Ordinal METAR Cloud Cover Classification in Sentinel-2 Satellite Data

Markus Götz*†, Erik Wessel†, Uğur Çayoğlu†, Babak Jahani‡¶, Charlotte Debus†, Jan Cermak‡§, Achim Streit*†,
*Helmholtz AI
†Scientific Computing Center (SCC)
‡Institute of Meteorology and Climate Research – Atmospheric Trace Gases and Remote Sensing
§Institute of Photogrammetry and Remote Sensing
Karlsruhe Institute of Technology (KIT)
76344 Eggenstein-Leopoldshafen, Germany
{firstname.lastname}@kit.edu
¶SRON Netherlands Institute for Space Research
2333 CA Leiden, the Netherlands

***Abstract*—Accurate cloud cover assessment is crucial in several fields such as weather forecasting, climate science, agriculture, or energy system planning, precipitation pattern forecasting and aiding in early detection of extreme weather events. Despite the crucial data that weather stations provide about sky cloud coverage, their measurements are geographically localized and thus lack spatial coverage. Meteorological satellites on the other hand offer great potential to address this limitation by continuously scanning large areas in short periods of time. This work proposes a novel approach for predicting cloud cover in global satellite images by leveraging ordinal point labels from ground-based weather stations (METAR), rather than relying on spatially resolved cloud masks, and demonstrates the effectiveness of this approach using a rank loss-based convolutional neural network of the EfficientNet family. The model is trained in transfer learning approach on a custom-collected data set across selected regions in the continental United States of America. Using station measurements only, we achieve an F_1-score of up to 0.6 and a ranked-within-1-accuracy ranging from 93.5% to 99.1%. Supplementing the data with labels created by visual inspection to correct for station-satellite mismatches improves scores to 0.75 and 98.4% to 100%. The results imply significantly improved cloud cover assessment in regions without weather stations, extending the capabilities to monitor localized cloud patterns.**

***Index Terms*—Satellite Images, METAR, Machine Learning, Convolution Neural Networks, Ordinal Classification, Sentinel-2**

I. Introduction

Cloud processes are among the key uncertainties in our present-day understanding of the Earth's weather and climate system [1]–[3]. In addition, humans depend on precise cloud pattern determination in a wide variety of areas, for example in agriculture or renewable energy planning [4]. It is therefore paramount to be able to monitor and forecast (long-term) cloud processes accurately. Simulating clouds, however, is non-trivial due to climate's chaotic nature. Small perturbations in initial conditions lead to vastly different outcomes [5] and thus introduce uncertainties. More so, anthropogenic changes to Earth's climate are directly affecting cloud patterns. Rising global temperatures for example increase atmospheric moisture and thus lead to different formation or stronger precipitation [6]. As a result, clouds remain the main reason for disagreements between climate models and observational data [5].

Consequently, the importance of satellite remote sensing data has increased significantly in recent years. Researchers now have access to large amounts of high-quality image data, facilitating data-driven analysis of Earth's atmosphere and surface, including clouds [7], [8]. Current data-driven and machine-learning-based approaches strongly focus on cloud segmentation, i.e., the creation of a per-pixel mask in (a series of) images. Albeit highly precise, the segmentation ground truth is costly to obtain and therefore usually geographically sparse, unable to fully capture cloud cover variances in their geographic heterogeneity [9]. Instead, it may be meaningful to reconsider cloud coverage prediction in a data-centric manner [10], i.e., systematically re-engineering input data to obtain higher modeling effectiveness and efficiency. One possible way to achieve this objective is to trade spatial labeling accuracy for global assessment coverage rate.

Hence, this work investigates the usefulness of predicting cloud cover in global Sentinel-2 satellite images by leveraging ordinal point labels from METAR measurements of ground-based, local weather stations. While spatially less resolved than cloud masks obtained from a meteorological satellite, weather stations have the inherent advantage of being globally distributed, available in high numbers with a high update frequency of one hour or less. Our contributions can be summarized as follows:

- the creation of a publicly available dataset for METAR cloud cover classification;
- a rank loss-based convolutional neural network for ordinal predictions using transfer learning;

Markus Götz and Erik Wessel contributed equally. Markus Götz is the corresponding author.

Manuscript received MM DD, YYYY; revised MM DD, YYYY.

- a detailed analysis of the predictive performance of the machine learning model.

The remainder of this paper is organized as follows. Section II introduces related work in machine learning cloud data analysis. In Section III we explain the collection methodology and characteristics of the cloud cover data. Subsequently, Section IV details the neural network architecture and the adjustments we made to its training procedure. In Section V an ablation study is present in which the impact of all modifications in relation to the predictive performance are evaluated. Finally, Section VI concludes the paper.

II. Related Work

Cloud detection in satellite images can be distinguished in two major methodological families. On the one hand, there are single-image algorithms that interpret individual images to obtain cloud information. The simplicity of single-image approaches tends to reduce implementation complexity, computational demands and make them easy to adapt to new data sources. Early cloud detection algorithms, e.g., Automated cloud cover assessment for Landsat TM images (ACCA) by Hollingsworth, Chen, Reichenbach, *et al.* [11] for Landsat satellite images, use physics-motivated thresholding rules on n-dimensional histograms. Similarly, the Fmask algorithm by Zhu and Woodcock [12] produces per-pixel masks for clouds, cloud shadows, and potentially snow in Landsat L1T imagery by binning manually engineered features of the spectral bands. In particular, they consider the Top Of the Atmosphere (TOA), Brightness Temperature (BA), Normalised Difference Vegetation Index (NDVI) and Normalised Difference Snow and Ice Index (NDSI) values incorporating geometrical arrangement. Building on this, Zhu, Wang, and Woodcock [13] introduce Fmask-2 mapping the approach to Sentinel-2 images and highlight the advantage of including additional spectral band for the detection of cirrus clouds in contrast to the thermal information provided by Landsat and earlier missions. Fmask-4 by Qiu, Zhu, and He [14] makes incremental improvements to this approach, achieving a per-pixel prediction accuracy of 94.59% for Landsat 8 and 94.30% for Sentinel-2 images.

ATCOR by Richter and Schläpfer [15] is a related thresholding approach on physical properties. Unlike the Fmask variants, ATCOR first classifies each pixel with respect to its land-, respectively cloud-cover, type leveraging spectral information in the visible, near-infrared shortwave infrared bands. Subsequent atmospheric corrections, using the radiative transfer equation, remove atmospheric scattering and absorption effects, followed by topographic correction accounting for differences in elevation and slopes of the observed terrain. ESA's Sen2Cor algorithm [16] focuses on the removal of atmospheric correction effects and is used to correct Sentinel-2 Level-1C images, i.e., co-located and calibrated top-of-atmosphere reflectances, to Level-2A images, i.e., first-order derived products. Sen2Cor classifies pixel values in several categories, including cloud and cloud shadow. In a study by Zekoll, Main-Knorn, Alonso, *et al.* [17] the performance differences between masking models are compared, reaching overall accuracies of 89% for Fmask, 91% for ATCOR and 92% for Sen2Cor on Sentinel-2 level-1C test data.

