کیلوهرتز انجام شده است؛ نویزهایی از جمله صدای عبور خودروها، صحبت افراد، صدای برگ‌ها، پژواک و بازتاب صوتی. استفاده شده است. به دلیل توان پروازی بالاتر پهپادDJI Mavic ، مدت زمان داده‌های ضبط‌شده برای این مدل نسبت به سایر پهپادها بیشتر است. داده‌ها شامل ۱۶۸۷۰ قطعه‌ی یک‌ثانیه‌ای از صداهای پهپادها در فاصله‌هایی بین ۲ تا ۱۵۰ متر می‌باشد. فرآیند برچسب‌گذاری به‌صورت دستی و با شروع ضبط پس از روشن شدن پهپاد و توقف ضبط پیش از خاموش شدن آن انجام شده است. سناریوهای مختلف پروازی نظیر ایستایی، نزدیک شدن، عبور عرضی، تغییر ارتفاع و سرعت و چرخش حول میکروفون‌ها در حین ضبط در نظر گرفته شده‌اند. همچنین، برای ارزیابی روی پهپاد ناشناخته، پهپاد Race را به‌صورت عمدی تغییر دادند؛ باتری و پره‌های آن تعویض شد و ولتاژ از ۱۱/۱ به ۱۴/۸ ولت افزایش یافت. این تغییرات منجر به تغییر در سرعت و صدای تولیدی پهپاد شد. برای این پهپاد تغییریافته، در مجموع ۱۳۶۵ ثانیه داده صوتی ضبط شد. در جدول (۱)، مدت زمان داده برای پهپادها و صداهای محیطی ارائه شده است.

برتری این مجموعه داده جمع‌آوری شده نسبت به دیگر منابع صوتی، این است که صداهای ضبط شده در فضای آزاد و دارای نویز می‌باشند، نه در محیط بسته و عاری از نویز. همچنین، بر خلاف سایر منابع صوتی که صدای پهپاد در نزدیکی میکروفون‌ها می‌باشد، در این مجموعه‌داده تلاش شده است تا

سناریوهای واقع‌گرایانه‌تری شامل فاصله‌های مختلف و نویزهای محیطی گوناگون لحاظ شود؛ این امر موجب افزایش چالش در فرآیند شناسایی و ارزیابی دقیق‌تر عملکرد مدل در شرایط واقعی شده است.

**جدول (۱):** اطلاعات مجموعه داده [۱۶]

| پهپاد/محیط | مدت زمان داده (ثانیه) |
|---|---|
| DJI Mavic | ۸۳۸۵ |
| Syma X5 | ۳۱۸۰ |
| Race | ۲۷۰۰ |
| پهپاد ناشناخته | ۱۳۶۵ |
| نویز محیطی | ۲۶۰۵ |

یکی دیگر از منابع موجود در زمینه صوت پهپاد، مربوط به [۲۰] با ۲۴ نوع پهپاد می‌باشد که شامل هشت پهپاد اصلی از نوع DJI Mini 2 و شانزده پهپاد باز مونتاژ با ترکیب بدنه مشابه پهپادهای اصلی و پره‌های جایگزین می‌باشند و در اتاقی با میکروفون‌های AT2050 در دو فاصله ۱ و ۵ متری و نرخ نمونه‌برداری ۴۴/۱ کیلوهرتز ضبط شده‌اند. برای هر پهپاد حدود ۱۰ دقیقه ضبط در دو فاصله، به‌طور کلی نزدیک به ۲۹۲۸۶ ثانیه داده خام ایجاد شده است.

در بسیاری از مجموعه‌داده‌های موجود، صداهای غیرپهپاد تنها شامل نویزهای محیطی ساده مانند باد، صحبت انسان یا سکوت هستند. این در حالی است که در محیط‌های واقعی، صداهای مشابه با پهپاد مانند هواپیماهای کوچک، هلیکوپترها یا دیگر وسایل دارای موتور نیز وجود دارند که ممکن است سامانه را دچار خطا کنند. بنابراین، صرف استفاده از نویزهای ساده به عنوان نمونه‌ی منفی در شناسایی پهپاد، منجر به ایجاد یک وظیفه‌ی ساده‌شده و غیرواقع‌گرایانه برای مدل می‌شود. در این شرایط، مدل ممکن است به جای یادگیری تفاوت‌های دقیق صوتی بین پهپاد و دیگر منابع صوتی، صرفا حضور یا عدم حضور یک نویز خاص را یاد بگیرد.

