Convolutional Recurrent Neural Networks for the Classification of Cetacean Bioacoustic Patterns

Dimitris Makropoulos[1,2], Antigoni Tsiami[1], Aristides Prospathopoulos[2], Dimitris Kassis[2], Alexandros Frantzis[3], Emmanuel Skarsoulis[4], George Piperakis[4] and Petros Maragos[1]

1. National Technical University of Athens; Greece 2. Hellenic Centre for Marine Research (HCMR) , Greece ; 3. Foundation for Research and Technology-Hellas , Greece ; 4. Pelagos Cetacean Research Institute, Greece

Introduction

❑ **Objectives :**

- Implementation of DL techniques to categorize biosignals generated by two cetacean species:
- Sperm whales *(Physeter macrocephalus)*
- Striped dolphins *(Stenella coeruleoalba)*

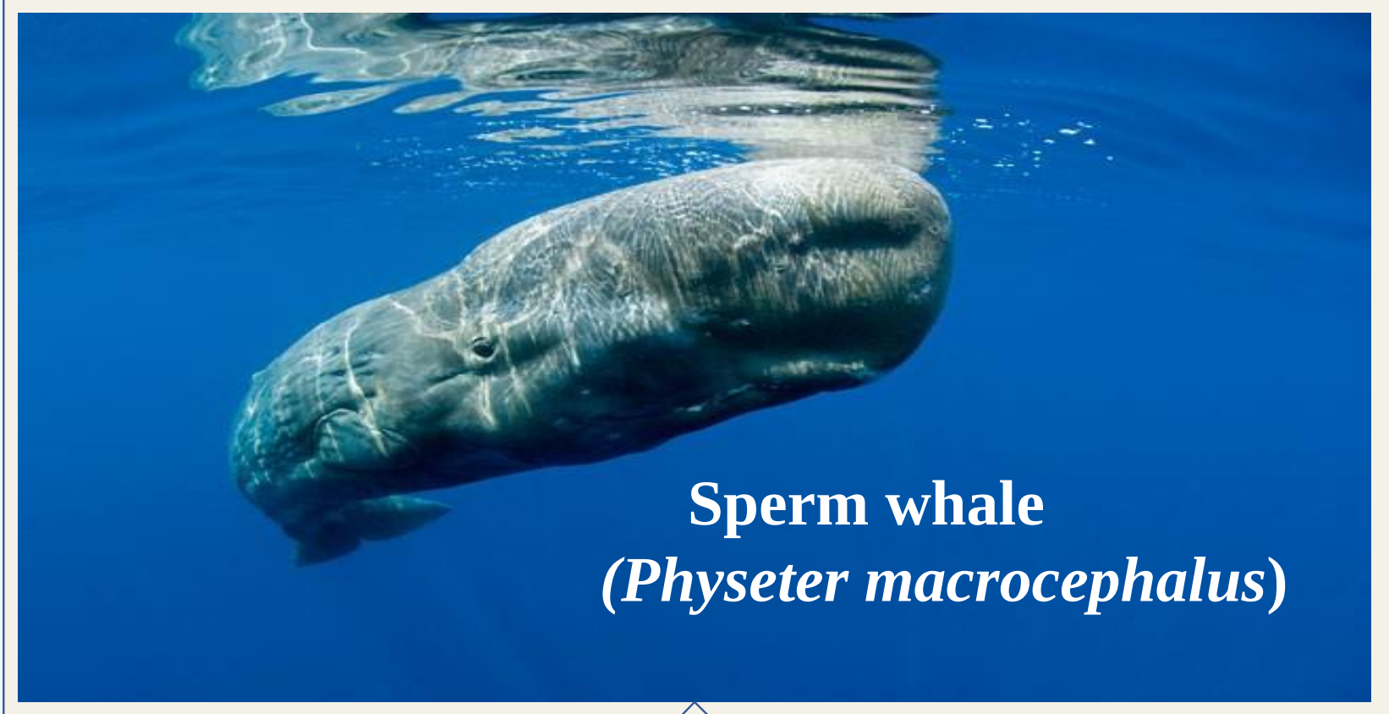
❑ **Motivation :**

- Build a recognition tool for protection of endangered species.

