Increasing the Robustness of Model Predictions to Missing Sensors in Earth Observation

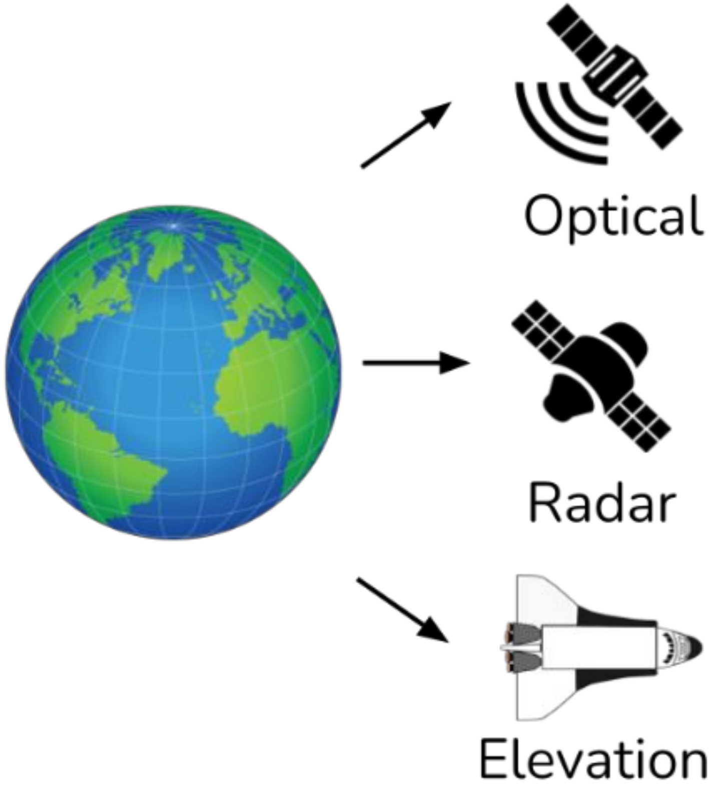
Francisco Mena, Diego Arenas, Andreas Dengel

University of Kaiserslautern-Landau | DFKI

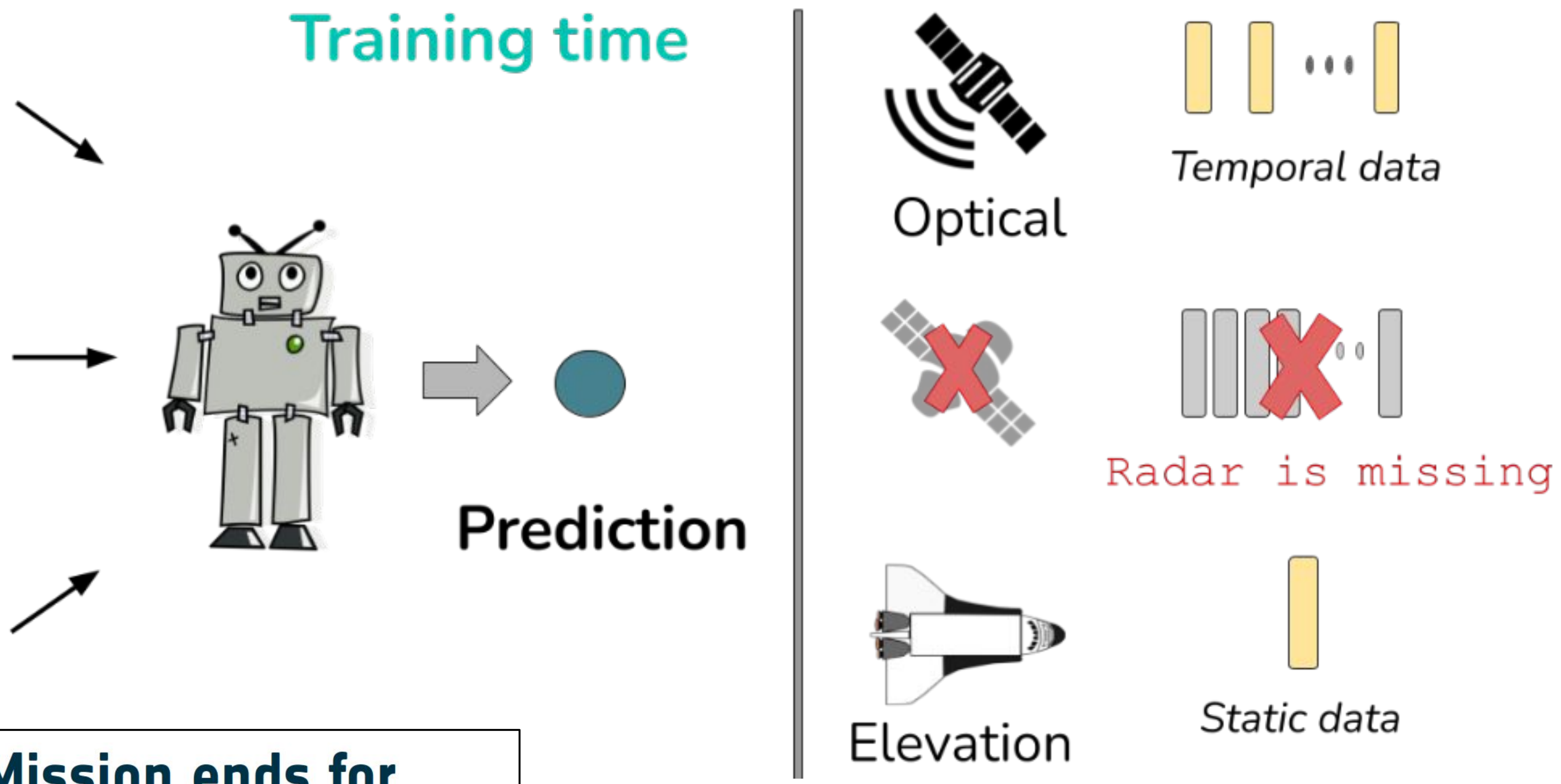
MACLEAN
MACHINE LEARNING FOR EARTH OBSERVATION

