If There's a Whale There's a Way

Whale and Dolphin Identification using Machine Learning and Deep Learning

Machine Learning Lecture – May 12th, 2022

Maren Rieker, MDS 2023

Victor Möslein, MDS 2023

Dinah Rabe, MPP 2022/ MDS 2023

Reed Garvin, MDS 2023

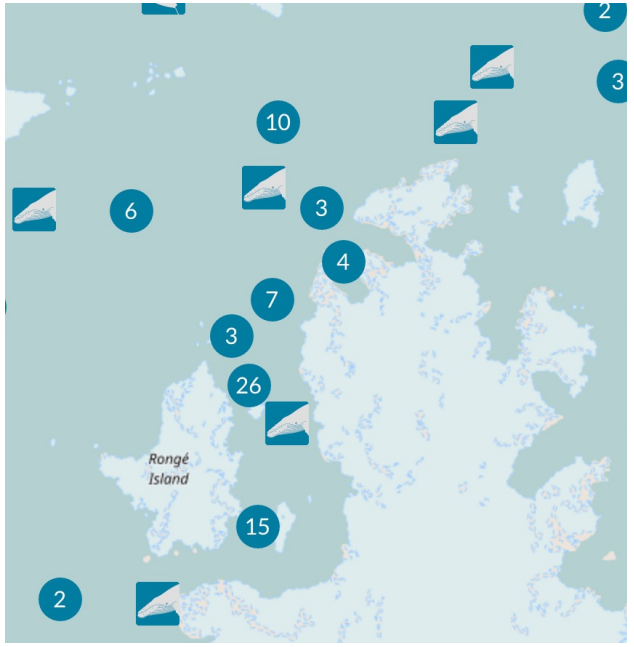
Hertie School

Happy Whale: A Collaborative Research Site

The goal of our project

Users **submit photos** of marine mammal encounters

Identification of species and individual animals by their unique markings



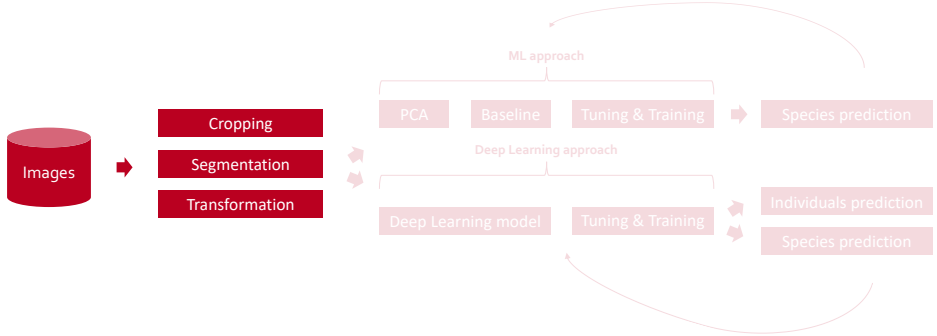
Animals are **tracked around the globe**

[1] Image Source: https://happywhale.com/home

Hertie School

Data and Preprocessing

Using image data



kaggle

51,033 images of whale and dolphin fins

Labels including animal IDs and species

Cropping into quadratic images of fins

Segmentation to remove background

Size reduction to reduce image size

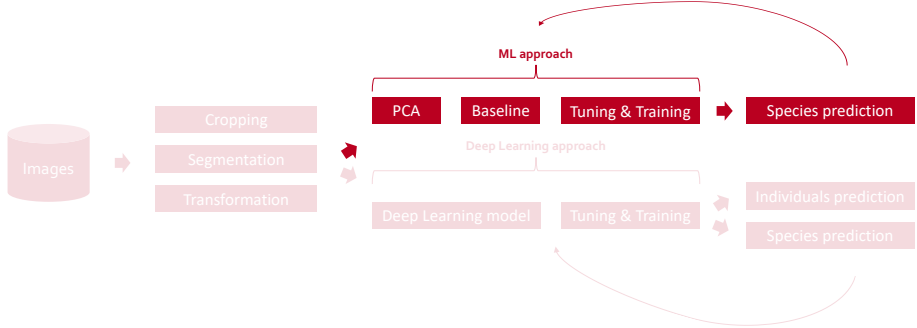
Transformation into number array

```
[[254., 254., 254., ..., 254., 254., 254.],
 [254., 254., 254., ..., 254., 254., 254.],
 [254., 254., 254., ..., 254., 254., 254.],
 ...,
 [254., 254., 254., ..., 254., 254., 254.],
 [254., 254., 254., ..., 254., 254., 254.],
 [254., 254., 254., ..., 254., 254., 254.]]
```

