NORSK REGNESENTRAL
NORWEGIAN COMPUTING CENTER

fm4cs

Foundation Models for Climate and Society (FM4CS)

Arnt Salberg
Norwegian Computing Center

esa
Φ-lab

NVE

Danish Meteorological Institute

Polar View

UiT The Arctic University of Norway

METEO ROMANIA

Objective

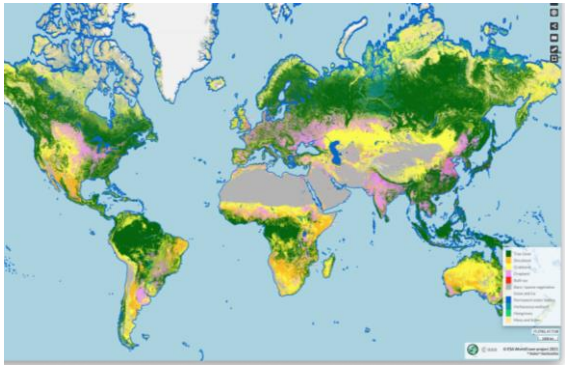
The main objective is to unleash the potential of self-supervised learning and large-scale multimodal Foundation Models for EO use cases related to climate science and society applications and services.

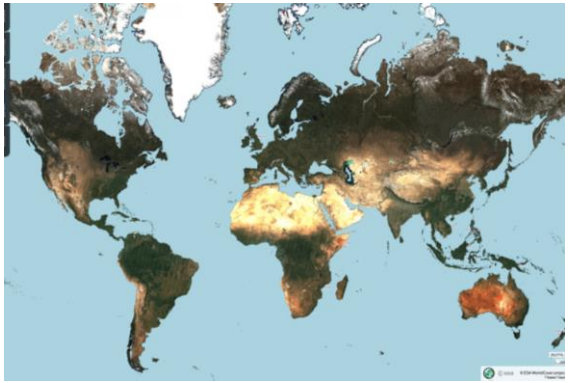
FM4CS use-cases:

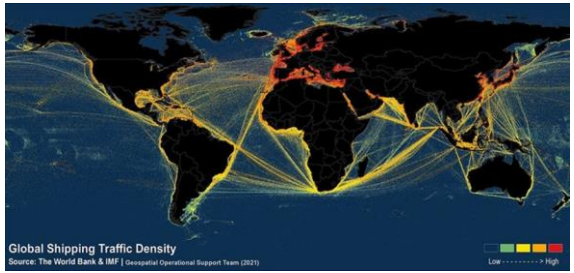
Use-case	Description	Task	Output resolution	Sensor	
Snow	Mapping of snow	Pixelwise regression	250m	S1, **S3 SLSTR**	
Sea-ice	Mapping of sea-ice concentration	Pixelwise classification	250m	**S1**, S3	
Ice-berg	Detection of ice-bergs	Object detection	10m	**S1**, S2	
Drought	Vegetation drought resilience	Pixelwise regression	250m	**S3 OLCI & SLSTR**	
Flood	Mapping of flooded areas	Pixelwise classification	10m	**S1**, S2	
Wetland	Mapping of mire areas	Pixelwise classification	10m	S1, **S2**	

Dataset

- Need a highly diverse training dataset to boost the training of the Foundation Model
- We use Sentinel-2 grid as spatial sampling unit
- Stratified sampling approach to ensure that diversity of land covers are included in the training data.



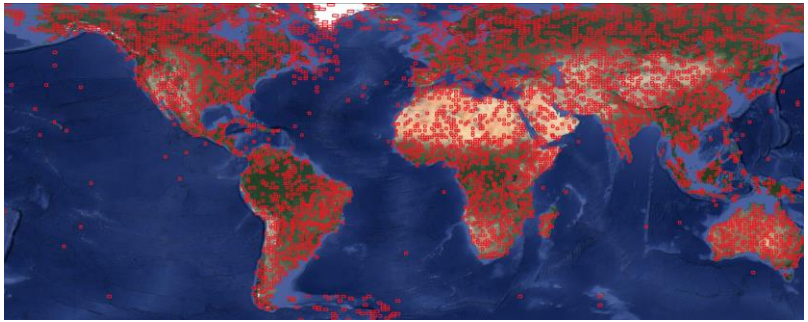




fm4cs

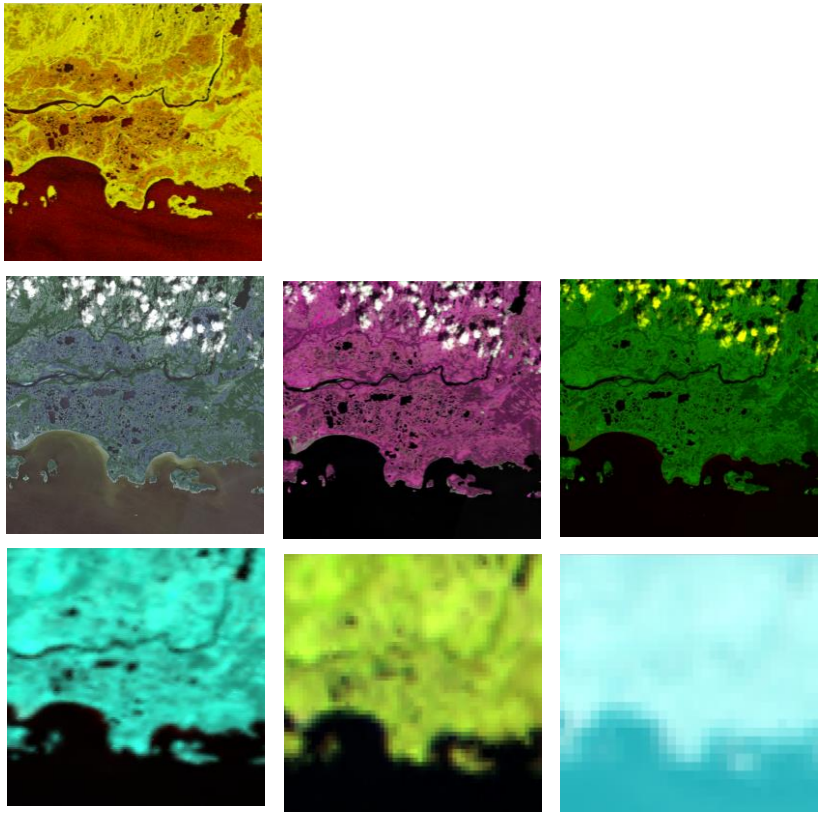
Dataset

- Selected more than 6000 different Sentinel-2 tiles.
- More than 18000 tile and date combinations, totals to more than 20 TB.
- For a given Sentinel-2 tile and time instant:
  - select S1, S2 and S3 SLSTR and S3 OLCI that is present.



