Sounds of Nature

Semester Project Presentation

09/03/2023

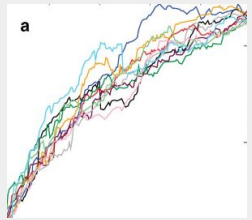
Amar Meddahi, Edgar Remy, Fabio Pereira de Araujo,
Linda Hammami, Sarah Chougar and Younes Boutiyarzist

Uncovering the Secrets of Nature

Ecologist experts record the bioacoustic activity for months

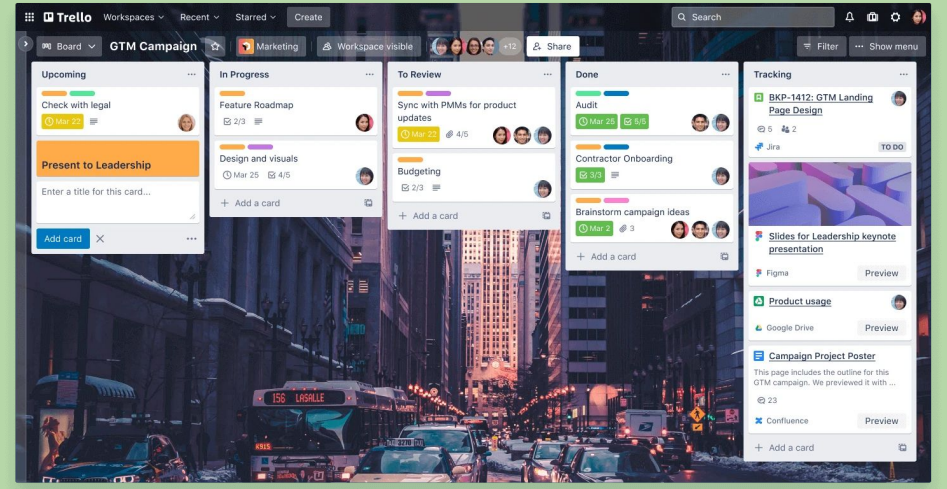
Recognition of species and groups of species

Soundscapes Recognition and Environmental Analysis



Synergizing Efforts: Structuring Effective Teamwork

Remote and asynchronous communication



Agile Working Framework (1 week sprint)

Taking Control of Your Day with Self-Management

Outline

- Recognition of species and groups of species
- Soundscapes Recognition and Biodiversity Analysis

Recognition of species and groups of species

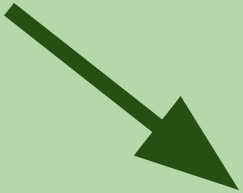
General Context

- In this project, artificial intelligence is used to solve an ecological problem using a system that recognizes species and groups of species in nature, especially the vocalizations of animals.

Objective

- The main goal is to develop a classifier capable of recognising animal species by their sounds:

Based on the animal's vocalizations, an AI algorithm can determine the species.



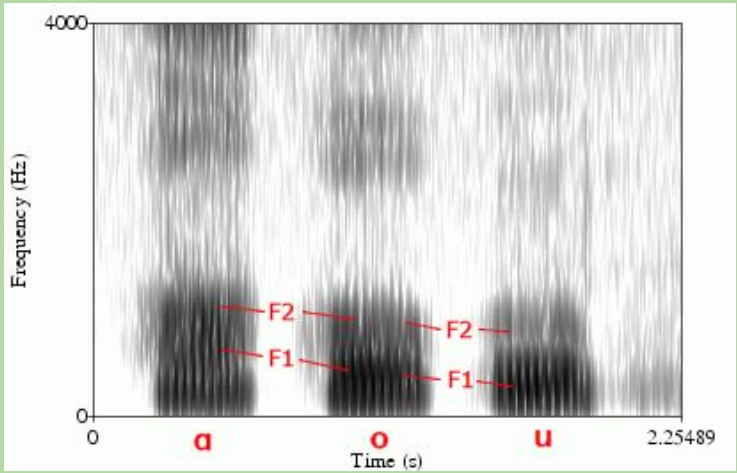
Conservation

Ecosystem Health

Datasets

Audio files

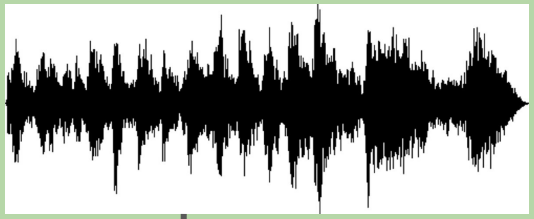
Spectrograms



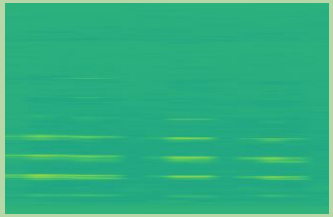
Preprocessing

Audio dataset

- Audio files < 5 seconds are repeated several times to reach 5 seconds
- Resampling (22,000 Hz)



Audio converted into spectrograms



Linear spectrogram

Mel spectrogram

Data Augmentation

First Method

Solving the problem of unbalanced classes :

Adding the noise extracted from part 1 of the project to the samples of underrepresented classes

= > This method is applied before starting the training of the models and directly to the audio files.

Second Method

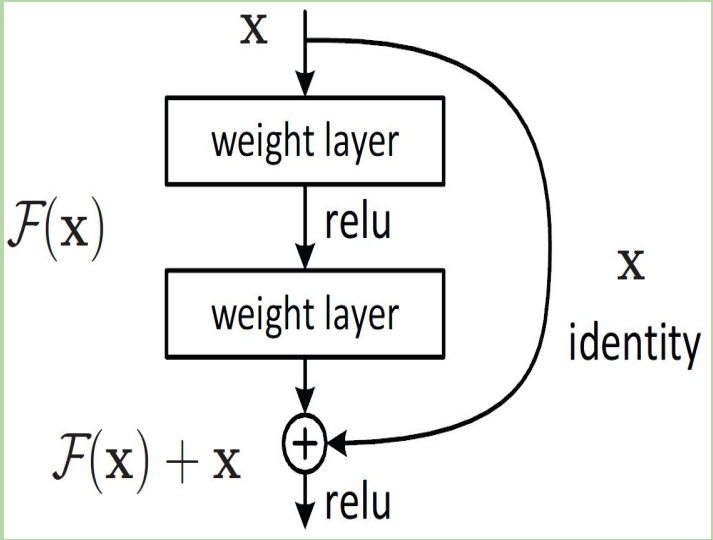
Adding new samples to all the classes equally :

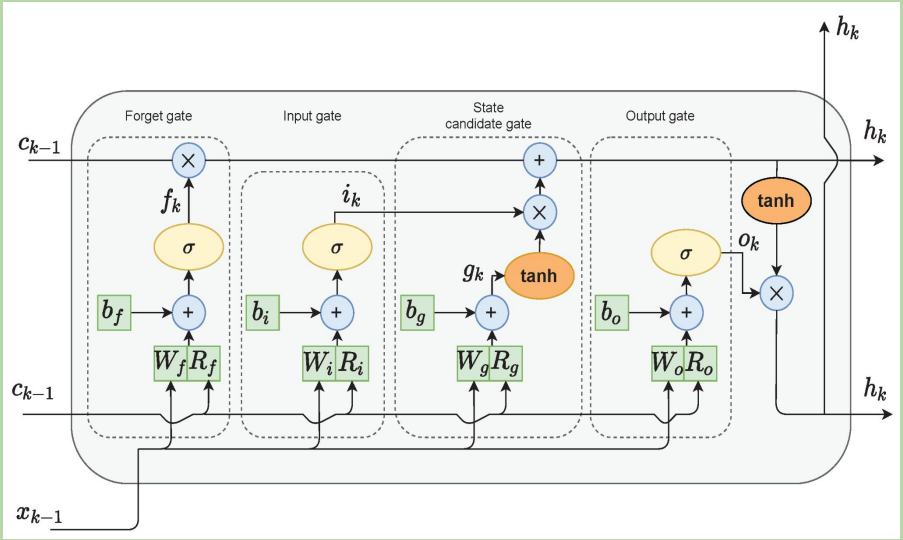
Using frequency shift augmentations on the spectrograms

= > This method is used during the training of the models and on the spectrograms.

