Tree Detection and Characterization in Diverse Agroforestry Systems using UAV Imagery and DL

Nagashree M. K[1]*, Satisha G.C[3], Rama Subramoniam S[1], Satish Kumar B[2], Narendran J[2], Maddi Reddy S. S[4], Chamundeshwari D[2], Premananda B. D[4], Hebbar R.[1], & Srivatsav S. K[5]

[1]Regional Remote Sensing Centre-South, NRSC, ISRO, Marathahalli, Bengaluru, Karnataka;
[2]Aerial Services and Digital Mapping Area, NRSC – Hyderabad.
[3]Indian Institute of Horticulture Research, Bengaluru, Karnataka.
[4]University of Agriculture Sciences, Bengaluru.
[5]Regional Remote Sensing Centre, NRSC, Delhi, ISRO
*Corresponding author e-mail: nagashree_tr@nrsc.gov.in

Abstract

Detection and counting of trees play an important role in biomass & yield estimation, agroforestry diversification, carbon sequestration and other environmental applications. Recently, high resolution UAV and satellite images have been widely used for tree detection and to estimate tree height, canopy cover with deep learning algorithms demonstrating its immense potential for counting trees with promising accuracies. In the present study, DL based algorithm was applied for detection of trees in diverse agroforestry systems covering two study areas viz., Niragunupal watershed (594 ha) Chikkaballapur district and ICAR-IIHR campus (273 ha) Hessaraghatta, Bengaluru, Karnataka. These study areas are characterized by the presence of diverse horticulture plantations like coconut, guava and mango and dominant forest species are tamarind, pongamia, acacia, agave, prosopis, etc. UAV images in RGB mode were acquired during February and March, 2023 with 2.5 cm spatial resolution and processed photogrammetrically for generating 5 cm Digital Surface Model (DSM), Digital Terrain Model (DTM) and ortho-mosaic. In-season ground truth collected for selected of tree species along with geotagged photographs and relative tree height. Retina Net pre-trained object detection model has been re-trained with 223 annotated training samples collected from Niragunupal study area. DTM and DSM were used for estimating the relative height of trees. The re-trained model used for inferencing for detection and counting of trees in both the study areas. The results revealed that the accuracy was 89.1% and 85.3 % with Kappa of 0.88 and 0.72, respectively for Niragunupal and IIHR test sites. The total number of trees estimated was 39544 in Niragunupal watershed and 22426 in ICAR-IIHR study areas. The analysis indicated that the tree heights ranged from 3.0m to 20.0 m with average height of 5.71m and the canopy cover area ranged from 3.0 m2 to 90 m2 for both study areas. It was observed that model trained for Niragunupal watershed could be used for inferencing in IIHR test site due to similar terrain conditions and diversity. Further work is in progress for development of deep learning model for discriminating different horticulture and forest tree species.

Keywords Tree Counting, UAV Images, DTM, Deep Learning

Introduction

The individual tree crown (ITC) segmentation algorithm based on aerial images is a prerequisite for understanding tree growth, tree species competition, and biomass assessment. (Zhou, et al. 2020). It is an automatic procedure that allows the detection of the position, the size, and the shape of trees in a remote sensing data. This procedure is extremely useful in ecological studies as it allows researchers to analyse a forest in its primary element, the tree or group of trees either linear or block plantations. The identification of individual tree-crowns (ITC) is an important research topic in forestry, remote sensing and computer vision [Zhen, et al 2016]. Acquiring individual tree information is beneficial for forest growth assessment and sustainable forest management (Zhou et al., 2020).

Traditional methods for ITC detection, image classification, segmentation, etc. applied to very high-resolution remote sensing imagery have been shown to struggle in disparate landscape types or image resolutions due to scale problems and information complexity (Pleșoianu et al 1998). Deep learning promised to overcome these shortcomings due to its superior performance and versatility, proven with reported detection rates of around 90%. However, such models still find their limits in transferability across study areas, because of different tree conditions (isolated trees vs. linear or block trees) and resolutions of the input data. Applications for tree detection and counting offer a powerful solution to the challenges of precision agriculture, enabling plantations to increase productivity and sustainability while reducing costs and manual labour. It has made enormous strides in a range of computer vision tasks but requires significant amounts of training data. Objective of this study is to use-deep learning algorithm for tree detection and characterization in two diverse agroforestry systems.

