

Innovation Lab - Final Presentation

Climex

Andreas, Carolin, Guoren, Henri and Yao

18th March 2021



Message from outer space

A nighttime satellite view of Earth from space, showing city lights and auroras.

Agenda

Our journey today.



Problem

Data
Challenges
Task



Solution

Modeling
Interpretable ML
Pipeline & Packaging



Workflow

Development Process
Collaboration
Scrum Workflow



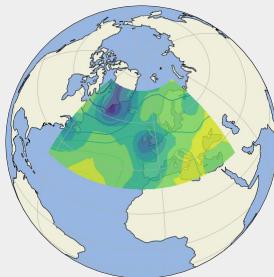
Conclusions & further research

Data

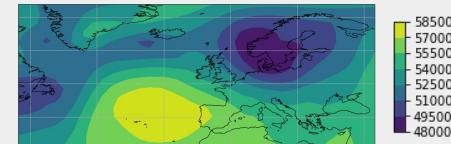
Our data consists of two air pressure levels over Europe of 110 years that are labelled daily.

Spatial

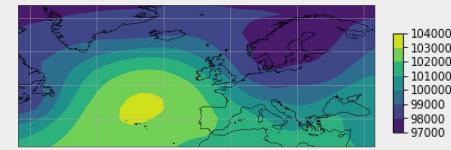
**two
air pressure
levels over
Europe**



mslp = mean sea
level pressure



z500 =
geopotential
height at 500 hPa



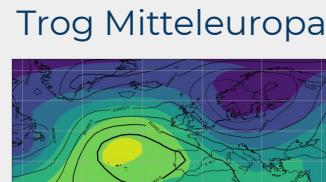
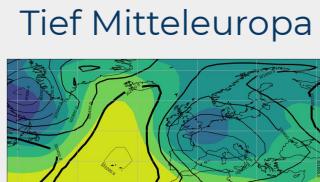
Temporal

**3-hourly
resolution
from
1900 to 2010**



Label

**different humans
decided
daily labels**

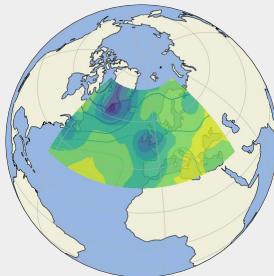


Data

Our data consists of two air pressure levels over Europe of 110 years that are labelled daily.

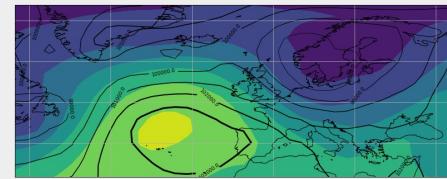
Spatial

two air pressure levels over Europe



mslp = mean sea
level pressure

z500 =
geopotential
height at 500 hPa



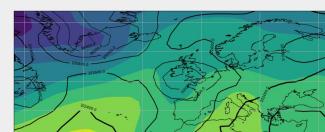
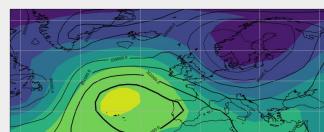
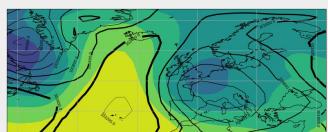
Temporal

3-hourly resolution from 1900 to 2010



Label

different humans decided daily labels

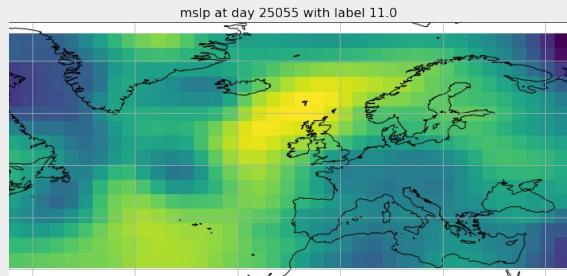


Data Quality

Example images:min distance between group of label 11-0 using neural network last layer feature as dataset.

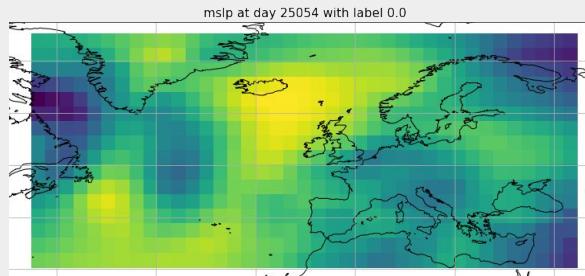
Tief Mitteleuropa

mslp

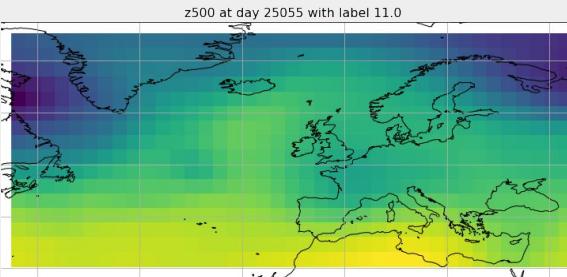


Residual

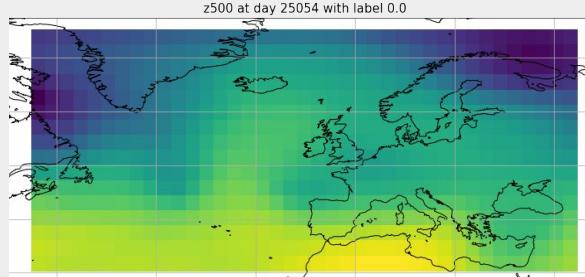
mslp at day 25054 with label 0.0



z500

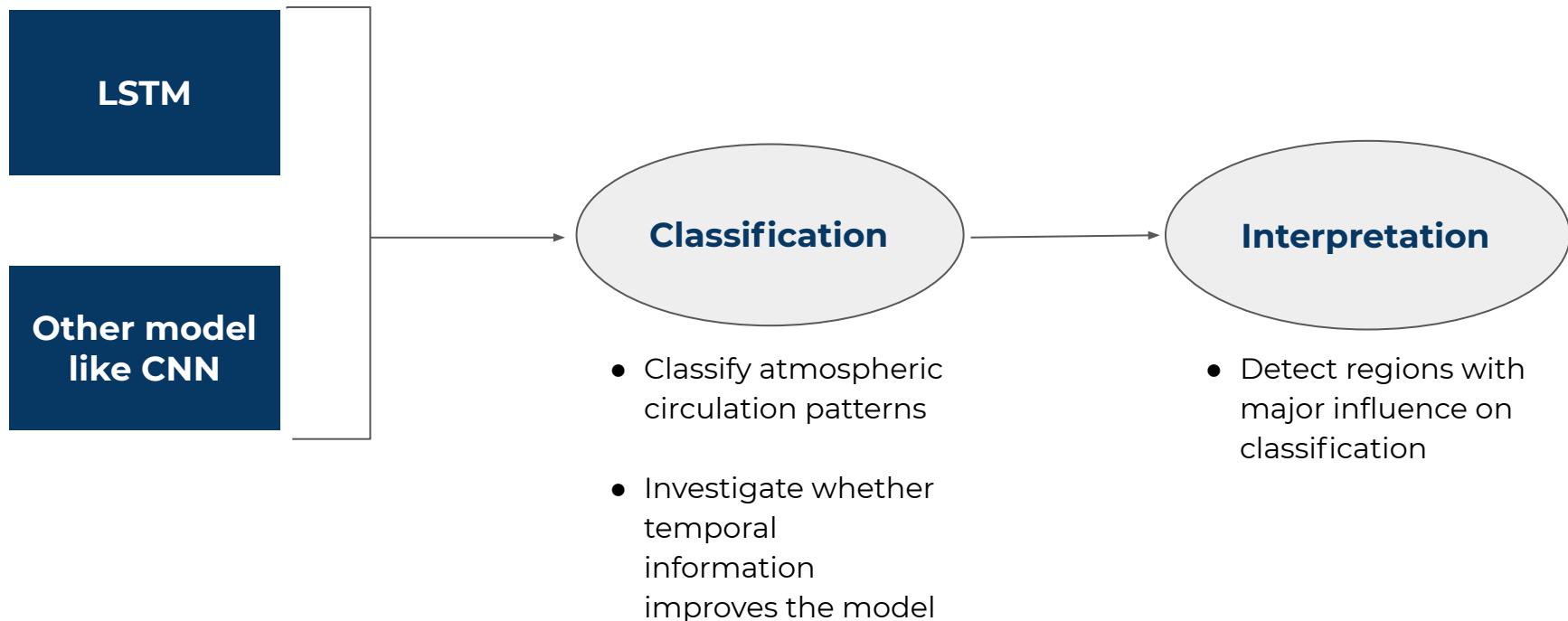


z500 at day 25054 with label 0.0



Project Summary - Task

Classification of atmospheric circulation patterns using Deep Learning.



