SYDE 461 Canopii

Team 19: Derin Denizkusu, Ethan Liang, Kevin Luong, Jessie Won

Abstract – The City of Kitchener has created a Sustainable Urban Forest Strategy aimed towards improving their urban forest to create a more sustainable city [1]. This plan, however, lacks a maintained tree inventory inhibiting effective long-term planning, especially in private areas. The project objective is to create a product used by residents in order to expand the existing private tree inventory through facilitating data collection. The proposed solution is an application that extracts tree species and location from images and exposes it to the city. The primary engineering components are the user interface and backend system. Currently, initial interface wireframes are completed and in the process of usability testing for feedback. The current backend system uses convolutional neural networks to vectorize images of leaves, and simple classification methods such as including nearest neighbor and K-means, for data extraction. Strong conclusions cannot yet be drawn from the results and efficacy of the classification. The recommended next steps are data acquisition for trees in southern Ontario, usability testing, continued algorithm development, and alternate data input explorations.

Keywords – Tree mapping, urban forestry, sustainability.

I. INTRODUCTION AND IMPACTS

The City of Kitchener has created a Sustainable Urban Forest Strategy (SUFS) aimed towards "planning, engaging, maintaining, protecting and planting Kitchener's urban forest" to combat climate change and increase their urban sustainability [1]. A gap in their existing program is the lack of tree inventory on private lands, which composes 56% of their urban forest [2]. This results in the city having a Low rating on "understanding of privately owned trees across the urban forest" due to having "no information about privately owned trees" [2]. This limited data prevents the city from using modelling software such as i-Tree Eco, which calculates the benefits of the urban forest and facilitates long-term planning [2].

The current process for updating a tree inventory is individual manual tree tagging, with limited software, survey wheels, tape measures, and tree identification keys [3]. Existing tree inventory solutions include ArcGis tree inventory and TreesCount!, however these either lack government integration or do not automate identification or data collection [3, 4]. Existing technologies with automatic identification such as PlantSnap and LeafSnap lack location tagging and keep their data private [5, 6]. Aerial surveillance also exists as a method of data collection, such as correlating LiDAR and ground data to obtain descriptions of height, DBH, and tree species but, no public product for this exists [7,8].

The project objectives are to leverage these existing technologies to build a product that can be used by residents with access to privately owned trees to automatically extract tree data and update the current inventory. This would, at a minimum, work to expand the database with information of private trees, which requires species and location data. Further development includes additional data extraction features such as diameter breast height (DBH) and crown health, which would better enable the use of environmental modelling software such as i-Tree eco [9]. Potential impacts of this product include the loss of jobs as a result of the automation of manual data collection, and privacy violation as the city may use the product on private land without full consent [10]. If successful, this project should help demonstrate the value of urban forests and prompt more effective planning to improve urban canopy in Kitchener. Successful urban forestry has numerous impacts such as improved drainage and reduced air pollutants [11,12]. Furthermore, the solution will engage local residents, which has proven to result in numerous local benefits, including a deeper understanding of sustainability issues, reduced long-term costs, and an increase in community capacity to mitigate climate change [13].

II. PROJECT SCOPE

The system diagram below shows the project scope and how it will be integrated with the current situation of concern (SOC). The right section shows the current system, where the City of Kitchener is manually adding data from their urban forest into their database. They do not have enough data to use their desired environmental modelling software, i-Tree Eco, and cannot perform extensive analyses [14].

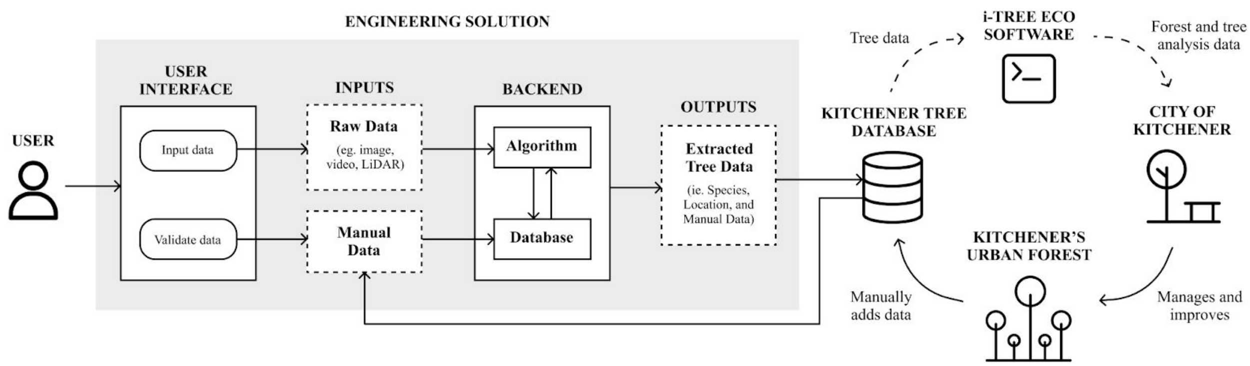

Fig. 1. Systems diagram of project scope

To improve upon this, the project can be summarized in the following situation impact statement: *Design a product that is to be used by residents in Kitchener-Waterloo with access to privately owned trees in order to tag them and improve the city's tree inventory with data on tree species and location.* This solution, as shown in the diagram, will extract information about tree attributes (output) from a provided data source (input) to update the existing tree inventory. The main system components are the user interface (UI) where a user can input data, and the backend, which accepts these inputs and extracts information using an algorithm and database. This output populates to Kitchener's tree inventory, and is compatible with the current database, and off-the-shelf tools, namely i-Tree Eco. Other off-the-shelf solutions such as phones, cameras, and LiDAR will be utilized as data input sources. The UI and backend are designed to be compatible with these inputs, along with manual data inputs, and will leverage existing software technologies such as Keras.

