Universal Journal of Computational Analysis **10 (2022), 18-24**
www.papersciences.com EU Academy of Sciences (EUAS) e-Symposium

Use of Artificial Intelligence and Sampling Methodologies to Investigate the Occurrence and Severity of COVID-19 Outbreaks

Edward McBean
University of Guelph
Guelph, Ontario
Canada
emcbean@uoguelph.ca

Abstract

The Severe Acute Respiratory Syndrome Coronavirus 2 (or COVID-19) pandemic has challenged medical systems around the globe to the brink of collapse. A critical feature relevant to the degree to which medical systems are constrained is the availability of accurate predictions of upcoming needs for space for hospitalization and intensive care units (ICU), as well as patient death rates. All these predictions are relevant to assess whether the medical systems will be overwhelmed with patients, including the degree of surges of COVID-19 caseloads, to establish if the medical systems will be able to handle the extremes as well as to still be able to respond to other, ongoing hospital needs. To provide such, this paper describes several alternative models including Artificial Intelligence and Survival Statistics models being developed to assess caseloads, thereby providing insights into future magnitudes of caseloads. Specific metrics currently in development and some completed, namely, accuracy, recall, and precision using examples associated with several variants of the COVID-19 are described, where these describe measures of accuracy in the making of the predictions. Emphasis of the specific impacts of the Delta variant with specifics on death rates and ICU are indicated as being more severe than those resulting from the Omicron variant, despite Omicron's high transmissibility.

Key Word and Phrases
XGboost, Kaplan-Meier, Artificial Intelligence, Omicron Variant, Delta Variant, Survival Analysis, COVID-19.

1. Introduction

The worldwide, rapid spread of the Severe Acute Respiratory Syndrome Coronavirus 2 (or COVID-19) pandemic has affected many millions of individuals and caused unprecedented medical challenges. The results are causing healthcare services to be strained in terms of capacities and personnel, due to the enormous pressures. With the impact of the COVID-19 crisis being felt on a global scale and being collectively frustrated by the limited ability to predict future changes in the virus trajectory surges, questions and concerns as to how to respond have continued to cause great anguish.

The needs to make predictions have had huge impacts on choices surrounding behavioral directives (e.g. wearing of masks, and ability to attend sports events related to physical proximity), influence decisions made for individual country's healthcare systems. In Canada, we knew there was real potential that Canada's hospital system could become overwhelmed, and this possibility underlined the need to have available, forecasting models to provide the medical profession and the associated government entities, but how to create information assemblies with the needed information (e.g., for people who test positive for COVID-19, how many people are likely going to be admitted to hospitals, how many will require intensive care or ventilators, and ultimately, how many deaths would emerge?), all with the aim to assist in there being information in a format to make informed decisions. With COVID-19 patients being so numerous (3.4 million cases in Canada, or 9% of Canadian population, with 58.7 million tests performed [1] in combination with the fluctuations and surges in numbers impacting the spaces available, assurance that useful information could be obtained in a manner that would facilitate appropriate actions were critical.

In response, the opportunities provided by modeling using Artificial Intelligence and Survival Statistics models have proven of substantial merit, providing the opportunities to identify from massive databases (e.g. in excess of 3.4 million positive cases contracting the COVID-19 virus), the databases enabled guidance on various interrelated parameters. In those contexts, published papers describing just some of the gains made in Canada through use of these powerful modeling structures, enabling insights that no human would be able to identify, are referred to below.

2. Best Insights

With millions of cases of COVID-19 in Ontario residents ending up requiring COVID-19 testing, space needs for hospitalization, some patients moving to Intensive Care Units (ICU), and unfortunately, many also resulting in deaths, the medical systems in Canada were faced with many tradeoffs having to be made during the ongoing surges of the COVID-19 virus. The challenge for researchers was to improve the understanding of how, for example, the environment, people's behavioral patterns, people's age, and many other individual attributes (e.g. co-morbidities such as presence/absence of chronic dementia, complex pneumonia, immunocompromised individuals, etc.) were influencing the escalation of caseloads with many requiring access to the hospital system. In this context, at the University of Guelph, we explored an array of parameter estimation technologies to determine the optimal population segmentation strategies and results, developing an array of models as briefly described in the following sections.

