TreeAI: a global database for tree species annotations and high-resolution imagery

Mirela Beloiu Schwenke[1], Zhongyu Xia[1], Yan Cheng[2], Arthur Gessler[1,3], Teja Kattenborn[4], Xinlian Liang[5], Clemens Mosig[6], Stefano Puliti[7], Nataliia Rehush[3], Lars T. Waser[8], Verena C. Griess[1] & Martin Mokros[9]

[1] Department of Environmental Systems Science, Institute of Terrestrial Ecosystems, ETH Zurich, 8092 Zurich, Switzerland (mirela.beloiu@usys.ethz.ch), (zhongyu.xia@usys.ethz.ch), (verena.griess@usys.ethz.ch)
[2] University of Copenhagen, Copenhagen, Denmark (yach@ign.ku.dk)
[3] Forest Dynamics, Swiss Federal Institute for Forest, Snow and Landscape Research WSL, 8903 Birmensdorf, Switzerland (arthur.gessler@wsl.ch), (nataliia.rehush@wsl.ch)
[4] Department for Sensor-based Geoinformatics, University of Freiburg (teja.kattenborn@geosense.uni-freiburg.de)
[5] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China (xinlian.liang@whu.edu.cn)
[6] Remote Sensing Centre for Earth System Research, Leipzig University (clemens.mosig@uni-leipzig.de)
[7] Norwegian Institute for Bio-economy Research (NIBIO) National Forest Inventory, Norway (Stefano.Puliti@nibio.no)
[8] Remote sensing group, Swiss Federal Institute for Forest, Snow and Landscape Research WSL, 8903 Birmensdorf, Switzerland (lars.waser@wsl.ch)
[9] Department of Geography, University College London, United Kingdom (m.mokros@ucl.ac.uk)

Keywords: tree diversity; deep learning; forest monitoring; RGBI bands; open benchmark datasets

Abstract:

Accurate and scalable tree species identification remains a critical challenge for global forest monitoring and management. Despite the increasing availability of remotely sensed data, the lack of standardized, high-quality ground truth datasets limits the potential of supervised machine learning models in capturing the tree diversity of forest ecosystems across different environmental and geographic contexts. Prior studies have highlighted the need for global-scale, high-resolution datasets to develop robust algorithms capable of capturing the diversity of forest ecosystems (Lines et al., 2022; Beloiu et al., 2023; Weinstein et al., 2024).

Towards a benchmark dataset for tree species identification in high-resolution aerial imagery. To address this critical gap, we introduce the TreeAI database (link), an open-access dataset designed to support advanced research in tree species identification and forest dynamics. The database comprises 52 datasets contributed by researchers from 29 countries, representing 61,158 annotated trees across 5,000 ha of forest ecosystems (Fig. 1) and it is still growing. Notably, 47 of these datasets are publicly available, making the TreeAI database one of the most extensive global resources for tree-level species annotations from earth observations.

The TreeAI database provides annotations paired with high-resolution imagery (RGB and near-infrared bands at 1–10 cm spatial resolution, with an average of 3.5 cm). This high spatial detail enables the precise characterization of individual trees and facilitates the development of computer vision models optimized for tree species identification. The database offers three key advancements. First, its global representation spans diverse ecosystems, climates, and species, enhancing its applicability across regions. Second, the inclusion of centimeter-scale orthophotos ensures sufficient detail for identifying subtle differences between species. Finally, its community-driven design fosters ongoing contributions and ensures a dynamic dataset that evolves with the field's needs.

Preliminary tree species identification analysis using deep learning algorithms conducted for Switzerland, with very heterogeneous forest ecosystems and challenging topography, yielded promising results. The average F1-score for nine common species was 0.72, with *Larix* spp., *Picea abies*, and *Tilia* spp. exceeding 0.80 (Fig. 2). The mean average precision (mAP) across all the species was 0.76. These findings underscore the potential of the TreeAI database to advance species identification research, particularly when integrated with cutting-edge machine learning techniques. To further harness TreeAI's potential, a scientific competition will be launched in 2025, challenging participants to develop deep-learning algorithms that maximize tree species identification accuracy across a broad range of forest ecosystems.

