Short-term PV power output prediction using convolutional neural network: learning from an imbalanced sky images dataset via sampling and data augmentation

Yuhao Nie[1], Ahmed Zamzam[2], Adam Brandt[1]

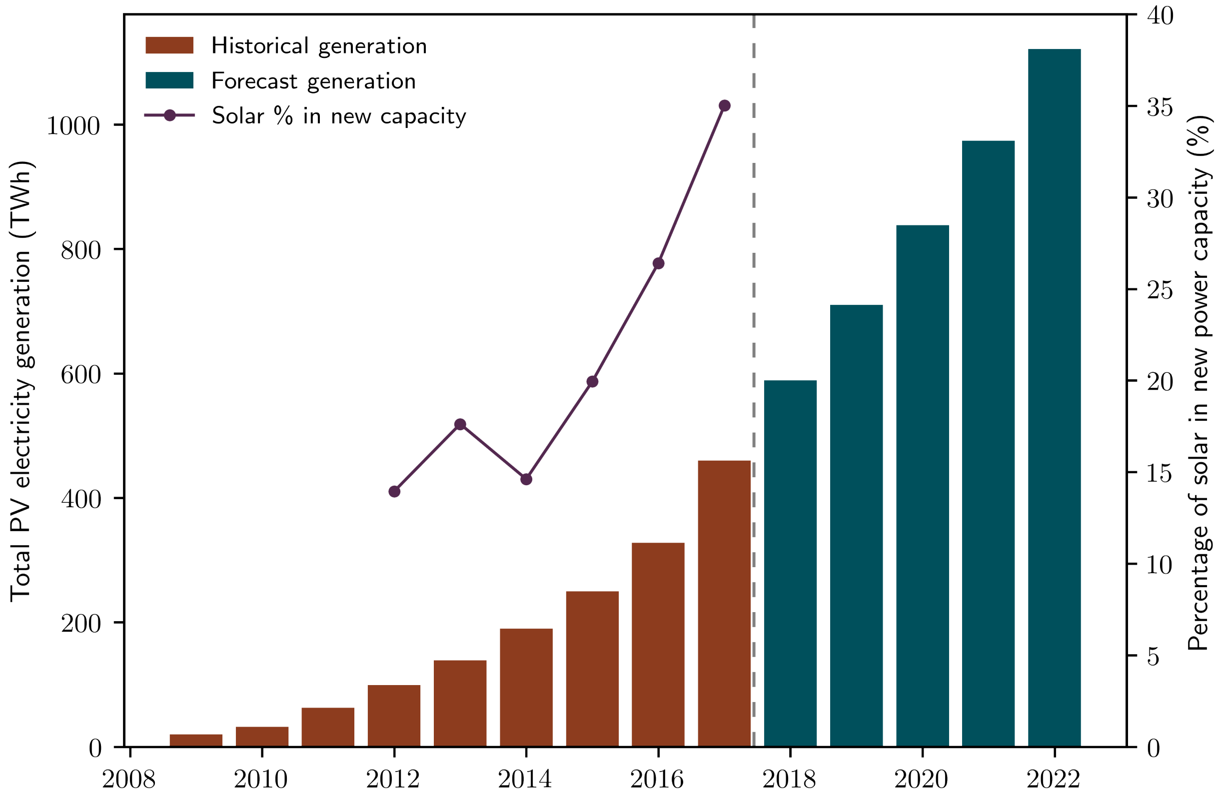
[1] Department of Energy Resources Engineering, Stanford University

[2] National Renewable Energy Laboratory

NeurIPS 2020 Workshop
Tackling Climate Change with Machine Learning
December 11, 2020

Decarbonize global energy supply and solar PV growth

- Global energy supply accounts for 25% of global greenhouse gas (GHG) emissions
- Integrating renewables is promising to reduce the GHG emissions from power generation



Solar photovoltaic (PV)

