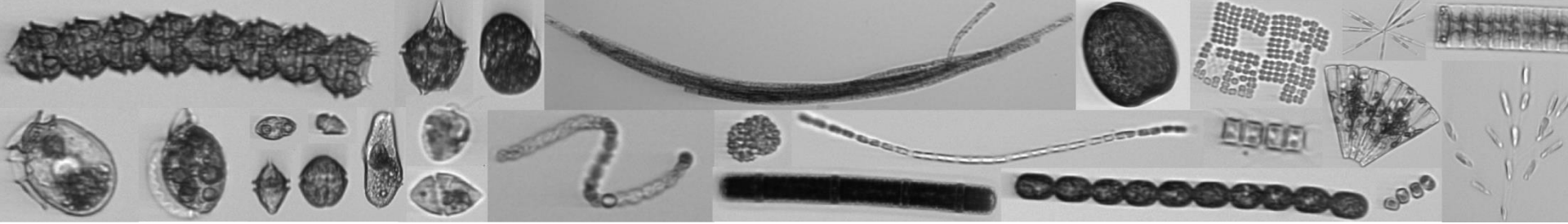


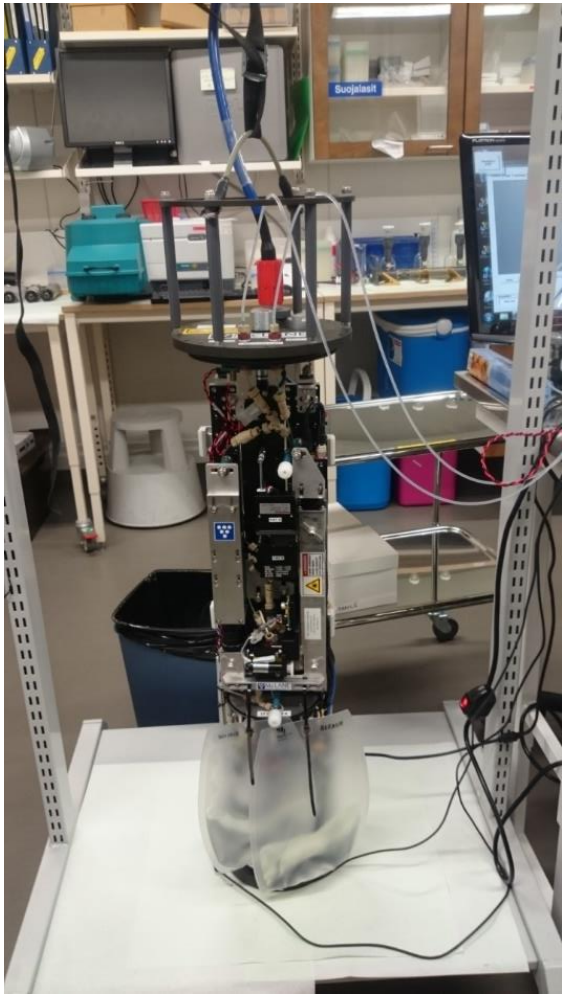
Using IFCB to monitor filamentous cyanobacteria in the Baltic Sea
-
Automated data pipeline and CNN based classification system

Kaisa Kraft, Otso Velhonoja, Jukka Seppälä, Heidi Hällfors, Sanna, Pasi Ylöstalo, Sami Kielosto, Sirpa Lehtinen, Johanna Oja, Timo Tamminen
Finnish Environment Institute SYKE, Marine Research Centre
25.8.2022

FinMARI Finnish Marine Research Infrastructure
SYKE
JERICO-S3
plus AQUACOSM
TAH TIINA JA ANTTI HERLININ SÄÄTIÖ
Academy of Finland project FASTVISION
LUT University
FMI

