| **TinyML을 통한 소리 인식 및 위치 탐지**  **요 약**  TinyML 모델을 학습시켜 드론에 장착하여 사고 피해자들의 소리를 감지하고 위치를 탐지하는 시스템을 구현한다. Sound Classification과 Sound Localization을 혼합하여 Edge Impulse 플랫폼을 사용하여 모델을 학습한다. 학습한 모델을 sony spresense에 배포하고 드론에 장착시켜 소리를 인식하고 위치를 탐지시킨다. |
| --- |

# 1. 서론

## 1.1. 연구배경

2023년 3월 튀르키예 대지진으로 8만명 이상의 사람이 목숨을 잃었다. 이러한 재난 상황에서 생존자를 빠르게 구조하는 것은 매우 중요하다. 기존의 드론 기술은 이미 재난 구조 작업에 활용되고 있지만, 생존자의 위치를 정확하게 파악하는 것이 여전히 어려운 문제가 있다. 기존의 드론 기술은 대부분 카메라나 적외선 카메라를 이용하여 생존자를 탐지하는데, 적외선 카메라의 경우 높은 가격과 복잡한 사용법 때문에 일반적인 상황에서 사용하기 어려운 단점이 있다. 또한, 카메라로만 생존자를 탐지하는 경우 생존자의 음성이나 소리를 감지하는 것이 어렵기 때문에, 이러한 상황에서는 생존자의 위치를 정확하게 파악하기 어려운 경우가 발생할 수 있다.

따라서, 이러한 문제점을 TinyML 기술을 활용하여 해결하고자 한다. TinyML 기술은 저전력, 저비용, 소형 디바이스에 적합한 머신 러닝 기술로, 드론과 같은 소형 디바이스에서도 사용이 가능합니다. 이를 활용하여 드론이 생존자의 음성이나 소리를 감지하여 위치를 파악하고, 이를 기반으로 생존자를 신속하게 구조할 수 있는 시스템을 개발할 것이다. 이러한 시스템은 재난 구조 작업에서 매우 유용하게 활용될 수 있으며, 이에 따라 많은 인명 구조 작업에서 적극적으로 활용될 것으로 기대된다.

