Communication Papers of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 77–82 DOI: 10.15439/2024F5990 ISSN 2300-5963 ACSIS, Vol. 41

Combining Local and Global Weather Data to Improve Forecast Accuracy for Agriculture

František Koutenský*, Jiří Pihrt*, Miroslav Čepek*, Vojtěch Rybář*, Petr Šimánek*,
Michal Kepka†, Karel Jedlička‡, and Karel Charvát§
*Faculty of Information Technology, Czech Technical University in Prague
Thákurova 9, Praha 6, Czech Republic
Email: {koutefra, pihrtjir, miroslav.cepek, vojtech.rybar, petr.simanek}@fit.cvut.cz
†Lesprojekt s.r.o.
Martinov 197, Záryby, Czech Republic
Email: kepka@lesprojekt.cz
‡Wirelessinfo
Cholinská 1048/19, Litovel, Czech Republic
Email: jedlicka@wirelessinfo.cz
§HELP SERVICE – REMOTE SENSING s.r.o.
Husova 2117, Benešov, Czech Republic
Email: charvat@hsrs.cz

***Abstract*—Accurate local weather forecasting is vital for farmers to optimize crop yields and manage resources effectively, but existing forecasts often lack the precision required locally. This study explores the potential of combining data from local weather stations with global forecasts and reanalysis data to improve the accuracy of local weather predictions. We propose integrating the HadISD data set, which contains data from 27 stations in the Czech Republic, with the Global Forecast System predictions and ERA5-Land reanalysis data. Our goal is to improve 24-hour weather forecasts using Multilayer Perceptrons, CatBoost, and Long Short-Term Memory neural networks. The findings demonstrate that combining local weather station data with global forecasts improves the accuracy of weather predictions in specific locations. This advancement holds promise in optimizing agricultural practices and mitigating weather-related risks in the region.**

I. Introduction

ACCURATE weather forecasting is crucial for farmers to make informed decisions that optimize crop yields and manage resources efficiently. However, available weather forecasts often lack the precision required for agricultural planning, leading farmers to invest in their own weather stations. This study explores the potential of combining data from local weather stations with global forecasts to improve local weather predictions.

As local weather station data could suffer from various inconsistencies, we propose testing this idea by integrating the local weather station data (HadISD data set [1], [2]), with the Global Forecast System (GFS) predictions. In addition, we incorporate ERA5-Land reanalysis data to introduce information on weather conditions in surrounding areas. Our goal is to improve the accuracy of 24-hour weather forecasts by evaluating three machine learning techniques: Multilayer Perceptons (MLP) [3], gradient-boosting regression trees method CatBoost [4], and Long Short-Term Memory (LSTM) [5] neural network.

In this study we focus on the Czech Republic, utilizing data from 27 stations recorded in the HadISD data set within the country and close neighborhood. We supplement this data with GFS forecasts, which provide weather predictions on a 0.25-degree grid resolution, corresponding to an approximately 27.8 km $\times$ 27.8 km area in Central Europe. The GFS model predicts various meteorological parameters at different atmospheric levels, offering a comprehensive data set for creating our machine-learning models.

To further improve our predictions, we employ the ERA5-Land data set, renowned for its high-accuracy reanalysis data. Recognizing the latency in the availability of ERA5-Land's data, we trained a U-Net [6] model to map the GFS forecast data to the ERA5-Land's high-resolution grid. This approach enables us to generate ERA5-Land-like predictions in near real-time, potentially enhancing the accuracy of our weather forecasts.

Our methodology involves training and comparing the performance of CatBoost, MLP, and LSTM machine learning techniques against two baseline models, the raw GFS predictions, and the last measured values from the stations. The training data set is constructed using weather data from 2022, pairing each station's observations with corresponding GFS grid data. The models are then validated using data from 2023.

In this paper, we present a detailed analysis of our approach, including data pre-processing, model architectures, training processes, and evaluation metrics. We discuss the performance improvements achieved by integrating ERA5-Land predictions.

Our findings demonstrate that combining local weather station data with global forecasts and incorporating ERA5-

Land reanalysis data can substantially improve the accuracy of weather predictions in specific locations. This advancement holds significant promise for optimizing agricultural practices, mitigating weather-related risks, and ultimately enhancing food security in the region.

