

Radio Frequency-based Drone Detection and Classification Using Discrete Fourier Transform and LightGBM

Ki-Hyeon Sung*, Soo-Jin Lee**

*Graduate Student, Dept. of Defense Science, Korea National Defense University, Nonsan, Korea

**Professor, Dept. of Defense Science, Korea National Defense University, Nonsan, Korea

[Abstract]

In this study, we proposed an efficient model that can detect and classify the drones and related devices based on radio frequency signals. In order to increase the applicability in the battlefield, proposed model was designed to be lightweight, to ensure rapid detection and high detection accuracy. Data preprocessing was performed by applying a Discrete Fourier Transform (DFT) that is faster than Hilbert-Huang Transform (HHT). We adopted the LightGBM model as the learning model, which can be easily used by non-professionals and guarantees excellent performance in terms of classification speed and accuracy. CardRF dataset was used to verify the performance of the proposed model. As a result of the experiment, the accuracy of 3 classes classification for detecting and classifying drones, WiFi, and Bluetooth device was 99.63% when the number of sample points was set to 100k and 99.40% when set to 500k during the data preprocessing with DFT. And, in the 10 classes classification for 6 drones, 2 Bluetooth devices, and 2 WiFi devices, the accuracy was 95.65% for 100k and 96.83% for 500k, confirming significantly improved detection performance compared to previous studies.

▶ **Key words:** Drone, Radio Frequency, Discrete Fourier Transform, LightGBM, CardRF Dataset

[요약]

본 연구에서는 드론 및 관련 장치들로부터 생성되는 무선주파수 신호를 기반으로 드론을 탐지하고 기종을 식별하는 모델을 제안하였다. 전장 환경에서의 활용 가능성을 높이기 위해 모델은 경량화와 신속한 탐지를 최우선으로 고려하되 높은 탐지 정확도도 보장할 수 있도록 설계하였고, 데이터 전처리는 Hilbert-Huang 변환보다 처리 속도가 더 빠른 이산 푸리에 변환으로 수행하였다. 학습 모델은 비전문가도 쉽게 사용할 수 있고 분류 속도 및 정확도 측면에서도 탁월한 성능을 발휘하는 LightGBM 모델을 채택하였다. 제안하는 모델의 성능 검증은 공개 드론 무선주파수 데이터셋인 CardRF dataset을 활용하여 수행하였다. 실험결과 드론, WiFi 및 Bluetooth 장치 3종을 탐지 및 식별하는 3 클래스 다중분류의 정확도는 이산 푸리에 변환으로 전처리를 수행하는 과정에서 샘플 포인트 수를 100k로 설정한 경우 99.63%, 500k로 설정한 경우 99.40%로 나타났다. 드론 6종, Bluetooth 장치 2종 및 WiFi 장치 2종에 대한 10 클래스 다중분류 실험에서는 샘플 포인트 수가 100k인 경우 95.65%, 500k인 경우 96.83%를 달성하여 이전 연구 대비 상당히 향상된 탐지성능을 보임을 확인하였다.

▶ **주제어:** 드론, 무선주파수, 탐지 및 식별, 이산 푸리에 변환, LightGBM, CardRF Dataset

-
- First Author: Ki-Hyeon Sung, Corresponding Author: Soo-Jin Lee
 - *Ki-Hyeon Sung (brightness79@gmail.com), Dept. of Defense Science, Korea National Defense University
 - **Soo-Jin Lee (cyberma@gmail.com), Dept. of Defense Science, Korea National Defense University
 - Received: 2024. 07. 26, Revised: 2024. 08. 28, Accepted: 2024. 09. 25.

I. Introduction

드론(drone)이라고도 알려진 무인항공기(Unmanned Aerial Vehicles, UAVs)의 사용은 근래 획기적으로 증가하고 있다. 글로벌 드론 시장은 최근 몇 년 동안 급속하게 성장하면서, 2023년 300억 6천만 달러 규모였던 시장이 2024년에는 348억 2천만 달러 규모로 성장할 것으로 예상된다.[1]

드론의 활용 분야도 환경 모니터링, 통신 및 교통관리, 건설관리 및 농업분야 등 매우 다양해지고 있다.[2] 그러나 드론 활용의 급증으로 인해 저고도 공역에서의 규정 위반이나 각종 불법행위가 발생하고 있으며, 테러 공격까지도 자행되고 있다.[3][4] 더욱이 소형 드론은 기존 레이더를 활용한 방공망으로 탐지가 쉽지 않기 때문에 군사적 용도로서도 매우 유용하게 활용되고 있다.

실제로 2022년 초 발발해 현재까지 진행되고 있는 러시아와 우크라이나의 전쟁에서 우크라이나는 Bayraktar TB2 등 군사용 드론을 적극적으로 사용하고 있으며, 정찰뿐 아니라 무장 공격 등 여러 임무를 성공적으로 수행하고 있는 것으로 알려져 있다.[5][6] 또한, 전황의 변화에 따라 러시아의 방공망을 뚫고 사용할 수 있는 소형 드론을 활용하기 위해 민간에서 사용하는 상업용 드론을 일회성 공격용으로 개조하여 저렴한 비용으로 높은 정밀도의 공격을 수행하였다.[7]

북한은 2014년부터 대한민국을 정찰하기 위해 다양한 드론을 사용해 왔으며, 상용 GPS를 사용하고 레이더 방공망에 잘 식별되지 않는 형태의 소형 드론을 2022년 12월 서울까지 침투시키는 등 다양한 방식으로 대한민국의 안보를 위협하고 있는 중이다.

이러한 드론에 의한 위협에 효율적으로 대응하기 위해 최근에는 드론을 조기 탐지 및 식별하기 위한 연구들이 활발하게 진행되고 있다. 드론 탐지 및 식별 연구는 크게 오디오 기반 탐지, 광학영상 및 적외선 기반 탐지, RADAR 기반 탐지, 및 무선주파수(Radio frequency) 기반 탐지 로 구분할 수 있다.[8] 이 중 무선주파수에 기반한 탐지는 드론의 크기나 형태, 작동상태에 큰 영향을 받지 않고, 가시선 여부와도 상관없이 감지가 가능하며, 현재 날씨나 주변 환경에 영향을 받지 않는다는 장점을 가지고 있다. 그리고 드론이 발생시키는 전자파 신호를 수신하여 탐지하는 수동 탐지이기 때문에 은밀성이 있고 실시간 탐지가 가능하다.

