ALF-Score: A Predictive, Personalized, Transferable and Network-Based Walkability Scoring System

by

© *Ali M. S. Alfosool*

A dissertation submitted to the School of Graduate Studies

in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Department of Computer Science

Memorial University of Newfoundland

December 2021

St. John's Newfoundland

Abstract

Measuring the environments around us, including cities, roads, and social environments is crucial to understanding human behaviour. This knowledge can help us predict how aspects of the environment influence behaviour and health. Walkability is a popular measure of the environment used to describe various aspects of the built and social environment associated with physical activity and public health. Most existing methods are missing or underutilizing some crucial parameters that substantially impact measuring accurate walkability scores. For instance, road network structure is an integral part of mobility and should be an essential part of walkability but is not widely discussed in existing methods. Moreover, most walkability measures provide area-based walkability, or their scores are distributed with low spatial resolution. Additionally, individuals' opinions are not considered when measuring walkability. Furthermore, walkability is subjective, and although multiple definitions of walkability exist, there is no single agreed-upon definition. Existing measures take a one-size-fits-all approach without providing any personalization based on users' perspectives, leaving much more desired. In this research, Active Living Feature Score or ALF-Score [1] is proposed, which is a novel approach to measure walkability scores more accurately and efficiently while addressing existing limitations. ALF-Score incorporates road network structure to derive features such as network science centralities and network embedding, which are crucial in understanding the road structure better. ALF-Score utilizes user opinion to build a high-confidence ground-truth used to generate models capable of estimating walkability scores based on user opinion.

[1] https://alfscore.com/

By incorporating machine learning approaches in its pipelines, ALF-Score achieved a much higher granularity and higher spatial resolution of walkability scores at point level. ALF-Score introduced two new methods: 1) a combined graph reduction and reconstruction technique that focuses on reducing the number of nodes in a road network achieving an average of 77% reduction while preserving the core structure of the road network, and 2) the Generalized Linear Extension of Partial Orders or GLEPO, which enables the conversion of relative rankings to absolute scores. Moreover, ALF-Score+ extends ALF-Score by incorporating user demographics such as age and gender to capture profile clusters that help provide personalized walkability scores suitable for varying individual profiles. Additionally, ALF-Score++ improves the overall scalability of ALF-Score and further extends this measure by incorporating transferability to allow reusability of already-learned knowledge and previously detected patterns as a base for further and continued learning to help reduce training time, improve prediction accuracy, reduce resource consumption, and lower the number of labels needed for training. Most importantly, ALF-Score++ enables zero-user-input application, which allows predicting walkability scores for any location on the road without training models for that particular region with a low transferability loss of 13.28 units using Deep Neural Network approaches, and a direct training loss of only 4.56 units using shallow learners (MAE on a scale of 0-100).

Contents

List of Tables

List of Figures

List of Algorithms

Acknowledgements

I would like to express my sincere gratitude to my supervisor Dr. Yuanzhu Chen, and co-supervisor Dr. Daniel Fuller. Their continued support of my work, patience, motivation, and most importantly, guidance and expertise helped me navigate through the challenges of my research and turn them into this successful work.

My sincere thanks also go to the Department of Computer Science, School of Human Kinetics and Recreation, School of Graduate Studies, Wireless Networking and Mobile Computing Laboratory (WineMocol) and Built Environment and Active Populations (BEAP) Lab at the Memorial University of Newfoundland.

I am grateful and indebted to my family, especially my parents, for all their unimaginable sacrifices on a daily basis. Their unbounded support of all my decisions gave me the strength to believe in myself when no one else did and the courage to venture into worlds most only dream about. To my father, a great healer of bodies and souls, a philosopher and a teacher to many, who passed away in December of 2018, knowing that you believed in me, gave me all the strength I ever needed to get back up every time life knocked me down. I will never forget your beautiful smiles... Your words and wisdom flow through me.

To my beautiful mother, who has been a warrior for most of her life, I am an

explorer because of you... You have never stopped fueling my curiosity, empathy and kindness. You are the strongest, most determined and kindest person I have ever known. You have been to the dark, yet you shine bright... From you and him is the man I am today, who prays you both will always be proud of.

To my life partner, thank you for believing in me and making me a better person. Thank you for being so patient with me. Thank you for all the hot meals for the kids when I couldn't! You are the biggest star in my life with the kindest heart who has changed many lives for the better. You are a role model, and I am always proud of you and grateful to be by your side. Of course, we have a couple of kids together too, so it's really out of my hands at this point...! Thank you, kids, for all the background singing (hint hint!) these past few years! Your cuteness and love have no measure and pulled me through many unhappy times... believe me, it's true... I have pictures to prove!

Thank you to my intelligent brother for keeping me company and offering me wonderful, intriguing and challenging conversations, and thank you to my kindest little sister who never stopped to amaze me and for teaching me that kindness knows no bounds...

Chapter 1

Introduction

1.1 Motivation

Worldwide physical inactivity is associated with 9% of all premature mortality (5.3 million deaths per year), 6% of the burden of coronary heart disease, and 7% of type 2 diabetes. If the population meeting physical activity guidelines increased by 10%, more than 533,000 deaths could be averted every year worldwide [71]. In Canada, self-report data suggest that 60% of the population meet physical activity guidelines of 150 minutes of moderate to vigorous physical activity per week (MVPA), while device-based estimates using research-grade accelerometers suggest that only 39% of Canadians meet physical activity guidelines [27]. There is a clear need to increase physical activity among the population. Yet, how to increase physical activity among the population remains a challenge. Canada's physical activity rates have been relatively stable for the past 15 years [64]. The built environment, defined as "man-made or modified structures that provide people with living, working, and recreational spaces [3]", represents an essential method with the potential to increase physical activity at the population level in Canada and worldwide. For example, significant changes to the built environment in transportation interventions suggest that new cycling infrastructure and public transit improvements can improve population health. Overall, meta-analyses indicate that installing new public transit in a city increases physical activity overall [134, 60]. For cycling infrastructure, there are clear physical activity benefits [109, 68]. Today, as local and federal governments respond to COVID-19, built environment changes are unfolding quickly in an attempt to redesign cities with the health of residents in mind [36].

In addition to increasing physical activity and overall health, there is also a need to address the pressing issue of climate change. Built environment and policy changes can increase physical activity, which has essential health benefits and reduce greenhouse gas emissions. For example, Maislish et al. using a case study from San Francisco, showed that increasing median daily walking and cycling from 4 to 22 minutes by improving public transit and walkability reduced the burden of cardiovascular disease and diabetes by 14% (32,466 Disability Adjusted Life Years (DALYs), and decreased Green House Gas Emissions by 14% [81]. Minor changes to our mobility could have essential impacts on physical activity and climate change.

Unfortunately, our transportation environments, particularly our road networks, strongly encourage us to use personal motor vehicles rather than more physically active and sustainable transportation modes. Over 200 million drivers [122] are in the United States alone, averaging 76.4% [20] of people commuting to work. In Canada, 74% [84] of commuters, 11.4 million, drive to work. According to the 2009 National Household Travel Survey (NHTS) [122], less than 4 percent of commuting

Americans walk or cycle to work. In a similar survey [84] conducted in Canada in 2011, 7 percent of commuting Canadians walk or cycle to work, both representing a tiny percentage. Analyzing and studying road importance will help researchers better understand the city structure's underlying factors to improve city planning and increase the rate of people walking and cycling to promote a healthier and more active lifestyle and a cleaner and safer environment. According to a study [32] published by Statistics Canada, 12.6 million Canadians reported in 2016 to have commuted by car to work, with an average commute duration of 24 minutes. The median distance to the workplace is 8.7 kilometres. Furthermore, close to 1 million car commuters spent at least 60 minutes travelling to work. Trips to work, grocery stores, hospitals or even casual jogs and road trips mainly occur on the road network. Roads are how we connect. From family visits to business trades, from a local alley to an autobahn, roads and the road networks are crucial parts of our daily lives providing access to social activities, employment, education and health. They bring us critical social benefits and contribute to our economic development. The importance of roads and how they impact our lives is an already-established fact. Measuring this importance is one of the primary objectives of this research. With the increase of publicly available geographic information systems, road network data is now freely accessible from numerous sources. Due to this availability, road network data can be used as a good source for various types of analysis like road importance [21, 124, 5, 63] road characteristics [86, 82] city planning [17, 139, 116] creating walkability or bikeability scores or examining the association between different health conditions such as diabetes and obesity of local communities with their walkability scores [135, 66, 116].

Some previous works, such as Winters et al. [133] and Glazier et al. [51], have been inspirations for this research. These works use road scores, partly based on road connectivity, for various analyses. Specifically, Winters et al. created Bike Score. This measure calculates the bikeability of multiple cities, uses road connectivity and focuses on finding associations between urban bikeability and cycling behaviour, aiming to determine if Bike Score was associated with between and within-city variability in cycling behaviour of 24 different North American cities. Glazier et al. compared previously published walkability measures related to transportation behaviours, obesity and diabetes, in Toronto, by factoring in population density, residential density, availability of salable destinations and street connectivity. The latter found individuals who live in more walkable areas are over twice as likely to walk, cycle or use public transport and are significantly less likely to drive or own a vehicle compared with those living in less walkable areas, which are up to one third more likely to be obese or have diabetes.

Walkability is a measure that many researchers have used to operationalize characteristics of the environment that support walking. Walkability [117] is a term used to describe aspects of the built and social environment with significant population-level impacts on physical activity and health. Several systematic reviews have shown that walkability in a neighbourhood is associated with more physical activity among both children [39], adults [95], and older adults [11, 126]. Knowing how walkable an area is based on walkability scores is essential for public health, urban transportation and engineering research planning. Walkability scores provide a necessary insight into how the environment around us influences our behaviour and health.

1.2 Walkability

Walkability has been widely used in the academic literature since the late 1990s by urban planners, the general public, and researchers [77, 74, 83, 42, 75, 43, 136, 120] when they began to examine the association between walkability and various factors including travel behaviour, urban design, real estate, physical activity, and obesity. Although there are multiple operational definitions of walkability in the literature [56, 44, 48, 54, 97, 41, 110] there is no single agreed-upon conceptual definition of walkability. Grant (2013) defines walkability as an "excellent shorthand for good urban design" [54]. Wikipedia defines walkability as "a measure of how friendly an area is to walking" [128]. Wang and Yang, in their review and bibliometric analysis, define walkability as "the extent to which the built environment is friendly to people who walk, which benefits the health of residents and increases the liveability of cities" [126]. These definitions are rather broad to encompass all relevant environmental features that may lead to active living and not guide measurement. Simply put, walkability is a way to show how walkable/connected/accessible our surroundings are concerning walking. Researchers typically use walkability as a measure to operationalize characteristics of roads combined with other attributes, including population density, access to shops and services, and safety, among others. Knowing how walkable an area is (i.e. walkability score) is an essential factor in everyone's lives, especially how the spread of the COVID-19 implicated restrictions and limitations to how and when we can/should go outside. Walkability scores provide essential insight into the roads and neighbourhoods around us. However, using the road structure as nodes is not widely discussed in existing methods.

1.3 Related Work

Several existing walkability measures provide walkability scores for Canada, each with different strengths and limitations. Approaches have typically relied on using self-reports from individual participants to understand how walkable an area might be. For example, according to Carr et al. [23], previous efforts (before 2011) to measure neighbourhood walkability have primarily relied on self-reported data, using time-intensive and costly measures. For example, Duncan et al. [37] focused on certain buffer distances within the street networks such as 400, 800, 1600 meters. The buffer distance was ignored in most existing measures and was deemed essential. The second approach uses geographic information systems (GIS), which combine different environment features like the number of street intersections or the population density to create neighbourhood-level walkability measures. While several city-specific walkability measures have been developed, two prominent, national-level walkability measures are available in Canada: Can-ALE and Walk Score. For comparison, the Canadian Active Living Environments measure (Can-ALE) [21] and Walk Score [124] were chosen as they are both commonly used by researchers and end-users alike. These measures each have different strengths and limitations.

1.3.1 Can-ALE

Can-ALE [Figure 1.1 right] is one of the most prominent and publicly available walkability metrics. It is an area-based measure at the Dissemination Area (DA) level. A DA is a small, relatively stable geographic unit composed of one or more adjacent dissemination blocks and is the smallest standard geographic area for which all census data are disseminated [112, 111]. DAs' physical size may vary depending on their geographical location, but they usually have a population of between 400 to 700 persons. Canada is divided into DAs covering all its provinces and territories, and there are approximately 54,000 DAs in Canada. Can-ALE, widely used for walkability scores in Canada, utilizes intersection density and dwelling density measures to devise a walk score for each Canadian DA. The measure also includes the number of Points of Interests (POI) within a 1KM buffer around the DA's centroid. According to *Canadian Active Living Environments Database (Can-ALE) User Manual & Technical Document* [22] "Can-ALE measures are based on one-kilometre, circular (Euclidean) buffers drawn from the centre points (centroids) of dissemination areas (DAs)". Can-ALE's only measure of road network importance is a count of the number of three (or more) way intersections per square kilometre of the buffer around a dissemination area's centroid.

For the 2016 Can-ALE dataset, authors produced the following measures for all DAs in Canada:

- Intersection density
- Dwelling density
- Points of interest
- Transit measure (transit stops for all DAs within Census Metropolitan Areas (CMAs) in Canada)

Intersection Density measure is defined as "directness and connectedness of streets and/or paths through a community" which was "done by counting the number of three or more way intersections within a one-kilometre buffer of the dissemination area (DA) centroid". Authors used both Road Network Files by Statistics Canada and OpenStreetMap roads and footpaths for this purpose, and from both datasets "limited-access highways (e.g., freeways, 400-Series Ontario Highways, Quebec Autoroutes) and highway entrances and exits are removed from the files before calculating intersection density, as these roads typically restrict active transport". Off-road footpaths and recreational trails were added to the files.

Dwelling Density measure is defined as "the average dwelling density of the DAs in the buffer area" which is "a common measure of active living environments and strongly correlates with active transportation rates".

Furthermore, the Points of Interest measure is defined as "the number of points of interest (POIs) in a one-kilometre buffer around the DA centroids," and "include a wide range of potential walking destinations (e.g., parks, schools, shops, places of business, landmarks, etc.)". Can-ALE includes almost all POI keys extracted from OSM under the Key "amenities". However, excluded alpine huts, caravan sites, wayside crosses, and other features that likely have no relationship with walking.

Transit measure is defined as "presence of public transit stops in the community" which involves a process similar to that of POI extraction. It was noted that this measure only covers DAs within Census Metropolitan Areas (CMAs).

1.3.2 Walk Score

Walk Score [124] [Figure 1.1 left] is another well-known walkability tool that uses a proprietary method that has several features, including population density and road metrics such as block length and intersection density. Its authors use a patented system and analyze hundreds of walking routes to nearby amenities. Although it is based on a closed-source system using proprietary methods, it is unclear if the scores are calculated for every location instead of general areas that provide distributed scores down to locations within each region. Their walkability score is awarded based on the distance to amenities of specific categories (eg. Grocery stores, coffee shops, restaurants, movie theatres, schools, parks, libraries, book stores, fitness centres, drug stores, hardware stores, clothing/music stores). Amenities within a 5-minute walk (.25 miles) are given maximum points. They use a decay function to devise what score should be given to more distant amenities, where no score is given for amenities located farther than a 30-minute walk. Walk Score also measures pedestrian friendliness by analyzing population density and road metrics such as block length and intersection density. Their data sources include Google Maps, Factual, Great Schools, Open Street Map, the U.S. Census, Localeze, and places added by the Walk Score user community. Walk Score is a rank between 0-100 where:

- 90–100 describes "Walker's Paradise" where daily errands do not require a car
- 70–89 describes "Very Walkable" where most errands can be accomplished on foot
- 50–69 describes "Somewhat Walkable" where some errands can be accomplished on foot
- 25–49 describes "Car-Dependent" where most errands require a car, and
- 0–24 describes "Car-Dependent" where almost all errands require a car

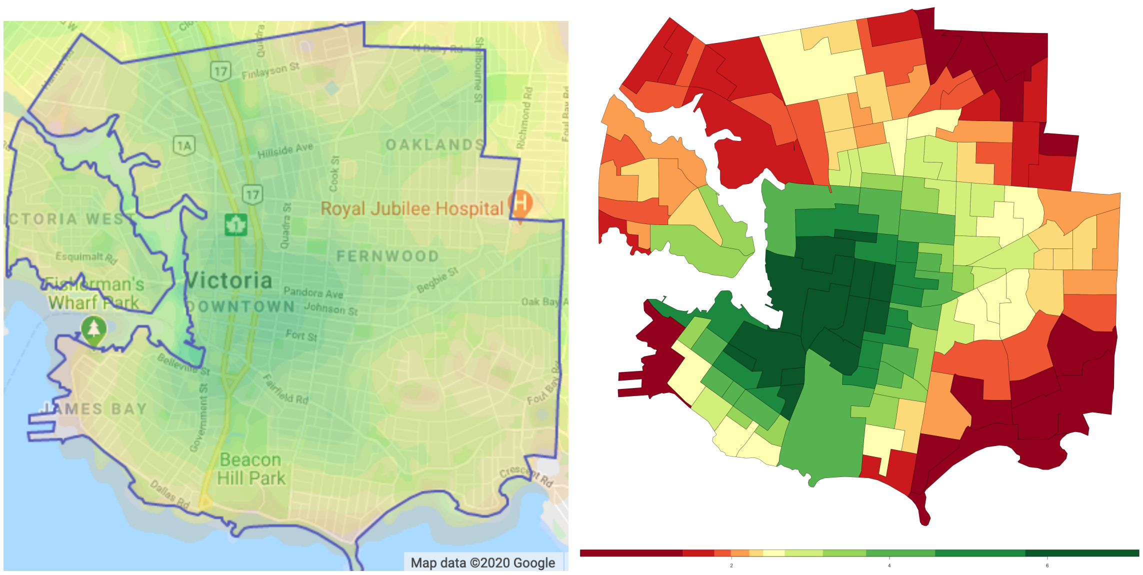

Figure 1.1: Walk Score [124] Walkability score of Victoria, BC (left - Screenshot taken from walkscore.com). Dark green is the most walkable, dark yellow/orange is the least walkable. Can-ALE Walkability score of Victoria, BC (right - generated through RStudio [103] Version 1.2 from rstudio.com), assigned by dissemination area. Dark green is the most walkable, dark red is the least walkable.

1.3.3 Limitations of previous approaches

There are several limitations regarding previous approaches that aimed to measure walkability. Both Walk Score and Can-ALE are heavily used/cited [28, 58, 29, 46]. But there are some noticeable drawbacks and opportunities for improvements. These limitations are significant and likely the result of limited interdisciplinary work between computer science, public health, and urban planning. The critical limitations of previous research works include:

1. Incomplete use of road structures
2. Lack of predictive models
3. Low spatial resolution
4. Lack of user opinion
5. Lack of personalization
6. Limited transferability to new cities

The following section will discuss the limitations of previous walkability measures. Some limitations are independent, while others may be directly or indirectly associated with other limitations. Specifically, predictive models, spatial resolution and personalization tend to go hand in hand. Using a predictive approach to estimate walkability scores allows training models on personalized demographics and generating walkability scores at higher spatial resolution.

1.3.3.1 Incomplete use of road structures

Road networks have been used in many analytical studies [18, 25, 138] to associate road structure with various topics. For example, R.S. Mahabir et al. studied the impact of road networks on the distribution of dengue fever cases [80] based on the data observed in 1998 in Trinidad and West Indies. Their result represented "the first evidence of dengue cases being found restricted between forested areas and major highways" which they found very useful in planning and implementing mitigation strategies to control dengue and Aedes aegypti mosquitoes. But the importance of road networks has been well-known for thousands of years. Chengjin Wang et al. studied the evolution of road networks in China between 1600 BC to 1900 AD [125] and suggested that road networks were used for major nationwide socio-economic exchanges long before modern transportation emerged and have been in long-term development and are continuously expanding. According to the authors, this expansion of road networks with continuous change in accessibility and coverage is characterized by a "core-peripheral configuration" closely associated with natural conditions and national defence and warfare. Furthermore, they found the development of the road networks are associated with "centralization of national power, national land governance, postal transport, the transport of specialized cargos, and international trade".

The importance of road networks should undoubtedly be considered when creating a walkability metric. A simple description of walkability is a way to show how walkable are the roads around us when it comes to walking around neighbourhoods. The influence of the roads themselves towards these scores is a significant component that differs from metric to metric. Road structure and connectivity play essential roles in how neighbourhoods are connected. Although there is an indirect use of the road structure (i.e. intersection density), in some of these measures (such as Can-ALE), road network and road structure as nodes are not fully utilized. Walk Score does not appear to use this directly, but it was not possible to verify that due to its closed-source design. However, Can-ALE's only influence of roads was accumulated in the intersection density parameter, the number of intersections with three or more ways per square km within the buffer zone. Let us imagine two locations, each with five intersections per square kilometre. One of these locations is on the edge of a small industrial region, while the other is within some of a large city's busiest and most popular attraction sites. Would it be meaningful if their road-structure-based scores carry the same weight due to having the same number of intersections within the buffer zone?

Just as road structure matters, network science approaches such as centralities and network embedding can extract valuable information from road networks and are just as important. These approaches are currently missing from existing walkability measures, yet they are integral parts of our mobility and should be essential for measuring walkability. Additionally, computation for large road networks has proven to be a time-consuming task that poses an additional challenge if left unchecked.

Road networks and other geographical-based data are expanding every day, while there is still a significant gap in processing this ocean of data. Lack of accessibility to computational resources such as superclusters or high-speed workstations to the general public (where only small groups such as researchers may have access to these resources) has led to the need for finding alternative ways to produce similar results with significantly lower computational requirements.

1.3.3.2 Lack of predictive models

Existing walkability measures are developed as stand-alone calculations. They are not designed to predict scores, and calculated scores may not be used in different areas than they have initially been calculated for. One reason behind this is that some of these research studies come from areas that tend not to consider using predictive machine learning approaches [47]. As a result, existing measures are generally computed once and only recalculated after a significant change in the data is observed or a substantial improvement to the approach is made (for example, Can-ALE in 2006 and 2016). The results may turn outdated due to the changes in the road structures and POI depending on many variables such as population, region size, and urbanization level. Walkability measures have not used predictive machine-learning-based approaches such as supervised and semi-supervised learning to build predictive models capable of predicting walkability scores in areas where data may not be available. Building predictive models enable us to address this challenge by utilizing smaller datasets that rely on multiple input features, including the road network, POI, and user data, learning from various elements, and projecting and accurately predicting walkability scores for large road networks, cutting the computational time and required technical resources to a fraction of that achieved through some traditional methods.

1.3.3.3 Low spatial resolution

Two aspects limit the spatial resolution of traditional walkability measures: 1) reliance on census-based geographies, including dissemination areas, and 2) use of buffers to calculate summary measures. As is evidenced in Figure 1, using dissemination areas does not capture the local feel of walkability for someone on the road. Dissemination Areas are on average 1-2 city blocks and can be much larger as population density decreases. As a result, walkability measures that use dissemination areas as their unit of analysis may mask a significant variation in walkability within each dissemination area. For instance, a drawback of Can-ALE is the 1KM radius limitation. There are cases where population density has caused the DA structure to fall shorter or much greater than the 1KM radius. This phenomenon causes the capture of either too many outside-DA variables or not enough parameters, among other issues, all leading to inaccurate results. Additionally, a 1KM radius may not be considered walkable for many people. Walk Score uses a decay function to address this.

1.3.3.4 Lack of user opinion

In most existing measures, user opinion is not used. These measures typically follow a one-size-fits-all approach mainly influenced by the researchers' perspective instead of the end-users view [12]. As a result, their walkability scores may be biased by characteristics deemed by researchers and developers as more or less critical. User opinion is a crucial yet missing piece that will substantially impact how the walkability scores are derived and assigned, reflecting various commonly observed patterns seen in people's daily lives.

1.3.3.5 Lack of personalization

Existing measures are not flexible and generalized to address all users and parameters with many assumptions. None of the mentioned methods utilize user opinion as one of their internal factors to derive their walkability scores nor try to bind their scores with a personalization component that would influence or represent the changing needs of individuals and various groups. Moreover, existing methods are typically calculated for the entire network only once. Walkability should be user-opinion-based and personalized to represent the population and valuable for various groups with different criteria.

Existing walkability measures are typically generalized with no consideration for individuals' preferences, opinions, or how they interact with the environment. Walkability scores generated by generalized algorithms are the same for everyone regardless of the individual. However, people live in different environments and have other priorities and preferences for what they would consider as walkable based on their personal needs. A single parent who is unemployed and has a low income with young children may likely have a completely different set of priorities and preferences for what they would consider a walkable environment compared to that of a higher-income working couple with no child. Similarly, a young single student with no child may find areas close to their school, library, fast-food restaurants, and bus stop more walkable. In contrast, a professional working parent may consider areas closer to parks and playgrounds, schools and daycares, grocery stores, and hospitals or clinics as more walkable. There is a strong need to develop a personalized walkability measure that is accurate and efficient. A personalized walkability measure can address many existing limitations and contribute new data and methods that have not been used in the field of walkability research development.

1.3.3.6 Limited transferability to new cities

Walkability metrics are city-specific and require all necessary data for a given city to be available. Some cities may often lack the required data (e.g. POI, population density, etc.) needed to calculate walkability scores using traditional measures. Furthermore, minor differences in the data may make comparisons between cities very complex. None of those existing methods, as mentioned earlier, have predictive capabilities that would allow them to fully utilize available features to build models that learn from data and can estimate scores for locations never seen before or have the ability of continuous and transferable learning, instead of calculating walkability score for the entire network, which is a very time and resource-consuming task.

1.3.3.7 Summary of limitations

These limitations leave a gap in our ability to measure and understand the actual impacts of walkability. Network-based walkability utilizes road network structure and embeds a solid link to these crucial components currently missing from most existing methods. It combines features that tie each element to a specific weight using machine learning methods to build predictive models that generate network-based walkability scores. Furthermore, walkability is very subjective, and network-based walkability brings to the table a systematic metric that uses user opinion and enables personalization capability to address each individual's profile.

1.4 Objectives

As an overall goal, this research aims to create a novel, reliable, precise, efficient and widely applicable measure of walkability that can help better understand and plan for city structures, improve overall health and physical activity at the population level, reduce traffic congestion, fight climate change and enhance sustainable transportation [45]. Neither Can-ALE nor Walk Score provides user opinion input, predictive models, personalization, or the use of road network structure as nodes. This research works towards addressing these missing components and aims to improve the accuracy and feasibility of walkability scores by using features based on road network structures and points of interest to generate models capable of estimating network-based walkability scores for any point within a road network and of doing so Active Living Feature Score (ALF-Score) was developed. ALF-Score is a measure that provides a faster and better walkability scoring system that aims to fill in the gap and address significant limitations of existing measures by introducing the use of opinion-based crowd-sourced user parameters in conjunction with road network characteristics to provide scores that better represent individuals' perspective of walkable areas while utilizing road network structure to improve accuracy and spatial resolution.

This research utilizes road networks as nodes to better understand the underlying structure that impacts the walkability score of every single location and applies it in such a fashion that brings out the uniqueness of all locations based on road structure and network connectivity. But furthermore, it shifts away from DA-based and radius-based approaches towards a point-based system that uses road networks and various POI categories. This approach allows for higher spatial resolution and potentially unique scores for every point on the road. Moreover, scores can be personalized to individuals' unique profile features. This personalization is derived from user-collected data used towards user profiling and building machine learning models that are flexible enough to bend the curve to find the most relevant and suitable predictions for each user based on their profiles.

One of ALF-Score's significant contributions is using user opinion data as a feature to influence walkability scores using what general users think of the region. This approach paves the path to build a predictive and personalized walkability measure that utilizes user demographics and profiles and user opinions to build walkability models capable of estimating walkability scores associated with individuals' profiles for any point within a road network currently missing from existing methods.

It is important to emphasize that this research does not aim to add yet another component to existing metrics. Instead, the goal is to reconstruct how the entire process works. This research utilizes road network structure, user opinion, machine learning, personalized models based on user demographics, transferability of learned models and various other methods to generate well-trained, accurate and transferable personalized models that predict walkability scores based on each user profile for any point within a road network at point-level with high spatial resolution, anywhere and anytime.

This work uses interdisciplinary research from computer science and public health to address the limitations of some previous research studies. This interdisciplinary research is not without challenges outlined in the next section.

1.5 Challenges

1.5.1 Incomplete use of road structures

The first step is to embed road network and network science approaches into this research to address the lack of network features as nodes and build a more accurate walkability scoring system. Furthermore, various machine learning techniques will be used that utilize road network data and other essential elements. Moreover, users' opinions are embedded to reflect a real-world perception of people's views into the generated scores. Additionally, a pipeline will be put together that is dedicated to producing highly accurate, predictive, robust, personalized and network-based walkability scores based on each user.

POIs, when used without consideration for road connectivity, distance, frequency, and relevance (i.e. weighting less relevant POIs with the same score as those more relevant) will lead to higher inaccuracies and less relevancy to varying POI categories and individual users. For example, for some users, grocery stores, daycares, and hospitals may be much more important, whereas, for some others, public transit stops and restaurants may be considered more important. Each category and subcategory will have a significant role for individuals with various requirements, needs, and preferences. Furthermore, POI connectivity on the road network, such as direct distancing as opposed to those separated by freeways or those in a different DA that fall within the 1KM radius of the centroid, plays a crucial role in determining a more suitable and personalized approach to measuring walkability scores. In most existing measures, many features that do not carry the same weight contribute equally towards measuring walkability scores, leading to less accurate and less reliable final scores.

1.5.2 Lack of predictive models

In existing works, when changes are observed in any of the data parameters or if the authors decide to update their walkability scores, the process requires a complete re-run of the entire computational operation. A time-consuming, resource-intensive process and the results may quickly get outdated with any neighbourhoods' structural shifts, which may be due to physical, geographical, environmental, ecological, sociological and economic changes. A predictive approach that could allow for continuous learning and avoid complete re-runs of the entire process is currently missing from existing measures.

1.5.3 Low spatial resolution

Measures such as Can-ALE are area-based, which means a large area (in the case of Can-ALE, a dissemination area or DA) is associated with a single walkability score. Although DAs vary in size, each has a population of between 400 to 700 people. A DA in a dense and heavily populated area may maintain many roads, buildings, shops, or other points of interest within a small area. However, a DA could also be isolated entirely, covering an extensive geographical area with a segregated population and a significantly smaller number of POIs. DAs may be completely different in structure, available resources and accessibility. Having a single score representing the walkability of the entire DA leads to a deficient representation and will not be meaningful for everyone. This lack of representation is because all points within each DA will always carry the same walkability score regardless of their specific location, structure or conditions.

1.5.4 Lack of user opinion

Without user opinion, walkability scores may be generated through a one-size-fits-all approach that is influenced and biased by the person(s) who decides what features or characteristics should be used and what priorities are given to each feature. However, when user opinion is used, every user is entitled to their personal opinion, and the expectation is that these opinions will vary. Hence, if a user deems certain features as more important to them, such as restaurants or bus stops, it does not mean every other user will conform to this importance of features. Although people tend to follow routines and have shown similar patterns, there will likely be individuals who may have similar and different opinions. Dealing with this difference of opinion is a significant challenge that will be addressed.

1.5.5 Lack of personalization

When it comes to personalization, there are several different challenges to consider. The first challenge is that a considerable amount of data is required to create meaningful features, user-opinion database and user profiles. Data that can be burdensome to collect from users. An additional challenge is that when user profiles are created and reduced to cluster profiles, each cluster profile may be associated with only a fraction of the user data from the original data submission pool. Each cluster profile needs to maintain enough labelled data for training and testing purposes to build machine learning models specific to each profile's demographics.

A personalized walkability pipeline needs to train its models so that they conform to different user profiles. How these profiles are defined is crucial to understanding how users are grouped based on factors that influence their perception of walkability. As a result, the first goal of building user profiles is to collect appropriate data and identify patterns from a diverse set of users. These user profiles are then analyzed and grouped into a further abstraction layer, called a cluster profile. These cluster profiles are then fed into the extended ALF-Score pipeline to generate personalized walkability models for each cluster profile. The precision of the data collection, sample size and the conflicts found among user opinions pose additional challenges that will be addressed in this research.

The aim is to take a significant step further from being a user-opinion-based pipeline and move to a personalized system. The pipeline considers user demographics such as age, gender, preferred walkable distance, profession and other profile parameters that allow the predictive models to estimate walkability scores based on specific user profiles with personalized, predictive scores for each user. Finally, a thorough exploration of the scalability and transferability of this pipeline will be done by applying the personalized, predictive models to user-based data collected from a larger city in Canada (Montréal QC).

1.5.6 Limited Transferability to New Cities

Due to the structure of how existing measures compute their scores and the lack of full incorporation of machine learning or utilizing predictive models, their pipelines lack transferability, which means if scores for one city are measured, they may not be applied to other cities directly or indirectly to help towards generating new scores. Furthermore, there are no known pre-trained walkability models available for public use to facilitate the transferability of existing measures.

1.6 Scope of Research Project

This work builds on the theory that accurate walkability scores should be easily and readily accessible for every single point within the road network, with the scores being predicted instead of calculated and personalized to individuals' profiles instead of generic one-size-fits-all to best represent the most accurate walkability scores as perceived by each individual. This research provides a practical approach by redesigning how walkability scores should be measured from the ground up. New and improved algorithms are developed to utilize many datasets and derive the most relevant feature sets. Furthermore, as road connectivity and structure play crucial roles in measuring walkability, this research extends their use to utilize them fully. Road connectivity and structure are used in various ways, such as road centrality metrics and road embedding representations.

This approach does not rely on region-based computations such as DA-based, nor does it depend on fixed-term limitations such as using data within a 1-km radius of the centroid of a region or applying the same weight to all POI categories. This research does not make any assumptions on what POI categories are more important nor treat them equally. This research does not provide a generalized metric that produces the same results for everyone. Instead, this research aims to build a personalized approach that uses users' opinions and profiles to understand each individual's needs and build models that are capable of predicting results that may be unique for each individual. Unlike Can-ALE and Walk Score, which follow rule-based deterministic approaches, this research develops a model-based approach to consider different interactions in different scenarios. Furthermore, various machine learning algorithms are utilized for training models that are capable of predicting high-resolution walkability scores based on users' profiles for any given point within the model's known road network, as well as utilizing transfer learning to induce transferability and Zero-User-Input capabilities to predict accurate and personalized walkability scores for any point within any given road network that was never seen before by the model.

There are four phases in this research. The first phase (Chapter 3 - Graph Reduction and Reconstruction) involves an in-depth examination of road network structures and graph reduction methods for road networks. This work was published and presented at the 2020 International Conference on Computing, Networking and Communications [5]. The published paper is slightly revised for flow in this dissertation document, and the introduction is shortened. The second phase (Chapter 4 - ALF-Score: a Predictive Network-Based Walkability System) focuses on building a better and faster way to predict walkability scores with point-based high resolution while addressing some of the challenges of scalability when working with large networks. Chapter 4 was submitted for publication and is currently under review. The third phase (Chapter 5 - ALF-Score+: Personalization of a Predictive Walkability System) focuses on the inclusion of user demographics in addition to user opinion to allow user profiling and personalization to associate users' preferences and characteristics with their preferred walkability scores to be able to predict personalized walkability scores for each individual. Chapter 5 was submitted for publication and is currently under review. The last phase (Chapter 6 - ALF-Score++: Transferability of a Predictive Walkability System) investigates and implements measures to scale up the entire processes in phase 1 through phase 3, so they can apply to much larger cities with a focus on implementing transferability measures that allow transferred learning where global models can be trained based on a small set of user data from only a few cities, capable of predicting accurate walkability scores for any point in any city. Chapter 6 was submitted for publication and is currently under review.

1.7 Research Questions

MAIN How walkable is our surroundings, and how to accurately and efficiently measure walkability scores personalized for different individuals?

- Why does it matter? Walkability is crucial as it is directly associated with mobility and affected by our surroundings and living environment. There is no agreed-upon unified measure of walkability, and current measures are limited and generalized at best. Having the ability to generate accurate walkability scores based on individuals' needs can help improve the overall population's health by promoting an increase in outdoor activities, understanding city structure, assisting with better planning of new city structures, reducing air pollution by encouraging walking and cycling instead of vehicular use and incentivizing people to walk more often. Every individual has their criteria of what constitutes as walkable. Current measures are generic, do not consider personalization and are not spatially high-resolution. They do not consider some of the most important factors such as road networks, road centralities and similarity patterns leading to imposed limitations and restrictions within their metrics.

SUB 1 How to calculate network centrality for very large networks using limited resources (e.g. on personal computers) in a fraction of the time needed to do so using conventional algorithms?

- Why does it matter? Large road networks have millions of nodes and edges, which lead to higher complexity and slower computation. Some computations could take days, weeks or months to complete on some personal computers, which restrict the use of road networks if the required resources are not met.

SUB 2 How to accurately predict walkability score for any given location using road network data as nodes and user opinion?

- Why does it matter? Conventional approaches do not fully utilize road network connectivity as nodes, which is essential in determining walkability. Newer roads, alleys, walkways, etc. that are added continuously are not ranked, and the scaled-down ranks are arbitrarily assigned. Conventional approaches do not have any self-learned components, and small changes in data/method require the entire calculation re-processed. They are very time-consuming and resource-intense and require large datasets to calculate walkability scores. They require too many manually-performed and segregated components and yield lower accuracy with no flexibility in their pipelines. Furthermore, user opinion is not used in most existing measures, making them generic one-size-fits-all methods influenced mainly by the researchers' perspectives instead of the end-users.

SUB 3 How to personalize walkability scores based on specific user profiles, using small user-sample datasets?

- Why does it matter? The mere fact that walkability remains a very subjective matter points to the need for a personalized walkability measure where each individual can enjoy personalized scores that are most suitable to their daily needs and preferences, based on their profile and demographics. Most previous research studies are generalized, with none focusing on user input or personalization and lack high-resolution walkability scores that vary from person to person as their criteria change.

SUB 4 How to implement transferability in walkability using transfer learning to allow estimation of walkability scores for any never-seen-before locations?

- Why does it matter? Data acquisition is costly and time-consuming, so is recalculating walkability scores for new locations. With today's demand, there is a growing need for re-using existing knowledge to save time and resources and improve performance. For this research to be applicable and useable in various scenarios, its pipelines need to work seamlessly over small datasets and be capable of using pre-trained models in conjunction with new data to learn new patterns. More importantly, the ability to transfer previously learned knowledge of small cities to predict walkability scores of different and larger cities with no user data will provide a tremendously important component to this research.

1.8 Contribution & Potential Outcomes

This research addresses one of the most prominent challenges among existing walkability measures as previously outlined. It mainly focuses on three variations of ALF-Score as a path to building a faster and better walkability scoring system using predictive and user-based personalized parameters to promote a healthier lifestyle and raise awareness about the impacts of city structure on our daily lives. However, the underlying approach is not limited to user-based personalized scores. Still, it is customizable on other parameters such as road classification, connectivity distribution, population exposure, local centres (hubs), weather information, etc. Suppose one was to modify the parameters of this pipeline. In that case, one could generate predictive and personalized models based on traffic congestion, area population, number of traffic lights, etc., to produce scores matching the requirements of their study.

ALF-Score is a network-based walkability measure that utilizes road network structure alongside user opinion and other features through machine learning approaches to build predictive models capable of generating high spatial resolution and network-based walkability scores. Below is a list of the main contributions of this research:

- Building an entirely new metric - ALF-Score (Active Living Feature Score) - and ability to measure much more precise, personalized, predictive, transferable and network-based walkability scores.
- Building new algorithms that allow for significant reduction of graph nodes while preserving the structure of the graph intact leading to much smaller networks capable of representing their much larger original networks.