II. Related Research

Weather forecasting is a well-explored area of research with numerous methodologies and models developed to improve prediction accuracy. The integration of local weather station data with global models is promising in improving forecast precision, particularly in agricultural contexts.

a) Global Forecast Models and Their Limitations: Global forecast models such as the Global Forecast System (GFS) from the National Centers for Environmental Prediction (NCEP) and the European Centre for Medium-Range Weather Forecasts (ECMWF) provide detailed predictions on a global scale. These models use numerical algorithms and vast amounts of atmospheric data to produce forecasts at various temporal and spatial resolutions. There is a large body of literature devoted to improving the global model using in-situ data, (e.g. [7], [8]) However, the coarse resolution of these models often limits their utility for local weather predictions, which are crucial for agricultural decision-making.

b) Machine Learning in Weather Forecasting: Machine learning has become increasingly prominent in weather forecasting due to its ability to handle large data sets and capture complex, nonlinear relationships within the data. Various ML techniques, including neural networks, support vector machines, and ensemble methods, have been applied to enhance forecast accuracy. For example, Multilayer Perceptron (MLP) models have been used to predict temperature [9] and precipitation with notable success. More recently, gradient-boosting algorithms like CatBoost have demonstrated superior performance in regression tasks related to weather prediction [10]. Various concepts have been used for weather forecasting based on weather station data: a 2D-convolutional LSTM in [11], Temporal Convolutional Network (TCN) in [12], and Copulas in [13], to name a few. Large convolutional neural networks (CNNs) were used for global machine learning weather forecasting [14].

c) Hybrid Approaches Combining Global and Local Data: Several studies have explored hybrid approaches that combine global forecast data with local observations to improve prediction accuracy. These methods often involve the use of statistical downscaling or machine learning models to integrate diverse data sources. For instance, a study by [15] demonstrated the effectiveness of combining a global climate model with local weather station observations using an ML model. Similarly, the other works [16]) highlighted the benefits of integrating ERA5 reanalysis data with local meteorological data to refine precipitation forecasts.

d) Application of U-Net for Spatial Predictions: The U-Net [6] architecture, initially developed for biomedical image segmentation, has been adapted for various geospatial applications, including weather forecasting. U-Net's ability to capture spatial hierarchies and produce high-resolution output maps makes it suitable for transforming coarse global forecast data into fine-scale local predictions. Recent studies have successfully employed U-Net to downscale climate model outputs, demonstrating significant improvements in prediction accuracy and spatial resolution [17].

e) Focus on Agricultural Applications: The intersection of weather forecasting and agriculture has been a focal point for research aimed at enhancing food security and optimizing resource management. Accurate local weather predictions can help farmers make timely decisions regarding planting, irrigation, and harvesting, thereby improving crop yields and reducing losses. Multiple research papers ([18], [19]) have emphasized the potential of combining local weather station data with advanced modeling techniques to support precision agriculture.

III. Data

In this work, we utilize three data sets, each serving a different purpose:

- **HadISD** [2], [1]: This data set is used to extract hourly records from 27 weather stations situated randomly across the area of the Czech Republic. These records contain the variables we aim to predict (temperature, dew point, wind speed), as well as additional variables like cloud coverage, precipitation depth across multiple periods (1h, 2h, ..., 24h), and sea level pressure.
- **GFS** [20]: Unlike HadISD's station-specific data, the GFS (Global Forecast System) data set delivers broader area weather predictions with various frequencies and forecast lead times, encompassing a comprehensive range of atmospheric variables at various altitudes. This feature-rich data set serves both as a baseline for our predictions and as a source for enhancements, utilizing every available feature across all altitude levels.
- **ERA5-Land** [21]: The ERA5-Land data set is a reanalysis tool, meaning it does not provide real-time data but rather offers a retrospective view of land variables over several decades. As a reanalysis data set, ERA5-Land integrates model data with historical observations using the laws of physics to create a globally consistent and comprehensive data set. This characteristic makes it ideal for understanding past climate conditions but limits its use for immediate weather events. We utilize the ERA5-Land data set to train an additional model that can generate features from the GFS data set. The details of this technique will be discussed further in the text.

Having HadISD and GFS data sets, for our experiments, we mapped a GFS rectangle to each HadISD station and merged the data sets accordingly. The final combined data set, consisting of GFS and HadISD data sets, contains 157 features, where 25 of them come from HadISD and 132 of them are from GFS.