최근 들어서는 인공지능 기술의 발전에 힘입어 머신러닝과 딥러닝 모델 등을 기반으로 무선주파수 감지를 수행하여 드론을 탐지 및 식별하는 연구가 활발하게 진행되고

있으며, 이러한 연구를 지원하기 위해 다양한 드론 무선주파수 데이터셋도 계속해서 발표되고 있다. 그러나 기존 연구들의 대부분은 강력한 딥러닝 모델을 사용하여 높은 정확도를 달성하는데 초점을 맞추고 있다. 반면 전장환경에서 절대적으로 요구되는 신속한 탐지나 자원제약적 특성을 고려한 탐지모델 경량화 및 업데이트의 용이성 등은 크게 고려하지 않기 때문에 군사 목적으로 활용하기 위해서는 개선이 필요하다.

이에 본 논문에서는 무선주파수를 기반으로 드론을 탐지함에 있어 전장환경에서의 활용 가능성을 높이기 위해 모델 경량화와 신속한 탐지를 최우선으로 고려하되 높은 탐지 정확도도 보장할 수 있는 모델을 제안한다. 데이터 전처리 과정에서는 비교적 속도가 느린 Hilbert-Huang 변환(Hilbert-Huang Transform, 이하 HHT)이 아니라 이산 푸리에 변환(Discrete Fourier Transform, 이하 DFT)을 적용을 고려한다. 학습 모델은 비전문가도 쉽게 활용할 수 있어 모델 업데이트 등 유지보수가 용이할 뿐만 아니라 분류 속도 및 정확도 측면에서도 탁월한 성능을 발휘하는 LightGBM 모델을 활용한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서 본 연구의 모델 구축 및 성능평가에 활용된 CardRF dataset에 대해 살펴보고, 드론의 무선주파수를 기반으로 인공지능 모델을 접목해 드론 탐지를 시도했던 연구들을 정리한다. 3장에서는 데이터 전처리 과정을 포함한 드론 탐지 모델의 구축 절차와 실험과정을 설명하고, 결과를 분석한다. 마지막으로 4장에서 연구 결과를 요약하고 결론을 맺는다.

II. Preliminaries

1. CardRF dataset

CardRF dataset는 무선주파수 기반 UAV 탐지 및 식별을 연구하기 위해 2021년에 수집 · 공개된 데이터셋이다. 데이터가 수집된 장소는 미국 노스캐롤라이나주 롤리의 Mid Pines Road에 있는 Lake Wheeler 사이트(야외환경)이며, 총 6종의 드론, 5종의 Bluetooth 장치 및 2종의 Wifi 장치에서 수집된 신호를 포함하고 있다. 세부적인 현황은 Table. 1에서 보는 바와 같다.

데이터 수집 과정에서 드론은 신호탐지 시스템으로부터 8~12미터 이격된 범위 내에서 다양한 모드로 비행했으며, 2.4GHz 대역 주파수를 수집하기 위하여 24 dBi 2.4 GHz 그리드 포물선 안테나와 2.4 GHz 대역 통과 필터를 사용하였다. 무선주파수는 20Gs/s의 속도로 샘플링하였다.

Table 1. Detailed list of devices in CardRF dataset

Device	Manufacturer	Model
UAV	DJI	Phantom 4
		Inspire
		Matrice 600
		Mavic Pro 1
	Beebeerun	FPV RC drone mini quadcopter
	3DR	Iris FS-TH9x
Bluetooth	Apple	Iphone 6S
		Iphone 7
		Ipad 3
	FitBit	Charge3 smartwatch
	Motorola	E5 Cruise
WiFi	Cisco	Linksys E3200
	TP-link	TL-WR940N

CardRF dataset는 실생활에서 쓰이는 다양한 무선통신 기기 신호와 다양한 드론 모델별 상이한 비행모드에 따라 발생하는 무선주파수를 포함하고 있다. 또한, 동일한 시기에 공개된 다른 드론 무선주파수 데이터세트와 달리 야외환경에서 수집되었기 때문에 실제 전장환경에서 드론 탐지 및 식별을 위한 연구에 가장 적합한 데이터세트라고 할 수 있다.[9]

2. Related works

드론 탐지를 위해 기계학습이나 딥러닝 모델 등을 활용하는 연구는 크게 4가지 형태로 구분하여 진행되고 있다.[10] 첫 번째는 레이더를 이용해 드론 존재 여부를 파악하거나 조류와 드론을 식별하는 연구이다. 이 경우 사용된 레이더 기종에 따라 탐지 거리에 차이가 발생할 수 있고, 탐지 가능 거리도 달라진다. 두 번째는 본 연구처럼 드론에서 방출되는 무선주파수를 이용해 드론의 존재 여부와 기종 등을 식별하는 연구이다. 이러한 접근방법은 드론과 제어기 사이에 상호 통신이 없는 GPS 기반의 자율형 드론 탐지는 제한될 수 있으며, 특히 전자기파 간섭이 증가하는 도시환경에서는 탐지능력이 심각하게 저하될 수 있다.[11] 세 번째는 드론 동작 시 발생하는 소리를 이용해 탐지하는 연구이다. 네 번째는 이미지나 영상, 열감지 센서를 활용해 드론을 탐지하는 연구이다.

4가지 형태의 접근방법 중 무선주파수를 기반으로 드론을 탐지하는 연구에서 인공지능 기반 모델은 매우 다양한 방식으로 활용되고 있다. 그러나 본 절에서는 제안하는 접근방법과의 비교를 위해 동일한 데이터세트를 활용했던 연구들을 중점적으로 고찰하고, 이어서 주목할만한 연구성과를 달성한 일부 연구에 대해 정리한다.

Olusiji O. Medaiyese 등[12]은 환경요인이나 채널효과로 인한 신호대 잡음비(Signal-to-Noise Ratio)의 변화에 강력한 시스템을 만들기 위해 SDAE(Stacked Denoising Autoencoder)를 사용하였다. 그리고 HHT와 웨이블릿 패킷 변환(Wavelet Packet Transform, 이하 WPT)으로 신호에서 특성을 추출하고, XGBoost 모델을 학습 모델로 활용하였다. SDAE의 경우 0.00793의 재구성오류를 달성하였고, 드론 신호 탐지에서 95.7%, 드론 기종 분류에서 83.3%의 정확도를 달성하였다.

Olusiji O. Medaiyese 등은 후속 연구[13]에서 특성추출에 Harmonic Wavelet Transform(HWT)을 적용하고, 분류모델로는 KNN, SVM 및 앙상블 등 기계학습 알고리즘을 사용하였다. 또한, WPT를 통해 시간/주파수 도메인 에너지분포 스펙트로그램을 추출하고, 이를 이미지로 학습하는 SqueezeNet 전이학습모델을 사용하였다. 실험 결과 드론 탐지에서는 99.9%, 드론 기종 분류에서는 90.9%의 정확도를 달성하였다.

