The Influence of Grouping Spatio-Temporal Features on XAI: Case Study with FogNet, a 3D CNN for Coastal Fog Prediction

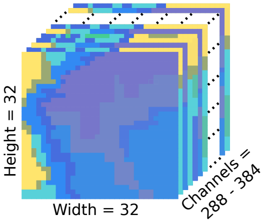
Evan Krell, Hamid Kamangir, Waylon Collins,
Scott A. King, Philippe Tissot

ai2es.org

This material is based upon work supported by the National Science Foundation under award 2019758
Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

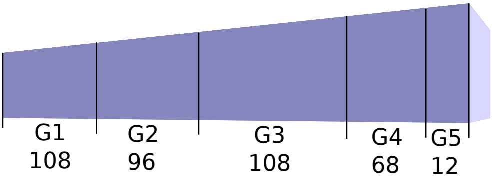
Outline

FogNet: 3D CNN for Coastal Fog Forecasting



- 3D convolutional neural network (CNN) with attention, dense block, & dilated convolution [1]
- High performance: beats NOAA's operational `High Resolution Ensemble Forecast (HREF)`
- Input data: spatio-temporal raster of metocean variables
- https://gridftp.tamucc.edu/fognet/

Physics-based channel groups



G1 wind
G2 turbulence kinetic energy & humidity
G3 lower atmosphere thermodynamic profile
G4 surface atmospheric moisture & microphysics
G5 sea surface temperature

EXplainable Artificial Intelligence (XAI)

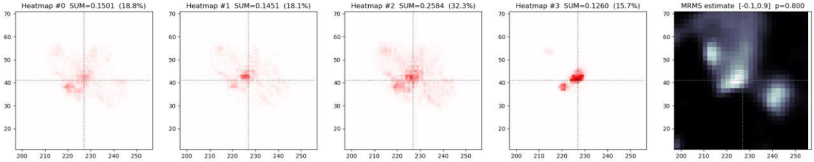
Motivating Scenario

- Forecaster has tools & models they rely on and understand
- Their tools suggest **no fog**
- Researcher's new model (performs well on test data): **yes fog**
- I suspect that the forecaster will want to know *why?*
  $\rightarrow$ what information is the model's decision based on?

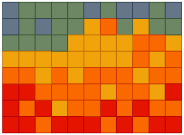
Post-hoc XAI methods

- Trained model $\rightarrow$ how do the input predictors influence model performance?
- **Feature importance:** rank predictors based on their influence on performance
- **Feature effect:** how much does each predictor contribute to a model decision

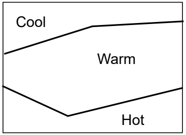
Example (Hilburn et al. [2])



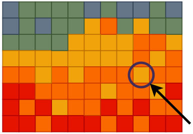
XAI: Challenges with Correlated Features



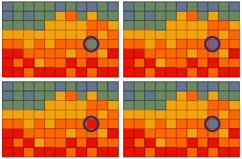
1. Input raster



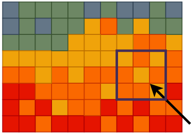
2. Matches learned feature



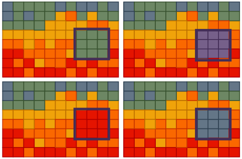
3. What is the influence of this cell on performance?



