Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Classifying type-specific green space using satellite imagery and LiDAR to assess associations with diabetes prevalence

Scarlett Rakowska

scarlett.rakowska@mail.utoronto.ca

University of Toronto https://orcid.org/0000-0001-5105-0102

Haoxuan Ge

University of Toronto https://orcid.org/0000-0002-9936-0860

Matthew Adams

University of Toronto https://orcid.org/0000-0001-7236-8419

Research Article

Keywords: urban green space, LiDAR, satellite imagery, diabetes prevalence, classification

Posted Date: May 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9580646/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Urban green space is often measured by using greenness indices, such as the Normalized Difference Vegetation Index, but these indices do not distinguish among green space types. This study integrated 3-metre spatial resolution multispectral imagery and LiDAR-derived normalized Digital Surface Models (nDSM) to classify type-specific green spaces across Mississauga, Ontario, and examine their associations with diabetes prevalence. We tested eight combinations of spectral bands, vegetation and water indices, and structural data using an object-based random forest and found that the best-performing model included all the bands, indices, and nDSM, achieving 88% overall classification accuracy. The classification identified deciduous trees, coniferous trees, shrubs, grass, and marsh/wetland, which were then linked to census tract-level diabetes prevalence using local and neighbourhood beta regression models by sex and age group, and adjusted for socioeconomic covariates. Deciduous trees, coniferous trees, and shrubs were more consistently associated with lower diabetes prevalence, particularly among adults aged 45 years and older. NDVI-based models showed weaker associations with diabetes prevalence after adjustment, while type-specific models retained associations. These findings suggest that type-specific green space and structure provide more informative measures of exposure than greenness alone. Integrating high spatial resolution multispectral imagery and LiDAR provides a methodological framework for health research examining relationships between urban green space and chronic diseases.

1. Introduction

Type 2 diabetes is a serious and rapidly growing public health challenge, and has been a global burden for decades, with an estimated 589 million adults aged 20–79 years living with diabetes in 2014 (International Diabetes Federation, 2025). It is projected to reach 853 million by 2050, with urban residents experiencing the highest burden (International Diabetes Federation, 2025). Lifestyle habits, environmental exposures, and demographics, such as heat, air pollution, noise, and obesogenic food environments, can accelerate type 2 diabetes onset (Bilal et al., 2018; Fazli et al., 2019; Guarnieri & Balmes, 2014; Hill-Briggs et al., 2021; James et al., 2015; Ruvuna & Sood, 2020). Neighbourhood environments are an important determinant of type 2 diabetes, as modifiable environmental features, such as green space, may help prevent its onset (Hill-Briggs et al., 2021; James et al., 2015).

Increased exposure to green spaces is often an intervention that improves environmental, social, and health benefits (Hartig et al., 2014). In this study, green space refers to vegetated land within urban areas, including trees, shrubs, grass, and marsh/wetlands. Green space is a climate-resilient intervention strategy that may contribute to human health (Fong et al., 2018; James et al., 2015). Green spaces reduce heat and improve air quality, which may lower the risk of type 2 diabetes by reducing physiological strain on the cardiovascular and metabolic systems (James et al., 2015; Zhang et al., 2023). Trees and shrubs can reduce temperature through shade and evapotranspiration, capture airborne pollutants such as fine particulate matter ($PM_{2.5}$) and nitrogen oxides (NO_x), and reduce noise (Hartig et al., 2014; Nguyen et al., 2021). Green space promotes physical activity (walkability), which can reduce obesity and type 2

diabetes, and improve insulin sensitivity (Fong et al., 2018; Ismail et al., 2021; James et al., 2015). Green space is also associated with improved mental health, particularly by reducing stress, anxiety, and depression, while elevating mood and promoting calmness (James et al., 2015). However, not all green space is equal. Tree canopies can effectively reduce temperatures and capture air pollutants, whereas grass and shrubs can support physical activity but provide limited shade (Hartig et al., 2014; Nguyen et al., 2021). This heterogeneity in green space suggests that the type and structure of green space, rather than its amount alone, may be critical for understanding diabetes risk.

Systematic reviews emphasize that green space lacks a standardized definition, particularly because satellite-derived measures often cannot capture green space type or structure (Ccami-Bernal et al., 2023). Most epidemiological studies assessing green space rely on green space metrics that do not distinguish among green space characteristics, notably satellite-derived Normalized Difference Vegetation Index (NDVI) or broad land-cover classes. NDVI is a spectral greenness index that uses near-infrared and red bands to estimate live green biomass (Purevdorj et al., 1998). NDVI is widely used in health and environmental studies due to its global availability, long-term temporal consistency, and straightforward computation and interpretation (Martinez & Labib, 2023), including mortality and cardiometabolic outcome studies (Crouse et al., 2017; James et al., 2015). However, NDVI has limitations for health research, as it identifies areas rather than distinguishing between green space types, such as tree canopy versus grass. To identify the structure of green space, we require structural height (Ccami-Bernal et al., 2023; De La Fuente et al., 2020; Patino et al., 2021). Areas with similar NDVI values may have different ecological functions and implications for exposure relevant to health (e.g., canopy versus grass) (James et al., 2015). Recent literature has emphasized that the morphology and structure of green spaces, such as height, may be important for health outcomes because they influence shading, cooling, air pollution filtration, and noise reduction (Hartig et al., 2014; Nguyen et al., 2021; Wang et al., 2024).

Evidence supports the protective role of green spaces in the urban context against type 2 diabetes; however, existing analyses use varying definitions of green spaces (Ccami-Bernal et al., 2023; De La Fuente et al., 2020). One systematic review identified seven of 19 studies evaluating green space exposure and type 2 diabetes, using NDVI to measure green space, and found that people living closer to green space had a lower risk of developing type 2 diabetes (De La Fuente et al., 2020). Similarly, a systematic review of 13 cohort studies using broader green space exposure definitions, including NDVI, EVI, vegetation continuous field (VCF), and green space, either with or without various buffer sizes, also suggests that exposure to green space could be a protective factor against the development of type 2 diabetes (Ccami-Bernal et al., 2023). Several studies in their review showed that exposure to green spaces decreases the incidence of type 2 diabetes, but not all showed significant differences (Ccami-Bernal et al., 2023). The review also highlighted the need for subgroups by age and sex and noted that there is no standardized definition of green space, particularly because satellite measurements do not capture its type or structure (Ccami-Bernal et al., 2023). This may explain why findings in the green space literature are sometimes mixed, as different definitions of green space exposure can yield different conclusions. For instance, one cross-sectional study investigated the relationship between neighbourhood green space and type 2 diabetes, using the percentage of green space within a 3 km

radius of participants' home neighbourhoods, and found a negative association between neighbourhood green space and type 2 diabetes (Bodicoat et al., 2014). However, another study using NDVI as a neighbourhood greenness proxy examined the relationship between the built environment, and cardiovascular diseases and diabetes mellitus, and found that the presence of urban green space did not have a clear relationship with diabetes, but only cardiovascular diseases, and that NDVI cannot measure all forms of natural space or distinguish among green space types (Patino et al., 2021). These inconsistencies suggest that incorporating green space height and structural complexity may provide more information for green space exposure measures than greenness alone.

Light Detecting and Ranging (LiDAR), a laser profiling and scanning system used to measure surface elevations on land and water, collects a cloud of laser range measurements to calculate the three-dimensional coordinates of surface features (Yan et al., 2015). Unlike NDVI, LiDAR-derived height can distinguish tall vegetation from short ground cover, thereby improving land use and land cover (LULC) classification (Jombo et al., 2022; Yan et al., 2015), but it is rarely used in epidemiological and health research. Previous studies that combine high-resolution multispectral imagery with LiDAR to distinguish among green space types have focused on improving classification accuracy rather than examining health outcomes, highlighting a gap in applying these methods to epidemiological research (Jombo et al., 2022; Yan et al., 2015). Numerous studies have shown that LiDAR-derived height features can successfully distinguish tall vegetation from low vegetation areas (Charaniya et al., 2004; Hecht et al., 2008; Huang et al., 2013; Jombo et al., 2022; Yan et al., 2015). Specifically, object-based image analysis and ensemble classifiers are well-suited for integrated datasets from heterogeneous urban environments and have achieved high classification accuracies in previous studies (Hamedianfar et al., 2014; Hartling et al., 2019; Hellesen & Matikainen, 2013; Jombo et al., 2022). For example, tree species in South Africa were classified with high overall accuracy (94–97%) by integrating WorldView-2 satellite imagery and LiDAR using an object-based random forest classifier, with LiDAR as the most important predictor (Jombo et al., 2022). Similarly, high overall accuracy (up to 83%) for tree species classification in a United States metropolitan area was achieved using WorldView-3 and LiDAR, with findings highlighting that spectral and structural information improve mapping in an urban context (Hartling et al., 2019). Despite advances in remote sensing, to our knowledge, no study has integrated these methods to examine health outcomes, including type 2 diabetes, and NDVI continues to dominate as the primary metric for green space exposure (Helbich, 2019; Labib et al., 2020). However, evidence suggests that lower odds of incident diabetes were associated with increases in tree canopy but not in total green space, indicating that total green space may not capture green space types (Astell-Burt & Feng, 2020). These findings suggest that improving the measurement of green space exposure may help clarify associations with diabetes prevalence and whether these associations differ across population subgroups.

Research has shown that the relationship between green space and health may differ across population subgroups (Ccami-Bernal et al., 2023). A recent scoping review has suggested that older adults are more vulnerable to environmental stressors, like heat and air pollution (Montoro-Ramírez et al., 2025), while studies reported that older adults spend more time in their residential neighbourhood relying on nearby amenities due to physical decline, cognitive impairment, and decreasing social ties (Firdaus, 2017; Hu et

al., 2018). Due to older adults' smaller daily activity spaces, neighbourhood characteristics are critical to investigate (Firdaus, 2017; Hu et al., 2018). Many studies combine adults into a single group and adjust for age and sex, rather than assessing age-specific and sex-specific associations (Astell-Burt & Feng, 2020; Bodicoat et al., 2014). Moreover, many studies rely on geographic buffers around residents, which does not align with the administrative units used in public health (Astell-Burt & Feng, 2020; Bodicoat et al., 2014). Systematic reviews have emphasized the need for age-specific and sex-specific subgroup analyses in future studies (Ccami-Bernal et al., 2023).