Syed Samiul Alam 등[14]은 정규화된 데이터세트를 활용해 다양한 수준의 노이즈가 미치는 영향을 평가하기 위해 0~25db의 노이즈신호를 삽입하였다. 탐지 및 식별을 위한 모델은 다중 퍼셉트론을 기반으로 풀링 계층(pooling layer)과 합성곱 계층(convolutional layer)을 결합한 딥러닝 신경망을 구성하였으며, 드론 탐지 정확도는 30db 기준 98.64%, 0db 기준 93.81%를 달성하였다.

Xiang Yan 등[15]은 데이터 전처리 및 특성추출 시 단 시간 푸리에 변환(Short Time Fourier Transform)을 활용하여 실수부와 허수부를 산출했으며, 신경망 속도 등을 이유로 데이터의 측정속도를 60MHz로 재샘플링하였다. 모델은 트랜스포머 기반으로 특성을 토큰화하여 분류하는 'SignalFormer'를 개발하여 적용하였으며, 0db 노이즈 기준으로 드론 탐지에서 98.47%의 정확도를 달성하였다.

M. F. Al-Sa'd 등[16]은 2019년 발표된 DroneRF dataset를 활용한 연구를 실시하였다. DFT를 통해 전처리한 데이터에 합성곱 계층을 적용하여 특성을 추출했으며, 4개 층으로 구성된 딥러닝 모델을 적용하여 드론 탐지에서 99.7%, 기종 분류에서 84.5%의 정확도를 달성하였다.

M. S. Allahham 등[17]도 DroneRF 데이터세트를 활용하였다. 웨이블릿 변환(Wavelet Transform, 이하 WT)을 통해 특성을 추출하였고, 압축샘플링을 통해 데이터의 차원을 감소시켰다. 이후 VGG-16 기반 전이학습 모델을 통해 드론 탐지는 100%, 드론 기종 분류는 94.6%의 정확도를 달성하였다.

III. The Proposed Scheme

1. Experiment Preparation

1.1 Data preprocessing

공개된 드론 무선주파수 데이터세트는 수집된 신호 데이터를 샘플링하여 시간 영역(time domain)으로 나열해 놓은 형태이다. 이러한 신호자료를 분석하여 기계학습이 가능한 특성을 추출하는 전처리 과정이 반드시 필요하며, 주로 DFT나 WT, HHT 등이 사용된다.[18]

DFT는 시계열 특성을 가지고 있는 주파수 신호를 연속된 사인파와 코사인파로 분해하여 분석하는데 사용되며, 각 주파수 성분의 원 신호에 대한 기여도와 주기적 패턴을 식별할 수 있다. 시계열적 특성을 가지는 주파수 신호를 각 주파수 성분별 특성으로 나누기 때문에 기계학습을 위한 특성추출에도 유용한 방식으로서, Eq.(1)과 같이 정의된다.[19]

$$y_i(m) = \left\| \sum_{n=1}^N x_i(n) e^{-\frac{j2\pi k(n-1)}{N}} \right\| \quad \text{Eq.(1)}$$

Eq.(1)에서 $y_i(m)$ 과 $x_i(n)$ 은 각각 산출된 특성벡터와 입력된 주파수 데이터를 의미한다. m 은 변환할 주파수의 해상도를 의미하며, n 은 시계열에 따라 샘플링된 주파수 데이터를 의미한다. DFT 시 주파수 입력값은 시계열에 따라 입력되며, 각 주파수 성분별 특징값이 실수성분과 허수성분으로 추출되면 그 절대값을 취하여 특성으로 사용한다.[20]

HHT는 자연계에서 발생하는 비선형, 비정상 신호를 분석하는데 사용되는 방법이다. 시간에 따라 주파수 성분이 변화하거나, 전통적인 신호 처리방법으로 분석이 어려운 신호를 분석하는데 유용하다. 주로 기상자료나 심전도 등 자연계에 존재하는 각종 시계열 데이터 분석에 널리 사용되고 있다. HHT는 원 신호를 경험적 모드 분해법(Empirical mode decomposition, 이하 EMD)으로 분해하여 일정 기준을 충족하는 내재모드함수(Intrinsic Mode Function, 이하 IMF)를 다수 추출한다. 이렇게 추출된 IMF는 힐버트 변환(Hilbert Transform)을 통해 주파수, 진폭 등 통계적 특성으로 변환되며, 변환된 $H(x(t))$ 는 Eq.(2)와 같이 표현된다.[21]

$$H(x(t)) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(r)}{t-r} dr \quad \text{Eq.(2)}$$

DFT는 주파수 특성추출을 위하여 사용되는 기본적인 방법이고, 신호처리 및 통신, 음향학 등 많은 분야에서 광범위하게 사용되는 주파수 분석방법이다. 다른 알고리즘에 비해 수학적으로 간단하고 구현이 용이하며, 고속 푸리에 변환(Fast Fourier Transform) 알고리즘을 사용할 경우 계산 복잡도가 $O(M \log N)$ 으로 줄어드는 장점이 가진다. 그러나 주파수의 시간에 대한 데이터가 손실되며 비선형 신호에 대하여 효과적이지 못하다는 단점이 있다.

HHT는 DFT와 비교했을 때 주파수의 시간 성분을 세밀하게 추적할 수 있고, 비선형 신호에도 강하다.[21] 이는 주파수 호핑 등 보안성을 강화한 최근 드론의 무선주파수 사용 방식을 고려할 때 큰 강점이 될 수 있다. 그러나 기본적으로 HHT는 DFT에 비해 계산량이 많으며 변환에 오랜 시간이 걸린다는 단점이 있다.

1.2 Model

본 연구에서는 드론 탐지 및 식별을 위한 모델로 LightGBM 모델을 활용하였다. LightGBM은 Microsoft사에서 개발한 고성능 그래디언트 부스팅 기반 모델로 기존 모델에 비해 처리 속도가 빠르고 효율적인 메모리 사용이 가능하며 대규모 데이터세트에 대해서 효과적인 성능을 보인다. 특히 과적합을 예방하기 위해 학습잔차를 비교, 샘플링하여 학습성능을 효율화하는 GOSS(Gradient-based One-Side Sampling) 알고리즘과, 데이터세트 특성 간 충돌 여부를 파악하여 적은 데이터손실로 특성을 결합하여 학습속도를 향상시키는 EFB(Exclusive Feature Bundling)를 사용하여 기존 그래디언트 부스팅 모델 대비 최대 20배 빠른 속도를 보이면서도 정확도를 유지하거나, 오히려 높은 정확도를 보이는 강력한 모델이다.[22]

이러한 LightGBM을 드론 탐지 및 식별을 위한 모델로 사용한 이유는 다음과 같다.