## 1.2. 연구목표

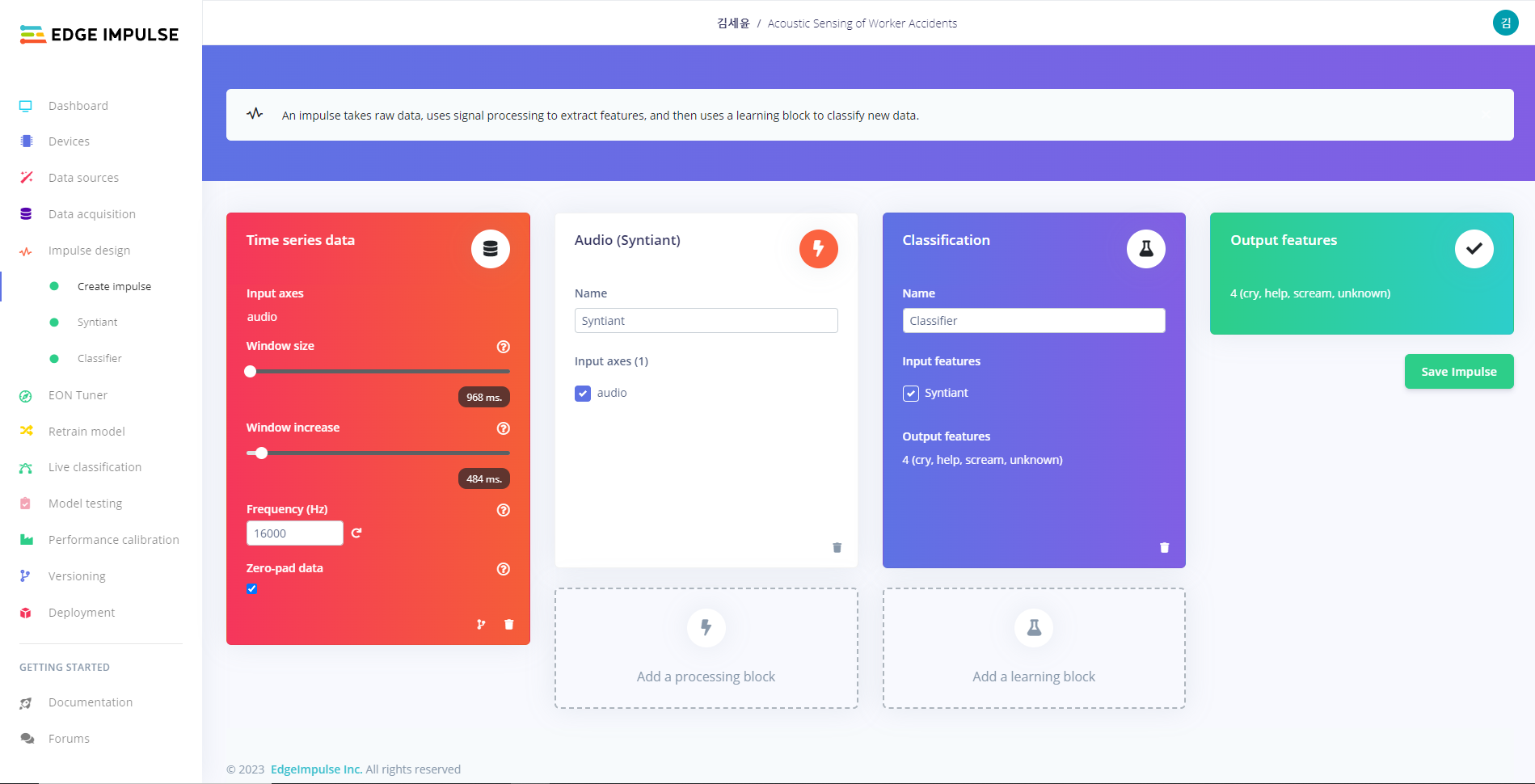
본 연구의 목표는 Edge Impulse 플랫폼을 활용하여 TinyML 모델을 학습시켜, 드론에 장착하여 사고 피해자들의 소리를 감지하고 위치를 탐지하는 시스템을 구현하는 것이다. 이를 통해 재난 구조 작업에서 드론이 보다 정확하고 신속하게 생존자를 찾을 수 있으며, 재난 구조 작업의 성공 확률을 높일 수 있도록 하는 것이 이번 연구의 핵심 목표이다.

# 2. 관련연구

## 2.1. TinyML

TinyML은 작고 저전력인 인공지능 모델을 만들고 배포하는 기술로, 임베디드 시스템에서도 활용할 수 있다. 이를 통해 기기 내부에서 지능적인 의사 결정을 수행하고, 외부 서버와의 통신을 최소화하여 보안성을 높이고 전력 소비를 줄일 수 있다. TinyML 기술은 IoT, 스마트 홈, 건강 모니터링 등 다양한 분야에서 활용될 수 있으며, 본 프로젝트에는 드론의 임베디드 시스템으로 사용한다. 이를 통해 작고 강력한 인공지능 모델을 개발하고 배포함으로써, 보다 효율적이고 안전한 임베디드 시스템을 구현할 수 있다.

### 2.1.1. Edge Impulse



<그림 1 Edge Impulse>

Edge Impulse는 작고 효율적인 TinyML 모델을 만들기 위한 종합적인 개발 플랫폼이다 [1]. 사용자는 Edge Impulse 웹 사이트에서 데이터 수집부터 모델 학습까지 전 과정을 진행할 수 있으며, 학습시킨 모델을 다양한 디바이스와 통합이 가능하다. 또한, Edge Impulse를 통해 데이터 전처리, 모델 최적화, 배포 등의 과정을 자동화하여, 비전문가도 쉽게 인공지능 모델을 개발하고 배포할 수 있다. 이를 통해, IoT, 스마트 홈, 로봇, 드론 등 다양한 임베디드 시스템에서 강력하고 효율적인 인공지능 모델을 쉽게 개발할 수 있다.

## 2.2. 기존 소리 인식 및 탐지 연구

### 2.2.1. Sound Classification

Sound classification은 소리의 특징을 분석하여 해당 소리의 종류를 분류하는 인공지능 기술 중 하나다 [2, 3]. 관련 연구에서는 CNN(Convolutional Neural Network)과 TDSN(Tensor Deep Stacking Network)을 훈련하기 위해 환경 소리의 스펙트로그램 이미지를 사용하여 Sound Classification을 구현했다.