Challenges

Challenges for both data processing and modelling.

Labelling

Manual labelling?
Label simulation?

- Similar feature values but different labels.

Spatial & Temporal dependence

Spatial dependence

- Similar to images, but not images. Solutions for image classification need to be adapted.

Temporal dependence

- Data splitting need to be careful.
- Take this dependency into the model.

Highly imbalanced data



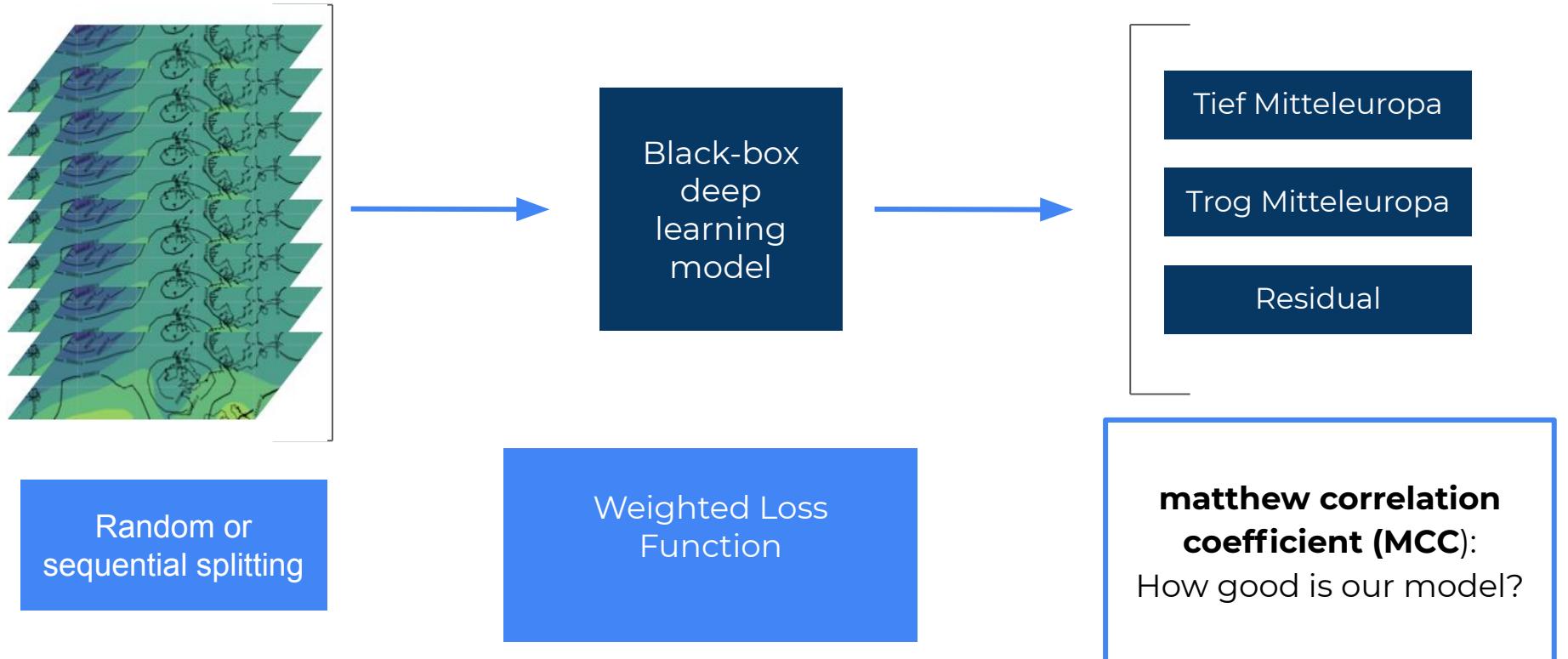


Model Setup And Evaluation



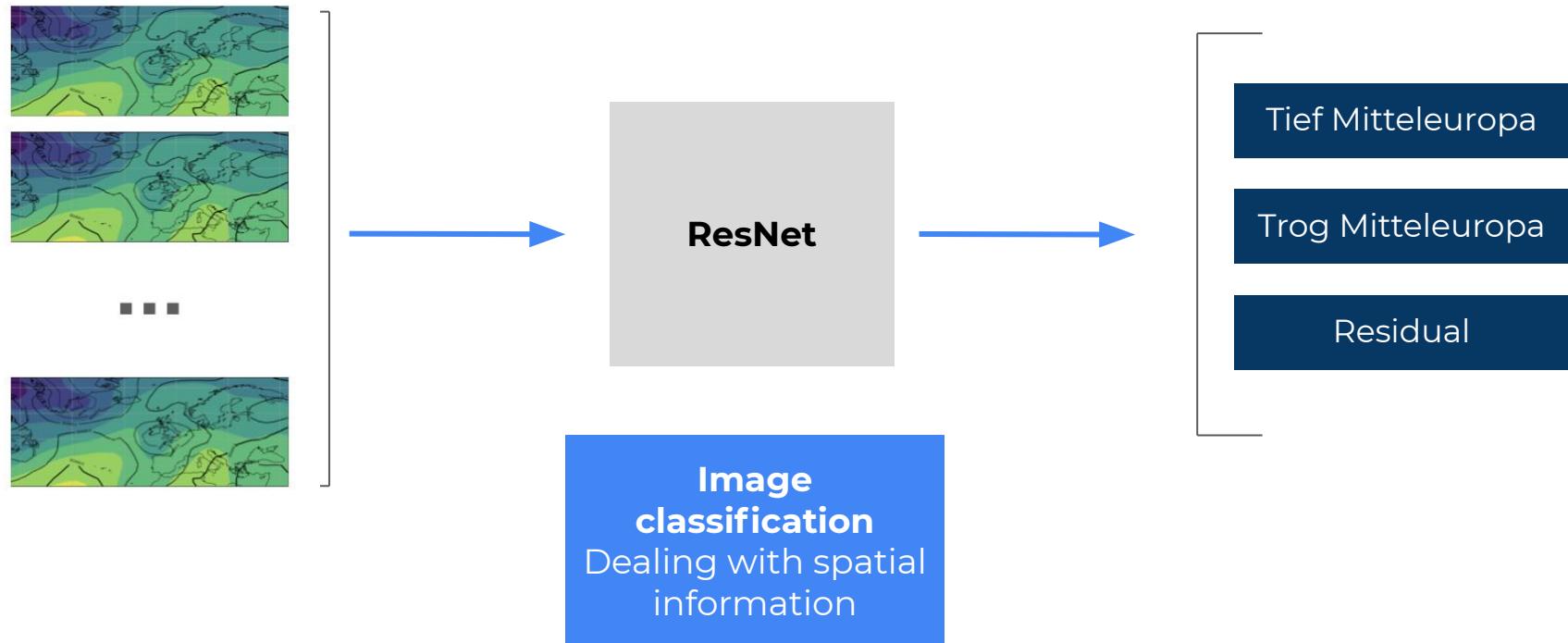
Modelling

- highly imbalance data -> weighted loss function & MCC as evaluation score
- temporal dependency -> different train/test/val splitting