On the other hand, multi-view algorithms facilitate cloud detection through observation of the same spatial area over a certain time period of time which needs to be short enough so that changes in the cloud properties are negligible, to leverage cloud dynamics in contrast to Earth's static surface. Overall multi-image algorithms tend to provide better results, i.e., are more accurate and robust, as they take into account the bidirectional reflective properties of the observed scene. However, they usually have a higher computational demand and necessitate intricate data curation due to their higher complexity, e.g., providing a constant frequency time dimension without missing data. Nonetheless, thanks to their robustness, the interest for such approaches seems to be increasing in the atmospheric science community, as Multi-Angle Polarimeters (MAP) are already included in a number of newly-launched and planned for additional future satellite missions, e.g., Plankton, Aerosol, Cloud, Ocean Ecosystem (PACE) [18], Coperni-

Approach	Data Source	Granularity	Throughput	Advantages	Disadvantages
Ours	satellite images (visible and SWIR), METAR reports	per-image $32\,\mathrm{km}\times32\,\mathrm{km}$ (Medium)	High	+ scalablability, near-real-time + open data only + no per-pixel annotation required + only images needed inference	- no per-pixel data - irregular METAR station distribution and varying quality
Fmask 4.0	Landsat/Sentinel-2 images	per-pixel (High)	Low	+ high accuracy, mature method + spectral-contextual features	- bright surface misclassification - computationally intensive
ATCOR	satellite images	per-pixel (High)	Low	+ atmospheric correction + high accuracy + terrain elevation considered	- relies on additional accurate atmospheric data - computationally intensive
Sen2Cor	Sentinel-2 images	per-pixel (High)	Low	+ cloud and atmospheric correction + scene classification	- limited to Sentinel-2 - computationally intensive
MACCS-MAJA	multi-temporal satellite images	per-pixel (High)	Very low	+ atmospheric correction + high-altitude compensation + cirrus cloud detection	- high complexity - requires accurate timing data - computationally intensive
MTCD	multi-temporal satellite images	per-pixel (High)	Very low	+ higher accuracy for areas with frequent cloud cover	- requires specific timescale - struggles with sudden clear sky - computationally intensive

TABLE I: Comparison of cloud detection approaches from satellite images.

cus Anthropogenic Carbon Dioxide Monitoring (CO2M) [19], and Metop-SG satellite [20].

The multi-sensor atmospheric correction and cloud screening (MACCS) [21] for example is a multi-temporal approach that is able to identify clouds and cloud shadows next to other semantic image segments, while also considering atmospheric corrections. Building on that, the MACCS-ATCOR joint algorithm (MAJA) by Lonjou, Desjardins, Hagolle, *et al.* [22] combines MACCS and ATCOR [15]. MAJA is more than the sum of its parts, as it additionally includes correction methods for haze and cirrus clouds. Furthermore, it also introduces a way to mix image sets of different missions, with the goal of combining data from the Sentinel-2 and Landsat 8 satellites.

The Multi-Temporal Cloud Detection (MTCD) model developed by Hagolle, Huc, Pascual, *et al.* [21] uses satellite images at constant viewing angles to better differentiate between cloudy and clear sky conditions per pixel. It leverages imagery from the FORMOSAT-2 and Landsat missions and focuses on abrupt changes in reflectance values between pixels, additionally linearly correlating with neighboring pixels. In order to handle landscape change scenarios, e.g., melting snow fields, the algorithm necessitates an adequate timescale able to captures these gradual changes. This timescale varies strongly based on local conditions, such as the observed season or short-term changes in land formations. Inaccurate capturing will degrade temporal consistency and leads to incorrect interpretations of the segments.

In this study we will limit ourselves to a single-view study first due to its lower analysis complexity. However, it is in principle conceivable to expand the proposed work to a multi-view setting and use insights from, e.g. MAJA or MTCD. A comparison of modeling approaches for detecting cloud coverage can be found in Table I. The novelty of our approach is its use of ordinal point labels instead of costly per-pixel cloud segmentation. For these ordinal measurements, we use widely available weather station cloud cover information in the form of METARs. During inference, only the satellite image slices are required, since the ordinal METAR cloud cover labels are derived by the model. Although our approach does not provide the same level of detail as others, it is scalable and can be used for near-real-time applications over a broad area by processing it in tiles of $32\,\mathrm{km}\times32\,\mathrm{km}$. In this way, our approach highlights a cost-effective and efficient solution for cloud cover detection that is even applicable to run on consumer-grade hardware.

III. METAR Cloud Cover Dataset

A. Sentinel-2 Satellite Imagery

The Sentinel-2 mission is a satellite-based Earth observation program developed by the European Space Agency in line with the Copernicus program. It collects high-resolution images on a global scale using a pair of satellites orbiting Earth on opposing sides. The program has a revisit-frequency of five days. Overall, the multi-spectral instruments measure 13 channels, covering the range from visible to short-wave infrared light at different resolutions. The highest resolution of 10 meters-per-pixel is reserved for the visible and near-infrared light bands. Additionally, there are six bands at a resolution of 20 meters-per-pixel, with four in the visible and near-infrared (VNIR) range and two in the short-wave infrared (SWIR) range for applications like clouds, snow, and ice detection as well as vegetation observation. The remaining three bands are provided at 60 meters-per-pixel and focus on the applications of cloud screening, atmospheric correction and cirrus detection.

In the following, we will utilize Sentinel-2 images as input data to a data-driven artificial intelligence model to assess cloud coverage. Out of the various post-processing levels of the offered product, we base our work on the most raw, i.e., Level-1C. It encompasses steps, such as radiometric corrections, refinement of geometric viewing model, resampling with geometric interpolation, conversion of values to reflectance, and generation of masks.

B. METAR Reports

Meteorological Aerodrome Report (METAR) is an internationally standardized format for reporting weather observations. It is mainly used by pilots, airports and aerodromes as well as meteorologists to obtain information on weather conditions at certain locations. However, METAR also has use cases outside of flight operation including, but not limited to, infrared meteorology [25], human sensory perception [26] or health issue due to urban overheating [27]. A map of the global distribution of METAR stations and for selected regions can be seen in Figure 1. Here, we process METAR data for satellite product training and evaluation similar to Cermak and Bendix [28] and Jahani, Karalus, Fuchs, *et al.* [29].

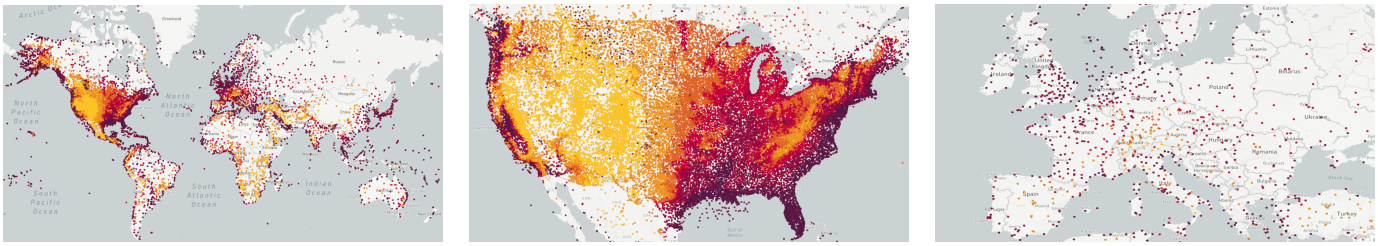
METAR records are usually issued half-hourly to hourly intervals and provide localized environmental conditions such as temperature, wind directions and speeds, precipitation, visibility, as well as cloud cover and height. An example METAR string is shown below with the cloud cover information highlighted in bold:

```
>> METAR KSFO 281356Z 26011KT 10SM  FEW002 BKN006 ...
   ...  OVC008 12/10 A2993
```

TABLE II: METAR cloud cover codes for different conditions and their human-readable interpretation.