به این منظور، برای افزایش دشواری و واقع‌گرایی مسئله‌ی تشخیص صوتی پهپاد، مجموعه داده [۲۸] را به داده‌ها اضافه می‌کنیم. مجموعه داده [۲۸] شامل ۱۲/۴ ساعت صوت ثبت‌شده از ۶۲۵ نمونه پروازی مرتبط با ۳۰۱ هواگرد منحصربه‌فرد است. بخش بزرگی از داده‌ها مربوط به هواپیماهای توربوفن دوموتوره مانند بوئینگ ۸۰۰-۷۳۷ و همچنین شامل داده‌هایی از هواپیماهای توربوپراپ، پیستونی، هلیکوپترها و مدل‌های چهار موتوره است.

## ۴- نتایج آزمایشات

در این بخش، نتایج به‌دست‌آمده از ارزیابی مدل پیشنهادی روی مجموعه‌داده‌های مختلف ارائه شده است تا میزان دقت شناسایی، طبقه‌بندی و عملکرد در مواجهه با پهپادهای ناشناخته و یا صداهای محیطی مشابه پهپاد بررسی شود.

### ۴-۱- تنظیمات شبکه

کلیه آزمایش‌ها با استفاده از کارت گرافیک NVIDIA Tesla T4 انجام شده‌اند که با ۱۶ گیگابایت حافظه و معماری Turing، برای پردازش‌های یادگیری عمیق نسبتا سبک و متوسط بهینه شده است. پیاده‌سازی مدل با استفاده از کتابخانه PyTorch صورت گرفت و مدل در تمام آزمایشات تنها در یک اپوک آموزش داده شده است و اندازه‌ی هر دسته[1] برابر با ۳۲ در نظر گرفته شد. برای پیش‌پردازش داده‌های صوتی، از نرخ نمونه‌برداری ۱۶ کیلو هرتز و مدت زمان ۱ ثانیه برای هر نمونه استفاده شد. ویژگی‌های ورودی به مدل به‌صورت طیف‌نگاشت مل با ۹۶ فیلتر مل استخراج شدند. همچنین، برای محاسبه‌ی طیف فوریه، اندازه‌ی پنجره برابر با ۱۰۲۴ و گام حرکت برابر با ۲۵۶ نمونه در نظر گرفته شد. همچنین برای بهینه‌سازی[2] پارامترها، از بهینه‌ساز Adam با نرخ یادگیری 1e-4 استفاده شد. تابع ضرر کل نیز به صورت فرمول (۱۵) تعریف می‌شود:

$$\mathcal{L}_{total} = \mathcal{L}_{det} + \mathcal{L}_{cls} \quad (۱۵)$$

که در آن، $L_{det}$ تابع ضرر باینری BCEWithLogits برای شناسایی (تشخیص وجود یا عدم وجود پهپاد) و $L_{cls}$ تابع ضرر CrossEntropy برای طبقه‌بندی نوع پهپاد است.

### ۴-۲- مقایسه انواع ویژگی‌های صوتی روی مجموعه داده [۱۶] (حالت پایه - آموزش ۸۰٪)

در گام نخست، به‌منظور ارزیابی عملکرد مدل پیشنهادی و مقایسه نتایج آن با مقاله [۱۶]، از ارزیابی متقابل ۵-بخشی[3] بر روی مجموعه‌داده‌ی معرفی‌شده در همان مقاله استفاده کردیم. در هر تکرار، داده‌ها به نسبت ۸۰ درصد برای آموزش و ۲۰ درصد برای ارزیابی تقسیم شدند. برای فراهم‌سازی ورودی مناسب به

[1] Batch Size
[2] Optimization
[3] 5-Fold Cross-Validation

شبکه‌ی عصبی سبک‌وزن طراحی‌شده، از مجموعه‌ای از ویژگی‌های صوتی شامل FFT، طیف‌نگاشت لگاریتمی مل و MFCC بهره گرفته شده است. در این پژوهش، هر یک از این ویژگی‌های استخراج‌شده به‌صورت جداگانه به‌عنوان ورودی به مدل داده شدند و عملکرد مدل در هر حالت به‌طور مستقل ارزیابی شد. در جدول (۲) و (۳)، نتایج این مرحله از آزمایشات نشان داده شده است. این رویکرد به ما امکان داد تا مقایسه‌ای دقیق میان تاثیر هر نوع نمایش صوتی بر دقت شناسایی و طبقه‌بندی پهپادها داشته باشیم.

