Proceedings of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 485–492 DOI: 10.15439/2024F5446 ISSN 2300-5963 ACSIS, Vol. 39

AI-Based Spatiotemporal Crop Monitoring by Cloud Removal in Satellite Images

Jiří Pihrt*, Petr Šimánek*, Alexander Kovalenko*, Jiří Kvapil†, and Karel Charvát‡
*Faculty of Information Technology, Czech Technical University in Prague
Thákurova 9, Praha 6, Czech Republic
Email: petr.simanek@fit.cvut.cz

†Lesprojekt s.r.o.
Martinov 197, Záryby, Czech Republic
Email: jiri.kvapil@lesprojekt.cz

‡HELP SERVICE – REMOTE SENSING s.r.o.
Husova 2117, Benešov, Czech Republic
Email: charvat@hsrs.cz

***Abstract*—Efficient crop monitoring and crop dynamics forecasting leveraging diverse satellite and point data are described. UnCRtainTS neural network architecture is utilized for cloud removal in satellite imagery which overcomes an issue in crop monitoring. Combining optical (Sentinel-2) and radar (Sentinel-1) satellite data improves the robustness and accuracy of the model in terms of satellite image reconstruction and vegetation index estimation. However, available soil-type geographical data and land surface analysis products, do not improve prediction accuracy significantly.**

I. Introduction

PRECISION agriculture [16] aims to maximize the output from farming by defining the precise and sufficient amounts of inputs like water, fertilizer, pesticides, etc. at the correct time to the crop for increasing its productivity and maximizing its yields. However, this approach increases the vulnerability of crop health, in the case of input errors [3] as the goal of precision agriculture is to use minimally required amount of inputs. Therefore, efficient and timely crop monitoring is essential to detect harmful patterns that may emerge under certain conditions, potentially compromising the harvest. Factors influencing crop development in the early stages are numerous and complex[9], often requiring expert knowledge to interpret and understand their interactions. These intricacies can lead to important relationships being overlooked, emphasizing the need for advanced monitoring techniques.

In recent years, machine learning has achieved significant success in remote sensing imagery [28],[2], particularly in crop detection using multispectral satellite data [29]. This technology allows for high-accuracy crop identification, providing valuable insights into agricultural practices and crop health. However, despite these advancements, cloud cover [29] remains as a significant obstacle to efficient crop monitoring. This issue is particularly relevant in regions like Central Europe, where average yearly cloud coverage can exceed 50%[1]. In such conditions, the efficiency of monitoring crops using satellite images is severely compromised, making it nearly impossible to rely solely on crop detection by remote sensing imagery.

Another critical challenge in agriculture is forecasting harvest dynamics [13]. External factors such as temperature, wind, and precipitation affect crop growth in various ways. These environmental variables can negatively impact crop development, which, if not mitigated timely, can lead to significant yield losses. To address these issues, treatments such as adjusting soil fertilization or irrigation can be applied to mitigate the adverse effects [1]. Understanding how crops grow and respond over time to varying environmental conditions allows for the timely implementation of interventions, thereby reducing the risk of damage and optimizing crop health [8].

Moreover, precise detection of crop dynamics in response to external interventions, such as fertilizing, irrigation, or the application of pesticides and fungicides, can lead to more efficient use of these treatments [26]. This precision agriculture approach can result in significant resource savings, such as conserving water and using fertilizers more reasonably. This helps in limiting the use of pesticides and fungicides to the minimum required for effective pest and disease management, promoting more sustainable agricultural practices.

To overcome the limitations posed by cloud cover and to enhance the forecasting of harvest dynamics, it is essential to integrate machine learning techniques with diverse data sources. This integration can develop robust models capable of handling incomplete or obscured data and improving the resilience and effectiveness of agricultural monitoring systems. This paper aims to investigate these methods, offering solutions to improve crop monitoring and forecasting in the face of frequent cloud cover and other environmental challenges. To

[1]https://www.dwd.de

predict and forecast crop health we monitored the normalized difference vegetation index (NDVI) which is a standard metric for quantifying the health and density of vegetation using spectral satellite data from red and near-infrared bands.

As a solution to the abovementioned problems, we explored the possibility of forecasting crop dynamics in high cloud coverage conditions. We leveraged a combination of Satellite (Sentinel 2 [15]) and radar data (Sentinel 1 [24]) and point data, such as soil type and land surface analysis (LSA) to develop a model for the next frame prediction to provide cloudless prediction that learns crop development pattern dynamics. By leveraging advanced technologies, more sustainable and efficient agricultural practices can be achieved, to secure a more stable food supply.

