Towards crop yield prediction using Automated Machine Learning

Jonathan Heil[1], Juan Manuel Valencia[1] and Anthony Stein[1]

Abstract: Recently, several Machine Learning models for crop yield prediction have been introduced in literature. The models differ in the underlying methodological approaches and show variations in the temporal and spatial resolution of the databases. For the creation of the models, a deep understanding of Machine Learning is required. Therefore, Automated Machine Learning, which aims to automate the creation process of Machine Learning models, offers a promising solution as an easy entry point in Machine Learning for crop yield prediction to non-professionals. Based on publicly available data for weather, phenological and yield observations, in this work, we created a dataset for winter wheat and winter barley on Germany's regional districts level. Furthermore, an initial evaluation of four state of the art Automated Machine Learning frameworks and three baseline models has been conducted. The results showed almost always significantly better performance of models created by Automated Machine Learning.

Keywords: crop yield prediction, Automated Machine Learning, open source data, winter wheat, winter barley

1 Introduction

The prediction of crop yields can play an important role to prevent yield failures and therefore contributes to the second Sustainable Development Goal "Zero Hunger" of the United Nations [Un15]. For instance, farmers can make use of crop yield predictions as a decision support tool to recognise potential losses and take actions to save the harvest [Ru19]. Furthermore, an accurate prediction of crop yields enables the different levels in the value chain to react to potential yield shortages or bumper crops, for example with marketing activities [Sh20]. Moreover, crop yield predictions also serve for researchers to examine influential factors such as weather conditions on crop yields [We20]. Several Machine Learning models have been introduced for the prediction of crop yields [vKC20]. The models vary in their underlying methodological approaches as well as the used data, e.g. Convolutional Neural Networks [Sr22] or Random Forest models [Vo19], weather station data [We20] or satellite images [Pe22]. However, strong expertise in Machine Learning and a deep understanding of the data is required to build reliable models. Accordingly, a high amount of manual effort for constructing a crop yield prediction model is needed. Therefore, Automated Machine Learning (AutoML) is a suitable solution to overcome the problems of expertise and time-consuming model construction. The idea

[1] University of Hohenheim, Department of Artificial Intelligence in Agricultural Engineering, Garbenstraße 9, 70599 Stuttgart, jonathan.heil@uni-hohenheim.de; juanmanuel.valenciapineda@uni-hohenheim.de, anthony.stein@uni-hohenheim.de

behind AutoML is to automate the processes to build a Machine Learning model [Fe20]. This involves typical steps in the Machine Learning process, such as preprocessing of the data, model selection or the optimization of hyperparameters. AutoML frameworks showed promising results in benchmark tests on classification and regression problems [Er20].

The heterogeneous and sophisticated research area of crop yield prediction exposes as a well-suited use case for AutoML: (1) Open access databases for yield observations [St22], climate data [De22b; De22a] and satellite images [Pe22] are available. (2) Researchers pursue different aims in their studies. Some are using the Machine Learning models to search for influential factors on crop yields [We20], others try to improve the models' prediction accuracies [Sr22]. AutoML could help the researchers without a deep knowledge in Machine Learning to build prediction models and thereby to focus on their specific research task [Er20]. In addition, researchers with sufficient knowledge of Machine Learning can build prediction models for benchmarking or even improving their own models by means of AutoML. But not only researchers will profit. If AutoML exposes as an easy and well performing entry point for applying Machine Learning by non-professionals, farmers in the first place without a deep knowledge in Machine Learning will be enabled to build advanced prediction models. Therefore, the aim of this study is to reveal the potential and suitability of AutoML when applied to crop yield prediction. The next chapter provides an overview of crop yield prediction models using Machine Learning. Following, the data acquisition process, the data analyzation and the investigated AutoML methods are described in detail. The experimental setup and the results are presented in the fourth chapter and discussed in chapter five. Finally, we draw a conclusion in the last chapter.