We identified the most significant features for our weather prediction task by leveraging the CatBoost model (described later) and computing its SHAP values [22]. These values were

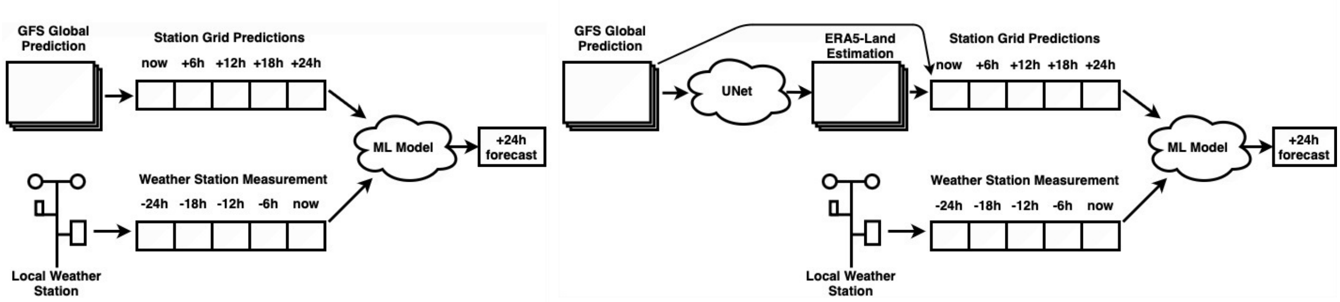

Fig. 1. Illustration of data flows in our experiments. The left schema illustrates a setup where we forecast local weather with the use of weather station data and GFS forecasts. The right schema shows how we improve the accuracy of the forecast with the superresolution model estimating the ERA5-Land reanalysis based on GFS data.

averaged across all target features (temperature, dew point, wind speed) and all stations. Below are the identified features, along with their origins and a brief explanation of each:

- **Helicity** (GFS, height above ground layer): Measures the potential for rotation in the atmosphere, which can be important for predicting severe weather events.
- **Temperature** (GFS, surface): A fundamental parameter, that influences various atmospheric processes and weather conditions.
- **Precipitable water** (GFS, atmosphere single layer): Representing the total atmospheric water vapor, is essential for forecasting precipitation and humidity levels.
- **Dew point** (HadISD): Dew point temperatures from station data indicate the atmospheric moisture content, aiding in humidity and fog predictions.
- **Minimum temperature** (GFS, height above ground): Minimum temperature above ground level might help in identifying cold spells and frost conditions.
- **Temperature** (HadISD): Observed temperatures from station data, a direct measurement of local weather conditions, is crucial for accurate forecasting.

IV. Experiment Design

The model takes the last 24 hours of weather data from the station and the weather forecast for the next 24 hours for the corresponding grid. Based on this, the model creates a prediction of selected weather parameters 24 hours from now. The left side of Figure 1 illustrates the idea.

The training data set is constructed from weather data from selected weather stations combined with its GFS rectangle for the year 2022. The next step is to split the whole history of station data and GFS predictions into time windows containing inputs and corresponding target values from the HadISD station. For the validation of the model, we used data from 2023.

As the experimental results below show, the models can improve the accuracy of the local prediction. Since the GFS data set has low spatial resolution and is known to have limited accuracy in predictions, we decided to explore possibilities to incorporate another data set, with better accuracy and higher spatial resolution. The ERA5-Land data set [21], [23], which is suitable for our case, is a reanalysis of past weather conditions and it is assessed to be a good approximation of the actual weather. However, the ERA5-Land predictions are available with considerable delay and thus it is not possible to use it directly.

To tackle this issue, a model to estimate the ERA5-Land values was developed. We use a subset of the GFS data set as coarse grid input and a U-Net architecture to create a fine-grained grid estimating ERA5-Land. In the training process of the U-Net superresolution model, GFS predictions and corresponding ERA5-Land data for the years 2015 to 2021 are utilized. The data were split time-wise to prevent information leaks. The earlier 80% are used for training and the remaining 20% are used for validation. The incorporation of the ERA5-Land data set into our AI forecast is illustrated on the right side of Figure 1.

