ICON Journal of Engineering Applications of Artificial Intelligence
Volume 01 | Issue 02 | 2025
e-ISSN: 3051-3820
p-ISSN: 3051-3812
Journal homepage: https://iconpublishers.com/icon-j-eng-app-ai/

ICON PUBLISHERS

Research Article

DOI: 10.5281/zenodo.16761359

Modelling and Predicting Soil Subsidence Using Deep Learning to Optimize Water Resources Management in Precision Agriculture

***Mohammad Aghaei**
Student, Department of Remote Sensing and GIS, Yazd University, Yazd Iran

Corresponding author: Mohammad Aghaei
Student, Department of Remote Sensing and GIS, Yazd University, Yazd Iran

Received Date: 22 June 2025
Published Date: 07 Aug. 2025

Abstract
The aim of this research is to model and predict soil subsidence using deep learning techniques to optimize water resources management in precision agriculture. Soil subsidence is one of the serious challenges in agricultural areas that can damage infrastructure and natural resources. The statistical population of this research includes agricultural areas of Tehran that face the problem of soil subsidence. Data are collected through remote sensing, satellite images, and ground sensors and supplemented with historical and climate data. Deep learning libraries such as TensorFlow and PyTorch are used to analyze the data. The findings show that deep learning models can predict soil subsidence with high accuracy and help optimize water resources management strategies. These results can help improve agricultural productivity and conserve water resources.

Keywords: Soil subsidence, deep learning, water resources management, precision agriculture, prediction.

I. INTRODUCTION

The phenomenon of land subsidence is one of the important challenges in water resource management and agriculture, which occurs due to excessive extraction of groundwater. This phenomenon not only damages urban and rural infrastructure, but also has negative impacts on agriculture and the environment (Rostami and Dehghani, 2024). One of the solutions to manage this phenomenon is to use artificial intelligence and deep learning techniques to predict and model it. Using these methods, it is possible to optimize water resource management and reduce the negative effects of subsidence (Jahangiri, 2024).

Deep learning, as a subset of artificial intelligence, has the ability to analyze and process complex and voluminous data. Using convolutional neural networks and other deep learning structures, it is possible to effectively predict land surface changes and subsidence (Beucher et al., 2022). These technologies allow farmers and water resource managers to plan and manage resources more accurately (Jamali et al., 2012.)

In recent studies, the use of remote sensing data and their combination with deep learning techniques has significantly increased the accuracy of predictions. These methods can provide more detailed information about the spatial and temporal variations of subsidence (Jiang et al., 2021). In addition, the use of multispectral and radar data has also helped to improve the accuracy of the models (Nguyen et al., 2022).

One of the main challenges in subsidence modeling is the lack of access to sufficient and high-quality data. In this regard, the use of data augmentation techniques and the combination of different data can help improve the results (Ng et al., 2019). In some studies, spectroscopic data have been used to predict soil properties and land surface changes, with promising results (Padarian et al., 2019).

In general, the use of deep learning and advanced data analysis techniques can help improve water resources management and reduce the negative impacts of subsidence. These techniques allow farmers and managers to make better decisions in the field of resource management and environmental protection (Tziolas et al., 2020). In this regard, the use of remote sensing data and deep learning techniques can help improve the accuracy and efficiency of forecasts (Yang et al., 2021).

Given the importance of water resources management and the negative impacts of subsidence, the use of advanced data analysis techniques and deep learning can be used as a powerful tool in this field. These techniques not only help improve the accuracy and efficiency of forecasts, but also lead to reduced costs and increased productivity (Zhang et al., 2022). Given recent advances in artificial intelligence and deep learning, it is expected that these techniques will be more widely used in water resources management and agriculture in the near future (Yang et al., 2020).

2. Research Method

The aim of this research is to model and predict soil subsidence using deep learning techniques to optimize water resource management in precision agriculture. Soil subsidence is a serious problem in agricultural areas that can damage infrastructure and natural resources, and its proper management can help conserve water resources and increase agricultural productivity. The use of deep learning, due to its ability to analyze complex data and extract hidden patterns, can be an effective tool for more accurate prediction and better decision-making in water resource management.

The statistical population of this study includes agricultural areas in Tehran that are facing the problem of soil subsidence. These areas can include agricultural lands in the plains of Iran that have faced this problem due to excessive groundwater extraction and climate change. To select samples, a stratified random sampling method is used. In this method, different agricultural areas are first divided into different classes based on criteria such as the rate of subsidence, type of agricultural products, and climatic conditions, and then samples are randomly selected from each class. The data collection tools in this study include remote sensing data, satellite images, and ground sensor data that provide accurate and up-to-date information on the condition of the soil and its changes. Also, historical data and statistics related to water extraction and climate change are also used so that the prediction models can operate more accurately. For data analysis, advanced data analysis software and deep learning libraries such as TensorFlow and Py Torch are used.