Code	Interpretation	Oktas
NCD	Nil cloud detected due to an error in the sensors	nan
CLR	No clouds below 12 000 ft (≈3.700 m) (only USA)	0
FEW	Few	1–2
SCT	Scattered	3–4
BKN	Broken	5–7
OVC	Overcast, i.e., full cloud coverage	8

These values are communicated in eights (oktas) of cloud occupation of the sky above the reporting entity. The three letters indicate how much of the sky is covered by clouds, while the last three digits indicate the base cloud height in multiples of $100\,\mathrm{ft}$, ($\approx30\,\mathrm{m}$), above ground level. In the example, there are few clouds at $200\,\mathrm{ft}$ ($\approx61\,\mathrm{m}$), broken clouds at $600\,\mathrm{ft}$ ($\approx183\,\mathrm{m}$), and an overcast sky at $800\,\mathrm{ft}$ ($\approx244\,\mathrm{m}$). While the coverage values usually refer to the conditions



(a) World (b) Continental USA (c) Europe

Fig. 1: Maps with all weather stations providing METAR information listed in the IEM ASOS-AWOS service by the University of Iowa [24]. Coloring indicates elevation above ground level, darker is lower and brighter higher.

directly above the station, the lack of a detailed reporting guideline may also lead to spatial offsets. An exhaustive description of METAR cloud coverage is depicted in Table II. In the following we will use the METAR codes, ranked flow lowest to highest cloud coverage, as ordinal classes to be predicted.

C. Data Compilation

We compile the data for this work from multiple sources. The input data for the model are Sentinel-2 Level-1C satellite image products obtained from the Copernicus Open Access Hub [30], while the prediction targets are METAR reports obtained from the Iowa State University, Iowa Environmental Mesonet (IEM) – ASOS-AWOS database [24]. On a technical level, each data retrieval client is designed as a separate service within a microservice architecture.

The data access workflows first query metadata before downloading the actual data product. In case of Sentinel-2, each spectral band can be retrieved in JPEG-2000 format. The crawler obtains square patches with an image height, respectively width, corresponding to a parametric geospatial search diameter and the center pixel registered on the issuing METAR station. METAR itself is retrieved as densely folded data packages for individual dates. In a post-processing step, the data is unfolded, duplicates removed, missing dates interpolated, and finally cached in a local database.

In line with this study, we focus our data compilation on sub-regions of the continental United Stated of America due to 1) a high density of METAR observation stations (see Figure 1), 2) a limit in special extent and therefore download data [1] as well as 3) a high variability of topological and climate conditions in accordance with the Köppen-Geiger classification [31], [32]. In particular, we consider the regions California, Florida, Montana, Texas, and Washington due to their distinct features. For each issued METAR report of a station in these regions, we obtain the time-wise closest Sentinal-2 product with a maximum temporal offset of $\pm 30\,\mathrm{min}$. A grand total of 12,749 image-annotation pairs are retained in the dataset.

IV. Cloud Cover Assessment

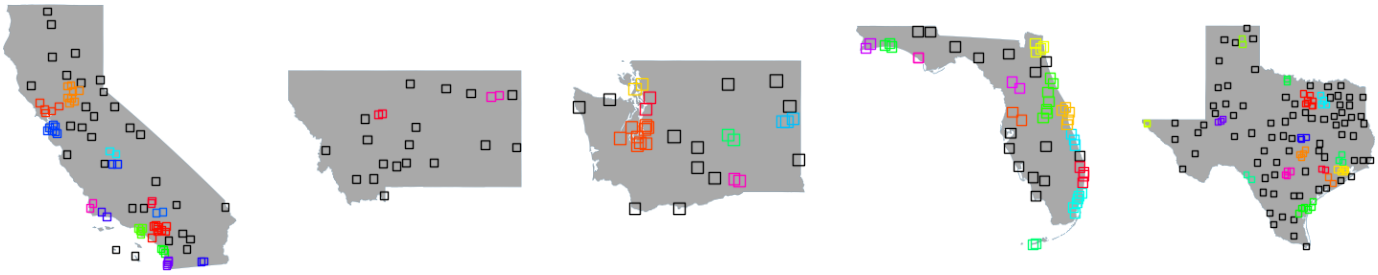
A. Modeling with Convolutional Neural Networks

Convolutional Neural Networks (CNN) are the de facto standard for supervised image classification problems using deep learning approaches. In this type of artificial feed-forward neural network, image features extractors, the convolution filters or kernels, are learned during the training stage and then used to derive rich representations for an attached classifier. Various CNN architectures have been proposed over the last decade. The current state-of-the-art is the EfficientNet model family [33], [34]. Overall, these have high training speed and parameter efficiency compared to previous models. This is achieved through a combination of training-aware neural architecture search and data-dependent hyperparameter scaling.

For all experiments in this work, we use the small EfficientNetv2s with roughly 20 million parameters and six convolutional blocks [34]. In a transfer learning approach, we initialize the models for this study with the weights from an image classifier pretrained on RGB images of the ImageNet-1k dataset [35]. This requires us to map the multi-spectral Sentinel-2 satellite images to the three-channel true color image input of the CNN. For this, we use the bands B4, B3 and B2 with central wave lengths of $664.6\,\mathrm{nm}$, $559.8\,\mathrm{nm}$ and $492.7\,\mathrm{nm}$ respectively, corresponding to the colors red, blue, and green of the human visual spectrum [30]. During fine-tuning of the pretrained model, all weights are unfrozen and may be adjusted during the training procedure. The classification layer is randomly reinitialized, following a He-initialization [36], and the initial output dimension for 1,000 classes reduced to five for all considered classes. We optimize the parameter with the Adam optimizer [37] with a learning rate $lr = 10^{-3}$ and momentum $m = 0.9$ at a batch size of 48 images.

Unlike traditional image classification problems in which the classes to be predicted have no inherent relation to one another, the METAR cloud cover codes form a partial ordering (see Section III). Hence, for our model we do not minimize the categorical cross-entropy as the target loss function. Instead, if a data sample x belongs to class c, it is automatically also classified into all lower-ordered categories $(0, 1, \ldots, c-1)$ as well. Hence, the target vector of the class SCT for example is $y = (\mathrm{CLR} = 1, \mathrm{FEW} = 1, \mathrm{SCT} = 1, \mathrm{BKN} = 0, \mathrm{OVC} = 0)$,

[1] Both services are quota limited.



(a) California (b) Montana (c) Washington (d) Florida (e) Texas

Fig. 2: Visualization of the output of the bias determination algorithm. Black patches indicate no overlaps with neighboring weather stations. Different colors indicate joint cluster regions.

where for all index positions $0 \leq i \leq (c = 2)$ the vector is set to 1 and for all higher-ordered classes to 0. During training, we minimize the mean-squared-error between the neural network's output $\hat{y}$ and the aforementioned target vector y. During inference time an actual cloud cover class is determined as the highest index of $\hat{y}$ exceeding the cut-off threshold of 0.5 irrespective of potential intermediate gaps on lower indices. Similar to prior work, e.g. NNRank [38], we find this approach to consistently performs better compared to a one-hot-encoded target class vector or a regression of the class rank.

B. Bias-free Sampling

We divide the data with a typical 80:10:10 percentage split for training, validation, and test set respectively. Therein, it is important to avoid information leakage, i.e., the use of information during model training which would not be expected to be available at prediction time as it would cause predictive scores to overestimate the model's utility when run at inference time [39]. This may happen, if two satellite image tiles spatially overlap when centered around two neighboring METAR stations.

Hence, we adopt a sampling strategy that maintains information independence by assigning jointly biased image sets exclusively to either one of the three dataset splits [40]. Practically, this is achieved via a modified DBSCAN clustering [41], [42] with the Chebychev distance metric and search radius ε equal to the maximum of the width and height of the satellite image patches. A visualization of the colored clusters identified in this manner is depicted in Figure 2.