Using an object-based random forest approach, we generate eight band combinations that combine multispectral bands, vegetation and water indices, and the normalized Digital Surface Models (nDSM) to map type-specific green spaces and quantify the proportion of each green space type at the census-tract level. In this study, we address these gaps by integrating 3-metre spatial resolution PlanetScope multispectral imagery with LiDAR-derived nDSM to classify type-specific green spaces across census tracts in Mississauga, Ontario. We then examine whether associations between type-specific green spaces and type 2 diabetes prevalence differ across local and neighbourhood levels, by age group and sex, and compare these results with NDVI-only models to assess whether a single greenness index is sufficient to capture these associations with diabetes. This study provides new evidence on how type-specific green spaces may contribute to diabetes prevalence by offering a methodological framework for health research.

2. Methods

2.1. Study Area

Mississauga, Ontario, Canada (43.5° N, 79.6° W) is a large suburban municipality located on the north shore of Lake Ontario, one of the five Great Lakes of North America, bordering the provincial capital of Toronto to the east. Mississauga is the third-most populous municipality in Ontario, covering approximately 293 km^2 with a population of 717,961 and a highly diverse socioeconomic composition (Statistics Canada, 2016, 2021). The built environment is predominantly auto-oriented, with seven major highways, and a mix of residential, commercial, and industrial land uses. Mississauga has over 1000 parks and open spaces, including along the Credit River (City of Mississauga, 2022a, 2022b). Mississauga, Ontario, is a large, auto-oriented, heterogeneous suburban city within the Greater Toronto Area with an uneven distribution of green space and rising rates of type 2 diabetes, making it a relevant study area for examining how the type and structure of green space relate to the diabetes burden at the neighbourhood scale.

Peel Region, which includes Mississauga, has been described as an example of 'hyper-diversity', referring to the region's pronounced heterogeneity in sociodemographic, socioeconomic, and lifestyle characteristics across its population (Dean et al., 2018). Peel Region experiences one of the highest burdens of type 2 diabetes in Ontario. About one in 10 adults in Peel has diabetes, and this number is projected to increase to one in six by 2025 (City of Mississauga, 2024). Among adults aged 45–64, one in

six has diabetes, rising to one in three after age 65 (City of Mississauga, 2024). The age-standardized prevalence of diabetes among adults aged 20 + in 2020 was 15.5% in Peel, higher than the provincial (9.8%) and national (10%) age-standardized prevalence (Diabetes Canada, 2022). It is estimated that there will be 100,000 new cases of diabetes in Peel between 2020 and 2030 (Smith, 2025). These patterns in both environmental exposures and social determinants of health make Mississauga, Ontario, an ideal study area to examine how green spaces relate to chronic disease risk. There is a research need to map green space types by integrating spectral and structural data, and to examine their associations with diabetes prevalence at local and neighbourhood levels, by sex and age group, while accounting for sociodemographic and economic factors. In this study, we analyzed 136 census tracts in Mississauga to examine the relationship between green space and diabetes prevalence. Figure 1 displays the 136 census tracts and land use classification in Mississauga, Ontario.

2.2. Multispectral data

A multispectral satellite image was generated by mosaicking seven PlanetScope surface-reflectance scenes that cover the entire city of Mississauga (Planet Labs PBC, 2021). These scenes were acquired on August 4, 2021, from 15:19 UTC to 15:22 UTC, at an average off-nadir angle of 3.7°. This acquisition date was selected to capture peak summer vegetation and minimal cloud cover (0–4%). The PlanetScope satellite imagery has a spatial resolution of 3-metres and includes eight bands: coastal blue (431 to 452 nm), blue (465 to 515 nm), green (513 to 549 nm), green II (547 to 583 nm), yellow (600 to 620 nm), red (650 to 680 nm), red edge (697 to 713 nm), and near infrared (NIR) (845 to 885 nm), provided in WGS 1983 UTM Zone 17N. These bands include information relevant to vegetation differentiation. The coastal blue and blue bands help separate water from vegetation, the green and green II bands capture chlorophyll reflectance from short vegetation, the yellow band helps with pigment and stress, the red band captures chlorophyll absorption, the red edge band can distinguish canopy structure, and the NIR band aids in leaf density and canopy cover (Demir et al., 2008; Fang et al., 2018; Gitelson et al., 1996; Hartling et al., 2019; Omer et al., 2016; Waser et al., 2014; Wężyk et al., 2019). Planet radiometrically calibrated, orthorectified, atmospherically corrected, and scaled the scenes to ensure that reflectance values were comparable across bands. The seven-band images were mosaicked in ArcGIS Pro v.3.6.0 to produce a single multispectral image, providing spectral information for classifying green space types.

2.3. LiDAR data

The airborne LiDAR data were collected by the City of Mississauga in 2021 (City of Mississauga, 2021). The LiDAR point cloud had a density of 23.5 points per square metre and was provided in WGS 1984 UTM Zone 17N. A Digital Terrain Model (DTM) was derived from ground-classified points, and a Digital Surface Model (DSM) was derived from first-return points, capturing upper surfaces, using ArcGIS Pro v.3.6.0. Both rasters were generated at a 3-metre spatial resolution to match the PlanetScope imagery. Spatial alignment between datasets was verified by visually inspecting features, and no mismatch was observed. A normalized Digital Surface Model (nDSM) was then created by subtracting the DTM from the DSM to obtain above-ground features, or object heights, representing the tops of surfaces such as tree canopies and building roofs. The nDSM provided structural information to complement the spectral information

from PlanetScope bands, improving the classification of green space types, particularly distinguishing between trees and grass, thereby increasing classification accuracy.

2.4. Indices

To enhance classification accuracy, eight vegetation and water indices were calculated from the PlanetScope multispectral bands using ArcGIS Pro v.3.6.0. Previous studies have shown that incorporating vegetation and water indices improves LULC classification accuracy, particularly for high spatial resolution imagery (Jombo et al., 2022; Yan et al., 2015). These indices highlight differences in chlorophyll absorption, pigment content, canopy density, and water presence that cannot always be captured by spectral imagery alone, thereby improving the classification accuracy of green space types.

The classification indices were selected based on previous research (Jombo et al., 2022). These included the normalized difference vegetation index (NDVI), which distinguishes vegetation from non-vegetation based on red absorption and NIR reflectance (Demir et al., 2008; Fang et al., 2018; Omer et al., 2016). The normalized difference vegetation index – green/red ratio (NDVI-GR) uses the green band to improve vegetation separation (Gitelson et al., 1996; Hartling et al., 2019). The normalized difference vegetation index – yellow (NDVI-Y) uses the yellow band to differentiate vegetation using leaf pigment and stress (Hartling et al., 2019; Waser et al., 2014). The normalized difference red edge index (NDRE) uses the red edge band to distinguish between canopy types (Wężyk et al., 2019). The green normalized difference vegetation index (GNDVI) uses the green and NIR reflectance to identify chlorophyll content (Omer et al., 2016; Wężyk et al., 2019). The enhanced vegetation index (EVI) improves vegetation detection in dense canopy areas by correcting for soil and atmospheric influences (Huete et al., 2002). The soil-adjusted vegetation index (SAVI) minimizes soil influence to better distinguish sparse vegetation (Huete, 1988). The normalized difference water index (NDWI) identifies water bodies, reducing confusion between water and vegetation (McFeeters, 1996). Together, these indices provide additional information that will strengthen classification accuracy and improve differentiation of green space types.

2.5. Data integration

The eight PlanetScope multispectral bands, the eight derived indices, and the LiDAR-derived nDSM were integrated into a single multiband raster dataset at 3-metre spatial resolution using ArcGIS Pro v.3.6.0. All datasets were in WGS 1984 UTM Zone 17N prior to integration. The integrated dataset provides spectral and structural information, and serves as predictor variables for LULC classification. Table 1 displays the multiband raster bands.

Table 1
Bands used in classifications.

Band name	Wavelength / Equation
Coastal blue	431 to 452 nm
Blue	465 to 515 nm
Green	513 to 549 nm
Green II	547 to 583 nm
Yellow	600 to 620 nm
Red	650 to 680 nm
Red edge	697 to 713 nm
Near infrared (NIR)	845 to 885 nm
Normalized Difference Vegetation Index (NDVI)	$\frac{NIR-Red}{NIR+Red}$
Normalized Difference Vegetation Index – Green/Red Ratio (NDVI-GR)	$\frac{NIR-(Green+Red)}{NIR+(Green+Red)}$
Normalized Difference Vegetation Index – Yellow (NDVI-Y)	$\frac{NIR-Yellow}{NIR+Yellow}$
Normalized Difference Red Edge Index (NDRE)	$\frac{NIR-RedEdge}{NIR+RedEdge}$
Green Normalized Difference Vegetation Index (GNDVI)	$\frac{NIR-Green}{NIR+Green}$
Enhanced Vegetation Index (EVI)	$2.5 \times \frac{NIR-Red}{NIR+6\times Red-7.5\times Blue+1}$
Soil Adjusted Vegetation Index (SAVI)	$\frac{NIR-Red}{NIR+Red+0.5}$
Normalized Difference Water Index (NDWI)	$\frac{Green-NIR}{Green+NIR}$
nDSM (normalized Digital Surface Model)	$DigitalSurfaceModel\,(DSM) - DigitalTerrainModel\,(DTM)$

2.6. Classification

Supervised classification was performed in ArcGIS Pro v.3.6.0 using an integrated multiband raster dataset combining PlanetScope multispectral bands, the LiDAR-derived nDSM, and derived indices.

Reference samples were created by visually interpreting features using Google Maps, Google Earth, and Google Street View to ensure that training polygons accurately represent on-the-ground features. These polygons served as training samples for each LULC class.

An object-based approach was selected because it groups neighbouring pixels with similar characteristics into larger and more meaningful segments. At finer spatial resolutions, such as PlanetScope's 3-metre spatial resolution, object-based image analysis classifies objects into polygons more accurately by using extracted attributes, such as shape, contrast, texture, and areal information, thereby improving overall accuracy compared to traditional pixel-based approaches (Jombo et al., 2022). Commonly used pixel-based approaches and classification methods may produce low accuracies, even with high spatial resolution data, since they introduce the "nugget" effect of salt-and-pepper noise (Jombo et al., 2022; Yan et al., 2015).

