Project Report: AI-Driven Mapping of Forest Biodiversity

Institution: Duke University
Supervisors: Dr. Tong Qiu, Dr. Zhuohong Li
Team Members: Hashim Alrefaei,Georgiy Zemlevskiy, Sophie Mao,Alex Zhang, Danial Murong

---

1. Objective

To construct an automated workflow for generating training data for a tree species classification deep learning model using data from the NEON (National Ecological Observatory Network) terrestrial forest sites. This enables biodiversity assessment and supports conservation efforts.

---

2. Key Innovations

- Leveraged deep learning models over traditional ensemble methods.
- Replaced simple bounding boxes with precise crown delineations.
- Developed automation tools to streamline dataset creation and model training.

---

3. Methodology Overview

Label Generation

- Extracted tree species, trunk location, height, and crown diameter using NEON Vegetation Structure data.
- Filtered out trees that were dead, damaged, or not visible from aerial view.

Tree Crown Delineation

- Used SAM2 (Segment Anything Model) on RGB imagery to create bounding boxes from the DeepForest dataset.
- Labeled known trees with species classes, unknown trees as "unknown," and others as "background."

Input Concatenation

- Converted NEON LiDAR point clouds to Canopy Height Models (CHM).
- Combined CHM with hyperspectral bands (reduced from 426 to 150 via Kullback–Leibler divergence).
- Concatenated inputs into a unified training format.

Model Training

- Trained a hybrid CNN-ViT semantic segmentation model (Parformer) using the generated dataset.
- Evaluated performance using metrics like Overall Accuracy, IoU, and Kappa score.

---

4. Individual Contributions

All student contributed to all aspects on the project but some concentrated more on something.

- **Georgiy Zemlevskiy**: Led the automation of dataset generation, including label extraction and hyperspectral data preprocessing pipelines.
- **Hashim Alrefaei**: Focused on crown delineation using SAM2 and developed the final segmentation model (Parformer).
- **Sophie Mao**: Worked on input concatenation strategies and explored techniques to improve overall model accuracy.

---

5. Results

- Achieved **Overall Accuracy**: 0.9877
- **Mean IoU**: 0.7958
- **Kappa (avg)**: 0.8891
- **Freq. Weighted IoU**: 0.9770
- Demonstrated strong performance on ~20x20m plots from Mountain Lake Biological Station (MLBS), VA.

---

6. Next Steps

- Automate bounding box generation using DeepForest for scalability.
- Expand Parformer training across all NEON sites.
- Experiment with alternative model architectures.
- Further optimize the end-to-end workflow for large-scale biodiversity mapping.