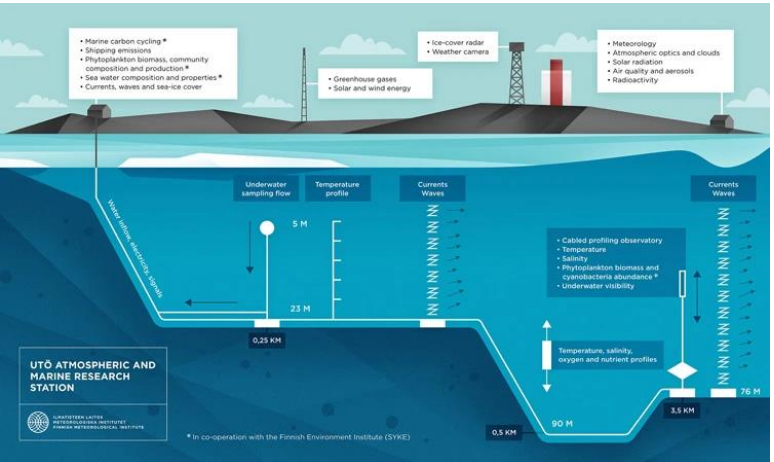
Imaging FlowCytobot - IFCB

- Imaging flow cytometer
- Takes images of phytoplankton cells and colonies inside the size range of 10-150µm
- Can be operated remotely mounted on a flow through system at a research station (like in Utö), ship of opportunity, put to measure directly to the sea or used in a laboratory environment etc..
- Takes a sample of 5ml with approx. 20 min interval
- The camera is triggered by chlorophyll-*a* or scatter
- Even as many as ~30 000 high resolution images / hour
- 150 µm mesh in IFCB inlet to prevent it from clocking
- Data analyzed with image recognition algorithm

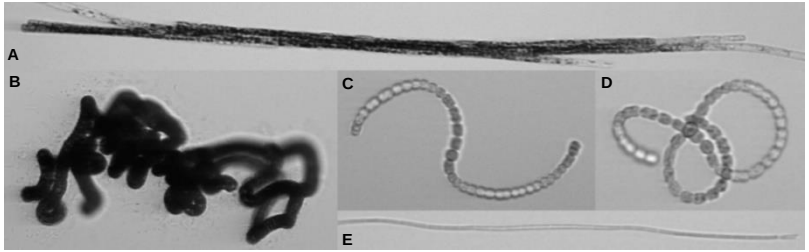


UTÖ Marine Research Station joint with **Finnish Meteorological Institute**

- Underwater pump with inlet at 5 m depth, 250 m offshore
- Water distributed to different sensors inside the station cabin
- Represents pelagial community of a mixed surface layer in brackish environment (salinity ~ 6 psu)
- Multiple parallel measurements from sea to atmosphere
- Continuous imaging flow cytometer observations from multiple years
- Light microscopy samples of cyanobacteria bloom in 2018

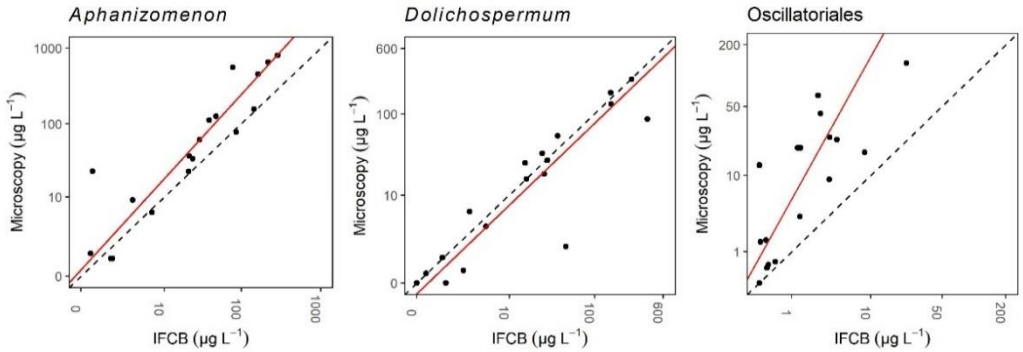


Cyanobacteria blooms in the Baltic Sea



Common filamentous species

- *Aphanizomenon flosaquae* (A)
- *Dolichospermum* spp. (C,D)
- *Nodularia spumigena* (B)
- Oscillatoriales (E)