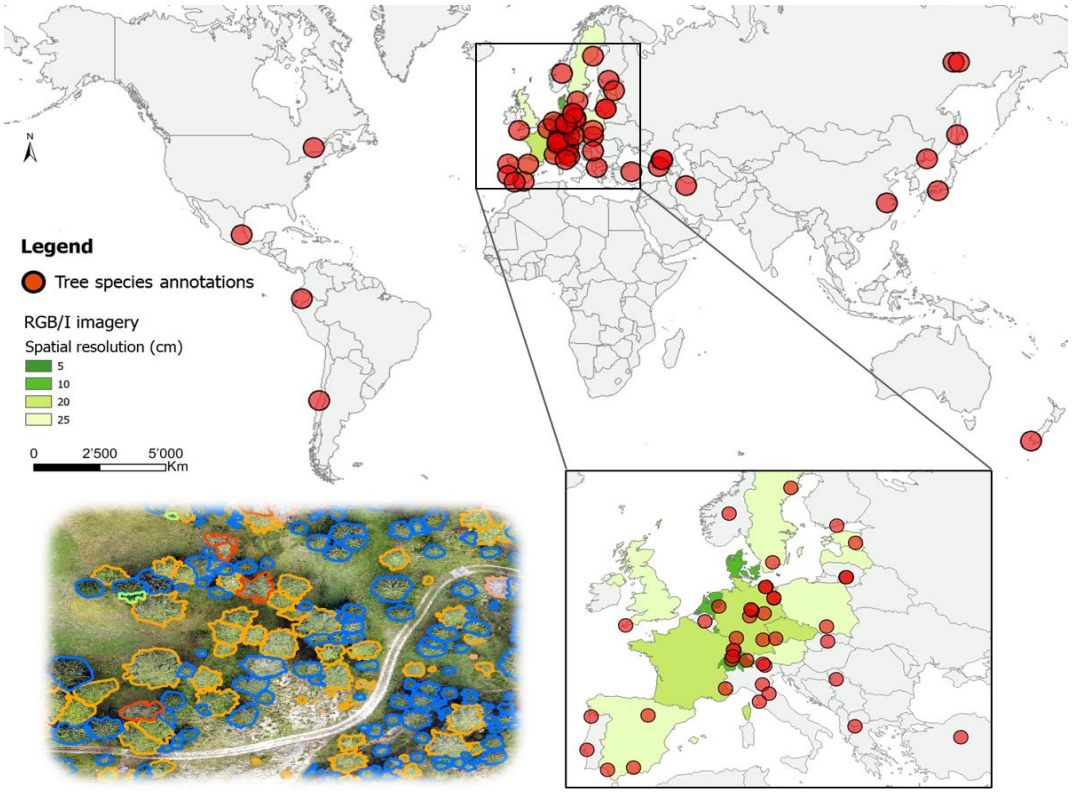

Figure 1. Spatial distribution of the current dataset with the European countries or regions providing high-resolution imagery. An example of individual canopy delineation is also presented.

The impact of a global database for tree species annotations. The TreeAI database serves as a benchmark dataset for advancing artificial intelligence models, enabling automated forest inventory systems. This capability allows for the creation of high-resolution maps detailing tree species distributions, which can be used by researchers and practitioners for applications such as forest management, biodiversity monitoring, and ecosystem conservation. Moreover, the dataset complements existing National Forest Inventory (NFI) data, providing additional resources for point-based regional studies and enhancing ecological research at finer scales.

TreeAI offers researchers a standardized and globally representative dataset that facilitates ecological studies on forest dynamics, biodiversity, and tree growth. The availability of detailed species annotations supports the development of generalized models capable of analyzing diverse biomes, thereby improving our understanding of forest composition, biomass, and carbon stock dynamics at the ecosystem level. For practitioners, the enhanced species identification capabilities offered by TreeAI underpin data-driven forest management strategies. These tools are instrumental in monitoring forests and mitigating climate change impacts, aligning with global efforts toward sustainable forestry. Finally, the database promotes the refinement of AI models for practical forestry applications, fostering innovation in open science and collaborative research.

Further needs and collaboration potential. To maximize the utility of the TreeAI database, future efforts should focus on expanding its geographic and tree species coverage. Collaborations with the global scientific community are essential to incorporating data from underrepresented regions, such as tropical and boreal forests, which remain inadequately sampled in existing datasets. Additionally, integrating TreeAI with Earth observation platforms, such as Planet Scope, Sentinel-2, and GEDI, could enable large-scale analyses that combine ground-level detail with satellite-derived information (Kattenborn et al., 2019; Schiefer et al., 2023; Grabska-Szwagrzyk et al., 2024). Further advancements will also explore methods to enhance data accessibility and interoperability, ensuring that the database meets the evolving needs of its users. Feedback from the broader forestry community will be

instrumental in shaping these developments, with an emphasis on addressing challenges related to data standardization, processing efficiency, and algorithm performance.