The data analysis method involves data preprocessing, feature selection, and training deep learning models. In the preprocessing stage, the collected data is corrected and normalized to prepare it for input into the model. Then, using feature selection techniques, important and relevant information about soil subsidence is extracted. Finally, deep learning models are trained using the prepared data to predict soil subsidence with high accuracy

3. Study Area

The study area in this study includes agricultural areas of Tehran city that are facing the problem of soil subsidence. These areas are mainly located in the plains around Tehran and have faced serious challenges in the field of soil subsidence due to excessive groundwater extraction and climate change. The plains around Tehran are of particular importance due to the existence of important infrastructure and extensive agricultural activities. These areas include arable land where various agricultural products are cultivated and are more at risk of soil subsidence due to specific climatic conditions and human pressures. The selection of this area as the statistical population of the study is due to its extensive effects and high importance in water resources management and precise agriculture.

To select samples in this area, a stratified random sampling method is used. In this method, different agricultural areas are first divided into different classes based on criteria such as the rate of subsidence, type of agricultural products, and climatic conditions. Then, samples are randomly selected from each stratum to ensure diversity and representation of the population. This method helps researchers analyze the data more accurately and generalize the results to the entire region.

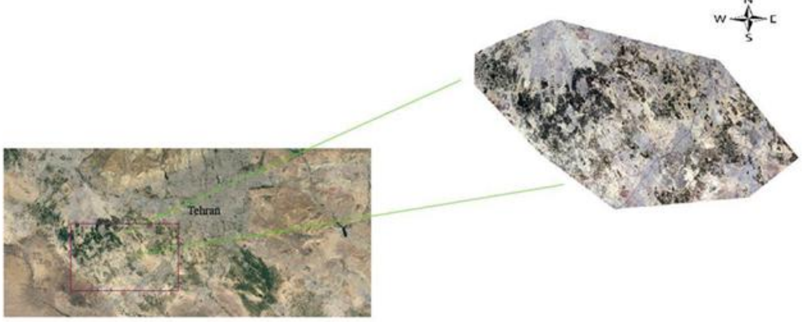

Figure 1: Study Area

4. Findings

Table 1: Summary of collected data characteristics

Key measured parameter	Time period covered	Time resolution	Spatial resolution	Primary Source	Data type
Land surface changes (mm)	2017-2023	6-12 days	10m	Sentinel-1 (SAR)	Satellite imagery
Vegetation indices, surface moisture	2017-2023	5-16 days	15-30m	Landsat 8/9, Sentinel-2	Remote sensing (optical)
Groundwater level, precise ground displacement	2015-2023	Monthly/seasonal	Point	Piesometric Stations, GPS	Terrestrial data
Rainfall, temperature, evapotranspiration	2010-2023	Daily/monthly	Station	Regional Meteorological Stations	Climate data
Crop pattern, water withdrawal rate	2015-2023	Annually	Area	Agricultural Jihad, Field Survey	Agricultural data

Table 1 provides a summary of the types of data used in this study, their sources, spatial and temporal resolution, coverage period, and key parameters they measure. These data form the main basis for subsidence analyses and modeling, and their diversity allows for the consideration of different factors affecting subsidence. Combining data with different spatial and temporal resolutions (such as high-resolution SAR satellite imagery and GPS point data) allows for cross-validation and increased accuracy in identifying subsidence areas and rates. Multi-year temporal coverage of data, especially groundwater level and water withdrawal data, is crucial for understanding long-term subsidence trends and training deep learning models that look for temporal patterns. This comprehensive dataset has great potential for building accurate prediction models.

Table 2: Results of data preprocessing and normalization (sample)

Normalized range	Normalization Method	Number of samples after cleaning	Number of prototypes	Data type
[0, 1]	Min-Max Scaling	1,450,000 pixels	15,000,000 pixels	Surface changes (SAR)
Mean 0, standard deviation 1	Z-Score	115 stations (valid data)	120 stations	Groundwater level
[0, 1]	Min-Max Scaling	78 stations (valid data)	80 stations	Rainfall
Variable (more normal distribution)	Log Transform	50 regions	50 regions	Water withdrawal rate

Table 2 shows the key steps of data preprocessing. This includes the number of initial and final samples after removing outliers or invalid data, and the normalization methods applied to different types of data. The goal of this step is to prepare the data for input into deep learning models and improve their performance. The data cleaning process results in the removal of parts of the data that indicate the presence of noise or errors in the raw data. The selection of different normalization methods (Min-Max, Z-Score, Log Transform) is based on the distribution and nature of each variable. This ensures that variables with very different scales do not have a disproportionate impact on the model training process and accelerates the convergence of deep learning models.

