EMS Annual Meeting 2022 | Bonn, Germany | EMS2022-522

A deep learning approach to identify sea breezes

Shalenys Bedoya-Valestt[1], Pablo Rozas-Larraondo[2], Cesar Azorin-Molina[1], Carlo Cafaro[3], Luis Gimeno[4], Lorenzo Minola[5,1], Jose A. Guijarro[6], Rober Dunn[3], Enric Aguilar[7] and Manola Brunet[7]

shalenys.bedoya@csic.es @climatoclab

Main goal

Explore CNNs vs. classical ML methods to take full advantage of the capabilities of DL/ML models to learn from sea breeze features and create a global identification method

Data and methods

Datasets (1981-2019)

Reanalysis: ERA5-Land hourly data
Observations: HadISD 3.1.1
[1,2]Sea breeze database
[3]JC weather types

Initial features (10)

Feature creation +(7)

Wind speeds, wind direction and humidity (MetPy)
Usable variables (cyclical encoding): wind direction, time of day and day of year

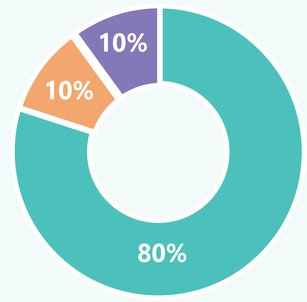
Feature selection

Model-based feature selection (RF) and PCA analysis were made without an improvement in the model performance

Transformations

MinMax scaling

Cross-validation



Training, Validation
Test

*Ramdonly splitted over 19-years 2000-2019

Intel(R) Core(TM) i9-10900
CPU @ 2.80GHz 2.81 GHz
128 GB RAM
NVIDIA GeForce RTX 3090 Ti

Baseline model

Input data
↓
Conv1D (24)
Batch Norm.
ReLU
↓
Conv1D (48)
Batch Norm.
ReLU
↓
Conv1D (96)
Batch Norm.
ReLU
↓
Conv1D (192)
Batch Norm.
LeakyReLU
↓
Flatten
↓
Dense (192)
Dropout (0.5)
ReLU
↓
Dense (96)
Dropout (0.5)
ReLU
↓
Dense (48)
Dropout (0.5)
ReLU
↓
Dense (24)
Dropout (0.5)
ReLU
↓
Dense (2)
Softmax
↓
Binary classification

Hyperparameters
Optimizer: Adam
Learning rate/decay: 0.001
Loss: cat. crossentropy
Batch size: 500
Epochs: 500/350

Results

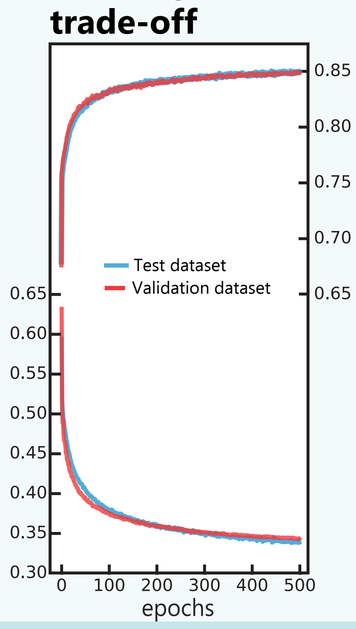
Individual station

Alicante, Spain
longitude: -0.6
latitude: 38.3

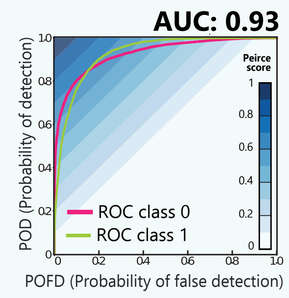
Data shapes:
train: (140256, 18)
validation: (17531, 18)
test: (17532, 18)

Good performance on Train, valid and test set

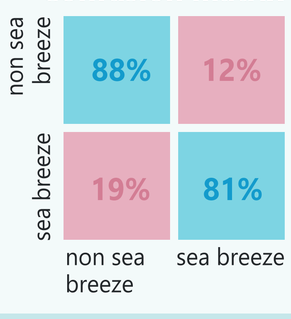
Learning curve trade-off



AUC: 0.93

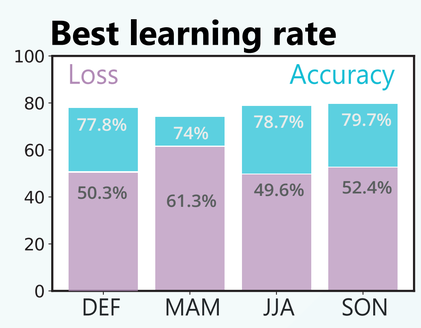


Confusion matrix



Predict on unseen data (1981-1999)

