Precision Agriculture (2021) 22:1617–1633
https://doi.org/10.1007/s11119-021-09803-0

Object-level classification of vegetable crops in 3D LiDAR point cloud using deep learning convolutional neural networks

Reji Jayakumari[1] · Rama Rao Nidamanuri[1] · Anandakumar M. Ramiya[1]

Accepted: 9 March 2021 / Published online: 26 March 2021

Abstract
Crop discrimination at the plant or patch level is vital for modern technology-enabled agriculture. Multispectral and hyperspectral remote sensing data have been widely used for crop classification. Even though spectral data are successful in classifying row-crops and orchards, they are limited in discriminating vegetable and cereal crops at plant or patch level. Terrestrial laser scanning is a potential remote sensing approach that offers distinct structural features useful for classification of crops at plant or patch level. The objective of this research is the improvement and application of an advanced deep learning framework for object-based classification of three vegetable crops: cabbage, tomato, and eggplant using high-resolution LiDAR point cloud. Point clouds from a terrestrial laser scanner (TLS) were acquired over experimental plots of the University of Agricultural Sciences, Bengaluru, India. As part of the methodology, a deep convolution neural network (CNN) model named CropPointNet is devised for the semantic segmentation of crops from a 3D perspective. The CropPointNet is an adaptation of the PointNet deep CNN model developed for the segmentation of indoor objects in a typical computer vision scenario. Apart from adapting to 3D point cloud segmentation of crops, the significant methodological improvements made in the CropPointNet are a random sampling scheme for training point cloud, and optimization of the network architecture to enable structural attribute-based segmentation of point clouds of unstructured objects such as TLS point clouds crops. The performance of the 3D crop classification has been validated and compared against two popular deep learning architectures: PointNet, and the Dynamic Graph-based Convolutional Neural Network (DGCNN). Results indicate consistent plant level object-based classification of crop point cloud with overall accuracies of 81% or better for all the three crops. The CropPointNet architecture proposed in this research can be generalized for segmentation and classification of other row crops and natural vegetation types.

Keywords Object based crop classification · LiDAR point cloud · Deep learning networks · Crop height modelling · 3D segmentation

✉ Rama Rao Nidamanuri
rao@iist.ac.in

[1] Department of Earth and Space Sciences, Indian Institute of Space Science and Technology, Thiruvananthapuram 695 547, India

Introduction

Remote sensing data have been widely used for crop identification and mapping. Modern crop management practices require knowledge of within-field spatial variance of the crop attributes. Various studies have demonstrated the possibility of high-resolution crop classification using multispectral (Handique et al. 2017; Ozdarici-Ok et al., 2015) and hyperspectral remote sensing data (Avola et al. 2019; Zhang et al. 2016). However, the scale of crop discrimination is limited to discriminating crop type, at the most, at field level. In general, the optical imaging sensors respond mainly to the spectral-horizontal attributes of crop fields. The fast-evolving LiDAR remote sensing, in principle, records vertical-structural attributes of crop fields at plant to patch level and is a potential technology for plant-level crop discrimination (Murray et al. 2020). LiDAR (light detection and ranging) remote sensing is a 3D laser scanning technology capable of providing both the horizontal and the vertical features of surface objects. Recent trends indicate that LIDAR remote sensing is one of the most promising technologies that can be operated in industrial, law enforcement, and natural resource management applications (Bellakaout et al. 2016; Sun et al. 2017). LiDAR remote sensing data, generally termed as LiDAR point cloud, records the height of surface objects as the primary measurement entity, which in turn makes it possible to extract various 3D geometrical features. However, due to the typical scene-based sparseness and unstructured nature of the LiDAR point cloud (Eckart et al., 2018; Lawin et al. 2017) establishment of a surface object's identity in LiDAR point cloud requires applying a host of processing and 3D point cloud labelling algorithms.

Deep learning neural network has been evolving as a general method for prediction and labelling of various types of data and applications. The Convolutional Neural Network (CNN) based deep learning model has recently been successfully used for semantic segmentation of 2D imagery (Liu et al. 2019; Varfolomeev et al. 2019). While the Deep Convolutional Neural Network (DCNN) model is amenable to directly ingest a rasterized point cloud as input, the unstructured nature of the 3D point cloud of natural landscapes limits the direct application of DCNN for classification or segmentation in a 3D perspective. PointNet (Qi et al. 2017a) is a broad-based improvement in the DCNN framework that offers a unified architecture for direct ingestion of LiDAR point cloud into deep learning architecture. The distinct feature of PointNet structure is the usage of shared multi-layer perceptron (MLP) which converts the input into a set of features in feature space for each point. For semantic segmentation, in addition to the generation of global features from the max-pooling layer, the shared MLP is used to produce per-point features. However, PointNet architecture fails to capture the local information of its immediate neighbourhood. As an improvement, in PointNet++ (Qi et al. 2017b), an extension of the PointNet, introduced the possibility of extracting neighbourhood information in a hierarchical feature learning fashion. Improvements of extending the architecture of PointNet to processing of point cloud of individual buildings have also been reported (Soilán et al. 2019). Semantic segmentation of LiDAR point cloud of vegetable crops is a challenging task due to the complex object-background interactions and thresholds of height differences required for normalization and filtering of LiDAR point cloud. There are no studies that assessed the potential of CNN based deep learning approaches for semantic segmentation of crops. The objective of this research is the assessment of high resolution LiDAR point cloud for object-based plant level classification of three important vegetable crops (tomato, cabbage, and eggplant) using deep neural network based method. Apart from exploring the potential of deep CNN for 3D classification of crops, the work presents a methodological framework for direct segmentation and classification of 3D point cloud. In contrast to the use of typical scientific method, which specifies hypotheses, the research approach used in this work is a heuristic engineering method composed of defining the problem, proposing a solution and validating the solution (Koen, 1988). Accordingly, a CNN based deep learning methodology was developed and implemented on the LiDAR point cloud for semantic segmentation and classification of the vegetable crops considered. The performance of the proposed methodology was assessed by comparison of the results with the ground truth information.

