UNIVERSITÄT BERN
OESCHGER CENTRE CLIMATE CHANGE RESEARCH

Machine learning-based recognition of hailstorm events using the hail reports collected by the MeteoSwiss smartphone application

Andrey Martynov

Institute of Geography and Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland

Bern Winter School on Machine Learning 2019

Motivation

- Summer hailstorms over Switzerland cause considerable damage to the property, crops, real estate, etc.
- In Switzerland there are almost no ground-based hail observations
- Main source of information about hailstorms: MeteoSwiss radars
- Independent observations would be useful for validating radar data, etc.
- MeteoSwiss application: > 50,000 reports since 2015. **Can they be used as "ground truth" data? Can they be used alone for detecting hailstorms?**

Bilder: 20 minuten, Leser-Reporter

UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

Research framework

Research questions:

- How do hailstorm reports correlate with the MeteoSwiss radar data?
- Can hail report data alone be used for determining the time and location of hailstorms with accuracy, comparable with that of MeteoSwiss radar data?

We use the machine learning approach by training a neural network with hail reports, using MeteoSwiss radar data as validation. The value of validation accuracy that can be reached reflects the link between radar and "ground" observations.

This is a pilot project, realized within the framework of the Bern Winter School on Machine Learning, but it will be (most probably) continued as part of our normal research activity.

Source data

UNIVERSITÄT
BERN
OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

- Available data:

  - gridded MeteoSwiss radar observations: 2.5 minutes.

    710*640, 1 km res, CH1903. NetCDF format

    Field: Probability of Hail (POH). Usual threshold: 80%

  - Hail reports sent from the MeteoSwiss application for smartphones:

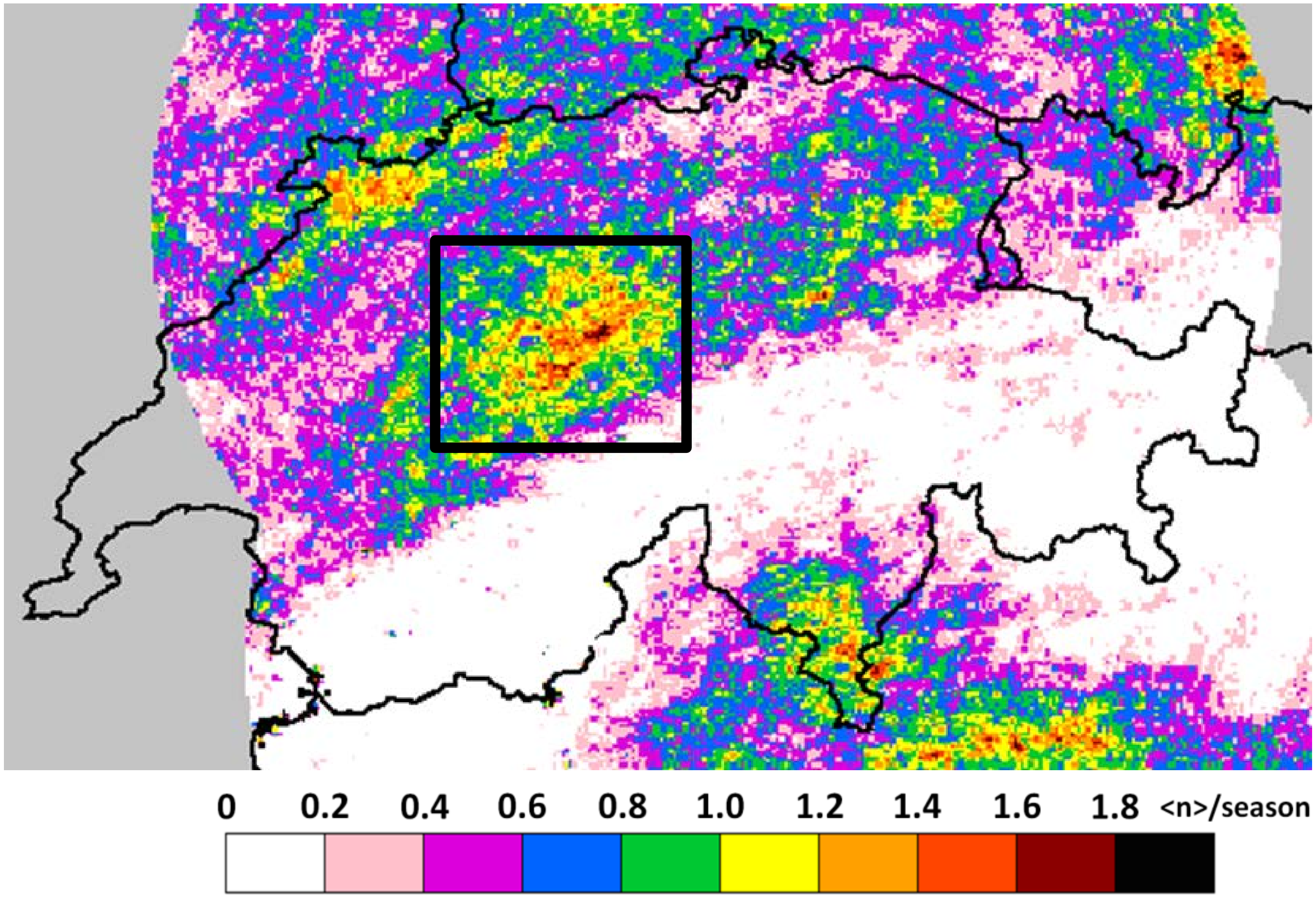
    2015-2018, in total: 58659 reports.

