PREDICTING THE SURGE: FORECASTING ONTARIO'S CHANGING MENTAL HEALTH NEEDS

Kristina L. Kupferschmidt[1,2], Cody Kupferschmidt[1], Joshua A. Skorburg[3], Branka Agic[4], Tara Elton-Marshall[4,5], Hayley Hamilton[4], Gina Stoduto[4], Katherine Vink[6,7], Samantha Wells[4,8], Christine Wickens[4,8], Graham W. Taylor[1,2]
[1] School of Engineering, University of Guelph, Guelph, Canada
[2] Vector Institute, Toronto, Canada
[3] Department of Philosophy, University of Guelph
[4] Centre for Addiction and Mental Health (CAMH), Toronto, Canada
[5] School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada
[6] Ministry of Children and Family Development, Government of British Columbia, Victoria, Canada
[7] Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Canada
[8] Dalla Lana School of Public Health, University of Toronto, Toronto, Canada
kupfersk@uoguelph.ca, kupfersc@uoguelph.ca, skorburg@uoguelph.ca, branka.agic@camh.ca, tara.eltonmarshall@camh.ca, hayley.hamilton@camh.ca, gina.stoduto@camh.ca, katherine.vink@gov.bc.ca, samantha.wells@camh.ca, christine.wickens@camh.ca, gwtaylor@uoguelph.ca

ABSTRACT

We implemented several state-of-the-art Machine Learning (ML)-based forecasting techniques to predict traffic to the Connex Ontario platform, a government-funded helpline and referral service for individuals struggling with mental health and addiction in Ontario, Canada. This research aims to serve as a proof of concept to demonstrate the value of developing forecasting models that incorporate alternative datasets collected from platforms other than electronic health records. Preliminary findings for the test period of March 1, 2020 to Feb 1, 2023 suggest that sophisticated ML models are able to harness covariates to improve accuracy for the tasks of producing 1-, 4-, or 12-week forecasts for the volume of mental health outreach. These forecasts have the potential to provide clarity regarding fluctuations in mental health service needs and allow policy-makers to make data-driven decisions, such as proactively allocating resources to reduce the strain on Ontario's mental health care system.

1 INTRODUCTION

In Canada, access to mental health care was problematic before the emergence of COVID-19. In 2018, the Canadian Community Health Survey found that 43.8% of survey respondents (n=5.3 million) felt that their mental health needs were either unmet or only partially met.[1] Timely access to treatments remains a primary concern; many Canadians do not have access to a primary care provider and many of those who do are unable to access that provider on short-notice, resulting in increased dependency on acute care services such as emergency departments.[2] The COVID-19 pandemic further exacerbated the fragility of Canada's mental health care system. The early phases of the pandemic resulted in increases in symptoms of poor mental health, demonstrated by rising numbers of outreaches to mental health supports from local branches of the Canadian Mental

[1] https://www150.statcan.gc.ca/n1/pub/82-625-x/2019001/article/00011-eng.htm
[2] https://www.cihi.ca/sites/default/files/document/mental-health-chartbook-report-2019-en-web.pdf

Health Association (CMHA)[3] and reports in nationwide surveys (Wickens et al., 2023).[4] To address the growing demand for online mental health services associated with the COVID-19 pandemic, governments at both the provincial and federal level expanded financial supports for mental health platforms to facilitate online counselling services, virtual mental health care, and crisis outreach (Moroz et al., 2020).

The need to optimize access to mental health services to proactively address societal changes is not inherently different from challenges faced by acute care health services during the height of the COVID-19 pandemic. Forecasting the magnitude and timing of service utilization during infection surges allowed policymakers to consider the limitations of the healthcare system and act proactively to prevent situations where service needs exceed availability. Similar predictions of the magnitude and timing of increases in care needs are critical for the mental health system. Accurate predictions of future demand for mental health services have the potential to inform efficient allocation of resources, educate the implementation of new community services such as mental health promotion and e-mental health outreach, and prioritize care for those known to be at increased risk (Canadian Mental Health Association, 2020).