V. Methodology

Our experiments can be divided into two stages. The aim of the first stage was to train a model that could transform GFS predictions to be closer to ERA5-Land. The variables we focused on included temperature, dew point, and the u and v components of wind.

We approached this problem as an image-to-image translation task. For this purpose, we employed the U-Net [6] convolutional neural network. The name is inspired by its U-shaped architecture shown in Figure 2. It can be described as a symmetrical encoder-decoder architecture, consisting of a contracting path to capture context and a symmetric expanding path that enables precise localization, further enhanced by the usage of skip/shortcut connections.

First, GFS data are interpolated to match the grid of ERA5-Land. Both are then cropped to 128x64 pixels, which is enough to cover the Czech Republic, and the dimensions are divisible by 16, which is required by U-Net. The variables are concatenated in the channel dimension and normalized. U-Net is trained with 4 input channels and 4 output channels, as well as other hyperparameters listed in Table I. Example prediction is shown in Fig. 3. The output of a trained U-Net is then denormalized, and time series for each station is generated by interpolating the variables at specific coordinates of the station.

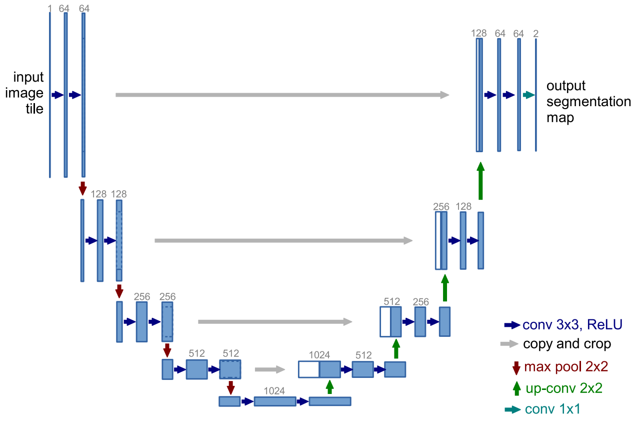

Fig. 2. U-Net architecture [6] used to estimate ERA5-Land like weather forecasts from GFS predictions. The exact hyperparameters of the U-Net architecture are outlined in Table I.

TABLE I
U-NET TRAINING HYPERPARAMETERS

Hidden dimensions	[32, 64, 128, 256, 512]
Batch size	16
Optimizer	Adam
Learning rate	1×10^{-3}
Loss function	MAE

The goal of the first stage was to create more reliable features by utilizing the GFS and ERA5-Land data sets for subsequent weather forecasts at specific stations. This process effectively generates 4 new features, which are then incorporated into the data set used for station forecasts.

The second stage involved comparing models on our enhanced station data set, which was augmented by the U-Net-generated ERA5-Land predictions. Since we are dealing with time series data, we utilized the LSTM (Long Short-Term Memory) network [5], known for its ability to capture long-term dependencies and patterns in sequential data. The LSTM model is based on a sequence-to-sequence (seq2seq) architecture. The model consists of an encoder that encodes the input time series sequence into a fixed-length vector representation and a decoder that generates the predicted output sequence based on the encoded representation. The encoder takes an input sequence of a given length and produces hidden states and cell states that capture the temporal dependencies in the data. The decoder takes the last hidden state of the encoder as its initial hidden state and generates the output sequence recursively. The input sequence consists of four previous time steps of a 10-time series, selected based on SHAP [24] analysis of a CatBoost model. The LSTM seq2seq model takes this input and regressively generates predictions for the next four time steps, effectively making a 24-hour forecast for the desired quantities.

Alongside LSTM, we also examined MLP (Multi-Layer Perceptron) [3] and CatBoost [4] models. MLP, a class of feedforward artificial neural networks, consists of multiple layers of nodes, with each node connected to every node in the subsequent layer. It is particularly adept at capturing complex relationships in the data through its dense connections and non-linear activation functions. MLPs are typically composed of an input layer, one or more hidden layers, and an output layer. Each node (or neuron) in a layer applies a weighted sum of the inputs followed by a non-linear activation function, which allows the model to learn and represent complex functions. MLPs are effective in scenarios where the relationship between inputs and outputs is highly non-linear and intricate, making them suitable for various predictive modeling tasks in weather forecasting.