    Fields: - anonymous ID
    - Coordinates in CH1903
    - Hail size (5 categories from "no hail" to largest, changed in 2018)
    - Event date and time, submission date and time, etc.

Study domain

UNIVERSITÄT BERN
OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

Domain of interest: the hot spot in Berner Prealps, 50 x 50 km



Multiannual hail climatology according to Nisi et al. 2014

Data preparation I

UNIVERSITÄT
BERN
OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

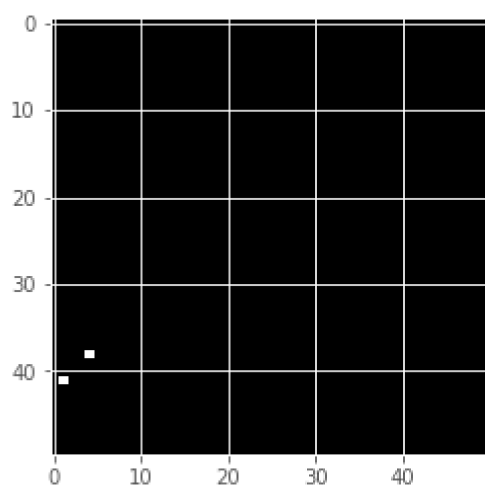
Data between Julian days 100 and 300 have been retained (warm period)

Hail reports:

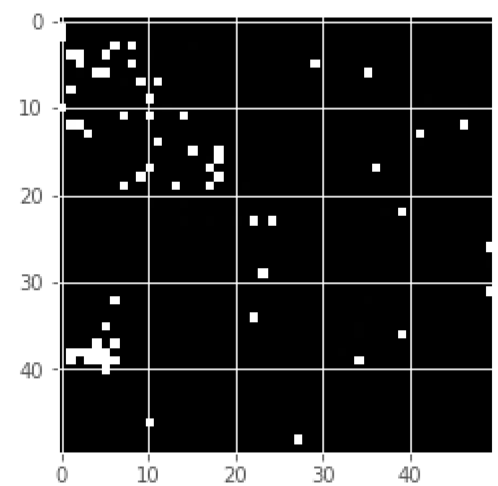
- Only reports within the study domain (shell script)
- "No hail" and largest size: filtered out (shell script)
- Hourly and daily 50x50 PNG files created out of reports:

  black if no hail, white if hail reported (Jupyter Notebook)
- PNG files converted to JPG (shell script)

"No hail"



"Hail"



Data preparation II

Data between Julian days 100 and 300 have been retained (warm period)

Radar data:

- Only data within the study domain (shell script + cdo (Climate Data Operators))
- Hourly and daily aggregations of 2.5 minute snapshots (shell + cdo)
- For all hours and days, when hail was reported:
  - maximum value of POH within these periods retained (shell script)

If for a given period, when hail was reported, the POH exceeded 80%: "Hail" period

Else: "No hail".