III. ENGINEERING METHODS OF ANALYSIS AND DESIGN

As shown in the high-level system diagram above, there are two primary components to the engineering solution: the UI and the backend system. The critical requirements of these two components are shown below. The overall engineering analysis, design, and testing plans are defined given these requirements, and benchmarks have been broken down into stages of low-fidelity prototypes (LFPs), medium-fidelity prototypes (MFPs), and a high-fidelity prototype (HFP).

User Interface (UI)

- A user must be able to input data
- A user must be able to validate results

Backend System

- The system must accept an input
- The system must extract location and species data
- The system must output results compatible with Kitchener's tree inventory

A. *User Interface Design and Testing Methods*

There are two critical requirements for the UI: data input and result validation. The user flow for this is shown on the right.

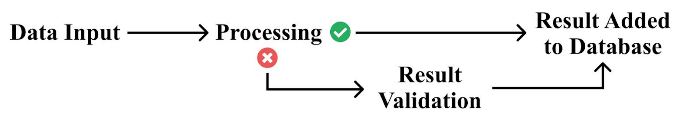

Fig. 2. User flow for the UI critical requirements

The UI will be the entry point for all inputs into the solution. Data sources being investigated are LiDAR, ground photography or video data. The initial engineering focus is on image data, as proven solutions for species recognition exit, greater potential for social engagement, less ethical and privacy concerns, and existing engineering resources. The final decision between which data source to choose will rely on a computational decision matrix based on the efficacy of data collection, increased work required by the city. and the cumulation of other benefits unique to either source. Regardless, these inputs will be processed by the backend system, and result validation is only required in the case of a processing error. The design and testing plan are the same for both requirements, and the tools used for the design are Notability, for sketches, and Figma, for wireframes and mockups. Tools that will be used to facilitate testing are InVision and Principle, for interactive prototyping.

The design plan is to build a modular interface compatible with any input, to allow for the team to make progress while research and engineering is conducted to decide the input. A mobile-first approach is being used, to ensure clarity and simplicity in all features, and responsiveness across screen sizes [15]. An iterative design methodology is also being followed, which is a cyclic process of prototyping, testing, and analyzing to refine the design [16]. It is critical that testing is done at every stage to get constant feedback. The primary validation method will be in-person usability testing, for which the protocol is [16]:

1. Identify the goal of the test and prepare respective prototypes, tasks, and questionnaires
2. Introduce yourself, the project and purpose, allow user to ask questions, and build a rapport
3. Conduct test
    a. Read the user the scenario and their task
    b. Share prototype with user and ask them to think aloud as they walk through it
    c. Record the user's actions, comments, questions, and body language
4. Share a post-test survey with reflective questions (eg. how easy was..., did you notice…, etc.)

The benchmarks for various stages of UI engineering are in the table below. The team will use React Native to develop the application in the MFPs stage, as the framework allows for portability of the application to browser, iOS, and Android platforms [17]. This will also enable modular development, so the team can build reusable components. Jest will be used for unit and integration at all stages of development to validate the functionality of the application [18]

Table 1. Stages and Benchmarks for the User Interface

Stage	Benchmarks
LFPs (Late November)	● Mobile mockups of ideal single and mass data input flows ● Completed user testing for data validation layout and workflow ● Comprehensive research on alternate data inputs, especially LiDAR
MFPs (February)	● Cross-platform mockups of all single and mass data input flows ● Wireframes of non-critical features (eg. view inventory, edit, etc.) ● Cross-platform application that can upload data and take photos
HFP (Symposium, Early March)	● Cross-platform application which supports communication with server for processing, storing and uploading tree images and feature vectors ● Comprehensive UI unit & integration test suite using Jest

B. *Backend System Design and Software Architecture*

Software design success is measured by quality attributes including maintainability, correctness, reliability and efficiency [19]. To ensure success, the software is split into components developed for the user's device and a backend system. Each component will be maintained individually to meet product revision quality attributes. The UI, image recognition system and data storage will all be kept separate. The system must be efficient enough for users to receive real-time leaf species recognition feedback and thus, a local, condensed, image recognition solution is hosted on the user's device to reduce calls to the backend server, which has the complete classifier to handle failures. To maintain reliability, the application is separate from this server, which hosts a database. This is also separate from the Kitchener database.

In the standard data flow as shown in Fig. 3, the user will interact with the device UI (1), which relays information to and from the device's storage (6). When the user inputs an image, it is pre-processed to meet neural network input expectations, such as size and format (2), then sent through to the deep neural network (3). This produces a vector representation of the image, which is inputted into the device-hosted classifier which determines the species by comparing it against a small dataset on the device(4). If this fails, the vector representation is sent to the full classifier hosted on the backend which compares the vector against all vectors existing in the database(5). When an image is successfully classified, the results are sent to the database and the classifier data is updated with the new representation (7, 8). If an image cannot be classified, it is uploaded to be reviewed manually (9). In this case, the image from local storage is sent to be stored so that the algorithm can be improved (10). The results database is connected to update Kitchener's tree inventory (11).