The proposed methodology includes adaptation of an enhanced deep convolutional neural network capable of combining local and global geometrical features for semantic segmentation of vegetable crops of crops, and fine-tuning the hyper-parameters to increase efficiency in predicting semantic labels in a lesser number of training iterations. High-density LiDAR point cloud of tomato, cabbage and eggplant crops grown on the experimental plots laid out as per a factorial experiment were acquired using a long-range high density terrestrial laser scanner. The pre-processed point clouds were segmented and classified in 3D perspective for plant-level crop classification using a modified version of the PointNet deep convolution neural network (CNN) model. The results validated using cross-validation and also compared with two popular models suggest consistent crop discrimination at the plant level.

Methodology

Data: crop growing experiment and TLS point cloud acquisition

A multi-crop crop growing experiment was set up on the farms of the University of Agricultural Sciences, Bengaluru from January–May, 2017. Figure 1 shows the general layout of the crop growing experiment. Three different vegetable crops—cabbage (Brassica

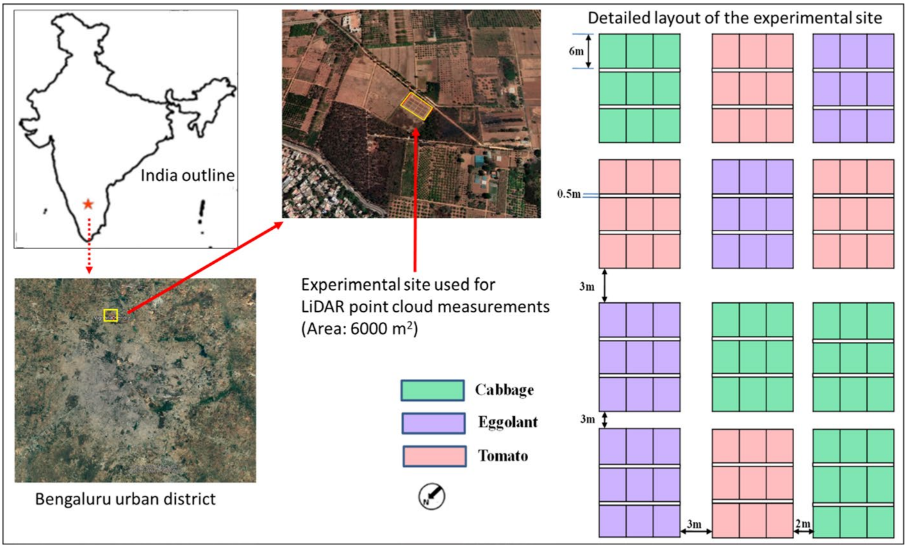

Fig. 1 Study site location and experimental layout of the crop growing experiments for TLS point cloud acquisition (imagery: extracted from Google Earth)

oleracea L.), eggplant (Solanum melongena L.) and tomato (Solanum lycopersicum L.) were grown on plots laid as per a factorial experiment. These are some of the widely cultivated crops in southern India. The cultivars chosen are the typical varieties that farmers use in Karnataka, India. The crops were watered through a drip irrigation system.

The experimental set up contained 12 plots of 12 m × 18 m size, each of which was further sub-divided into three subplots of size 6 m × 12 m with four replications, resulting in a total of 36 subplots. Three levels of chemical (N) fertilizer: low, medium, and high were supplied randomly to the sub-plots. The rate of N fertilizer for different treatment levels was calculated on a per-row basis. Corresponding to the 'medium' N level in this work, 46 kg N ha^{-1}as urea for tomato, 60 kg N ha^{-1} for cabbage, and 50 kg N ha^{-1}for eggplant. 'High N' referred to 50% more and 'low N' to 50% less compared to the standard recommendation. Apart from the N fertilizer, phosphorus (P) and potassium (K) at the rate of 17.5 kg P ha^{-1}and 19.9, 41.5, and 16.6 kg K ha^{-1}for tomato, cabbage, and eggplant were applied at sowing. The crops were grown on soil ridges and watered through a drip irrigation system, which contained a pipe network laid on the soil ridges and positioned closer to the plants. As a management practice for minimizing the lodging, tomato plant rows were supported by sticks.

A terrestrial laser scanner (TLS) mounted on a movable-height adjustable tripod (Faro Focus350^{S}, USA) was used for capturing the 3D point clouds of the crops grown. The TLS, which operates in the near-IR region of the electromagnetic spectrum, was tuned and programmed to acquire data at an average point density of 20,000 points/m^2 and 4 mm point spacing. To cover the entire agricultural field and record returned pulses from most of the crop plants, the TLS instrument was positioned at multiple locations in the study area. The crop-target laser pulses from 16 scans were co-registered to generate LiDAR point cloud of the experimental layout. An outline of the methodological process flow is shown in Fig. 2.

Point cloud processing and reference data generation

Point cloud filtering

To enhance the crop plant geometrical representation and minimize the inherent surface interference, a ground filtering procedure was applied on the point cloud. The ground filtering results in the separation of point cloud into two categories, ground and non-ground. For this, a triangulation-based algorithm suggested by (Axelsson, 2000) was chosen. In this algorithm, an iterative TIN (triangular irregular network) is constructed based on the seed points selected based on the heuristically identified grid size. During each iteration, points are added to the TIN surface based on the distance to the plane surface and subtended node