Related Work

- Main idea: Convert biosignals into time-frequency representations generating an image dataset.
- Two main alternatives to spectrogram representations: Use either raw waveform as input or apply traditional ML techniques.

Analysis of patterns on time and frequency domain



Clicks:
Centroid frequency:
15 kHz
Duration of 20 ms-30 ms

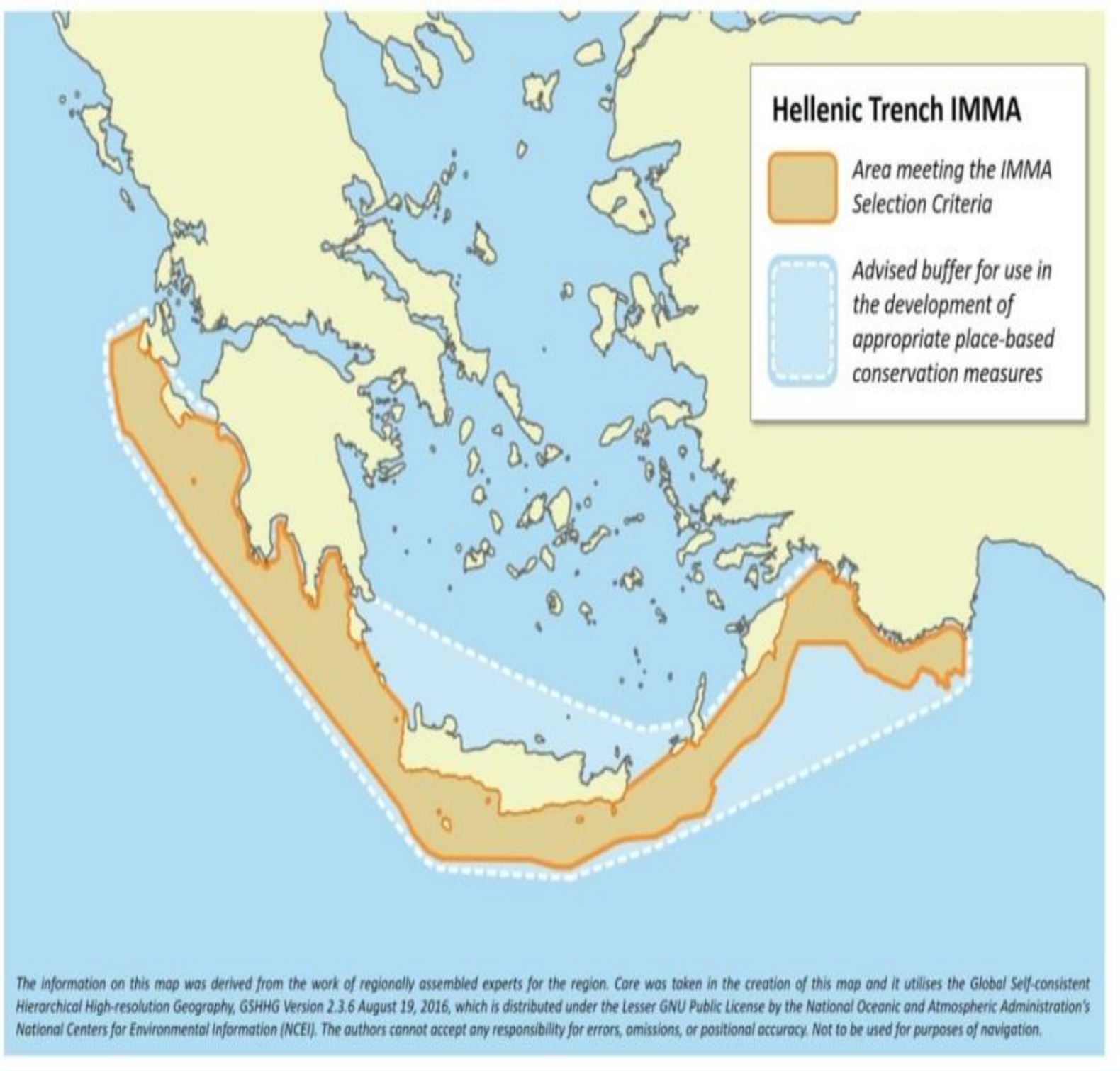
Codas:
Centroid frequency:
5 kHz

Clicks:
Frequencies
over 100kHz

Whistles:
Frequencies from 2 to 30 kHz
Duration of 0.5 ms-4 s

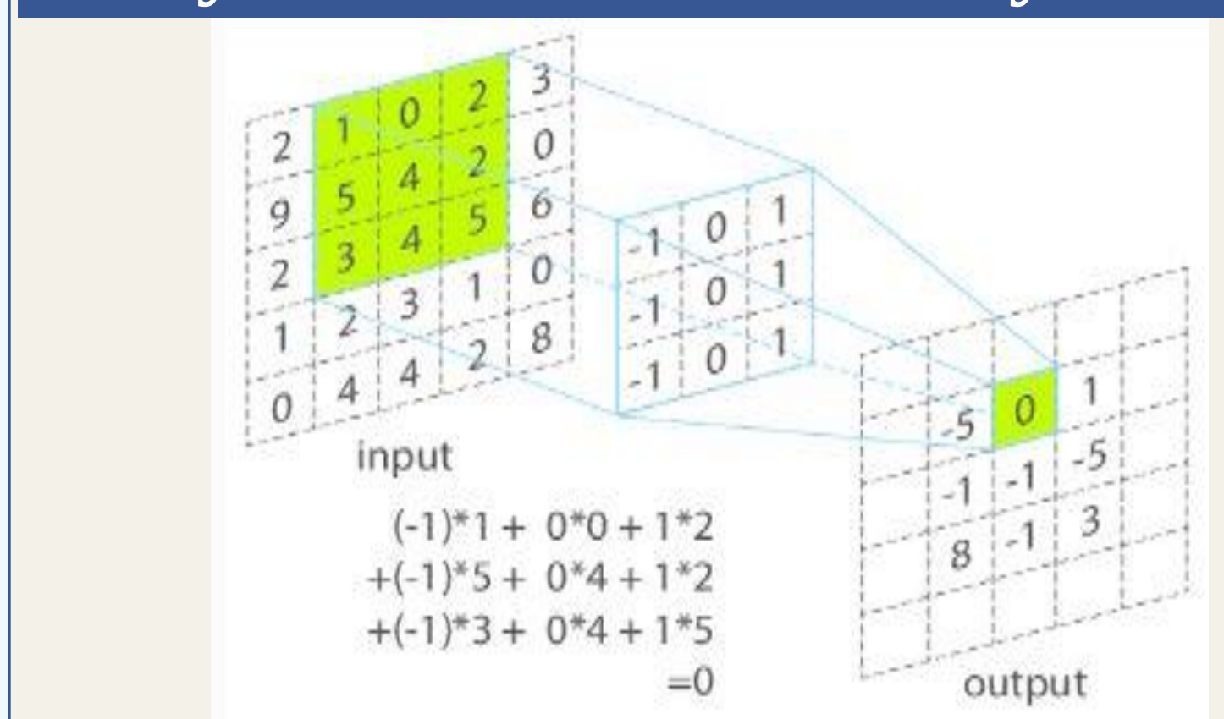