یکی از جنبه‌های مهم در ارزیابی عملکرد و دقت مدل‌های مختلف، استفاده از ماتریس درهم‌ریختگی[1] است که هر کلاس توسط یک سطر و یک ستون نمایش داده می‌شود، به‌طوری‌که یکی نمایانگر برچسب‌های واقعی (کلاس هدف) و دیگری نشان‌دهنده‌ی برچسب‌های پیش‌بینی‌شده (کلاس خروجی) است. از این ماتریس، معیارهای کلیدی دقت[2]، صحت[3]، پوشش[4] و F1-score (ترکیب متعادلی بین معیارهای دقت و صحت) برای ارزیابی عملکرد مدل در وظیفه‌ی شناسایی استخراج می‌شوند. TP، TN، FP و FN به ترتیب نمونه‌های مثبت صحیح (پیش‌بینی درست وجود پهپاد)، نمونه‌های منفی صحیح (پیش‌بینی درست عدم وجود پهپاد)، نمونه‌های مثبت کاذب (پیش‌بینی نادرست وجود پهپاد در حالی‌که وجود ندارد)، نمونه‌های منفی کاذب (پیش‌بینی نادرست عدم وجود پهپاد در حالی‌که وجود دارد) می‌باشند.

$$Accuracy = {TP + TN}/{TP + TN + FP + FN} \quad (۱۱)$$

$$Precision = {TP}/{TP + FP} \quad (۱۲)$$

$$Recall = {TP}/{TP + FN} \quad (۱۳)$$

$$F1 - Score = {2TP}/{2TP + FP + FN} \quad (۱۴)$$

[1] Confusion Matrix
[2] Accuracy
[3] Precision
[4] Recall

**جدول (۲):** نتایج طبقه‌بندی روی مجموعه داده [۱۶] با نسبت ۸۰ درصد آموزش و ۲۰ درصد ارزیابی

| دقت | مدل |
|---|---|
| ۹۳/۶ | [۱۶] |
| ۶۶ | شبکه سبک‌وزن (FFT) |
| ۹۶/۹۹ | شبکه سبک‌وزن (LogMel) |
| **۹۹/۲۱** | شبکه سبک‌وزن (MFCC) |

**جدول (۳):** نتایج شناسایی روی مجموعه داده [۱۶] با نسبت ۸۰ درصد آموزش و ۲۰ درصد ارزیابی

| مدل | دقت | صحت | پوشش | F1 |
|---|---|---|---|---|
| [۱۶] | ۹۷/۷ | ۹۸/۶ | ۹۸/۷ | ۹۸/۶ |
| شبکه سبک‌وزن (FFT) | ۹۴/۵ | ۹۷/۰۲ | ۹۶/۴۹ | ۹۶/۷۴ |
| شبکه سبک‌وزن (LogMel) | ۹۷/۹۱ | ۹۸/۳۸ | ۹۹/۱۹ | ۹۸/۷۷ |
| شبکه سبک‌وزن (MFCC) | **۹۹/۵۵** | **۹۹/۶۶** | **۹۹/۷۹** | **۹۹/۷۳** |

## ۴-۳- آزمایش در شرایط کمبود داده (آموزش ۲۰٪)

یکی از چالش‌های اصلی داده‌های صوتی پهپاد، کمبود داده‌های آموزشی متنوع و کافی است. برای مقابله با چالش کمبود داده، ما نسبت معکوس ۲۰ درصد برای آموزش و ۸۰ درصد برای ارزیابی را نیز بر روی مجموعه‌داده مقاله [۱۶] اعمال کردیم. این آزمایش به منظور بررسی عملکرد مدل در شرایطی با داده‌های آموزشی محدود انجام شد که نتایج در جدول (۴) و (۵) نشان داده شده است.