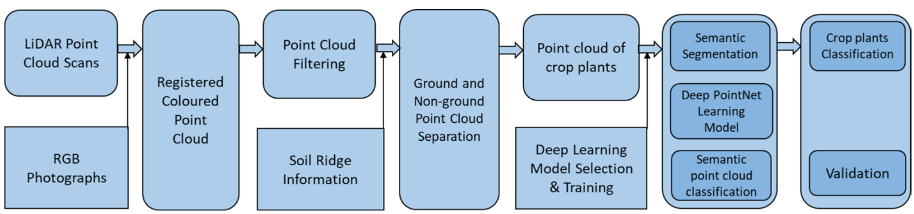

Fig. 2 Outline of the methodological process flow

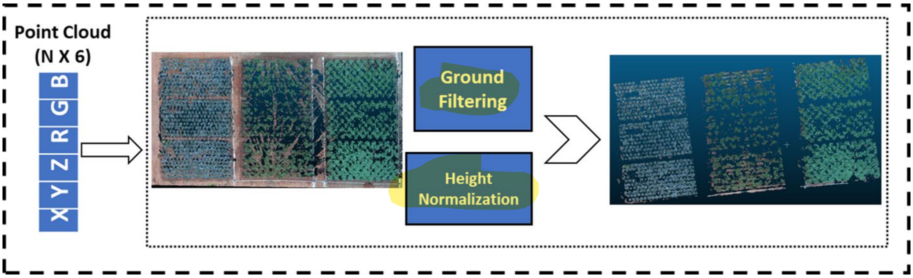

Fig. 3 Visualization of spatial filtering of TLS point cloud for segregating point cloud of crop plants

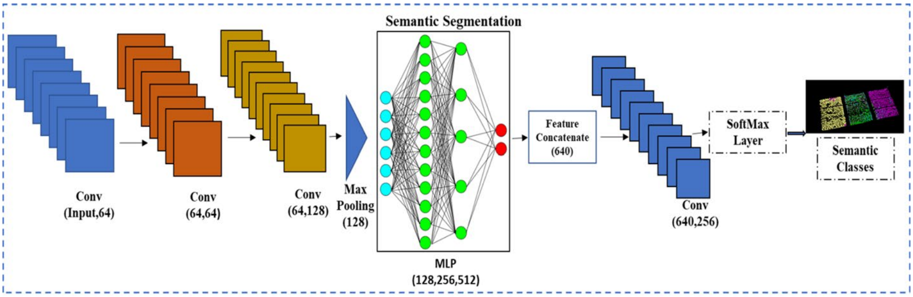

Fig. 4 The deep neural network architecture for semantic crop segmentation

angles of TIN facets. This process filters out the ground points retaining only the non-ground points used in further analysis.

At some instances in the experimental plots, the plant heights were only comparable to that of the soil ridges and plant support sticks. Hence, there was a possibility of the loss of point cloud of crops in the ground filtering processes. This problem was mitigated by optimizing the ground filtering process to retain the point cloud of both ground and non-ground wherever the confusion persisted. The height of the crops was normalized to the local ground surface of the experimental plots. Figure 3 presents a visualization of the process flow of the point cloud filtering and the resultant point cloud. Based on expert analyses, six information classes were identified for the 3D segmentation and classification—three types of crop plants and three types of field infrastructure (pipes, sticks, and residual soil ridges) elements for classification. While our aim was to classify the point cloud for crops only, the categories of field infrastructure were also included in the classification to account for the overlapping and height-confusion in the laser returns of crop plants and field infrastructure.

Segmentation and classification using deep learning methods

An outline of the deep learning architecture adapted for point cloud classification is shown in Fig. 4. The methodological architecture has two functions: segmentation, and classification. For the segmentation, the PointNet deep learning architecture was modified to enhance object representation and learning, and feature derivation for handling the discrete TLS point cloud. The PointNet is a recent deep learning neural network model with a unified architecture for various applications in segmentation, object classification, scene semantic parsing using point cloud (Qi et al. 2017a). The PointNet network learns a collection of points that selects representative points by randomly picking a set of points and visualizing the activation regions for them. Classification is done by the fully connected layers of the network by aggregating the optimal values into global descriptors for predicting per point labels.The PointNet has been applied for successful indoor scene segmentation, urban scene segmentation, using the point cloud produced from Kinect and airborne laser scanner (Lowphansirikul et al. 2019). In contrast, the TLS point clouds of crop plant objects lack regular and well quantifiable shape and its related geometrical attributes. Further, the plant height and ground surface profile changes are comparable with the minimum measurement range of the laser scanner. Combined with the possibility of ground interference and the shorter plant height, 3D segmentation of crops is a challenging task and the inferences and methodological procedures from other application domain may not be directly applicable. Therefore, the architecture of the PointNet model for enhancing the learning of local and overall crop structure in the TLS point cloud was modified and named the improved model as 'CropPointNet' in this work.

A k-dimensional LiDAR point cloud represented as $X=\{x_1, x_2,\ldots x_n\}\subseteq R^k$, where $k=6$ represents the 3-dimensional coordinates of the point cloud (x,y,z) and colour intensity (RGB) of each point is the input for the deep CNN model. The rationale behind the exclusion of the normalized position coordinates, unlike the PointNet, is that in the case of 3D agricultural LiDAR point cloud, this mode of representation does not enhance the semantic interpretation of LiDAR point cloud. The point cloud was fed into the CropPointNet model by random sampling of the point cloud.

Similar to the image domain, where the convolution process extracts the features building up a denser feature map, the point cloud convolution in the CropPointNet generates a feature map with more features extracted in the forward pass. The input point clouds were given to a convolutional neural network that produces 64-dimensional features. These features were further given to the convolutional layer resulting in 512-dimensional feature extraction. In each layer before the activation function, batch normalization (BN) was performed (as shown in Eq. 1) which can work around the covariance shift for reducing the overfitting of the data and dispense an implicit effect of regularisation. When the pointwise convolution occurs, the output of each mini-batch of the previous layer was scaled with a mean-centred towards the standard deviation. Thus, batch normalization adds standard deviation and mean as two trainable parameters to the layer. During training time, a batch normalization layer calculates the mini-batch mean (μ_{MB}) and variance (σ^2_{MB}). Accordingly, the layer inputs were normalized using the expression

$$\overline{u_i} = \frac{u_i - \mu_{MB}}{\sqrt{\sigma^2_{MB} + \in}} \tag{1}$$

where, u_i is the mini-batch during training. Finally, a scale and shift is applied to Eq. 2 to obtain z_i, the output from the layer i,

