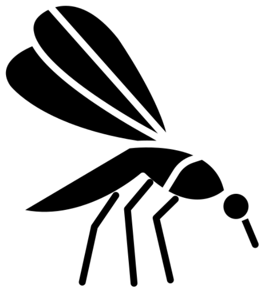
Mahidol University

MosWing: a Noise-Robust Mosquito Wingbeat Detection Model

Phuriwat Angkoondittaphong
Napahatai Sittirit
Danaidech Ardsamai

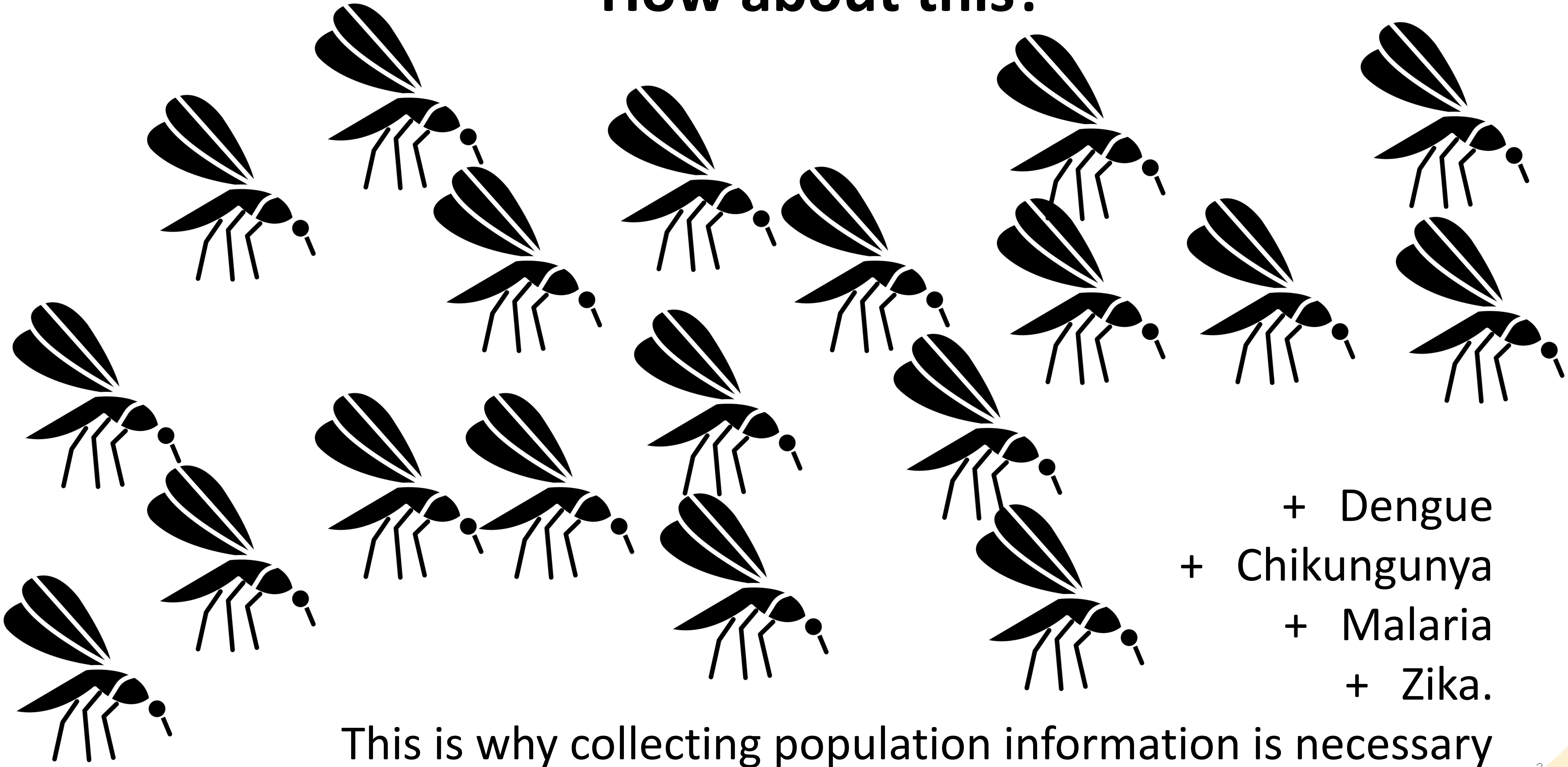
Wisdom of the Land

How do you handle this puny mosquito?