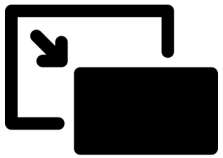

Hertie School

The Machine Learning Approach

Machine Learning Models

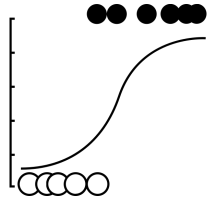


1



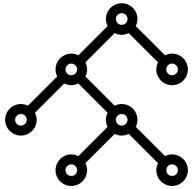
Principal Component Analysis (PCA) to reduce size of data set

2



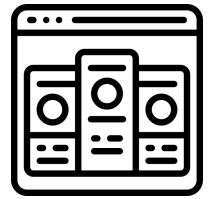
Baseline Logistic Regression model

3



Tuning and Training of Random Forest and XGBoost models

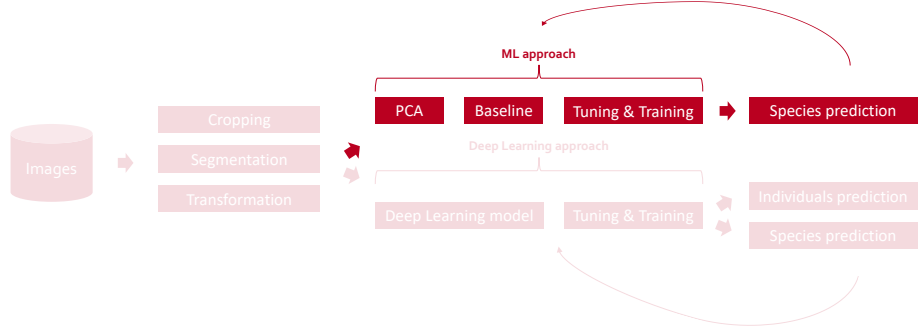
4



Evaluation and Comparison along performance metrics

Hertie School

Results and Limitations

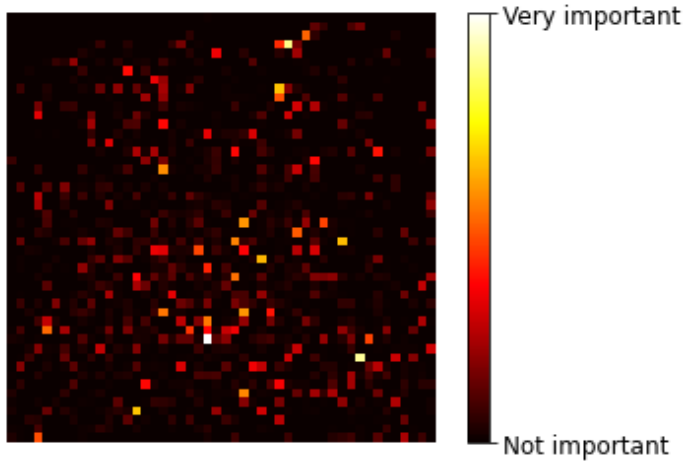


Complexity of data set

- **Low prediction accuracy** for all Machine Learning models

17%

Feature importance

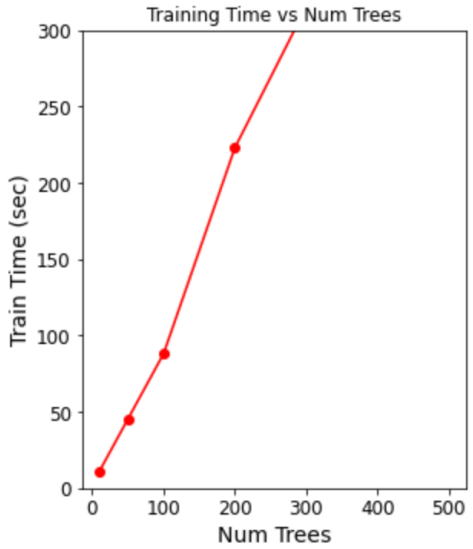


Size of data set

- **High training time** makes it infeasible to optimize models

>20h[1]