$$z_i = \gamma \overline{u_i} + \beta \tag{2}$$

with the γ and β learned during the training process with the other parameters in the network. To handle non-linearity in the network, a computationally efficient activation function -rectified linear unit (ReLU) was used, defined as

$$f(x) = \max(0, x) \tag{3}$$

Mathematically, the CropPointNet network can be represented as:

$$Y_i = \mathrm{Conv}^{(\mathrm{i})}\left(\mathrm{X_j}\right) \in \mathrm{R^K} \tag{5}$$

$$Y_i^{(F_i)} = \max\left(0, BN\left(Y_i\right)\right) \tag{6}$$

where Conv refers to the convolutional layers, i refers to the i$^{\text{th}}$ layer of the network, j denotes the j$^{\text{th}}$ 3D point of the point cloud, BN refers to the batch normalization, max is the activation function after the convolution and $Y_i^{F_i}$ refers to the feature map generated from the i$^{\text{th}}$ layer. The CropPointNet network was designed to learn both the point level features and the features obtained from the pooling layer. The max-pooling method was used to project the most stimulated presence of a feature. Following this, the MLP (multi-layer perceptron) maps the features to the next level of 512 dimensions, where the same weights are shared. This is the subsequent step for the extraction of global features from the point cloud. The point level features and the features from the max pooled layer were concatenated together to obtain rich contextual and denser feature representations. These concatenated features form a robust map for further semantic interpretation of the data. Finally, all the points in the dataset were labelled with a SoftMax activation layer, which varies from 1 to 6.

Comparative performance

Comparing the relative performance of the CropPointNet architecture, the point cloud was classified using two other methods: (i) the standard PointNet architecture and, (ii) Dynamic Graph CNN for Learning on Point Clouds (DGCNN) (Wang et al. 2018). DGCNN is a dynamically updated graph neural network constructed based on the neighbourhood graph to exploit the inherent geometric structure of the 3D point clouds. DGCNN performs operations similar to convolution on points defining neighbourhoods that are connected by graph edges. An edge convolution operator called EdgeConv takes care of the translation-invariance. The model learns to cluster points semantically while updating the graph at each layer. For training the DGCNN, 0.01 and 0.92 were chosen as the learning rate and momentum for batch normalization, respectively. Based on heuristics the batch size as 16, the number of epochs as 50, and the number of iterations for feature learning as 20 were calculated.

Validation and statistical test of significance

Ground truth imagery was generated based on the expert-guided segmentation categorizing the point cloud into six classes—cabbage, tomato, eggplant, soil, stick, and irrigation pipe. The training data were created by a random sampling of the ground truth LiDAR point cloud. To make the model robust to any point perturbations during the model training, the data samples were augmented and opted for jittering of the coordinates. The performance of the model was validated using the k-fold cross-validation with k = 5. During the training, the network was optimized using adam optimizer (Kingma & Ba, 2014) with a learning rate of 0.001, data size for minimum-batch of 16, the momentum of 0.92, and 20 epochs of iteration. Based on the results of cross-validation, accuracy assessment metrics, overall accuracy and per-class accuracy were computed. Besides, the robustness of the model predictions was also assessed, accounting for imbalance in the class distributions, using non-parametric accuracy metrics—Precision, Recall, and F1- Score given by

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{6}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{7}$$

$$\text{F1-Score} = 2\left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}\right). \tag{8}$$

Precision is the ratio of correctly predicted ground truth labels to the total ground truth labels. Recall calculates the ratio of predicted ground truth labels to the total number of ground truth labels. F1-Score provides a balance between Precision and Recall in the case of imbalance in the class distribution. A model with nominal values of 0.7 for Precision and 0.5 for Recall and F1-Score is considered to be a good performing.

Besides, the statistical significance of the accuracy estimates was also assessed using a z- test (Johnson et al. 2000), computed as

$$z = \frac{A_1 - A_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \tag{9}$$

where, A_1 and A_2 are the overall accuracy estimates computed for a pair of classification methods.

The variance of the estimated accuracy is:

$$\sigma^2 = \frac{A(1 - A)}{n} \tag{10}$$

where n is the number of test samples.

Results and analysis

Results of the object-based point cloud classification of the vegetable crops using the CropPointNet deep learning architecture are presented in Fig. 5 and the corresponding accuracy estimates in Tables 1 and 2.

As evident from Fig. 4, the individual plant level discrimination of the tomato, cabbage, and eggplant is unambiguous. The crop objects in the point cloud were classified with the best overall accuracy of 81.5% by the CropPointNet model. The PointNet and DGCNN models yielded an overall accuracy of 55% and 66.5%, respectively (see Table 1). However, the quality of discrimination varied by the crops and the model. Cabbage and eggplant crops were discriminated accurately by the CropPointNet and DGCNN models (see Fig. 5a, c). The discrimination of tomato crop was very low by all the three models, despite

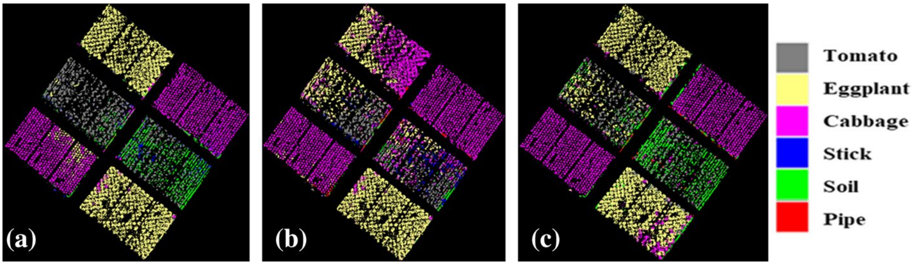

Fig. 5 Semantic segmentation using **a** CropPointNet, **b** Pointnet, and **c** DGCNN

Table 1 Per-class accuracy estimates (%) for both the crop and non-crop categories in the LiDAR point cloud from all the three deep learning neural network models

Class	CropPointNet	PointNet	DGCNN
Tomato	65	60	61
Eggplant	88	69	83
Cabbage	91	72	82
Stick	79	21	00
Soil	72	41	81
Pipe	70	50	76

Table 2 Overall accuracy of the classification of point cloud for vegetable crops from all the three deep learning neural network models

Models	Overall Accuracy (%)
CropPointNet	81.5
PointNet	55.2
DGCNN	66.5

very high inter-crop discrimination. Most of the tomato crop discriminations are underestimations with the apparent confusion with residual soil ridges (see Fig. 5b). Results from the PointNet model show a substantial misclassification wherein the cabbage crop was overestimated at the cost of eggplant followed by tomato.