R-CNN

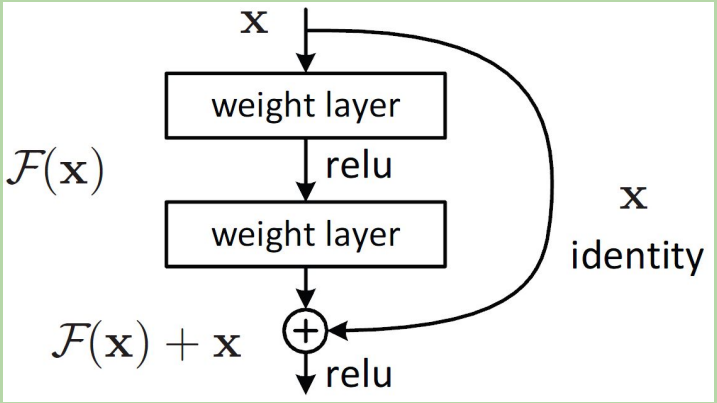
- **Resnet** : A block of Convolutions
- **LSTM** : Memory cells

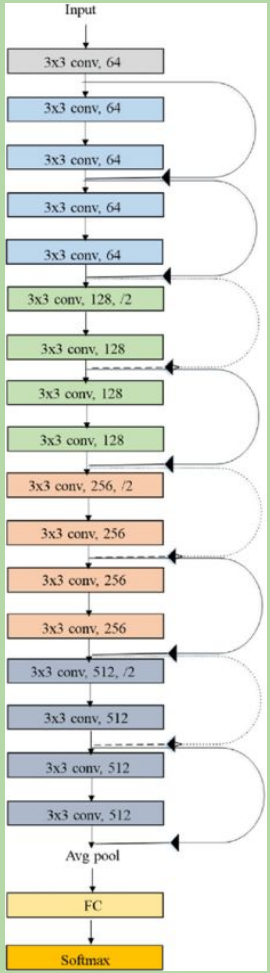




ResNet18

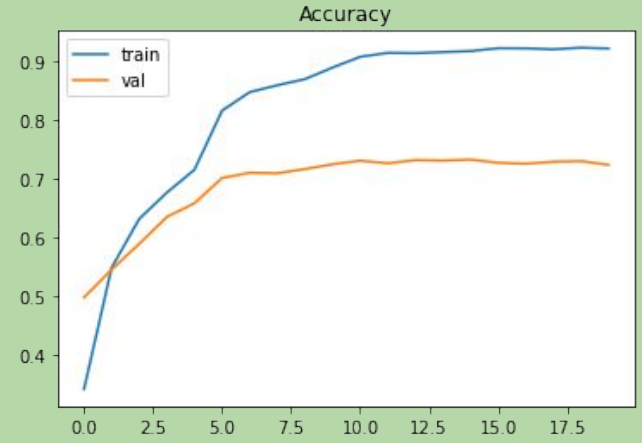
- **Number of parameters** :
  11,210,883