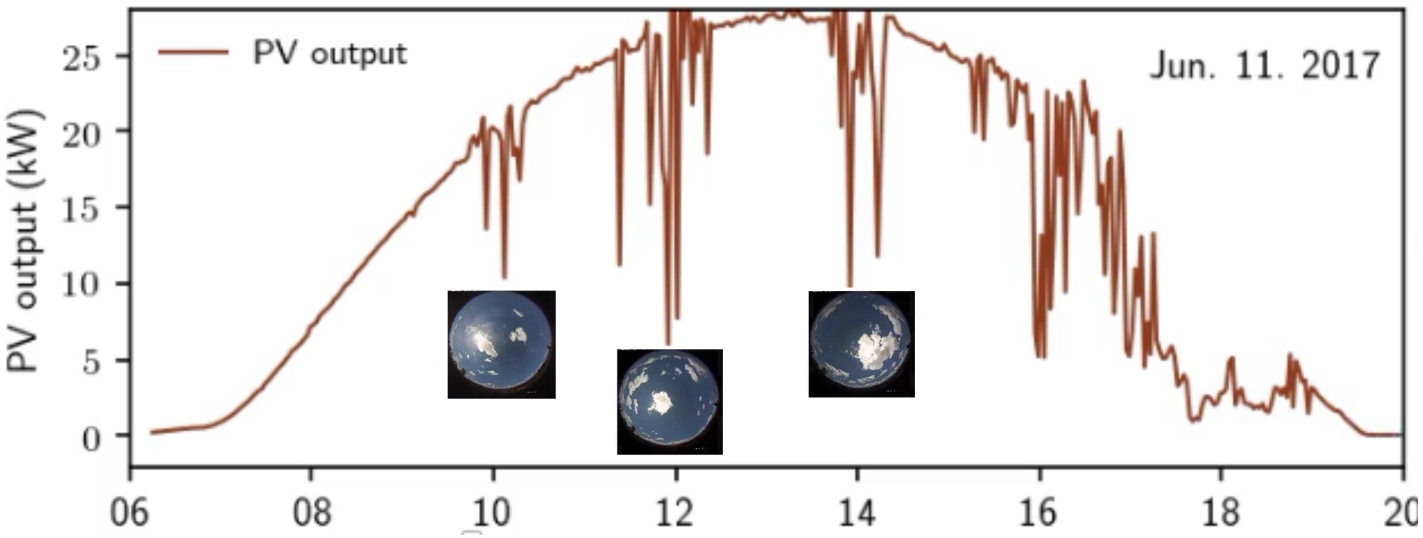
- 36% compound annual growth rate in **power generation** for past 5 years
- Accounts for 35 % of global annual addition to **power capacity** in 2017, more than fossil and nuclear combined

1. IEA. World energy outlook (weo) 2017. Technical report, IEA Paris, France, 2018
2. REN21. Renewables 2018 global status report. Technical report, Renewable Energy Policy Network for the 21st Century, 2018.

PV integration challenged by solar intermittency

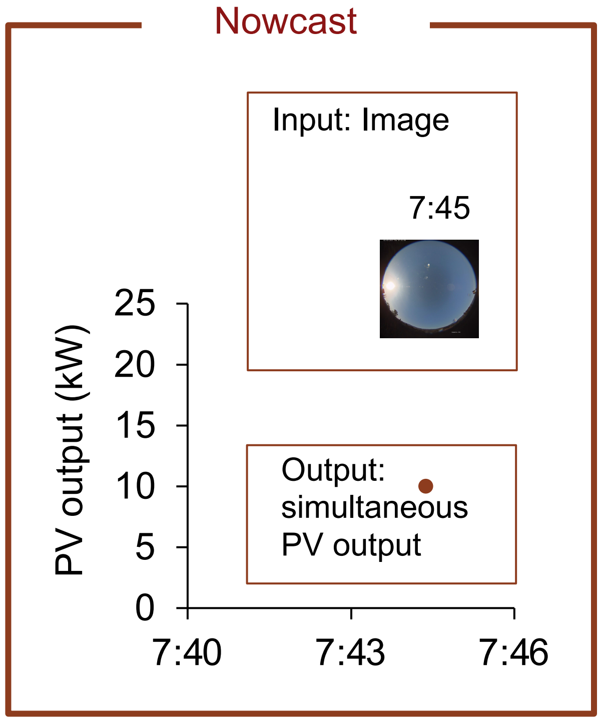
- Large-scale PV integration is challenged by solar intermittency
- 70%~80% power loss within in a few minutes in partly cloudy day
- Strong fluctuation in power generation is mainly caused by short-term cloud events
- The need for accurate and reliable power forecasting, especially under cloudy conditions

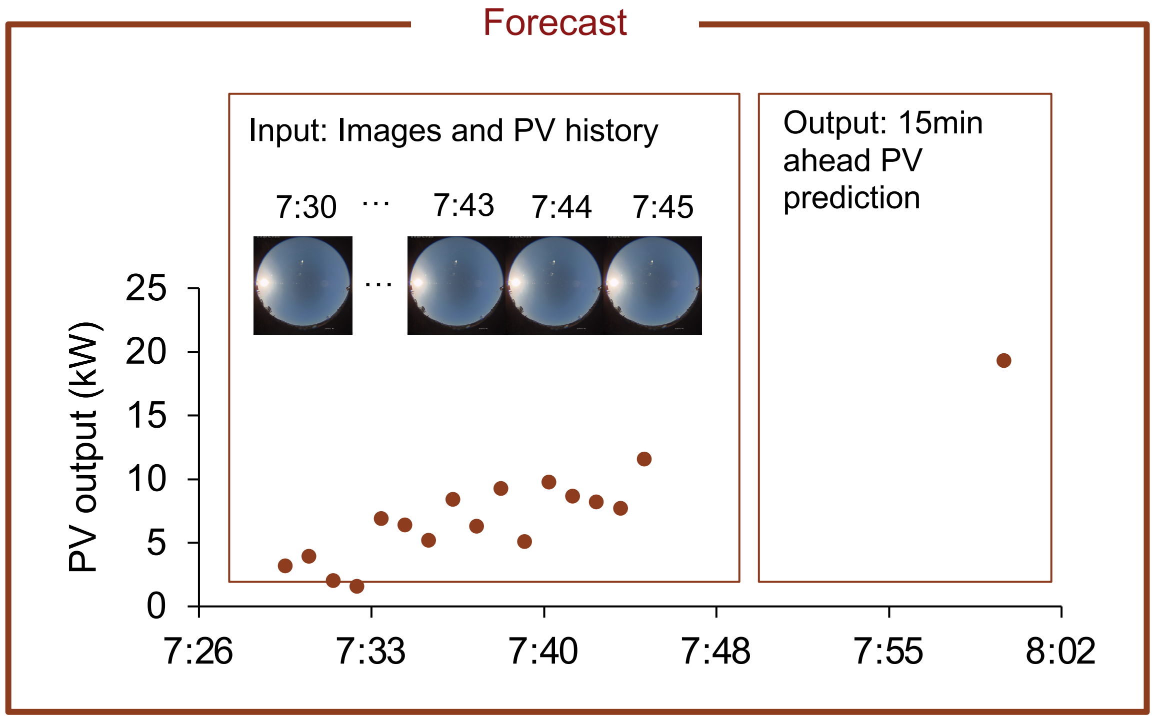
PV output profile for a 30-kW rooftop PV system on Stanford campus