2 Crop yield prediction using Machine Learning

The number of publications about the prediction of crop yields using Machine Learning increased since 2013 with a peak in 2018 [vKC20]. Subsequently, we describe several studies, subdivided in Shallow Learning and Deep Learning approaches.

Shallow Learning approaches: A global prediction of yields for spring wheat, maize, soybeans and rice was conducted in [Vo19]. The data set combined yield observations and climate data from 1961 to 2008. A special focus was on climate extremes such as anomalies in temperature and precipitation. A Random Forest model served for the prediction of the crop yields. A crop yield prediction with focus on the influence of weather, soil and phenological stages was carried out in [We20]. A crop modelling approach as well as a Support Vector Machine were implemented for the prediction task. The data set comprises on crop yield observations, phenological phases, weather data and soil structures on a scale of the regional districts of Germany in the period from 1998 to 2018. The authors extracted the years with the highest yield losses and analysed potential factors causing the deficits. The climatic features seem to have an impact on yield losses, especially in the Support Vector Machine. Nevertheless, some years showed an overall decline of yields, whereas in other years the yields varied depending on the crops and regions. In [Ka20], Sentinel 2 [EE22] data were used to calculate several Vegetation Indices such as Normalized Difference Vegetation Index and Enhanced Vegetation Index. The crop yields were predicted with the indices using an Ordinary Least Squares Regression. Likewise, vegetation indices by means of Sentinel 2 data were calculated and combined with soil data for the yield prediction of soybeans for Austria in [Pe22]. Afterwards five different models, e.g. a Random Forest model and a Support Vector Machine, were trained to predict the yields. A Stochastic Gradient Descent model showed the lowest Mean Absolute Errors and Root Mean Squared Errors. A crop yield prediction for Algeria was carried out in [Me21]. The authors used preprocessed satellite data from the Anomaly Hotspots of Agricultural Production program [Re17]. The yield observations were provided by the agricultural ministry of Algeria. For the prediction task, several Machine Learning models were tested against two non-ML benchmarks. The first benchmark stated a congruence between the yield for a specific county and the yield mean of all counties. The second benchmark model based on a linear correlation between the Normalized Difference Vegetation Index and the yield. The prediction accuracy of the Machine Learning models varied during the vegetation period and, especially in the beginning of the year, the difference in the prediction accuracy between the Machine Learning models and the first benchmark model were negligible.

Deep Learning approaches: In [KWA19], a combination of a Convolutional Neural Network and a Recurrent Neural Network is introduced for the prediction of corn and soybean yields within the spatial area of the US-Corn Belt. The data set consisted of weather data, soil conditions, agricultural treatments and yield observations from 1980 to 2018. The presented approach outperformed the baseline models significantly. Another Deep Learning approach on base of genotype, weather, soil and yield data is presented in [KW19]. The authors predicted the yield potentials for different corn hybrids. Therefore, two Neural Networks were trained. The first Neural network served for the prediction of the yield of all hybrid crops at different locations, the second Neural Network was specified for the yield prediction of one particular corn hybrid. The yield potentials of the different corn hybrids were estimated by the difference of the results from the two Neural Networks. In [NNL19], multispectral and RGB images of barley and wheat were taken at different dates in the vegetation period. The yield data was captured during the harvest with a sensor mounted at the harvest combine. The authors discussed various settings for the architecture of a Convolutional Neural Network and stuck to already implemented Convolutional Neural Networks such as in [KSH17]. Predictions based on RGB images showed a lower Mean Absolute Errors than predictions based on multispectral images. In [Ma20], an unmanned aerial vehicle was equipped with multispectral, panchromatic and thermal sensors. The features of the multimodal data were fused using a Deep Neural Network. The crop yields were predicted with another Deep Neural Network and compared with baseline models.