In the context of mental health, the task of dataset curation for forecasting is complex, as there is no centralized source of data that can be used. Although much of the predictive machine learning-based research for mental health applications to date use electronic health records (EHRs), there are associated risks (Chekroud et al., 2021). For example, EHRs often only include interactions with licensed professionals (e.g. physicians), while in reality, mental health interventions can involve a range of professionals including clinical psychologists, psychotherapists, and social workers. Surveys regarding mental health and mental health service utilization also have limitations, such as convenience sampling (Patten et al., 2021). Thus, the use of EHRs or any other single data source is likely to provide an incomplete view of the total usage of mental health care services and alternative data sources should be explored.

In recent years there have been major advances in leveraging machine learning (ML) and deep learning-based models to produce forecasts (Makridakis et al., 2023). Basic univariate forecasting, which can be completed using either classical or ML-based training methods, involves predicting future values for a single target time-series solely based on past values from the same series. However, these models are unable to capture effects from external factors that influence the target time-series. In particular, univariate systems have shown limited effectiveness in forecasting the rapidly changing, non-stationary time series that characterize mental health trends during the COVID-19 pandemic period. This limitation is notable in the context of the pandemic's complex and evolving impact on mental health dynamics.

To address this, covariates or global training objectives can be integrated into univariate models, allowing for the inclusion of external variables such as economic, social, or environmental indicators that are believed to influence the target time-series (Januschowski et al., 2020). Despite the greater capacity of sophisticated ML-based global models, it is important to note that they do not universally outperform simpler models such as pure univariate models in all forecasting tasks (Makridakis et al., 2023).

The objective of this research was to compare the performance of ML-based time-series forecasting models with other statistical techniques for the task of predicting the number of individuals requiring mental health care from alternative-care platforms in Ontario for forecast horizons of 1, 4, and 12 weeks. Specifically, experimentation was completed for the time period spanning the COVID-19 pandemic, taking into account outbreak trends related to COVID-19 as well as pandemic-specific search terms. Long term, we aim to validate models allowing for meaningful probabilistic estimates that can be used by decision-makers, in both policy and clinical roles to proactively plan for upcoming service utilization needs.

[3]https://cmho.org/covid-19-mental-health-impacts/

[4]https://ontario.cmha.ca/news/1-in-4-ontarians-access-mental-health-help-the-highest-rate-during-the-pandemic/

2 METHODOLOGY

2.1 DATASETS

Several datasets were selected to complete the forecasting tasks and be included as either a (1) target variable that is being predicted by the forecasting model, or (2) external covariates to provide additional context to the model when predicting the target variable. All covariates were chosen in consultation with team collaborators who are also mental health researchers and service providers and were selected as indicators of social and population factors thought to exacerbate mental health conditions.

Target Dataset - Connex Ontario - Connex Ontario is an information and referral service for caretakers of individuals or individuals struggling with mental health, addiction, or problem gambling. This dataset was selected as a proxy for the number of individuals accessing alternative help platforms for mental health challenges. For this study data were filtered to only include platform interactions related specifically to mental health, excluding addiction and problem gambling. This dataset includes interaction-level data of historical chat service utilization (online, phone-based, and email) which was aggregated by date to represent the number of interactions related to mental health for each week.

COVID-19 Ontario Ministry of Health Public Data - The Ontario Ministry of Health hosts an online collection of COVID-19 datasets that compile all publicly reported data related to COVID-19 testing and hospitalizations on a daily basis. We used Daily Intensive Care Unit (ICU) Admissions of patients with COVID-19.[5] This was selected rather than Daily COVID-19 cases to account for changes in eligibility for Polymerase Chain Reaction (PCR) testing in the general population resulting in a significant shift in daily confirmed cases.

Google Trends Data - Historical frequency of Google search terms for specific time frames and geographic locations are available through the Google Trends platform. The use of this data has been successful for related predictive tasks such as suicide rates and some influenza epidemics (Barros et al., 2019; Ginsberg et al., 2009). To evaluate the impacts of different stressors we defined 5 general themes (Table A.2) of search terms : (1) mental distress, (2) economic stressors, (3) social stressors, (4) treatment-seeking, and (5) COVID information. With the exception of common illness information, these search terms were defined and validated by Knipe et al. (2020). To adjust to the Canadian context, several search terms were updated for regional vernacular (e.g. *ei* and *employment insurance* for welfare) and results were filtered for IP addresses from Ontario, Canada.

