Innovative Technologies for Pan-Canadian Surveillance of Climate Change

Ruth Rebutoc, MSc.

Joon Lee, Ph.D.

Submitted to Health Canada

August 30, 2019

Table of Contents

List of Tables

Executive Summary

The purposes of this environmental scan are:

- To provide an overview of the technologies of artificial intelligence (AI), machine learning, and big data;
- To explain how these technologies can be used to monitor and forecast the public health effects of climate change; and
- To describe the current (and near future) state of the human resources (HR) and data requirements necessary for this goal.

In 2008, the Government of Canada published a report (Lemmen, Warren, Lacroix, & Bush, 2008) describing the current and expected impacts of climate change across the country. This report acknowledged that global temperatures have reached a tipping point making many of the consequences unavoidable. As a result, the report included a table of public health risks (see Table 1) including the spread of vector-borne illness; heat-related morbidity and mortality caused by more frequent heatwaves; severe storms resulting in flooding, hurricanes, and wildfires; and air quality impacts from increased pollen and wildfire emissions.

Understanding the catalysts and effects of these events requires access to and analysis of multiple, diverse datasets. For example, monitoring wildfire emissions and forecasting how they will spread downwind requires data from satellites and air quality monitors, as well as weather forecasts. This paper examines the potential of machine learning to facilitate this data integration process due to its ability to analyze big data; find patterns, categorize information, or make predictions; learn from the information; and provide near real-time feedback (Alpaydin, 2014).

Results from the environmental scan indicated the use of machine learning in in research related to the health effects of climate change is relatively new and more advanced in some areas than others. For example, the use of machine learning to track wildfire emissions is more advanced than in research tracking Lyme disease because of access to vast amounts of data available in the former area (e.g. satellite images, ground-level radar, air quality monitors), which can be used to train machine learning algorithms.

Being able to conduct this research will require very specific domain knowledge. For example, in reference to wildfire emissions, researchers would need to know: what climate variables will increase the dispersal of emissions; what concentration of emissions is harmful to people with respiratory illness (and the general public) and when is this level likely to be reached; what types and sources of data are needed; and what category of machine learning algorithm (i.e. supervised, unsupervised, etc.) is best suited to answer these questions. It is improbable that any one person will have expertise in all three domains. Therefore, this environmental scan included a search of the Canadian Occupational Projection System (COPS), as well as Canadian and United States' (U.S.) universities in order to assess the available HR capacity. The COPS search predicted a deficit of data scientists and computer programmers until at least 2026 (Government of Canada, 2019a). Additionally, no university offered a program with a focus in all three disciplines. However, a few Canadian universities (such as the University of Toronto) have begun to offer degrees in public health with the option to minor or specialize in climate change or environmental studies. The top ten Canadian universities also offer programs and/or courses in data science, AI and machine learning - presenting the possibility of modifying existing programs to incorporate all three topics.

The rest of the paper is organized as follows: Section 1 contains the Introduction, Methods, and technology portions (i.e. AI, machine learning, big data). The Introduction briefly explains the public health risks of climate change in Canada and the possibility of incorporating machine learning to better understand the problems. Next, the Methods section describes the database and grey literature search. The Methods section is followed by a section describing AI, machine learning, and big data including definitions, basic principles, and examples of their use outside healthcare and climate change research.

Section 2, is organized by individual public health risk as designated by the Government of Canada (Public Health Agency of Canada, 2013a) (see Table 1).). In addition to these risks, Food Security was added because of the consequences currently being faced by Aboriginal communities in Canada (particularly the Arctic) such as the early breakup of sea ice. This section was designed to function as separate papers describing the individual public health risk, current method of monitoring, and potential of machine learning to improve on current methods along with the necessary data and infrastructure. If machine learning research was discovered in the database search, detail was included about the algorithm(s) used and the types and sources of data. This detail was given to serve as a template for evaluation and use by future researchers. The number of subsections for a given risk differs depending on the amount of research available on monitoring or use of machine learning, and level of detail included in that research.

Finally, Section 3 explains the infrastructure and HR requirements, and gives the authors' recommendations. Based on the findings from the environmental scan, recommendations for the design, development, and implementation of the surveillance system are provided. The section detailing the HR requirements discusses the current and near future capability of the Canadian workforce to use machine learning in climate change and public health research. This section also contains a discussion about relevant programs and course-work offered by the top ten Canadian universities, and competition by the private sector for these graduates.

1.1 Introduction

In 2016, the World Health Organization declared climate change, "the greatest threat to global health in the 21st Century" (World Health Organization, 2016). Climate change refers to a significant and sustained change in global average weather patterns. Previously known as global warming, the term climate change has been adopted in recognition of the vast array of weather-related changes that are becoming the new normal including increased seasonal temperatures, more intense storms, and changing patterns of precipitation (Public Health Agency of Canada, 2013a).

By the end of the century, global temperatures are expected to be 2° Celsius to 6° Celsius higher than at present (National Aeronautics and Space Administration, 2012). While the impact will be global, the effects of climate changes will vary by region. Given Canada's vast land mass and topography, the country is expected to face an array of consequences. Canada occupies 9.985 million km^2, is bordered on three sides by ocean (i.e. Atlantic, Pacific, Arctic), and spans 41 degrees of latitude from 42°N to 83°N (Encyclopaedia Britannica, 2019; Séguin et al., 2008). In general, Canadian provinces are expected to experience milder winters, warmer summers, and changing patterns of precipitation including droughts and severe storms (Public Health Agency of Canada, 2013a). More specifically, while temperatures across the country are expected to rise, the greatest warming will be seen in the Prairie provinces and Arctic region. The Arctic is predicted to see the greatest temperature rise accompanied by loss of permafrost and sea ice. Prairie provinces can also expect hotter summers, periods of drought, and periods of intense precipitation with increased risk of flooding. Sea level rise on the East and West coast will increase the risk of coastal flooding, while the West coast is expected to experience more drought-like conditions through the summer months (Public Health Agency of Canada, 2013a) .

The problems created by these changes are significant given that societies are adapted to exist within long-standing patterns of weather. Our buildings, infrastructure, livelihoods, culture, and types and sources of food are influenced by the local climate.

In *From Impacts to Adaptation: Canada in a Changing Climate* (Lemmen et al., 2008), the Government of Canada stated, the "impacts of changing climate on many physical and biological systems, such as ice and snow cover, river, lake and sea levels, and plant and animal distributions, are unequivocal" (Lemmen et al., 2008, p. 8). This report provided comprehensive evidence of the effects of climate change at regional and national levels, including specific impacts to public health.

Ultimately, the information contained in this report led to a list of four categories of public health risk (see Table 1) (Public Health Agency of Canada, 2013a). Warming temperatures are expected to facilitate the spread of infectious diseases as mosquitos and ticks move farther north. Water- and food-borne pathogens are also expected to increase, for example warming water temperatures promote the growth of toxic algae plumes, while warming air temperatures aid the growth and survival of E. coli (Public Health Agency of Canada, 2019). As temperatures rise, so too will cases of heat stroke and asthma as populations acclimated to temperate climates experience more and longer heat waves. Asthma and allergy sufferers will also be affected indirectly as warmer temperatures expand the growing season and increase the allergenicity of some pollen (Héguy et al., 2008). Finally, extreme heat, pollen, and wildfires are all expected to negatively impact air quality, increasing adverse outcomes for people with chronic respiratory illnesses (Gerardi & Kellerman, 2014).

As acknowledged in *From Impacts to Adaptation* (Lemmen et al., 2008), these consequences are already being experienced and are expected to worsen across Canada. Many of these changes cannot be prevented, therefore, the key challenge for provincial/territorial and federal governments, as well as healthcare professionals will be knowing how to mitigate the consequences. In this paper, AI, big data, and machine learning are presented as options for monitoring and forecasting some of the consequences of climate change in order to mitigate the impacts to public health.

Big data refers to the large volume of diverse, quickly changing data (Barton, 2019), while AI and machine learning are branches of computer science associated with being able to process and analyze such information in an "intelligent" manner by using methods such as reasoning or trial-and-error (Abedini et al., 2015; Singh, Martins, Joanis, & Mago). Machine learning algorithms may be particularly useful for monitoring and forecasting the effects of climate change because they are used for solving problems, categorizing information, and making predictions. These algorithms also have the potential to quickly analyze large quantities of data and provide vital information in near real-time. As an example, satellites have become an attractive option for finding and monitoring wildfires because, unlike ground-based sensors, they can scan Canada's vast landmass multiple times a day. Machine learning algorithms are already being used to analyze this information in order to determine fire intensity and forecast the distribution of smoke downwind (Yao, Raffuse, et al., 2018).

This research aims to address four objectives set by the Statement of Work from the Climate Change and Innovation Bureau of Health Canada. Briefly these were to: 1) provide an overview of big data, AI, and machine learning; 2) explain how these technologies are currently being used, or are proposed for use in other sectors and/or government applications; 3) describe how the health effects of climate change are currently being monitored in Canada; 4) provide recommendations for how machine learning, AI, and big data could be used to improve these methods; and 5) outline the broad human resource (HR) and infrastructure requirements necessary for this task.

Table 1. ***Health Risks Associated with Climate Change***

Infectious Disease	Extreme Weather Events	Higher Temperatures	Air Quality
Water- and food-borne	Severe Storms	Heat Stroke	Heat
Vector	Wildfires	Pollen, Allergens, Asthma	Pollen
	Hurricanes and Flooding		

1.2 Methods

Based on the objectives outlined in the statement of work, and identified risks to public health, specific combinations of key words were used to search targeted sources of grey literature, and the following electronic databases: Scopus, PubMed, CAB Abstracts, and Environment Complete.

The search strategy, including selection of databases was developed in consultation with a University of Calgary health sciences librarian and an expert in the field of machine learning. The search was limited to English-language articles with full text online published between January 1, 2000 and August 1, 2019. Conference abstracts captured through the database search were also included when relevant. The following keywords were used, in varying combinations, to search the databases: machine learning, big data, public health, climate change or global warming (including words specific to Government of Canada identified risks) and food shortage, both with and without Canada. Searches were limited to the title, abstract, and keywords of each article. To see a list of the database results see Table 2.

Table 2. ***Database Results***

Public Health Risks of Climate Change in Canada[1]	Database Results
Infectious Disease: Vector, food- and water-born, West Nile, Lyme	69
Extreme Weather: Floods, Hurricanes, Wildfires	754
Higher Temperatures: Heat Event, Heat Stroke, High Temperatures, Heat Wave	1022
Air Quality: Pollen, Aeroallergens, Allergens, Asthma, Respiratory Illness	310
Food Security:	129

[1] In all cases, for risk related results, the keywords were paired with “climate change” and/or “global warming”, and Canada/Canadian

The inclusion of grey literature, including news articles, was also guided by the Canadian Agency for Drugs and Technologies in Health (CADTH) grey literature checklist (Canadian Agency for Drugs and Technologies in Health, 2019), and the University of Toronto grey literature searching guide: “How to Find and Document Grey Literature” (University of Toronto, 2019). The search of grey literature was targeted – based on the research questions.