Cabbage and eggplant crops are discriminated with the highest accuracy of 91% and 88% respectively (see Table 2). For the non-crop classes (residual soil ridges, pipe, and stick) the CropPointNet model yielded the relatively higher per-class accuracies (about 70%) even though substantially lower than the accuracy of the crops. The discrimination of non-crop classes by the PointNet model is negligible (see Table 2). Cabbage and eggplant were classified with a marginal difference of accuracy (3%). However, the Precision scores of cabbage and eggplant differed substantially (12%, and 21%) from the CropPointNet model and PointNet model respectively. Eggplant and cabbage were classified with Precision of 91% and 79% respectively. Compared to the accuracy metric, the Precision score matches closely with the quality of crop classifications accounting for the false positives in the cabbage (see Fig. 5). The Recall and F1-Score substantiate the matching of crop classifications with the ground truth data.

Table 3 Object-to-object comparison of CropPointNet model predictions and ground truth for accuracy estimation—Precision, Recall, and F1-score

Class	Precision	Recall	F1 Score
Tomato	0.79	0.85	0.81
Eggplant	0.91	0.88	0.89
Cabbage	0.79	0.91	0.85
Stick	0.79	0.79	0.79
Soil	0.71	0.76	0.73
Pipe	0.91	0.70	0.79

Overall, the scores of Precision, Recall, and F1-Score are very closer to the overall and per-class accuracies estimated from the *k*-fold cross-validation. Compared to the PointNet model, results from the CropPointNet model are much superior by magnitude and consistency of all the accuracy metrics across the crop and non-crop classes (see Tables 1 through 3). The results of the statistical z-test computed between the performance of the CropPointNet with PointNet and DGCNN models yielded a z-score of 6.31 and 3.81, respectively. At the 95% confidence interval, the computed z-scores are higher than the tabulated value of 1.96, confirming that the superior classification performance the CropPointNet is not by chance. This observation confirms the potential impact of the point cloud sampling strategy introduced in the training stage of deep learning networks in the CropPointNet architecture.

The discrimination of the non-crop geometrical objects (return pulses from irrigation pipes, support sticks, and residual soil ridges) from the crop plants in the scene is a vital requirement for automatic crop scene analysis. Figure 6 visualizes the segmentation of a sub-plot for each crop. The crops are segmented at plant level exhibiting distinct separability with other geometrical features such as support sticks. Segmentation results from the CropPointNet model (Fig. 6a) are unambiguous compared to the relatively confused discrimination of tomato plants (Fig. 6b, c).

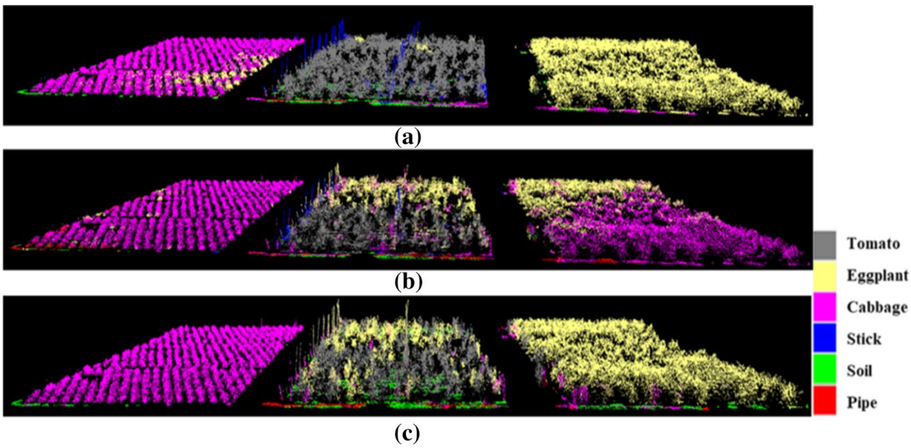

Fig. 6 Visualization of the crop segmentations at sub-plot level depicting the separation of crop plants from non-crop geometrical features from the models **a** CropPointNet, **b** PointNet, and **c** DGCNN

Spatial conformity of the predicted crop plant objects

The statistical accuracy metrics analyses for assessing the performance of the crop discriminations show consistently good accuracy in the point cloud classification for crop discrimination at the semantic point labelling level. However, as the deep learning model architectures implemented in this research are object-based segmentation and classification approaches, in principle, it is possible to evaluate the accuracy by matching at crop plant object level as well. The level of completeness and correctness of the object-based labelled crop plants was assessed using a matching area-based accuracy metric, Intersection over Union (IoU) expressed as

$$\mathrm{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{11}$$

where FN is false negative, FP is false positive, and TP is true positive for the respective information classes considered. The mean IoU (mIoU) was computed from the IoU values of all the six classes. The IoU values of the crop and non-crop information classes are shown in Table 4. The IoU values of each information class and the mIoU exceed the threshold (Meng et al. 2019) IoU value indicating the consistency of plant-to-plant level classification results.

