

CE-merkityn lääketieteellisen ohjelmistosensorin vaikutus COVID-19 pandemian ennustamisessa: rekisteritutkimus käyttäen koneoppimismenetelmiä

Leevi Limingoja

Lääketieteen kandidaatti

Kansanterveystieteen osasto

Helsinki 11.2.2022

Tutkielma

leevi.limingoja@helsinki.fi

Ohjaaja: Tea Lallukka

HELSINGIN YLIOPISTO

Lääketieteellinen tiedekunta

Tiivistelmä

Johdanto: Ottaaksemme kantaa tämänhetkiseen COVID-19 pandemiaan, sekä muihin tuleviin pandemioihin, tarvitsemme robustia, reaaliaikaista ja populaatiotasosta datan keräystä ja analysointia. Nopea ja kattava populaatiotason tieto raportoitujen oireiden trendeistä tarjoaa aikaisemman katsauksen viruksen leviämisen etenemisessä ja auttaa ennustamaan terveydenhuollon tarvetta.

Tavoite: Tutkimuksen tavoitteena oli käyttää CE-merkattua verkkopohjaista oirearviokyselyä, ©Omaoloa, ja validoida sen data kansallisen COVID-19 liittyviä potilaskontakteja vasten ja ennustaa pandemian etenemistä Suomessa.

Menetelmät: Datamme koostui reaaliaikaisista ©Omaolon oirearviokyselyistä (414 477 kappaletta), sekä Terveyden ja hyvinvoinnin laitoksen hoitoilmoitusrekisteritiedoista, joka sisältää COVID-19 liittyvät päivittäiset hoitokontaktit. Data on 16.3.2020 – 15.6.2020 kattaen Suomen ensimmäisen pandemia-aallon. Oirearviokyselyt antavat käyttäjälleen lääketieteellisesti pätevän algoritmin pohjalta todennäköisyyden COVID-19 taudista ja tarjoaa porrastetun ohjeistuksen siitä, kuinka toimia jatkossa. Koulutimme lineaariregressio ja XGBoost malleja yhdessä piirre-esivalintamenetelmien (F-pisteitys ja keskinäinen informaatio) kanssa ennustamaan hoitokontakteja aina viikko kerrallaan yhden viikon päähän.

Tulokset: Mallimme ylsivät 24 % - 36.4 % prosentuaaliseen keskiarvovirheeseen (MAPE), kun ennustettiin kansallisesti päivittäisiä potilaskontakteja. Paras tulos saavutettiin yhdistämällä ©Omaolon ja aikaisempien hoitoilmoitusjärjestelmän tiedot yhteen. Paras ennuste saavutettiin käyttämällä lineaariregressiota ja keskinäistä informaatiota piirre-esivalintamenetelmänä.

Johtopäätökset: Tarkkoja ennustuksia COVID-19 liittyvistä hoitokontakteista on mahdollista tehdä hyödyntämällä sekä ©Omaolon oirearviokyselyjä, että kansallisia hoitoilmoitusrekisteritietoja. Täten voidaan todeta, oirearviokyselyjä on mahdollista käyttää pandemian ennustamiseen, ja siten terveydenhuollon kuormituksen ennustamiseen myös tulevissa pandemioissa.

Avainsanat: COVID-19, koneoppiminen, regressioanalyysi, pandemiat

Impact of a CE-Marked Medical Software Sensor on COVID-19 Pandemic Progression Prediction: a Register Study Using Machine Learning Methods

Leevi Limingoja¹, BM; Kari Antila², PhD; Vesa Jormanainen^{1,3}, MD, MSc; Joel Röntynen⁴, PhD; Vilma Jägerroos⁴, MSc; Leena Soinen⁵, MD, PhD, eMBA; Hanna Nordlund⁶, CSC; Kristian Vepsäläinen^{3,7}, MSc; Risto Kaikkonen⁴, MSoc; Tea Lallukka¹, PhD

¹Department of Public Health, University of Helsinki, Helsinki, Finland

²Solita Oy, Tampere, Finland

³Finnish Institute for Health and Welfare, Helsinki, Finland

⁴Solita Oy, Helsinki, Finland

⁵Digifinland Oy, Helsinki, Finland

⁶IT Centre for Science, Helsinki, Finland

⁷University of Eastern Finland (UEF), Kuopio, Finland

Corresponding Author:

Leevi Limingoja, BM

Department of Public Health

University of Helsinki

P.O. Box 20 (Tukholmankatu 8 B), FIN-00014 University of Helsinki, Finland

Helsinki,

Finland

Phone: + 358 400344644

Email: leevi.limingoja@helsinki.fi

Abstract

Background: To address the current COVID-19 and any future pandemic, we need a robust, real-time, and population-scale collection and analysis of data. Rapid and comprehensive knowledge on the trends in reported symptoms in populations provides an earlier window into the progression of the viral spread and helps to predict the needs and timing of professional healthcare.