Problem formulation: Nowcast vs Forecast

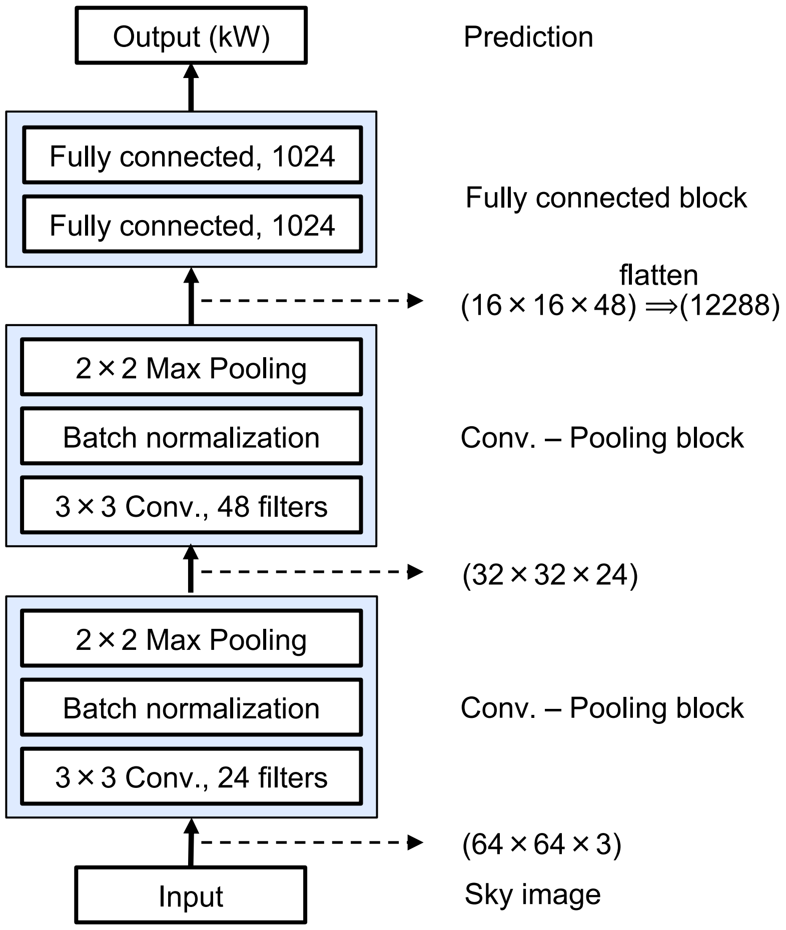
- Nowcast – Given sky images, predict concurrent PV output
- Forecast – Given sky images and PV output history, predict PV output 15 min ahead into the future