We classified nine LULC classes representing Mississauga: water, paved surface/road, building, bare soil, deciduous tree, coniferous tree, shrub, grass, and marsh/wetland. A random forest classifier with 100 decision trees and a maximum depth of 30 was used for supervised classification, as it is well-suited for segmented and multispectral images, can handle many and correlated predictors, and is robust to overfitting (Belgiu & Drăguţ, 2016). The integrated imagery was segmented at the highest spectral and spatial detail and the minimum segment size. The segments were characterized using active chromaticity colour and mean digital numbers.

To evaluate the different use of predictors, we ran eight random forest classifications beginning with standard bands and progressively adding complex additional bands: (i) RGB (red, green, blue)-NIR; (ii) RGB-NIR + nDSM; (iii) RGB-NIR + extended bands (coastal blue, green II, yellow, red-edge); (iv) RGB-NIR + extended bands + nDSM; (v) RGB-NIR + all indices; (vi) RGB-NIR + all indices + nDSM; (vii) RGB-NIR + extended bands + all indices; and (viii) RGB-NIR + extended bands + all indices + nDSM. This stepwise approach allowed us to isolate the contribution of spectral and structural information.

For each of the eight random forest classifications, we assessed accuracy using 500 stratified random points distributed across the LULC classes. Confusion matrices were generated to compute overall accuracy, producer's accuracy, user's accuracy, and kappa statistic. The classification with the highest overall accuracy was then used in further analyses to examine the associations between green space types and diabetes prevalence. Variable importance was extracted to identify the most influential predictors, and cross-validation rates were recorded.

2.7. Diabetes prevalence

Diabetes prevalence data were obtained from the Institute for Clinical Evaluative Sciences (ICES), which includes 2020 prevalence rates aggregated by Mississauga census tracts, per 100 people (Institute for Clinical Evaluative Sciences, 2020). Diabetes cases were based on physician diagnoses and hospital records. Although the data include the three primary forms of diabetes, type 1, type 2, and gestational diabetes, approximately 90% are type 2 diabetes cases (Ong et al., 2023). Type 2 diabetes is a chronic, progressive condition characterized by impaired insulin production or action, which regulates blood sugar

(serum glucose) (Diabetes Canada, 2025). The data includes diabetes prevalence per census tract for males and females, and by age group (20+, 20–44, 45–64, and 65+).

2.8. Covariates

Based on a recent systematic review of green space exposure and type 2 diabetes incidence, we adjusted for four covariates: educational level (percent with a bachelor's degree or higher), occupation (percent unemployed), marital status (percent married), and income (median after-tax income in 2019) (Ccami-Bernal et al., 2023). These covariates were the most common across the 13 studies reviewed, each appearing in at least five studies (Ccami-Bernal et al., 2023). Previous evidence also highlights age and sex as important confounders; however, in our study, we accounted for these by fitting separate age-specific and sex-specific models (Ccami-Bernal et al., 2023). Unlike many cohort studies that adjust for individual-level factors, such as body mass index (BMI) or alcohol use, our covariates were measured at the census-tract level, reflecting neighbourhood contexts. The datasets for the covariates used in this study were extracted from the Peel Region Data Portal, based on the 2021 Canadian Census at the census tract level (Peel Region, 2021).

2.9. Statistical analyses

Beta regression models were used to examine the relationship between diabetes prevalence and green space types, controlling for covariates across 136 census tracts in Mississauga. Beta regression is suitable for continuous proportion-based data, such as diabetes prevalence expressed as the number of cases per 100 people. The diabetes prevalence data were divided by 100 to bound the data from 0 to 1 (Ferrari & Cribari-Neto, 2004; Smithson & Verkuilen, 2006). Then, a beta regression transformation was conducted using the following equation:

$$y'' = \frac{y' \times (N-1) + 0.5}{N}$$

1

,

where

y" = transformation;

y' = dependent variable; and

N = sample size.

The transformed diabetes prevalence was used as the dependent variable in all models. The independent variables were the proportions of each green space type (deciduous trees, coniferous trees, shrubs, grass, marsh, and all green spaces combined) and the mean NDVI, each in its own model.

For each green space variable (proportion deciduous trees, coniferous trees, shrubs, grass, marsh, all green spaces combined, and mean NDVI), we fit unadjusted and adjusted beta regression models by sex (male and female) and age group (20+, 20–44, 45–64, and 65). Adjusted models included covariates of percent with a bachelor's degree or higher, percent unemployed, percent married, and median after-tax income, to control for confounding factors. In addition, corresponding neighbourhood models using queen's contiguity were fit for the unadjusted and adjusted models. Multicollinearity was assessed for the adjusted models using the Variance Inflation Factor (VIF), with a threshold of VIF > 5. Statistical modelling was conducted in R v.4.2.2 (Posit team, 2025). The models were evaluated based on the positive or negative direction of the coefficient estimates for the green space predictors and their p-values. The coefficient estimates, errors, number of observations, pseudo R^2, Akaike information criterion (AIC), Bayesian information criterion (BIC), and log-likelihood were also recorded.

3. Results

3.1. Segmentation, classification accuracy, and variable importance

The overall accuracy of the eight classifications ranged from 75% to 88%. The baseline model using red, green, blue, and NIR bands achieved the lowest overall accuracy of 75%, which increased to 84% when the nDSM was added. Classifications without the nDSM achieved overall accuracies of 75%-76%, whereas models with the nDSM achieved 83%-88%. The shrub class benefited most from the nDSM, with accuracy increasing by 10%-40%. The best-performing classification, which included all bands, indices, and the nDSM, achieved an overall accuracy of 88%, and we selected it for the remaining analysis.

The best-performing classification, paved surfaces/road, buildings, bare soil, coniferous trees, grass, and marsh/wetlands achieved the highest accuracies (96%, 91%, 90%, 90%, 100%, and 90%, respectively). In contrast, water, deciduous trees, and shrubs had the lowest accuracies (77%, 70%, and 50%, respectively). Table 2 displays the overall accuracy and the accuracy of the LULC classes for all eight classifications. The highlighted values display the highest accuracy in each class.

Based on the best-performing classification, the nDSM was the most important variable for classifying LULC classes, as shown in Fig. 2, highlighting the strong contribution of vertical structure in differentiating green space types. Following the nDSM, the red-edge band, GNDVI, and NDWI were also important. In contrast, the least important bands were green, green II, and NDRE. The cross-validation rate was 99.%. The variable importance and cross-validation rate for all eight band combinations are displayed in Table S1.

Table 2
Overall accuracy of each classification combination for each land use land cover (LULC) in percentage (%).

Overall Accuracy	**RGB-NIR**	**RGB-NIR + nDSM**	**RGB-NIR + extended bands**	**RGB-NIR + extended bands + nDSM**	**RGB-NIR + indices**	**RGB-NIR + indices + nDSM**	**RGB-NIR + extended bands + indices**	**RGB-NIR + extended bands + indices + nDSM**
	75	**84**	**76**	**83**	**75**	**84**	**75**	**88**
1) Water	74	76	73	73	76	75	75	77
2) Paved surface / Road	72	91	78	92	75	93	74	96
3) Building	78	89	78	84	73	83	75	91
4) Bare soil	100	80	90	100	100	90	100	90
5) Deciduous tree	40	70	90	90	70	100	70	70
6) Coniferous tree	80	70	60	80	70	80	70	90
7) Shrub	0	20	0	10	0	30	10	50
8) Grass	100	100	91	100	97	100	88	100
9) Marsh / Wetland	80	100	100	80	80	100	100	90
Kappa Statistic	68	80	70	79	68	80	69	84

3.2. Classification map

The most complex combination was the best-performing classification, which will proceed in the analysis, with an overall accuracy of approximately 88%, which included the eight multispectral bands (coastal blue, blue, green, green II, yellow, red, red edge, NIR), the eight indices (NDVI, NDVI-GR, NDVI-Y, NDRE, GNDVI, EVI, SAVI, NDWI), and the nDSM. A classification map of LULC classes was created, comprising nine classes (water, paved surface/road, building, bare soil, deciduous trees, coniferous trees, shrub, grass, and marsh/wetland). Of the nine classes, five are considered green space types for this study (deciduous trees, coniferous trees, shrubs, grass, and marsh/wetland). The most common LULC class is paved surface/road, covering approximately 35% of the area. Grass is the most common green space type, covering approximately 17% of the area, while shrubs are the least common, covering 0.12% of the area. Table 3 displays the area of coverage for each LULC class, expressed in square kilometres and as a

percentage, along with descriptive statistics on the area coverage of each LULC class within census tracts. The classes are distributed across the entire city, but their spatial distribution varies. Green space is more prominent in the southern half of the city, whereas buildings and paved surfaces dominate the northern half. Grass cover often follows road networks and paved surfaces. Water bodies are concentrated in the southern half of the city and along the Credit River corridor, which flows from the northwest toward Lake Ontario. Figure 3 displays the eight LULC classes in Mississauga census tracts.

Table 3
Area of coverage and descriptive statistics of the area coverage of each land use land cover (LULC) class in census tracts are shown as proportions.

Water	**Km^2**	**Percentage (%)**	**Minimum (km^2)**	**Mean (km^2)**	**Median (km^2)**	**Maximum (km^2)**	**Standard Deviation (km^2)**
	1.38	**0.47**	**0.00**	**0.005**	**0.001**	**0.52**	**0.009**
Paved surface / Road	102.67	34.92	0.05	0.29	0.29	0.58	0.11
Building	51.33	17.46	0.04	0.19	0.18	0.40	0.07
Bare soil	2.05	0.70	0.00	0.003	0.0009	0.04	0.006
Deciduous tree	44.95	15.29	0.04	0.19	0.18	0.42	0.07
Coniferous tree	38.12	12.97	0.03	0.16	0.16	0.40	0.07
Shrub	0.35	0.12	0.00	0.001	0.001	0.003	0.0009
Grass	49.35	16.79	0.04	0.15	0.13	0.47	0.08
Marsh / Wetland	3.81	1.30	0.00	0.01	0.007	0.08	0.02
Total area coverage	294.01						

3.3. Diabetes prevalence and covariates

Diabetes prevalence varies across Mississauga's census tracts. On average, 3.23 (SD = 0.69) of 100 adults aged 20–44 have diabetes, while that average increases to 17.22 (SD = 3.82) per 100 adults aged 45–65, and continues to increase to 35.35 (SD = 5.41) per 100 adults aged 65+. By sex, the mean prevalence is 14.96 (SD = 2.12) per 100 for males aged 20 + and 13.23 (SD = 2.22) per 100 for females aged 20+. Table 4 presents the descriptive statistics of diabetes prevalence by census tract, by sex and age group. Figure 4 illustrates the spatial variation in diabetes prevalence for males and females across age groups. Sociodemographic and economic factors also vary in Mississauga census tracts. On average, 37% of residents hold a bachelor's degree or higher, 9% of residents are unemployed, 53% of residents are married, and the median after-tax income is approximately $35,000. Table 5 presents

descriptive statistics for the covariates included in the adjusted models, which account for sociodemographic and economic factors.