Materials and Methods

Study sites: Niragunupal watershed covering an area of 594 ha located in the Bagepalli taluk of Chikkaballapur district in Karnataka was chosen as the first study area shown Figure 1. The area comprises of land cover features such as built-up, barren, agricultural fields, hilly areas and water bodies etc. The major crop grown is sunflower followed by maize, potato, carrot, mulberry, etc. The dominant tree and shrub species in the study area are Tamarind, Pongamia, Acacia, Agave, Prosopis, etc. and Coconut, Guava and Mango are the plantation crops scattered in the area. This watershed can be categorized under Agro Silvi agroforestry system.

ICAR-IIHR Campus located in Hessaraghatta, Bengaluru is taken up as second study area also as validation site. Here dominant tree species are Pseudo Ashoka, Rain tree, Pongamia and silver oak etc. Figure 2 and plantation crops like Mango, Sapota, Guava, Jackfruit and other horticulture crops. This agroforestry system can be classified as Silvi-Horti. UAV datasets were acquired and analysis ready data including RGB, DSM & DTM Images were used for this study. Characteristics of UAV dataset is given below Table 1(a).

UAV data acquisition & Processing: UAV data acquisition was carried over the study area using Trinity F90+ and Sony RX1RII RGB digital camera at a flying height of 120m with spatial resolution of 1.5 cm with forward overlap of 70% and side overlap of 70%. Table 1(b). Prior to UAV surveys, a reconnaissance survey of the study area and its surroundings was carried out to observe the topography and obstructions such as cell towers, power lines or other objects for safety of the UAV.

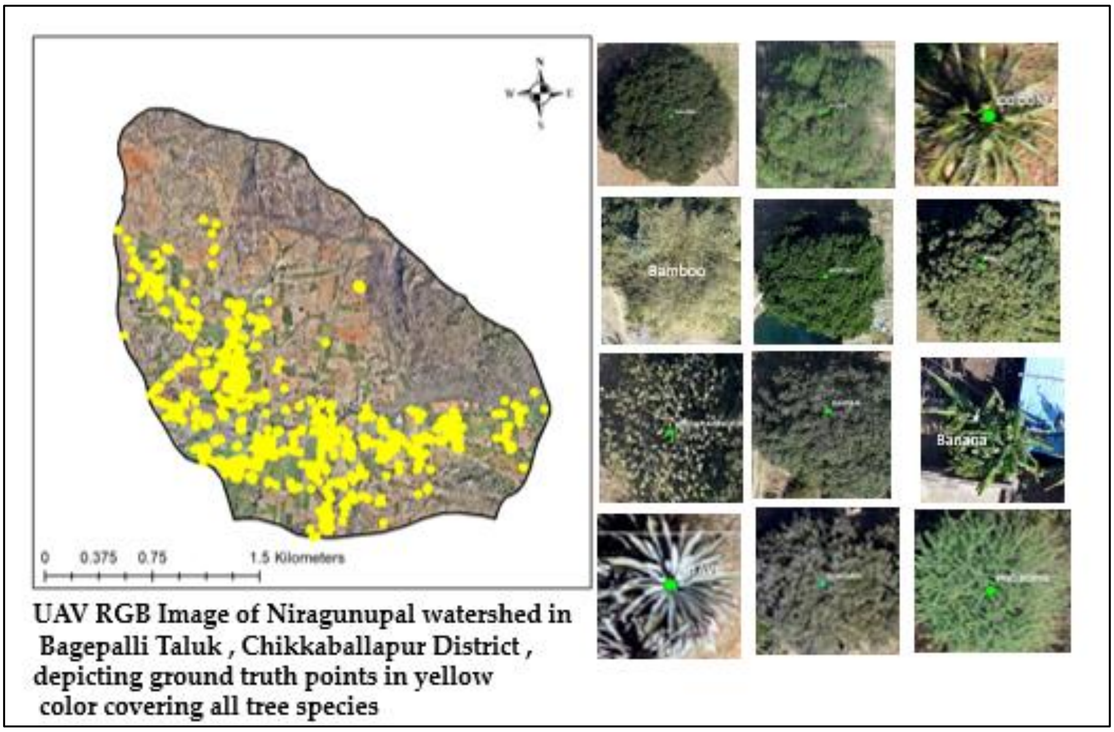


Fig. 1 Niragunupal watershed with GT Points.