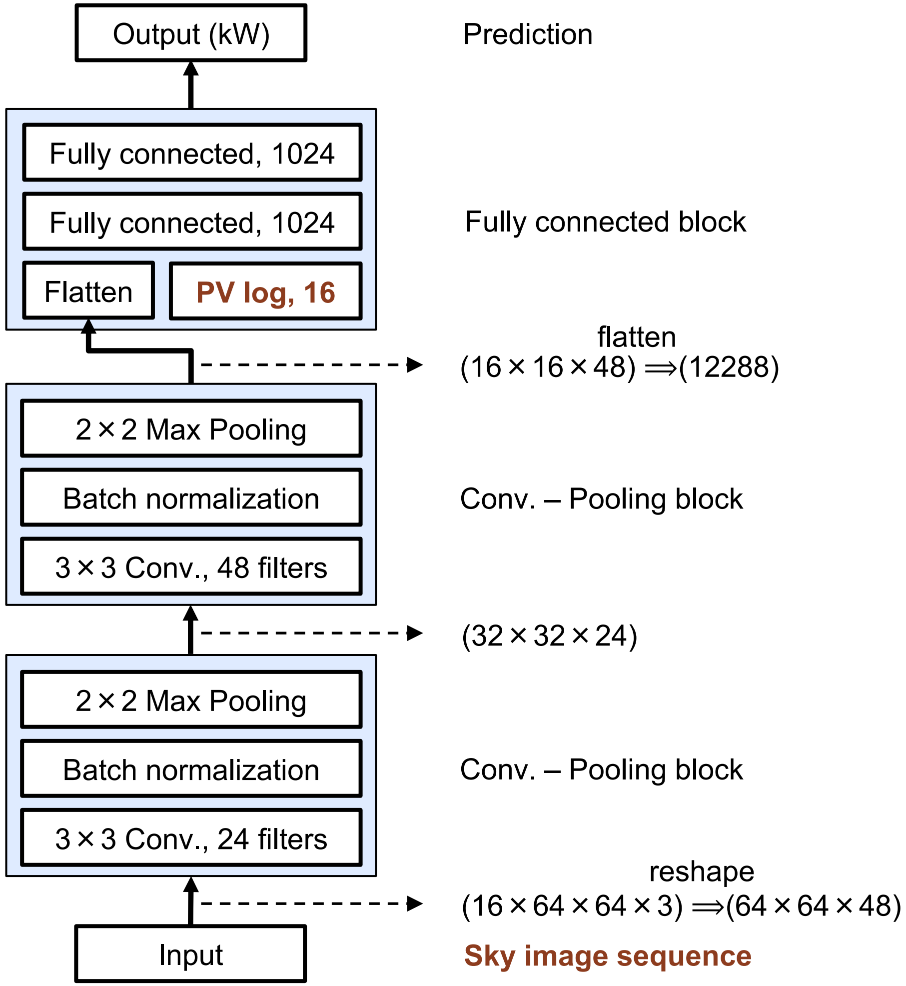


Baseline model Architecture

SUNSET - Nowcast



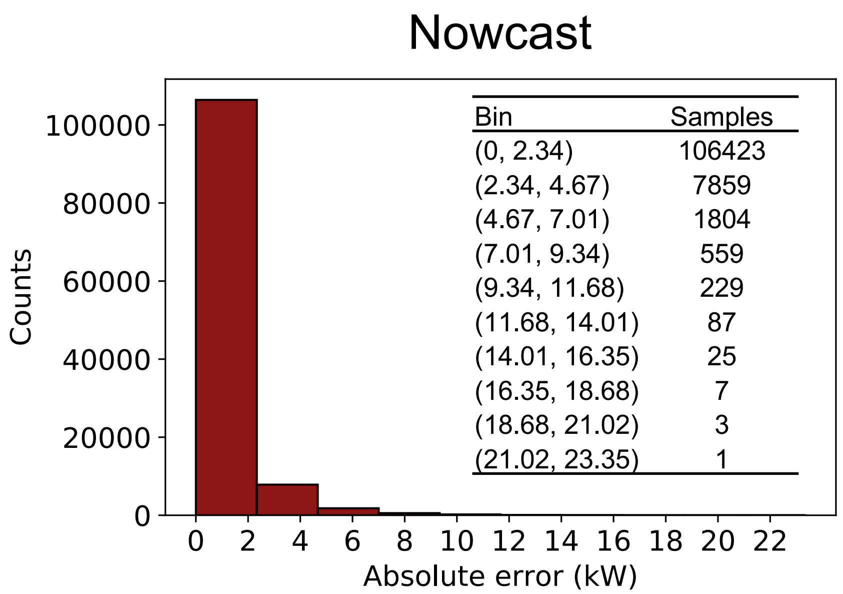
SUNSET - Forecast

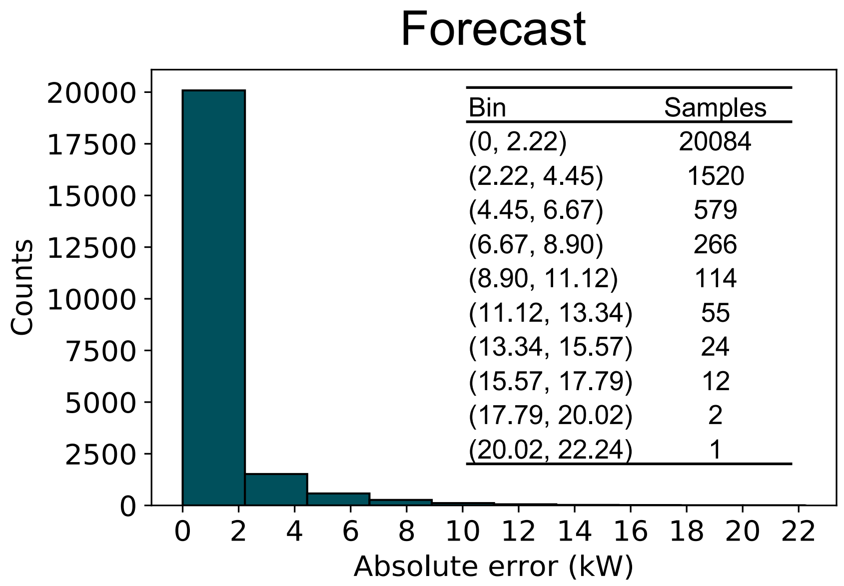


Dataset

1. Sky images and PV output dataset
   - Entire dataset: Mar. 2017 to Nov. 2019
     - Data frequency: 1 min
     - Sky images: 64×64
     - PV data: 30 kW system, ~125 m from the camera