Mahidol University

How about this?



+ Dengue
+ Chikungunya
+ Malaria
+ Zika.

This is why collecting population information is necessary

Existing Methods for Gathering Mosquito Population Information

Manual counting

- Labor-intensive and expensive
- Requires expert for counting

Existing Methods for Gathering Mosquito Population Information

Manual counting

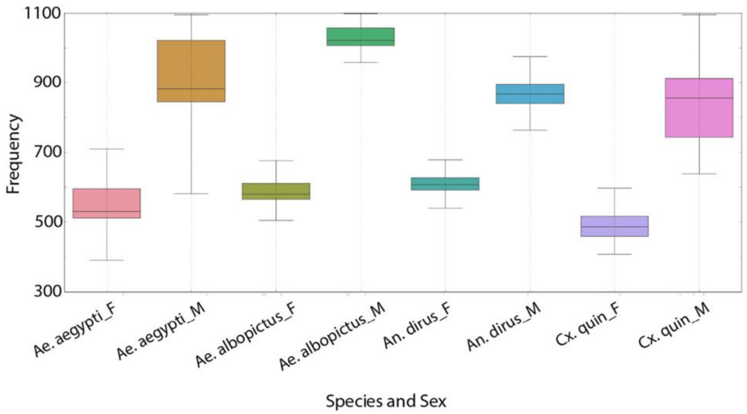
- Labor-intensive and expensive
- Requires expert for counting

Machine Learning Models

- ***Does not need the expert onsite***
- ***Based on acoustic data*** or based on optical data[5]

Commonly used features in ML:

- Fundamental frequency [7]
- Feature extraction from frequency (e.g. MFCC) [8]
- Spectrograms [9]



Issues in Existing Mosquito Detection/Classification Models

Model performance significantly dropped under noisy environment [12]

Mahidol University

Problem statement

1. **Separation of detector and classifier** leads to the loss of temporal information in wingbeat sounds and over-computation in the classifier.

1. When mosquito sound is **interrupted by other noises**, the current model's performance drops significantly.

Objective

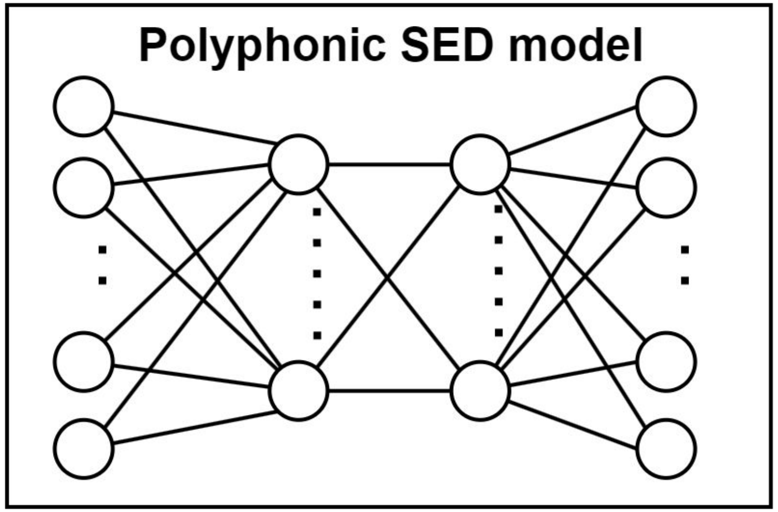
To develop **a machine learning model** that can **identify periods of mosquito presence** and classify its species and sex from wingbeat sounds, along with being **robust to noisy environments**.