Subsequently, a greedy assignment strategy places all identified clusters in the respective subset with the largest number of missing items with respect to the desired split fraction first. Then, all remaining patches with no overlaps, colored in black in Figure 2 are assigned randomly in stratified manner, i.e., the resulting class distribution in each of the splits is roughly equal.

C. Satellite Image Preprocessing with the Normalized Difference Snow Index (NDSI)

One of the crucial challenges in cloud cover prediction in satellite images is to distinguish between the clouds themselves and high-reflective surfaces with similar properties in the visual light spectrum, e.g., ice and snow covered land, as well as deserts [11], [12], [16], [43], [44]. A common approach is to preprocess the data in the visual light spectrum by multiplying it with the *Normalized Difference Snow Index (NDSI)* [45] as follows:

$$NDSI = \frac{X_G - X_S}{X_G + X_S} \tag{1}$$

$$X^* = [X_R, X_G, X_B] \cdot (1 - NDSI) \tag{2}$$

where X is a number of (hyper-)spectral input images, R, G and B denote the red, green and blue color channels in the visual light spectrum ($400\,\mathrm{nm}$ to $700\,\mathrm{nm}$) and S the shortwave infrared (SWIR) band, around $1610\,\mathrm{nm}$, B11 for Sentinel-2. The resulting images X^* structurally resemble the originals while drastically improving contrast between clouds and the surface. Examples of pre-processed NDSI images are depicted in Figure 3. This approach is particularly useful for feeding the multispectral satellite images in the transfer learning approach to deep neural networks that have been pre-trained on RGB images only (cf. Section IV).

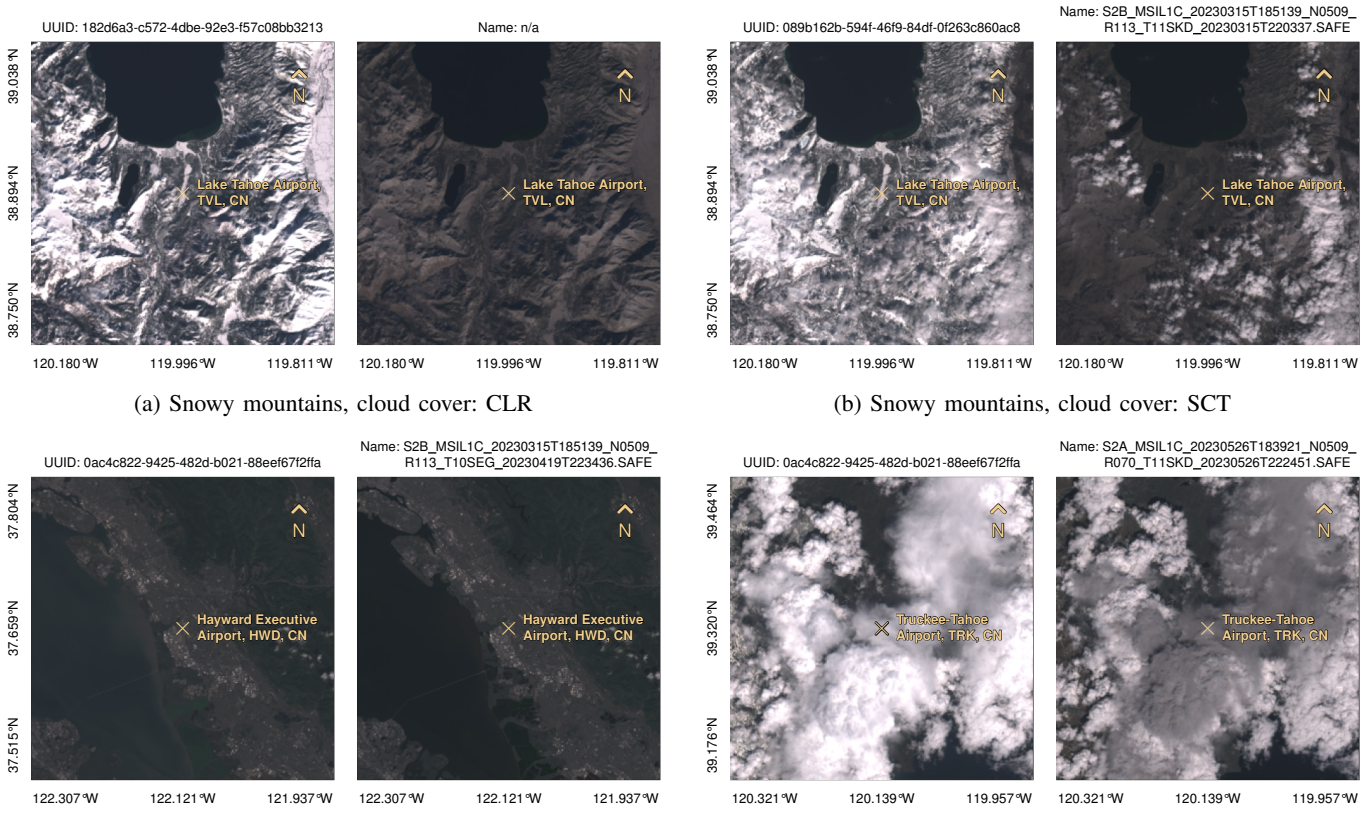
V. Experimental Evaluation

A. Implementation

The experiments were run on the HDFML cluster system comprised of 15 compute nodes with commodity components. Each node is equipped with two 12-core Intel Xeon Gold 6126 CPUs, 206 GB of DDR3 main memory, and four NVIDIA Tesla V100 SXM2 GPUs with 32 GB VRAM per card. The GPUs communicate node-internally via an NVLink interconnect. The system is optimized for GPUDirect communication across node boundaries, supported by 2x Mellanox 100 Gbit EDR InfiniBand links. The corresponding software is written in the Python programming language [46] and is publicly available from the authors' source code repository [3]. It contains a microservice-based backend for data acquisition and preprocessing of both, Sentinel-2 satellite images and METAR annotations. The neural network is implemented in PyTorch [47]. The checkpoints for the trained models can be found online[4]. The model weight files start

[3]https://github.com/ErikWessel/MSCC/

[4]https://doi.org/10.5281/zenodo.14750977



(a) Snowy mountains, cloud cover: CLR

(b) Snowy mountains, cloud cover: SCT

(c) Snowless coastline, cloud cover: FEW

(d) NDSI discolored clouds, cloud coverage: BKN

Fig. 3: Examples of Sentinel-2 image patches centered above a weather station providing METAR cloud cover information. For each image set, left: default RGB product, right: RGB product multiplied with the NDSI value. All images are supplied in the dataset of this paper.

with `efficientnet_v2_s_final_16km_300_ndsi_ord_regr_automatic_labels.pt` has been used for all experiments in Section V-C and `ord_regr_manual_labels.pt` for all experiments in Section V-D.