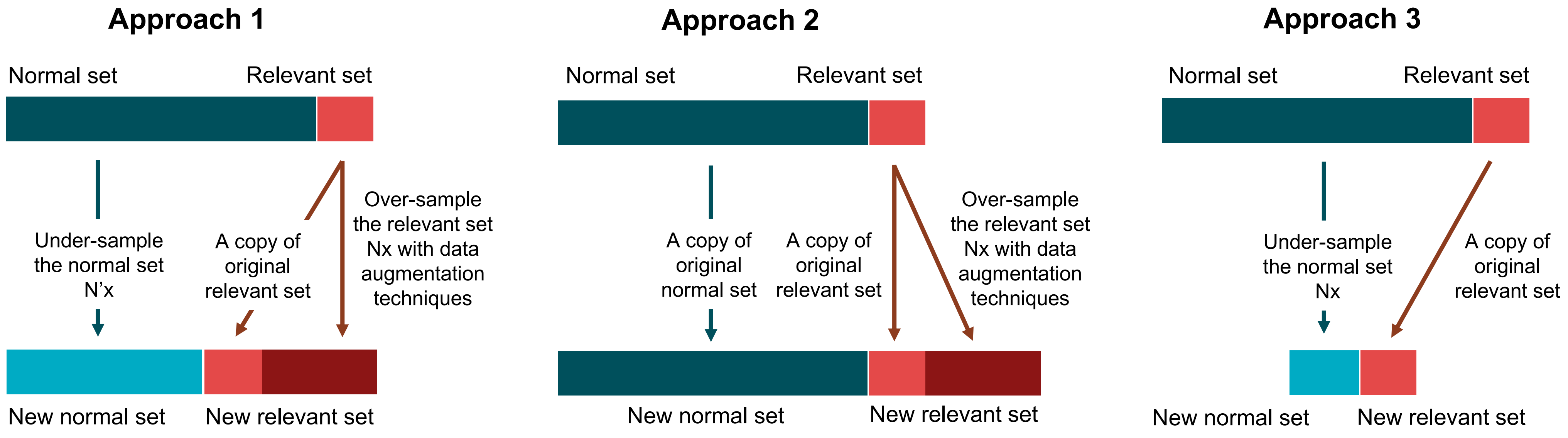
2. Partition of dataset for model development and test
   - Test set
     - 20 days from the entire dataset (10 sunny days + 10 cloudy days)
   - Development set
     - For nowcast, 36% random samples
     - For forecast, 18% random samples
     - Imbalance
       - Normal set: easy samples (mostly sunny): 90%
       - Relevant set: hard samples (cloudy): 10%





Methodology

- Sampling approaches



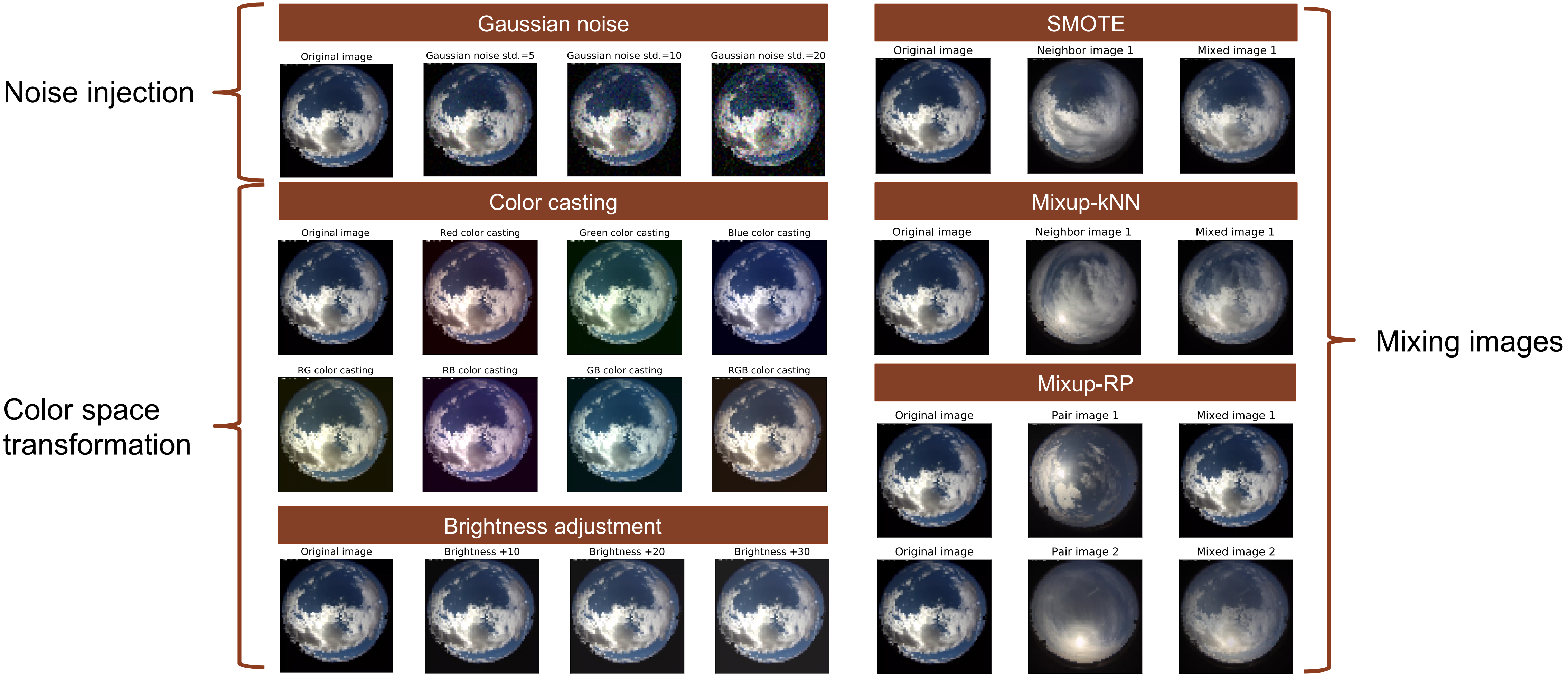
- Under-sample normal set
- Over-sample relevant set
- Keep dataset size unchanged