- Utilizing interpolation to estimate node centrality for road nodes and reconstructing full node centrality measures for an entire network by going through the reduced networks.
- Providing a new approach that significantly speeds up similar processes compared to existing methods and ability to generate similar results as in traditional methods but in a much shorter period and requiring much less technical resources such as the necessary processing power in the form of superclusters, GPUs and very powerful servers.
- Applying reduction and reconstruction techniques to St. John's, NL road network and ability to match the result with those calculated using traditional methods, yet in a fraction of the time.
- Building a pipeline towards generating a step-by-step approach to produce predictive models capable of predicting walkability scores.
- Building predictive models ready to provide instant predictions for any given point with an outcome of predictive regression results.
- Building a completely new method of converting relative rankings to absolute rankings which can be very much useful in various scopes and fields such as ranking every member of one team (4-6 members) among the team and converting it such that the numbers measure up to everyone across all teams, or ranking 5 locations concerning one another and converting it such that each locations' relative rank measures up to all locations ever ranked on the same system.
- Predicting walkability score for the entire city of St. John's, NL, with results verifiable via visual maps.

- Exploring, integrating and profiling user characteristic measures based on collected data.
- Building a pipeline to personalize predictive models based on users' profiles to achieve the ability to predict personalized walkability scores based on individual profiles.
- Scaling ALF-Score pipeline to utilize a much smaller fraction of user data towards much larger regions.
- Integration of transferability into ALF-Score so models trained on small cities can be used to predict walkability scores for larger cities with zero-user-input.
- Building predictive and personalized models for the city of Montréal, QC and generating walkability scores for the entire city of Montréal to show the power of transferability achieved in the methods, approaches and models using user data.
- Predicting walkability scores for the entire city of Kingston, ON using the zero-user-input approach.
- Predicting walkability scores for the entire city of Vancouver, BC using the zero-user-input approach.

Chapter 2

Background

This chapter presents background information on the technical aspects of this research. Because this work is highly interdisciplinary, it is vital to establish a common language to understand better the methods used in this work. Specifically, this section reviews graph theory and machine learning concepts using road networks and walkability research examples.

2.1 Graph Theory Fundamentals

Graph G is defined as $G = (V, E)$ where V is a list of vertices/nodes and E is a list of links/edges. Each vertex represents a point on the road, and each edge represents a road connection where two vertices are directly connected. The degree of a vertex is the number of edges that are incident to the vertex, and a path in a graph is a sequence of edges that joins a set of vertices. The shortest path is the path(s) between two vertices in a graph where the sum of the edge weights is minimized. Road network G is considered undirected, but this may be changed depending on the completeness of data tags. For example, one-way streets represent directed vertices, while two-way streets represent undirected vertices. Reduced graph G_r is defined such that $G_r = (V_r, E_r)$ where V_r is a subset of V but E_r will have less edges while containing some new links.

There are many ways to represent a graph data structure, such as an adjacency matrix and an adjacency list. An adjacency matrix, a 2-dimensional array of size $|V| \times |V|$, carries a simple implementation and has a faster data access of $O(1)$ time as well as when removing, editing and checking whether an edge exists between vertex i and j, i.e. $G[i][j] = 1$. However, it takes $O(|V|^2)$ space regardless of whether the network is sparse or otherwise, as there is always an entry dedicated to each vertex's association with all other vertices in the graph. An adjacency list is an array of size $|V|$ where the value of cell i represents a list of vertices adjacent to the i^{th} vertex. Adjacency list saves space for sparse networks taking $O(|V| + |E|)$. For a dense network, a worst-case scenario of $O(|V|^2)$ time can be expected, and queries such as finding if an edge between two vertices exists take $O(|V|)$ time. An adjacency list is used since road networks are generally sparse and computer memory is a concern.

2.1.1 Road Network

Road networks are interconnected roads designed to accommodate vehicles and pedestrian traffic. According to Urban Securipedia [121], road networks consist of "a system of interconnected paved carriageways designed to carry buses, cars and goods vehicles". Furthermore, road networks form the "most basic level of transport infrastructure within urban areas, and will link with all other areas, both within and

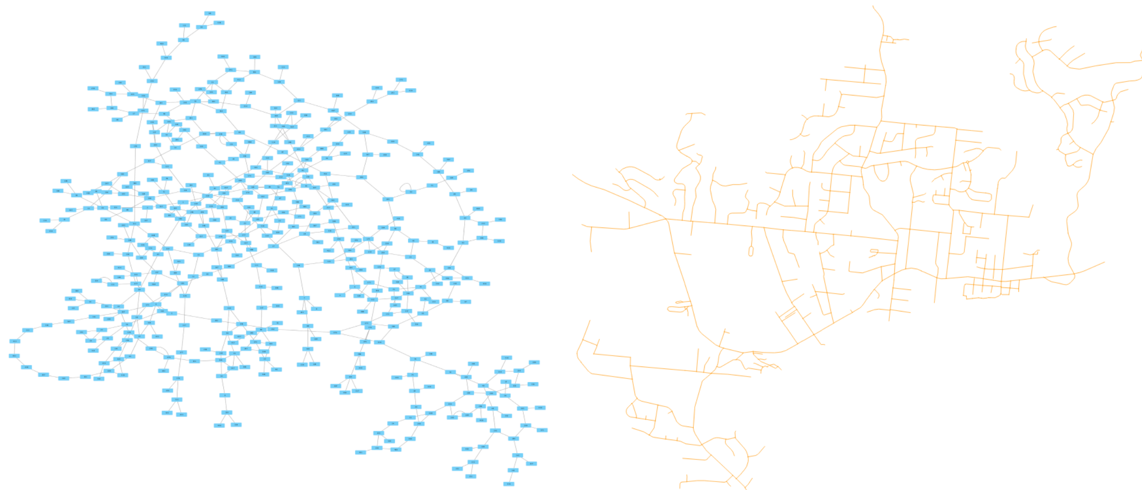

Figure 2.1: A neighbourhood in Victoria, BC. Left: Neighbourhood's road network visualized as a complex network by node connectivity (generated by Cytoscape [30]), Right: Neighbourhood's road network visualized as how they are physically structured within the city (generated by QGIS [98]).

beyond the boundaries of the urban area". Urban Securipedia divided road networks into eight parts: intersections (controlled or uncontrolled intersections, roundabouts), urban roads, rural roads, motorways, bicycle lanes, footpaths and pedestrian areas, pedestrian crossings, bridges and tunnels.

A Road network is a form of a complex network where nodes refer to physical geographical points and edges are the connections between two points. For example, if two points are directly and physically connected, there will be an edge between them. When multiple edges are connected, they form roads. Road maps are typically converted into a road network structure for computational purposes. Road networks can be represented as graphs.

2.1.2 Complex Networks

A complex network is a set of many connected nodes that interact in different ways [102] and are connected via links (edges). Some examples for complex networks are:

- Social networks [132]. For example [102]:
  - Friendship - where two people are connected if they are friends
  - Scientific - where two scientists are connected if they have been coauthors in any paper
  - Family - where two people are connected if they belong to the same close family
- Technological networks [130]. For example [102]:
  - Internet - where two computers are connected if they are in the same domain
  - WWW - where two web pages are connected if there is a link from one to the other
  - Words - where two words are connected if they are synonyms
- Biological networks [131]. For example [102]:
  - Proteic - where two proteins are connected if they participate in the same metabolic path
  - Genetic - where two genes are connected if one regulates the expression of the other
  - Ecologic - where species are connected if they have a predator-prey relationship

According to [31], complex networks are "graphs that depart substantially from regular or statistically regular graphs". Furthermore, while a good idea about "the local connectivity of a simple network" can be obtained by "considering only its vertices' degree (number of neighbours/connected vertices)", more "complex networks will demand the specification of many more additional respective properties or features".

2.1.3 Centrality

Centrality quantifies how important vertices (or edges) are in a network. Social network analysts, in particular, have expended considerable effort in studying centrality. There are many mathematical measures of vertex centrality that focus on different concepts and definitions of what it means to be central in a network [88].

A component of this research is the ability to find the importance of each node with respect to all other node in the network based on their connectivity and centrality measures. Road importance can represent its control over the connectivity of the entire network, defining each road's influence on traffic flow, drivability and walkability. In some centrality measures such as Degree centrality where $x_i = \sum_j A_{ij}$ [88], and PageRank where $x_i = \alpha \sum_j A_{ij} \frac{x_j}{\delta_j^+} + \beta$ [88], higher centrality values correspond to higher importance, whereas some other measures such as Closeness centrality where $x_i = \frac{1}{n} \sum d_{ij}$ [88], d_{ij} being the distance from vertex j to vertex i, smaller centrality values represent higher importance. Here, x_i is the centrality of vertex i, δ_j is vertex j's out degree and α and β are positive constants with α usually equal to 0.85 and β equal to 1, but can be fine-tuned as needed.

Betweenness centrality, $x_i = \sum_{st} \frac{n_{st}^i}{g_{st}}$ [88], with n_{st}^i representing if vertex i lies on the shortest path from s to t and g_{st} the total number of shortest paths from s to t, measures the extend to which a vertex lies on shortest paths between other vertices, quantifying the importance of the roads based on their control over the connectivity and flow. Vertices with high betweenness centrality may have a stronger influence within the network. They may control a higher traffic rate, and any disruption to these roads may cause some repercussions concerning traffic flow. Closeness centrality measures the mean distance from a vertex to all other vertices. PageRank, (which is a variation of Katz $x_i = \alpha \sum_j A_{ij}x_j + \beta$ [88]), uses the number of incoming edges to outgoing edges and the importance of each vertex pointing to and from a vertex to converge on a score. Another centrality measure, Hyperlink-Induced Topic Search (HITS), also known as hubs and authorities, is used in this research. HITS is generally used on directed graphs and measures each vertex based on two factors: its authority centrality x_i and its hub centrality y_i where $x_i = \alpha \sum_j A_{ij}y_j$ and $y_i = \beta \sum_j A_{ji}x_j$ [88] where x_j is vertex j centrality. A good hub points to many good authorities; a good authority is pointed to by many hubs.

Centrality Measure	***Computational Complexity***
Degree Centrality	$O(\|V\|)$
Eigenvector Centrality	$O(\|V\| + \|E\|)$
PageRank Centrality	$O(\|V\| + \|E\|)$
Katz Centrality	$O(\|V\| \times (\|V\| + \|E\|))$
Closeness Centrality	$O(\|V\| \times (\|V\| + \|E\|))$
Betweenness Centrality	$O(\|V\|^2 \log \|V\| + \|V\| \times \|E\|)$

Table 2.1: Centrality Computational Complexity

For centrality measures such as betweenness and closeness centralities, where they compute the shortest path for all nodes, the computational time gets more complex as the network increases in size. Considering the computational complexity of a famous betweenness centrality algorithm by Brandes [19] which is $O(|V| \times |E|)$ when applied to an unweighted graph and $O(|V|^2 \log |V| + |V| \times |E|)$ when applied to a weighted graph and the computational complexity of closeness centrality based on an algorithm by Sariyüce et al. [105] which is $O(|V| \times (|V| + |E|))$, given an extensive network, centrality computation is a lengthy and time-consuming task, such as the downtown Toronto road network with close to half a million nodes and almost 1 million edges, which required over five days to processes. (The computation was performed on a 2012 MacBook Pro with 8GB of RAM, i7 CPU with four cores, and an SSD drive.) Large graphs generally require extensive calculations, which come with expensive resource requirements, potentially causing additional costs and delays in processing and analysis. To address this challenge and make it more feasible to process larger networks, a reduction and reconstruction process [5] was developed that can significantly reduce the number of nodes in a network while preserving its core structure essential to compute its centrality.

2.1.4 Network Embedding

Network representation [9] learning, also known as network embedding, aims to generate numerical representations for nodes in a network to preserve its structures while allowing for great abstraction, which is one of the several ways used to preserve the road structure mainly by incorporating the resulting reduced network into a feature list. Essential use for network embedding is its excellent representation of road structure when used as a feature set in machine learning methods. At the same time, its dimensionality reduction gives the ability to significantly reduce networks' size while maintaining its core structure intact. Various methods allow embedding road networks into numerical representations while lowering the network dimension and preserving its structure¿ This research uses two specific methods for learning continuous feature representations of nodes in road networks:

node2vec [55] - based on a biased random walk procedure

struc2vec [101] - which uses a hierarchy to measure node similarity at different scales

2.2 Machine Learning

Machine learning (ML), a branch of Artificial Intelligence (AI), focuses on applications [62] that learn from experience and are capable of continuous learning and improving their prediction accuracy over time. Unlike traditional rule-based approaches, what differentiates machine learning from traditional approaches is how machine learning algorithms are trained and tested by going through various datasets that allow the algorithms to discover patterns and determine essential features to build models. There are various types of machine learning, such as supervised learning, semi-supervised learning, unsupervised learning and reinforcement learning. However, supervised learning and semi-supervised learning are among more popular approaches [57]. When it comes to machine learning, data plays an important role. The type of expected output and whether there is enough labelled data can help determine the appropriate methods. For example, the regression method is commonly used as a supervised learning approach to predict the numerical or continuous output. On the other hand, classification is typically used when the desired outcome is categorical or discrete. Semi-supervised learning is a great way to utilize labelled and unlabelled data by learning patterns in the feature set. Unsupervised learning is typically applied where there is little to no labelled data.

Linear Regression is a supervised machine learning algorithm where the predicted output is continuous and with a constant slope [85]. A random forest is a meta estimator that fits several classifying decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting [107]. A decision tree builds regression or classification models in the form of a tree structure, and it breaks down a dataset into smaller and smaller subsets. At the same time, an associated decision tree is incrementally developed while the final result is a tree with decision nodes and leaf nodes [104]. Gradient boosting is yet another machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees where it builds the model in a stage-wise fashion as other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function. [129].

2.2.1 Transfer Learning

Transfer learning is the process of re-utilizing the knowledge learned from other related tasks and has become very popular in recent years, especially in deep learning. In many machine learning approaches solving a single task at hand has been the main focus. Still, in more recent years, the development of approaches that allow for knowledge transfer has become a very popular focus [119]. As with most real-world problems, specifically in machine learning, collecting labelled data is a time-consuming, expensive [115] and difficult task. Transfer learning uses the knowledge learned from previous problems to solve new but related problems [143]. As a result of its approach, transfer learning can help reduce training time, resources, and the required labelled data [79] and improve overall accuracy. Weiss et al. [127] provide a much more formal definition of transfer learning as the following: "given a source domain $\mathcal{D}_{\mathrm{S}}$ with a corresponding source task $\mathcal{T}_{\mathcal{S}}$ and a target domain $\mathcal{D}_{\mathrm{T}}$ with a corresponding task $\mathcal{T}_{\mathcal{T}}$, transfer learning is the process of improving the target predictive function $f_T(.)$ by using the related information from $\mathcal{D}_{\mathrm{S}}$ and $\mathcal{T}_{\mathcal{S}}$, where $\mathcal{D}_{\mathrm{S}} \neq \mathcal{D}_{\mathrm{T}}$ or $\mathcal{T}_{\mathcal{S}} \neq \mathcal{T}_{\mathcal{T}}$".

The type of transfer learning used in this research is the transfer of model parameters. The general idea behind transfer learning is to apply a previously trained model on labelled data (in the case of supervised learning) to another similar task with little data available. Instead of starting from scratch, start with some existing knowledge. Transfer learning is typically used in computer vision. For example, the weights of a model trained to detect apples could be transferred for another task of detecting fruits. In this case, instead of training the new model to detect apples from scratch, the knowledge about detecting apples is transferred, and the algorithm now looks to learn how to detect other fruits. Transfer learning is a technique that requires significantly less data for training and will also speed up the training process [50].

There are a few approaches to transfer learning and to name a few, feature extraction, training a generalized model, and the use of pre-trained models are among these options. When it comes to feature extraction, determining the best representation for the problem at hand is a crucial task which if done correctly, can often lead to much better and more accurate results. Carefully selected features can often lead to a robust and well-generalized model applicable to various related problems. Another approach using transfer learning to solve a task with insufficient data is to train models designed for a similar task with an abundance of data. These models can then act as a starting point to address the original task. To highlight the difference with other approaches, to solve a given task A using this technique, training on a similar task B is performed. Once satisfied with the model, transfer and reuse of this knowledge in task A may begin. Furthermore, using already available pre-trained models is yet another common approach. There are countless pre-trained models available online that provide ready weights for many popular tasks such as classifying types of images, object detection and object tracking. It is important to highlight that this approach only requires access to a previously trained model and not the entire dataset. Goodfellow, in his book [53] further discussed two extreme forms of transfer learning, namely: 1) one-shot learning - which only one labelled example of the transfer task is given while, 2) zero-shot learning, which has no labelled example given.

Being able to generate reproducible and transferable predictive walkability models is an important component of which this research addresses in two ways: 1) by gaining the ability to utilize previously learned knowledge when directly generating walkability scores for new cities (zero-user-input approach), 2) by using this previously learned knowledge as a base to train new models (combined approach) which can lead to reduced training time, improved accuracy, reduced resource consumption, and reduction in the labels required for supervised learning. A well-generalized model will have the capability of transferring its knowledge to various cities never seen during its training to generate accurate walkability scores in a fraction of the time without the need for any new user input within the target city.

Transfer learning falls under-representation learning intending to use the same representation in various tasks. According to Ian Goodfellow [53], transfer learning can be viewed as of particular multi-task learning that typically revolves around supervised learning. Although, transfer learning can also be used to solve unsupervised learning tasks. Transfer learning aims to take advantage of previously trained models and extract knowledge useful in new tasks. However, transfer learning is also beneficial to generate predictions in another environment directly and for other tasks without any more learning needed [53]. Many surveys on transfer learning explore various parts of this domain, focusing on its potentials, advancements, and gaps, such as one by Pan et al. [92] published in 2010 and a more recent survey by Zhang et al. [140] published in 2019.

2.2.2 Machine learning formulation

Machine learning formulation of walkability in a network-based context pursues the following structure:

Problem: Training a machine-learned model that predicts a numeric value defining how walkable different points within the road network are. The feature set will need to include road parameters extracted from the road network. A continuous multilabel regression, which predicts walkability scores based on node characteristics.

Outcome: The ideal outcome is to avoid calculating walkability scores individually for every single point within the road network and have the ability to accurately predict point-specific scores from only a small sample of crowd-sourced labelled data while incorporating user opinion and having the ability to influence the outcome as user demographics change. The output of this machine learning pipeline would be trained models, and the output of these models is walkability score predictions for given nodes. Walkability score is defined as how walkable the user will find the selected point, an output normalized between 0 and 100.

Input: A feature set containing various features derived from road network data (centrality, road embedding), GIS data (POIs), as well as user demographics (user-defined and system-defined). Please refer to the Data section (2.3) for an in-depth description.

Output:

1. Walkability metric, a global scalar that is consistent across all road nodes that have user input. But it is also coherent across all road nodes with and without user input labels.
2. Walkability function $w : V \Longrightarrow R^+$
3. Performance metric of w: consistency with user rankings R.

   - Specifically, consider a user ranking $r \in R$ involving a set of 5 nodes $\{v_1, v_2, v_3, v_4, v_5\}$.

   - Ranking r gives each node v_i $(i = 1, 2, 3, 4, 5)$ a rank denoted o_i^r, (i.e. the "order" of node v_i as in user ranking.)

- Similarly, walkability w would also imply for v_i $(i = 1, 2, 3, 4, 5)$ a rank denoted o_i^w, (i.e. the "order" of node v_i as suggested by w.)

4. The loss function of w with respect to r, or "inconsistency" between r and w, can be defined as

$$L(r, w) = \sum_{i=1}^{5} |o_i^r - o_i^w|$$

- (Note there can be other ways to define such an inconsistency, e.g. l_2-norm or "out-of-order" counts.)

- The loss function with respect to R is aggregate of $L(r, w)$, say summation:

$$L(R, w) = \sum_{j=1}^{l} L(r_j, w)$$

- That is, the total amount of inconsistency of w as compared to all user-provided rankings.

(Note again that the aggregate function can take other forms, too, such as mean or multiplication.)

5. The goal is to find w that minimizes $L(R, w)$, the amount of inconsistency between w and R.

- A conversion from "relative" rankings to "absolute" scores among all user-provided data with as little discrepancy among R as possible is a crucial step)

Success and Failure Metrics: Accuracy, efficiency and consistency; neither would be good enough without the other. In terms of accuracy, the success metric in this component is based on how accurate the results are based on visual verification and cross-validation with user inputs. In terms of efficiency, if the total time required to produce predictions based on trained models would be less than the computation time using the traditional walkability measures, then the model would be efficient. When it comes to consistency, spatial consistency is measured through mapping and visual verification of projected scores. As for user ranking consistency, a metric that measures consistency or lack thereof (inconsistency) has been defined and developed, which is a crucial step to determine if the user-driven ground truth remains in the same order as users ranked it after being processed through the relative to absolute conversion (GLEPO algorithm), to ensure the full utilization of user-labelled data and least amount of introduced bias.

Usage: The output of the model, walkability scores, is robust and much more refined with a much higher spatial resolution that can be used to create an interactive interface that would allow users to visually view walkability scores for various regions at the desired resolution. Furthermore, due to the predictive nature of the models, there would be no need to calculate the walkability scores for an entire network. Models trained on only a small subset, as needed, can produce predictive scores for the entire network in a fraction of the time.

Alternatives: If machine learning were not used, the approach to this issue would be very different and tedious, involving many manual and redundant tasks. For example, walkability scores would have significantly lower spatial resolution and little reliance on user opinion. It would follow a rule-based approach producing generic scores, making it very difficult to incorporate personalization and requiring complete recalculation for every cycle generation.

Using machine learning allows for incorporating various vital features such as road network data and moving to an efficient approach generating point-based walkability scores with much higher refined distribution when compared to the traditional methods such as Can-ALE, where it provides only area-based (i.e. DA-level) scores. Furthermore, when done correctly, using different machine learning algorithms allows the machine to decide what features are more relevant to enhance the selection process further and produce more accurate and highly efficient results. Without the machine learning pipeline, this process would be a time-consuming task that would not be as accurate or efficient.

When it comes to utilizing machine learning, there are many considerations to ensure the accuracy and efficiency of the models. A generalized approach to train models over a set of data while testing the models over a different and unseen dataset can produce more accurate prediction results. However, there are various challenges in keeping a machine learning model as generalized as possible. For example, lack and bias in the data, incorrectly labelled data, etc., can lead to modelling errors. Overfitting and underfitting, are also two widespread modelling errors that occur when the model's primary function is too closely fit the data (too well-learned) for the case of overfitting, or in the case of underfitting, when the function does not capture the prominent patterns in the data (not enough learning). Overfitting results in high accuracy when applied to previously seen data but significantly lower accuracy when applied to unseen data. Underfitting results in unpredictable output. In both cases, low generalization of the models leads to unreliable predictions.

There are many ways to address these challenges, one of which is data split. The core concept in data split is to split the labelled data into two datasets: 1) training set and 2) testing set. Typically a 70-30 or an 80-20 percent split is commonly seen. However, this research has experimented with various other variations, such as a 90-10 and a 60-40 percent split. In the case of an 80-20 percent split, 80 percent of the labelled data is assigned to the training set while the labels from the remaining 20 percent are extracted to form a testing set where the trained model will perform prediction on. The prediction results are then compared to the actual results to determine a baseline on the accuracy when compared to the accuracy of the same model performing on the training set.

The dataset used to train models is likely a relatively small set. Therefore, by taking an even smaller training set (due to cluster profiling), there is a risk of losing essential patterns and trends within the dataset, increasing the error rate in the models. K-fold cross-validation is yet another statistical method commonly used to test and estimate the accuracy of machine learning models on new data, [65] which is typically done by splitting the dataset into k sets. One set is selected as the test data, and the remaining sets are chosen as the training data. Once the model is built, it is evaluated on the test dataset. The procedure repeats for every set in k to ensure the entire dataset is utilized towards building and evaluating the models.

There are many methods to determine the accuracy of machine learning models when it comes to predicting the desired output. Such methods include 1) Mean Absolute Error (MAE) with $MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$, where x_i is the actual value, y_i is the prediction, and n is the total number of data points, 2) Mean Squared Error (MSE) with $MSE = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2$, 3) Coefficient of Determination, with $R^2 = 1 - \frac{RSS}{TSS}$ where sum of squares of residuals or $RSS = \sum_{i=1}^{n} (y_i - x_i)^2$, total sum of squares or $TSS = \sum_{i=1}^{n} (x_i - \bar{x})^2$, and $\bar{x}$ is the mean value of actual labels, and 4) Root Mean Squared Error (RMSE), with $RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}}$. In this research, both MAE as well as RMSE is used.

Moreover, various machine learning techniques have been applied and compared based on multiple feature set combinations to find the most suitable technique and feature set combination that results in the best prediction accuracy predicting walkability scores. The following methods were used in both supervised and semi-supervised environments to train shallow and deep models:

- Random Forest Regressor
- Linear Regression
- Decision Tree Regressor
- Gradient Boosting
- Polynomial Features (Non-Linear approach)
- Lasso CV
- Label Propagation
- Label Spreading
- Logistic Regression
- K-Means Clustering
- PCA
- t-SNE
- Multi-Layer Perceptrons (MLP)

2.3 Scalability

In 1990, a paper published by Mark D Hill [59] suggested there is no "generally-accepted definition" for scalability, likely due to the misuse of the term at the time, mainly for its marketing potentials. He then challenged the technical community to define scalability or stop using it to describe systems. André B Bondi, in a paper [16] published in 2000, suggested that scalability is "a desired attribute of a network, system or process". Bondi argued that systems with poor scalability might engage in repeated wasteful activities, consuming and wasting processing time and resources. Ensuring the ALF-Score pipeline does not engage in repeated wasteful activities is one of the sub-objectives of this research, which is particularly important since road networks can vary in size, with some cities being very small (e.g. with a population of a few hundred). In contrast, some other cities could be huge and dense (e.g. Tokyo, Japan, with over 37 million people in just one city). The Table 2.2 shows a list of various cities used in this research alongside their network size, number of POIs, population and total land area size. Processing data from St. John's, NL, as opposed to data from Toronto, ON, will have a significantly different resource requirement and time consumption due to the change in the size of the city leading to an extended set of complexities. If the algorithms are not optimized, this difference in requirements may lead to the infeasibility of the research. Various cities, including those mentioned in the Table 2.2, have been experimented with within this research. However, only St. John's NL, Kingston ON, Vancouver BC, and Montréal QC are highlighted.

City	# of Nodes	# of Edges	# of POIs	Population	Total Land Area
Victoria, BC	6,770	8,593	3,318	85,792	19.47 km^2
Kingston Metro, ON	3,427	4,769	813	161,175	1,906.82 km^2
St. John's Metro, NL	5,364	6,851	592	205,955	804.63 km^2
Vancouver Metro, BC	45,125	60,299	13,321	2,463,431	2,878.52 km^2
Montréal Metro, QC	76,663	114,414	10,045	4,247,000	4,604.26 km^2
Toronto Metro, ON	479,520	Over a million	23,930	6,417,516	5,905.71 km^2

Table 2.2: List of road networks for various cities with their network and POI sizes that have been experimented with within this research. Nodes and edges are extracted from road networks. Population density and the total land area information are excerpted from Wikipedia.

2.4 Data

There are several data types required for this work. Individual user profile data, road network data (road structure, edge list, node/edge centrality, node embedding), and POI. Since road networks are accessible through many different sources such as OpenStreetMaps (OSM) [91] and Statistics Canada [113], different data formats are also available for processing. Among these formats, many files are binary or text-based. User readability and processing speed are two main factors when selecting the appropriate format. Binary-based files tend to be faster when processed, whereas text-based files are more user-readable through conventional software. However, the main concern when selecting the most appropriate data source and format in this research is data comprehensiveness and accuracy.

The road network data were extracted from both sources, OSM and Statistics Canada (referencing the Census year 2016). OSM (typically with `.osm` as its file format) follows an `XML` scheme containing `nodes` and `ways` elements. The `nodes` elements define points in space, consisting of their unique id, latitude and longitude, with an optional third dimension, altitude, which could be included depending on the completeness of the data. The `ways` elements define linear features and area boundaries as an ordered list of `nodes` elements' unique ids with specific property tags such as `highway` [90], `barrier`, `amenity`, `name`, `oneway` or `landuse`. Statistics Canada however, depending on the year the road network data was compiled, may provide three data formats: 1) ArcGIS `.shp`, 2) Geography Markup Language `.gml`, and 3) MapInfo `.tab`. For this research, Esri vector shapefile is utilized, which stores the location, shape, and attributes of geographic features and several other fields. For example, `TYPE` which defines the type of the road such as `CRES` referring to crescent, or `DIRECTION` defining the direction of the road, where applicable. Many other variables such as Census subdivision types, Census metropolitan area, or census agglomeration are also available to provide an in-depth description of each road. There are a few differences between the two data sources: formatting, information included, completeness, comprehensiveness, recency, or reliability. However, one of the main differentiators between the two data sources is that OpenStreetMap utilizes crowdsourcing, which considers each person living in a community an expert of their local surroundings and collects information from a large group of individuals from various communities. The road data was used to build various datasets derived directly or indirectly from the road network structure. For instance, a comprehensive node list and an edge list were derived from the graph used to generate a complete set of centrality measures and road embedding representations. Road importance and graph reduction have been explored and explained in [5]. Here is an example of variables available through extracted road networks:

```
NGD_UID, TYPE, DIR, AFL_VAL, ATL_VAL, CSDUID_L, CSDTYPE_L,
RANK, CLASS, ...
```

The POI data for this research was exclusively extracted from OSM. There are 3 main elements defined in OSM generated POI files: `nodes`, `ways` and `relations`. Nodes define a point on the map, each with a specific key assigned to it. From various existing keys, the "amenity" key [90] maintains the most relevant set of POIs. There are numerous categories such as Sustenance, Education, Transportation, Financial, Healthcare, Entertainment, Arts & Culture, and Others. However, since there are only eight main categories deemed relevant, it was decided that each of the eight relevant keys would be used as a feature on its own. Here is an example of a single entry of a feature set containing a node id, latitude and longitude and various other POI features:

```
777 -0.039065 0.836415 -0.598678 -0.346100 -0.001968 0.010741
6,-52.7321222911614,47.5687221942087,0,0,0,0,1,0,0,1,4,5,
0,0,0,0,0,0,1,0,0,1,0,0,1,1,3,0,0,1,0,2,1,1,0,3,10,0,0,0,0,0,0,0,0, ...
```

Additionally, a crucial component of this research revolves around crowd-sourced user data, which contains specific user labels (opinion) and demographics. A web interface was built to enable data collection for the required information from volunteer participants to achieve this. Two data categories were collected: 1) user-defined features containing seven variables, and 2) system-defined features containing six variables. User-defined features consist of walkability ranking, preferred walkable distance, age group, gender, if the user lives alone, if the user has children, and occupation. The system-defined features are public IP address, public port number, user device language, user browser type, user operating system and time of submission.

The web interface displays an interactive map with 5 locations marked as $\{A, B, C, D, E\}$. Every time the web page is reloaded, 5 locations are randomly selected from a pool of points on the road network within a specific geographic area. Before each data submission, users can adjust the ranking of the 5 locations by reordering them based on the user's personal perceived relative walkability. This approach allows users to use their knowledge of the city while exploring a diverse set of potential locations in the region. The relative rankings are then submitted as groups of 5 locations to the server for downstream processing. Note that given the random nature of the web tool, the group of five locations provided by the web tool can overlap across different users and even various submissions of the same user but not within the same submission. These overlaps provide common references among submission groups and the needed data that can help distinguish between users' perceptions and preferences. The challenge here is to balance out the data from users to yield a walkability measure that can produce specific results for each user cluster. Moreover, the web interface allows volunteer participants to zoom in and out, pan around the map and view various POIs as automatically provided by Google Maps API. Here are three examples of variables available through the crowd-sourcing platform:

— Predefined data: data extracted from a node list on the server such as Node ID, Longitude, Latitude:

```
850, -52.7098539006027, 47.6177887084157
```

— User-defined: Data entered by users such as Walkability Order (relative ranking), Preferred Walkable Distance, Age Group, Gender, If User Lives Alone, If User Have Children, Occupation:

`3, 1000, 50-59, Man, No, Yes, Executive`

— User extracted: data from a user that was not explicitly entered such as Public IP Address, Public Port Number, User Device Language, User Browser Type, User OS, Time of Submission:

```
192.168.0.1, 12345, en-CA-en-US, Mozilla/5.0 (Windows NT 10.0; ),
AppleWebKit/Safari/537.36, 2020-05-26 22:27:34
```

Here is an example of a single road embedding entry generated by node2vec:

```
777 -0.039065 0.836415 -0.598678 -0.346100 -0.001968 0.010741
-0.150576 0.114083 -0.704087 -0.285862 0.447592 -0.345718
0.405268 0.116926 0.151894 0.051038 0.203209 ...
```

Here is an example of a single node centrality entry generated for "AverageShortestPathLength", "BetweennessCentrality", "ClosenessCentrality", "ClusteringCoefficient", "Degree", "Eccentricity", "IsSingleNode", "name", "NeighborhoodConnectivity", "NumberOfDirectedEdges", "NumberOfUndirectedEdges", "PartnerOfMultiEdgedNodePairs", "Radiality", "SelfLoops", "Stress", "TopologicalCoefficient":

```
‘‘42.90749064’’, ‘‘3.7453E-4’’, ‘‘0.02330595’’, ‘‘0.0’’, ‘‘2’’, ‘‘96’’,
‘‘false’’, ‘‘1’’, ‘‘2.0’’, ‘‘2’’, ‘‘0’’, ‘‘0’’, ‘‘0.66740087’’, ‘‘0’’,
‘‘830096’’, ‘‘0.5’’
```

In total, 680 features are included in this research. The following are the input parameters:

1. The primary road network data $G = (V, E)$ covers St. John's, NL and is undirected and unweighted (but weights are calculated when needed). $|V| = 5{,}364$ nodes and $|E| = 6{,}851$ edges (refer to Table 2.2 for the size of various other road networks used in this research). However, this reflects the entire network. This network has eight components with a small subset of nodes and edges (23 nodes, 16 edges) belonging to 7 small components, whereas the giant component maintains 5,341 nodes and 6,835 edges. Only the giant component is used; therefore, $|V| = 5{,}341$ nodes and $|E| = 6{,}835$ edges. Figure 4.2 also shows the region covered by the road network data. Road node features, a mix of numeric, ordinal, and categorical data. First, *POIs* with 530 features. These features are derived from 8 separate OSM categories with 53 total POI subkeys. Every single key contributes to 10 features. Each feature reflects on a specific geometrical/distance range to show whether that particular key falls within the associated distance of any point within the network. The geometrical distance ranges from 200m to 2,000m with 200m increments (based on popular user walkable distance entries). There are 530 POI features for every single node on the road network. *Complex Networks* which contributes ten features, namely: Betweenness centrality, Closeness centrality, clustering coefficient, degree, eccentricity, neighbourhood connectivity, stress, topological coefficient, average shortest path length, and radiality. Finally, *Road Embedding* which contributes 128 features based on 128 reduced dimensions.

2. There are 13 variables collected from users (such as location labels, user-defined

and system-defined variables), and 11 are used as user profiling features. Figure 2.2 shows a representation of location data collected from the participating users in St. John's, NL. Purple points represent unique nodes, while red, yellow and green points represent the low, medium and high frequency of location overlaps among user submissions. In various experiments, user data from approximately 80 unique users was used with $|S| = 409$, where S represents all user submissions. Each S_j represents a user submission where $(j = 1, 2, \ldots, |S|)$. Each S_j user submission consists of 5 road nodes $\langle v_1, v_2, v_3, v_4, v_5 \rangle$. Groups of 5 node rankings per submission were decided to be used. With few nodes (i.e. 2 locations), there is not enough data to capture variation and relevancy in user rankings. With too many nodes (i.e. n > 10), posing cognitive challenges is a risk when users rank too many locations. In total, 1,050 user-submitted ranks (852 unique locations) from the city of St. John's, NL and 995 user-submitted ranks (824 unique locations) from the city of Montréal, QC from around 80 participants were included. The outcome of this research is influenced by user representation. Therefore, it is essential to highlight that due to limited access to volunteers, the participant list comprises people who are educated and financially secure and does not include vulnerable or other minority populations. When possible, the inclusion of more diverse volunteers is highly recommended to reduce bias and improve user representation and personalization. "Anchor" [1] locations appearing in two or more submissions that connect/associate these submissions together are used to establish a global view

[1] ALF-Score uses the presence of these "anchor" points to create a unified list (using GLEPO - Generalized Linear Extension of Partial Orders [6]) that is used to build generalized walkability

of user opinion. Each road node v_i $(i = 1, 2, \ldots, 5)$ has a unique relative ranking r_{v_i} between $\{1, 2, \ldots, 5\}$. Order of r_{v_i}, the relative ranking, is of utmost importance as it defines how users perceive the walkability of each v_i with respect to one another. Virtual links [6] are also utilized in this pipeline.

2.5 Summary

This section reviewed graph theory and machine learning, focusing on road networks and transfer learning, followed by the required data and the methods used to collect them. The goal of this section is to provide the fundamental knowledge needed to understand the underlying structure of this research better and to be able to set up and recreate the experiments performed in the later chapters of this research while exploring some of the introductory yet important and relevant concepts. Understanding the base anatomy of referenced fields such as complex networks, graph theory, and machine learning is a crucial step to accurately and efficiently process each step in this research. Furthermore, since machine learning has been a core component to build various predictive models throughout this research, the data structure and feature sets are also equally important and carefully defined in this section. Additionally, since transfer learning is one of the finalizing components of this research and relies on pre-trained models, it is important to have a solid foundation on how transfer learning works, why it is important in this research and what it requires to function.

scores. On the one hand, these potential conflicts are sources of inconsistency. On the other hand, their existence means there are individual differences between various users that should be accounted for when generating walkability measures.

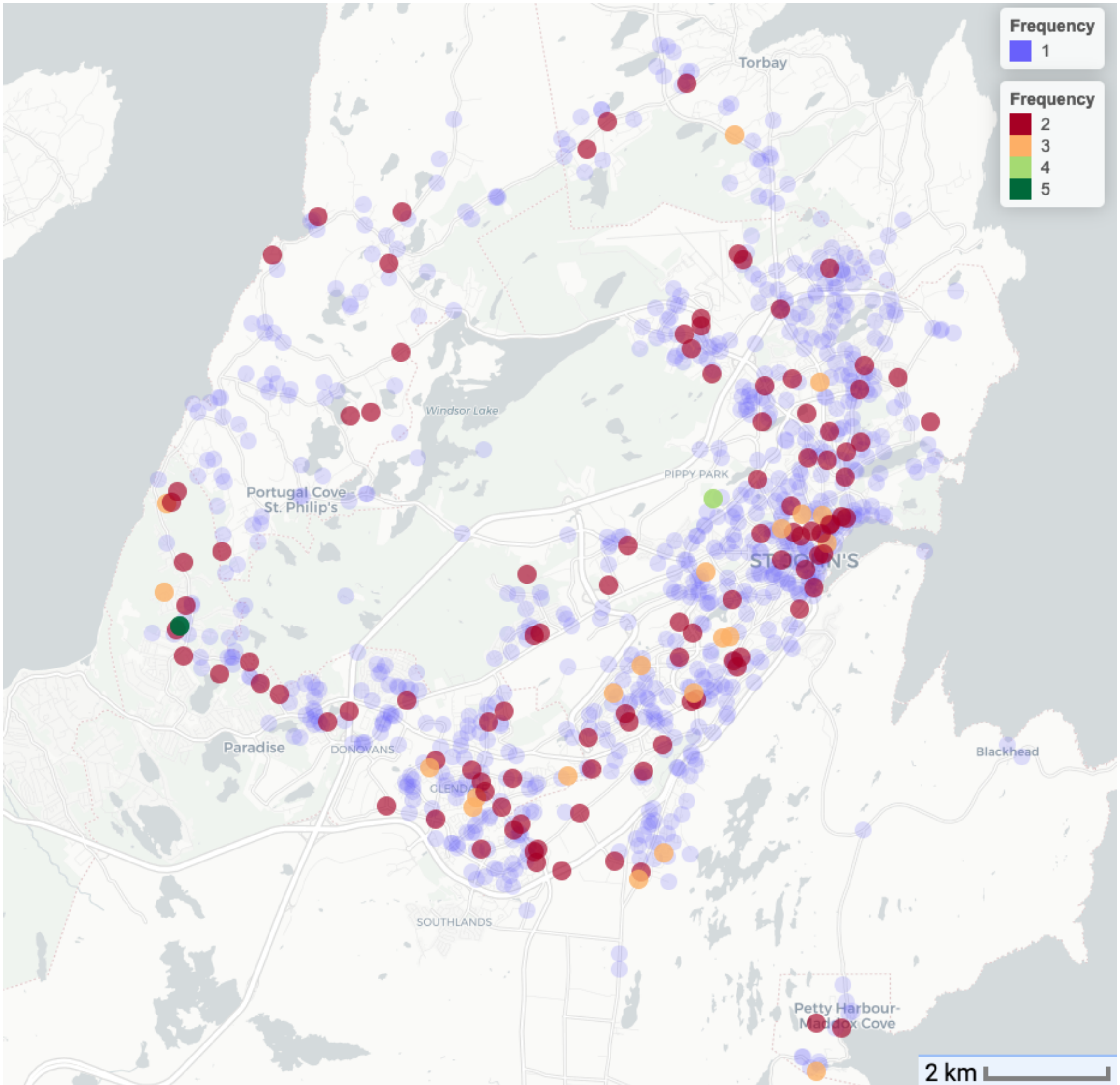

Figure 2.2: There are 1050 user-submitted locations with 852 unique locations covering most of the road network in St. John's, NL. This figure shows a large portion of the crowd-sourced data collected through the web-tool platform. Purple points represent unique nodes, whereas red, yellow and green points represent the low, medium and high frequency of user opinion overlaps respectively across user-submitted data. Generated through RStudio [103] Version 1.2 from rstudio.com.