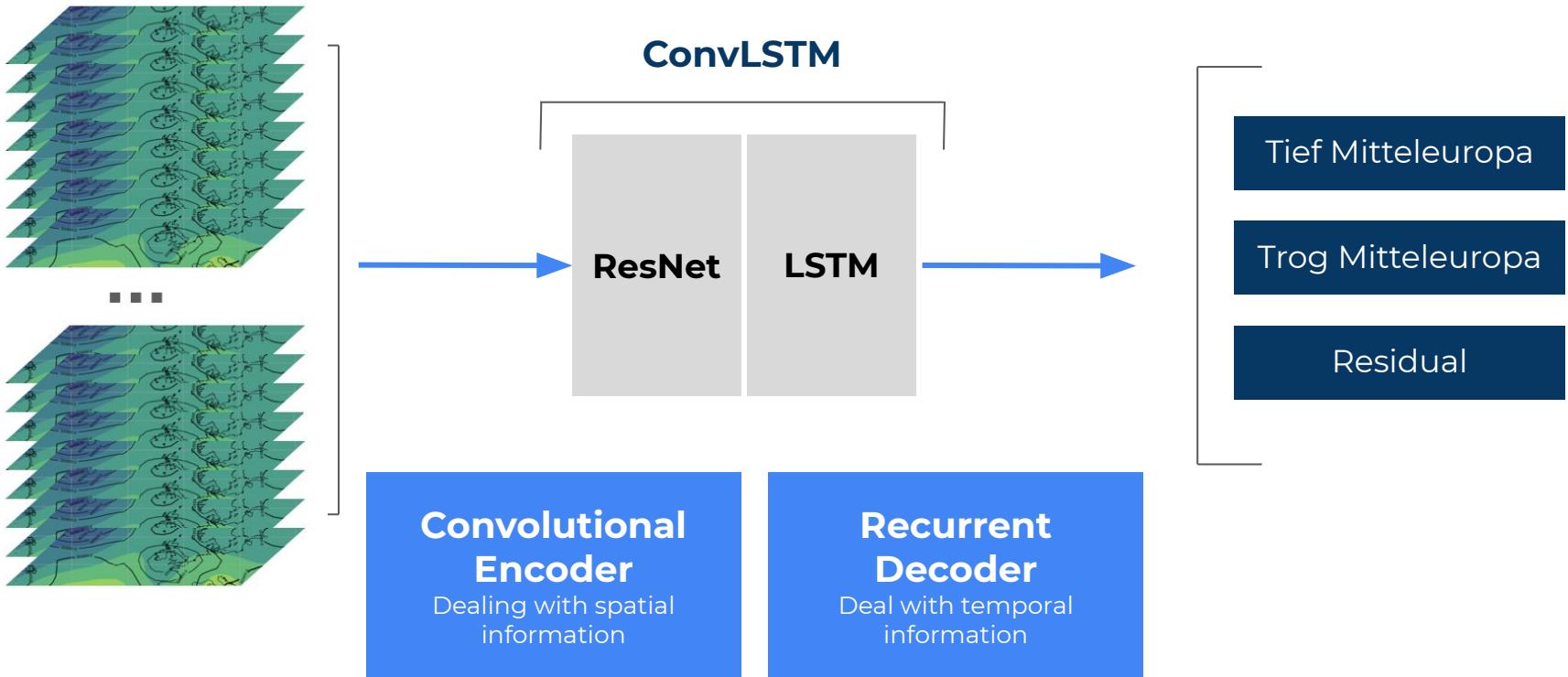
Modelling

Consider Spatial Info



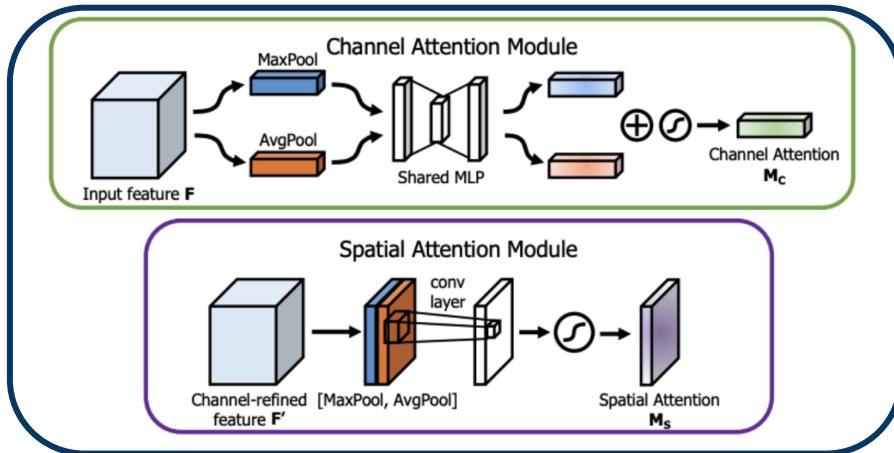
Modelling

Consider both Spatial and Temporal Info



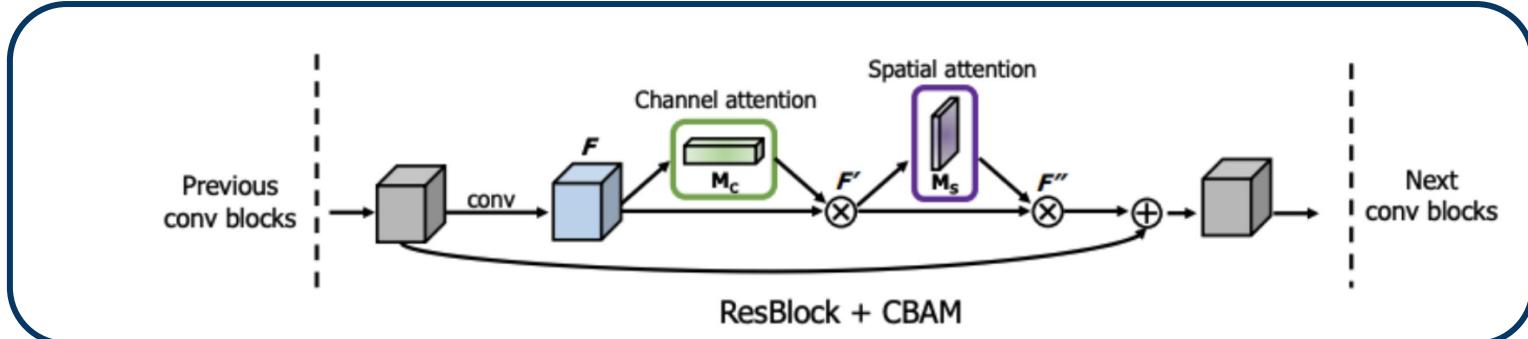
Modelling

Add some Attention might help



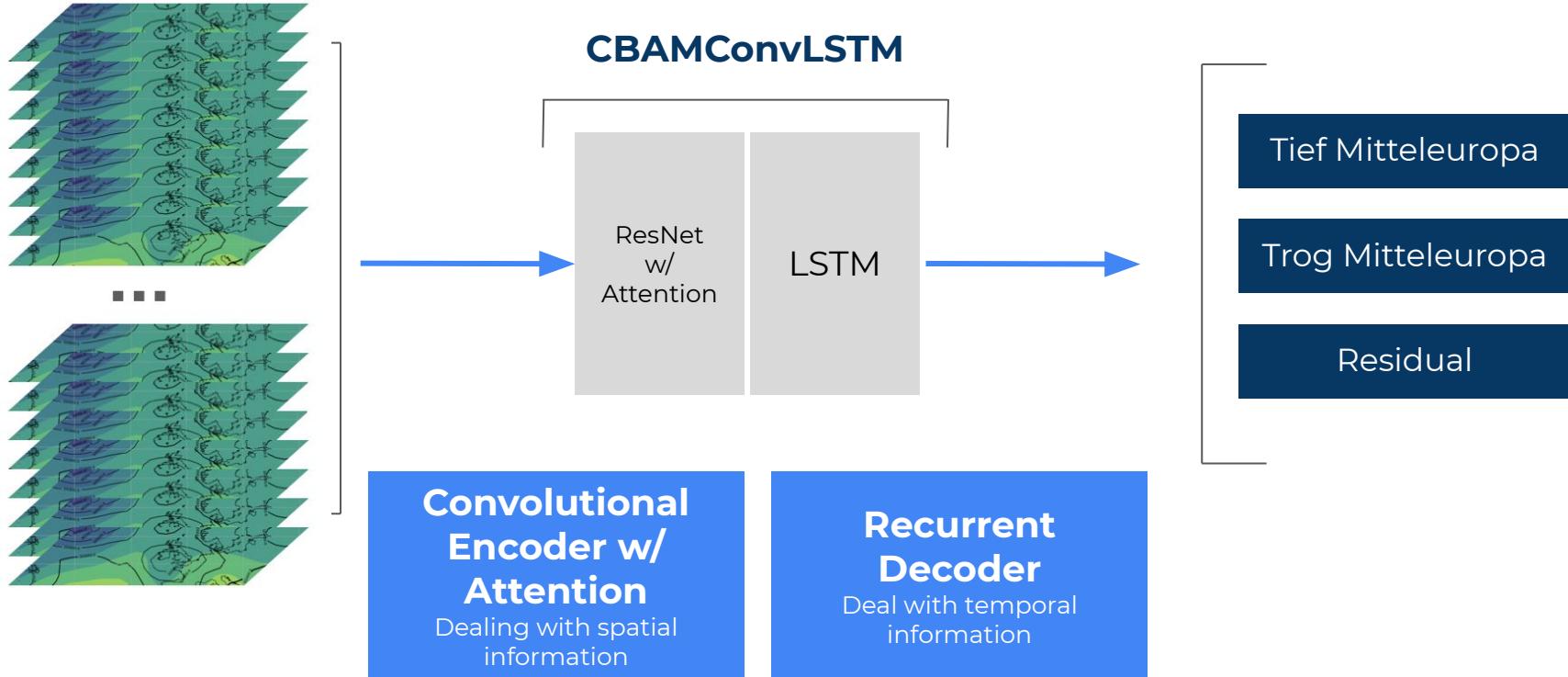
CBAM block

channel attention & spatial attention



Modelling

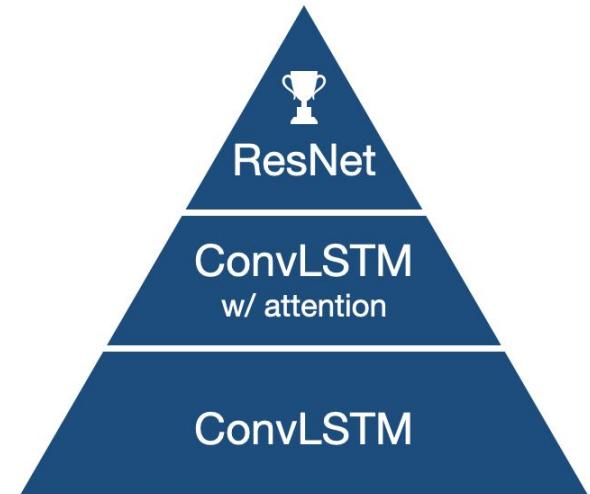
Integrate CBAM into ResNet



Modelling

ResNet wins!

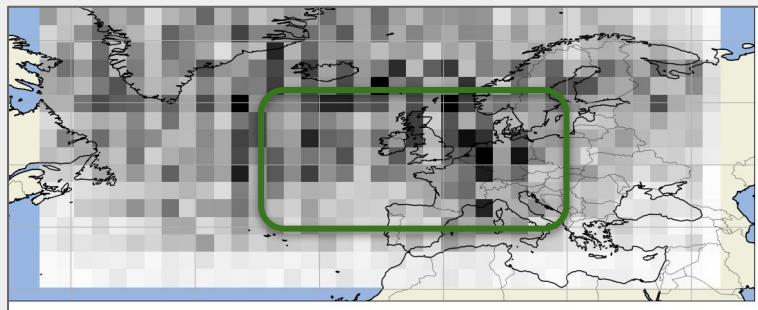
MCC		Sequential Splitting	Random Splitting
ResNet		0.300	0.330
