|  |
| --- |
| **Mapping forest type combining field plot and high-resolution data with machine learning in the boreal forest of Interior Alaska** |
| The boreal biome, the largest terrestrial biome on Earth, is increasingly vulnerable to climate change. This biome is warming twice as rapidly as the global average. Climate change has increased the temperature, frequency, severity, and extent of fires, which is resulting in forest type conversion and species ranges to shift. The increasing pace and scale of disturbances and rapid shifts in dominant forest type necessitate precise mapping to better understand ecosystem response to climate change and support effective management strategies.  While most previous vegetation mapping efforts have utilized coarse to moderate resolution satellite data, there has yet to be a spatially extensive and high-resolution forest type map. In this study, we present a framework to generate high-resolution forest type maps using a combination of field and aerial imagery in the Tanana unit of Interior Alaska. Our dataset is composed of high-resolution (1 m) canopy height, various vegetation indices derived from hyperspectral, and topographic variables including elevation, slope, aspect, and solar radiation collected by NASA Goddard’s Lidar, Hyperspectral and Thermal Imager (G-LiHT) and field data collected by Forest Inventory and Analysis program (FIA).  We classified forest types at three different levels i) forest and nonforest, ii) hardwood, softwood, and nonforest, and iii) major forest types such as birch, black spruce, white spruce and nonforest. To achieve this objective, we applied and compared the performance of a convolutional neural network (CNN) with a XGBoost model. In this framework, we also studied the contribution of combinations of different data modalities to influence classifier accuracy in identifying forest types. We used SHapley Additive exPlanations (SHAP) method for understanding the importance of topographical factors that are related to forest distribution.  We found that the CNN model outperformed XGBoost model across all forest type classifications in terms of overall accuracy and macro average F1 score. The overall accuracy of CNN model was 93.06% for forest and nonforest, 82.59% for hardwood, softwood and nonforest, and 74.74% for birch, black spruce, white spruce and nonforest. In addition, we found canopy height and digital terrain model were the most important variables for all classifications. Further, we found several vegetation indices to promote detection including Anthocyanin Reflectance Index (ARI1) which was useful for differentiating between forest and nonforest. While vegetation indices such as Photochemical Reflectance Index (PRI), Pigment Specific Normalized Different (PSND), Gitelson and Merzlyak (GM1) and DATT2 were useful for differentiating between hardwood and softwood.  The CNN models were best for classification of boreal forest types. Additionally, we found that elevation was the most important topographical factor for driving forest type distribution. Further, we found that vegetation indices such as PRI, PSND, GM1 and DATT2 were more useful for differentiating between boreal forest types. We aim to use the high-resolution forest type map to produce wall to wall maps in the study area in future. The development of these kinds of frameworks are crucial for operationalizing remote sensing data in biodiversity and forest monitoring, which is especially important in large, remote areas such as boreal forests of interior Alaska. |