

# Enhancing Bark Classification with Boosted Support Vector Machines Using Bagging and Feature Selection Techniques

Gokul Kottilapurath Surendran<sup>1</sup>, Martin Mokros<sup>1,2,3</sup> Jozef Vybostok<sup>3</sup>

<sup>1</sup> Czech University of Life Sciences, Prague; <sup>2</sup> Univeristy College London; <sup>3</sup> Technical University in Zvolen, Slovakia

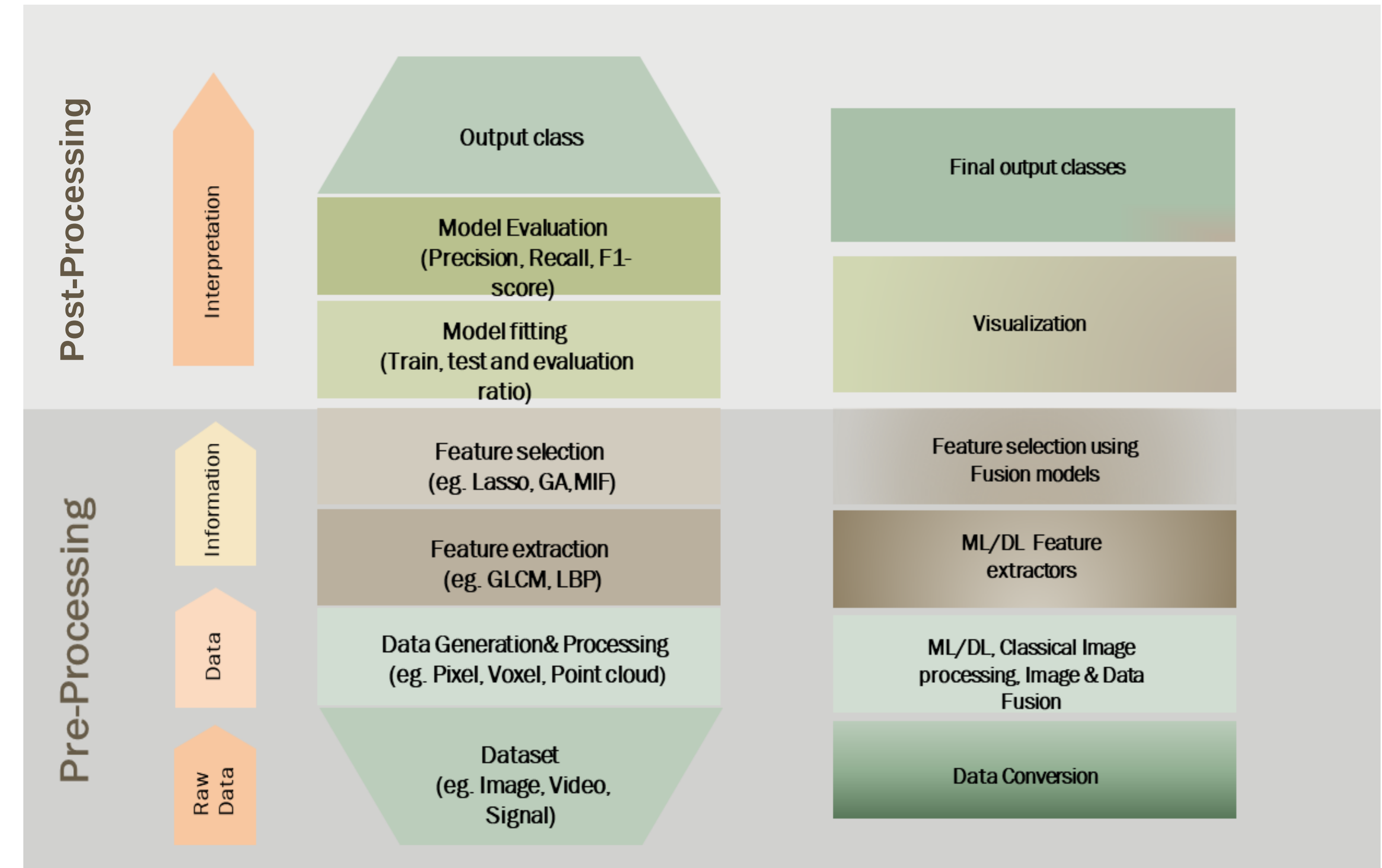
### CONTRIBUTIONS

The objective was to enhance bark classification accuracy through the refinement of the support vector machine using bagging and feature selection techniques, leading to the achievement of an impressive accuracy rate of 95%