II. Related Research

Remote sensing using multispectral satellite data has been widely explored for its applications in agriculture. Multispectral imagery allows for the identification of various crop types and their health status based on their spectral signatures. Studies such as those by Peña-Barragán et al. [14] and Quan et al. [17] have demonstrated the effectiveness of using multispectral data for crop classification and harvest health monitoring. Additionally, dash et al. [4] explore the effectiveness of unmanned aerial vehicles (UAV) and satellite imagery for monitoring forest health, specifically focusing on mature *P. radiata* trees. This research under controlled experimental conditions shows that both UAV and satellite sensors can detect plant stress, as evidenced by deviations in spectral indices and strong correlations with field observations.

However, as it is mentioned above the issue of cloud cover remains a significant barrier. Techniques to mitigate this include the use of cloud masking algorithms and temporal interpolation methods. For example, one of the early works by Zhu et al. [30] presents the Fmask algorithm, which effectively identifies and masks clouds in Landsat imagery, enabling clearer analysis of vegetation.

To overcome the limitations of optical data due to cloud cover, researchers have increasingly turned to radar imagery, which can penetrate clouds and provide consistent data. Sentinel-1 synthetic aperture radar (SAR) data, for instance, has been successfully integrated with optical data to improve crop monitoring. The study by Veloso et al. [25] reports the use of Sentinel-1 and Sentinel-2 data fusion for crop-type mapping, demonstrating improved accuracy and reliability.

Later advances of cloud removing algorithms included using generative adversarial networks for filmy cloud removal on satellite imagery with multispectral conditional [7], using a deep residual neural network and SAR-optical data fusion [12], spatiotemporal generative networks [22], enhanced cloud removal with global-local fusion leveraging data from synthetic aperture radar [27], as well as using uncertainty quantification for cloud removal in optical satellite time series [6], which is the current state-of-the-art method to the best of our knowledge.

Regarding crop classification and crop dynamics monitoring, machine learning techniques, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown great promise in enhancing the analysis of remote sensing data. The work by Russwurm and Körner [19] employed RNNs to analyze time-series data from Sentinel-2. Forecasting crop dynamics involves predicting how crops will develop over time, considering various environmental factors. Models such as those developed by Han et al. [8] leverage weather data and soil conditions to predict crop yields, providing valuable insights for proactive agricultural management.

Moreover, hybrid models combining machine learning with mechanistic crop models offer enhanced predictive capabilities. For example, Shahhosseini et al. [23] integrated a machine learning approach with the Agricultural Production Systems sIMulator (APSIM) crop model, resulting in improved predictions of crop growth and yield under varying environmental conditions. The precise detection and analysis of crop responses to interventions such as fertilization and irrigation can lead to more efficient resource use. Studies such as Lobell et al. [11] have shown that remote sensing can effectively monitor crop responses to nitrogen application, optimizing fertilizer use and enhancing sustainability. Recent advancements in IoT and sensor networks further support this precision agriculture approach, as demonstrated by Liakos et al. (2018) [10], who reviewed the integration of IoT in agriculture for real-time monitoring and decision-making.

III. Data

Integrating data from multiple sources in machine learning provides numerous advantages, particularly in fields that require comprehensive analysis and forecasting. Combining diverse data sets introduces greater variability and richness, which helps machine learning models capture a wider range of patterns and relationships. This approach leads to more robust and generalized models capable of making accurate predictions in varied scenarios. Additionally, using multiple data sources helps mitigate biases present in individual datasets, resulting in a more balanced representation of the underlying phenomena and reducing both bias and variance in the model's predictions.

In the context of crop dynamics forecasting, the integration of data from satellites, radars, and point sources such as temperature, wind, and precipitation provides significant benefits. Satellite imagery offers valuable insights into crop health and development through multispectral data, but its effectiveness is often limited by cloud cover. Radar data, which can penetrate clouds, ensures continuous monitoring regardless of weather conditions. While radar data provides structural information about crops, it may lack the spectral details available in satellite images. By combining these sources, a more complete and reliable view of crop conditions can be achieved.

It is worth noting that there are publicly available datasets designed for the cloud removal task using machine learning, the most widely utilized currently being SEN12MS-CR-

TS [5]. This dataset contains pairs of spatially aligned Sentinel 1 and 2 images. However, the Sentinel 2 images are of the L1C product. For surface analysis, the atmospherically corrected L2A product is more suitable. Another limitation is the temporal resolution of this dataset, consisting of only 30 samples per year, which is lower than the actual acquisition rate. High intervals between frames in the input sequence may cause missing valuable information in the skipped samples, such as recent cloudless images. Furthermore, the overall dataset contains diverse regions globally. However, this work was specifically aimed at agricultural lands in the Czech Republic. For these reasons, we opted to create our own dataset for this work.