ConvLSTM	1 day	0.197	0.193
	3 days	0.231	0.203
ConvLSTM w/ Attention	1 day	0.220	0.240
	3 days	0.245	0.208



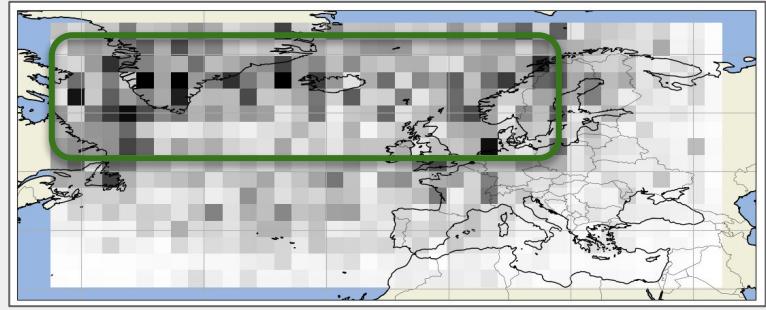
Interpretable Machine Learning

The importance of regions differs between the classes.

Tief Mitteleuropa



Trog Mitteleuropa



Conclusion:

- No object detection → whole picture is important for classification
- Different regions are important for different Großwetterlagen
 - for Tief Mitteleuropa: focus on middle of Europe and the Northeastern Atlantic
 - for Trog Mitteleuropa: focus on Scandinavia, Iceland and Greenland