The process was iterative. Once the initial database search was complete, the reference section of some articles was explored, and targeted searches were conducted in some journals as well as Google Scholar when more detail was required about a particular topic. These articles are not included in the numbers shown on the PRISMA chart, but all articles cited are listed in the References section.

1.3 Artificial Intelligence, Machine Learning and Big Data

Big data is a term used to describe a data set that is "diverse, complex, disorganized, massive, and multimodal" (National Institutes of Health, 2019). More simply, big data has often been associated with "5 Vs" that refer to the *volume* of data included in a data set, the *veracity* (i.e. accuracy and credibility) of the sources, the *velocity* at which this data is generated and changed, the *variety* or diversity of sources, and the *value* the data adds to the analysis (Anuradha, 2015). Depending on the topic of investigation, big data might also include sources like social media posts, audio, video, images, and documents such as medical records. Given the dynamic nature of such data, time often becomes an additional variable.

The challenges associated with big data have fuelled the rapid growth of data science as a field in recent years. Data science brings together computer science and statistics to extract useful information and knowledge from data (Provost, 2013). Since the complexity and scale of big data preclude the use of conventional methods and software, data science focuses on utilizing advanced computing techniques such as cloud, parallel, and high-performance computing. Furthermore, since most big data sets are collected from the real world with minimal data cleansing, an important component of data science is to address the "messiness" of big data, which includes missing data, outliers, errors, and discrepancies. Another major component of data science is visualization, where the goal is to effectively communicate data characteristics and information extracted from the data in an easy-to-understand, visual manner.

Big data and data science are closely related to machine learning and AI, which fall under the umbrella of computer science. AI, in the simplest terms, is technology that has been programmed to, and is capable, of problem-solving techniques associated with *human* intelligence such as reasoning, pattern recognition, the ability to adapt to new circumstances by drawing on and generalizing from stored knowledge, and learning through trial and error (Schmelzer, 2019). Common application areas include natural language processing (enabling computers to analyze human language data), image recognition, and/or robotics (Lavigne, Mussa, Creatore, Hoffman, & Buckeridge, 2019).

To date, machine learning has been one of the most effective and successful ways of creating AI systems, although AI can be developed without any machine learning (e.g. rule-based systems). The widespread success of machine learning in many practical AI applications has made machine learning the dominant methodological choice in modern AI research, which has led to the terms AI and machine learning being used interchangeably. The strength of machine learning is largely attributable to its paradigm of letting a machine (such as a computer) learn from data on its own how to do a given task without explicit programming. Because machine learning heavily depends on the quality and quantity of the data it learns from, big data and data science are enormously important for its success. Machine learning, like statistics, can be used to create a model used for predictions or inferences. However, unlike statistics, machine learning does not need to begin with underlying assumptions about the data, such as it being evenly distributed around a mean, or that there is a relationship between the independent and dependent variables. The drawback of these assumptions is of course, if they are wrong, the results are likely misleading or erroneous.

Machine learning can be described as a four-step process whereby a computer 1) analyzes data, 2) “learns” from the data 3) automatically generates an algorithm based on the “knowledge” gained from the data, and 4) uses the algorithm to solve a problem, make a prediction, or categorize information. The data, as mentioned above, are often referred to as “big data” because they are usually collected in bulk from multiple, disparate sources. The system then “mines” the data looking for patterns, trends, or similar processes to inform the algorithm (Alpaydin, 2014).

While there are many different kinds of machine learning techniques, deep learning (particularly convolutional artificial neural networks) has been the most successful and popular technique in recent years, largely thanks to the substantial performance improvements it has brought to computer vision (e.g., object recognition in photos) and natural language understanding (e.g., digital assistants such as Siri and Alexa) applications (Lecun, 2015). Deep learning is a modern extension of artificial neural networks characterized by more sophisticated model architectures. Deep learning requires a large amount of data and substantial computing power; both of these criteria have only started to be satisfied in recent years thanks to the emergence of big data and powerful, yet affordable, computing resources.

The algorithms used to solve problems, make predictions, or categorize information are broadly part of three divisions of machine learning: supervised, unsupervised, and reinforcement learning (Vu et al., 2018). Supervised machine learning is used for regression or classification tasks. In this process, an algorithm is used to either assign input data to known categories (classification) or map the continuous relationship between input and output data (regression).

The process is called supervised machine learning because a person knowledgeable about the problem, first “supervises” the system by feeding it sample data that are labelled with the correct answers or output. Through this method, the system “learns” to create an algorithm that will generalize from the training data to previously unseen examples. Once the algorithm is established a set of unlabelled examples is entered, this time without the corresponding output in order to test the algorithm. This machine learning technique is used, for example, in the detection of credit card fraud. In this example, credit card transactions previously determined to be fraudulent (or not) would be used to train the system to recognize features of a transaction that make it fraudulent, such as purchases in New York and Australia on the same day.

In unsupervised machine learning, there are no pre-determined labels or relationships between input and output data. Instead, unsupervised machine learning is used to find latent patterns, associations, or structure within large data sets. (Alpaydin, 2014). This form of machine learning is chosen when researchers do not know the answer to the problem and/or, the amount of data is too large and complex to be labeled prior to training the algorithm. A familiar example of unsupervised machine learning is Amazon product recommendations. A person who buys a baby crib, bottles, and diapers might see recommendations for car seats below each item or when completing the purchase. These recommendations were automatically generated by the algorithm as it learned to recognize items commonly purchased together.

Finally, in reinforcement learning, the system is given a goal and a set of parameters, then allowed to use a process of trial and error to achieve that goal. Reinforcement learning is often described as an agent that must make decisions or choose a course of action in an environment. Each time a *correct* decision is made, or a goal is achieved (either the overall goal or a step in the process) the system receives a reward in the form of positive feedback. The aim is to achieve the objective by following an optimal sequence of actions that lead to the maximum reward. For this reason, when a sequence of steps or decisions is used, the feedback is tied to the entire sequence rather than individual steps within it. This process of learning through trial and error is by definition iterative, and in the case of reinforcement learning, often requires large amounts of data. For this reason, reinforcement learning is said to “hunger” for data. Indeed, it is often difficult to provide enough data to satisfy the algorithm. This is why reinforcement learning has mostly been limited to problem domains where data can be simulated such as board games (e.g. chess, Go), or robotics.

AI, machine learning, and big data technologies have come a long way and continue to advance rapidly. In medicine, for example, AI and machine learning have been some of the hottest research topics in recent years as demonstrated by high-profile op-ed articles (Darcy, Louie, & Roberts, 2016; Hinton, 2018; Naylor, 2018; Shortliffe & Sepulveda, 2018). However, it is important to note that what these technologies can realistically achieve today is still quite limited. Not all problems can be solved by these technologies, especially when the problem requires general AI (i.e., truly human-like intelligence with the ability to solve a wide range of problems) rather than narrow AI that can only perform well in a specific task (e.g., identify tumors in MRI images). In addition, regardless of problem domain, large-scale quality data must exist and be accessible before machine learning and AI research can begin. This is why within medicine the specialties with large-scale digital data, usually imaging data due to the success of deep learning in computer vision (e.g., ophthalmology (Gulshan et al., 2016), dermatology (Esteva et al., 2017), pathology (Janowczyk & Madabhushi, 2016)), are further ahead in harnessing AI.

In public health, social media and Internet data in general have also been extensively researched for population health surveillance (e.g., (J. Liu, Weitzman, & Chunara, 2017; Nguyen et al., 2016), most notably the Google Flu Trends (Ginsberg et al., 2009), despite being widely criticized for subsequent poor performance and eventually discontinued (Lazer, Kennedy, King, & Vespignani, 2014).

2 Risks of Climate Change to Public Health in Canada

2.1 Vector-borne Disease

Climate conditions such as average seasonal temperature and rainfall are determinants of the lifecycle and distribution of water-, food-, and vector-borne pathogens. As such, any change in these climate variables can augment a stage in the lifecycle of these pathogens. For example, increases in precipitation and temperature have been shown to increase the reproductive capacity and survivability of water- and food-borne pathogens, as well as hosts of vector-borne disease (Semenza et al., 2012). According to the Public Health Agency of Canada (PHAC), five pathogens, including the Norovirus (commonly referred to as the stomach flu) and Salmonella, are responsible for over 90 percent of water- and food-borne illnesses (Public Health Agency of Canada, 2019) infecting over 4 million Canadians each year. Generally causing enteric illness with symptoms of nausea, vomiting, diarrhea, headache, fever, chills, and muscle pain, most cases are mild and resolve within a few days (Public Health Agency of Canada, 2006). However, some cases lead to hospitalization and even death. Like vector-borne diseases, the lifecycle and range of these pathogens are influenced by climate variables including warming air and water temperatures, increased precipitation, and severe storms (Public Health Agency of Canada, 2019). Rising seasonal temperatures also have the potential to expand the range and viability of vectors, and food- and water-borne pathogens outside historically endemic areas (Public Health Agency of Canada, 2013a, 2013b). This is evidenced by the spread of Lyme disease and the West Nile virus in Canada (Gasmi et al., 2018; McPherson et al., 2017; Public Health Agency of Canada, 2017; Sonenshine, 2018).

2.1.1 Lyme Monitoring. No research was found using machine learning methods to monitor the spread of Lyme disease in or outside Canada. In Canada, Lyme disease is most often monitored through passive surveillance. This system relies on members of the public, clinicians, or veterinarians submitting ticks to laboratories for testing or the reporting of human Lyme disease cases by provincial and territorial public health organizations. Some active surveillance does occur in provinces with a higher prevalence of the blacklegged tick. Active surveillance involves the attempted collection of ticks through methods such as “drag sampling” in areas considered to be ecologically viable for the species (Clow et al., 2017). However, neither of these methods would produce enough data for use with machine learning. Additionally, if 20 percent of the ticks tested in an area are positive for Lyme, the public health agency may stop testing and simply treat persons or pets who are bitten with antibiotics (as is currently the case in Ottawa) (Rabson, 2019).

2.1.2 West Nile Potential of Machine Learning. Researchers are beginning to solve the problem of insufficient data in research on the West Nile Virus (WNV) using machine learning. Only four studies were found (one without full text online) and all were conducted in the U.S. One study (Young, 2013), used a cubist machine learning method, remote sensing data, environmental and climate variables, and the theory of landscape epidemiology to predict the incidence rate of WNV across the entire U.S. The cubist method uses inductive learning to create rule-based decision trees from the data (Young, 2013). Landscape epidemiology describes how the temporal dynamics of host, vector, and pathogen populations interact spatially within a permissive environment to enable transmission. GIS data consisted of topographic information such as the amount and type of vegetation in an area and elevation. Environmental variables included urban vs agricultural land cover. While climate data consisted of temperature and precipitation. The strength of the model varied dramatically over the study period (2003 to 2008). However, overall the study found "temperature, precipitation, elevation [the vegetation index], and land cover" were correlated (0.86) with new rates of WNV (Young, 2013, p. 247).