Conclusion. The TreeAI database represents a transformative step in addressing the challenges of tree species identification and forest monitoring. A rigorous quality check of uploaded data and continuous support through cloud-based solutions ensure reliability and long-term usability, making it a robust, dynamic platform. By providing globally representative, high-resolution resources, it bridges critical gaps in existing datasets and supports the development of innovative deep-learning applications. As an open-access platform, TreeAI fosters collaboration across disciplines, promoting advancements in forest science that are essential for sustainable ecosystem management and climate change adaptation.

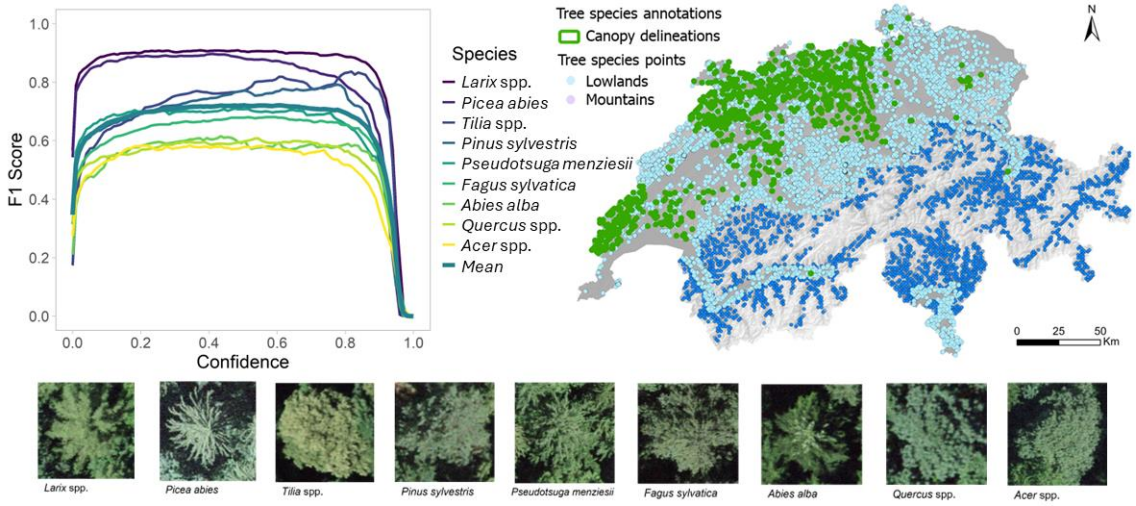

Figure 2. Preliminary results of nine common tree species identified in Switzerland.

References

Beloiu, M., Heinzmann, L., Rehush, N., Gessler, A., & Griess, V.C. (2023) Individual Tree-Crown Detection and Species Identification in Heterogeneous Forests Using Aerial RGB Imagery and Deep Learning. *Remote Sensing*, 15(5), 1463. https://doi.or/10.3390/rs15051463

Grabska-Szwagrzyk, E., Tiede, D., Sudmanns, M., & Kozak, J. (2024) Map of forest tree species for Poland based on Sentinel-2 data. *Earth System Science Data*, 16(6), 2877–2891. https://doi.or/10.5194/essd-16-2877-2024

Kattenborn, T., Lopatin, J., Förster, M., Braun, A.C., & Fassnacht, F.E. (2019) UAV data as alternative to field sampling to map woody invasive species based on combined Sentinel-1 and Sentinel-2 data. *Remote Sensing of Environment*, 227, 61–73. https://doi.or/10.1016/j.rse.2019.03.025

Lines, E.R., Allen, M., Cabo, C., Calders, K., Debus, A., Grieve, S.W.D., et al. (2022) AI applications in forest monitoring need remote sensing benchmark datasets. https://doi.or/10.48550/arXiv.2212.09937

Schiefer, F., Schmidtlein, S., Frick, A., Frey, J., Klinke, R., Zielewska-Büttner, K., et al. (2023) UAV-based reference data for the prediction of fractional cover of standing deadwood from Sentinel time series. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 8, 100034. https://doi.or/10.1016/j.ophoto.2023.100034

Weinstein, B.G., Marconi, S., Zare, A., Bohlman, S.A., Singh, A., Graves, S.J., et al. (2024) Individual canopy tree species maps for the National Ecological Observatory Network. *PLOS Biology*, 22(7), e3002700. https://doi.or/10.1371/journal.pbio.3002700