Ref. Ares(2025)9982446 - 18/11/2025

REFOREST

Organisation: Czech University of Life Sciences Prague

D4.2

Predictive tool for carbon and biodiversity assessment

Date 10.11.2025

Doc. Version 05

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101060635 (REFOREST). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

UK Research and Innovation

UK partners are funded by UK Research and Innovation (UKRI) under UK government's Horizon Europe funding guarantee [grant number 10039700].

Document Control Information

Settings	Value
Deliverable Title	Predictive tool for carbon and biodiversity assessment
Work Package Title	Monitoring and verification
Deliverable number	4.2
Description	Layperson description of the development, application and deployment of an online calculator predicting carbon and biodiversity status as a function of geography
Lead Beneficiary	CZU
Lead Authors	CZU – Martin Lukáč
Contributors	ORC – Colin Tosh, CZU – Martin Mokroš, Arunima Singh, Kottilapurath Surendran Gokul
Submitted by	Martin Lukáč
Doc. Version (Revision number)	05
Sensitivity (Security):	Low
Date:	10/11/2025

Document Approver(s) and Reviewer(s):

NOTE: All Approvers are required. Records of each approver must be maintained. All Reviewers in the list are considered required unless explicitly listed as Optional.

Name	Role	Action	Date
Colin Tosh	ORC – WP4 deputy lead	Approved	12/11/2025

Document history:

The Document Author is authorised to make the following types of changes to the document without requiring that the document be re-approved:

- Editorial, formatting, and spelling
- Clarification

To request a change to this document, contact the Document Author or Owner.
Changes to this document are summarised in the following table in reverse chronological order (latest version first).

Revision	Date	Created by	Short Description of Changes
05	10/11/2025	CZU – M. Lukáč	Revised Deliverable to address Review Board comments
04	21/10/2024	CZU – M. Lukáč	Section 4 updated
03	17/10/2024	CZU – M. Mokroš, K. S. Gokul	Section 2 & 3 updated
02	16/10/2024	ORC – C. Tosh	Section 2 – content update
01	29/09/2024	CZU – M. Lukáč	Initial version

Configuration Management: Document Location

The latest version of this controlled document is stored in https://czuvpraze.sharepoint.com/teams/fld-t-reforest/Sdilene%20dokumenty/Forms/AllItems.aspx

Nature of the deliverable		
R	Report	
DEC	Websites, patents, filing, etc.	
DEM	Demonstrator	x
O	Other	

Dissemination level		
PU	Public	x
CO	Confidential, only for members of the consortium (including the Commission Services)	

ACKNOWLEDGEMENT

This report forms part of the deliverables from the REFOREST project which has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101060635. The Community is not responsible for any use that might be made of the content of this publication.

More information on the project can be found at: http://agroreforest.eu/

COPYRIGHT

DISCLAIMER

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This deliverable presents the development of an online predictive tool that estimates soil carbon stocks and biodiversity levels across European landscapes using satellite imagery, ground observations, and machine learning. The tool builds upon the neural network datasets and workflows developed in Deliverable 4.1, transforming them into a functional and accessible platform for stakeholders, including farmers, landowners, researchers, and policymakers.

At its core, the predictive tool employs a Convolutional Neural Network (CNN) trained on extensive datasets that combine Sentinel-2 satellite imagery, the LUCAS 2018 soil dataset, and biodiversity observations from the Global Biodiversity Information Facility (GBIF). The model learns complex spatial and spectral patterns that relate to carbon and biodiversity indicators, enabling accurate predictions even in regions not directly sampled by field surveys.

An interactive online calculator allows users to obtain instant predictions for any geographic location in Europe by simply entering map coordinates. The web interface is designed for ease of use on both computers and mobile devices, ensuring broad accessibility without the need for specialised software or technical expertise.

The tool's accuracy and robustness are enhanced by combining datasets from multiple regions and by integrating advanced training methods that prevent overfitting and improve generalisation. The approach is further strengthened through the incorporation of UAV-based high-resolution imagery, which increases spatial detail and captures fine-scale environmental variation.

Beyond its immediate application, the predictive tool provides a methodological foundation for integrating carbon and biodiversity monitoring into agroforestry, regenerative agriculture, and environmental policy frameworks. It supports evidence-based decision-making for climate mitigation, soil restoration, and biodiversity conservation, contributing directly to the objectives of the European Green Deal and EU Soil Mission.

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

Abbreviation	Definition
CNN	Convolutional neural network
GSD	Ground sampling distance
SOC	Soil organic carbon
UTM	Universal transverse mercator
ECMWF	European center for medium-range weather forecasts
TOA	Top Of Atmosphere
MSI	Multi spectral instrument
UAV	Unmanned aerial vehicles

1. INTRODUCTION

The carbon and biodiversity predictive tool described in this document builds upon the technology and solutions introduced in D4.1, specifically the Neural Network Training Dataset. The predictive tool utilises a neural network model trained on satellite imagery and a verified dataset of ground observations. Data sets from Czechia and the rest of Europe (LUCAS 2018 topsoil data set) have been used to create the soil carbon model. A slightly different approach has been employed for biodiversity, utilising a country-specific approach. GBIF (plant species richness) and satellite data were acquired from Spain, Sweden, the Netherlands, and Switzerland.