Objective: The objective of this study was to use a CE-marked medical online symptom checker service, ©Omaolo, and validate the data against the national demand for COVID-19-related care to predict the pandemic progression in Finland.

Methods: Our data comprised real-time ©Omaolo COVID-19 symptom checker responses (414,477 in total) and daily admission counts in nationwide inpatient and outpatient registers provided by the Finnish Institute for Health and Welfare (THL) from March 16th to June 15th, 2020 (the first wave of the pandemic in Finland). The symptom checker responses provide self-triage information input to a medically qualified algorithm that produces a personalised probability of having COVID-19, and provides graded recommendations for further actions. We trained linear regression and XGBoost models together with F-score and mutual information feature pre-selectors to predict the admissions once a week, one week in advance.

Results: Our models reached a MAPE (mean absolute percentage error) between 24.2% and 36.4% in predicting the national daily patient admissions. The best result was achieved by combining both ©Omaolo and historical patient admission counts. Our best predictor was linear regression with mutual information as the feature pre-selector.

Conclusions: Accurate short-term predictions of COVID-19 patient admissions can be made, and both the symptom check questionnaires and the daily admissions data contribute to the accuracy of the predictions. Thus, symptom checkers can be used to estimate the progression of the pandemic, which can be considered when predicting the healthcare burden in a future pandemic.

KEYWORDS

healthcare; health technology assessment; machine learning; COVID-19; COVID-19 forecasting; pandemic; health technology; digital health; Online symptom checker; health data; admission data; viral spread

1 INTRODUCTION	1
1.1 OBJECTIVES	2
2 METHODS	2
2.1 DATA	2
2.1.1 ©Omaolo	2
2.1.2 National registers Hilmo, Avohilmo and Paavo.....	3
2.2 PREDICTING THE DAILY USE OF HEALTHCARE RESOURCES	3
2.3 PROFILING THE MOTIVES OF THE ©OMAOLO COVID-19 SYMPTOM CHECK USERS..	4
3 RESULTS	5
3.1 PREDICTING THE DAILY USE OF HEALTHCARE RESOURCES	5
3.2 PROFILING THE MOTIVES OF THE ©OMAOLO COVID-19 SYMPTOM CHECK USERS	11
4 DISCUSSION.....	12
4.1 PRINCIPAL RESULTS	12
4.2 PROFILING THE MOTIVES OF THE ©OMAOLO COVID-19 SYMPTOM CHECK USERS	13
4.3 CONSIDERATION OF OTHER SOURCES	13
4.4 STRENGTHS AND LIMITATIONS.....	14
4.5 CONCLUSIONS.....	17
ACKNOWLEDGMENTS	17
CONFLICTS OF INTEREST	17
ABBREVIATIONS	17
REFERENCES	18
MULTIMEDIA APPENDIX 1 : ©OMAOLO QUESTIONNAIRE.....	21

1 Introduction

The rapid spread of the SARS-CoV-2 virus leading to a pandemic presented challenges for nationwide assessment of the progression of the COVID-19 pandemic [1]. The virus was first discovered in Wuhan, China in December 2019 and COVID-19 was declared a pandemic by the World Health Organisation in March 2020 [2-4]. In Finland, cases started to appear in late February 2020, and the Finnish government announced a national lockdown in mid-March 2020 to slow the viral spread and to protect risk groups [1].

Digital health technology tools such as symptom checkers have been used in different countries (e.g., Finland, France, Israel, Italy, the Netherlands, UK, and USA) as self-triage tools for possible SARS-CoV-2 infections [5-11]. In Finland, a COVID-19 symptom checker was added to a pre-existing national, CE-marked medical symptom checker service, ©Omaolo [5,12]. The web-based symptom checker provides the user with advice on further actions based on a medically approved algorithm. There have been studies about how well symptom checkers perform as clinical tools [13], but their data's potential on predicting epidemic progression, to our knowledge, has not yet been studied. Having real-time comprehensive data on reported symptom trends could provide an earlier window into the viral spread and help predict the burden of professional healthcare.

To study if the data collected by the ©Omaolo service and the national care notification registers could be used to predict pandemic progression in Finland, we used the methods of machine learning. Perhaps the best potential of machine learning over more traditional methods lies in its ability to better adapt to the data and, thus, to the evolution of the underlying phenomenon. With large datasets like the ©Omaolo COVID-19 symptom checker responses, machine learning may also uncover more complex associations between the factors contributing to the predicted outcomes. Machine learning models can also be trained and re-trained along the way to reveal how the significance of the individual input variables for making the predictions will change over time and to become a more accurate predictor as more data is collected.