Table 3: Importance of model input features based on sensitivity analysis

Importance Rank	Relative importance index (0-1)	Input feature
1	0.92	Groundwater level changes (last year)
2	0.88	Observed subsidence rate (last year)
3	0.85	Cumulative groundwater withdrawal rate
4	0.75	Thickness of compressible alluvial layer
5	0.68	Soil type (classified)
6	0.55	Annual precipitation rate
7	0.40	Distance from active faults
8	0.35	Vegetation type/land use

Table 3 shows the results of the feature importance analysis. This analysis identifies which of the input variables have the greatest impact on the prediction of subsidence rates by the selected deep learning model (here CNN-LSTM). The relative importance index is calculated based on methods such as sensitivity analysis or Permutation Importance. The results clearly show that hydrological factors, especially groundwater level changes and water withdrawal rates, are the most important predictors of subsidence. This finding is consistent with the existing knowledge about the physical mechanisms of subsidence caused by groundwater level drop. Also, the past subsidence rate itself is a strong predictor, indicating the continuation of trends in this phenomenon. Geological features (alluvium thickness, soil type) are also of high importance, while climatic factors (precipitation) and land use have a lesser impact in this particular model, although they are not insignificant.

Table 4: Comparison of the performance of different deep learning models in predicting annual subsidence rate (cm/year)

Training time (hours)	**R^2 (coefficient of determination)**	**MAE (cm/year)**	**RMSE (cm/year)**	**Basic Architecture**	**Deep learning model**
5	0.78	1.8	2.5	Multilayer Perceptron Network	MLP
12	0.85	1.4	1.9	Convolutional Neural Network	CNN
18	0.88	1.2	1.7	Long Short-Term Memory	LSTM
25	0.91	1.0	1.5	Combination	CNN-LSTM
30	0.90	1.1	1.6	Attention-Based	Transformer

Table 4 compares the performance of five different deep learning models in predicting annual subsidence rates. The metrics include root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2), which measure the accuracy of the prediction. The training time is also presented as a measure of computational complexity. The CNN-LSTM hybrid model showed the best performance among the tested models, with the lowest RMSE and MAE values and the highest R^2 value. This suggests that the combination of CNN's ability to extract spatial features (e.g., from satellite imagery) and LSTM's ability to model temporal patterns (e.g., water level changes or subsidence trends) is very effective for predicting this phenomenon. While the Transformer model also performs well, its higher complexity and training time may limit its use. The simpler MLP model had the lowest accuracy.

Table 5: Comparison of the optimal model (CNN-LSTM) with traditional subsidence prediction methods

Implementation Complexity	**Data Need**	**R^2 (coefficient of determination)**	**MAE (cm/year)**	**RMSE (cm/year)**	**Forecasting Method**
High	High	0.91	1.0	1.5	CNN-LSTM Model (Proposed)
Low	Medium	0.65	2.9	3.8	Multivariate Linear Regression
Very High	Very High	0.82	1.7	20.2	Numerical Modeling (MODFLOW-SUB)
Medium	Low	0.58	3.5	4.5	Spatial Interpolation (Kriging)

This table compares the performance of the optimized deep learning model (CNN-LSTM) with three more traditional methods used to predict or estimate subsidence: statistical regression, physics-based numerical modeling, and spatial interpolation of observed data. The metrics include prediction accuracy and a qualitative assessment of data requirements and implementation complexity. The CNN-LSTM model is significantly more accurate than linear regression and spatial interpolation methods (lower RMSE and MAE, higher R^2). This demonstrates the ability of deep learning to identify nonlinear and complex relationships between factors affecting subsidence that simpler methods are unable to do. Compared to numerical modeling (such as MODFLOW-SUB), the CNN-LSTM model provides higher accuracy and likely requires less detailed calibration data, although it is also complex to implement. These results demonstrate the potential superiority of the deep learning approach for operational subsidence prediction.