- CNN, 전이학습이나 트랜스포머 등 딥러닝 모델을 사용했던 기존 무선주파수 분석방법들과 비교했을 때, 학습 및 식별 속도가 빨라 신속한 처리를 요하는 전장환경에 적합하다.
- 군사작전에서 사용할 드론 탐지 및 식별을 위한 장비는 경량화가 필수인데, 이를 위해서는 LightGBM 모델의 효율성과 메모리 최적화 능력이 필요하다.
- 무선주파수 신호 데이터는 그 특성상 대용량 데이터 세트가 사용될 수 밖에 없는데, LightGBM 모델은 대용량 데이터세트에 대해서도 효율적으로 작동할 수 있다.
- LightGBM 모델은 성능 최적화 등이 쉬워서 차후 모

델 유지보수 및 성능개량이 용이하다.

- LightGBM 모델은 EFB를 사용하여 다수의 특성에 대해서 빠른 처리속도를 보여준다. 이러한 장점은 DFT의 주파수 해상도 문제를 해결하는데 유용하다.

2. Experiment

실험은 2가지 형태로 구분하여 진행하였다. 먼저 드론, Bluetooth 및 WiFi 3종의 장치로부터 생성되는 무선주파수를 탐지 및 식별하는 실험(3 클래스 다중분류)을 진행하였다. 이어서 3종의 장치별 세부 모델(기종)을 탐지 및 식별하는 실험(10 클래스 다중분류)을 진행하였다.

원본 CardRF 데이터세트에는 3개사 6종의 드론, 3개사 5종의 Bluetooth 장치 및 2개사 2종의 WiFi 장치에 대한 무선주파수 신호가 포함되어 있다. 드론과 관련해서는 3DR사 Iris FS-TH9x 기종을 제외한 5종에 대해서는 드론 자체에서 발생하는 신호와 드론 컨트롤러에서 발생하는 신호를 함께 제공한다. 그러나 본 연구의 비교 대상으로 삼은 이전 연구들은 드론 컨트롤러에서 발생하는 무선주파수 신호는 제외하고 실험을 진행하였다. 따라서 본 연구에서는 정확한 성능 비교를 위해 컨트롤러 신호는 배제하였다. 다만 드론 자체에서 발생하는 신호를 제공하지 않는 Iris FS-TH9x 기종에 대해서는 컨트롤러 신호를 포함해 실험을 진행하였다. 또한, Bluetooth 장치와 관련해서도 Ipad 3 및 Iphone 7은 제외하고 Iphone 6S 기종에 대한 데이터만 활용하고, Motorola E5 Cruise 기종을 제외하여 이전 연구들과 동일하게 데이터세트를 구성한 후 드론 6종, Bluetooth 장치 2종 및 WiFi 장치 2종에 대한 3 클래스 및 10 클래스 다중분류 실험을 실시하였다.

실험과정은 Fig. 1에서 보는 바와 같다. 원본 데이터세트를 7:3의 비율로 학습 및 테스트 데이터세트로 분리한 후 DFT로 전처리한 데이터세트와 HHT로 전처리한 데이터세트를 생성하였다. 생성된 2종의 데이터세트를 개별적으로 LightGBM 모델에 학습시킨 후 각각의 테스트 데이터세트를 대상으로 성능을 측정하고 결과를 비교하였다.

DFT로 전처리를 수행하는 경우에는 데이터세트 신호를 100k개 및 500k개 샘플 포인트씩 분할하여 세그먼트를 분리하였으며, 샘플 포인트를 2가지 경우로 구분한 이유는 다음과 같다.

먼저 100k개는 CardRF 데이터세트를 생성한 저자들이 권고한 바를 따름으로써 이전 연구들과 정확한 성능 비교를 수행하기 위함이다. 추가로 500k개로 분할한 이유는 DFT 수행 시 샘플의 수가 증가하면 주파수 해상도가 높아져서 잡음과 특성 주파수를 분리하는 정확도가 향상될 수

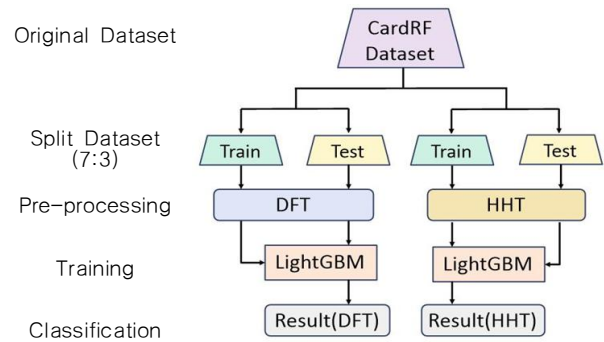


Fig. 1. Experiment Process

있다는 선행연구[23] 결과를 참고하여 샘플 포인트 변화에 따른 탐지 및 식별 성능향상 효과를 확인하기 위함이다.

분할 기준주파수 기준점은 512개를 적용하였고, 추출된 $y_i(m)$ 의 실수 성분과 허수 성분을 각 세그먼트의 특성 벡터로 사용하였다. 샘플 포인트를 100k개와 500k개로 설정하고 DFT로 전처리를 수행한 후 최종적으로 생성된 학습 및 테스트 데이터세트의 세부 현황은 Table 2에서 보는 바와 같다.

Table 2. Dataset after DFT pre-processing

# of Sample	Device	Model	Train	Test
100k	UAV	Phantom	4,375	1,875
		Inspire	4,375	1,875
		Matrice 600	8,750	3,750
		Mavic Pro 1	4,375	1,875
		Beebeerun	8,750	3,750
		Iris FH-TH9x	8,750	3,750
	Blue tooth	Iphone 6S	8,750	3,750
		Charge3	8,750	3,750
	WiFi	Linksys E3200	8,750	3,750
		TL_WR940N	8,750	3,750
500k	UAV	Phantom	875	375
		Inspire	875	375
		Matrice 600	1,750	750
		Mavic Pro 1	875	375
		Beebeerun	1,750	750
		Iris FH-TH9x	1,750	750
	Blue tooth	Iphone 6S	1,750	750
		Charge3	1,750	750
	WiFi	Linksys E3200	1,750	750
		TL_WR940N	1,750	750

특성은 총 1,024개가 추출되었으며, 각 특성은 숫자로 표현되어 있다. 샘플 포인트 수를 100k개로 설정한 경우 특성 벡터 길이는 36,863~45,055바이트이며, 500k개로 설정한 경우에는 36,863~40,959바이트로 확인되었다.