Table 4
Descriptive statistics of diabetes prevalence per 100 people per census tract by sex and age group.

Male	Minimum	Mean	Median	Maximum	Standard Deviation
Age 20+	7.80	14.96	14.80	20.70	2.12
Age 20–44	0.80	3.04	3.10	4.70	0.73
Age 45–64	6.40	19.04	18.95	31.60	4.09
Age 65+	20.90	38.77	38.80	52.90	5.69
Female					
Age 20+	6.50	13.23	13.30	18.60	2.22
Age 20–44	1.30	3.40	3.40	6.10	0.91
Age 45–64	3.80	15.46	15.35	27.40	3.89
Age 65+	15.70	32.49	32.60	44.80	5.73

Table 5
Descriptive statistics of covariates included in the adjusted models to account for sociodemographic and economic factors.

Proportion bachelor's degree or higher	Minimum	Mean	Median	Maximum	Standard Deviation	NA
	0.20	**0.37**	**0.37**	**0.56**	**0.08**	**11**
Proportion unemployed	0.05	0.09	0.09	0.13	0.01	11
Proportion married	0.33	0.52	0.53	0.65	0.06	11
Median after-tax income ($)	23,600	34,965	34,000	62,400	5,902	11

3.4. Statistical analyses

We used beta regression models to estimate diabetes prevalence and green space types at the census-tract level, considering local and neighbourhood exposures, by sex and age group. Negative exposure coefficients indicate lower diabetes prevalence with greater exposure to green space exposure, whereas positive coefficients indicate higher diabetes prevalence with greater exposure. In the local unadjusted beta regression models, a higher proportion of green space was associated with lower diabetes prevalence for deciduous trees, coniferous trees, shrubs, marsh, all green spaces combined, and NDVI, across both sexes and all age groups. Grass was not associated with diabetes prevalence.

The neighbourhood unadjusted beta regression models showed broadly similar results, where higher proportions of green space types were associated with lower diabetes prevalence for deciduous trees, coniferous trees, shrubs, marsh, all green spaces combined, and NDVI, across both sexes, in adults aged 45–64 and 65+. Whereas the age group 20–44 showed mixed sex associations, deciduous trees and marshes were associated with both sexes. However, coniferous trees, shrubs, all green spaces combined, and NDVI were associated only with females. Similar to the local unadjusted models, grass showed no association.

After adjusting for covariates, including educational level, occupation, marital status, and income, many associations attenuated. In the local adjusted beta regression models, deciduous trees, coniferous trees, and all green spaces combined retained negative associations for ages 45–64 and 65+, for both sexes. Shrubs also retained negative associations among adults aged 45–64, for both sexes. Meanwhile, NDVI only retained negative associations with males 65+. Grass and marsh showed no association.

The neighbourhood-adjusted models showed broadly similar results; however, sex differences were more apparent, and some positive associations also appeared. Deciduous tree retained negative associations for ages 45–64 and 65 + for both sexes, while coniferous tree only retained negative associations for ages 45–64 for both sexes and females 65+. Shrubs retained negative associations for both sexes, aged 45–64 and females aged 20–44 and 65+. All green spaces combined were associated with negative associations for females aged 45–64 and 65+. Meanwhile, NDVI retained associations for both sexes aged 45–64 and for females aged 65+. However, positive associations also emerged for the adjusted neighbourhood models. A higher proportion of green space types was associated with higher diabetes prevalence for coniferous trees, marsh, and all green spaces combined among males aged 20–44, and for grass among females aged 45–64.

These results indicate that deciduous trees, coniferous trees, and shrubs are negatively associated with diabetes prevalence, particularly among adults aged 45+. In contrast, grass and marsh showed inconsistent results.

Compared with the type-specific models, the NDVI-only models show negative associations with diabetes prevalence in the unadjusted models, consistent with those observed in the type-specific models. However, after adjusting for covariates, the NDVI associations attenuated and were observed only in some sexes and age groups. Only males aged 65 + were associated with NDVI in the local adjusted models, and only males and females aged 45–64, and females 65 + were associated with NDVI in the neighbourhood adjusted models. In contrast, the type-specific models continued to show strong negative associations, particularly for deciduous trees, coniferous trees, and shrubs. Figure 5 presents all model results, showing the direction of association (negative/positive) and significance by sex and age group. Supplementary Data S1 presents the coefficient estimates, errors, number of observations, pseudo R^2, Akaike information criterion (AIC), Bayesian information criterion (BIC), and log-likelihood.

4. Discussion

This study addresses whether type-specific green space and structure provide additional information beyond the greenness index, NDVI, for understanding diabetes prevalence. We investigated the relationship between diabetes prevalence and the proportions of green space types across 136 Mississauga census tracts, by sex (male/female) and age group (20+, 20–44, 45–64, and 65+), using 2020 (diabetes prevalence) and 2021 (classification) data, to test whether type-specific and structure-specific green space is more informative than NDVI. Overall, we found that deciduous trees, coniferous trees, and shrubs are more consistently associated with lower diabetes prevalence than NDVI alone, and that these associations are most evident among adults aged 45 + across both sexes. These findings support the literature that not all green space is equal and that green space structure may be more relevant than greenness alone. By integrating multispectral imagery with LiDAR-derived nDSM, we were able to distinguish among green space types and better understand their associations with diabetes prevalence. These findings suggest that green space structures provide additional information beyond total greenness for understanding diabetes prevalence. By creating individual models for each green space type, we found that tree-based and shrub-based green spaces yielded more consistent associations than NDVI-only models, even after controlling for sociodemographic and economic factors.

4.1. Integrated classification

Our integrated object-based random forest classification produced an overall accuracy of 88% across nine LULC classes, including five type-specific green space classes. This high accuracy aligns with previous research showing that object-based segmentation of integrated spectral and structural data has been effective in achieving high classification accuracies in heterogeneous urban environments (Hamedianfar et al., 2014; Hartling et al., 2019; Hellesen & Matikainen, 2013; Jombo et al., 2022; MacFaden et al., 2012; Zhou, 2013).

We used LiDAR data to better differentiate between green space types, such as grass and tree canopy, by integrating PlanetScope multispectral imagery with LiDAR-derived nDSM. This green space differentiation enables more precise exposure measurements for epidemiological analysis. In our study, we found that the nDSM was the most important variable for classification, and among the eight classification band combinations, those that included the nDSM achieved an increase in overall accuracy of 7–12%. This suggests that LiDAR-derived nDSM helps distinguish structural green space types that appear spectrally similar, thereby strengthening the measurement of green space exposure. Studies that rely solely on vegetation indices, such as NDVI, do not capture vertical height, thereby reducing classification accuracy. From the classification model that included RGB-NIR, extended bands, and nDSM to one that incorporated the indices, the overall accuracy increased by 4%. Although there was little increase in overall accuracy across the other classification models when incorporating the indices, future research may investigate which indices increase accuracy for specific LULC classes. These findings support that structural height is important for differentiating green space types that NDVI cannot capture.

4.2. Modelling against diabetes prevalence

We estimated diabetes prevalence and green space types at the census-tract level, accounting for local and neighbourhood exposures, by sex and age group. We found that green spaces are associated with lower diabetes prevalence in Mississauga, aligning with existing literature showing that people living in greener areas have a lower risk of type 2 diabetes (Ccami-Bernal et al., 2023; De La Fuente et al., 2020). Our study distinguished green space types to provide a clearer picture, rather than focusing on total green space and greenness indices. We found that diabetes prevalence is associated with deciduous trees, coniferous trees, and shrubs, but not with grass, marshes, and NDVI alone, particularly among adults aged 45+, a finding that many studies have not examined, as they combine green space types when associating green space with health outcomes, such as diabetes prevalence. These findings support the literature indicating that definitions of green space exposure vary and that subgroup analyses can identify informative patterns.

Our findings align with the idea that taller, denser structural vegetation is relevant to type 2 diabetes, as it provides shade and cooling, improves air quality, and reduces stress. Tree canopy is also associated with lower urban heat and air pollution, both of which have been linked to an increased risk of type 2 diabetes (Gao et al., 2022; Yang et al., 2020; Yin et al., 2024). In contrast, low vegetation, such as grass, provides little shade, cooling, and air filtration (Yin et al., 2024). Because our classification used LiDAR-derived height and multispectral data, we were able to distinguish among green space types and better understand how these types may affect diabetes prevalence. NDVI attenuated its associations after adjusting for covariates. Studies relying on NDVI alone may miss type-specific associations. These differences in how exposure is defined may explain why findings in the green space literature are sometimes mixed. Our findings suggest that distinguishing among green space types can yield more consistent associations with diabetes prevalence. Overall, different definitions of green space exposure can influence conclusions.

We also built neighbourhood models using queen's contiguity to better represent spatial context, because people are exposed not only to their census tract but also to adjacent tracts, which often share environmental, health, and social characteristics. Our neighbourhood models suggest that these associations are present not only within census tracts, but also within broader neighbourhoods. We also found positive associations between type-specific green spaces and diabetes prevalence in the adjusted neighbourhood models for men aged 20–44 in areas with higher coniferous trees and marsh cover, and for women aged 45–64 in areas with higher grass cover. These results suggest that neighbourhood green space is associated with other neighbourhood features, such as lower walkability, higher traffic volumes, and less healthy food environments. These neighbourhoods may be more car-dependent, have poorer street connectivity and fewer walkable destinations, and include less expensive multi-unit housing, which may attract individuals already at higher risk of diabetes, such as low-income individuals, even after adjusting for covariates.

We also built adjusted models, controlling for covariates such as education, unemployment, marital status, and income. Although many associations with diabetes prevalence were attenuated relative to the unadjusted models, associations with tree-based and shrub-based green space types remained

statistically significant among adults aged 45+, whereas those with grass and marsh green space types did not. In contrast, the NDVI models retained only limited associations, suggesting that NDVI loses its effectiveness as a proxy for green space after adjusting for sociodemographic and economic factors, whereas type-specific green spaces exhibit stronger associations with diabetes prevalence.