2.1 Artificial Intelligence (AI) Models to Identify Important Co-Mobidities associated with COVID-19

At an early stage of the COVID-19 caseloads in 2020 in Ontario, to improve the assessments of the magnitudes of hospitalizations, patients needing Intensive Care Units and in excess of 37,000 deaths in Canada [2], we utilized confirmed cases (e.g. timing of reporting date of infection), information about infected individuals by including age, gender, location, income level (as available), travel history etc. (utilizing a total of 24 metrics determined from OHDP data, and utilizing various AI options employing XGBoost, Artificial Neural Networks, and Random Forest, these models have been able to translate confirmed cases to the quantity of admitted people to the hospitals and the number of deaths were developed, the models were able to predict mortality and recovery of COVID-19 patients with high degrees of accuracy [3]-[5]. In the early stages of the pandemic, the most important parameters were found to be age, test date, sex, presence/absence of chronic dementia and residency in Long Term Care (LTC) facilities. The importance of 'test date' was determined to be attributable to the learning curve ongoing within the medical profession on how to treat patients with COVID-19 indicating that the later that a patient became ill, the less likely was the occurrence of death. Residency in LTCs has continued to be important in terms of death rates throughout the duration of the pandemic.

These types of models have proved to have the ability to determine the set of characteristics (i.e. segment) of the population that showed a higher propensity for obtaining COVID-19. With slight modifications, this AI model was able to be leveraged as part of a risk classifier for individuals based on demographic information coupled with extent and severity of their symptoms showing that AI models could deal with enormous quantities of data and were demonstrated using SHAP values, showed that consistently, the most important features influencing the likelihood of death of a positive COVID-19 case were related to age, test date, sex, hypertension, LTC resident, chronic dementia, etc. [3].

2.2 Assessing the Influence of Alternative Approaches to Open Up the Economy

With huge impacts to the economy, there were many pleas to 'open up' the economy; however, this involved the potential for the impact from opening up the economy would cause a resurgence of COVID-19 caseloads. For example, concerns regarding the impact of opening daycares and day camps were sought, to alleviate challenges (in particular, for families with young children, to allow the parents to work while their children were in school and/or day care). In this context, a SEIR Model ('S' being the fraction of susceptible individuals, 'E' the fraction of exposed individuals, 'I' the fraction of infective individuals, and 'R' the fraction of recovered individuals) [6],[7] were utilized to assess the potential spread of COVID-19 caseloads, to reflect how caseload numbers might reverberate through the population in response to the opening of daycare and summer day camps. This type of modeling was able to characterize actions involving reduction of the capacity (e.g. lowering student numbers in a classroom) and contact rates through social distancing protocols, and hence, the additional caseloads that might have otherwise occurred, could reduce the caseload growth for Toronto by 88%, highlighting the importance of managing social distancing protocols as effective measures to help control the spread of COVID-19.

2.3 A Survival Analysis Approach to Evaluate the Likelihood of Testing Positive for COVID-19 following Alternative Vaccinations

Analyses of the effectiveness and longevity of COVID-19 has become the forefront of pandemic-related research, where around the world, countries have strongly encouraged vaccination as the most assured method to curtail the need for stringent public health measures. Survival analysis is a familiar statistical concept in the many realms of medicine, wherein this modeling procedure has been used to calculate the probability of survival (from a defined event) using time-to-event data. In COVID-19 research, use of applications of survival analysis have been rarely published (although applied in other applications, e.g. [8], [9], despite having significant potential to advise the direction of the pandemic in many facets. One of these facets is the use of survival estimates to understand the probability of an individual testing positive for the virus in populations of vaccinated individuals. The results of these analyses demonstrate significant potential to indicate the merits of vaccines. For example, [10] have examined use of Kaplan-Meier protocols, a form of survival analysis (as a means of incorporating right-censored information), to provide a statistical approach to improve the understanding of time-to-event likelihoods of occurrence of positivity to the COVID virus. Using applications of epidemiology and the study of vaccines, survival analyses were used to quantify the probability of testing positive.