The primary inputs include the most recent Sentinel-2 L2A frames, consisting of all 12 spectral bands. Additionally, we incorporate the most recent Sentinel-1 GRD (IW mode) frames, specifically the VV and VH polarization bands, which are orthorectified and terrain-corrected to ensure spatial alignment with other sources. These datasets are essential for capturing both optical and radar imagery, providing a comprehensive view of the agricultural landscape.

Furthermore, we have experimented with integrating various EUMETSAT LSA products. Among these, we used MDIDSSF, MDMETv3, and MNSLF, which are available at much lower spatial resolution than Sentinel images. For these products, we utilized single values nearest to the region of interest to supplement the primary data. The temporal resolution for these inputs is one day, ensuring that we capture daily variations in the land surface conditions.

Additionally, we considered geographical data from the Czech Republic's VUMOP mapping service (https://mapy.vumop.cz/), specifically the soil types layer. This layer, available in the EPSG:5514 coordinate system for the Czech Republic, required reprojection to align with our other datasets. Being categorical data, it was necessary to convert it into a one-hot encoded format to be usable in our models. It is important to note that this data is static, and lacking temporal variation, but it provides valuable contextual information about soil types.

For training purposes, we utilized the Scene Classification Layer (SCL) from all Sentinel-2 frames. This layer is useful for masking out clouds and cloud shadows, ensuring target ground truth data is cloudless. Additionally, we experimented with a binary mask to exclude non-agricultural land from the analysis, derived from a layer available on the VUMOP mapping service. This mask helps in focusing the model's attention solely on agricultural areas, potentially improving the relevance and accuracy of the inferences.

Using the data sources described previously, we create a specialized dataset for this work. Regions of interest are sampled uniformly in the bounding box of the Czech Republic. After removing regions with little to no agricultural land (mapped by VUMOP) or missing data, the number of regions is 212. We use data from the year 2022 for training and 2023 for validation. Only data from April to October are used, as winter months are less useful for crop monitoring.

All spatial data inputs were standardized to a 10-meter resolution to maintain consistency across datasets. For datasets with lower resolutions, we upscaled them to meet this standard. The dataset contains and therefore our models were trained using input frames of 256x256 pixels, allowing for efficient processing while maintaining high spatial detail.

IV. Experiment Design

For a cloud removal task using machine learning, training samples consist of a cloudless Sentinel-2 target frame, and data that the model uses as input, most importantly a sequence of the most recent Sentinel-2 frames. To generate such samples, we generate sequences of Sentinel-2 frames using a sliding window and pick the last frame as the target. The sample is dropped if the target frame contains clouds or cloud shadows, as classified in SCL. Additional input data, such as a sequence of most recent Sentinel-1 frames, are added depending on the experiment. In our experiments, we standardized the number of recent frames to 5, though this parameter is adjustable depending on the specific requirements of future studies.

The neural network architecture we opted to use is UnCRtainTS [6], as it currently is the state-of-the-art method for the cloud removal task with a publicly available implementation. UnCRtainTS is an attention-based convolutional network, as described in Fig. 1. However, in our experiments, we used the mono-temporal version, which excludes the temporal aggregator. The multi-temporal version requires a significant amount of memory during training. Therefore, due to hardware limitations, we used the mono-temporal version. The advantage of this version is that it allows for more straightforward aggregation of additional data which is either static or not temporally aligned with the Sentinel-2 frames. It is worth noting that Sentinel-1 frames are also not temporally aligned with Sentinel-2.

To confirm the effectiveness of UnCRtainTS on our dataset, we also compare it to a baseline U-Net [18], a widely used image-to-image convolutional network.

For both UnCRtain and U-Net, input sequences are flattened and all inputs are concatenated in the channel dimension. In the case of point data, each value first has to be upscaled into a 256x256 image, so that it can be concatenated with the other images.

Although UnCRtainTS can be used with a negative log-likelihood (NLL) loss function for uncertainty estimation as in the original paper, we used classic regression loss functions (MSE, MAE), to focus on more accurate predictions instead of uncertainty quantification. All models in the experiments are trained using the Adam optimizer with a learning rate of 0.001.