Statistics Canada - Consumer Price Index - Statistics Canada publishes the consumer price index (CPI), a commonly used measure of the cost of living for Canadians, monthly. CPI was chosen as a covariate to represent fluctuations to cost of living thought to aggravate mental illness. It includes eight major components including food, shelter, household operations, and others.[6] CPI values for each component as well as an overall value are published with monthly resolution. In this study, all data were filtered to exclusively include CPI values for Ontario.

2.2 MODELS

The performance of ten models was evaluated for their ability to produce forecasts with the objective of minimizing mean absolute percent error (MAPE). All forecasts were univariate in nature but included five statistical and five ML-based models (Table A.1). To compare traditional and ML-based approaches, several baseline time-series techniques were evaluated, including variants of naïve, AutoRegressive Integrated Moving Average (ARIMA), and exponential smoothing with trend and seasonality (ETS). The included ML-based techniques spanned multiple approaches including global and local training paradigms both with and without the inclusion of covariates. All experiments were implemented using the Autogluon library (v0.8.2).[7]

[5] `https://data.ontario.ca/dataset/covid-19-cases-in-hospital-and-icu-by-ontario-health-region`

[6] `https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1810000401`

[7] `https://github.com/awslabs/autogluon`

Forecasts for each model were evaluated over seven simulated test windows, all of which fell within a larger test period of 2020-03-01 to 2023-02-01. These test windows were chosen to evaluate model performance during subsequent waves of the COVID-19 pandemic, with much stronger containment measures being present during earlier test periods (Table A.3). Each test window began approximately from the initial onset of the COVID-19 wave (as defined by the Ontario Health Ministry). A rolling evaluation approach was employed meaning that forecasts for all horizons were produced for each day within the test window. To make each prediction, the model was provided with the context of the preceding 52 weeks to the prediction date (Table A.3). Model parameters were trained based on a training data set consisting of historical data from the two years prior (104 weeks) for each of the test windows. Model performance was evaluated using symmetric mean absolute percent error (sMAPE). The performance for each of the seven test windows was evaluated independently for each of the three forecasting horizons (i.e. 1, 4, and 12 weeks).

3 Results and Discussion

3.1 Overall model performance

Initial results suggested that in many scenarios the ML-based models were able to outperform traditional statistical approaches (Table 1). For the 12-week forecasting horizon, the ML-based models outperformed statistical approaches for every test window, as well as being the best performing for six out of seven test windows for the 1- and 4-week tasks. When the average performance across all test windows was considered, the statistical AutoETS model performed best for the 1- and 4-week tasks (equal MAPE to DeepAR and AutoARIMA).

Table 1: Mean absolute percent error (sMAPE) values for each test period and forecasting horizon

1-Week Forecasting Horizon								
Model	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Overall(μ)
DLinear	0.134	0.121	0.127	0.144	**0.127**	0.202	0.197	0.150
DeepAR	0.231	0.155	**0.111**	**0.132**	0.138	0.209	0.218	0.170
PatchTST	0.283	0.121	0.119	0.137	0.133	**0.159**	0.204	0.165
RecursiveTabular	0.133	**0.106**	0.138	0.163	**0.127**	0.177	0.243	0.155
TemporalFusionTransformer	0.210	**0.106**	0.117	0.141	0.137	0.166	**0.182**	0.151
AutoARIMA	0.154	0.109	0.118	0.140	0.136	0.176	0.203	0.148
AutoETS	0.136	0.112	0.117	0.145	0.142	0.178	0.199	**0.147**
ETS	0.133	0.111	0.119	0.146	0.143	0.183	0.203	0.148
SeasonalNaive	**0.131**	0.112	0.147	0.161	0.169	0.204	0.254	0.168
4-Week Forecasting Horizon								
Model	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Overall(μ)
DLinear	0.297	0.207	0.117	0.162	0.153	0.213	0.153	0.186
DeepAR	0.286	**0.141**	**0.090**	0.158	0.133	0.194	0.214	**0.174**
PatchTST	0.340	0.147	0.102	**0.140**	0.143	0.177	0.229	0.183
RecursiveTabular	0.298	0.200	0.128	0.171	0.156	0.251	0.154	0.194
TemporalFusionTransformer	0.320	0.164	0.092	0.171	**0.119**	**0.163**	0.211	0.177
AutoARIMA	0.306	0.159	0.098	0.155	0.144	0.188	0.169	**0.174**
AutoETS	**0.274**	0.167	0.106	0.160	0.157	0.203	**0.148**	**0.174**
ETS	0.274	0.177	0.114	0.159	0.157	0.212	0.151	0.178
SeasonalNaive	0.287	0.180	0.140	0.183	0.173	0.231	0.170	0.195
12-Week Forecasting Horizon								
Model	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Overall(μ)
DLinear	**0.278**	0.218	0.107	0.194	0.134	0.165	0.246	**0.192**
DeepAR	0.318	0.219	0.118	0.181	0.182	**0.144**	0.302	0.209
PatchTST	0.335	**0.146**	0.143	0.171	**0.126**	0.201	0.272	0.199
RecursiveTabular	0.297	0.234	0.228	0.186	0.158	0.380	0.314	0.257
TemporalFusionTransformer	0.341	0.401	**0.083**	**0.137**	0.149	0.193	**0.215**	0.217
AutoARIMA	0.311	0.206	0.112	0.192	0.132	0.221	0.261	0.205
AutoETS	0.299	0.198	0.120	0.195	0.138	0.245	0.257	0.207
ETS	0.298	0.263	0.137	0.187	0.136	0.291	0.285	0.228
SeasonalNaive	0.300	0.189	0.149	0.212	0.170	0.250	0.297	0.224