ECML PKDD 2024

Multi-sensor models in Earth Observation



Missing sensor data



Mission ends for Copernicus Sentinel-1B satellite

https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1/Mission_ends_for_Copernicus_Sentinel-1B_satellite

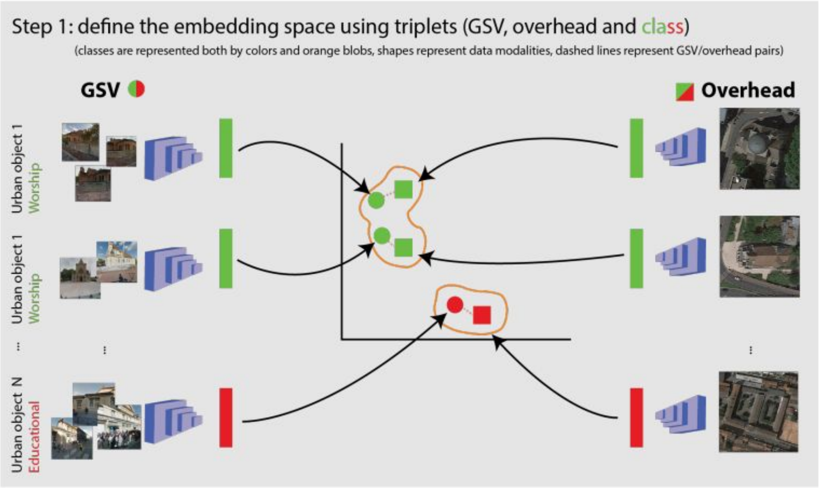
How to handle missing sensors in models at prediction time?

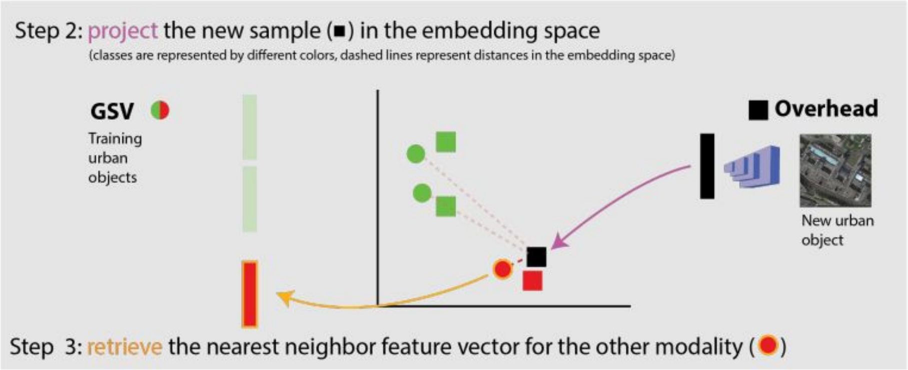
How to increase the prediction robustness of model to missing sensors?

How to handle missing sensor data?

1. Impute: replace the input features of the missing sensors with a numerical value.
   a. Random [1] or zero [2] value.

2. Exemplar: replace the features of the missing sensors with existing values observed in the training set, based on available sensors and a shared space [3].