**جدول (۴):** نتایج طبقه‌بندی روی مجموعه داده [۱۶] با نسبت ۲۰ درصد آموزش و ۸۰ درصد ارزیابی

| دقت | مدل |
|---|---|
| ۴۸/۴۶ | شبکه سبک‌وزن (FFT) |
| ۸۱/۷۳ | شبکه سبک‌وزن (LogMel) |
| **۹۴/۱۵** | شبکه سبک‌وزن (MFCC) |

**جدول (۵):** نتایج شناسایی روی مجموعه داده [۱۶] با نسبت ۲۰ درصد آموزش و ۸۰ درصد ارزیابی

| مدل | دقت | صحت | پوشش | F1 |
|---|---|---|---|---|
| شبکه سبک‌وزن (FFT) | ۸۵/۸۴ | ۸۸/۶۹ | ۹۵/۴۶ | ۹۱/۹۳ |
| شبکه سبک‌وزن (LogMel) | ۸۸/۵۸ | ۸۸/۸۱ | ۹۹/۰۱ | ۹۳/۶۲ |
| شبکه سبک‌وزن (MFCC) | **۹۵/۷۳** | **۹۵/۹۳** | **۹۹/۱۷** | **۹۷/۵۲** |

### ۴-۴- آزمایش شناسایی پهپاد ناشناخته

به منظور ارزیابی توانایی تعمیم مدل در شرایط واقعی، عملکرد حاصل از این دو تقسیم‌بندی آموزش (۸۰-۲۰ آموزش-ارزیابی و بالعکس) را بر روی پهپادهایی که در مرحله‌ی آموزش حضور نداشتند (پهپادهای ناشناخته) نیز مورد بررسی قرار دادیم. در جدول (۶)، نتایج را با مقاله [۱۶] مقایسه کرده‌ایم.

**جدول (۶):** نتایج شناسایی پهپاد ناشناخته از مجموعه داده [۱۶]

| دقت | مدل |
|---|---|
| ۹۷/۳ | [۱۴] |
| ۹۳/۷۹ | شبکه عمیق سبک‌وزن (FFT) (۲۰-۸۰ آموزش-ارزیابی) |
| ۹۸/۷۳ | شبکه عمیق سبک‌وزن (LogMel) (۲۰-۸۰ آموزش-ارزیابی) |
| ۹۶/۰۷ | شبکه عمیق سبک‌وزن (MFCC) (۲۰-۸۰ آموزش-ارزیابی) |
| **۹۹/۲۹** | شبکه عمیق سبک‌وزن (FFT) (۸۰-۲۰ آموزش-ارزیابی) |
| ۸۱/۷۳ | شبکه عمیق سبک‌وزن (LogMel) (۸۰-۲۰ آموزش-ارزیابی) |
| ۹۴/۱۵ | شبکه عمیق سبک‌وزن (MFCC) (۸۰-۲۰ آموزش-ارزیابی) |

### ۴-۵- ارزیابی در محیط‌های واقعی‌تر با داده‌های غیرپهپادی

ماهیت ساده و کنترل‌شده نویزهای محیطی باعث شد که تشخیص صدای پهپاد از سایر صداها کار دشواری نباشد. به همین دلیل، در گام بعدی برای ارزیابی مدل در شرایط چالشی‌تر و شبیه‌تر به دنیای واقعی، از مجموعه‌داده‌های محیطی متنوع شامل صداهای وسایل نقلیه هوایی مانند هواپیما و هلیکوپتر نیز استفاده کردیم. این داده‌ها از مقاله [۲۸] گرفته شده‌اند که شامل صداهای واقعی ضبط‌شده از وسایل نقلیه هوایی در شرایط محیطی گوناگون هستند. مدل پیشنهادی را بر روی مجموعه‌ی ترکیبی از داده‌های قبلی یعنی مجموعه داده [۱۶] و داده‌های غیر پهپاد جدید، با نسبت ۲۰ درصد برای آموزش و ۸۰ درصد برای ارزیابی مورد آزمایش قرار دادیم تا مطابق نتایج در جدول (۷) و (۸) توانایی مدل در تشخیص صدای پهپاد در میان صداهای مشابه و گمراه‌کننده سنجیده شود.