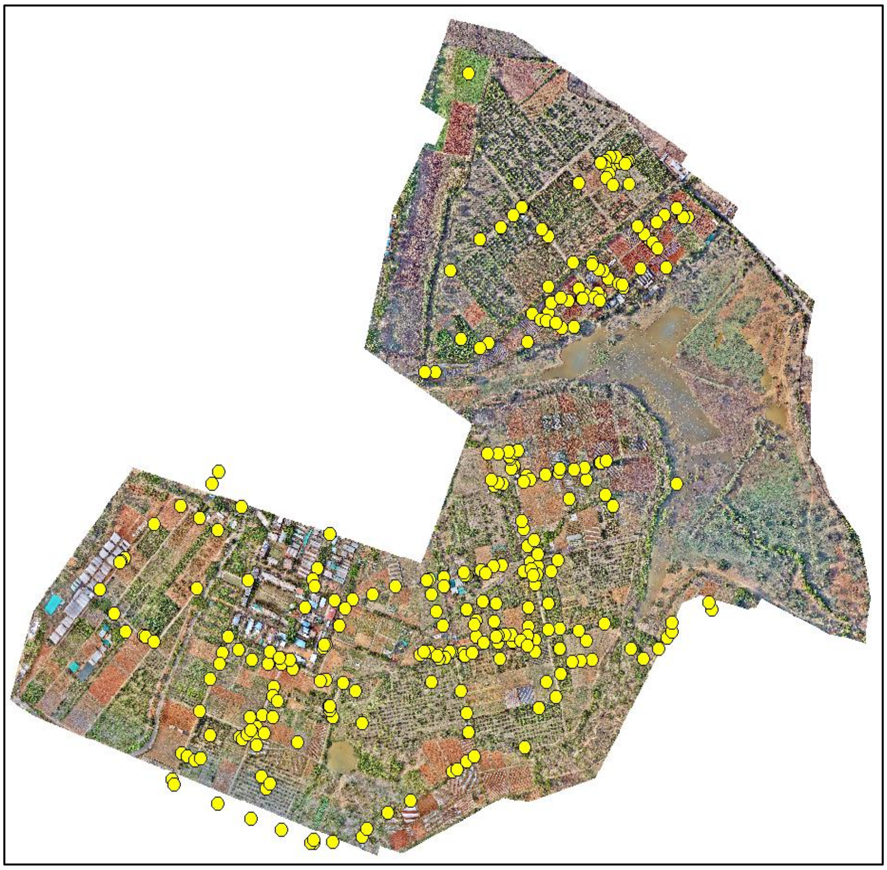


Fig. 2 IIHR Campus Hessaraghatta with GT points.

Table 1(a) Details of UAV RGB, DTM & DSM data.

Data/Sensor	Resolution/ Parameters
Digital Camera RGB Data	2.5 cm
Digital Surface model	5 cm
Digital Terrain model	5 cm
Direction	Nadir in bi-directional mode
Flying Height	120 m
Acquisition period	04 and 05 Feb. 2023 for Bagepalli and 27 to 29 March 2023 for ICAR-IIHR Campus

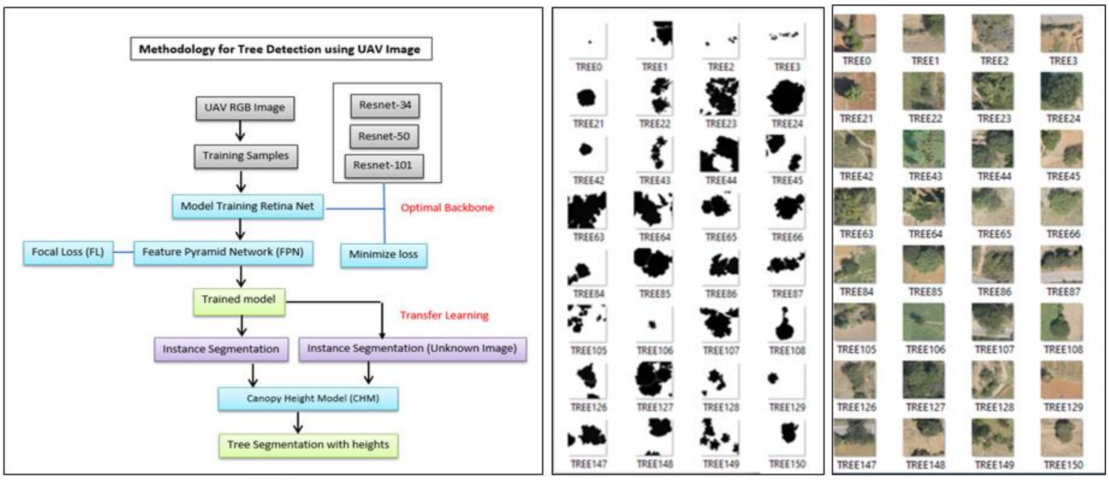
Table 1(b) Flight parameters during the UAV data collection.