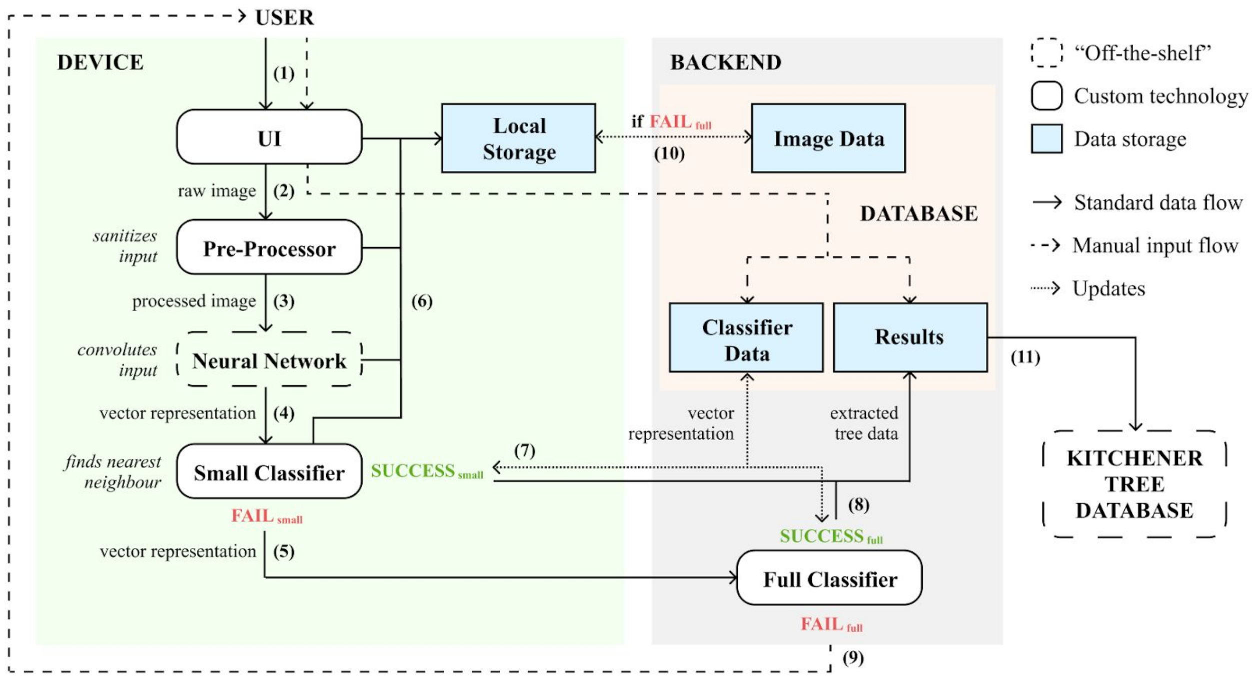

Fig. 3. Software architecture design diagram

To build a tree species classifier, a pre-trained convolutional neural network (CNN) will be coupled with a classifier, as done in prior plant recognition systems [20, 21, 22]. A preliminary solution will be built using features obtained from one of the final layers in a CNN, such as InceptionV3 or Densenet. The classifier will be built using a Euclidean distance measure to find the nearest feature vector with a known classification. With more data, a neural network or another classification method can be used. The selection of the technologies and CNNs will be based on accuracy. The collection of location data will be conducted through a device's geolocation functionality. The design for this component has not been defined, and is not a priority until the MFPs stage. Benchmarks for the other stages are below.

Table 2. Stages and Benchmarks for the Backend System

Stage	Benchmarks
LFPs (Late November)	● Compile a sufficient dataset, including training and test data for a classifier with 60%[1] accuracy for the five most common trees in the region ● Build, test, and compare at least 3 different species recognition systems
MFPs (February)	● Classify tree species in southwest Ontario with 60% accuracy[1] ● Include location tagging from image meta data and location services
HFP (Symposium, Early March)	● Server hosting image processing algorithm, 80% accuracy [1] ● Database storing vectorized images, unclassifiable images, and tree data ● API endpoints exposing results to city

[1] Based off of a study on the efficacy of an image based classification of plant genus [23]

As benchmarks and decision making are mostly accuracy-based, the testing protocol to evaluate classification accuracy is to separate the collected data into a training and testing set, with a data split of either 80-20 or 70-30 as per previous image recognition work with plant data [20, 21, 22]. The classifier will be built using the training set, and the classification accuracy will be calculated using the ratio of correct classifications to all classifications on the testing set. Independent components of the image recognition pipeline are validated with manual testing to ensure that problems in the pipeline can be isolated with ease. Further in-person testing will also be conducted by knowledgeable expert users, such as tree planting volunteers, who will use the solution and manually validate the species classification of the images they capture, following the usability testing protocol aforementioned.

IV. PROJECT OUTCOMES TO DATE

The team is on track for creating LFPs for late November. Formal mockups are required to complete the UI LFPs, and usability testing will be conducted for feedback. The classification systems are dependent on additional data collection, after which various deep networks and classification prototypes can be evaluated.

A. *User Interface Outcomes*

Initial wireframes for both critical requirements have been completed. Both flows are below, where Take Photo, is a single entry input, and Upload Images, is a bulk data input.