(Kraft et al. 2021 Front Mar Sci)

Image analysis was performed using the publicly available code by H. Sosik et al. (Sosik & Olson 2007)

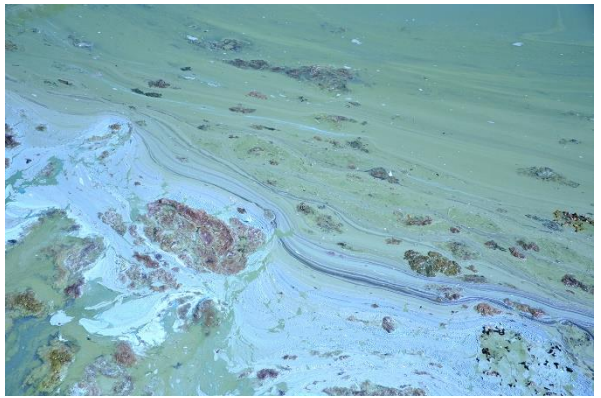
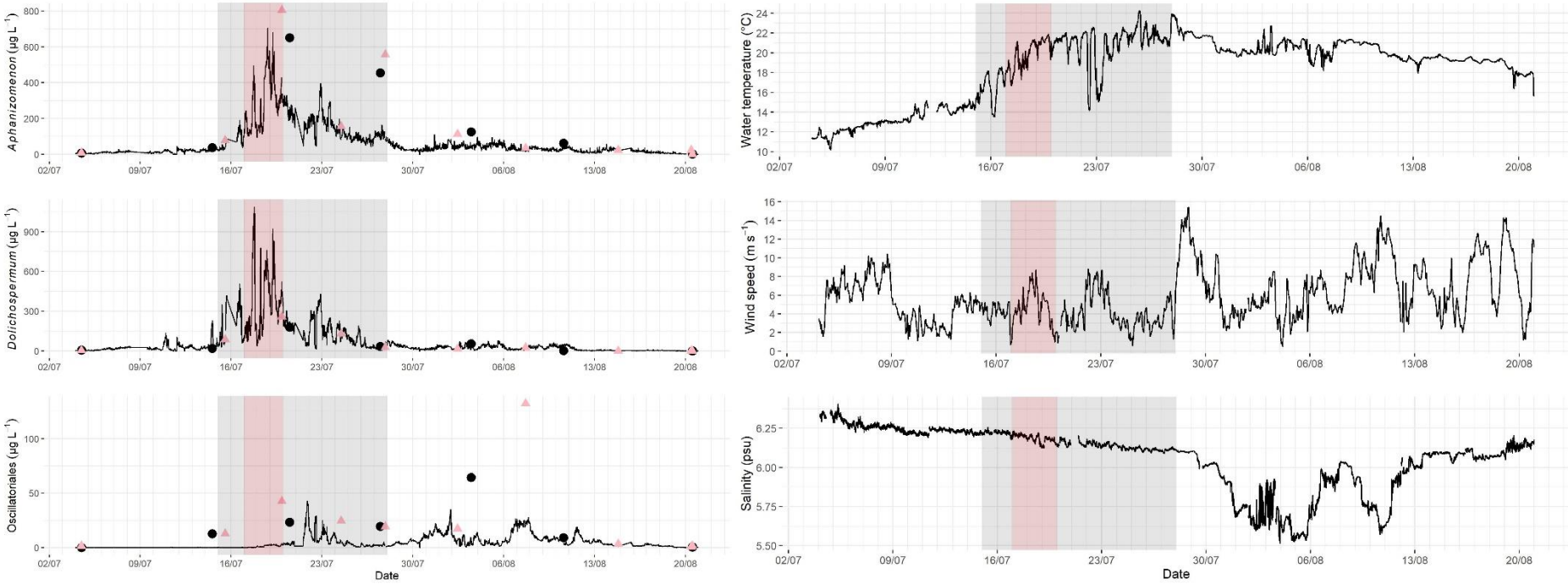