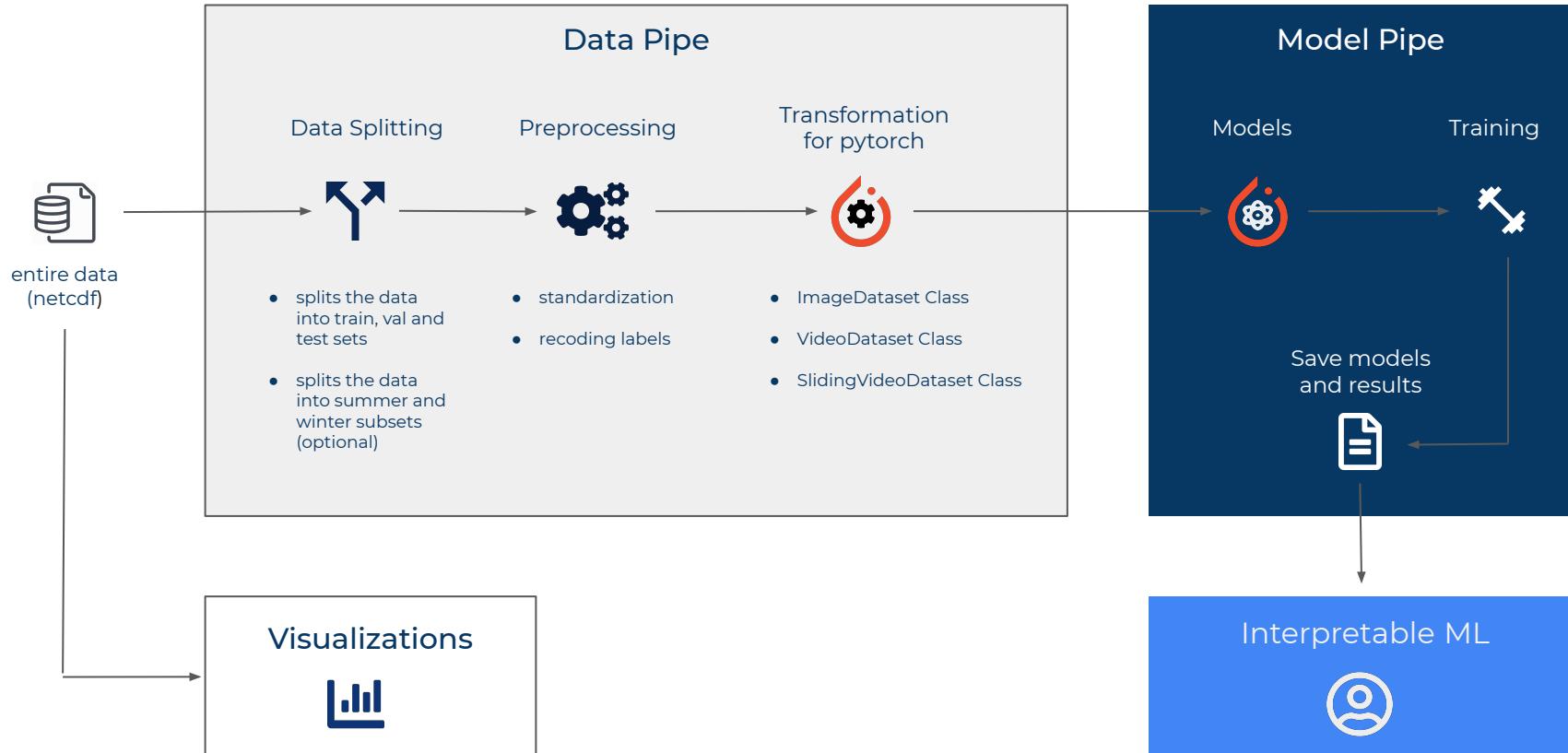


Pipeline And Packaging



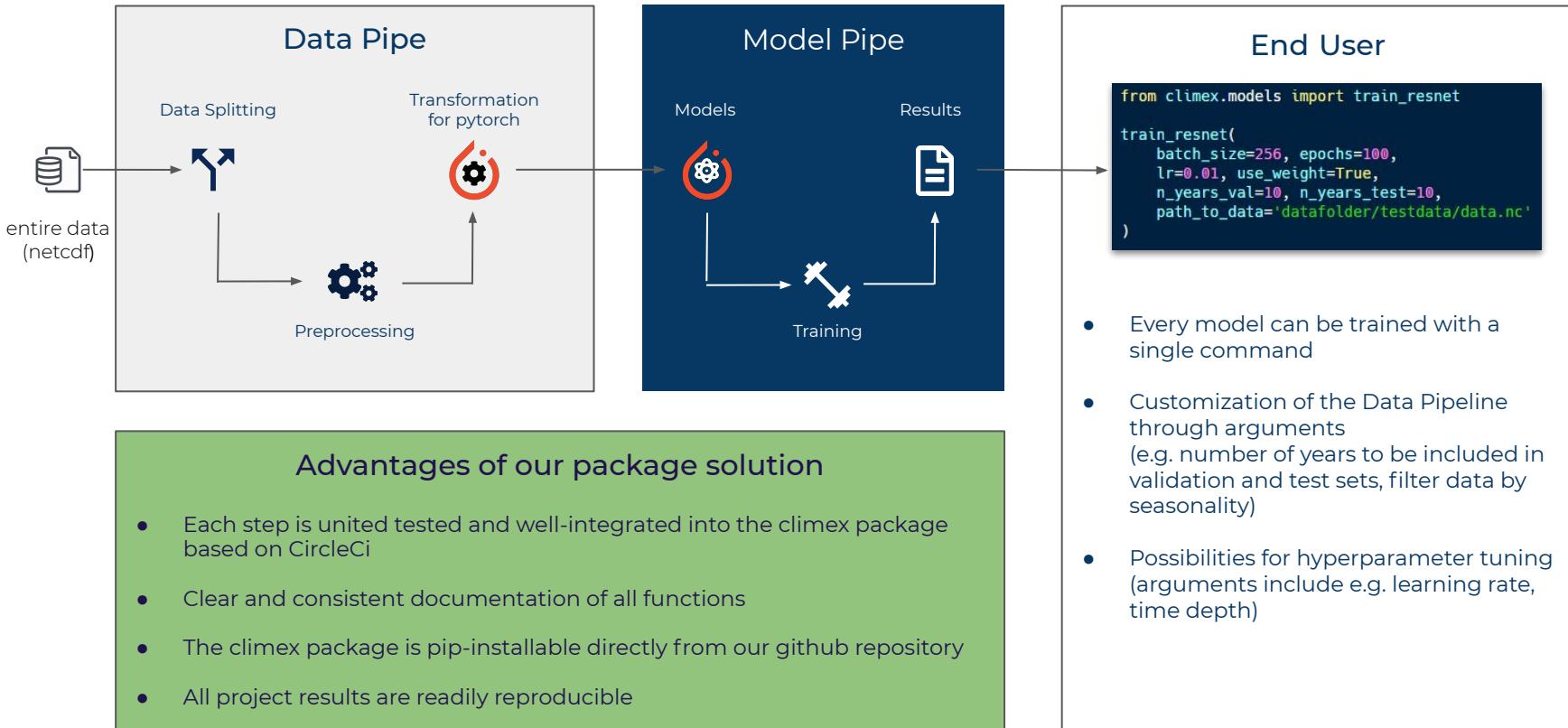
Pipeline and Packaging

Our consistent end-to-end pipeline avoids redundancies and creates a very clear structure for our project



Pipeline and Packaging

Our pip-installable climex package makes all of our results reproducible in an intuitive and well-tested manner