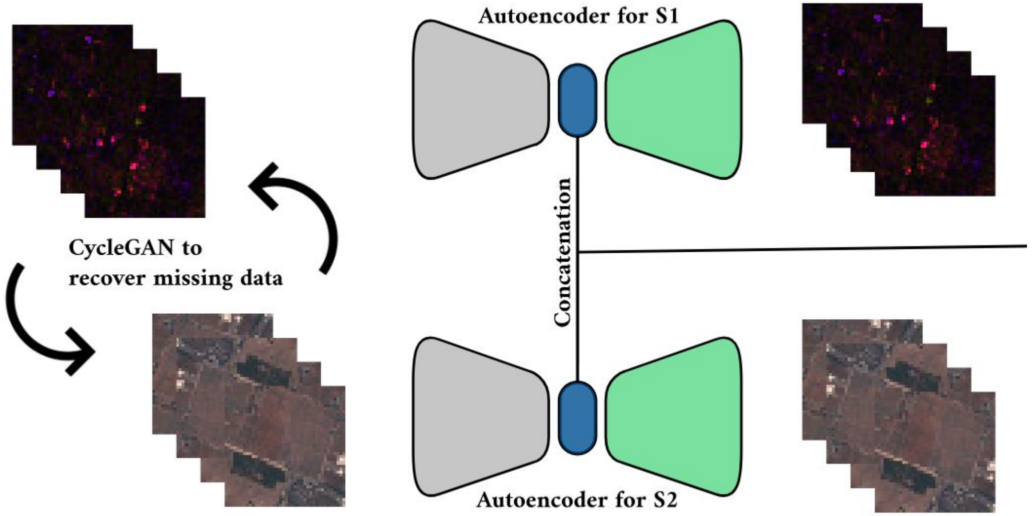




[1] Srivastava, N., & Salakhutdinov, R. R. (2012). Multimodal learning with deep boltzmann machines. Advances in neural information processing systems.
[2] Hong, D., Gao, L., Yokoya, N., Yao, J., Chanussot, J., Du, Q., & Better, B. Z. M. D. M. (2020). Multimodal Deep Learning Meets Remote-Sensing Imagery Classification.
[3] Srivastava, S., Vargas-Munoz, J. E., & Tuia, D. (2019). Understanding urban landuse from the above and ground perspectives: A deep learning, multimodal solution. Remote sensing of environment.

How to handle missing sensor data?

3. Reconstruction: replace the features of the missing sensors with a model that learned to reconstruct [1].



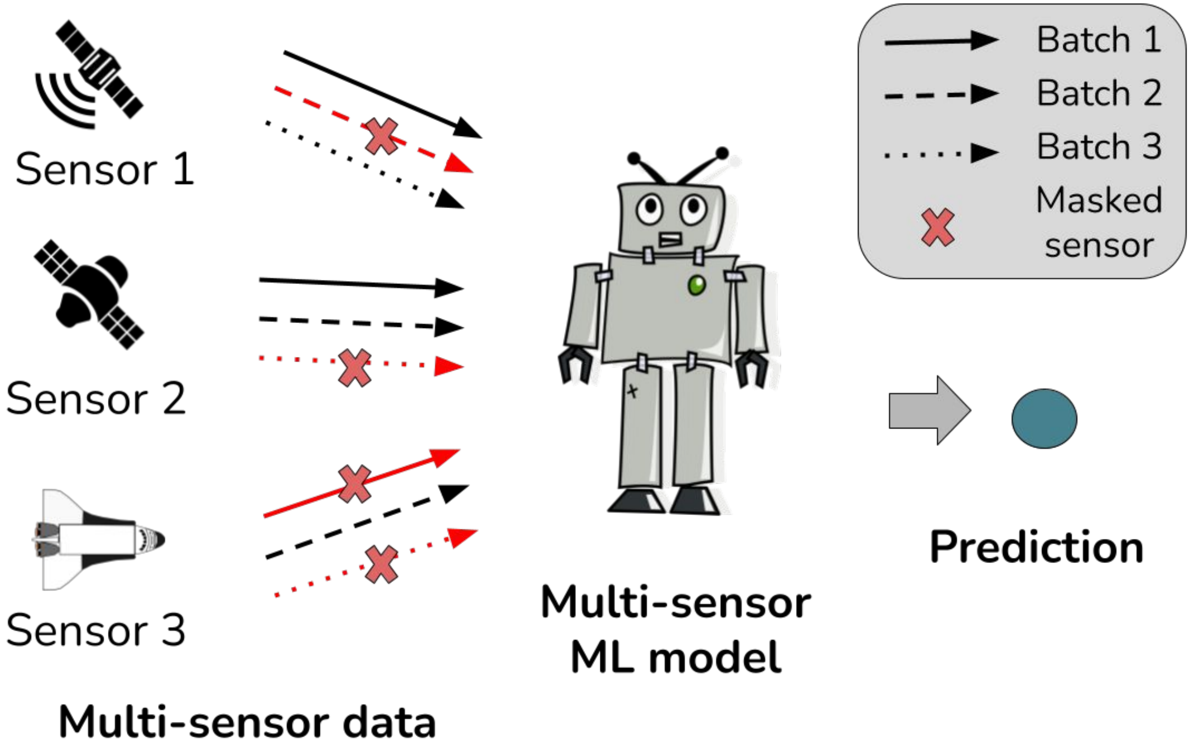
4. Ignore: omit the features of the missing sensors by using a dynamic merge function [2].

[1] Efremova, N., Seddik, M. E. A., & Erten, E. (2021). Soil moisture estimation using Sentinel-1/-2 imagery coupled with cycleGAN for time-series gap filing. IEEE Transactions on Geoscience and Remote Sensing.
[2] Mena, F., Arenas, D., Charfuelan, M., Nuske, M., & Dengel, A. (2024). Impact Assessment of Missing Data in Model Predictions for Earth Observation Applications

Two multi-sensor models using the Impute and Ignore techniques

Input Sensor Dropout

- ➢ Motivated by Temporal Dropout [1].
- ➢ We introduce **Sensor Dropout** (SensD).
- ➢ Dropout: zero-masking (Impute)