Hourly data: 1363 "no hail" hours, 143 "hail" hours

Daily data: 389 "no hail" days, 67 "hail" days

Most works have been performed using daily data.

Neural network

UNIVERSITÄT BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

The design of the network is based on the Tutorial V design:

- distinguishing between two labels: "hail" and "no hail" ("German and Italian")
- JPG images used as input for the network
- Pre-trained model, allowing for fine-tuning of the model on hail data
- Supervised learning
- Initial structure of the network: fully connected, 2 layers (512/2, sigmoid/softmax)

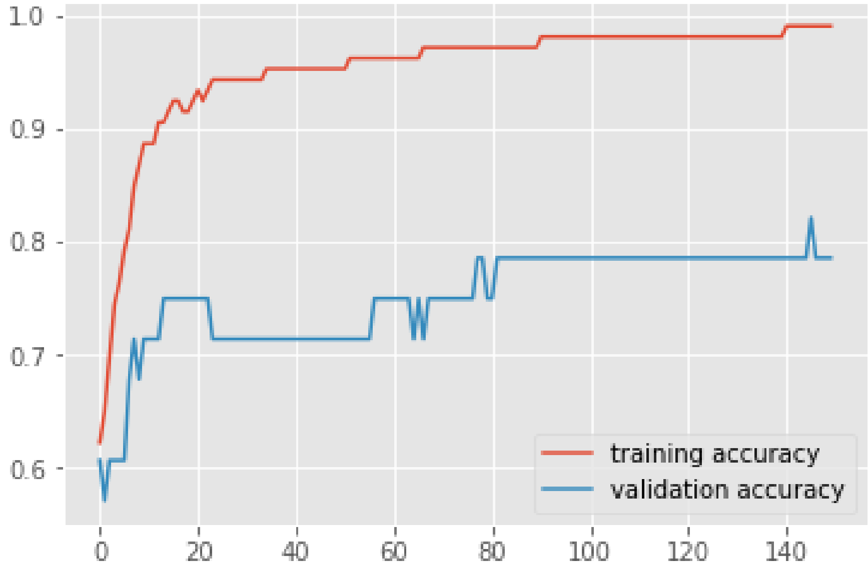
- Tests and variations:
    - Modified parameters in the initial model: N1, learning rate
    - Additional fully connected layers

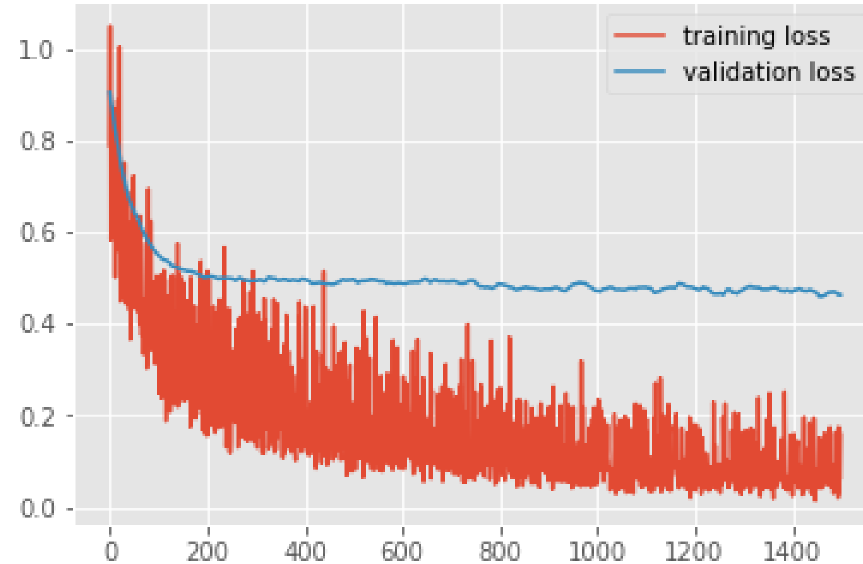
Computations: 64-CPU server *climcal4* of GIUB.