Discussion

Technological approaches for plant-level crop discrimination and phenotyping are vital for precision crop monitoring. Satellite remote sensing data has been extensively used for crop type mapping over large areas. Depending upon the spatial resolution, generally, the crop type maps are generated at a field or coarser scale. Rapid developments in the miniaturization of remote sensors (multispectral imaging sensors and 3D LiDAR scanners) and drones and other terrestrial senor platforms have given rise to the on-demand within field/patch level remote sensing data acquisition capability. This evolving capability offers farmers geo-tagged information on crops at individual plant/patch level and initiate targeted management practices. Remote sensing data acquisition from a terrestrial platform (ground and drone) is especially valuable in countries like India where the average farm size is only 1 ha. Due to the potential of offering structural information of crop plants, technologies for processing and analysing LiDAR point cloud acquired from terrestrial platforms are vital for remote sensing applications in areas where farm sizes are small and the crop type diversity is very high.

Table 4 Object-to-object comparison of the reference PointNet model predictions and ground truth for accuracy estimation—Precision, Recall, and F1-score

Class	Precision	Recall	F1-Score
Tomato	0.69	0.60	0.64
Eggplant	0.48	0.69	0.56
Cabbage	0.38	0.90	0.54
Stick	0.84	0.21	0.33
Soil	0.95	0.41	0.57
Pipe	0.72	0.50	0.59

The terrestrial laser scanner (TLS) has promising applications in agriculture. Semantic segmentation and classification in 3D perspective are common processes in the application of LiDAR point cloud data for crop discrimination. An improved deep learning convolutional neural network architecture, 'CropPointNet' for the segmentation and classification of TLS point cloud for the discrimination of three different vegetable crops (cabbage, tomato, and eggplant) has been implemented in this work. The LiDAR point clouds were acquired using a terrestrial laser scanner on the reference crop growing plots established on the experimental farms of the University of Agricultural Sciences, Bengaluru. The raw point clouds were spatially filtered for segregating the point cloud of the crop plants. The unorganized and discrete point cloud of the crops in the original 3D perspective were processed for semantic segmentation and classification of the vegetable crops at the plant level. This was implemented by adapting and improving a deep learning convolutional neural network (CNN) architecture (PointNet) developed for the segmentation of common indoor objects with low-density point cloud generated in a typical computer vision application. The deep CNN model, "CropPointNet" has improved training strategy for handling high-density point clouds such as the TLS point cloud used in this study and compact-guided feature mapping architecture for faster convergence and error handling ability. The performance of the 3D semantic object-based classification of the crops was assed rigorously using a host of statistical, and non-parametric accuracy estimation measures both at the overall and per-class level. Besides, the results from CropPointNet model were compared with that of two deep CNN models PointNet and DGCNN. Results suggest the possibility of plant-level discrimination of the crops chosen with an overall accuracy of 81%. The accuracy of the crop discrimination by the CropPointNet model is superior and statistically significant. At the individual crop level, the soil and plant structure substantially influence the accuracy and spatial conformity of the crops. Both the CropPointNet and DGCNN models have consistently discriminated the cabbage and eggplant crops with accuracy close to 90%. As evident from Fig. 5 and 6, a point cloud of the plots of tomato suffered a massive loss of point cloud in the ground filtering process. As a result, the quality of tomato crop classification is relatively inferior, even though the CropPointNet model performed relatively better compared to the PointNet and DGCNN models. Relating to the state-of-the-art in this evolving domain of remote sensing technologies for plant-level segmentation and classification the significance of the results is described in the following sub-sections.

There have been some works in the literature on the classification of crops using the LiDAR point cloud. Weiss et al. (2010) and Weiss and Biber (2011) reported laboratory experiments on the plant species classification in 3D LiDAR point cloud using multiple supervised machine learning algorithms. They reported the best overall accuracy of 98% from the Logistic Model Trees algorithm. In a similar study, Paulus et al. (2013, 2014) analyzed 3D LiDAR point cloud for the classification of different plant organs (leaf, stem) of grapevines and barley plant using a statistical surface-feature approximation of histogram using a robust estimation technique, RANSAC. These studies suggest the successful discrimination of plant organs with an accuracy of about 96%. Similar to the TLS instrument used for data acquisition used in this experiment, Murray et al. (2020) quantified plant structure of orchids. They suggest superior results from TLS point cloud and are better than the results from proximal photogrammetry. In Jin et al. (2018), the authors applied a faster region-based convolutional neural network (R-CNN) based deep learning model for segmentation of the TLS point cloud of maize plants. The methodology employed a region growing procedure on the 2D depth imagery obtained from the conversion of the 3D point height into intensity before application of the R-CNN deep learning model for segmentation. The best estimates of Precision and Recall estimates are about 93%. Despite using different methods of classification and crop type, these studies have demonstrated the successful plant level segmentation and classification of crops using LiDAR point cloud with accuracies comparable or slightly better than our results. However, the principal difference and contribution of this research are the comprehensive experimental setup and evaluating a potential automated method for object-based crop classification. The state-of-the-art studies, including the afore-cited, have acquired point cloud data under a controlled experimental setup—mostly indoor or laboratory-controlled with smooth soil surfaces and involving only one or a few crop species or plant organs. Further expanding the state-of-the-art, this study has been carried out on multiple crops under open natural environmental conditions subjected to the crop growth, soil, and surface landscape variability common to operational regional agricultural practices. Results from this work, thus, present a broader spectrum of the potential and challenges of applying 3D LiDAR remote sensing data for object-based crop classification from a realistic landscape perspective.