CatBoost, a gradient-boosting algorithm, is well-suited for this task because it handles categorical data well and mitigates the problem of overfitting. It builds an ensemble of decision trees where each new tree is trained to correct the errors of the previous ones, leading to improved accuracy. CatBoost is particularly effective in scenarios where the relevance of historical data might vary, providing robust predictions even when the importance of past data fluctuates. One of its key advantages is its ability to handle categorical features natively, without requiring extensive preprocessing or encoding, which simplifies the model training process and enhances performance. Additionally, CatBoost incorporates ordered boosting and other advanced techniques to reduce overfitting and improve the generalization of the model.

For each of these three models, we developed two versions: (a) one trained on GFS and station data, and (b) another trained on GFS, ERA5-Land predictions, and station data. This protocol was designed to assess the impact of ERA5-Land predictions, which we hypothesized might play a crucial role in enhancing the accuracy and reliability of the weather forecasts. The Figure 1 illustrates both versions.

VI. EXPERIMENTAL RESULTS

Table II summarises our results. We present results for three models in two versions, as mentioned above. In addition, two additional baseline techniques were added for comparison: using the GFS forecast directly and assuming the weather in 24 hours will be the same as the current conditions. The forecast results are presented for three weather parameters: Temperature, Dew Point, and Wind Speed, all measured at a height of 2 meters above ground. The tables show the mean absolute error (MAE) of all examined methods for all target parameters. The metric is calculated over all 27 selected weather stations for predictions in the year 2023.

Table II shows that the direct GFS forecast has the second worst accuracy. It is on average more than $2°C$ off the actual measured temperature. Using actual current weather achieves slightly worse accuracy with $2.5°C$ average absolute error. Similar results are seen for the Dew Point and Wind Speed.

The results for our models show that we can greatly improve over the GFS forecasts as well as the last value prediction. The models built on top of the GFS predictions and weather recorded by a weather station achieve better accuracy. The best model, CatBoost, achieves mean absolute errors of $1.07°C$ in Temperature, $1.04°C$ in Dew Point, and 1.01 m/s in

TABLE II
ACCURACY OF PRESENTED METHODS AND COMPARISON TO BASELINE MODELS. THE VALUES SHOW THE MEAN ABSOLUTE ERROR BETWEEN THE MODEL FORECAST AND ACTUALLY MEASURED VALUE. ALL FORECASTS ARE 24 HOURS INTO THE FUTURE.

Method	**Temperature** 2m above ground MAE [$°C$]	**Dew Point** 2m above ground MAE [$°C$]	**Wind Speed** 2m above ground MAE [m/s]
Last value (persistence)	2.53	2.54	1.83
Direct GFS Prediction	2.18	1.38	2.04
LSTM with GFS	1.20	1.14	1.12
MLP with GFS	1.16	1.13	1.09
CatBoost with GFS	1.07	1.04	1.01
LSTM with GFS and estimated ERA5-Land	1.15	1.23	1.12
MLP with GFS and estimated ERA5-Land	1.14	1.15	1.05
CatBoost with GFS and estimated ERA5-Land	**1.06**	**1.02**	**1.01**