Parameter	Value
Flight Altitude	120 m
Side Overlap	70%
Forward Overlap	70%
Camera trigger speed	1.1 sec
Number of flight lines	67
Average Ground Sampling Distance (GSD)	1.55 cm

Methodology: Deep-learning (DL) algorithms, which learn the representative and discriminative features in a hierarchical manner from the data, have recently become a hotspot in the machine-learning area and have been introduced into the geoscience and remote sensing (RS) community for RS big data analysis. (L. Zhang. Et al 2016). The site-specific crop mapping is important as it helps the agronomists to develop estimations for yield, biomass, vigor, crop cover, etc. for multiple crops discretely. (Latif et.al 2019). In the last two decades, various semi- and fully-automatic algorithms have been developed for individual tree detection and crown delineation. However, even if one method is best for a specific application, it may not be optimal for other situations. (Zhen. et al 2016). The workflow for identifying and locating trees comprises six steps: (i) Input UAV RGB Image (ii) Manual digitization of trees as a basis for the creation of training data and as test data (ground truth features) (iii) Generating training data (iv) Training deep learning model (v) Inferencing or detecting objects (vi) Transfer Learning (vii) Accuracy assessment.

Generation of training data: The challenge for applying deep learning to natural systems is the need for large training datasets. (Weinstein et al 2018). UAV RGB images of 2.5 cm spatial resolution was used for manually on-screen digitization. 223 training samples as Images and corresponding labels generated as shown in Fig. 3(b) for first study site. These training samples form an important input for generation of DL model. This is possible because the distances between the trees are relatively large, and the trees are therefore usually recognizable as individual objects. The flow chart of the deep learning tree-crown detection method depicted in Figure 3(a). Gray boxes indicate input and blue boxes indicate intermediary data products. Purple boxes indicate processing and analysis stages. Green boxes indicate final outputs.

Training Deep learning Model: Retina Net seems to be the best effective algorithm in Deep learning for objects detection. (B. S. Reddy et al. 2022). A pre-trained Retina Net one-stage object detection model is used in this study for detection of trees. The training of the new model requires heavy computational work and is time-consuming. Pre-trained models re-trained with additional training samples.



(a) (b)

Fig. 3 (a) The flowchart for tree detection. (b) Training data with Images & labels.

The one-stage Retina Net network architecture as shown in Figure 4. This model addresses the challenges of imbalanced data and objects of different sizes. It accomplishes this through a unique architecture that uses a Feature Pyramid Network (FPN) and Focal Loss function. The model architecture consists of a backbone network (Res Net), a feature pyramid network (FPN), and two task-specific sub networks for classification and regression. The backbone network is responsible for extracting features from the input image. The FPN combines these feature maps to construct a pyramid of multi-scale feature maps that are used to detect objects of different sizes. The classification sub network uses the feature maps produced by the FPN to classify objects into different classes, while the regression sub network refines the bounding box coordinates of the objects. It also uses anchor boxes, which are predefined bounding boxes of different sizes and aspect ratios, to detect objects at different locations and scales. The Feature Pyramid Network addresses the issue of objects of different scales by generating a feature pyramid with different scales and resolutions. Focal loss addresses the issue of imbalanced data by assigning higher weights to hard examples, which are objects that the model is struggling to detect. The focal loss function works by down weighting the loss assigned to well-classified examples and up-weighting the loss assigned to misclassified examples. This approach is more effective than traditional cross-entropy loss, which assigns equal weight to all examples regardless of their difficulty.

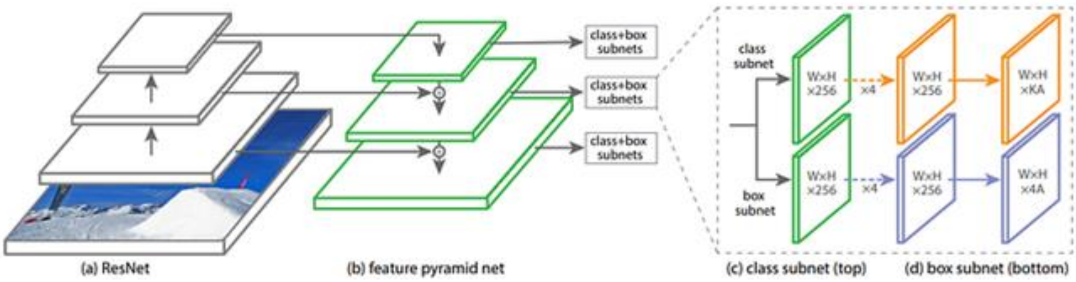


Fig.4 Network architecture of Retina net.