Training time of Random Forest model

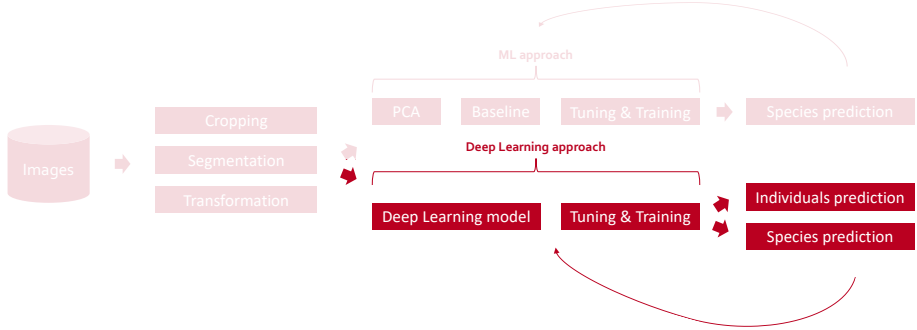


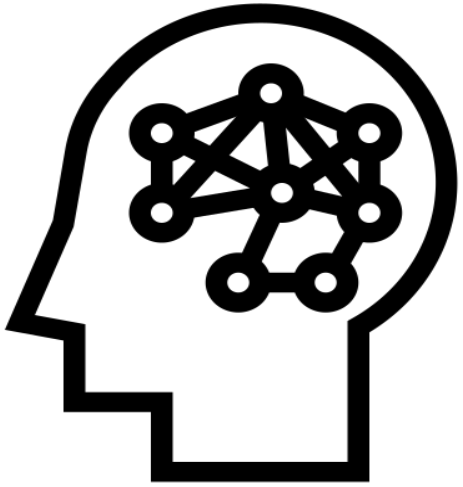
[1] For Randomized Search Cross Validation Hyperparameter Tuning of Random Forest and XGBoost with Early Stopping models using sklearn's RandomizedSearchCV method on the training data set after PCA (30,000 images with 502 features each)