Chapter 3

Graph Reduction and Reconstruction

This chapter involves an in-depth examination of road network structures, graph reduction and reconstruction methods for road networks. This work was published and presented at the 2020 International Conference on Computing, Networking and Communications [5]. The published paper is slightly revised for flow in this dissertation document, and the introduction is shortened.

Over a decade ago, a significant issue with computing measures such as walkability was the lack of data. As we walk through the age of data explosion and exploration, lack of data, a long-standing matter, is no longer an issue. Fortunately, road network data is abundantly available today from various sources such as Statistics Canada [113] and OpenStreetMap [91] making analytical studies on road networks much more accessible. However, as data expands, analyzing larger regions requires much more processing power and computational time. Especially when working with networked data, as the availability of data increases, so does the connectivity and complexity of the network.

However, since most popular and existing algorithms to measure importance are already very efficient yet too complex to process large networks, researchers are yearning for alternative methods to decrease the long computational time required to analyze large networks. Specifically, when analyzing road networks, although processing small neighbourhoods may be accomplished quickly, when processing larger regions such as a city, province/state or even an entire country, the complexity of the network requires much more than a day-to-day personal computer and much longer than mere hours to process the network. For example, when considering the computational time required to calculate the betweenness centrality, with the complexity of $O(|V|^2 \log |V| + |V| * |E|)$, for a road network covering downtown Toronto which contains 479,520 nodes and close to 1 million edges, on a 2012 MacBook Pro with 8GB of RAM, an SSD drive and an Intel Core i7 quad-core processor clocking at 2.7GHz, an approximate five days (or 121 hours) of processing was needed to complete the task. This research approaches this problem with a simple yet powerful solution: a reduction-reconstruction hybrid system.

This work incorporates graph reduction and centrality interpolation while utilizing some already-efficient complex networks centrality algorithms to produce ready-to-analyze road centrality scores for the entire given dataset while reducing the required computational time compared to the conventional algorithms that do not use reduction. Furthermore, the produced road scores can be applied to non-network characteristics such as amenities and POIs, elevation, road type, road condition and road structure to have accurate walkability scores.

To produce accurate road centrality scores based on their connectivity and to enhance computational time for road network analysis and reduce network's complexity, the network is reduced (Section 3.1) while preserving its original structure. The reduced graph's centrality scores are then calculated and used to reconstruct the original network using interpolation (Section 3.1). During the reduction step, the number of nodes in the graph decreases, which helps reduce the complexity of the road network and the computational time when measuring centrality. This reduction is particularly important when using centrality algorithms with higher computational complexity, such as those calculating the shortest paths [88] between all nodes in the network, which will require longer computational time. As this graph reduction algorithm results in a smaller network by removing many "less-important" nodes from the original graph, centrality scores computed for this reduced graph do not cover the entire original road network, leaving some informational gaps. To address this shortcoming, linear interpolation techniques are used to fill in the blanks. The final road scores from these techniques will cover all the nodes from the entire road network, yet the computational time is noticeably reduced.

3.1 Graph Reduction

When working with reduction, a balance between properties to retain and those to reduce needs to be maintained to keep the network's structure intact and avoid loss of crucial information, which could produce unusable results. Road networks generally contain multiple points within each road. These points have the following degree properties:

1. degree of 1: representing the beginning or the end of the road,
2. degree of 2[1]: representing points on the road that are only connected to two other points on the same road,
3. degree of 3 or more: representing points where the road intersects with other roads.

Nodes with a degree of 1 and those with a degree of 3 or more hold the key to preserving the structure of the network. Connected nodes with a degree of 2 that are within the same road, in most cases, have a closely associated importance which can be interpolated (Section 3.1) from nearby nodes. This definition of properties is used to choose what nodes to remove and what nodes to retain to ensure the removal does not impact the network's core structure.

In Algorithm 1, *Graph reduction*, vertices v with a degree of 2 are split into two path searches beginning from each of v's two degrees: v_0 and v_1. Each path includes continuous occurrences of immediate vertices with a degree of 2. Once a vertex with a degree other than two is reached, it will be selected as the *head* (path with v_0) or *tail* (path with v_1) vertex, and the search stops. All nodes within these paths are removed, and an edge between *head* and *tail* is added to represent the reduced path.

[1]if the road ends to or begins from another road, that point will be shared by the secondary connected `ways` element.

Algorithm 1 Graph Reduction

Input: G

Output: G_r

```
for all vertices v in G do
    if v not reduced then
        if v has degree of 1 then
            tail = v_0 {in this case v's only degree}
            if v_0 not reduced then
                [tail, weight] = EndFinder(G, v_0, v)
                addEdge (v, tail, weight) to G_r
        else if v has degree of 2 then
            mark v as reduced
            head = v0
            tail = v1
            if v0 not reduced then
                [head, head_weight] = EndFinder(G, v0, v)
            if v1 not reduced then
                [tail, tail_weight] = EndFinder(G, v1, v)
            weight = head_weight + tail_weight
            addEdge (head, tail, weight) to G2
        else
            for all neighbours n of v do
                if n not reduced AND n's degree != 2 then
                    addEdge (v, n, 1) to G2
```

Algorithm 2, *EndFinder*, a variation of the depth-first search algorithm, takes in graph G, *start* and *root* vertices, and finds a path, if any, where all its vertices have a degree of 2, and returns its end-point vertex. To verify that the reduction algorithm maintains the original structure of the road network, small real-world road networks were collected (such as downtown St. John's, St. John's east end and St. John's west end), processed and reduced. The original and reduced networks were physically visualized on the map, for example, Fig. 3.1, to visually inspect the networks' structure. To further verify the accuracy of the reduction algorithms, numerous artificial graphs were created, for example, Fig. 3.2, to consider different scenarios and special cases. Each artificial graph is carefully designed and measured to act as a benchmark and compare its non-reduced and reduced results.

Algorithm 2 End Finder

Input: G, *start*, *root*

Output: [end vertex where path reduction ends, weight]

```
add start to stack
mark root as visited
weight = 0
while stack is not empty do
    vertex = stack.pop
    increase weight by 1
    if vertex is not visited then
        if vertex degree != 2 then
            return [vertex as end point, weight]
        mark vertex visited & reduced
        update stack with vertex degrees that are not visited
```

Once reduction is complete, G_r will contain a similar number of roads but fewer vertices. Any centrality calculation performed on the reduced graph G_r should be noticeably faster than those performed on the original graph G. The proposed reduction method, on average, reduces the number of nodes by 77%, and although the

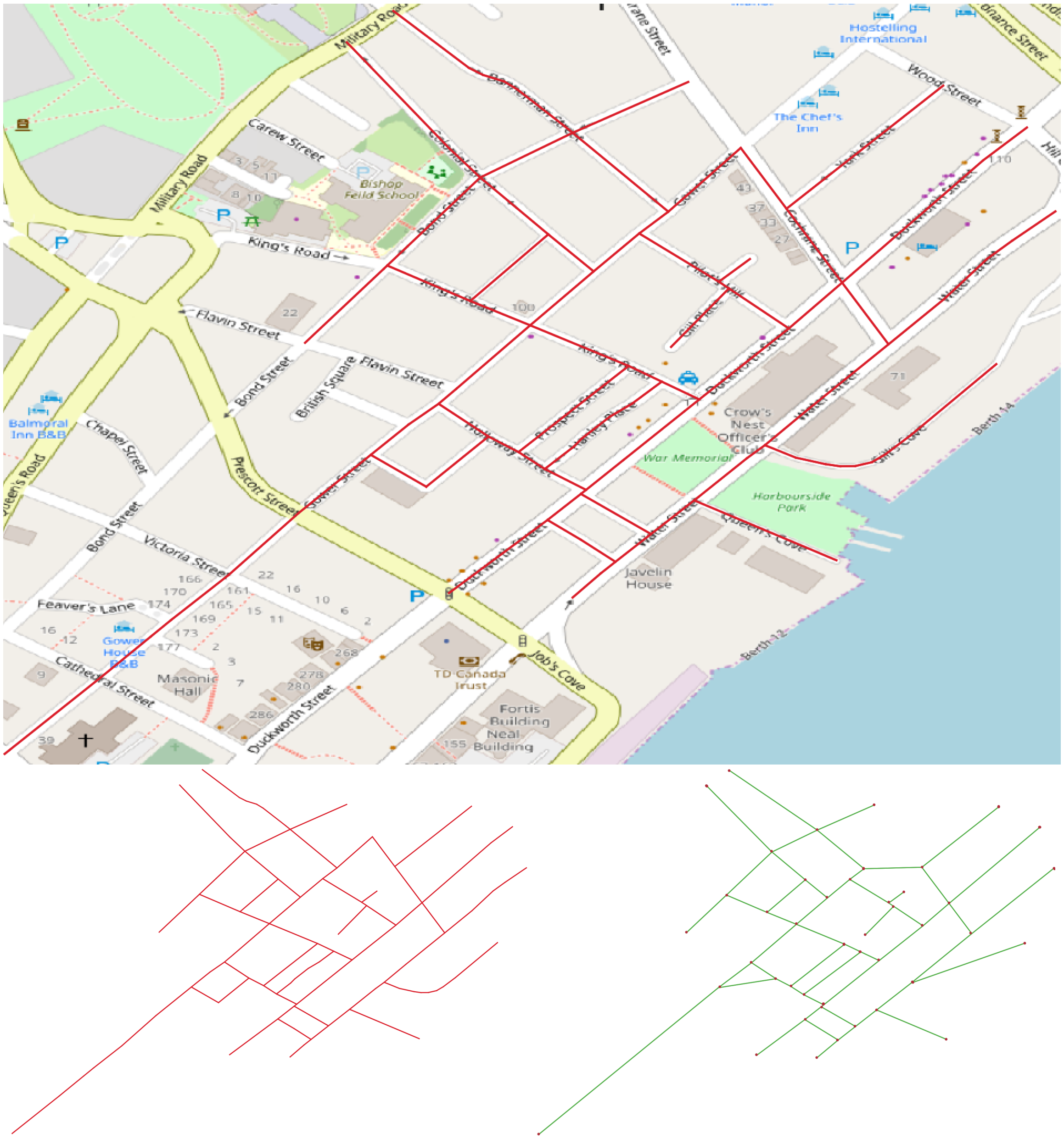

Figure 3.1: Original road network (left) vs. reduced graph (right). The reduced graph still preserves the original graph's structure, including all roads; however, there is a loss of information such as road curvature refinement caused by removing nodes from the original graph. This figure covers a small section of downtown St. John's with 30 roads and 85 nodes. The reduced network, while maintaining 30 roads, contains only 44 nodes. The top image represents the original graph over a standard OSM map layer. Generated by QGIS [98].

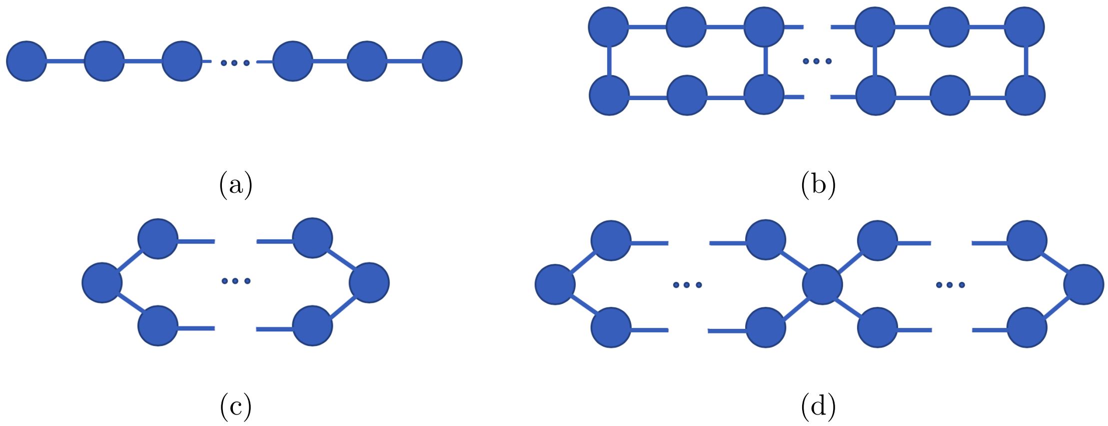

Figure 3.2: Sample artificial graphs. 3.2a: a single acyclic path with 1000 nodes. 3.2b: two paths each with 501 nodes, a total of 1002 nodes. Every second node of the two paths is linked. 3.2c: a single cyclic path with 1000 nodes. 3.2d: two cyclic paths that share a single node. Each cycle has 500 nodes with a total of 999 nodes. The Figure is drawn by the author.

structure of the graph is kept mostly intact (Fig. 3.1), there is a significant loss of vertex data. Due to this loss of vertex data, reduced vertices should be placed back after the reduction is complete and have their centrality scores approximated to reconstruct a complete graph. To take advantage of the reduced complexity and yet produce centrality scores for a complete graph, an interpolation method is used to reconstruct the reduced graph.

3.2 Graph Reconstruction

Graph reconstruction is a way of constructing new data points and reconstructing removed or missing data points, within the range of a set of known data points using interpolation. There are various types of interpolation such as polynomial interpolation [13] and spline interpolation [33, 106]. Linear interpolation [15] is used in this work to simplify the computation yet reflect the weight geographical distance carries over centrality computations. Given two coordinates (χ_1, μ_1) and (χ_2, μ_2), the μ interpolation for some point χ is defined as $\mu = \mu_1 + (\chi - \chi_1)\frac{\mu_2 - \mu_1}{\chi_2 - \chi_1}$. To interpolate centrality χ_i where i is a reduced vertex, given two of its closest "un-reduced" neighbors lying on the same road as i, on each side, the following is used:

$$\chi_i = \frac{\delta_{(i,n_1)}}{\delta_{total}} \times \zeta_{n_1} + \frac{\delta_{(i,n_2)}}{\delta_{total}} \times \zeta_{n_2}$$

where $\delta_{(i,n_1)}$ is the distance between nodes i and n_1, $\delta_{(i,n_2)}$ is the distance between nodes i and n_2, and δ_{total} is the total distance between n_1 and n_2, ζ_{n_1} is n_1's centrality and ζ_{n_2} is n_2's centrality. The distance between i and each end-point n_1 and n_2 determines the nodes' influence on i's interpolated centrality. The closer each node is to i the more influence it will have in interpolating i's centrality. There are 3 steps to measure the distance used in this interpolation:

1. computing the physical distance of each neighbouring vertices within the same road during the data processing step (graph construction),
2. calculating the total distance between the head vertex and the tail vertex as a path going through all points connecting them, during the reduction step,
3. computing individual distances for specific vertices on the same path during interpolation step.

3.3 Results

Table. 3.1 includes a small sample of datasets used for testing and verification. The algorithms can reduce the number of nodes, on average, by 77% while preserving the original network's structure, and the application of most centrality measures such as betweenness centrality and closeness centrality performed over the reduced network is completed considerably faster. Additionally, the interpolated centrality scores showed 89% accuracy on average. A few select test networks were chosen to verify the accuracy of the interpolated values. They had various centrality measures computed for both their original and reduced forms and compared the final results. The entire graph reduction and reconstruction process take $O(n^2)$.

Road network	***Number of roads***	***Number of nodes***	***Number of nodes after reduction***	***reduction***
Artificial-A	n/a	1002	498	50.3%
Artificial-B	n/a	999	1	99.9%
Artificial-C	n/a	1000	1	99.9%
Artificial-D	n/a	1000	2	99.8%
Partial Toronto	32,450	142,267	51,123	64%
St. John's metro	11,016	105,475	14,034	86.7
City of St. John's	3,442	21,330	4434	79.2%

Table 3.1: Reduction results

Table 3.2 shows the betweenness and closeness centralities applied to three different road networks. As evident, the computational time for centrality measurement has been reduced significantly using the reduction and reconstruction techniques. As

Road network	***Betweenness on original***	***Betweenness on reduced***	***Closeness original***	***Closeness* reduced**
Partial Toronto	121:00 hrs	6:08 hrs	32:54 hrs	1:52 hrs
St. John's metro	41:00 hrs	2:09 hrs	9:55 hrs	0:35 mins
St. John's City	01:22 hrs	00:04 mins	00:26 mins	00:01 min

Table 3.2: Reduction Computational time

linear interpolation methods were used, interpolating centrality scores for the reduced vertices is done very quickly and in linear time and the added time due to the interpolation process is almost negligible. Fig. 3.3 shows two heat maps generated for St. John's downtown road network. The heat-map on the left is generated from the original network (full coverage). In contrast, the heat-map on the right is generated from the reconstructed version of the reduced graph (with significantly smaller coverage) by applying interpolation to measure the missing centrality scores. The figure shows that the interpolated results are very similar to the original results.

Fig. 3.4 represents four different centrality measures computed for the city of St. John's. Betweenness centrality scores appear to be the closest to potential walkability scores, which are measured based on locally observed data. This similarity is likely due to how Betweenness centrality measures importance by highlighting each road's role in allowing traffic to pass from one part of the road network to another. Furthermore, in Fig. 3.5 bottom left plot, which represents centrality correlation for a small road network, as expected, there is a constant and noticeable positive correlation between most centrality measures. However, in Fig. 3.5 bottom right plot, which represents the centrality correlation of a larger road network, little to no correlation between these centrality measures is observed. This finding holds for many different sample

Figure 3.3: Comparison between the original graph (top left), the reduced graph (top right), and the complete interpolated centrality after reduction and reconstruction (bottom). Heat maps represent road scores for downtown St. John's and are closely similar, albeit one generated through reduction and reconstruction. (Generated by QGIS [98])

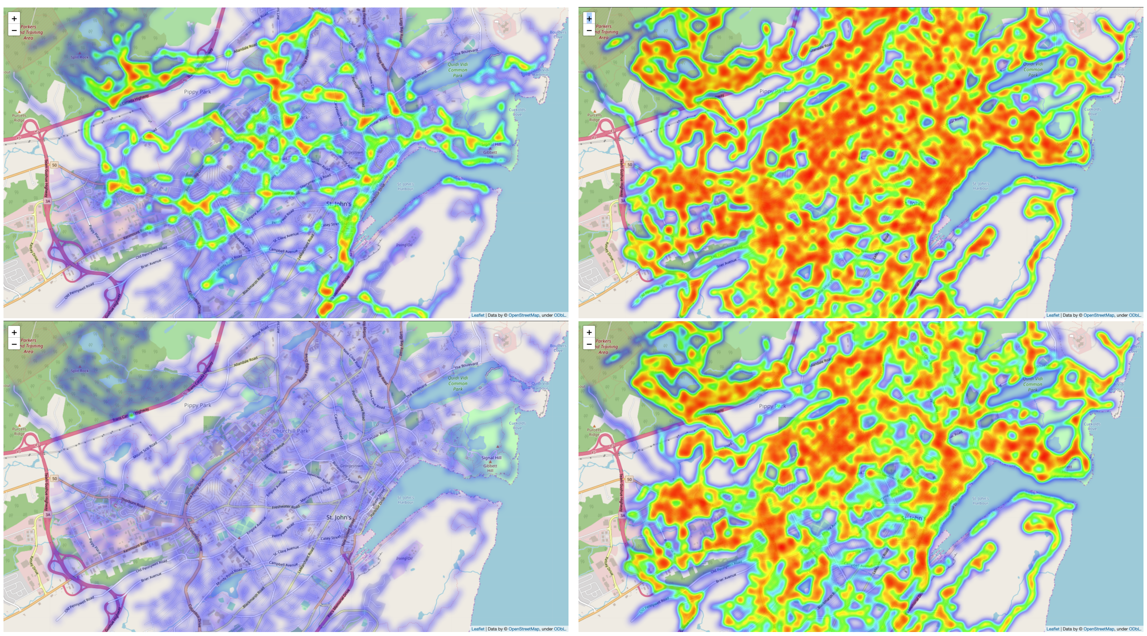

Figure 3.4: Top left: Betweenness centrality. Top right: Closeness centrality. Bottom left: hubs and authority centrality. Bottom right: page rank centrality. Visualized through Google Map.

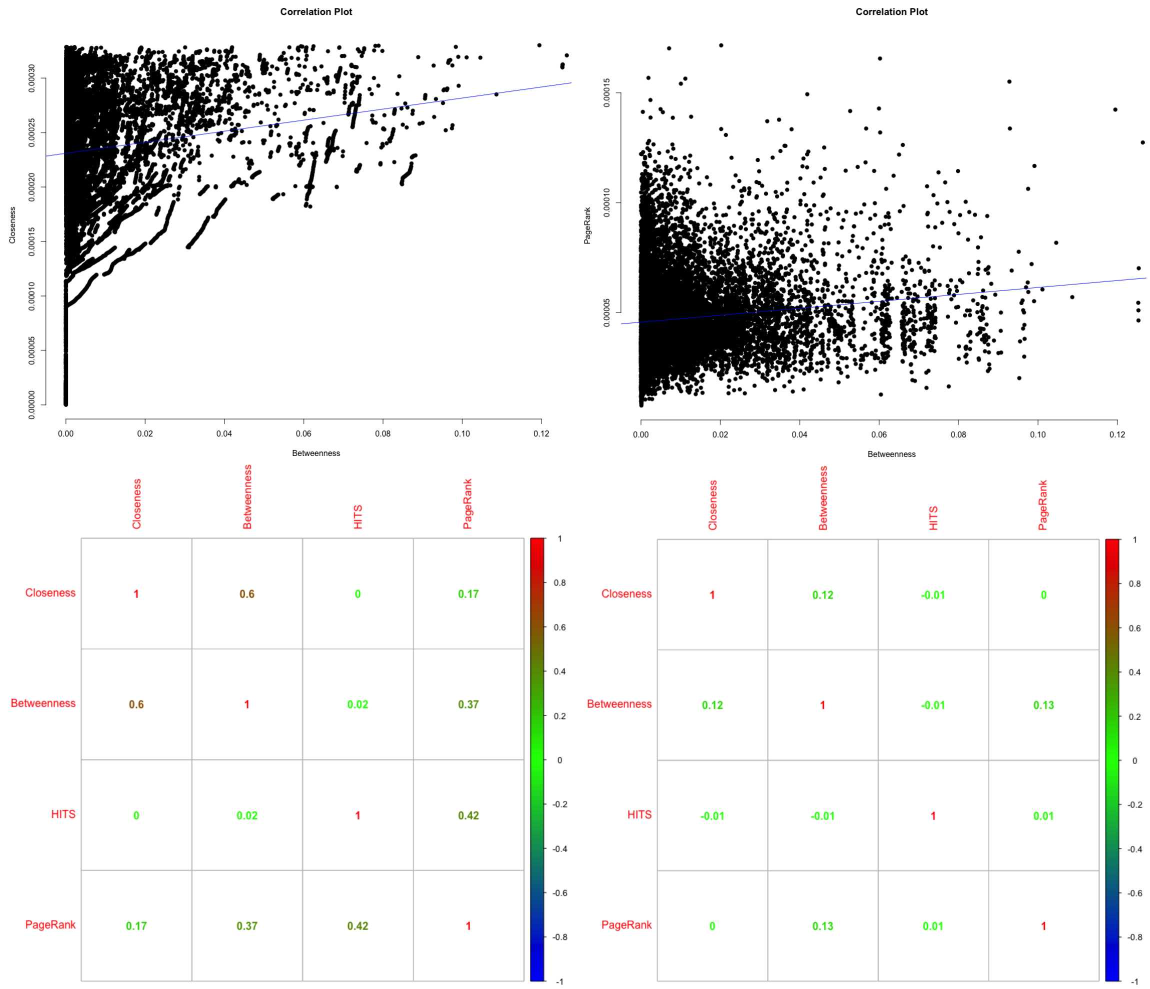

Figure 3.5: Top left: Betweenness vs Closeness centrality correlation comparison. The X-axis represents Betweenness centrality scores, while Y-axis represents Closeness centrality scores. Top right: Betweenness vs PageRank centrality correlation comparison (By Cytoscape [30]). X-axis represents Betweenness centrality scores while Y-axis represents PageRank centrality scores. Bottom left: correlation for a small road network of 85 nodes. Bottom right: correlation for a larger road network of 21,330 nodes (By RStudio [103])

.

test road networks covering small and larger sections of St. John's and Toronto. This phenomenon was an expected side-effect of the number of connected and disconnected components in each network. Some centrality measures, such as Closeness centrality, use the inverse of the distance, which could be infinite in the case of disconnected components turning the inverse to a very small number. Many small hand-picked road networks contained only a single connected component, such as multiple partial road networks covering downtown St. John's, east end and west end of St. John's, all containing only a single connected component. However, this was not the case for larger networks such as the road network for the entire city of St. John's, which contains 49 components.

Although one may assume a road network should always be a single component, making all roads connected, there are special cases such as when certain roads are not connected to the rest of the road network, such as race tracks, runways, roads in small islands that are only connected to the mainland by ferry lines, etc. The giant components of the networks where all the roads are connected were selected to address this. For example, an instance of Newfoundland's road network containing 21,330 nodes within 49 components was selected and determined to hold the majority of its nodes within its giant component, leaving only a few hundred nodes distributed among 48 components. To verify if using a single giant component will address the low and sometimes non-existing correlation between different centrality measures in larger networks, road data was filtered to contain the nodes from only the giant component. However, similarly, low correlation results between various centrality measures were observed even after removing all disconnected components. Notably, in the case of the partial road network for the province of Newfoundland, as seen in

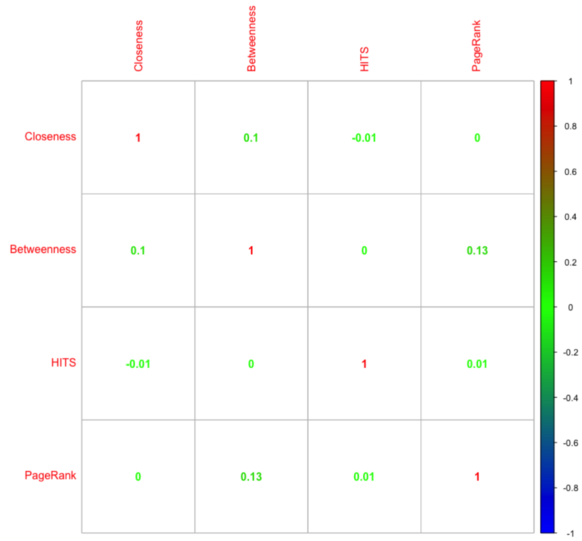

Figure 3.6: Correlation list for a larger road network containing 21,330 nodes. This figure is based only on the single giant component of the network that contains 20,975 nodes before reduction and 4,206 nodes after reduction. (Generated by RStudio [103])

Fig. 3.6, although the correlation is still similar, a small decrease is noticeable for the single giant component.

Road network	***Reduction time***	***Interpolation time***
Partial Toronto	01:33 hr	00:15 mins
St. John's metro	00:36 mins	00:07 mins
St. John's City	00:03 mins	00:01 mins

Table 3.3: Interpolation Computational time

Using the reduction algorithms in this research, the ability to reduce a significant number of nodes in various road networks was achieved, which led to a substantial reduction in the required computational time and saved countless hours, possibly days of computation, as evident from Table 3.2. Although to ensure accurate results are produced, the network structure needs to be kept intact during the reduction step, limiting vertex reduction. However, a significant reduction of 77% on average was achieved. Furthermore, to produce results for the entire network, specifically to address the removed vertices, a variation of linear interpolation was used, maintaining an accuracy of 89% on average. Though interpolation requires additional time to compute, the computational time for interpolation (Table 3.3) is almost negligible since it runs in linear time.

Chapter 4

ALF-Score: a Predictive Network-Based Walkability System

This chapter involves an in-depth definition of the ALF-Score walkability measure, which utilizes road network structure and machine learning to generate spatially high-resolution walkability scores derived from user opinion. This work has been submitted for publication. The paper submitted for publication is slightly revised for flow in this dissertation document, and the introduction is shortened.

Since there is no known predictive walkability measure, this research will devise an entirely new approach to measure walkability scores to bring a fresh take on how people consider locations walkable or otherwise. This approach allows the use of various essential features, currently not utilized, to help better understand our surroundings and map different locations together based on their similarities and characteristics. A predictive approach helps move into the highest possible spatial resolution, point—level, and ease into downscaling and upscaling processes when applied to various resolutions such as partial road networks, DAs, neighbourhoods, districts, etc.

One of the key elements of predictive walkability lies in its contribution, which opens the door to many new possibilities such as the capability of building a platform to provide personalized walkability based on individual's profiles (explored in the next chapter) where users' observations and demographics can be infused with individuals' opinions into building predictive and personalized models. These models can identify user patterns and compensate for the output to best fit the user's profile to bypass traditionally time-consuming and resource-intensive processes.

The overall ALF-Score pipeline (shown in Figure 4.1) requires various data inputs and processes. However, since ALF-Score is a network-based walkability measure, an essential step in this pipeline is to utilize road networks and other road characteristics better. To this end, ALF-Score incorporates a map database derived from the road network and POI inputs extracted from OpenStreetMap and Statistics Canada. Other GIS features such as road embedding and various centrality measures were later generated from the road network structure and used as additional road-network-based features. In addition, crowd-sourced user opinion is collected through the web-based data collection platform, specifically developed for this purpose. The web tool allows the collection of user opinions within groups of 5 locations, all of which are only relative among their respective groups. The crowd-sourced data is then processed through the Generalized Linear Extension of Partial Orders (GLEPO) algorithm in such a manner that the relative structure of each submission is converted into a global view within all user submissions. GLEPO's output of user ranking alongside the GIS-derived features are then fed into the supervised machine learning pipeline to train predictive models capable of estimating walkability score for any given point within the road network that falls within the map database coverage.

4.1 Crowd-Sourcing Platform

To further expand on the crowd-sourcing platform and the collected data, which play crucial roles in providing the required user opinion, first, how the data collection takes place is explored, followed by the reasons behind this approach. To have the ability to collect accurate user data and ensure data completeness while reducing bias, a web interface was developed capable of collecting various information from volunteer users such as walkability order of 5 randomly selected locations (relative ranking among only the 5 locations), users preferred walkable distance, their age group, gender, if users live alone, have children, their occupation, and a few other browser-agent details that are publicly available (such as users browser type, operating system, etc.). The web interface shown in Figure 4.2 top displays an interactive map with five randomly selected locations marked as $\{A, B, C, D, E\}$. Every time the web page is reloaded, Five randomly selected locations are chosen from a given geographic area. The coverage is tied to a pool of nodes derived from the map database, specifically the road network structure of the region. Users can zoom in and out, pan around each area and see various POIs, landmarks and region-specific information set automatically by Google Maps within the display area.

Various randomization methods are introduced to ensure an evenly distributed coverage of node labels and user opinions across the given geographical area. Users can reorder the ranking of the 5 locations by moving them up or down based on

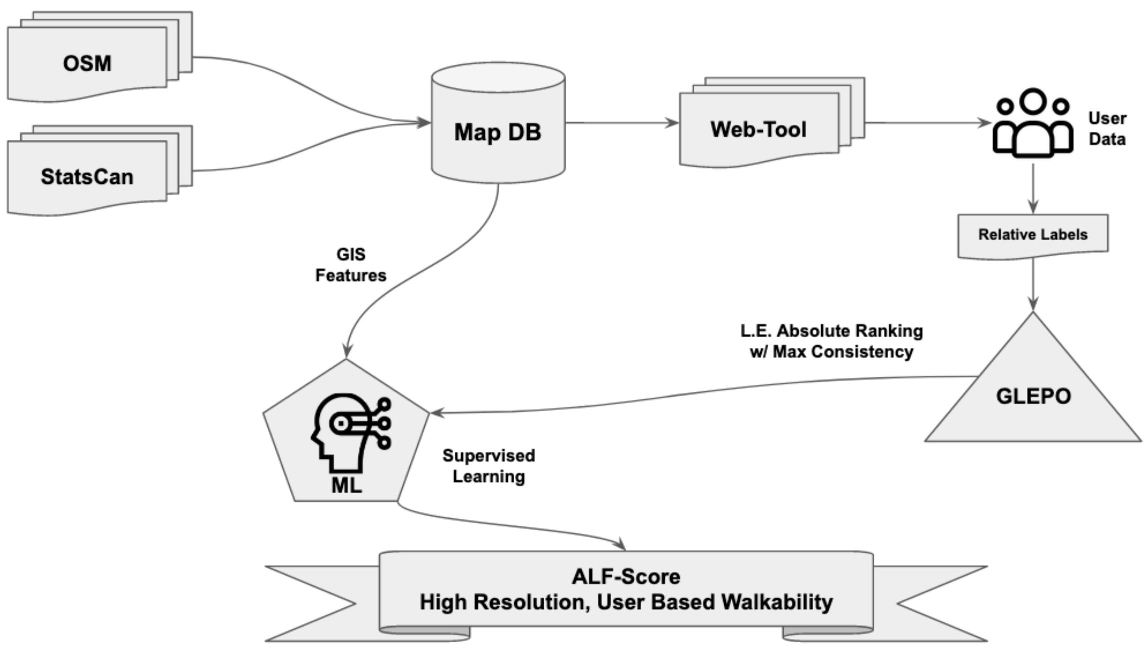

Figure 4.1: ALF-Score utilizes various GIS features such as road network structure, POI, and features derived from road networks such as various centrality measures and road embedding. GLEPO's linear extension of user opinions that produces a global view of relative user opinions is then aligned with the GIS features as an input to the supervised machine learning processes. Walkability estimates produced through trained models will have a high spatial resolution, represent users' opinions, and provide better insights into different regions and neighbourhoods. (Figure is drawn by the author.)

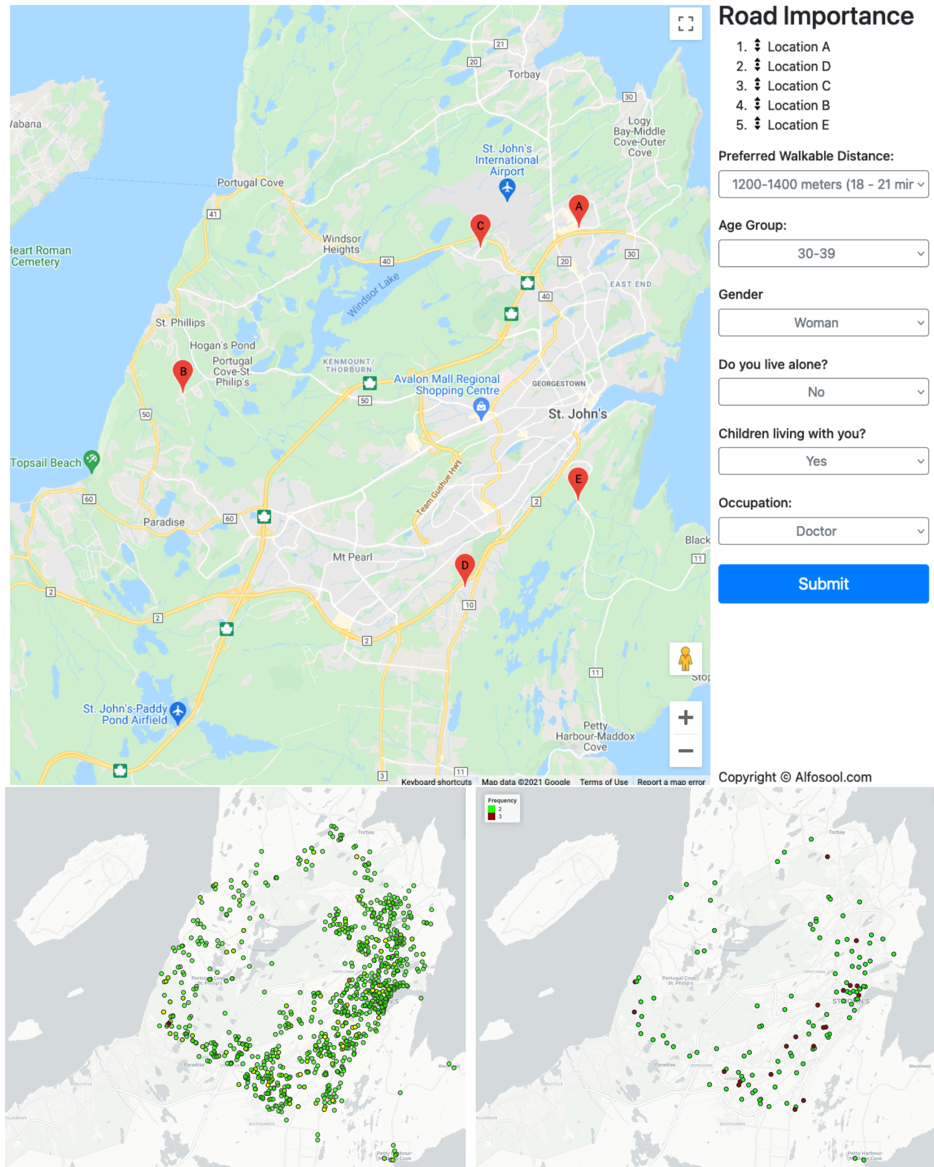

Figure 4.2: The interactive web-based data collection platform (top - Map generated by Google Map, screenshot taken from alfosool.com) has been deployed with road data from various cities in Canada. Displayed here is the city of St. John's, NL. Total 1050 (bottom left[1]) rankings were received from participants showing a well-distributed data collection. Some locations were ranked multiple times (bottom right[2]) by various participants. It is observed that the maximum number of conflicts in this scenario is 3 per location with very little occurrence. ([1],[2]: Maps generated through RStudio [103] Version 1.2 using mapview package from rstudio.com.)

how they perceive the relative walkability of the shown locations among each set of 5 locations. This approach reduces a potential cognitive challenge and allows users to use their intellect and knowledge of the city while exploring a diverse set of possible locations in the region. The relative rankings are then submitted as groups of 5 locations to the server for downstream processing. Given the random-selection nature of the web tool, the 5 (unique) locations randomly selected by the web tool can overlap across different sets, users or even different submissions of the same user. On the one hand, the overlaps provide common references among submission groups. On the other hand, the submission groups will likely accumulate perceived conflicts across their relative rankings. In Figure 4.2 bottom right, all locations with two or more associated submissions leading to conflicts are highlighted.