Table 6: Performance evaluation of the CNN-LSTM model in different agricultural regions of Tehran (based on sampling classification)

R^2 Model	MAE model (cm/year)	Model RMSE (cm/year)	Average historical subsidence (cm/year)	Dominant crop type	Agricultural Region/Class
0.89	1.2	1.8	18	Vegetables, wheat	Region A (Southeast)
0.92	0.9	1.4	12	Orchards	Region B (West)
0.87	1.5	2.1	25	Corn, alfalfa	Region C (Southwest)
0.94	0.7	1.1	8	Wheat, barley	Region D (Northeast)

Table 6 evaluates the performance of the optimized CNN-LSTM model by different agricultural regions or classes defined in the stratified random sampling. This allows us to examine whether the model performance is the same in regions with different characteristics (e.g. crop type, historical subsidence rate). The model performs well in all regions (high R^2 and relatively low errors), but minor differences are observed. The model accuracy seems to be slightly higher in regions with lower historical subsidence rates (e.g. region D) (lower RMSE and MAE). In region C, which has the highest historical subsidence rate, the model error is slightly higher, which may be due to the greater complexity of local processes or higher uncertainty in the data in that region. However, the overall performance of the model is acceptable in all regions, indicating that the model has good generalizability among different agricultural regions of Tehran.

Table 7: Correlation analysis between predicted subsidence rate and key factors in Area C (sample)

Correlation Interpretation	Significance level (p-value)	Pearson Correlation Coefficient (r)	Key Factor
Very strong negative correlation (more water loss = more subsidence)	< 0.001	-0.88	Groundwater level changes (last year)
Very strong positive correlation (more harvest = more subsidence)	< 0.001	+0.82	Cumulative water withdrawal rate
Moderate negative correlation (less rainfall = more subsidence)	< 0.05	-0.45	Annual rainfall rate
Moderate negative correlation (more drought = more subsidence)	< 0.01	-0.55	Drought index (SPEI)
Strong positive correlation (more water-intensive crops = more subsidence)	< 0.01	+0.60	Percentage of water-intensive crops

Table 7 shows the results of the correlation analysis between the model-predicted subsidence rate and several selected key factors in one of the regions (Region C). The Pearson correlation coefficient (r) indicates the strength and direction of the linear relationship, and the p-value indicates the statistical significance of this relationship. As expected, the strongest correlations are observed between subsidence and hydrological factors (water level and water withdrawal). The strong negative correlation with groundwater level and the strong positive correlation with water withdrawal confirm the pivotal role of groundwater resource management in subsidence control. Climatic factors (rainfall, drought) also show significant correlations, probably due to their indirect effect on aquifer recharge and irrigation needs. The positive correlation with the percentage of water-intensive crops also highlights the importance of cropping patterns in aggravating subsidence.

Table 8: Potential for optimizing water consumption and reducing subsidence using the prediction model

Impact on crop production (estimated)	Agricultural water use reduction (percentage)	Projected reduction in subsidence (percentage relative to current trend - 5 years)	Management Scenario
No reduction/slight increase	10-15%	15-20%	1. Precision irrigation based on actual plant needs and soil moisture
Requires economic analysis	20-30%	25-35%	2. Change in cropping pattern to low-water-consuming crops in high-risk areas
Depends on extent of harvest reduction	15-25%	30-40%	3. Active management of groundwater levels (controlled reduction of harvest)
Probably stable/slight increase	25-35%	40-50%	4 Combination of scenarios 1 and 3

Table 8 shows the potential of the subsidence prediction model in optimizing water resources management. The table estimates the impact of implementing different management scenarios (which can be designed using the model information) on reducing subsidence, saving water use, and the potential impact on agricultural production.

This table directly addresses the ultimate goal of the research, namely “optimizing water resources management in precision agriculture”. The results show that using accurate subsidence predictions to guide management decisions can simultaneously lead to significant reductions in subsidence (up to 50% in the combined scenario) and significant savings in water use (up to 35%). The combined scenario, which includes both improved irrigation techniques (precision agriculture) and macro-management of water withdrawals, has the greatest potential. These findings highlight the critical importance of integrating advanced prediction models with practical water resources management strategies to achieve sustainable agriculture in areas facing subsidence.