Neural network performance

A typical example: learning rate=0.00001, initial TIV network

(fully connected, 2 layers (512/2 neurons, sigmoid/softmax)





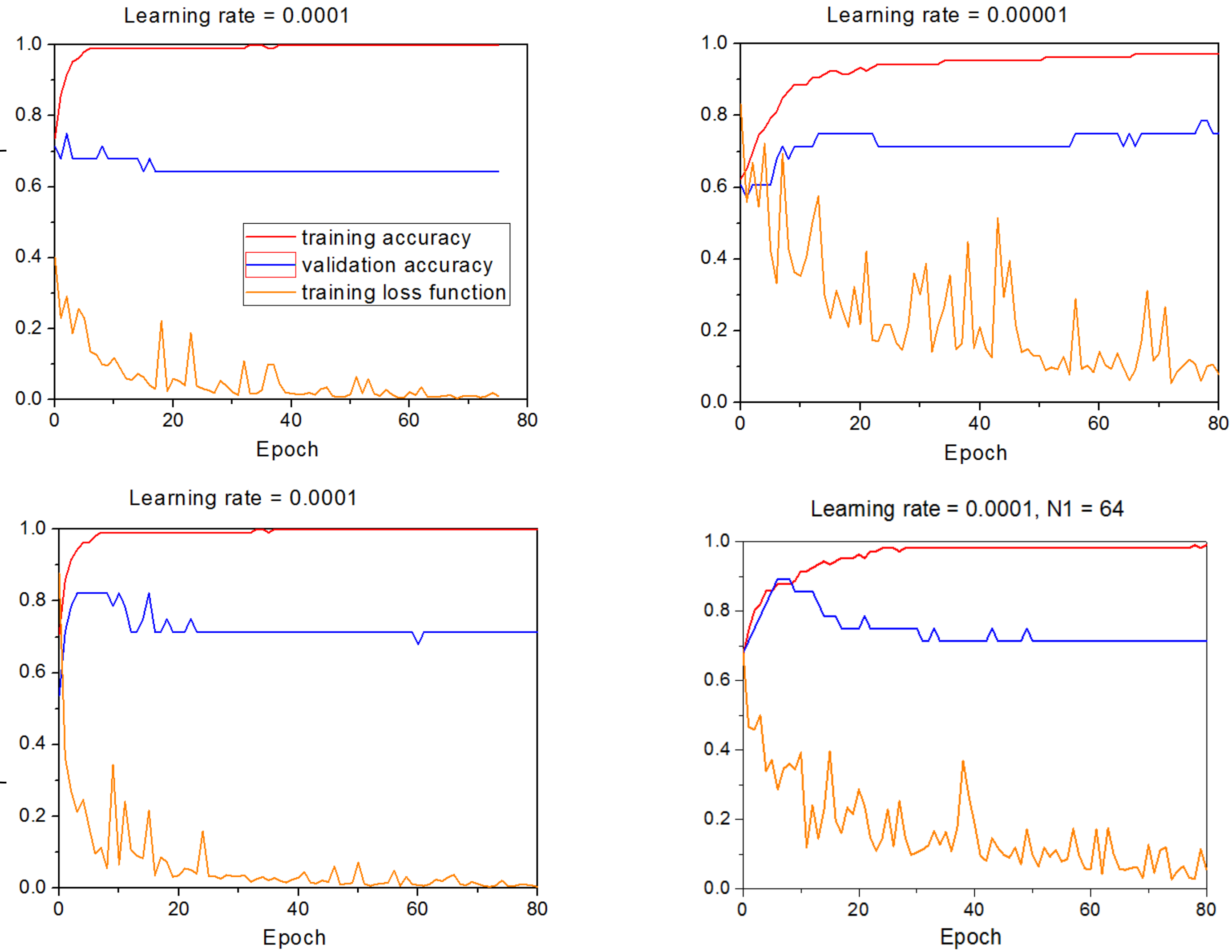
The best validation accuracy values are ~0.8, i.e. in this case.