V. Results

The experiments test various input sources and hyperparameters. Unless stated otherwise, by default the model is the following: UnCRtainTS architecture, MAE loss function, batch size = 6, MAE loss function, and inputs are Sentinel-1 and Sentinel-2 sequences.

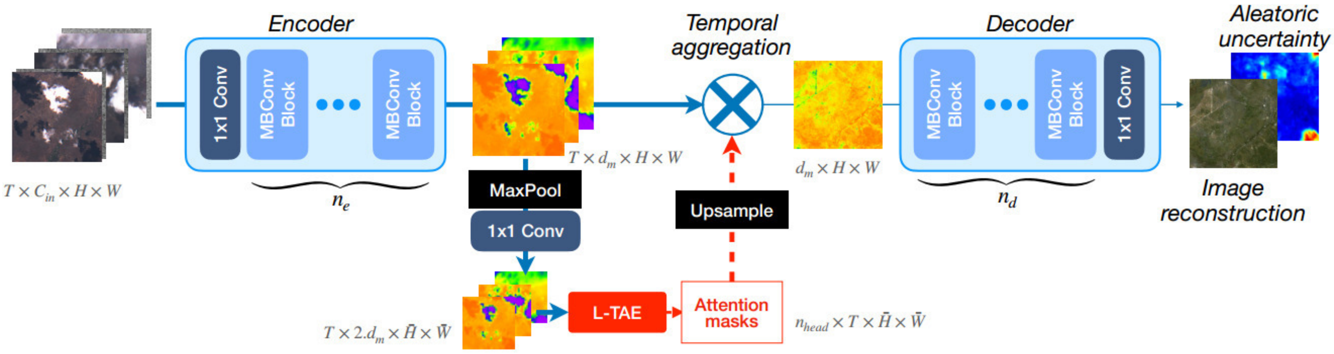

Fig. 1. UnCRtainTS [6] architecture. The network has three main parts; an encoder and decoder consisting of MBConv [21] blocks processing feature maps at full input resolution, and an attention-based temporal aggregator computing an attention mask by applying an L-TAE [20] to downsampled feature maps.

The MAE and MSE metrics, reported in each experiment, are computed on the validation set, exclusively on agricultural land pixels, averaged across all 12 Sentinel-2 L2A bands.

Table I compares UnCRtainTS against several baselines. *Most recent cloudless* simply repeats the most recent fully cloudless Sentinel-2 frame. *Mosaicking* uses the SCL cloud mask to repeat the most recent unobstructed frame for each pixel independently. This uses more recent data but suffers from cloud and cloud shadow artifacts due to imperfect cloud detection, which is mostly not an issue in fully cloudless frames.

The importance of Sentinel-1 and Sentinel-2 input sequences is demonstrated in Table II. It is clear that using both data sources is crucial for good results. Table III experiments with several batch sizes.

Table IV shows that training with MAE loss function outperforms MSE, even on the MSE metric. Additionally, training on all pixels, rather than just those mapped as agricultural land, slightly improves performance. This suggests that incorporating more diverse data makes the model more robust in the target domain (agricultural land) as well. It could also be caused by false positives in the VUMOP mask, and the model trained on all data handling these outliers better.

Experiments including the additional inputs (soil types and LSA point data) are shown in Table V. Interestingly, the inclusion of these additional data sources did not lead to any significant improvement in the model's performance. This suggests that the current model already captures the essential features required for accurate predictions, or that the added data may require further preprocessing or different integration methods to be beneficial.

In the context of crop monitoring, there are metrics useful for quantifying crop health, such as NDVI (normalized difference vegetation index). NDVI is computed from multi-spectral satellite imagery as

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{Red}}{\mathrm{NIR} + \mathrm{Red}}.$$

For Sentinel-2, NIR is band 8 and Red is band 4. Table VI shows that optimizing NDVI directly (optimizing $loss(\mathrm{NDVI}(x), \mathrm{NDVI}(y))$ instead of $loss(x, y)$) leads to significantly more accurate NDVI predictions. The disadvantage is that the model no longer predicts the raw Sentinel-2 bands. For applications where that is needed as well, separate models for each task can be trained.

Example predictions using the default model are visualized in Fig. 2. Example NDVI predictions over time using the model trained with NDVI MAE loss function are shown in Fig. 3.