Multi-label classification model

Input

x

Is a sound wave



Output

$\hat{y}$ + BCE Loss

Is a prediction probability of each event at each time frame

$$x \in [-1, 1]^{sr \times d}$$

Where d is duration in seconds of the sound wave and sr is sampling rate
sr = 8khz, d = 10 seconds

$$\hat{y} \in [0, 1]^{n_{seg} \times n_{class}}$$

Where n_seg is number of segment and n_class is number of classes
n_seg = 10 segments, n_class = 6

Baseline - CRNN Sound Even Detection (SEDNet)

SEDNet [13] is a polyphonic sound event detection model

- Receives input as log mel-band energy (mel-band spectrogram)
- Log mel-band energy does not design to convey some fine, high-resolution features that can differentiate classes.

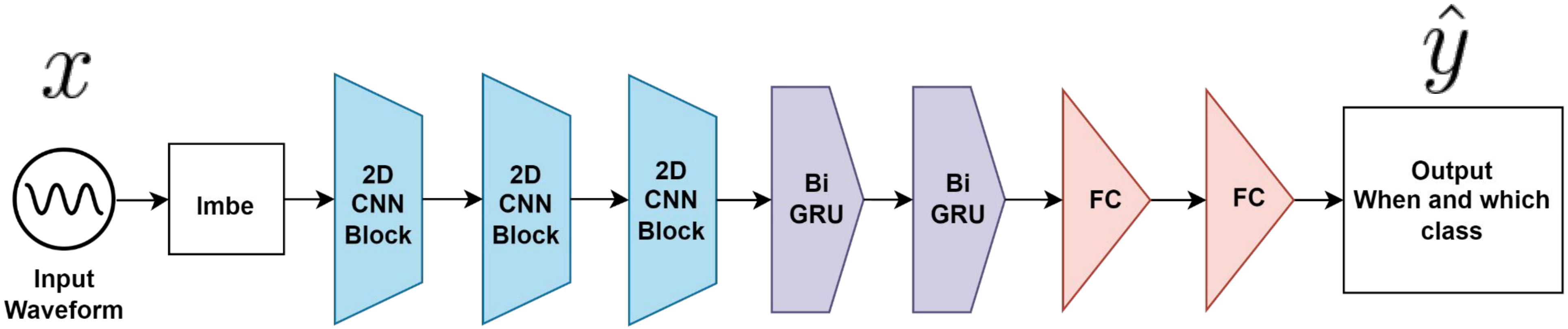


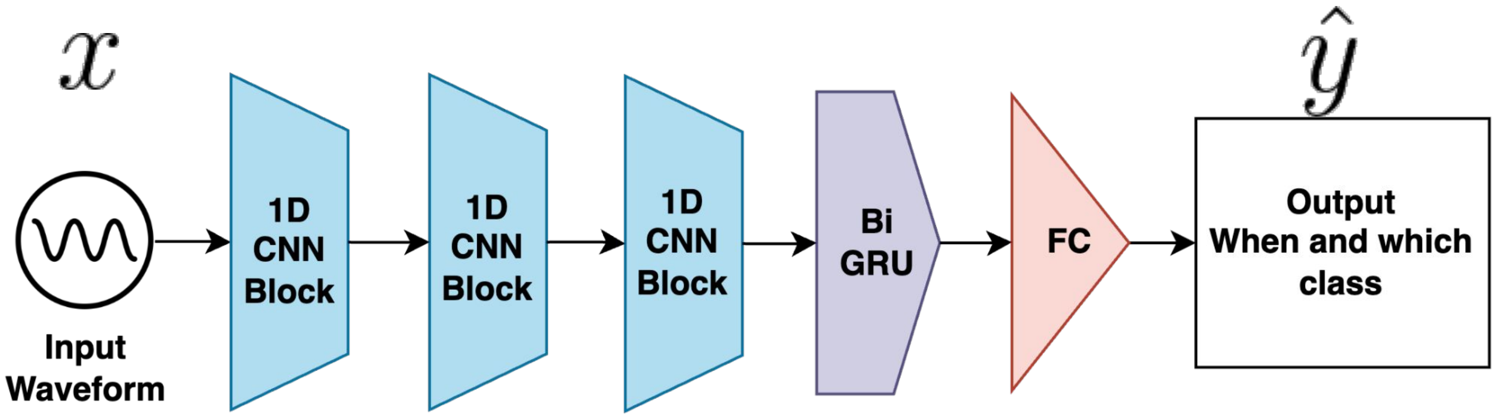
Diagram of SEDNet model architecture

Our Proposed Model (1DCRNN)

- A polyphonic sound event detection model that **processes raw audio signal**
- Inspired by SEDNet with major difference in **dismissing spectrogram**