Burst Pulse Clicks

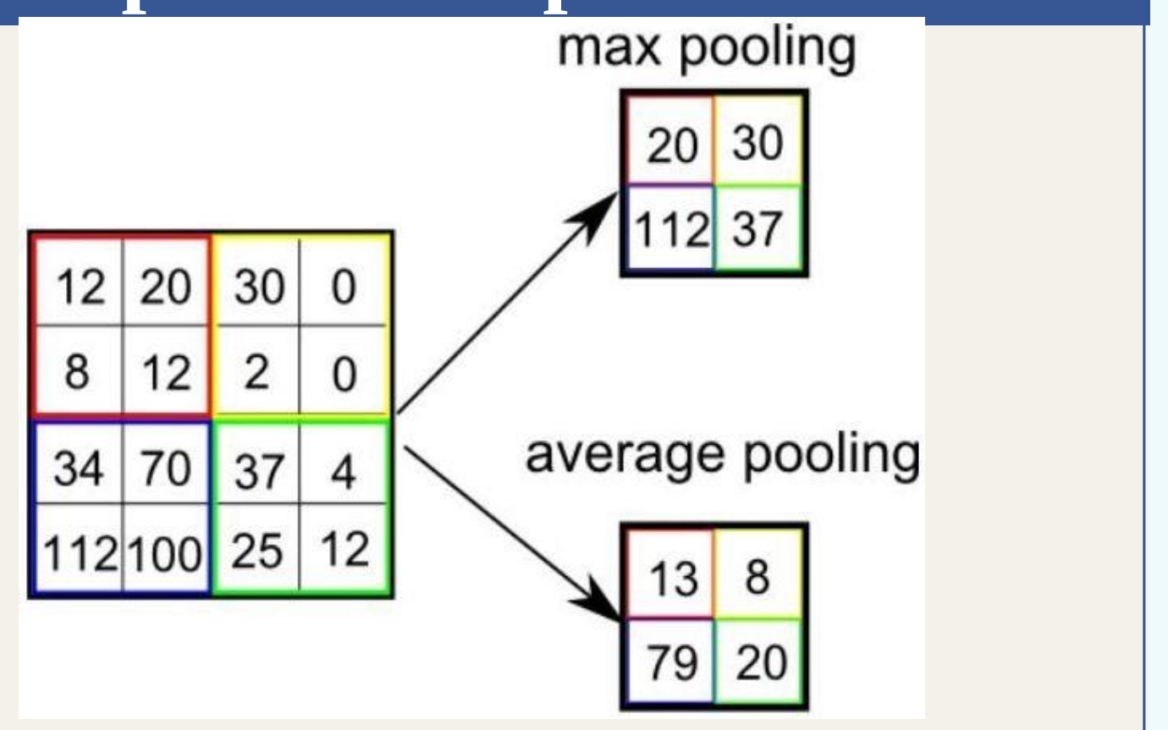
Origin of data: Hellenic Trench


Hellenic Trench IMMA
Area meeting the IMMA Selection Criteria
Advised buffer for use in the development of appropriate place-based conservation measures
The information on this map was derived from the work of regionally assembled experts for the region. Care was taken in the creation of this map and it utilises the Global Self-consistent Hierarchical High-resolution Geography, GSHHG Version 2.3.6 August 19, 2016, which is distributed under the Lesser GNU Public License by the National Oceanic and Atmospheric Administration's National Centers for Environmental Information (NCEI). The authors cannot accept any responsibility for errors, omissions, or positional accuracy. Not to be used for purposes of navigation.

1. A passive acoustic listener (PAL) was deployed at Pylos at 500 m depth from September 2008 to November 2009.
2. Another PAL was deployed in the Bay of Sougia at a depth of 100 m in summer 2020 and 2021.
3. Data were collected using a towed array during cetacean surveys along the Hellenic Trench.