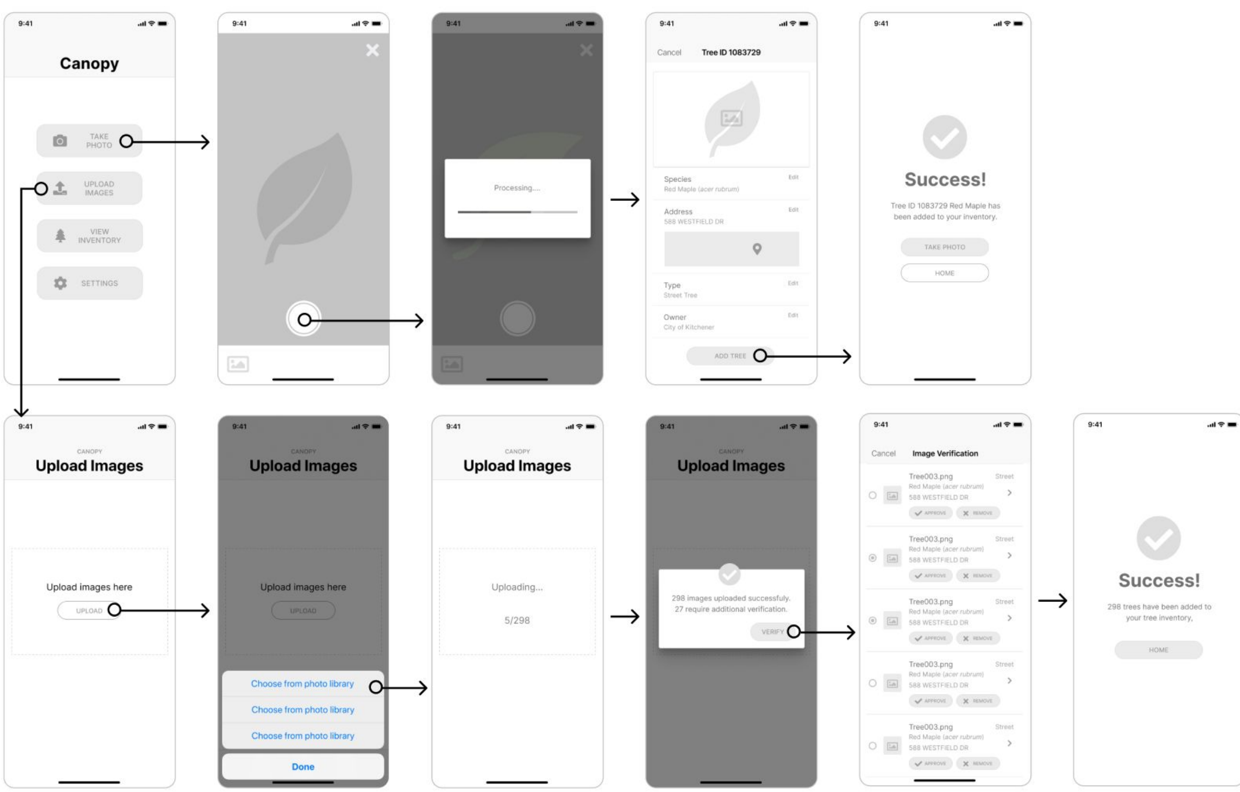

Fig. 4. Workflow for ideal single data input and bulk data input with processing error

Both flows are simple as they are mostly automated, with the most complexity in the case of a processing error or if the user wants to perform manual validation. This is not a concern for a single entry input since the data of one tree can easily be displayed on a screen. This is much more complex when verifying multiple entries due to a large amount of fields and an unclear hierarchy of data. Some explorations are shown below, however, testing and additional research needs to be done to determine the most important fields per entry, what data is extractable, what data must be manually entered, and what the most common use cases are. For example, the method through which users correct errors must still be determined. Depending on the accuracy of the algorithm, the actions of “Approve”, “Edit”, and potentially “Remove” will vary in frequency and importance. Other considerations are bulk changes, for example changing the tree type for multiple trees or approving all images that require validation.

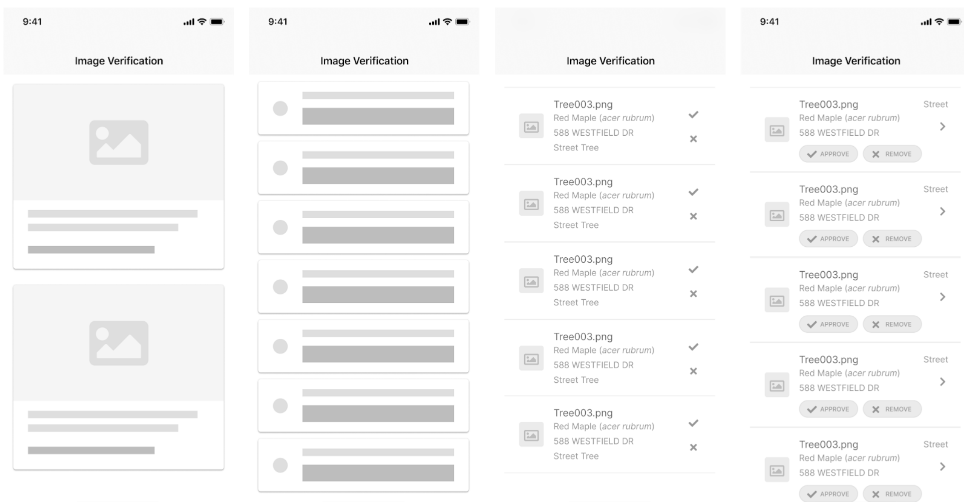

Fig. 5. Wireframe explorations for mass input verification

B. *Backend System Outcomes*

To date, the team has implemented an end-to-end tree species classifier which ingests an image of a tree leaf, transforms it into the correct size and resolution, feeds it into a CNN, and classifies the tree species based on similarity to leaves in a training set. The training set is composed of a compilation of public datasets and cross-referenced with an index of trees native to Ontario [24].

To determine the feasibility of image processing a solution to the problem space, it was necessary to prove that CNNs (InceptionV3 and Densenet) were capable of extracting differentiable features from the leaves. The team was able to complete this, and next, was required to determine the best neural net for the project. This is defined as the one which maximizes logical distance between images of different species and minimizes distance between images of the same species. As the output of a CNN is a multi-dimensional vector, principal component analysis (PCA) was used to reduce the dimensionality in order to create a 3D visualization [25]. This determined a low percentage of total variance in the data, 13.9% and 21.3% for InceptionV3 and Densenet respectively, which while unrepresentative of the full characteristics, was important in establishing a mental model of the data for the team.

