SOIL ANALYSIS VIA REMOTE SENSING AND ARTIFICIAL INTELLIGENCE FOR PRECISION REGENERATIVE AGRICULTURE

by

Takoda Chance Kemp, Bachelor of Geographic Analysis (Hons.)

A thesis presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Spatial Analysis

in the program of

Spatial Analysis

Toronto, Ontario, Canada, 2022

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

Abstract

Soil electrical conductivity maps were generated for greenspace in the Greater Toronto Area using a conditional generative adversarial network, which is a form of deep learning where one neural network is used to train another. The results of the analysis show that the model can accurately predict soil conductivity 34.6% of the time. It could possibly be strengthened with the inclusion of more electromagnetic bands in the supervised classifications used to train the network, such as the infrared spectrum, as well as Light Detection and Ranging data. This three-dimensional imagery should be considered, as the model is not optimized when soil is obscured by foliage. Generally, these two datatypes are commercially available, and commonly used for the analysis of greenspace. Microdrones can potentially be equipped with computer vision enabled sensors operating this neural model to iteratively analyse soil types, and complete aerial cropping.

Dedication

For my mother, and the communities that have supported my growth. I would like to thank my benevolent colleagues, and instructors that I have had the sincerest, and warmest pleasure of working with. Thank you, Dr. K. Wayne Forsythe for being an amazing supervisor, and sparking my interest in remote sensing, and Mr. Mark Hoel, for inspiring me to become a geographer.

Table of Contents

List of Tables

List of Figures

List of Acronyms

Acronym	Definition
cGAN	Conditional generative adversarial network
DSM	Digital Surface Model
GIS	Geographic Information System
GPS	Geographic Positioning System
LiDAR	Laser imaging, detection, and ranging
RGB	Red, Green, Blue
TEX	Texture
UAV	Unmanned Aerial Vehicle

1 Introduction

A changing climate has prompted many to seek alternative methods of production, housing, and security. A stable agricultural sector must tolerate the changes in the biosphere related to human influence, like changing climate patterns, and extreme weather events, which may introduce uncertainty to plant production. Currently, artificial intelligence has allowed, through machine learning, the ability to effectively cluster entities, or features with a certain number of dimensions into a pre-defined number of groups (Chiang and Mirkin, 2007). This analysis aims to classify the electrical conductivity of soil in three parks in Toronto, using drone imagery, then internalize those results into a generative model to reduce the time investment of future analyses.

Farming involves many variables. There are many plants that require specific care, and climate. Generally, practices that increase the viability of all species of produce are especially valuable. Broadly speaking, these techniques increase the availability of nutrients in the soil to a plant, and are mostly used by industry professionals, or those that grow produce for profit. Around the world there are numerous climatic patterns to consider, and countless permutations of crop cover. Theoretically, a statistical method that allows a drone operator to quickly realize the qualities of their ground conditions, at least iteratively, would allow for large-scale improvements in efficiency.

This is especially useful for farming practices but can also be applied in other facets of Geography – like planning and conservation. Farmers understand well that the factors important to the health of a plant are found in the soil, surrounding geography, and climate. These factors are generally the volume of nutrients, the availability of light and water, and finally, the potential of hydrogen (pH) in the soil. Furthermore, contemporary research has found that, in certain scenarios, self-introduced bacteria may also play a helpful role in sustainable agricultural production (Kumar et al., 2016; Ramakrishna et al., 2019). This is also true for mycorrhizal fungi, which have a symbiotic relationship with plant roots, and give plants nutrients for carbon (Rillig et al., 2016). Another recent discovery in agricultural science is humic and fulvic acids, which increase the availability of nutrients for a plant's roots (Canellas et al., 2015).

Anecdotally, farmers in the Bahamas have reported soil analysis technology as having a large cost barrier. The ability to classify all agriculturally viable soil using a small amount of data will be an important asset for researchers and agriculturalists. An agriculturalist, in this case, would

be a person that produces plants for sale or horticultural reasons. These farmers would no longer need to spend that capital on analyses, and can instead use it elsewhere, like on the importation of seeds, fertilizer, or increasingly energy efficient capital assets, like solar panels. The model can be altered for aesthetically pleasing endemic plants, or profitable produce – so that pollution, and food security can be addressed, respectively.

How this all relates to machine learning is within the realm of remote sensing. These days, open-source tools are available for anyone that seeks to create high-resolution maps of their land and perform supervised k-means cluster analyses of their imagery. In any event, soil samples of nutrients and pH can be taken from parcels of the study area. Human significant pH ranges, which are those that can grow useful plants, can be cartographically illustrated over the study area. Finally, through supervised classifications, in both free, and enterprise Geographic Information System (GIS) software, these human significant ranges may highlight growing substrates that are suitable for a national park, or agricultural production, and marginal urban greenspace. Gone is the need to sample every possible parcel of soil to figure out what it requires for a healthy plant. Currently, machine learning through k-means clustering analysis of readily available images allows for the extrapolation, and illustration of soil factors from relatively small sample areas to large swaths of similar landscape.

In Canada, a licence is not required to fly a drone below 250 grams, as they are considered 'micro-drones', and not included in the basic, or advanced operation rules of heavier unmanned aerial vehicles (UAVs) (Transport Canada, 2021). Truly, this is an opportune moment in time for any researcher interested in capturing high-resolution imagery with a relatively low barrier of entry.

These maps would be immensely useful for re-cropping efforts, may they be manual, or robotic. Furthermore, this creates the possibility of aerial cropping, soil amendment, and fertilization, through the autonomous programming of payload carrying drones to spot significant areas and sow the relevant nutrients, endemic crops, bacteria, fungi, beneficial acids, and pH.

1.1 Research Goals

The steps required to create a relatively high-quality image of agriculturally viable landscape will first be explored. Then, a robust soil sampling method for a supervised k-means

clustering classification will be employed. In essence, K-means clustering analysis is a machine learning statistical method where entities are grouped into categorized clusters given shared factors (Askari et al., 2014; Kim et al., 2021) This classification is conducted on the natural colour bands of the electromagnetic spectrum, as well as statistical indices such as Principal Component Analysis (PCA). Lastly, a conditional generative adversarial network (cGAN) will be trained and evaluated for is efficacy in generating soil classification maps from natural colour images, which will allow analysts to forgo the k-means clustering classification step, and employ the model in novel areas, saving time, and, potentially, financial capital, within the realm of agricultural analysis. Therefore, this research design will answer the question of whether it is possible to use artificial intelligence to classify, and then generate illustrations of soil substrate quality using personal microdrones, and publicly available environmental spatial data. It is hoped that factors such as local conditions of flora, soil type, and the degree to which a greenspace is managed will be internalized by the cGAN model.

The following list distills directives related to achieving the goal of using cGANs to model soil conductivity. This variable was chosen because of its ease of attainment, and the amount of information that can be inferred from the single unit. Here, soil electrical conductivity corresponds to "crop yields, crop suitability, availability of plant nutrients, and the activity of soil microorganisms" (Ou et al., 2019).

Itemized Goals:

- Design a methodology that relates image mosaics to soil conductivity readings, while mitigating spatial autocorrelation.
- Test the efficacy of using machine learning to classify soil imagery using k-means clustering.
- Attempt to internalize a k-means clustering algorithm into a cGAN and cross-reference the generated images with ground truth data. Then, infer if this deep learning method is a viable alternative to manual soil collection in novel areas.
- Ultimately, create a cGAN that is able to categorize soil electrical conductivity from natural colour images, and potentially be placed onto drones to lower variable costs for farmers and urban greenspace planners.

1.2 Study Area

Figure 1-1 illustrates the study areas, which are three parks in central Toronto, Canada. The three parks that will be analysed are Riverdale Parks West, and East, and Don Valley Park.

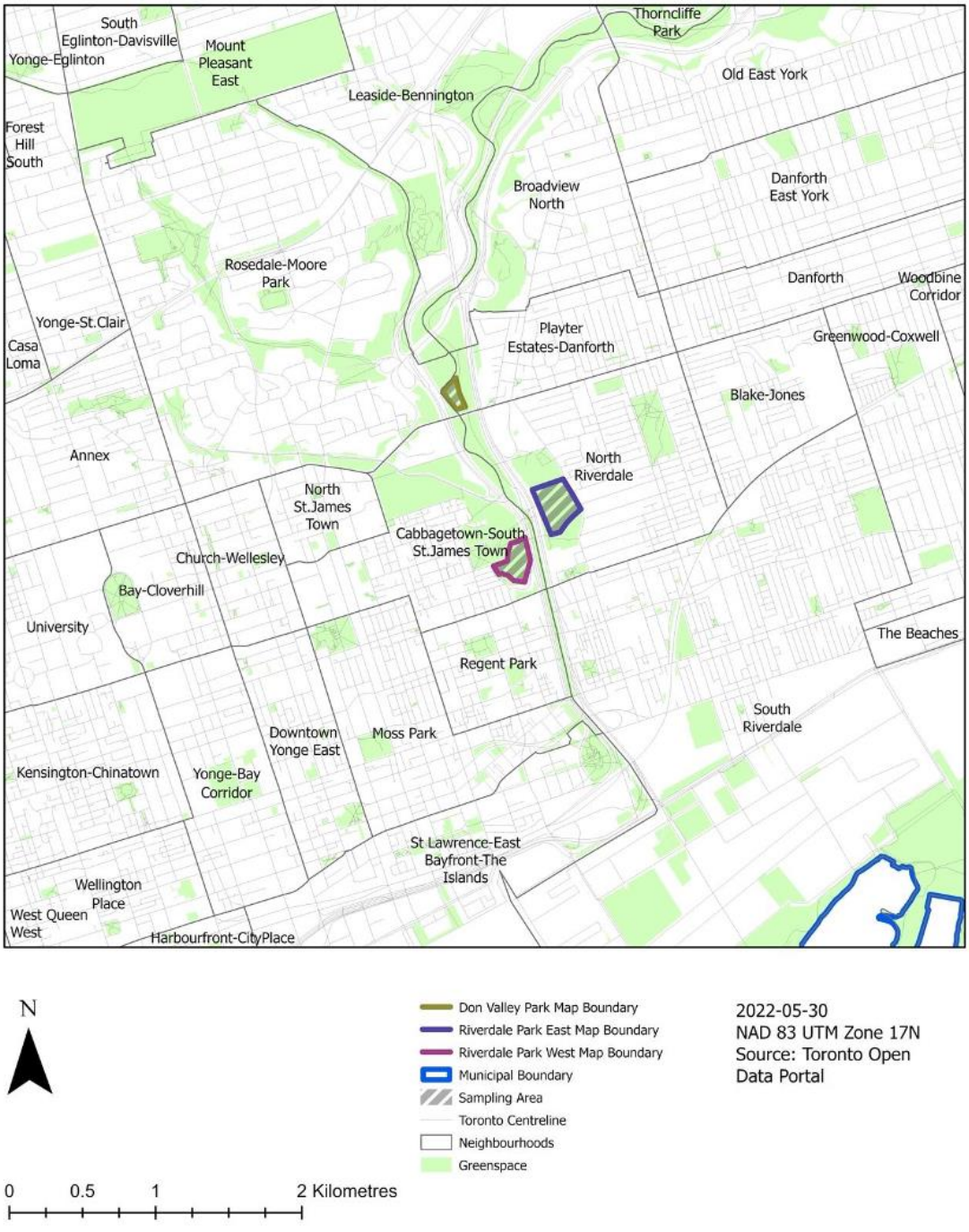


Figure 1-1 Sampling Areas in the Context of Central Toronto

Only a short distance from downtown Toronto, these three parks are centrally located within the city and are close to bus and subway routes. Riverdale Park West most directly serves the residential neighbourhood of Cabbagetown, though it is within walking distance of Yonge, and Bloor Streets. Riverdale East, which is located directly East of Riverdale Park West, serves the community of Riverdale North and is slightly further from Bloor Street. These two parks are quite well managed, as people regularly use these two locations for recreational activities. Hence, the quality of the soil for human significant, agricultural, or horticultural purposes, may be higher than in unmanaged areas. Lastly, Don Valley Park, is north of Bloor Street, and serves the community of the western Danforth. It is neither an actively managed woodland, nor greenspace. Unlike Riverdale Park West, which is known for its wetlands, and Riverdale Park East, which "has developed into a mature maple-ash woodland" (Toronto, 2022a), Don Valley Park contains large amounts of endemic stinging nettle, which is not very conducive to recreational activity.

These areas were chosen because of their proximity to the central downtown area of Toronto, which kept transit costs for the researcher within budgetary limits. Secondly, they are marginal urban greenspaces that could potentially be the site of future agricultural initiatives that make use of remote analysis and cropping. Finally, they are large open spaces with much soil to be analysed, where the drone has a low probability of coming in close proximity to structures, and retains the ability to capture high resolution images.

2 Literature Review

2.1 Our Current Understanding of Soil Sampling

A precision agriculturalist must preoccupy themselves with accurate soil sampling, despite the high cost of equipment, and time investment required. Currently, machine learning is the most effective method of predicting the qualities of untested soil parcels with limited sample data (Zhang et al., 2021). Alas, including "abundant environmental covariate information" in conjunction with a semi-supervised classification produces higher quality results (Zhang et al. 2021). These types of covariate environmental data include "elevation (m), slope gradient (%), plan curvature, profile curvature, relative position index (%), and topographic wetness index", of which, many are available to Canadian researchers (Zhang et al., 2021). Such covariate data could be used in conjunction with local pH, electrical conductance, or temperature values to increase classification accuracy (Zhang et al., 2021).

How soil traverses agriculturally variable land helps researchers and farmers understand the distribution of elements needed to increase crop yield (Menzies Pluer et al., 2020). Unmanned aerial vehicles can be used to segment differences within a digital surface model to predict soil characteristics over larger areas (Menzies Pluer et al., 2020). This use of machine learning for predicting soil quality is supported by Jia et al. (2021) who recommend the use of high-resolution images captured from aerial devices to be used as sample data.