Sound classification을 통해, 본 프로젝트에서는 사고 현장에서 살려 달라는 소리나 울음 소리 등을 인식할 수 있게 모델을 학습시킨다. 이를 구현하기 위해서는, 먼저 해당 소리 데이터를 수집하고, 이를 토대로 모델을 학습시켜야 한다. 학습된 모델은 드론 등의 임베디드 시스템에 적용되어, 실시간으로 소리를 분류하고, 사고 피해자들의 상황 파악 및 구조 작업에 활용할 수 있다. 이를 통해, 재난 현장에서 빠른 대응과 효과적인 구조 작업을 수행할 수 있게 된다.

### 2.2.2. Sound Localization

Sound localization[4]은 소리가 발생한 위치를 인식하는 인공지능 기술 중 하나다. 관련 연구에서는 CDR(coherent-to-diffuse power ratio)의 추정치를 기반으로 한 ASN(acoustic sensor network)에서 음원 위치 파악을 조사한다. 실제 음향과 마이크 쌍 사이의 거리를 통해 위치를 예측한다. 획득된 거리 추정치는 음원의 위치를 ​​추정하기 위해 알려진 상대 노드 위치를 갖는 ASN에 의해 ​​사용된다.

Regression을 사용하여 학습된 모델은 소리의 특성을 기반으로 해당 소리가 발생한 위치를 예측할 수 있다. Sound localization은 음성 인식, 화상 회의 등의 다양한 분야에서 활용될 수 있으며, 이를 위해 다양한 학습 방법이 사용된다. 본 프로젝트에서는, 드론의 임베디드 시스템이 소리를 인식하고, 그 소리가 발생한 위치를 파악하여, 구조 작업에 적극 활용한다.

Regression은 sound localization 중 하나로, 학습된 모델은 입력 데이터와 출력 데이터 간의 관계를 파악하여, 새로운 입력 데이터가 주어졌을 때, 해당 출력 데이터를 예측할 수 있도록 합니다.

### 

### 2.2.3. Integration of Sound Classification & Localization

Sound classification은 소리의 종류를 분류하는 작업에 초점을 두고 있으며, Sound localization은 소리의 발생 위치를 파악하는 작업에 초점을 두고 있다. 그러나 최근에는 이 두 작업을 함께 처리하는 연구가 늘어나고 있다.

예를 들어, Sound classification과 Sound localization을 모두 다루는 연구로는 SeldNet [5], CRNN [6] 등이 있다. 이번 프로젝트에서는 이와 같은 기존 연구들에서 사용된 모델을 참고하여, 드론의 임베디드 시스템에서 Sound classification과 Sound localization을 함께 처리하려고 한다.

이와 같이 기존의 Sound classification과 Sound localization을 혼합하여 사용한 논문 연구를 참고하여, 이번 프로젝트에서는 드론에 이를 적용하여 Sound classification과 Sound localization을 함께 처리할 것이다.

# 3. 프로젝트 내용

## 3.1. 시나리오

### 3.1.1. 데이터 수집

데이터셋 수집 단계에서는 기존에 수집된 소리 데이터셋을 참고하여 우리가 필요한 소리를 포함하고 있는 데이터셋을 수집하는 것이 중요하다. 이를 위해 공개된 오픈 데이터셋[7]이나 경연 대회[8] 등에서 제공하는 데이터셋을 사용한다.

| **Category** | **Number of Samples** |
| --- | --- |
| alarm | 50개 |
| crying baby | 40개 |
| crash | 50개 |
| barking dog | 45개 |
| running engine | 39개 |
| female scream | 45개 |
| female speech | 40개 |
| burning fire | 51개 |
| footsteps | 42개 |
| knocking on door | 40개 |
| male scream | 31개 |
| male speech | 40개 |
| ringing phone | 40개 |
| piano | 42개 |

<표 1 데이터셋1[7]>

이 데이터셋에서 있는 여러 종류의 소리 중에서 우리가 필요로 하는 살려달라는 소리나 우는 소리 등을 찾아내어 추출한다. 나머지 소리는 noise로 라벨링을 한다. 이 과정에서는 데이터셋에서 사용할 소리에 대한 정보와 해당 소리의 레이블링 정보를 모두 수집하여 Edge Impulse에 명시해야 한다.