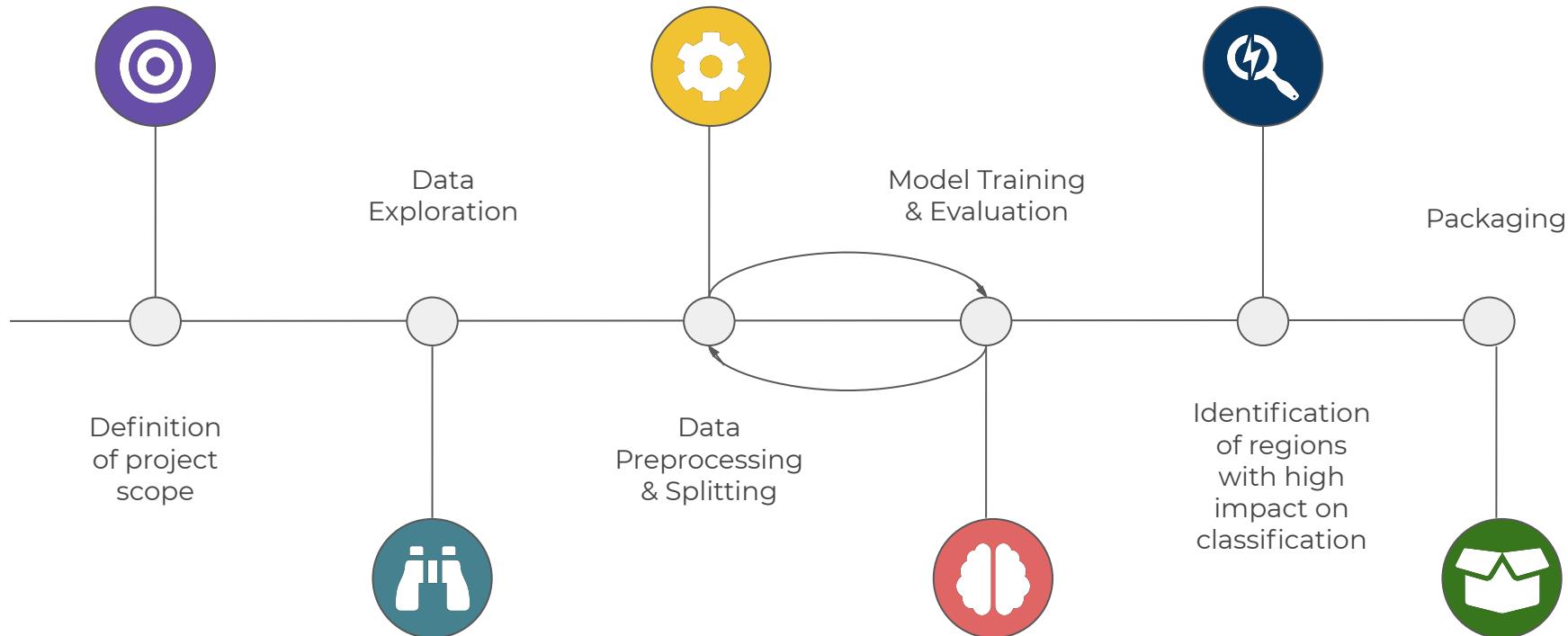


Workflow & Team Coordination



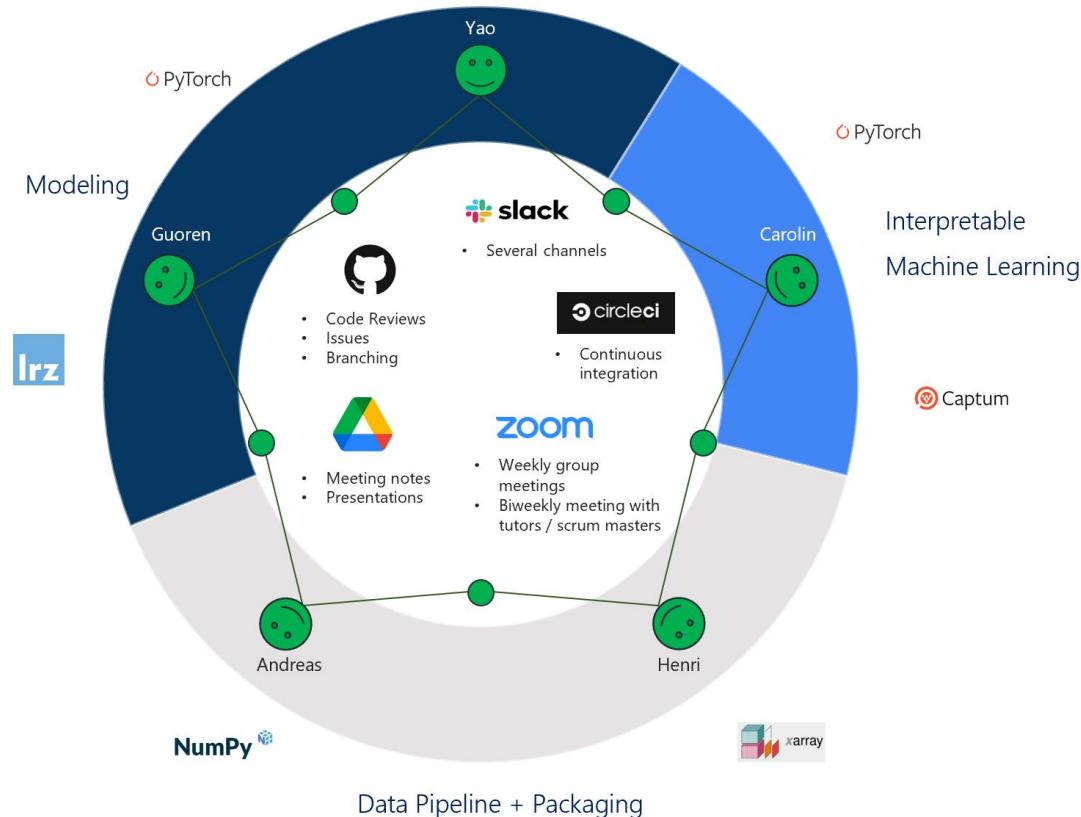
Development process

The common definition of milestones forms the basis for tracking our development process and planning the weekly sprints



Collaboration

The continuous common integration of the individual parts of our specialists resulted in a successful and well-integrated solution



Scrum Workflow

Some parts of the Scrum workflow worked very well from the beginning, while others were quickly learned and improved throughout the Innolab project

Scrum Workflow

Some parts of the Scrum workflow worked very well from the beginning, while others were quickly learned and improved throughout the Innolab project

Project Board

The screenshot shows a project board interface with a backlog section. The backlog contains the following items:

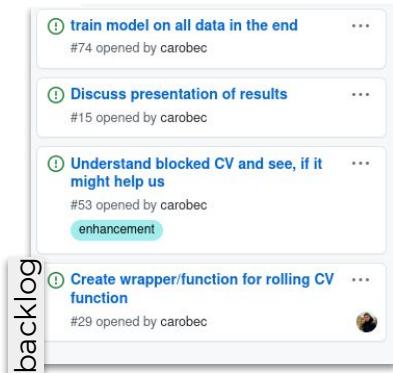
- train model on all data in the end ...
#74 opened by carobec
- Discuss presentation of results ...
#15 opened by carobec
- Understand blocked CV and see, if it might help us ...
#53 opened by carobec
enhancement
- Create wrapper/function for rolling CV function ...
#29 opened by carobec

A small profile picture of a person is visible next to the last task.

Scrum Workflow

Some parts of the Scrum workflow worked very well from the beginning, while others were quickly learned and improved throughout the Innolab project

Project Board



Learnings & Conclusions

- + Issues and project board
- A bigger picture of the project

Scrum Workflow

Some parts of the Scrum workflow worked very well from the beginning, while others were quickly learned and improved throughout the Innolab project

Project Board

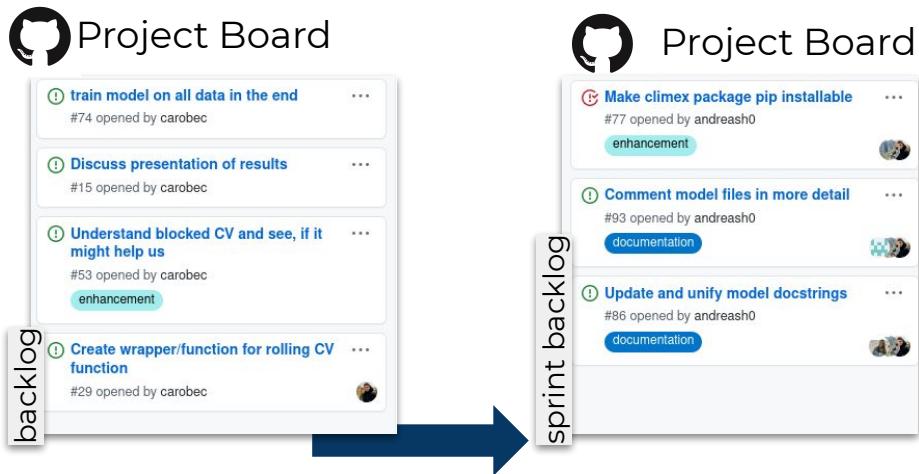
The screenshot shows a project board interface with a backlog section. The backlog contains the following items:

- train model on all data in the end ...
#74 opened by carobec
- Discuss presentation of results ...
#15 opened by carobec
- Understand blocked CV and see, if it might help us ...
#53 opened by carobec
enhancement
- Create wrapper/function for rolling CV function ...
#29 opened by carobec

A small profile picture of a person is visible next to the last task.

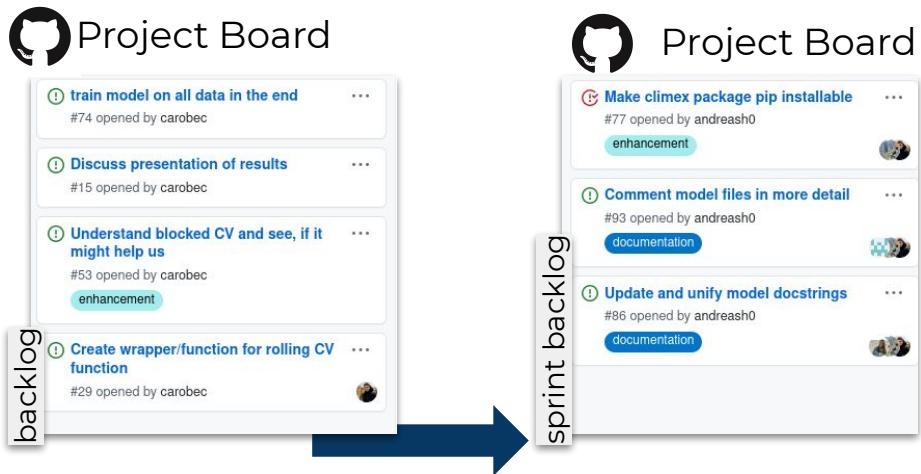
Scrum Workflow

Some parts of the Scrum workflow worked very well from the beginning, while others were quickly learned and improved throughout the Innolab project



Scrum Workflow

Some parts of the Scrum workflow worked very well from the beginning, while others were quickly learned and improved throughout the Innolab project