The decision to collect the data labels as a combination of relative rankings and not absolute scores (e.g. ranking every 5 locations between a fixed range such as 0-100 where the same rank may apply to other locations) was made to reduce bias, conflict and variability in the data while incentivizing the participants to make a precise decision to determine which of the given 5 locations would be most or least walkable and order the locations as deemed appropriate. While absolute scores are easier to use, each individual may have a completely different rationale for why a location has been ranked the way it is, such as a walkability score of 65 and not 40. Utilizing relative rankings takes that factor out and helps narrow the focus on the most important factor: whether location A is more walkable than location B for a given individual. This way, there will be a concise approach to determine each individual's perception towards walkability when comparing different points together over the responses of all users. Conflicts are unavoidable and lead to inconsistencies among user opinions. The challenge here is to balance the opinions collected from users to yield a walkability metric that represents an average user opinion. User opinions collected maintain a unique form observing only relative orders within each submission of 5 locations and therefore do not represent the global view among the overall data. Due to this and the nature of the conflict resolution, the problem remains in the NP-complete space, and therefore, there is a need to devise a new approach to handle conflicts and represent user opinions within a global perspective.

4.2 Generalized Linear Extension of Partial Orders - GLEPO

Generalized Linear Extension of Partial Orders or GLEPO, is an algorithm devised and developed in this research specifically to process the relative approach of user opinion label collection. GLEPO produces a generalized list of all user opinions in total order and absolute ranks to represent relative ranks of small submissions within a global observation of the overall data. This conversion is especially important as the users rank locations within small groups of 5 nodes. The order/rank of the nodes in these small groups remains unique and only relative within the same group of 5 locations. Rankings o_i^r between $1-5$ are relative only within their own set of 5 nodes where i ranges between 1 and the total number of nodes in a single submission (5). The o_i^r relative rankings similarly apply to all user submissions within S. Each S_j submission maintains five nodes holding a unique rank between $1-5$ where j ranges between 1 and the total number of submissions. o_i^r is completely localized at this stage and in no relationship with nodes from other submissions. To correctly utilize user data in conjunction with their submitted rankings and find the missing link between various submissions, an approach was devised to evaluate and establish a unified relationship between all nodes within the user-submitted data.

The first step of GLEPO is to group the submissions into multiple lists S_j containing 5 nodes $\{v_1, v_2, v_3, v_4, v_5\}$ each. The next step is to detect anchor nodes. Anchor nodes are defined as those commonly reoccurring nodes that naturally repeat in various submissions. For instance, in the submissions $x = \{n_1, n_2, n_3, n_4, n_5\}$ and $y = \{n_6, n_7, n_3, n_8, n_9\}$, the node n_3 would be considered as an anchor node defining a connection between submissions x and y. GLEPO algorithm uses anchor nodes to define connections between two or more submissions. The reason behind this approach is to narrow down an approximate positioning between various nodes if there is an anchor node(s) connecting them in two or more submissions. Submissions may have none, one or more anchor nodes. For instance, in the example above n_3 is the only anchor node, but it also happens to be in the centre of both submissions x and y. Therefore, $\{n_1, n_2\}$ must fall before $\{n_8, n_9\}$ and similarly $\{n_6, n_7\}$ must fall before $\{n_4, n_5\}$.

The main routine (Algorithm 4) iterates through the user data to form one possible variation of a newly sorted global list. When an anchor node is detected, the submission containing the anchor node is passed along to the $addToSorted()$ function to be evaluated and decide where the node entries within that submission should be placed. This process helps appropriately address the anchor nodes' associated relationship regarding nodes' ranks while sorting the entries based on their user-submitted order. However, when multiple anchor nodes are detected, special conditions may be invoked within the $addToSorted()$ function that determines the best course of action.

GLEPO's sorting subroutine (Algorithm 5) parses the passed submission where one (or more) anchor node(s) has been detected. The core of this algorithm revolves around determining the number of anchor nodes, their positions within the current submission and associated position in the partially sorted list, and invoking the appropriate case associated with each condition. There are four main conditions, two of which are associated with detecting only a single anchor node positioned either at the beginning or the end of the submission in process. The third condition is applied when the anchor node is not at the beginning or the end of the submission in-process and applies to two cases: 1) if only a single anchor is detected, or 2) multiple anchors detected but the anchor in the process is the last. The final case accounts for all other conditions. Each case determines the segment or segments of the submission in the process that requires insertion into the global list. These segments are then passed to another subroutine for insertion into the list.

Algorithm 3 Grouping user data by submission

Input: Original User Submission Data

Output: User Data Organized By Submission

initialize *subs_grouped* as an empty list
for every entry in user data **do**
    if *subs_grouped* contains an item with similar submission_id **then**
        append the new row to the existing sub list
    else
        append the new row to a new sub list.
return *subs_grouped* list

Algorithm 4 GLEPO: Main Routine

Input: User data organized by submission: *subs_grouped*

Output: List of sorted user data entries: *sorted_entries*

```
initialize sorted_entries as an empty list
initial randomization by shuffling subs_grouped
for every submission group sub_g in subs_grouped do
    if sorted_entries is empty then
        for every submission sub in sub_g do
            append sub to sorted_entries
    else
        for every submission sub in sub_g do
            if submission node exists in sorted_entries then
                add submission to list of anchor nodes.
        if anchor node(s) detected then
            pass sub_g and anchor list to addToSorted() function
        else
            pass sub_g to FindVLink() function
return sorted_entries list
```

Due to the structure of the data input and the nature of the algorithm, and if any variation in the order the input is fed into the algorithm is detected, a potentially different order of the global list is expected after every run of the algorithm. For example, the order of submissions and the nodes can affect the decision-making processes of the algorithm when conflicts are detected. Due to this, two randomization components were introduced to the process. The first randomization component is applied within the main subroutine (Algorithm 4), which randomly shuffles the order of appearance in all submission inputs. The second randomization component is invoked through the sorting subroutine (Algorithm 5), which randomizes the locations

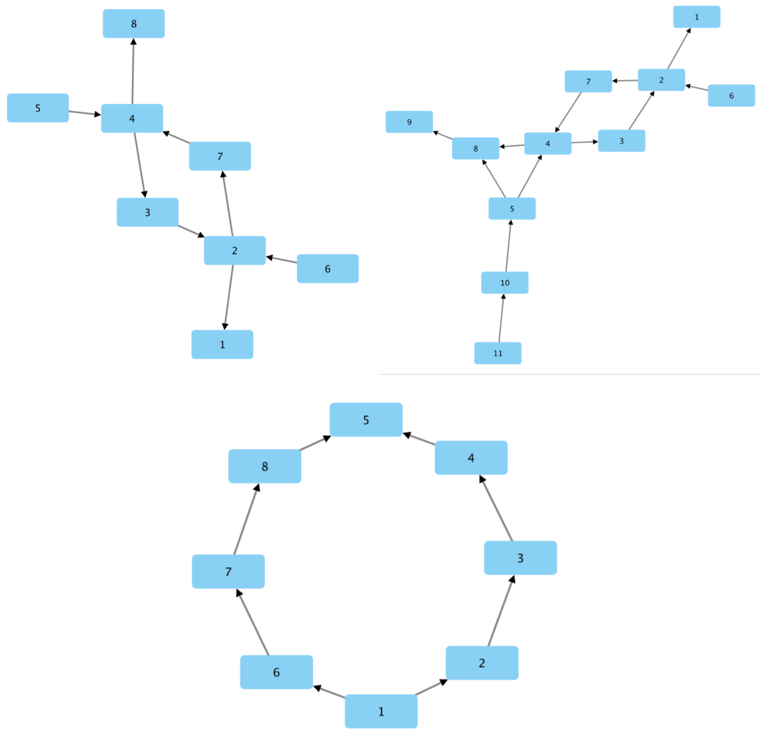

Figure 4.3: Three examples of many artificial networks created to represent various possible scenarios of user submissions and conflicts. Left: graph of 2 submissions (each with five nodes) containing two anchor nodes, '4' and '2', forming a loop. Center: graph of 3 submissions with four anchor nodes. Right: graph of 2 submissions containing two anchor nodes '1' and '5' falling on two extreme ends. (Figures are drawn by the author.)

that nodes are inserted into the global list by determining all appropriate positions and randomly choosing one to insert the node in the process. Further details on how the randomized insertion is used can be found in the Algorithm 6. With the randomization introduced, the entire GLEPO process requires running multiple iterations until the order of nodes within the global list converges. This convergence is associated with the order of how each node appears within the global list in each run. The main process averages every node's order throughout all iterations and determines the final order under which each node falls after multiple iterations. Due to this approach, various nodes may have a similar ranking which is the expected result. GLEPO has been tested over numerous numbers of iterations such as 10, 30, 50, 80 and 100 iterations. It was observed that with 30 iterations and onward, the order of nodes appearing in the generated global lists starts to converge and yield a stable result.

Algorithm 5 GLEPO: Sorting Subroutine

Input: Submission group *sub_g* alongside its list of anchor node(s)

Output: Updates: *sorted_entries*

```
set current pointer to 0
for every anchor node do
    if it's the only anchor and is positioned at the beginning of sub_g then
        set segment to all 4 elements to its right
        pass segment to RandomizeInsertion() function
    else if it's the only anchor and is positioned at the end of sub_g then
        set segment to all 4 elements to its left
        pass segment to RandomizeInsertion() function
    else if it's the last anchor but positioned in the middle of sub_g then
        set left segment to all elements to its left
        set right segment to all elements to its right
        pass left segment to RandomizeInsertion() function
        pass right segment to RandomizeInsertion() function
        update current pointer
    else
        set segment to all elements between current pointer and anchor's position
        pass right segment to RandomizeInsertion() function
        update current pointer
return sorted_entries list
```

An important step here is to ensure GLEPO can handle special cases, no matter how rare their occurrence may be. To this end, various artificial networks were created to present different scenarios within the user submission dataset. Each of these artificial networks was converted into a graph and processed by GLEPO. The resulting output was tested and verified to ensure the algorithm always returns results with the least possible inconsistency (defined later in this chapter). In Figure 4.3, three artificial networks are demonstrated to verify GLEPO's accuracy and consistency. A few possible (and correct) outcomes are observed for each graph. For example, in Figure 4.3 right, node 5 is the highest-ranked node among the two submissions, whereas node 1 is the lowest-ranked node among the two. Therefore, based on users' opinions and the limited information given by only these two submissions, node five should be ranked the highest and node one the lowest. In this example, all other nodes could form several different orders as long as they remain true to their respective submission. To name a few possible outcomes $[1, 2, 3, 4, 6, 7, 8, 5]$, $[1, 6, 7, 8, 2, 3, 4, 5]$ and $[1, 2, 6, 3, 7, 4, 8, 5]$ are all potential lists. Node 1 maintains the least rank, while node 5 is the highest in all generated outcomes. It is further observed that nodes from each submission always follow their original order of $[2, 3, 4]$ and $[6, 7, 8]$. It is important to reiterate the significance of the randomization component and how it will help form a uniform distribution of nodes into positions within GLEPO's output that are most appropriate, especially when there is no anchor node to establish a direct connection between various nodes.

To ensure the sorted list remains true and consistent with the users' submitted rankings, its consistency to the original user data is examined. Each user submission consists of 5 nodes $\langle v_1, v_2, v_3, v_4, v_5\rangle$, where each node v_i $(i = 1, 2, 3, 4, 5)$ has a ranking r denoted o_i^r, ("order") between 1 and 5. The rank o_i^r defines users' assigned order to each node concerning the other four nodes within that submission. When GLEPO sorts the nodes from all submissions into a unified global list, the most natural approach to check for any inconsistencies is to compare and verify the order of all nodes in their original user submissions with that appearing in the global list. Suppose the order of the five nodes in each original submission remains intact and is the same as that appearing within the global list. In that case, the nodes within that submission are considered consistent with the original user's opinion. However, if the order differs, all out-of-order nodes are counted, amounting to a value representing its inconsistency, which is done by pair-wise comparison of each submission appearing in two separate lists: 1) a list in the original order submitted by the user, 2) a list in the order appearing in the sorted list.

Based on the earlier definition of anchor nodes, submissions containing one or more anchor nodes become anchored. In contrast, those submissions without an anchor node are defined as Constrain-Free submissions. Based on this definition, two possible challenges are faced with GLEPO on how well-anchored the submissions are concerning user data: 1) over constraining, 2) under constraining. Over-constraining occurs with excessive conflicts and inconsistencies through anchor nodes when adding new nodes to the general list. For example, when the majority of the submissions have multiple conflicts with multiple submissions, over-constraining is faced. With more crowd-sourced user opinion comes more potential conflicts and disagreements. A natural solution to this is the embedded randomization of GLEPO over multiple iterations to ensure a good distribution within the input set and the node insertion.

Algorithm 6 GLEPO: Randomized Insertion Subroutine

Input: Current list of sorted entries, minimum insertion range, maximum insertion range, list of elements to insert

Output: Updates: *sorted_entries* by randomly inserting new elements within the given range

if the size of the list of elements to insert is less than the range between min and max **then**
  for the size of the list of elements to insert, generate random numbers within min and max range
  sort the generated random indexes
  for a range between 0 and the size of the list of elements to insert **do**
    insert an element of the list into *sorted_entries* at the position within the generated numbers
else
  for the size of the range between min and max, generate random numbers within the size of the list of elements to insert
  sort the generated random indexes
  for a range between 0 and the size of the range between min and max **do**
    insert an element of *sorted_entries* at selected position into the list of elements to insert at the position within the generated numbers
  replace elements of *sorted_entries* within min and max range with an updated list of elements to insert
return *sorted_entries* list

Under constraining, on the other hand, occurs when there is an insufficient number of connections (or anchors) among user submissions. This lack of anchored submissions leads to an inability to establish accurate associations among submissions. These associations enable the algorithm to present the data within a global list that reflects the global view of all users concerning one another. The majority of the user submissions within the crowd-sourced data used in this research do not have an anchor node and are constraint-free, leading to under-constraining. Two main approaches are considered when addressing this challenge, both of which are complementary and can be done concurrently. The first is the collection of more user data. Second, through creating virtual anchor nodes or virtual links. The virtual link is defined as an artificial connection between two or more submissions created specifically to associate constraint-free submissions that do not have any naturally-occurring anchor nodes. Virtual link utilizes geographical proximity to establish a virtual connection between a single submission and the partially complete global list by determining the closest node(s) (or shortest physical distance) between the two. The algorithm 7shows the approach taken to create virtual links. This subroutine is called through the main routine whenever a submission without an anchor node is detected.

Algorithm 7 GLEPO: Virtual Link Subroutine

Input: Current list of sorted entries, submission group *sub_g*, distance matrix of all nodes

Output: Create a virtual link for the given *sub_g* and updates *sorted_entries* accordingly

set the desired threshold - maximum distance to consider a virtual link
initialize main shortest distance to a very high number
for every element in *sorted_entries* **do**
    find element's distance to all 5 nodes within *sub_g*
    if shortest distance among the 5 is shorter than main shortest distance **then**
        set main shortest distance to new distance
        set main shortest node to associated node
if main shortest distance < threshold **then**
    treat main shortest node as an anchor node
    pass *sub_g* and main shortest node to *addToSorted*() function
else
    pass the entire *sub_g* to *RandomizeInsertion*() function
return *sorted_entries* list

4.3 Active Living Feature Score - ALF-Score

The Active Living Feature Score or ALF-Score is a novel measure of walkability that aims to create a better scoring system capable of generating walk scores at a much higher spatial resolution (point level) that are representative of user opinions and more informative to most individuals. ALF-Score takes a user-centric approach instead of the traditional researcher-centred approach. ALF-Score aims to utilize road network structure and characteristics better and derive node features defined by various road characteristics and their direct and indirect associations that can consider how the underlying road structure can influence neighbourhood and city structure and walkability scores. ALF-Score employs user-based parameters such as users' opinions of walkability to represent users' perception better and provide user-derived scores of what individuals perceive as various degrees of walkability.

At the core of the ALF-Score pipeline lies various machine learning approaches and methods to train machine-learned models capable of estimating numeric values defining how walkable specific points on the map are, based on various feature combinations that include road network, road embedding and POI features. This approach would help estimate walkability scores based on node characteristics. The input maintains a combination of various features derived from road structure and POIs generated by different methods while user opinion acts as training labels. It is essential to highlight that raw user opinions are not directly used as labels but instead processed through GLEPO with the resulting list representing user opinions within a global view among all participating users, which are used as the labels for various machine learning techniques. The ideal outcome is to avoid calculating walkability scores individually for every single point on the map and have the ability to accurately estimate point-specific scores from only a small sample of crowd-sourced labelled data. The output of the machine learning pipeline would be a trained model, and the model's output will be walkability score predictions for given nodes. The score will define how walkable users may find the given node (location), an output normalized between 0-100.

When it comes to the machine learning pipeline, and since the goal is to estimate continuous numerical output (walkability scores) based on numerical features (some converted from other types such as categorical, ordinal, etc.), this would make up for a node regression problem. To address a regression problem, there are many possible approaches such as linear regression, random forest, support vector regression (SVR), and multi-layer perceptron neural network (MLP), which are explored in this research. The prediction input variables describing data features are represented as $\{x_1, x_2, x_3, \ldots, x_n\}$ while the output is represented by y. To train the model, labelled sets are defined as $\{features, label\} : \{x, y\}$. Given an unlabelled set $\{features, ?\} : \{x, ?\}$, the expectation from trained models is to produce y'. The models, defined by internal parameters learned through the process, map unlabelled example sets to predicted value y'. Additionally, since some of the features are not numerical (e.g. categorical or ordinal), one hot encoding has been used where applicable to convert non-numerical features into appropriate numerical entries. The developed machine learning pipeline has five primary components: 1) GLEPO, 2) consistency measure and verification, 3) feature selection, 4) model training, and 5) model validation.

The model's output, walkability scores for given points within the dataset, is robust and refined with a much higher spatial resolution than other walkability measures. The nature of the output and its predictive structure can be used to create interactive interfaces that allow users to visually view walkability scores for any selected areas with as high as point-level resolution. Furthermore, due to the nature of the model, there is no need to calculate the walkability score for the entire network. Predictions for small subsets or even the entire network, as needed, can be made available in a fraction of the time required to calculate the walkability scores using traditional methods.

The desired walkability function is a global scalar which is consistent across all nodes with user input. But it is also coherent across all nodes without user input.

This walkability function is defined as $w : V \Longrightarrow R^+$ where V is a set of vertices and R is a set of user rankings. The performance metric of w is based on consistency with user rankings R. Specifically, the performance metric considers user ranking $r \in R$ involving a set of 5 nodes $\{v_1, v_2, v_3, v_4, v_5\}$. Ranking r gives each node v_i ($i = 1, 2, 3, 4, 5$) a rank denoted by o_i^r (i.e. the "order" of node v_i as they appear in user ranking). Similarly, walkability w would also imply v_i ($i = 1, 2, 3, 4, 5$) a rank denoted by o_i^w (i.e. the "order" of node v_i as suggested by w). Furthermore, the loss function of w for r, or the "inconsistency" between r and w, can be defined as

$$L(r, w) = \sum_{i=1}^{5} |o_i^r - o_i^w|$$

while the loss function concerning R is aggregate of $L(r, w)$:

$$L(R, w) = \sum_{j=1}^{l} L(r_j, w)$$

That is, the total amount of inconsistency of w as compared to all user-provided rankings. The aggregate function can take other forms, too, such as mean or multiplication. There can also be other ways to define such an inconsistency measure, e.g. l_2-norm or "out-of-order" counts, or Kendall rank correlation coefficient [69]. The goal is to determine the most appropriate w that minimizes $L(R, w)$, the amount of inconsistency between w and R. A conversion from "relative" ranking only associated within a small localized set of nodes provided by a single user to a "global" list of scores which represents relativity among all users and their provided opinions within the network with as little discrepancy among R as possible is a crucial step.

Algorithm 8 Inconsistency Loss Function

Input: *sorted_grouped* list where it maintain the original order as well

Output: Percentage of inconsistency over all user data and its sorted counterpart

initialize *count_mis* as 0
initialize *count_sub* as 0
for every submission group *sub_g* in *sorted_grouped* **do**
    initialize $list1$ as *sub_g* with user-submitted orders
    initialize $list2$ as *sub_g* with GLEPO sort orders
    initialize *count* as 0
    for every $index1$ between 0 and length of *sub_g* **do**
        if $index1$ is not the last index **then**
            for every $index2$ between $index1 + 1$ and length of *sub_g* **do**
                retrieve $index3$ value at position $index2$ of $list2$
                retrieve the $index4$ of the order at position $index3$ of $list1$
                if $index1$ is greater than $index4$ **then**
                    increase *count* by 1 unit
    add *count* to *count_mis*
    increase *count_sub* by 10 units
set loss percentage to $(count\_mis \times 100) \div count\_sub$
return loss percentage

The improved computational complexities for various algorithms in the pipeline are highlighted in Table 4.1.

Various machine learning techniques are applied and compared based on various feature set combinations. The goal is to find the most suitable technique and feature combination set that produces the most appropriate models predicting accurate walkability scores. The techniques that were used in both supervised and semi-supervised environments were: 1) random forest, 2) linear regression, 3) decision tree, 4) support

Table 4.1: The computational complexity of various algorithms in the pipeline.

Method	***Big-O***
Distance Matrix Measurement	$O(3n^2)$
AddToSort	$O(10n)$
RandomizedInsertion	$O(n)$
Virtual Link Creator	$O(4n)$
GLEPO	$O(3n^2)$

vector regression (SVR), 5) gradient boosting, 6) polynomial features (a non-linear approach), 7) lasso CV and 8) multi-layer perceptron neural network (MLP). Feature combinations used are: 1) only POI features, 2) only network features (centrality measures), 3) only road embedding features, 4) POI + road network features, 5) POI + road embedding features, 6) road network + road embedding features, 7) all features.

It is important to point out that ALF-Score is a novel and unique approach with no similar measure for direct comparison. The success metric when it comes to determining the accuracy of generated models is based on how close the results are to user opinions and the general knowledge of the area, based on two main approaches: 1) using validation and test sets to verify the models, 2) visual inspection and verification based on local knowledge of various cities. As with most machine learning approaches, various challenges are observed, especially when keeping the models as generalized as possible. For example, insufficient data, bias in data collection or processing, and incorrectly labelled data are challenges that can lead to modelling errors. Overfitting and underfitting are also two very common modelling errors that occur when the model's primary function is too closely fit the data (too well-learned) for the case of overfitting or when the function does not capture the prominent patterns in the data (not enough learning) in the case of underfitting. Overfitting leads to high accuracy in training results (previously seen data) but leads to significantly lower accuracy when applied to new (unseen) data. Underfitting, on the other hand, leads to unpredictable outputs. In both cases, low generalization of the models leads to unreliable predictions. Furthermore, one of the main challenges this research faces is that the crowd-sourced data is relatively small. Two directly associated issues are: 1) potential bias in representation due to lack of variety in participating user groups, 2) lack of appropriate coverage when it comes to relative walkability scores associated with various road network structures.

There are many approaches to address these challenges. One of these approaches used in this research is data split. The core concept of data split is to separate the labelled data into two datasets: 1) training set 2) testing set. For example, in an 80-20 percent split, 80 percent of the labelled data is assigned to the training set while the remaining 20 percent form the test set. Typically a 70-30 or an 80-20 percent split is commonly seen, but various other variations such as a 90-10 and a 60-40 percent splits are experimented with. It is important to highlight that the labels from the test set are never shown to the algorithms while training the models. The prediction results are then compared to the actual results to determine a baseline on the accuracy and performance of various models. Furthermore, a well-balanced split can make a huge difference in a well-generalized model, especially when labels are scarce. For instance, when working with a small set of data, if too much or too little is assigned to either the training or the test set, the model may not perform well due to a lack of observable patterns. Additionally, Mean Absolute Error (MAE) and Root Mean Squared Error

(RMSE) were used to measure each model's error. Mean Absolute Error (MAE) is defined as

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

where x_i is the actual value, y_i is the prediction, and n is the total number of data points. Root Mean Squared Error (RMSE), on the other hand, is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}}$$

K-fold cross-validation [65] is used as yet another method to address this challenge. K-fold cross-validation is typically done by splitting the dataset into k smaller sets. One set is selected as the test dataset, and the remaining sets are combined as one training dataset. Once the model is built, it is evaluated on the test dataset. The procedure repeats for every other set in k to ensure the entire dataset is utilized towards building and evaluating the model.

4.4 Results

In total, 668 features (excluding the 12 demographic parameters used in the later chapters) are used to train various models, with some models trained with a specific subset of these features. This is done to examine and compare the resulting accuracy and determine the effects of each feature set on generating walkability scores closest to that of users' opinion baseline. The following input parameters are used in the machine learning pipeline:

1. Road network data $G = (V, E)$ which covers the city of St. John's, NL. G is connected, undirected and unweighted. $|V| = 5{,}341$ nodes and $|E| = 6{,}835$ edges. Road node features, a mix of numeric, ordinal, and categorical data. 1) POI features with 530 features. These features are derived from 8 separate OSM categories with 53 total subkeys. Every single key contributes to 10 features. Each of these features reflects on a specific geometrical distance range showing whether that particular key falls within the associated distance of any node within the network. The geometrical distance ranges from 200 meters to 2,000 meters with 200-meter increments. These 530 features are extended to every node on the road network. 2) Network features consisting of 10 features, namely: Betweenness centrality, Closeness centrality, clustering coefficient, degree, eccentricity, neighbourhood connectivity, stress, topological coefficient, average shortest path length and radiality. 3) Network embedding with 128 features using node2vec [55] technique.
2. Crowd-sourced data. At this stage in the research, there are approximately 40 unique users with $|S| = 210$, where S represents all user submissions and each S_j where $(j = 1, 2, \ldots, |S|)$ represents a user submission. Each S_j user submission consists of 5 road nodes $\langle v_1, v_2, v_3, v_4, v_5 \rangle$. This data collection approach was chosen (rankings of 5 nodes per submission) with careful considerations. With a small number of nodes (i.e. 2 locations), there is not enough relativity among submitted nodes. With too many nodes (i.e. 10 locations), there is a risk of posing cognitive challenges when users decide rankings between many locations. To ensure sufficient relativity among submitted data yet avoid cognitive challenges, there need to be enough location nodes to create a balanced

association among various nodes. There are a total of 1,050 user-submitted locations among their relative ranks. Of these 1,050 locations, there are 852 unique locations. Leaving 198 "anchor" locations appearing in two or more submissions. Each road node v_i $(i = 1, 2, \ldots, 5)$ has a unique relative ranking r_{v_i} between $\{1, 2, \ldots, 5\}$. The order of r_{v_i}, which is a relative ranking, is of utmost importance as it defines how users perceive the walkability of each v_i concerning one another.

Using the Generalized Linear Extension of Partial Orders (GLEPO) algorithm, users' relative opinions (among groups of 5 locations) were successfully converted into globally relative scores (among all submissions). GLEPO maintains a high consistency of 98.24% on average throughout the conversion. Furthermore, after numerous variations and experimentation, it was determined that the best results were produced using the randomized version of the algorithm with at least 30 iterations. Virtual link, a method developed to establish associations between user submissions based on their geographical distance, was enabled when the best results were produced.

Compared to Can-ALE, a clear variation is observed, especially when producing fine-tuned high spatial resolution ground truth. As observed in Figure 4.4, while Can-ALE (left) is dedicating a single and spatially low-resolution walkability score to each large region (DAs), users have different and varying opinions (center) ranging from low to high walkability associated to different points on the map that may fall within the same region. Therefore, GLEPO provides a more representative and high spatial resolution ground truth. This research does not imply that Can-ALE is incorrect but rather points out some of its limitations and shortcomings due to its characteristics. Since Can-ALE is area-based, its scores provide a spatially low-resolution coverage.

Since user opinion is not used in its pipeline, Can-ALE scores do not represent users' opinions. As observed in GLEPO's result, not only do the volunteer participants have a clear difference of opinion with that of Can-ALE's provided walkability scores but there is also a difference in opinion among participants themselves, which opens up the possibility for future research on applying GLEPO and ALF-Score to different subgroups of users (personalization which is explored in the next chapter).

Although GLEPO's results provide an overall well-represented user opinion baseline to walkability scores while establishing a ground-truth for ALF-Score machine learning pipeline as its y label vector, GLEPO's main contributions may be extended to a vast range of researches such as enabling accurate use of crowd-sourced data while reducing bias as a result of the difference in opinion.

Figure 4.4 further expands on how predicted walkability scores generated by ALF-Score (right) compare with that of Can-ALE (left) and how ALF-Score can provide high spatial resolution walkability scores instead of existing area-based walkability measures. High variation was found in ALF-Score walkability scores while observing an even distribution within the selected region. For instance, the closer to the Kelsey Drive region, the more walkable ALF-Score becomes. (There are numerous amenities within the vicinity of Kelsey Drive.)

GLEPO's output was used to train the supervised machine learning models as a regression problem on 668 features iterating over multiple variations of data combination subsets. Six primary techniques were used, namely: 1) Random Forest Regression, 2) Linear Regression, 3) Support Vector Regression (SVR), 4) Gradient Boosting, 5) Decision Tree Regression, and 6) Multi-layer Preceptions (MLP). Random Forest outperformed all other techniques in terms of performance and accuracy

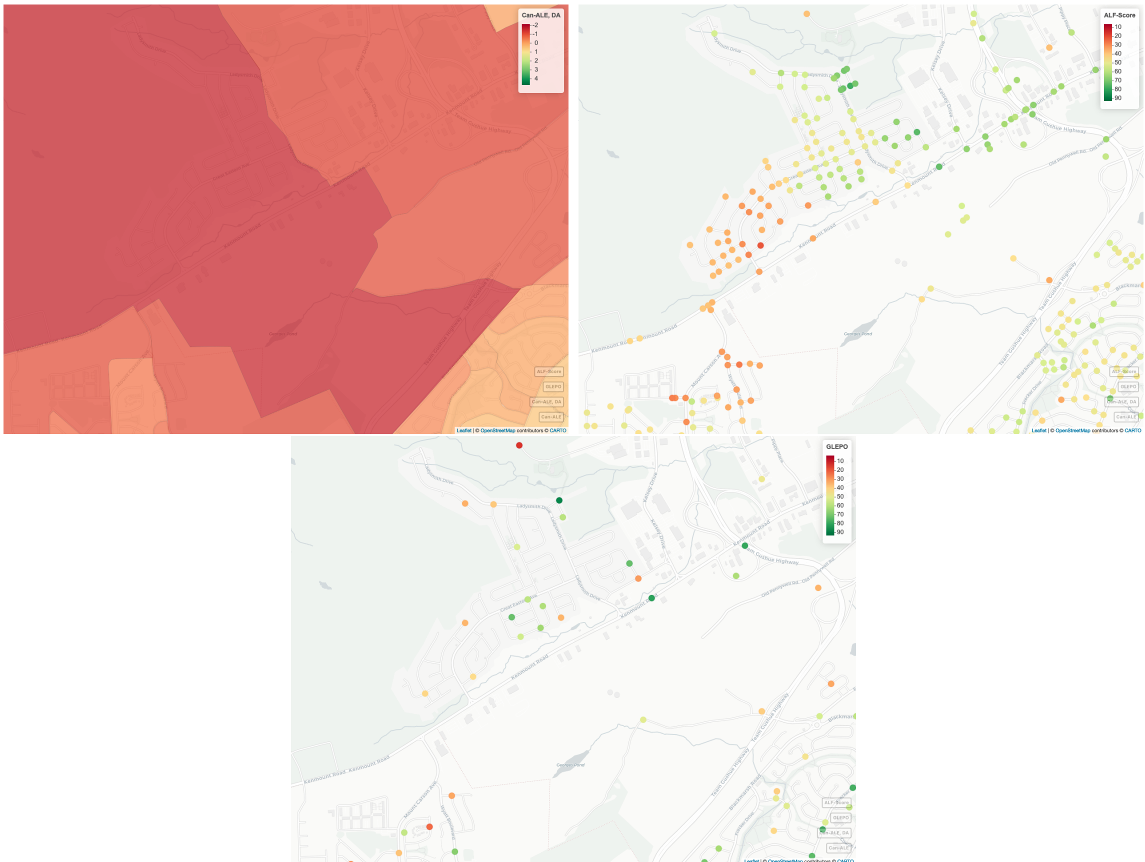

Figure 4.4: Great Eastern Ave. in St. John's, NL. Top Left: Can-ALE, Top Right: ALF-Score, Bottom: GLEPO. A score variation ranging between 16-81 is observed within GLEPO, which provides a user-level insight about how users perceive their neighbourhood in their own opinion, as opposed to what Can-ALE suggests of a neighbourhood to its users. Furthermore, a variation ranging between 20-70 is observed in ALF-Score, which provides high spatial resolution scores specific to each point instead of a single score for an entire area observed in Can-ALE. Note: Can-ALE colours are slightly dimmed due to the adjusted opacity to visualize street overlay.

by achieving a top prediction accuracy of 87.49% using all features. Table 4.2 takes a closer look at all feature combinations used in various experiments followed by their best-achieved accuracy. All accuracy values are representative of the best-recorded accuracy over numerous runs. Models trained on only POI features were observed to perform relatively similar to that of those models trained on POIs plus network features and POIs plus road embedding. However, models trained on network features combined with network embedding perform better than those using POIs (except for linear regression). The highest accuracy was observed when all features were used together, which presents a convincing argument in contrast to some other walkability measures' hypotheses, which assume high importance towards POI features. Furthermore, it also conveys an important message that road network structure plays a crucial role in measuring walkability. The improvement in accuracy observed after adding POI features to the network and road embedding features contributes to the complimentary positions POIs and road network features have concerning one another, where combined can provide a more in-depth understanding of our surroundings.

Technique \ Set	***POI***	***POI + Net.***	***POI + Emb.***	***Network + Embedding***	***All***
Random Forest	77.36	79.49	79.40	81.07	**87.49**
Linear Regression	51.15	54.57	61.74	30.86	**72.21**
SVM	66.47	68.50	66.35	75.38	**76.36**
Gradient Boosting	60.30	59.16	58.17	68.75	**68.78**
Decision Tree	65.33	67.52	64.98	75.89	**76.35**
MLP	69.86	70.94	73.18	74.63	**79.87**

Table 4.2: Exploration of various machine learning techniques and feature combinations with their top-performing accuracy.

Figure 4.5 shows how ALF-Score (center) compares with two prominent walkability measures, namely Can-ALE (right) and Walk Score (left), for the walkability scores of the city of St. John's, NL (Canada). The walkability scores produced by Can-ALE vary by region (DAs), where all nodes within each DA carry the same score. Some of these regions may cover much larger geographical areas, whereas other regions may cover a higher population density within a much smaller geographical area. This variation significantly reduces these scores' accuracy, relevancy, and spatial resolution. As observed in Figure 4.5 center, this issue is no longer the case with ALF-Score walkability ranks. They are much more refined with a clear variability among the scores for different nodes within the same region and significantly higher spatial resolution when compared to Can-ALE. Furthermore, ALF-Score walkability ranks represent users' opinions with higher accuracy associated with users' perception of walkability in different regions. The accuracy of the predicted ALF-Score ranks was successfully verified based on user-opinion-based ground-truth and local knowledge of the city of St. John's.

Furthermore, it was observed that areas with greater population density are assigned with higher Can-ALE scores, which may not always be how users perceive walkability. In Figure 4.6 which represents two examples where ALF-Score and Can-ALE do (right) and do not agree (left), it is observed that Can-ALE fail to identify regions such as local in-city parks and trails as walkable areas as observed in Figure 4.6 left. The Signal Hill area is a well-known and commonly visited area by the local community, especially hikers, runners, joggers and families. There are numerous amenities nearby, such as convenience stores, restaurants, coffee shops and gas stations, and there are multiple bus stops. Yet, the area was considered as not walkable

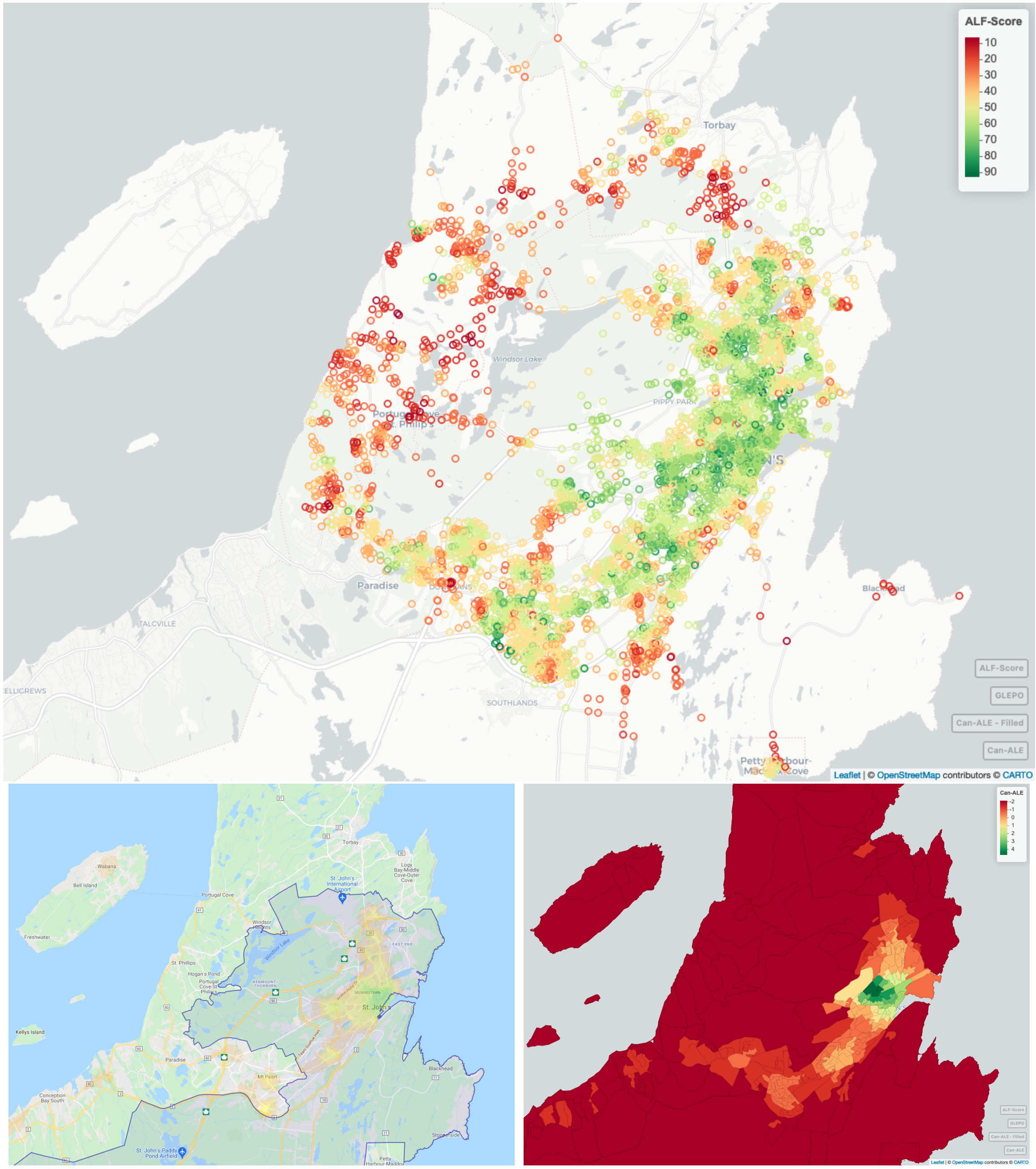

Figure 4.5: Walk Score [124] Walkability score of St. John's, NL (bottom left - screenshot from walkscore.com), followed by Can-ALE Walkability score (bottom right) and ALF-Score walkability of the same region (top). Dark green is most walkable, dark red is least walkable. Walk Score is observed not to have sufficient data to cover the entire region while Can-ALE is observed to be overly generalized. Generated by RStudio [103].