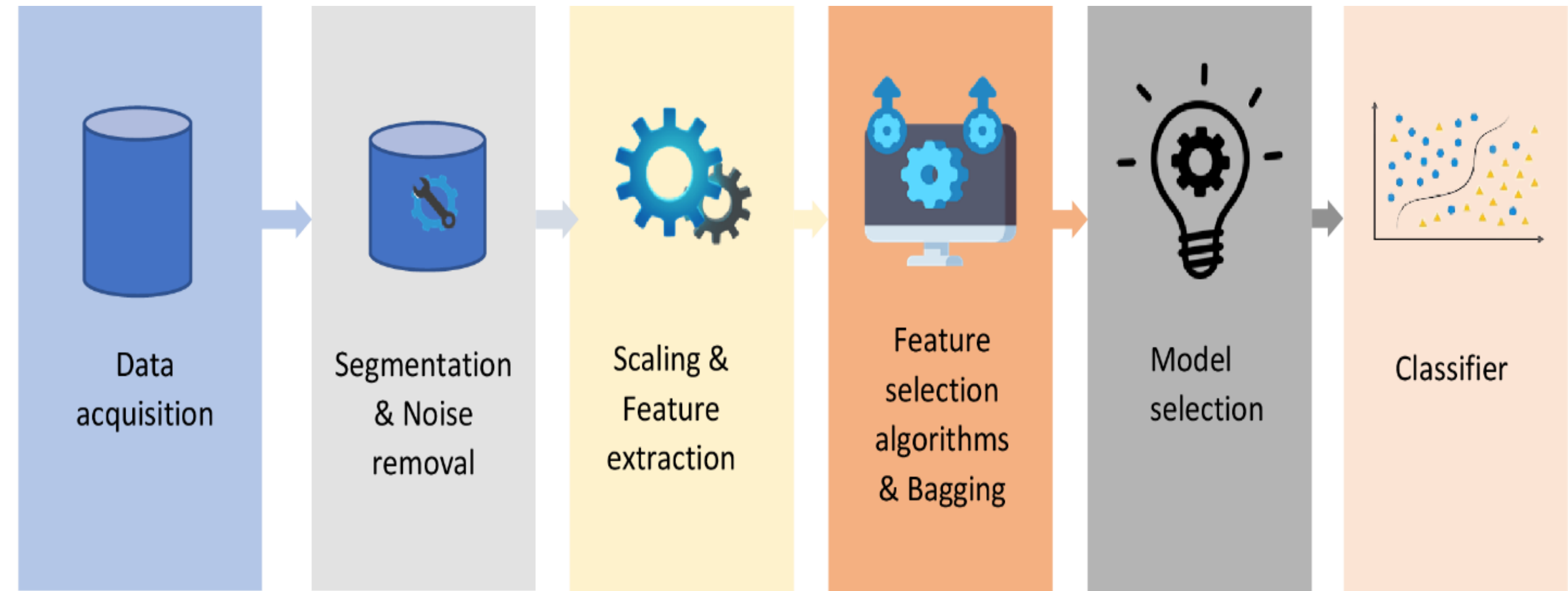
### WHAT, WHY, HOW

- What ?** - Advanced machine learning algorithm that provides higher precision in comparison to CNN
- Why ?** - To achieve improved outcomes while working within constrained computational resources
- Employing traditional machine learning algorithms for smaller datasets,, rather than relying solely on neural networks.
  - Explore potential avenues for enhancing traditional algorithms
  - Find out the best feature extractor (GLCM, CNN)
- How ?** - Feature selection algorithms (Lasso, Mutual Information, Feature selection transformer, GA)
- Bagging classifier

### GENERAL ARCHITECTURE: MACHINE LEARNING



### PROPOSED WORKFLOW



- A. The Bark Dataset contains 527 images of European beech (*Fagus sylvatica* L.), Sessile oak (*Quercus petraea* (Matt.) Liebl.), Norway spruce (*Picea abies* (L.) H. Karst.), and European silver fir (*Abies alba* Mill.); captured in Včelien, Banskobystrický, Slovakia, in June 2022.
- B. Performed image segmentation and effectively eliminated outliers and extraneous elements from the dataset.
- C. We derived features utilising convolutional neural networks and grey-level co-occurrence matrices.
- D. Employed feature selection algorithms and bagging techniques to ascertain the optimal classification approach.
- E. Choosing the optimal model.
- F. Evaluating the accuracy of the SVM classifier in comparison to alternative algorithms.