A large dataset of approximately 200GB of Sentinel-2 MSI Level-1C imagery for Chechia and the Netherlands has been acquired. A database of approximately 20,000 topsoil SOC (soil organic carbon) 500 m x 500 m multispectral images, centred at the soil carbon sample point, has been acquired from Czechia, and a similar dataset has been acquired from the rest of Europe. Biodiversity has been assessed on a larger scale due to the sparsity of the GBIF data set at a lower scale. Images fed into the neural net represented 5 km x 5 km squares of the landscape. We have proceeded without bias correction as correlations with road and population density only explain very small amounts of data variance. Approximately 15,000 image snippets have been used from Spain, approximately 15,000 from Sweden, approximately 4,000 from the Netherlands, and the 1,000-image dataset from Switzerland was used only for validation. The geographic range of data will eventually expand across Europe to enhance the generalisation of the predictive tool based on satellite imagery being developed.

We aimed to develop an actionable and accessible online predictive tool that leverages cutting-edge satellite imagery technology. This tool is designed for use by various stakeholders, including farmers, landowners, researchers, policymakers, environmental managers, and land use planners, providing critical insights into carbon stock and biodiversity levels based on remote sensing data. Due to the nature of the underlying data and its improving quality and availability, we will continue improving the models to enhance their performance. The user interface is unlikely to change, but the model performing the calculations will continue to improve over time.

To further enhance this tool's predictive accuracy and overall performance, we are expanding its technological framework to incorporate near-field remote sensing data, particularly that captured via Unmanned Aerial Vehicles (UAVs). By integrating UAV-derived data, which offers higher spatial resolution and greater detail compared to satellite imagery, we can significantly improve the granularity of our models. This will enable us to refine the tool's performance by capturing more localised environmental variations and fine-scale landscape features, which are crucial for precise carbon sequestration and biodiversity pattern predictions.
This additional layer of high-resolution data will enable us to optimise the neural network's training process, reducing uncertainties and improving generalisation across various ecosystems and land use types. Ultimately, incorporating UAV-based remote sensing will enable more robust predictions, providing stakeholders with a more reliable and detailed understanding of carbon dynamics and biodiversity at both broad and fine scales, thereby enhancing decision-making for sustainable land management and climate action.

2. PREDICTIVE TOOL DEVELOPMENT

The carbon and biodiversity predictive tool described in this document builds upon the technology and solutions introduced in D4.1, Neural Network Training Dataset.

2.1 Training Dataset Description

Remote Imagery:
Around 200GB of Sentinel-2 MSI remote imagery for the Netherlands and the Czech Republic has been downloaded using Google Earth Engine. The dataset consists of ortho-images covering 100 km² tiles in UTM/WGS84 projection, with radiometric measurements at Top of Atmosphere (TOA) reflectance. The imagery has been resampled at different native resolutions, 10, 20, and 60 meters, across various spectral bands, particularly B1, B2, and B3 in the visible and infrared spectra. Additional processing includes the use of Land/Water masks, Cloud Masks, and ECMWF data, which cover factors such as the total ozone column, water vapour, and mean sea level pressure. This initial dataset, featuring images from the summer of 2022, is sufficient for the current phase of workflow development, CNN training, and optimisation. However, as the project progresses, the imagery database is expected to expand to include other EU countries, supporting the broader generalisation of the predictive tool.

Carbon Data:
To optimise remote image sampling and CNN training for soil carbon predictions, the key dataset comprises approximately 20,000 topsoil SOC (%) field samples collected between 2016 and 2021 across Chechia, as well as a further 20,000 LUCAS images covering the rest of Europe. The Central Institute for Supervising and Testing in Agriculture, CZ samples, provide a robust, high-quality dataset for the project's initial phase. The locations of these samples have been mapped and utilised to anchor the CNN's training process for predicting carbon stocks based on satellite imagery. After optimising the model with the Czechia dataset, training was extended to other European regions. The LUCAS 2018 topsoil dataset, which includes around 20,000 data points, was incorporated to further generalise the model across the EU, ensuring that the predictive tool can accurately estimate carbon stocks across various landscapes. Randomly selected 10% of each dataset was set aside for validation to predict data that the network has not seen. This procedure ensures that overfitting does not occur: if error reduction continues for the training data set, but flattens off for the data set the network has not seen, the training stops as the model is simply extracting features in a particular data set and is no longer generalising.