Another study (Keyel et al., 2019) used a random forest model to examine the relationship between 66 climate variables (e.g. temperature and precipitation) and 21 non-climate variables (e.g. agricultural vs urban land cover, population density) on the incidence rate of WNV in New York and Connecticut. The study found climate variables were the most highly correlated with new cases of WNV, specifically the average and low temperatures from July through September.

Finally, a study conducted in Chicago (Gardner et al., 2013) used regression tree and random forest to identify natural and artificial features around storm water catch basins that make them more likely to have mosquito larvae. Data included water samples from 15 catch basins in four Chicago municipalities. Water was tested for larvae and chemistry. In addition, a survey was taken of the type, height, and density of trees and shrubs as well as their distance to the catch basis. This study showed a positive correlation between the variables of tree density, shrub height, and water high in ammonia and/or nitrates and mosquito larvae. While the initial study is labour-intensive, the goal was to establish parameters that could be used for more targeted eradication of mosquitoes in the future.

2.1.3 Water- and Food-borne Illness – Potential of Machine Learning. Intense precipitation can lead to manure or sewage runoff into drinking or recreational water sources, which can result in gastroenteric illness from E. coli or *Giardia duodenalis*. However, no study used machine learning techniques to forecast this problem.

Only one study was found that used machine learning methods to monitor food- or water-borne illness. This study used a random forest model to forecast the incidence of toxic algae blooms off the coast of British Columbia (B.C.). Warming waters are contributing to the frequency of these blooms, which are known to carry a neurotoxin. The algae are absorbed by shellfish in the area increasing the likelihood that it could be ingested by humans leading to paralytic shellfish poisoning, a potentially fatal illness (Finnis, Krstic, McIntyre, Nelson, & Henderson, 2017). The shellfish are regularly monitored at multiple sites around the island. Therefore, the data from samples collected from 2002 to 2012 were analyzed along with environmental variables such as air temperature, sea surface salinity, sun light penetration through the water, upwelling (i.e. a rising of cooler nutrient-rich water), and sea surface temperature. These were not the only environmental variables, but these were the variables that had the most influence on the model. For more information see Finnis et al., (2017).

2.2 Air Quality: Heat and Aeroallergens

"Extreme heat is a leading cause of illness and death from weather-related hazards in Canada" (McLean et al., 2018). Symptom of heat related illnesses directly related to heat exposure include muscle cramps, heat rash, exhaustion, fainting, and heat stroke. Indirectly, extreme heat has also been correlated with urinary and kidney stones, stroke, sudden infant death syndrome, and adverse outcomes for persons with respiratory or heart disease.

Climate change is expected to increase the "frequency, length and severity" of extreme heat-events (Health Canada, 2011). Compounding this problem, heat-related mortality exists on a curve. Effects of a heat wave occur on the first day of the event, but heat-related illnesses and mortality may persist for three days after and up to 15 days for people with respiratory illness (Gachon et al., 2016). Additionally, persons residing in cities with less seasonal variability, are susceptible to heat stress at lower temperatures because they have less experience with periods of extremely hot weather. For example, one study found the temperature threshold in Calgary, for heat-related morbidity began at 20° Celsius. Similarly, areas of northern British Columbia saw susceptibility rise at 14° Celsius, while the temperature threshold in Windsor was 27° Celsius (McLean et al., 2018).

In addition to extreme heat events, the global temperature rise associated with climate change has also lengthened the growing season of pollen producing plants and increased the amount of pollen produced, as well as the allergenicity of some pollen (Héguy et al., 2008). This is another public health concern because numerous studies have shown a correlation between elevated pollen concentrations and increased emergency hospitalizations, particularly for children under 10, due to asthma-related symptoms (Gerardi & Kellerman, 2014; Héguy et al., 2008).

2.2.1 Heat Monitoring. As a result, Environment and Climate Change Canada (ECCC) has developed 22 regional thresholds at which heat warnings are issued to the public (Environment and Climate Change Canada, 2010). Alerts are triggered when temperatures are predicted to be 5° Celsius to 10° Celsius above average highs for two or more days. The thresholds also take nighttime lows into consideration, as well as health-related impacts and socio-economic vulnerabilities (Gachon et al., 2016; Johnson, 2018). The forecasts are produced using high resolution models and real-time analysis by meteorologists. Experienced meteorologists can easily detect weather patterns indicative of a heatwave allowing ECCC to anticipate a high temperature event up to 10 days in advance at a 25km resolution, down to 10km within two days of the event (Gachon et al., 2016).

A similar system was implemented across Quebec in 2010. Developed to provide "real-time Surveillance and Prevention of the impacts of Extreme Meteorological Events (SUPREME) on public health" (Toutant, Gosselin, Belanger, Bustinza, & Rivest, 2011). The model uses open source software to collect meteorological and health data, analyze the impact of forecasted meteorological events on public health, and disseminate warnings to public health officials, and emergency management coordinators (Gachon et al., 2016). Forecasts from Environment Canada are assessed using a 3-day weighted average and compared to regional thresholds established by the Quebec National Institute of Public Health. If a forecast meets or exceeds a threshold, an alert is automatically sent to the relevant municipal and public health authorities. SUPREME is considered relatively reliable based on an evaluation from 2010 to 2015, which found, of the 77 alerts issued, 44 correctly forecast episodes of extreme heat. In total 93 heat waves occurred during this period, meaning the system also missed issuing 16 alerts (Gachon et al., 2016).

Meteorological data includes forecasts collected hourly from Environment Canada of the minimum and maximum temperature, and humidity over the following 24, 48, and 72 hour periods; as well as air quality data from the Quebec Ministry of Sustainable Development Environment and Parks (Toutant et al., 2011). Health data is drawn from daily emergency department and hospital admissions, and a daily health report (Toutant et al., 2011). The model also collects geospatial data including Landsat images to identify urban heat islands, and census data (e.g. sociodemographic information, age, population density) for the purposes of identifying vulnerable populations. SUPREME integrates this information into a portal that users can query with any of the above variables. For example, to identify vulnerable populations, a user could search for urban heat islands in proximity of populations 75 years or older; or places capable of providing relief such as public buildings, cooling centres, pools, and beaches.

2.2.2 Heat - Potential of Machine Learning and Big Data. Outside Canada, the European Heat Wave Pipeline was recently developed "for predicting and mapping heat waves"

(Gobbi, Alikadic, Ylinen, Angaramo, & Furlanello, 2017, p. 3734). Input for the model includes a rasterized map of Europe showing a 14-day forecast of the maximum temperature - provided daily by the Finnish Meteorological Institute. (Maximum temperatures are calculated using an ensemble prediction model with a spatial resolution of 18 km.). In addition, 32 years of daily maximum temperature data from the European Climate Assessment and Dataset are used to establish heat wave thresholds. Together this information is used to produce a grid of thresholds that are assigned to 'Local Administrative Units'. These sources of information are fed into an algorithm that can map a heat wave, including information on "duration and intensity", to a specific, local administrative unit for the following 14-day period. Since the indications of a heat wave are mapped to a specific location and threshold, the data can be compared across regions and timeframes. Efficacy was demonstrated when the system was able to predict the 2017 European heat wave. A comparison of the raster maps produced by the European Heat Wave Pipeline with alerts issued by Meteoalarm (the forecasting and alerts system of The Network of European Meteorological Service), showed the system was able to accurately predict the length, location, and magnitude of the heatwave (Gobbi et al., 2017).

The process was designed to be relatively straightforward, easily replicable, and inexpensive. Both historical and forecast data for the study were provided as netCDF files. These are common file formats for accessing and sharing multidimensional data making them ideal for forecasts that often include temperatures, precipitation, humidity, wind speeds, etc. The data was extracted from these files and uploaded to PostgreSQL – a free, opensource database. Scripts were written in R, a free statistical software system, and run on the PostgreSQL database using a "standard lap top" (Gobbi et al., 2017, p. 2). The authors stated extraction of the forecasts from the Finnish Meteorological Institute took approximately five hours followed by one hour to calculate the heatwave forecast. The thresholds were calculated once a year and took approximately one day.

Another study (Wang et al., 2019) conducted in seven cities across China employed a random forest model to improve the prediction of heat stroke using meteorological data including the previous one to three days' maximum temperature and relative humidity as well as the average temperature and relative humidity over the past five years, structural data such as the percentage of impermeable area, and socio-economic variables including the number of houses with air conditioners and proportion of rural versus urban households. These variables were examined alongside Internet searches for the phrase "heat stroke". Meteorological and search engine variables were divided into timeframes of one to five days (i.e. relative humidity of the last 5 days) and collected during the summer months[2] for the years, 2012 to 2015. Variables were entered using a one-day lag with the goal of creating lead time of at least a day for the prediction of heat stroke. The authors found that the variables of maximum temperature, relative humidity, and Internet searches for "heat stroke" all on the previous day, contributed the most to the model.

It appears researchers are just beginning to use machine learning and big data to improve prediction of heat waves and the resulting public health effects. However, the detailed explanation of the types and sources of data in the above studies, increase the possibility of replication. In particular, the authors of the European Heat Wave Pipeline study stress that, "The whole pipeline can easily be reproduced in any spatial resolution, geographic location, or time period if both historical data and forecast temperatures are available" (Gobbi et al., 2017). Environment Canada provides daily and sometimes hourly weather data from all of its weather stations (n=8,771). Historical data are also available, in some cases as far back as 1840 (Government of Canada, 2019b). This data includes weather information such as the minimum and maximum temperature, humidity, and precipitation. Environment Canada also provides seven-day forecast data. This information is available in comma separated value and XML format for easy incorporation into statistical packages such as R as in the Gobbi et al., (2017) study.

Apart from similar weather and climate data, the Wang et al., (2019) study included population data such as GDP per capita, population density, and proportion of residents 65 and older. Comparable data is available from Statistics Canada. For a detailed list of comparable types and sources of data, see Table 2.

---

[2] Actual months not specified

Table 3. ***Variables and Sources of Data – Extreme Heat***

Data	**Source**
Census and Population-level Population Density GDP per Capita Urban vs Rural Households Proportion of Population ≥65 Internet Penetration Air Conditioners per 100 Households	Statistics Canada
Meteorological Temperature (max. min. avg.; historical, forecast) Humidity (historical, forecast) Wind Speed	Environment Canada

Finally, the inclusion of Internet keyword queries to enhance syndromic surveillance has been in use for over a decade (Olson, 2013). A recent (2013) survey by Statistics Canada found that 70 percent of people with a home Internet connection report using it for health-related queries (Canada, 2013). The current version of Google's search engine query tool is Google Trends.[3] This tool displays the popularity of a search term or phrase for a specified timeframe (e.g. 2000 to August 2019) and location (e.g. worldwide, Canada, Alberta, etc.). The tool also generates a list of related words along with their popularity within the same time and location. Google Trends also has the ability to compare the popularity of words or phrases, for example: signs of heat stroke compared to symptoms of heat stroke to understand which phrase is more common. The results can be downloaded in comma separated value format for easy upload into statistics packages such as SPSS or SAS. Graphs and maps displaying the results can also be embedded into any HTML page and kept current by linking the display to Google Trends (Google Trends, 2019).