| **Category** | **Number of Samples** |
| --- | --- |
| static\_task | 300개 |
| flight\_task | 36개 |
| dev\_static | 3개 |
| dev\_flight | 5개 |
| static\_task\_final | 20개 |
| flight\_task\_final | 1개 |
| SPCUP19\_ground\_truth\_static.mat. | 1개 |
| SPCUP19\_ground\_truth\_flight.mat. | 1개 |
| SPCUP19final\_ground\_truth\_static.mat. | 1개 |
| SPCUP19final\_ground\_truth\_flight.mat. | 1개 |

<표 2 데이터셋2[8]>

Sound Localization을 학습시키기 위해 방향 정보를 가진 데이터셋을 활용한다. mat파일 형태로 벡터 정보값을 가지고 있으며 이를 라벨링하여 모델에 방향 정보를 학습시킨다.

이렇게 수집된 데이터셋을 이용하여 모델 학습을 진행한다.

### 3.1.2. Edge impulse를 통한 학습

데이터 전처리, 모델 학습, 모델 평가 단계는 Edge Impulse를 통해 진행한다.

첫 번째로 데이터 전처리 단계에서는 수집된 데이터셋에서 노이즈 제거 및 스케일링, 필터링, 주파수 변환 등의 전처리 과정이 진행됩니다. 이 과정에서 전처리된 데이터셋을 Edge Impulse에 업로드하여 모델 학습을 준비한다.

두 번째로 모델 학습 단계에서는 Edge Impulse에서 제공하는 다양한 알고리즘을 이용하여 소리 분류 및 위치 인식 모델을 학습합니다. 이 과정에서 모델 하이퍼 파라미터를 조정하여 최적의 모델을 구축한.

세 번째로 모델 평가 단계에서는 학습된 모델을 테스트 데이터셋으로 검증하여 모델의 성능을 평가합니다. 이 과정에서 정확도, 정밀도, 재현율, F1-score 등의 평가 지표를 사용하여 모델의 성능을 분석합니다. 이를 통해 모델의 성능을 개선하고 최종 모델을 선정한다.

### 3.1.3. Device 설계

프로젝트에 임베디드 시스템으로 학습한 모델을 넣는 device로는 sony spresense[9]를 사용한다. 사용할 device는 여러 대의 마이크를 장착하여 소리를 인지함과 동시에 각 마이크의 데이터를 분석하여 소리의 방향을 감지한다.

### 3.1.4. Device에 Deploy

학습한 모델을 export하여 디바이스에서 실행할 수 있는 형태로 변환한다. 디바이스는 컴퓨터와 연결하고 Edge Impulse에서 감지한 디바이스에 학습한 모델을 이식한다.

### 3.1.5. Drone에 Test

이식한 디바이스를 드론에 장착하여 소리를 인지하고 위치를 탐지하는 것을 테스트한다.

## 3.2. 요구사항

### 3.2.1. 데이터에 대한 요구사항

프로젝트의 목적은 드론이 사람들의 소리를 탐지하고 그들의 위치를 찾는 것이다. 따라서 데이터셋은 이러한 목적을 충족시키기 위해 충분한 양의 다양한 소리 샘플을 제공해야 한다. 또한, 방향 정보를 가지고 있어야 하며, 이를 통해 드론이 소리가 발생하는 방향을 인식할 수 있어야 한다. 데이터셋은 가능한 한 실제 상황과 유사하도록 구성되어야 하며, 이를 위해 경연 대회나 기존 연구에서 사용되었던 데이터셋을 활용한다.

### 3.2.2. Edge Impulse에 대한 요구사항

프로젝트에서 사용할 Edge Impulse 모델은 classification과 localization 두 가지 기능을 모두 수행해야 하므로, 모델링 과정에서 이를 동시에 구현할 수 있는 기능이 필요하다. 이를 위해 Edge Impulse 플랫폼에서는 이 두 가지 기능을 하나의 모델에서 모두 구현할 수 있는 기능을 제공하고 있으며, 프로젝트에서도 이를 활용할 것이다.