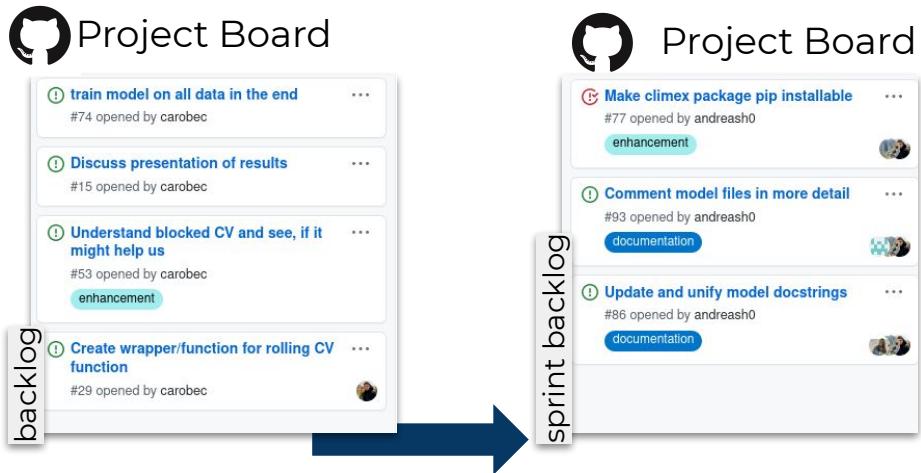


Learnings & Conclusions

- + Short term goals and more granular specification

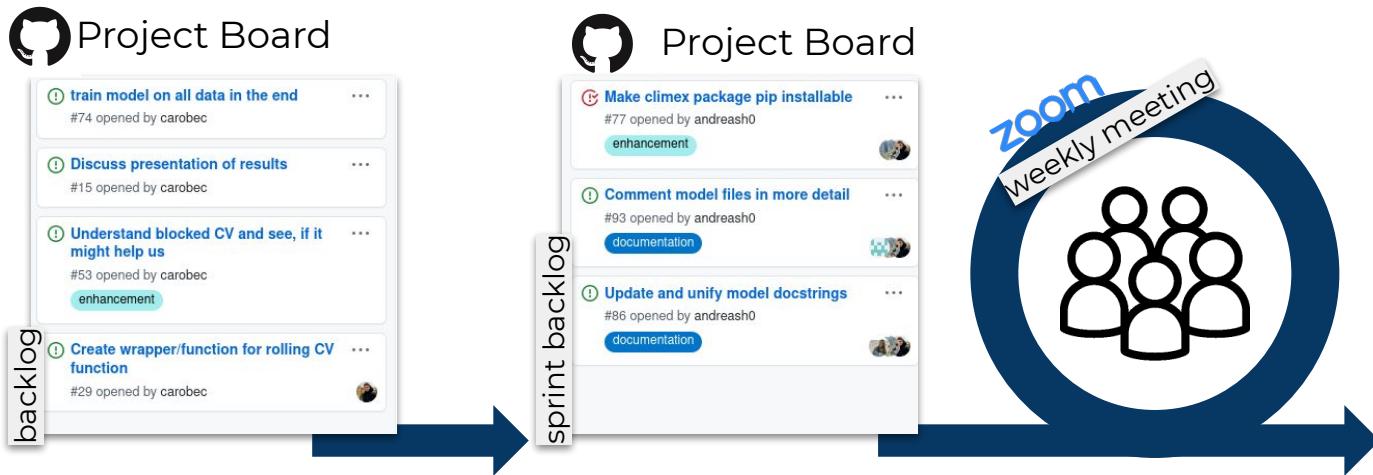
Scrum Workflow

Some parts of the Scrum workflow worked very well from the beginning, while others were quickly learned and improved throughout the Innolab project



Scrum Workflow

Some parts of the Scrum workflow worked very well from the beginning, while others were quickly learned and improved throughout the Innolab project

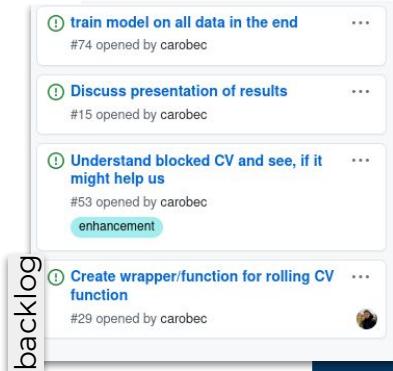


Scrum Workflow

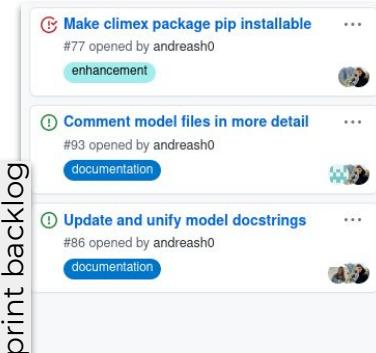
Some parts of the Scrum workflow worked very well from the beginning, while others were quickly learned and improved throughout the Innolab project



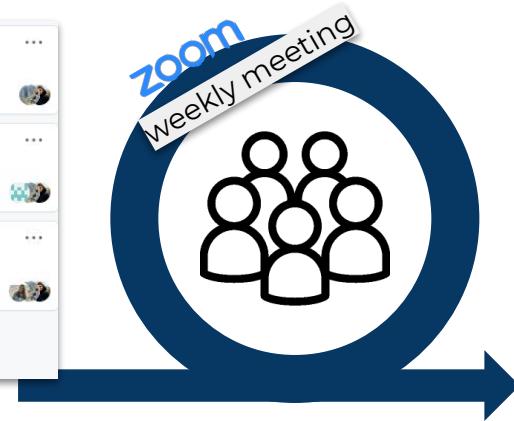
Project Board



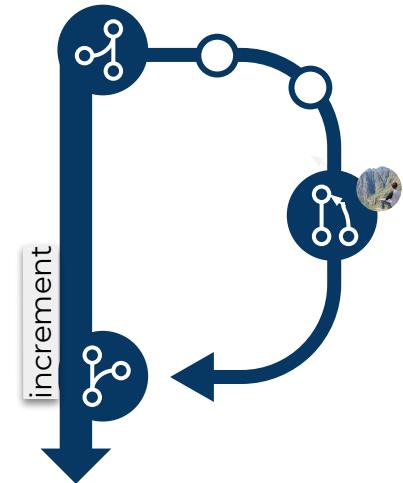
backlog



sprint backlog



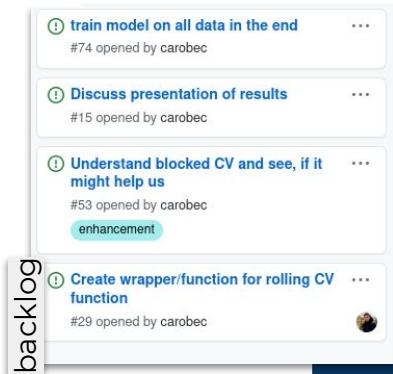
-o-Git Workflow



Scrum Workflow

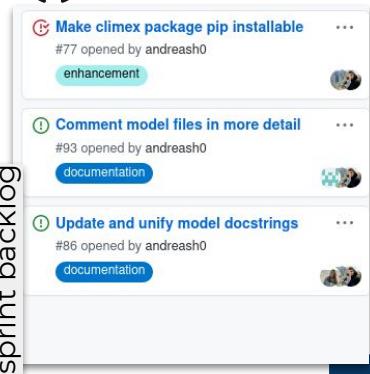
Some parts of the Scrum workflow worked very well from the beginning, while others were quickly learned and improved throughout the Innolab project

Project Board



backlog

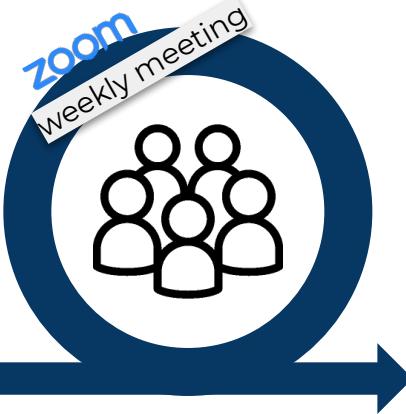
Project Board



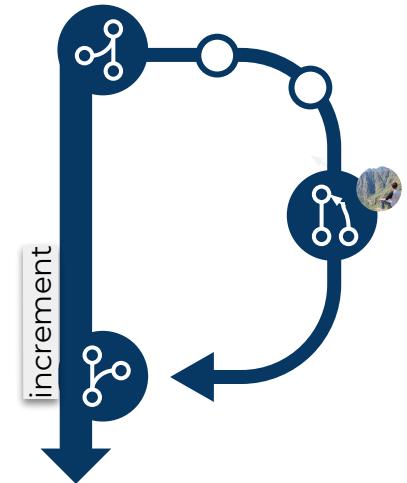
sprint backlog

Learnings & Conclusions

- + Git as organization tool
- + branching as save and secure foundation for commits and increments of the project
- Unified style guide and better communication



-o-Git Workflow

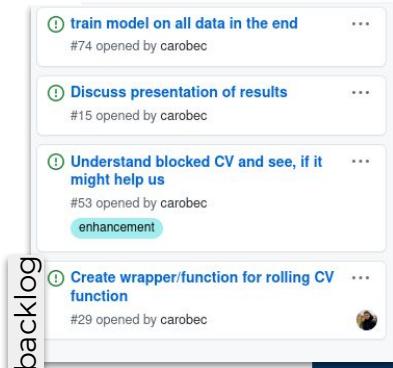


Scrum Workflow

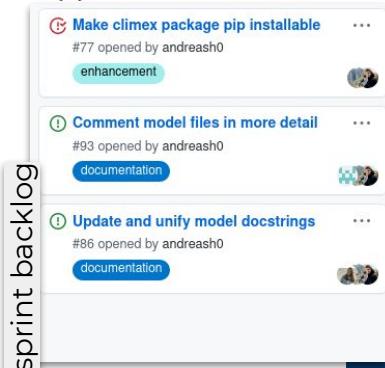
Some parts of the Scrum workflow worked very well from the beginning, while others were quickly learned and improved throughout the Innolab project