3 Materials and methods

3.1 Data acquisition

The dataset created in this work is inspired by [We20], where open access data were used to predict crop yields and examine influential factors on crop yield losses. The data stems from two main providers: The federal and state statistical offices, specifically from "Regionaldatenbank Deutschland" [St22], and the DWD German Weather Service [De22a; De22b]. The regional database contains information for yearly crop yields in dt/ha of the two selected crops winter wheat and winter barley and a spatial scale corresponding to the regional districts of Germany, e.g. "Landkreis Osnabrück". The selected time period ranged from 1999 to 2021. Two types of stations of the DWD were used: *weather stations* [De22a] and *phenological stations* [De22b]. The weather database provides historical observations of more than 1000 stations distributed all over Germany. Nevertheless, we used those stations (853) reporting data within the time period from 1999 to 2021. Six weather variables with daily resolution were selected: maximum temperature (°C), minimum temperature (°C), relative humidity (%), sunshine duration (hrs), wind speed (m/s) and precipitation (mm). The phenological stations were a total of 6503 and after filtering the specified time period, 1809 were selected for winter wheat and 1833 for winter barley. These stations provide the day of the year (DOY) for BBCH-stages which is referred as Phase ID.

For obtaining the final dataset for crop yield, a preprocessing step has been conducted. First, we considered only yields corresponding to the deeper division level in the regional database, i.e. for counties (or districts), and avoiding aggregated information present in the data in a city, state or national level. A five-digit code represents the district depth, with the first two digits corresponding to the state number, the next one the government district and the last two digits to the district number. A total of 471 districts were available. In sum 17.6% of the data was missing for the 23 years and the two selected crops. For dealing with missing values we applied two operations. Initially, an average of the surrounding districts that belong to the same government district and to the same year has been used to impute the missing values. This reduced the percentage of missing data to 5%. Second, the mean historical yield value per district was used to fill those remaining gaps, leaving 0.28% of missing values, which correspond to two districts that did not contain information, and therefore were removed from the final dataset.

For the case of weather data, after filtering the recording time range that fits the crop yield data, we resampled the data with a seven days (weekly) average for all the variables, except for precipitation, in which a weekly sum was performed. The 52nd week was calculated with eight or nine days to cover the whole 365 or 366 days of the year, depending on whether it was a leap year or not. Every week of each variable equals a different feature for the dataset. The number of official districts can change depending on the year of their definition. Therefore, we used a shapefile with the districts division of 2021 [Op22], representing the administrative areas within Germany [Bu22]. The total number of districts in the shapefile was 401. Using the metadata information of the DWD weather stations, the latitude and longitude of the selected weather stations were obtained. Then, we took the shapefile and assigned each weather station to a district. 347 districts had at least one weather station, and from the 853 stations, 842 were assign to a district. Some districts were assigned several stations. Four of the eleven stations, that were not assigned to a district, were located on the North Sea and therefore removed. The remaining stations were not assigned due to uncertainties of the order of decimals in the definition of the latitude and longitude of the stations with respect to the georeferencing of the districts shapes, or vice versa, resulting in the stations being at the boundary of a district. Therefore, the distance method of Pythons Geopandas module was used to assign them to the closest shape [Ke22]. Also, a mean operation was applied to those districts with more than one station to obtain a final value per district.

A similar process was carried out to assign the phenological stations with the same shapefile. Phenological stations are ascribed districts wherein volunteer workers observe BBCH-stages of crops. The metadata did not contain all the stations that were in the winter wheat and winter barley datasets. A total of 1852 stations were available in the winter barley dataset, 19 were not found in the metadata, and 3 were not assigned to any district. For winter barley, 349 districts had at least one phenological station. In the case of winter wheat, 1830 stations were available but 21 were not included in the metadata, and 3 not assigned into a district. 352 districts had at least one phenological station for winter wheat. Finally, we joined the crop yield observations with the weather and phenological datasets. The weather stations had also missing values, at least one missing value was found in approximately 55% of the weather dataset, meaning that at least one week of any variable did not have a recorded value. The phenological data for winter barley presented 26% of all rows with missing values, which accounts for the years in specific districts where no phenological data was recorded, while 5% contained at least one missing phase ID. For winter wheat 25% of the data was missing, and 7% had at least one missing value. Therefore, a five-year-mean value for each feature of the district was calculated for imputing the missing values to cover the regional characteristics without an overweighting of influences by climate change.