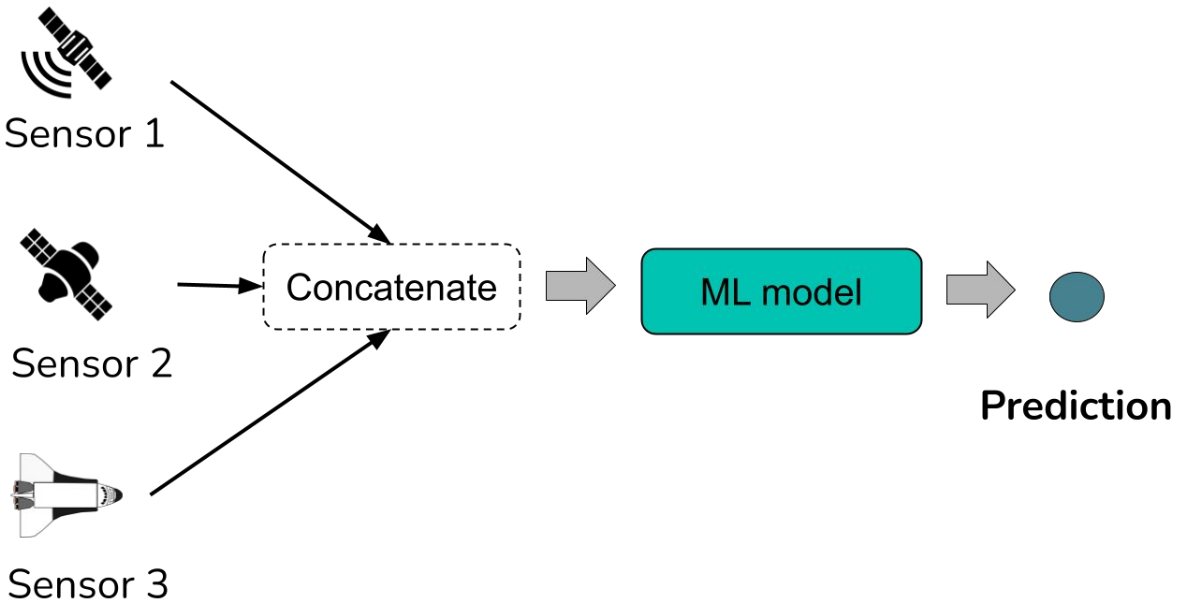


- ➢ Hyper-parameter:
  - ○ (SensD) dropout ratio
  - ○ No ratio: list all combinations of missing sensors, and select one randomly.

[1] Garnot, V. S. F., Landrieu, L., & Chehata, N. (2022). Multi-modal temporal attention models for crop mapping from satellite time series. ISPRS Journal of Photogrammetry and Remote Sensing.

Input Sensor Dropout

- Motivated by Temporal Dropout [1].
- We introduce **Sensor Dropout** (SensD) at input level.
- Dropout: zero-masking (Impute)



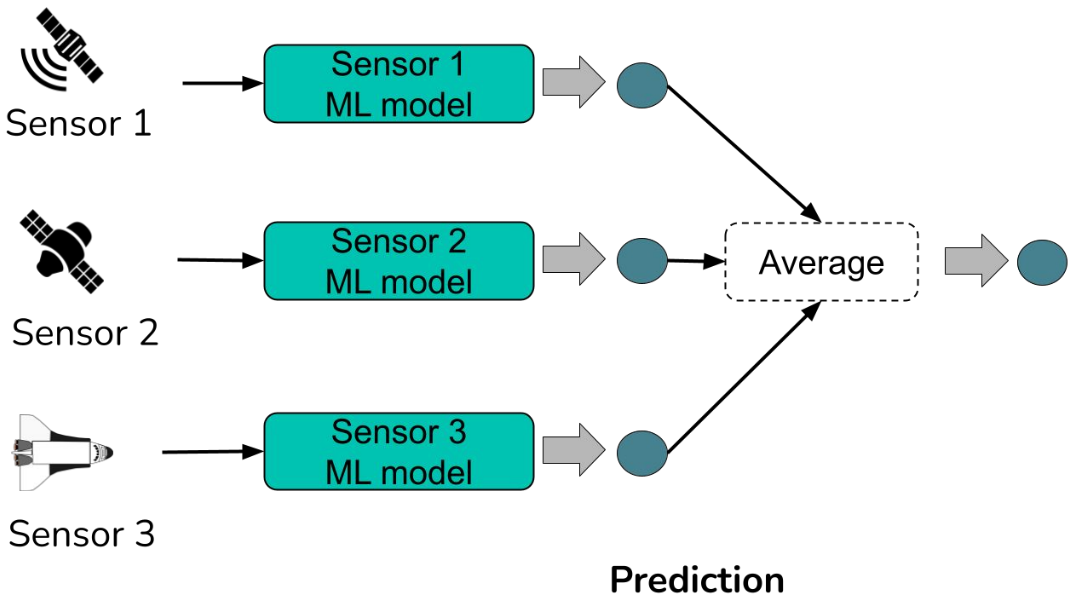
- Input-level fusion ML model.

Simulating missing sensor data during training.

[1] Garnot, V. S. F., Landrieu, L., & Chehata, N. (2022). Multi-modal temporal attention models for crop mapping from satellite time series. ISPRS Journal of Photogrammetry and Remote Sensing.

Ensemble Sensor Invariant

- Motivated by sensor invariant models [1].
- Sensor invariant: sharing layers between sensor-dedicated models.
- We extend the ensemble-based model [2].
  - Independent per-sensor learning.
  - Aggregation during inference (Ignore).