B. Metrics

To evaluate our model we employ a mixture of metrics for arbitrary and ordinal classification problems. For the former we determine the true positive (TP), false positive (FP) and false negative (FN) rates. These allow us to derive the precision $P = TP/(TP + FP)$, i.e., the proportion of positive predictions that were actually correct, and the recall $R = TP/(TP + FN)$, i.e., the fraction of positives that have been correctly identified. The harmonic between the two quantities is the F_1 score and assess overall classification quality, ranging from 0 to 1, higher is better:

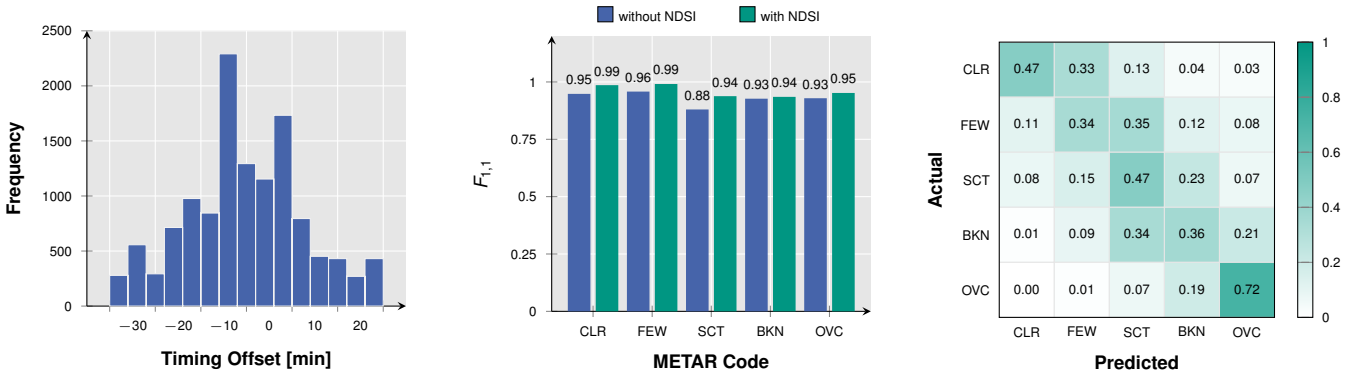
$$F_1 = 2\frac{P * R}{P + R} \tag{3}$$

In ordinal classification problems, a model's prediction may be incorrect if compared directly against the desired output, but still within an acceptable margin given the entire partial class ordering. In our case, for example, predicting scattered cloud coverage in a satellite image patch with only a few clouds could be considered reasonably accurate. Ranked-within-n-accuracy ACC_n is a family of scoring metrics that allows a wider range of ordinal outputs to be considered correct [48]. A formal definition can be found in Equation (4) and Equation (5), where Y is the ground-truth, $\hat{Y}$ a model's prediction and N their cardinality.

$$within_n(y, \hat{y}) = \begin{cases} 1 & |\arg\max_i[y_i = 1] - \\ & \quad \arg\max_j[\hat{y}_j = 1]| \leq n \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$ACC_n(Y, \hat{Y}) = \frac{1}{N}\sum_{i=1}^{N} within_n(Y_i, \hat{Y}_i) \tag{5}$$

It is clear from this formulation that ranked-within-0-accuracy ACC_0 defaults to the standard accuracy metric. Similarly, we can derive a modified version of the $F_{1,n}$ score, that uses a $within_n$ for precision P_n and recall R_n for $n > 0$. Hence, for the purpose of this work we will consider the ranked-within-1 F-score ($F_{1,1}$) to be our default scoring metric.



(a) Distribution of closest timing offset between satellite images and METAR.

(b) $F_{1,1}$ values per class with and without NDSI preprocessing on METAR labels.

(c) Confusion matrix for ACC_0 with NDSI preprocessing.

Fig. 4: Evaluation of the EfficientNetv2 ordinal regression model on the raw METAR ground truth.

C. Results on METAR Labels

The best performing model purely based on METAR reports is presented in Figure 4. It showcases that it is in principle feasible to assess cloud coverage purely based on these label with high $F_{1,1}$ scores. Yet, accurate assessment, i.e., without any ordinal tolerance, is somewhat challenging with ACC_0 scores between 34% and 72%. While not displayed here, pure classification and class regression approaches fare even worse with $F_{1,1}$ scores as low as 0.651 and 0.871 and ACC_0 between 19% and 75%, respectively 7% and 48%.

A notable hyperparameter of the model training and prediction pipeline is the size of the satellite images patches. The originals are taken at a resolution of $10\,\mathrm{m}$ per pixel for a radius of $16\,\mathrm{km}$ resulting in a resolution of $3200\,\mathrm{px}\times3200\,\mathrm{px}$. For this study, we found it sufficient to downsample all images to $300\,\mathrm{px}\times300\,\mathrm{px}$. Higher resolution did not result in any noticeable increase in predictive performance, tested at $800\,\mathrm{px}\times800\,\mathrm{px}$ and the full $3200\,\mathrm{px}\times3200\,\mathrm{px}$ widths and heights.

On image sizes of $300\,\mathrm{px}\times300\,\mathrm{px}$ with three channels an inference step needs to perform 5.36 GMAC (giga multiply-and-accumulate operations) or around 10.72 GFlops (giga floating-point operations per second). On a single A100 GPU this results roughly in a throughput of 320 images per second without mixed-precision or pruning. Consequently, the proposed pipeline is suitable for high-performant in the loop predictions.

Leveraging NDSI-adjusted color channels consistently improve the model's predictive performance of METAR reports (see Figure 4 center). The $F_{1,1}$ values increase by up to six percentages points without any further changes. Equally, ACC_0 values in the confusion matrix, not display here, align closer to the main diagonal. While not the best performing models overall, the use of NDSI-adjusted data is equally useful for the classification and class regression approaches.

D. Results with Manual Relabeling

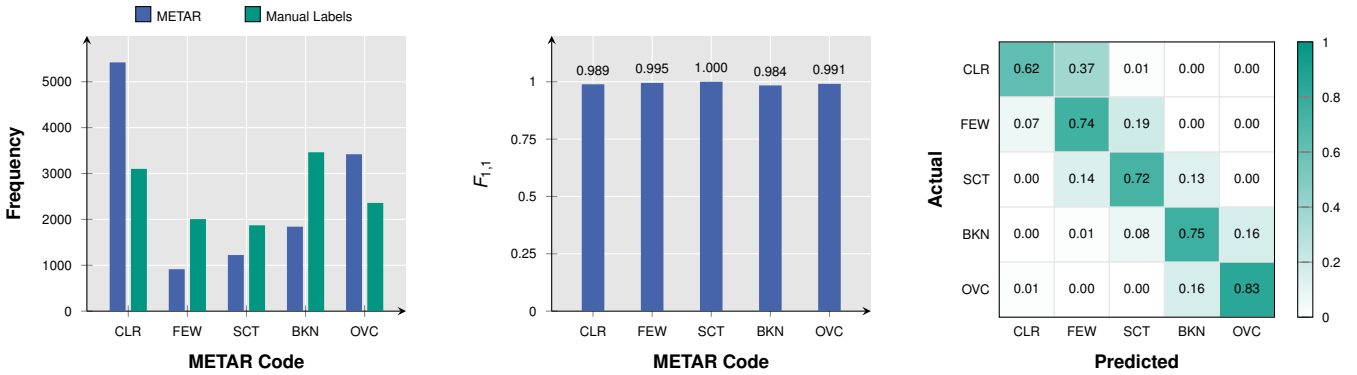
To judge the data quality of the METAR reports as well as their suitability to judge top-down cloud cover in satellite images from the ground, we additionally labeled all 12,749 images in the dataset manually [5]. The EfficientNet model checkpoints trained on the automatically obtained as well as the manually relabeled data is online (cf. Section V-A). We find that there is a non-negligible difference between these ground truth values, as can be seen in Figure 5a. Training on manual labels further improves the model's performance, as illustrated in Figure 5b and Figure 5c. $F_{1,1}$ scores are close to ideal, i.e., at 1.0, while ACC_0 now ranges between 62% and 83%. Notably, with very few exceptions, all predictions are in the band diagonal of the confusion matrix with bandwidth one.