Biodiversity Data:
For biodiversity predictions, CNN training procedures followed a similar approach to the soil carbon process, using remote imagery to identify landscape features correlated with plant diversity at multiple spatial scales. The initial focus was on the Netherlands, the smallest national dataset in the study, using vascular plant species occurrence data from the GBIF database. Data has been analysed using multiple linear regression to investigate potential sampling bias, especially concerning population density and road occurrence, which are common factors that might skew species richness data. We have combined the Spanish, Swedish, and Dutch datasets, but have retained 10% of each for validation purposes. The Swiss data was not presented to the network and represented the "ultimate validator": data that the network has not seen and is from a different country. While

population density and road occurrence were significant predictors of plant species richness, the model's overall goodness of fit was very low, indicating minimal real-world impact from these factors. Given this poor fit, further CNN training proceeded without significant bias correction. However, additional statistical analyses were conducted to confirm this finding before expanding the geographic range of biodiversity data to other EU countries, ensuring that the model generalises well across diverse regions.

2.2 Neural Network Training

Neural network training for our carbon stock and biodiversity predictive tool is a process that involves feeding the model a comprehensive dataset of satellite imagery and ground-truth field observations, allowing it to learn complex patterns and correlations between remote sensing data and the target variables (carbon and biodiversity levels).

Input Data (Satellite Imagery, as described above)
The primary input to the neural network comes from Sentinel-2 MSI (MultiSpectral Instrument) satellite images. These images are captured at multiple spatial resolutions (10, 20, and 60 meters) and comprise different spectral bands, including visible and infrared. The GEOTIFF images obtained in this step contain 13 bands or layers each, representing a different spectral frequency, mostly in the invisible spectrum. The main image used for analysis is the median of all images obtained in 2022, which facilitates generalisation rather than using, for example, summer images only. For carbon modelling, all bands of the relevant spectral range were used in combination with additional satellite-derived data, including Land/Water masks, Cloud masks, and environmental data from ECMWF (European Centre for Medium-Range Weather Forecasts), which provide information about atmospheric conditions such as ozone concentration and sea level pressure.

CNN Architecture and Training

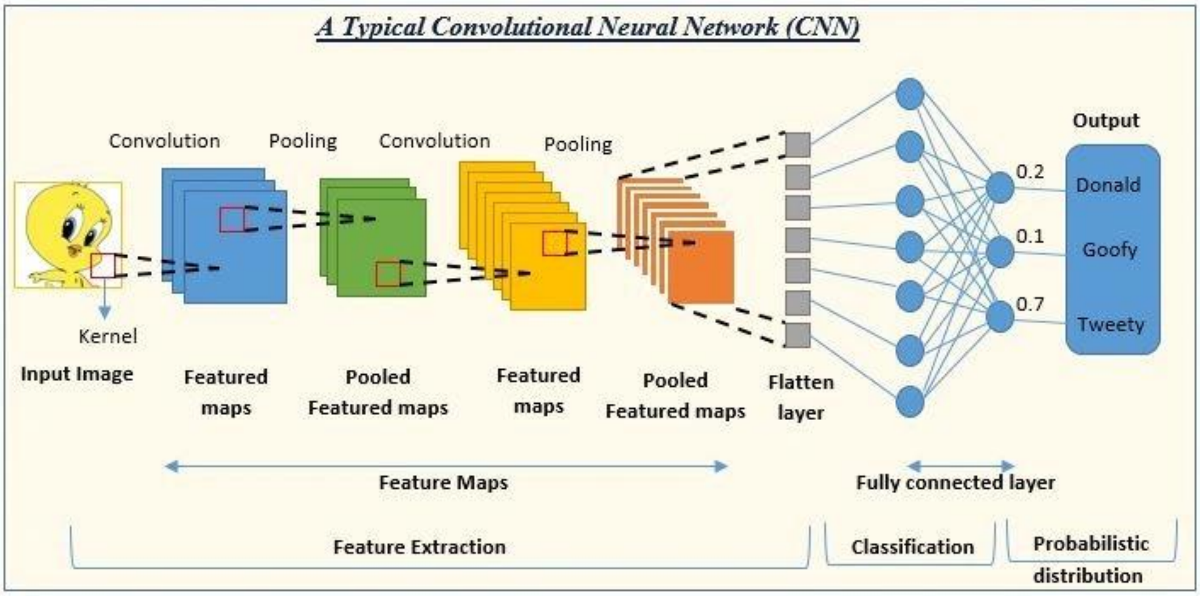


Figure 1: A typical convolutional neural network (Image taken from https://www.analyticsvidhya.com/blog/2022/01/convolutional-neural-network-an-overview/, accessed 16/10/24. Actual network dimensions used in this work are illustrated in the text.)

The foundation of the predictive tool is the Convolutional Neural Network (CNN) architecture, specifically chosen for its ability to process and analyse spatial data, such as satellite imagery, efficiently. CNNs are particularly effective in recognising patterns in grid-like data structures, making them ideal for this application, where pixel-level information from satellite images is critical in predicting environmental variables like carbon stock and biodiversity.

A CNN consists of several layers of interconnected neurons designed to extract hierarchical features from the input data. Our model's primary input is derived from Sentinel-2 MSI satellite imagery, which comprises multiple spectral bands at varying spatial resolutions. These spectral bands capture different properties of the Earth's surface, such as visible light and infrared radiation, enabling the CNN to identify patterns related to vegetation, soil, and other landscape features indicative of soil carbon content or plant biodiversity. The spectral information, such as Bands B1, B2, and B3 in the visible and infrared spectrum, is particularly useful for detecting chlorophyll concentration in plants or surface reflectance in soils, which correlate strongly with the carbon and biodiversity metrics.