TABLE I
EVALUATION OF DIFFERENT METHODS

Method	MAE	MSE
most recent cloudless	0.0371	0.00338
mosaicking	0.0386	0.00388
U-Net	0.0278	0.00174
UnCRtainTS	**0.0255**	**0.00153**

TABLE II
EVALUATING CONTRIBUTIONS OF SENTINEL INPUT SEQUENCES

Model	Sentinel-1	Sentinel-2	MAE	MSE
UnCRtainTS	✓		0.0375	0.00300
UnCRtainTS		✓	0.0281	0.00191
UnCRtainTS	✓	✓	**0.0255**	**0.00153**

VI. LIMITATIONS

Despite the promising results and potential applications of our crop monitoring system, several limitations must be acknowledged. Cloud coverage and data availability remain significant challenges. Although combining Sentinel-1 and

TABLE III
EVALUATION OF THE MODEL WITH VARYING BATCH SIZE

Model	batch size	MAE	MSE
UnCRtainTS	4	0.0256	0.00159
UnCRtainTS	6	**0.0255**	**0.00153**
UnCRtainTS	8	0.0257	0.00156

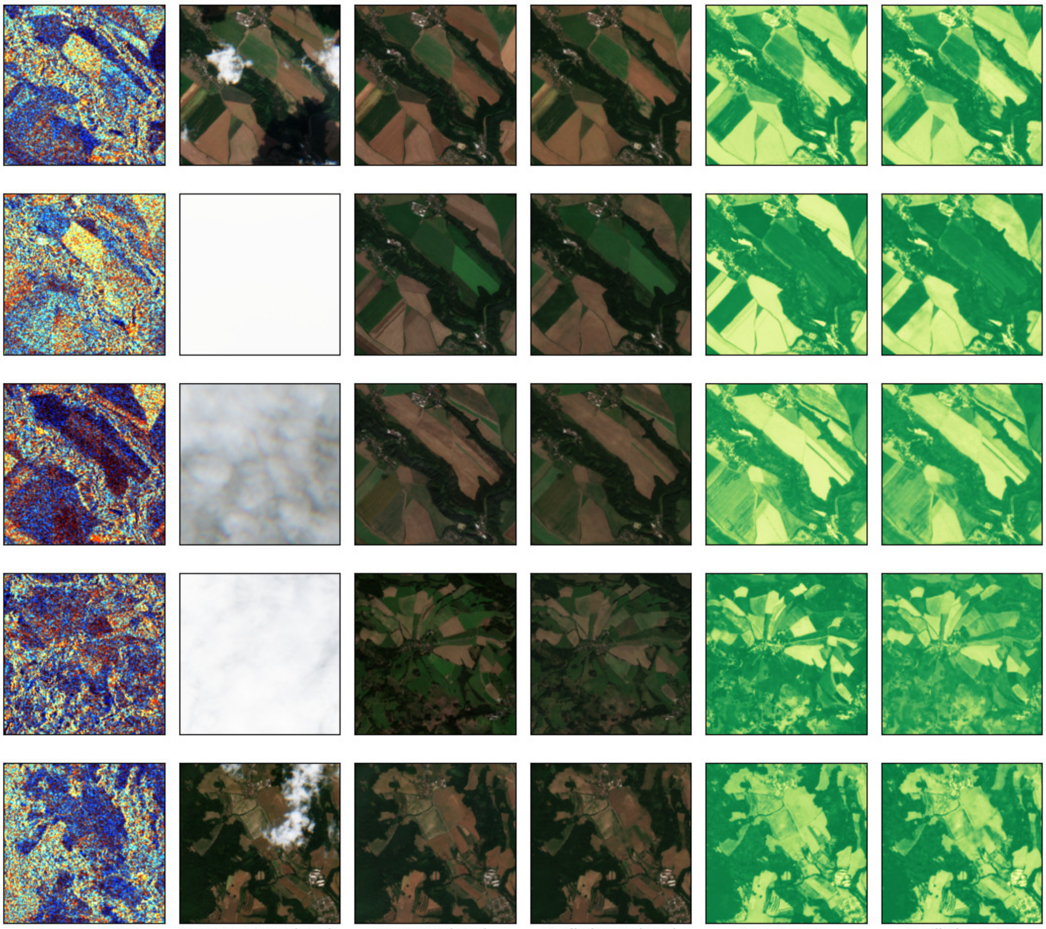

Fig. 2. Prediction examples. Rows: different samples, first three are from the same region of interest. Columns: visualization of the most recent input Sentinel-1 frame; visible channels of the most recent input Sentinel-2 frame; visible channels of the target and prediction; visualization of NDVI (0-1 maps to yellow-green) of the target and prediction.