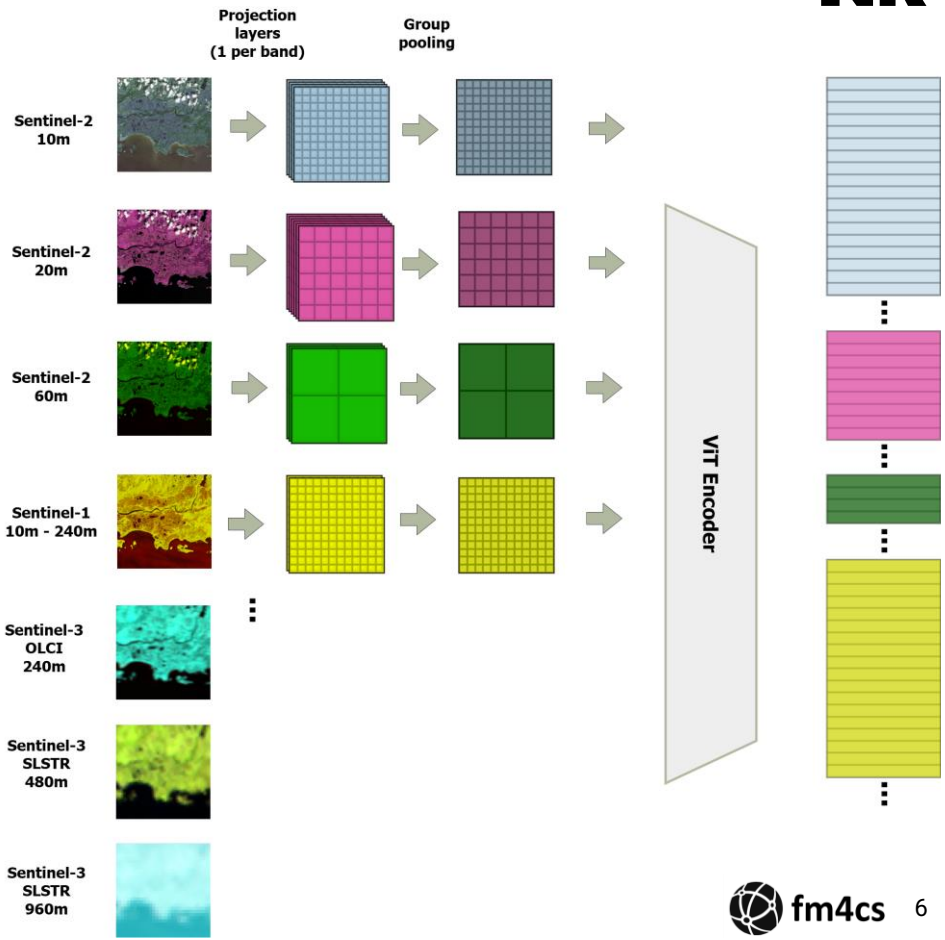
Dataset

- Sentinel-1 SAR
  - GRD / 2 bands
  - VV/HV or HH/HV
  - Multiple resolutions
- Sentinel-2 MSI
  - Level 2A / 12 bands
  - 10m / 20m /60m
- Sentinel-3 OLCI
  - Level 1C / 21 bands
  - 300 m
- Sentinel-3 SLSTR
  - Level 1C / 9 bands
  - 500m / 1000m



FM encoder architecture

- ViT is the core processing module (12 or 24 layers).
- Handles multiple sensors at their native resolution.
- Different patch projection layers for different sensors bands.
- Group pooling merges sensor specific patches.
- Group specific encoding is used to differentiate between the different sensor groups.
- All sensors have the same ground cover.
- Ground cover is randomly selected from a set of candidates (1000m - 32000m).
- Patch size is randomly selected from 6 to 32.



FM architecture – flexible patch size (FlexiViT approach)

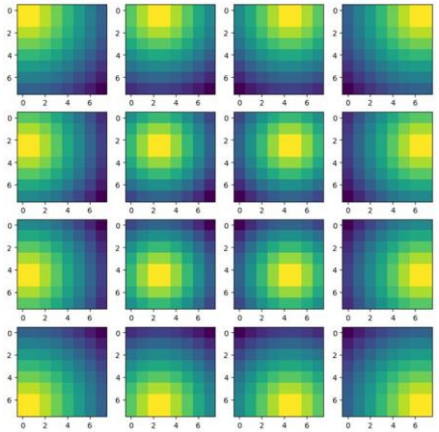
- Randomized patch size during training
- Patch size encoding used to differentiate between differentiate patch sizes ($ps \cdot GSD/(6 \cdot 10)$)
- Interpolates the weights in the patch projection layer. Extended to per group.
- Extended FlexiViT to work in a masked image modelling setup (loss).
- Budget in terms of number of tokens.
- Can in principle use different patch size per group.

FM architecture – 2D-ALiBi

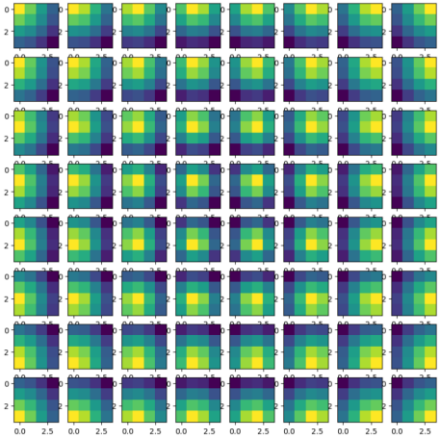
2D-ALiBi – biases the self-attention matrix based on the distance between token's ground cover distance.

$$a_{hij} = \sqrt{d} \cdot q_{hi} \cdot k_{hj} - \frac{distance(c_i, c_j)}{max(p)} \cdot m(h)$$

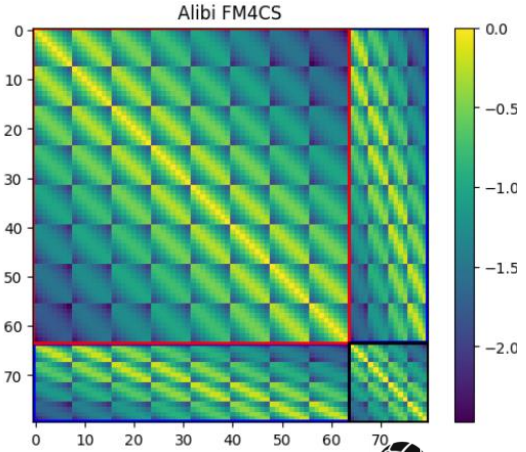
How 4 20m patches relates to 8 10m patches