2.2 Using GIS to Analyse Remotely Sensed Imagery

Using enterprise GIS applications like ArcGIS, or the specialised photogrammetry analysis software CATALYST Professional, arithmetic functions can be employed on publicly available datasets of satellite imagery, like the United States Geological Survey's Earth Explorer, to extrapolate agricultural conditions like pH and topsoil qualities (Esri, 2022; PCI Geomatics Enterprises, Inc., 2021b; U.S. Geological Survey, 2022). Alas, Earth Explorer does not provide centimetre scale imagery for free, as would be available with personal microdrone imagery, though they do make available Landsat-9 data without charge, which provides 30m spectral bands of the electromagnetic spectrum, including short wave infrared, and near infrared (U.S Geological Survey, 2020). These electromagnetic bands can be used to arithmetically interpolate factors such

as wetness, and vegetation indices in applications such as CATALYST Professional (PCI Geomatics Enterprises, Inc., 2021a; PCI Geomatics Enterprises, Inc., 2021b).

For this analysis, the best attainable spatial resolution was five cm, which is higher than publicly available satellite imagery online. Xu et al. (2018) supports the use of "Very High Spatial resolution images" such as WorldView-2 (spatial resolution 1.84 metres) to "improve soil model performance". In this case, the microdrone can attain an even higher spatial resolution.

2.3 Issues Regarding Spatial Autocorrelation

Spatial autocorrelation remains an issue in many environmental models where the location of samples cannot be explicitly chosen by a researcher, especially where there is a diversity of pattern (Hu et al., 2020). Where spatial autocorrelation is present, the axiom that samples are independent of one another is no longer true, which introduces bias into statistical analysis, and can result in the erroneous rejection of the null hypothesis (Dormann et al., 2007).

Tools like Dronelink allow a researcher to create flight plans that when commenced autonomously control the drone by precisely guiding them to planned locations where imagery can be acquired (Dronelink, 2022). The grid of a potential flight plan can be utilized to quickly find and define soil collection sites as it can be used to sample points in the study area from a grid (Morrison et al., 2008). In this regard, any grid superimposed over the study area where latitude and longitude coordinates can be parsed is useful (Morrison et al., 2008). This way, the independence of samples can be assured in the data collection stage.

The number of samples to acquire is a concern. There is much variability in GIS literature as most studies regard separate features. For analysis that falls within a broad range of country-level guidelines, Huuskonen and Oksanen (2018) reiterate a minimum temporal resolution of five years, and a spatial resolution of 10 hectares, but concede that this is not enough for precision agriculture. Regarding ecological niche models, a sample size of 10-20% of total localities is shown to produce "similar models" (Boria and Blois, 2018).

In this case, they recommend sampling from "internally consistent" areas that have been split into management zones, while limiting the required training set (Huuskonen and Oksanen,

2018). Moreover, Huuskonen and Oksanen (2018) propose that these maps can be used with smart glasses in augmented reality suites to aid farming endeavours, where different types of categorizations may be projected over the ground to help an operator complete a large variety of tasks.

2.4 Effective Agricultural Regeneration

Giller et al. (2021) express grave concern for the health of future agriculture and soil. There are many qualms derived from poor food system health, including "hunger, poverty, and obesity" (Giller et al., 2021). An industrialized agriculture's over dependence on chemical fertilizers, and pesticides has begotten "poor quality food, environmental degradation, biodiversity loss, [and] exploitive labour relations and animal welfare" (Giller et al., 2021).

That being said, "Regenerative Agriculture" without taking in account the local context may prove problematic (Giller et al., 2021). Their paper uses an "agronomic perspective" which is defined as a "perspective steeped in the use of plant, soil, ecological, and system science to support the production of food, feed, and fiber in a sustainable manner" (Giller et al., 2021). Potentially, this agronomic perspective can extend to the production of the aesthetic greenspace with the employment of machine learning models. Some papers discuss greenspace's demarcated positive effects on urban health outcomes (Markevych et al., 2017; Vries et al., 2003).

Moreover, other artificial intelligence research involving drones regards the amendment of already existing plants. Here, Chen and Li (2019) bring attention to the fact that 1/3 of the world's food supply "relies on animal pollination". They theorize an autonomous pollination effort that is operated through microdrones, and employs artificial intelligence (Chen and Li, 2019). In the case of the United Kingdom alone, artificial intelligence endeavours "could add $USD232BN" to their gross domestic product in the next decade (Chen and Li, 2019). Now, artificial intelligence, especially through artificial pollination, can improve "efficiency in our current farming methods to increase production and reduce wastage" while utilizing relatively less natural capital (Chen and Li, 2019). In essence, plans exist to create robotics enabled smart-farming initiatives stemming from the statistics of artificial intelligence.

2.5 Artificial Intelligence in Sensor-Enabled Robotics for Farming

Computer-vision enabled microdrones can be used to recognize plants for pollination, to iteratively make up for the gap in pollination supply (Chen and Li, 2019). As pollination is "responsible for 90% of the living things on our planet", it will be important to combat the pressing realities of pollination shortage due to pandemic colony collapse disorder, which affects the Western honey bee, and the overall increased demand for pollination services due to the general increase in agricultural activity (Chen and Li, 2019). The supervised classification of different ecosystems can allow these pollination operations to be tailor fit to certain geographies, and specific fauna.

Ahn et al. (2018) also support precision farming with AI driven robots due to growing food security concerns. They postulate that autonomous robots would be especially important in areas where dense brush prevents a global positioning system (GPS) signal from reaching the device (Durand-Petiteville et al., 2018 in Ahn et al., 2018). Computer vision would be especially important for such an endeavour, as the logical next step, harvesting, has been made possible as well, with lettuce "center and stem" through a "peeling and suction mechanism" recognition system having achieved "100% and 81% accuracy respectively" (Hughes et al., 2018 in Ahn et al., 2018). Furthermore, researchers have attained the function of fruit ripeness recognition with deep convolutional neutral networks, in the case of certain peppers, which help in increasing the efficiency of organizing labor efforts (Halstead et al., 2018 in Ahn et al., 2018).

This computer vision would employ a cGAN, which is a computationally intensive task that involves using a neural network to train another neural network (TensorFlow Developers, 2022). This technology can "accurately characterize complex heterogeneous spatial structures" and "can be reused effectively to generate multiple-scale spatial structures" (Chen et al., 2022). In essence, this tool is instrumental in achieving the ability to generate novel soil classification maps from natural colour imagery in a short amount of time. Where soil is heterogenous, this tool may provide time savings.

2.6 Greenspace and Health

Jim and Chen (2006) have calculated a "conservative estimate" of peoples "willingness-to-pay" to use urban greenspaces, in Guangzhou City, China, which is RMB17.40 every month. This is reminiscent of online subscription services and could prove to be a reasonable form of income for municipalities should people come to value greenspace cooperatively with policy makers. People in this area are accustomed to an entry-fee, with the figure "significantly associated with income" (Jim and Chen, 2006). In any event, any potential tax, or cooperative payment scheme could scale with income to provide proportional value proposition.

Vries et al. (2003) state that "people living in urban areas are generally found to be less healthy than people living in more rural areas", with much of "this phenomenon… related to the greenness of people's living environment". People are shown to partake in healthy activities like "recreational walking and cycling" in natural environments rather than urban environments (Vries et al., 2003).

Using computer-vision enabled microdrones, greenspace can be nurtured in urban microcosms to create a healthier human habitat. Behaviours may be altered by the presence of these greened zones, and "the absence of pollution", in the sense that spontaneous recreational activities are encouraged, with people's "mood[s] and ability[s] to concentrate" showing improvements with even only "pictures of natural settings" (Driver et al., 1991; Hartig et al., 1996; Ten Wolde, 1999 in Vries et al., 2003).

The ability for a microdrone to both assess and amend soil quality would greatly assist the ability to meet environmentally conscious goals. Ultimately, Markevych et al. (2017) postulate relationships between urban health, and greenspace will be improved should active efforts focus on three main goals. Ultimately, drone agriculture may be a viable route toward securing these objectives. The first of these objectives is the introduction of greenspace to reduce the incidence of pollution exposure (Markevych et al., 2017). Secondly, marginal greenspace that has suffered from physiological stress should be amended, so that "psychophysiological stress" seen in "hormonal, cardiovascular, and musculoskeletal parameters" can be reduced (Markevych et al., 2017). Finally, areas where greenspace-encouraged sporadic recreational activities occur should be maintained, as "social cohesion within the neighbourhood is related to human health and

wellbeing" (Rios et al., 2012; Fone et al., 2014 in Markevych et al., 2017). In essence, microdrones equipped with computer vision can strengthen these relationships.

2.7 Acidic, Bacterial, and Fungal Soil Amendments to Mitigate Food Scarcity, Plant Disease, and the Runaway Effects of Agricultural Regeneration

Despite anecdotal reports subtropical areas of the New World experiencing large barriers to entry regarding soil analysis, small farms in Europe are well documented as less productive and requiring more costs at the government level (Toma et al., 2021). However, this type of small agriculture is noted as being "an important buffer against poverty for lower income rural households" while maintaining the natural land, community identity/heritage, and biodiversity (Toma et al., 2021).

Alas, the transition from purely sustenance farming to productive farming puts a strain on this conservation, as there is a clear tension between "food availability" and "food system stability" (Toma et al., 2021). Ultimately, there is a clear need to educate younger generations on the value of small farms, and therefore sustainable agriculture (Toma et al., 2021). As micro-drones are a new consumer product, educating prospective farmers in their agricultural uses may prove an effective tactic in increasing the incidence of urban, and small-scale intensive agriculture.

2.7.1 Micro Drones and Seed Pods

As more people adopt farming techniques to live more resiliently, continued sustainable stewardship of greenspace should employ technical analysis applicable to the smaller, marginal settings of urban and household-rooftop producers. This would derive macro-scale improvements in food production, and security, potentially leading to systems for development in autonomous agricultural production.

Mohan et al. (2021) propose that the use of drones to provide reforesting services can help introduce plant biodiversity into a drying area, strengthen natural capital for indigenous people, combat the incidence of soil erosion, mitigate disease, and take carbon out of the atmosphere to fight runaway climate change. These drone-based planting efforts can crop zones that are generally

hard to reach, or considerably labour-intensive like, “post-wildfire reforestation, mangrove reforestation, forest reforestation after degradation, weed eradication, and desert greening.” (Mohan et al., 2021). These proposed strategies can “reduce the lag time between forest stands” and “accelerate” the attainment of “ecological objectives” (Mohan et al., 2021). Satellite imagery can then be utilized to assess the success of such efforts, as well as recognize the areas needing artificial cropping services (Mohan et al., 2021). Mohan et al. (2021) recognize that local “biotic, and abiotic” conditions, as well as the “understanding of species-specific biological traits” are crucial in the enhancement of artificial seeding technology.

With the implementation of machine learning, and drones into farming practices over a long period of time, runaway effects should be considered. Researchers must consider that haphazard changes to larger swaths of arable land can be dangerous, and introduce new forms of plant disease, infestation, and unintended soil changes. Even though deep learning and remote sensing allow farmers to impose precise, and practical changes on the environment based on what is significant to human needs, the long-term sustainability of imposed effects should be considered.

Mohan et al. (2021) describe a “seed pod” which involves the encapsulation of a “processed” seed with “beneficial materials that aid in its dispersal and increase the regeneration success”. In any event, drones would be equipped with these seed pods, or “vessels”, and sent to “microsites” which are defined as “a small portion within an environment… that can facilitate seed germination”. These seed pods would then be altered for “condition-specific needs” and include different mixtures of “rooting substrate, nutrients, phytohormones, and mycorrhizal and bacterial symbionts”. Humic and fulvic acid could also be included in the mix, to stimulate “lateral root growth” and their uptake of nutrients, as well as “aid in the development of sustainable intensification” as per the findings of Canellas et al. (2015). Figure 2-1 shows a seed pod design that is scalable.

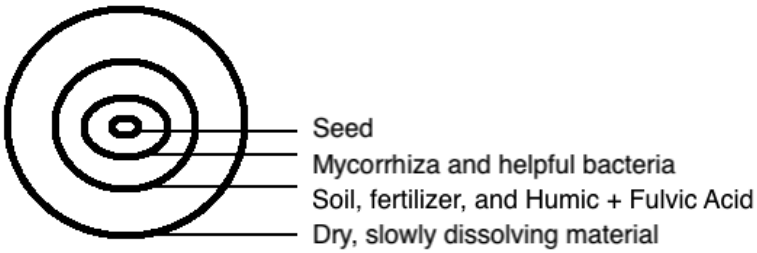


Figure 2-1 Potential Seed Pod Design

2.8 Deep Learning with Conditional Generative Adversarial Networks for Farming

Lohit (2021) describe a process where reforestation jobs can be completed "nine times faster than other human planting systems" by using drones and deep learning analysis. This is done by tasking the drone to "hover on the deforested land… then shoot at seed at that particular location" (Lohit, 2021). Deep learning classifies the study area into five zones, "1. Road 2. Deforested Land 3. Forested Land 4. Rocky Terrain, [and] 5. Buildings" (Lohit, 2021). Python, and TensorFlow are used to create these classes, while mission planning software is used to calibrate and direct the drone to areas of interest (Lohit, 2021).

Uddin et al. (2019) illustrates covariate interpolative data, in the form of Principal Component Analysis, as also being helpful in strengthening classification models. Such a function reduces the number of dimensions, or factors (such as elevation, and reflection) present into variables that do not correlate highly with one another, to potentially increase the significance of the output model (Uddin et al., 2019).