4. XAI via Feature Replacement: single-cell change still matches feature --> minimal impact on performance

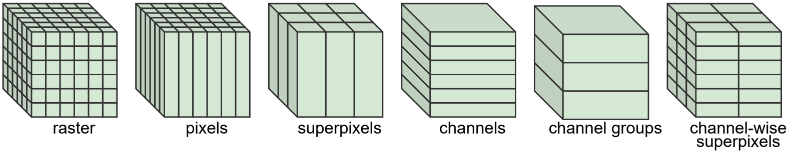


5. What is the influence of this superpixel?



6. Replacing larger region --> break up learned feature --> could change model decision

Some feature grouping schemes



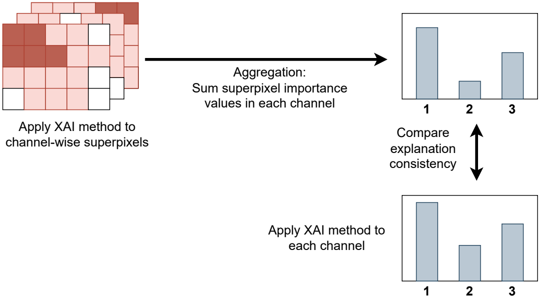
More granular groups

- More specific explanations
- Spatio-temporal autocorrelation problem

Coarser groups

- Potentially vague insights
- May join correlated values

Idea: analyze consistency among explanations to guide XAI interpretation



Methods: Feature Importance

- Global methods → how did feature influence model performance?
  - **Permutation Feature Importance (PFI):** replace feature with permuted values [3]
  - **LossSHAP (LS):** approximate Shapley values . . . combinatorial complexity [4]
  - **Group-hold-out (GHO):** entirely remove feature & retrain model [5]

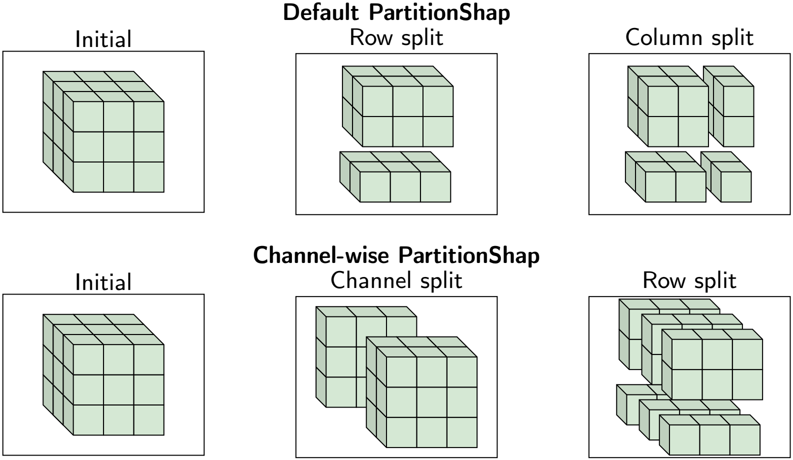
PFI (tabular example)

Height at age 20 (cm)	Height at age 10 (cm)	...	Socks owned at age 10
182	155	...	20
175	147	...	10
...	...	...	...
156	142	...	8
153	130	...	24

Image from [6]

Methods: Feature Effect

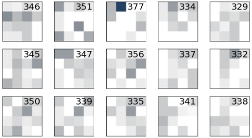
- Local methods → how did feature influence specific model decision?
  - **Channel-wise PartitionSHAP (CwPS):** approximate Shapley values for superpixels within each raster channel [7]



Our SHAP fork with CwPS:

github.com/conrad-blucher-institute/partitionshap-multiband-demo

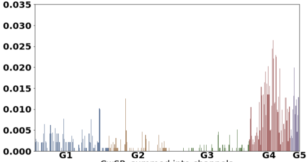
Results: Feature Importance



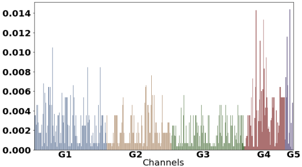
Top 15 channel-wise superpixels

XAI on 3 feature grouping schemes:

1. Channel groups (physics-based)
2. Channel-wise
3. Channel-wise superpixels (CwSP)

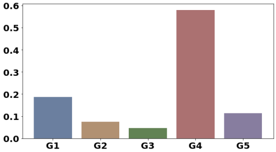


CwSP, summed into channels

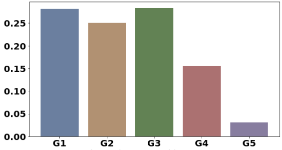


Channels

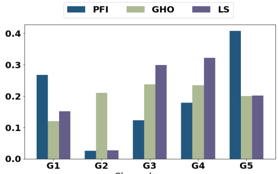
Column: XAI method applied
Row: level of aggregation