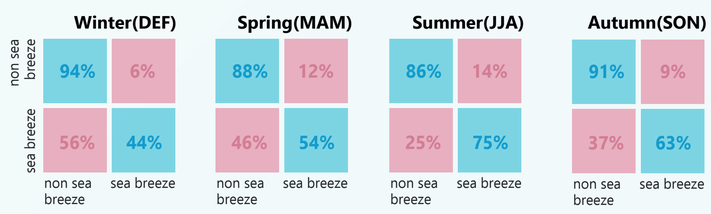
Best learning rate



Overall rates
loss: 53.4%
accuracy: 77.54%
SB days prediction: 63%

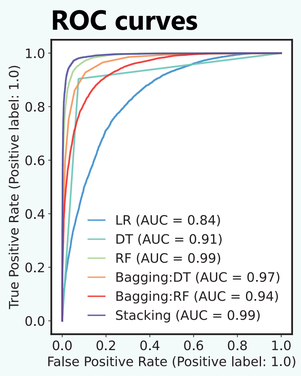
Less learning rates on seasonal datasets

especially in wintertime

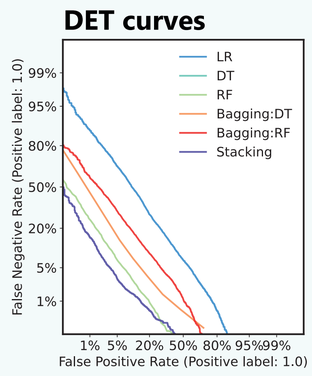


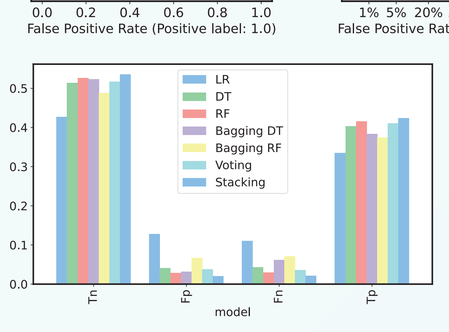
Learning with ML methods

ROC curves



DET curves

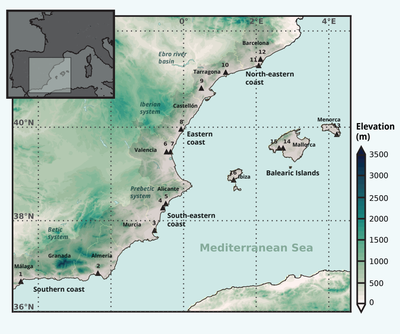




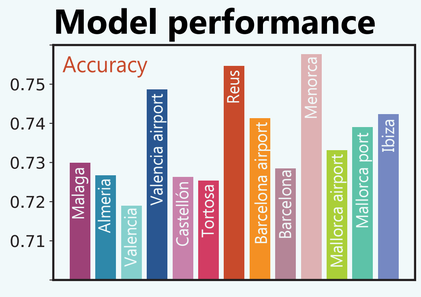
Stacking method scores
loss: 4.1%
accuracy: 95.9%
SB days prediction: 95.5%

Best performance with stacking ensemble learning

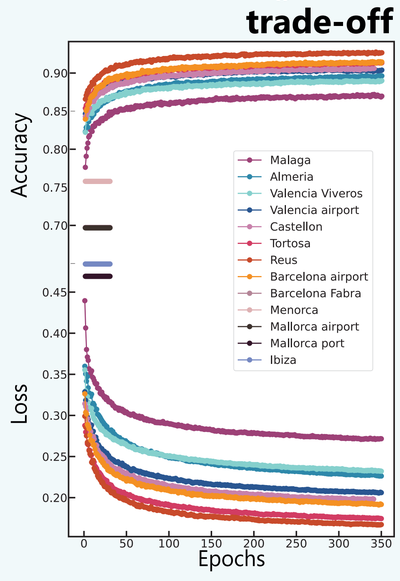
Evaluate the base model learning in other locations



Model performance



Learning curve trade-off



Good performance in Mediterranean sites

What's next?

1. Continue exploring 2D and 3D convolutional neural networks - weather and terrain features
2. Improve baseline model with multi-input features and recurrent learning
3. Train model with global sea breeze database (requires previous SB identification)

References: [1]Azorin-Molina et al. (2011a). ***Theor Appl Climatol.*** 106:531–546. [2]Bedoya-Valestt et al. (2022). ***Clim. Dyn.*** In press. [3]Azorin-Molina et al. (2011b). ***Int J Climatol.*** 31: 468 – 486

FUNDING: We acknowledge supports from the Ramon y Cajal Fellowship (RYC-2017-22830), and the Spanish Ministry of Science, Innovation and Universities (RTI2018-095749-A-I00)

[1]CIDE (CSIC-UV-Generalitat Valenciana), [2]ANU College of Science, [3] MetOffice UK, [4] EPhysLab (CIM-UVIGO), [5] DIST, [6] Retired from AEMET, [7] C3 (URV)