In our study, we fit sex-specific and age-specific beta regression models, whereas most past studies combine all adults and often adjust for sex and age. This approach allowed us to examine whether associations differed across demographic groups, rather than treating them as confounders or combining groups into a single model. Older adults are physically and psychologically vulnerable, spending more time in their residential neighbourhood due to physical decline, cognitive impairment, and decreasing social ties (Firdaus, 2017; Hu et al., 2018). They are also more vulnerable to environmental stressors, like heat and air pollution (Montoro-Ramírez et al., 2025). Our results show that tree and shrub green spaces are associated with lower diabetes prevalence among adults aged 45+, even after adjusting for covariates, suggesting that canopy-rich areas that support walkability, cooling, air pollution filtration, and stress reduction may be relevant for older adults.

Our results show that trees and shrubs are the most consistently associated green space types with lower diabetes prevalence, particularly among adults aged 45+, across both sexes, even after adjusting for covariates, in a local and neighbourhood context in Mississauga. These findings suggest that relying solely on NDVI may obscure associations between green space types and diabetes prevalence, providing guidance for future research. These findings highlight the need to move beyond greenness indices, such as NDVI, to type-specific green space exposures when studying chronic diseases. This study supports the use of type-specific and structure-specific green space exposure measurements when examining chronic disease outcomes because they do not sum across different green space types into a single greenness value. These findings suggest that tree-dominated and shrub-dominated green spaces may be particularly relevant when investigating diabetes prevalence.

4.3. Strengths and limitations

To our knowledge, this is the first study to integrate high spatial resolution multispectral imagery and LiDAR-derived heights to classify type-specific green spaces associated with diabetes prevalence. Our study integrated 3-metre spatial resolution PlanetScope imagery and LiDAR-derived nDSM to distinguish green space types, such as deciduous trees, coniferous trees, shrubs, grass, and marsh, and to assess whether a single greenness proxy, such as NDVI, is sufficient to capture associations with diabetes prevalence in Mississauga, Ontario. Since NDVI cannot distinguish between vegetation types, the methodological advancement of using structural information allows for the differentiation of green space types which may be more relevant to health. By combining spectral and structural information, we improve the measurement of green space exposure by identifying green space types and associating them with diabetes prevalence. We also fit beta regression models at local and neighbourhood scales by sex and age group, both unadjusted and adjusted for covariates. Our study supports improving the measurement and interpretation of green space exposure across subgroups and spatial contexts.

This is a cross-sectional study, which means we cannot establish causal relationships between green space types and diabetes prevalence, and selective migration is possible. We also lacked access to individual-level data; therefore, exposures were based on census tract-level data, which introduces the uncertain geographic context problem (UGCoP) and the modifiable areal unit problem (MAUP). Our study assumes that residents are primarily exposed to green space within and around their census tract, while daily activity spaces are not being considered. These boundaries may differ by age and sex. Finally, although our final classification model achieved high overall accuracy, some classes, such as shrubs, had lower accuracies, which may lead to misclassification and attenuation of associations for those green space types. In addition, our classification was based on a single-date image acquired during peak summer conditions, which captures full vegetation cover but does not account for seasonal variation, potentially affecting the transferability of the classification across years. Nonetheless, integrating high spatial resolution satellite imagery and LiDAR enables type-specific green space mapping and provides a methodological framework for future epidemiological studies. These findings may also inform urban planning and public health policy by suggesting that the structure of green space may be more relevant in urban environments than greenness alone. Future studies can implement longitudinal designs and daily activity spaces to strengthen inference.

5. Conclusion

Overall, this study used random forest supervised classification with multispectral satellite imagery and LiDAR to classify green space types in Mississauga census tracts and to associate them with diabetes prevalence. We found that combining satellite bands, calculating vegetation and water indices, and calculating LiDAR-derived nDSM performed the highest overall accuracy (approximately 88%). Our results show that deciduous trees, coniferous trees, and shrubs were most consistently associated with lower diabetes prevalence, particularly among adults aged 45+, even after adjusting for covariates ($p < 0.05$). In contrast, grass and marsh showed less consistent associations, and NDVI-only models showed weaker associations once covariates were included. Our findings suggest that urban diabetes prevention strategies may benefit from structurally complex vegetation, such as trees and shrubs, rather than grass cover. This highlights the importance of not relying solely on overall greenness, such as the NDVI, but rather investigating type-specific green spaces in future epidemiological studies.

Declarations

Acknowledgements

I acknowledge the support and funding provided by the University of Toronto’s Novo Nordisk Network for Healthy Populations, which was partially funded by a donation from Novo Nordisk to the University of Toronto. Novo Nordisk was not involved in the design, conduct, analysis, or reporting of this research project.

The diabetes prevalence data were provided for this project by the Ontario Community Health Profiles Partnership (OCHPP) via datasets at ICES.

This research was also supported by the Ontario Graduate Scholarship.

References

1. Astell-Burt, T., & Feng, X. (2020). Urban green space, tree canopy and prevention of cardiometabolic diseases: A multilevel longitudinal study of 46 786 Australians. *International Journal of Epidemiology*, *49*(3), 926–933. https://doi.org/10.1093/ije/dyz239
2. Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011
3. Bilal, U., Auchincloss, A. H., & Diez-Roux, A. V. (2018). Neighborhood Environments and Diabetes Risk and Control. *Current Diabetes Reports*, *18*(9), 62. https://doi.org/10.1007/s11892-018-1032-2
4. Bodicoat, D. H., O'Donovan, G., Dalton, A. M., Gray, L. J., Yates, T., Edwardson, C., Hill, S., Webb, D. R., Khunti, K., Davies, M. J., & Jones, A. P. (2014). The association between neighbourhood greenspace and type 2 diabetes in a large cross-sectional study. *BMJ Open*, *4*(12), e006076. https://doi.org/10.1136/bmjopen-2014-006076
5. Ccami-Bernal, F., Soriano-Moreno, D. R., Fernandez-Guzman, D., Tuco, K. G., Castro-Díaz, S. D., Esparza-Varas, A. L., Medina-Ramirez, S. A., Caira-Chuquineyra, B., Cortez-Soto, A. G., Yovera-Aldana, M., & Rojas-Rueda, D. (2023). Green space exposure and type 2 diabetes mellitus incidence: A systematic review. *Health & Place*, *82*, 103045. https://doi.org/10.1016/j.healthplace.2023.103045
6. Charaniya, A. P., Manduchi, R., & Lodha, S. K. (2004). Supervised Parametric Classification of Aerial LiDAR Data. *2004 Conference on Computer Vision and Pattern Recognition Workshop*, 30–30. https://doi.org/10.1109/CVPR.2004.446
7. City of Mississauga. (2021). *City of Mississauga LiDAR Dataset* [Dataset].
8. City of Mississauga. (2022a). *2021 Existing Land Use* [Dataset]. https://data.mississauga.ca/datasets/mississauga::2021-existing-land-use/explore?location=43.609579%2C-79.674600%2C11.90
9. City of Mississauga. (2022b). *Parks Plan*. https://www.mississauga.ca/wp-content/uploads/2022/03/10085651/Parks-Plan.pdf
10. City of Mississauga. (2024). *Diabetes: A Growing Concern*. https://www.mississauga.ca/projects-and-strategies/city-projects/cities-for-better-health/
11. Crouse, D. L., Pinault, L., Balram, A., Hystad, P., Peters, P. A., Chen, H., Van Donkelaar, A., Martin, R. V., Ménard, R., Robichaud, A., & Villeneuve, P. J. (2017). Urban greenness and mortality in Canada's largest cities: A national cohort study. *The Lancet Planetary Health*, *1*(7), e289–e297. https://doi.org/10.1016/S2542-5196(17)30118-3

12. De La Fuente, F., Saldías, M. A., Cubillos, C., Mery, G., Carvajal, D., Bowen, M., & Bertoglia, M. P. (2020). Green Space Exposure Association with Type 2 Diabetes Mellitus, Physical Activity, and Obesity: A Systematic Review. *International Journal of Environmental Research and Public Health*, *18*(1), 97. https://doi.org/10.3390/ijerph18010097
13. Dean, J., Regier, K., Patel, A., Wilson, K., & Ghassemi, E. (2018). Beyond the Cosmopolis: Sustaining Hyper-Diversity in the Suburbs of Peel Region, Ontario. *Urban Planning*, *3*(4), 38–49. https://doi.org/10.17645/up.v3i4.1700
14. Demir, N., Poli, D., & Baltsavias, E. (2008). *Extraction of Buildings and Trees Using Images and LIDAR Data* [Application/pdf]. https://doi.org/10.3929/ETHZ-B-000011960
15. Diabetes Canada. (2022). *Diabetes in Canada—2022 Backgrounder*. https://www.diabetes.ca/DiabetesCanadaWebsite/media/Advocacy-and-Policy/Backgrounder/2022_Backgrounder_Canada_English_1.pdf
16. Diabetes Canada. (2025). *Type 2 diabetes*. https://www.diabetes.ca/about-diabetes/type-2
17. Fang, F., McNeil, B. E., Warner, T. A., & Maxwell, A. E. (2018). *Combining high spatial resolution multi-temporal satellite data with leaf-on LiDAR to enhance tree species discrimination at the crown level*. 208166 Bytes. https://doi.org/10.6084/M9.FIGSHARE.6917456
18. Fazli, G. S., Moineddin, R., Bierman, A. S., & Booth, G. L. (2019). Ethnic differences in prediabetes incidence among immigrants to Canada: A population-based cohort study. *BMC Medicine*, *17*(1), 100. https://doi.org/10.1186/s12916-019-1337-2
19. Ferrari, S., & Cribari-Neto, F. (2004). Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics*, *31*(7), 799–815. https://doi.org/10.1080/0266476042000214501
20. Firdaus, G. (2017). Built Environment and Health Outcomes: Identification of Contextual Risk Factors for Mental Well-being of Older Adults. *Ageing International*, *42*(1), 62–77. https://doi.org/10.1007/s12126-016-9276-0
21. Fong, K. C., Hart, J. E., & James, P. (2018). A Review of Epidemiologic Studies on Greenness and Health: Updated Literature Through 2017. *Current Environmental Health Reports*, *5*(1), 77–87. https://doi.org/10.1007/s40572-018-0179-y
22. Gao, D., Friedman, S., Hosler, A., Sheridan, S., Zhang, W., & Lin, S. (2022). Association between extreme ambient heat exposure and diabetes-related hospital admissions and emergency department visits: A systematic review. *Hygiene and Environmental Health Advances*, *4*, 100031. https://doi.org/10.1016/j.heha.2022.100031
23. Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, *58*(3), 289–298. https://doi.org/10.1016/S0034-4257(96)00072-7
24. Guarnieri, M., & Balmes, J. R. (2014). Outdoor air pollution and asthma. *The Lancet*, *383*(9928), 1581–1592. https://doi.org/10.1016/S0140-6736(14)60617-6
25. Hamedianfar, A., Shafri, H. Z. M., Mansor, S., & Ahmad, N. (2014). Improving detailed rule-based feature extraction of urban areas from WorldView-2 image and lidar data. *International Journal of*