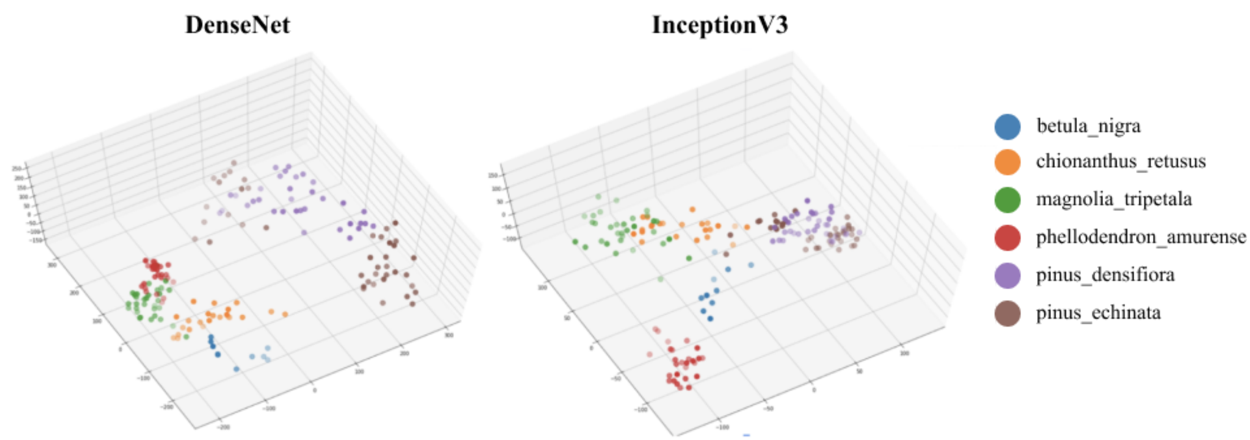

Fig. 6. Comparison of DenseNet and Inception V3 CNN PCA

As shown above, both CNNs were capable of separating leaf samples from different species by distance, meaning CNNs warrant further exploration. To further validate this, the team performed unsupervised clustering with K-means to determine if the vectors of the neural nets could group similar species together. Clustering performance for each species is shown below, and the number of samples successfully recognized as belonging to the same cluster was detailed.

Table 3. K-means Clusters for Two CNNs

Species	Densenet Percentage in same cluster	Inception v3 Percentage in same cluster
Betula nigra	0/10 (0 %)	10/10 (100%)
Chionanthus Retusus	23/23 (100%)	17/23 (73.9%)
Magnolia Tripetala	30/32 (96.8%)	21/32 (65.6%)
Phellodendron Amurense	27/27 (100%)	27/27 (100%)
Pinus Densiflora	21/34 (61.8%)	21/34 (61.8%)
Pinus Echinata	27/42 (64.3%)	27/42 (64.3%)
TOTAL	**120/168 (71.4%)**	**123/168 (73.2%)**

Finally, the team created a pipeline to test and classify actual images of leaves. This was done using K-nearest neighbor to predict the species of the images. For a test set of two tree species native to Ontario, the classifier using Densenet features correctly classified 6/9 test leaves while the classifier using InceptionV3 features correctly classified 3/9 leaves. Based on the small volume of results, it is currently not possible to establish an accurate assessment of K-NN clustering performance and further work is needed.

C. *Research Outcomes*

As a potential alternate data source, the team performed extensive research on LiDAR, a remote sensing technology which transmits laser pulses from a moving vehicle and combines this data with GPS, inertial measurement units and aerial imagery to gain a comprehensive understanding of the topography [26, 27]. Key findings showed that LiDAR is reasonably priced with some estimates of airborne LiDAR costing around $200 per square kilometer scanned, and for Kitchener, would cost ~$27,000 with price decreasing as larger areas are scanned [28, 29]. In 2014, Kitchener authorized a LiDAR assessment to determine the current state of the urban tree canopy, but the data is private and this process has not been repeated since [30]. Research by Kim et al. showed that LiDAR data can be classified to determine the most significant features in differentiating between trees, roads and buildings, as well as the efficacy of a K-means and Spectral Angle Mapper classification method [31]. Granted access to Kitchener's study, these methods could potentially be used to identify trees in Kitchener. Once this has been done, many features of the trees can also be extracted from LiDAR data such as tree height, species, and even individual tree growth over time [32, 33, 34]. This concludes that LiDAR is a promising potential solution for the project scope, and should continue to be pursued.

V. CHALLENGES AND CONCLUSIONS DRAWN FROM OUTCOMES TO DATE

Based on current results and efforts to date, the project is on track for meeting the team's goals for SYDE 461. Various LFPs are on track to be completed by late-November but initial project goals set have been modified to reflect the availability of resources to the team. Specifically, extracting size data of trees has been removed as a critical requirement, as it is not required by Kitchener's database and adds significant complexity. This may become a future feature depending on feedback, additional research, and timelines. The scope of the LFP has been modified to shift UI engineering work to the MFP. Focus on the LFP is on UI and algorithm design to achieve a more holistic understanding of user needs and workflows. The backend LFP benchmark of obtaining datasets and performing research is still on track, but unforseen blockers are affecting progress towards the algorithm specific LFP benchmarks, namely data storage and image resolution. Extensions to other solutions, especially LiDAR, are still being explored. While research has shown that LiDAR is a promising alternative to image recognition, image processing will continue to be the focus of the project due to the reasons aforementioned. If the team is unable to overcome these challenges, this solution and it's benchmarks may be revised.