Photo by Lauri Laakso

High frequency phytoplankton observations connected with changes in the environment can help to understand mechanisms of the blooms

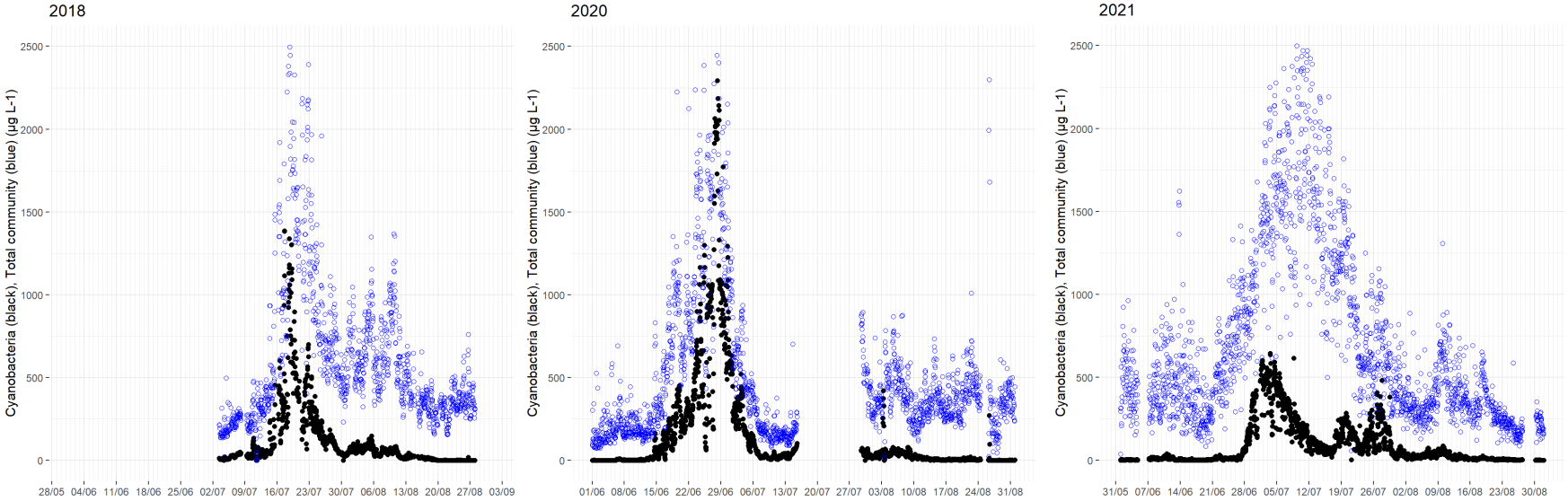


bloom period highlighted in grey (biomass > approximately 100 µg L-1)
bloom peak period highlighted in red (the days with the highest peaks)
light microscopy counts ●flow-through ▲pump inlet