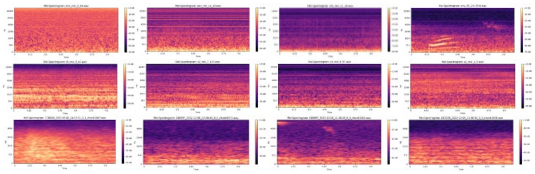
**جدول (۷):** نتایج شناسایی روی مجموعه داده [۱۶] و [۲۸] با نسبت ۲۰ درصد آموزش و ۸۰ درصد ارزیابی

| دقت | مدل |
|---|---|
| **۹۷/۹۵** | شبکه سبک‌وزن (MFCC) |

**جدول (۸):** نتایج طبقه‌بندی روی مجموعه داده [۱۶] و [۲۸] با نسبت ۲۰ درصد آموزش و ۸۰ درصد ارزیابی

| F1 | پوشش | صحت | دقت | مدل |
|---|---|---|---|---|
| ۹۶/۲۴ | ۹۷/۷۳ | ۹۴/۹۷ | ۹۷/۸۵ | شبکه سبک‌وزن (MFCC) |

برای بررسی دقیق‌تر ویژگی‌های صوتی استخراج‌شده از نمونه‌های داده، شکل (۴) تغییرات انرژی در حوزه زمان-فرکانس را نمایش می‌دهند و نشان می‌دهند که صداهای مربوط به پهپادها الگوهای طیفی منحصربه‌فردی دارند که می‌توانند برای مدل‌های یادگیری عمیق قابل تفکیک باشند. همچنین می‌بینیم که داده‌های غیر پهپادی از مقاله [۲۸] نسبت به داده‌های محیطی مقاله [۱۶]، دارای شباهت ساختاری و طیفی بیشتری به داده‌های پهپاد می‌باشند که منجر به عملکرد بهتر مدل در تشخیص نمونه‌های پهپادی مشابه گردد و در عین حال بر چالش تفکیک میان کلاس‌های محیطی و پهپادی تاکید کند.



**شکل (۴):** الگوهای طیفی داده‌های محیط (ردیف اول) و پهپاد (ردیف دوم) از مجموعه داده [۱۶] داده‌های غیر پهپادی (ردیف سوم) از مجموعه داده [۲۸]

### ۴-۶- ارزیابی محاسباتی و کارایی زمان اجرا

علاوه بر آزمایشات فوق، اطلاعات ارائه‌شده در جدول (۹) نشان می‌دهد که زمان مورد نیاز برای استخراج ویژگی‌های فرکانسی و انجام شناسایی و طبقه‌بندی صوتی توسط شبکه پیشنهادی ما کاملا قابل‌قبول و مناسب برای کاربردهای بلادرنگ است. این موضوع بیانگر کارایی بالا و کاربردپذیری عملی مدل ما در شرایط واقعی و در مواجهه با پهپادهای مختلف است.

سبک‌وزن بودن مدل، الزامات محاسباتی را در سطح پایینی نگه می‌دارد و آن را برای کاربردهای بلادرنگ و سامانه‌های نهفته مناسب می‌سازد. در همین راستا، زمان کل اجرای هر مرحله

شامل استخراج هر یک از ویژگی‌های صوتی، واردسازی آن به شبکه و استخراج ویژگی‌ها تا پیش از لایه‌ی شناسایی/طبقه‌بندی نیز اندازه‌گیری شده است، تا میزان بار محاسباتی و کارایی مدل در سناریوهای عملی با دقت ارزیابی گردد.