- Copy normal set
- Over-sample relevant set
- Increase dataset size

- Under-sample normal set
- Copy relevant set
- Reduce dataset size

Methodology

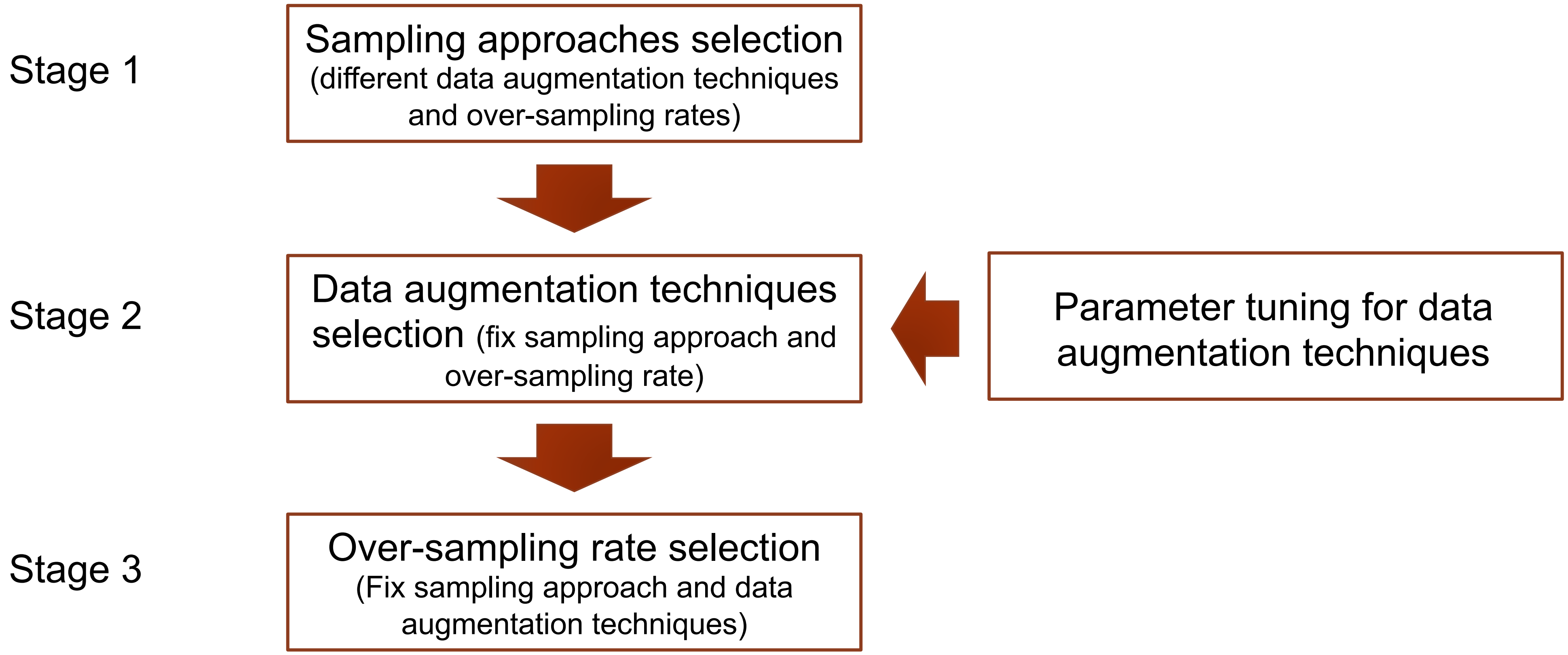
- Data augmentation techniques



Stanford University

Experiments and results