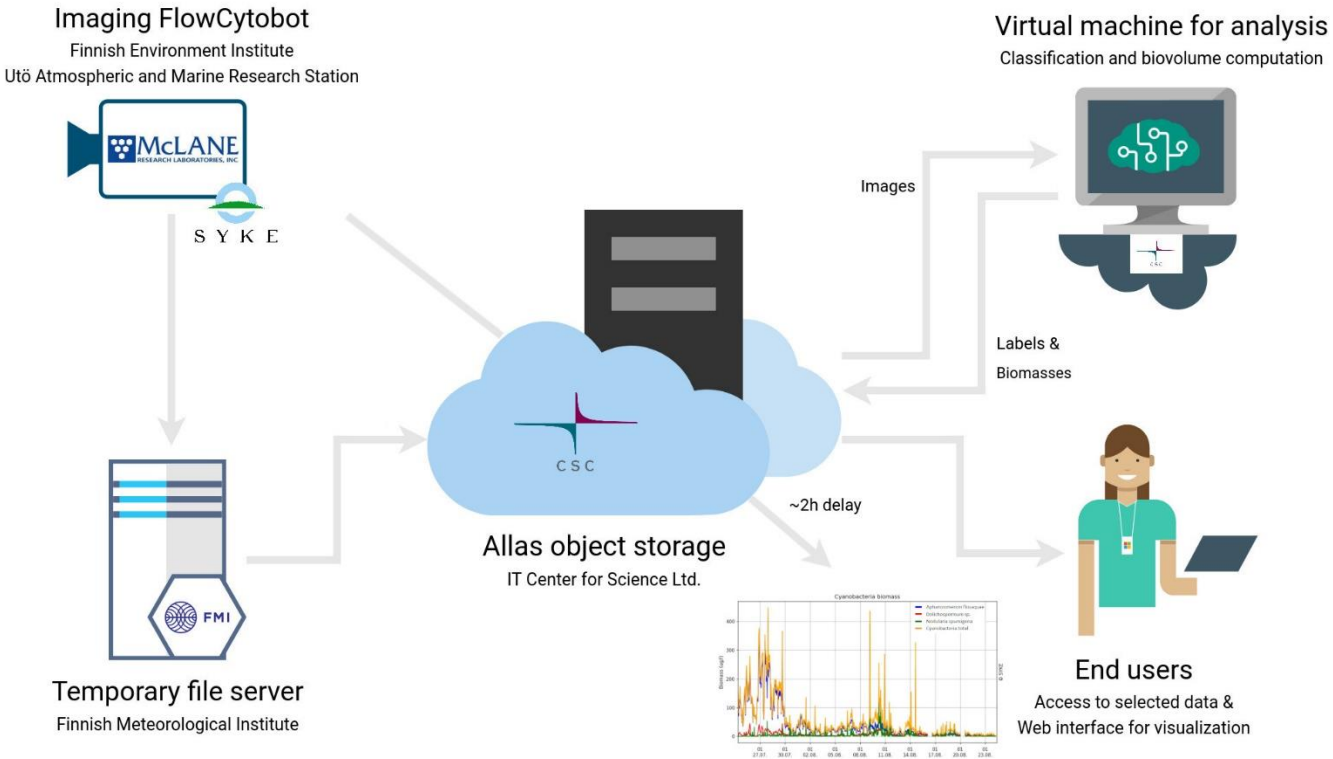
-> species dynamics were connected to changes in the water mass

(Kraft et al. 2021 Front Mar Sci)

Cyanobacteria blooms in 2018, 2020, 2021



Near-real time classification via Convolutional Neural Networks



- Training data set
- http://doi.org/10.23728/b2share.abf913e5a6ad47e6baa273ae0ed6617a
- Evaluation data set
- http://doi.org/10.23728/b2share.7c273b6f409c47e98a868d6517be3ae3
- Visualization still under development
- https://plankton.live/
- https://ifcb.plankton.live/timeline?dataset=uto
- https://swell.fmi.fi/hab-info/