3.2 Data description

The yearly mean yield values on national scale for winter wheat oscillate around 62 dt/ha and 83 dt/ha, while for winter barley they vary between 51 dt/ha and 75 dt/ha. Even if there have been some years with a marked decrease in the production of both crops (e.g. 2003, 2007, 2011, 2018), the general trend is positive. The regression slope for winter wheat is less distinct with a value of 0.3, while the winter barley has shown a higher increase through the years with a slope of 0.48. The mean yield during this 23-years period is approximately 73 dt/ha for winter wheat and 65 dt/ha for winter barley, respectively. The mean temperature values range between 8 degree Celsius and 10.7 degree Celsius, showing with a minimum in 2010 and a maximum in 2018. The mean temperature does not exhibit a significant trend, with magnitude of 0.02 in the slope of the regression line. For the case of average precipitation over Germany, the values oscillate between 350 mm and 670 mm. The global minimum occurred in 2018. However, another significant local minimum happened in 2003. The maximum mean precipitation value occurred in 2002. Opposite to temperature, the trend for precipitation is evident, showing a decrease during the last two decades. The occurrence of the phenological stages for winter wheat and winter barley were time-displaced. For winter wheat the sowing period (phase 10) starts later (277-283) than for winter barley (260-267). Consequently, the rest of the stages also appear later. If the sowing happened later in the year, the rest of the stages are deferred as well and vice versa. This is the case for years like 2001, 2007, 2013 or 2014 in winter barley crops. The years 2003 and 2018 show the opposite behaviour for winter barley, having an earlier sowing period and nevertheless the steam elongation stage occurred later.

3.3 Automated Machine Learning methods

In agricultural sciences, few experiments with AutoML have been conducted, e.g. for the identification of weeds [Es21] or the identification of pest aphid species [Ha19]. AutoML aims to automate tedious tasks in the construction and optimization of Machine Learning models, e.g. the preprocessing of data, the selection of suitable models or the optimization of hyperparameters. Several frameworks have been introduced offering an end-to-end-solution for applying Machine Learning. Therefore, we compared four different state of the art AutoML frameworks: AutoGluon, Auto-sklearn, H2O and TPOT. For this study we made use of the 2020 introduced AutoGluon Tabular in the version 0.5.2 [Er20]. AutoGluon Tabular includes datatype specific preprocessing and an ensemble method with stacking of multiple layers which are trained sequentially. TPOT relies on the scikit-learn package and makes use of genetic programming for their algorithm and hyperparameter optimization [Ol16]. We used the version 0.11.7. Auto-sklearn was first introduced in 2015 [Fe19]. Meanwhile, an updated version with improvements, for example in the model selection process, is available [Fe20]. For our experiments, we used the currently latest version 0.15.0 [Ma22]. Auto-sklearn is based on Bayesian optimization methods. The framework provides 15 different models from the Python package scikit-learn. Auto-sklearn includes Meta Learning based on meta-features, which are computed for every input dataset. Moreover, the ensemble building process is conducted by the ensemble selection method [Fe19]. H2O AutoML is available since 2017 [LP20]. The H2O AutoML framework draws on models from H2O, which is a Java-based, in-memory machine learning platform [H222]. H2O makes use of fast random search for their algorithm selection and hyperparameter optimization. For our experiments, we used the currently latest version 3.38.0.1. All four frameworks have in common to offer different configuration parameters, for example for excluding specific machine learning models.