As portrayed in [10], the results have been used to indicate the probability of testing positive for COVID-19 given a populations' vaccination status. For example, for the population of Ontario, COVID-19 testing data enabled the demonstration that for individuals with two doses of a vaccine during a peak period of Delta variant cases, those receiving a COVID-19 vaccine (two doses for a population of ~600,000 individuals and ~50,000 three-dose recipients), that vaccines of Moderna and Pfizer provided very good protection from COVID-19 infection, showing low positivity for 38 day periods (plus an assumed 14 day lag period following the injection of a vaccine into an individual, where the majority of vaccines are effective, and provided better than 95% protection against testing positive). The evaluated probabilities of testing positivity align with the publically-reported vaccine efficacies of the mRNA vaccines, supported that resolution using Kaplan-Meier methods in determining vaccine benefits is a justifiable and useful approach in addressing vaccine-related effectiveness in the COVID-19 landscape [10].

2.4 Assessment of the Relative Impacts Arising from the Delta Variant relative to the Omicron Variant

North America, and as evident also in many other countries, an array of variants of COVID-19 have evolved and, in many respects, were responsible for successive surges in caseloads. Perhaps the most virulent examples are that of first the Delta variant, which was then later followed by the Omicron variant, as being the most egregious examples (see: Figure 1). The evidence is now quite clear that the initial predictions that while the Omicron variant is considerably more transmissible in comparison with the Delta variant, the health impacts of the Omicron have to date been substantially less impactful, where a summary of the relative magnitudes of ICU and deaths are summarized by the available databases in Ontario have demonstrated [11].

Uses of AI models (as of March 2022) are also demonstrating their ability to assess the progression of caseloads to the medical system, arising from the individual variants [11]. Given the sizes of the databases, the ability of AI models to assess the implications of alternatives Hospitalization, ICU, and deaths associated with alternative variants, the results are able to attain high resolution regarding all of the following: accuracy (both positive and negative to the total number of cases), recall (true positive divided by the summation of true positive + false negative), precision (true positive divided by the summation of true positive + false positives) options. The ability to evaluate the implications of these massive databases is apparent, given their ability to assess the merits of various actions, and the importance of individual attributes relative to one another, and develop important insights.