Remote Sensing, *35*(5), 1876–1899. https://doi.org/10.1080/01431161.2013.879350

26. Hartig, T., Mitchell, R., De Vries, S., & Frumkin, H. (2014). Nature and Health. *Annual Review of Public Health*, *35*(1), 207–228. https://doi.org/10.1146/annurev-publhealth-032013-182443
27. Hartling, S., Sagan, V., Sidike, P., Maimaitijiang, M., & Carron, J. (2019). Urban Tree Species Classification Using a WorldView-2/3 and LiDAR Data Fusion Approach and Deep Learning. *Sensors*, *19*(6), 1284. https://doi.org/10.3390/s19061284
28. Hecht, R., Meinel, G., & Buchroithner, M. F. (2008). Estimation of Urban Green Volume Based on Single-Pulse LiDAR Data. *IEEE Transactions on Geoscience and Remote Sensing*, *46*(11), 3832–3840. https://doi.org/10.1109/TGRS.2008.2001771
29. Helbich, M. (2019). Spatiotemporal Contextual Uncertainties in Green Space Exposure Measures: Exploring a Time Series of the Normalized Difference Vegetation Indices. *International Journal of Environmental Research and Public Health*, *16*(5), 852. https://doi.org/10.3390/ijerph16050852
30. Hellesen, T., & Matikainen, L. (2013). An Object-Based Approach for Mapping Shrub and Tree Cover on Grassland Habitats by Use of LiDAR and CIR Orthoimages. *Remote Sensing*, *5*(2), 558–583. https://doi.org/10.3390/rs5020558
31. Hill-Briggs, F., Adler, N. E., Berkowitz, S. A., Chin, M. H., Gary-Webb, T. L., Navas-Acien, A., Thornton, P. L., & Haire-Joshu, D. (2021). Social Determinants of Health and Diabetes: A Scientific Review. *Diabetes Care*, *44*(1), 258–279. https://doi.org/10.2337/dci20-0053
32. Hu, H., Cao, Q., Shi, Z., Lin, W., Jiang, H., & Hou, Y. (2018). Social support and depressive symptom disparity between urban and rural older adults in China. *Journal of Affective Disorders*, *237*, 104–111. https://doi.org/10.1016/j.jad.2018.04.076
33. Huang, Y., Yu, B., Zhou, J., Hu, C., Tan, W., Hu, Z., & Wu, J. (2013). Toward automatic estimation of urban green volume using airborne LiDAR data and high resolution Remote Sensing images. *Frontiers of Earth Science*, *7*(1), 43–54. https://doi.org/10.1007/s11707-012-0339-6
34. Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*(1–2), 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2
35. Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, *25*(3), 295–309. https://doi.org/10.1016/0034-4257(88)90106-X
36. Institute for Clinical Evaluative Sciences. (2020). *Institute for Clinical Evaluative Sciences* [Dataset]. https://www.ices.on.ca
37. International Diabetes Federation. (2025). *IDF Global Clinical Practice Recommendations for Managing Type 2 Diabetes*. https://idf.org/media/uploads/2025/04/IDF_Rec_2025.pdf
38. Ismail, L., Materwala, H., & Al Kaabi, J. (2021). Association of risk factors with type 2 diabetes: A systematic review. *Computational and Structural Biotechnology Journal*, *19*, 1759–1785. https://doi.org/10.1016/j.csbj.2021.03.003
39. James, P., Banay, R. F., Hart, J. E., & Laden, F. (2015). A Review of the Health Benefits of Greenness. *Current Epidemiology Reports*, *2*(2), 131–142. https://doi.org/10.1007/s40471-015-0043-7

40. Jombo, S., Adam, E., & Tesfamichael, S. (2022). Classification of urban tree species using LiDAR data and WorldView-2 satellite imagery in a heterogeneous environment. *Geocarto International*, *37*(25), 9943–9966. https://doi.org/10.1080/10106049.2022.2028904
41. Labib, S. M., Lindley, S., & Huck, J. J. (2020). Spatial dimensions of the influence of urban green-blue spaces on human health: A systematic review. *Environmental Research*, *180*, 108869. https://doi.org/10.1016/j.envres.2019.108869
42. MacFaden, S. W., O'Neil-Dunne, J. P. M., Royar, A. R., Lu, J. W. T., & Rundle, A. G. (2012). High-resolution tree canopy mapping for New York City using LIDAR and object-based image analysis. *Journal of Applied Remote Sensing*, *6*(1), 063567–1. https://doi.org/10.1117/1.JRS.6.063567
43. Martinez, A. D. L. I., & Labib, S. M. (2023). Demystifying normalized difference vegetation index (NDVI) for greenness exposure assessments and policy interventions in urban greening. *Environmental Research*, *220*, 115155. https://doi.org/10.1016/j.envres.2022.115155
44. McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, *17*(7), 1425–1432. https://doi.org/10.1080/01431169608948714
45. Montoro-Ramírez, E. M., Parra-Anguita, L., Álvarez-Nieto, C., Parra, G., & López-Medina, I. M. (2025). Climate change effects in older people's health: A scoping review. *Journal of Advanced Nursing*, *81*(12), 8066–8079. https://doi.org/10.1111/jan.16270
46. Nguyen, P.-Y., Astell-Burt, T., Rahimi-Ardabili, H., & Feng, X. (2021). Green Space Quality and Health: A Systematic Review. *International Journal of Environmental Research and Public Health*, *18*(21), 11028. https://doi.org/10.3390/ijerph182111028
47. Omer, G., Mutanga, O., Abdel-Rahman, E., & Adam, E. (2016). Empirical Prediction of Leaf Area Index (LAI) of Endangered Tree Species in Intact and Fragmented Indigenous Forests Ecosystems Using WorldView-2 Data and Two Robust Machine Learning Algorithms. *Remote Sensing*, *8*(4), 324. https://doi.org/10.3390/rs8040324
48. Ong, K. L., Stafford, L. K., McLaughlin, S. A., Boyko, E. J., Vollset, S. E., Smith, A. E., Dalton, B. E., Duprey, J., Cruz, J. A., Hagins, H., Lindstedt, P. A., Aali, A., Abate, Y. H., Abate, M. D., Abbasian, M., Abbasi-Kangevari, Z., Abbasi-Kangevari, M., Abd ElHafeez, S., Abd-Rabu, R., … Vos, T. (2023). Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: A systematic analysis for the Global Burden of Disease Study 2021. *The Lancet*, *402*(10397), 203–234. https://doi.org/10.1016/S0140-6736(23)01301-6
49. Patino, J. E., Hong, A., Duque, J. C., Rahimi, K., Zapata, S., & Lopera, V. M. (2021). Built environment and mortality risk from cardiovascular disease and diabetes in Medellín, Colombia: An ecological study. *Landscape and Urban Planning*, *213*, 104126. https://doi.org/10.1016/j.landurbplan.2021.104126
50. Peel Region. (2021). *Data Portal* [Dataset]. https://data.peelregion.ca
51. Planet Labs PBC. (2021). *Planet Application Program Interface: In Space for Life on Earth* [Dataset]. https://api.planet.com

52. Posit team. (2025). *RStudio: Integrated Development Environment for R* [Computer software]. Posit Software, PBC. http://www.posit.co/

53. Ruvuna, L., & Sood, A. (2020). Epidemiology of Chronic Obstructive Pulmonary Disease. *Clinics in Chest Medicine*, *41*(3), 315–327. https://doi.org/10.1016/j.ccm.2020.05.002

54. Smith, L. (2025). *Peel Public Health addresses increase in diabetes through a focus on prevention*. https://peelregion.ca/press-releases/peel-public-health-addresses-increase-diabetes-through-focus-on-prevention#:~:text=–%20April%2025%2C%202025.,help%20prevent%20type%202%20diabetes.

55. Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, *11*(1), 54–71. https://doi.org/10.1037/1082-989X.11.1.54

56. Statistics Canada. (2016). *2016 Census—Boundary File* [Dataset]. https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-2016-eng.cfm

57. Statistics Canada. (2021). *Census Profile, 2021 Census of Population* [Dataset]. https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/index.cfm?Lang=E

58. Wang, H., Gholami, S., Xu, W., Samavatekbatan, A., Sleipness, O., & Tassinary, L. G. (2024). Where and how to invest in greenspace for optimal health benefits: A systematic review of greenspace morphology and human health relationships. *The Lancet Planetary Health*, *8*(8), e574–e587. https://doi.org/10.1016/S2542-5196(24)00140-2

59. Waser, L., Küchler, M., Jütte, K., & Stampfer, T. (2014). Evaluating the Potential of WorldView-2 Data to Classify Tree Species and Different Levels of Ash Mortality. *Remote Sensing*, *6*(5), 4515–4545. https://doi.org/10.3390/rs6054515

60. Wężyk, P., Hawryło, P., Szostak, M., Zięba-Kulawik, K., Winczek, M., Siedlarczyk, E., Kurzawiński, A., Rydzyk, J., Kmiecik, J., Gilewski, W., Szparadowska, M., Warchoł, A., & Turowska, A. (2019). Using LiDAR Point Clouds in Determination of the Scots Pine Stands Spatial Structure Meaning in the Conservation of Lichen Communities in "Bory Tucholskie" National Park. *Archives of Photogrammetry, Cartography and Remote Sensing*, *31*(1), 85–103. https://doi.org/10.2478/apcrs-2019-0007

61. Yan, W. Y., Shaker, A., & El-Ashmawy, N. (2015). Urban land cover classification using airborne LiDAR data: A review. *Remote Sensing of Environment*, *158*, 295–310. https://doi.org/10.1016/j.rse.2014.11.001