Inoue (2020) describes soil analysis using drones as a "smart farming" technique that "support[s]… efficient, sustainable, and profitable crop production". Inoue (2020) states that 10m or higher spatial resolution is appropriate for "agricultural regions of relatively small farmlands", which corresponds with the needs of precision regenerative agriculture. A "constellation" of satellites or drones would support the statistical tools that work in conjunction with "big-data science" and seemingly provide the highest possible spatial and temporal resolutions for "big data

for agriculture". Spectral analysis, and indices can be used to define and standardize the optimal amount of fertilizer content present in produce to provide farmers precise knowledge of how much fertilizer is needed, as well as the best time to harvest (Inoue, 2020). These indices include the normalized difference spectral, and vegetation indices (Inoue, 2020).

Inoue (2020) supports these claims with the real-life example of France's FARMSTAR service, which is "used by 18,000 farmers" who control 800,000 hectares of agricultural land. This service attains the imagery, and provides diagnostic information to these agriculturalists, which consists of "consultation for optimizing practical management of wheat, rapeseed, and other crops" Inoue (2020). The use of this service has shown a demarcated increase in productivity of "150-250 Euro" per hectare, per year, "a 10-15%" increase of viable crop growth, and a "10-17%" reduction in fertilizer application – all for "10 Euro" per hectare per year (Inoue, 2020).

Tamayo et al. (2020) incentivise using many drones to analyse agriculturally viable land with a theorized modular design. This design includes sun-powered recharging stations for drones limited by battery life, which is a common concern for the "surveillance of large, inaccessible areas", and would be included as a stop within its flight plan during the analysis mission (Tamayo et al., 2020). Tamayo et al. (2020) explore different configurations of recharging pads to optimize flight time.

Chen et al. (2022) show that conditional generative adversarial networks are generally suited to "data-driven geosciences". For this paper, a cGAN is used to extrapolate soil conductivity data to larger areas. Interestingly, pattern simulations have been composed for geological, and hydrological processes using these deep neural networks, with satellite imagery being used recently to simulate fluvial structures (Chen et al., 2022).

2.9 Conclusion of Literature Review and Current Research Gaps

There is much to be discovered regarding optimal soil sampling rates, as there is probably not one concrete method that is applicable to all forms of productive agricultural operations. These forms would be of natural park conservation, agricultural analysis, and the reintroduction of endemic biodiversity into marginal urban greenspace. These types of zones pose different characteristics of accessibility, biodiversity, and human significance, which would alter the

constraints of any study. Alas, the present model will attempt to create a robust method of soil factor attainment, with the goal of large-scale standardization of human-significant soil characteristics for the internalization of categorizations into a deep learning model.

In this case, those interested in precision regenerative agriculture would only need to procure a natural colour map of their study area and submit it into a one-click program for a categorized image. This would provide farmers who lack the finance or knowledge for such analysis, the ability to avoid the burdens of having to attain a personal drone capable of capturing covariate environmental data. This may involve licensing in some jurisdictions, due to the attachment of heavy spectral imaging equipment. Having to produce introspective factors such as normalized difference indices or principal components, and the need to complete their own k-means clustering classifications can also be avoided.

The reality is that these advanced geospatial services are generally available to those in more economically developed nations. It is imperative, the mitigation of the digital divide of precision regenerative agriculture tools, by the provision of easily accessible, open-source software.

3 Methodology of Mapping, Sampling, and Supervised Classification

Various marginal urban greenspaces in the City of Toronto, namely parks, are mapped using a DJI Mavic Mini 1 in conjunction with the flight planning software Dronelink's mapping function. This drone weighs 249 grams, has a 12-megapixel camera, and can capture 4000x2250 images at a 16:9 aspect ratio in red, green, and blue (DJI, 2022). These maps will be used to create supervised classifications of soil electrical conductivity with each image mosaic. These classifications will be used in conjunction with the natural colour imagery to create a generative model that converts the image mosaics into classified images relatively quickly. Illustrated in Figure 3-1, this model allows a drone operator to iteratively categorize the soil in novel areas without having to train and classify for each new area's dataset.

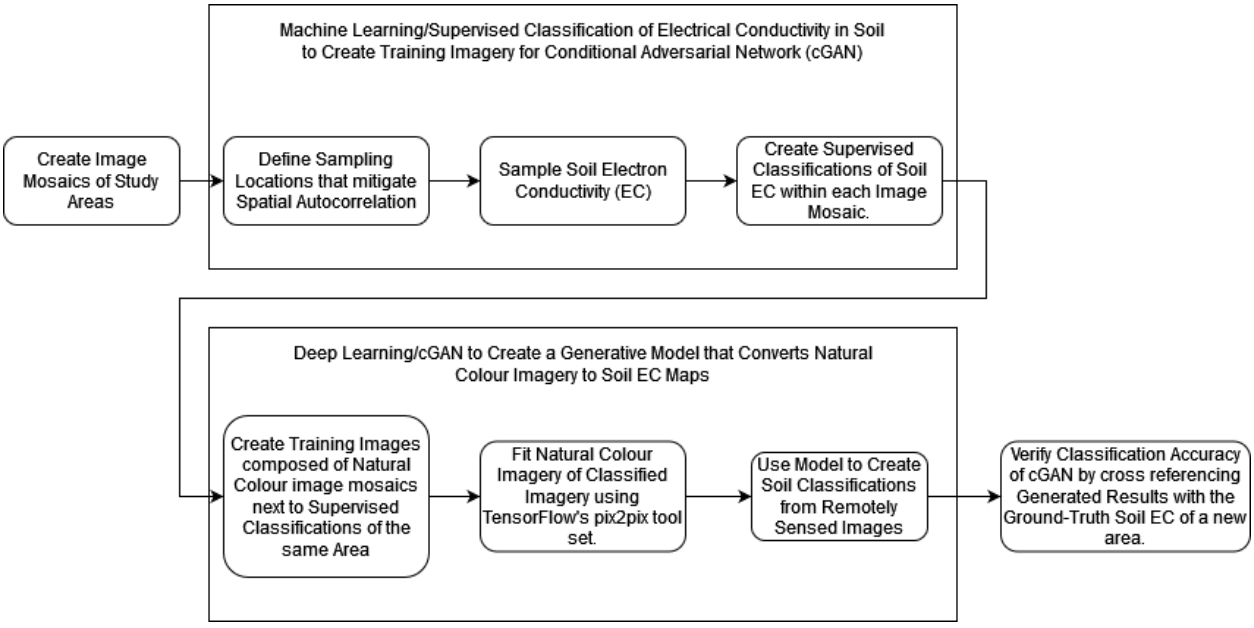


Figure 3-1 Model of Data Collection, Model Training, and Post Classification Accuracy Assessment

Dronelink allows an operator to specify the overlap of each image taken, so an overlap of 80% is used. Flight durations are planned with battery life in mind, due to the drone's small size, so operations are generally kept short – which limits the number of images that can be captured. This is one of the limitations of using a microdrone for photogrammetry. Also, the Mavic Mini

remains stable “at in wind speeds of up to 8m/s”, so it is imperative that weather in clear even more than other remote sensing contexts (DJI, 2022).

Supervised classifications require training areas to relate broad regions of electromagnetic reflectance to others. Training areas for this analysis are created from the ground reading of soil electrical conductivity. To define these areas, points from a flight plan’s grid are taken to produce a soil sampling square, or polygon, if near an edge. These areas are chosen from a grid of squares, modulated by the boundary of the study area, to passively avoid spatial autocorrelation in the data collection phase.

If, per chance, there are too many grid squares to reasonably take soil samples from, due to a large area requiring many images, every other square may be sampled, or every third, and so on, Ultimately, this method of modulating the latitudinal and longitudinal grid squares of the flight plan to find training areas is supported by Morrison et al. (2008), who state that “grid-based systematic designs were more efficient and practically implemented than the others”, when considering the sampling of areas where “natural variability is high”.

For consistency, images and soil samples were acquired at approximately midday during clear weather conditions. These images and soil samples must be captured on a same-day basis. This is due to the high-cost of purchasing high resolution satellite imagery. Purchased images may also not correspond to the day when soil samples were acquired. Micro-drones allow a researcher to create very high-quality image mosaics of any area of interest for only the cost of a computer and a drone.

For the supervised classifications, their kappa coefficients are indicators of “agreement between the predicted and experimental results” and should be considered (Li and Zhang, 2017). The first study area, Riverdale Park East, is illustrated in Figure 3-2. Figure 3-3 shows the flight path that was taken by the drone during the mapping mission.