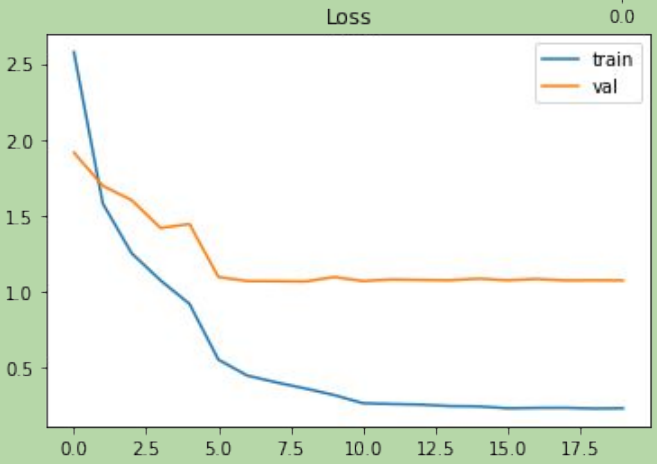




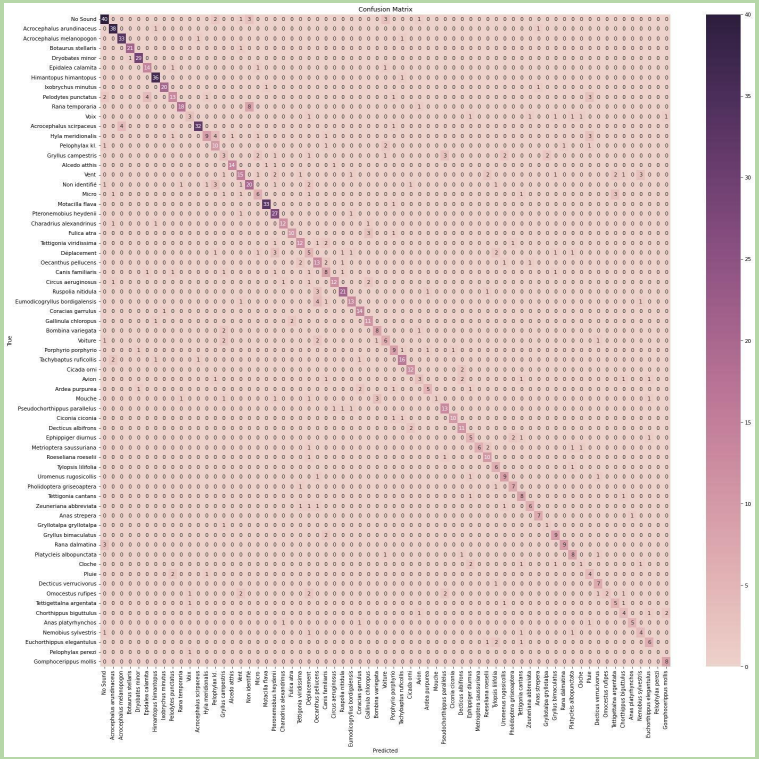
The best Model : ResNet18

	Accuracy
Train	**0.92**
Val	**0.73**
Test	**0.72**

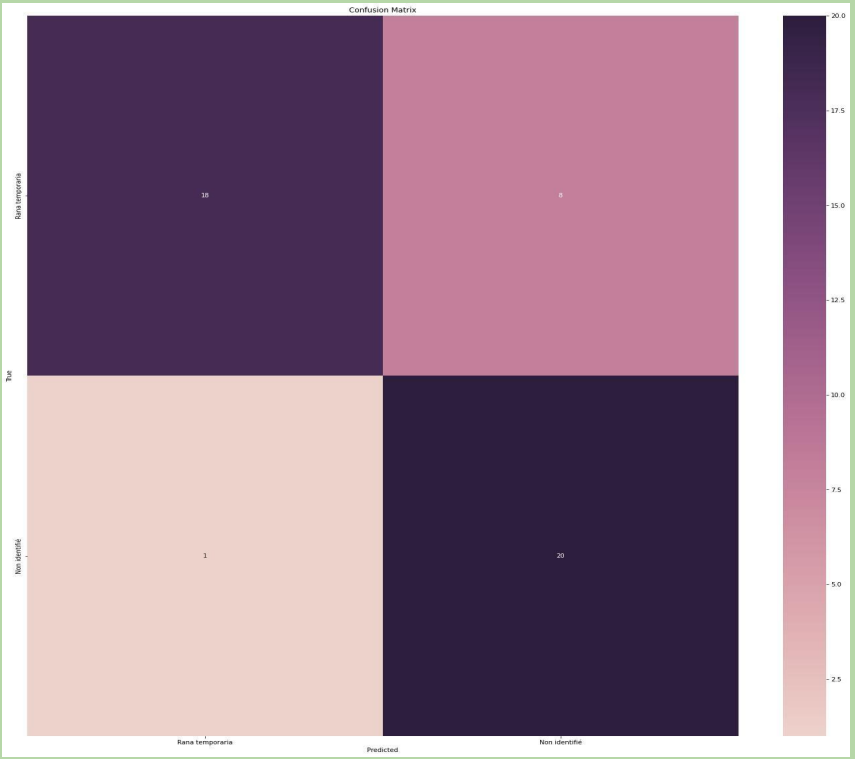




Confusion Matrix

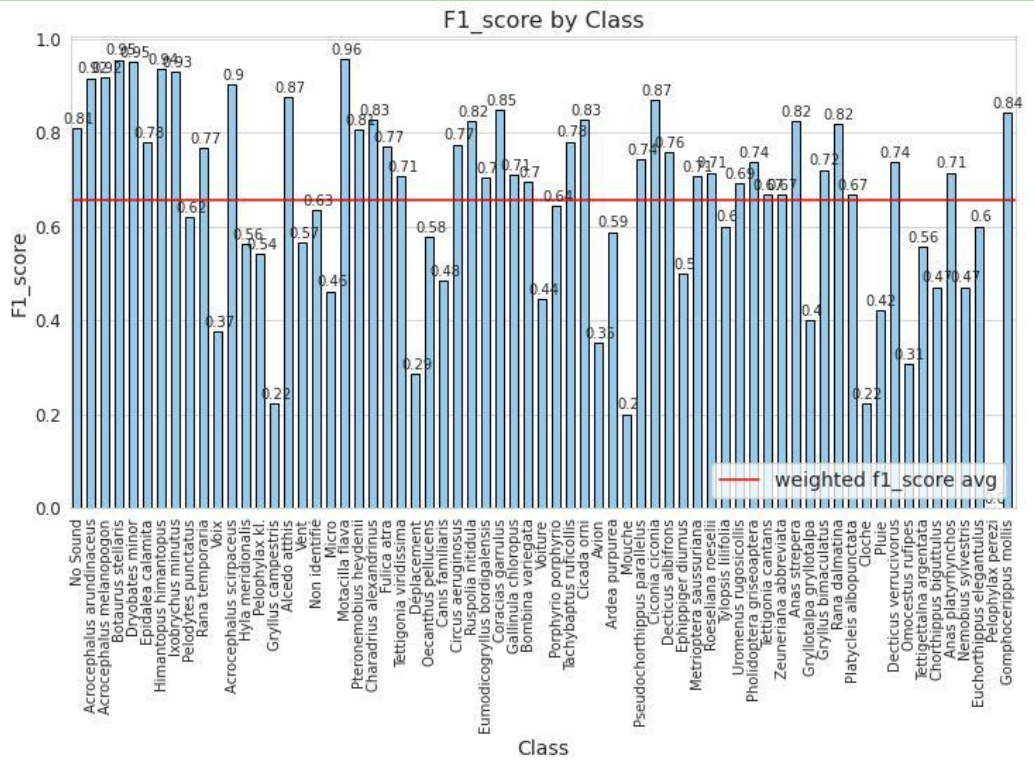


Confusion matrix



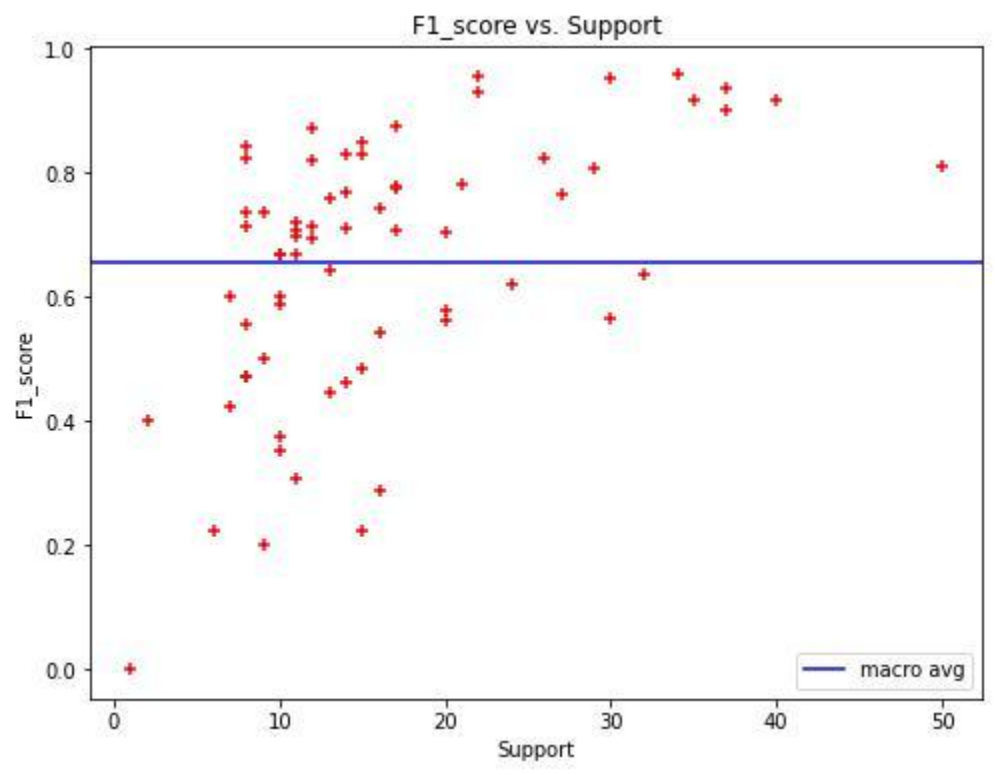
Relevant confusion matrix