UNIVERSITÄT BERN
OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

Sensitivity I: 1 FC layer

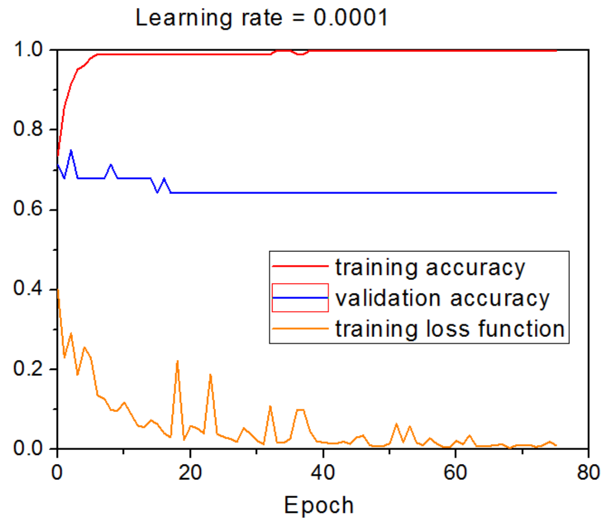
- Higher learning rate or higher N1: more rapid saturation of the training accuracy, but often worse validation accuracy.

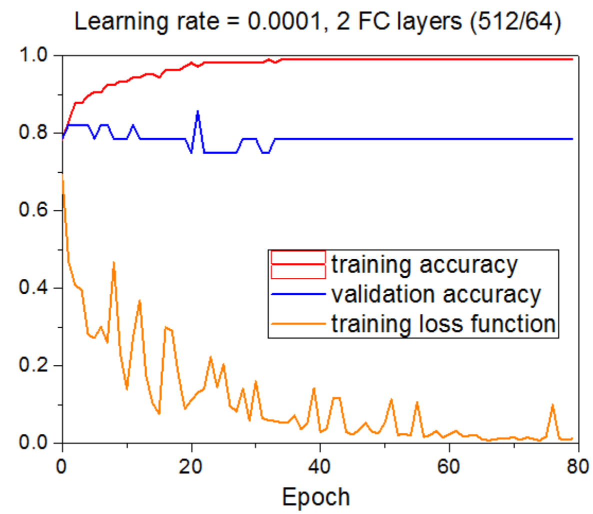
Two realisations

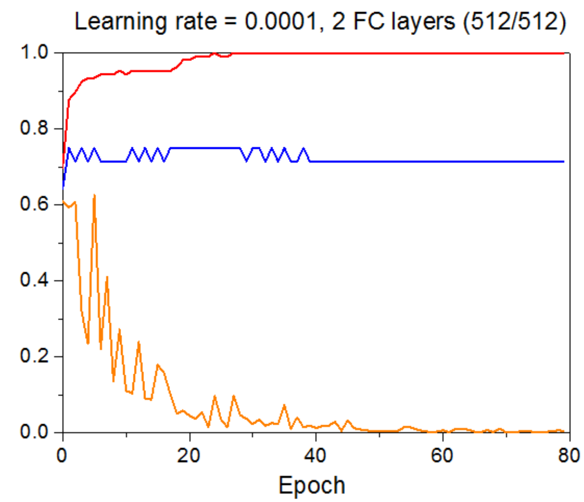


Sensitivity II: 1 FC layer

- 2 FC layers: slow convergence of training accuracy, saturation of validation accuracy







Manual check of the validation efficiency:

Out of 389 "no hail" days: 329 correct (84.5%)

(only 67 "no hail" days were used in training)

Out of 67 "hail" days: 62 correct (92.5%)

Discussion

UNIVERSITÄT BERN

OESCHGER CENTRE CLIMATE CHANGE RESEARCH

- Choice of the domain: large in comparison with a hailstorm, small enough for not containing many events in parallel;

- Reported hail size not taken into account; only the presence or absence of hail is examined;

- Daily data: allow for formation "swaths" of reports along the trajectories of hailstorms - distinguishable patterns.

- The "internal" variability of results, caused by random permutations of training data, is considerable and is comparable to sensitivity to model parameters.

- Maximum validation accuracy: 0.8.

Conclusion: the detection of hailstorms with a neural network, using only crowd-sourced hail reports, matches with the radar-based predictions in ~85-90% of cases. A good omen for further studies.