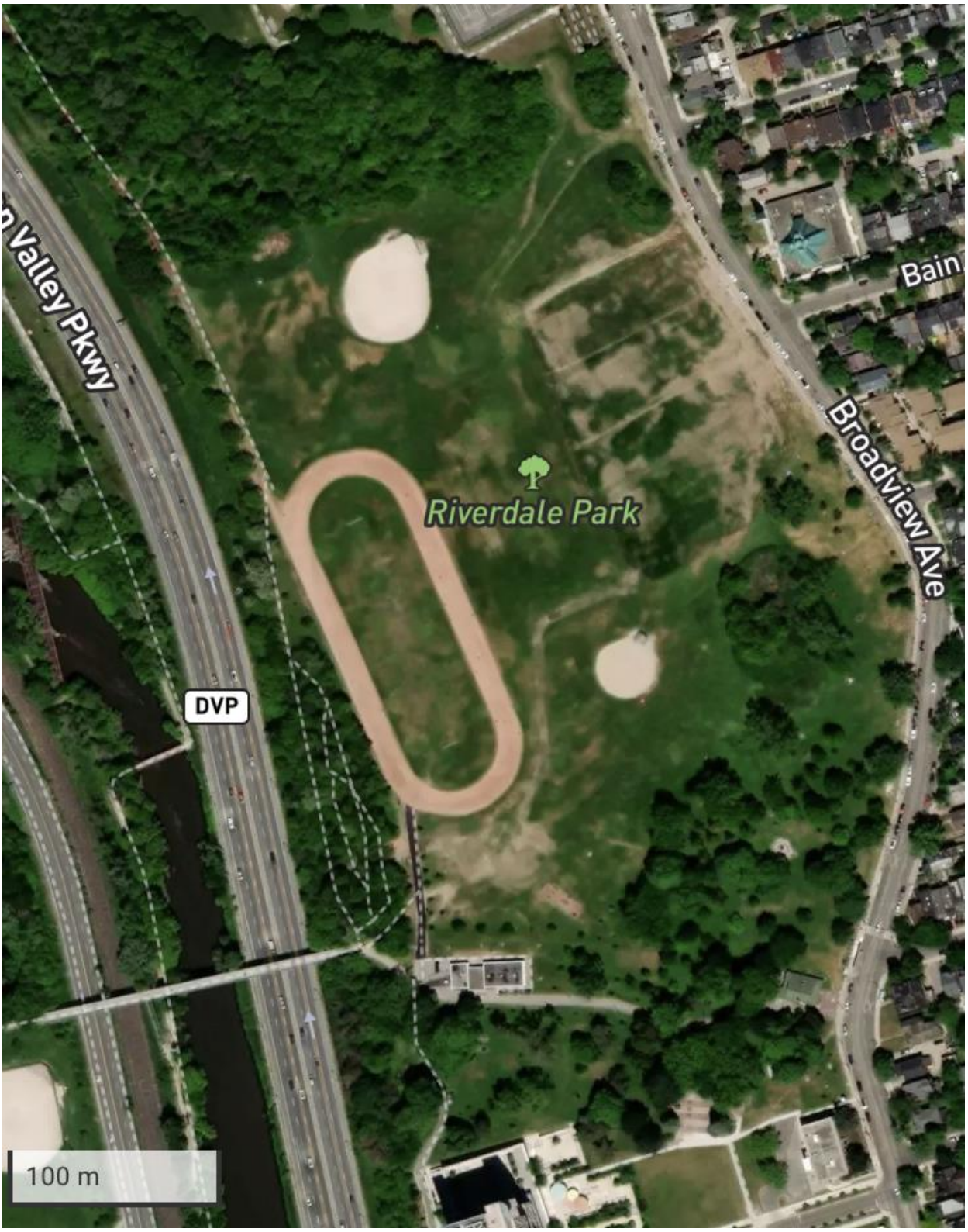


Figure 3-2 Image of Study Area – Riverdale Park East

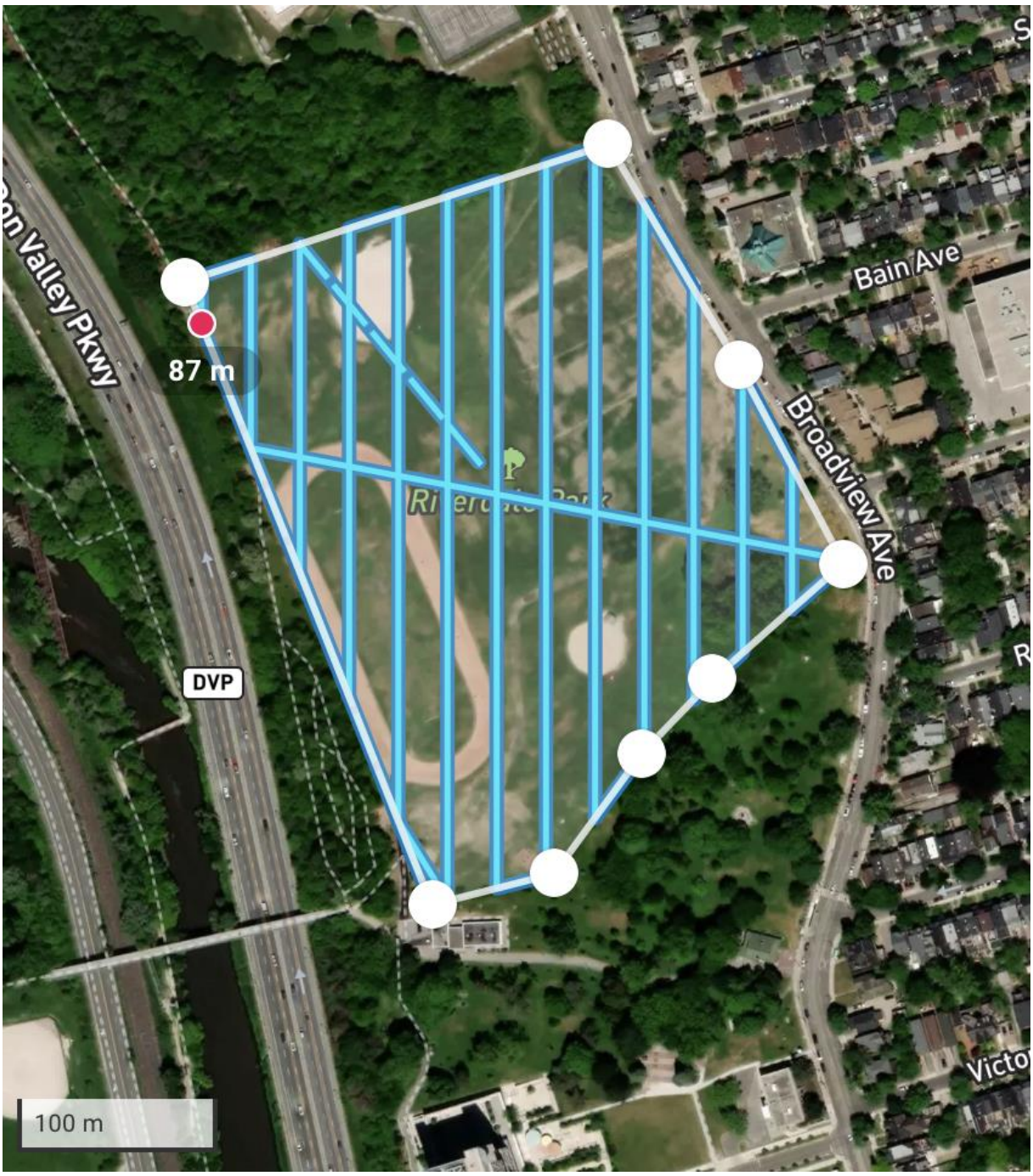


Figure 3-3 Riverdale Park East Flight Plan

The training sites created in ArcGIS Pro, were imported into CATALYST Professional. Figure 3-4 shows the approximate mapping areas, and sampling sites in a larger context. Open

Drone Map, the application used to create the image mosaics of the study areas, also outputs digital surface models of each area. (OpenDroneMap 2022). These will have the effect of strengthening model fit.

Figure 3-4 Sampling and Mapping Sites in the City of Toronto. (Toronto, 2019; 2022; 2022b)

In order to state there is a relationship between the electromagnetic reflectance values present in the natural colour imagery, and the generated classifications of the cGAN, spatial autocorrelation must not be present. The Moran's I statistic can be used to calculate if these clusters are statistically significant to determine the degree of spatial autocorrelation (Zhang et al., 2008).

Regarding the image data that was captured for this analysis, the only sensors available on the Mavic Mini are correspond with the Red, Green, and Blue bands of the electromagnetic spectrum (DJI, 2022). Including an infrared camera on this micro would go over the 250g weight limit (Transport Canada, 2021).

On the Dronelink mission planner, the mapping mission can be set to grid mode, which conventionally creates a much longer flight plan in duration. The software was used to derive the latitudes and longitudes of each training vector thus enabling a grid-based sampling methodology that avoids higher spatial autocorrelation in the presence of "natural variability" (Morrison et al., 2008). Figure 3-5 illustrates these sample sites with green dots. Furthermore, the grid squares from where the sample sites are selected are marked with larger orange dots.

Fortunately, spatial autocorrelation can be seemingly avoided in this analysis as both sample sites, as well the locations where images were acquired, can be actively collected in a grid pattern.

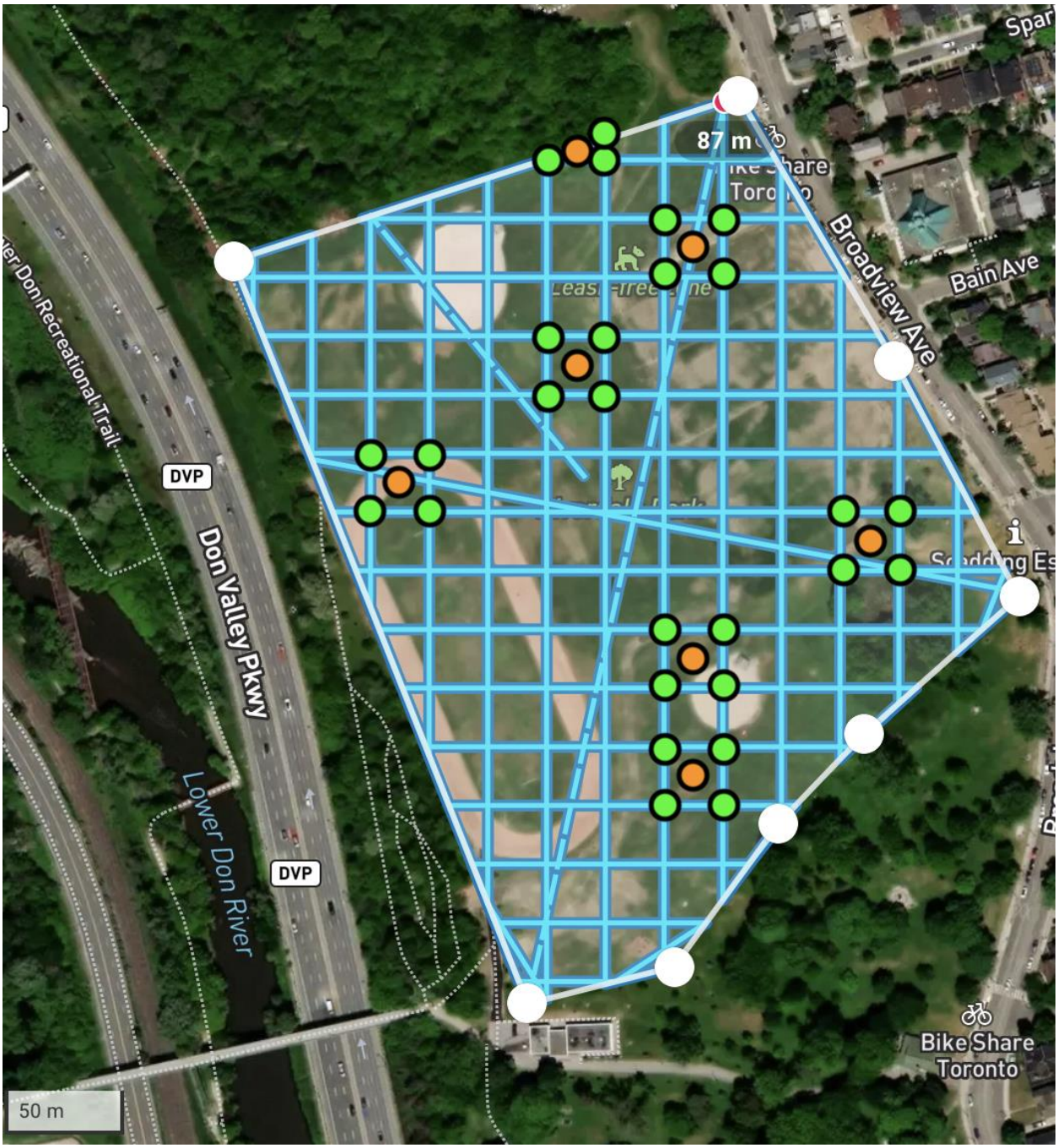


Figure 3-5 Flight plan, in grid mode, with superimposed green dots representing sample sites, and larger orange dots marking the grid squares from where they are attained.

Starting from the first, northernmost, and triangular polygon created by the grid, the latitude, and longitude coordinates of each vector of every 20th polygon are plotted as a soil sampling points. This area contains 27 eligible vectors/sampling sites from seven modulated grid

squares. As an example, Table 3-1 shows the training coordinates that were sampled in Riverdale Park West. They are hierarchically labelled under the modulated grid square that they are the vectors of. That is, squares of equal area superimposed over the study area, and soil samples are taken from the corners of each square. Spatial correlation is ultimately avoided because the squares are generally equally spaced out from one another.

Table 3-1 Sample site vectors for Riverdale Park's training areas which have been attained from the mission planning interface.

```
1
a 43.671103 -79.355424
b 43.671034 -79.355703
c 43.671039 -79.355416
2
a 43.670811 -79.355123
b 43.670816 -79.354814
c 43.670597 -79.355107
d 43.670589 -79.354804
3
a 43.670354 -79.355720
b 43.670359 -79.355404
c 43.670147 -79.355720
d 43.670152 -79.355426
4.
a 43.669934 -79.356638
b 43.669934 -79.356329
c 43.669711 -79.356630
d 43.669716 -79.356329
5
a 43.669716 -79.354200
b 43.669700 -79.353884
c 43.669488 -79.354207
d 43.669499 -79.353906
6
a 43.669254 -79.355110
b 43.669260 -79.354802
c 43.669052 -79.355103
d 43.669052 -79.354802
7
a 43.668835 -79.355110
b 43.668835 -79.354809
c 43.668622 -79.355110
d 43.668617 -79.354809
```

For this analysis, an electrical conductivity reading was taken at the ground. Subsequently, for the supervised classification, the electrical conductivity measurements corresponded with a training area that was a two-metre buffer around the sampling site. This seemed to be a generally homogenous area of soil across the three mapped areas and produced classification accuracies that were acceptable to continue the analysis.

To create the highest resolution image mosaics, drone altitude and image overlap were calibrated based on the findings of Dandois et al. (2015), who report in their modelling of forest structure that "no significant differences were observed at different levels of lighting, altitude, and side overlap". The highest quality "estimates of canopy height… were obtained under optimal conditions of clear lighting and high image overlap (>80%)" (Dandois et al., 2015). Therefore, drone altitude for each study area was based on the avoidance of structures, while images were always captured with at least an 80% overlap.

Figure 3-6 illustrates the second study area, Riverdale Park West, with training sites, and their relative grid squares marked in similar colouring.

Figure 3-6 Image of Study Area – Riverdale Park West, with training areas

Accessibility and clearance of hilly terrain are considered with knowledge of the area. Therefore, this flight is to be conducted at 100m instead of ~80m. Nevertheless, seven training sites were attained (as with the first study area). Alas, the northernmost potential training site, from

where the grid is modulated to avoid spatial autocorrelation, was chosen because it contains the first ground-accessible sampling vectors/sites. In this case, the more northern intersections are available for aerial analysis, but are blocked on the ground by path, and wall. Here, every eighth grid square was chosen to be a potential training site. Finally, Figure 3-7 outlines the flight plan and mission details for this site.

Figure 3-7 Riverdale Park West Flight Plan with mission details. (100m with 80% image overlap)

The third, and final study area is Don Valley Park, an area situated just north of the Don Valley Parkway. Figure 3-8 illustrates this study area and flight plan, while Figure 3-9 shows where soil will be sampled. In this case, every 4th grid square is chosen to attain 18 sample vectors.

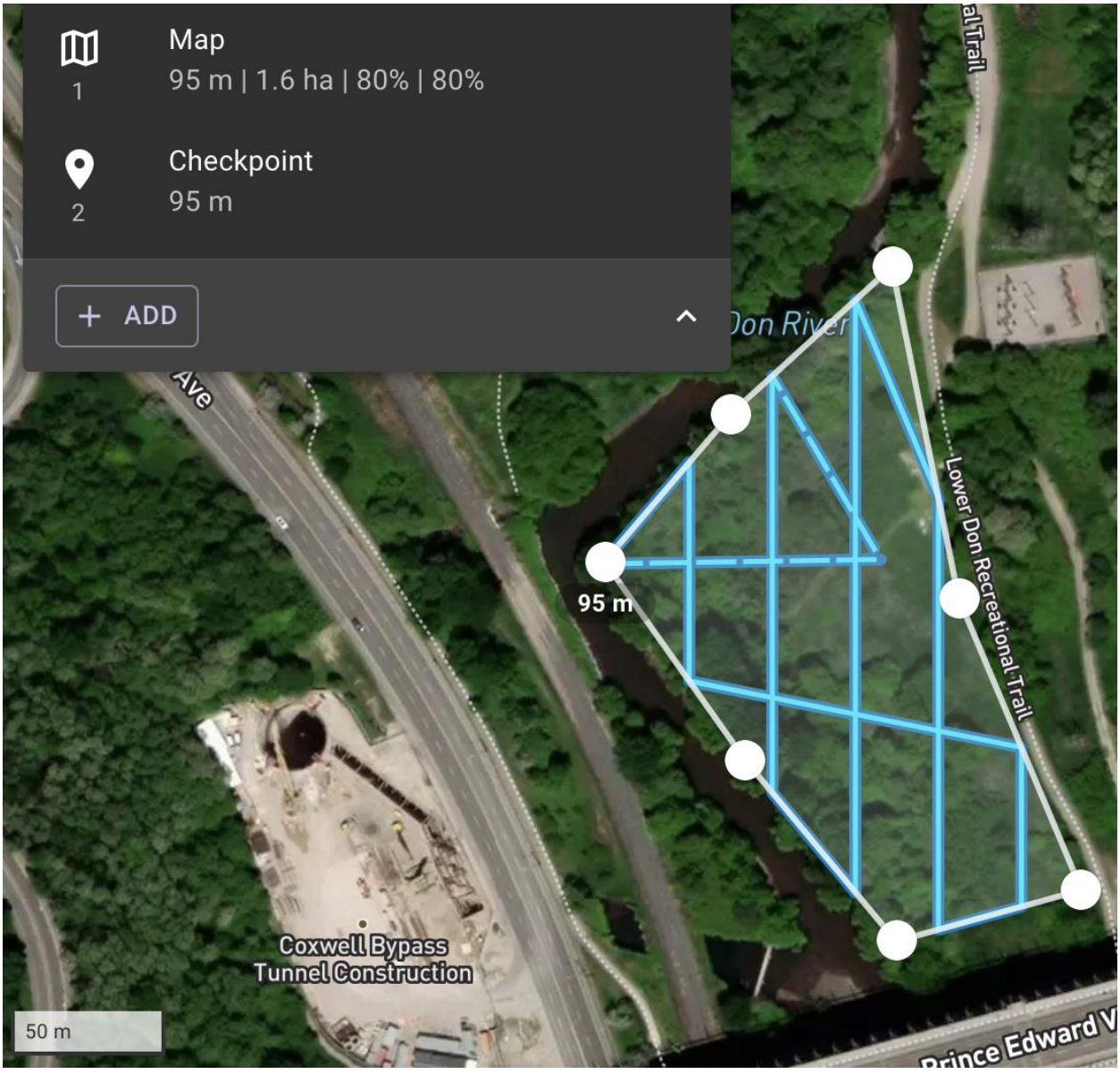


Figure 3-8 Don Valley Park Flight Plan and Mission Details (95m with 80% image overlap)