F1-score by class

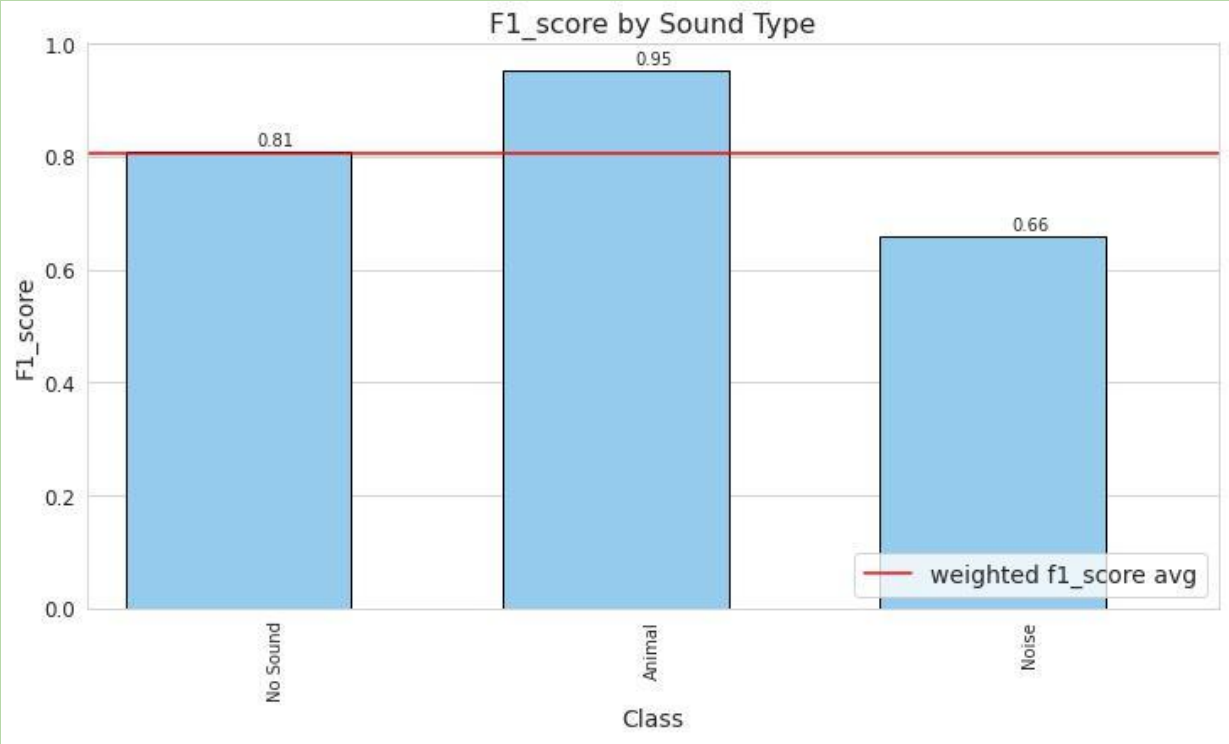


Class	F1-score (avg)
Animal	**0.69**
Noise	**0.41**

F1-score vs support



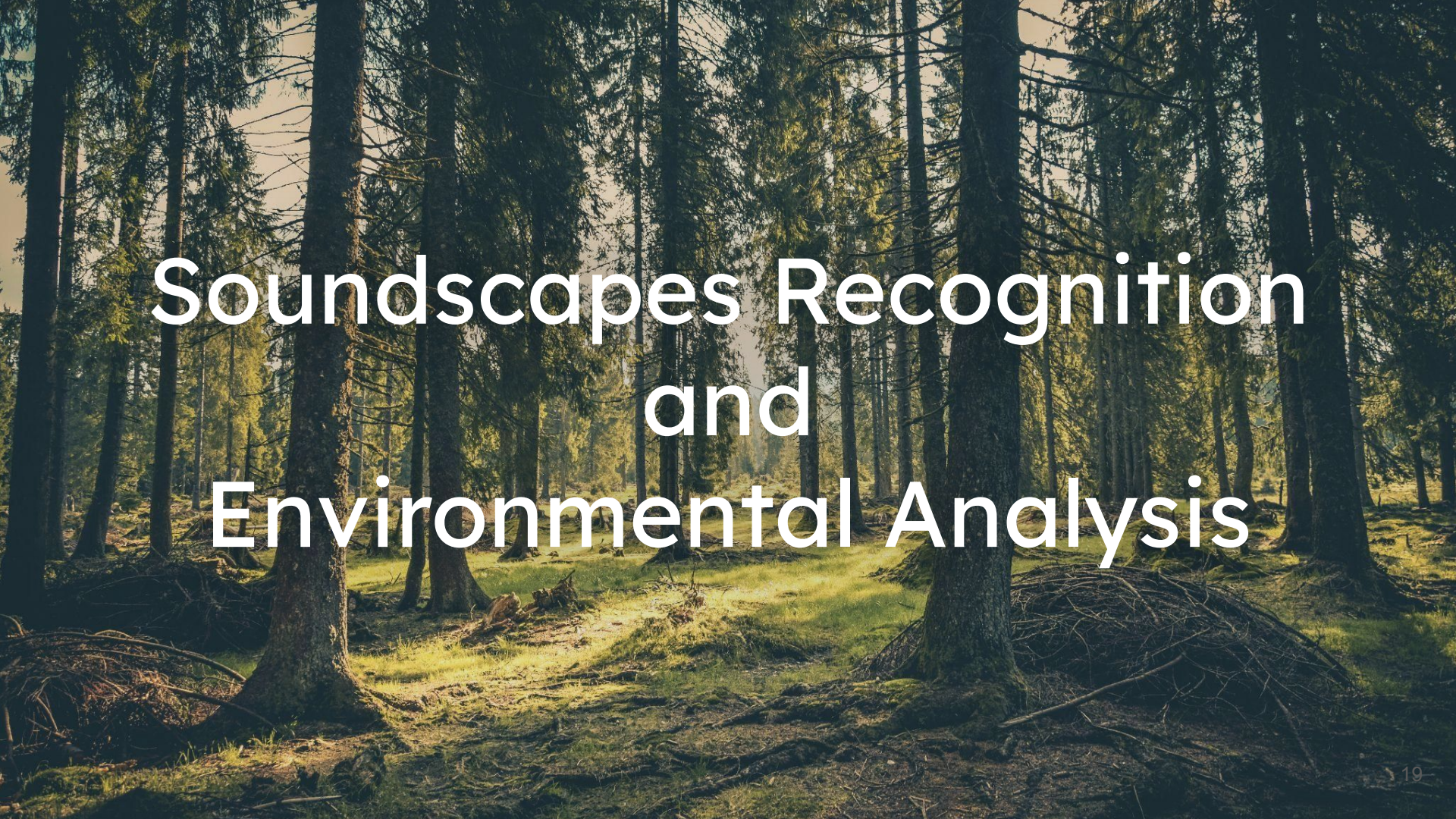
F1-score by Sound-Type



Perspectives

What are the relevant research perspectives?

- Hierarchical loss.
- Updating the softmax function to handle audios that contains multiple species.