CwSP, summed into groups



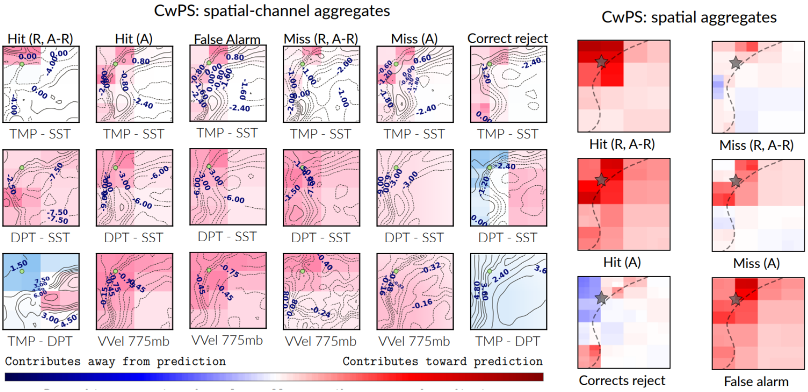
Channels, summed into groups



Channel groups

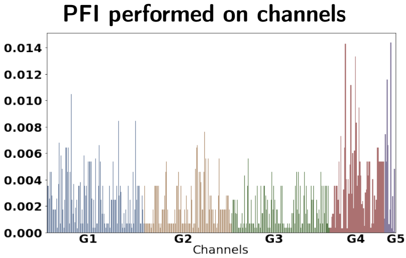
Results: Feature Effect

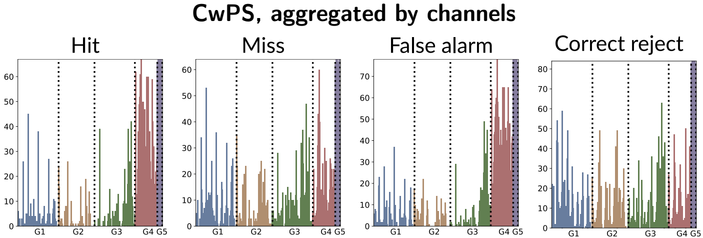
- **Channel-wise PartitionSHAP**: computed **293** local explanations
- 67 hits, 64 misses, 78 false alarms and 84 random correct rejects.
- Hit/miss fog types: *Advection* **(A)** and *Radiation, Advection-Radiation* **(R, A-R)**



Results: Comparing Feature Importance & Effect

- Feature importance: which features improve performance?
- Feature effect: which features are used for decisions?
- Group 5 has dominant effect, but less importance
  $\rightarrow$ influence both correct and incorrect outcomes



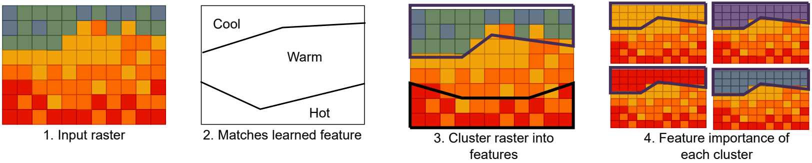


Conclusions & Future Work

- Consistencies emerge, but explanations sensitive to feature grouping
  - Removing a superpixel may not remove enough of the information to change a forecast, where removing the whole channel (or group) does
- Strong influence from coastline, especially near KRAS (target location)
  $\rightarrow$ perhaps point-based models effective (i.e. autoencoder)?
- Group 5 discrepancy explained?
  - Feature importance: moderate importance
  - Feature effect: dominating channels
  - But since **G5** channels used also for misses & false alarms
    $\rightarrow$ expected to lower model performance

Future Work

Data-driven feature groups



But in 3D, and we have interleaved mix
of temporal (time steps) and spatial (altitudes)

References

1. Kamangir, Hamid, et al. "FogNet: A multiscale 3D CNN with double-branch dense block and attention mechanism for fog prediction." Machine Learning with Applications 5 (2021): 100038.
2. Hilburn, Kyle A., Imme Ebert-Uphoff, and Steven D. Miller. "Development and interpretation of a neural-network-based synthetic radar reflectivity estimator using GOES-R satellite observations." Journal of Applied Meteorology and Climatology 60.1 (2021): 3-21.
3. McGovern, Amy, et al. "Making the black box more transparent: Understanding the physical implications of machine learning." Bulletin of the American Meteorological Society 100.11 (2019): 2175-2199.
4. Covert, Ian, Scott Lundberg, and Su-In Lee. "Feature removal is a unifying principle for model explanation methods." arXiv preprint arXiv:2011.03623 (2020).
5. Au, Quay, et al. "Grouped feature importance and combined features effect plot." Data Mining and Knowledge Discovery 36.4 (2022): 1401-1450.
6. https://www.kaggle.com/code/dansbecker/permutation-importance
7. Kamangir, Hamid, et al. "Importance of 3D convolution and physics on a deep learning coastal fog model." Environmental Modelling & Software (2022): 105424.