We do not assume a systematic or intentional misreporting by the METAR stations. In fact, most mismatches are usually a singular ordinal class off and the average mismatch frequency per station is only around 15.9% as can be seen in Figure 6. We attribute this difference majorly to the large time offset of a maximum of 30 minutes between the Sentinel-2 satellite image capture and the corresponding METAR report. In this time frame, clouds may have undergone significantly changes, e.g., due to strong wind. It is, however, noteworthy that the raw METAR reports compared to the manual relabeling have a tendency to favor the outer extremes (see Figure 5, left). This may be attributed to selection biases of the human operated reporting systems [49].

E. Limitations

One of the main limitations of the presented approach is its applicability to short-term monitoring tasks. These depend directly on the revisiting frequency of the leveraged satellite programs and their data. In its current form, each Sentinel-2 satellite revisits a geolocation every 10 days resulting in a combined constellation revisit of 5 days. For tasks where a higher frequency is necessary, alternative data sources are more suitable. A satellite program offering these capabilities is ESA's ICEYE SAR constellation [50] with a revisiting duration of a single day.

[5]https://doi.org/10.5281/zenodo.14691473



(a) Cloud coverage distribution in METAR and manually relabeled.

(b) $F_{1,1}$ values per class with NDSI preprocessing on manual labels.

(c) Confusion matrix for ACC_0 with NDSI preprocessing.

Fig. 5: Evaluation of the EfficientNetv2 ordinal regression model on the manually labeled ground truth.

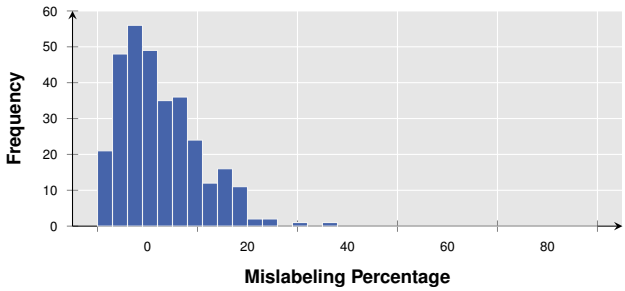


Fig. 6: Labeling mismatches frequency between METAR and satellite images per weather station.

VI. Conclusion

In this work, we have evaluated the viability of assessing cloud cover in satellite observations by supplementing them with ground-based METAR weather station measurements. One of the key objectives of the study was the evaluation of the generalization capabilities of the approach to data-sparse regions. We have phrased this problem as an ordinal classification problem of satellite image patches and modeled it using deep convolutional neural networks of fine-tuned architectures from the EfficiNet family.

There are several conclusions that we can draw from this work. First, ground-based labels are strongly informative for cloud cover assessment in top-down satellite products, confirmed by high predictive F_1 and ACC_1 performance metrics. Second, direct leveraging traditional preprocessing algorithms, e.g., NDSI, can significantly improve a model's forecasting capabilities, increase convergence rate and reduce computational cost. This is an interesting find because it highlights the potential of compressing additional information through manual feature engineering into deep learning transfer approaches without requiring to explicitly modify input dimensionality. Third, a coarse-grained image resolution of $300\,\mathrm{px}$, corresponding to a spatial resolution of $32\,\mathrm{km}\times32\,\mathrm{km}$, is sufficient to assess cloud cover information well. A finer resolution does not notably improve prediction metrics. At the same time it also does not reduce it, making it viable for high-resolution meteorological studies. Finally, the quality of METAR-provided varies strongly, partially containing inaccurate information. While deep neural networks are somewhat robust against such data noise, it demonstrates the necessity to carefully cleanse and refine METAR information before using it as ground-truth. Our study has demonstrated that cloud cover label corrections can account for more than 10 percentage points improvement.

There are several avenues for future work of this study. There are several possibilities to branch out in terms of data beyond the investigated continental United States of America, as well as the low-revisiting Sentinel-2. Furthermore, we have not yet considered local measurement information with respect to wind speed and direction. While potentially also false, this kind of data may hold increase forecast robustness and allow for internal and external cloud dynamics understanding. Additionally, this approach may be suitable for retrieving cloud fractions at the sub-pixel level based on newly-launched and planned Earth-observing satellite missions carrying Multi-Angle Polarimeter, such as PACE, CO2M, and Metop-SG. The MAPs intended for these satellite missions have resolutions a lot coarser than Sentinel-2, due to different mission objectives. Yet, they observe each pixel from multiple angles and for each of them measures the reflectance and polarization, providing information on the angular and spectral characteristics.

Acknowledgments

This work is supported by the Helmholtz AI platform, the HAICORE@KIT and the Helmholtz Data Federation (HDF) grant.

References

[1] J. R. Norris, R. J. Allen, A. T. Evan, M. D. Zelinka, C. W. O'Dell, and S. A. Klein, "Evidence for climate change in the satellite cloud record," *Nature*, vol. 536, no. 7614, pp. 72–75, Aug. 2016, ISSN: 1476-4687. DOI: 10.1038/nature18273. [Online]. Available: https://doi.org/10.1038/nature18273.

[2] V. Masson-Delmotte, P. Zhai, S. Pirani, *et al.*, "IPCC, 2021: Summary for policymakers. in: Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change," 2021. DOI: 10.1017/9781009157896.

[3] M. Caretta, A. Mukherji, M. Arfanuzzaman, *et al.*, "Water," in *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, H. O. Pörtner, D. C. Roberts, M. Tignor, *et al.*, Eds. Cambridge, UK and New York, USA: Cambridge University Press, 2022, pp. 551–712, ISBN: 9781009325844. DOI: doi:10.1017/9781009325844.006.

[4] R. Rawshan Ara Begum, R. Lempert, T. B. E. Ali, *et al.*, "Point of departure and key concepts," in *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, H. O. Pörtner, D. C. Roberts, M. Tignor, *et al.*, Eds. Cambridge, UK and New York, USA: Cambridge University Press, 2022, pp. 121–196, ISBN: 9781009325844. DOI: 10.1017/9781009325844.003.

[5] P. Arias, N. Bellouin, E. Coppola, *et al.*, "Technical summary," in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, V. Masson-Delmotte, P. Zhai, A. Pirani, *et al.*, Eds. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2021, pp. 33–144. DOI: 10.1017/9781009157896.002.

[6] B. O'Neill, M. van Aalst, Z. Zaiton Ibrahim, *et al.*, "Key risks across sectors and regions," in *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, H. O. Pörtner, D. C. Roberts, M. Tignor, *et al.*, Eds. Cambridge, UK and New York, USA: Cambridge University Press, 2022, pp. 2411–2538, ISBN: 9781009325844. DOI: 10.1017/9781009325844.025.2412.

[7] Z. Zhu, M. A. Wulder, D. P. Roy, *et al.*, "Benefits of the free and open landsat data policy," *Remote Sensing of Environment*, vol. 224, pp. 382–385, Apr. 2019, ISSN: 0034-4257. DOI: 10.1016/j.rse.2019.02.016.

[8] O. Dubovik, G. L. Schuster, F. Xu, *et al.*, "Grand challenges in satellite remote sensing," *Frontiers in Remote Sensing*, vol. 2, 2021, ISSN: 2673-6187. DOI: 10.3389/frsen.2021.619818.

[9] A. M. Wilson and W. Jetz, "Remotely sensed high-resolution global cloud dynamics for predicting ecosystem and biodiversity distributions," *PLoS biology*, vol. 14, no. 3, e1002415, 2016. DOI: 10.1371/journal.pbio.1002415.