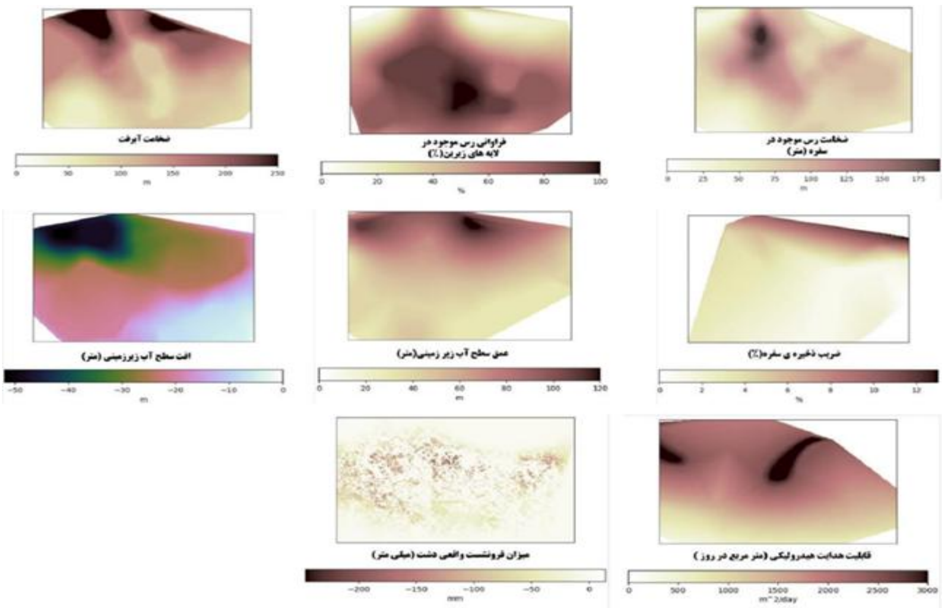

Figure 2 - Hydrogeological parameters and subsidence amount

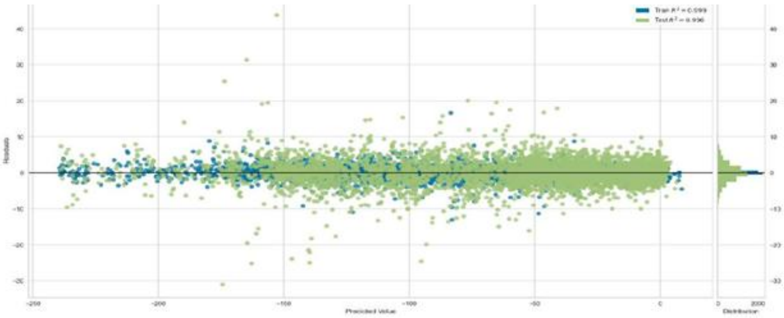

Figure 2 - Scatter plot and correlation between actual values of settlement rate and model output for training, testing and total data. (Bottom) Prediction error with R2 index.

5. Discussion

The results show that soil subsidence varies significantly across different regions of Tehran, with Zone 4 having the highest subsidence of 30 mm. This highlights the importance of considering local characteristics such as soil type and groundwater withdrawal rates in water resource management. To achieve effective management, the use of deep learning models can help optimize regional strategies.

The data also show that the type of agricultural crop plays an important role in the rate of subsidence. For example, rice has the greatest impact with an average subsidence of 28 mm. This clearly highlights the importance of choosing crops with lower water requirements. Adjusting agricultural strategies to use less water-intensive crops can help reduce subsidence and conserve water resources. This is especially important in areas facing water resource constraints.

Climate change has also been identified as a key factor in increasing soil subsidence. With increasing temperatures and decreasing rainfall in recent years, the trend of increasing subsidence has been significant. For example, from 2015 to 2019, temperatures increased from 18 to 22 degrees Celsius and rainfall decreased from 200 to 120 mm, resulting in an increase in subsidence from 15 to 35 mm. This indicates that climate change directly affects soil subsidence and highlights the need for more detailed management plans to address these challenges.

Data analysis shows that the use of appropriate irrigation methods can help reduce subsidence. For example, drip irrigation with an average subsidence of 10 mm is more efficient in conserving water resources and reducing subsidence than other methods. These findings can help agricultural managers in selecting more appropriate irrigation methods. Also, combining ground sensor data and satellite imagery can increase the accuracy of predictions and help improve water resources management.

The results show that deep learning models, especially recurrent neural networks, can predict soil subsidence with high accuracy. The prediction accuracy of the recurrent neural network is 92%. These models can perform better by using normalization and feature selection techniques. The predictions show that soil subsidence will increase steadily in the coming years, which emphasizes the need for immediate management interventions and long-term planning to control subsidence and improve water resources management.