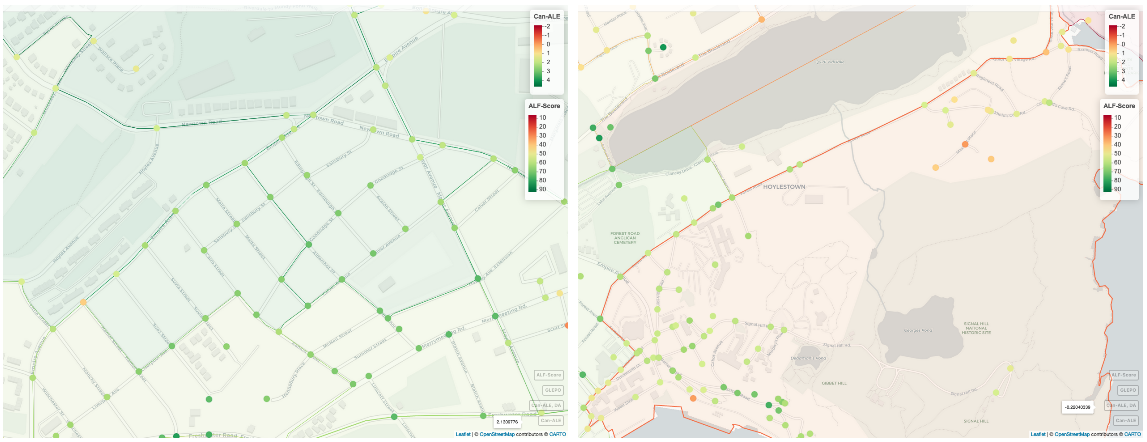

Figure 4.6: Examples of where ALF-Score and Can-ALE do and do not agree. Each DA polygon is represented by a single Can-ALE value, while each circle represents an ALF-Score rank. Left: Downtown St. John's, NL, shows a strong agreement among the two measures. Right: Signal Hill region, St. John's, NL, shows a strong disagreement between the two measures. Note: Can-ALE is represented as an overlay with a small opacity/alpha and is visualized with significantly lighter colours due to this transparency. Legends and borderlines represent the actual colours. (Maps generated through RStudio [103] Version 1.2 using mapview package from rstudio.com.)

by Can-ALE. The analysis suggests that Can-ALE's approach is missing some of the important features and area characteristics that people consider important such as being near ponds, trails, and national monuments. Based on various experiments, ALF-Score is observed to do a much better user representation and provide a more in-depth ranking based on different road characteristics, feature sets and users' opinions.

4.5 Discussion

The purpose of this chapter is to fill in the gap and address some of the most prominent challenges observed in existing walkability measures by introducing a novel approach to measure walkability score with a high spatial resolution, precision, accuracy and efficiency. Strong indicators of walkability can be derived from road importance which requires extensive utilization of road network structure. However, there is very little use of road network structure, if at all, in most existing walkability measures. This lack leads to a significant gap in utilizing one of the most important walkability factors and, hence, lower accuracy. Although an important consideration is that road network structure can be quite large and creating a graph structure while calculating graph-based features (such as centralities and network embedding) pose various challenges (which is addressed in the previous chapter [5]), ALF-Score walkability scores using road network structure show a substantial improvement, especially when the road network structure is combined with user opinion. Road network structure provides crucial information about neighbourhoods and cities. Furthermore, road network structure is also a conduit to propagate the importance of POIs. It was shown that incorporating user opinion and perception as one of the crucial contributing components in the pipeline can significantly improve the overall relevancy of the generated scores.

Additionally, using crowd-sourced data in ALF-Score opens the door to many new possibilities, such as building a platform to provide personalized walkability based on user demographics, which will be challenging without user opinion. Users' observations and demographics can be included in personalized models that can identify user patterns and generate models to best fit users' profiles.

Furthermore, collecting crowd-sourced data is expensive, time-consuming, limited and resource-intensive. It is difficult to recruit a diverse group of volunteer participants who have the appropriate knowledge of the city and are willing to help with providing their walkability scores that could amount to sufficient data to be used as ground truth. It is also essential to consider that walkability is subjective, and people will have different opinions. Specifically, each user is entitled to their personal opinion regarding walkability scores (among other things). Everyone may and likely will have a different opinion on how walkable they may consider different locations. This expected variation can potentially lead to noticeable inconsistencies within the user labels. This challenge was addressed in two phases 1) developing a web-based crowd-sourcing tool to maximize the data calibration and accuracy of collected data, 2) using GLEPO to process the collected data further to ensure a fair distribution within localized relative scores and accurate globalized representation among all users. Accuracy of ground-truth data is crucial as it represents the population through only a small sample size, and as more data is collected, accuracy will likely improve.

The concept and processes behind GLEPO and the data collection methods used in this chapter can be applied to many other fields when it comes to utilizing opinion-based data as it allows for the collection of small and relative sets leading to reduced bias and presumptive measures, such as defining a hard-coded range for walkability. For instance, if users rank walkability scores for locations by assigning a number between 0-100, locations will get varying scores associated with them by different users. This assignment is dependent on how users define the numerical range and how this assignment aligns, or otherwise, with that of the researcher(s). Utilizing relativity instead of absolute scores allows for a reduction in the imposed cognitive challenge, which is done by assigning a representative score for each location where users use relative ordering by indicating if they consider a location more or less walkable than another location without giving a fixed numerical value. GLEPO will then take the responsibility of globalizing the relative ranks to reflect how they would rank concerning all other user submissions. This approach gives researchers the flexibility to maintain their rationality while preserving the integrity of users' opinions.

To tie in with the previous points, since ALF-Score relies on crowd-sourced data, it is likely that only a small amount of data may be available in the initial stages of the research. This crowd-source data will play a crucial role in two key ways: 1) to represent user opinion of a much larger population and 2) to represent nodes of a much larger geographical region. Based on the experimentation and results in this chapter, the ALF-Score pipeline has proven to accurately generate walkability models capable of estimating walkability scores for various cities using only a small set of user data.

Furthermore, the spatial resolution of walkability was significantly improved by utilizing road networks, user opinions and machine learning approaches that help generate point-level walkability scores of all nodes within a given road network, as opposed to area-based approaches found in some existing measures. This higher spatial resolution of walkability provides the much-needed depth for various analyses in many research studies that require fine-tuned and specific scores for various locations within proximity that are situated within the same DA.

An important factor about this research is its interdisciplinary contributions, specifically in Computer Science and Public Health. Interdisciplinary research studies are crucial and can provide practical applications to many important and immediate challenges at hand. Day-to-day users are among the intended users of ALF-Score, but they are not the only intended users. ALF-Score aims to be used as a tool that allows researchers in the public health sector, especially the epidemiologist, to understand better how walkability scores are associated with our health, such as physical activity, obesity, and diabetes. Technology is advancing every day, and what better use of it than to improve people's lives and health. This kind of practical interdisciplinary research can truly and positively impact the world.

Chapter 5

ALF-Score+: Personalization of a Predictive Walkability System

This chapter involves an in-depth definition of the ALF-Score+ walkability measure, which is an extension of the previously defined ALF-Score with a focus on embedding personalization based on user demographics and opinions. This work has been submitted for publication. The paper submitted for publication is slightly revised for flow in this dissertation document, and the introduction is shortened.

ALF-Score+ is an extension of ALF-Score, which utilizes user-defined and system-defined demographics to create individual profiles to develop profile clusters. The ALF-Score pipeline then uses user labels and profile clusters to further process and generate machine learning predictive models capable of estimating personalized walkability scores specific to each profile cluster. ALF-Score pipeline requires various data inputs. Since ALF-Score is a network-based walkability measure, an important step is to use road networks and other road characteristics, which ALF-Score+ also inherits.

A map database [Figure 5.1] is derived from road networks and POI inputs extracted from both OpenStreetMap and Statistics Canada to help enrich the feature set towards achieving high-resolution walkability scores. The introduction of user profiling and profile clustering into the ALF-Score+ pipeline helps assign users to distinctive clusters that could represent the majority of users within that cluster. Each cluster is then used to train specific machine learning models that best represent the users within that cluster and can estimate walkability scores influenced by users' opinions and demographics.

5.1 Cluster Profiling

People tend to have many similarities in their daily routines such as work hours, mealtime, profession, preferences or hobbies [141, 94, 52, 67]. For instance, many people work regular shifts between 9 am and 5 pm, with most taking their lunch break at or around noon. Furthermore, many of these patterns may be linked to specific demographics such as age, gender, or personal circumstances. For example, picking up children from childcare or school daily. With that in mind, cluster profiling aims to establish common points among various users that can be associated with how each user interacts with their surroundings and perceives walkability.

Each user submission contains five entries, with every entry maintaining metadata for its associated location, user's relative ranks, user-defined and system-defined demographics. Public IP addresses alongside submission timestamps are associated with each entry, and their combination provides unique identifiers distinguishing each submission. Various submissions associated with each user help will estimate user profiles

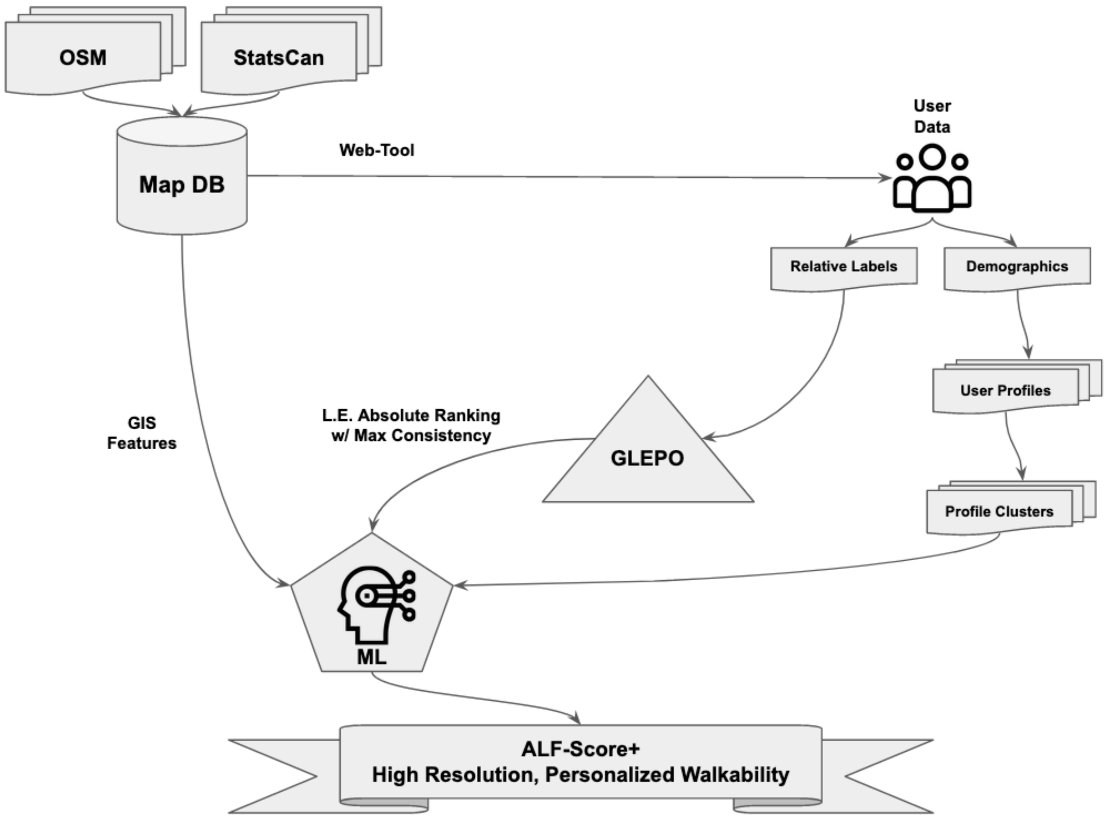

Figure 5.1: ALF-Score+ utilizes various GIS features such as POI and road network data (and derived features such as centralities and road embedding) and features engineered from their combination or subset. Aligned with GLEPO's linear extension of user labels, user demographics from two categories (user-defined & system-defined) are processed into user-profiles and then grouped into a few select clusters. k personalized walkability models are produced using the machine learning pipeline for k profile clusters, taking GLEPO's output alongside the profile clusters. Figure is drawn by the author.

related to each user.

The next step is to group these user profiles into clustered profiles. However, when it comes to clustering user profiles, an important component is determining the correct number of clusters to represent all participating users best. To this end, two various k estimation methods are used, namely the elbow method [87] and the silhouette coefficient [108]. Both methods utilize k-means as their base clustering algorithm, with the elbow method measuring the distortion over a various number of clusters. At the same time, the other compares the silhouette coefficient for each cluster. Lower distortion in the elbow method and higher silhouette coefficient in the silhouette method represent better fits using each method [Figure 5.2].

Another method used in this research to better understand the data and ensure clusters capture the right set of data is the t-distributed stochastic neighbour embedding (t-SNE). t-SNE is a common method typically used for visualizing high-dimensional data [26, 89, 142] by giving each point a location in a two or three-dimensional map. t-SNE was used to reduce 11 user features into a 2D space for visualization and verification. Each cluster is then associated with a specific colour determined by the k-means method. Once the correct number of clusters is determined, and t-SNE visualization has been verified, each user is assigned to a particular cluster. The user data is then segmented based on clusters for processing through the ALF-Score+ machine learning pipeline. The Figure 5.3 generated by t-SNE shows a 2-dimensional representation of 11 demographic features (6 user-defined and five system-defined) collected from the volunteer participants. A perplexity of 10 is used for this visualization; however other variations have also been experimented with. It is important to highlight that other techniques available for dimensionality reduction, such as Principal Component Analysis (PCA), have also been experimented with within this research.

5.2 An Extension to ALF-Score

To further explore how and where ALF-Score+ extends the original ALF-Score, first, the data processing stage is explored. Once data collection and cleaning steps are complete, the pipeline will proceed to the verification step to ensure data completeness. Specifically, since ALF-Score+ uses user-defined and system-defined parameters where each user may have submitted through a different device (PC, Mac, Linux, iOS, Android, etc.) or platform (Chrome, Safari, Firefox, etc.), the metadata will need to be checked to ensure validity and accuracy. To this end, data entries are converted into dataframes [93] and various data cleaning methods [123] were applied. These methods include but are not limited to addressing empty cells, removing duplicates, verifying appropriate data, confirmation of appropriate data formats and visualization of GEO-based data such as POIs and road networks to ensure accurate and matching geo-projections and transformations are used. Once the data has been cleaned and verified, entries are sent to the Generalized Linear Extension of Partial Orders (GLEPO) [6] to process users' relative ranking and generate ground truth for the machine learning pipelines. GLEPO works by structuring the data entries into separate lists of submissions. Anchor nodes for each submission (where applicable) are found. GLEPO's sorting algorithm associates these localized submissions to build a unified list that globalizes (among volunteer participants) users' local and relative rankings into absolute scores. A normalization function is used to keep scores between

0-100. Since there may be many 'right ways' of ordering the final list, GLEPO's path selection and submission selection are completely randomized with multiple passes of GLEPO to ensure the results are always highly consistent. Virtual link [6] was enabled for the experiments performed in the personalization chapter.

User demographics, profile data, labels and how each user ranks different locations are crucial input components of ALF-Score+. To stay true to users' opinions, ALF-Score+ inherits the incorporated consistency measure part of the GLEPO algorithm. Specifically, all submissions go through a pair-wise verification process that compares the five user-defined relative ranks of each submission in the original order submitted by the user to that of the same 5 locations in order of their appearance in the final globalized list (output of GLEPO). For instance, a submission for 5 locations $\{A, B, C, D, E\}$ may carry a relative ranking of $\{5_A, 4_B, 3_C, 2_D, 1_E\}$. These 5 locations may appear in a different order in a final list such as $\{..., A, ..., B, ..., E, ..., C, ..., D, ...\}$ and therefore, a globalized ranking of $\{5_A, 4_B, 3_E, 2_C, 1_D\}$. GLEPO's consistency measure determines if the processed ground truth remains true to that provided by the users. Furthermore, with each run of randomized GLEPO, a slightly different output is generated due to its randomized nature. Multiple passes of GLEPO are processed to generate the final globalized list of ranks that shows the most commonly associated ranks for all user-ranked nodes. Furthermore, Kendall rank correlation coefficient [2] is also used to further verify the consistency of the original relative ranks to that found in the final globalized list.

At this point, user labels are ready to be processed in the ALF-Score+ machine learning pipeline. The next step requires preparing the demographic data to ensure each feature is accurately and appropriately processed and accounted for. An important note is that each feature carries a different type of information, possibly of different data types (i.e. categorical, ordinal, numerical, etc.) and may have a different priority or weight, which entails converting some of the features. For instance, time entries are converted into their epoch representation (number of seconds since 1 January 1970), while certain other features, for example, the categorical and ordinal features, are encoded. One-hot encoding is used to encode features from user-defined and system-defined datasets such as profession, age group, whether user lives alone, operating system, system language, and other features. Once the encoding is complete, this user-derived feature set is ready to be processed by the cluster profiling pipeline. The machine learning pipeline will also utilize this user-based feature set as a combination of and in conjunction with other road-network-based features.

5.3 Experimental Set-up and Results

Three separate instances of data collection web tools associated with St. John's NL, Vancouver BC, and Montréal QC were created to help crowd-source user opinions and user demographics. This chapter focuses on data collected for the city of St. John's, NL. In the experiments performed to estimate the number of clusters (k), the elbow method and the silhouette coefficient were applied with a range between 1 and 21 clusters. `kmeans++` algorithm was used as their base as well as the core clustering algorithm, with `n_init` set to 10 (number of times the k-means algorithm will be run with different centroid seeds), `max_iter` set to 300 (maximum number of iterations of the k-means algorithm for a single run), `random_state` set to 42 (determines random number generation for centroid initialization to make the randomness deterministic), and `tol` set to $1e04$ (relative tolerance with regards to Frobenius norm of the difference in the cluster centers of two consecutive iterations to declare convergence).

Using the elbow method and the silhouette coefficient and based on the data collected from the city of St. John's, it was determined that six distinguished clusters [Figure 5.2] would best represent the volunteer participants. To further visualize these clusters and verify the association between users placed into the same cluster, 11 user features for all volunteer users were reduced to a 2-dimensional representation using t-SNE with `n_components` set to 2 (dimension of the embedded space) while `init` was set to `random` (initialization of embedding) with a `perplexity` of 10 (related to the number of nearest neighbours that is used in other manifold learning algorithms with larger datasets usually requiring a larger perplexity). Figure 5.3 shows the t-SNE visualization of crowd-sourced user demographics (user-defined combined with system-defined) with six distinguished clusters separated by six different colours.

5.4 Results

The study in this chapter includes $n = 40$ users (from the city of St. John's, NL). Among the users were $n = 20$ (50 %) women. The average age of participants was 48.6 (standard deviation = 17.1). The most commonly reported walkable distance was 800-1000 meters. Ten (25 %) of participants were living alone, while 14 (35 %) participants had children living in their homes, with the average number of children being 2.6 (standard deviation = 1.2). Finally, the most commonly reported professions were Retired $n = 8$ (20 %), Professor $n = 4$ (10 %), and Nurse $n = 4$ (10 %). Table 5.1 shows an overview of the user data.

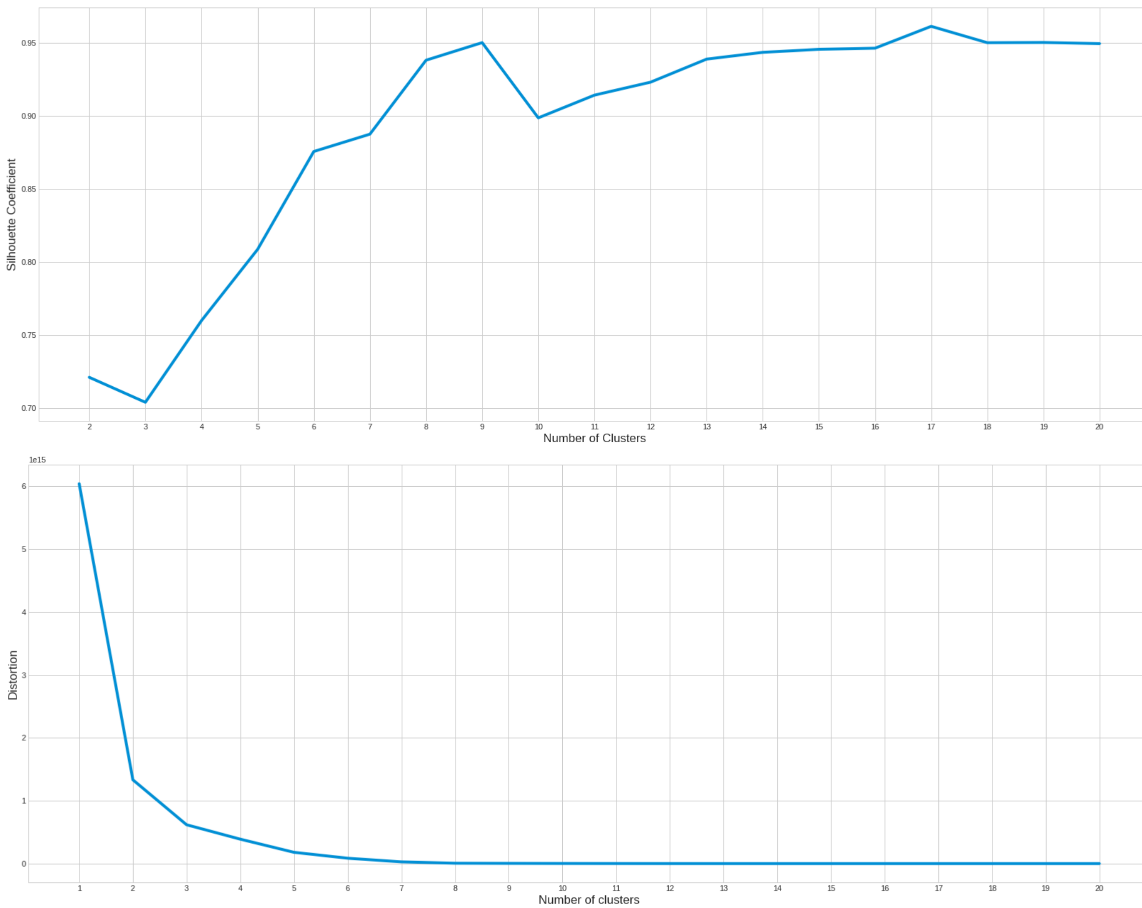

Figure 5.2: The Elbow method and silhouette coefficient determine the most appropriate number of clusters. In the top figure, it can be observed the silhouette coefficient increases between 3 clusters and 9 clusters, with 6 clusters being pivotal with a very small variation with that of 9 clusters. The bottom figure, elbow method, shows that the k-means distortion value drops significantly as the number of clusters is increased to 3, with the distortion value converging around 8 clusters with 6 clusters being pivotal (similar to that of the top figure) with a very small variation with that of 8 clusters. (Generated by Matplotlib [61])

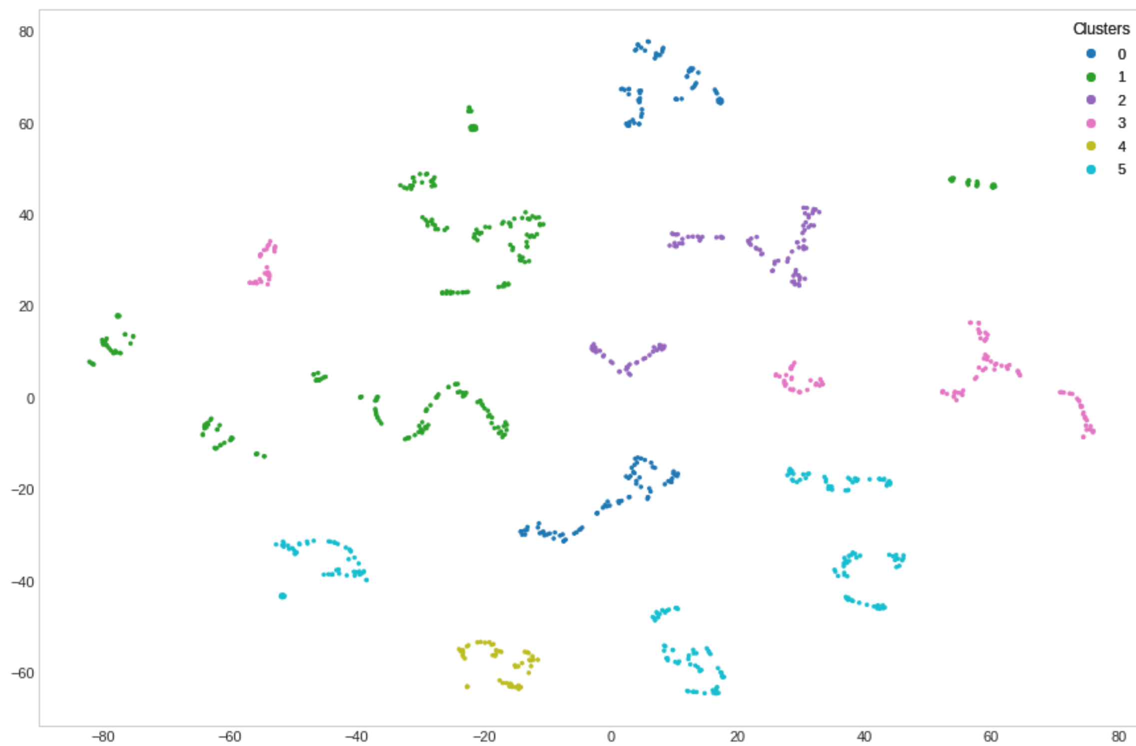

Figure 5.3: 2-dimensional t-SNE representation of user entries on demographics and system features. Eleven features are reduced to 2 features. The perplexity is set to 10, and the colours are associated with 6 clusters determined by the k-means algorithm. (Generated by Matplotlib [61])

Table 5.1: User submissions and feature statistics for the city of St. John's. Each submission maintains five entries containing five unique locations. Different submissions may randomly include the same location.

Number of	***Value***
Submissions	210
Entries	1050
Unique Locations	895
Estimated Participants	40
Demographic Features	6(+1)
Systems Features	5(+1)
Region	St. John's

ALF-Score+ maintains accuracy of 90.48% on average over at least 35 separate experiments with the best case of 93.70%. Specific accuracy of each cluster varies within each run (e.g. Table 5.2) and changes through different experiments; however, ALF-Score+'s overall accuracy on average is consistently and noticeably higher than that of ALF-Score. This improvement in accuracy is attributed to the user-focused approach of ALF-Score+. It is also essential to highlight the significant drop in the size of the training sets compared to the original used in ALF-Score. Although a smaller number of data entries in each training set was used to train models specific to each cluster profile, the trained models captured more in-depth characteristics of each profile cluster due to a better alignment and synchronicity of data entries in each cluster. It is believed that ALF-Score+ accuracy can be improved significantly with more data. In the previous chapter detailing ALF-Score [6], random forest [8] regressor was deemed a better choice among other methods. ALF-Score+ inherits this regressor. Furthermore, Mean Absolute Error (MAE) and Root Mean Squared

Error (RMSE) [24] were both used to measure how close the fitted line is to the actual data points' labels.

Table 5.2: Accuracy for the ALF-Score+ pipeline has increased to an average of 90.48%, with a best-case of 93.70% compared to that of ALF-Score. It can be observed that although each cluster uses a smaller data size for training since the set is highly focused on a single cluster of users that are deemed similar, the estimation accuracy is maintained and improved. Accuracy is expected to improve with more data.

Cluster #	***Training Size***	***Test Size***	***Accuracy***
0	112	23	93.70%
1	131	27	90.20%
2	278	56	91.42%
3	209	42	88.46%
4	149	30	86.58%
5	112	23	92.49%

Furthermore, a clear variability in the results of ALF-Score when compared to the personalized models trained on specific profile clusters generated by ALF-Score+ can be observed. Figure 5.4 shows the Pearson correlation coefficients [14] between ALF-Score generalized walkability predictions and ALF-Score personalized walkability predictions for all six profile clusters generated by ALF-Score+.

5.5 Analysis

One of the most critical components of this research is user demographics (user-defined and system-defined). However, selecting the right features to collect has been

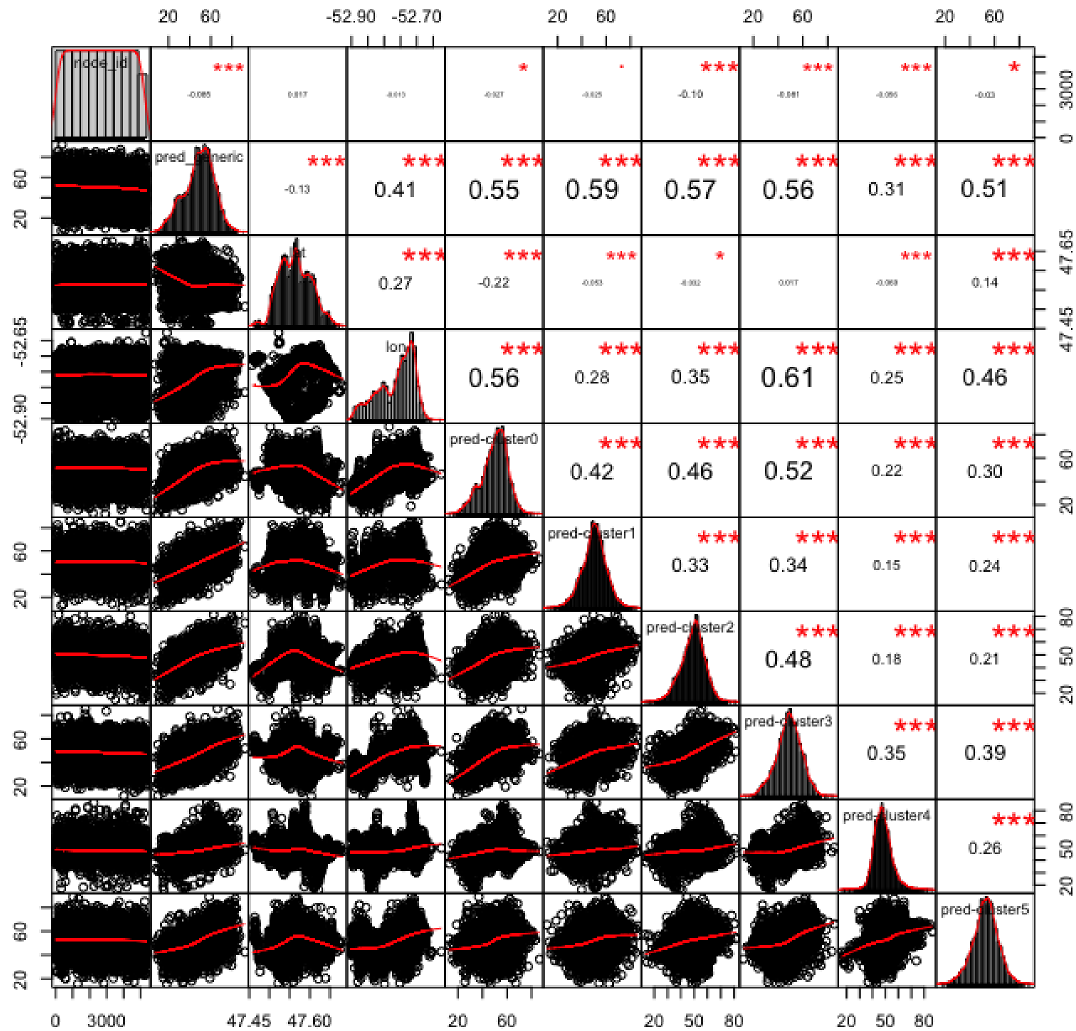

Figure 5.4: Histogram Pearson correlation coefficients for the generalized ALF-Score prediction scores and the personalized ALF-Score prediction scores for all six profile clusters generated by ALF-Score+. (Generated by RStudio [103])

detrimental to the success of this research. Various experiments have been performed outside of those stated here that considered different variations of user data, with some key features removed or encoded differently. The results were not feasible and led to a strong case of the importance of selecting the right features and how well the data is cleaned and prepared. Ultimately, in most research studies similar to this that heavily rely on data, specifically crowd-sourced data, not only does selection of the appropriate features matter, the number of data samples, their consistency, relevancy and accuracy will also play critical roles towards the succession of the research. Furthermore, when it comes to user-defined data, measures must be taken to ensure users are well informed of what they are expected to do and what they should expect. The collection of user data is a time-consuming and challenging task. In case of a lack of necessary features or incorrect data collection, the recollection of user data is an even more time-consuming and challenging task.

Through visual inspection of the elbow method and silhouette coefficient, it was determined that a 6-cluster solution is optimal (see Figure 5.2) based on the user data collected from the city of St. John's. Examination of the 6 clusters suggests that each user profile cluster appears to be consistent with how different groups experience walkability. The list below shows the predominant groups represented in each cluster:

- Cluster 0: Young men with no children, partly living alone, either students or unemployed, perceive walkable distances as less than 800 meters. (see Figure 5.5 right)
- Cluster 1: Women in their 30's who don't live alone and have children, and perceive walkable distances as greater than 1000 meters.

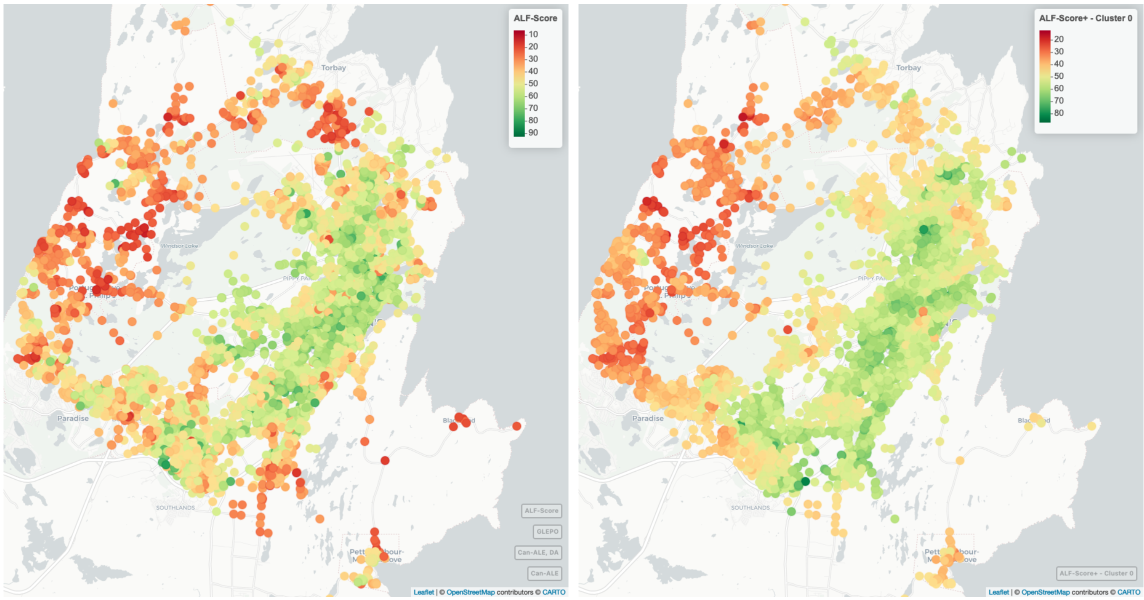

Figure 5.5: Comparison between the generalized ALF-Score (left) and ALF-Score+ for cluster 0 (right). A clear variability can be observed that is dependent on the demographics of cluster 0. (Generated by RStudio [103].)

- Cluster 2: Students in their 30's who do not live alone and do not have children, perceive walkable distances as greater than 1500 meters.
- Cluster 3: Female professionals in their 20's and 30's who do not live alone and have no children, perceive walkable distances as greater than 1500 meters. (see Figure 5.6 left)
- Cluster 4: Professionals and retired people in their late 40's who perceive walkable distances as greater than 1000 meters.
- Cluster 5: Retired professionals in their late 60's up to 80 who perceive walkable distances as less than 800 meters. (see Figure 5.6 right)

Figure 5.5 shows a comparison of the results generated by ALF-Score (left) and ALF-Score+ (right and specific to cluster 0). The ALF-Score results are produced

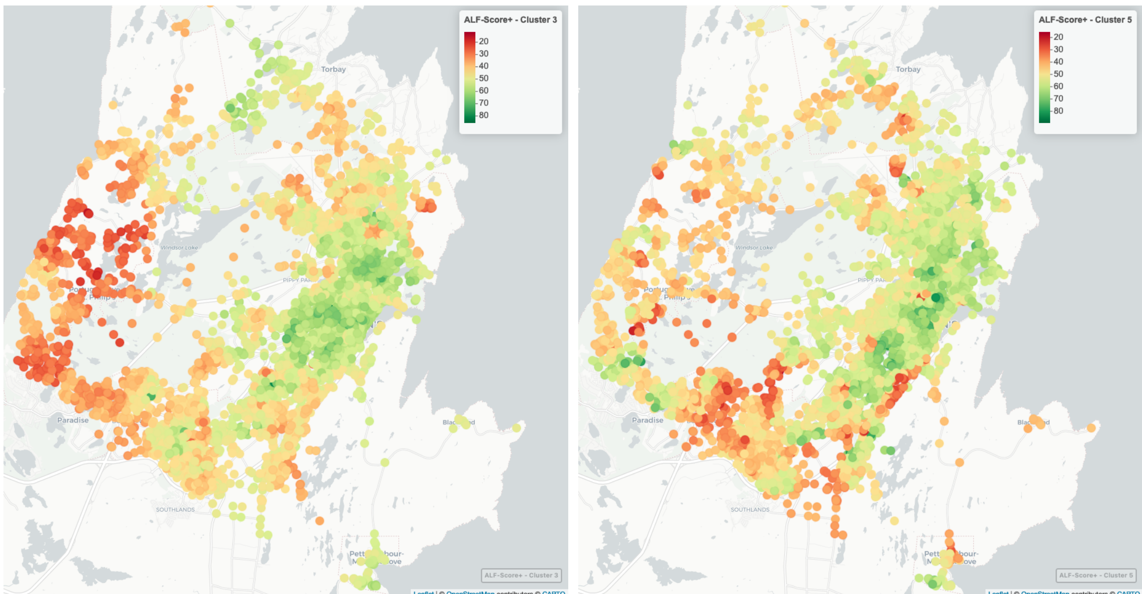

Figure 5.6: A comparison between cluster 3 (left) and cluster 5 (right) of the ALF-Score+. A noticeable variability in the walkability of the two clusters can be observed. Cluster 3 appears to be mainly walkable around and within the downtown area with no significantly walkable (dark green) locations. Cluster 5 appears to have higher overall walkability (darker green); however, with increased coverage of more walkable areas around the southeast region. Additionally, fewer locations with highly unwalkable scores (dark red) can be observed in cluster 5. (Generated by RStudio [103].)

without explicitly using the clustering or user profile cluster-specific models. Visual comparison of the maps shows that in cluster 0, more areas are considered walkable within the three main cities of St. John's, Mount Pearl and Paradise, with most "out-of-town" areas regarded as not walkable. This observation is aligned with the findings of researchers who have determined younger people tend to drive less and walk more [38] due to various reasons [70] such as lower income, higher living costs and being more health-conscious.

In Figure 5.6 differences in the walkability predictions of two different clusters produced by ALF-Score+ can be observed. Cluster 3, which represents female professionals in their 20's and 30's, shows an increase among less and highly-walkable locations with only key areas (i.e., around downtown St. John's and some local parks) considered as highly walkable. However, from cluster 5, which represents retired and older adults, it is observed that, on average, the city is predicted to be more walkable with a decrease in unwalkable areas. Simultaneously, cluster 5 appears not to have many highly-walkable areas, other than a few small areas such as parts of the downtown region. However, more "out-of-town" areas are considered as walkable as opposed to cluster 3.

Further in-depth analysis of clusters 0 and 5 shows substantial variability in some neighbourhoods and communities as personal characteristics, demographics and preferences of each cluster profile change. In cluster 0 [Figure 5.7 top], which represents young men living alone, most of the area is considered reasonably walkable. In cluster 5 [Figure 5.7 bottom], which represents retired and older adults, the entire area is assigned with low walkability scores. A clear variation between ALF-Score+ generated scores for different cluster profiles can be observed through these analyses. While both measures (ALF-Score and ALF-Score+) follow the same pipeline and logic, each cluster's characteristics influence the final decision-making process of ALF-Score+, making its results unique and personalized. The ALF-Score+ method can capture and distinguish between these variations among users and user profiles to generate highly personalized walkability scores for users based on their profile cluster characteristics in addition to an overall predictive accuracy improvement to that of ALF-Score.