- Uses **recurrent** layer to handle sequential time-series data

Advantages:

- Solve spectrogram issue of overlooking some discriminators
- Use less computational time



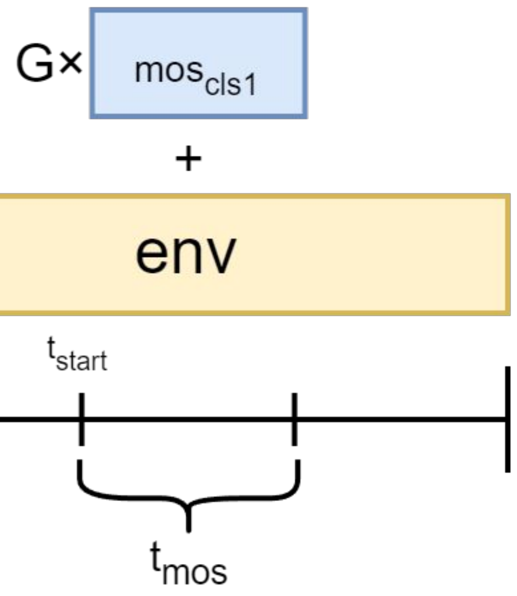
1DCRNN model architecture

Noise Simulation

Process of Overlaying mosquito sounds onto background noises:

1. Each chuck of generated sound consists of:
   a. *10 seconds of background*
   b. *1-2 randomly overlayed mosquito sounds,* each no more than 2 seconds
2. *Apply a gain factor* G to vary the volume (amplitude)

of the mosquito sounds



Mahidol University

Data Preparation

Mosquito:

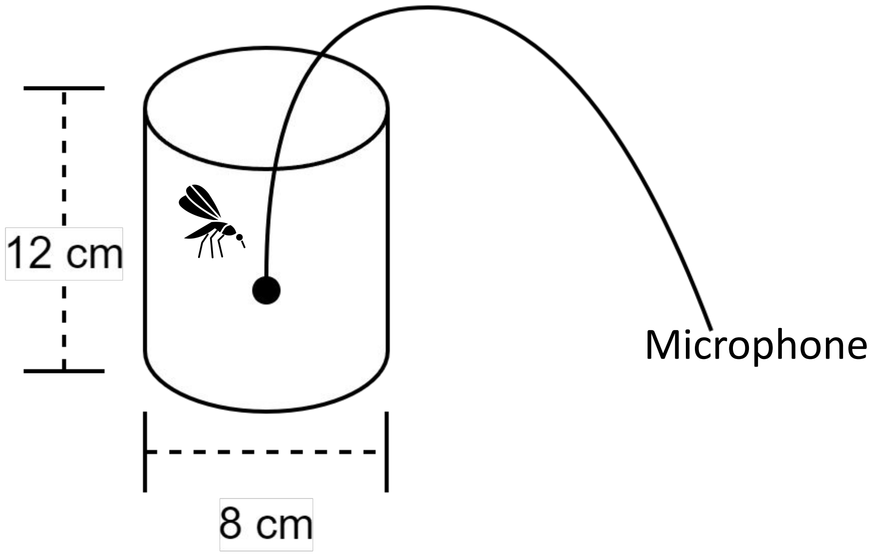
- 2 Sets of recordings retrieved from MIRU, Home-recorded and Studio-recorded.
- 5 species, each 2 sexes
- **Recording Process:**
    - put the mosquito in the small cylindrical container with a microphone
    - recorded for a few minutes then cut only the mosquito presence into small (0.5 - 2 seconds) cut files