Instance Segmentation or detecting objects: In recent years, instance segmentation has become a key research area in computer vision. Instance segmentation technology not only detects the location of the object but also marks edges for each single instance, which can solve both object detection and semantic segmentation concurrently (Rabi Sharma. Et.al 2022). Deep learning inference is the process of using a trained DL model to make predictions against UAV RGB Image. Instance Segmentation integrates an object detection task where the goal is to detect an object and its bounding box prediction in an image and semantic segmentation task, which classifies each pixel into predefined categories

Transfer Learning: Transfer Learning is usually applied when there is a new dataset smaller than the original dataset used to train the pre-trained model (Larsen-Freeman, D. 2013). Transfer learning is a Machine Learning technique whereby a model is trained and developed for one task and is then re-used on a second related task (Hussain, et al 2018). It's currently very popular in deep learning because it can train deep neural networks with comparatively little data. DL Model generated for Nirupangal watershed is used for ICAR IIHR Campus without any training samples.

Accuracy assessment: Accuracy assessment or validation is done for DL model and Inference output. Accuracy metrics for DL model is measured using precision and recall scores. Precision measures the percentage of predictions made by the model that are correct. Recall measures the percentage of relevant data points that were correctly identified by the model. Precision is the ratio of the number of true positives to the total number of positive predictions. The model detected 100 trees, and 89 were correct, the precision is 89 percent. The overall accuracy of the classified image compares how each of the objects is classified versus the definite tree obtained from their corresponding ground truth data. Producer's accuracy measures errors of omission, which is a measure tree types can be classified. User's accuracy measures errors of commission, which represents the likelihood of a classified tree cover matching the actual GT. The error matrix and kappa coefficient have become a standard means of assessment of image classification accuracy. About 223 ground truth features collected for Niragunupal watershed were manually mapped. Accuracy assessment is carried out by using the collected ground truth information and additional ground truth points taken for validation. Depending on the input training samples, the detected trees and the results of the accuracy metrics vary.

Spatial Analysis: The Inference layer generated for both study sites have been used to detect height of trees in meters and canopy cover in sq. meters. The Canopy Height Model (CHM), represents the heights of the trees on the ground. We can derive the CHM by subtracting the ground elevation from the elevation of the top of the surface (or the tops of the trees). Canopy height is generated by subtracting the DSM (Digital Surface model) values with a DTM (Digital Terrain Model). The DTM and DEM are of 5m resolution. Canopy height presented good results for tree height estimation as shown in Figure 5. This includes the actual heights of trees, builds and any other objects on the earth's surface

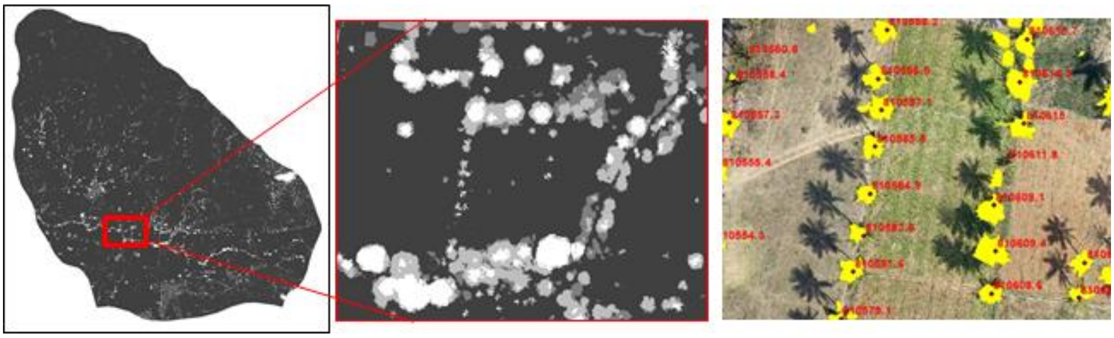


Fig. 5 CHM-Tree height estimation (using UAV DSM & DTM) (middle image shows the zoomed view showing different tree heights.

Results and Discussion

DL model that is generated using Retina net object detection model Niragunupal watershed has model accuracy of 89.1% Figure 6. Transfer learning is a very popular technique in deep learning where existing pre-trained models are used for new tasks. Basically, the idea is to take one neural network that we trained for one specific task and use it as a starting point for another task. The trained model of Niragunupal watershed is used for Inferencing ICAR IIHR RGB Image. Model accuracy found to be 85.3%. Figure 6. It has been found that model generated through Transfer learning is suited for medium height tree detection which can further be improved by additional training samples. Total No of 39544 trees have been detected for the first study site and 22426 trees were detected for second study site. Spatial analysis performed on Inference DL output and Canopy height model output as shown in Table 2(a) and 2(b).