### ACKNOWLEDGEMENT

This abstract is based upon work from COST Action 3DForEcoTech, CA20118, supported by COST (European Cooperation in Science and Technology). We express our gratitude to IGA-3168 (FLD, CZU) for their generous support in funding this research and APVV-20-0391 Monitoring of forest stands in three-dimensional space and time by innovative close-range approaches for funding in data collection.

### OBJECTIVES

1. To minimize the **cost and scientific effort** needed for species classification
2. Deliver **superior outcomes** using traditional algorithms
3. Examine potential opportunities to augment the current algorithms

Gokul  
WhatsApp contact



### RESULTS

Achieved enhanced accuracy through the utilization of Support Vector Machines, employing:

1. **CNN feature extractor** – Successfully achieved improved accuracy by utilizing CNN as the feature extractor.
2. **Feature selection algorithms** – The feature selection approach enhanced the algorithm’s capacity to yield superior outcomes.
3. **Bagging** – Contributed a modest enhancement to the accuracy of the algorithms, albeit not to the extent seen with feature selection.

| Method                                                    | Training accuracy (%) | Testing accuracy (%) |
|-----------------------------------------------------------|-----------------------|----------------------|
| SVM/Mutual information f. selection (GLCM f. ext./scaled) | 95                    | 91                   |
| DT/Bagging classifier (CNN f. ext.)                       | 100                   | 89                   |
| <b>SVM/Lasso f. selection/CNN f. ext.</b>                 | <b>100</b>            | <b>94</b>            |
| RF/Bagging classifier (CNN f. ext.)                       | 99                    | 84                   |
| SVM/Bagging classifier (CNN f. ext.)                      | 95                    | 89                   |

| Method                                            | Training accuracy (%) | Testing accuracy (%) |
|---------------------------------------------------|-----------------------|----------------------|
| SVM/CNN f. ext.                                   | 100                   | 91                   |
| <b>SVM/F. selection transformer (CNN f. ext.)</b> | <b>100</b>            | <b>95</b>            |
| SVM/GA f. selection (CNN f. ext.)                 | 100                   | 89                   |
| SVM/Recursive f. elimination (CNN f. ext.)        | 100                   | 93                   |
| SVM/Lasso f. selection (GLCM f. ext./scaled)      | 83                    | 80                   |

For the ultimate model, we opted for CNN as the feature extractor, yielding an extraction of 12,800 features. Subsequently, we employed premier feature selection algorithms- such as the **Feature Selection Transformer, Lasso Feature Selection**, and **Recursive Feature Elimination**- resulting in an accuracy exceeding 90%. The pinnacle of accuracy, reaching 95%, was attained using the Feature Selection Transformer model.

| Class               | Precision | Recall      | F1-Score |
|---------------------|-----------|-------------|----------|
| European Beech      | 1.00      | 1.00        | 1.00     |
| European Silver fir | 0.83      | 1.00        | 0.91     |
| Norway Spruce       | 0.95      | 1.00        | 0.98     |
| Sessile Oak         | 1.00      | 0.81        | 0.90     |
| <b>Accuracy</b>     |           | <b>0.95</b> |          |

### FUTURE WORK

In the future, we plan to:

1. **Barknet 2.0 Europe**- Generate a dataset named ‘Barknet 2.0 Europe,’ derived from the collection of bark images.
2. **Close-range Datasets** – Utilizing other close-range datasets for classification purposes.
3. **Natural-Inspired algorithms** – To explore the application of various Natural-Inspired algorithms for the purpose of species classification.

### LITERATURE REFERENCES

- [1] Mokros, et al. <https://doi.org/10.5194/egusphere-egu23-14332>
- [2] M. Carpentier et al. <https://doi.org/10.1109/IROS.2018.8593514>
- [3] Kim et al. <https://doi.org/10.1038/s41598-022-08571-9>

### ACKNOWLEDGEMENT

This abstract is based upon work from COST Action 3DForEcoTech, CA20118, supported by COST (European Cooperation in Science and Technology). We express our gratitude to IGA-3168 (FLD, CZU) for their generous support in funding this research and APVV-20-0391 Monitoring of forest stands in three-dimensional space and time by innovative close-range approaches for funding in data collection.



Ceská zemědělská univerzita v Praze  
**Fakulta lesnická a dřevařská**



COST Action 3DForEcoTech  
"From 4D to 3D: A new paradigm in forest monitoring and management"