62. Yang, B.-Y., Fan, S., Thiering, E., Seissler, J., Nowak, D., Dong, G.-H., & Heinrich, J. (2020). Ambient air pollution and diabetes: A systematic review and meta-analysis. *Environmental Research*, *180*, 108817. https://doi.org/10.1016/j.envres.2019.108817

63. Yin, Y., Li, S., Xing, X., Zhou, X., Kang, Y., Hu, Q., & Li, Y. (2024). Cooling Benefits of Urban Tree Canopy: A Systematic Review. *Sustainability*, *16*(12), 4955. https://doi.org/10.3390/su16124955

64. Zhang, L., Wang, L., Wu, J., Li, P., Dong, J., & Wang, T. (2023). Decoding urban green spaces: Deep learning and google street view measure greening structures. *Urban Forestry & Urban Greening*, *87*, 128028. https://doi.org/10.1016/j.ufug.2023.128028

65. Zhou, W. (2013). An Object-Based Approach for Urban Land Cover Classification: Integrating LiDAR Height and Intensity Data. *IEEE Geoscience and Remote Sensing Letters*, *10*(4), 928–931. https://doi.org/10.1109/LGRS.2013.2251453

Figures

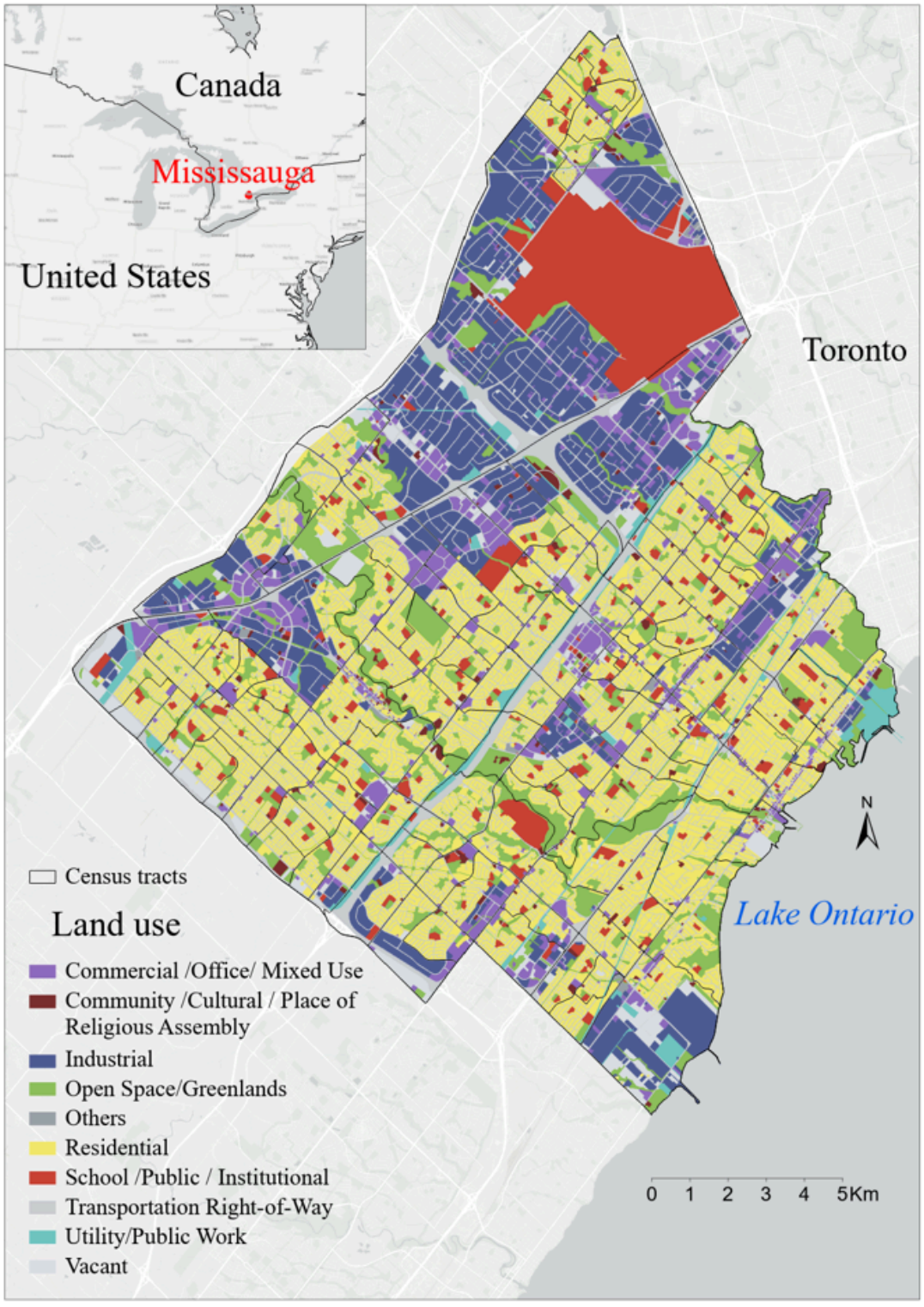


Figure 1

Study area of the 136 Mississauga census tracts with land use classification.

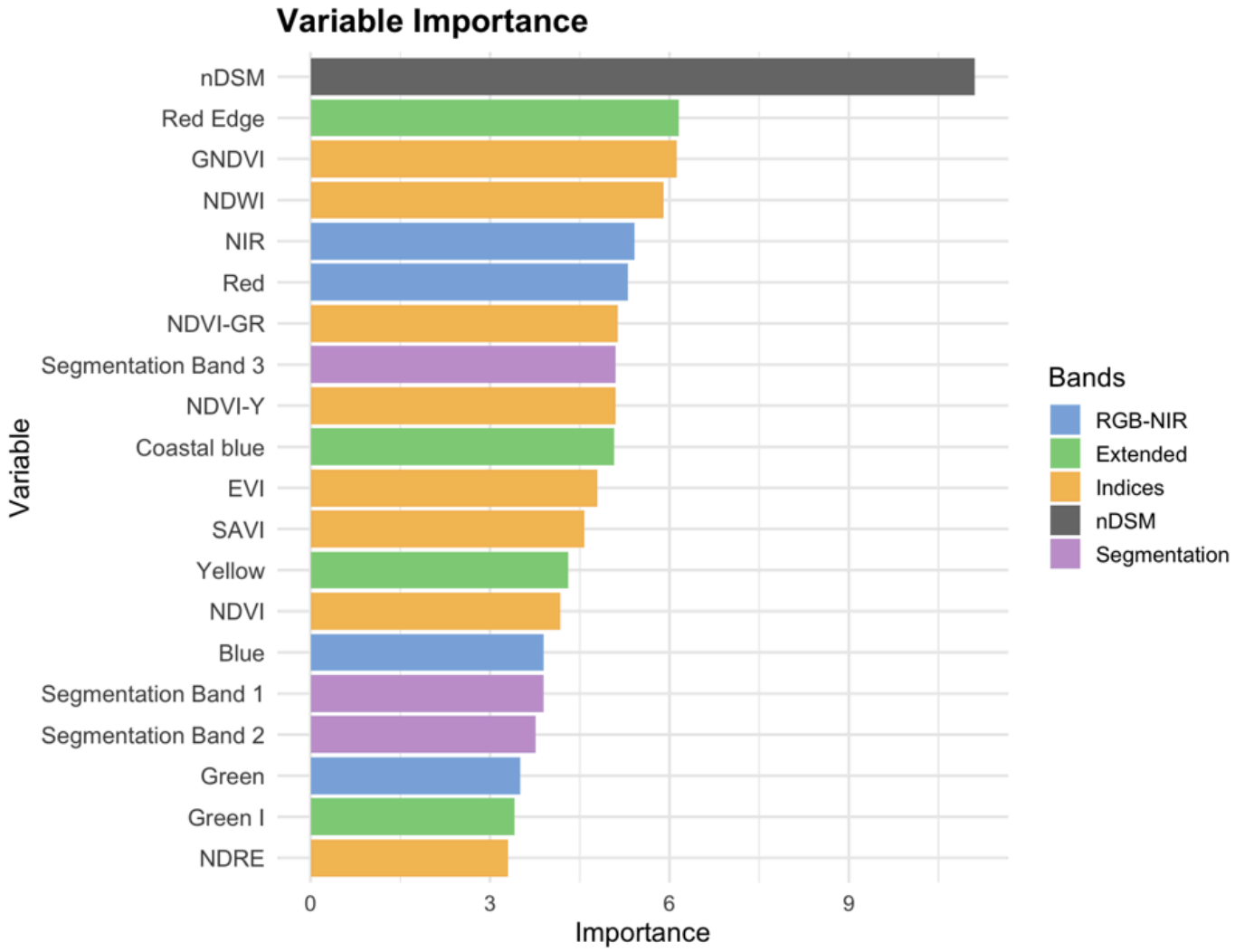


Figure 2

Variable importance of the classification model, including RGB-NIR, extended bands, indices, and normalized Digital Surface Model (nDSM).