4 Experiments

4.1 Experimental setup

As we stated in the motivation that AutoML opens the opportunity to quickly construct Machine Learning models without the lengthy and knowledge intensive process of tuning, the configuration parameters of all AutoML frameworks have been left in the default settings with one exception. We restricted the time limit for training to 30 minutes per run due to limited resources. The experiments were carried out on a notebook equipped with an AMD Ryzen 9 5900HX CPU, 32 GB RAM and a NVIDIA GeForce RTX 3080 GPU. The AutoML frameworks investigated in this study make use of stochastic algorithms for the inner optimization process. Therefore, 30 runs for every framework per dataset were conducted to produce statistical expressive results. The datasets winter wheat and winter barley were split into a training set and a testing set. The AutoML frameworks include validation steps such as k-fold cross validation. Therefore, a separate validation set was not necessary. For training, the years from 1999 to 2018 were used. The hold-out periods from 2019 to 2021 built the testing set. For benchmark models, we followed the work of [Sr22], which published a crop yield prediction on an equal dataset [We20]. Therefore, we built a Random Forest model (RF), a Support Vector Regression model (SVR) as well as a Deep Neural Network (DNN) with the same hyperparameters as in [Sr22]. Other models, such as in the second chapter presented, showed too big differences in the time periods, the crop types, the research areas, the data and the evaluation to act as benchmark. For evaluating the models, we computed the metrics Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and the Concordance Correlation Coefficient (CCC). The results were tested on statistical significance by means of an analysis of variance and pairwise comparisons with f-tests. Few results were not normally distributed. Therefore, an additional Friedman-test and pairwise comparisons with Mann-Whitney-U-tests were conducted. The significance level was set to 0.05.

4.2 Results

In Tab. 1, the mean values over 30 runs of the evaluation metrics based on the predictions for the winter wheat testing dataset are presented. The small letters indicate whether the mean values for all 30 runs had significant differences (different small letters column-wise) or not (same small letters column-wise). The analysis of variance and the Friedman-test stated that within the evaluation metrics are significant differences. The SVR produced with an average RMSE of 10.90 and an average MAE of 8.60 the worst predictions. The CCC values for SVR were close to zero, so we rounded them. The lowest average RMSE and average MAE was achieved by H2O. In the pairwise comparison to the other three AutoML frameworks, H2O showed significantly lower RMSE and MAE values. Whereas the highest average CCC occurred for AutoGluon. The pairwise comparison confirmed, that AutoGluon has a significantly higher CCC than Auto-sklearn, H2O and TPOT.

Framework/Model	RMSE (+/- 1SD)	MAE (+/- 1SD)	CCC (+/- 1SD)
AutoGluon	8.88[a] (+/- 0.24)	7.04[a] (+/- 0.23)	**0.65[a] (+/- 0.02)**
Auto-sklearn	8.92[a] (+/- 0.33)	7.00[a] (+/- 0.32)	0.50[b] (+/- 0.03)
H2O	**8.59[b] (+/- 0.12)**	**6.68[b] (+/- 0.10)**	0.52[c] (+/- 0.02)
TPOT	9.34[c] (+/- 0.62)	7.28[a] (+/- 0.60)	0.54[d] (+/- 0.05)
Benchmark RF	9.63[d] (+/- 0.04)	7.55[c] (+/- 0.04)	0.28[e] (+/- 0.01)
Benchmark SVR	10.90[e] (+/- 0.02)	8.60[d] (+/- 0.02)	0.00[f] (+/- 0.00)
Benchmark DNN	10.16[f] (+/- 0.68)	8.01[e] (+/- 0.64)	0.43[d] (+/- 0.04)

Tab. 1: Means of the evaluation metrics of all 30 runs based on the winter wheat testing dataset

The average evaluation metrics for the predictions based on the winter barley training dataset are displayed in Tab. 2. The two benchmark models SVR and DNN showed again a significant difference to the AutoML frameworks. The benchmark SVR model had the highest average RMSE and average MAE as well as the lowest CCC. In the pairwise comparison, the MAE values did not significantly differ between RF and Auto-sklearn. The lowest RMSE and MAE values as well as the highest CCC were provided by H2O.