https://github.com/veot/syke-pic, Kraft et al. 2022 Front Mar Ski

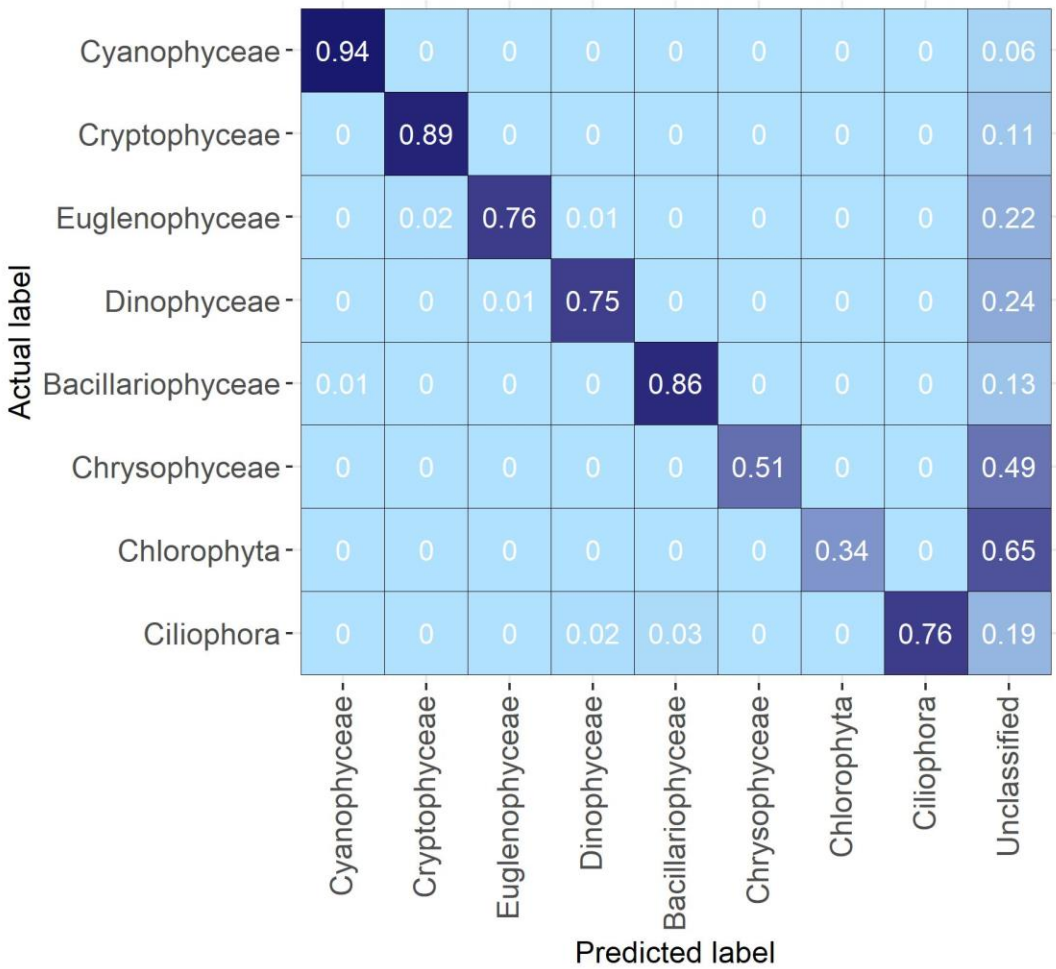
Classifier performance

- Kraft et al. 2022 Front Mar Sci
- Weighted F1-score of test data of our labeled image data set 0.95
- Weighted F1-score of our evaluation data (59 natural samples annotated entirely) 0.83
- Class-specific thresholds are used for filtering out the unidentifiable images, determined with our test set complemented with unidentifiable images (initial situation -> to be finetuned as data accumulates)
- ND in the Evaluation Data means “Not Determined”, the metrics were not calculated for classes with <10 images