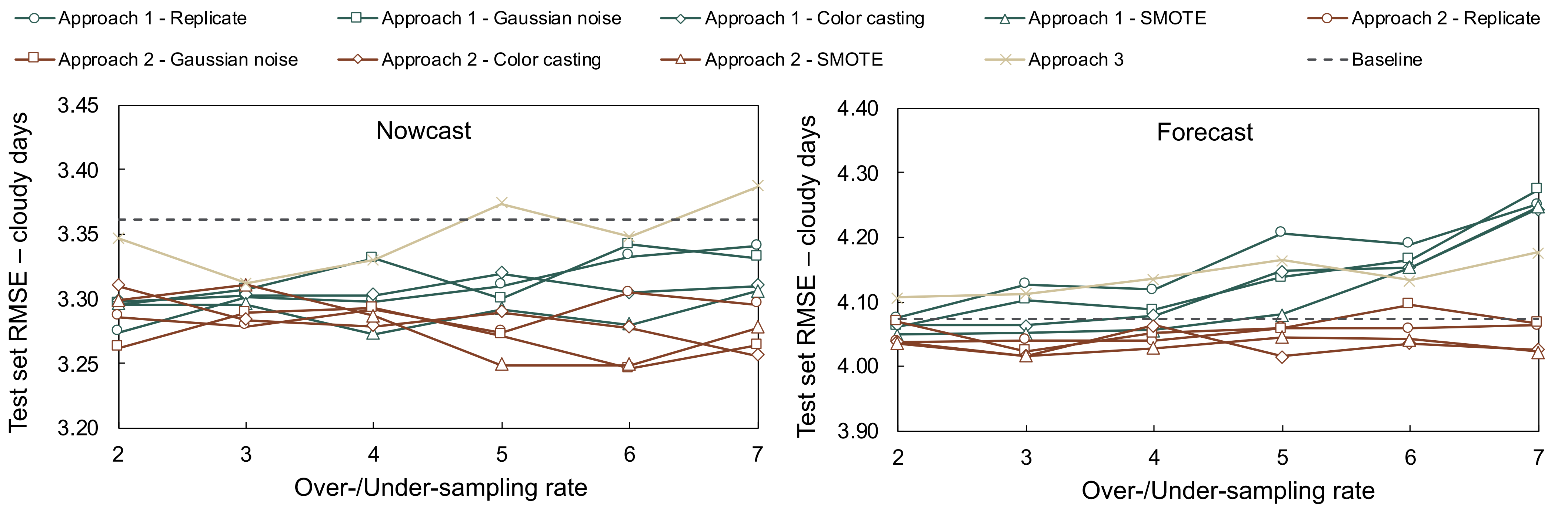
- Experiment design – three-stage selection process



Stanford University

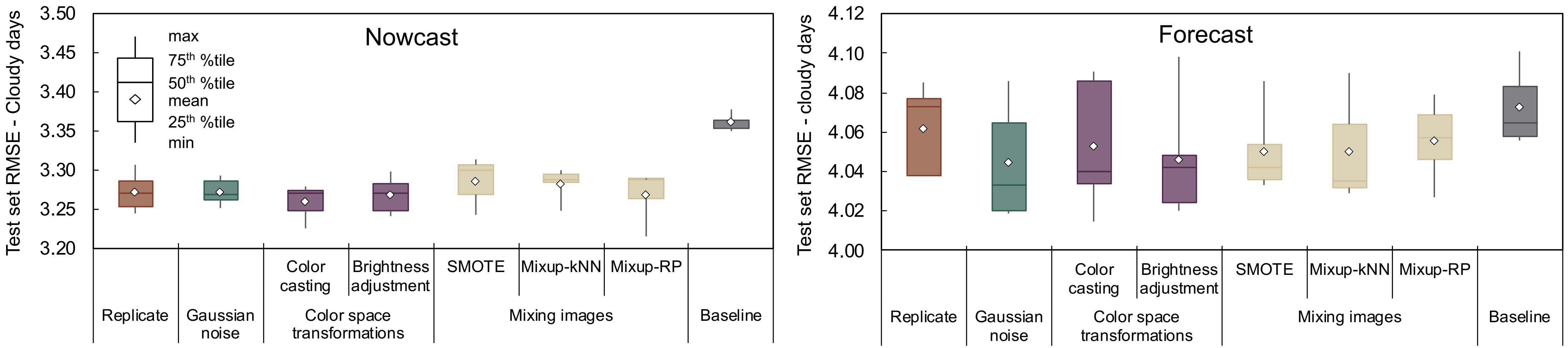
Experiments and results

- Optimal sampling approach selection
  - Sampling approach 2 is consistently better for both nowcast and forecast models