Ground truth

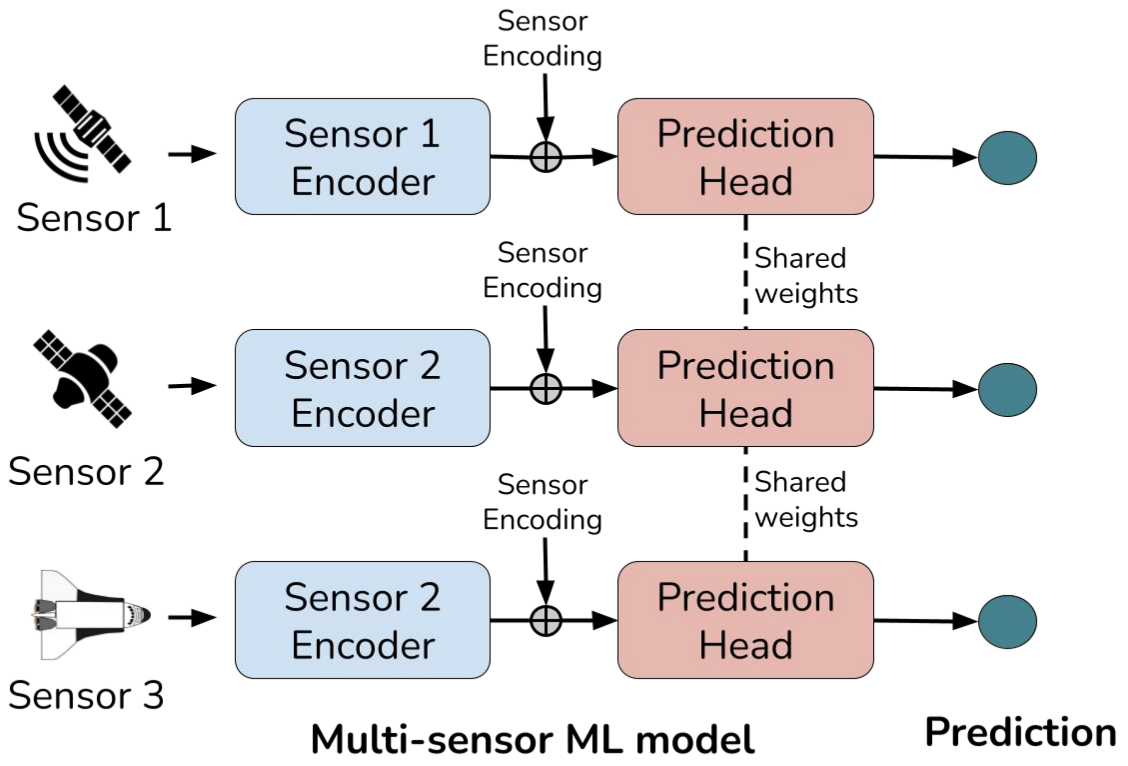
$$\hat{\Theta}_s = \arg\min_{\Theta_s} \sum_{i=1}^{N} \mathcal{L}(y^{(i)}, \hat{y}_s^{(i)})$$

Prediction from sensor "s"

[1] Francis, A., Mrziglod, J., Sidiropoulos, P., & Muller, J. P. (2021). Sensei: A deep learning module for creating sensor independent cloud masks. IEEE Transactions on Geoscience and Remote Sensing.
[2] Mena, F., Arenas, D., Charfuelan, M., Nuske, M., & Dengel, A. (2024). Impact Assessment of Missing Data in Model Predictions for Earth Observation Applications.

Ensemble Sensor Invariant

- ➢ We extend the ensemble-based model by using a **shared prediction head**.
- ➢ Sensor-dedicated encoders are required to match the same spatio-temporal-spectral dimension.
- ➢ Multi-sensor learning (parameters attached in the prediction head).



Ground truth

$$\hat{\Theta} = \arg \min_{\{\Theta_s\}_{s \in \mathcal{S}}} \sum_{i=1}^{N} \mathcal{L}(y^{(i)}, \hat{y}_s^{(i)}) \ .$$

Prediction from sensor “s”

Validation Datasets

- Pixel-wise datasets with temporal features.

Predictive Task	Samples	Where	Temporal length	Temporal sensor	Static sensor
Crop-type classification [1]	69,000	Global	12 months (monthly)	optical (Sentinel-2, 12) radar (Sentinel-1, 2) weather (ERA5, 2)	topographic (SRTM's DEM, 2)
Live fuel moisture content [2]	2,578	USA	4 months (monthly)	optical (Landsat 8, 8) radar (Sentinel-1, 3)	topographic (NED's DEM, 2) soil (UNASP, 3) land cover (GlobCover, 12) canopy height (LIDAR)
Particulate matter 2.5 [3]	167,309	China	3 days (hourly)	atmosphere conditions (3), dynamics (4), precipitation (2)	-

[1] Tseng, G., Zvonkov, I., Nakalembe, C. L., & Kerner, H. (2021, October). Cropharvest: A global dataset for crop-type classification. In Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track.
[2] Rao, K., Williams, A. P., Flefil, J. F., & Konings, A. G. (2020). SAR-enhanced mapping of live fuel moisture content. Remote Sensing of Environment.
[3] Chen, S.: PM2.5 Data of Five Chinese Cities. UCI Machine Learning Repository (2017).