Mosquito Species	Male(files)	Female(files)
An. Dirus	178	152
Cx. Quin	113	120
Ae. Albopictus	218	128
Ae. Aegypti	152	166
An. Minimus	50	37

Number of wingbeat cut files of home-recorded

Mosquito Species	Male(files)	Female(files)
An. Dirus	220	286
Cx. Quin	370	385
Ae. albopictus	271	280
Ae. Aegypti	300	303
An. Minimus	30	33

Number of wingbeat cut files of studio-recorded



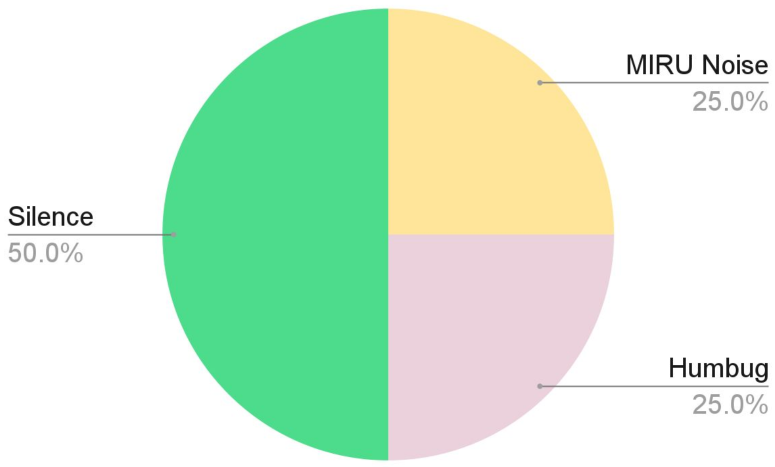
Process of recording mosquito sound

Dataset

Training: 40,000 seconds

Validating/Testing: 20,000 seconds

- Environmental sound
  - **Type 1 MIRU**: 25%
  - **Type 2 HumbugDB**: 25%
  - **Type 3 Silence**: 50%
- Mosquito wingbeat sound
  - From MIRU 2 sets of recordings.



The dataset only contain **the mosquito species that coexist** in the real geographical location

- **Habitat A**: Anopheles and Culex genus (6 classes)
- **Habitat B**: Aedes and Culex genus (6 classes)

Final Result: F1 per class

Model	Habitat A Results					
	An.Mini.M	An.Diru.M	Cx.Quin.M	An.Mini.F	An.Diru.F	Cx.Quin.F
SEDNet (baseline)	0.000	0.591	0.458	0.070	0.520	0.772
1DCRNN	0.000	0.423	0.465	0.353	0.566	0.671

Model	Habitat B Results					
	Ae.Aeg.M	Ae.Albo.M	Cx.Quin.M	Ae.Aeg.F	Ae.Albo.F	Cx.Quin.F
SEDNet (baseline)	0.181	0.360	0.453	0.572	0.514	0.756
1DCRNN	0.391	0.449	0.532	0.591	0.759	0.773

- Result is competitive in habitat A, 1DCRNN is better in habitat B
    - 1DCRNN are able to learn An.Minimus.F, while SEDNet can’t

Final Result: F1 for detection

Model	Habitat A Results		
	F1	P	R
SEDNet (baseline)	0.860	0.878	0.842
1DCRNN	0.877	0.940	0.822

Model	Habitat B Results		
	F1	P	R
SEDNet (baseline)	0.876	0.847	0.908
1DCRNN	0.936	0.967	0.906

1DCRNN is much better at detection than SEDNet
SEDNet has higher recall, but is only slightly better

Mahidol University

Thank you
Q & A