Experiments and results

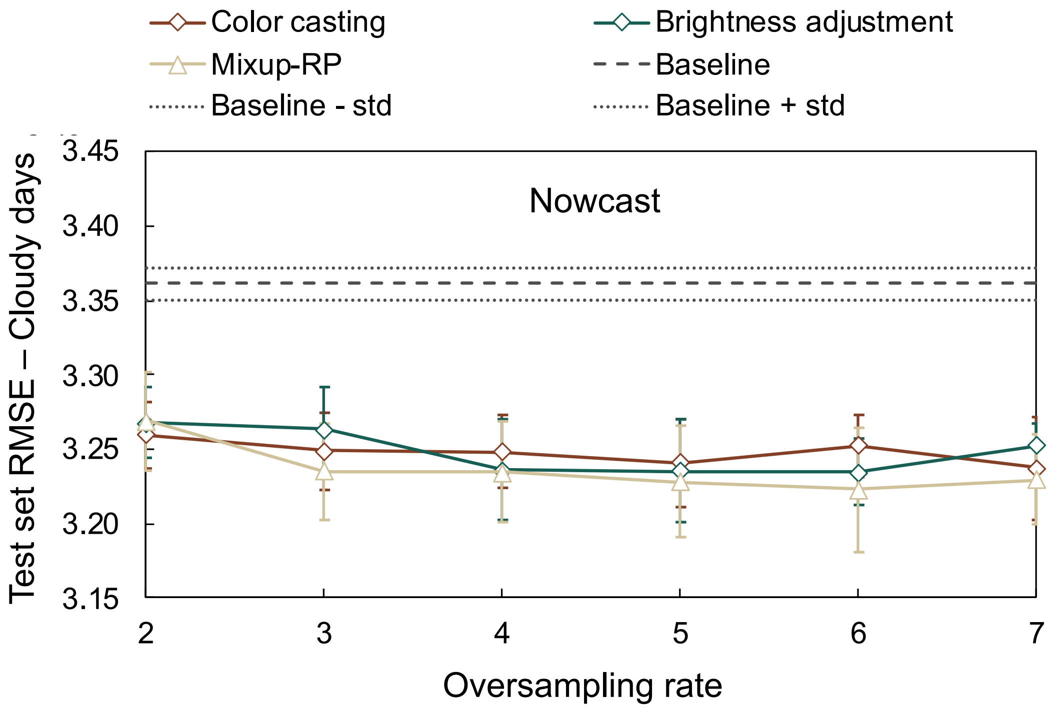
- Top-3 data augmentation techniques
  - Nowcast: Color casting, Brightness adjustment and Mixup-RP
  - Forecast: Gaussian noise, Brightness adjustment and Mixup-kNN

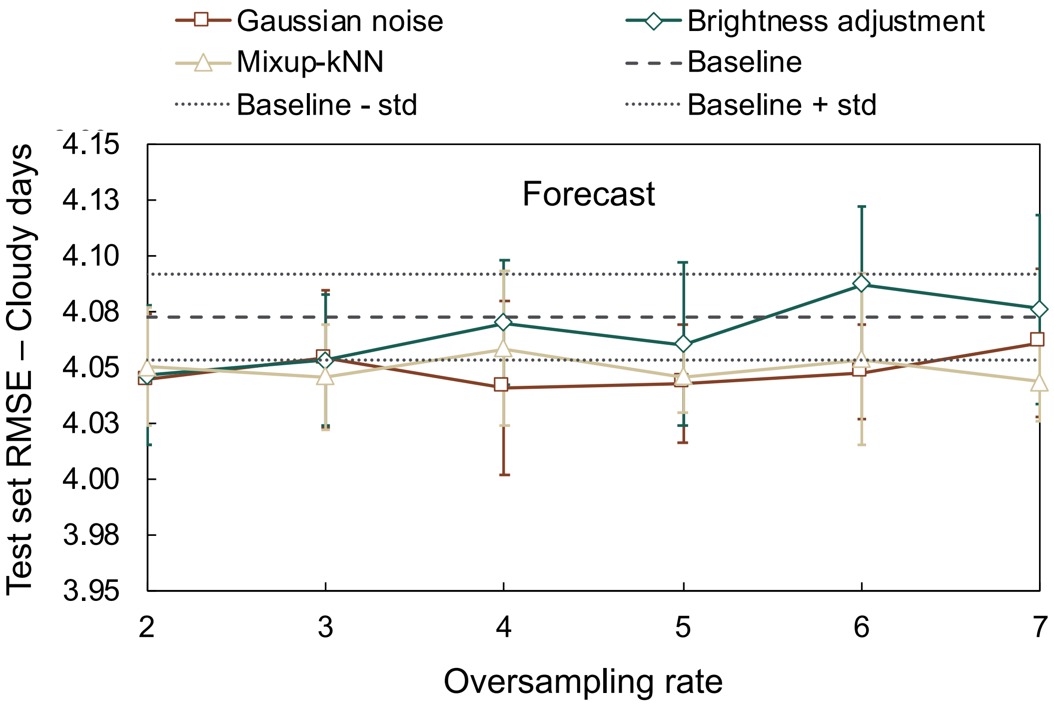


Stanford University

Experiments and results

- Optimal oversampling rate selection
  - For nowcast, increasing oversampling rate can help improve the model performance
  - For forecast, no close relationship found between the oversampling rate and the model performance





Summary

- This study examines the efficacy of applying different sampling and data augmentation approaches to an imbalanced sky images dataset for two PV output prediction tasks
- For nowcast, sampling and data augmentation can effectively enhance the model performance
- For forecast, sampling and data augmentation only improve the model performance limitedly
- Oversampling by expanding the original imbalanced dataset is the best among the three studied sampling approaches
- Increase oversampling rate can help improve the nowcast model performance but not for the forecast model
- Augmentation techniques need to be re-evaluated for application in other tasks with imbalanced images dataset