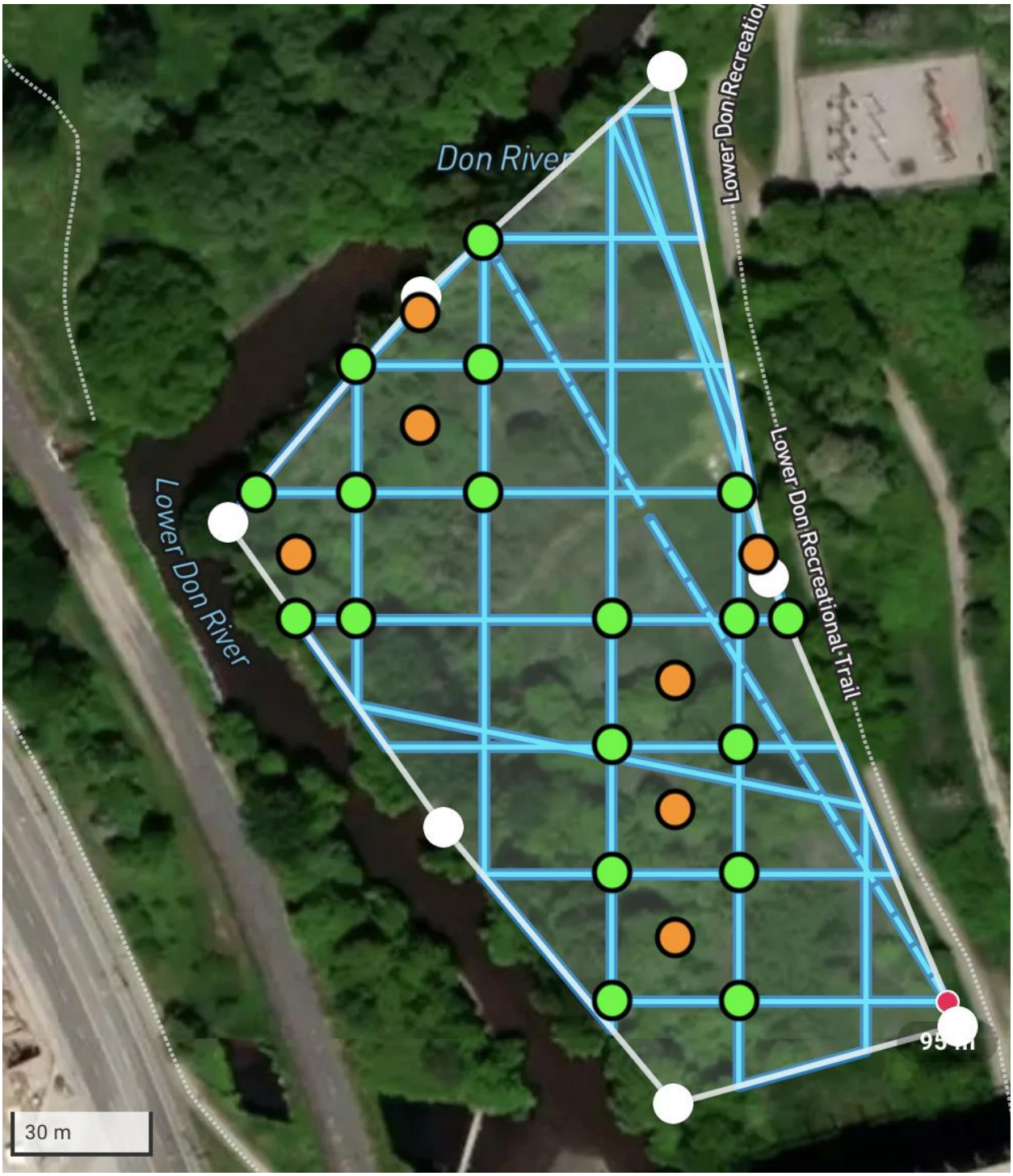


Figure 3-9 Don Valley Park Soil Sampling Sites

3.1 Training Sites, EC Readings, and Supervised Classification

At each training site, a soil electrical conductivity probe was inserted 7.62 cm (3”) into the soil. The reading is given 10 seconds to stabilise, and the result is joined to the latitude and longitude coordinate in the database. An attempt was made to collect all readings as quickly as possible, to reduce environmental variability due to changes in weather.

Additionally, a texture layer, showing the homogeneity of the red band, is created for each raster with a search radius of 10 metres, as, at this scale, that search radius resulted in less noise, and a higher classification accuracy. Then, a Principal Component Analysis is completed using these five bands, with an output of three eigenchannels. These artificial bands are used to increase the classification accuracy of the supervised classification.

Finally, three supervised classifications were completed with and without the supplemental bands, in order of their creation, to find the highest classification accuracies and Kappa Coefficients. For this classification, both the orthophotos, and digital surface models of the three zones are merged into single raster datasets.

Here, the electrical conductance values are categorized by quantiles into three groups. These groups, Low, Medium, and High, are coloured with Red, Orange, and Green, respectively. For added accuracy, ‘Built-Area’, and ‘Water’ classes were created to mask these areas that cannot support plant life. Figure 3-10 illustrates the training sites, to scale, that will be used for the supervised soil classification of Riverdale Park East.

Figure 3-10 Example of training areas, to scale, consisting of 2m buffers around each sample region (Toronto, 2019; 2022; 2022b)

3.2 Deep Learning Generation of Soil Quality Maps from Natural Colour Imagery

Finally, an attempt was made to internalize the model into a cGAN. Using deep learning software, the natural colour maps, created with drone imagery, will be fit to the categorized maps, created in GIS software. If successful, one natural colour map can be used to instantly create one precise soil classification map.

Using TensorFlow, once the natural colour images have been classified, a training dataset is created for an image-to-image generative model (TensorFlow Developers, 2022). In this case, TensorFlow's pix2pix toolset, which is a conditional generative adversarial model, is used (TensorFlow Developers, 2022; TensorFlow, 2022). Figure 3-11 shows an example of the training data used for this model. Ultimately, this model creates soil quality maps from natural colour images by internalizing the difference between the photogrammetry, and supervised classifications.

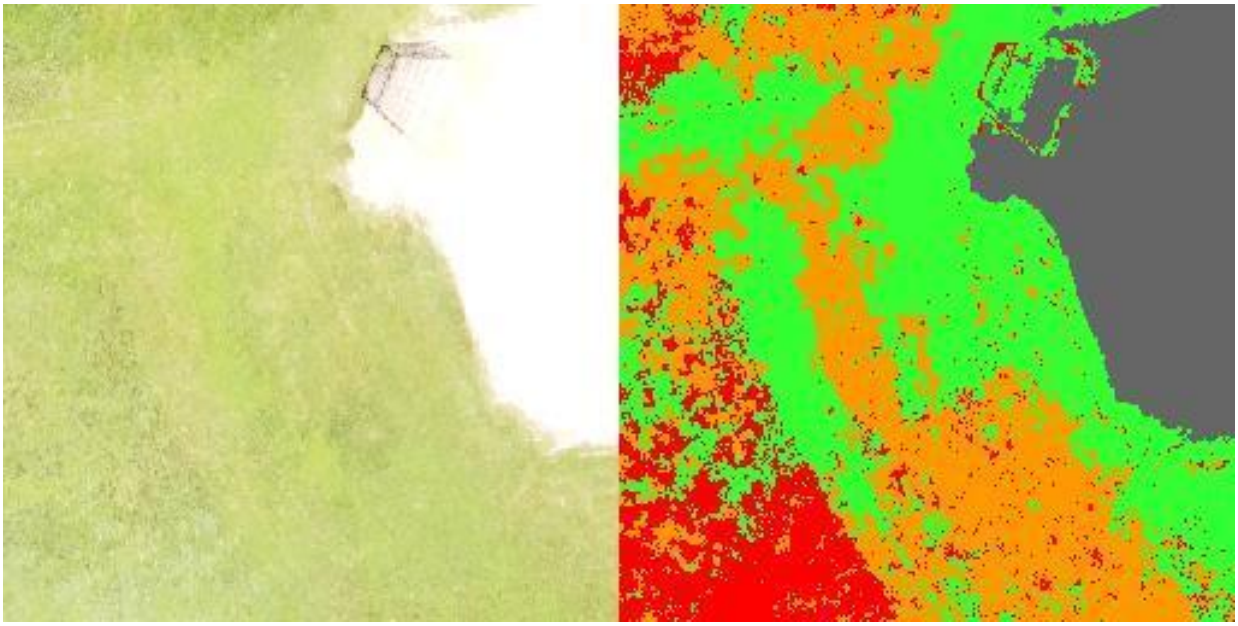


Figure 3-11 Example of training image used for generative image model

For the generative model, 210 training images were created from the natural colour and classified images of the three study areas. These images are flipped vertically, and horizontally to triple the size of the training set. This allowed the cGAN to internalize soil characteristics while

ignoring the directional cardinality of the images. The training set consists of 1:500 and 1:900 scale images. Maps created from the generative model should aim to for input images of these scales.

Finally, another image mosaic was created to test the prediction accuracy of the generative model. This site is sampled like those of the training set but is used solely to cross reference the measured soil qualities with those generated using TensorFlow.

In essence, this cGAN operation will result in a geospatial "image synthesis" that aims to produce below-topsoil predictions similar to "subsurface geological models" seen in certain modern geospatial analyses that employ similar techniques (Azevedo et al. (2020) in Chen et al., 2022). This image synthesis, or generation, results in soil classification maps as depicted on the right in Figure 3-11. These geospatial models are generated using only natural colour images, such as the left side of Figure 3-11, as input data. Alas, "large-scale structures" require "deeper neural network architecture and a larger number of features" which require a relatively large amount of graphical processing (Chen et al., 2022).

4 Analysis

The maps of Don Valley Park, Riverdale Parks East, and West were attained at a five cm spatial resolution and aggregated to a single raster file. Table 4-1 details the results of the supervised classifications. Here, the inclusion of the digital surface model layer improves overall classification accuracy. Table 4-1 details the total area, and spatial autocorrelation statistics for the three areas. Here, the highest coefficient is seen when all supplemental layers are included in the model.

Table 4-1 Supervised Classification Outcomes

	Average Accuracy	Overall Accuracy	Kappa Coefficient and Confidence Interval at 95%	Standard Deviation
Red + Green + Blue (RGB)	67.46 %	62.01 %	0.62725 +/- 0.00037	0.00019
RGB + Digital Surface Model (DSM)	71.38 %	64.60 %	0.64738 +/- 0.00047	0.00024
RGB + DSM + Homogeneity (TEX)	71.60 %	64.76 %	0.64836 +/- 0.00048	0.00024
RGB + DSM + TEX + PCA	70.34 %	64.76 %	0.65085 +/- 0.00041	0.00021

These figures imply that both the producer, and the model are reliable around 2/3 of the time. Here, the Kappa Coefficient is a measurement of interrater reliability between -1, and 1 (McHugh, 2012). In any event, a coefficient of 0.65, alongside the average, and overall accuracies of 70.34, and 64.76%, respectively, are indeed moderate. Li and Zhang (2017) consider the kappa coefficient to be moderate between 0.4, and 0.7, and “almost perfect” beyond that. Furthermore, Bogdan et al. (2022) consider kappa coefficient values in the 0.7-0.8 range to be a “Substantial agreement”. Table 4-2 shows the confusion matrix of the highest performing supervised classification.

Table 4-2 Confusion Matrix of Supervised Classification with DSM, TEX, and PCA layers

	Reference Data					
Classified Data	No EC	Low EC	Med/High EC	Built	Water	Totals
No EC	36.31	25.35	25.37	12.97	0	100
Low EC	16.67	61.06	22.25	0.03	0	100
Med/High EC	7.88	11.73	72.73	7.66	0	100
Built	1.19	0	1.08	97.72	0	100
Water	6.04	0	3.48	6.62	83.87	100
Totals	68.09	98.14	124.91	125	83.87	500

Within the confusion matrix of the most successful supervised classification, we can see that the Built, and Water layers have the highest correspondence with their reference data, at 97.72, and 83.87% respectively. These features are quite distinct from foliage, and generally homogenous, which contributes to this relatively higher percentage. Of the soil classes, Med/High EC has the highest percentage in the confusion matrix, at 72.73%, while No EC has the lowest percentage correspondence with its reference data, at 36.31%. Table 4-3 shows that the three sampling areas exhibit only negligible to mild spatial autocorrelation, though the confidence intervals range from 66% to 91%. Furthermore, the overall spatial autocorrelation statistic of all sites is 0.08 with a p-value of 0.31.