When considering only the early stages of the pandemic (test windows 1 and 2) the DeepAR model resulted in the lowest MAPE for the 4- and 12-week tasks. For the 1-week forecasting task, the RecursiveTabular model demonstrated the lowest error, however, it was only a marginal improvement over the seasonal naïve forecast.

The results of this study suggest that model strengths are highly dependent on the context in which they are deployed. For instance, larger models with more parameters such as DeepAR, PatchTST, and the Temporal Transfusion Transformer performed significantly worse for test window 1 where there was very little training data demonstrating the impacts of the COVID-19 pandemic. On the contrary, DLinear, a simplified linear ML model, had the best performance for the 12-week task. This may be due to the fact that was an insufficient amount of training data, or lack of relevant covariates, to successfully leverage highly parameterized models such as TemporalFusionTransformer.

While these results suggest that ML-based approaches may outperform classical time-series approaches on some test-windows of the Connex dataset, they do not guarantee generalization to different time-series or even different time periods. This motivates further exploration of such models in other datasets from similar mental health platforms to better understand the impacts of specific temporal contexts.

4 Conclusions and Recommendations

Preliminary results suggest that given the current dataset, ML-based models are able to produce superior forecasts when compared with basic statistical approaches for many of the evaluated tasks. However, our work to-date has not looked specifically at any elements related to demographic representation, health equity, or evaluating algorithmic fairness within the Connex Ontario dataset. Future research directions include conducting an in-depth fairness analysis using the Connex Ontario dataset and exploring covariate importance through the application of explainability techniques.

Acknowledgments

We acknowledge the support of the Center for Additions and Mental Health (CAMH) Foundation COVID-19 MH Resiliency and Coping Fund, the Natural Sciences and Engineering Research Council of Canada (NSERC) [PGSD3-519322-2018], the Ontario Government (ORF-RE) and Vector Institute. Resources used in preparing this research were provided, in part, by the Province of Ontario, the Government of Canada through CIFAR, and companies sponsoring the Vector Institute www.vectorinstitute.ai/#partners

References

Joana M Barros, Ruth Melia, Kady Francis, John Bogue, Mary O'Sullivan, Karen Young, Rebecca A Bernert, Dietrich Rebholz-Schuhmann, and Jim Duggan. The validity of google trends search volumes for behavioral forecasting of national suicide rates in ireland. *Int. J. Environ. Res. Public Health*, 16(17), September 2019.

Canadian Mental Health Association. Canada must act now to prevent echo pandemic of mental illness due to COVID-19 - CMHA national. `https://cmha.ca/news/canada-must-act-now-to-prevent-echo-pandemic-of-mental-illness-due-to-covid-19/`, April 2020. Accessed: 2020-5-2.