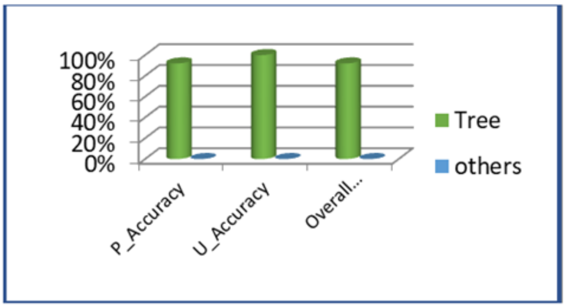
DL model and overall classification accuracies found encouraging. Most of the trees detected using Retina Net model indicated that the tree heights ranged 0 to 5 m was very high and trees height range between 5 to 10 meters were medium range. This clearly indicates that model best suited for medium tree height detection. Further the tree heights ranged from 3.0m to 20.0 m with average height of 5.71m and the canopy cover area ranged from 3.0 m2 to 90 m2 for both study areas. The accuracy metrics computed with respect to the two study sites are given in Fig. 6.

Table 2(a) Tree height and canopy area for Niragunupal watershed

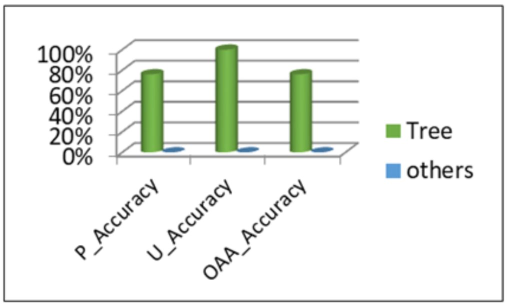
Tree height (m)		Canopy area	
0-5	34487	0-2	23291
5-10	3158	2-5	10376
10-15	963	5-11	3799
15-20	140	11-25	1485
20-25	20	25-72	590

Table 2(b) Tree height and canopy area for IIHR watershed.

Tree height (m)		Canopy area	
0-4	12623	0-5	14515
4-8	7526	5-10	5409
8-12	1547	10-19	2029
12-16	545	19-37	418
16-23	158	37-91	55



Niragunupal watershed



IIHR Campus, Bengaluru

Fig. 6 Classification accuracy.

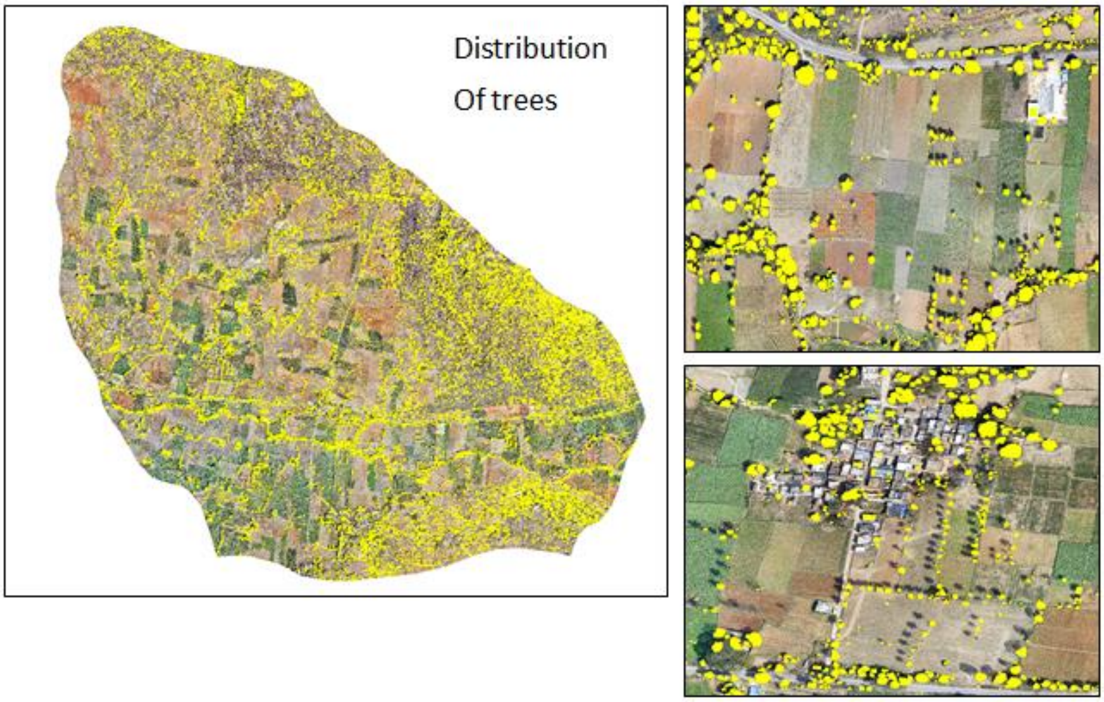


Fig. 7 Tree Segmentation for Niragunupal watershed using Retina net DL model for Niragunupal Watershed and zoomed views.