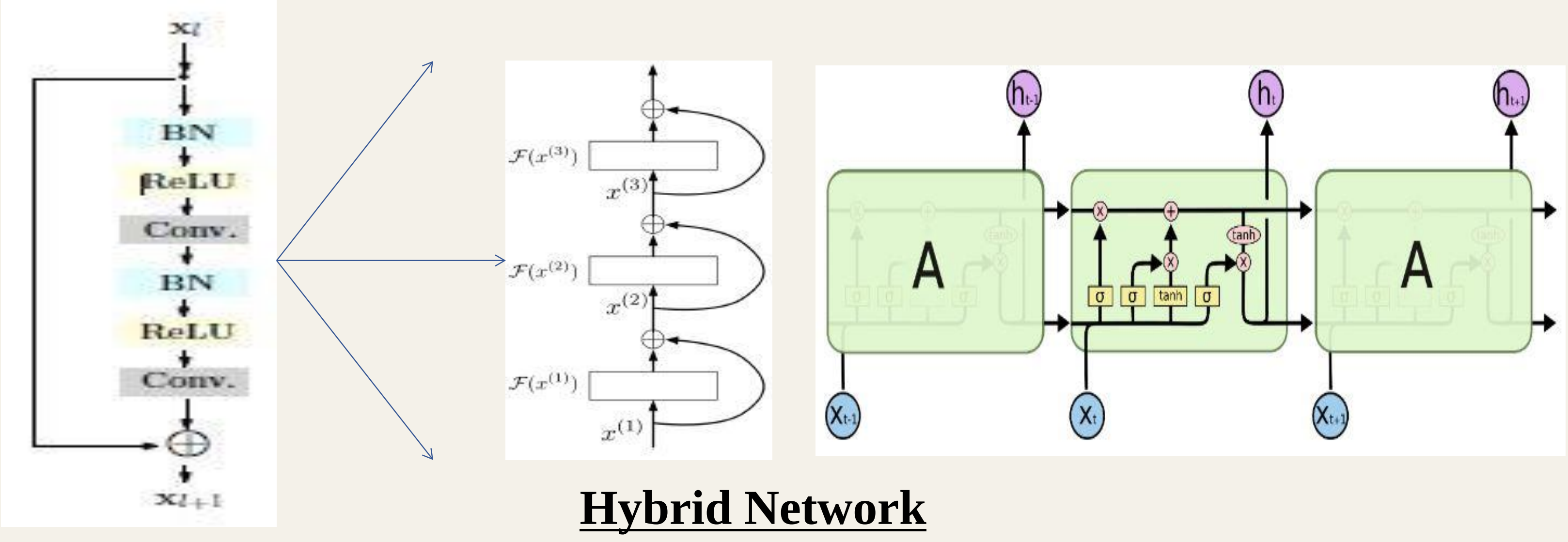
Key idea: Construct a hybrid extractor of spectro-temporal features





- Convolutional and pooling Layers: Feature extraction from 2D time-frequency representations.
- Investigate temporal dynamics utilizing RNN variants (LSTMs-GRUs).

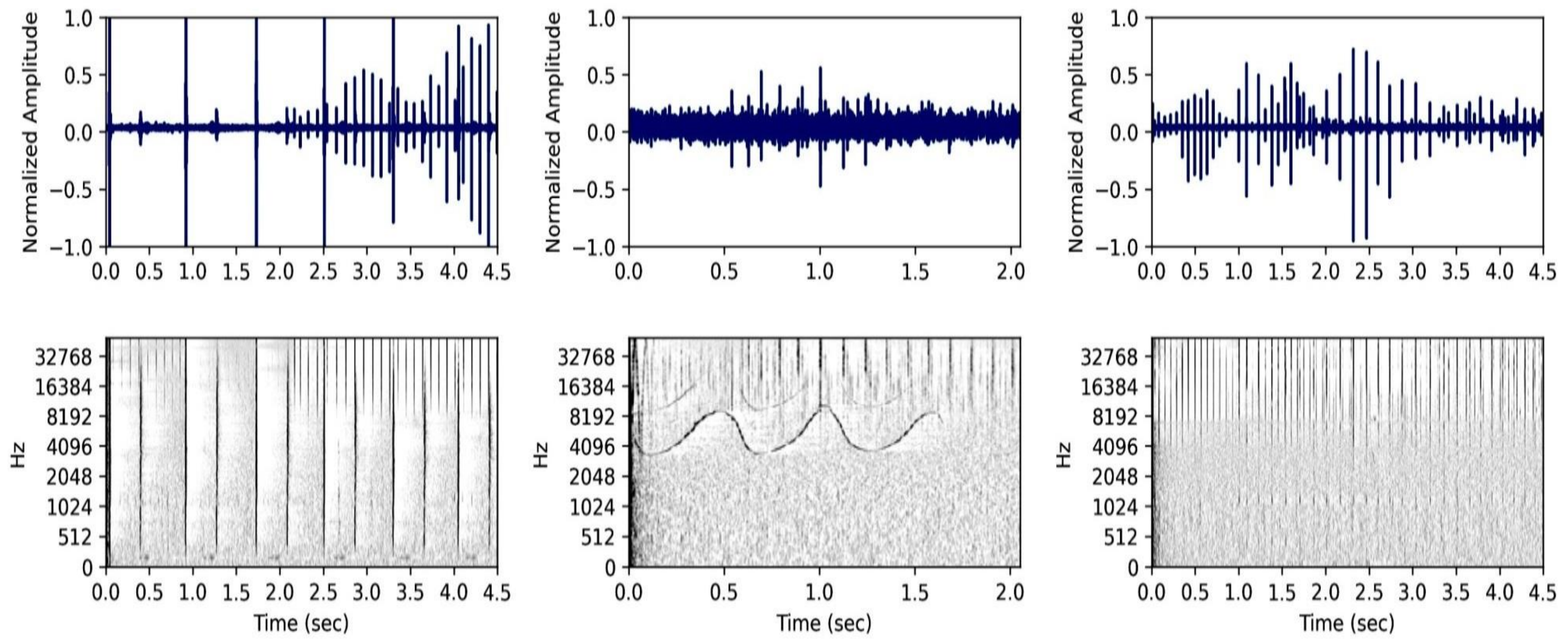
Components'analysis of a hybrid deep network