Table 4-3 Parameters of the Study Areas

Parameters					Results		
Study Area	Date of Analysis	Area (ha)	Area (sqkm)	Samples	Average Electrical Conductance	Moran's I of Sample Site	Moran's p-value
Riverdale Park East	2022, May 20th	~7.2	0.071	27	0.189	-0.079	0.33
Riverdale Park West	2022, May 30th	~4.3	0.042	23	0.278	0.33	0.09
Don Valley Park	2022, May 31st	~1.6	0.016	18	0.083	-0.04	0.87
All Sites	2022, May 20th – 31st	~13.1	0.13	68	0.191	0.08	0.31

Figure 4-1 illustrates a side-by-side comparison of the image mosaics and the soil/land classifications. The Red band shows areas classified as 'No EC', or no electrical conductivity. These areas are measured as zero (0) conductivity with the soil probe. The Low level, in orange, shows areas with 0.1 to 0.2 EC. Finally, the Medium/High aggregation is coloured in green, and shows areas with 0.3 electrical conductivity or higher.

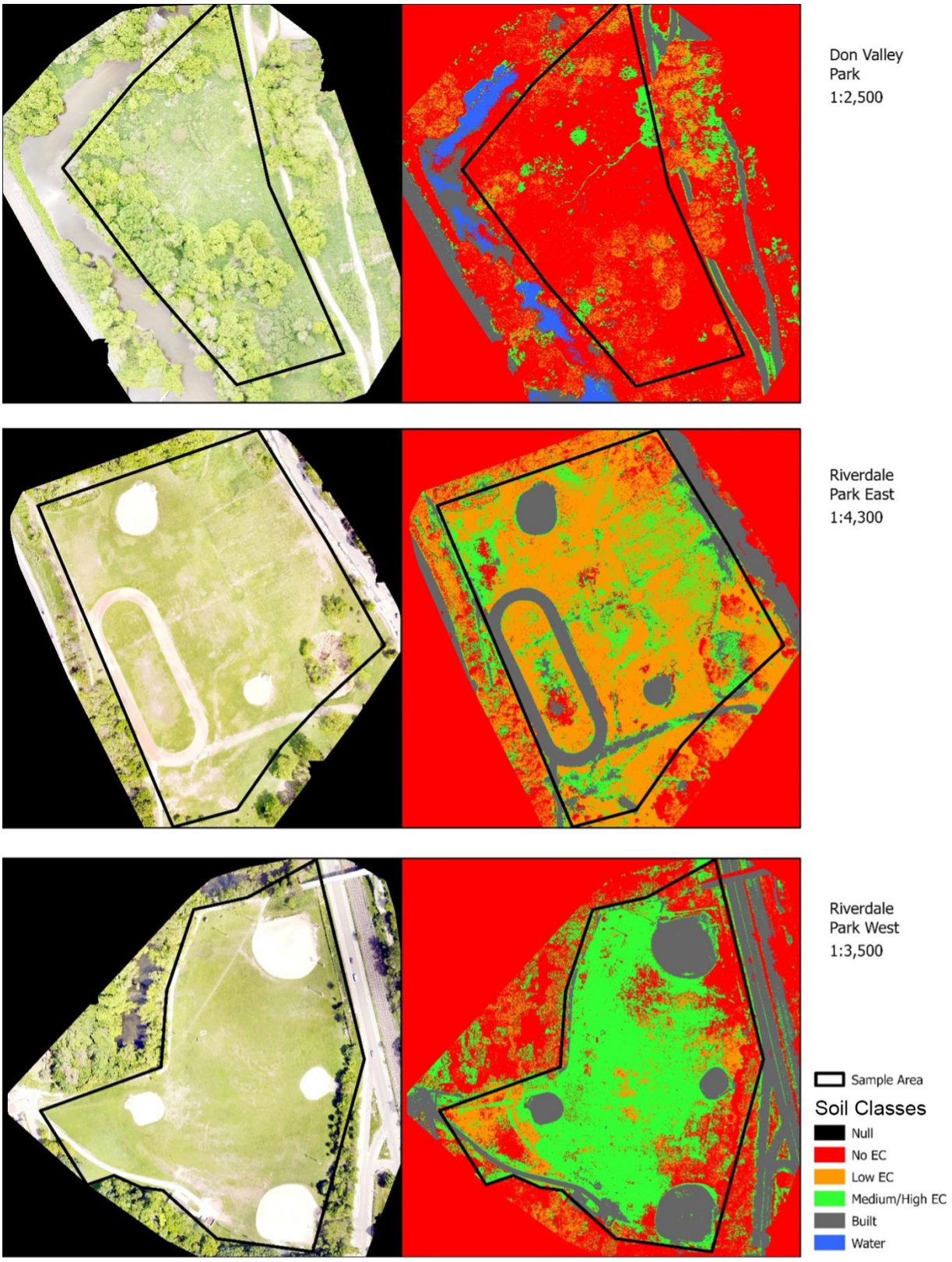


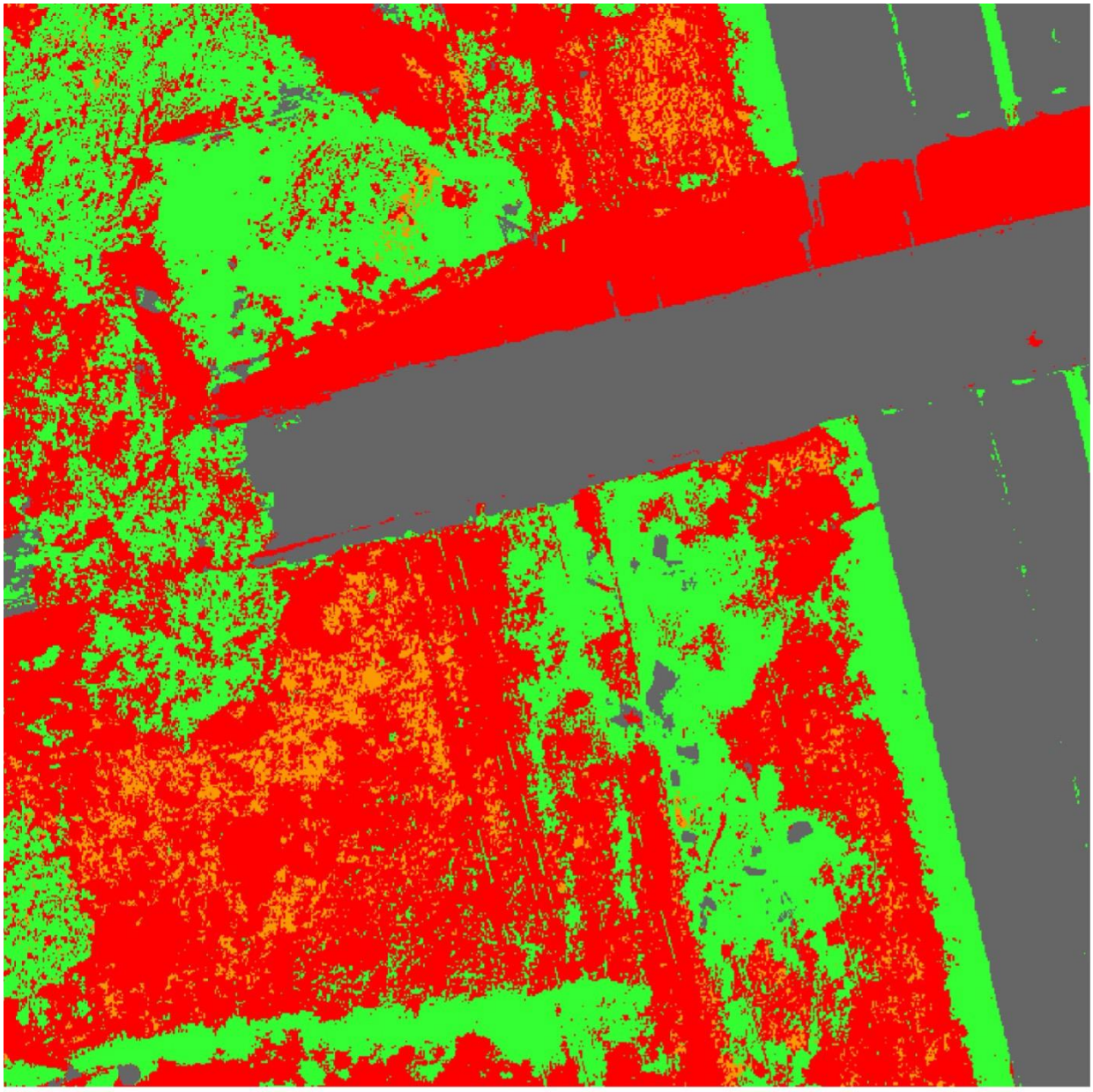
Figure 4-1 Side-by-side view of image mosaics, and Soil Classes (from supervised classification).

Initially, Figure 4-2 shows a natural colour image of the greenspace in question – something that a consumer drone can capture. Subsequently, Figure 4-3 is the result of the original supervised classification of the site. Finally, Figure 4-4 is the attempt that uses TensorFlow to create a supervised classification map from the natural colour image, using only training images like that of Figure 3-11. Figure 4-4 is shown in a lower resolution because this resolution, 256x256, is what the cGAN is calibrated for. For a higher resolution, more computational power is required to train and run the model for output.

1:500

2022-05-30

Figure 4-2 Natural Colour Image of Riverdale Park West, on May 30th, 2022



1:500

2022-05-30

Figure 4-3 Supervised Classification of Figure 4-2's Soil Quality created in CATALYST Professional

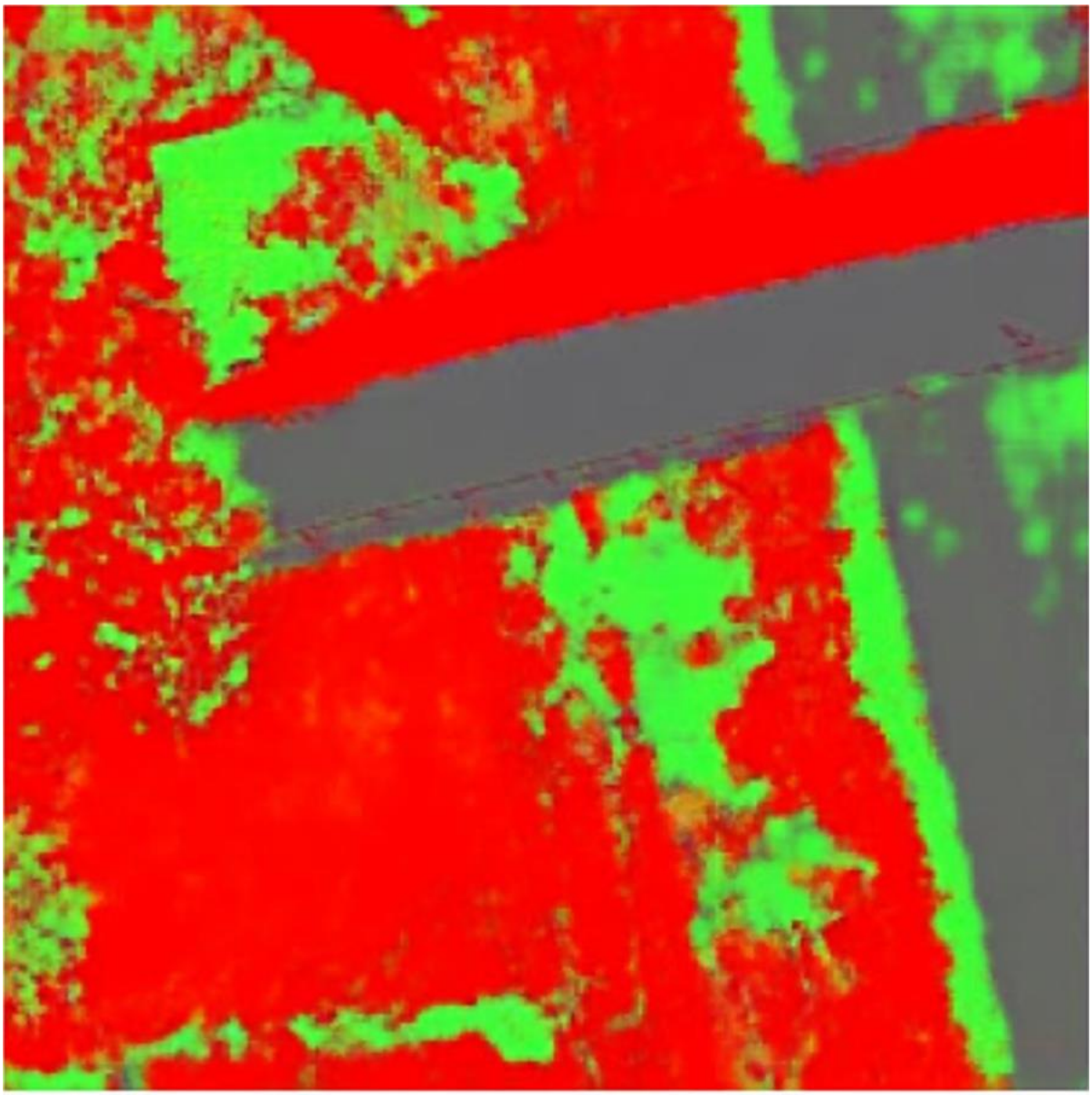


Figure 4-4 Generated Soil Classification Map using TensorFlow's pix2pix toolset.

Figure 4-5 shows a 1:500 scale image of High Park, during the cherry blossom season, acquired on May 5th, 2022. In this image, there are built, green, and barren areas. This is an image that was not included in the training set. Subsequently, it was tested with the generative model, which shown in Figure 4-6.