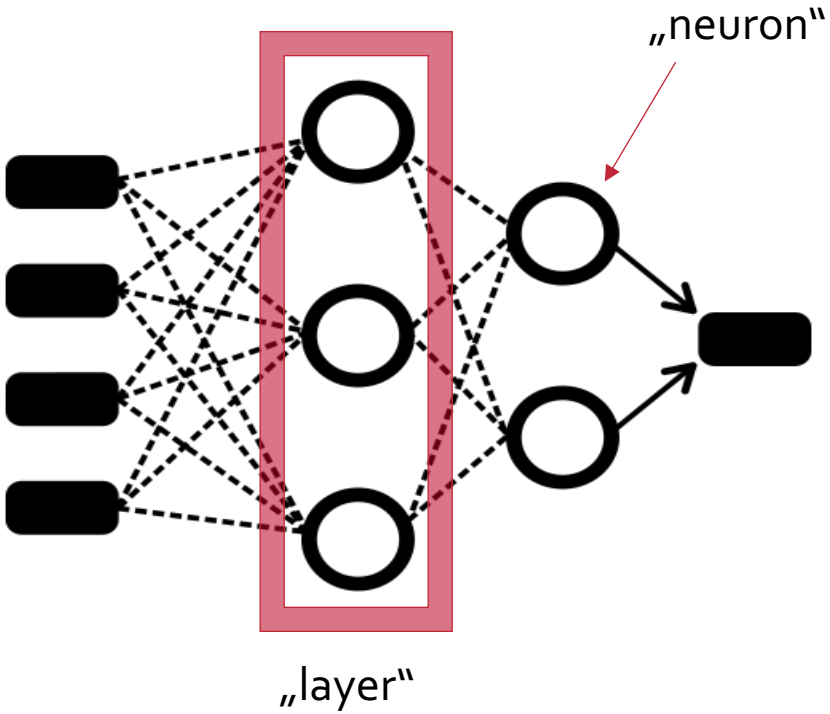
Hertie School

The Deep Learning Approach

Deep Learning Model

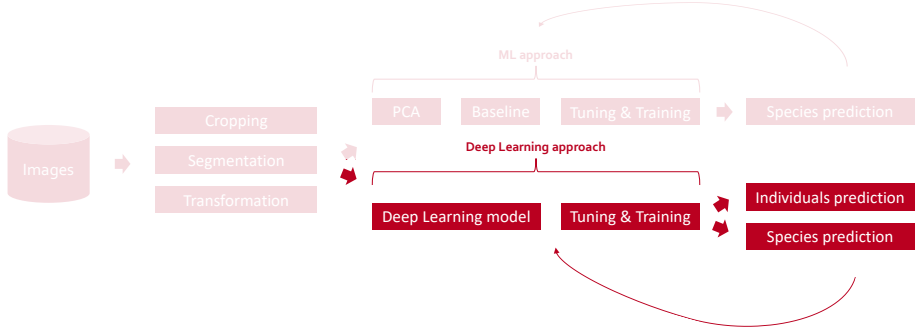




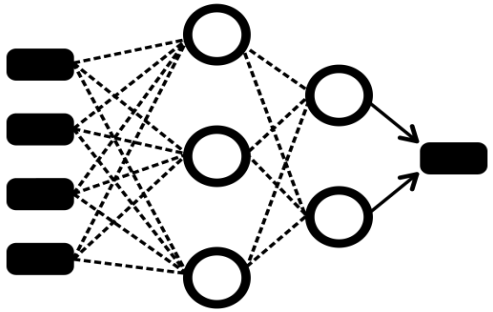


Hertie School

Deep Learning Model



1

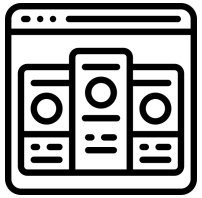


Baseline Untuned Convolutional Neural Network

2

Tuning and Training
Increasing the number of Epochs

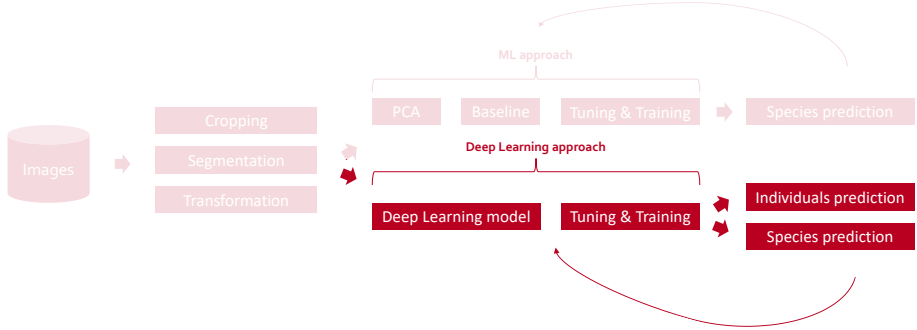
3



Evaluation and Comparison
along performance metrics

Hertie School

Results and Potential



- computational power
- automation

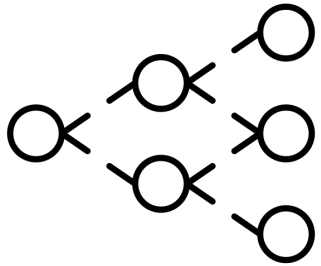
Low precision for **individual animals** **11%**

High accuracy for **species** **63%**

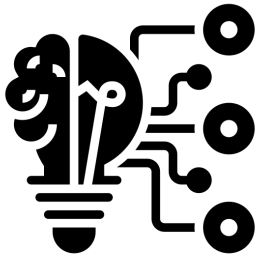
Hertie School

Analysis and Lessons Learned

Need for Advanced Deep Learning Methods



Conventional **Machine Learning** models **fail in complex classification tasks**, especially without the availability of supercomputing power



Advanced **Deep Learning** models are needed to accurately identify subtle differences in images and **assist human decision making**

Stay Happy, Whales!

References

References

A. Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. ” O’Reilly Media, Inc.”, 2019.

S. Ghosh, N. Das, I. Das, and U. Maulik. Understanding deep learning techniques for image segmentation. *ACM Comput. Surv.*, 52(4), aug 2019.

1.H. C. Huang, J. Joseph, M. J. Huang, and T. Margolina. Automated detection and identification of blue and fin whale foraging calls by combining pattern recognition and machine learning techniques. In *OCEANS 2016 MTS/IEEE Monterey*, pages 1–7. IEEE, 2016.

Dubois, D. and Owen, D., 2020. Understanding the Digital Ecosystem - Findings from the 2019 Federal Election. [online] Available at: <https://b1c9862c-6924-4cfd-9cbe-6c6f0144a777.filesusr.com/ugd/38105f_c2beb2fbbe5f46199fbc2f636ace59ee.pdf> [Accessed 21 November 2021].

Links

https://github.com/Whale- way/happy- whale

https://happywhale.com/home

https://www.kaggle. com/c/happy- whale- and- dolphin

https://keras.io/api/applications/ efficientnet/