How 8 10m patches relates to 4 20m patches



The FM4CS ALiBi matrix

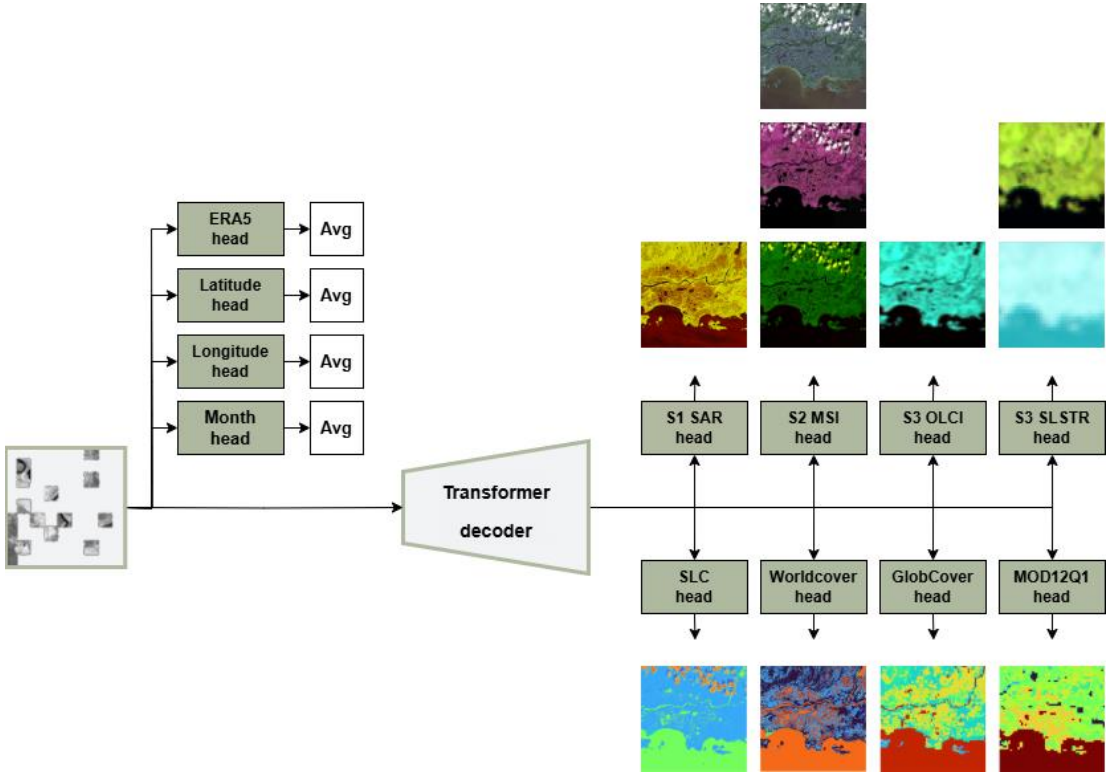


FM training

Masked image modeling principle

Pixel-level mapping of

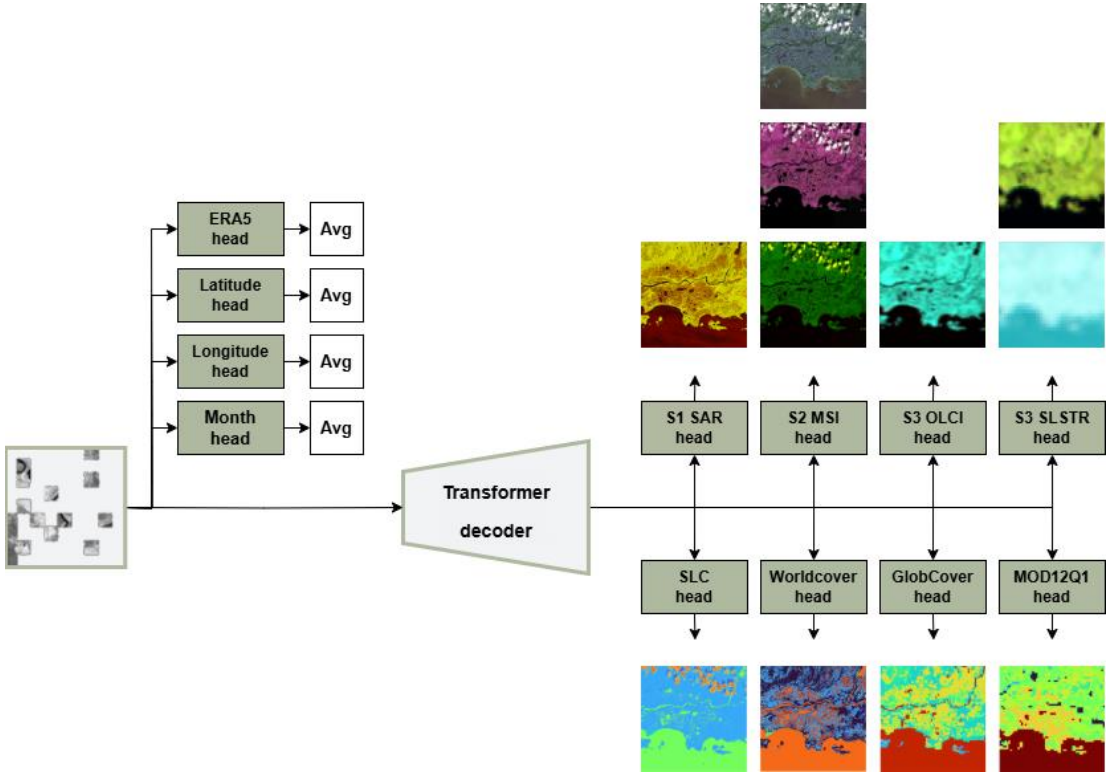
- Reconstructing all sensor bands
- Scene classification map (20m)
- ESA WorldCover (10m)
- ESA GlobCover (300)
- Modis MOD12Q1 (500m)