Framework/Model	RMSE (+/- 1SD)	MAE (+/- 1SD)	CCC (+/- 1SD)
AutoGluon	10.22[a] (+/- 0.29)	8.40[a] (+/- 0.29)	0.41[a] (+/- 0.02)
Auto-sklearn	10.44[b] (+/- 0.44)	8.57[ab] (+/- 0.46)	0.39[b] (+/- 0.04)
H2O	**9.57[c] (+/- 0.15)**	**7.70[c] (+/- 0.15)**	**0.43[c] (+/- 0.03)**
TPOT	10.07[a] (+/- 0.33)	8.11[d] (+/- 0.34)	0.32[d] (+/- 0.05)
Benchmark RF	10.63[d] (+/- 0.06)	8.60[b] (+/- 0.06)	0.20[e] (+/- 0.01)
Benchmark SVR	12.49[e] (+/- 0.03)	10.13[e] (+/- 0.03)	0.00[f] (+/- 0.00)
Benchmark DNN	12.03[f] (+/- 0.63)	9.81[f] (+/- 0.58)	0.31[d] (+/- 0.03)

Tab. 2: Means of the evaluation metrics of all 30 runs based on the winter barley testing dataset

5 Discussion

The results show significant differences in the evaluation metrics of the different AutoML frameworks for both datasets. Especially H2O performed well on the evaluation metrics RMSE and MAE. A reason for the superior results of H2O is deemed to the internally used and automated algorithm and hyperparameter selection process. H2O is based on a fast random search method [H222]. Auto-sklearn uses Bayesian Optimization and Meta-Learning [Fe19], TPOT an evolutionary method for the algorithm and hyperparameter selection process [Ol16]. The experiments with AutoGluon did not have a hyperparameter optimization because it is disabled in the default settings [Er20]. AutoGluon reached for winter wheat the highest CCC between the predicted and true values for the yield. In Fig. 1, the distribution of the true and predicted yields by H2O and AutoGluon for winter barley are displayed. Even though there was a significant difference in the pairwise comparison between H2O and AutoGluon for the CCC, the visual difference is not easy to recognize.

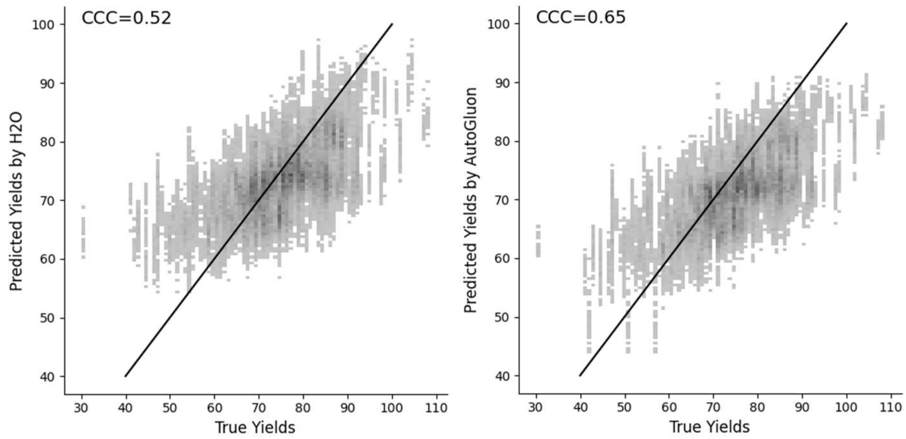

Fig. 1: Distribution of true and by H2O (l.) and AutoGluon (r.) predicted yields for wheat