An unexpected challenge the team is facing is the difficulty in communication with the City of Kitchener. The team has been in contact with David Schmitt, head of Urban Forestry, but the frequency of communication has dropped. This has made it increasingly difficult to tailor a solution directly addressing the needs and systems in place in Kitchener. It is also difficult to ascertain what tree data would be most helpful for the development of a useful urban forestry plan. The team has focused on species and location based off of the existing inventory and required fields by i-Tree Eco. This has not blocked progress and prototypes can still be developed, but without proper communication moving forward, the team cannot ensure the highest priority problems are being solved within the SOC.

A conclusion drawn from testing is the necessity for a more comprehensive training and testing dataset. This is required to increase the robustness of the algorithm. The current dataset only contains training data for 15 of the 62 tree species in southern Ontario, and the solution cannot identify trees which it has not been trained against. A full testing set is also lacking, which means validation of the classifier performance against benchmarks cannot be concluded. More data must be collected such that the LFP and MFP algorithms can be measured against the predetermined benchmarks.

Challenges were also faced in the current classification method. Two major problems are inefficient data storage solutions and varying different image resolutions. For a relatively small training dataset of 2130 images, the numerical representation of the images and extracted vectors currently require >30 and >80 times more storage respectively. Processing this large amount of data results in high RAM consumption which often crashes the process and requires restarts, erasing progress, and will incur significant data storage costs as engineering progresses. The other main challenge is handling images of different resolutions. The team's temporary solution to unblock progress uses one image resolution for its preliminary models, but is exploring multiple solutions such as smart cropping. If the image recognition system can not meet the stated benchmarks, alternate solutions would be used. An alternative is to provide the top five closest matching trees according to the software and rely on manual input from the user.

VI. RECOMMENDATIONS FOR PRACTICAL NEXT STEPS IN PROJECT

As communications with the City of Kitchener have been an unexpected challenge, the team has, and will continue to contact alternate individuals such as Anne Grant, manager of the UW Ecology Lab which includes a UW Urban Forest, and the City of Waterloo. The team will also pursue additional communication with neighbouring districts, such as Toronto, and Guelph, to get additional insights and feedback. If the challenges with David Schmitt are not resolved, the team will pivot focus to the City of Waterloo as an immediate next step, with Guelph and Toronto as back-up plans due to proximity and their preexisting plans on urban forestry management [14]. Engineering efforts may be required to adjust the output compatibility with new inventories and software.

Since it is inconclusive whether image processing is the best solution for species identification of private trees, and the LiDAR research conducted shows promising results, the next steps are to continue pursuing LiDAR as a data source. This includes continued communication with experts in LiDAR, specifically through research labs at the University of Waterloo. Regarding Kitchener's 2014 LiDAR study, the team will continue communications in hopes of gaining access to this data. If LiDAR becomes a viable solution, the team will work to blend multiple sources of data to get a more holistic understanding of the composition and change of tree canopy cover. If the team cannot obtain access to Kitchener's study, it is recommended to pursue alternate LiDAR data sources and data collection methods for testing and validation. In the case that the team cannot capitalize on using LiDAR effectively, it cannot be used by the City of Kitchener, or is otherwise demonstrated to be infeasible, then image processing will continue to be the team's primary focus unless alternate solutions are found.

In-person usability testing of the UI designs is a necessary immediate next step to get feedback and validate assumptions on the solution. This includes the need for manual validation, the steps to manual validation, and the intuitivity of single and bulk data inputs. Usability testing protocols are defined in Section III and these tests will be conducted after more structured mockups are designed based on the wireframes. Feedback from these tests will then be implemented in iterations on the mockups. In these designs, it is imperative to consider the target users: residents who own private trees, homes, or businesses. Next steps are to conduct research and better define a persona, as the proposed product will require a fundamental technological understanding and an interest in sustainability, both of which vary based on demographic. Prototypes should also be tested and reviewed with industry experts, including Igor Ivkovic, for software and UI design feedback, and David Schmitt, for feedback on compatibility, requirements, needs, and resources.

Similarly, additional testing needs to be conducted for the backend algorithm. As results are currently inconclusive, the testing protocol for function of the system documented in Section III should continue to be followed. In-person testing and review should also be conducted with industry experts including Professor Hamid Tizhoosh, for technical image-processing feedback. This will further identify areas of improvement and additional next steps. With respect to known challenges, namely inefficient storage, the team should explore alternate compression or filtering methods. Different classification methods such that storage of full vectors will no longer be required should also be explored. To address restrictions on aspect ratios, immediate next steps will be to explore cropping and content-aware fill.

Regarding insufficient data limiting the training and testing of the classification algorithm, immediate next steps are to collect more data. By training on images with different orientations, colors, sizes, etc., the algorithm can better classify a wider range of images. An increased number of testing samples will also allow for validation of the algorithm's robustness. Increasing the dataset can be done through modifying existing data and collecting new data. For the training dataset, existing images can be rotated, blurred, cropped or compressed to generate different representational vectors for better coverage. The testing set, however, requires real leaf images not derived from the existing image set. To find the required leaf image data, it can be gathered manually by taking pictures or through online resources. Collecting leaf images online can also be either a manual or an automatic algorithmic scraping solution. The two methods for online collection have different tradeoffs. An algorithmic method will be far faster and scalable at the cost of potentially low accuracy and high initial time investment. For a manual process involving tagging images by hand, accuracy can be guaranteed though progress will be slower and the scale may be insufficient to make improvements to the algorithm. As such, the team is recommended to experiment with automatic data collection first to determine the complexity and risk of the scraping algorithm. If a high quality solution cannot be developed with reasonable time investment, then the team will pivot and undertake a manual approach to data collection. This is to be completed by early-November to allow time for training and testing for the LFPs.