In the first layer of the CNN, convolutional filters are applied to the input image data. These filters, or kernels, are small matrices that "slide" across the image, capturing local features such as edges, textures, and spectral gradients. In our case, these features could represent variations in plant canopy structure, soil reflectance, or other land surface characteristics. The network employs multiple filters to extract various features at different levels of abstraction. For instance, early layers might detect basic patterns, like the boundary between vegetation and bare soil. In contrast, deeper layers might extract more complex features, such as the unique spectral signature of certain vegetation types that indicate higher carbon sequestration.

We have experimented with a range of network dimensions, from a simple six-layer network to a state-of-the-art ResNet-50 (50-layer network). A good balance of performance and computational efficiency appears to lie somewhere between these two points. The architecture used for the carbon analysis is as follows (in Python):

```
layers.Conv2D(32, (3, 3), activation='relu', input_shape=(79, 79, 13)),
 layers.MaxPooling2D((2, 2)),
 layers.Conv2D(64, (3, 3), activation='relu'),
 layers.MaxPooling2D((2, 2)),
 layers.Conv2D(128, (3, 3), activation='relu'),
 layers.MaxPooling2D((2, 2)),
 layers.Conv2D(256, (3, 3), activation='relu'),
 layers.Flatten(),
 layers.Dense(128, activation='relu'),
 layers.Dense(64, activation='relu'), # Additional dense layer
 layers.Dense(11, activation='softmax')
```

The code represents the foundation for a net similar to that in Figure 1, a diagram with 11 layers: 8 feature extraction and 3 fully connected layers. The net used for biodiversity was similar but scaled upwards for the larger images. We have also tested a range of optimisation algorithms and parameter values, where the Adam optimiser was set with a learning rate of 0.0001. Networks generally converged to solutions very quickly (within 10 epochs).

After each convolution operation, the output passes through a non-linear activation function, commonly the Rectified Linear Unit (ReLU), which adds non-linearity to the model. This is crucial because environmental data, such as carbon stock or biodiversity, often exhibit complex, non-linear relationships that need to be accurately captured. The CNN architecture also includes pooling layers, which downsample the data by reducing its spatial dimensions, allowing the network to become computationally efficient while retaining the most critical information.

The network builds increasingly abstract representations of the input imagery as the data passes through multiple convolutional and pooling layers. This process allows the CNN to learn basic landscape features and complex interactions between various spectral bands, atmospheric conditions, and land cover types associated with soil carbon and biodiversity.

Once the convolutional layers extracted these hierarchical features, the data was passed to a set of fully connected layers, which interpreted the high-level features learned by the earlier layers. These layers essentially function like a traditional neural network, where each neuron is connected to every neuron in the previous layer. Here, the network combined the features into a final prediction—either carbon stock levels or biodiversity richness at a specific location.

Training the CNN involved a process known as supervised learning, where the model learned to map the input data (satellite imagery) to the target variable (e.g., soil carbon or biodiversity) by minimising the difference between its predictions and the actual values observed in the ground-truth dataset. This was accomplished through an iterative training process, exposing the CNN to thousands of examples from the dataset. The model calculated its prediction error by comparing its outputs with the true values, using a loss function (such as Mean Squared Error for regression tasks). The loss function measures the discrepancy between the predicted values and the actual observations.

The training algorithm then employed a process called backpropagation, where the error was propagated back through the network to update the neurons' weights. This update process was managed by an optimisation algorithm (Stochastic Gradient Descent (SGD)), which adjusted the model's parameters to reduce the prediction error in subsequent iterations. Over time, the CNN learnt the complex relationships between the satellite-derived features and the target environmental variables, gradually improving its prediction accuracy.

One of the critical challenges in CNN training is ensuring that the model can generalise well to unseen data. In the case of carbon stock and biodiversity predictions, it is crucial that the network not only performs well on the training data from Chechia and the Netherlands but also can accurately predict outcomes for other regions in Europe, where environmental conditions may vary. The training process includes techniques such as cross-validation and regularisation to ensure this generalisation. Cross-validation involves partitioning the data into training and validation subsets to monitor the model's performance on unseen data during training, helping to prevent overfitting.

Generalisation and Bias Correction

In the initial training phase, the model was trained on data from Czechia for carbon and the Netherlands for biodiversity. However, additional datasets from across the EU, such as the LUCAS 2018 topsoil dataset, are introduced to ensure that the model generalises well across different landscapes. We used three procedures to ensure networks generalise well to relevant images they have not seen: 1) training nets on a wide range of images from many countries, 2) reserving validation data subsets from data sets to ensure the network is not overfitting and quantify how well they are performing on novel data, and 3) rotating images by a random number degrees between 0 and 360 at each presentation. While early analyses indicated that certain biases, such as population density and road occurrence, could influence species richness, their overall impact was found to be negligible. Therefore, CNN training was executed without major bias corrections, though further analyses will be conducted to confirm this.