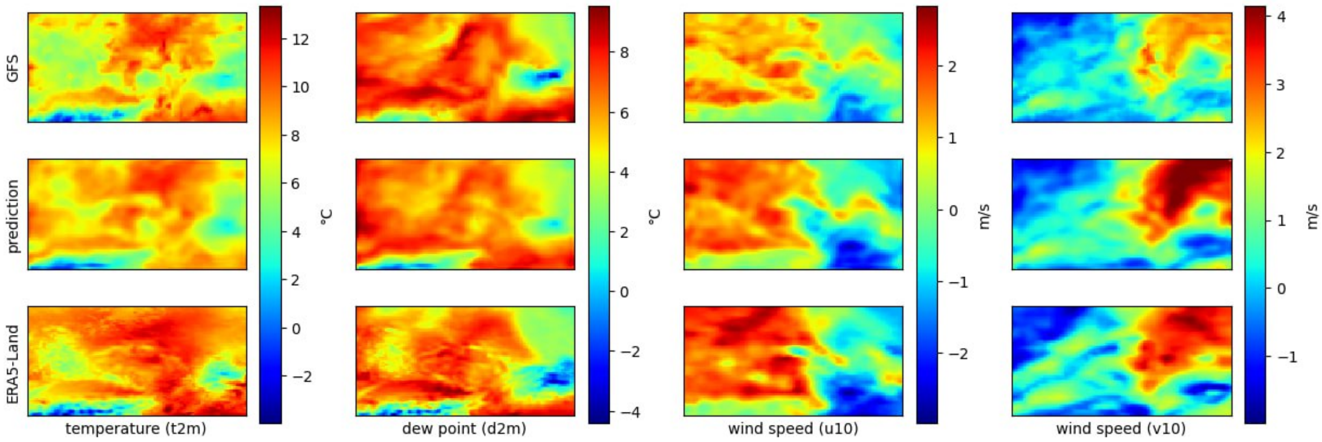

Fig. 3. Example of ERA5-Land data estimated by U-Net superresolution model from GFS predictions. The top row shows the GFS forecast, the middle row is the output of our U-Net model and the bottom row is ground truth ERA5-Land data.

TABLE III
MAE OF GFS FORECASTS AND U-NET MODEL ESTIMATING ERA5-LAND

Variable	GFS to ERA5-Land	Estimated ERA5-Land to ERA5-Land
Temperature [$°C$]	3.03	1.10
Dew Point [$°C$]	2.77	1.09
Wind Speed E-W [m/s]	1.47	0.46
Wind Speed N-S [m/s]	1.51	0.47

Wind Speed. Which represent around 50% improvement for Temperature, Dew Point, and Wind Speed respectively over GFS predictions.

The models incorporating the estimated ERA5-Land data set achieve slightly better results. Similar to no ERA5-Land data set, the CatBoost model achieves the lowest error. The mean absolute error for the Temperature is $1.06°C$ for the 24-hour forecast. The Dew Point forecasts show a mean absolute error of $1.02°C$ and the Wind Speed forecasts show a 1.05 m/s error.

Comparing the three selected ML-based models, the best accuracy is achieved by the CatBoost model. The difference in accuracy between MLP and LSTM is not so significant.

Figure 3 illustrates the results of the U-Net architecture when estimating the ERA5-Land weather reanalysis from GFS forecasts. We illustrate the results using the temperature, the dew point, and wind speed. The wind speed is shown as south-north and east-west components of the speed vector. The top row shows the original GFS forecasts, the bottom row shows the ERA5-Land targets and the middle row represents the U-Net estimate. Table III shows the mean absolute error between our estimated ERA5-Land values and the ground truth values. For comparison, the table also shows the mean absolute error between GFS and ground truth ERA5-Land values. The numbers show about 60% to 70% improvement in the estimation of the ERA5-Land value. This improvement supports the results presented earlier when the introduction of estimated ERA5-Land data improved the accuracy of the forecast.

VII. CONCLUSION

The experimental results confirm the viability and effectiveness of the proposed methodology in generating highly accurate localized forecasts. The AI-driven 24-hour predictions, which integrate GFS data with local measurements, demonstrate markedly superior accuracy compared to GFS alone. This enhanced precision empowers farmers to refine their planning processes, potentially leading to improved crop yields, more efficient fertilizer application, and strengthened food security.

Our findings indicate that the inclusion of estimated ERA5-Land data does not contribute significantly to model accuracy improvements.

Moving forward, research efforts will concentrate on minimizing the volume of historical data required from weather stations. These refined techniques will be implemented across approximately 200 weather stations in Czech and Slovakian vineyards[1].

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] R. J. H. Dunn, K. M. Willett, D. E. Parker, and L. Mitchell, "Expanding hadisd: quality-controlled, sub-daily station data from 1931," *Geoscientific Instrumentation, Methods and Data Systems*, vol. 5, no. 2, pp. 473–491, 2016. [Online]. Available: https://gi.copernicus.org/articles/5/473/2016/

[2] D. R. J. H., "Hadisd version 3: monthly updates," Hadley Centre Technical Note, Tech. Rep., 2019.

[3] S. Haykin, *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.

[4] A. V. Dorogush, V. Ershov, and A. Gulin, "Catboost: gradient boosting with categorical features support," 2018.

[5] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation."

[7] D. Salisbury, K. Mogensen, and G. Balsamo, *Use of in situ observations to verify the diurnal cycle of sea surface temperature in ECMWF coupled model forecasts*. European Centre for Medium Range Weather Forecasts, 2018.

