**ACSU**

**An Engineering Project in Community Service**

**Phase – I Report**

***Submitted by***

**SRI ABHISHEK MAMIDI- 19MIM10047**

**MOHITH SANKAR- 19MIM10098**

**BOMMISETTY KRISHNAH - 19MIM10090**

**RAJ KAMAL PATEL- 19BCG10030**

**SIDDHU CHELLURU- 19BCY10178**

**GHADIYARAM HANUMANTH SREENIVAS DIXIT- 19MIM10037**

**SANJAY S- 19BCE10265**

**PANDIPATI PAVAN- 19BCE10245**

***in partial fulfillment of the requirements for the degree of***

***Bachlore of Engineering and Technology***



**VIT Bhopal University**

**Bhopal**

**Madhyapradhesh**

**December,2021**



**Bonafide Certificate**

Certified that this project report titled **"ACSU"**is the bonafide work of SRI ABHISHEK MAMIDI- 19MIM10047,MOHITH SANKAR- 19MIM10098,BOMMISETTY KRISHNAH MAANAS- 19MIM10090,RAJ KAMAL PATEL- 19BCG10030 ,SIDDHU CHELLURU- 19BCY10178,GHADIYARAM HANUMANTH SREENIVAS DIXIT- 19MIM10037,SANJAY S- 19BCE10265,PANDIPATI PAVAN- 19BCE10245 who carried out the project work under my supervision.

This project report (Phase I) is submitted for the Project Viva-Voce examination held on …………..

**Supervisor**

**INTRODUCTION:**

Nowadays, the positive effect of trees in cities is widely accepted. Not only do trees influence a city’s climate positively, but they also over several economic benefits.

Trees lower the city’s temperature and therefore the energy consumption of air conditioning by providing shade and increasing evaporative cooling, reduce the chance of floods during severe rains by storing water in their root system, and clean the air by reducing the amount of carbon dioxide during photosynthesis. As urban trees typically grow as stand-alone objects, individual tree health condition data is required. The manual annotation of these individual trees is an inecient, time-consuming, and costly process, usually carried out in situ.

Hence, manual annotation by experts is not scalable to large areas, and the resulting data lacks timeliness. A recently upcoming approach to increase this capacity, is the involvement of citizen and is known as community-based monitoring (CBM).

**Motivation:**

**Objective:**

Keeping the disadvantages of the above situation in mind, it is very important to preserve trees and monitoring is one of the effective measures which can be implemented at a large scale. Which is why we are aiming to build a tree monitoring application using deep learning and google maps.

**Existing Work / Literature Review:**

* **Vision for trees: - Automated** tree delineation has been a recent topic of interest in research, while computer vision and machine learning approaches became more important in recent years. Tree detection, species recognition, and tree health assessment has been tackled using remote sensing data such as multi-spectral aerial or satellite images, hyperspectral data, and dense (full-waveform) Light detection and ranging (LiDAR) point clouds.
* **Tree detection: -** Most of the existing work addressed tree detection from LiDAR point clouds (Lahivaara et al.,2014, Zhang et al., 2014) or from a combination of LiDAR and aerial images (Qin et al., 2014, Paris and Bruzzone, 2015). LiDAR data provides direct height information and 3D-points, which are well-suited to distinguish trees from other objects and the ground. However, the acquisition of LiDAR data requires dedicated flight campaign’s, which are expensive. Lafarge and Mallet (2012) model cities from dense aerial LiDAR point clouds by simultaneously reconstructing buildings, trees, and ground surfaces.
* **Tree species classification: -** Remote sensing data with high spectral resolution or detailed geometrical information (full waveform LiDAR) is suitable to extract physical features for tree species classification. Tree species were automatically classified either from multi-spectral aerial or satellite images (Waseret al., 2011, Pu and Landry, 2012), hyperspectral data (Roth et al., 2015), or dens (full-waveform) LiDAR point clouds (Yao et al., 2012b). Hyperspectral data provides species-specific spectral patterns while full-waveform LiDAR species-specific waveforms (laser reflectance patterns) that depending on the shape of the canopy. Most of the work has been applying a standard classification pipeline encompassing two steps.
* **RegisTree project: -** RegisTree 3 is a collaboration between ETH Z¨urich and the California Institute of Technology (Caltech) investigating the detection and classification of public objects in publicly available imagery from Google Maps. Wegner et al. (2016) aimed to catalog the object of interest at city scale. They proposed a multi-view detection approach that combines information from multiple sources, such as maps, aerial images, and street-view images, by using a CRF. Faster-RCNN (Ren et al., 2015) a state-of-the-art object detection approach, is adopted to detect the object of interest in the street-view and aerial image domain, respectively.
* **Topic of the work (1- 2 Pages)**
* System Design / Architecture
* Working Principle
* Expected Results

**Theoretical principal**

This section provides an overview of CNN and shortly describes the theoretical background of the training process. First, the most important building blocks of CNNs are described.