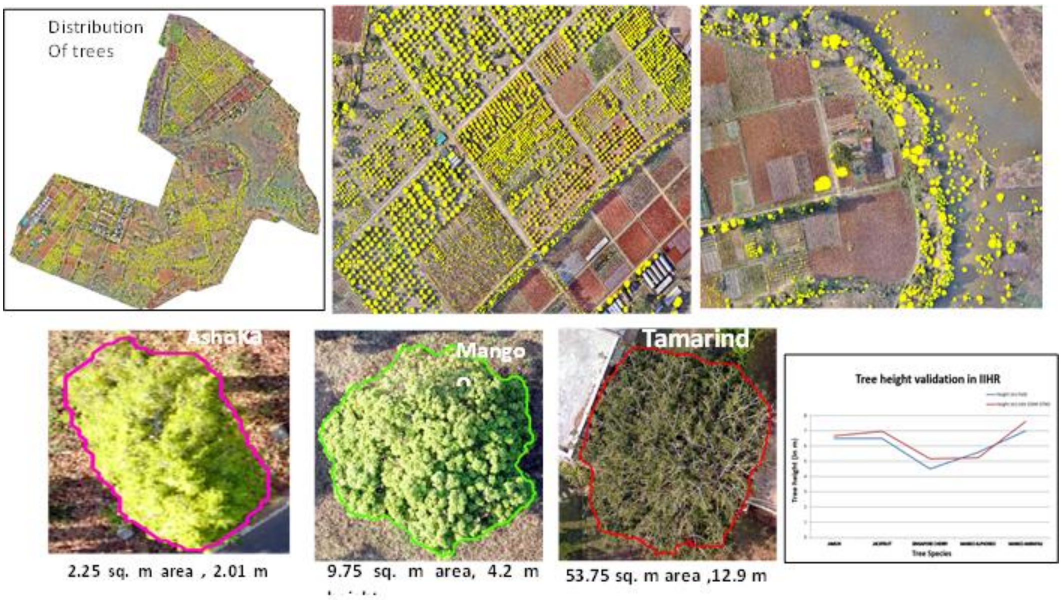


Fig. 8 Validation of DL Model on IIHR Campus, Bengaluru.