5.6 Discussion

ALF-Score+ is a walkability measure that builds walkability models capable of generating personalized walk scores based on each user and their profile. The ALF-Score+ pipeline needs to identify users and link them with the appropriate cluster profile that best represents their criteria to produce personalized walkability scores. Each user is then associated with the appropriate model to generate the relevant, personalized walkability scores. However, to profile users and build cluster profiles, clustering algorithms require specific user traits to distinguish various users and user groups. A straightforward approach to link users to appropriate cluster profiles is by asking the users to provide specific demographics about themselves, such as age group and gender (similar to the method used in this research to collect volunteers' data). However, in a production environment, an important challenge is raised due to privacy reasons. Unlike the data collection process in this study that involves volunteer participants willing to help with this research, everyday users may not be willing to disclose certain personal demographics about themselves, such as their age group or gender. It

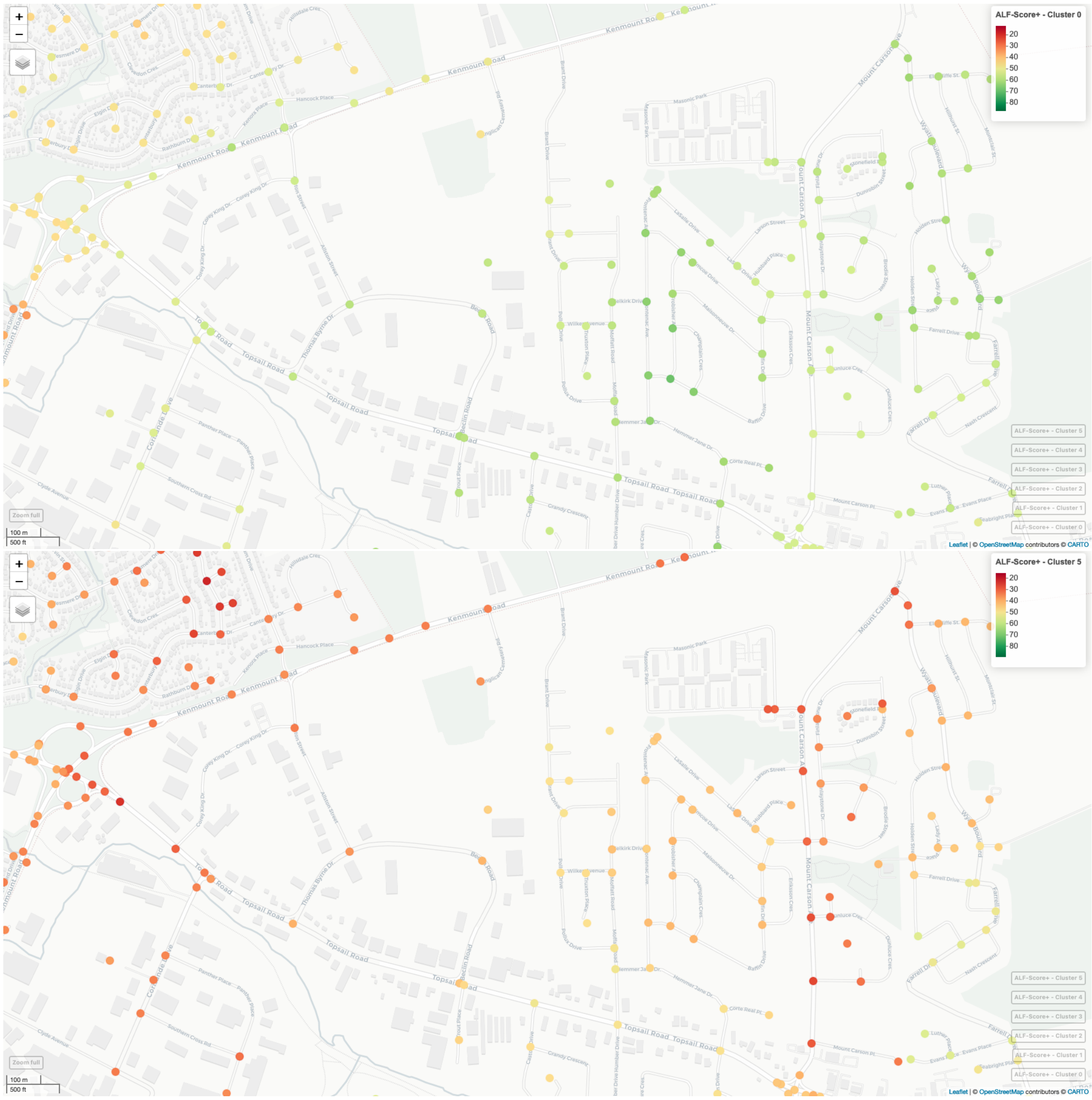

Figure 5.7: Personalized ALF-Score+ walkability for clusters 0 (top) and 5 (bottom) for the Topsail road, Kenmount Road and Mount Carson areas. A clear variation between the preferences of the two cluster profiles is observed. (Generated by RStudio [103].)

is challenging to profile users and associate them with the most representative profile cluster without such information. An appropriate step will need to be taken to ensure this important challenge is addressed accordingly. Building a reverse-profiling algorithm could be considered that instead of using user demographics, it would ask users to rank several locations based on their personal opinion. These rankings would then be reverse-profiled to estimate users' demographics and link to the most appropriate cluster profile. This approach will require an existing knowledge of a sample group of users, their opinions, and demographics.

Although personalized walkability may not have been considered in previous measures, it is clear that personalization is essential and should be considered when creating walkability measures. In this chapter, ALF-Score+ was introduced, which uses user demographics and system data to build user profile clusters that distinguish prominent profiles to build personalized machine-learned models that can accurately estimate spatially high-resolution walkability scores for individuals based on cluster profiles for each point within the road network. ALF-Score+ successfully achieved an accuracy of 90.48% on average with a best-case of 93.70%, with its generated scores accurately associated with the appropriate user groups.

It is crucial to keep in mind that volunteer data such as user opinion and demographics will significantly impact the outcome. Furthermore, algorithmic bias must be considered to avoid unintended use/misuse. Ensuring an unbiased selection of participants and inclusion of vulnerable and minority groups are crucial to providing a fair user representation. It is also equally important to highlight that while ALF-Score is meant for regular use, using profile clusters to derive specific walk scores could drive certain populations to certain areas. While this is unintended and could potentially be a non-issue, it can result in inequities across demographic groups if misused.

Chapter 6

ALF-Score++: Transferability of a Predictive Walkability System

This chapter involves an in-depth definition of the ALF-Score++ walkability measure, which is an extension of the previously defined ALF-Score with a focus on transferability and scalability. This work has been submitted for publication. The paper submitted for publication is slightly revised for flow in this dissertation document, and the introduction is shortened.

ALF-Score++ is the second extension of the ALF-Score pipeline, focusing on transferability. ALF-Score++ pipeline 6.1utilizes a map database containing road network data and POIs extracted from Statistics Canada [113] and OpenStreetMap (OSM) [91]. The map database feeds into two separate processes: 1) GIS feature extraction 2) user data extraction through the web-tool interface. The GIS feature extraction process extracts and generates the required features such as node lists, edge lists, various centrality measures, road embedding, and various POI features.

The output of this process is fed into the machine learning component as one of its three main input feature sets. The user data extraction process involves the web-tool interface that utilizes road data to feature various points on an interactive map where users provide their opinion and demographics. User data is broken into two separate processes, each resulting in a separate input to the machine learning component. The first process is the collection of users' opinions through the web tool, where users provide relative ranks for various points on the map. This process passes users' opinions to the Generalized Linear Extension of Partial Orders or GLEPO algorithm to convert users' relative ranks to a globalized rank among all submissions. The output of GLEPO is fed into the machine learning component as its second feature set. This input serves as the y label during the training and testing processes. The second process of user data revolves around users' demographics and is intended for personalization (excluded from this diagram). This process uses various clustering techniques and unsupervised learning methods to generate profile clusters. These profile clusters represent users deemed by the algorithm as similar. These profile clusters are then fed into the machine learning component as its third feature set. The machine learning component utilizes these three feature sets in conjunction with its internal transfer learning process, and the general flow is as follows. GIS features form a feature set associated with specific locations that have their ranks available through the GLEPO algorithm as its y label. These models will be trained on the data from only one specific city. The first round of models trained through a deep neural network technique will transfer their knowledge to the second round of training. Transfer learning utilizes appropriate layers (mostly top layers) within these models while replacing the output layer. The new data used in the transfer learning process will include features and user opinions from a second city. The output will be a more generalized model capable of transferring its knowledge to cities never seen before during its training process. The personalization process utilizes this transfer learning approach to do the same task but on each separate profile cluster, resulting in multiple models capable of predicting personalized walkability scores for cities seen or never seen by the algorithm. These personalized models will be generalized over users from multiple cities but are personalized to their demographics.

6.1 Data Preparation

The first step is to gather the feature set that includes POI, road embedding, and road network data to prepare the map database. The POI data is available freely through OpenStreetMap (OSM) [91]. Overpass-Turbo [100] was utilized with the help of a customized extraction code to extract OSM POIs from 53 unique amenity categories. Once complete, a new algorithm was devised that creates POI-based features for all nodes within the network. Below is an example of a single POI contained within a GeoJSON file extracted from OSM through Overpass-Turbo. Each POI point is divided into two parts: 1) description 2) geometry. Description contains the type and properties of the point, while the geometry contains the location's type as well as its coordinates:

- "type": "Feature", "properties": "@id": "node/1401297904",
  "amenity": "fire_station", "name": "Caserne 29 Rosemont"
  "geometry": "type": "Point", "coordinates": [-73.5762681, 45.5453509]

The role of the POI2Features algorithm (Algorithm 9) is to prepare the segregated

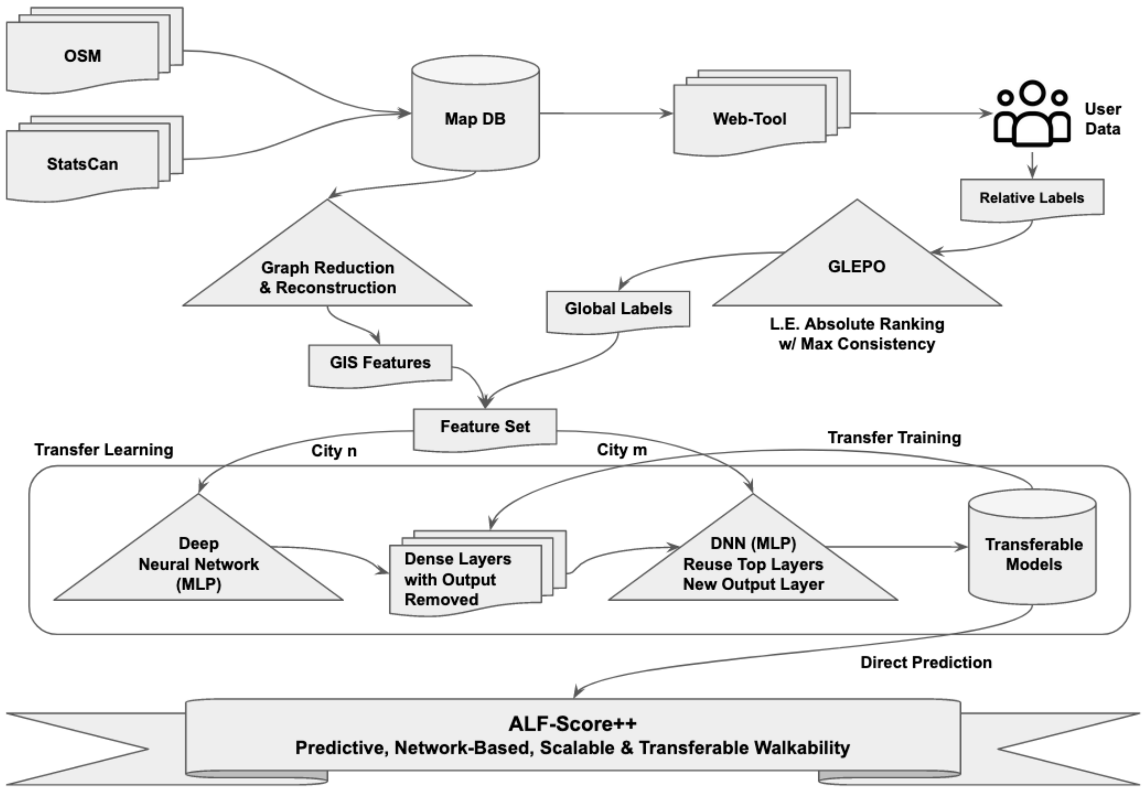

Figure 6.1: ALF-Score++ utilizes features similar to ALF-Score and ALF-Score+ such as road network structure, POI, centrality measures and road embedding. GLEPO's linear extension of user opinions that produces a global view of relative opinions is then aligned with the features as an input to the machine learning processes. Models trained by ALF-Score++ can be applied to cities seen or unseen by the algorithms during the training processes. Walkability estimates produced through trained models will have a high spatial resolution, represent users' opinions, and provide a better insight into different regions and neighbourhoods. (Figure is drawn by the author.)

POI data based on each specific node within a given road network. Ten separate distance ranges are created for each POI category. Each POI is represented by a node on the road, and the number of its appearances within these ten different distance ranges to every node within the road network is measured. Based on 53 amenity categories, a POI feature list containing 530 feature columns and n rows for the number of unique nodes in the road network was produced. Below is an example of one possible POI feature header structure followed by an example of a single entry for a unique node:

- node_id lon lat bar_200 bar_400 bar_600 ... bbq_200 bbq_400 bbq_600 ...

  317 -73.57113438 45.51020696 0 6 12 ... 14 11 12 ...

Algorithm 9 POI2Features

Input: GeoJSON POI List — List of Categories, Road Network Node List

Output: POI feature list for all nodes in road network within the range of 200m - 2,000m

initialize $feature\_list$ as an empty list
set headers in $feature\_list$ for 10 distance ranges
for every node in node list **do**
    get node latitude and longitude
    for every poi in poi list **do**
        get poi latitude and longitude
        find distance between node and poi
        if distance is falls within the predefined range **then**
            add 1 to node's feature in $feature\_list$ for the specific range
return $feature\_list$

While road network data is available freely from both OSM and Statistics Canada

[113], Statistics Canada was chosen to extract the road networks data in the form of ArcGIS Shapefile [40]. The 2016 Shapefile for Canada was extracted, which is the most recent Census year available. Furthermore, QGIS, [98] which is “a free and open-source cross-platform desktop geographic information system application that supports viewing, editing, and analysis of geospatial data” was used to extract road networks specific to four different cities of St. John’s NL, Montréal QC, Vancouver BC and Kingston ON from a single and large Shapefile containing the road network for entire Canada. All of the individual city sub-networks were further processed to build specific node lists, and edge lists that are used in the pipeline. “shp2graph” package [78] through R [99] was utilized to generate node lists and edge lists for various road networks which have been stored in the form of graphs. It is important to mention that the coordinate systems may differ depending on the data source and the format. For instance, some formats may present coordinates in UTM [114] or WGS [49] while others may present them in different coordinate systems. Appropriate conversions, projections and transformations, where applicable, may be required. As is the case with many research studies, it is crucial to maintain a unified unit of measurement throughout the research to avoid any unwanted disasters [73]. Individual city sub-networks are also processed to generate various complex network features such as different centralities and road embedding features for all road networks. Furthermore, graph reduction and reconstruction techniques [5] may be applicable when working with large networks. Additionally, the edge list for each city is processed through Cytoscape [30], which is “an open source bioinformatics software platform for visualizing molecular interaction networks and integrating with gene expression profiles and other state data”, to generate a list of network features. To generate

road embedding features, edge lists are processed through Node2vec [55], which is "an algorithm to generate vector representations of nodes on a graph". All features that are not numerical go through an encoding process called one-hot encoding to prepare the features necessary for the machine learning pipeline.

The next step needed to prepare the pipeline is the application of Generalized Linear Extension of Partial Orders or GLEPO [6]. GLEPO requires a few datasets such as users' opinions, node lists and a distance matrix connecting all the nodes within the road network. The overall GLEPO pipeline involves multiple algorithms such as *seperateBySub* which is used to prepare users' opinions into subsets that are suitable for processing. Various other algorithms such as *calculateDistance*, *FindDistance*, *addToSorted*, *FindVLink*, *RandomizeInsertion*, *normalize* and *GLEPO* are also used to further process user opinion and to convert their relative rankings into generalized scores which are globalized among all opinions. The output of the GLEPO pipeline is a generalized list of users' opinions which can be fed into the next component of the pipeline. This globalized list is crucial to the entire structure of ALF-Score as it plays an important role in ground truth used as machine learning labels.

6.2 Experiments

There are three main experimentation scenarios used to guide this research forward, and they are 1) the matching approach, 2) the combined approach, and 3) the zero-user-input approach.

Matching approach is a scenario where users' opinions from one specific city are used to train and test models that are applied to the same city. This approach creates an important base for the ALF-Score++ machine learning pipeline. It focuses on testing the feasibility and accuracy of the pipeline derived from users' opinions and feature sets belonging to the same city. For instance, using users' opinions and feature sets collected for the city of St. John's, NL, to train and test models on St. John's road network. Furthermore, this approach is used to test the scalability of the pipeline for very large road features and user opinion datasets.

Combined approach is an approach that focuses on the transferability of ALF-Score++ models. This approach uses data from multiple cities to train and test various models. These models can then be applied to cities either seen by the pipeline through the training process or cities never seen by the algorithms before. This approach aims to test and verify that transfer learning can improve the overall generalization of the models while broadening models' applicability. This scenario has multiple variations, specifically how training and test sets are selected. Two of the commonly used variations are random and semi-random selections. In the random selection, a typical 80-20% training-test distribution is used that includes data from two cities. In the semi-random approach, for example, a 50-50-100 distribution, the entire data from the second city is used in the training set alongside 50% randomly selected data from the first city, while the remaining 50% is maintained for the test set. There are other variations, and the models are tested in both cities.

Zero-user-input approach aims to use models that are previously trained on a specific city (or cities) to predict walkability scores of other cities. This approach uses predefined features and pre-trained models to generate walkability scores for points in cities never seen by the algorithms and is very important to help identify how applicable and transferable are the pre-trained models to data from unseen cities and whether the patterns observed and learned in different cities are similar and transferable to one another. Models in this scenario could have been trained on either a single city or be multi-city models. The models in this scenario can be applied to cities never seen before during the training process and those previously trained, making them very versatile.

6.3 Transfer Learning

ALF-Score [6] pipeline has been tested for various supervised and semi-supervised approaches and methods. However, the most promising shallow models are random forest, support vector machine (SVM), and decision tree, whereas the most promising deep model, was multi-layer perceptron neural network (MLP). These methods generated reasonably accurate results while random forest performed the best among all. Random forest was set up with 100 estimators (the number of trees in the forest), while the maximum depth of the tree was not limited. Most other parameters such as the number of jobs to run in parallel, the number of features to consider when looking for the best split and bootstrap sampling were set to scikit-learn's [96] default parameters. Random forest is an ensemble approach. Ensemble learners aim to use multiple weak learners to build a strong learner that performs very well, taking a divide and conquer approach. Random forest uses a standard decision tree that could be considered its weak learner. Multiple of these trees will then form a forest that can perform better as a group. Table 6.3 shows the difference in observed error between random forest using 100 weak learners and a single decision tree. Random forest performs significantly better. There are two specific functions in scikit-learn's random forest that, although not specifically labelled as transfer learning approaches, are geared toward transferring previously learned knowledge. These functions are *warm_start* and *partial_fit*. Warm start aims to fit an estimator repeatedly over the same dataset but with varying parameters. Using this approach, one can look at various parameters to improve performance while reusing the model learned from previous parameters to save computing resources and time. Warm start is typically used for fine-tuning the model parameters. On the other hand, partial fit aims to provide an online machine learning approach while maintaining fixed model parameters between calls by allowing for new data in every call, also known as a mini-batch. Online machine learning updates the predictor in sequential order as new data becomes available. This approach is the opposite of batch learning, where the training dataset never changes.

Furthermore, MLP was used to utilize deep learning, specifically as a doorway to transfer learning. In this research, transfer learning was approached under the assumption that previously trained models of similar tasks are available (through ALF-Score). The first step to initiate the transfer learning process is to import three sets of data: 1) previously trained MLP models, 2) GIS features such as POI, centrality and embedding features associated with the new city, 3) user data such as users' opinions and demographics associated with the new city. After a successful import of data, the usual data processing and preparation steps will need to be taken, such as dealing with incomplete entries and processing features through one-hot encoding, where applicable. TensorFlow [1] was used to facilitate MLP training and transfer learning processes and is "a free and open-source software library for machine learning and artificial intelligence" that enables us to apply various techniques with very efficient implementations. To set up TensorFlow for transfer learning, the first step is to create a Sequential model. Next, multiple Dense layers can be added as hidden layers. Each dense layer takes in a unit value and an activation function. The unit value, which is a positive integer, defines the dimensionality of the output space. The activation function [34] acts as a trigger based on the input values and fires only if the input exceeds a set threshold. In this setup, ReLU activation function [35] was used. If the input is negative, ReLU returns 0; otherwise, it will return the actual input. For the last layer that acts as the output layer, the unit is set to 1. It is common to see the Softmax activation function used in classification tasks for the last dense layer. However, since the task in this work is a regression problem, the linear activation function is used. At this point, the model needs to be compiled with the loss function, optimizer and metrics set. The loss function was set to *mean_absolute_error*, the optimizer to *adam* and the metrics to *mean_squared_error*. The last step is to fit the model by feeding the feature set followed by the labels and setting the number of epochs and the size of the validation split. Depending on the batch size, the number of epochs and the data size, the process may take a while. This process will result in a model trained on the {features, label}: {x, y} set.

# of Dense Layers	***Output Shape Range***	***Total Parameters***	***Optimizer***	***# of Epochs***
2	8-16	10,945	Adam	200
5	50-300	418,301	Adam	300
11	50-1,000	2,673,301	AdaMax	400
12	50-800	2,303,001	AdaMax	600

Table 6.1: Various deep neural network settings under which MLP and transfer learning were experimented with.

ALF-Score uses various combinations of dense layers and the number of neurons. Table 6.1 shows a brief set of example settings that are experimented with. To transfer the model generated/imported as above, the first step is to create a new Sequential model and copy the hidden layers desired from the original model over to the new model. In this process, the output layer is excluded. It is also important to ensure all transferred layers are frozen by setting them as non-trainable so the algorithm will not modify them. Next, a dense output layer is added to the new model with the unit set to 1 and the activation function set to linear. Finally the loss function is set to *mean_absolute_error*, the optimizer to *adam* and the metrics to *mean_squared_error* and compile and fit the new model. After a few iterations/epochs, unfreezing the reused hidden layers is an option to allow backpropagation to modify and fine-tune them and re-evaluate the performance. It is also suggested [53] to reduce the learning rate to avoid changes in fine-tuned weights when these layers are unfrozen. A good rule of thumb is to train the model for the new task for a few epochs while the reused layers are frozen. Then unfreeze the reused layers and continue to train with a reduced learning rate for further fine-tuning these layers. The learning rate is always an important variable to consider in transfer learning. If the learning rate is set too high, training may diverge, and if the learning rate is set too low, the processing speed will be very slow to reach convergence. Experimenting with various parameters may be an excellent approach to find the best setting that may be most appropriate in a particular task.

It is important to note that if the input data of the new task does not have the same shape structure as the data used in the original task, they will need to be processed to match the original size. However, that is not the case with ALF-Score++

since the structure of feature sets used for training various models remains the same. Additionally, according to Géron [50] "...transfer learning will work best when the inputs have similar low-level features". It is a good idea to replace the appropriate layers for the new task as they will likely be very different from the original task. For example, a voice recognition task will still need to produce the correct and valid words associated with its output layer. However, top layers may need to recognize words spoken by different people. In this case, reusing the top layers may be more useful [53]. Furthermore, the output layer of the original model will be replaced since it is no longer useful as an updated output using the new input is expected. It should be noted here that it is suggested that the more similar the tasks are, the more hidden layers may be used. For instance, in the case of ALF-Score++, since the original task is very similar to the new one, all hidden layers may be kept with only the output layer requiring replacement.

6.4 Results

In this work, ALF-Score++ successfully achieved the capability of transferability. First, using the newly collected user opinion data for the city of Montréal QC, a consistency of 99.6% was achieved during the GLEPO processing stage. While various feature combinations and machine learning techniques were experimented with, the lowest achieved MAE prediction error (matching approach) using a shallow model was achieved using random forest at 11.87 units (Figure 6.4 top left). In comparison, MLP was the best performing deep model with an MAE error of 13.87 units on a scale of 0-100. Table 6.2 shows an overview of the user data while Table 6.3 highlights some of the techniques and feature combinations used to generate ALF-Score for the city of Montréal using users' opinions collected from the same city.

Table 6.2: User submissions and feature statistics for the city of Montréal. Each submission maintains five entries containing five unique locations. Different submissions may randomly include the same location.

Number of	***Value***
Submissions	199
Entries	995
Unique Locations	824
Estimated Participants	40
Demographic Features	6(+1)
Systems Features	5(+1)
Region	Montréal

Technique	***POI***	***POI + Network***	***POI + Embedding***	***Network + Embedding***	***All***
Random Forest	19.65	18.20	17.13	15.47	**11.87**
MLP	26.65	24.08	23.44	23.56	21.91
SVM	29.03	31.04	29.78	23.63	21.74
Decision Tree	21.65	31.87	34.23	24.45	21.49

Table 6.3: Exploration of various machine learning techniques and feature combinations over an 80-20 data split (matching approach) for the city of Montréal, QC reflecting their top-performing accuracy. Results represent MAE error over a range of 0-100 units.

As explored in the background section, the goal of transfer learning is to take advantage of previously trained models. For instance, models trained based on the data

MLP	***POI***	***POI + Network***	***POI + Embedding***	***Network + Embedding***	***All***
St. John's (STJ on STJ (100%))(1)	27.55	26.22	22.23	21.91	17.88
Montréal (MTL on MTL (100%)(1)	26.65	24.08	23.44	23.56	21.91
STJ on MTL (100%)(3)	n/a	n/a	n/a	n/a	32.44
STJ on STJ (50%) + MTL (100%)(2)	26.87	25.10	23.55	19.31	15.77
STJ on STJ + MTL(2) ***(rand 80-20)***	25.87	23.74	21.45	20.23	**14.12**
MTL on STJ (100%)(3)	n/a	n/a	n/a	n/a	33.89
MTL on STJ + MTL(2) ***(rand 80-20)***	25.11	22.23	21.67	20.11	16.23
MTL on STJ (100%) + MTL (50%)(2)	27.67	24.86	14.43	21.51	16.73
MTL on STJ (100%) + MTL (80%)(2)	24.84	20.17	19.92	18.36	**13.87**
MTL on STJ (100%) + MTL (20%)(2)	29.66	25.34	25.73	22.89	18.34

Table 6.4: Exploration of the three experimentation approaches (1) Matching, (2) Combined and (3) Zero-user-input over five different feature combinations and two different data split approaches based on data from the cities of St. John's NL and Montréal QC. Results represent MAE error over a range of 0-100 units.

for St. John's NL in previous chapters to essentially extract their learned knowledge that could be useful when applied to training new models (combined approach) for new cities. However, transfer learning is also useful for directly generating predictions for new cities without further learning (zero-user-input approach). Zero-user-input was the first transfer learning approach experimented with within this work using the best model trained on the data for the city of St. John's by the random forest method to predict ALF-Score walkability for the city of Montréal (Figure 6.4 top right), which resulted in a correlation of 0.4 compared to the predictions generated by a model that was trained purely on Montréal's user data (Figure 6.4 top left). Furthermore, the second approach of using previously trained models towards training new models (MLP) (Figure 6.4 bottom left) led to a much higher correlation of 0.79 compared to the models trained only on the data from one city. It is observed that this approach can well utilize the transfer of previously learned knowledge in conjunction with a new learning task which enables new models to identify additional patterns that may

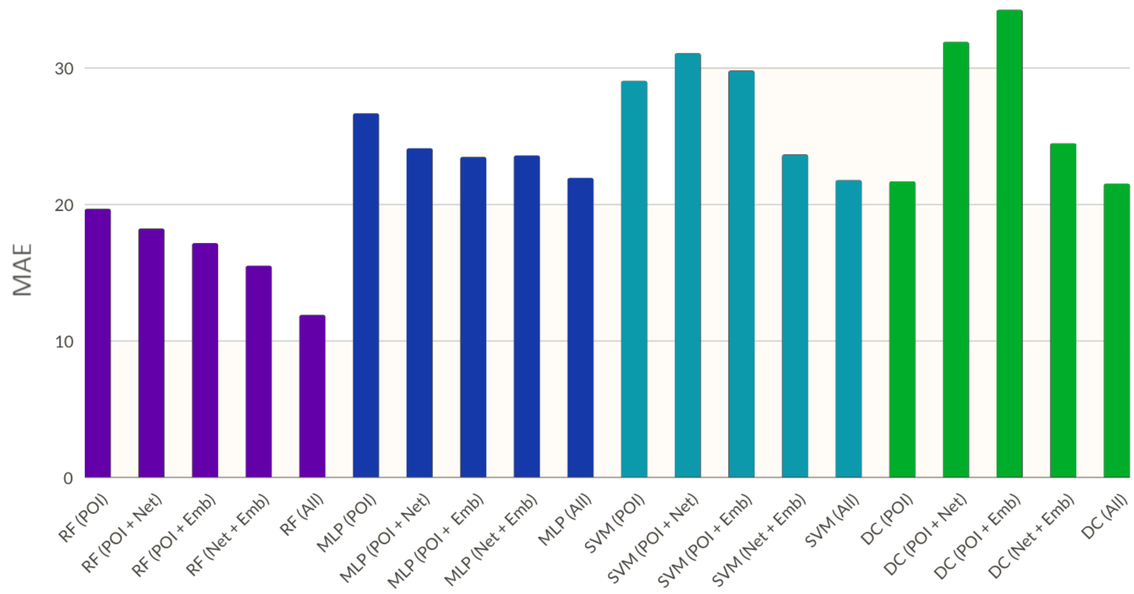

Figure 6.2: Experimentation results of four machine learning techniques over five feature combinations for Montréal, QC with an 80-20 percent data split. The bars represent MAE error over a range of 0-100 units. RF: Random forest, MLP: Multi-Layer Perceptrons, SVM: Support Vector Machine, DC: Decision Tree. RF provides the best performance overall. (Generated by Matplotlib [61])

have not been fully captured by models trained on small sets of user data.

It can be observed that among the top 150 features (out of 668 features), 128 of them belong to the road embedding feature list and account for all road embedding features, which highlights the importance of road embedding with regards to predicting walkability scores based on user-submitted ground truth. Additionally, among the top 150 features, only 14 belong to the POI feature list, which contributes to 530 features. Furthermore, among the top 150 features, six belong to centrality features out of the total of 10 centrality features.

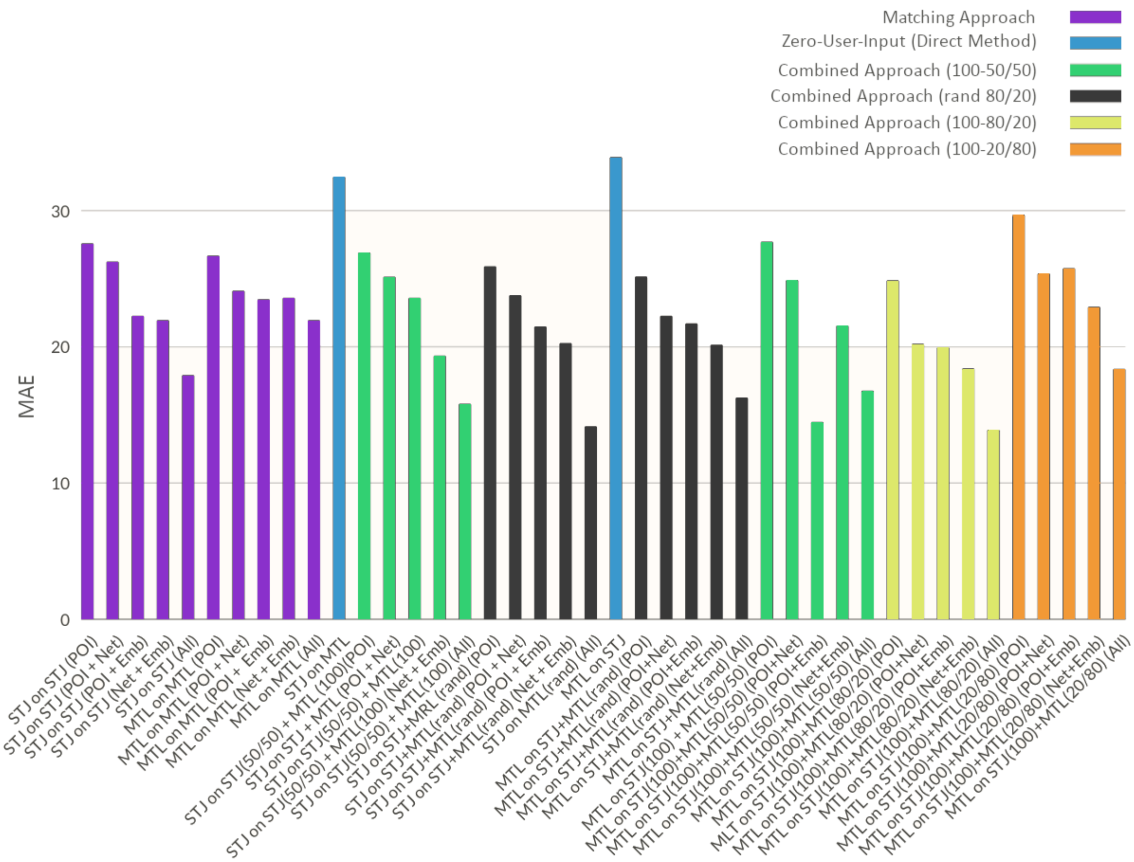

Figure 6.3: Exploration of 3 approaches (1) Matching, (2) Combined, (3) Zero-user-input. The combined approach is extensively tested with various conditions. One such condition is different ways of data split to better understand how the data affects the transfer of knowledge in transfer learning while providing solid training and test sets. The best performance was observed to be generated through a completely random selection into an 80-20 percent split. MTL on STJ reflects score predictions for Montréal based only on a model trained on St. John's. MTL on STJ+MTL, on the other hand, reflects score predictions for Montréal based on a transfer-learned model trained on both St. John's and Montréal. (Generated by Matplotlib [61])

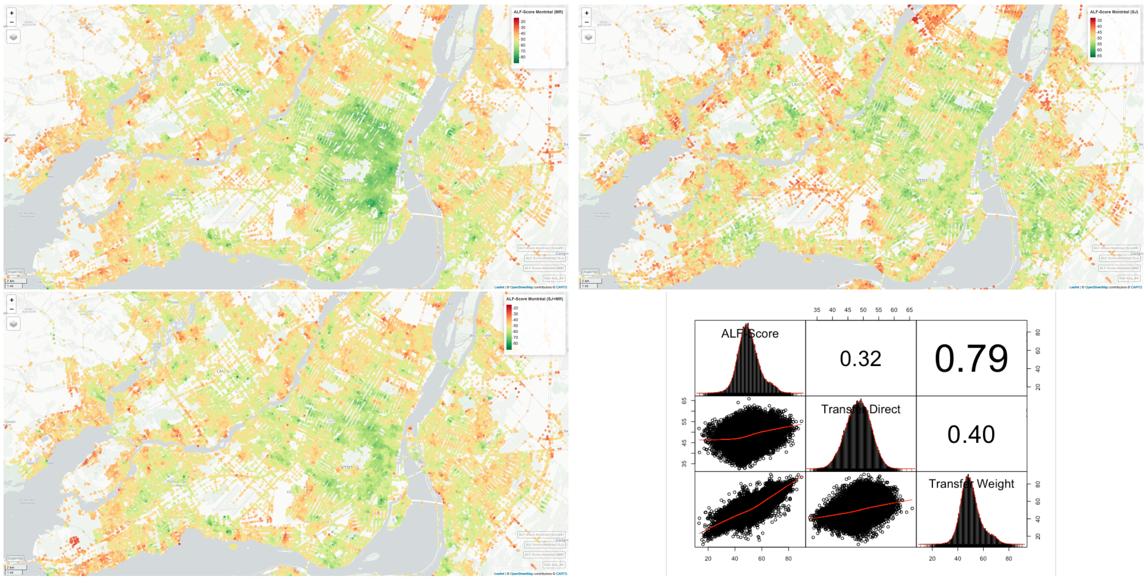

Figure 6.4: Walkability results were produced by three separate variations of ALF-Score and ALF-Score++ for the city of Montréal, QC and their correlation. Top left: predictions based on a model only trained for Montréal's user data (matching approach). Top right: predictions based on a transferred model only trained on St. John's user data (zero-user-input approach). Bottom left: predictions based on a model trained for Montréal's user data while having the previously trained weights for St. John's user data transferred in its transfer learned training process (combined approach). Bottom right: correlation between the three variations. The road network for Montréal maintains over 76 thousand nodes. ALF-Score walkability scores range between 0-100 units. This range can be adjusted if needed. (Generated by RStudio [103])

The road embedding features account for 0.778486799 importance over 128 features while representing only 19% of the overall features. The centrality features account for 0.039919843 importance over ten features, and the POI features account for 0.169245465 importance over 530 features while representing over 79% of the features.

Eccentricity accounts for the highest centrality importance among the ten features; however, it contributes almost 33% to the overall centrality importance, which is rather a significant amount when considering there are nine other centrality features. The highest-ranked POI is restaurants within 600 meters, contributing to almost 9% of all POI importance among 529 other POI features. Furthermore, it is very interesting to see 8 out of the top 10 POIs are either restaurants or cafes, while bars within 1,800 meters and benches within 1,800 meters amount to the remaining top 2 POIs. This may point to the possibility of many people seeking out places to socialize, with light entertainment and the possibility to gather with friends and family. Especially, since the user data in this research was collected post COVID-19 pandemic, this may show an underlying effect of the pandemic's isolation taking a toll on people's mental and physical health and as it changes people's priority and perception to place an important value on socializing.

The next step is to utilize the zero-user-input approach of the transfer learned model trained on the user data collected from the two cities of St. John's NL and Montréal QC which have different structures and apply this model directly to the third and fourth cities of Kingston ON and Vancouver BC, which the model has never seen before, to generate ALF-Score walkability. In Figure 6.7 ALF-Score walkability (right) is compared to Can-ALE scores (left) for the city of Kingston, ON. At

Feature	***Importance***
Eccentricity	0.01316184
Stress	0.004649347
Betweenness Centrality	0.004590322
Average Shortest Path Length	0.0043923
Topological Coefficient	0.003773664
Neighborhood Connectivity	0.003381009
Radiality	0.002233024
Closeness Centrality	0.001581954
Clustering Coefficient	0.001386535
Degree	0.000191574

Table 6.5: Feature importance for all centrality features (10 features in total) contributes to 4.1% of the total feature importance.

Feature	***Importance***
restaurant_600	0.014984423
bar_1800	0.010154083
cafe_1400	0.007755144
cafe_1600	0.007659399
cafe_2000	0.007125045
cafe_1800	0.005620239
restaurant_1000	0.005089702
restaurant_1400	0.004664054
restaurant_1400	0.004664054
bench_1800	0.003845234

Table 6.6: Feature importance for top 10 (from 530) POI features. The entire 530 features contribute 17.1% to feature importance.