FM training

Image-level prediction of

- ERA5
- Latitude
- Longitude
- Month

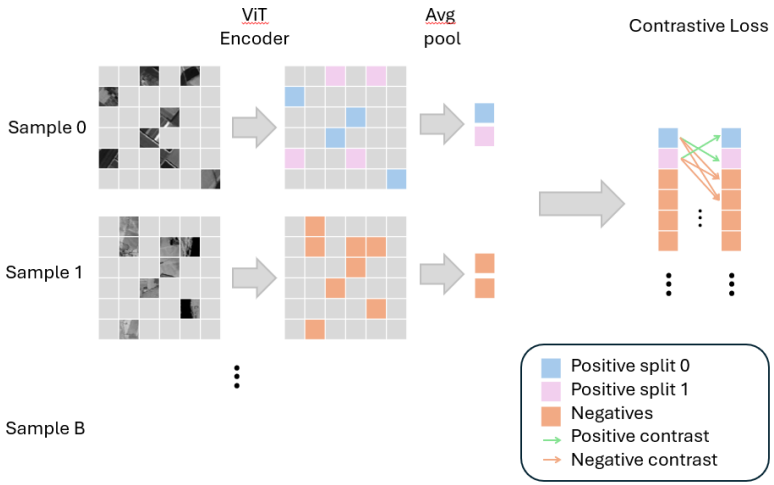


FM training

Constrastive loss:

$$\mathcal{L}_{snn} = -\frac{1}{B}\sum_{i=1}^{B} \log \frac{\sum_{i\neq j, y_i = y_j, j=1,\ldots,B} \exp\left(-f\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)/\tau\right)}{\sum_{i\neq k, k=1,\ldots,B} \exp\left(-f\left(\boldsymbol{x}_i, \boldsymbol{x}_k\right)/\tau\right)}$$

(Not sensor- or scale-invariant!)



FFT loss:

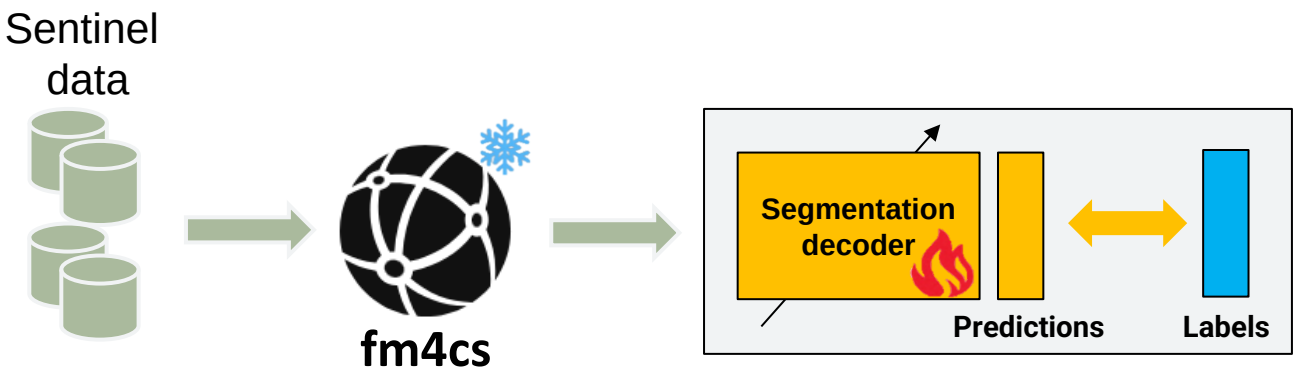
$$L_{FFT} = \frac{1}{P}\sum_{p=1}^{P} L_1\left(\left|F\left(y_p\right)\right|, \left|F\left(y_p'\right)\right|\right).$$

Kraus et al. (2024) observed a more stable training when utilizing the FFT loss.

Total loss:

$$L_{tot} = L_{MAE} + L_{CE} + L_{MSE} + L_{CON} + +L_{FFT}$$

Benchmarking



Benchmarking using the PANGAEA framework (Marsocci et al., 2025)

Preliminary benchmark results

100% training data

Model	HLS Burns	MADOS	PASTIS	Sen1Floods11	FBP	DynEarthNet	CropMap	SN7	AI4Farms
CROMA	82.42	**67.55**	32.32	90.89	51.83	38.29	49.38	59.28	25.65
DOFA	80.63	59.58	30.02	89.37	43.18	39.29	51.33	61.84	27.07
GFM-Swin	76.90	64.71	21.24	72.60	67.18	34.09	46.98	60.89	27.19
Prithvi	83.62	49.98	33.93	90.37	46.81	27.86	43.07	56.54	26.86
RemoteCLIP	76.59	60.00	18.23	74.26	**69.19**	31.78	52.05	57.76	25.12
SatlasNet	79.96	55.86	17.51	90.30	50.97	36.31	46.97	61.88	25.13
Scale-MAE	76.68	57.32	24.55	74.13	67.19	35.11	25.42	**62.96**	21.47
SpectralGPT	80.47	57.99	35.44	89.07	33.42	37.85	46.95	58.86	26.75
S12-MoCo	81.58	51.76	34.49	89.26	53.02	35.44	48.58	57.64	25.38
S12-DINO	81.72	49.37	36.18	88.61	51.15	34.81	48.66	56.47	25.62
S12-MAE	81.91	49.90	32.03	87.79	51.92	34.08	45.8	57.13	24.69
S12-Data2Vec	81.91	44.36	34.32	88.15	48.82	35.90	**54.03**	58.23	24.23
UNet Baseline	**84.51**	54.79	31.60	**91.42**	60.47	**39.46**	47.57	62.09	**46.34**
ViT Baseline	81.58	48.19	38.53	87.66	59.32	36.83	44.08	52.57	38.37
FM4CS	79.19	49.10	**39.99**	90.13	-	-	53.96	61.07	26.71