2.2.3 Air Quality Monitoring. The direct and indirect costs of asthma are estimated to be $2.1 billion annually in Canada. These costs are associated with physician services, hospitalizations, and medication costs (Asthma Canada, 2019). Despite the high cost of asthma-related symptoms, and the correlation with pollen, a recent Canadian study (Li et al., 2019) found tracking the "onset, duration and severity of [the pollen season] is difficult, because of

[3] Previous version of Google search engine query tools used in research on syndromic surveillance were: Google Adworks Keyword and Google Insight

insufficient stations and/or monitoring networks” (p. 267). Pollen counts for the majority of the Canadian population are provided by the Aerobiology Research Laboratories which operate 30 monitoring stations in and around major population centres. Samples from the stations are collected daily and sent for analysis. Pollen counts from this organization are used by The Weather Network, to provide information and issue alerts when the pollen concentration is expected to affect public health (Martins, 2018). The obvious drawbacks of this system are, the delay between collection, analysis and alerts; and the lack of monitoring in more rural areas.

2.2.4 Air Quality - Potential of Machine Learning. Only one article was found describing the use of machine learning to predict daily pollen count (Zewdie, Lary, Liu, Wu, & Levetin, 2019). In this study, NEXRAD weather radar data were fed into both an artificial neural network (ANN) and random forest algorithms to estimate daily ragweed pollen in a 300km x 300km area around the radar (located near the University of Tulsa, Oklahoma). NEXRAD is a system of approximately 160 weather radar located in the U.S. that collect data on precipitation, cloud cover, and wind speed and direction. These variables are known to affect the release and dispersal of pollen in the atmosphere. The two machine learning methods were used in order to develop and test different models for estimating pollen levels.

Artificial neural networks were designed to mimic the processing of information in the human brain. While individual neurons in the brain can pass information to any other nearby neuron, ANNs have individual layers. Each layer of artificial neurons will assign a weight estimating how likely the input is to complete a task or answer a question (e.g. does this image contain a human face?). The weights are based on training data previously fed to the system with the correct path or answer. Using the weights, the algorithm produces a response along with a probability indicating how “sure” the algorithm is of that response.

The NEXRAD weather radar near the University of Tulsa began observations in 1987, with observations from 1995 to the present accessible to researchers. This allowed the authors to train the algorithms using data from 1995 to 2014. Both algorithms created similar models displaying the concentration and distribution of pollen within the study zone.

2.2.5 Wildfire Emissions. Forest fires are increasing in frequency and severity across Canada representing a serious concern for public health. According to Natural Resources Canada, over the last ten years, forest fires have burned an average of three million hectares a year (Natural Resources Canada, 2019). In addition to the physical threat of fire, smoke emissions can be carried downwind for thousands of miles diminishing the air quality for populations both near and remote to the fire (Government of Canada, 2013).

Emissions from forest fire smoke include a variety of air pollutants including particulate matter. Particulate matter (specifically $PM_{2.5}$) is a mixture of natural and synthetic compounds in liquid or solid form. $PM_{2.5}$, which is often found in higher concentration in forest fire emissions, is commonly examined in studies of wildfire smoke and human health. The number designation (2.5) refers specifically to the size of the particles which are light enough to stay aloft for long periods of time increasing the likelihood that they will be inhaled (Yao, Brauer, Raffuse, & Henderson, 2018). The effects of this pollutant on public health have been widely examined with studies reporting increased dispensation of asthma-related medications, and increased hospitalizations and mortality from asthma, respiratory, and cardiovascular disease (Yao, Brauer, & Henderson, 2013).

2.2.6 Wildfire and Emissions Monitoring. As stated above, the frequency and intensity of forest fires is expected to grow, increasing smoke emissions and affecting air quality for populations both near and distant from the fire. Environment Canada uses the Air Quality Health Index (AQHI) to monitor air quality and issue alerts when emissions are high (Environment and Climate Change Canada, 2007). The AQHI measures air quality on a ten-point scale (1=low risk; 10=high risk) and alerts are issued along with messages for both the general population and those at greater risk of adverse health effects from smoke emissions. These alerts reflect current, local air quality levels. In an attempt to mitigate the most serious public health impacts of wildfire smoke, the Government of Canada also employs the FireWork Wildfire Smoke Prediction System. FireWork predicts how smoke from wildfires is expected to grow or travel during a 48-hour period. The forecasts are based on information regarding current fire hotspots, fuel type and availability; as well as meteorological forecasts and data on levels of $PM_{2.5}$ and ground-level ozone (O_3) (Government of Canada, 2013).

In addition to emissions forecasting, Natural Resources Canada developed the Fire M3 system for “Monitoring, Mapping, and Modeling” fire behaviour (Natural Resources Canada, 2019). To accomplish this goal, satellite images of fire hot spots are obtained from the United States Forest Service (USFS), the National Atmospheric and Space Administration (NASA), the National Oceanic and Atmospheric Administration (NOAA), and the University of Maryland (Natural Resources Canada, 2019). The satellite images are then analyzed alongside current and forecasted weather conditions, topography at and around the fire, and fuel type and availability. The combined information is used, not only to detect actively burning fires, but also to predict their course and estimate the total area burned on a daily and yearly basis.

In addition to FireWork and Fire M3, two wildfire smoke forecasting platforms, BlueSky and FireSmoke Canada operate mainly out of the University of British Columbia. BlueSky is a smoke forecasting system administered by the University of British Columbia. The system predicts how smoke will affect communities both adjacent to and distant from the fire by analyzing the smoke plume and forecasting its path over a 48-hour period (and up to 60 hours).

The system operates by measuring the initial height of the plume then using information from Natural Resources Canada about the fire including size, location, and forest type. This information is combined with meteorological forecasts from the University of British Columbia to predict the height of the plume downwind. Downwind height is important because the higher the emissions particles are injected into the atmosphere, the farther they can be carried and the greater the opportunity to affect air quality and public health. Forecast information (including predicted levels of $PM_{2.5}$) are available to "professionals in the air quality, health & safety, emergency management, and science & research communities as well as the public" through the FireSmoke Canada website. This site uses the data provided by BlueSky and integrates it with a geographic information system to offer interactive maps of the smoke forecasts.[4]

In 2013, to examine the efficacy of BlueSky forecasts, Yao et al. (2013) compared plume forecasts from BlueSky Canada with observations made by NOAA. The study found "modest agreement between BlueSky forecasts and NOAA observations. The authors also examined the association of BlueSky forecasts with an increase of asthma-related symptoms. Data from B.C. PharmaNet and the B.C. Medical Services Plan Billings database revealed a forecasted increase in $PM_{2.5}$ correlated with an increase in physician visits for asthma and an 8 to 12 percent increase in the dispensation of asthma-related medications.

FireSmoke is a portal where users can access and generate emissions forecasts. FireSmoke uses BlueSky Canada to produce hourly forecasts of surface-level $PM_{2.5}$ accumulations for the following 48 hours. In addition to forecasts, the portal offers users the ability to create emissions scenarios for controlled burns and wildfires.

[4] FireSmoke website: http://firesmoke.ca/forecasts/

2.2.7 Wildfires - Potential of Machine Learning. Given Canada's landmass, satellite monitoring represents a relatively inexpensive method of surveillance, particularly in remote or sparsely populated areas. Remote sensors can provide daily (sometimes twice-daily) detection and monitoring of wildfires. Satellites also have the ability to identify smoke plumes, which can indicate the location of a fire.

However, thick smoke or even cloud cover can obscure a fire causing them to go undetected in remote regions (Yao et al., 2013). Even twice daily monitoring means a fire may have been burning for several hours before being captured in a satellite image and analyzed by an algorithm. This information is limited further by satellites like the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO), which follow a narrow grid resulting in a 16-day cycle (Yao, Raffuse, et al., 2018). Finally, the actual size of the fires cannot be determined since satellites like CALIPSO capture images in squared kilometers (Yao, Raffuse, et al., 2018). As Yao et al., (2018) point out, most remote sensing and analysis platforms aboard satellites are also only capable of assessing the entire column of smoke. Whereas, a ground-level analysis of pollutants would provide more accurate information about respiratory risks. Additionally, many forecasting platforms use deterministic models to forecast the distribution of smoke from a fire. While these equations are being used effectively on platforms such as BlueSky, they require considerable expertise to operate.

Machine learning offers advantages over traditional deterministic models (like the one used by BlueSky) that can address some of these problems. Machine learning can automate manual input of variables by human experts and increase the amount of data included in a model by collecting up-to-date information from multiple sources and mining it for applicable data. These algorithms also have the ability to accommodate non-linear relationships between input and output variables, and factor in instability or unpredictability of the input data unlike deterministic models (Yao, Brauer, et al., 2018). This is important, for example because smoke plumes are dynamic.

Yao, et al. (2018) demonstrated these benefits by using a machine learning approach "to predict the minimum height of smoke in the atmosphere using variables that reflect fire activity, location, and meteorology" (p. 103). The study examined data from the province of B.C., which has experienced several record-breaking fire seasons over the last decade. Data were collected from April through September (i.e. wildfire season) for the years 2006 through 2015. The study aimed to improve the prediction of ground-level concentrations of smoke emissions to facilitate assessment of public health risks.

The machine learning model generated by the authors was able to improve predictions of ground-level emissions by incorporating near real-time, empirical variables about fire activity, geography, and meteorology. Meteorological data included wind variables and incorporated information about the month and time of day. Fire activity consisted of data on the intensity and location of a fire. While geographic variables included information on terrain, land use, and elevation.

In order to train and test the accuracy of the model, the study covered a period of 10 years from 2006 to 2015. This means retrospective data were used for the meteorological and fire-related variables. However, as the authors point out, weather forecasts are widely available and could be incorporated, and the retrospective fire data could be replaced with NASA's Fire Information for Resource Management System (FIRMS), which provides current measures of fire intensity.

When tested, the model explained 82 percent of ground-level emission observations by CALIPSO. The model was particularly adept at identifying smoke above a threshold not relevant to public health. The authors suggest the model could be operationalized, using near real-time data, to increase the accuracy and relevance of emissions predictions.

Only three studies examining the public health effects of wildfire smoke were found outside Canada. Mazzoni et al., (2007), aimed to estimate the vertical height of smoke plumes, while Zou et al., (2019) combined satellite data with a vulnerability model. Taking a more specific look at the link to public health, Reid et al., (2019) compared surface area concentrations of $PM_{2.5}$ and O_3 with daily counts of hospitalizations and visits to emergency departments for respiratory illnesses. The Mazzoni et al., (2007) and Zou et al., (2019) studies used MODIS (MODerate-resolution Imaging Sectroradiometer)[5] data from the Terra and Aqua satellites, focused on $PM_{2.5}$ pollution, and specified that vertical plume height was the critical variable for estimating the public health impacts of smoke emissions. Zou et al., (2019) used a random forest model like Yao et al. (2018), while Mazzoni et al., (2007) used support sector machine, a machine learning

[5] The MODIS remote sensing platform with a 1km resolution classifies each image taken as fire, cloud, water, or land. For more information see the Mazzonie et al., 2007 reference.

approach used to categorize or classify data. In this case, machine learning was used to classify pixels in satellite images representing one square kilometer as containing (or not) smoke plume. A few other differences included the concentrations of O_3 in the Reid et al., (2019) study, data on the vertical distribution of aerosols from remote sensors aboard the International Space Station in the Zou et al., (2019) study, and the use of multiple ground-based sensors in the Reid et al. (2019) study. Overall, there were more similarities related to types and sources of data, and methods of analysis, lending further credibility to the Yao et la. (2018) study.