Soundscapes Recognition
and
Environmental Analysis

Problem Statement

Given a bioacoustic activity

Task n°1: Soundscapes Recognition

Find the labels associated with a bioacoustic activity:

Tree line evolution

- Blue: forest that has descended in altitude since the 50's
- Red: forest that has increased in altitude since the 50's

Sensor position

- Estive: 200+m above the upper tree line
- Edge: At the tree line

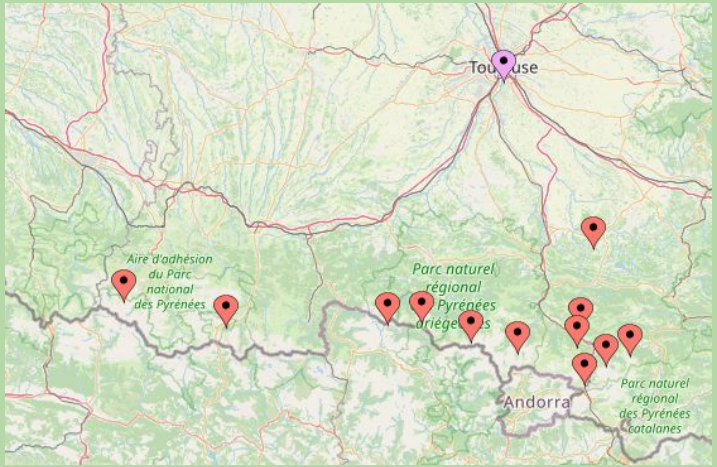
Task n°2: Environmental Analysis

Estimate ecosystem variables

Examples

- Distance to a road
- Average altitude
- Average temperature evolution

Overcoming the Challenges of Handling a Large Dataset



- Months of bioacoustic activity recordings (**1.1TB**)
- **16 different geographical** sites (red pins)

- One day of the same season chosen for each site
- Three 30-minute samples throughout the day
- **<u>New representation of the database (~10GB)</u>**

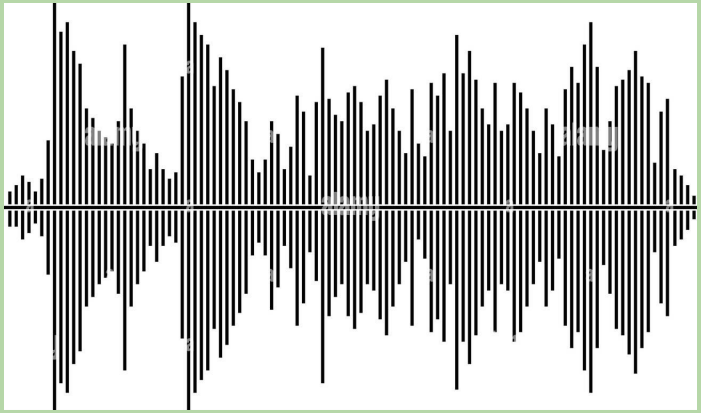
Index Approach

Examples

- **ACI →high values = indicate storms, intermittent rain drops falling from vegetation**
- **ACI → low values = recordings with consistent cicada noise**

Song Meter Mini

recording



processing (padding,
16 kHz resampling)

calculation of 60 indices for each minute

- temporal index
- frequential index

cleaning

3116 rows × 64 columns

Approach with Embedded Indexes

The OpenL3 Library

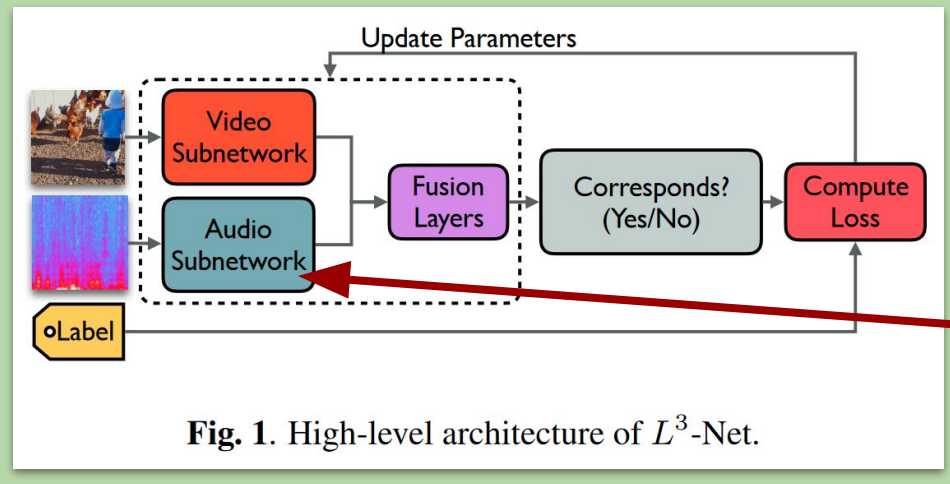
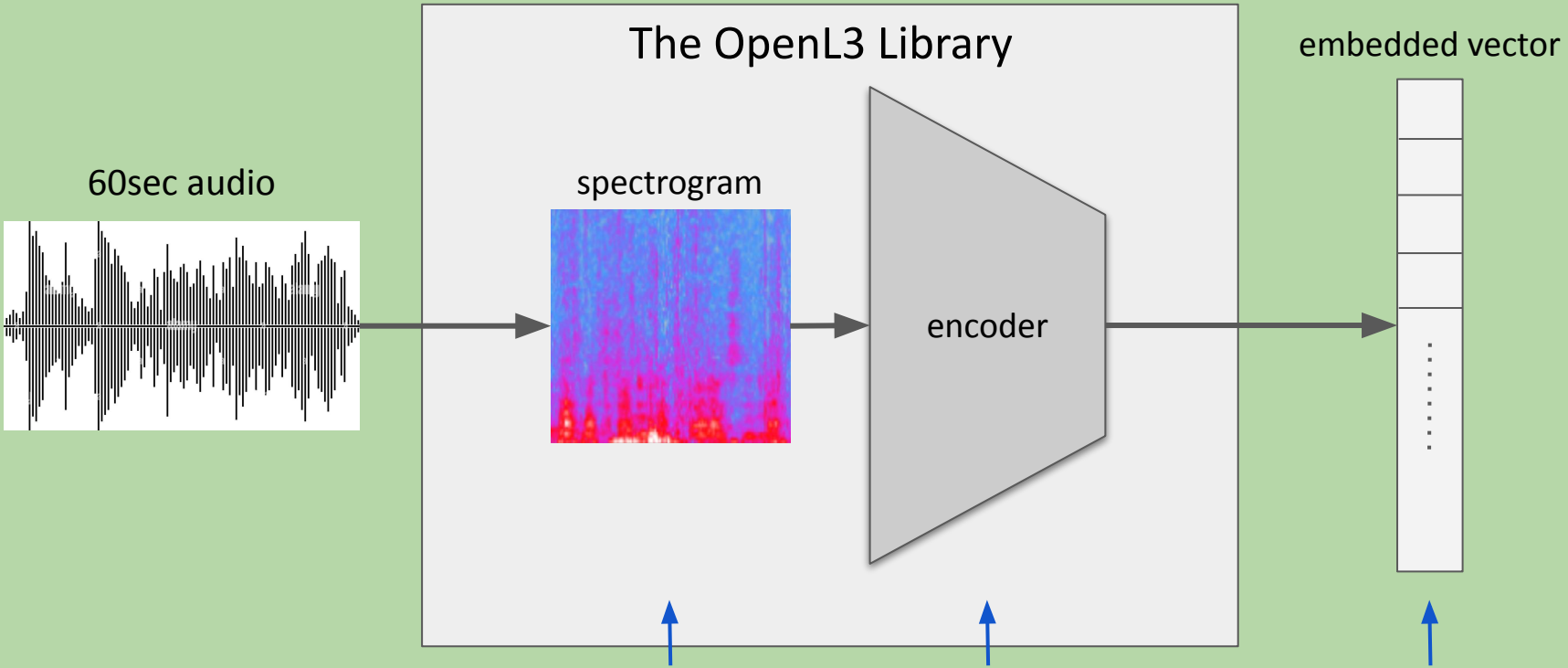


Fig. 1. High-level architecture of L^3-Net.