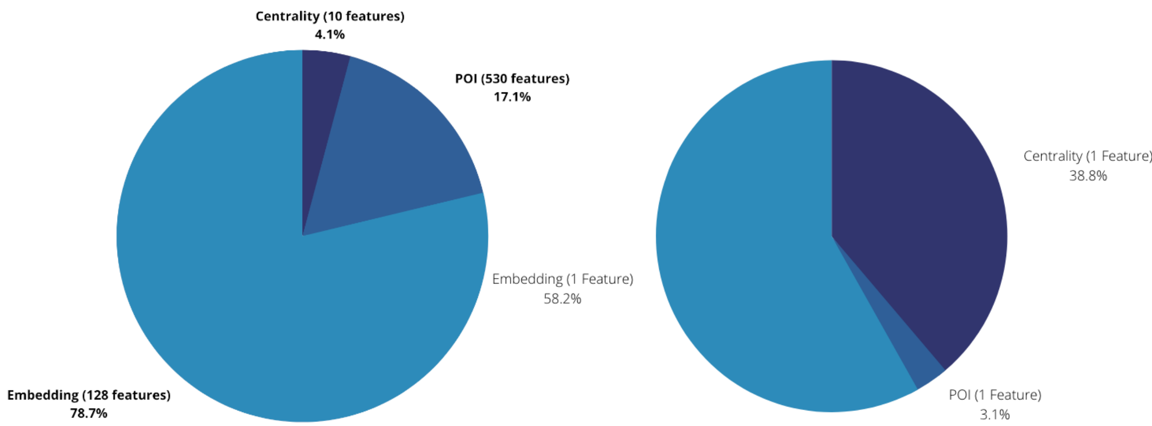

Figure 6.5: Total contribution to feature importance among 668 features is divided into three categories: 1) centrality, 2) POI, 3) road embedding. Left: Road embedding, while contributing to only 19% of the total features, accounts for 78.7% of the total feature importance, while centrality features contribute to 4.1% and POI features to 17.1% of the total feature importance. Right: when normalized to individual feature importance, the highest contribution is through embedding features where each feature contributes to 58.2% of the total embedding contribution of 78.7%, and each centrality feature contributes to 38.8% of the total centrality feature importance of 4.1, while each POI feature contributes to only 3.1 % of the total contributing feature importance of 17.1%. (Figures are drawn by the author.)

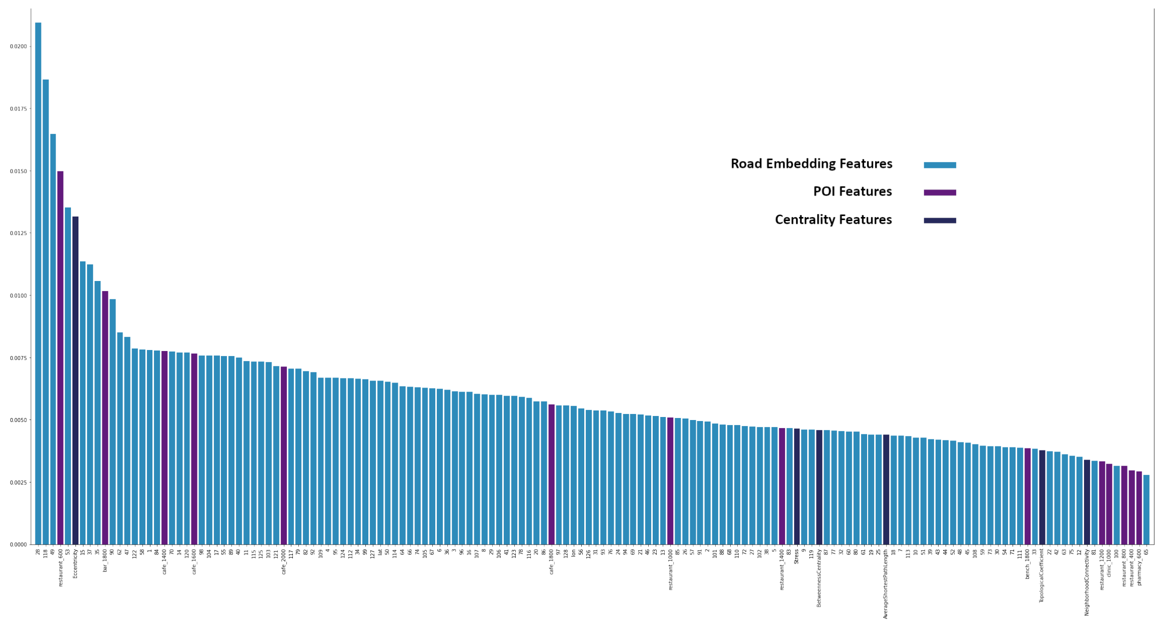

Figure 6.6: Top 150 features. While a noticeable difference is observed among the top 13 features, a steady trend is observed among most embedding features. Embedding feature importance accounts for most of the feature importance. It can also be observed that despite having the highest number of features (530), only a small number of POI features appear in the top 150 features. (Generated by Matplotlib [61])

the first glance, a clear variation in the spatial resolution between these two methods is observed with ALF-Score capturing the walkability of the region in a much greater depth. While Can-ALE shows some variation among different dissemination areas (DA), only the city center is highlighted with visible green and marked as walkable. Although ALF-Score++ agrees with Can-ALE with assigning higher walkability scores to the city center, the first major differentiator among the two is that in Can-ALE higher walkability is given to the central and highly populated areas of the city center whereas in ALF-Score++, while the central region is ranked with higher walkability, ALF-Score++ recognizes the core as slightly less walkable compared to locations surrounding the core of the city center. Specifically, ALF-Score++ favours waterfront walkways and paths as more walkable as opposed to Can-ALE. For instance, the area near Leon's Centre on Ontario Street is known to be a walkable area and is ranked with high walkability through ALF-Score zero-user-input approach whereas it is ranked with a significantly lower walkability score by Can-ALE.

Additionally, ALF-Score captured a cluster of greener/more walkable spots close to students' housing and living quarters near Queen's University. While this area is popular among many students, faculty and other members of the public, Can-ALE was unable to capture it due to its area-based structure and lower spatial resolution. Moreover, it is observed that various other areas that ALF-Score++ ranked as walkable Can-ALE failed to capture their actual walkability due to its lower spatial resolution and granularity. For instance, the Division St. — Dalton Ave. — Benson St. region (which falls under multiple DAs) is ranked with low walkability scores by Can-ALE whereas ALF-Score captured and distributed much more refined and relatable walkability scores to varying spots where there are many restaurants, stores and other popular places. Furthermore, the walkability of Point Frederick Peninsula (across the LaSalle Causeway bridge) is in the red zone of the Can-ALE scores while ALF-Score suggests the opposite for the region. This region houses multiple military campuses with varying facilities and is deemed walkable.

Figure 6.8 shows the ALF-Score++ walkability (right) compared to Can-ALE scores (left) for the city of Vancouver, BC. The ALF-Score++ for this region is generated based on a zero-user-input approach and similar to ALF-Score++ for Kingston, a high spatial resolution is observed as opposed to Can-ALE's low spatial resolution for the same area. To look further into this region, Can-ALE highlights the inner campus area (left side) of the University of British Columbia with light orange while the outer campus area (right side) remains darker orange. ALF-Score++ picks up on the fact that the right area should be more walkable due to bus stops and various facilities commonly used by students and staff. Additionally, North Vancouver's walkability appears not to have been captured by Can-ALE where its walkability for the region is ranging between dark orange and red. In contrast, ALF-Score was able to better capture various popular areas in North Vancouver that are walkable. Furthermore, the walkability for the Richmond area is barely captured by Can-ALE with mostly dark orange and red walkability. ALF-Score++ on the other hand can capture various walkable areas in that region. An interesting observation here is the similarity with zero-user-input walkability data generated for the city of Kingston. Can-ALE typically marks areas close to the water as less walkable whereas ALF-Score++ tends to object. ALF-Score++ results are positively associated with a collective knowledge of Vancouver and Kingston. ALF-Score++ utilizes its transferability capabilities to better understand the city structures and find patterns in various associated data to

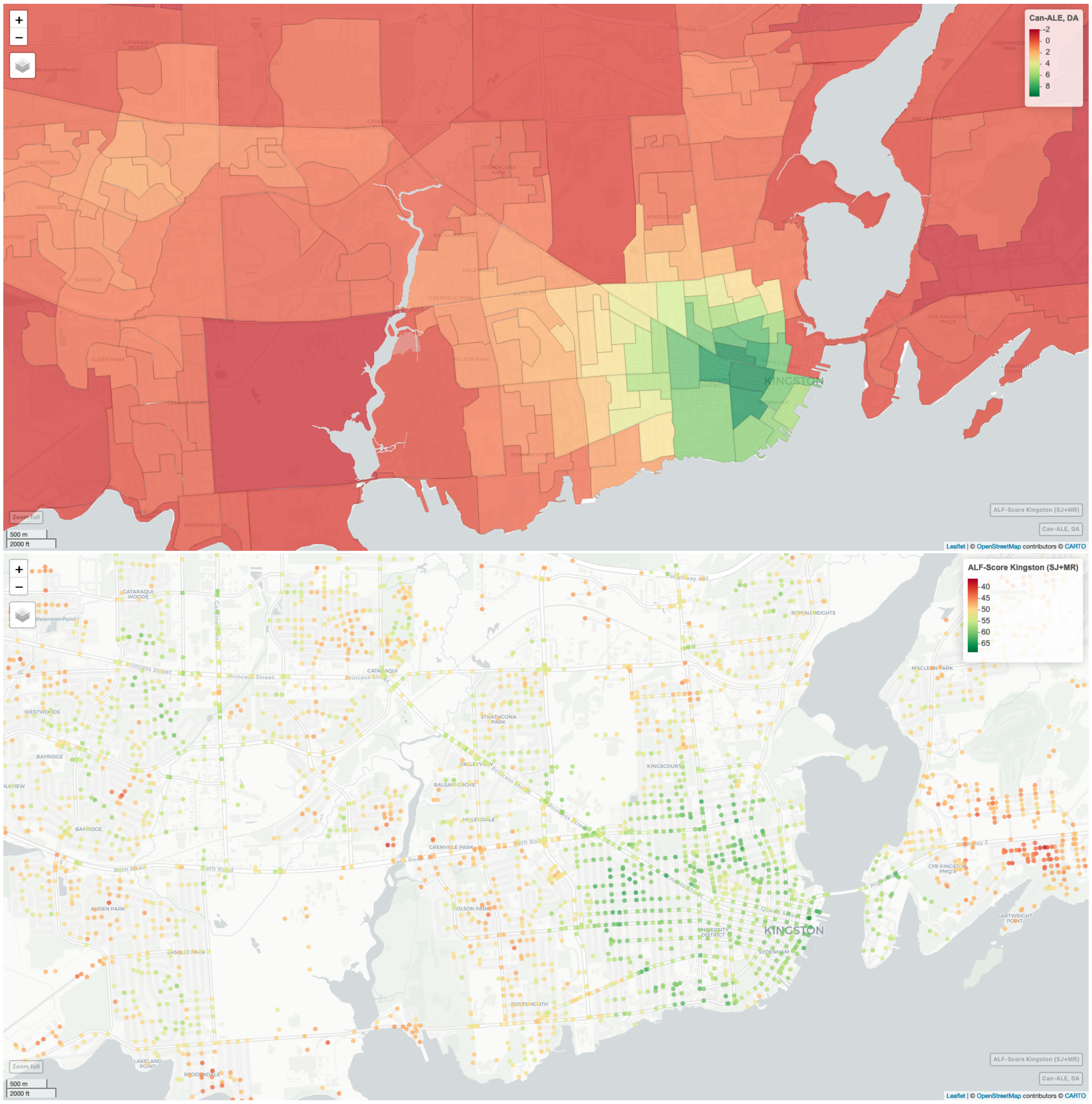

Figure 6.7: Left: Can-ALE for the city of Kingston, ON. Right: walkability results produced by ALF-Score++ for the city of Kingston, ON using a zero-user-input approach of a model trained through transfer learning based on user data from two cities of St. John's and Montréal. The road network for Kingston maintains over 3400 nodes. ALF-Score walkability scores range between 0-100 units. This range can be adjusted if needed. (Generated by RStudio [103])

generate zero-user-input walkability scores for virtually any location on the map.

As observed earlier, the combination of user data from just two cities of St. John's and Montréal allowed ALF-Score generate accurate walkability scores for cities never seen by its algorithms. Transfer learning has proven to work well in this application even with a small set of user data. Additionally, it is believed as more user data is accumulated, ALF-Score algorithms will be able to better capture various patterns in the data leading to improved accuracy.

In this work, the ALF-Score pipeline was reviewed, tested and improved to ensure it is scalable as data size increases. The pipeline was optimized to perform well while processing, training and predicting walkability scores for small and large cities alike. One of the major enhancements to the pipeline was improving the GLEPO algorithm such that the processing time is reduced. This reduction process went through multiple stages. In the initial trials, every iteration of GLEPO took approximately 17 minutes on a personal 2015 MacBook Air configured with a 2.2GHz dual-core Intel Core i7 (Turbo Boost up to 3.2GHz) with 4MB shared L3 cache and 8GB of 1600MHz LPDDR3 onboard memory. A typical run of the algorithm takes approximately 50 iterations totalling over 14 hours of operation. In the final stage of this improvement, the newly updated ALF-Score++ pipeline was able to process the same data using the same computer in just under 3 minutes per iteration, a reduction of almost 6 fold. A GLEPO run of 50 iterations will now only take 2.5 hours. Additionally, after rigorous experimentation and testings, it was determined that while the optimal number of iterations desired for the GLEPO algorithm is 50 iterations or more, the minimum number of iterations required to produce a consistent global list is 30 iterations which could lead to successful completion of the example task above within only 1.5 hours.

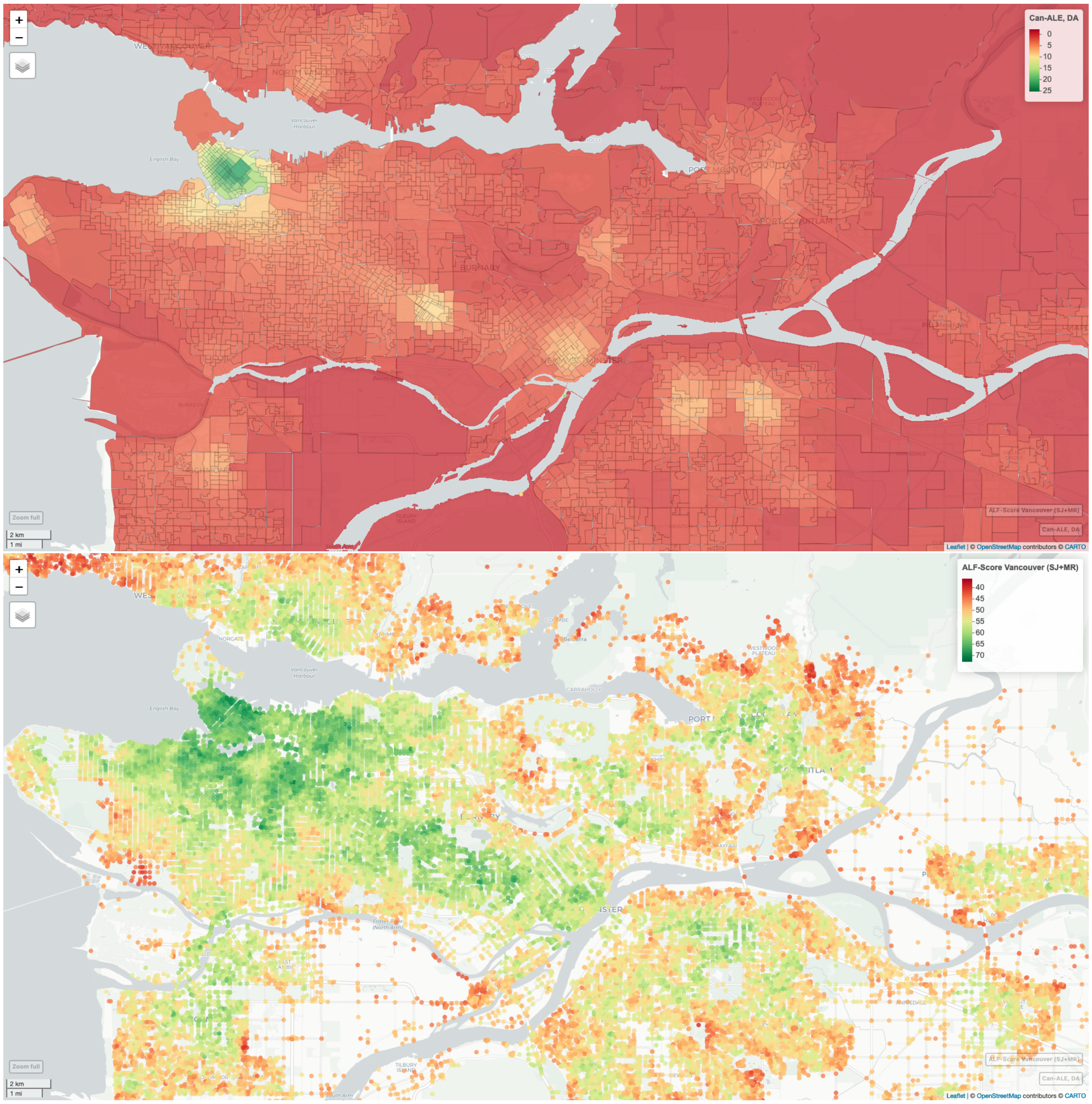

Figure 6.8: Left: Can-ALE for the city of Vancouver, BC. Right: walkability results produced by ALF-Score++ for the city of Vancouver, BC using a zero-user-input approach of a model trained through transfer learning based on user data from two cities of St. John's and Montréal. The road network for Vancouver maintains over 45 thousand nodes. (Generated by RStudio [103])

6.5 Personalization of Transferability

A side-effect of transfer learning is its generalization. Each city will have its range of walkability. Small cities may have a smaller range of walkability whereas larger cities may have a wider range of walkability. When models trained based on small and large cities are combined using transfer learning, the newly trained model will be more generalized. Although this generalization is very important to be able to take a zero-user-input approach to generate walkability scores for cities never seen by the algorithms, one must keep in mind that a balance of data must be maintained. As observed earlier (in Montréal's results), applying a model trained only on a small city might not capture the varying patterns of a larger city and vice versa. It is important to ensure the transfer learning process maintains a good balance of user data for training, such as using user data for small and large cities to build the base model with transferability capabilities. These small samples can prove to be invaluable in improving the overall quality and accuracy of pattern detection and prediction. Moreover, to further address the generalization phenomenon happening during the transfer learning phase, ALF-Score personalization extension (ALF-Score+) can be utilized to create personalized and transferable models that are generalized to various city structure patterns, yet personalized to specific individual profile clusters.

Personalization is not a necessary component of transferability. However, the personalization component can be combined with transferability to complement its pipeline. The personalization process is explained in detail in the previous chapter, ALF-Score+ [4], but certain approaches can be adjusted to allow compatibility with ALF-Score++ transfer learning capability. The first approach to personalization in the context of transferability is to combine the user demographics from all user data planned to be used in the process. This includes user data collected from all cities that will be included in the transfer learning process. In this work, user data from two cities of St. John's NL and Montréal QC is used. The combined user demographics are then processed by the ALF-Score+ profile clustering pipeline resulting in a list of appropriate clusters. After the right number of clusters is determined through the elbow method and the silhouette coefficient, assigned cluster numbers will be associated with each user entry specific to each city. For instance, if 5 clusters are deemed appropriate, each cluster may contain users from either or both cities of St. John's and Montréal. If user entry E_x^{STJ} from St. John's happen to be associated to cluster 1 while user entry E_y^{MTL} from Montréal is also associated to cluster 1, these cluster associations are retained within localized data for each city. Then the user data from each city is segregated based on the cluster association and will follow the MLP training approaches to train a base personalized model. Top hidden layers in these models can then be processed by the transfer learning pipeline to train new models.

Figure 6.9 shows the silhouette coefficient (top right) with a gradual increase in a span of 10 clusters between 2 and 12 clusters. A small variation is observed between 10 and 12 clusters with 12 clusters being the peak of the coefficient at 0.95. On the other hand, the elbow method (Figure 6.9 top left), shows the k-means distortion drop significantly as the number of clusters is increased to 4 followed by a gradual decrease in distortion converging between 9 and 11 clusters. Ten-cluster point is observed as a pivotal point in both methods and for this reason, 10 personalized profile clusters were generated. Figure 6.9 bottom shows user demographics features reduced to a

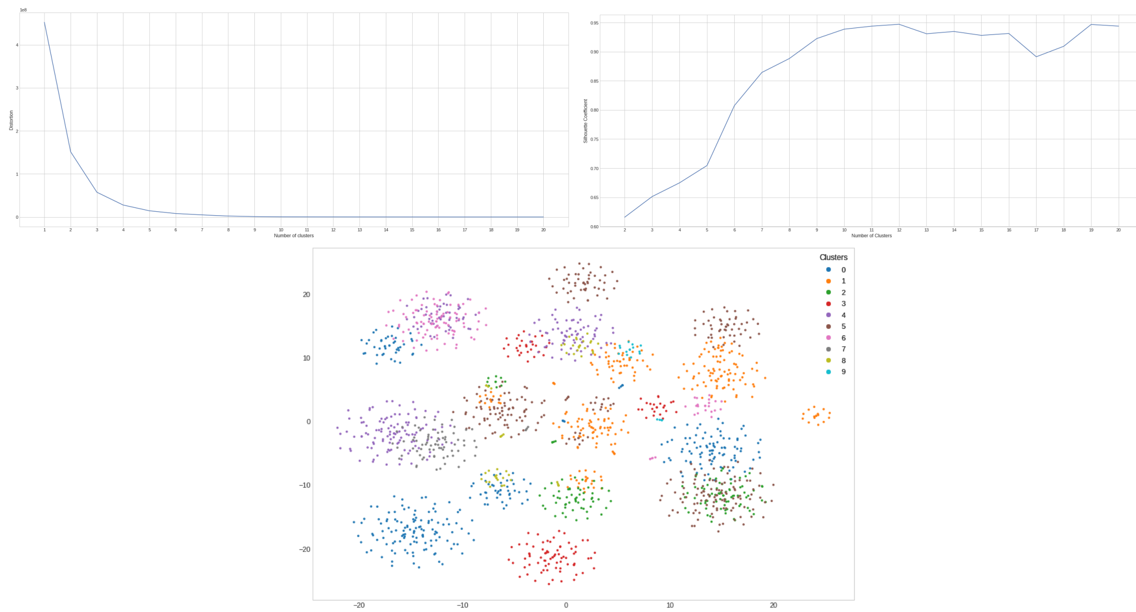

Figure 6.9: The Elbow method (top left) and silhouette coefficient (top right) are used to determine the most appropriate number of clusters. 2-dimensional t-SNE representation (bottom) of users over their demographics and system features. (Generated by Matplotlib [61])

2-dimensional representation using the t-distributed stochastic neighbour embedding (t-SNE).

From Table 6.7, it can be observed that the error rate has been decreased over the generalized models with the lowest error of 4.56 units for random forest and 13.28 units for MLP. It is important to note the columns Avg RF Error and Avg MLP Error in Table 6.7 represent average MAE error among two models: 1) the model trained on the subset of user data containing nodes associated to the relevant cluster that are only within St. John's, 2) the model trained on the subset of user data containing nodes associated to the relevant cluster that are only within Montréal. Furthermore, Transfer (MLP) column represents the best accuracy achieved through transfer learning using MLP. This transfer learning utilizes weights trained through model 1 as the base of the learning and continues the learning process using the user data that was used as input to model 2. An additional note is that cluster 9 contains only volunteer users from the Montréal data and the average accuracy columns for RF and MLP associated with this cluster only represent the accuracy for the models associated with this cluster that were trained on Montréal data.

Cluster #	***Training Size***	***Test Size***	***Avg RF Error***	***Avg MLP Error***	***Transfer (MLP)***
0	312	78	8.66	17.22	18.01
1	258	65	5.89	15.30	14.85
2	144	36	7.52	16.80	16.17
3	128	32	6.43	16.06	15.87
4	241	60	4.56	14.62	13.28
5	317	79	9.07	18.24	17.49
6	117	29	7.62	17.96	17.31
7	78	19	10.21	19.32	19.82
8	56	14	9.98	18.61	19.55
9	19	5	13.45	21.76	20.32

Table 6.7: Accuracy for personalized and transferable models generated by ALF-Score++ pipeline based on 10 profile clusters. A best-case of 4.56 MAE error units using random forest (RF) technique is observed while the MAE error for MLP has also been decreased with a best-case of 13.28 units. Although each cluster uses a smaller data size for training as opposed to the larger original dataset, since the sets are highly focused on specific clusters of users with similar opinions towards walkability, prediction accuracy was maintained and improved. Accuracy is expected to improve further with more data.

Figure 6.10 visualizes the prediction results for the city of Kingston ON, based on a specific personalized cluster (left) processed through zero-user-input transfer learning and compares with the results of a zero-user-input transfer learned model of the same city without any personalization (right). The personalized model is trained based on cluster # 4 over user data collected from both cities of St. John's NL and Montréal QC with a transfer learning error of 13.28 units (MLP). Cluster # 4 is associated with men in their 20's and 30's who do not live alone and mostly have no children, consider 1,200 to 1,400 meters as a walkable distance and with a common profession selected as Professional. A very noticeable difference is observed in how the algorithm assigns walkability scores based on various locations when personalized. Although, it is important to note that with more user data collected, the clusters will need to be redefined. Furthermore, cluster # 4 was chosen as it shows the most extreme variation to that generated by the generalized ALF-Score++ through transfer learning. This was done to maximize the visibility of the variation and to highlight the different results generated using a personalized cluster. It is worth noting that different clusters will have different results some closer to the generalized model while some others (such as cluster # 4) may have a clear variation. Stratification of the data means a sparser training set for each cluster profile however since the models are well defined, improved accuracy is observed. Finally, the application and visualization of transfer learning can be applied to all other individual profile clusters; however, are omitted for brevity.

It is important to note that as more user data is collected, to improve user demographics associativity with the profile clusters, clustering will need to be reapplied. However, transfer learned models that are trained based on personalized clusters can continue to be used as zero-user-input models to predict personalized walkability scores for cities never seen by the algorithms during their training cycle.

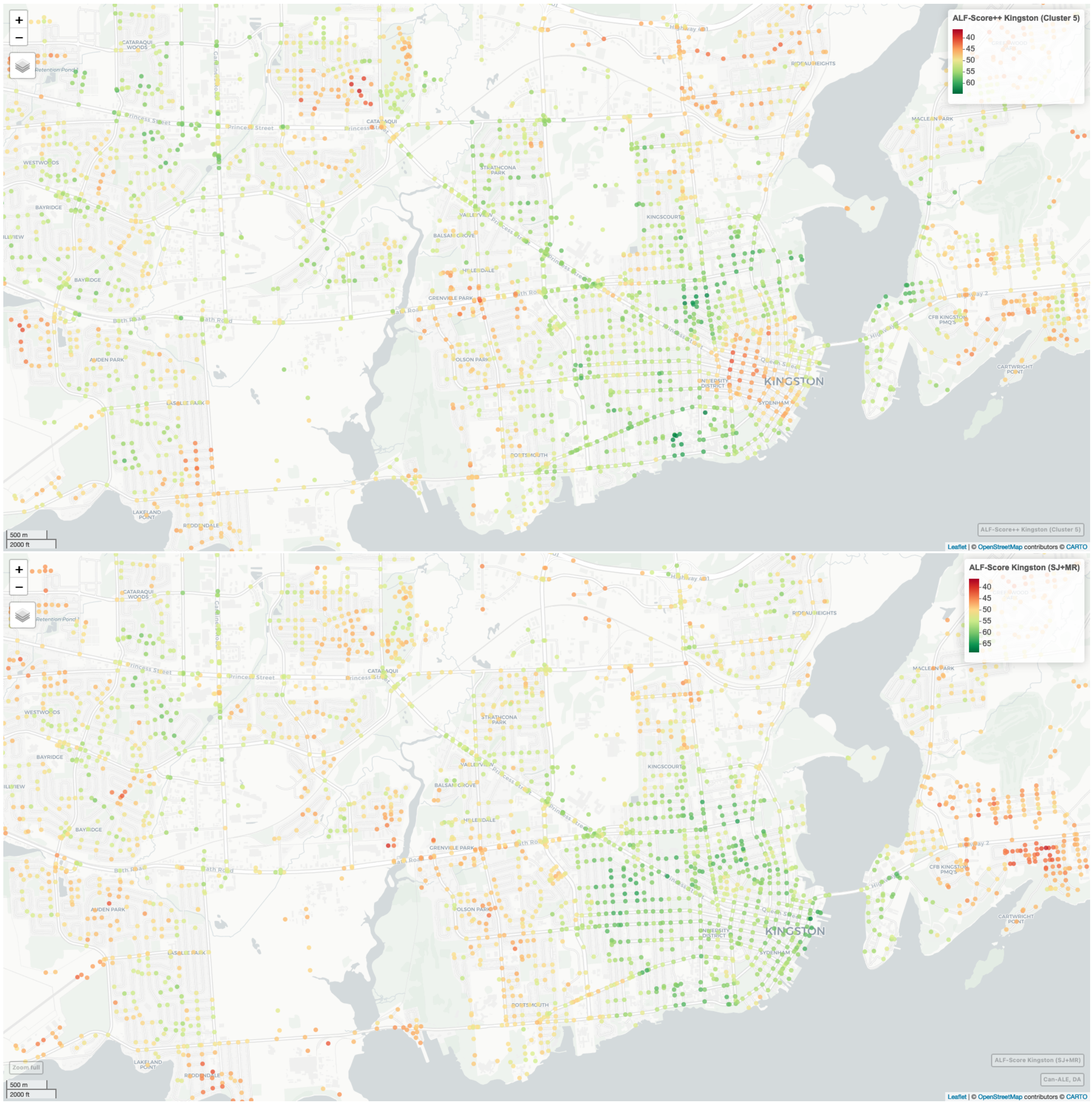

Figure 6.10: Left: Zero-user-input personalized and transferred predictions of cluster # 4 for the city of Kingston, ON. Right: ALF-Score++ zero-user-input transferred predictions without personalization for the city of Kingston, ON. (Generated by RStudio [103])

6.6 Discussion

The goal of this overall research is to explore how machine learning can be applied to the spatial domain with application in public health through generating relevant and meaningful walkability scores with the high spatial resolution based on a very small set of users' opinions. In this chapter, ALF-Score pipeline was improved and tested to be fully capable of scaling up and down to match the data based on the size of the city and user opinion data while performing promptly. Additionally, since the computational complexity of the pipeline is $O(n^2)$, It is expected that processing larger cities will perform reasonably and within the expected parameters. Moreover, ALF-Score++ was shown fully capable of processing data and generating models for the city of Montréal QC which is almost 16 times larger than that of the city of St. John's NL, within a timely fashion without requiring any extended resources while these models are capable of producing walkability scores with high spatial resolution compared to that of Can-ALE. Figure 6.11 shows a correlation comparison between ALF-Score++ walkability scores and Can-ALE walkability scores for four different cities in Canada.

Moreover, the power of transferability was observed to provide an upper hand to transfer the knowledge learned from small cities to predict accurate walkability scores for much larger cities. This leads to many advantages such as reduced resource requirement and reduced processing time while increasing the flexibility of applicability of the trained models. Furthermore, its application of zero-user-input transfer learning proved to be a huge success in predicting walkability scores for cities never seen by the algorithm before and without any prior knowledge about them while utilizing

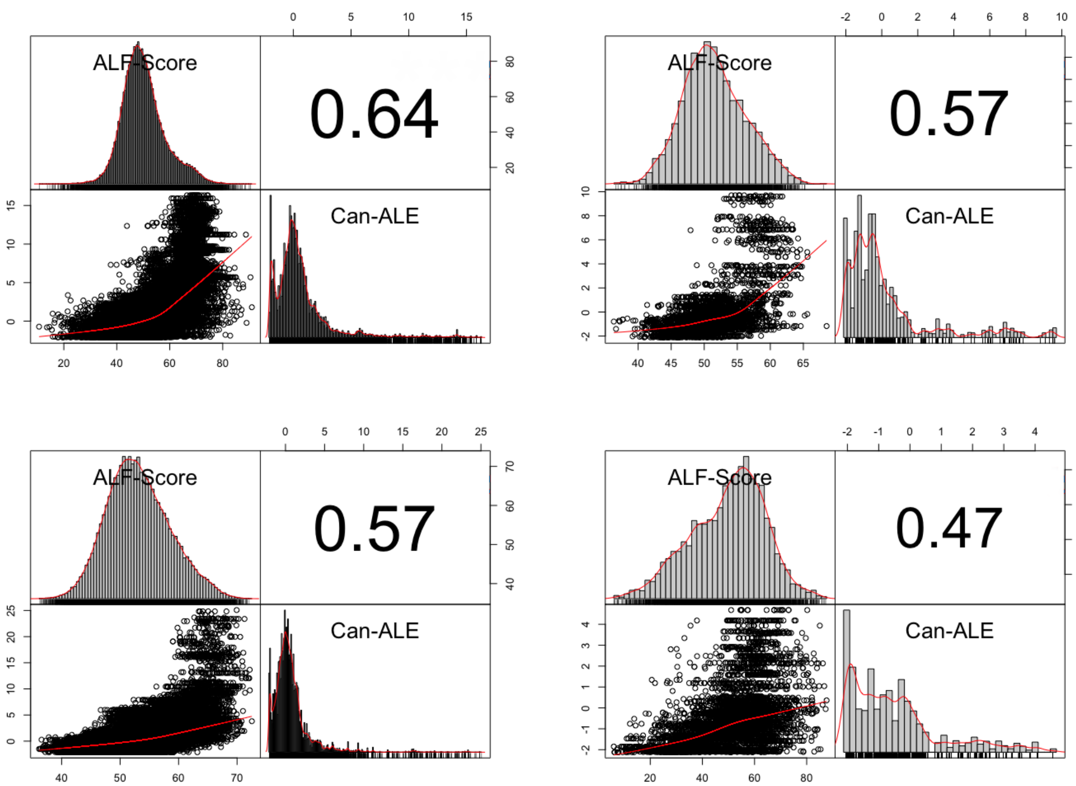

Figure 6.11: Correlation between ALF-Score++ and Can-ALE for four different cities. Top left: Montréal QC, Top right: Kingston ON, Bottom left: Vancouver BC, Bottom right: St. John's NL. (Generated by RStudio [103])

previously learned information and patterns. Of note, the transfer learning process was able to capture both relative and absolute differences between cities in terms of walkability. For example, the range of walkability scores assigned to Kingston based on the transfer learning was between 30 and 70, whereas scores generated for the city of Vancouver based on transfer learning ranged between 30 and 80 when St. John's and Montréal were used as training cities. Developing measures that capture both relative and absolute differences in cities has been an ongoing challenge, that transfer learning may be able to solve.

Furthermore, it was observed that adding a small set of users' opinions from a different region can lead to much higher pattern recognition by the models while allowing for a better generalization of these models. This generalization can therefore help capture various common patterns found in different cities without any actual prior knowledge about them.

Similar to many machine learning tasks, the ALF-Score++ pipeline can train more accurate models and benefit from more data. For instance, collecting small user data samples across various cities, towns and user groups could cover a much more diverse set of user demographics, user opinion, patterns, city and road structures leading to a well-representative model applicable to virtually any location within any road network. Given enough user information from a few select key cities in Canada (cities with varying structure and sizes), models generated through transfer learning of this data should be able to estimate accurate scores for any location in Canada. To generate new models, one does not need to rerun the entire process on the entire dataset. All that is required is to transfer the knowledge from previously trained models (which can be transfer-learned models themselves). ALF-Score++

pipeline can be adjusted to be fully capable of continuous learning. This could be particularly important as changes to road networks are detected. The network-based approach combined with continuous transfer learning can help models detect patterns associated with various regions, types of road and user demographics and provide accurate predictions for new roads and structures never seen by the model.

As the work went through the predictive process, a variation was observed between the performance of shallow and deep models. Throughout ALF-Score, random forest (a shallow model) was the preferred technique since 1) it performed best across all other techniques (shallow and deep) achieving an MAE error as low as 4.56 units, 2) its simplicity and powerful approach, 3) faster processing and prediction compared to MLP. Although, MLP (a deep model) is the main technique used in ALF-Score++ since deep models are preferred when it comes to transfer learning due to their layered structure. The lowest error achieved using MLP was 13.28 units. Although deep neural network techniques are generally expected to provide more accurate results, they require a tremendous amount of data. As it was observed, this could be an issue when dealing with only a small set of data like the one used in this research. It is believed that as more user data is accumulated, MLP's performance will improve.

According to the popular book by Aurélien Géron [50], it turns out that transfer learning does not perform very well given a small network. This is presumed to be because small networks learn few patterns while dense networks learn very specific patterns which may not be useful for other tasks. Géron suggests that "transfer learning works best with deep convolutional neural networks, which tend to learn feature detectors that are much more general". Furthermore, deep neural network methods are typically used when there is huge and unstructured data to process, for example classifying images. On the other hand, shallow models, for example, random forest, tend to be more useful to process structured data with many dependable features, for example predicting the weather. These models work well with smaller cases and larger attributes whereas, in the case of deep neural networks, much more cases and much fewer attributes are required. This is one of the reasons why higher error rates in ALF-Score were observed among the MLP models when compared to the shallow models, specifically random forest. Deep neural network models perform particularly well in vision and speech domains.

Lastly, it is important to highlight that there are other methods and approaches to transfer learning such as ensemble transfer learning approach [76] and boosting for transfer learning [137] and one should consider such methods and choose the most appropriate based on the task at hand.

Chapter 7

Conclusion

To conclude this dissertation, it is important to point out the importance of the interdisciplinary research between computer science and other fields, specifically public health. Technology is advancing every day, and what better use of it than to improve people's lives and health. This kind of practical interdisciplinary work can truly and positively impact the world. For example, worldwide physical inactivity is associated with 9% of all premature mortality (5.3 million deaths per year), 6% of the burden of coronary heart disease, and 7% of type 2 diabetes. If the population meeting physical activity guidelines increased by 10%, more than 533,000 deaths could be averted every year worldwide [71]. The built environment, defined as "man-made or modified structures that provide people with living, working, and recreational spaces" [3], represents an important method with the potential to increase physical activity at the population level in Canada and worldwide. Many research studies focus on how computer science can bring a wealth of possibilities to public health and health professions. For instance, Tolsgaard et al. [118] studied the role of data science and machine learning in health professions education. They conducted a review to explore "(1) published applications of data science and ML in HPE literature and (2) the potential role of data science and ML in shifting theoretical and epistemological perspectives in HPE research and practice". Furthermore, in another article by Barnes et at. [10], the authors made a case for Computational Health Science, an interdisciplinary collaborative work among health scientists, computer scientists, engineers, psychologists, and other social scientists. They defined computational health science as conducting "innovative research that will inform future practice directed at changing health behaviour through improved communication, networking, and social capital". Others, such as Kunkle et al. [72], call for action to address emerging health needs through integrating new and personalized technologies.