HHT 수행 시에는 각 EMD를 통해 2개의 IMF를 분리하였으며, 각각의 IMF에 대하여 힐버트 변환을 통해 진폭, 첨도, 피크 투 피크 등 14개의 특성을 추출하여 총 28개의

특성 벡터를 사용하였다. 이 과정에서 일부 파일에서는 IMF가 2개가 아닌 1개만 추출되는 경우가 확인되었다. IMF가 1개만 추출되면 28개의 특성을 추출할 수 없기에 학습에 부정적인 영향을 미칠 수 있어 제외하였으며, 그 결과 3가지 기종의 데이터 수에 미세한 차이가 발생했다.

샘플 포인트 수는 DFT로 전처리를 수행한 경우와 동일하게 100k개와 500k개로 설정하였고, 최종적으로 생성된 학습 및 테스트 데이터세트의 현황은 Table 3에서 보는 바와 같다. 생성된 특성 벡터의 길이는 샘플 포인트 수가 100k개인 경우 475~643 바이트이며, 500k개인 경우에는 503~643 바이트인 것으로 확인되었다. 추출되는 특성의 수가 28개이기 때문에 1,024개의 특성이 추출되는 DFT 전처리에 비해 특성 벡터의 길이가 상당히 짧다.

Table 3. Dataset after HHT pre-processing

# of Sample	Device	Model	Train	Test
100k	UAV	Phantom	4375	1875
		Inspire	4375	1875
		Matrice 600	8750	3750
		Mavic Pro 1	4375	1875
		Beebeerun	8750	3750
	Iris FH-TH9x	8394	3685	
	Blue tooth	Iphone 6S	8750	3750
		Charge3	8583	3660
	WiFi	Linksys E3200	8750	3750
		TL_WR940N	8750	3750
500k	UAV	Phantom	875	375
		Inspire	875	375
		Matrice 600	1750	750
		Mavic Pro 1	875	375
		Beebeerun	1750	750
	Iris FH-TH9x	1731	736	
	Blue tooth	Iphone 6S	1750	750
		Charge3	1750	750
	WiFi	Linksys E3200	1750	750
		TL_WR940N	1750	750

CardRF 데이터세트는 500만개의 데이터 샘플을 가진 복수의 파일들로 구성되어 있고, 데이터파일 한 개는 0.25 밀리초 동안 수집된 주파수를 포함하고 있다. 실험 진행 과정에서 파일을 모두 전처리하는데 소요된 시간은 샘플 포인트 500k개 기준으로 DFT는 3시간 48분이었으며, HHT는 12시간 15분이었다. 이러한 결과를 근거로 판단했을 때 데이터세트가 보다 커지면 DFT를 사용하는 것이 HHT에 비해 데이터 전처리 소요 시간을 획기적으로 감소시킬 수 있는 유용한 접근방법이 될 것이다.

각 데이터세트 별로 추출된 특성을 LightGBM 모델에 학습시킨 후 성능평가를 수행하였다. 이때 학습 데이터세트와 테스트 데이터세트는 7:3의 비율로 분할하였다. 모든

실험은 Intel i5-8300H CPU @ 2.30GHz, 16.0GB RAM, NVIDIA GeForce GTX 1050 VGA, Windows 10 운영체제에서 Jupyter notebook 환경으로 수행하였다.

3. Results and Analysis

전처리가 완료된 데이터세트를 LightGBM 모델에 적용하여 30회 반복 실험하였으며, 실험 시 적용된 주요 하이퍼파라미터는 Table 4에서 확인할 수 있다.

Table 4. Hyperparameter for LightGBM model

Hyperparameter	Value
epoch	2000
max leaves	4
learning rate	0.05
max depth	10
callbacks	30

실험 결과, DFT로 전처리를 수행하였을 때 샘플 포인트 수 100k개 기준으로 3 클래스 및 10 클래스 다중분류에서 각각 99.63% 및 95.65%의 정확도를 달성하였고, 500k에서는 99.40% 및 96.83%의 정확도를 달성하였다. 샘플 포인트 수를 100k에서 500k로 증가시켰을 때 3 클래스 다중분류 정확도는 0.23%P 감소하였지만, 10 클래스 다중분류 정확도는 1.18%P 증가하였다.

HHT로 전처리를 수행했을 때에는 샘플 포인트 100k개 기준으로 3 클래스 다중분류에서 99.10%, 10 클래스 다중분류에서 91.60%의 정확도를 달성하였다. 500k에서의 다중분류 정확도는 각각 99.28% 및 95.36%를 달성하였고, 3 클래스 다중분류보다 10 클래스 다중분류에서 성능이 더 크게 향상됨을 확인하였다. 세부적인 성능분석 결과는 Table 5와 Table 6에서 확인할 수 있다.

Table 5. Experimental results (100k)

Classification	Index	Result	
		DFT	HHT
3 classes	Accuracy	99.63	99.10
	Precision	99.63	99.17
	Recall	99.61	98.96
	F1 score	99.62	99.06
10 classes	Accuracy	95.65	91.60
	Precision	94.62	88.74
	Recall	94.62	88.01
	F1 score	94.59	88.30

Table 6. Experimental results (500k)

Classification	Index	Result	
		DFT	HHT
3 classes	Accuracy	99.40	99.28
	Precision	99.50	99.33
	Recall	99.28	99.15
	F1 score	99.39	99.24
10 classes	Accuracy	96.83	95.36
	Precision	95.39	93.39
	Recall	95.79	93.41
	F1 score	95.85	93.63

DFT로 전처리를 수행했을 때 소요시간은 훨씬 짧아지면서도 모든 성능평가 지표에서 HHT로 전처리를 수행한 경우보다 우수한 성능을 보였다. 이처럼 DFT가 HHT보다 우수한 성능을 달성한 이유는 추출되는 특성 수와 고차원 특성을 효율적으로 학습하는 LightGBM 모델의 장점 때문이다. 우선 HHT는 28개의 특성만을 추출하나 DFT는 1,024개의 특성을 추출하기 때문에 주파수 해상도가 높아 보다 정교한 학습이 가능하다. 그리고 본 연구에서 적용한 LightGBM 모델은 DFT로 추출한 고차원 특성을 EFB를 통해 손실을 최소화하면서 효율적으로 학습할 수 있다.

3 클래스 다중분류 실험에 대한 혼동행렬은 Fig. 2 및 Fig. 3에서 보는 바와 같다. 샘플 포인트의 수가 100k인 경우보다 500k인 경우에서 클래스 간 오분류가 확연하게 감소함을 확인할 수 있다.