Preliminary benchmark results

50% training data

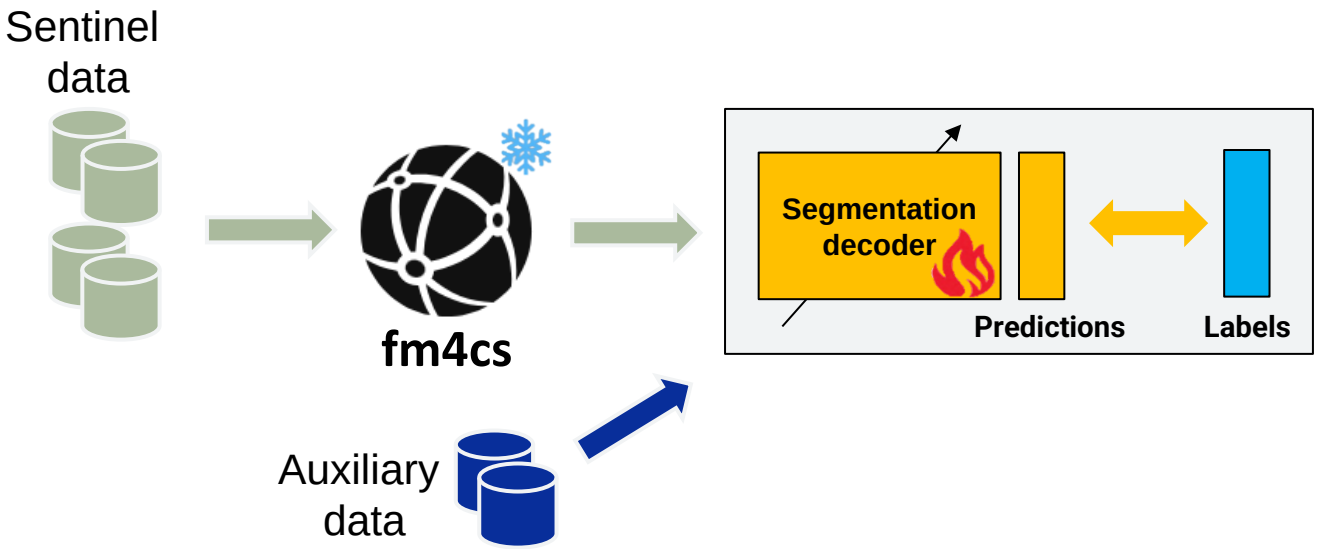
Model	HLS Burns	MADOS	PASTIS	Sen1Floods11	FBP	DynEarthNet	CropMap	SN7	AI4Farms
CROMA	81.52	57.68	32.33	90.57	48.01	38.30	42.20	59.31	28.19
DOFA	78.02	55.21	28.60	88.39	36.90	**39.20**	30.93	47.06	26.69
GFM-Swin	74.36	**63.37**	20.41	71.61	63.14	31.25	31.42	59.83	28.43
Prithvi	80.89	40.79	33.13	89.69	40.27	33.43	42.51	49.45	29.27
RemoteCLIP	74.28	53.26	17.46	71.67	**65.92**	30.91	36.3	50.83	25.11
SatlasNet	75.97	52.24	16.78	89.45	46.04	36.34	35.29	60.74	27.08
Scale-MAE	75.47	46.87	23.26	72.54	62.11	32.60	20.32	**61.24**	26.40
SpectralGPT	76.40	58.00	34.61	87.52	21.71	36.52	32.09	56.28	27.46
S12-MoCo	79.79	42.90	32.59	89.22	46.92	34.45	41.32	56.21	28.38
S12-DINO	80.12	40.42	35.71	88.93	44.85	32.76	31.13	55.14	25.68
S12-MAE	80.13	44.29	31.15	88.43	45.63	33.29	28.07	55.55	27.50
S12-Data2Vec	79.82	41.22	33.42	86.58	46.73	32.61	28.53	56.94	25.84
UNet Baseline	**82.39**	43.87	30.25	**90.91**	55.42	35.14	36.30	46.82	**45.02**
ViT Baseline	78.17	28.77	38.71	86.08	57.32	37.33	39.53	49.21	38.37
FM4CS	78.12	50.63	**38.85**	90.30	-	-	**56.27**	60.35	27.02