Baselines & Evaluation

- Impute
  - Input: as the counterpart of **ISensD** without SensD technique.
  - ITempD: input-level fusion model with temporal dropout.
- Exemplar
  - Feature: feature-level fusion model. Missing sensors are replaced with a similarity-based search at feature-level.
- Ignore:
  - Ensemble: as the counterpart of **ESensI** without the sensor invariant component.

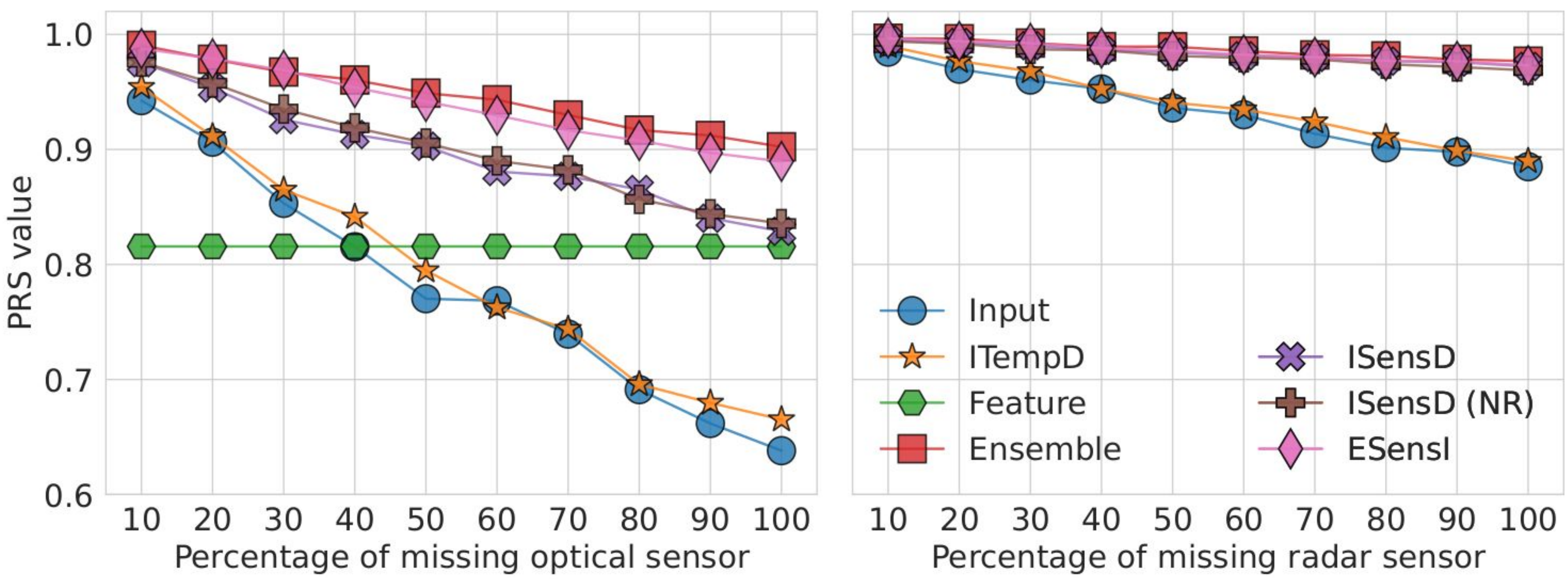
- 10-fold cross validation stimulating missing sensors in validation fold.
- Focus on the effect of missing temporal sensors.
- Performance Robustness Score [1]

$$\mathrm{PRS}(y, \hat{y}_{miss}, \hat{y}_{full}) = \min\left(1,\ \exp\left(1 - \frac{\mathrm{RMSE}(y, \hat{y}_{miss})}{\mathrm{RMSE}(y, \hat{y}_{full})}\right)\right)$$

[1] Heinrich, R., Scholz, C., Vogt, S., & Lehna, M. (2024). Targeted adversarial attacks on wind power forecasts. Machine Learning.

PRS on classification datasets

Crop-type classification

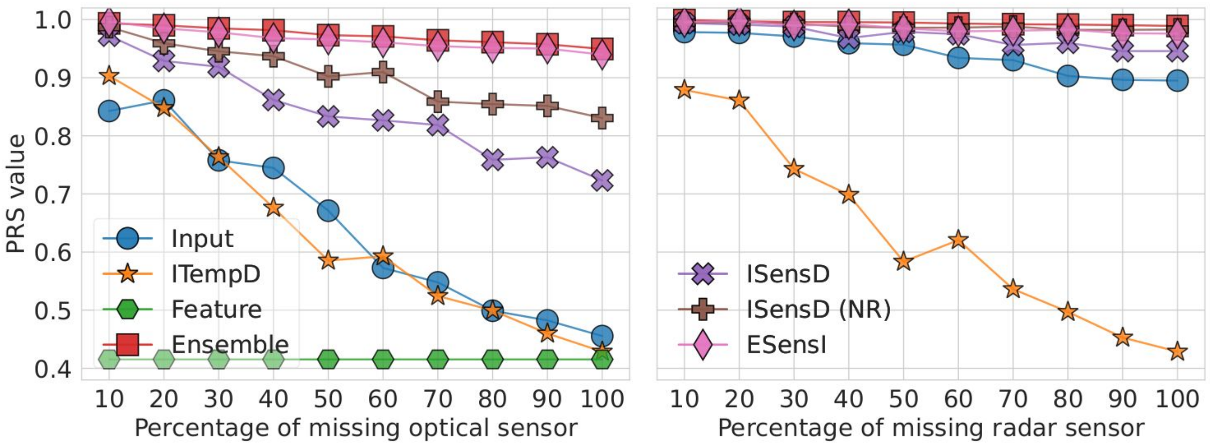


The decrease in PRS is mostly linear for all methods when the number of samples with missing sensors increases.