A comparison with prediction results of the above-mentioned publications in the second chapter is difficult to realize because of differences in the time periods, the crop types, the regional areas, the data types and structures as well as the evaluations of the models. The experimental setup with a 30-minute time limit only delivers a first indication. Therefore, further experiments with a wider range of time limits should be examined. As the frameworks base on different algorithm selection and hyperparameter optimization processes, the prediction results could vary on various run time limits. Additionally, the AutoML frameworks offer different configuration parameters for adjusting. In this study we did not change the default settings to examine an off-the-shelf usage of AutoML for non-professionals in Machine Learning. Further experiments with different settings of the configuration parameters are worth considering. The benchmark models revealed almost for every evaluation metric a significantly inferior result than the AutoML frameworks. Due to the lack of knowledge about the detailed preprocessing steps of the database in [Sr22], where we extracted the hyperparameters of the benchmark models, it is expectable that the benchmark models will have problems to fit our datasets. Surprisingly, the MAE for the predictions on the winter barley dataset did not significantly vary between RF and Auto-sklearn. Although we used the same initial data as in [Sr22], the preprocessing of the data captured a high amount of manual effort due to missing values, discrepancies in the regional districts compared to the third level of the nomenclature of territorial units for statistics (NUTS-3), which are used in [Sr22], and the assignment of the weather stations data to the corresponding districts. Therefore, we decided to elaborate on the data acquisition part in detail. The section about related work on crop yield prediction using Machine Learning exposed that several more open access databases, such as satellite images [Pe22], are available for crop yield prediction. For further studies these databases could serve for experiments with AutoML frameworks.

6 Conclusion

The aim of this work was an initial investigation of the suitability of AutoML for crop yield prediction. Thus, we created a dataset with winter wheat and winter barley yields, weather data and phenological data. Subsequently, we predicted the crop yields by means of four AutoML frameworks and three benchmark models. The experimental results showed a significantly improved performance in terms of RMSE, MAE and CCC for AutoML methods in comparison to hand-tuned RF, SVR and DNN models. There are two main contributions this study reveals: (1) AutoML can offer an easy entry point for the application of Machine Learning in crop yield prediction. But configuration parameters, such as the restriction of the runtime, have to be taken in consideration. Therefore, further experiments with changes in the configuration parameters are worth considering. (2) The data acquisition and preprocessing part played an important role in this study. Several challenges with the data, such as many missing values or the assignment of the weather stations, occurred. Therefore, we offer in this study a detailed description how we created and manipulated the data.

Acknowledgements: The research was conducted within the scope of NaLamKI [Bo23] and funded by the Federal Ministry for Economics and Climate Action with the funding code 01MK21003[A-J].

Bibliography

[Bo23] Bosse, S. et al.: Nachhaltige Landwirtschaft mittels Künstlicher Intelligenz – Ein plattformbasierter Ansatz für Forschung und Industrie. 43. GIL-Jahrestagung, 2023.

[Bu22] Bundesamt für Kartographie und Geodäsie: Verwaltungsgebiete 1:250000. https://gdz.bkg.bund.de/index.php/default/verwaltungsgebiete-1-250-000-stand-01-01-vg250-01-01.html, accessed 19 Oct 2022.

[De22a] Deutscher Wetterdienst: Historische Klimadaten. https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/kl/historical/, accessed 19 Oct 2022.

[De22b] Deutscher Wetterdienst: Phänologische Beobachtungen. https://opendata.dwd.de/climate_environment/CDC/observations_germany/phenology/annual_reporters/crops/historical/, accessed 19 Oct 2022.

[EE22] European Union; European Space Agency: Copernicus Open Access Hub. https://scihub.copernicus.eu/, accessed 30 Oct 2022.

[Er20] Erickson, N. et al.: AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data, 2020.

[Es21] Espejo-Garcia, B. et al.: Testing the Suitability of Automated Machine Learning for Weeds Identification. AI 1/2, pp. 34-47, 2021.

[Fe19] Feurer, M. et al.: Auto-sklearn: Efficient and Robust Automated Machine Learning. In (Hutter, F.; Kotthoff, L.; Vanschoren, J. Eds.): Automated Machine Learning. Methods, Systems, Challenges. Springer, Cham, pp. 113-134, 2019.

[Fe20] Feurer, M. et al.: Auto-Sklearn 2.0: Hands-free AutoML via Meta-Learning, 2020.

[H222] H2O.ai: Welcome to H2O 3. https://docs.h2o.ai/h2o/latest-stable/h2o-docs/welcome.html, accessed 19 Oct 2022.

[Ha19] Hayashi, M. et al.: Automated machine learning for identification of pest aphid species (Hemiptera: Aphididae). Applied Entomology and Zoology 4/54, pp. 487-490, 2019.