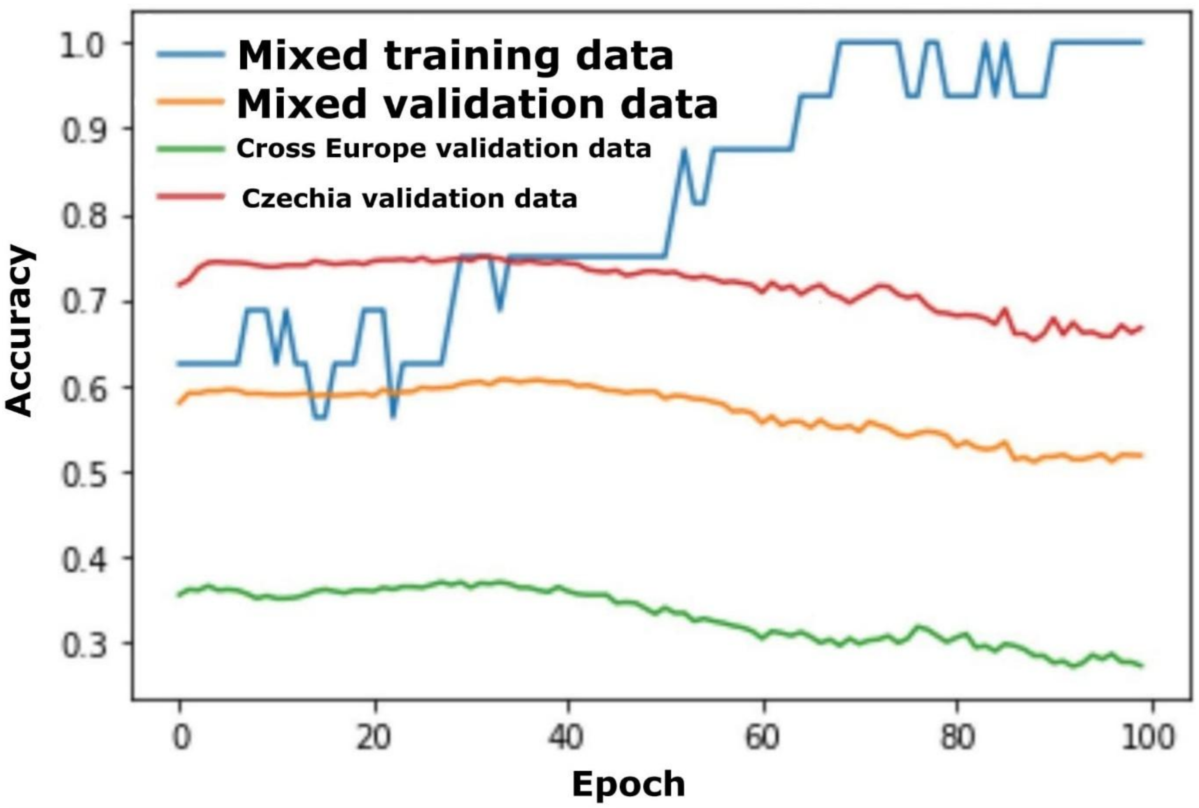


Figure 2: Soil carbon classification accuracy during training of the 11-layer CNN

Mixed training, shown in Figure 2, refers to a mixed training set consisting of images and soil carbon data from Czechia and the rest of Europe. To better explain the data shown in the Figure, the mixed validation data is data reserved from this training set that the network has not seen. The network's performance in classifying this data is evaluated at each epoch. Cross-European validation is the same procedure that uses data only from the LUCAS cross-European data set. Czechia validation is a type of validation that uses only a subset of data from the Czechia data set. The network achieves over 70% classification accuracy on unseen Czech data, predominantly due to the narrow range of environments sampled in this dataset (typically from fields of intensive agriculture). The networks classify correctly only 40% of the time for the cross-European LUCAS data, primarily due to the broad range of environments in this dataset, which spans forests, wetlands, and intensive farming. This figure will be improved considerably with a refinement of training, such as omitting the “distractor“ training set from Czechia during training. The training performance of the untrained network at time 0 falls within the range of 0.15 to 0.25. The downward slope of the validation data sets and the upward trajectory of the mixed training data are classic signs of network overfitting. In this case, the network would be utilised for predictive purposes at around epoch 25 before overfitting occurs.

2.3 Predictive Tool Development

In the next step, we developed an online calculator **https://soilcarbonestimator.fld.czu.cz/** that utilises the CNN model described earlier to predict carbon stock and biodiversity levels for any given location. No specialised software tool beyond a mainstream web browser is needed to access the tool; the interface supports any HTML-capable browser, including those on mobile phones. This tool allows users to interact with the underlying predictive model simply and intuitively, providing valuable environmental data without requiring specialised knowledge of remote sensing or machine learning techniques.

How the Online Calculator Works:

User Input

The user begins by entering the latitude and longitude coordinates of the location they are interested in. These geographic coordinates specify the precise point on the Earth’s surface where the user wishes to predict soil carbon levels or biodiversity.