TABLE IV
EVALUATING VARIOUS LOSS FUNCTIONS (*masked* LOSS IS COMPUTED ONLY ON PIXELS MAPPED AS AGRICULTURAL LAND)

Model	loss function	MAE	MSE
UnCRtainTS	MAE	**0.0255**	**0.00153**
UnCRtainTS	MAE masked	0.0258	0.00158
UnCRtainTS	MSE	0.0267	0.00161
UnCRtainTS	MSE masked	0.0291	0.00184

TABLE V
EVALUATING CONTRIBUTIONS OF AUXILIARY INPUTS (VUMOP SOIL TYPES LAYER AND LSA POINT DATA)

Model	soil types	LSA	MAE	MSE
UnCRtainTS			0.0255	0.00153
UnCRtainTS	✓		**0.0254**	**0.00153**
UnCRtainTS		✓	0.0258	0.00158
UnCRtainTS	✓	✓	0.0260	0.00160

TABLE VI
EVALUATING NDVI PERFORMANCE WITH VARYING LOSS FUNCTIONS

Model	loss	NDVI MAE	NDVI MSE
UnCRtainTS	NDVI MAE	**0.0631**	0.0105
UnCRtainTS	NDVI MSE	0.0647	**0.0098**
UnCRtainTS	MAE	0.0721	0.0122
UnCRtainTS	MAE	0.0801	0.0135

Sentinel-2 data helps mitigate cloud cover issues, there are still instances where data from both sources may be inadequate.

The temporal resolution of available data can also be limited. While we aimed for a daily temporal resolution for some inputs, Sentinel-2 images are often only available every 5 days for specific regions. This gap can lead to temporal inconsistencies and impact the monitoring of rapid changes in crop conditions.

The spatial resolution of some input data required upscaling to match the 10-meter resolution standard. This is particularly

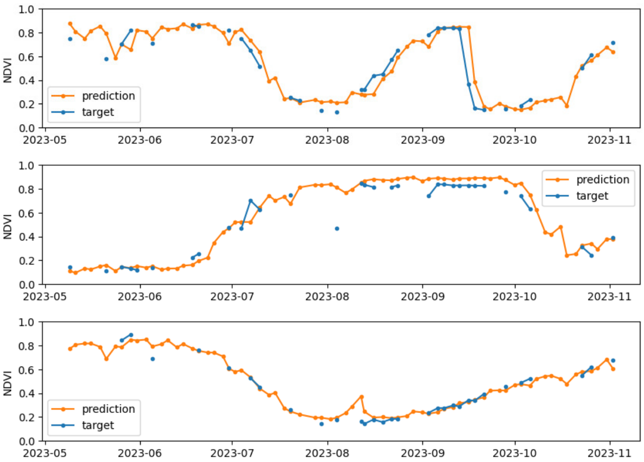

Fig. 3. Examples of NDVI prediction over time. This was acquired by running the model for each sample over the year for a given region and measuring NDVI at three handpicked locations (agricultural land). Target values obscured by clouds are regarded as missing values and not plotted. However, the cloud mask is not perfect and some target values may be affected by clouds.

true for the LSA products that originally had much lower resolutions. Upscaling may introduce artifacts or reduce the precision of the information, which can affect overall model performance.

While we have included optional data inputs such as additional LSA products and specific categorical layers, these have not yet demonstrated significant benefits in our current framework. Future work will continue to explore and validate the potential contributions of these additional datasets to further enhance the accuracy and robustness of the agricultural monitoring models.

The generalizability of our models to other regions, crop types, and environmental conditions is not guaranteed. Differences in agricultural practices, climatic conditions, and soil properties may require additional tuning and validation to ensure the models perform well in diverse settings.

The computational resources required for training and deploying advanced machine learning models, like UnCRtainTS, are substantial. High-performance computing environments or cloud-based solutions are necessary, which may not be accessible to all users, especially in resource-constrained settings. Errors in preprocessing steps, such as cloud masking or data alignment, can propagate through the modeling pipeline, affecting the final predictions' accuracy. Ensuring high-quality preprocessing is critical but can be challenging given the complexity and volume of the data.

Using point data, which includes measurements of environmental factors such as temperature, wind, and precipitation, could be critical for understanding crop dynamics but is not directly observable from remote sensing data. Including these point measurements could provide essential context for interpreting remote sensing data. For example, temperature affects plant metabolism and growth rates, and integrating temperature data helps understand the impact of these factors on crop development. Precipitation is vital for soil moisture and overall plant health, and combining precipitation data with satellite and radar imagery would help assess drought conditions or waterlogging. Wind influences pollination and the spread of pests and diseases, and wind data can presumably help predict potential pest outbreaks or physical damage to crops. However, these types of data were not available for the Czech Republic.