Preliminary benchmark results

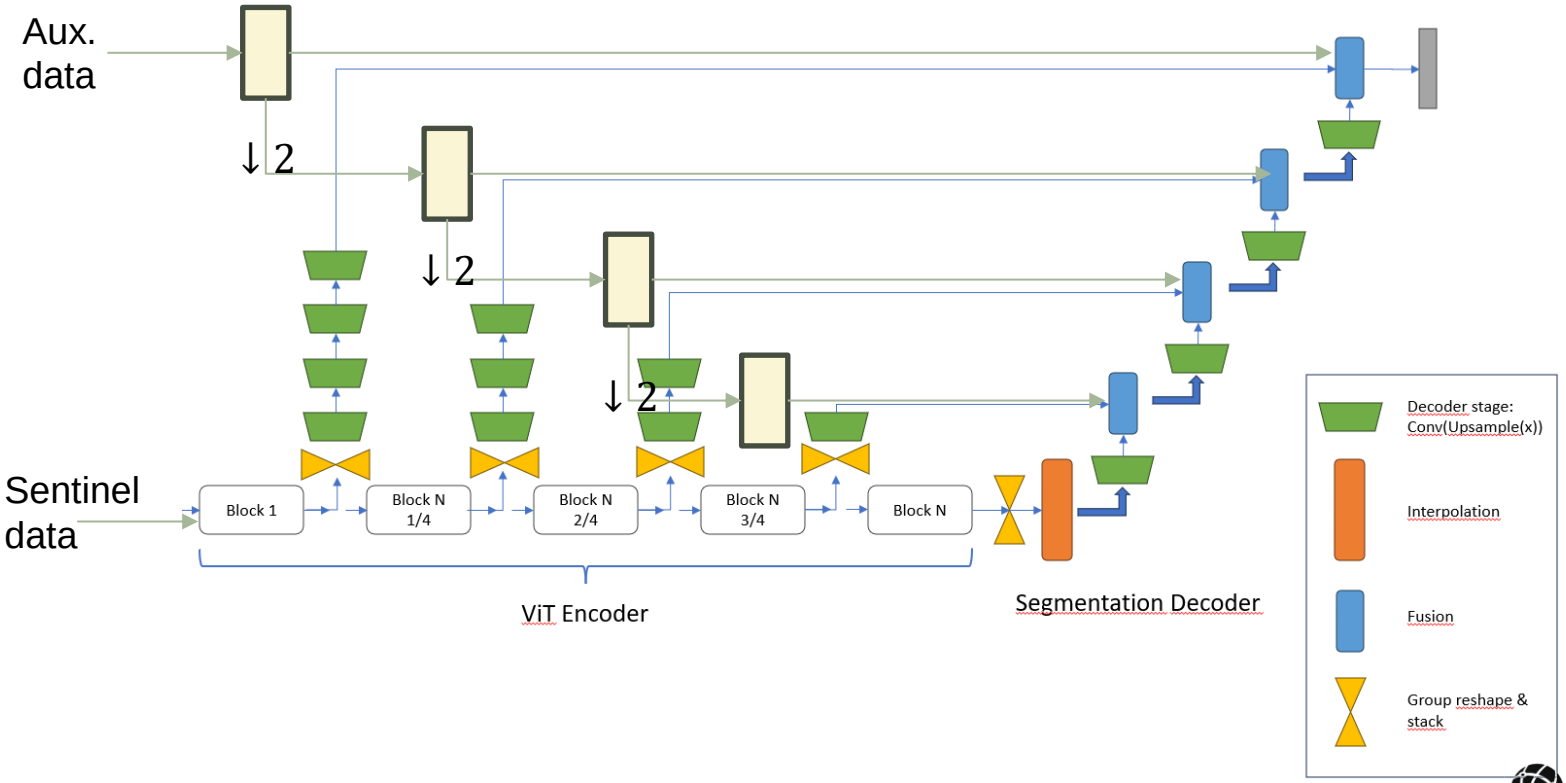
10% training data

Model	HLS Burns	MADOS	PASTIS	Sen1Floods11	FBP	DynEarthNet	CropMap	SN7	AI4Farms
CROMA	76.44	32.44	32.80	87.22	37.39	36.08	36.77	42.15	38.48
DOFA	71.98	23.77	27.68	82.84	27.82	**39.15**	29.91	46.10	27.74
GFM-Swin	67.23	28.19	21.47	62.57	55.58	28.16	27.21	39.48	32.88
Prithvi	77.73	21.24	33.56	86.28	29.98	32.28	27.71	36.78	35.04
RemoteCLIP	69.40	20.57	17.19	62.22	56.23	34.43	19.86	43.11	23.85
SatlasNet	74.79	29.87	16.76	83.92	37.86	34.64	29.08	49.78	13.91
Scale-MAE	75.47	21.47	22.86	64.74	48.75	35.27	13.44	49.68	26.66
SpectralGPT	**83.35**	20.29	34.53	83.12	39.51	35.33	31.06	36.31	37.35
S12-MoCo	73.11	19.47	32.51	79.58	35.57	32.24	36.54	49.46	37.97
S12-DINO	75.93	23.47	36.62	84.95	34.63	32.78	38.44	41.15	37.91
S12-MAE	76.60	18.44	31.06	84.81	35.56	30.59	35.29	40.51	23.60
S12-Data2Vec	74.38	17.86	33.09	81.91	37.27	33.63	34.11	40.66	22.85
UNet Baseline	79.46	24.30	29.53	**88.55**	52.58	35.59	13.88	46.08	34.84
ViT Baseline	75.92	10.18	38.44	81.85	**56.53**	35.39	27.76	36.01	**39.20**
FM4CS	76.37	**43.32**	**38.76**	87.93	-	-	**49.48**	**60.22**	25.36

Downstream task adaption

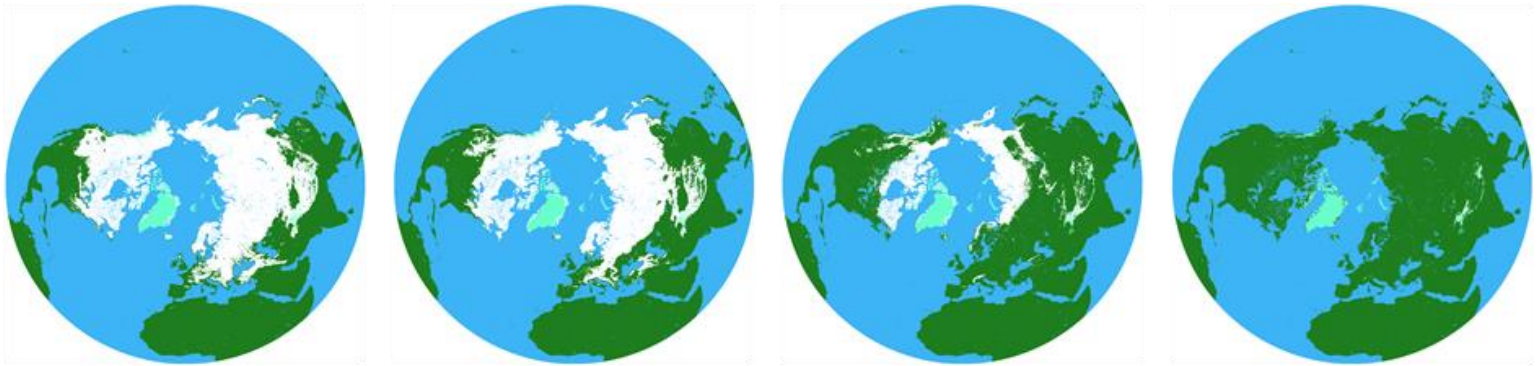


Downstream task – segmentation with auxilliary data



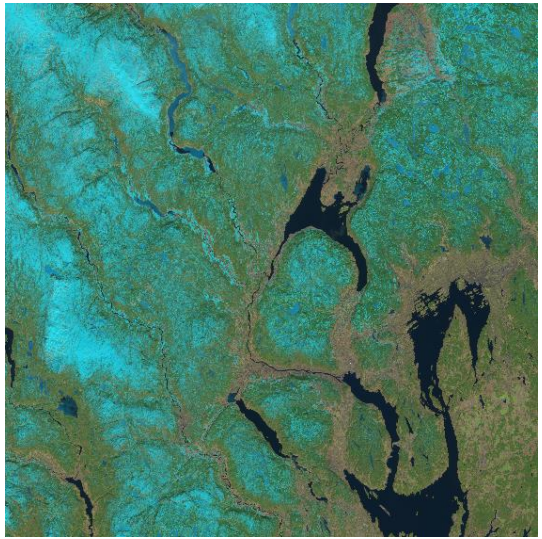
Snow mapping and monitoring

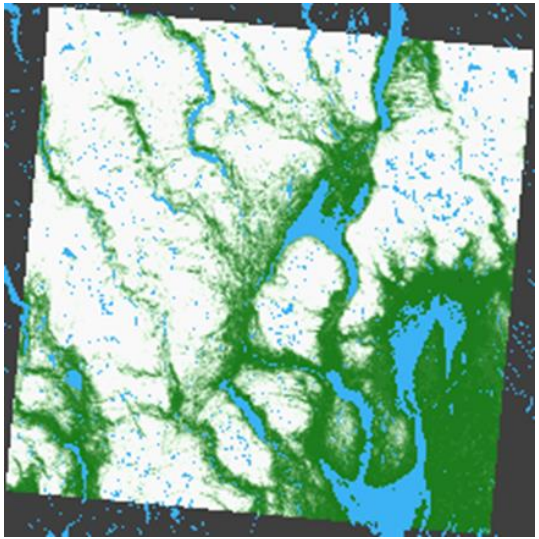
Snow cover helps regulate the Earth's surface and atmosphere and affects regional weather patterns.



Snow mapping and monitoring

Reference data created from S2 MSI snow products with physics-based algorithms combined with manual inspection and control to ensure high quality.