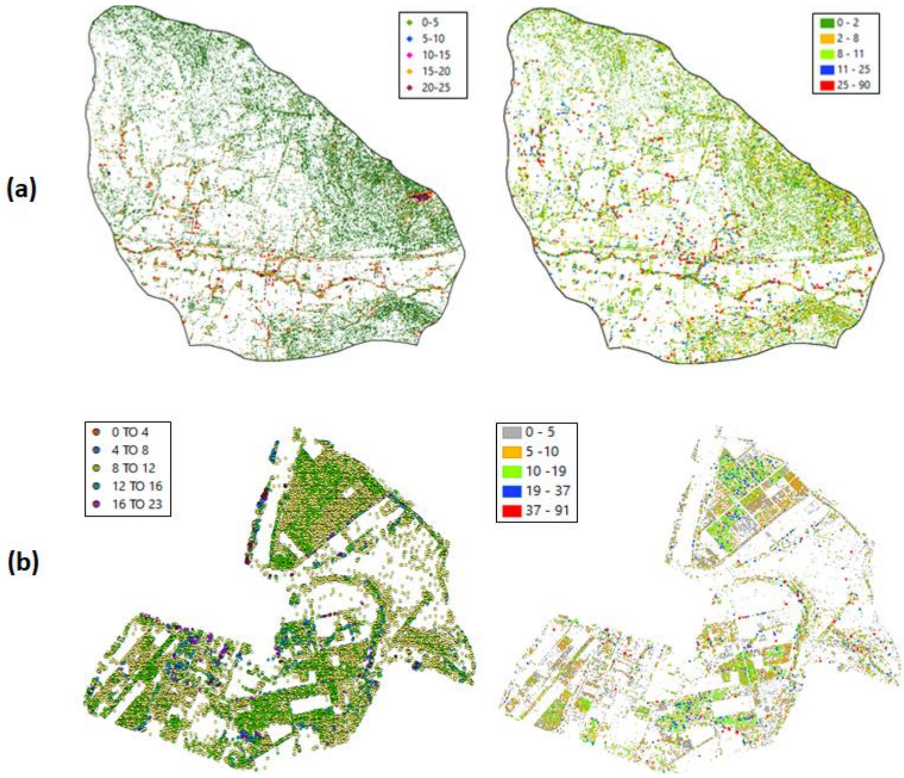


Fig. 9 (a) Niragunupal watershed – tree height and canopy cover (tree height in meters). **(b)** IIHR campus – tree height and canopy cover.

Conclusions and Recommendations

Instance segmentation still remains challenging in the field, such as segmenting small objects, aerial object segmentation, real-time segmentation using drone technology, etc. (Sharma, R., et al 2022). Object detection can be used in many areas to reduce human efforts and increase the efficiency of processes in various fields. From the experiments on our dataset, we obtained an overall classification accuracy of 89.1% for Niragunupal watershed**.** Transfer learning is a very popular technique in deep learning where existing pre-trained models are used for new tasks. Basically, the idea is to take one neural network that we trained for one specific task and use it for another area (IIHR Campus Bengaluru with an accuracy of 85%). The model can further be improved by attempting Ensemble Learning Methods for Deep Learning Neural Networks. Ensemble deep learning methods refer to training several deep learning models and combining some rules to make predictions. An ensemble can make better predictions and achieve better performance than any single contributing model. An ensemble can make better predictions and achieve better performance than any single contributing model. Further, tree layer can be readily used for preparing labelled training samples by assigning the species names in conjunction with collected GT. This study area has over 15 species of trees and shrubs which may be identified with these inputs and form an extended scope of this study.

Acknowledgments

Authors express their gratitude towards Director, National Remote Sensing Centre (NRSC), Hyderabad for the guidance, support and constant encouragement. The support extended by Aerial Survey and Digital Mapping Area, NRSC are also thankfully acknowledged.

References

B. S. Reddy, A. Mallikarjuna Reddy, M. H. D. S. Sradda, T. Mounika, S. Mounika and M. K, (2022). "A Comparative Study on Object Detection Using Retina net," *IEEE 2nd Mysore Sub Section International Conference (MysuruCon)*, Mysuru, India, 2022, pp. 1-6.

Pleșoianu, Alin-Ionut, LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). “Individual tree-crown detection and Species classification in very high-resolution remote sensing imagery using a deep learning ensemble Model." Remote Sensing 12.15 (2020): 2426.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998 Gradient Based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324. https://doi.org/10.1109/5.726791.

Hussain, Mahbub & Bird, Jordan & Faria, Diego. (2018). A Study on CNN Transfer Learning for Image Classification.

Latif Muhammad (2019). Multi-crop recognition using UAV-based high-resolution NDVI time-series. Journal of Unmanned Vehicle Systems 7.3 (2019): 207-218.

Zhen, Zhen & Quackenbush, Lindi & Zhang, Lianjun. (2016). Trends in Automatic Individual Tree Crown Detection and Delineation—Evolution of LiDAR Data. Remote Sensing. 8. 333. 10.3390/rs8040333.

Zhou, Y., Wang, L., Jiang, K., Xue, L., and Yun, T. (2020). Individual tree crown segmentation based on aerial image using super pixel and topological features. J. Appl. Remote Sens. 14:2.

Larsen-Freeman, D. (2013). Transfer of Learning Transformed. Language Learning, 63, pp.107-12

L. Zhang and B. Du, "Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art," (2016) in IEEE Geoscience and Remote Sensing Magazine, vol. 4, no. 2, pp. 22-40,

Sharma, R., Saqib, M., Lin, C.T. et al. A Survey on Object Instance Segmentation. SN COMPUT. SCI. 3, 499 (2022).

Citation

Nagashree, M.K., Satisha, G.C., Rama Subramoniam, S., Satish Kumar, B., Narendran, J., Reddy, M.S.S., Chamundeshwari, D., Premananda, B.D., Hebbar, R., Srivatsav S.K. (2024). Tree Detection and Characterization in Diverse Agroforestry Systems using UAV Imagery and DL. In: Dandabathula, G., Bera, A.K., Rao, S.S., Srivastav, S.K. (Eds.), Proceedings of the 43rd INCA International Conference, Jodhpur, 06–08 November 2023, pp. 109–118, ISBN 978-93-341-2277-0.