Additionally, ongoing research and collaboration with agricultural experts can help refine the models, improve data integration techniques, and enhance the system's robustness and applicability across different agricultural contexts.

VII. Conclusion

The integration of machine learning techniques with diverse data sources has demonstrated significant potential in enhancing crop monitoring and forecasting, particularly in the context of frequent cloud cover and other environmental challenges. Our study leveraged a combination of the most recent Sentinel-2 L2A frames, Sentinel-1 GRD frames, and various point data sources to develop a robust model for predicting crop dynamics under high cloud coverage conditions. By combining optical and radar imagery, we were able to create a more complete and reliable view of crop conditions, ensuring continuous monitoring and accurate forecasting regardless of weather conditions.

The results of our experiments indicate that the integration of multiple data sources, such as Sentinel-2 L2A frames, and Sentinel-1 GRD frames enhances the accuracy of the models and provides a comprehensive understanding of crop health and development in terms of predicting and forecasting the NDVI index. For instance, the use of radar data from Sentinel-1 complemented the optical data from Sentinel-2 by providing information even under cloud cover.

We also explored the inclusion of various LSA products and categorical data from the VUMOP mapping service, although their contributions to the overall model performance were not as significant. This highlights the importance of selecting relevant and high-quality data sources tailored to specific agricultural monitoring needs.

The next steps for advancing our crop monitoring system involve refining the existing models to improve performance and reduce computational requirements. This includes optimizing model architectures and experimenting with additional data sources, such as temperature, wind, precipitation, and/or soil moisture, as well as sensors and UAV imagery. Extensive field validation will be conducted to assess the accuracy and reliability of the system in real-world conditions, complemented by detailed error analysis to refine the models.

Acknowledgment

This project is funded with state support from the Technology Agency of the Czech Republic and the Ministry of Industry and Trade of the Czech Republic under the TREND Program.

References

[1] Emmanuel Abiodun Abioye et al. "Precision irrigation management using machine learning and digital farming solutions". In: *AgriEngineering* 4.1 (2022), pp. 70–103. DOI: https://doi.org/10.3390/agriengineering4010006.

[2] Paolo Bertellini et al. "Binary Classification of Agricultural Crops Using Sentinel Satellite Data and Machine Learning Techniques". In: *2023 18th Conference on Computer Science and Intelligence Systems (FedCSIS)*. IEEE. 2023, pp. 859–864. DOI: 10.15439/2023F1703.

[3] Nicolas R Dalezios et al. "Remotely sensed methodologies for crop water availability and requirements in precision farming of vulnerable agriculture". In: *Water resources management* 33 (2019), pp. 1499–1519. DOI: https://doi.org/10.1007/s11269-018-2161-8.

[4] Jonathan P Dash, Grant D Pearse, and Michael S Watt. "UAV multispectral imagery can complement satellite data for monitoring forest health". In: *Remote Sensing* 10.8 (2018), p. 1216. DOI: 10.3390/rs10081216.

[5] Patrick Ebel et al. "SEN12MS-CR-TS: A Remote Sensing Data Set for Multi-modal Multi-temporal Cloud Removal". In: *IEEE Transactions on Geoscience and Remote Sensing* (2022). DOI: 10.1109/TGRS.2022.3146246. URL: https://doi.org/10.1109/TGRS.2022.3146246.

[6] Patrick Ebel et al. "UnCRtainTS: Uncertainty quantification for cloud removal in optical satellite time series". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, pp. 2086–2096. DOI: 10.1109/CVPRW59228.2023.00202.

[7] Kenji Enomoto et al. "Filmy cloud removal on satellite imagery with multispectral conditional generative adversarial nets". In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2017, pp. 48–56. DOI: 10.1109/CVPRW.2017.197.

[8] Jichong Han et al. "Prediction of winter wheat yield based on multi-source data and machine learning in China". In: *Remote Sensing* 12.2 (2020), p. 236. DOI: https://doi.org/10.3390/rs12020236.

[9] Jay Ram Lamichhane et al. "Abiotic and biotic factors affecting crop seed germination and seedling emergence: a conceptual framework". In: *Plant and soil* 432 (2018), pp. 1–28. DOI: https://doi.org/10.1007/s11104-018-3780-9.

[10] Konstantinos G Liakos et al. "Machine learning in agriculture: A review". In: *Sensors* 18.8 (2018), p. 2674. DOI: 0.3390/s18082674.