1:500

2022-05-05

Figure 4-5 A 1:500 image of High Park on May 5th, 2022

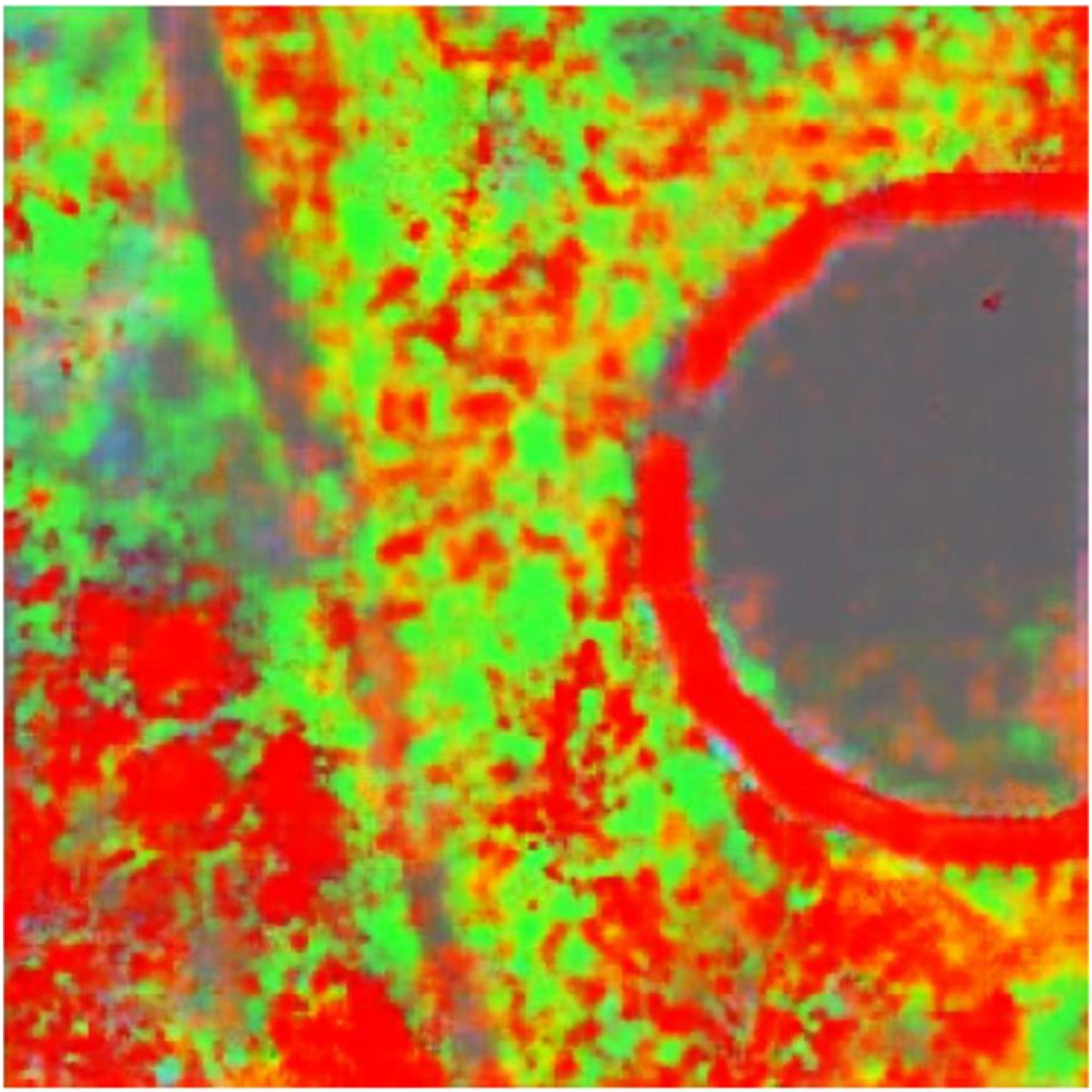


Figure 4-6 A soil classification map generated, using TensorFlow, from the High Park image mosaic.

Finally, an image mosaic with its own soil samples was compared against the results of generated images to provide a post-classification accuracy assessment of the soil classifier. Figure 4-7 shows an image mosaic of a portion of the Evergreen Brick Works, with electrical conductivity

measurements taken at 30 soil sampling points. In this case, the grid-based sampling methodology is also applied to derive soil conductivity validation/accuracy-assessment sites.

Figure 4-7 An image mosaic of Evergreen Brickworks acquired on June 8th, 2022, with the coordinates of the validation soil conductivity sites.

In the sample set, a few locations are inaccessible, or have no reading because the coordinate is in a large body of water, or is on top of a built path, made of inert material. No conductivity readings can be made at these locations.

The post-classification accuracy assessment provided by the validation points at Evergreen Brickworks shows a 33% classification accuracy attained by the generative classification model, with 10/30 generated soil electrical conductivities successfully cross-referenced and shown to the reflect the ground truth. When non-soil classes are removed from the analysis, the generative model exhibits a 9/26 success rate, or 34.6%. Finally, Table 4-4 shows the confusion matrix derived from the success of TensorFlow's generation of supervised classifications, including non-soil classes.

Table 4-4 Confusion Matrix of TensorFlow's Generated Supervised Classification of Soil EC

Soil Class	Ground Truth (n)	Generated (n)	Correct	Accuracy	User's Accuracy	Producer's Accuracy	F1
None	0	12	0	60%	0 (undefined)	0 (undefined)	0
Low	13	10	6	63.33%	0.6	0.46	0.52
Med/High	13	4	3	63.33%	0.75	0.23	0.35
Built	1	4	1	90%	0.25	1	0.4
Water	3	0	0	90%	0 (undefined)	0 (undefined)	0

This confusion matrix confirms that, using TensorFlow's cGAN model, 10 out of 30 generated images were successfully cross referenced as corresponding with the ground truth. The third column shows the count of correctly generated classifications by soil class. These results are interesting, as the model was trained using imagery from different locations. Ultimately, the "Low" electrical conductivity classification shows the highest F1 score within the confusion matrix, which is 0.52. The F1 score is the average of both the User's and Producer's accuracies. The combination of these two scores relates to the strength of the model for each classification from 0-1. Therefore,

the “Low” classification has the highest likelihood of being classified correctly by the model. Moreover, both the “None” and “Water” classes exhibit undefined User’s and Producer’s Accuracies due to there being no incidence of correct classification for these two. Additionally, no samples had zero conductivity, while no water classes were generated by TensorFlow for the test samples. Figure 4-8 shows a cartographical illustration of the generative model’s post classification accuracy assessment.

Figure 4-8 Results of post-classification accuracy assessment including Built, and Water classes.

5 Discussion and Conclusion

5.1 Discussion

The three sampling sites sum to a total of 0.13 square kilometres, whereas total greenspace in Toronto amounts to 121.12 square kilometres (Toronto, 2019). Therefore, for this analysis, 0.001% of total greenspace in Toronto is used as training data. However, if considering only the neighbourhoods that the sampling areas are in, which are Playter Estates-Danforth, Cabbagetown-South St. James Town, and North Riverdale, there are 0.66 square kilometres of greenspace in total. Table 5-1 shows the relative size of the training sample area when compared to the total amount of greenspace in the City of Toronto, and the local neighbourhood region.

These sampling regions do have natural variability but can be considered a coherent biome within a political boundary. The results of the supervised classification are most accurate within the mapped regions, then to a lesser extent in the local neighbourhood, and political regions, and then least accurate when used to generate statistics for peripheral areas. Ultimately, as proximity from the sampling areas increases, the reliability of the model decreases, and Table 5-1's statistics imply this. Importantly, while natural soil electrical conductivity data is noisy, coherence of data can be attained if sampling methodology remains constant.

Table 5-1 Relative Sample Size given the Local, and Broad Geographic Region

Parameters	Sampling Areas		
	Total City Greenspace (Toronto Greenspace Population)	Local Neighbourhood Greenspace (Local Greenspace Population)	Sample Greenspace
Area (sq km)	121.123	0.660	0.13
Relative Sample Size (%)	0.108%	19.740%	100%

It should be noted that the samples were taken at around midday on each respective day of analysis. Even though some areas look to be quite green and are classified as having a low electrical

conductivity, subtle soil conditions reflect factors like “structure, texture…, topography, and the quantity and distribution of crop residues” (Kitchen et al., 2003). Confusingly, in some areas, like Don Valley Park, there are many endemic plants, with dry soil, and seemingly no electrical conductivity. Furthermore, in Riverdale Park West, even though there is a large flat grassland, electrical conductivity seems to be moderate to high in many places, which is probably due to the local maintenance of this park as opposed to the wild growth of the former.

On the date of acquisition, the water bodies in these images were filled with plants. Interestingly, the areas in these water bodies with plants are classified as soil with electrical conductivity, instead of as water. As a result, inherent differences in local flora must be accounted for with a proportional increase in sample size compared to their variance. Also, the generated images produce tertiary hues at the borders between two classes which can introduce bias into the analysis when visually interpreted.

The images were all acquired at a five cm spatial resolution, but some artifacts persist near the edges of maps, especially near trees, due to there being fewer datapoints in those regions. Ultimately, any aerial cropping mechanism would be useless in these canopy regions given a, frankly, non-existent accuracy.

Given the current strengths of the model over bare soil, a drone can theoretically fly over, run the neural network using live images, and iteratively gain insight into the conductivity of the soil in almost real time. In essence, the speed of the model would be limited of the strength of the on-board computer’s central processing unit.

In any event, an overall accuracy of 34.6% leaves much room for improvement. In general, where the topsoil is obscured by dense foliage, grass, or canopy, the readings are not very accurate. The reliability of soil conductivity readings in each location may also be variable, as a single reading is taken from each coordinate. Finally, the readings were taken on the same day, though not at the same exact time. Ultimately, changes in weather may have occurred between the capture of the image mosaic, and the soil conductivity readings, weakening the cGAN’s ability to internalize soil conductivity.

In essence, the ability to analyse areas under the canopy, allowed by 3-dimensional inputs into the k-means clustering classification, may provide notable improvements in accuracy.

Furthermore, as only the red, green, and blue bands of the electromagnetic spectrum were included, different bands should be considered. Though not available for this analysis due to cost constraints, infrared cameras are commercially available for agricultural remote sensing and can be readily included as a factor.

5.2 Conclusion

For all intents and purposes, this generative model allows a soil analyst to run the same supervised classification on any natural colour image. Though, the results here correspond with an urban, and temperate city. Furthermore, training, and validation data are limited to a relatively green region in the summer. However, if the duration between soil samples correlates with the seasonal changes of the study area, soil quality can be approximated elsewhere in the region with a degree of confidence. In any event, the model can be perpetually updated with new training data.

This model is strong because it can be applied using open-source software, and microdrones, which both lower the barrier of entry. That being said, not all places have the same guidelines regarding UAVs. Furthermore, some steps of the process can be improved with stronger hardware, as a dedicated graphics processing unit would speed up the cGAN's training.

This model exhibits a final accuracy of 33% and is quite weak. This may be explained by a lack of input data, or training time. A large time investment must be made into the cross referencing of the final model output with the real world, so it is important that the dataset is trained on as much data as possible before being utilized in the real world.

For greenspace initiatives with a fixed geographic scope, this model can theoretically operate concurrently to a changing climate. In this case, past data could be discarded as quickly as a fixed temporal resolution of the soil data sampling grid. Whereas, for projects that potentially build large datasets over time, a relatively powerful computer would be advisable, with proportional upgrades to major components to preserve the speed of calculation.

Future researchers may explorer similar image to image transfer methodologies with a larger sample set, or with a different sampling, and training image methodology. Furthermore, the strength of the model's fit may also scale with the power of a dedicated graphics card. Conversely,

a more effective sampling methodology may beget models with higher correspondence between the natural colour, and generated imagery, and therefore require less time to calculate.

The results of this research should contribute to deep learning research with remote sensing. There is a large potential for the generation of high quality, and useful data from noisy inputs given the current artificial intelligence tools available.

5.3 Limitations

In general, when considering the generated imagery, areas that are obscured by large foliage beget a high degree of uncertainty for visual analysis. Higher resolution digital elevation models created with light imaging, detection, and ranging (LiDAR) that provide a factor related to soil conductivity that is not obscured by trees may be crucial in increasing generated image accuracy. In general, areas with sparser greenery showed the highest conductivity estimate accuracy. Conversely water showed a very low classification success, with many ponds being classified as having zero electrical conductivity.

The battery life of the microdrone is the limiting factor in this case, as both a longer flight duration, and a better resistance to weather, like wind, would be improved by a stronger battery. With these improvements larger, and higher quality maps can be obtained.

Training site accessibility is another limitation. Where there is a large body of water, it proves difficult, and sometimes impossible to get a local soil electrical conductivity reading. Lastly, the time in between mapping, and the collection of samples should be considered, as there is always a time delay between the two. Should a weather event, like rain, alter the composition of the soil, another orthophoto would be required to retain sampling accuracy.

These limitations may be attributed to the low sample sizes of the training, and validation sets. Ultimately, only four urban green spaces, which are each managed differently, are included in this model. The grid-based sampling methodology supports the large degree of “natural variability” of the included landscapes and can theoretically be applied to areas that can be iteratively included in the model (Morrison et al. 2008). The limiting factor with a larger sample

size would be the strength of the processing power needed to internalize the soil biomes into the model of the cGAN.

Where these generated images are successful creates an opportunity to understand how local factors of growth contribute to the phenomenology of local greenspace. There is an opportunity to attain a broader range of ground truth data to increase interpretability of marginal urban greenspace. Readily available infrared, and LiDAR cameras that can fit onto drones may provide a cost-effective solution of attaining another band of the electromagnetic spectrum for the analysis. Furthermore, research into other types of deep neural networks should be used in conjunction with soil analysis, to find the method which achieves the highest classification accuracy when cross-referenced with ground truth information. Finally, the model must be trained at a higher resolution, which would be possible with more computational power. This model is calibrated on a 256x256 pixel image, which may have introduced artifacts into the input image mosaics and decreased the strength of the cGAN's classification. Ultimately, stronger computers, or more processing time with the available resources would be required to see if higher resolution images provide improved post-classification accuracy results.