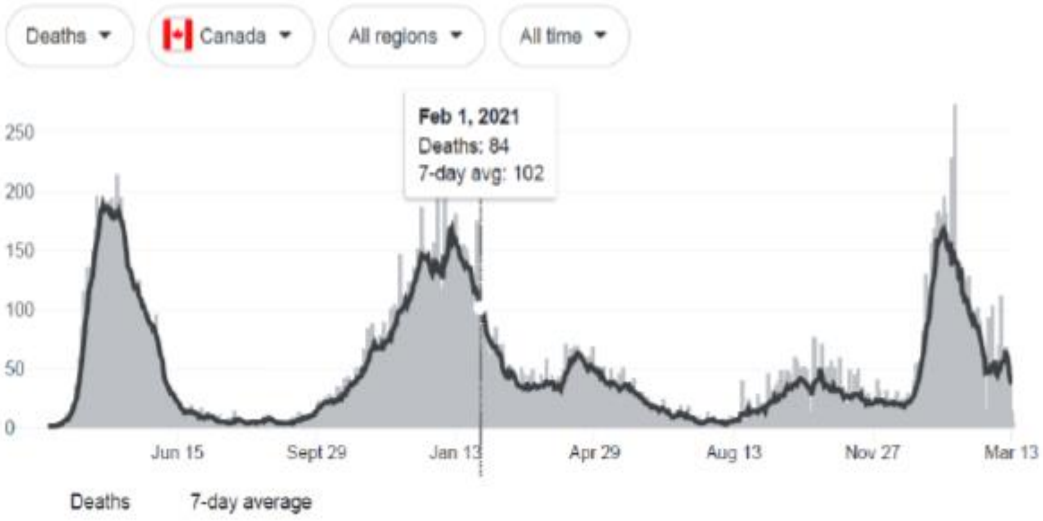

Fig. 1 Deaths from COVID-19 in Canada for Period of March 8, 2020 to March 13, 2022 [12]

2.5 Linkages with Wastewater-Based Epidemiology

The years of the pandemic harrowing nations and devastating economies are, unfortunately, only the next step in the issues that will continue to plague the world. Human migration has always meant disease transmission, but modern jet travel carries viruses globally within hours, meaning that an outbreak anywhere is an outbreak *everywhere*. Known in the field as "Wastewater-Based Epidemiology" (WBE), studying and monitoring what exactly is in wastewaters (SARS-CoV-2, norovirus, salmonella, and many more) can characterize epidemiology beyond individual case testing. Figures 2(a) and (b) show a photo of the exterior of the building and the floorplan layout of the facility situated adjacent to the Guelph wastewater treatment plant.

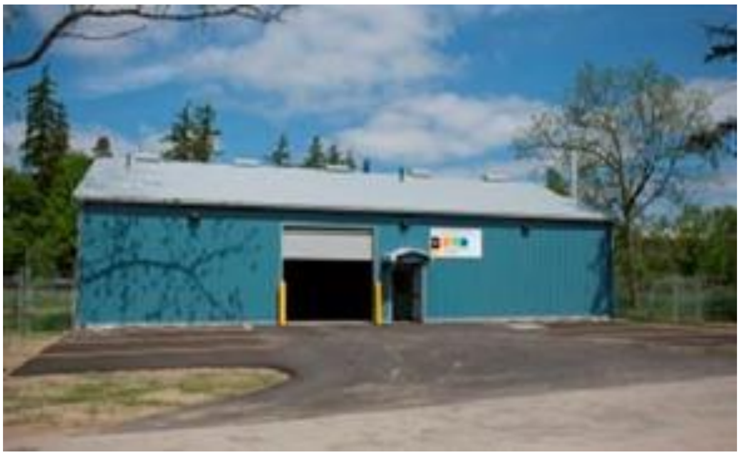

Fig. 2(a) Wastewater-Based Epidemiology Facility, Guelph, Ontario
(facility is adjacent to the wastewater treatment plant for the City of Guelph, allowing access to points of withdrawal of wastewater to be accessed for water quality testing)

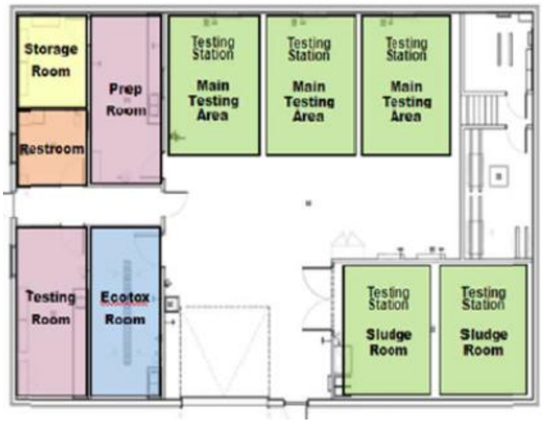

Fig. 2(b) Layout of Interior of WBE Facility

Success of wastewater epidemiology is in the forefront of understanding growth rates of pandemics. Ongoing research in WBE includes attention to duration of wastewater sampling using passive samplers for COVID-19 as described in [13], and assessing the values of different sampling methodologies (e.g. [14]) while we continue to explore surges and periods of subsidence in the number of positive cases of COVID-19 in wastewater, using the research capabilities such as the wastewater pilot facility at the University of Guelph (see: Figure 3).