[10] J. Jakubik, M. Vössing, N. Kühl, J. Walk, and G. Satzger, "Data-centric artificial intelligence," *arXiv preprint arXiv:2212.11854*, 2022.

[11] B. V. Hollingsworth, L. Chen, S. E. Reichenbach, and R. R. Irish, "Automated cloud cover assessment for Landsat TM images," in *Imaging Spectrometry II*, M. R. Descour and J. M. Mooney, Eds., International Society for Optics and Photonics, vol. 2819, SPIE, 1996, pp. 170–179. DOI: 10.1117/12.258064.

[12] Z. Zhu and C. E. Woodcock, "Object-based cloud and cloud shadow detection in landsat imagery," *REMOTE SENSING OF ENVIRONMENT*, vol. 118, pp. 83–94, 2012, ISSN: 0034-4257. DOI: 10.1016/j.rse.2011.10.028.

[13] Z. Zhu, S. Wang, and C. E. Woodcock, "Improvement and expansion of the fmask algorithm: Cloud, cloud shadow, and snow detection for landsats 4-7, 8, and sentinel 2 images," *REMOTE SENSING OF ENVIRONMENT*, vol. 159, pp. 269–277, 2015, ISSN: 0034-4257. DOI: 10.1016/j.rse.2014.12.014.

[14] S. Qiu, Z. Zhu, and B. He, "Fmask 4.0: Improved cloud and cloud shadow detection in landsats 4-8 and sentinel-2 imagery," *REMOTE SENSING OF ENVIRONMENT*, vol. 231, 2019, ISSN: 0034-4257. DOI: 10.1016/j.rse.2019.05.024.

[15] R. Richter and D. Schläpfer, "Atmospheric/topographic correction for satellite imagery," *DLR report DLR-IB*, vol. 565, 2005.

[16] M. Main-Knorn, B. Pflug, J. M. B. Louis, V. Debaecker, U. Müller-Wilm, and F. Gascon, "Sen2cor for sentinel-2," in *Image and Signal Processing for Remote Sensing XXIII*, SPIE, vol. 10427, 2017, pp. 37–48. DOI: 10.1117/12.2278218.

[17] V. Zekoll, M. Main-Knorn, K. Alonso, *et al.*, "Comparison of masking algorithms for sentinel-2 imagery," *Remote Sensing*, vol. 13, no. 1, p. 137, 2021. DOI: 10.3390/rs13010137.

[18] P. J. Werdell, M. J. Behrenfeld, P. S. Bontempi, *et al.*, "The plankton, aerosol, cloud, ocean ecosystem mission: Status, science, advances," *Bulletin of the American Meteorological Society*, vol. 100, no. 9, pp. 1775–1794, 2019. DOI: 10.1175/BAMS-D-18-0056.1.

[19] S. Lu, J. Landgraf, G. Fu, *et al.*, "Simultaneous retrieval of trace gases, aerosols, and cirrus using remotap—the global orbit ensemble study for the co2m mission," *Frontiers in Remote Sensing*, vol. 3, p. 914 378, 2022. DOI: 10.3389/frsen.2022.914378.

[20] R. Lang, G. Poli, B. Fougnie, *et al.*, "The 3mi level-1c geoprojected product–definition and processing description," *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 225, pp. 91–109, 2019. DOI: 10.1016/j.jqsrt.2018.12.022.

[21] O. Hagolle, M. Huc, D. V. Pascual, and G. Dedieu, "A multi-temporal and multi-spectral method to estimate aerosol optical thickness over land, for the atmospheric correction of formosat-2, landsat, venμs and sentinel-2 images," *Remote Sensing*, vol. 7, no. 3, pp. 2668–2691, 2015. DOI: 10.3390/rs70302668.

[22] V. Lonjou, C. Desjardins, O. Hagolle, *et al.*, "Maccs-atcor joint algorithm (maja)," in *Remote Sensing of Clouds and the Atmosphere XXI*, SPIE, vol. 10001, 2016, pp. 25–37. DOI: 10.1117/12.2240935.

[23] O. Hagolle, M. Huc, D. V. Pascual, and G. Dedieu, "A multi-temporal method for cloud detection, applied to formosat-2, venμs, landsat and sentinel-2 images," *Remote Sensing of Environment*, vol. 114, no. 8, pp. 1747–1755, 2010. DOI: 10.1016/j.rse.2010.03.002.

[24] Iowa State University, Department of Agronomy, *Iowa environmental mesonet – asos-awos-metar data*, online, 2023. [Online]. Available: https://mesonet.agron.iastate.edu/request/download.phtml.

[25] E. Rumi, D. Kerr, A. Sandford, J. Coupland, and M. Brettle, "Field trial of an automated ground-based infrared cloud classification system," *Meteorological Applications*, vol. 22, no. 4, pp. 779–788, 2015. DOI: 10.1002/met.1523.

[26] P. Nastos, A. Bleta, and I. Matsangouras, "Human thermal perception related to föhn winds due to saharan dust outbreaks in crete island, greece," *Theoretical and Applied Climatology*, vol. 128, pp. 635–647, 2017. DOI: 10.1007/s00704-015-1724-3.

[27] V. Gornyy, S. Kritsuk, I. S. Latypov, A. Manvelova, and A. Tronin, "Satellite mapping of urban air overheating risk (case study of helsinki, finland)," *Cosmic Research*, vol. 60, no. Suppl 1, S38–S45, 2022. DOI: 10.1134/S0010952522700058.

[28] J. Cermak and J. Bendix, "A novel approach to fog/low stratus detection using meteosat 8 data," *Atmospheric Research*, vol. 87, no. 3-4, pp. 279–292, 2008. DOI: 10.1016/j.atmosres.2007.11.009.

[29] B. Jahani, S. Karalus, J. Fuchs, T. Zech, M. Zara, and J. Cermak, "Algorithm for continual monitoring of fog life cycles based on geostationary satellite imagery as a basis for solar energy forecasting," *EGUsphere*, vol. 2024, pp. 1–22, 2024. DOI: 10.5194/egusphere-2023-2885.

[30] E. S. A. (ESA), *Copernicus open access hub*, online, 2023. [Online]. Available: https://scihub.copernicus.eu/.

[31] M. C. Peel, B. L. Finlayson, and T. A. McMahon, "Updated world map of the köppen-geiger climate classification," *Hydrology and Earth System Sciences*, vol. 11, no. 5, pp. 1633–1644, 2007. DOI: 10.5194/hess-11-1633-2007. [Online]. Available: https://hess.copernicus.org/articles/11/1633/2007/.

[32] H. E. Beck, N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, and E. F. Wood, "Present and future köppen-geiger climate classification maps at 1-km resolution," *Scientific Data*, vol. 5, no. 1, p. 180 214, Oct. 2018, ISSN: 2052-4463. DOI: 10.1038/sdata.2018.214. [Online]. Available: https://doi.org/10.1038/sdata.2018.214.

[33] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 97, PMLR, Jun. 2019, pp. 6105–6114.

[34] M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training," in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 139, PMLR, Jul. 2021, pp. 10 096–10 106. [Online]. Available: https://proceedings.mlr.press/v139/tan21a.html.

[35] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, IEEE, 2009, pp. 248–255. DOI: 10.1109/CVPR.2009.5206848.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034. DOI: 10.1109/ICCV.2015.1.

[37] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2017. arXiv: 1412.6980 [cs.LG]. [Online]. Available: https://arxiv.org/abs/1412.6980.

[38] J. Cheng, Z. Wang, and G. Pollastri, "A neural network approach to ordinal regression," in *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, IEEE, 2008, pp. 1279–1284. DOI: 10.1109/IJCNN.2008.4633963.