2.2.8 Wildfires – Machine Learning Infrastructure. Regarding the infrastructure necessary for monitoring and forecasting wildfire emissions, Yao et al. (2018), presented a template. The authors stated, “we were interested in developing an empirical model that could be operationalized in near real-time with readily available data” (p. 101). Therefore, data for each of the relevant variables was documented and where retrospective data were used, the authors proposed a near real-time alternative. Table 3 lists the category of variables (e.g. meteorological; fire activity) and their sources, as well as near real-time alternatives for those sources.

The authors used a random forests algorithm, which is one of the most commonly employed machine learning algorithms. Random forest is an ensemble of regression trees capable of accommodating non-linear relationships between input and output variables while allowing for complex interrelationships among the input data[6]. In other words, random forests can accept the volume and variety of big data and run an analysis even when the relationship between input and output variables is not well understood (Kühnlein, Appelhans, Thies, & Nauss, 2014).

A detailed explanation of the algorithm is documented in the article including how each variable was defined and assessed for value. Further details about the algorithm are also available in companion articles the authors wrote while conducting related research (Z Liu et al., 2005; Zhaoyan Liu et al., 2009; Omar et al., 2009; Vaughan et al., 2009; Vaughan, Winker, & Powell, 2005). This process ensures the most accurate model with the least number of variables.

[6] “Data cleaning and analysis was conducted in R Statistical Computing Environment (R Core Team, Vienna, Austria)” The data were fitted with the random forests algorithm, an ensemble of regression trees, each of which was constructed with a random subset of observations and a random subset of predictive variables. This machine learning approach can provide accurate prediction while being robust against overfitting, accommodating non-linear relationships between the dependent and independent variables, and accounting for complex interactions between the independent variables.” (Breiman, 2001)

As noted in the Human Resource Requirements section below, being able to monitor and forecast the public health effects of climate change requires specialized knowledge in at least two of the three topic areas. However, the Yao et al. (2018) detailed documentation of the process, makes it more likely that someone with the specific knowledge and skills could replicate the method and expand it beyond British Columbia.

Table 4. ***Variables and Sources of Data – Wildfire Emissions***

Data	**Source**	**Near Real-Time Alternative**
Satellite Imagery: Vertical distribution of smoke in the atmosphere	CALIPSO Satellite - NASA	
Satellite Imagery: Fire Activity (hot spot data and intensity)	MODIS - NASA	FIRMS -NASA
Meteorological Temperature Wind speed and dire	MODIS - NASA MERRA - NASA	Environment Canada
Geographic Location:	U.S. Geological Survey Earth Resources Observation and Science (EROS) Center $GTOPO_{30}$	

2.3 Severe Storms.

ECCC identifies severe storms, including hurricanes, floods, and wildfires as an ever-increasing threat of climate change.

2.3.1 Floods – potential of Machine Learning. Floods are reliant on precipitation which is notoriously difficult to forecast (Cao, 2018; Holmstrom, Liu, & Vo, 2016). Even small changes in any one of the numerous phases of the hydrological cycle can dramatically alter a forecast (Slingo, 2011). Furthermore, variables such as soil temperature, snow melt, runoff, and the amount of impermeable land are also responsible for flooding. According to a literature review documenting the use of machine learning in flood prediction, the practice only began gaining popularity in the last decade (since 2008) (Mosavi, Ozturk, & Chau, 2018). This review found ANNs were the most frequently used method with researchers citing their ability to construct fairly accurate models despite the non-linear relationship between rainfall and flooding. The authors also found researchers were beginning to experiment with hybrid machine learning methods, often by blending two methods. This was done to further improve accuracy, generalizability, prediction lead-time, and decrease costs. Ultimately the authors concluded, the use of machine learning in “flood prediction is quite young and in the early stage of advancement” (Mosavi et al., 2018, p. 27).

One study, conducted in Canada, examined the ability of a machine learning model to forecast flooding from the Englishman river on Vancouver Island, B.C. The authors used an ensemble of ANNs because this method was viewed as inexpensive, relatively easy to development and use, robust to variations in the hydrological cycle and accurate. Although the system under-forecast peak flows, overall it provided a good daily assessment of the threat of flooding from the river. One caveat the authors add, “there is no single “best” [hydrologic] model” (Fleming, Bourdin, Campbell, Stull, & Gardner, 2015, p. 504). This is due to the variability in the hydrologic cycle which is affected by things like topography and bodies of water. In this case, the “Pacific Ocean Void” was a confounding factor - so called due to a lack of satellites and ocean-level weather monitors (e.g. weather buoys).

2.3.2 Hurricanes. Hurricanes and hurricane force tropical storms only make landfall in Canada about once every three years (Environment and Climate Change Canada, 2012). Further, the cool waters off Canadian coasts drain the energy from these storms, so, “by the time that these hurricanes do impact Canada it’s mostly a huge rainstorm and not much of a windstorm”, according to meteorologist at the University of Toronto, Athena Masson (Rocha, 2017). However, as the climate warms these storms are expected to intensify and become more frequent. Therefore, a brief discussion will be presented here on the use of machine learning to predict their track and intensity.

Hurricanes – Potential of Machine Learning. Like floods, the use of machine learning in this field is new. The database search initially produced 33 articles. Of the 33, 14 were relevant and met the criteria for review. Eight of these were conference papers. In addition, the oldest article among the results was published in 2013. Most of the articles focused on how machine learning methods could be used to forecast the track or intensity of the hurricane and resultant precipitation. Possibly because the technology is so new in this area of research, multiple studies compared the predictive power (either to forecast intensity, wind speed, or precipitation) of machine learning methods with traditional numerical models.

One study found using publicly accessible data, employed support vector regression (SVR) to predict the intensity of a hurricane (defined by maximum sustained wind speed) based on infrared satellite images of the hurricane (Asif, Dawood, Jan, Khurshid, & DeMaria, 2018). The infrared images are able to capture the size and structure of the core of the hurricane, which is a reliable indicator of intensity, due to the color difference between the cool (blue) core and warmer (red) outer bands. The model was trained and tested using images of hurricanes from multiple satellites for the years 1978 to 2009. The images were stored in HURSAT-B1 – a publicly accessible database. The results proved the SVR model could more accurately predict the maximum wind speed than physically based models.

2.4 Food Security

"Canada does not worry about its food security" according to The Canada Country Study: Climate Impacts and Adaptation (Government of Canada, 1998). Indeed, as of 2017 Canada was exporting $56 billion in agricultural products each year (Canadian Agri-Food Trade Alliance, 2017). While food security might not be a problem nationally, like all the consequences of climate change in Canada, those on food security are regional and some places are already experiencing an impact.

Much of the concern around food security was focused on Canada's Indigenous population, particularly in the Arctic region. Here communities are already reporting an earlier breakup of sea ice, less permafrost, and lower river levels with less flow leading to stagnating water. Decreased permafrost is a problem because many communities have traditionally stored game in sections of permafrost. Therefore, the continuing melts will necessitate adaptation in the form of moving to places with access go refrigeration, or changing diets. Early sea ice breakup is a problem because it limits remote communities' access to fishing and hunting grounds as well as travel between communities. It can also impede access to foods imported by trucks on seasonal roads. This is a serious problem when a survey of Northern Aboriginal communities found, 96 percent of adults hunt, fish, or gather natural resources as a means of subsistence (Séguin et al., 2008). There have also been reports of people becoming stranded, injured, or drowning when caught off guard by earlier breakup of sea ice (Furgal & Seguin, 2006).

Aboriginal groups outside the arctic are also expected to be disproportionately affected by climate change in relation to food security. For example, groups on the North shore of Vancouver Island rely on a diet high in salmon, and shell fish such as clams (Talloni-Álvarez, 2019). At varying stages of their lifecycle, these marine species need access to the shoreline, which is being disrupted by rising sea levels cause by climate change. Also, as noted in Section

2.1.3 warming water temperature are also contributing to more frequent toxic algae blooms which get into the food supply when they are absorbed by shellfish (Finnis et al., 2017).

2.4.1 Food Security – Potential of Machine Learning. While AI and machine learning cannot solve all of these problems, a few articles were found using the technology to forecast the breakup of river ice in order to anticipate floods which often accompany this event. Although the focus was river ice, these studies hint at the suggestion that it may be possible to forecast the breakup of sea ice. (If we set aside for a moment the regional variations in the hydrological cycle.)

Two of these studies took place in Canada (A. W. Beaton, R. Corston, K. Kenny, F., 2019; Wei, 2018). One study used a stacking ensemble learning framework to identify the variables with the highest predictive power of ice breakup dates on the Athabasca River at Fort McMurray (Wei, 2018). The authors developed models of river ice breakup dates, specifically, Bayesian[7] Regulated Back-Propagation Artificial Neural Network (BRANN) and Adaptive Neuro Fuzzy Inference System. (However, the BRANN model out-performed the ANFIS, and so only the former will be discussed here.) BRANN describes the process of a system making a prediction, receiving feedback from incoming data, and then adjusting the conditional probabilities of future predictions. Data, consisting of weather-related or river ice variables. was collected from Alberta Environment and Parks, which maintains a multiyear database of observations of the Athabasca River around Fort McMurray. The authors used 36 years of observations. Weather-related variables included minimum, maximum, and average daily temperatures, precipitation, and snowpack; river ice variables included water level and flow, and ice thickness. Results of the study determined air temperature and river flow are the highest predictors of ice breakup.

Another study (A. Beaton, Whaley, Corston, & Kenny, 2019), used satellite imagery from the MODIS remote sensing platform and Google Earth Engine (GEE) to create a dataset of ice breakup dates including beginning, end, and duration for five rivers draining into Hudson or James Bay. This region is isolated from larger population centres, therefore when flooding occurs the First Nations communities who live in the area must be flown to safety. The authors used nearly 20 years of MODIS imagery showing river conditions and historical breakup dates.

---

[7] Bayesian Theorem is the process of adjusting conditional probabilities based on incoming information

GEE is a free, publicly accessible, cloud-based platform that has proven effective for analyzing remotely sensed data for a variety of applications (e.g. forestry, and agriculture). GEE was chosen because it does not require data to be downloaded into the system and allows users to quickly adjust algorithms to improve analysis or prediction (A. Beaton et al., 2019). Once the dataset was created, it was used to calculate probabilities of the beginning, end, and duration of ice breakup. This method was able to forecast breakup dates within -2.0 to 6.7 days of empirical data captured by Water Survey of Canada. The authors stress that the dataset can be used along with hydrological knowledge, and satellite data for prediction of breakup dates along these rivers.