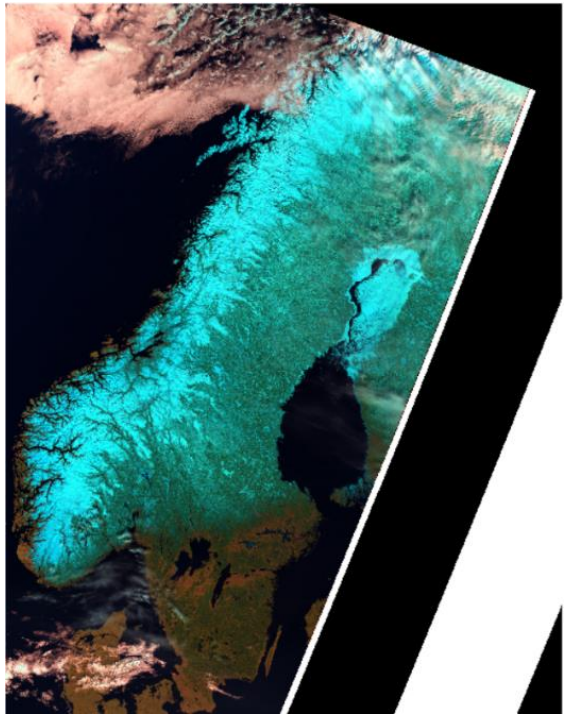


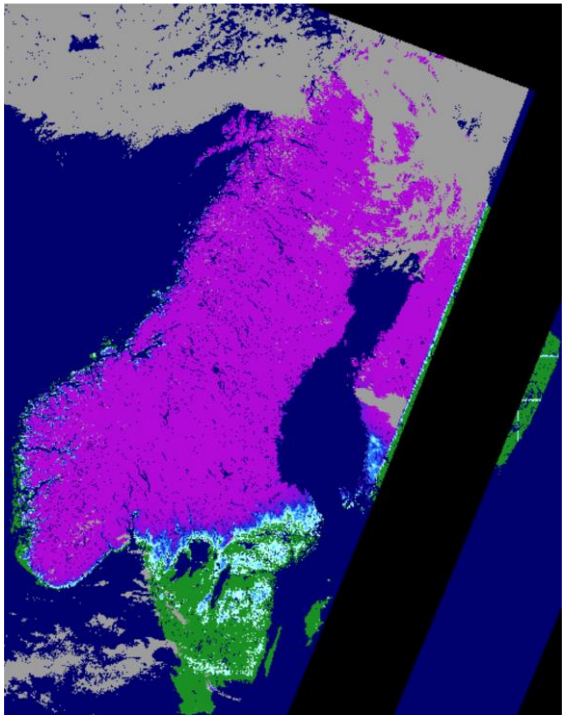
Target variables

- ***Snow cover fraction***
  - Earth's energy balances
  - Water resources
- ***Snow grain size***
  - Avalanche risk
  - Energy balance
- ***Snow surface wetness***
  - Snowmelt and runoff
  - Flood forcasting

Snow mapping and monitoring

Preliminary fractional snow cover results





Test MAE=2.73

Mapping of sea ice concentration

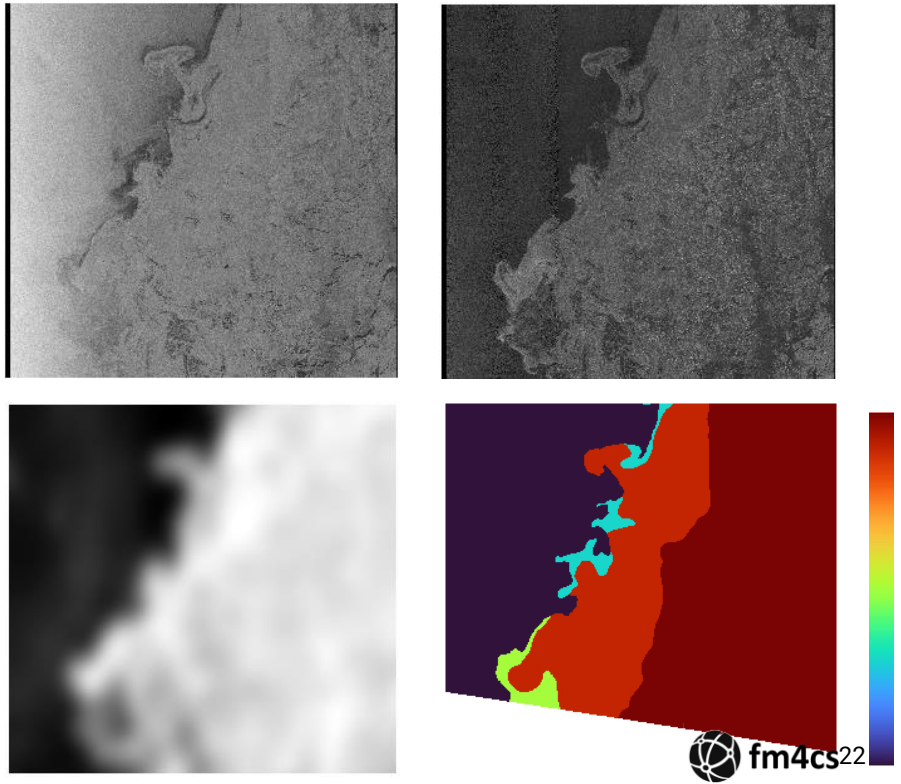
Danish Meteorological Institute

Sea ice plays a critical role in the global climate system and maritime operations.

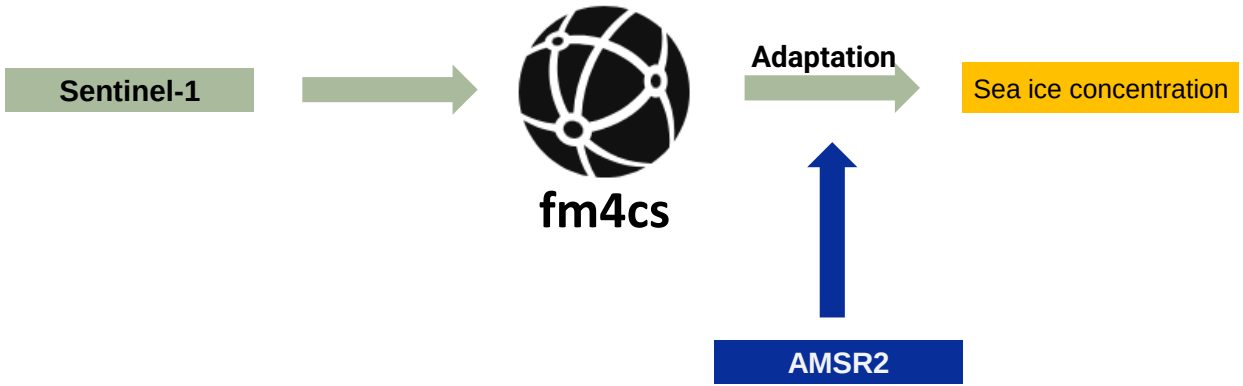
SIC indicates the percentage of sea ice coverage within each polygon and is the target variable.