[8] G. Balsamo, A. Agusti-Parareda, C. Albergel, G. Arduini, A. Beljaars, J. Bidlot, E. Blyth, N. Bousserez, S. Boussetta, A. Brown *et al.*, "Satellite and in situ observations for advancing global earth surface modelling: A review," *Remote Sensing*, vol. 10, no. 12, p. 2038, 2018.

[9] M. He, J. Qin, N. Lu, and L. Yao, "Assessment of era5 near-surface air temperatures over global oceans by combining modis sea surface temperature products and in-situ observations," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023.

[10] A. Callens, D. Morichon, S. Abadie, M. Delpey, and B. Liquet, "Using random forest and gradient boosting trees to improve wave forecast at a specific location," *Applied Ocean Research*, vol. 104, p. 102339, 2020.

[1] https://ekovin.cz/

[11] D. Kreuzer, M. Munz, and S. Schlüter, "Short-term temperature forecasts using a convolutional neural network — an application to different weather stations in germany," *Machine Learning with Applications*, vol. 2, p. 100007, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666827020300074

[12] P. Hewage, A. Behera, M. Trovati *et al.*, "Temporal convolutional neural (tcn) network for an effective weather forecasting using time-series data from the local weather station," *Soft Computing*, vol. 24, no. 24, pp. 16 453–16 482, 2020. [Online]. Available: https://doi.org/10.1007/s00500-020-04954-0

[13] S. C. Maina, D. Mwigereri, J. Weyn, L. Mackey, and M. Ochieng, "Evaluation of dependency structure for multivariate weather predictors using copulas," *ACM J. Comput. Sustain. Soc.*, vol. 1, no. 2, dec 2023. [Online]. Available: https://doi.org/10.1145/3616384

[14] K. Bi, L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian, "Accurate medium-range global weather forecasting with 3d neural networks," *Nature*, vol. 619, no. 7970, pp. 533–538, 2023.

[15] G. J. Anderson and D. D. Lucas, "Machine learning predictions of a multiresolution climate model ensemble," *Geophysical Research Letters*, vol. 45, no. 9, pp. 4273–4280, 2018. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL077049

[16] M. Cucchi, G. P. Weedon, A. Amici, N. Bellouin, S. Lange, H. Müller Schmied, H. Hersbach, and C. Buontempo, "Wfde5: bias-adjusted era5 reanalysis data for impact studies," *Earth System Science Data*, vol. 12, no. 3, pp. 2097–2120, 2020. [Online]. Available: https://essd.copernicus.org/articles/12/2097/2020/

[17] Y. Sha, D. J. Gagne II, G. West, and R. Stull, "Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. part ii: Daily precipitation," *Journal of Applied Meteorology and Climatology*, vol. 59, no. 12, pp. 2075–2092, 2020.

[18] A. Bramantoro, W. S. Suhaili, and N. Z. Siau, "Precision agriculture through weather forecasting," in *2022 International Conference on Digital Transformation and Intelligence (ICDI)*. IEEE, 2022, pp. 203–208.

[19] C. El Hachimi, S. Belaqziz, S. Khabba, and A. Chehbouni, "Towards precision agriculture in morocco: A machine learning approach for recommending crops and forecasting weather," in *2021 international conference on digital age & technological advances for sustainable development (ICDATA)*. IEEE, 2021, pp. 88–95.

[20] National Centers for Environmental Prediction, National Weather Service, NOAA, U.S. Department of Commerce, "Ncep gfs 0.25 degree global forecast grids historical archive," Boulder CO, 2015.

[21] J. Muñoz-Sabater, E. Dutra, A. Agustí-Panareda, C. Albergel, G. Arduini, G. Balsamo, S. Boussetta, M. Choulga, S. Harrigan, H. Hersbach *et al.*, "Era5-land: A state-of-the-art global reanalysis dataset for land applications," *Earth system science data*, vol. 13, no. 9, pp. 4349–4383, 2021.

[22] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.

[23] M.-S. J., "Era5-land hourly data from 1950 to present," Copernicus Climate Change Service (C3S) Climate Data Store (CDS), Tech. Rep., 2019.

[24] S. M. Lundberg and S. Lee, "A unified approach to interpreting model predictions," *CoRR*, vol. abs/1705.07874, 2017. [Online]. Available: http://arxiv.org/abs/1705.07874