3 Infrastructure and HR Requirements

3.1 Data and Computing Infrastructure Requirements

This environmental scan found the use of machine learning in research on the public health effects of climate change is relatively new. Additionally, because machine learning algorithms need ample data to solve problems or answer questions, its use is not widespread across all risks. This is likely why the use of machine learning is more advance in wildfire emissions forecasting than flood risk. As can be seen in Sections 2.3.3 to 2.3.5 emissions' research benefits from a wealth of data including satellite images of smoke plumes and fire hotspots, ground-level air quality monitors, and topographical data such as available fuel and water sources. By contrast, as explained in Section 2.3.1, multiple, difficult to monitor variables such as soil temperature and precipitation are responsible for making an area vulnerable to flooding resulting in a lack of data for use with machine learning algorithms.

However, some commonalities exist. Research in any of the risk areas (see Table 1) requires climate and/or weather data. Weather here, refers to current, recent past (i.e. less than 30 years)[8], and forecast data. While research related to the risks of extreme heat requires understanding heat thresholds, which necessitates decades of minimum and maximum daily temperatures or climate data. Precipitation forecasts were also relevant for research in all four risk categories. The link between precipitation and severe storms (such as floods and hurricanes), or even smoke emissions seems obvious, but intense precipitation can spread bacteria through

[8] Based on the definition of climate provided by NASA in, "NASA – What's the Difference Between Weather and Climate?" (National Aeronautics and Space Administration, 2015)

runoff into recreation or drinking water (vector-borne illness). Humidity and wind were also common independent variables across most risk-related areas of research.

Environment Canada provides seven-day forecasts, in addition to daily, and sometimes hourly weather data from all of its 8,771 weather stations. This information is publicly available in comma separated value and XML formats allowing it to be uploaded into statistical packages such as R and SPSS. Further, some meteorological data, including forecasts and warnings are available as RSS feeds allowing computer systems to automatically read and incorporate the information into algorithms (Toutant et al., 2011). Topographic data was also necessary for many of the studies. This information included variables such as water sources, vegetation, and proportion of non-permeable land. In addition to weather data, were population-level variables such as population density and mean age; and socio-economic factors such as rural versus urban population, literacy rate, percentage of homes with air conditioning, and household income. For a list of variables, frequently used in the studies reference here, see Table 4.

Table 5. ***Common Sources and Types of Data - Public Health Risks of Climate Change***

Data	Source	Study
Topography		
Vegetation Index	NASA – MODIS remote sending data	Infectious Disease, Severe Storms, Higher Temperatures, Air Quality
Elevation	US Geological Survey; Global Land Cover Facility SRTM	Severe Storms
Land Cover	National Land Cover Database - USGS	Infectious Disease, Severe Storms, Higher Temperatures, Air Quality
Soil Temperature		
Weather		
Temperature (min. max. avg.)	Environment Canada	Infectious Disease, Severe Storms, Higher Temperatures, Air Quality
Precipitation	Environment Canada	Infectious Disease, Severe Storms, Higher Temperatures, Air Quality
Humidity	Environment Canada	Infectious Disease, Severe Storms, Higher Temperatures, Air Quality
Wind Speed	Environment Canada	Severe Storms, Higher Temperatures, Air Quality
Census		
Population Density	Statistics Canada	Infectious Disease, Severe Storms, Higher Temperatures, Air Quality
GDP per Capita	Statistics Canada	High Temperatures
Urban vs Rural Households	Statistics Canada	Higher Temperatures, Air Quality
Proportion of Population ≥65	Statistics Canada	Higher Temperatures, Air Quality

Regarding health-related statistics (e.g. hospital admissions, medication dispensations, mortality, etc.), these variables were primarily used to establish correlations with climate change-related weather events. For example, emergency department visits, hospital admissions, and non-traumatic mortality were examined in relation to heat waves across Canada in order to establish varying heat wave thresholds (Gachon et al., 2016). Once the link had been established, and/or the threshold determined, the use of real-time health statistics is not necessary for forecasting climate change-related weather events or communicating with the public to mitigate the effects.

Furthermore, appropriate computing infrastructure is required to: 1) collect required data from various sources as automatically as possible; 2) link and pre-process all data; 3) train new or update existing machine learning models; 4) run the machine learning models on incoming data; and 5) display the resulting surveillance and/or predictive information, typically on a visual dashboard. The exact specifications of the computing infrastructure regarding computing power and data storage depend on: 1) the volume of the data; 2) the rate at which the data grow; 3) the frequency at which the machine learning models are required to generate output; and 4) the complexity of the machine learning model architecture. Multiple central processing units (CPUs) and graphical processing units (GPUs) are required to enable parallel computing and accelerate model training.

3.2 Human Resource Requirements

Understanding the public health impacts of climate change and being able to forecast adverse events using AI and machine learning requires three areas of specialized knowledge: data science, public health, and climate change. In most cases, employment in any one of these areas requires an advanced degree, but no degree program lies at the intersection of all three specializations. However, many universities offer programs in at least two of these disciplines.

As the applications for AI and machine learning expand across many disciplines, demand is outpacing supply. Advances in AI are growing rapidly, with some researchers predicting more developments in the next five years than over the last three decades (Teja, 2019). This demand means there is an increased need to train data scientists who can handle big data, as well as machine learning engineers capable of understanding and developing machine learning models. As a result, Canadian universities have begun investing more heavily in computer science programs. The University of Toronto currently ranks in the top 12 universities world-wide for computer science and information systems according to QS World and the Shanghai Global Ranking system. Two more, the universities of Waterloo and British Columbia make the top 50 (QS Top Universities, 2018; Shanghai Ranking, 2019). Table 5 lists the top ten universities for computer science in Canada as ranked by Maclean's. Each school offers master's- and doctorate-level degrees in computer science. These programs offer a core set of courses in AI and machine learning that set students up for careers in a wide array of industries including technology, healthcare, banking, and engineering. Many of these university programs offer the ability to specialize or minor in a related sciences field, in addition to co-op or internship options. These opportunities will broaden the skill set of graduates and make them more competitive as companies seek hands-on experience as well as a degree (Alini, 2018).

Provincial and federal governments are also investing in the fields of computer science and artificial intelligence. For example, the Alberta government recently announced the addition of 406 technology-related seats, including computer science programs, to post-secondary institutions across the province (Rieger, 2018).

The federal government is investing in computer science and AI with several high-profile commitments in recent years. In 2017, the federal government committed $125 million to the Pan-Canadian AI Strategy. Led by the Canadian Institute for Advanced Research, the project aims to increase the number of graduates with expertise across the spectrum of AI; to become thought leaders around the "economic, ethical, policy and legal implications of advances in artificial intelligence"; to maintain a nation-wide collaboration on AI; and to foster stronger connections between Canada's AI research hubs: Toronto, Montreal, and Edmonton (CIFAR, 2019). Finally, the federal government committed nearly $2 billion, in partnership with private industry, to operate "innovation supercluster". These superclusters are intended to foster industry-led research, learning, and development in AI.

While Canadian universities have increased investment in computer science programs, the Canadian Occupational Projection System (COPS) still predicts a deficit in technology professions such as computer programmers and data scientists through the year 2026 (Government of Canada, 2019a). Further, being able to model the public health effects of climate change requires expertise in at least one of the other fields. Researchers must know what algorithms would best answer the research questions; what data is necessary, where and how to access it; and how to process the information. As a brief example, in order to map the spread of smoke emissions, among the information Yao et al., (2018) had to know was how time of day and season affect temperature, because these variables affect the rise and fall of smoke in the atmosphere. Understanding the importance of these variables allowed them to be included in the algorithm resulting in a more accurate model of smoke dispersion.

Currently no university offers a program at the intersection of all three disciplines (i.e. AI or machine learning, public health, and climate change). Although all of the top ten Canadian computer science universities offer science-related minors, co-op, and internship opportunities, it will be nearly impossible to know how many graduates use these options to obtain skills in climate change, or public health research as most universities only track the number of graduates by faculty.

Government organizations will also have to compete with private industry for computer science graduates who have an average a salary between $55,000 and $70,000, and up to

$100,000 in their first year (Alini, 2018). Therefore, although interest and investment in computer science and AI are increasing, challenges will remain in harnessing these skills for use in climate change and public health research.

A list of the top ten U.S. universities for computer science programs is included in a separate table (Table 3) along with information about public health programs and any climate change activity by the institution. However, a recent (Spicer, Olmstead, & Goodman, 2018) report found one-in-four science, technology, engineering, and math graduates "from the Universities of Toronto, British Columbia and Waterloo…opted to work outside Canada" after graduation (p. 6). The inference of this report being, it is more likely Canadian STEM graduates are leaving to work in the U.S. than American STEM graduates are coming to work in Canada[9].

[9] A brief search was done, but no indication was found that Canadian technology companies are attracting U.S. graduates at a rate that would make attempts at recruitment a viable option at this time.

Table 6. Canadian Universities with Top 10 Computer Science Programs

University	Computer Science Program: AI and Machine Learning courses	Public Health MSc. PhD.	Climate Change
Toronto	Yes	Yes	Environment and Health Program – part of the School of Public Health
British Columbia	Yes	Yes	Institute for Resources, Environment and Sustainability; Member of PICS – MSc., PhD., and MA programs offered
Waterloo	Yes	Yes	Interdisciplinary Centre on Climate Change (master's program)
Montreal	Yes	Yes	Unsure
McGill	Yes	Yes	School of Environment (undergraduate degrees)
Alberta	Yes	Yes	Transdisciplinary Research network on Climate Change, Water Governance, and the Futures of Communities
Simon Fraser	Yes	Yes	Member of PICS
McMaster	Yes	Yes	Centre for Climate Change (not a degree program)
Ottawa	Yes	Yes	Office of Sustainability (not a degree program)
Carleton	Yes	Yes	Carleton Climate Commons Working Group (not a degree program)

Table 7. ***U.S. Universities with Top 10 Computer Science Programs***

University	Computer Science Program: AI and Machine Learning courses	Public Health MSc. PhD.	Climate Change
New York	Yes	Yes	Department of Environmental Studies – MSc., PhD. programs
Columbia	Yes[10]	Yes	Climate and health Program – Part of the Mailman School of Public Health
Washington	Yes	Yes	Center for Health and Global Environment (CHANGE) - One Undergraduate and one graduate level course on Climate change and health
California LA	Yes	Yes	Institute of Environmental Sciences – BS, PhD.
Princeton	Yes	Yes	Could not find a degree program
Harvard	Yes	Yes	Could not find a degree program
Berkeley	Yes	Yes	Energy, Climate, and Environment – Only a research group
Carnegie Mellon	Yes	Yes	Center for Engineering and Resilience for Climate Adaptation
Stanford University	Yes	Yes	Stanford Earth – supports MSc. And PhD research
Massachusetts Institute of Technology	Yes	Yes	Center for Global Change Science – Not a degree program

[10] Could not find a course on AI

4 Recommendations

In light of the very serious, and current public health consequences of climate change; in addition to the shortage of data scientists and experts in machine learning, we make the following recommendations.