Furthermore, measuring the environments around us, including cities, roads, and social environments, is crucial to help better understand human behaviour. This knowledge can help predict how aspects of the environment influence behaviour and health. Walkability is one measure of the environment used to predict health. As this research revolves around walkability, it is vital to define walkability since there are many operational definitions of walkability in the literature. Walkability is a concept that many researchers have used to operationalize characteristics of the environment that support walking. Simply put, walkability scores can tell how "walkable" the surroundings are. However, it is important to keep in mind that there is no single agreed-upon conceptual definition of walkability. For instance, some measures define walkability focusing on POIs, whereas others may consider a higher weight for population density. Similarly, some walkability measures deem an area-based approach feasible to calculate accurate walkability scores, whereas others may have a different

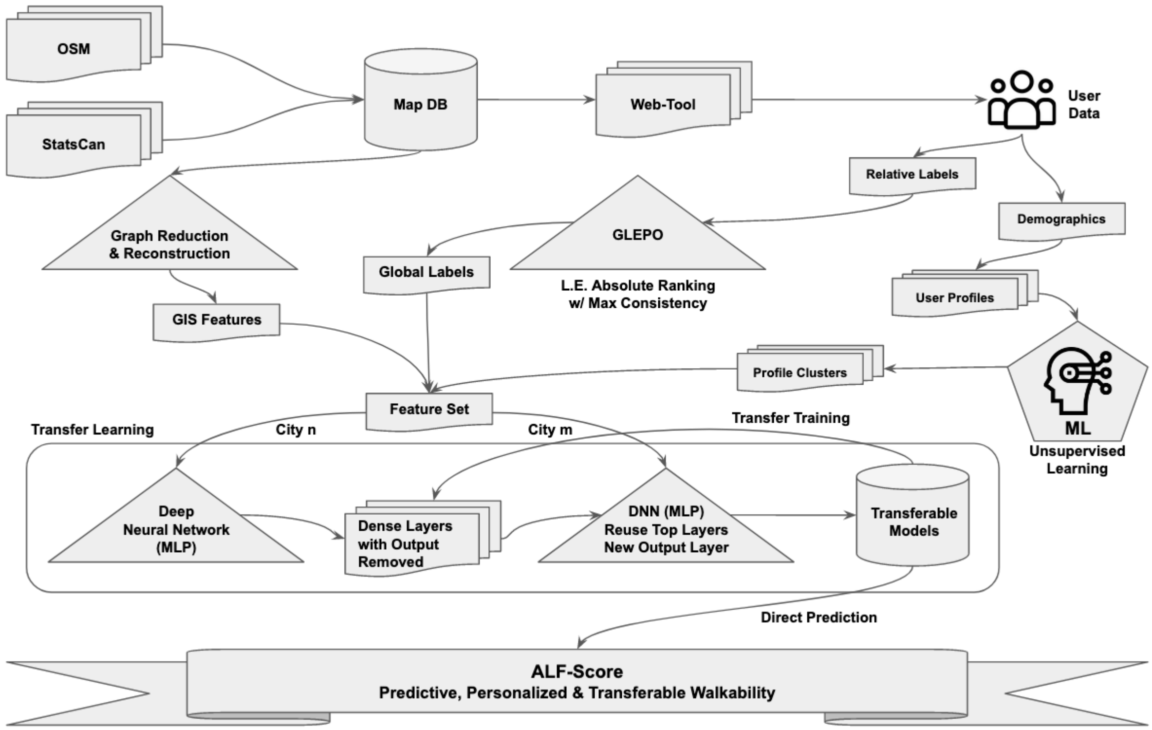

Figure 7.1: An overall look at the entire ALF-Score pipeline. (Figure is drawn by the author.)

acceptable policy. Figure 7.1 reflects on the overall structure of the entire ALF-Score pipeline.

The primary reason behind incorporating road networks into ALF-Score is that roads and road network structures are essential in people's daily lives. To better understand our surroundings and how aspects of the environments influence human behaviour, how people are connected to their surroundings should be considered. The importance of road networks has been well-known for thousands of years. Chengjin Wang et al. studied the evolution of road networks in China between 1600 BC to 1900 AD [125] and suggested that road networks were used for major nationwide socio-economic exchanges long before modern transportation emerged and have been in long-term development and are continuously expanding. This expansion of road networks with continuous change in accessibility and coverage is closely associated with natural conditions, national defence and "centralization of national power, national land governance, postal transport, the transport of specialized cargos, and international trade".

Road network data is abundantly available from various sources such as Statistics Canada [113] and OpenStreetMap [91]. However, as data expands, analyzing larger regions requires much more processing power and computational time. Especially when working with road networks, typically larger networks lead to higher connectivity between the nodes causing a higher complexity within the network. Most popular and existing algorithms to measure network importance are already efficient, yet due to the nature of the problem they are solving, they are still too complex to process large networks quickly and on personal computers without any specialized hardware.

Graph reduction and reconstruction techniques were proposed (and presented in Chapter 3) to address the existing gap. Its first phase reduces network size to a fraction of the original graph while maintaining the main structural characteristics. The reduced graph can then be processed much faster than the original graph and be used in various methods such as different centrality measures. In its second phase, the reduced graph is reconstructed using a variation of the linear interpolation technique. This phase reconstructs the original graph based on the partially calculated data measured in phase one and estimates the values for the "reduced" nodes.

This approach made it possible to reduce large road networks and decrease the required computational time, saving hours and possibly days, as evident from Table. 3.2. Although to ensure accurate results are produced, the network structure needs to be kept intact during the reduction process, limiting node reduction. However, on average, a 77% reduction of road nodes was achieved within various networks while preserving the original networks' structure. The application of most centrality measures such as betweenness centrality and closeness centrality performed over the reduced network was made considerably faster. It is worth mentioning that the second phase runs in linear time and the computational time for reconstruction (Table 3.3) is almost negligible.

After completing the graph reduction and reconstruction phase, it was possible to fully reduce and reconstruct networks to calculate various centrality measures for different cities in Canada for a fraction of the time needed using traditional methods. The next important step was to consider implementing user opinion as a guiding factor towards how ALF-Score walkability scores based on road network features are influenced by the general public's view, which brings up a significant gap in existing walkability measures, highlighting that most are one-size-fits-all heavily influenced by only a few people's perspectives (generally the researchers). Through user opinion, ALF-Score walkability represents user opinion while being more informative to most users by taking a user-centric approach instead of the traditional researcher-centred approach.

An interactive web interface was devised to reduce bias while collecting user opinion data, allowing users to provide their opinion in rankings between 5 randomly selected locations. The decision to collect the data labels as a combination of relative rankings and not absolute scores (e.g. ranking every 5 locations between a fixed range such as 0-100 where the same rank may apply to other locations) was made to reduce bias and conflict in the data while incentivizing the participants to make a precise decision to determine which of the given 5 locations would be most or least walkable and order the locations as deemed appropriate. While absolute scores are easier to use, each individual may have a completely different rationale for why a location has been ranked the way it is, such as a score of 65 and not 40. Utilizing relative rankings takes that factor out to focus on the most important factor: whether location A is more walkable than location B for a given individual. This way, there will be a concise approach to determine each individual's perception towards walkability when comparing different points together over other users' responses.

It is essential to highlight that only a small group of users provided user opinions and demographics in this research due to limited access to volunteer participants. To avoid bias and to be more representative of people, one should consider the inclusion of various groups and sub-groups of individuals such as minorities and those with disability.

Conflicts are unavoidable and lead to inconsistencies among user opinions. The challenge was to balance out the opinions collected from users to yield a walkability metric that is agreeable to most users. The collected user opinions maintain a unique form observing only relative orders within each submission of 5 locations and therefore do not represent the global view among all people. Due to this and the nature of the conflict resolution, the problem remains in the NP-complete space. Therefore, a new approach to handling conflicts and representing user opinions within a global perspective needs to be devised. Generalized Linear Extension of Partial Orders, or GLEPO, is an algorithm developed specifically to process this approach of user opinion label collection and to convert relative ranks to absolute orders. GLEPO produces a generalized list of all user opinions in total order and absolute ranks to represent relative ranks of small submissions within a global observation of the overall data.

Using the GLEPO algorithm, converting users' relative opinions among groups of 5 locations into globally relative scores (among all submissions) was successful with only a minor inconsistency. GLEPO maintains a high consistency of 98.24% throughout the conversion. Furthermore, after numerous variations and experimentation, it was determined that the best results were produced using the randomized version of the algorithm with at least 30 iterations. In comparison, 50 or more iterations were preferred. The virtual link was enabled when the best results were produced. GLEPO's output provides an overall well-represented user opinion baseline to walkability scores while establishing a ground truth for the ALF-Score pipeline. However, GLEPO's main contributions may be extended to an extensive range of researches requiring proper use of crowd-sourced user data while reducing bias due to differences in opinion.

A few alternative approaches were experimented with in the earlier stages of GLEPO's development before moving on to the final approach. One of the earlier versions of GLEPO, which was called R2A or Relative to Absolute, utilized Can-ALE data as its base. This approach took advantage of already established scores generated by Can-ALE and assigned these scores to their associated locations within the user data. The association was done by projecting each DA's score to all nodes within the DA. The algorithm would then rearrange the nodes based on the conflicts observed within the user opinion data. Each point within the resulting globalized list may have a completely different score than what it may have had within Can-ALE; however, the overall list will still maintain a score range similar to Can-ALE for the region.

This approach worked great as a starting point to establish a base algorithm to produce a based line to process the user opinion data. But the issue with this approach is that although the resulting scores for the nodes may have been different than that they would have been assigned through Can-ALE, the influence of this extension was observed within the range of the overall score list. For instance, if Can-ALE for a region $\mathcal{R}^1$ ranges between -2 and 2, this would have been the same range of score for R2A's globalized list for the same region; even though a given *node a* may have been ranked as -1 in Can-ALE and yet ranked as +1.5 by R2A. Furthermore, having other regions with different ranges in their overall Can-ALE scores included in the algorithm, for example, $\mathcal{R}^2$ ranging between -3 and 10, would further influence the overall baseline. Can-ALE data was removed in the later iterations, relying only on user-provided data. Some other earlier iterations of GLEPO also included variations with no randomization, no virtual link or without multiple iterations.

Since user opinion was fully processed into usable data, there are a few potential approaches to utilizing this user opinion data. Treating them as labels in a machine learning pipeline is an excellent approach chosen in this research. The reason behind choosing machine learning is that predictive models have considerable potential for flexibility and diversity of application. Through machine learning, this research foregoes the need to compute walkability scores for every single location but instead train models that can make intelligent estimations. This approach reduces score generation time and makes it possible and feasible to estimate scores for all points within the road network leading to a high point-level spatial resolution. Furthermore, machine learning approaches open the door to many possibilities, such as incorporating personalization through unsupervised learning methods and utilizing transferability to gain the ability to transfer previously learned ALF-Score walkability knowledge to various cities that were never involved in the training process.

Various ML techniques were applied and compared based on different feature set combinations to incorporate machine learning. The goal is to find the most suitable technique and feature combination sets that produce the most appropriate models predicting accurate walkability scores. The techniques that were used in both supervised and semi-supervised environments are: 1) random forest, 2) linear regression, 3) decision tree, 4) support vector regression (SVR), 5) gradient boosting, 6) polynomial features (a non-linear approach), 7) lasso CV and 8) multi-layer perceptron neural network (MLP). Feature combinations used are: 1) only POI features, 2) only network features (centrality measures), 3) only road embedding features, 4) POI + road network features, 5) POI + road embedding features, 6) road network + road embedding features, 7) all features.

Random forest outperformed all other techniques in terms of performance and accuracy by achieving a top prediction accuracy of 87.49% (ALF-Score) using all features. Models trained on only POI features performed relatively similar to those trained on POI plus network features and POI plus road embedding. However, models trained on network features combined with network embedding appeared to perform better than those using POI previously motioned. Furthermore, road network structure plays a crucial role in measuring walkability. Improvement in accuracy is observed after adding POI features to the network and road embedding features, which highlights the complementary position of POI to road network structure.

ALF-Score's use of user opinion and machine learning approach to estimate scores based on various features opens the possibility of introducing an extension to personalize this walkability measure. ALF-Score+ explores this extension by utilizing user-defined and system-defined user demographics to create individual profiles to develop profile clusters. The ALF-Score pipeline then uses user labels and profile clusters for further processing. It generates machine learning predictive models that estimate personalized walkability scores specific to each profile cluster. The introduction of user profiling and profile clustering into the pipeline helps assign users to distinctive clusters that represent the majority of users within that cluster. Each cluster is then used to train specific machine learning models that best represent the users within that cluster with estimated walkability scores influenced by their opinion.

Using the elbow method and the silhouette coefficient, it was determined that six distinguished clusters would best represent the users from the St. John's dataset. ALF-Score+ maintained an accuracy of 90.48% on average over at least 35 separate experiments with the best case of 93.70%. ALF-Score+'s overall accuracy on average is consistently and noticeably higher than that of ALF-Score. This improvement in accuracy is attributed to the user-focused approach of ALF-Score+ models; however, one should also note the significant drop in training size of personalized sets compared to the original ALF-Score experiments. Although smaller training sets were used to train each cluster profile's specific model, the trained models captured more in-depth characteristics of each profile cluster. Random forest regressor remained the top-performing technique.

The final phase of this research, ALF-Score++ [7], is the second extension of ALF-Score. It focuses on scalability and transferability of the overall pipeline to ensure 1) the feasibility of the flow and the algorithms within the pipeline regardless of the size of the city being processed, and 2) producing repurposable, reusable and transferable predictive walkability models. Most experiments in the previous chapters focused on utilizing smaller cities, such as the city of St. John's, NL, with only 5,364 nodes and 6,851 edges. Furthermore, a small set of user opinion data containing 1,050 user entries covering 895 unique locations was used. Feasibility of the flow and the algorithms within the pipeline is a crucial step to ensure the pipeline can handle road networks and user opinion datasets of varying sizes and can process them promptly. For example, the city of Montréal QC and its surroundings have around 76,663 nodes and 114,414 edges which is almost 16 times larger than the road network for the city of St. John's. Transfer learning is yet another missing component from many existing walkability measures. Producing transferable and reproducible predictive walkability models is an essential component of ALF-Score++, which allows this measure to utilize previously learned knowledge when generating walkability scores. This knowledge can be used to reduce future training time, required resources and labelled data, help improve the overall accuracy and provide estimated walkability scores for cities never seen before by the algorithms.

Finally, with ALF-Score++, transferability was explored. A way to use existing ALF-Score models as a base to transfer previously trained knowledge was proposed to speed up the processing time for new training and increase the accuracy of these models. Transferability was successfully achieved using ALF-Score++. The lowest prediction MAE error (matching approach) using random forest shallow model was at 11.87 units (Figure 6.4 top left) (and 4.56 when personalized). At the same time, MLP was the best performing deep model with an MAE error of 13.87 units (and 13.28 when personalized). Walkability scores for the city of Montréal were accurately predicted using user data associated with this city. However, walkability scores for the cities of Kingston and Vancouver were also estimated using models only trained on the user data from the cities of St. John's and Montréal. The performance was improved by combining transferability (ALF-Score++) with personalization (ALF-Score+) to generate transferable and personalized walkability scores, achieving an MAE error of 4.56 units for direct training using random forest on a personalized cluster and 13.28 units for transfer learning using MLP, over a value range of 0-100 units.

ALF-Score and its various extensions, such as ALF-Score+ and ALF-Score++, can be very beneficial and powerful tools for many people from various backgrounds working on different domains. Although ALF-Score can produce results specific to various parameters, such as demographics, to provide personalized walkability scores, the ALF-Score pipeline takes a generalized approach instead to allow addressing various issues that may not be directly related to walking or walkability. For instance, bikeability, school friendliness, transit friendliness, or even POI specialties based on different demographics and perceptions. Moreover, the pipeline may be capable of handling a wide variety of features and other types of networks instead of the road network. For example, subway networks. At its core, ALF-Score requires a vector of user ground truth labels alongside a list of features. ALF-Score uses its dedicated web tool to collect the ground truth labels and processes them through GLEPO to reflect relative to absolute conversion within a small group of users. However, the ALF-Score pipeline follows a black box system and works with any compatible input data regardless of how they were prepared. The ground truth data can be processed according to researchers' needs, and this step can be bypassed in the pipeline if needed. Although walkability scores generated by ALF-Score and its extensions rely on road network data, the generalization offered by their pipelines can be further distilled beyond road networks. Road network data is treated as any other feature and can be replaced with an appropriate feature based on the issue at hand and the research requirements. It is my genuine belief and hope that ALF-Score can open the door to many possibilities well beyond the scope covered in this research.

To conclude, ALF-Score and its extensions show tremendous potential in utilizing the science of complex networks, graph theory, and machine learning to act as a positive tool to solve real-world problems in varying fields such as computer science and public health. Although the contributions of this research to the literature are evident, I believe ALF-Score has the potential to continue to contribute to the world, people's health and well-being in many significant ways and that the possibilities of its application and use are endless.

I am hopeful for a brighter future!

Bibliography

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[2] H. Abdi. The Kendall rank correlation coefficient. *Encyclopedia of Measurement and Statistics. Sage, Thousand Oaks, CA*, pages 508–510, 2007.

[3] U. S. E. P. Agency. Basic information about the built environment. *United States Environmental Protection Agency*, 2020.

[4] Ali M. S. Alfosool, Daniel Fuller, and Yuanzhu Chen. ALF-Score+ — Personalization of a Predictive Network-Based Walkability Scoring System, October 2021. DOI: 10.31219 osf.io/qdtfj.

[5] Ali M. S. Alfosool, Yuanzhu Chen, and Daniel Fuller. Road Importance Using

Complex-Networks, Graph Reduction & Interpolation. In *2020 International Conference on Computing, Networking and Communications (ICNC)*, pages 855–859, 2020.

[6] Ali M. S. Alfosool, Yuanzhu Chen, and Daniel Fuller. ALF-Score — a Novel Approach to Build a Predictive Network-Based Walkability Scoring System, April 2021. DOI: 10.31219 osf.io/tcgqp.

[7] Ali M. S. Alfosool, Yuanzhu Chen, and Daniel Fuller. ALF-Score++ — Transferability of a Predictive Network-Based Walkability Scoring System, November 2021. DOI: 10.31219 osf.io/2xgwa.

[8] D. Amaratunga, J. Cabrera, and Y.-S. Lee. Enriched random forests. *Bioinformatics*, 24(18):2010–2014, 2008.

[9] N. Arsov and G. Mirceva. Network Embedding: An Overview. *arXiv preprint arXiv:1911.11726*, 2019.

[10] M. Barnes, C. Hanson, and C. Giraud-Carrier. The case for computational health science, 2018.

[11] D. W. Barnett, A. Barnett, A. Nathan, J. Van Cauwenberg, and E. Cerin. Built environmental correlates of older adults' total physical activity and walking: a systematic review and meta-analysis. *International journal of behavioral nutrition and physical activity*, 14(1):1–24, 2017.

[12] G. A. Battista and K. Manaugh. Using embodied videos of walking interviews in walkability assessment. *Transportation research record*, 2661(1):12–18, 2017.

[13] M. Ben-Or and P. Tiwari. A deterministic algorithm for sparse multivariate polynomial interpolation. In *Proceedings of the twentieth annual ACM symposium on Theory of computing*, pages 301–309. ACM, 1988.

[14] J. Benesty, J. Chen, Y. Huang, and I. Cohen. Pearson correlation coefficient. In *Noise reduction in speech processing*, pages 1–4. Springer, 2009.

[15] T. Blu, P. Thévenaz, and M. Unser. Linear interpolation revitalized. *IEEE Transactions on Image Processing*, 13(5):710–719, 2004.

[16] A. B. Bondi. Characteristics of scalability and their impact on performance. In *Proceedings of the 2nd international workshop on Software and performance*, pages 195–203, 2000.

[17] D. B. Bong, K. C. Lai, and A. Joseph. Automatic road network recognition and extraction for urban planning. *International Journal of Applied Science, Engineering and Technology*, 5(1):209–215, 2009.

[18] R. Borgoni, A. Gilardi, and D. Zappa. Assessing the risk of car crashes in road networks. *Social Indicators Research*, 156(2):429–447, 2021.

[19] U. Brandes. A faster algorithm for betweenness centrality. *Journal of mathematical sociology*, 25(2):163–177, 2001.

[20] U. C. Bureau. *2017 American Community Survey 1-Year Estimates*, 2017.

[21] Canadian Urban Environmental Health Research consortium (CANUE). *Canadian Active Living Environments Database (Can-ALE)*, 2016 (accessed May 3, 2020). `https://canue.ca/data/`.

[22] Canadian Urban Environmental Health Research consortium (CANUE). *Canadian Active Living Environments Database (Can-ALE) User Manual & Technical Document*, 2018 (accessed May 3, 2020). `http://canue.ca/wp-content/uploads/2018/03/CanALE_UserGuide.pdf`.

[23] L. J. Carr, S. I. Dunsiger, and B. H. Marcus. Validation of Walk Score for estimating access to walkable amenities. *British journal of sports medicine*, 45(14):1144–1148, 2011.

[24] T. Chai and R. R. Draxler. Root mean square error (RMSE) or mean absolute error (MAE). *Geoscientific Model Development Discussions*, 7(1):1525–1534, 2014.

[25] S. H. Chan, R. V. Donner, and S. Lämmer. Urban road networks—spatial networks with universal geometric features? *The European Physical Journal B*, 84(4):563–577, 2011.

[26] M. C. Cieslak, A. M. Castelfranco, V. Roncalli, P. H. Lenz, and D. K. Hartline. t-Distributed Stochastic Neighbor Embedding (t-SNE): A tool for eco-physiological transcriptomic analysis. *Marine genomics*, 51:100723, 2020.

[27] R. C. Colley, G. Butler, D. Garriguet, S. A. Prince, and K. C. Roberts. Comparison of self-reported and accelerometer-measured physical activity among canadian youth. *Health Rep*, 30(7):3–12, 2019.

[28] R. C. Colley, T. Christidis, I. Michaud, M. Tjepkema, and N. A. Ross. An examination of the associations between walkable neighbourhoods and obesity and self-rated health in Canadians. *Health reports*, 30(9):14–24, 2019.

[29] R. C. Colley, T. Christidis, I. Michaud, M. Tjepkema, and N. A. Ross. The association between walkable neighbourhoods and physical activity across the lifespan. *Health reports*, 30(9):3–14, 2019.

[30] C. Consortium. *Cytoscape: An Open Source Platform for Complex Network Analysis and Visualization*. Accessed Oct 2021, `https://cytoscape.org/`.

[31] L. da F. Costa. *What is a Complex Network? (CDT-2)*, 04 2018.

[32] T. Daily. Study: Long commutes to work by car. *Statistics Canada*, 2019.

[33] C. De Boor. Bicubic spline interpolation. *Journal of mathematics and physics*, 41(1-4):212–218, 1962.

[34] I. Deep AI. *What is an Activation Function?* Accessed Oct 2021, `https://deepai.org/machine-learning-glossary-and-terms/activation-function`.

[35] I. Deep AI. *What is ReLu?* Accessed Oct 2021, `https://deepai.org/machine-learning-glossary-and-terms/relu`.

[36] F. Diaz, J. Abbassi, D. Fuller, and E. n. Diab. Canadian transit agencies response to covid-19: Understanding strategies, information accessibility and the use of social media. *Transportation Research Record*, 2020.

[37] D. T. Duncan, J. Aldstadt, J. Whalen, S. J. Melly, and S. L. Gortmaker. Validation of Walk Score® for estimating neighborhood walkability: an analysis of four US metropolitan areas. *International journal of environmental research and public health*, 8(11):4160–4179, 2011.

[38] T. Dutzik, J. Inglis, and P. Baxandall. Millennials in motion: Changing travel Habits of young Americans and the implications for public policy. *U.S. PIRG Education Fund, Frontier Group*, 2014.

[39] S. D'Haese, G. Vanwolleghem, E. Hinckson, I. De Bourdeaudhuij, B. Deforche, D. Van Dyck, and G. Cardon. Cross-continental comparison of the association between the physical environment and active transportation in children: a systematic review. *International journal of behavioral nutrition and physical activity*, 12(1):1–14, 2015.

[40] I. Environmental Systems Research Institute. *What is a shapefile?* Accessed Oct 2021, `https://desktop.arcgis.com/en/arcmap/10.3/manage-data/shapefiles/what-is-a-shapefile.htm`.

[41] R. Ewing and R. Cervero. Travel and the built environment: a meta-analysis. *Journal of the American planning association*, 76(3):265–294, 2010.

[42] R. Ewing and S. Handy. Measuring the unmeasurable: Urban design qualities related to walkability. *Journal of Urban design*, 14(1):65–84, 2009.

[43] R. Ewing, S. Handy, R. C. Brownson, O. Clemente, and E. Winston. Identifying and measuring urban design qualities related to walkability. *Journal of Physical Activity and Health*, 3(s1):S223–S240, 2006.

[44] A. Forsyth. What is a walkable place? The walkability debate in urban design. *Urban design international*, 20(4):274–292, 2015.

[45] A. Forsyth. What is a walkable place? The walkability debate in urban design.

Urban design international, 20(4):274–292, 2015.

[46] R. Fry, A. Akbari, S. Mah, and N. Ross. Measuring Active Living Environments: An international comparison between Canada and Wales. *International Journal of Population Data Science*, 3(2), 2018.

[47] D. Fuller, R. Buote, and K. Stanley. A glossary for big data in population and public health: discussion and commentary on terminology and research methods. *J Epidemiol Community Health*, 71(11):1113–1117, 2017.

[48] L. Gauvin, L. Richard, C. L. Craig, M. Spivock, M. Riva, M. Forster, S. Laforest, S. Laberge, M.-C. Fournel, H. Gagnon, et al. From walkability to active living potential: an "ecometric" validation study. *American journal of preventive medicine*, 28(2):126–133, 2005.

[49] G. Geography. *World Geodetic System (WGS84)*. Accessed Oct 2021, `https://gisgeography.com/wgs84-world-geodetic-system/`.

[50] A. Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, 2019.

[51] R. H. Glazier, M. I. Creatore, J. T. Weyman, G. Fazli, F. I. Matheson, P. Gozdyra, R. Moineddin, V. K. Shriqui, and G. L. Booth. Density, destinations or both? A comparison of measures of walkability in relation to transportation behaviors, obesity and diabetes in Toronto, Canada. *PloS one*, 9(1):e85295, 2014.

[52] M. C. Gonzalez, C. Hidalgo, and A.-L. Barabasi. Understanding Individual Human Mobility Patterns. *Nature*, 453:779–82, 07 2008.

[53] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT press, 2016.

[54] B. Grant. Getting to great places: how better urban design can strengthen San Jose's future. *Spur report. San jose*, 2013.

[55] A. Grover and J. Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864, 2016.

[56] S. Hajna, N. A. Ross, S. J. Griffin, and K. Dasgupta. Lexical neutrality in environmental health research: Reflections on the term walkability. *BMC public health*, 17(1):1–4, 2017.

[57] Harshit Khandelwal. *A Closer Look into the Major Types of Machine Learning Models*, 2019 (accessed Dec 11, 2020). `https://becominghuman.ai/a-closer-look-into-the-major-types-of-machine-learning-models-77164a47012?gi=cb94b231462f`.

[58] T. Herrmann, W. Gleckner, R. A. Wasfi, B. Thierry, Y. Kestens, and N. A. Ross. A pan-Canadian measure of active living environments using open data. *Health reports*, 30(5):16–26, 2019.

[59] M. D. Hill. What is scalability? *ACM SIGARCH Computer Architecture News*, 18(4):18–21, 1990.

[60] J. A. Hirsch, D. N. DeVries, M. Brauer, L. D. Frank, and M. Winters. Impact

of new rapid transit on physical activity: a meta-analysis. *Preventive medicine reports*, 10:184–190, 2018.

[61] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[62] IBM Cloud Education. *Machine Learning*, 2020 (accessed Dec 11, 2020). `https://www.ibm.com/cloud/learn/machine-learning`.

[63] E. Ivanová and J. Masarova. Importance of road infrastructure in the economic development and competitiveness. *ECONOMICS AND MANAGEMENT*, 18, 08 2013.

[64] I. J. Janine Clarke, Rachel Colley and M. S. Tremblay. Accelerometer-measured moderate-to-vigorous physical activity of canadian adults, 2007 to 2017. *Statistics Canada Health Reports*, 2019.

[65] Jason Brownlee. *A Gentle Introduction to k-fold Cross-Validation*, 2020 (accessed Nov 17, 2020). `https://machinelearningmastery.com/k-fold-cross-validation/`.

[66] P. Jia, X. Pan, F. Liu, P. He, W. Zhang, L. Liu, Y. Zou, and L. Chen. Land use mix in the neighbourhood and childhood obesity. *Obesity Reviews*, 2020.

[67] B. Jiang, J. Yin, and S. Zhao. Characterizing the human mobility pattern in a large street network. *Physical Review E*, 80(2):021136, 2009.

[68] M. Kärmeniemi, T. Lankila, T. Ikäheimo, H. Koivumaa-Honkanen, and R. Korpelainen. The built environment as a determinant of physical activity: a system-

atic review of longitudinal studies and natural experiments. *Annals of behavioral medicine*, 52(3):239–251, 2018.

[69] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.

[70] N. J. Klein and M. J. Smart. Millennials and car ownership: Less money, fewer cars. *Transport Policy*, 53:20–29, 2017.

[71] H. W. Kohl 3rd, C. L. Craig, E. V. Lambert, S. Inoue, J. R. Alkandari, G. Leetongin, S. Kahlmeier, L. P. A. S. W. Group, et al. The pandemic of physical inactivity: global action for public health. *The lancet*, 380(9838):294–305, 2012.

[72] S. Kunkle, G. Christie, D. Yach, and A. M. El-Sayed. The importance of computer science for public health training: an opportunity and call to action. *JMIR public health and surveillance*, 2(1):e10, 2016.

[73] S. Lilley. Lost in translation. In *System Failure Case Studies*, volume 3. National Aeronautics and Space Administration, 2009.

[74] T. A. Litman. Economic value of walkability. *Transportation Research Record*, 1828(1):3–11, 2003.

[75] T. A. Litman. *Economic value of walkability*. Victoria Transport Policy Institute Canada, 2017.

[76] X. Liu, Z. Liu, G. Wang, Z. Cai, and H. Zhang. Ensemble transfer learning algorithm. *IEEE Access*, 6:2389–2396, 2017.

[77] R. H. Lo. Walkability: what is it? *Journal of Urbanism*, 2(2):145–166, 2009.

[78] B. Lu. *shp2graph: Convert a SpatialLinesDataFrame Object to an 'igraph'-Class Object*, 2018. R package version 0-5, `https://CRAN.R-project.org/package=shp2graph`.

[79] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, and G. Zhang. Transfer learning using computational intelligence: A survey. *Knowledge-Based Systems*, 80:14–23, 2015.

[80] R. Mahabir, D. Severson, and D. Chadee. Impact of road networks on the distribution of dengue fever cases in trinidad, west indies. *Acta tropica*, 123(3):178–183, 2012.

[81] N. Maizlish, J. Woodcock, S. Co, B. Ostro, A. Fanai, and D. Fairley. Health cobenefits and transportation-related reductions in greenhouse gas emissions in the san francisco bay area. *American journal of public health*, 103(4):703–709, 2013.

[82] M. M. A. Manan, A. Várhelyi, A. K. Çelik, and H. H. Hashim. Road characteristics and environment factors associated with motorcycle fatal crashes in Malaysia. *IATSS research*, 42(4):207–220, 2018.

[83] K. Manaugh and A. El-Geneidy. Validating walkability indices: How do different households respond to the walkability of their neighborhood? *Transportation research part D: transport and environment*, 16(4):309–315, 2011.

[84] S. C. Minister of Industry. *2011 National Household Survey (NHS)*, 2013.

[85] ML Glossary - Read The Docs. *Linear Regression*, 2017 (accessed Dec

11, 2020). https://ml-cheatsheet.readthedocs.io/en/latest/linear_regression.html.

[86] P. Morency, J. Archambault, M.-S. Cloutier, M. Tremblay, and C. Plante. Major urban road characteristics and injured pedestrians: A representative survey of intersections in Montréal, Quebec. *Canadian journal of public health*, 106(6):e388–e394, 2015.

[87] Mudgalvivek - Medium. Machine Learning: Clustering: Elbow method, 2020 (accessed Feb, 2021). https://medium.com/@mudgalvivek2911/machine-learning-clustering-elbow-method-4e8c2b404a5d.

[88] M. Newman. *Networks: An Introduction*. Oxford University Press, Inc., New York, NY, USA, 2010.

[89] F. H. M. Oliveira, A. R. Machado, and A. O. Andrade. On the use of t-distributed stochastic neighbor embedding for data visualization and classification of individuals with Parkinson's disease. *Computational and mathematical methods in medicine*, 2018, 2018.

[90] OpenStreetMap contributors. Key:amenity - OpenStreetMap Wiki, 2020 (accessed May 3, 2020). https://wiki.openstreetmap.org/wiki/Key:amenity.

[91] OpenStreetMap contributors. *OpenStreetMap*, 2020 (accessed May 3, 2020). https://www.openstreetmap.org.

[92] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[93] pandas development team. pandas.DataFrame, 2021 (accessed March, 2021). `https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html`.

[94] T. Paul, K. Stanley, N. Osgood, S. Bell, and N. Muhajarine. Scaling Behavior of Human Mobility Distributions. In J. A. Miller, D. O'Sullivan, and N. Wiegand, editors, *Geographic Information Science*, pages 145–159, Cham, 2016. Springer International Publishing.

[95] J. Paulo Dos Anjos Souza Barbosa, P. Henrique Guerra, C. de Oliveira Santos, A. P. de Oliveira Barbosa Nunes, G. Turrell, and A. Antonio Florindo. Walkability, overweight, and obesity in adults: a systematic review of observational studies. *International journal of environmental research and public health*, 16(17):3135, 2019.

[96] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[97] Public Health Ontario. Walkability. `https://www.publichealthontario.ca/en/BrowseByTopic/HealthPromotion/Pages/Walkability.aspx`, 2018.

[98] QGIS Development Team. *Quantum Geographic Information System (QGIS)*. Open Source Geospatial Foundation, 2020 (accessed May 3, 2020).

[99] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. https://www.R-project.org/.

[100] M. Raifer. *Overpass turbo*. Accessed Oct 2021, http://overpass-turbo.eu/.

[101] L. F. Ribeiro, P. H. Saverese, and D. R. Figueiredo. struc2vec: Learning node representations from structural identity. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 385–394, 2017.

[102] Rogelio A. Hernandez-Lopez. Complex networks and collective behavior in nature, 2010 (accessed May 3, 2020). http://web.mit.edu/8.334/www/grades/projects/projects10/Hernandez-Lopez-Rogelio/structure_1.html.

[103] RStudio. *RStudio — Open source & professional software for data science teams*, (accessed May 3, 2020). https://www.rstudio.com/.

[104] Saed Sayad. *Decision Tree - Regression*, 2020 (accessed Dec 11, 2020). https://www.saedsayad.com/decision_tree_reg.htm.

[105] A. E. Sariyüce, K. Kaya, E. Saule, and Ü. V. Çatalyiirek. Incremental algorithms for closeness centrality. In *2013 IEEE International Conference on Big Data*, pages 487–492. IEEE, 2013.

[106] I. J. Schoenberg. *Cardinal spline interpolation*, volume 12. Siam, 1973.

[107] scikit-learn. *RandomForestRegressor*, 2020 (accessed Dec 11, 2020). https://scikit-learn.org/stable/modules/generated/sklearn.

ensemble.RandomForestRegressor.html.

[108] scikit-learn Machine Learning in Python. Selecting the number of clusters with silhouette analysis on KMeans clustering, 2020 (accessed Feb, 2021). https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html.

[109] M. Smith, J. Hosking, A. Woodward, K. Witten, A. MacMillan, A. Field, P. Baas, and H. Mackie. Systematic literature review of built environment effects on physical activity and active transport–an update and new findings on health equity. *International journal of behavioral nutrition and physical activity*, 14(1):1–27, 2017.

[110] S. Spoon. *What defines walkability: Walking behavior correlates*, 2005.

[111] Statistics Canada. Dissemination area (DA), 2020 (accessed Dec 9, 2020). https://www12.statcan.gc.ca/census-recensement/2011/ref/dict/geo021-eng.cfm.

[112] Statistics Canada. Dissemination area: Detailed definition, 2020 (accessed May 3, 2020). https://www150.statcan.gc.ca/n1/pub/92-195-x/2011001/geo/da-ad/def-eng.htm.

[113] Statistics Canada. Statistics Canada: Canada's national statistical agency, 2020 (accessed May 3, 2020). https://www.statcan.gc.ca/eng/start.

[114] U. G. Survey. *What does the term UTM mean? Is UTM better or more accurate than latitude/longitude?* Accessed Oct 2021, https://www.usgs.gov/faqs/

`what-does-term-utm-mean-utm-better-or-more-accurate-latitudelongitude`.

[115] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu. A survey on deep transfer learning. In *International Conference on Artificial Neural Networks*, pages 270–279. Springer, 2018.

[116] Z. Tian, L. Jia, H. Dong, F. Su, and Z. Zhang. Analysis of urban road traffic network based on complex network. *Procedia engineering*, 137:537–546, 2016.

[117] M. Tobin and et al. Active Living Environments: Re-Framing and Defining the Concept of Walkability. *BMC Public Health*, 2020.

[118] M. G. Tolsgaard, C. K. Boscardin, Y. S. Park, M. M. Cuddy, and S. S. Sebok-Syer. The role of data science and machine learning in health professions education: practical applications, theoretical contributions, and epistemic beliefs. *Advances in Health Sciences Education*, pages 1–30, 2020.

[119] L. Torrey and J. Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI Global, 2010.

[120] L. Tseng, P. Heffernan, and D.-T. Lee. Two-guard walkability of simple polygons. *International Journal of Computational Geometry & Applications*, 8(01):85–116, 1998.

[121] Urban Securipedia. Road network, 2013 (accessed May 3, 2020). `https://securipedia.eu/mediawiki/index.php/Road_network`.

[122] F. H. A. F. U.S. Department of Transportation. *2009 National Household Travel*

Survey (NHTS), 2011.

[123] W3Schools. Pandas - Cleaning Data, 2021 (accessed March, 2021). https://www.w3schools.com/python/pandas/pandas_cleaning.asp.

[124] Walk Score (R). *Walk Score (R)*, 2020 (accessed May 3, 2020). https://www.walkscore.com/.

[125] C. Wang, C. Ducruet, and W. Wang. Evolution, accessibility and dynamics of road networks in china from 1600 bc to 1900 ad. *Journal of Geographical Sciences*, 25(4):451–484, 2015.

[126] H. Wang and Y. Yang. Neighbourhood walkability: A review and bibliometric analysis. *Cities*, 93:43–61, 2019.

[127] K. Weiss, T. M. Khoshgoftaar, and D. Wang. A survey of transfer learning. *Journal of Big Data*, 3(1):1–40, 2016.

[128] Wikipedia. Walkability. https://en.wikipedia.org/wiki/Walkability, 2018.

[129] Wikipedia. *Gradient boosting*, 2020 (accessed Dec 11, 2020). https://en.wikipedia.org/wiki/Gradient_boosting.

[130] Wikipedia. Biological network, 2020 (accessed May 3, 2020). https://en.wikipedia.org/wiki/Biological_network.

[131] Wikipedia. Computer network, 2020 (accessed May 3, 2020). https://en.wikipedia.org/wiki/Computer_network.

[132] Wikipedia. Social network, 2020 (accessed May 3, 2020). https://en.wikipedia.org/wiki/Social_network.

[133] M. Winters, K. Teschke, M. Brauer, and D. Fuller. Bike Score®: Associations between urban bikeability and cycling behavior in 24 cities. *International journal of behavioral nutrition and physical activity*, 13(1):18, 2016.

[134] C. Xiao, Y. Goryakin, and M. Cecchini. Physical activity levels and new public transit: a systematic review and meta-analysis. *American journal of preventive medicine*, 56(3):464–473, 2019.

[135] H. Xue, X. Cheng, P. Jia, and Y. Wang. Road network intersection density and childhood obesity risk in the US: a national longitudinal study. *Public health*, 178:31–37, 2020.

[136] K. YADA, Y. INABA, and Y. ISEDA. Study on relation between walkability and joints of block pavement. *JOURNAL OF PAVEMENT ENGINEERING, JSCE*, 2:189–196, 1997.

[137] Y. Yao and G. Doretto. Boosting for transfer learning with multiple sources. In *2010 IEEE computer society conference on computer vision and pattern recognition*, pages 1855–1862. IEEE, 2010.

[138] B.-L. Ye, W. Wu, and W. Mao. A method for signal coordination in large-scale urban road networks. *Mathematical Problems in Engineering*, 2015, 2015.

[139] N. J. Yuan, Y. Zheng, and X. Xie. Segmentation of urban areas using road networks. *Microsoft Corp., Redmond, WA, USA, Tech. Rep. MSR-TR-2012-*

65, 2012.

[140] L. Zhang and X. Gao. Transfer adaptation learning: A decade survey. *arXiv preprint arXiv:1903.04687*, 2019.

[141] R. Zhang, K. G. Stanley, D. Fuller, and S. Bell. Differentiating Population Spatial Behavior Using Representative Features of Geospatial Mobility (ReFGeM). *ACM Trans. Spatial Algorithms Syst.*, 6(1), Feb. 2020.

[142] W. Zhu, Z. T. Webb, K. Mao, and J. Romagnoli. A deep learning approach for process data visualization using t-distributed stochastic neighbor embedding. *Industrial & Engineering Chemistry Research*, 58(22):9564–9575, 2019.

[143] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2020.