	Training	Validation	Test				2021 data			
Class / taxonomic group	N	Thre	N	Pr	Re	F1	N	Pr	Re	F1
Oscillatoriales	2664	0.31	888	0,99	1,00	0,99	3893	0,98	0,98	0,98
Monoraphidium contortum	196	0.69	66	0,98	0,98	0,98	439	0,99	0,96	0,97
Skeletonema marinoi	2477	0.46	825	1,00	0,99	0,99	7402	0,99	0,94	0,97
Heterocapsa triquetra	1966	0.39	655	0,98	0,97	0,97	2267	0,92	0,95	0,94
Cryptophyceae / Teleaulax sp.	4098	0.53	1366	0,96	0,97	0,96	16952	0,97	0,90	0,93
Aphanizomenon flosaquae	4193	0.24	1398	0,97	1,00	0,98	1849	0,87	0,98	0,92
Peridiniella catenata chain	116	0.7	38	0,97	1,00	0,99	89	0,99	0,87	0,92
Dolichospermum sp. / Anabaenopsis sp.	7368	0.38	2456	0,98	0,99	0,98	790	0,88	0,96	0,92
Pauliella taeniata	71	0.62	24	1,00	0,96	0,98	56	0,96	0,86	0,91
Oocystis sp.	505	0.5	169	0,88	0,93	0,90	161	0,91	0,89	0,90
Mesodinium rubrum	679	0.44	227	0,96	0,95	0,96	560	0,92	0,86	0,89
Melosira arctica	26	0.3	8	0,73	1,00	0,84	58	0,85	0,91	0,88
Dolichospermum sp. / Anabaenopsis sp. coiled	1502	0.41	501	0,93	0,96	0,95	70	0,74	0,99	0,85
Eutreptiella sp.	1348	0.43	450	0,95	0,94	0,94	1678	0,90	0,76	0,83
Licmophora sp.	44	0.43	15	1,00	0,80	0,89	78	0,90	0,77	0,83
Nodularia spumigena	101	0.32	34	0,80	0,94	0,86	62	0,80	0,85	0,83
Heterocapsa rotundata	368	0.56	123	0,84	0,90	0,87	2609	0,89	0,70	0,78
Ceratoneis closterium	27	0.41	9	1,00	1,00	1,00	75	0,68	0,91	0,78
Peridiniella catenata single	539	0.52	180	0,89	0,97	0,93	222	0,75	0,81	0,78
Pennales thick	126	0.37	42	0,93	0,88	0,90	1088	0,72	0,85	0,78
Thalassiosira levanderi	1522	0.63	508	0,95	0,95	0,95	2008	0,87	0,68	0,77
Chaetoceros sp. chain	829	0.51	277	0,93	0,95	0,94	693	0,76	0,77	0,76
Centrales	288	0.51	96	0,98	0,89	0,93	92	0,77	0,68	0,72
Dinophysis acuminata	130	0.68	44	0,98	0,91	0,94	17	0,79	0,65	0,71
Pennales thin	469	0.29	156	0,96	0,99	0,97	334	0,61	0,84	0,71
Cyclotella choctawhatcheeana	61	0.47	21	0,89	0,81	0,85	199	0,92	0,57	0,71
Gymnodiniales	41	0.29	14	0,92	0,86	0,89	38	0,78	0,64	0,70

Table contains results of evaluation data for F1 > 0.7

Group specific confusion of our evaluation data



Kraft et al. 2022 Front Mar Ski

What can we identify? ~50 categories, ~60000+ annotated images

Taxons / groups so far identified with IFCB from the Baltic Sea (work ongoing)

Cyanophyceae
Aphanothece paralleliformis
Chroococcales
Chroococcus spp.
Merismopedia spp.
Snowella spp. / *Woronichinia* spp.
Oscillatoriales
Aphanizomenon flosaquae
Aphanizomenon spp.
Dolichospermum spp. /*Anabaenopsis* spp.
Nodularia spumigena

Cryptophyceae
Cryptomonadales
Teleaulax spp.

Dinophyceae
Prorocentrum cordatum
Dinophyceae (under & over 20µm)
Dinophysis acuminata
Dinophysis norvegica
Gymnodiniales
Gymnodinium spp.
Gymnodinium like cells
Gyrodinium spp.
Heterocapsa rotundata
Heterocapsa triquetra
Amylax triacantha
Gonyaulax verior
Peridiniella catenata
Protoceratium reticulatum

Chrysophyceae
Dinobryon balticum
Uroglenopsis spp.
Apedinella radians
Pseudopedinella spp.