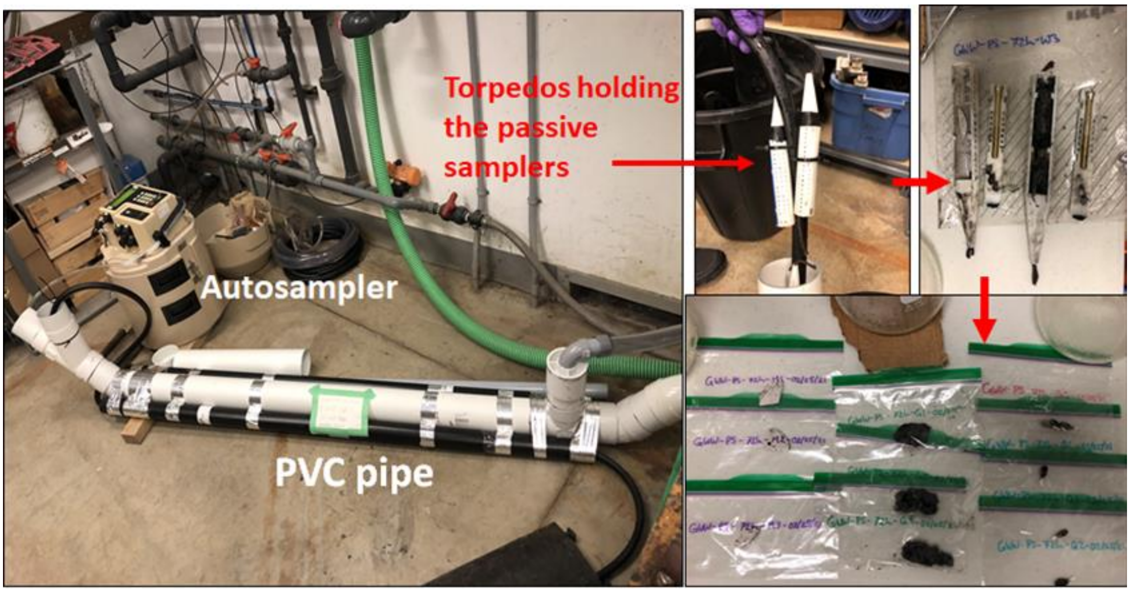

Fig. 3 Experimental Set-up of COVID-19 Wastewater Research

3. Conclusions

The challenges of the COVID-19 pandemic continue but the data are showing that declines in severity of the variants of the virus are evident. From the basis of the enormous quantitative information datasets which have been assembled, and AI and Statistical Survival modeling procedures are enabling the assessment, improvement in understanding, and ability to inform the caseloads associated with the medical systems.

Specifics of the ongoing learning are showing:

(i) The XGBoost AI algorithm is proving very effective at deriving the importance of alternative co-morbidities and assessing the directions/influencing of caseload trajectories. It is noted that specific attention needs to be paid to collinearities in the AI analyses since not appropriately reflecting this can significantly influence the importance of various attributes;

(ii) Survival Analysis modeling can provide important insights into right-censored data as a general methodology, and provides the opportunity to predict failure probabilities (e.g. estimates of testing positive for COVID-19);

(iii) SEIR models can be effective at assessing the impacts of alternative strategies to contain outbreaks of the virus, assuming that accurate measures of contact can be estimated;

(iv) Wastewater-based epidemiology will continue to grow in utilization since many types of waterborne pathogens (and drugs) can be identified in the feces/urine of populations. While not able to identify specific individuals, WBE can be used to identify the growth/decay/resurgence within contributing populations and hence, provide excellent insight into pandemic patterns.