Spatial filtering and crop point cloud generation

Unlike the indoor scene dataset, the crops are affected by various outdoor factors such as terrain undulation, soil background, wind, and other agricultural infrastructure elements. Therefore, the data has to be pre-processed before ingestion into the deep neural network for training. The complex tasks in pre-processing are the ground filtration and establishment of a normalized surface. The nature of soil preparation for planting crops determines the critical separation line between soil and plant point cloud. For example, the presence of soil ridges and furrows in this study has created a continuous undulation of the soil surface, complicating the ground filtering process and demanding human intervention. When the raise of the soil ridges is comparable to the height of crops, the ground filtering process leads to the loss of point cloud of plants. It has been observed that the filtering approach has wrongly identified about 15% of the point cloud of the crop plants as ground points. Additional features that reduce the quality of point cloud are the plant or patch level agricultural infrastructure such as drip irrigation pipes and support sticks for minimizing plant lodging. This support system for tomato and lack of distinct height profile with crop plant has enhanced the overlap in the height of crop plants and non-crop geometrical features. Faithful retrieval of the point cloud of the crops in the filtering process has yielded only about 50% of the point cloud for tomato. The lack of distinct canopy architecture of the tomato crop at the fully grown stage and the reduced density point cloud of tomato crop are the primary reasons for the low accuracy of the crop discrimination. However, the relatively consistent and higher scores of the non-parametric accuracy metrics (Table 5) indicate two distinct features of the crop classification. First, the reasonably well discrimination of the crops indicated by the confusion matrix based the overall accuracy is supported with a similar level of perception. Second, in contrast to the relatively low accuracies showed (Table 1), the quality of discrimination of the non-crop classes is also high (0.79 and better for some classes) and comparable to that of the crops. Compared to the estimate of sampling-based computation of the accuracy metrics (Table 1), estimates from the plant-to-plant matching based accuracy measures emphasize the number of correct predictions. The apparent differences in the accuracy estimates of crop and non-crop categories are due to the uncertainties in the ground filtering.

The estimates of the IoU which quantify the crop plant-to-plant level conformity between the predicted and ground truth suggest relatively higher discrimination of

Table 5 Estimates of the per-class and mean IoU of object-oriented classification of vegetable crops. Values greater than 50% indicate consistent matching of predictions of crop plants with ground truth

Class	CropPointNet	DGCNN	PointNet
Tomato	80.66	70.71	59.74
Eggplant	81.58	80.42	68.07
Cabbage	83.54	90.47	74.27
Stick	65.88	0.00	25.65
Soil	63.68	74.01	41.26
Pipe	61.32	70.10	48.68
mIoU	**72.78**	**64.28**	**52.94**

the crops with the mean IoU (mIoU) scores exceeding 50% by all the three models (see Table 4) (72.78%, 64.28%, and 52.94% respectively). However, by the stability in the discrimination of both crop and non-crop information classes, the CropPointNet model meets the accuracy thresholds (see Table 4). The substantial changes in the accuracy of the tomato crop suggest the need for choosing an appropriate discrimination model for crop classification under skeletal point density.

Role of crop plant canopy shape

The ability of deep learning models to predict crop patterns at the scale of individual plant objects in the 3D point cloud has also been affected by the shape of crop canopy (see Fig. 5, 6). While the apparent uniform spatial distribution and higher accuracy of cabbage and eggplant suggest fair discrimination, the relatively lower classification accuracy and the substantial gaps in the spatial distribution of the tomato crop indicate the models' sensitivity to the crop canopy. Cabbage plant has a specific geometrical structure with a compact spherical mass of smooth or dense-leaved heads enfolded over each other. Due to this, all three models are effective in identifying the crop. In the case of tomato, due to irregular canopy shape and lodging at the mature growth stage, the stalk overlaps with neighbouring plants. This overlapping further enhances the confusion in the geometrical features of the crops leading to sub-optimal feature learning in the discrimination model. As a result, tomato was confused with partly with eggplant and residual soil ridges leading to lower classification accuracies by the PointNet and DGCNN models. However, compared to the performance of the PointNet and DGCNN models, the CropPointNet model indicates a relatively better accuracy for the tomato crop due to the reduction of false positives from the residual soil ridges.

The deep learning-based CropPointNet model presented in this work was applied on three vegetable crops with complicated canopy structures and including the non-crop geometrical features such as residual soil and crop management infrastructure. The model has the potential to apply on other vegetable or field crops for plant or patch level discrimination in 3D geometrical perspective, amenable for application in precision crop management. The crop plant level object-based classification is useful for generating reference crop libraries for further use in automated object-level crop discrimination. We suggest exercising caution in generalizing the model functionality on other standing crops. A limitation of this work is the usage of the point cloud from a single phenological stage. The phenological growth stage also influences crop discrimination. In our future work, we would like to assess the point cloud for crop discrimination at different phenological growth stages. Also, we recommend further studies across different regions, crops and phenological stages.

Conclusions

Plant level object-based crop discrimination is vital for precision monitoring and management of crop growth, health and nutrient status. 3D LiDAR point cloud has been fast emerging as a general remote sensing data source for high resolution field level mapping and modelling of crops. The discrimination of crops at the plant level is determined mainly by the sophistication of the data and the classification model used. The potential of LiDAR point cloud acquired from a terrestrial laser scanner for plant-level discrimination of three common vegetable crops-, tomato, cabbage and eggplant has been studied. A modified general deep learning convolutional neural network model (CropPointNet) has been adopted for efficient feature learning and object-based segmentation of crop plant objects in the point cloud. The results suggest unambiguous plant level crop discrimination with the best accuracy of about 90% for eggplant and cabbage and 70% for tomato. Contextual soil management factors and overlapping plant canopies are the main factors in reducing the spatial completeness and quality of crop discrimination. Synchronized with a sophisticated ground filtering process, the CropPointNet deep learning model is a flexible supervised model for object-based segmentation and classification of LiDAR point cloud for high-resolution crop discrimination.

Acknowledgements This work has been carried out as part of an Indo-German research cooperation (DBT: The Rural-Urban Interface of Bengaluru- a Space of Transitions in Agriculture, Economics, and Society; DFG: Research Unit FOR2432/1) funded by the Department of Biotechnology (DBT), Government of India and the German Research Foundation (DFG), Germany. We gratefully acknowledge the financial support from DBT and DFG provided in the form of a research grant (DBT: DBT/IN/German/DFG/14/BVCR/2016; DFG: WA 2135/4-1 and BU 1308/13-1).

Authors' contributions R R Nidamanuri supervised the research and partly wrote the manuscript; A M Ramiya guided the implementations; J Reji wrote the codes, implemented the work processes, and partly wrote the manuscript.