**جدول (۹):** میانگین زمانی استخراج ویژگی و استنتاج مدل (بر حسب میلی‌ثانیه)

| مدل | میانگین زمان استخراج ویژگی | میانگین زمان استنتاج مدل |
|---|---|---|
| شبکه عمیق سبک‌وزن (FFT) | ۱۱/۸۳ | ۲/۱۷ |
| شبکه عمیق سبک‌وزن (LogMel) | ۱۹/۸۲ | ۲/۳۲ |
| شبکه عمیق سبک‌وزن (MFCC) | ۲۲/۴۸ | ۳/۵۹ |

## ۵- نتیجه‌گیری و کارهای آینده

در این پژوهش، یک سامانه‌ی سبک‌وزن و کارآمد برای شناسایی و طبقه‌بندی صوتی پهپادها با بهره‌گیری از شبکه‌های عصبی عمیق طراحی و ارزیابی شد. مدل پیشنهادی، با وجود حجم بسیار کم (تنها حدود ۰/۶۲ میلیون پارامتر)، موفق شد دقت شناسایی را تا ۹۹/۵۵ درصد و دقت طبقه‌بندی را تا ۹۹/۲۱ درصد افزایش دهد. آزمایش‌ها نشان دادند که این مدل نه تنها در تشخیص دقیق انواع مختلف پهپادهای شناخته‌شده عملکرد قابل‌قبولی دارد، بلکه در مواجهه با پهپادهای ناشناخته نیز توانایی تفکیک مناسبی از کلاس‌های غیرپهپادی مانند صداهای محیطی، هلیکوپتر، هواپیما و دیگر منابع صوتی مشابه از خود نشان می‌دهد.

برای تقویت عمومی‌سازی مدل، در مرحله ارزیابی، داده‌های صوتی غیرپهپادی از منابع متنوع، از جمله صداهای هلیکوپتر و ترافیک هوایی به مجموعه‌داده افزوده شد تا چالش تفکیک بین منابع صوتی مشابه افزایش یابد.

از جمله مسیرهای آینده برای توسعه این سامانه می‌توان به موارد زیر اشاره کرد: استفاده از داده‌های چندکاناله و روش‌های مکان‌یابی و تشکیل پرتو برای بهبود دقت در شرایط محیطی پیچیده که با استفاده از آرایه‌ای از میکروفن‌ها، امکان تمرکز بر سیگنال‌های صوتی در یک جهت خاص را فراهم می‌سازد و در نتیجه به تقویت سیگنال‌های پهپاد و حذف نویزهای محیطی کمک می‌کند؛ به‌کارگیری روش‌های یادگیری انتقالی و یادگیری تطبیقی برای افزایش توان مدل در شناسایی پهپادهای ناشناخته در محیط‌های جدید؛ طراحی نسخه‌های بهینه‌سازی‌شده برای اجرا روی سخت‌افزارهای کم‌مصرف مانند سامانه‌های تعبیه‌شده؛ گسترش مجموعه‌داده با استفاده از داده‌های جمع‌آوری‌شده در محیط‌های واقعی و متنوع برای افزایش تعمیم‌پذیری مدل.

## ۶- مراجع


[۱] پیکام، علیرضا، شاهبندرزاده، حمید، "اولویت بندی عوامل موثر بر عملکرد پهپادها در صحنه‌ی نبرد ناهمتراز آینده با استفاده از فرآیند تحلیل سلسله مراتبی فازی"، نشریه پدافند غیر عامل، دوره ۱۱، شماره ۱، صفحه ۵۱-۶۱، ۱۳۹۹.
https://dor.isc.ac/dor/20.1001.1.20086849.1399.11.1.5.4

[۲] امیرزاده، مجید، حسینی مرادی، سیدعلی، قبادی، نادر، "تشخیص به‌موقع پرنده‌های هدایت‌پذیر از دور چند بال چرخان با استفاده از الگوریتم YOLOv5 بهینه‌سازی‌شده"، نشریه علوم و فناوری‌های پدافند نوین، دوره ۱۴، شماره ۱ - صفحه ۱۱-۲۲، ۱۴۰۲.
https://dor.isc.ac/dor/20.1001.1.26762935.1402.14.1.2.2

[۳] بهرامی، محمد، اصغری، امیر، بینش مروستی، محمدرضا، انصاریان، سجاد، "ارائه یک روش بهبودیافته تشخیص هواپیمای بدون سرنشین با استفاده از یادگیری عمیق جهت افزایش سرعت تشخیص"، نشریه پدافند الکترونیکی و سایبری، دوره ۱۱، شماره ۱، صفحه ۸۱-۹۶، ۱۴۰۲.
https://dor.isc.ac/dor/20.1001.1.23224347.1402.11.1.7.8