- **contains an encoder** that is trained
- the encoder summarizes audio (as spectrograms) **into a vector** of **smaller dimension**

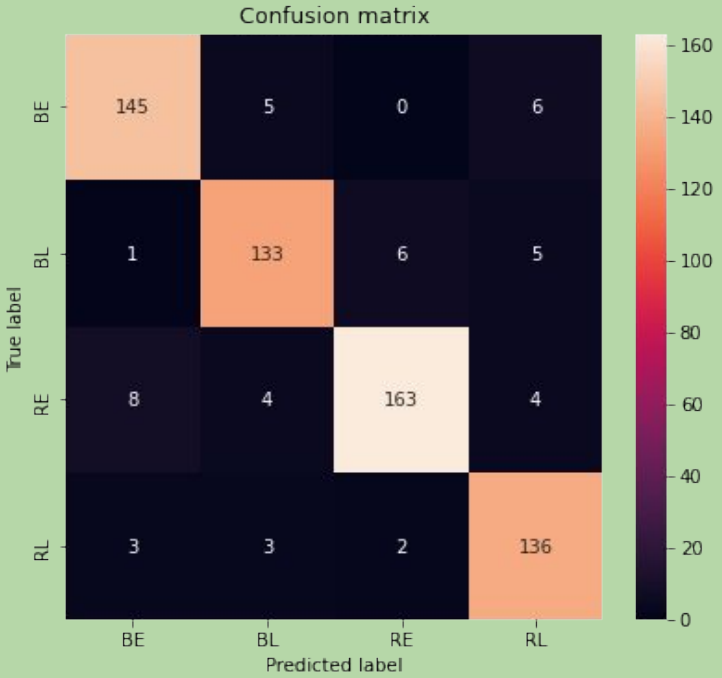
Source: Look, Listen and Learn More: Design Choices for Deep Audio Embeddings
Jason Cramer, Ho-Hsiang Wu, Justin Salamon, and Juan Pablo Bello.
IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pages 3852-3856, Brighton, UK, May 2019.