* **Convolutional Neural Networks (CNNs)**

CNNs are a special type of neural networks well-suited to learn directly from images. Therefore, CNNs do not only learn to solve a task (i.e., classification), but also learn their features directly from the ground truth image data. The core concept behind CNNs is the sharing of parameters. Thus, deeper networks with a larger number of layers can be trained, successfully dealing with the large amount of input data. CNN takes an input volume (i.e. RGB image) and transforms it with a series of layers to an output (i.e. array of class scores in a classification), by modelling a function.



CNN architectures usually follow this pattern

* **Fully-connected layer (FC)**

Fully-connected layers are the standard layers known from the initial idea of neural networks. FC layers take an input array and produce an output array. As the layer name postulates, an FC layer connects all input neurons to output neurons by calculating the scalar product between a learned weight vector and the input array. In CNN architectures, this layer type is usually applied at the top of the network to transform the input to the desired number of outputs

* **Convolutional layer (CONV)**

Convolutional layers are used to implement the core concept of CNNs, In contrast to FC layers, connecting all inputs to all outputs, CONV layers only consist of local connections between input volume and output neurons. Parameter sharing of neurons is implemented as convolution. Therefore, a convolutional layer consists of a set of learnable filters, which have a small spatial extent but have the same depth as the input volume. Therefore, a single filter is a volume of parameters analogous to the weight vector in FC layers. Parameter sharing takes place when the volumetric filter is convolved with the input volume to create an output.

* **Pooling Layer (POOL)**

Pooling layers do not contain any learnable parameters, but are applied to reduce the spatial extent of the input. A common pooling strategy is max pooling, that down-samples the input by convolving a max kernel with the input. A 2 x 2 kernel with stride 2 will reduce the spatial extent by factor 2 whereas the depth will be preserved. Reducing the spatial extent allows to increase the depth by learning more filters per CONV layer.

**Methodology**:

The purpose of this thesis is to investigate a system that can automatically predict the health condition of individual trees from street-level imagery. This task is approached by training deep convolutional neural networks (CNNs) using automatically generated ground truth data. This section first describes the generation of the required ground truth image data by combining available tree inventories with images from Google street-view. Second, the chosen CNN architecture based on the VGG-16 is presented.

Extracting image patches from Google street-view panoramas

For each ground truth tree location with a given annotation date, a corresponding street-level image was downloaded from Google street-view by carrying out the following sequence of operations

1. Crawling metadata of the nearest seed panorama:
2. Crawling metadata of linked panoramas
3. Filtering linked panoramas by distance Filtering linked panoramas by distance
4. Selecting the panorama of interest
5. Downloading image data
6. Extracting image patch at tree location
7. Assigning image label from tree inventory annotations

**CNN for tree health classification**

The chosen approach for tree health classification adopts the VGG-16. This network is used because the training for tree health condition can profit from a pre-trained tree species classification model This pre-trained model is expected to be a proper starting point, as it was trained also on tree images from Google street-view. Therefore, the pre-trained features from the species classifier are transferred to the new task of health classification.

**CONCLUSION: -**

Within this work, a CNN based approach has been presented to successfully predict an individual tree’s health condition given a street-level image. The best performance was achieved by starting from a tree species classifier and fine-tuning the model with a limited number of survey images annotated by experts (166 images per health category). By predicting a binary health condition (good vs. poor) an overall accuracy of > 80% was reported. Furthermore, the direct prediction of the level of health on a 3-point scale achieved > 60% overall accuracy.

The achieved performance of > 80% overall accuracy reveals that standard street-level RGB images provide scient evidence to train the adopted CNN for tree health prediction. The performance depends on the consistency of the health labels as well as on the quality of the image data. Training from noisy crowd-sourced tree health data in combination with images from Google street-view resulted in slightly lower performance ranging from 0.73 to 0.78 OA, depending on the temporal agreement criteria between image and annotation date. The performance was improved by relaxing the temporal agreement criteria, which introduced additional noise, but resulted in a larger amount of ground truth images (>4,000 training samples per class).

* **Reference:**

Deep learning and Google Maps for tree monitoring