### 3.2.3. Device에 대한 요구사항

디바이스는 Edge Impulse와 호환성이 있어야 하며, Edge Impulse에서 학습한 모델을 디바이스에 deploy할 수 있어야 한다. 또한, 디바이스는 운용할 수 있을 정도의 하드웨어 성능을 가지고 있어야 한다. 이는 디바이스에서 inference를 수행하는 데 필요한 계산 능력과 메모리 용량을 의미한다. 마지막으로 디바이스에는 마이크가 장착되어 있어야 하며, 이를 통해 주변 소리를 수집할 수 있어야 합한다. 이러한 디바이스 요구 사항을 충족시키지 않으면, 모델을 실제로 실행하여 실시간으로 소리를 분류하거나 위치를 파악하는 것이 어려울 수 있다.

# 4. 향후 일정

4월에 sound classification과 sound localization 모델을 학습할 예정이다. 이를 위해 Edge Impulse를 이용하여 전처리, 모델링, 평가를 진행할 것이다. 이후에는 5월에 디바이스에 학습한 모델을 deploy하고, 드론에 이를 장착하여 테스트할 계획이다. 이를 통해 소리를 이용한 드론 제어 시스템의 구현을 위한 기술적인 기반을 마련하고, 이를 실제로 적용함으로써 성능을 검증할 것이다.

# 

# 5. 결론 및 기대효과

이 프로젝트는 딥러닝 기술을 활용하여 드론 사고 발생 시, 사고 피해자들의 위치를 빠르고 정확하게 파악할 수 있는 솔루션을 제공하는 것을 목표로 하고 있다.

데이터셋 수집부터 모델 학습까지 edge computing 환경에서 진행되어 성능과 속도 면에서 우수한 결과를 기대할 수 있다. 또한 디바이스에 deploy하여 드론에 적용함으로써, 사고 발생 시 보다 빠르고 정확한 대응이 가능해지며, 이는 인명 구조에 큰 도움이 될 것이다.

데이터셋의 요구사항을 충족시키기 위해 방향 정보를 포함하여 수집하였으며, Edge Impulse에서는 classification과 localization을 동시에 모델링하여 효율적인 모델 학습을 할 수 있도록 구성하였다. 또한 디바이스에 대한 요구사항 역시 Edge Impulse와의 호환성과 하드웨어 성능이 충분하도록 설정하여 문제 없이 작동할 수 있도록 구성하였다.

이를 통해 드론 사고 발생 시, 사고 피해자 위치 파악에 대한 대처 능력이 향상되어 빠른 대응이 가능해지며, 이는 인명 구조에 큰 도움이 될 것이다. 이 프로젝트를 통해 딥러닝 기술이 실제 사회에서 활용될 수 있는 예시를 보여주고, 드론 사고 발생 시 대처 능력을 향상시켜 인명 구조 분야에 기여할 수 있는 선구적인 결과물이 될 것이다.

# 6. 참고문헌

[1] Edge Impulse, <https://www.edgeimpulse.com/>

[2]A. Khamparia, D. Gupta, N. G. Nguyen, A. Khanna, B. Pandey and P. Tiwari, "Sound Classification Using Convolutional Neural Network and Tensor Deep Stacking Network," in IEEE Access, vol. 7, pp. 7717-7727, 2019, doi: 10.1109/ACCESS.2018.2888882.

[3] Sajjad Abdoli, Patrick Cardinal, Alessandro Lameiras Koerich, End-to-end environmental sound classification using a 1D convolutional neural network, Expert Systems with Applications, Volume 136, 2019,

Pages 252-263, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2019.06.040.

(<https://www.sciencedirect.com/science/article/pii/S0957417419304403>)

[4] A. Brendel and W. Kellermann, "Distributed Source Localization in Acoustic Sensor Networks Using the Coherent-to-Diffuse Power Ratio," in IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 1, pp. 61-75, March 2019, doi: 10.1109/JSTSP.2019.2900911.

[5] Sharath Adavanne, Archontis Politis, Joonas Nikunen, Tuomas Virtanen, "Sound Event Localization and Detection of Overlapping Sources Using Convolutional Recurrent Neural Networks" , <https://doi.org/10.48550/arXiv.1807.00129>

[6] Sławomir Kapka, Mateusz Lewandowski, "Sound source detection, localization and classification using consecutive ensemble of CRNN models", <https://doi.org/10.48550/arXiv.1908.00766>

[7]DCASE2023, <https://dcase.community/challenge2020/task-sound-event-localization-and-detection>

[8] Signal Processing Cup, <http://dregon.inria.fr/datasets/signal-processing-cup-2019/>

[9] Sony Spresense, https://developer.sony.com/develop/spresense/specifications