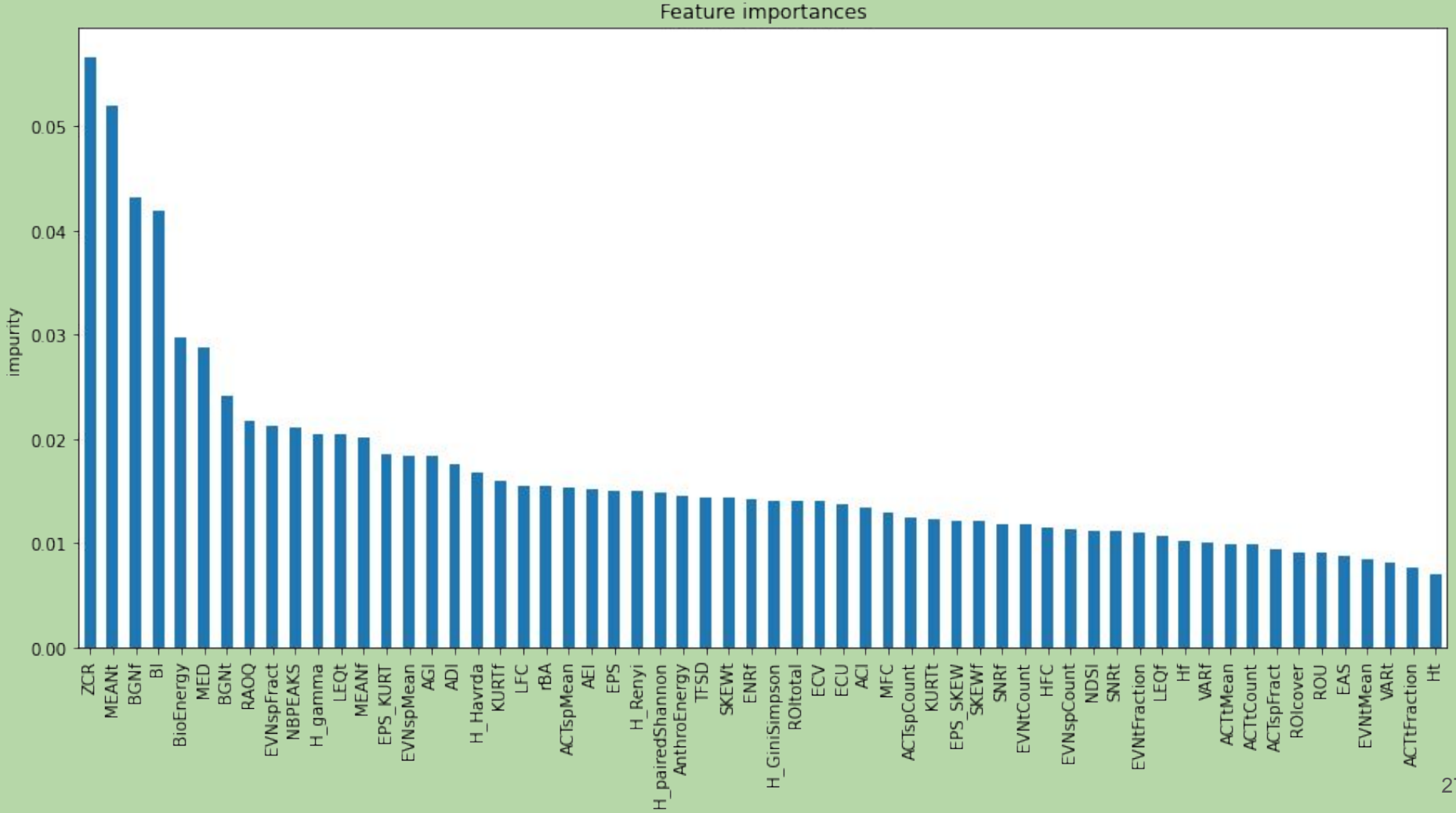
Approach with Embedded Indexes

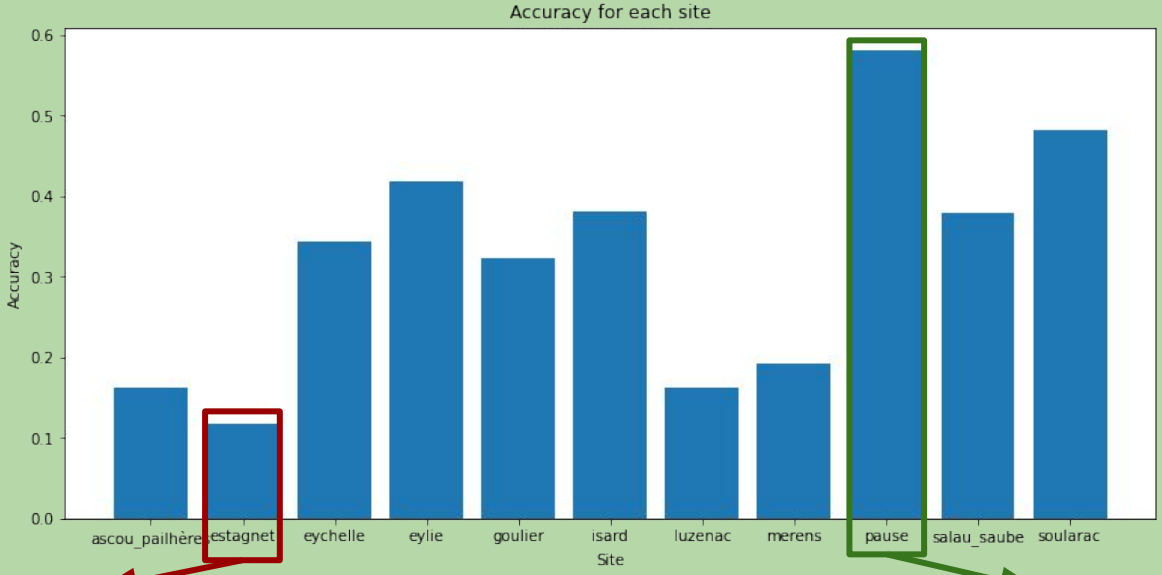


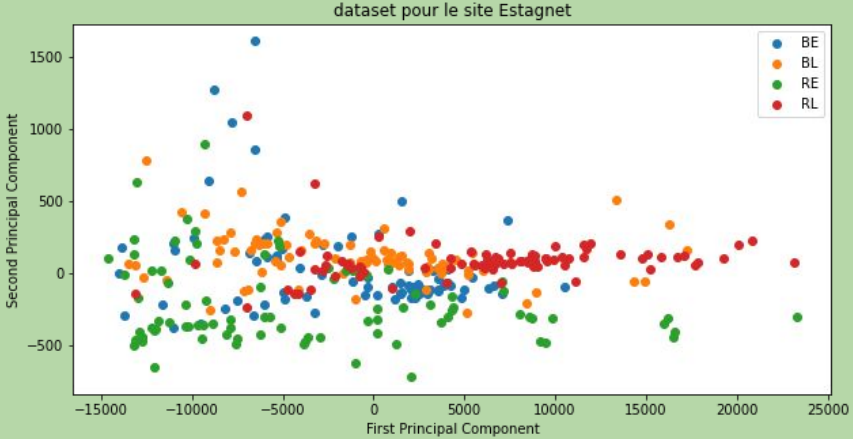
Analysis of index approach

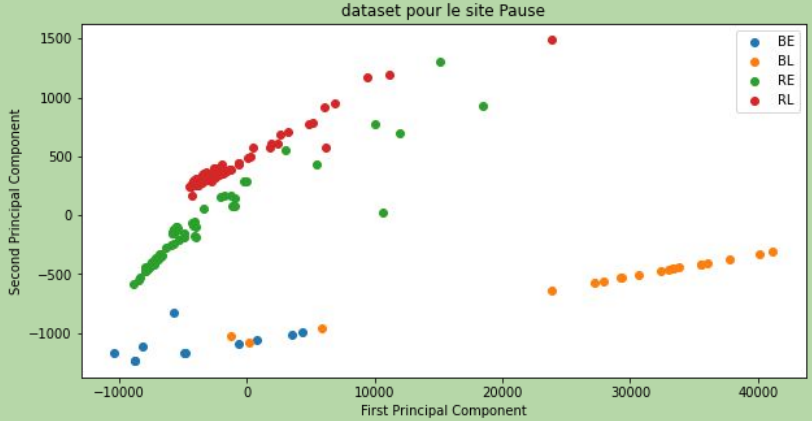
Methods	Accuracy
K-Nearest Neighbors	81.4 %
Support Vector Machine (SVM)	62.3 %
Kernel SVM	72.7 %
Naive Bayes	31.8 %
XGboost	83.9 %
Multilayer perceptron (MLP)	**91.8 %**




Feature importances
impurity
0.00
0.01
0.02
0.03
0.04
0.05
ZCR
MEANt
BGNf
BI
BioEnergy
MED
BGNt
RAOQ
EVNspFract
NBPEAKS
H_gamma
LEQt
MEANf
EPS_KURT
EVNspMean
AGI
ADI
H_Havrda
KURTf
LFC
rBA
ACTspMean
AEI
EPS
H_Renyi
H_pairedShannon
AnthroEnergy
TFSD
SKEWt
ENRf
H_GiniSimpson
ROItotal
ECV
ECU
ACI
MFC
ACTspCount
KURTt
EPS_SKEW
SKEWf
SNRf
EVNtCount
HFC
EVNspCount
NDSI
SNRt
EVNtFraction
LEQf
Hf
VARf
ACTtMean
ACTtCount
ACTspFract
ROIcover
ROU
EAS
EVNtMean
VARt
ACTtFraction
Ht

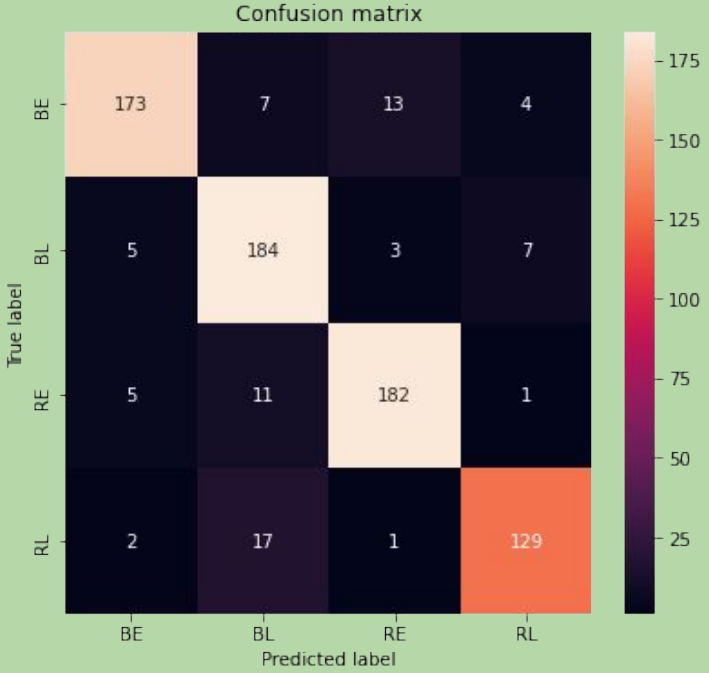

Accuracy for each site
Accuracy
0.0
0.1
0.2
0.3
0.4
0.5
0.6
ascou_pailhère
estagnet
eychelle
eylie
goulier
isard
luzenac
merens
pause
salau_saube
soularac
Site

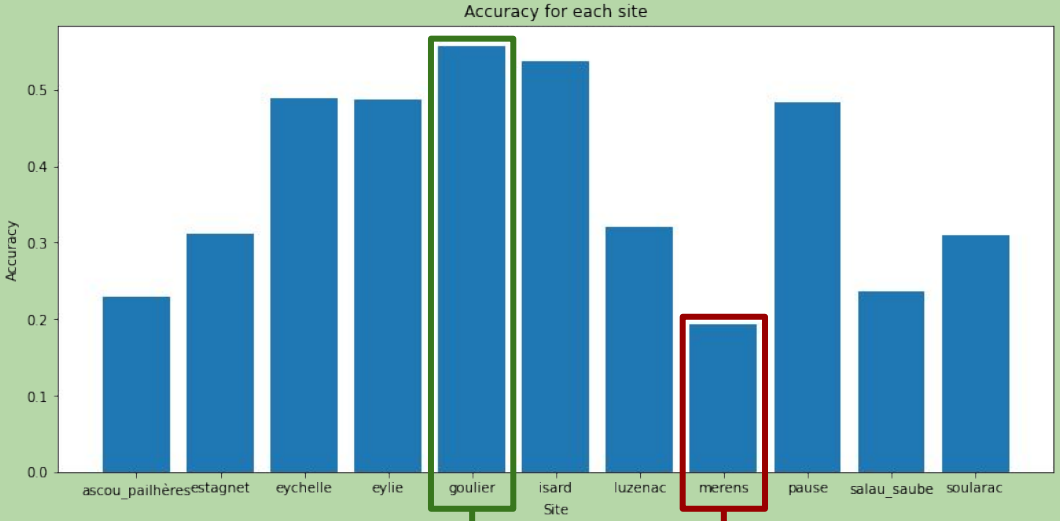

dataset pour le site Estagnet
BE
BL
RE
RL
Second Principal Component
First Principal Component