[Ke22] Kelsey Jordahl et al.: geopandas/geopandas: v0.12.1. Zenodo, 2022.

[KSH17] Krizhevsky, A.; Sutskever, I.; Hinton, G. E.: ImageNet classification with deep convolutional neural networks. Communications of the ACM 6/60, pp. 84-90, 2017.

[KW19] Khaki, S.; Wang, L.: Crop Yield Prediction Using Deep Neural Networks. Frontiers in plant science 10, p. 621, 2019.

[KWA19] Khaki, S.; Wang, L.; Archontoulis, S. V.: A CNN-RNN Framework for Crop Yield Prediction. Frontiers in plant science 10, p. 1750, 2019.

[LP20] LeDell, E.; Poirier, S.: H2O AutoML: Scalable Automatic Machine Learning. 7th ICML Workshop on Automated Machine Learning, 2020.

[Ma20] Maimaitijiang, M. et al.: Soybean yield prediction from UAV using multimodal data fusion and deep learning. Remote Sensing of Environment 237, p. 111599, 2020.

[Ma22] Machine Learning Professorship Freiburg: auto-sklearn. https://automl.github.io/auto-sklearn/master/, accessed 19 Oct 2022.

[Me21] Meroni, M. et al.: Yield forecasting with machine learning and small data: What gains for grains? Agricultural and Forest Meteorology 308-309, p. 108555, 2021.

[NNL19] Nevavuori, P.; Narra, N.; Lipping, T.: Crop yield prediction with deep convolutional neural networks. Computers and Electronics in Agriculture 163, p. 104859, 2019.

[Ol16] Olson, R. S. et al.: Automating Biomedical Data Science Through Tree-Based Pipeline Optimization. In (Squillero, G.; Burelli, P. Eds.): Applications of Evolutionary Computation. Springer International Publishing, Cham, pp. 123-137, 2016.

[Op22] Opendatasoft: Shapefile Germany. https://data.opendatasoft.com/explore/dataset/georef-germanykreis%40public/export/?disjunctive.lan_code&disjunctive.lan_name&disjunctive.krs_code&disjunctive.krs_name&disjunctive.krs_name_short&refine.lan_code=04, accessed 19 Oct 2022.

[Pe22] Pejak, B. et al.: Soya Yield Prediction on a Within-Field Scale Using Machine Learning Models Trained on Sentinel-2 and Soil Data. Remote Sensing 9/14, p. 2256, 2022.

[Re17] Rembold, F. et al.: ASAP - Anomaly hot Spots of Agricultural Production. 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), 2017.

[Ru19] Rupnik, R. et al.: AgroDSS: A decision support system for agriculture and farming. Computers and Electronics in Agriculture 161, pp. 260-271, 2019.

[Sh20] Sharma, R. et al.: A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. Computers & Operations Research 119, p. 104926, 2020.

[Sr22] Srivastava, A. K. et al.: Winter wheat yield prediction using convolutional neural networks from environmental and phenological data. Scientific reports 1/12, p. 3215, 2022.

[St22] Statistische Ämter des Bundes und der Länder: Erträge ausgewählter landwirtschaftlicher Feldfrüchte. https://www.regionalstatistik.de/genesis//online?operation=table&code=41241-01-03-4&bypass=true&levelindex=0&levelid=1662987040832#abreadcrumb, accessed 19 Oct 2022.

[Un15] United Nations: Transforming our World: The 2030 Agenda for Sustainable Development. https://sdgs.un.org/sites/default/files/publications/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf, accessed 13 Sep 2022.

[vKC20] van Klompenburg, T.; Kassahun, A.; Catal, C.: Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture 177, p. 105709, 2020.

[Vo19] Vogel, E. et al.: The effects of climate extremes on global agricultural yields. Environmental Research Letters 5/14, p. 54010, 2019.

[We20] Webber, H. et al.: No perfect storm for crop yield failure in Germany. Environmental Research Letters 10/15, p. 104012, 2020.