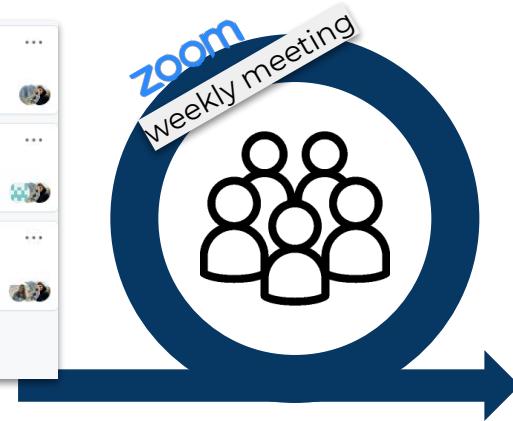
Project Board



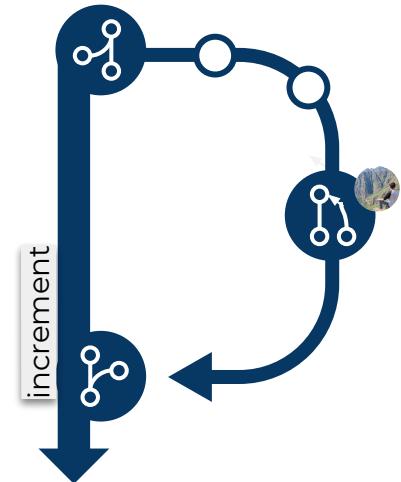
backlog



sprint backlog



-o-Git Workflow

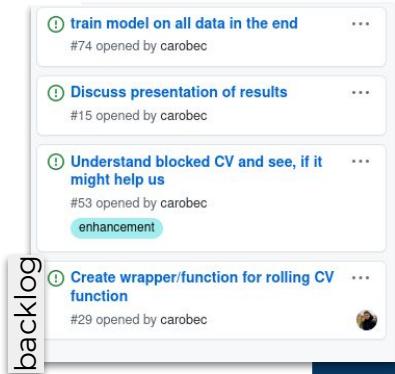


Scrum Workflow

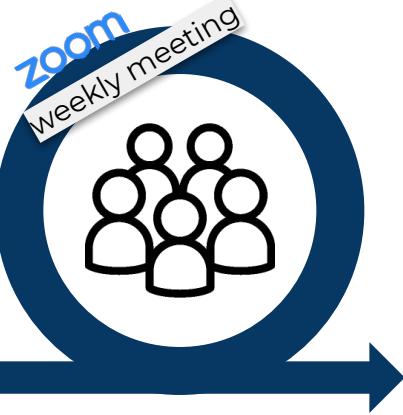
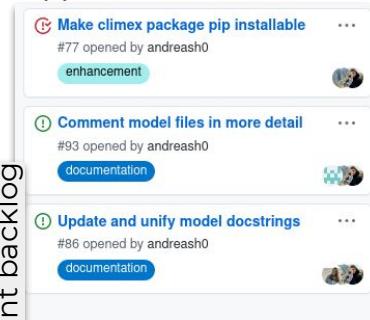
Some parts of the Scrum workflow worked very well from the beginning, while others were quickly learned and improved throughout the Innolab project

-o-Git Workflow

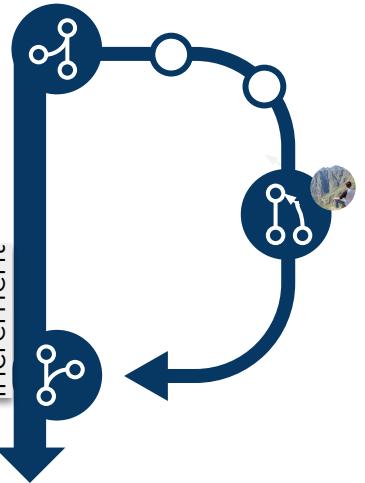
Project Board



Project Board



increment



Continuous Integration &
Peer Review



defensive checks (circle ci)



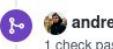
andreash0

review requests changes



henrifnk

changes uploaded



andreash0

pull request merged

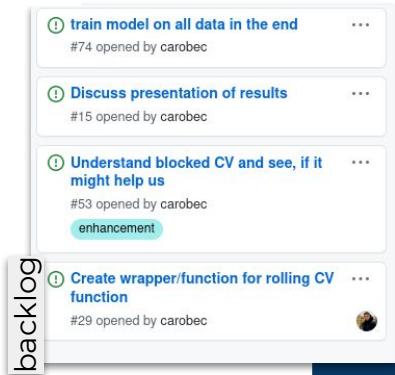
1 check passed

Scrum Workflow

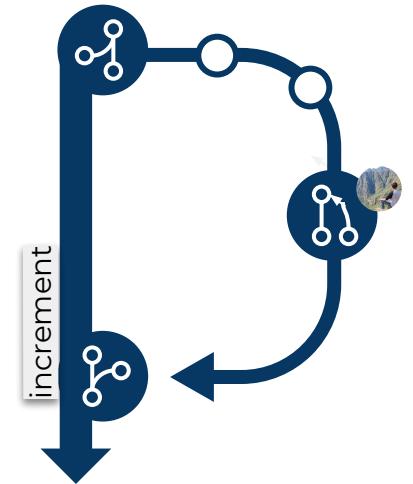
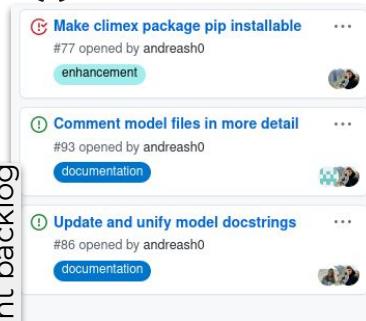
Some parts of the Scrum workflow worked very well from the beginning, while others were quickly learned and improved throughout the Innolab project

-o-Git Workflow

Project Board



Project Board



defensive checks (circle ci)

review requests changes

changes uploaded

pull request merged

Conclusions and further research

Further research and experiments could be done in the fields of the Data Pipeline, Modelling and Evaluation

Conclusions and further research

Further research and experiments could be done in the fields of the Data Pipeline, Modelling and Evaluation

Data Pipeline



Temporal information

Problem: LSTM seems not able to capture the whole temporal structure of the input data.

Approach: Add informations about long term effects to the LSTM.

Solutions:

- calculate different models per season
- embed information about the timestamp of a observation to the decoder

Conclusions and further research

Further research and experiments could be done in the fields of the Data Pipeline, Modelling and Evaluation

Data Pipeline



Temporal information

Problem: LSTM seems not able to capture the whole temporal structure of the input data.

Approach: Add informations about long term effects to the LSTM.

Solutions:

- calculate different models per season
- embed information about the timestamp of a observation to the decoder

Modelling



Vision Transformer (ViT)

Problem: Find new architectures as comparison to ResNet

Approach: Fit other models that consider spatial information.

Solutions:

- The Transformer is a new model that comes from NLP
- The scores are comparable to those from ResNet

Conclusions and further research

Further research and experiments could be done in the fields of the Data Pipeline, Modelling and Evaluation

Data Pipeline



Temporal information

Problem: LSTM seems not able to capture the whole temporal structure of the input data.

Approach: Add informations about long term effects to the LSTM.

Solutions:

- calculate different models per season
- embed information about the timestamp of a observation to the decoder

Modelling



Vision Transformer (ViT)

Problem: Find new architectures as comparison to ResNet

Approach: Fit other models that consider spatial information.

Solutions:

- The Transformer is a new model that comes from NLP
- The scores are comparable to those from ResNet

Evaluation



Probability detection

Problem: hard label classification might lead to very insecure predictions.

Approach: predict class probabilities for each observation.

Solutions:

- Visualize the probabilities in time to detect areas with high uncertainty
- Try to fit a DL model with a certainty threshold



Thank You For Your Attention