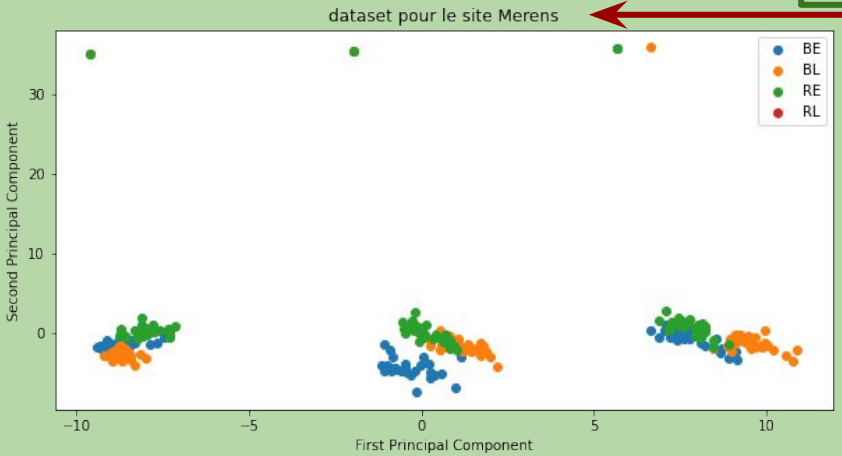

dataset pour le site Pause
BE
BL
RE
RL
Second Principal Component
First Principal Component

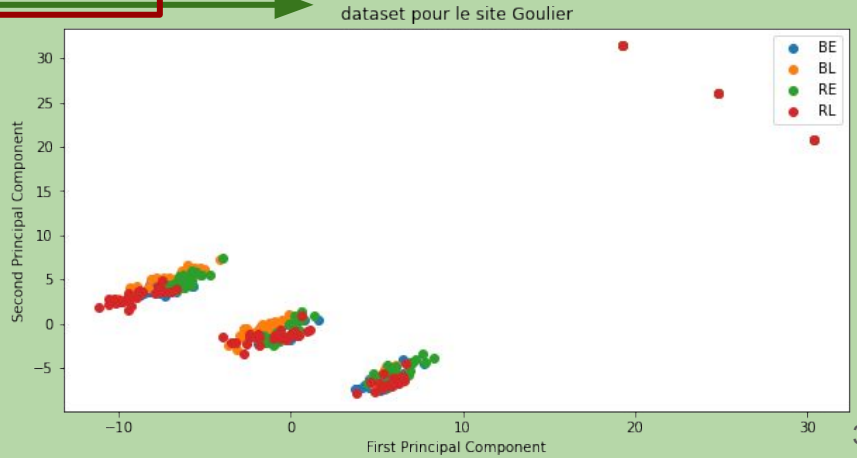
Analysis of embedded indexes approach

Methods	**Accuracy** (mel128 / mel256)
K-Nearest Neighbors	86.6 % / 83.5 %
Support Vector Machine (SVM)	75.1 % / 67.6 %
Kernel SVM	85.2 % / 81.0 %
Naive Bayes	44.1 % / 45.3 %
XGboost	74.1 % / 65.6 %
Multilayer perceptron (MLP)	**89.8 % / 84.1 %**



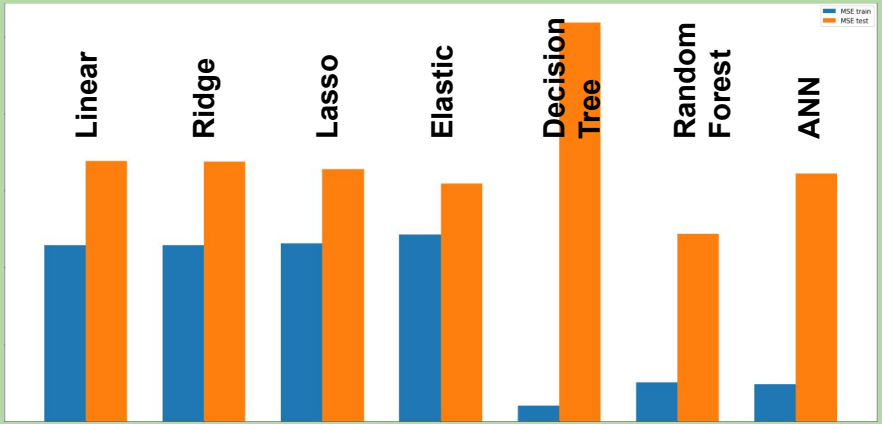

Accuracy for each site
Accuracy
0.0
0.1
0.2
0.3
0.4
0.5
ascou_pailhères
estagnet
eychelle
eylie
goulier
isard
luzenac
merens
pause
salau_saube
soularac
Site


dataset pour le site Merens
BE
BL
RE
RL
Second Principal Component
First Principal Component

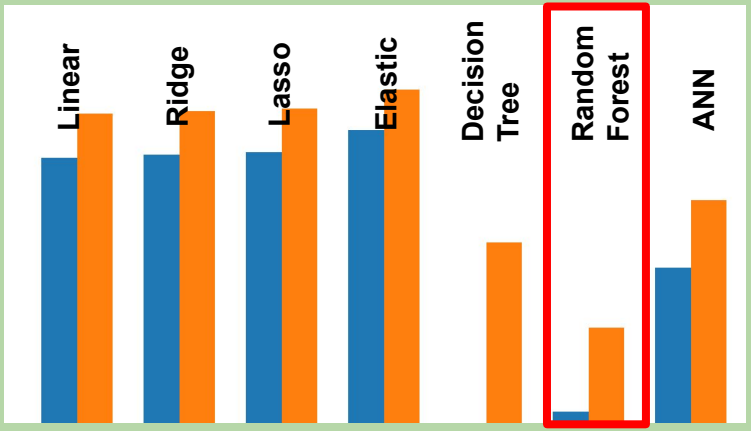

dataset pour le site Goulier
BE
BL
RE
RL
Second Principal Component
First Principal Component

Environmental Analysis: Choice of model and representation

MEAN MSE Benchmark (Blue: Train; Orange: Test)



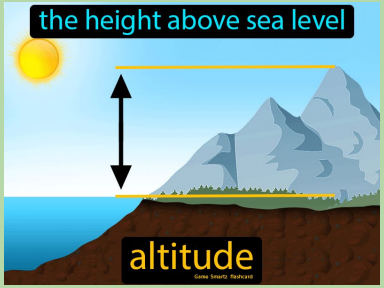
Latent representation



Acoustic indices

Best configuration

Environmental Analysis: Best configuration Evaluation



Average altitude (with an average accuracy of ~50m)

Slope (with an average precision of ~2°)

Distance to a trail (with an average accuracy of ~2m)

dominant species 87% ACC (3 categorical labels)

Perspectives

What are the relevant research perspective?

- More accurate latent space representation
- Train/Test systems on a wider range of data (geographic locations, seasons, weather, time of day, etc.)
- Cross-referencing results to understand environmental changes (global warming, increasing forests etc.)


Questions