As the project scope has been modified and reduced, there are additional non-critical requirements and features which the team hopes to complete by the end of March. These include extracting DBH and size data from trees, editing entries, and viewing all trees in the database. However, these are currently out of scope, and depending on progress, feedback, and additional research, may be included in scope for the MFPs and HFP in 462. The next steps are to continue with the current plan, and reassess the project over the next two months.

ACKNOWLEDGEMENTS

The team would like to thank Professor Igor Ivkovic for providing mentorship and Professor Hamid Tizoosh for his technical guidance. The team would also like to acknowledge David Schmitt, for his insights into Kitchener's urban forestry landscape, Dr. Sean Geobey for his help with understanding the sustainability space, and Dr. Kate Mercer for her feedback on project research and problem space. Lastly, the team would like to thank Professor Matthew Borland for his instruction throughout this course.

REFERENCES

[1] “Kitchener Sustainable Urban Forest Strategy 2019-2039,” Kitchener Urban Forestry. [Online]. Available: https://www.kitchener.ca/en/resourcesGeneral/Documents/INS_PARKS_Sustainable_Urban_Forest_Strategy_DRAFT_SPREAD.pdf. [Accessed: 01-Nov-2019].

[2] “Sustainable Urban Forest Report Card,” City of Kitchener, 2017. [Online]. Available: https://www.kitchener.ca/en/resourcesGeneral/Documents/INS_PARKS_Sustainable_Urban_Forest_Report_Card_2017.pdf. [Accessed: 01-Nov-2019].

[3] “New York City Street Tree Map,” NYC Parks, 2019. [Online]. Available: https://tree-map.nycgovparks.org. [Accessed: 26-Oct-2019].

[4] “ArcGIS Online,” *arcgis.com*. [Online]. Available: https://www.arcgis.com/index.html. [Accessed: 01-Nov-2019].

[5]“PlantSnap,” earth.com, 2019. [Online]. Available: https://www.plantsnap.com. [Accessed: 01-Nov-2019].

[6] N. Kumar, P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. C. Lopez, and J. V. B. Soares, “Leafsnap: A Computer Vision System for Automatic Plant Species Identification,” *Computer Vision – ECCV 2012 Lecture Notes in Computer Science*, pp. 502–516, 2012. [Accessed: 01-Nov-2019].

[7] Aerial Forest Inventory System, by J. Lyle. (2011, Jun. 16). *Patent US20130211721A1*. Accessed on: Sep. 26, 2019. [Online]. Available: https://patents.google.com/patent/US9063544. [Accessed: 01-Nov-2019].

[8] Forest Inventory Assessment Using Remote Sensing Data, by Z. Parisa. (2011, Dec. 22). *Patent US9063544B2*. Accessed on: Sep. 26, 2019. [Online]. Available: https://patents.google.com/patent/US20130211721A1/en. [Accessed: 01-Nov-2019].

[9] “i-Tree Eco.” [Online]. Available: https://www.itreetools.org/tools/i-tree-eco. [Accessed: 02-Nov-2019].

[10] “The Impact of Automation on Employment - Part I,” NCCI Holdings Inc., 10-Oct-2017. [Online]. Available: https://www.ncci.com/Articles/Pages/II_Insights_QEB_Impact-Automation-Employment-Q2-2017-Part1.aspx. [Accessed: 01-Nov-2019].

[11] “Stormwater to Street Trees, Engineering Urban Forests for Stormwater Management,” Sep-2013. [Online]. Available:

https://www.epa.gov/sites/production/files/2015-11/documents/stormwater2streettrees.pdf. [Accessed: 31-Oct-2019].

[12] Agriculture Organization of the United Nations, Urban and Peri-urban Forestry. [Online]. Available: http://www.fao.org/forestry/urbanforestry/87029/en/. [Accessed: 01-Nov-2019].

[13] D. Bergstrom, K. Rose, J. Olinger, and K. Holley, “The Sustainable Communities Initiative: The Community Engagement Guide for Sustainable Communities,” *Journal of Affordable Housing & Community Development Law*. [Accessed: 01-Nov-2019].

[14] “Developing a Sustainable Urban Forest Program,” City of Kitchener, 2017. [Online]. Available: https://www.kitchener.ca/en/resourcesGeneral/Documents/INS_OPS_UrbanForestry_Developing-a-sustainable-urban-forest-program.pdf. [Accessed: 31 Oct 2019].

[15] “Responsive Web Design Tenets,” *Building Responsive Data Visualization for the Web*, pp. 24–69, 2015. [Accessed: 01-Nov-2019].

[16]J. Nielsen, "Iterative user-interface design," in Computer, vol. 26, no. 11, pp. 32-41, Nov. 1993. [Accessed: 01-Nov-2019].

[17] “React Native · A framework for building native apps using React,” *React Native Blog ATOM*. [Online]. Available: https://facebook.github.io/react-native/. [Accessed: 03-Nov-2019].

[18] “Jest · Delightful JavaScript Testing,” *Jest*. [Online]. Available: https://jestjs.io/. [Accessed: 03-Nov-2019].

[19] J. A. Mccall, P. K. Richards, and G. F. Walters, “Factors in Software Quality. Volume I. Concepts and Definitions of Software Quality,” Nov. 1977. [Accessed: 01-Nov-2019].

[20] A. K. Reyes, J. C. Caicedo, and J. E. Camargo, “Fine-tuning Deep Convolutional Networks for Plant Recognition,” *Laboratory for Advanced Computational Science and Engineering Research*. [Accessed: 01-Nov-2019].

[21] M. M. Ghazi, B. Yanikoglu, and E. Aptoula, “Plant identification using deep neural networks via optimization of transfer learning parameters,” *Neurocomputing*, vol. 235, pp. 228–235, 2017. [Accessed: 01-Nov-2019].