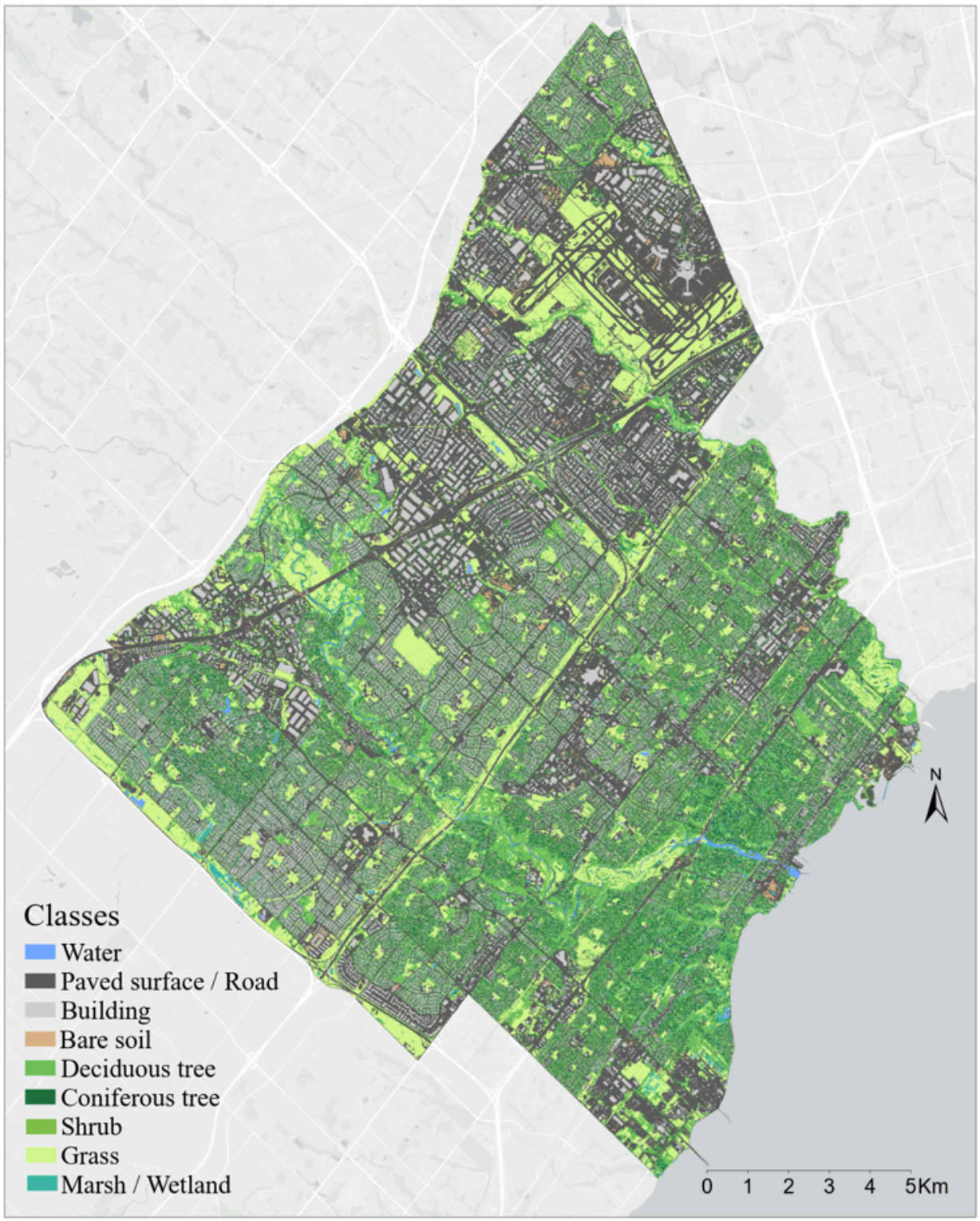


Figure 3

Classification map for land use land cover (LULC) in Mississauga, Ontario, using a random forest classifier.

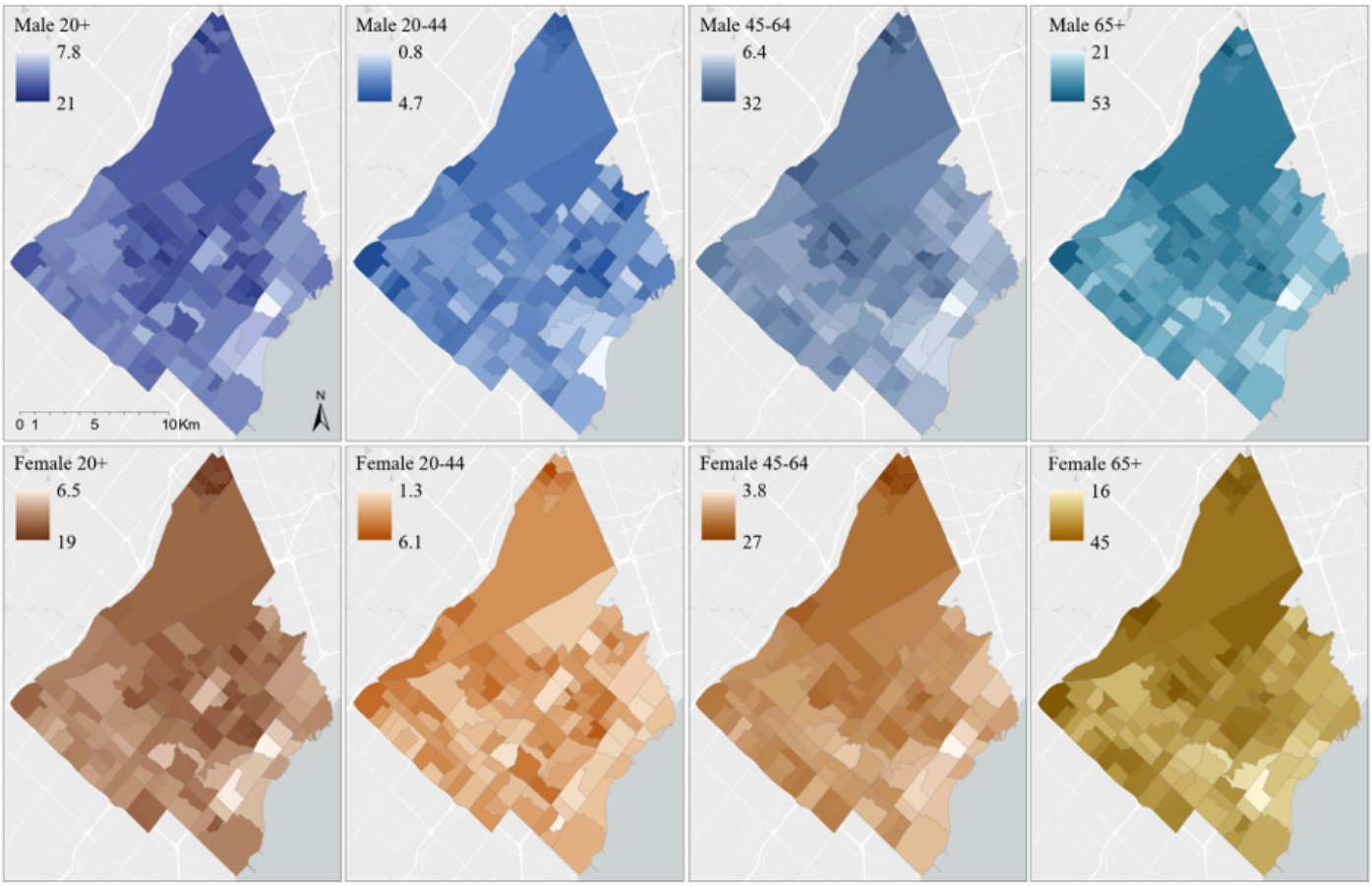


Figure 4

Diabetes prevalence per 100, by sex and age group in Mississauga census tracts.

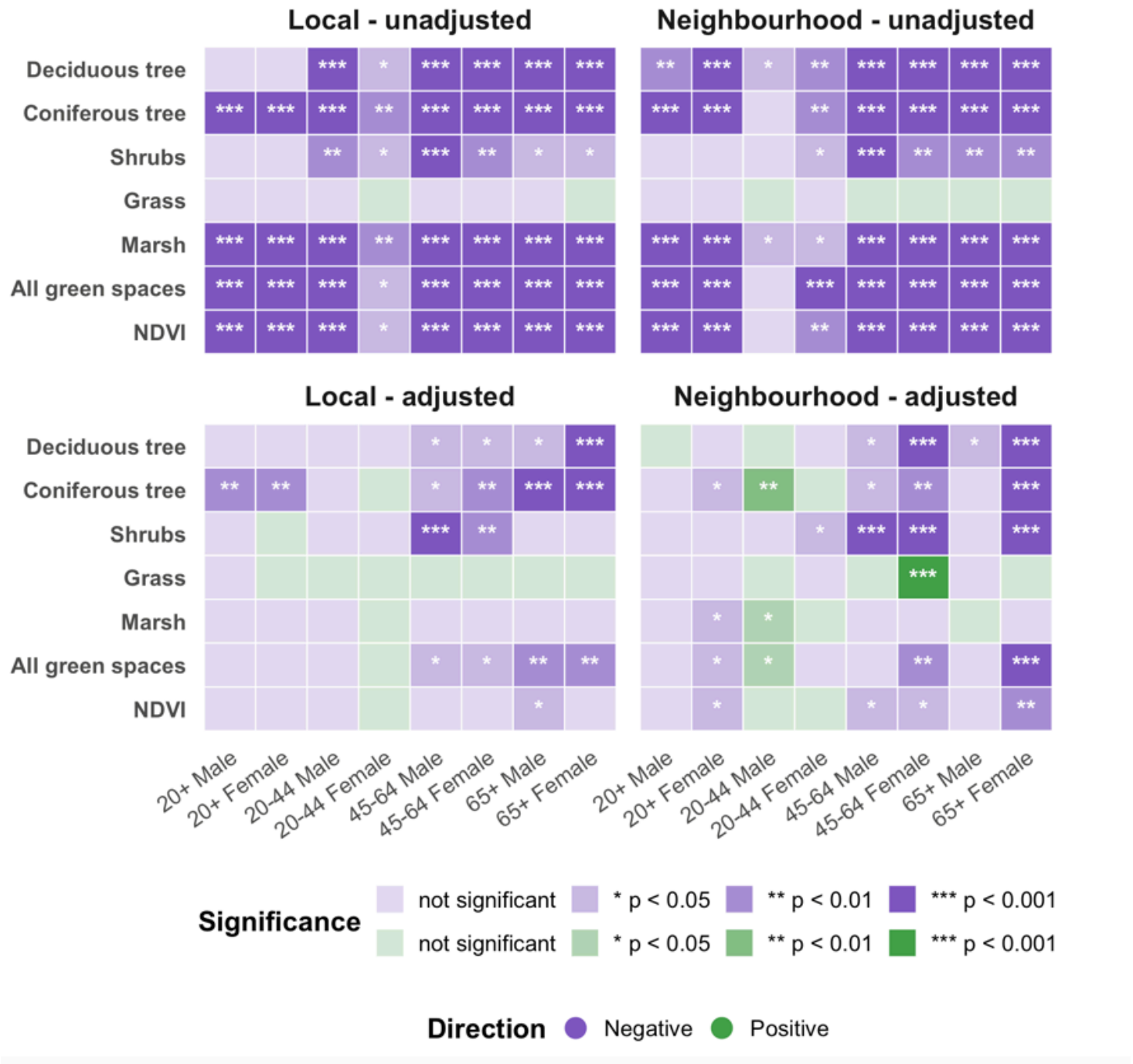


Figure 5

Direction and significance of associations between type-specific green spaces and diabetes prevalence across local and neighbourhood beta regression models, by sex and age group.