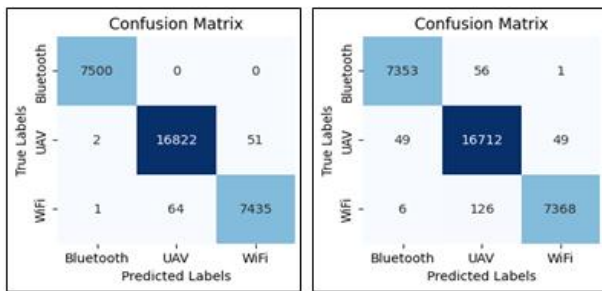


Fig. 2. Confusion Matrix of 3 classes classification (100k)

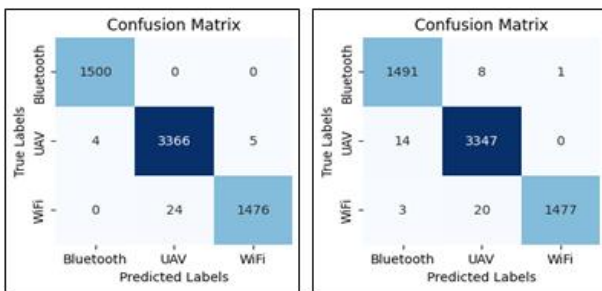


Fig. 3. Confusion Matrix of 3 classes classification (500k)

10 클래스 다중분류 실험에 대한 혼동행렬은 Fig. 4 및 Fig. 5에서 보는 바와 같다. Fig. 4는 샘플 포인트의 수를 100k로 설정하고 전처리를 수행한 후의 다중분류 실험에 대한 혼동행렬로서, 드론 기종 분류에서 다수의 오분류가 발생하였음을 확인할 수 있다.

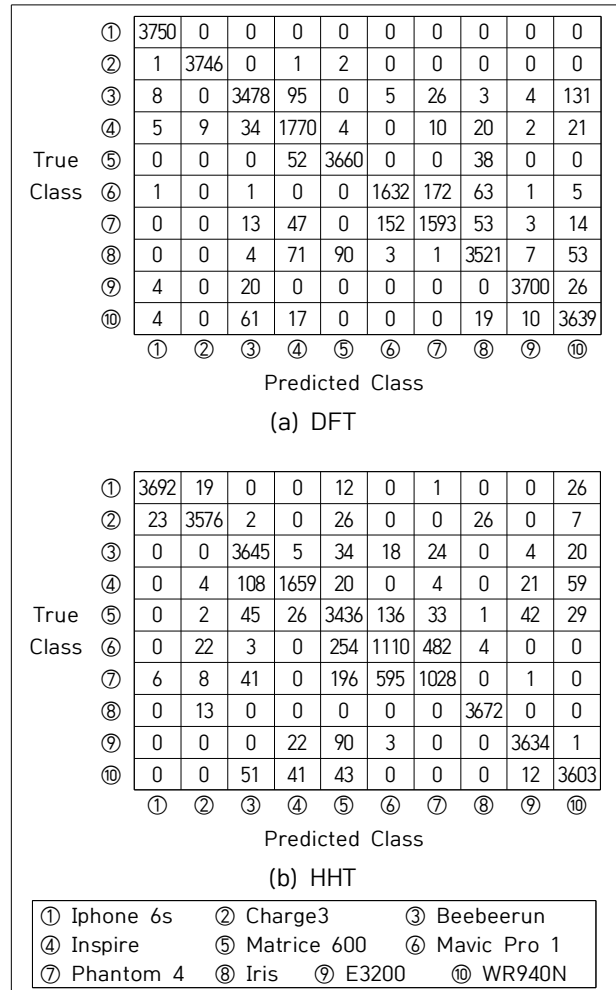


Fig. 4. Confusion Matrix of 10 classes classification (100k)

Fig. 5는 샘플 포인트의 수를 500k로 설정하고 전처리를 수행했을 때의 다중분류 실험에 대한 혼동행렬을 보여주고 있다. 샘플 포인트의 수를 100k로 전처리를 수행한 경우보다 클래스별 오분류가 확연하게 감소하였음을 확인할 수 있다. 그러나 DJI사 드론 2종(⑥Phantom 4 및 ⑦Mavic Pro 1)에 대해서는 여전히 DFT 및 HHT 모두에서 오분류가 다수 발생하였다.

본 연구와 같이 CardRF 데이터세트를 활용했던 선행연구들과의 성능 비교 결과는 Table 7에서 보는 바와 같다. 드론, Bluetooth 및 WiFi 3종의 장치에 대한 무선주파수 탐지 및 식별 성능은 선행연구에서 제시된 최고 성능에는

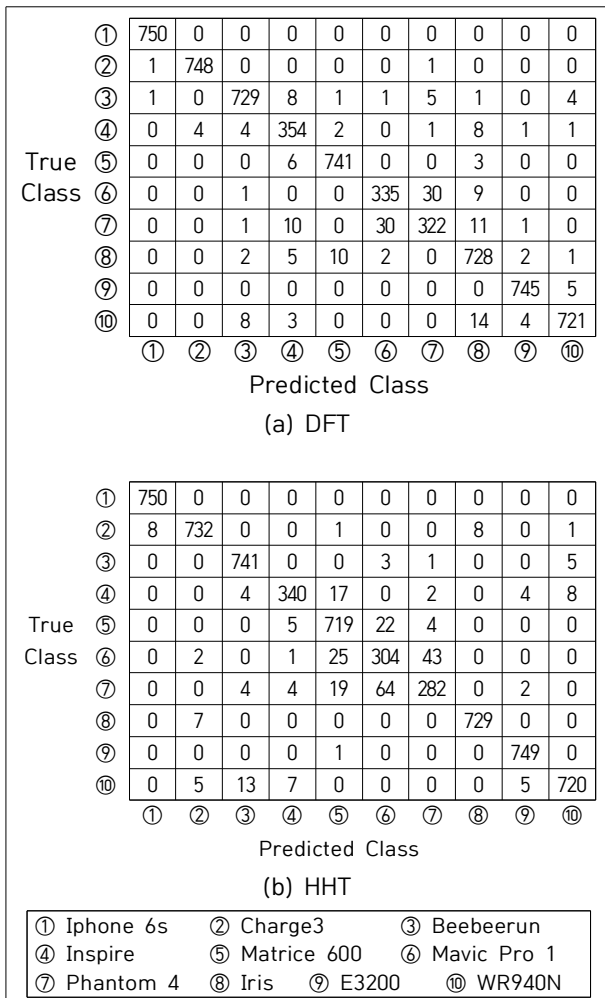


Fig. 5. Confusion Matrix of 10 classes classification (500k)