Funding Stated in the Acknowledgements.

Data availability As per the regulations of our institute, we can share the data acquired in this research upon individual request until November 2021. Thereafter, we will be hosting the data on a commonly accessible platform with DOI.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Avola, G., Di Gennaro, S. F., Cantini, C., Riggi, E., Muratore, F., Tornambè, C., & Matese, A. (2019). Remotely sensed vegetation indices to discriminate field-grown olive cultivars. *Remote Sensing, 11*(10), 1242.

Axelsson, P. E. (2000). DEM generation from laser scanner data using adaptive TIN models. *International Archives of the Photogrammetry and Remote Sensing., 33,* 110–117.

Bellakaout, A., Cherkaoui, M., Ettarid, M., & Touzani, A. (2016). Automatic 3D extraction of buildings, vegetation and roads from LIDAR data. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences., 41,* 173–180.

Eckart, B., Kim, K. & Jan, K. (2018). EOE: Expected overlap estimation over unstructured point cloud data, In Proceedings - 2018 International Conference on 3D Vision, 3DV 2018. Institute of Electrical and Electronics Engineers Inc. 747–755. Doi:https://doi.org/10.1109/3DV.2018.00090

Handique, B. K., Khan, A. Q., Goswami, C., Prashnani, M., Gupta, C., & Raju, P. L. N. (2017). Crop discrimination using multispectral sensor onboard unmanned aerial vehicle. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences., 87*(4), 713–719.

Jin, S., Su, Y., Gao, S., Wu, F., Hu, T., Liu, J., Li, W., Wang, D., Chen, S., Jiang, Y., & Pang, S. (2018). Deep learning: individual maize segmentation from terrestrial lidar data using faster R-CNN and regional growth algorithms. *Frontiers in plant science, 9,* 866.

Johnson, R. A., Miller, I., & Freund, J. E. (2000). *Probability and statistics for engineers*. Pearson Education.

Kingma, D.P. & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv: 1412.6980.

Koen, B. V. (1988). Toward a definition of the engineering method. *European Journal of Engineering Education, 13*(3), 307–315. https://doi.org/10.1080/03043798808939429

Lawin, F. J., Danelljan, M., & Felsberg, M. (2017). Deep projective 3D semantic segmentation. In F. J. Lawin, M. Danelljan, & M. Felsberg (Eds.), *Lecture Notes in Computer Science* (pp. 95–107). Springer Verlag.

Liu, Y., Piramanayagam, S., Monteiro, S. T., & Saber, E. (2019). Semantic segmentation of multisensor remote sensing imagery with deep ConvNets and higher-order conditional random fields. *Journal of Applied Remote Sensing, 13*(1), 016501.

Lowphansirikul, C., Kim, K.S., Vinayaraj, P. & Tuarob, S. (2019). 3D Semantic segmentation of large-scale point-clouds in urban areas using deep learning. In Proceedings of the IEEE 11th International Conference on Knowledge and Smart Technology, 23–26 Jan. 2019. https://doi.org/10.1109/KST.2019.8687813.

Meng, Q., Hashimoto, Y. & Satoh, S.I. (2019). Fundus image classification and retinal disease localization with limited supervision. In Asian Conference on Pattern Recognition. Springer

Murray, J., Fennell, T. H., Blackburn, G. A., Whyatt, J. D., & Li, B. (2020). The novel use of proximal photogrammetry and terrestrial LiDAR to quantify the structural complexity of orchard trees. *Precision Agriculture, 21,* 473–483.

Ozdarici-Ok, A., Ok, A., & Schindler, K. (2015). Mapping of agricultural crops from single high-resolution multispectral images Data-driven smoothing vs. parcel-based smoothing. *Remote Sensing, 7*(5), 5611–5638.

Paulus, S., Dupuis, J., Mahlein, A. K., & Kuhlmann, H. (2013). Surface feature-based classification of plant organs from 3D laser scanned point clouds for plant phenotyping. *BMC Bioinformatics, 14*(1), 238.

Paulus, S., Dupuis, J., Riedel, S., & Kuhlmann, H. (2014). Automated analysis of barley organs using 3D laser scanning: An approach for high throughput phenotyping. *Sensors, 14*(7), 12670–12686.

Qi, C.R., Su, H., Mo, K. & Guibas, L. J. (2017a). PointNet: Deep learning on point sets for 3D classification and segmentation. In Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. 77–85. Doi:https://doi.org/10.1109/CVPR.2017.16

Qi, C. R., Yi, L.,. Su, H., & Guibas, L. J. (2017b). PointNet++: Deep hierarchical feature learning on point sets in a metric space. In Proceedings of the 1st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

Soilán, R. M., Lindenbergh, R., Riveiro, R. B., & Sánchez, R. A. (2019). Pointnet for the automatic classification of aerial point clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 4,* 445–452.

Sun, S., Li, C., & Paterson, A. H. (2017). In-field high-throughput phenotyping of cotton plant height using LiDAR. *Remote Sensing*. https://doi.org/10.3390/rs9040377

Varfolomeev, I., Yakimchuk, I., & Safonov, I. (2019). An application of deep neural networks for segmentation of microtomographic images of rock samples. *Computers, 8,* 72.

Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M. & Solomon, J.M. (2018). Dynamic graph cnn for learning on point clouds. arXiv preprint, arXiv: 1801.07829.

Weiss, U., & Biber, P. (2011). Plant detection and mapping for agricultural robots using a 3D LIDAR sensor. *Robotics and autonomous systems, 59*(5), 265–273.

Weiss, U., Biber, P., Laible, S., Bohlmann, K. & Zell, A. (2010). Plant species classification using a 3D LIDAR sensor and machine learning. In Proceedings of 2010 Ninth International Conference on Machine Learning and Applications, 339–345. IEEE.

Zhang, X., Sun, Y., Shang, K., Zhang, L., & Wang, S. (2016). Crop classification based on feature band set construction and object-oriented approach using hyperspectral images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9*(9), 4117–4128.