[4] Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., & Xie, S., "A convnet for the 2020s," In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 11976-11986, 2022. https://doi.org/10.1109/CVPR52688.2022.01167

[5] Dong, Y., Wu, F., Zhang, S., Chen, G., Hu, Y., Yano, M., Sun, J., Huang, S., Liu, F., Dai, Q. and Cheng, Z.Q., "Securing the Skies: A Comprehensive Survey on Anti-UAV Methods, Benchmarking, and Future Directions," Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 6659-6673. 2025. https://doi.org/10.48550/arXiv.2504.11967

[6] M. A. Alqodah, M. Tahsin, M. H. Omari, M. M. Matalgah and D. Harrison, "RF-Based Lightweight Machine Learning for Comprehensive Drone Activity Classification," 2024 2nd International Conference on Artificial Intelligence, Blockchain, and Internet of Things (AIBThings), Mt Pleasant, MI, USA, pp. 1-5, 2024. https://doi.org/10.1109/AIBThings63359.2024.10863363

[7] P. Podder, M. Zawodniok and S. Madria, "Deep Learning for UAV Detection and Classification via Radio Frequency Signal Analysis," 2024 25th IEEE International Conference on Mobile Data Management (MDM), Brussels, Belgium, pp. 165-174, 2024. https://doi.org/10.1109/MDM61037.2024.00040

[8] Sazdic-Jotic B, Andric M, Bondzulic B, Simic S, Pokrajac I, "FLEDNet: Enhancing the Drone Classification in the Radio Frequency Domain, " Drones, 9(4), 243, 2025.