Table 7. Performance comparison

Classification		Accuracy					
		Proposed		Olusiji et al. [12]	Olusiji et al. [13]	Alam et al. [14]	Yan et al. [15]
		DFT	HHT				
3 classes	100k	99.63	99.10	90.4	99.9	98.64	98.47
	500k	99.40	99.28				
10 classes	100k	95.65	91.60	67.86	90.9	80.6	-
	500k	96.83	95.36				

미치지 못하였지만 다른 연구보다는 향상되었다. 그러나 드론 탐지에서 최고 성능을 달성했던 선행연구[13]의 경우 다양한 기계학습 알고리즘을 분류모델로 적용하면서 전이 학습까지 수행했지만, 본 연구는 LightGBM 모델만 단독으로 사용하고도 99% 이상의 성능을 달성했다는 점에서 큰 의미를 가진다고 볼 수 있다.

드론 기종을 포함한 10종의 장치에 대한 탐지 및 분류 성능은 선행연구 대비 크게 향상되었다. 샘플 포인트의 수를 500k개로 설정하고 DFT로 전처리를 수행한 후 학습한

경우 가장 우수한 성능을 보였으며, 이전까지 가장 우수한 결과를 보였던 연구[13]보다 성능이 5.93%p 향상되었다. 이러한 결과를 통해 DFT를 활용한 주파수 특성추출과 LightGBM 모델의 결합이 드론과 관련된 무선주파수 탐지 및 분류에서 유용한 대안이 될 수 있음을 알 수 있다.

IV. Conclusions

드론에 의한 위협이 점점 더 심각해지고 있는 상황에서 효과적인 대응을 위해서는 드론의 조기 탐지가 매우 중요하다. 이에 본 논문에서는 인공지능 기술을 기반으로 드론 및 관련 장치에서 발생하는 무선주파수를 탐지 및 식별할 수 있는 방안을 제시하였다.

탐지 및 식별 모델 구축 과정에서는 신속성과 정확성, 그리고 경량화를 중요시하는 전장 환경에서의 적용 가능성을 높이는 것을 최우선으로 고려하여 전처리 기법 및 학습 모델을 선정하였다. 우선 데이터 전처리 과정에서는 주파수 신호 데이터에서 특성을 추출하는 다양한 방법들 중 처리 속도가 빠른 DFT를 적용하였다. 그리고 DFT에 비해 비교적 처리 속도가 느린 HHT를 추가로 적용하여 성능을 비교하였다. 학습 모델은 분류 속도 및 정확성, 유지보수 용이성 등을 종합적으로 고려하여 기계학습 모델인 LightGBM 모델을 채택하였다.

제안하는 접근방법의 효과성을 검증하기 위한 데이터셋트는 공개된 드론 무선주파수 데이터셋트 중 실내 실험실이나 전자파가 차폐된 무선 챔버 등에서 수집되지 않고, 야외환경에서 수집되어 가장 현실성있는 데이터셋트라고 할 수 있는 CardRF 데이터셋트를 활용하였다. CardRF 데이터셋트에 대해 DFT와 HHT를 적용해 주파수 특성을 추출하는 전처리 과정에서는 샘플 포인트의 수를 100k와 500k로 구분하여 수행하였으며, LightGBM 모델을 이용해 학습을 수행한 후 성능을 분석하였다.

샘플 포인트의 수를 100k개로 설정하고 DFT로 전처리한 경우 3 클래스 다중분류에서는 99.63%, 10 클래스 다중분류에서는 95.65%의 정확도를 달성하였다. HHT로 전처리를 수행한 경우에는 3 클래스 다중분류에서 99.10%, 10 클래스 다중분류에서 91.60%의 정확도를 보였다. 그리고 주파수 특성 추출 과정에서 노이즈에 대한 내성을 높이기 위해 샘플 포인트의 수를 500k로 증가시켜 DFT로 전처리를 수행한 경우에는 3 클래스 다중분류 정확도가 미세하게 낮아진 99.40%로 나타났지만, 10 클래스 다중분류 정확도는 96.83%로 1.18%p 향상되었다. 샘플 포인트 수

를 500k개로 설정하고 HHT로 전처리를 수행한 경우에는 3 클래스 다중분류 성능이 크게 향상되지 않았지만, 10 클래스 다중분류 성능은 3.76%p로 크게 향상되었다.

DFT와 HHT로 전처리를 수행하는 과정에서 소요되는 시간은 DFT가 HHT에 비해 획기적으로 짧았으며, 그럼에도 불구하고 DFT로 전처리를 수행한 경우가 모든 성능평가 지표에서 우수한 결과를 보였다. 그리고 동일한 데이터 세트를 활용했던 이전 연구들과 성능을 비교해 본 결과, 3 클래스 다중분류 성능은 유사하거나 미세하게 향상되었지만 10 클래스 다중분류 성능은 크게 향상되었다.

한편 연구를 진행하면서 이전 연구와 비교하여 데이터 전처리 속도와 학습 모델의 분류 속도가 어느 정도 향상되었는지에 대해서는 직접적인 비교 분석을 실시하지 못하였다. 전처리 속도는 파라미터 값의 설정에 따라 얼마든지 달라질 수 있으며, 논문에서 제시된 DFT와 HHT의 전처리 속도 비교 결과 역시도 절대적인 비교 결과로 보기는 어렵다. 따라서 향후 연구에서는 다양한 인공지능 알고리즘을 활용한 드론 탐지 및 식별 모델을 직접 구축하고, 전처리 방법도 다양하게 적용하여 처리 속도를 비교하면서 전장 환경에 최적화될 수 있는 드론 탐지 및 식별 모델을 탐색해 나갈 예정이다.