Hybrid Network

Include both a convolutional and a sequential component

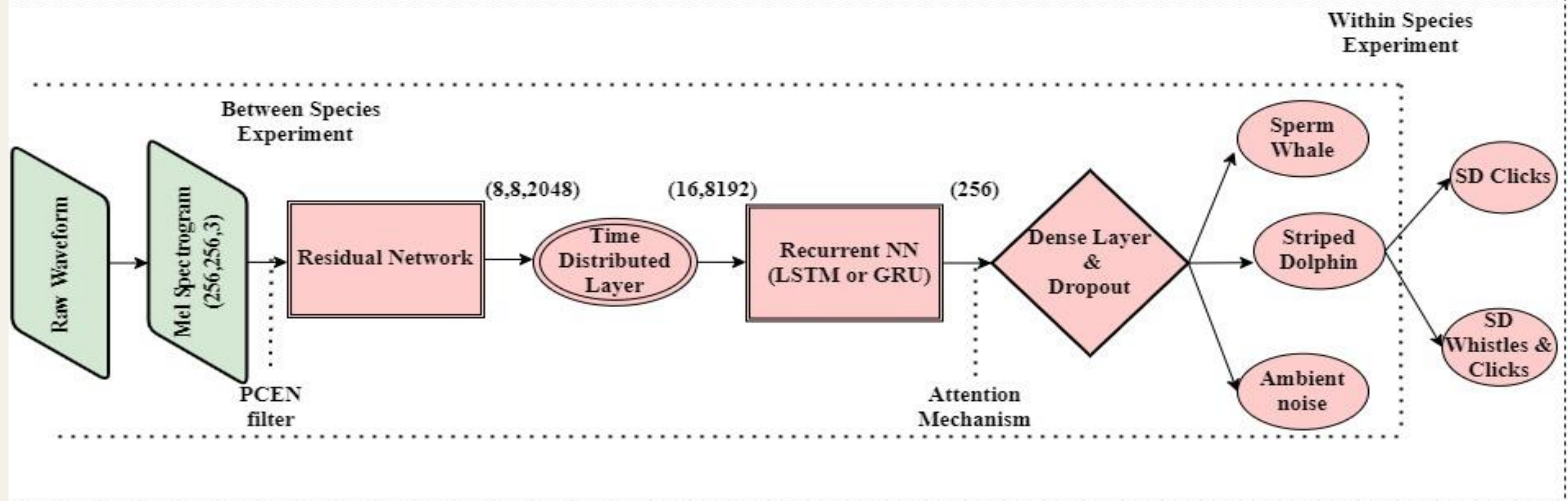
Signal processing



Raw waveforms (up) and spectrograms (down) of sperm whale clicks (left) vs striped dolphin clicks and whistles (center) and striped dolphin clicks (right)

- We apply a high-pass Butterworth filter with a low-frequency cutoff at 1 kHz.
- Mel Spectrograms: Compute the discrete Fourier transforms (DFT) over every windowed signal. Square modulus of DFT.
- Per Channel Energy Normalization (PCEN) to suppress stationary, narrowband electronic noise and enhance contrast between background and foreground transient events.

Proposed CRNN Architecture



Design of experiments & results

1. Between species experiments

•291 recordings of sperm whale calls

•284 striped dolphin calls.

•90 files of ambient noise.