[11] David B Lobell, Kenneth G Cassman, and Christopher B Field. "Crop yield gaps: their importance, magnitudes, and causes". In: *Annual review of environment and resources* 34 (2009), pp. 179–204. DOI: https://doi.org/10.1146/annurev.environ.041008.093740.

[12] Andrea Meraner et al. "Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 166 (2020), pp. 333–346. DOI: https://doi.org/10.1016/j.isprsjprs.2020.05.013.

[13] Michele Meroni et al. "Yield forecasting with machine learning and small data: What gains for grains?" In: *Agricultural and Forest Meteorology* 308 (2021), p. 108555. DOI: https://doi.org/10.1016/j.agrformet.2021.108555.

[14] José M Peña-Barragán et al. "Object-based crop identification using multiple vegetation indices, textural features and crop phenology". In: *Remote Sensing of Environment* 115.6 (2011), pp. 1301–1316. DOI: 10.1016/j.rse.2011.01.009.

[15] Darius Phiri et al. "Sentinel-2 data for land cover/use mapping: A review". In: *Remote Sensing* 12.14 (2020), p. 2291. DOI: https://doi.org/10.3390/rs12142291.

[16] Francis J Pierce and Peter Nowak. "Aspects of precision agriculture". In: *Advances in agronomy* 67 (1999), pp. 1–85. DOI: https://doi.org/10.1016/S0065-2113(08)60513-1.

[17] Yinghui Quan et al. "A novel feature extension method for the forest disaster monitoring using multispectral data". In: *Remote Sensing* 12.14 (2020), p. 2261. DOI: https://doi.org/10.3390/rs12142261.

[18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: (May 2015). DOI: 10.48550/ARXIV.1505.04597. arXiv: 1505.04597 [cs.CV].

[19] Marc Rußwurm and Marco Körner. "Self-attention for raw optical satellite time series classification". In: *ISPRS journal of photogrammetry and remote sensing* 169 (2020), pp. 421–435. DOI: https://doi.org/10.1016/j.isprsjprs.2020.06.006.

[20] Vivien Sainte Fare Garnot and Loic Landrieu. "Lightweight Temporal Self-Attention for Classifying Satellite Images Time Series". In: *arXiv preprint arXiv:2007.00586* (2020). DOI: https://doi.org/10.48550/arXiv.2007.00586.

[21] Mark Sandler et al. "Mobilenetv2: Inverted residuals and linear bottlenecks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 4510–4520. DOI: 10.1109/CVPR.2018.00474.

[22] Vishnu Sarukkai et al. "Cloud removal from satellite images using spatiotemporal generative networks". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2020, pp. 1796–1805. DOI: 10.1109/WACV45572.2020.9093564.

[23] Mohsen Shahhosseini et al. "Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt". In: *Scientific reports* 11.1 (2021), p. 1606. DOI: https://doi.org/10.1038/s41598-020-80820-1.

[24] Ramon Torres et al. "GMES Sentinel-1 mission". In: *Remote sensing of environment* 120 (2012), pp. 9–24. DOI: https://doi.org/10.1016/j.rse.2011.05.028.

[25] Amanda Veloso et al. "Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications". In: *Remote sensing of environment* 199 (2017), pp. 415–426. DOI: https://doi.org/10.1016/j.rse.2017.07.015.

[26] Jonathan Weekley, Joseph Gabbard, and Jerzy Nowak. "Micro-level management of agricultural inputs: emerging approaches". In: *Agronomy* 2.4 (2012), pp. 321–357. DOI: https://doi.org/10.3390/agronomy2040321.

[27] Fang Xu et al. "GLF-CR: SAR-enhanced cloud removal with global–local fusion". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 192 (2022), pp. 268–278. DOI: https://doi.org/10.1016/j.isprsjprs.2022.08.002.

[28] Qiangqiang Yuan et al. "Deep learning in environmental remote sensing: Achievements and challenges". In: *Remote Sensing of Environment* 241 (2020), p. 111716. DOI: https://doi.org/10.1016/j.rse.2020.111716.

[29] Chongyuan Zhang, Afef Marzougui, and Sindhuja Sankaran. "High-resolution satellite imagery applications in crop phenotyping: An overview". In: *Computers and Electronics in Agriculture* 175 (2020), p. 105584. DOI: https://doi.org/10.1016/j.compag.2020.105584.

[30] Zhe Zhu, Shixiong Wang, and Curtis E Woodcock. "Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images". In: *Remote sensing of Environment* 159 (2015), pp. 269–277. DOI: https://doi.org/10.1016/j.rse.2014.12.014.