Manually created maps of SIC are used to train a decoder for mapping SIC in Sentinel-1 data.

- Sensor: Sentinel 1 (~250m)
- Auxiliary data: AMSR2
- Severe class imbalance

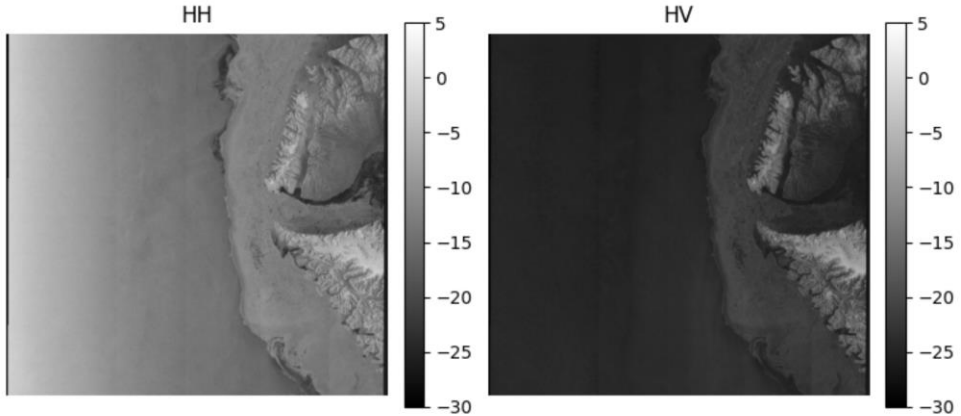


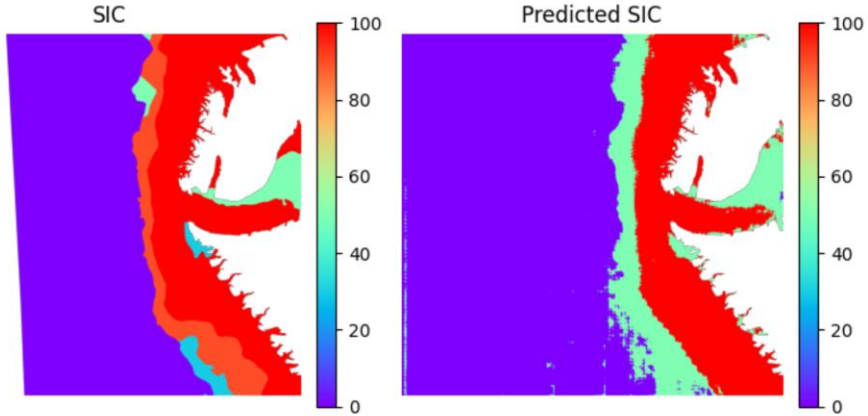
Use-case: sea ice



Preliminary sea ice results

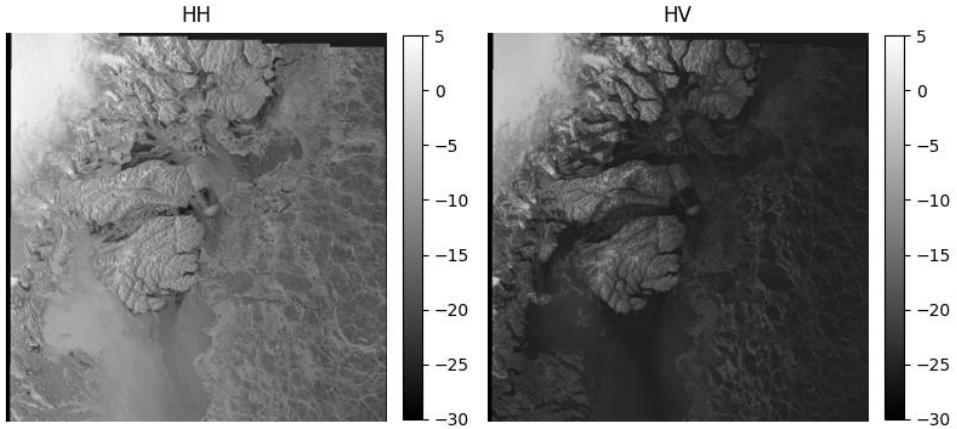
Overall accuracy 3 classes	83%
Water	94%
SIC 10 – 90 %	73%
SIC 100%	86%

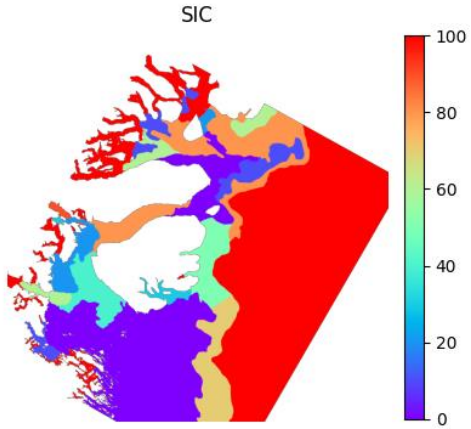


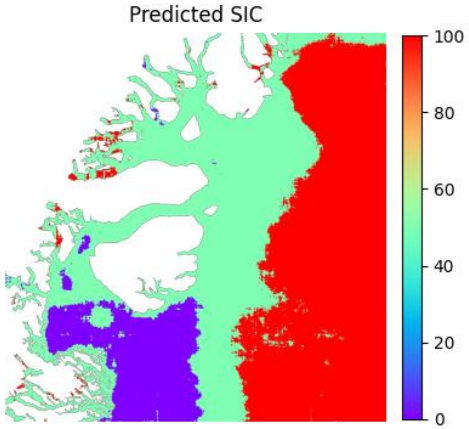


Preliminary sea ice results

Overall accuracy 3 classes	83%
Water	94%
SIC 10 – 90 %	73%
SIC 100%	86%







Conclusions

- We propose one model for analyzing satellite data provided by Sentinel-1 SAR, Sentinel-2 MSI, Sentinel-3 OLCI and Sentinel-3 SLSTR of land and ocean environments.
- The FM4CS model supports sensor resolutions from both SAR and multispectral data, with vast resolution differences from 10m to 1000m.
- The model can analyze date from one satellite sensor or a combination of sensors.
- The core of the model is a ViT. However, it is highly flexible and supports different input ground covers and patch sizes.
- The model (Base and Large versions) will be made freely open for download and use soon.