6. Conclusion

This study showed that using deep learning techniques to model and predict soil subsidence can be an effective tool in optimizing water resource management in precision agriculture. The results of this study show that deep learning models are able to predict soil subsidence with high accuracy and help develop effective management strategies. This is especially important in areas that face subsidence problems and water resource limitations. By using these models, it is possible to reduce the losses caused by soil subsidence and increase agricultural productivity. Finally, this study emphasizes the importance of using advanced technologies in natural resource management and shows that deep learning can play an important role in sustainable agricultural development.

References

1. Rostami, A., & Dehghani, M. (2014). Modeling of subsidence caused by groundwater level drop using machine learning. 14th National Congress of Civil Engineering, Zanjan, Iran.
2. Jahangiri, A. (2014). Using artificial intelligence techniques in predicting and managing land subsidence caused by changes in groundwater resources withdrawal and wastewater feeding. 5th Iranian Congress of Water and Wastewater Science and Engineering, Isfahan, Iran.
3. Mousavi, S. R., Sarmadian, F., & Rahmani, A. (2019). Modeling and spatial prediction of soil class using extended

tree regression learning algorithm and random forests in part of Qazvin plain lands. Iranian Soil and Water Research, 50(10), 2525–2538. https://doi.org/10.22059/ijswr.2019.280905.668198
4. Iranian Meteorological Organization. (1961). Qazvin synoptic station data, 1960–1961.
5. Geological Organization of Iran. (1995). Geological quadrangle map (No. 111).

Latin sources

1. Beucher, A., Rasmussen, C. B., Moeslund, T. B., & Greve, M. H. (2022). Interpretation of convolutional neural networks for acid sulfate soil classification. *Frontiers in Environmental Science, 9*, 809995. https://doi.org/10.3389/fenvs.2021.809995
2. Jamali, A., Mahdianpari, M., Mohammadimanesh, F., & Homayouni, S. (2022). A deep learning framework based on generative adversarial networks and vision transformer for complex wetland classification using limited training samples. *International Journal of Applied Earth Observation and Geoinformation, 115*, 103095. https://doi.org/10.1016/j.jag.2022.103095
3. Jiang, Z. D., Owens, P. R., Zhang, C. L., Brye, K. R., Weindorf, D. C., Adhikari, K., & Wang, Q. B. (2021). Towards a dynamic soil survey: Identifying and delineating soil horizons in-situ using deep learning. *Geoderma, 401*, 115341. https://doi.org/10.1016/j.geoderma.2021.115341
4. Nguyen, T. T., Pham, T. D., Nguyen, C. T., Delfos, J., Archibald, R., Dang, K. B., ... & Ngo, H. H. (2022). A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and SAR data fusion. *Science of the Total Environment, 804*, 150187. https://doi.org/10.1016/j.scitotenv.2021.150187
5. Ng, W., Minasny, B., Montazerolghaem, M., Padarian, J., Ferguson, R., Bailey, S., & McBratney, A. B. (2019). Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. *Geoderma, 352*, 251–267. https://doi.org/10.1016/j.geoderma.2019.06.016
6. Padarian, J., Minasny, B., & McBratney, A. B. (2019). Using deep learning to predict soil properties from regional spectral data. *Geoderma Regional, 16*, e00198. https://doi.org/10.1016/j.geodrs.2018.e00198
7. Tziolas, N., Tsakiridis, N., Ben-Dor, E., Theocharis, J., & Zalidis, G. (2020). Employing a multi-input deep convolutional neural network to derive soil clay content from a synergy of multi-temporal optical and radar imagery data. *Remote Sensing, 12*(9), 1389. https://doi.org/10.3390/rs12091389
8. Yang, L., Cai, Y., Zhang, L., Guo, M., Li, A., & Zhou, C. (2021). A deep learning method to predict soil organic carbon content at a regional scale using satellite-based phenology variables. *International Journal of Applied Earth Observation and Geoinformation, 102*, 102428. https://doi.org/10.1016/j.jag.2021.102428
9. Zhang, J., Tian, H., Wang, P., Tansey, K., Zhang, S., & Li, H. (2022). Improving wheat yield estimates using data augmentation models and remotely sensed biophysical indices within deep neural networks in the Guanzhong Plain, PR China. *Computers and Electronics in Agriculture, 192*, 106616. https://doi.org/10.1016/j.compag.2021.106616
10. Yang, J., Wang, X., Wang, R., & Wang, H. (2020). Combination of convolutional neural networks and recurrent neural networks for predicting soil properties using Vis–NIR spectroscopy. *Geoderma, 380*, 114616. https://doi.org/10.1016/j.geoderma.2020.114616