Data Retrieval

Once the user submits the coordinates, the calculator retrieves the relevant satellite imagery data corresponding to that location from the underlying image database. This includes Sentinel-2 MSI imagery that has been processed and stored as part of the dataset described earlier. The imagery for the location will contain the specific spectral bands and radiometric information necessary for prediction.

Model Inference

The satellite data for the selected coordinates is then passed into the trained CNN model. Using the features it has learned from its training on soil carbon and biodiversity datasets (such as spectral reflectance patterns, atmospheric conditions, and land cover characteristics), the CNN processes the imagery data and generates a prediction.

For carbon predictions, the model will analyse the spectral and spatial features of the landscape (such as vegetation density, soil properties, and land use patterns) to estimate Soil Organic Carbon (SOC) at the specified location. The model will assess landscape diversity, including plant species richness, by utilising learned correlations between remote image features and biodiversity indicators to make biodiversity predictions.

Prediction Output

The online calculator then displays the selected location's predicted carbon stock (typically in units such as tons of carbon per hectare) and biodiversity level (which may be represented by species richness or a biodiversity index). This output provides users with a quantifiable estimate of the environmental variables they are interested in, enabling them to assess the carbon sequestration potential or biodiversity health in a specific area.

Additional Features

To enhance the user experience and provide contextual information, the calculator may also display visual representations, such as a map overlay with the selected coordinates and relevant satellite imagery, or additional data, including temperature, precipitation, and land use type, which could affect carbon and biodiversity predictions. Users may also be able to export this data for further analysis or decision-making in land management or environmental monitoring. Neural network training for our carbon stock and biodiversity.

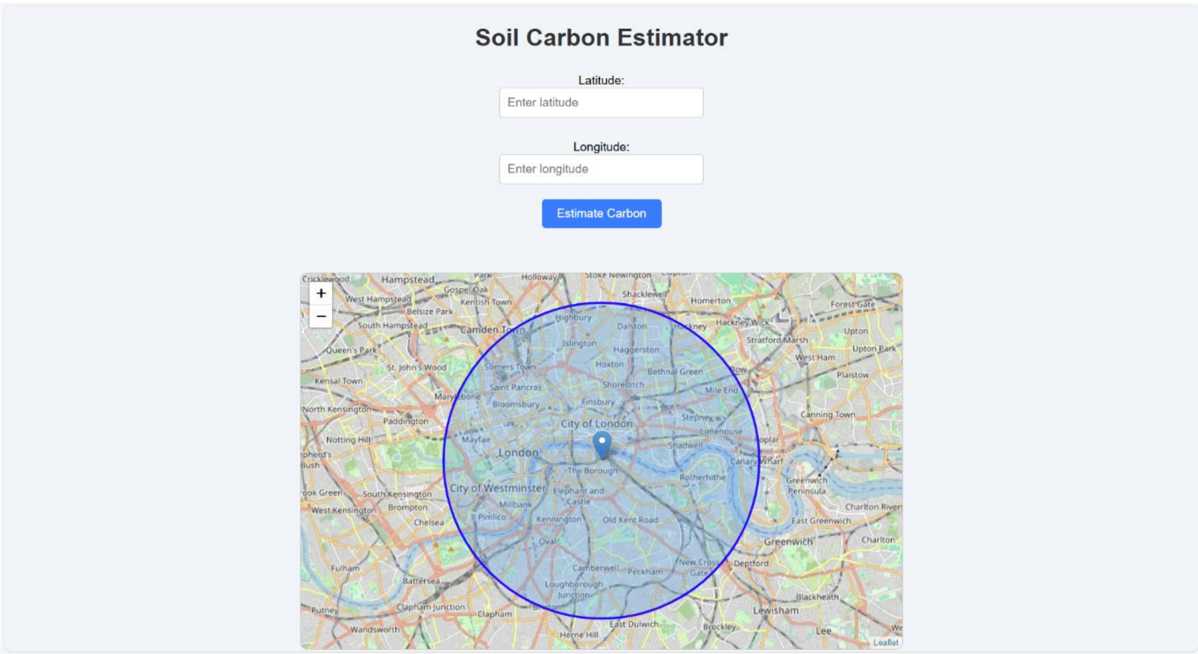


Figure 3: Online Soil Carbon Estimator tool

3. FURTHER DEVELOPMENT OF THE PREDICTIVE TOOL BEYOND THIS DELIVERABLE

The following section describes a workflow in addition to the Grant Agreement. Developments in image capture and analysis technology, concurrent with our work on the CNN, enabled this additional workflow. A separate nationally funded project involving several ReForest contributors yielded an interesting development that we could directly apply to enhance the accuracy of the carbon and biodiversity predictive tool.