[39] S. Kaufman, S. Rosset, C. Perlich, and O. Stitelman, "Leakage in data mining: Formulation, detection, and avoidance," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 6, no. 4, pp. 1–21, 2012. DOI: 10.1145/2382577.2382579.

[40] J. Lange, G. Cavallaro, M. Götz, E. Erlingsson, and M. Riedel, "The influence of sampling methods on pixel-wise hyperspectral image classification with 3d convolutional neural networks," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 2018, pp. 2087–2090. DOI: 10.1109/IGARSS.2018.8518671.

[41] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *kdd*, vol. 96, 1996, pp. 226–231.

[42] M. Götz, C. Bodenstein, and M. Riedel, "Hpdbscan: Highly parallel dbscan," in *Proceedings of the workshop on machine learning in high-performance computing environments*, 2015, pp. 1–10. DOI: 10.1145/2834892.2834894.

[43] H. Choi and R. Bindschadler, "Cloud detection in landsat imagery of ice sheets using shadow matching technique and automatic normalized difference snow index threshold value decision," *Remote Sensing of Environment*, vol. 91, no. 2, pp. 237–242, 2004, ISSN: 0034-4257. DOI: 10.1016/j.rse.2004.03.007.

[44] J. H. Jeppesen, R. H. Jacobsen, F. Inceoglu, and T. S. Toftegaard, "A cloud detection algorithm for satellite imagery based on deep learning," *Remote Sensing of Environment*, vol. 229, pp. 247–259, 2019, ISSN: 0034-4257. DOI: 10.1016/j.rse.2019.03.039.

[45] D. K. Hall, G. A. Riggs, and V. V. Salomonson, "Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data," *Remote sensing of Environment*, vol. 54, no. 2, pp. 127–140, 1995. DOI: 10.1016/0034-4257(95)00137-P.

[46] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009, ISBN: 1441412697.

[47] A. Paszke, S. Gross, F. Massa, *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, vol. 32, Curran Associates, Inc., 2019.

[48] L. Gaudette and N. Japkowicz, "Evaluation methods for ordinal classification," in *Advances in Artificial Intelligence*, Springer, 2009, pp. 207–210. DOI: 10.1007/978-3-642-01818-3_25.

[49] J. Chen, H. Dong, X. Wang, F. Feng, M. Wang, and X. He, "Bias and debias in recommender system: A survey and future directions," *ACM Transactions on Information Systems*, vol. 41, no. 3, pp. 1–39, 2023. DOI: 10.1145/3564284.

[50] E. S. A. (ESA), *Esa iceye constellation*, online, 2023. [Online]. Available: https://www.eoportal.org/satellite-missions/iceye-constellation#performance-specifications.

Markus Götz received his B.Sc. and M.Sc. degrees in IT-System Engineering from the University of Potsdam, Germany, in 2010 and 2014 respectively, with intermediate stays the Blekinge Tekniska Högskola, Sweden and CERN, Switzerland. Since 2017 he holds a Ph.D. degree in Computational Engineering from the University of Iceland in conjuction with the Juelich Supercomputing Centre, Germany. Currently, he is a postdoctoral researcher at the Scientific Computing Center, Karlsruhe Institute of Technology, Germany as the project manager for the Helmholtz Analytics Framework and the head of the HelmholtzAI consultants team. In line with his work, he focuses on applied artificial intelligence and data analysis on high-performance cluster systems to work on the grand challenges in the natural sciences. Markus Götz's research interests include machine learning, global optimization as well as parallel algorithm engineering. He is a member of the IEEE.

Erik Wessel studied Computer Science at Karlsruhe Institute of Technology (KIT), where he earned both his B.Sc. and M.Sc. degrees. His master's thesis forms the basis of this paper. Currently, he is an IT-Consultant at Exxeta. His research interests include artificial intelligence, data analytics, software engineering and architecture, computer graphics and visualization, as well as natural sciences like meteorology and climatology.

Uğur Çayoğlu is a computer scientist specializing in data structures, algorithms, and data compression, with a particular focus on applications for climate data. He earned his B.Sc. and M.Sc. in Computer Science from Karlsruhe Institute of Technology (KIT), Germany, where he also completed his Ph.D., which explored advanced methods for data compression tailored to large-scale climate datasets. During his occupation at KIT, Uğur Çayoğlu served as the team lead of the Simulation and Data Life Cycle Lab (SDL) Earth System Sciences, where he led research initiatives on efficient data handling and processing for climate modeling and simulations. After transitioning to the private sector, Uğur Çayoğlu took on the role of project lead in the automotive industry, where he applies his expertise in data compression and algorithm design to tackle complex challenges in data-intensive applications. His current work focuses on optimizing data pipelines, enhancing system efficiency, and integrating advanced computational methods into automotive technologies.

Babak Jahani is an atmospheric scientist specializing in the remote sensing of atmospheric composition. He is currently affiliated with the Earth observation group of SRON Netherlands Institute for Space Research, Leiden, the Netherlands. He earned a B.Sc. in Agriculture Engineering-Water Science Engineering from Shahid Chamran University of Ahvaz, Iran (2015), and an M.Sc. in Agriculture Engineering-Water Resources Engineering from the University of Tabriz, Iran (2017). His PhD, completed at the Universitat de Girona, Spain (2021), focused on the cloudy-to-cloud-free transition zone. From 2021 to 2023, he was a postdoctoral researcher at Karlsruhe Institute of Technology (KIT). Babak Jahani's research interests include clouds, aerosols, weather forecast, and Remote Sensing.

Charlotte Debus studied Physics at the University of Heidelberg, Germany. After her M.Sc. degree in 2012, she did here PhD in the area of Medical Physics at the German Cancer Research Center (DKFZ) form 2013 – 2016. After a 2-year Post-Doc period at the DKFZ, she was a research associate at the German Aerospace Institute, from 2019 – 2020, where she worked on machine learning methods and high performance computing. Since 2022 Charlotte is leading a junior research group on robust and efficient AI.

Jan Cermak is a professor of geophysical remote sensing at Karlsruhe Institute of Technology (KIT). He works at the interface of remote sensing and climate research with a particular interest in clouds. He studied geography and in 2002 received an M.A. degree in Environment and Development from the University of London, UK (King's College London and School of Oriental and African Studies). His PhD, conferred in 2006 by the University of Marburg, Germany, focused on satellite remote sensing of fog. After a postdoc period at ETH Zurich, Switzerland, he became a professor of climatology at Ruhr-Universität Bochum, Germany, in 2011. Since 2016, he is at KIT, where he currently leads the Institute of Meteorology and Climate Research – Atmospheric Trace Gases and Remote Sensing.

Achim Streit is one of the directors of the Scientific Computing Center (SCC) at the Karlsruhe Institute of Technology (KIT). He is also a Professor for Distributed and Parallel High-performance Systems at KIT's department of Informatics. In 1999, he received a Diploma in Computer Science from the University of Dortmund, Germany and in 2003 a Ph.D. degree in the same subject from the University of Paderborn, Germany. Afterwards Achim Streit led the Federated Systems and Data Division at the Juelich Supercomputing Centre, Germany. He initiated and chaired several national and international research initiatives within the Helmholtz association (e.g., Helmholtz Data Federation and Helmholtz Information & Data Science Academy (HIDA)) on the national level (e.g., NFDI4Ing and NFDI-MatWerk) and the European level (e.g., EUDAT and EOSC). His research interests include high-performance and data-intensive computing, Big Data and federated data management, data analytics as well as job scheduling and resource management for parallel and distributed systems.