https://doi.org/10.3390/drones9040243
[9] S. Pant, M. Manning, J. Laliberte, P. Sevigny, S. Rajan and B. Balaji, "Investigating Radar Micro-Doppler Signatures for Drone Payload Detection," 2025 25th International Conference on Digital Signal Processing (DSP), Pylos (Messinia, Southwest Peloponnese), Greece, pp. 1-5, 2025. https://doi.org/10.1109/DSP65409.2025.11075208
[10] J. J. M. de Wit and L. de Martín, "Emerging Trends in Radar: Drone Characterization Using Deep Learning on Micro-Doppler Data," in IEEE Aerospace and Electronic Systems Magazine, vol. 40, no. 6, pp. 48-53, 2025. https://doi.org/10.1109/MAES.2025.3546151
[11] Gong, J., Li, D., Yan, J., & Kong, D., "Comparative Micro-Doppler Signal Detection in L-Band and X-Band Drone Detection Systems, ", International Conference on Autonomous Unmanned Systems, pp. 291-300, 2024. https://doi.org/10.1007/978-981-96-3592-4_30
[12] Islam, S.B., Chowdhury, M.E., Hasan-Zia, M., Kashem, S.B.A., Majid, M.E., Ansaruddin Kunju, A.K., Khandakar, A., Ashraf, A. and Nashbat, M., "VisioDECT: a novel approach to drone detection using CBAM-integrated YOLO and GELAN-E models," Neural Computing and Applications, 1-24, 2025. https://doi.org/10.1007/s00521-025-11448-3
[13] Liu, Z., An, P., Yang, Y., Qiu, S., Liu, Q. and Xu, X., "Vision-based drone detection in complex environments: a survey," Drones, 8(11), p.643, 2024. https://doi.org/10.3390/drones8110643
[14] Seidaliyeva, U., Ilipbayeva, L., Taissariyeva, K., Smailov, N. and Matson, E.T., "Advances and challenges in drone detection and classification techniques: A state-of-the-art review," Sensors, 24(1), p.125, 2023. https://doi.org/10.3390/s24010125
[15] S. Al-Emadi, A. Al-Ali, A. Mohammad and A. Al-Ali, "Audio Based Drone Detection and Identification using Deep Learning," 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC), Tangier, Morocco, pp. 459-464, 2019. https://doi.org/10.1109/IWCMC.2019.8766732 Available: https://github.com/saraalemadi/DroneAudioDataset
[16] Najafi, J., Mirzakuchaki, S. & Shamaghdari, S., "Autonomous Drone Detection and Classification Using Computer Vision and Prony Algorithm-Based Frequency Feature Extraction, " J Intell Robot Syst 111, 8, 2025. https://doi.org/10.1007/s10846-024-02216-x Available: https://github.com/Jafar-Najafi/Audio-Files
[17] S. Yuan et al., "MMAUD: A Comprehensive Multi-Modal Anti-UAV Dataset for Modern Miniature Drone Threats," 2024 IEEE International Conference on Robotics and Automation (ICRA), Yokohama, Japan, pp. 2745-2751, 2024. https://doi.org/10.1109/ICRA57147.2024.10610957
[18] Alla, I., Olou, H.B., Loscri, V. and Levorato, M., "From sound to sight: Audio-visual fusion and deep learning for drone detection, " Proceedings of the 17th acm conference on security and privacy in wireless and mobile networks, pp. 123-133, 2024. https://doi.org/10.1145/3643833.3656133
[19] Yi, W., Choi, J.W. and Lee, J.W., "Sound-based drone fault classification using multitask learning, " arXiv preprint arXiv:2304.11708, 2023. https://doi.org/10.48550/arXiv.2304.11708 Available: https://zenodo.org/records/7779574
[20] Diao, Y., Zhang, Y., Zhao, G. and Khamis, M., "Drone authentication via acoustic fingerprint," Proceedings of the 38th Annual Computer Security Applications Conference, pp. 658-668, 2022. https://doi.org/10.1145/3564625.3564653 Available: https://researchdata.gla.ac.uk/1348/
[21] Casabianca, P., & Zhang, Y., "Acoustic-based UAV detection using late fusion of deep neural networks," Drones, 5(3), 54, 2021. https://doi.org/10.3390/drones5030054
[22] T. Li, Z. Huang, X. Zhai and S. Wang, "A Lightweight UAV Audio Detection Model Based on Multiscale Feature Fusion," 2023 5th International Academic Exchange Conference on Science and Technology Innovation (IAECST), Guangzhou, China, pp. 1524-1528, 2023. https://doi.org/10.1109/IAECST60924.2023.10503266
[23] S. S. Katta, S. Nandyala, E. K. Viegas and A. AlMahmoud, "Benchmarking Audio-based Deep Learning Models for Detection and Identification of Unmanned Aerial Vehicles," 2022 Workshop on Benchmarking Cyber-Physical Systems and Internet of Things (CPS-IoTBench), Milan, Italy, pp. 7-11, 2022. https://doi.org/10.1109/CPS-IoTBench56135.2022.00008
[24] Wang, M. Y., Chu, Z., Ku, I., Smith, E. C., & Matson, E. T., "A 15-category audio dataset for drones and an audio-based uav classification using machine learning," International Journal of Semantic Computing, 18(02), 257-272, 2024. https://doi.org/10.1142/S1793351X24300048
[25] M. Ohlenbusch, A. Ahrens, C. Rollwage and J. Bitzer, "Robust Drone Detection for Acoustic Monitoring Applications," 2020 28th European Signal Processing Conference (EUSIPCO), Amsterdam, Netherlands, pp. 6-10, 2021. https://doi.org/10.23919/Eusipco47968.2020.9287433
[26] Utebayeva, D., Ilipbayeva, L., & Matson, E. T. "Practical study of recurrent neural networks for efficient real-time drone sound detection: A review," Drones, 7(1), 26, 2022. https://doi.org/10.3390/drones7010026
[27] J. Kim, Q. Zhang, E. T. Matson and M. Y. Wang, "Improving Drone Classification with Audio-Derived Visual Features: A Vision Model Comparison," 2024 Eighth IEEE International Conference on Robotic Computing (IRC), Tokyo, Japan, pp. 41-45, 2024. https://doi.org/10.1109/IRC63610.2024.00013
[28] Downward, B., & Nordby, J., "The AeroSonicDB (YPAD-0523) dataset for acoustic detection and classification of aircraft", arXiv preprint arXiv:2311.06368, 2023. https://doi.org/10.48550/arXiv.2311.06368