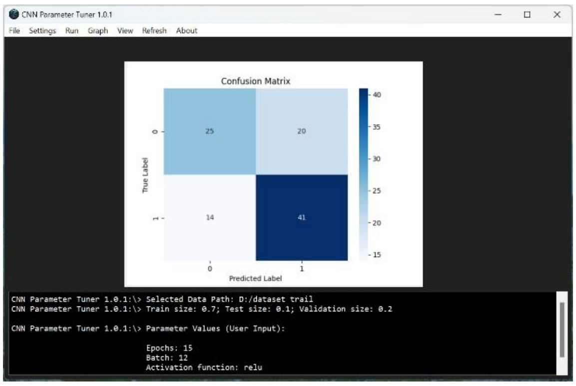
3.1 Brief Description of the Novel CNN Parameter Tuner

Our study applies machine learning and deep learning techniques to classify tree species based on bark images (https://zenodo.org/records/13881151). The study uses two primary datasets: one from the Slovak Republic and the other from the Czech University of Life Sciences. The datasets include tree bark images from seven species captured using high-quality digital cameras. Each tree is represented by multiple images taken from different angles, with at least 60% overlap between consecutive images. Images are segmented into two categories: normal cropped (containing more of the bark area) and exact cropped (containing a smaller, high-quality region-specific portion of the bark).

The images undergo manual segmentation to remove anomalies such as moss, tree leaves, or glare. The segmented images are divided into training and testing sets, with 75% of the data used for training. Additionally, a "Non-linear" dataset is created by applying transformations like swirl and image warping to assess the robustness of machine learning algorithms in handling complex patterns. The Gray Level Co-occurrence Matrix (GLCM) is a key feature extraction technique which analyses the spatial relationships between pixel brightness values in grayscale images. Specific textural features like contrast, dissimilarity, homogeneity, and energy are extracted for classification.

The study tests a variety of classical machine learning algorithms (such as k-Nearest Neighbours, Decision Trees, Random Forest, and Support Vector Machines) as well as neural networks like the Convolutional Neural Network (CNN) and Multilayer Perceptron (MLP). The performance of these algorithms is compared using different parameter values. To optimise the performance of each algorithm, a grid search approach is applied to identify the best hyperparameters for each model. Parameters such as the number of epochs, batch size, filters, and pooling layers are fine-tuned for CNN models. The models' performance is evaluated based on accuracy, precision, recall, and F1-scores. The results show that CNN outperforms classical algorithms in terms of accuracy and robustness, particularly when handling datasets such as the Non-linear one.

This methodology demonstrates how the integration of high-quality bark images and advanced machine learning techniques can enhance tree species classification, with a strong emphasis on data quality, preprocessing, and model fine-tuning.



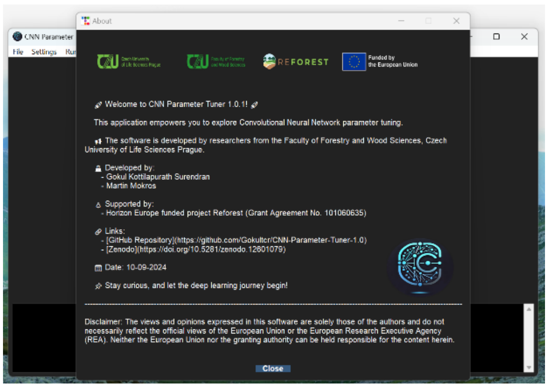


Figure 4: CNN Parameter Tuner tool

3.2 Application Of The Novel Tuner To Carbon And Biomass Predictor

The methodology outlined above for tree species classification can be effectively adapted to enhance the accuracy and performance of the carbon and biodiversity prediction tool. The approach begins with improving the data acquisition process. In the tree species study, high-quality bark images were captured at multiple angles, and this method can be extended to satellite and UAV imagery for predicting carbon and biodiversity. The predictive tool can gain more granular and comprehensive input data by capturing detailed landscape features, such as vegetation density and soil patterns, through various spectral bands and resolutions. This will enable the model to capture finer environmental details, which are essential for accurately predicting carbon stocks and biodiversity levels.

In the next stage, data preprocessing is critical for removing artefacts and anomalies, similar to the segmentation used in the tree bark images. Satellite imagery must be processed to eliminate irrelevant elements, such as seasonal anomalies or shadows, that could skew the predictions for the carbon and biodiversity tool. Applying manual or automated segmentation techniques will help focus the analysis on the most important landscape features, such as trees, shrubs, hedgerows, or even grazing animals, improving the data quality fed into the model.

The feature extraction techniques used in the tree species study, like the Grey-Level Co-occurrence Matrix (GLCM), can also be adapted for carbon and biodiversity predictions. The model can extract key spectral and textural properties from satellite and UAV imagery in this context. For example, physical vegetation features can be extracted and used to predict soil carbon content and biodiversity levels. By focusing on these detailed landscape features, the model will be better equipped to detect environmental patterns influencing carbon sequestration and biodiversity status. The predictions will be validated against ground observations from existing AF systems.

In the same way that the CNN in the tree species study learned from bark images, the model for carbon and biodiversity predictions can be trained on UAV imagery combined with satellite observation and ground-truth datasets. The convolutional layers of the CNN will extract relevant features from the imagery, while the fully connected layers will generate predictions for carbon stock and biodiversity levels. This deep learning approach will enable the model to capture complex, non-linear relationships between the landscape features and environmental variables, leading to more accurate predictions across different ecosystems. Parameters such as the number of convolutional layers, filters, pooling layers, and batch size can be fine-tuned to achieve the best possible performance for specific datasets. This will enable the model to generalise more effectively across different geographic regions and ecosystems, thereby reducing the risk of overfitting and enhancing the reliability of the predictions.