Diatomophyceae
Chaetoceros spp.
Chaetoceros danicus
Chaetoceros similis
Chaetoceros subtilis
Chaetoceros throndsenii
Coscinodiscus granii
Cylindrotheca closterium
Cyclotella choctawhatcheeana
Melosira arctica
Skeletonema marinoi
Thalassiosira baltica
Thalassiosira levanderi
Diatoma tenuis
Licmophora spp.
Nitzschia paleacea
Pauliella taeniata

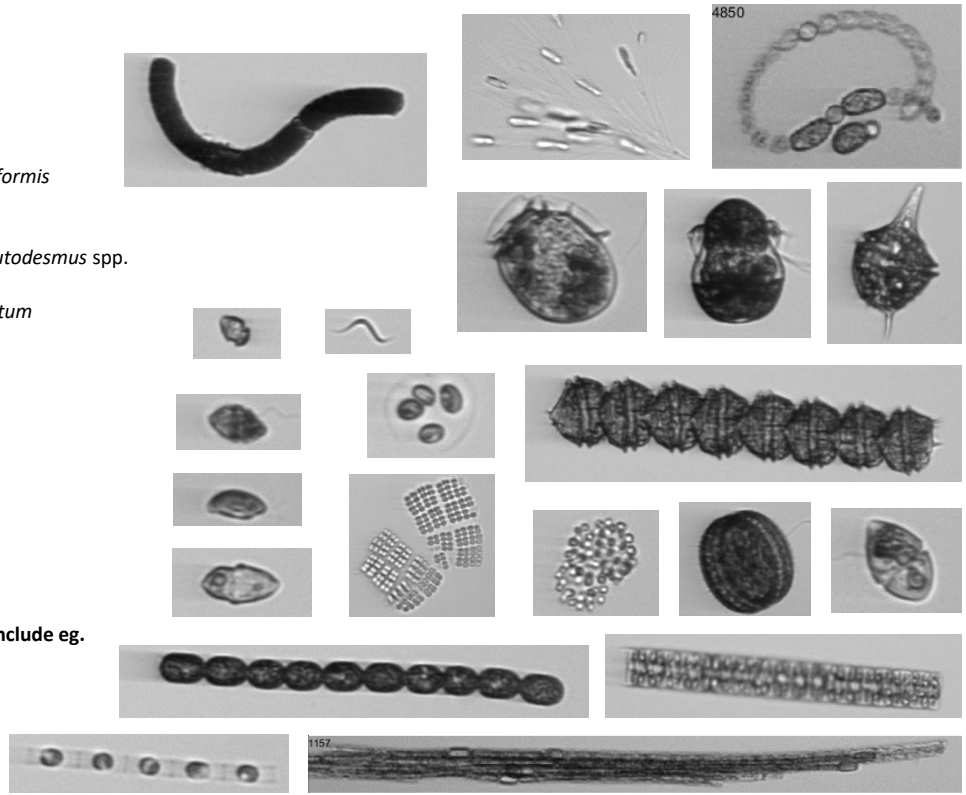
Euglenophyceae
Eutreptiella spp.

Chlorophyta
Cymbomonas tetramitiformis
Pyramimonas spp.
Chlorococcales
Scenedesmus spp. / *Acutodesmus* spp. / *Desmodesmus* spp.
Monoraphidium contortum
Binuclearia lauterbornii
Oocystis spp.

Katablepharis remigera
Flagellates
Nanoplankton

Ciliates
Strombidium spp.
Mesodinium rubrum

Additional categories include eg.
akinetes
heterocytes
resting stages
cysts



References and sources

- Kraft et al 2021. First application of IFCB high-frequency imaging-in-flow cytometry to investigate bloom-forming filamentous cyanobacteria in the Baltic Sea. Frontiers in Marine Science, 282.
- Kraft et al. 2022. Towards operational phytoplankton recognition with automated high-throughput imaging, near real-time data processing, and convolutional neural networks. Frontiers in Marine Science, 9.
- Sosik & Olson 2007. Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry. Limnology and Oceanography: Methods, 5(6), 204-216.
- Olson & Sosik 2007. A submersible imaging-in-flow instrument to analyze nano-and microplankton: Imaging FlowCytobot. Limnology and Oceanography: Methods, 5(6), 195-203.

- https://en.ilmatieteenlaitos.fi/uto
- https://www.finmari-infrastructure.fi/field-stations/uto-fmi/
- https://mclanelabs.com/imaging-flowcytobot/