2. Within species experiments

•284 dolphin calls are divided into two classes of 135 clicks and 149 whistles and clicks.

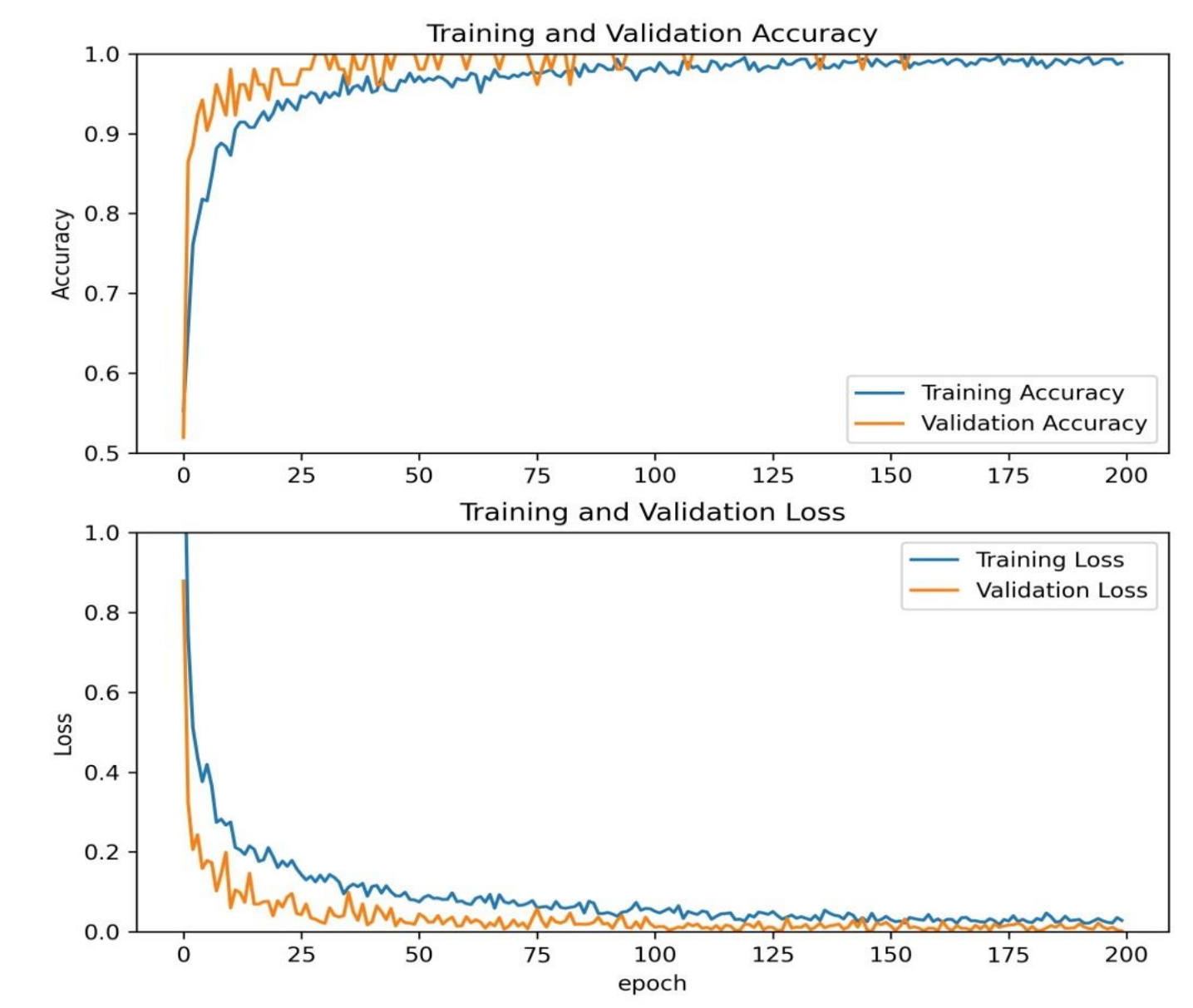
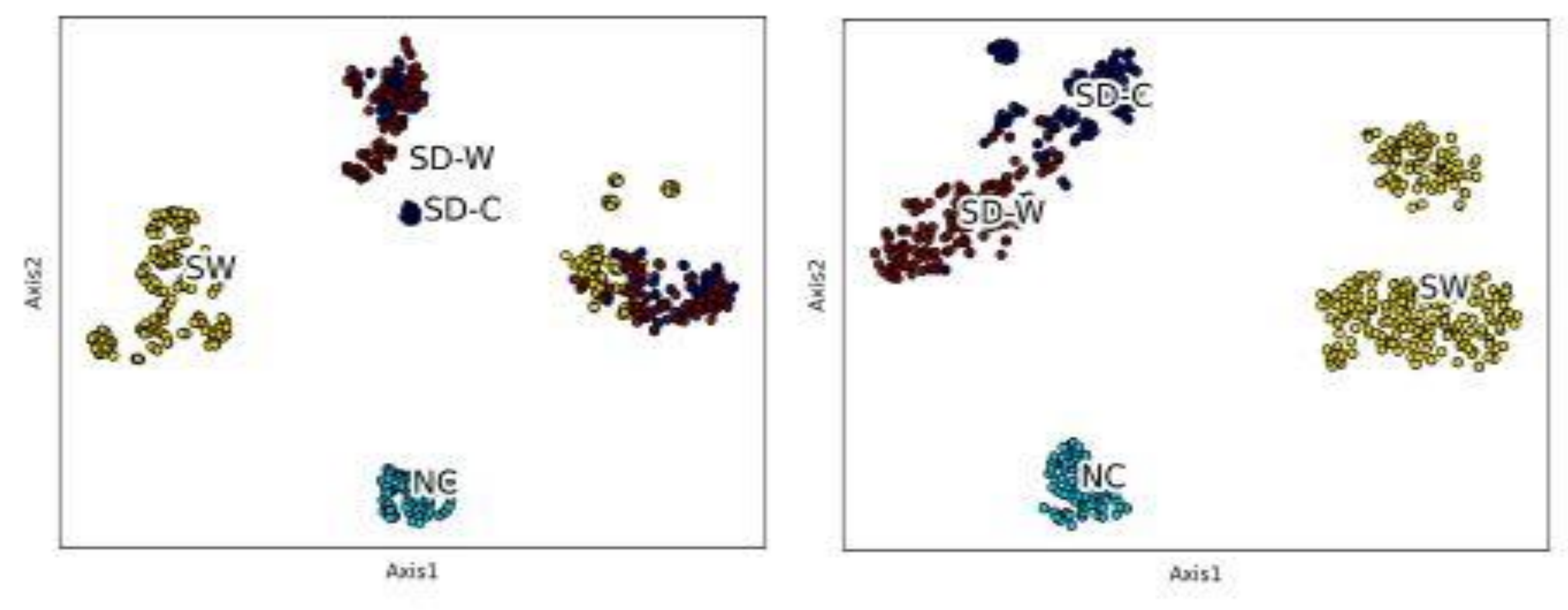


Table 1. Performance of different NN architectures

Models	Results on a test set (Mean values)		
	Parameters	*Accuracy*	*Precision*
MFCC-SVM (RBF)	-	83.0%	73.4%
MFCC-kNN	-	75.45%	73.4%
ResNet	1.0M	87.0%	84.7%
ResNet-LSTM	9.77M	91.3%	89.9%
ResNet-BiLSTM	18.5M	90.1%	89.1%
ResNet-GRU	7.6M	90.9%	89.8%
ResNet-BiGRU	14.2M	88.7%	88.0%
ResNet-LSTM-Attention	9.8M	90.4%	89.9%
Parallel ResNet-LSTM	8.2M	89.2%	88.7%

Visualization of features in low dimensional space



(a) Cepstral features (b) CRNN features

SW: Sperm whale clicks, NC: No Clicks,
SD-C, SD-W: Striped dolphins Clicks & Whistles.

Conclusions

(a) baseline DL models outperform traditional ML methods;
(b) hybrid networks achieve higher accuracies than baseline ResNets;
(c) bidirectional networks do not increase performance;
(d) all architectures have succeeded to solve a - between species- classification problem while hybrid architectures have demonstrated advantages on differentiating intraclass overlapping patterns.