The method of handling non-linear and complex data through transformations, such as image warping, can also be applied in this context. The model can learn to handle heterogeneous environments where carbon and biodiversity predictions may vary greatly by simulating complex environmental patterns- such as crop or woody vegetation. Training the model on this complex data type will make it more robust and adaptable, ensuring accurate predictions even in regions with challenging landscape features or land use types.

Finally, the validation process can follow a similar approach as in the tree species study, using metrics such as accuracy, precision, recall, and F1-scores to assess the model's performance. By incorporating these techniques into the carbon and biodiversity prediction tool, the model can process high-resolution satellite and UAV imagery more effectively, extract the key spectral and spatial features that influence carbon and biodiversity, and use optimised machine learning and deep learning models to provide more accurate and reliable predictions. Additionally, the tool will be able to generalise across diverse landscapes and ecosystems, making it a valuable resource for stakeholders involved in environmental management, conservation, and policy.

3.3 Usability and Further Development of the Predictor Tools

The predictive tool is designed for a broad range of users engaged in land management, research, and environmental governance. In the current form, researchers and educators can draw on the underlying datasets, algorithms, and validation workflows to support studies in remote sensing, soil science, and biodiversity monitoring. With the integration of these tools and workflows into decision-support systems, farmers and landowners can utilise them to estimate the carbon sequestration potential and biodiversity value of their land, thereby supporting informed decisions on tree planting, agroforestry design, and soil management. Advisors and consultants can integrate the tool into farm advisory services, sustainability audits, and climate-action planning, using its outputs to guide practical interventions. Policy makers and environmental agencies can utilise the spatial predictions to inform land-use planning, environmental reporting, and the design of payment schemes for ecosystem services, including carbon and biodiversity credits. Finally, NGOs and civil society organisations may be able to employ the tool as an accessible communication and engagement platform, encouraging public participation in conservation and regenerative agriculture initiatives.

To make the tool fully accessible to these diverse groups, further development should focus on enhancing its user interface, interoperability, and interpretability. A simplified dashboard with visual indicators, map-based outputs, and exportable reports would allow non-technical users to interpret results without prior experience in remote sensing. Integration with mobile devices and low-bandwidth environments would support farmers and advisors working in the field. Creating multi-language interfaces and providing clear guidance materials, case studies, and user tutorials would make the tool more inclusive across regions. Technical development should also include API connectivity with existing platforms such as FarmTree, the ReForest Living Lab data hub, or upcoming CAP monitoring systems to enable data exchange and reduce duplication of effort. Finally, establishing user feedback mechanisms through Living Labs and stakeholder workshops will ensure that future versions of the tool evolve in response to practical needs and improve usability. Future iterations should also include a training and demonstration module, potentially co-developed within the ReForest Living Labs.

4. Conclusions

The carbon and biodiversity predictive tool developed in this deliverable represents a step forward in using remote sensing and artificial intelligence to support data-driven environmental management.

By merging high-resolution satellite and UAV imagery with verified field observations, the model delivers reliable, spatially explicit estimates of soil carbon stocks and biodiversity patterns across diverse European landscapes.

The tool showcases how complex data and modelling techniques can be integrated into a practical decision-support platform accessible to end-users through a simple web interface. This open-access approach allows land managers and policymakers to integrate carbon and biodiversity assessments directly into planning, management, and monitoring processes, supporting transparency and informed decision-making.

Ongoing model refinement will focus on expanding geographic coverage, incorporating additional environmental variables, and improving the CNN architecture through advanced parameter tuning and hybrid learning strategies. These efforts will further enhance the precision, scalability, and interoperability of the predictive framework.

Within the broader ReForest context, this tool supports several key project objectives: it enables the quantification of ecosystem services in agroforestry systems, facilitates monitoring and verification for carbon farming and biodiversity credits, and contributes to the development of integrated digital platforms for Living Labs and policy analysis.

The predictive tool demonstrates how combining Earth observation, artificial intelligence, and open science can transform the way Europe monitors, values, and manages its natural capital. It provides a replicable and scalable foundation for future applications in sustainable land use, climate action, and ecosystem restoration across Europe and beyond.

APPENDIX 1: REFERENCES AND RELATED DOCUMENTS

ID	Reference or Related Document	Source or Link/Location
I	A Forestry investigation: Exploring factors behind improved Tree species Classification using Bark images. Gokul Kottilapurath Surendran, Martin Mokros, Deekshitha, Martin Lukac, Jozef Vybostok	Paper under 3rd review in Ecological Informatics at the time of release of the deliverable
II	Online Soil Carbon Estimator tool	https://soilcarbonestimator.fld.czu.cz/
III	CNN Parameter Tuner tool	https://zenodo.org/records/13881151