REFERENCES

- [1] The Business Research Company, "UAV Drones Global Market Report 2024", <https://www.researchandmarkets.com/report/unmanned-aerial-vehicles-uavs-drones>
- [2] K. Telli, O. Kraa, Y. Himeur, A. Ouamane, and M. Boumehraz, "A Comprehensive Review of Recent Research Trends on Unmanned Aerial Vehicles (UAVs)", *Systems*, 11, 400, Nov 2023. DOI:10.3390/systems11080400
- [3] Anna Jackman, "Police Drones: Uses, Challenges, Futures", <https://research.reading.ac.uk/drone-geographies/wp-content/uploads/sites/271/2023/09/Police-report.pdf#:~:text=URL%3A%20https%3A%2F%2Fresearch.reading.ac.uk%2Fdrone>
- [4] O. H. Dahle, J. Rydberg, M. Dullweber, N. Peinecke and A. A. A. Bechina, "A proposal for a common metric for drone traffic density", *International Conference on Unmanned Aircraft Systems (ICUAS)*, pp. 64-72, Dubrovnik, Croatia, Jun 2022. DOI: 10.1109/ICUAS54217.2022.9836143.
- [5] Seth Cropsey, "Drone Warfare in Ukraine: Historical Context and Implications for the Future," <https://www.hoover.org/research/drone-warfare-ukraine-historical-context-and-implications-future>
- [6] Lauren Kahn, "How Ukraine Is Using Drones Against Russia," <https://www.cfr.org/in-brief/how-ukraine-using-drones-against-russia>
- [7] Kristen D. Thompson, "How the Drone War in Ukraine Is Transforming Conflict," <https://www.cfr.org/article/how-drone-war-ukraine-transforming-conflict>
- [8] M. H. Rahman, Mohammad A. S. Sejan, M. A. Aziz, R. Tabassum, Jung-In Baik, and Hyoung-Kyu Song, "A Comprehensive Survey of Unmanned Aerial Vehicles Detection and Classification Using Machine Learning Approach: Challenges, Solutions, and Future Directions", *2024 Remote Sensing* 16, no. 5: 879., Mar 2024. DOI:10.3390/rs16050879
- [9] Olusiji O. Medaiyese, M. Ezuma, A. Lauf, and A. Adeniran, "Cardinal RF (CardRF): An Outdoor UAV/UAS/Drone RF Signals with Bluetooth and WiFi Signals Dataset", *IEEE Dataport*, Jul 2022. DOI:10.21227/1xp7-ge95
- [10] U. Seidaliyeva, L. Ilipbayeva, K. Taissariyeva, N. Smailov, and E. Matson, "Advances and Challenges in Drone Detection and Classification Techniques: A State-of-the-Art Review", *Sensors* 24(1):125, Dec 2023. DOI: 10.3390/s24010125.
- [11] N. Al-Iqubaydhi, A. Alenezi, T. Alanazi, A. Senyor, N. Alanezi, B. Alotaibi, M. Alotaibi, A. Razaque, and S. Hariri, "Deep learning for unmanned aerial vehicles detection: A review", *Computer Science Review*, Vol. 51, No. C, Jun 2024. DOI: 10.1016/j.cosrev.2023.100614.
- [12] Olusiji O. Medaiyese, M. Ezuma, A. P. Lauf, and A. A. Adeniran, "Hierarchical Learning Framework for UAV Detection and Identification", *IEEE Journal of Radio Frequency Identification*, vol. 6, pp. 176-188, Mar 2022. DOI: 10.1109/JRFID.2022.3157653.
- [13] Olusiji O. Medaiyese, M. Ezuma, A. P. Lauf, and Ismail Guvenc, "Wavelet transform analytics for RF-based UAV detection and identification system using machine learning", *Pervasive and Mobile Computing*, Volume 82, 101569, Jun 2022. DOI:10.1016/j.pmcj.2022.101569.
- [14] Alam, Syed Samiul, Arbil Chakma, Md Habibur Rahman, Raihan Bin Mofidul, Md Morshed Alam, Ida Bagus Krishna Yoga Utama, and Yeong Min Jang, "RF-Enabled Deep-Learning-Assisted Drone Detection and Identification: An End-to-End Approach", *Sensors* 23, no. 9: 4202, April 2023. DOI:10.3390/s23094202
- [15] Yan, Xiang, Bing Han, Zhigang Su, Jingtang Hao., "SignalFormer: Hybrid Transformer for Automatic Drone Identification Based on Drone RF Signals", *Sensors* 23, no. 22: 9098, November 2023. DOI:10.3390/s23229098
- [16] Mohammad F. Al-Sa'd, Abdulla Al-Ali, Amr Mohamed, Tamer Khattab, and Aiman Erbad, "RF-based drone detection and identification using deep learning approaches: An initiative towards a large open source drone database", *Future Generation Computer Systems*, 100: 86-97, November, 2019. DOI:10.1016/j.future.2019.05.007

- [17] M. S. Allahham, T. Khattab, and A. Mohamed, "Deep Learning for RF-Based Drone Detection and Identification: A Multi-Channel 1-D Convolutional Neural Networks Approach", 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIOT), pp. 112-117, Doha, Qatar, February, 2020. DOI:10.1109/ICIOT48696.2020.9089657.
- [18] Z. Oubrahim, A. Yassine, B. Mohamed, and O. Mohammed, "Power Quality Disturbances Characterization Using Signal Processing and Pattern Recognition Techniques: A Comprehensive Review", *Energies* 16, no. 6: 2685, Mar 2023. DOI:10.3390/en16062685
- [19] O. Brandes, J. Farley, M. Hinich, and U. Zackrisson, "The time domain and the frequency domain in time series analysis", *The Swedish Journal of Economics*, 25-42, Mar 1968. DOI:10.2307/3438983
- [20] I. Nemer, S. Tarek, A. Irfan, Ansar Ul-Haque Yasar, and Mohammad A. R. Abdeen., "RF-Based UAV Detection and Identification Using Hierarchical Learning Approach", *Sensors* 21, no. 6: 1947, Mar 2021. DOI:10.3390/s21061947
- [21] Norden E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing H. Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung and Henry H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis", *Proc. R. Soc. Lond. A*.454903-995 March 1998. DOI:10.1098/rspa.1998.0193
- [22] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu, "LightGBM: A highly efficient gradient boosting decision tree", *Advances in neural information processing systems* 30, 2017.
- [23] Ç. Candan, "Analysis and Further Improvement of Fine Resolution Frequency Estimation Method From Three DFT Samples", *IEEE Signal Processing Letters*, Vol. 20, No. 9, pp. 913-916, Sep 2013. DOI: 10.1109/LSP.2013.2273616.

Authors



Ki-Hyeon Sung received the B.S. degrees in Industrial Engineering from the Air Force Academy Korea. He is currently a graduate student in the Department of Defense Science, Korea National Defense University.

His research interests include artificial intelligence, machine learning, and the security of drones and satellites.



Soo-Jin Lee received B.S., M.S. and Ph.D. degrees in Computer Science from Korea Military Academy, Yonsei University and Korea Advanced Institute of Science and Technology(KAIST) in 1992, 1996 and 2006.

He is currently a professor of the Department of Defense Science, Korea National Defense University from 2006. His research interests include National Cybersecurity Policy, Intrusion Detection System, Mobile Network Security, Machine Learning, Encryption theory and applications.