[22] H. Esmaeili and T. Phoka, “Transfer Learning for Leaf Classification with Convolutional Neural Networks,” 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2018. [Accessed: 01-Nov-2019].

[23] M. Seeland, M. Rzanny, D. Boho, J. Wäldchen, and P. Mäder, “Image-based classification of plant genus and family for trained and untrained plant species,” *BMC Bioinformatics*, vol. 20, no. 1, Mar. 2019. [Accessed: 01-Nov-2019].

[24] “The Tree Atlas: Southwest region,” *Ontario.ca*. [Online]. Available: https://www.ontario.ca/environment-and-energy/tree-atlas/ontario-southwest/. [Accessed: 01-Nov-2019].
[25] H. Abdi and L. J. Williams, “Principal component analysis,” *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010. [Accessed: 01-Nov-2019].

[26] “Fundamentals about lidar,” *What is lidar data?-Help | ArcGIS Desktop*. [Online]. Available: https://desktop.arcgis.com/en/arcmap/10.3/manage-data/las-dataset/what-is-lidar-data-.htm. [Accessed: 01-Nov-2019].

[27] US Department of Commerce and National Oceanic and Atmospheric Administration, “What is LIDAR,” *NOAA's National Ocean Service*, 01-Oct-2012. [Online]. Available: https://oceanservice.noaa.gov/facts/lidar.html. [Accessed: 01-Nov-2019].

[28]“Lidar Pricing.” [Online]. Available: https://www.michigan.gov/documents/cgi/Lidar_Pricing_SOM_CSS_409728_7.pdf. [Accessed: 02-Nov-2019].

[29] Statistics Canada, “Census Profile, 2016 Census Kitchener - Cambridge - Waterloo [Census metropolitan area], Ontario and Ontario [Province],” *Census Profile, 2016 Census - Kitchener - Cambridge - Waterloo [Census metropolitan area], Ontario and Ontario [Province]*, 09-Aug-2019. [Online]. Available: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CMACA&Code1=541&Geo2=PR&Code2=35&Data=Count&SearchText=kitchener&SearchType=Begins&SearchPR=01&B1=All&TABID=1. [Accessed: 01-Nov-2019].

[30] J. O’Neil-Dunne, “Urban Forestry Tree Canopy Report,” University of Vermont Spatial Analysis Lab, 2016 .[Online]. Available: https://www.kitchener.ca/en/resourcesGeneral/Documents/INS_OPS_UrbanForestry_Tree-Canopy-Report-Kitchener.pdf. [Accessed: 31 Oct 2019].

[31] A. M. Kim, R. C. Olsen, and F. A. Kruse, “Methods for LiDAR point cloud classification using local neighborhood statistics,” *Laser Radar Technology and Applications XVIII*, 2013. [Accessed: 01-Nov-2019].

[32] H.-E. Andersen, S. E. Reutebuch, and R. J. Mcgaughey, “A rigorous assessment of tree height measurements obtained using airborne lidar and conventional field methods,” Canadian Journal of Remote Sensing, vol. 32, no. 5, pp. 355–366, 2006. [Accessed: 01-Nov-2019].

[33] S. Kim, G. Schreuder, R. J. Mcgaughey, and H.-E. Andersen, “Individual Tree Species Identification Using LiDAR Intensity Data,” American Society for Photogrammetry and Remote Sensing, Apr. 2008. [Accessed: 01-Nov-2019].

[34] S. Krause, T. G. Sanders, J.-P. Mund, and K. Greve, “UAV-Based Photogrammetric Tree Height Measurement for Intensive Forest Monitoring,” Remote Sensing, vol. 11, no. 7, p. 758, 2019. [Accessed: 01-Nov-2019].

TEAM MEMBER CONTRIBUTIONS TO ENGINEERING ANALYSIS & DESIGN

Contributions to Team Project Engineering Analysis and Design		
Team Member	**Description of Contributions, including Tools used.**	**Outcomes**
Ethan Liang	● Worked with Derin to create software architecture diagram on paper ● Diagnosed and resolved software issues classification issues in the algorithm ● Set up the development tools for the project including Google Collaboratory and designed the temporary pipeline	● Created the species classification component of the project. Built the infrastructure to retrieve and save images/feature data with Pillow and pickle and implemented PCA visualization, K-means clustering analysis using scikit-learn
Derin Denizkusu	● Created machine learning infrastructure through Keras for project ● Worked on creating preliminary leaf image classification system as well as researching and planning complete image recognition solution ● Created software architecture design for complete solution ● Worked with Ethan to create software architecture diagram on paper	● Created machine learning infrastructure through Keras for project
Kevin Luong	● Worked with Jessie on collecting data for training and test dataset, including finding existing data sets, finding list of trees in southern Ontario to focus, and finding individual images of specific species ● Conducted research into feasibility of LiDAR ● Assisted Derin and Ethan in theoretical validation of pattern recognition technologies	● Generated dataset for testing and training of image processing algorithms ● Research outcomes determining LiDAR is a feasible alternative data source
Jessie Won	● Defined UI design guidelines and methodologies to be used for the project ● Defined critical requirements, stages, and benchmarks for the UI ● Defined usability testing protocol to be used for design and all other in-person testing ● Worked with Kevin on collecting data for training and test dataset, including finding existing data sets, finding list of trees in southern Ontario to focus, and finding individual images of specific species	● Created all concept sketches, using Notability, and wireframes and mockups using Figma ● Created all diagrams (ie. problem scope, user flow, software architecture), tables, and visuals, also using Figma ● Generated dataset for testing and training of image processing algorithms