4.1 Design, Development, and Implementation of the Pan-Canadian Surveillance System

This report has outlined the major public health risks related to climate change, what kinds of data are currently collected for surveillance purposes, and how machine learning and AI have been utilized in and outside Canada. The first step is to prioritize the public health risks, which can be facilitated by data availability and accessibility, since a prerequisite of any machine learning based system is an appropriate data set for training and a data pipeline for deployment in practice. Some public health risks, such as wild fires, benefit from rich satellite imaging data. Many weather and climate variables (e.g., temperature, humidity, precipitation) are also readily available. On the contrary, Lyme disease surveillance currently relies on manual reporting where electronic data capture is limited and substantial time lags exist.

Data linkages between climate and/or weather and health data should also be considered. Once required data are identified, discussions with the data custodians (e.g., Environment Canada, Public Health Agency of Canada) regarding data access should follow if the data are not public and Health Canada does not currently have access. If required data are not currently collected, investment in data collection infrastructure (e.g., ground-level sensors, computerized reporting mechanism) would be inevitable.

With respect to machine learning, given the most successful machine learning applications to date across diverse disciplines, the most promising areas would apply deep learning, particularly convolutional neural networks, to imaging data. Therefore, any application that can benefit from satellite images (e.g., wild fires) should be given priority. Machine learning also holds promise for analyzing and extracting information from complex, high-dimensional, time-varying data.

Predictive or classification models via supervised learning seem most relevant for surveillance purposes. Unsupervised learning can also be useful in understanding and characterizing complex data sets, which can guide supervised learning. Given the passive nature of surveillance, it is anticipated that reinforcement learning would be least utilized, as it is typically used to identify optimal sequences of actions that lead to the maximum reward, requiring a large number of interactions between the agent (who is taking actions) and environment in the training data.

4.2 Expanding Human Resource Capacity

The Government of Canada and provincial governments should fund university programs at the intersection of all three disciplines of data science, public health, and climate change. As displayed in Table 5, a few Canadian universities have begun establishing public health degree programs with specializations in climate change or environmental studies. Additionally, all of the top 10 universities offer courses in AI and machine learning. Therefore, the federal and provincial governments should offer funding to universities that modify their public health programs to include courses on AI, machine learning, and big data. For example, the University of Toronto has a public health program with a specialization of "Environment and Health" in addition to courses (unrelated to the program) on data science, AI, and machine learning. Funding is necessary to establish collaboration between departments including interdepartmental leaders who can establish appropriate degree criteria. To be effective, the program would need several courses on AI and machine learning. Another approach would be to train data scientists in public health and/or climate change by adding additional courses or creating a specialization track in existing data science, computer science, or AI programs.

A second option (either in lieu of or in addition to the one above), is for the governments to fund continuing education courses in data science, AI and machine learning with a preference for online learning. This option would allow persons with degrees in public health and/or environmental studies to augment their expertise. Massive open online course providers (MOOCs) such as Coursera and edX also provide courses in data science, machine learning, and AI for free.

Prior to pursuing any of these recommendations, a market analysis should be conducted to assess interest by the universities, availability of qualified instructors, and interest level of prospective students.

References

Abedini, M., Chen, Q., Codella, N. C., Garnavi, R., Sun, X., Celebi, M. E., . . . Marques, J. S. (2015). Accurate and scalable system for automatic detection of malignant melanoma. *Dermoscopy Image Analysis, 293343*.

Alini, E. (2018). Hot Jobs: The $100K entry-level job you can get here in Canada.

Alpaydin, E. (2014). *Introduction to machine learning*: Cambridge, Massachusetts : MIT Press; Third edition.

Anuradha, J. (2015). A brief introduction on Big Data 5Vs characteristics and Hadoop technology. *Procedia computer science*, *48*, 319-324.

Asif, A., Dawood, M., Jan, B., Khurshid, J., & DeMaria, M. (2018). PHURIE: hurricane intensity estimation from infrared satellite imagery using machine learning. *Neural Computing and Applications*, 1-14.

Asthma Canada. (2019). Asthma Facts and Statistics. Retrieved from https://asthma.ca/wp-content/uploads/2019/02/Asthma-101.pdf

Barton, A. J. (2019). Big Data. *Journal of Nursing Education, 55*(3), 123-124. doi:10.3928/01484834-20160216-01

Beaton, A., Whaley, R., Corston, K., & Kenny, F. (2019). Identifying historic river ice breakup timing using MODIS and Google Earth Engine in support of operational flood monitoring in Northern Ontario. *Remote sensing of environment, 224*, 352-364.

Beaton, A. W., R. Corston, K. Kenny, F. (2019). Identifying historic river ice breakup timing using MODIS and Google Earth Engine in support of operational flood monitoring in Northern Ontario.

Canada, S. (2013). Individual Internet use and e-commerce, 2012.

Canadian Agency for Drugs and Technologies in Health. (2019). Grey Matters: a practical tool for searching health-related grey literature | CADTH.ca. Retrieved from https://www.cadth.ca/resources/finding-evidence/grey-matters

Canadian Agri-Food Trade Alliance. (2017). Agri-Food Exports. Retrieved from http://cafta.org/agri-food-exports/

Cao, S. (Producer). (2018, 2018-03-07). Here's Why Predicting the Weather Is Still So Difficult. *Technology*. Retrieved from https://observer.com/2018/03/weather-forecast-predictions-still-difficult/

CIFAR. (2019). Pan-Canadian Artificial Intelligence Strategy. Retrieved from https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy

Clow, K. M., Ogden, N. H., Lindsay, L. R., Michel, P., Pearl, D. L., & Jardine, C. M. (2017). The influence of abiotic and biotic factors on the invasion of Ixodes scapularis in Ontario, Canada. *Ticks and tick-borne diseases, 8*(4), 554-563.

Darcy, A. M., Louie, A. K., & Roberts, L. W. (2016). Machine Learning and the Profession of Medicine. *Jama, 315*(6), 551-552. doi:10.1001/jama.2015.18421

Encyclopaedia Britannica. (2019). Canada | History, Geography, & Culture.

Environment and Climate Change Canada. (2007). About the Air Quality Health Index - Canada.ca.

Environment and Climate Change Canada. (2010). Criteria for public weather alerts.

Environment and Climate Change Canada. (2012). Canadian Hurricane Centre: frequently asked questions.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115-118. doi:10.1038/nature21056

Finnis, S., Krstic, N., McIntyre, L., Nelson, T. A., & Henderson, S. B. (2017). Spatiotemporal patterns of paralytic shellfish toxins and their relationships with environmental variables in British Columbia, Canada from 2002 to 2012. *Environmental research, 156*, 190-200.

Fleming, S. W., Bourdin, D. R., Campbell, D., Stull, R. B., & Gardner, T. (2015). Development and Operational Testing of a Super-Ensemble Artificial Intelligence Flood-Forecast Model for a Pacific Northwest River. *JAWRA Journal of the American Water Resources Association, 51*(2), 502-512.

Furgal, C., & Seguin, J. (2006). Climate Change, Health, and Vulnerability in Canadian Northern Aboriginal Communities *Environ Health Perspect* (Vol. 114, pp. 1964-1970).

Gachon, P., Bussières, L., Gosselin, P., Raphoz, M., Bustinza, R., Martin, P., . . . Jeffers, S. (2016). Guide to identifying alert thresholds for heat waves in Canada based on evidence.

Gardner, A. M., Anderson, T. K., Hamer, G. L., Johnson, D. E., Varela, K. E., Walker, E. D., & Ruiz, M. O. (2013). Terrestrial vegetation and aquatic chemistry influence larval mosquito abundance in catch basins, Chicago, USA. *Parasites & vectors, 6*(1), 9.

Gasmi, S., Bouchard, C., Ogden, N. H., Adam-Poupart, A., Pelcat, Y., Rees, E. E., . . . Koffi, J. K. (2018). Evidence for increasing densities and geographic ranges of tick species of public health significance other than Ixodes scapularis in Québec, Canada. *PloS one, 13*(8), e0201924.

Gerardi, D. A., & Kellerman, R. A. (2014). Climate change and respiratory health. *Journal of occupational and environmental medicine, 56*, S49-S54.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature, 457*(7232), 1012-1014. doi:10.1038/nature07634

Gobbi, A., Alikadic, A., Ylinen, K., Angaramo, F., & Furlanello, C. (2017). *A heat wave forecast system for Europe.* Paper presented at the 2017 IEEE International Conference on Big Data (Big Data).

Google Trends. (2019). Google Trends.

Government of Canada. (1998). *The Canada country study: Climate impacts and adaptation*. Retrieved from http://publications.gc.ca/collections/Collection/En56-119-7-1998E.pdf.

Government of Canada. (2013). FireWork: National Wildfire Smoke Model - Environment Canada.

Government of Canada. (2019a). Canadian Occupational Projection System (COPS). Retrieved from http://occupations.esdc.gc.ca/sppc-cops/l.3bd.2t.1ilshtml@-eng.jsp?lid=16&fid=1⟨=en

Government of Canada. (2019b). Stations Results - Historical Data. Retrieved from http://climate.weather.gc.ca/historical_data/search_historic_data_stations_e.html?searchType=stnProv&timeframe=1&lstProvince=&optLimit=yearRange&StartYear=1840&EndYear=2019&Year=2019&Month=7&Day=30&selRowPerPage=25

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., . . . Webster, D. R. (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *Jama, 316*(22), 2402-2410. doi:10.1001/jama.2016.17216

Health Canada. (2011). Communicating the Health Risks of Extreme Heat Events.

Héguy, L., Garneau, M., Goldberg, M. S., Raphoz, M., Guay, F., & Valois, M.-F. (2008). Associations between grass and weed pollen and emergency department visits for asthma among children in Montreal. *Environmental Research, 106*(2), 203-211.

Hinton, G. (2018). Deep Learning-A Technology With the Potential to Transform Health Care. *Jama, 320*(11), 1101-1102. doi:10.1001/jama.2018.11100

Holmstrom, M., Liu, D., & Vo, C. (2016). Machine learning applied to weather forecasting: Stanford University.

Janowczyk, A., & Madabhushi, A. (2016). Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *J Pathol Inform, 7*, 29. doi:10.4103/2153-3539.186902

Johnson, L. (2018). Heat health risk prompts change to Environment Canada warning system | CBC News.

Keyel, A. C., Timm, O. E., Backenson, P. B., Prussing, C., Quinones, S., McDonough, K. A., . . . Andreadis, T. G. (2019). Seasonal temperatures and hydrological conditions improve the prediction of West Nile virus infection rates in Culex mosquitoes and human case counts in New York and Connecticut. *PloS one, 14*(6), e0217854.

Kühnlein, M., Appelhans, T., Thies, B., & Nauss, T. (2014). Improving the accuracy of rainfall rates from optical satellite sensors with machine learning—A random forests-based approach applied to MSG SEVIRI. *Remote Sensing of Environment, 141*, 129-143.

Lavigne, M., Mussa, F., Creatore, M. I., Hoffman, S. J., & Buckeridge, D. L. (2019). *A population health perspective on artificial intelligence.* Paper presented at the Healthcare management forum.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). Big data. The parable of Google Flu: traps in big data analysis. *Science, 343*(6176), 1203-1205. doi:10.1126/science.1248506

Lecun, Y. B., Yoshua. Hinton, Geoffrey. (2015). Deep learning. *Nature, 521*(7553), 436-444. doi:doi:10.1038/nature14539

Lemmen, D. S., Warren, F. J., Lacroix, J., & Bush, E. (2008). From impacts to adaptation: Canada in a changing climate. *Government of Canada, Ottawa*.