Adam M Chekroud, Julia Bondar, Jaime Delgadillo, Gavin Doherty, Akash Wasil, Marjolein Fokkema, Zachary Cohen, Danielle Belgrave, Robert DeRubeis, Raquel Iniesta, et al. The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*, 20(2):154–170, 2021.

Jeremy Ginsberg, Matthew H Mohebbi, Rajan S Patel, Lynnette Brammer, Mark S Smolinski, and Larry Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232): 1012–1014, February 2009.

Tim Januschowski, Jan Gasthaus, Yuyang Wang, David Salinas, Valentin Flunkert, Michael Bohlke-Schneider, and Laurent Callot. Criteria for classifying forecasting methods. *Int. J. Forecast.*, 36 (1):167–177, January 2020.

Duleeka Knipe, Hannah Evans, Amanda Marchant, David Gunnell, and Ann John. Mapping population mental health concerns related to COVID-19 and the consequences of physical distancing: a google trends analysis. *Wellcome Open Res*, 5:82, June 2020.

Spyros Makridakis, Evangelos Spiliotis, Vassilios Assimakopoulos, Artemios-Anargyros Semenoglou, Gary Mulder, and Konstantinos Nikolopoulos. Statistical, machine learning and deep learning forecasting methods: Comparisons and ways forward. *J. Oper. Res. Soc.*, 74(3):840–859, March 2023.

Nicholas Moroz, Isabella Moroz, and Monika Slovinec D'Angelo. Mental health services in canada: Barriers and cost-effective solutions to increase access. *Healthc. Manage. Forum*, 33(6):282–287, November 2020.

Scott B Patten, Stanley Kutcher, David Streiner, David Gratzer, Paul Kurdyak, and Lakshmi Yatham. Population mental health and COVID-19: Why do we know so little? *Can. J. Psychiatry*, 66(9): 782–784, September 2021.

Christine M Wickens, Veda Popal, Venesa Fecteau, Courtney Amoroso, Gina Stoduto, Terri Rodak, Lily Y Li, Amanda Hartford, Samantha Wells, Tara Elton-Marshall, et al. The mental health impacts of the covid-19 pandemic among individuals with depressive, anxiety, and stressor-related disorders: A scoping review. *Plos one*, 18(12):e0295496, 2023.

A APPENDIX

Table A.1: Model comparison table

Model Name	**Model Type**	**Training Paradigm**	**Accepts Covariates?**
Seasonal Naïve	Statistical	Local	✗
ARIMA	Statistical	Local	✗
ETS	Statistical	Local	✗
AutoETS	Statistical	Local	✗
DLinearModel	Machine Learning	Local	✗
DeepAR	Machine Learning	Global	✓
RecursiveTabular	Machine Learning	Global	✓
PatchTST	Machine Learning	Global	✗
TemporalFusionTransformer	Machine Learning	Global	✓

Table A.2: Themes of Google trends search terms used as covariates in forecasting models

Theme	**Terms defined by Knipe et al. (2020)**	**Additional terms to adjust to target task**
Mental distress	depression, anxiety, suicide, fear, loneliness	
Economic stressors	eviction, mortgage loan, unemployment, food bank, welfare	mortgage, ei, employment insurance
Social stressors	pharmacy, education, abuse, alcohol drink, divorce	alcohol
Treatment seeking	cognitive behavioral therapy, self-care, counselling, crisis hotline, mindfulness	cbt, self-care, helpline, therapy
Illness information	covid, sick, symptoms, cough, fever, flu, rsv	

Table A.3: Test and training window dates

Test Window	**Test Start Date**	**Test End Date**	**Context Start Date**	**Context End Date**
Test Window 1	2020-03-01	2020-08-30	2019-03-03	2020-02-23
Test Window 2	2020-09-27	2021-03-28	2019-09-29	2020-09-20
Test Window 3	2021-03-14	2021-09-12	2020-03-15	2021-03-07
Test Window 4	2021-09-05	2022-03-06	2020-09-06	2021-08-29
Test Window 5	2021-12-12	2022-06-12	2020-12-13	2021-12-05
Test Window 6	2022-04-03	2022-10-02	2021-04-04	2022-03-27
Test Window 7	2022-07-31	2023-01-29	2021-08-01	2022-07-24