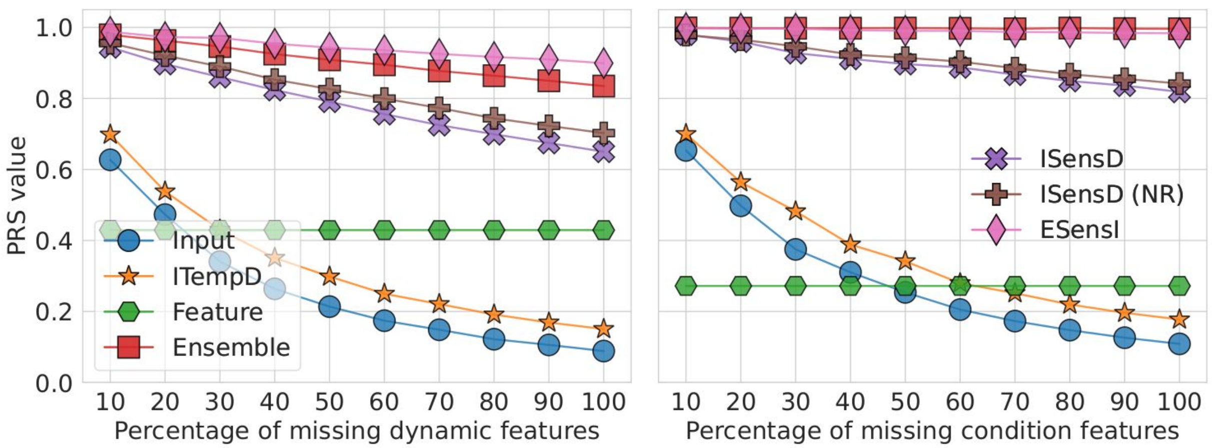
PRS on regression dataset

- In regression, the robustness effect is more variable.

Live fuel moisture content



Particulate matter 2.5



Input-level fusion models have an exponential decay in PRS

Full-sensor predictive performance

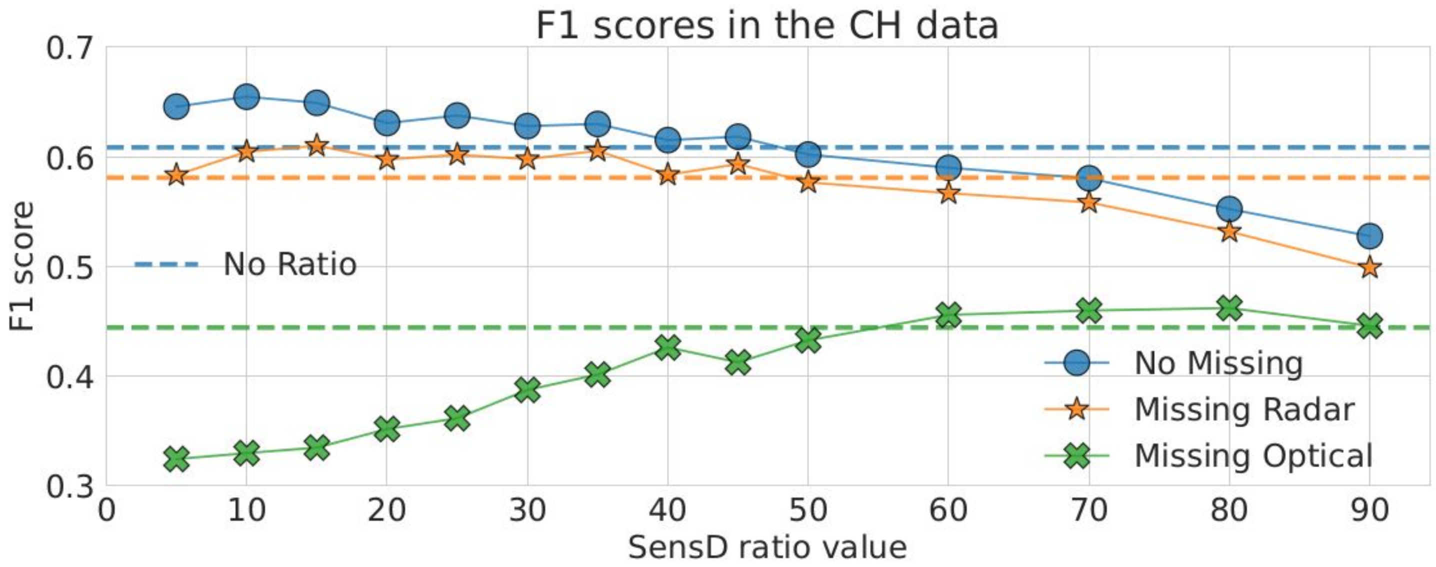
Table 1: Predictive performance on a full-sensor evaluation, i.e. no missing features, on different datasets. The F1 is shown in classification and R^2 in regression.