Li, X., Zhou, Y., Meng, L., Asrar, G., Sapkota, A., & Coates, F. (2019). Characterizing the relationship between satellite phenology and pollen season: A case study of birch. *Remote sensing of environment, 222*, 267-274.

Liu, J., Weitzman, E. R., & Chunara, R. (2017). Assessing Behavioral Stages From Social Media Data. *CSCW Conf Comput Support Coop Work, 2017*, 1320-1333. doi:10.1145/2998181.2998336

Liu, Z., Omar, A., Hu, Y., Vaughan, M., Winker, D., Poole, L., & Kovacs, T. (2005). CALIOP algorithm theoretical basis document. Part 3: Scene classification algorithms. *NASA-CNES document PC-SCI-203*.

Liu, Z., Vaughan, M., Winker, D., Kittaka, C., Getzewich, B., Kuehn, R., . . . Hostetler, C. (2009). The CALIPSO lidar cloud and aerosol discrimination: Version 2 algorithm and initial assessment of performance. *Journal of Atmospheric and Oceanic Technology, 26*(7), 1198-1213.

Martins, D. (2018). Pollen season looms. Who'll get it worst this year? Retrieved from http://www.theweathernetwork.com/news/articles/pollen-season-whos-getting-it-worse/97973

Mazzoni, D., Logan, J. A., Diner, D., Kahn, R., Tong, L., & Li, Q. (2007). A data-mining approach to associating MISR smoke plume heights with MODIS fire measurements. *Remote Sensing of Environment, 107*(1-2), 138-148.

McLean, K., Stranberg, R., MacDonald, M., Richardson, G., Kosatsky, T., & Henderson, S. (2018). Establishing Heat Alert Thresholds for the Varied Climatic Regions of British Columbia, Canada. *International journal of environmental research and public health, 15*(9), 2048.

McPherson, M., García-García, A., Cuesta-Valero, F. J., Beltrami, H., Hansen-Ketchum, P., MacDougall, D., & Ogden, N. H. (2017). Expansion of the Lyme disease vector Ixodes scapularis in Canada inferred from CMIP5 climate projections. *Environmental health perspectives, 125*(5), 057008.

Mosavi, A., Ozturk, P., & Chau, K.-w. (2018). Flood prediction using machine learning models: Literature review. *Water, 10*(11), 1536.

National Aeronautics and Space Administration. (2012). Global Warming. doi:https://earthobservatory.nasa.gov/

National Aeronautics and Space Administration. (2015). What's the Difference Between Weather and Climate? doi:http://www.nasa.gov/mission_pages/noaa-n/climate/climate_weather.html

National Institutes of Health. (2019). What is Big Data? | Data Science at NIH.   Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/

Natural Resources Canada. (2019). Fire Monitoring, Mapping, and Modeling (Fire M3). Retrieved from http://cwfis.cfs.nrcan.gc.ca/background/summary/fm3

Naylor, C. D. (2018). On the Prospects for a (Deep) Learning Health Care System. *Jama, 320*(11), 1099-1100. doi:10.1001/jama.2018.11103

Nguyen, Q. C., Li, D., Meng, H. W., Kath, S., Nsoesie, E., Li, F., & Wen, M. (2016). Building a National Neighborhood Dataset From Geotagged Twitter Data for Indicators of Happiness, Diet, and Physical Activity. *JMIR Public Health Surveill, 2*(2), e158. doi:10.2196/publichealth.5869

Olson, D. R. K., K. J. Paladini, M. Viboud, C. Simonsen, L. (2013). Reassessing Google Flu Trends Data for Detection of Seasonal and Pandemic Influenza: A Comparative Epidemiological Study at Three Geographic Scales *PLoS Comput Biol* (Vol. 9).

Omar, A. H., Winker, D. M., Vaughan, M. A., Hu, Y., Trepte, C. R., Ferrare, R. A., . . . Rogers, R. R. (2009). The CALIPSO automated aerosol classification and lidar ratio selection algorithm. *Journal of Atmospheric and Oceanic Technology, 26*(10), 1994-2014.

Provost, F. F., Tom. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *https://home.liebertpub.com/big*. doi:10.1089/big.2013.1508

Public Health Agency of Canada. (2006). Enteric disease: A major health concern in Canada - Canada.ca.

Public Health Agency of Canada. (2013a). Climate change and public health fact sheets - Canada.ca.

Public Health Agency of Canada. (2013b). Preventative Public Health Systems and Adaptation to a Changing Climate Program - Canada.ca.

Public Health Agency of Canada. (2017). West Nile virus 2017.

Public Health Agency of Canada. (2019). Microbial foodborne diseases due to climate change - Canada.ca.

QS Top Universities. (2018). Computer Science & Information Systems.

Rabson, M. (2019, 2019-07-18). Lyme-infected ticks are so common in parts of Canada, testing no longer done. *The Globe and Mail Canada*. Retrieved from https://www.theglobeandmail.com/canada/article-lyme-infected-ticks-are-so-common-in-parts-of-canada-testing-no-2/

Reid, C. E., Considine, E. M., Watson, G. L., Telesca, D., Pfister, G. G., & Jerrett, M. (2019). Associations between respiratory health and ozone and fine particulate matter during a wildfire event. *Environment international, 129*, 291-298.

Rieger, S. (2018). Alberta adds 406 seats to train students in face of projected tech worker shortage | CBC News.

Rocha, R. (2017). These are all the hurricanes that hit Canada since 1900.

Schmelzer, R. (2019). Amazon Dives Deep into Reinforcement Learning.

Séguin, J., Berry, P., Bouchet, V., Clarke, K.-L., Furgal, C., Environmental, I., & MacIver, D. (2008). Human health in a changing climate: a Canadian assessment of vulnerabilities and adaptive capacity. *Human health in a changing climate, 1*.

Semenza, J. C., Herbst, S., Rechenburg, A., Suk, J. E., Höser, C., Schreiber, C., & Kistemann, T. (2012). Climate change impact assessment of food-and waterborne diseases. *Critical reviews in environmental science and technology, 42*(8), 857-890.

Shanghai Ranking. (2019). Global Ranking of Academic Subjects 2019 - Computer Science and Engineering.   Retrieved from http://www.shanghairanking.com/Shanghairanking-Subject-Rankings/computer-science-engineering.html

Shortliffe, E. H., & Sepulveda, M. J. (2018). Clinical Decision Support in the Era of Artificial Intelligence. *Jama, 320*(21), 2199-2200. doi:10.1001/jama.2018.17163

Singh, M., Martins, L. M., Joanis, P., & Mago, V. K. Building a cardiovascular disease predictive model using structural equation model & fuzzy cognitive map.

Slingo, J. P., T. (2011). Uncertainty in weather and climate prediction *Philos Trans A Math Phys Eng Sci* (Vol. 369, pp. 4751-4767).

Sonenshine, D. (2018). Range expansion of tick disease vectors in North America: implications for spread of tick-borne disease. *International journal of environmental research and public health, 15*(3), 478.

Spicer, Z., Olmstead, N., & Goodman, N. (2018). Reversing the brain drain: where is Canadian STEM talent going. *Munk School of*.

Talloni-Álvarez, N. E. S., Rashid U. Le Billon, Philippe. Cheung, William W.L. (2019). Climate change impact on Canada's Pacific marine ecosystem: The current state of knowledge.

Teja, S. (2019). Here's Why Canada Can Win The AI Race.

Toutant, S., Gosselin, P., Belanger, D., Bustinza, R., & Rivest, S. (2011). An open source web application for the surveillance and prevention of the impacts on public health of extreme meteorological events: the SUPREME system. *International Journal of Health Geographics, 10*(1), 39. doi:doi:10.1186/1476-072X-10-39

University of Toronto. (2019). Research guides: Searching the Literature: A Guide to Comprehensive Searching in the Health Sciences: Grey Literature. doi:https://guides.library.utoronto.ca/c.php?g=577919&p=4123572

Vaughan, M. A., Powell, K. A., Winker, D. M., Hostetler, C. A., Kuehn, R. E., Hunt, W. H., . . . McGill, M. J. (2009). Fully automated detection of cloud and aerosol layers in the CALIPSO lidar measurements. *Journal of Atmospheric and Oceanic Technology, 26*(10), 2034-2050.

Vaughan, M. A., Winker, D. M., & Powell, K. A. (2005). CALIOP algorithm theoretical basis document, part 2: Feature detection and layer properties algorithms. *Rep. PC-SCI, 202*, 87.

Vu, M.-A. T., Adalı, T., Ba, D., Buzsáki, G., Carlson, D., Heller, K., . . . Widge, A. S. (2018). A shared vision for machine learning in neuroscience. *Journal of Neuroscience*, *38*(7), 1601-1607.

Wang, Y., Song, Q., Du, Y., Wang, J., Zhou, J., Du, Z., & Li, T. (2019). A random forest model to predict heatstroke occurrence for heatwave in China. *Science of the Total Environment, 650*, 3048-3053.

Wei, S. (2018). River ice breakup timing prediction through stacking multi-type model trees. *Science Direct*.

World Health Organization. (2016). WHO | WHO calls for urgent action to protect health from climate change – Sign the call. *WHO*. doi:/entity/globalchange/global-campaign/cop21/en/index.html

Yao, J., Brauer, M., & Henderson, S. B. (2013). Evaluation of a wildfire smoke forecasting system as a tool for public health protection. *Environmental health perspectives, 121*(10), 1142-1147.

Yao, J., Brauer, M., Raffuse, S., & Henderson, S. B. (2018). Machine Learning Approach To Estimate Hourly Exposure to Fine Particulate Matter for Urban, Rural, and Remote Populations during Wildfire Seasons. *Environmental science & technology, 52*(22), 13239-13249.

Yao, J., Raffuse, S. M., Brauer, M., Williamson, G. J., Bowman, D. M., Johnston, F. H., & Henderson, S. B. (2018). Predicting the minimum height of forest fire smoke within the atmosphere using machine learning and data from the CALIPSO satellite. *Remote sensing of environment, 206*, 98-106.

Young, S. G. T., Jason A. Cothren, Jackson. (2013). A remote sensing and GIS-assisted landscape epidemiology approach to West Nile virus.

Zewdie, G. K., Lary, D. J., Liu, X., Wu, D., & Levetin, E. (2019). Estimating the daily pollen concentration in the atmosphere using machine learning and NEXRAD weather radar data. *Environ Monit Assess, 191*(7), 418. doi:10.1007/s10661-019-7542-9

Zou, Y., O'Neill, S. M., Larkin, N. K., Alvarado, E. C., Solomon, R., Mass, C., . . . Shen, H. (2019). Machine Learning-Based Integration of High-Resolution Wildfire Smoke Simulations and Observations for Regional Health Impact Assessment. *International journal of environmental research and public health*, *16*(12), 2137.

Appendix A: PRISMA Chart