Potentially, the inclusion of infrared and LiDAR datasets may tackle the issues seen with this current methodology. Hypothetically, where there is a clear line of site to the ground, without much greenery to obscure the topsoil, the model would perform better. Where the line of site is broken, accuracy sharply decreases, hence the need for three-dimensional data under the canopy, and near to the ground. In any event, this type of analysis likely requires datasets that include the highest resolution input imagery available, higher-dimensional data, as well as a broader range of the electromagnetic spectrum which can be measured by researchers utilizing a drone. Finally, soil conductivity readings must be captured as soon as possible after the images are attained to mitigate subtle changes in weather which may affect soil conditions. While this model's success rate is 34.6% overall, its strengths may help researchers understand the datatypes, and practices that can be used in the future to increase efficacy.

References

Ahn, H. S., Sa, I., and Dayoub, F. (2018). Introduction to the special issue on precision agricultural robotics and autonomous farming technologies. *IEEE Robotics and Automation Letters, 3(4),* 4435-4438. https://doi.org/10.1109/LRA.2018.2871803

Askari, G., Li, Y., and MoezziNasab, R. (2014;2018;). an adaptive polygonal centroidal voronoi tessellation algorithm for segmentation of noisy sar images. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences., XL-2/W3, 65-68. https://doi.org/10.5194/isprsarchives-XL-2-W3-65-2014

Bogdan, P., Nedeff, V., Barsan, N., Panainte-Lehadus, M., Chitimus, D., Nedeff, F., and Tomozei, C. (2022). Applications of the GIS and remote sensing methods in the land use and cover maps. case study: Roman city and the surroundings of the northern side area. Paper presented at the 1-5. https://doi.org/10.1109/EEAE53789.2022.9831397

Canellas, L.P., Olivares, F.L., Aguiar, N.O., Jones, D.L., Nebbioso, A., Mazzei, P., and Piccolo, A. (2015). Humic and fulvic acids as biostimulants in horticulture. *Scientia Horticulturae, 196,* 15-27. https://doi.org/10.1016/j.scienta.2015.09.013

Chen, Q., Cui, Z., Liu, G., Yang, Z., and Ma, X. (2022). Deep convolutional generative adversarial networks for modeling complex hydrological structures in monte-carlo simulation. Journal of Hydrology (Amsterdam), 610, 127970. https://doi.org/10.1016/j.jhydrol.2022.127970

Chen, Y. and Li, Y. (2019). Intelligent autonomous pollination for future farming -A micro air vehicle conceptual framework with artificial intelligence and human-in-the-loop. *IEEE Access, 7,* 1-1. https://doi.org/10.1109/ACCESS.2019.2937171

Chiang, M.M. and Mirkin, B. Experiments for the number of clusters in K-means. Progress in artificial intelligence (pp. 395-405). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-77002-2_33

Dandois, J., Olano, M., and Ellis, E. (2015). Optimal altitude, overlap, and weather conditions for computer vision UAV estimates of forest structure. *Remote Sensing, 7(10),* 13895-13920. https://doi.org/10.3390/rs71013895

DJI (2022). Mavic Mini. Retrieved May 19th, 2022 From https://www.dji.com/ca/mavic-mini

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Daniel Kissling, W., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M., and Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography (Copenhagen), 30(5),* 609-628. https://doi.org/10.1111/j.2007.0906-7590.05171.x

Dronelink (2022). Dronelink. Retrieved May 13th, 2022 From https://www.dronelink.com/

Esri (2022). Tasseled Cap Function. ArcGIS Pro. Retrieved May 13th, 2022 From https://pro.arcgis.com/en/pro-app/latest/help/analysis/raster-functions/tasseled-cap-function.htm

Giller, K.E., Hijbeek, R., Andersson, J.A., and Sumberg, J. (2021). Regenerative agriculture : An agronomic perspective. *Outlook on Agriculture, 50(1),* 13-25. https://doi.org/10.1177/0030727021998063

Hu, L., Chun, Y., and Griffith, D.A. (2020). Uncovering a positive and negative spatial autocorrelation mixture pattern: A spatial analysis of breast cancer incidences in broward county, florida, 2000–2010. *Journal of Geographical Systems, 22(3),* 291-308. https://doi.org/10.1007/s10109-020-00323-5

Huuskonen, J. and Oksanen, T. (2018). Soil sampling with drones and augmented reality in precision agriculture. *Computers and Electronics in Agriculture, 154,* 25-35. https://doi.org/10.1016/j.compag.2018.08.039

Inoue, Y. (2020). Satellite- and drone-based remote sensing of crops and soils for smart farming - a review. *Soil Science and Plant Nutrition (Tokyo), 66(6),* 798-810. https://doi.org/10.1080/00380768.2020.1738899

Jia, X., Cao, Y., O'Connor, D., Zhu, J., Tsang, D.C.W., Zou, B., and Hou, D. (2021). Mapping soil pollution by using drone image recognition and machine learning at an arsenic-contaminated agricultural field. Environmental Pollution (1987), 270, 116281. https://doi.org/10.1016/j.envpol.2020.116281

Jim, C.Y. and Chen, W.Y. (2006). Recreation–amenity use and contingent valuation of urban greenspaces in guangzhou, china. *Landscape and Urban Planning, 75(1),* 81-96. https://doi.org/10.1016/j.landurbplan.2004.08.008

Li, J., and Zhang, Y. (2017). GIS-supported certainty factor (CF) models for assessment of geothermal potential: A case study of tengchong county, southwest china. Energy (Oxford), 140, 552-565. https://doi.org/10.1016/j.energy.2017.09.012

Kim, M., Kim, D., Eom, H., You, S.H., Jeong, J., and Woo, S. (2021). Segmentation of marine forecast zone in korea by using wave data: Based on K-means clustering algorithm. Journal of Coastal Research, 114(sp1), 251-255. https://doi.org/10.2112/JCR-SI114-051.1

Kumar, M., Tomar, R.S., Lade, H., and Paul, D. (2016). Methylotrophic bacteria in sustainable agriculture. *World Journal of Microbiology and Biotechnology, 32(7),* 120-9. https://doi.org/10.1007/s11274-016-2074-8

Kitchen, N.R., Drummond, S.T., Lund, E.D., Sudduth, K.A., and Buchleiter, G.W. (2003). Soil electrical conductivity and topography related to yield for three contrasting Soil–Crop systems. *Agronomy Journal, 95(3),* 483-495. https://doi.org/10.2134/agronj2003.4830

Markevych, I., Schoierer, J., Hartig, T., Chudnovsky, A., Hystad, P., Dzhambov, A.M., de Vries, S., Triguero-Mas, M., Brauer, M., Nieuwenhuijsen, M.J., Lupp, G., Richardson, E.A., Astell-Burt, T., Dimitrova, D., Feng, X., Sadeh, M., Standl, M., Heinrich, J., and Fuertes, E. (2017). Exploring pathways linking greenspace to health: Theoretical and methodological guidance. *Environmental Research, 158,* 301-317. https://doi.org/10.1016/j.envres.2017.06.028

McHugh, M.L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22(3),* 276-282. https://doi.org/10.11613/BM.2012.031

Menzies Pluer, E.G., Robinson, D.T., Meinen, B.U., and Macrae, M.L. (2020). Pairing soil sampling with very-high resolution UAV imagery: An examination of drivers of soil and nutrient movement and agricultural productivity in southern ontario. Geoderma, 379, 114630. https://doi.org/10.1016/j.geoderma.2020.114630

Mohan, M., Richardson, G., Gopan, G., Aghai, M.M., Bajaj, S., Galgamuwa, G.A.P., Vastaranta, M., Arachchige, P.S.P., Amorós, L., Corte, A.P.D., de-Miguel, S., Leite, R.V., Kganyago, M., Broadbent, E.N., Doaemo, W., Shorab, M.A.B., and Cardil, A. (2021). UAV-supported forest regeneration*: Current trends, challenges and implications. Remote Sensing (Basel, Switzerland), 13(13),* 2596. https://doi.org/10.3390/rs13132596

Morrison, L.W., Smith, D.R., Young, C.C., and Nichols, D.W. (2008). Evaluating sampling designs by computer simulation: A case study with the Missouri bladderpod. *Population Ecology, 50(4),* 417-425. https://doi.org/10.1007/s10144-008-0100-x

OpenDroneMap. (2022). OpenDroneMap. Retrieved May 25th, 2022 From https://www.opendronemap.org

Ou, I., Tsai, K., Chu, Y., and Liao, Y. (2019). Self-sustaining soil electrical conductance measurement using a DC-DC power converter. *IEEE Sensors Journal, 19(22),* 10560-10567. https://doi.org/10.1109/JSEN.2019.2931180

PCI Geomatics Enterprises, Inc. (2021a). NDVI. CATALYST. https://catalyst.earth/catalyst-system-files/help/references/pciFunction_r/easi/E_ndvi.html

PCI Geomatics Enterprises, Inc. (2021b). TASSEL. CATALYST. https://catalyst.earth/catalyst-system-files/help/references/pciFunction_r/modeler/M_tassel.html

Purwanto, Yaumidin, U.K., Yuliana, C.I., Nurjati, E., Rahmayanti, A.Z., Cahyono, B.D., and Novandra, R. (2021). Urban farming and food security: Household's adaptive strategy to COVID-19 crises. IOP Conference Series. *Earth and Environmental Science, 892(1),* 12070. https://doi.org/10.1088/1755-1315/892/1/012070

Ramakrishna, W., Yadav, R., and Li, K. (2019). Plant growth promoting bacteria in agriculture: Two sides of a coin. *Applied Soil Ecology: A Section of Agriculture, Ecosystems & Environment, 138,* 10-18. https://doi.org/10.1016/j.apsoil.2019.02.019

Rillig, M.C., Sosa-Hernández, M.A., Roy, J., Aguilar-Trigueros, C.A., Vályi, K., and Lehmann, A. (2016). Towards an integrated mycorrhizal technology: Harnessing mycorrhiza for sustainable intensification in agriculture. *Frontiers in Plant Science, 7,* 1625-1625. https://doi.org/10.3389/fpls.2016.01625

Tamayo, L.V., Thron, C., Fendji, J.L.K.E., Thomas, S., and Förster, A. (2020). Cost-minimizing system design for surveillance of large, inaccessible agricultural areas using drones of limited range. *Sustainability (Basel, Switzerland), 12(21),* 8878. https://doi.org/10.3390/su12218878

TensorFlow Developers. (2022). TensorFlow (v2.8.2). Zenodo. https://doi.org/10.5281/zenodo.6574269

TensorFlow (2022). Tensorflow Core: pix2pix: Image-to-image translation with a conditional GAN. Tensorflow. https://www.tensorflow.org/tutorials/generative/pix2pix

Toma, I., Redman, M., Czekaj, M., Tyran, E., Grivins, M., and Sumane, S. (2021). Small-scale farming and food security – policy perspectives from central and eastern europe. *Global Food Security, 29,* 100504. https://doi.org/10.1016/j.gfs.2021.100504

Toronto. (2019). Parks. Open Data Portal. Retrieved July 23rd, 2022 From https://open.toronto.ca/dataset/parks/

Toronto. (2022). Neighbourhoods. Open Data Portal. Retrieved April 22nd, 2022 From https://open.toronto.ca/dataset/neighbourhoods/

Toronto. (2022a). Riverdale Park East. The City of Toronto – Parks. Retrieved August 2, 2022 From https://www.toronto.ca/data/parks/prd/facilities/complex/343/index.html

Toronto. (2022b). Toronto Centreline (TCL). Open Data Portal. Retrieved May 30th, 2022 From https://open.toronto.ca/dataset/toronto-centreline-tcl/

Transport Canada (2021). Find your category of drone operation. Government of Canada. Retrieved May 12th, 2022, from https://tc.canada.ca/en/aviation/drone-safety/learn-rules-you-fly-your-drone/find-your-category-drone-operation#micro-drones

Uddin, M.P., Mamun, M.A., and Hossain, M.A. (2019). Effective feature extraction through segmentation-based folded-PCA for hyperspectral image classification. *International Journal of Remote Sensing, 40(18),* 7190-7220. https://doi.org/10.1080/01431161.2019.1601284

United States Geological Survey – USGS. (2020). Landsat 9 Fact Sheet 2019-2008. https://doi.org/10.3133/fs20193008

United States Geological Survey – USGS. (2022) EarthExplorer. Retrieved May 13th, 2022 From https://earthexplorer.usgs.gov/

Vries, S.D., Verheij, R.A., Groenewegen, P.P., and Spreeuwenberg, P. (2003). Natural environments - healthy environments? an exploratory analysis of the relationship between greenspace and health. *Environment and Planning. A, 35(10),* 1717-1731. https://doi.org/10.1068/a35111

Xu, Y., Smith, S.E., Grunwald, S., Abd-Elrahman, A., Wani, S.P., and Nair, V.D. (2018). Estimating soil total nitrogen in smallholder farm settings using remote sensing spectral indices and regression kriging. Catena (Giessen), 163, 111-122. https://doi.org/10.1016/j.catena.2017.12.011

Zhang, C., Luo, L., Xu, W., and Ledwith, V. (2008). Use of local moran's I and GIS to identify pollution hotspots of pb in urban soils of galway, ireland. The Science of the Total Environment, 398(1), 212-221. https://doi.org/10.1016/j.scitotenv.2008.03.011

Zhang, L., Yang, L., Ma, T., Shen, F., Cai, Y., and Zhou, C. (2021). A self-training semi-supervised machine learning method for predictive mapping of soil classes with limited sample data. *Geoderma, 384*, 114809. https://doi.org/10.1016/j.geoderma.2020.114809