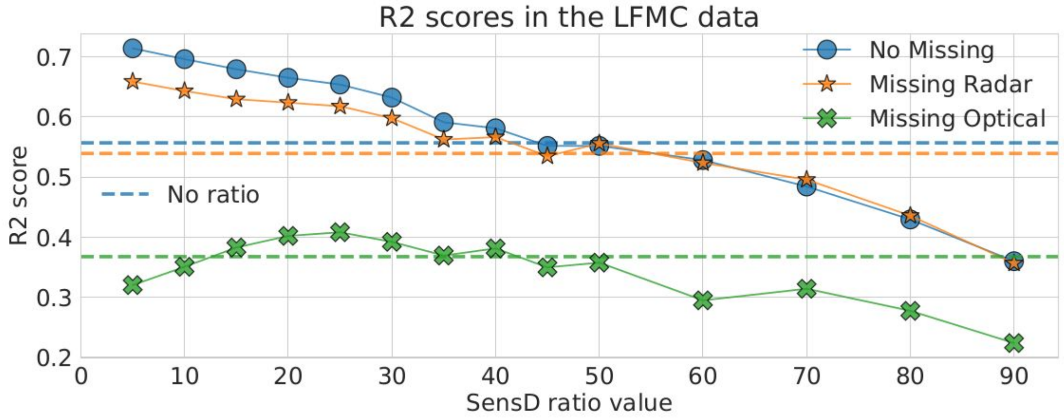
Dataset	Input	ITempD	Feature	Ensemble	ISensD	ISensD (NR)	ESensI
CH	**0.648**	0.642	0.576	0.605	0.598	0.611	0.615
LFMC	0.714	**0.717**	0.650	0.313	0.655	0.545	0.326
PM25	**0.917**	0.882	−0.354	0.308	0.525	0.496	0.232

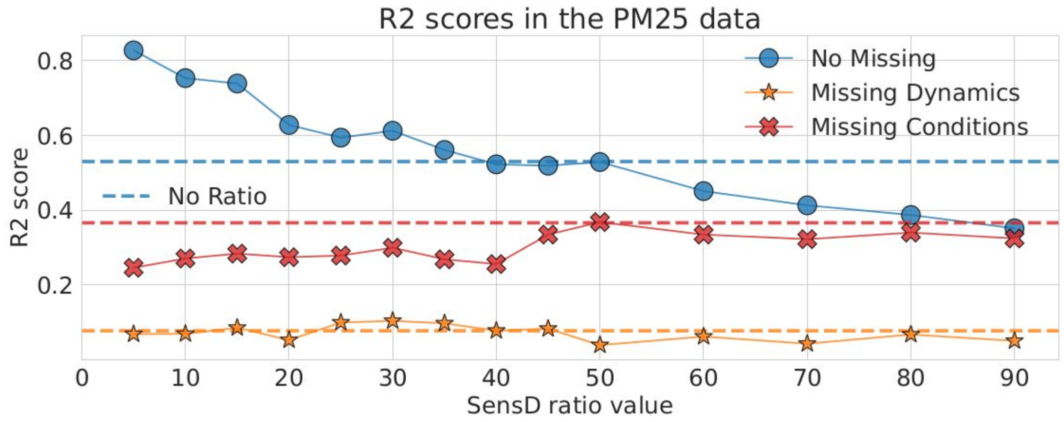
- ISensD performs worse than its counterpart, Input.
- ESensI outperforms its counterpart, Ensemble, in two out of three datasets.
  - Overall performance is quite low.

Ablation on ISensD dropout ratio

- The best ratio value depends on the dataset and sensor missing.







Ablation ESensI

- The best configuration of ESensI method depends on the dataset.

Table 2: Predictive performance results of different configurations in the ESensI model. The results are obtained in the full-sensor evaluation.

Shared weights	Sensor encoding	Normalize	Others	CH (F1)	LFMC (R^2)	PM25 (R^2)
-	-	-		0.603	0.313	0.308
✓	-	-		0.595	0.281	**0.281**
✓	✓	-		**0.606**	0.309	0.274
✓	✓	✓	add	0.598	**0.317**	0.241
✓	✓	✓	concat	0.595	0.297	0.196

Conclusion

- 💡 We introduce two methods that increase the robustness of its counterpart to missing sensors.

- 💡 Limitation in the full-sensor scenario.
  - ➢ ISensD
    - ■ Full-sensor scenario is a minority.
  - ➢ ESensI
    - ■ Individual predictions in regression are not effective.

- 💡 The effectiveness of the SensD value is highly variable, as it depends on the dataset and the missing sensor.
  - ➢ The "no ratio" version is a good parameter-free alternative.

- 💡 Future work: combine good components of both approaches.

Increasing the Robustness of Model Predictions to Missing Sensors in Earth Observation

Francisco Mena, Diego Arenas, Andreas Dengel

University of Kaiserslautern-Landau | DFKI

ECML PKDD 2024

Implementation

- For input-level fusion methods.
  - Temporal features are aligned, and static are repeated across time.
- For sensors with temporal features: 1D CNN.
  - Similar results are observed with other architectures.
- For sensors with static features: MLP.

- Architecture configuration: 2 layers, 128 units, 20% dropout.
- Optimization: ADAM, 128 batch size.