Acknowledgements

Supporting agencies: This research was supported in part by ICES, which is funded by an annual grant from the Ontario Ministry of Health (MOH) and the Ministry of Long-Term Care (MLTC). The research was supported by the Ontario Health Data Platform (OHDP), a Province of Ontario initiative to support Ontario's ongoing response to COVID-19 and its related impacts. The opinions, results and conclusions reported in this paper are those of the authors and are independent from the funding sources. No endorsement by the OHDP, its partners, or the Province of Ontario is intended or should be inferred.

The financial support of the Natural Sciences and Engineering Research Council of Canada (NSERC) via 400677, 401636 and 401641 are gratefully acknowledged. Funding from the University of Guelph Leadership Chair program is also gratefully acknowledged.

References

1. *Government of Canada, "COVID-19 Daily Epidemiology Update"*, https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html, *March 14, 2022.*
2. *Johns Hopkins University, Visual Dashboard (desktop):* https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6, *March 13, 2022.*
3. *Snider B., Phillips P., MacLean A., McBean E., Gadsden A., Yawney J.,* "Artificial Intelligence to Predict the Risk of Mortality from COVID-19; Insights from a Canadian Application", *IEEE Canadian Journal of Electrical and Computer Engineering, 2020*
4. *Snider B., McBean E., Yawney J., Gadsden A., Patel B.*, "Identification of Variable Importance for Predictions of Mortality from COVID-19 Using AI Models for Ontario, Canada", *Frontiers in Public Health, Infectious Diseases – Surveillance, Prevention and Treatment, **9**(2021), 1-8.*
5. *Snider B., Patel B., McBean E.*, "Insights into Co-Morbidity and Other Risk Factors related to COVID-19 within Ontario, Canada", *Frontiers in Artificial Intelligence, **4**(2021), 1-8.*
6. *Snider B., Patel B., McBean E.*, "Asymptomatic Cases, the Hidden Challenge in Predicting COVID-19 Caseload Increases", *Infectious Disease Reps, **13**(2021), 340-347.*
7. *Snider B., Phillips P., McBean E., Yawney J., Gadsden S.A*., "Influence of Opening Up Day-Care and Day Camps on Resurgence Potential of COVID-19 Pandemic: Assessing Infectivity Potential from Youth in Ontario, Canada", *IEEE Transactions on Computational Social Systems, **8**(2021), 1052-1060.*
8. *Henderson R., Jones M., Stare J.*, "Accuracy of point predictions in survival analysis", *Statist. Med., **20**(2001), 3083-3096.*
9. *Kenah E., Britton T., Halloran E., Longini I.,* "Molecular Infectious Disease Epidemiology: Survival Analysis and Algorithms Linking Phylogenies to Transmission Trees", *PLoS Compt Biol., **12**(2016), p.e1004869.*
10. *Hilal W., Chislett M., Snider B., McBean E., Yawney J., Gadsden A.,* "A Survival Analysis Approach Evaluating the Likelihood of Testing Positive for COVID-19 post-vaccination", *IEEE Access, under review, 2022.*

11. *Hilal W., Chislett M, Snider B., McBean E., Yawney J., Gadsden A.,* "Use of AI to Assess the Relative Impacts of the Delta and Omicron Variants on Rates of Hospitalization, ICU, and Death", *IEEE Access, under review, 2022.*
12. *Johns Hopkins University, website, March 13, 2022.*
13. *Habtewold J., McCarthy D., McBean E., Law I., Goodridge L., Habash M., Murphy H.*, 2022. "Passive Sampling, a Practical Method for Wastewater-Based Surveillance of SARS-CoV-2", *Environmental Research, 204-112058, 2022.*
14. *Jiang A., Nian F., Chen H., McBean E., "Passive Samplers, an Important Tool for Understanding the COVID-19 Pandemic – with Insights to the outbreak in Nanjing of July 2021"*, Environmental Reviews. Springer, *2022.*