1.1 Objectives

The study objective was to assess if a nationwide symptom checker can be used as a predictive tool in estimating the national progression of the COVID-19 pandemic and healthcare admissions by utilising machine learning methods.

2 Methods

2.1 Data

2.1.1 ©Omaolo

The COVID-19 epidemic in Finland started in mid-March 2020. On March 16, 2020, the Finnish Government announced a state of emergency due to the COVID-19 epidemic and consequently implemented several physical distancing measures aimed at slowing the spread and protecting risk groups [1]. Part of the national response was the ©Omaolo COVID-19 web-based symptom self-assessment tool, a CE-marked medical device [5,12]. It was launched for use March 16, 2020 and was published in the two national languages (Finnish and Swedish) and later also in English. The COVID-19 symptom checker functioned as any other symptom checker in ©Omaolo and was jointly developed by DigiFinland Oy, Duodecim Publishing Company LTD, Finnish Institute for Health and Welfare (THL), Solita Oy and Mediconsult Oy.

In the symptom checker, the user answers a set of pre-defined, expert-validated questions. As a result, it returns self-triage information on how to proceed with one's situation. The progress of filling in the questionnaire from start to finish is recorded to the log files of the service. The respondent has a choice to answer anonymously without including one's personal information in the process. The questionnaire itself includes several background questions such as age, postal code, gender and reason for filling in the questionnaire, existing medical conditions, whether the respondent has had close contact with a COVID-19-positive person, whether the respondent or close contacts have been ordered to quarantine by a physician, where the respondent thinks they may have caught the virus and what kind of work the respondent does in regards to contacts with others (Multimedia Appendix 1: ©Omaolo questionnaire). The questionnaire has been updated several times during the pandemic to better coincide with the latest COVID-19 research [5].

During the study period, a total of 547,428 responses were submitted to ©Omaolo. Of these, the contents of 132,951 responses were unsaved due to technical reasons. Almost all the unsaved responses were submitted prior to March 28, 2020, when ©Omaolo was yet not configured to save the anonymous responses. A small number of the anonymous responses were not saved during short maintenance breaks throughout the period. Accounting for these losses, a total of 414,477 responses were available for analyses. Care reminders, the self-triaged recommendations for care as given by the Duodecim EBMEDS service [5], were available for all submitted responses, including the unsaved ones. The data were pseudo-anonymised prior to the analyses.

2.1.2 National registers Hilmo, Avohilmo and Paavo

We used the established, national care notification registries Hilmo and Avohilmo [14] of the Finnish Institute for Health and Welfare (THL) to estimate the demand of COVID-19-related care. These registers contain structured inpatient (Hilmo) and outpatient (Avohilmo) records from all public and private specialist care hospitals in Finland. These records were combined in the data pre-processing stage and will be called Hilmo collectively in the following sections. The data were anonymised of all identifiers before use.

As a supplementary source, we used the publicly available version of the Paavo register maintained by Statistics Finland [15]. Among other variables, Paavo contains basic demographics of Finnish citizens based on the postal code of their residence. We used these data to identify and rectify the regional bias in age distribution of the ©Omaolo users. These data were anonymous at the source.

2.2 Predicting the daily use of healthcare resources

We chose our study period to be from March 16, 2020, the release of the COVID-19 symptom check, to June 15, 2020, the approximate beginning of a period of low activity in the pandemic following the first wave. For the predictions, we used two regressors: linear regression [16] and XGBoost regression [17]. The reasoning behind selecting two regressors was to compare a simple and traditional method (linear regression) to a modern option (XGBoost regression) that has many hyperparameters that can be learned from the

data. Both methods were implemented with three feature pre-selection strategies: a human expert (KM. A.), F-score [16] and mutual information [18]. All regressor feature selector combinations were tested separately resulting in six different machine learning models.

We chose a scenario where the number of daily, COVID-19-related healthcare admissions as extracted from the Hilmo register was predicted one week ahead, every week on Wednesdays. This follows a hypothetical scenario where the resources for the following seven days would be decided mid-week (on Wednesday) to give two full days to prepare for the weekend, for example by reassessing the need for extra resources and personnel.

For training, testing and validation of the models, we used time series nested cross-validation [19]. This strategy was chosen to ensure that the model is trained and tested with samples independent from the validation set, thus, no information from the samples past the prediction point was used. During cross-validation, the set of features (and other hyperparameters in case of XGBoost), with which the regressor best generalizes its predictions to unseen data in terms of average prediction error on different validation sets, is selected for a given regressor.

We also chose to train the classifiers with Hilmo and ©Omaolo first separately, and then combined. This was to test how much, if any, the results would improve if the data from both sources were used.

2.3 Profiling the motives of the ©Omaolo COVID-19 symptom check users

The ©Omaolo COVID-19 symptom checker achieved considerable popularity immediately after its release. Tens of thousands of responses per day were submitted during the first week. The submission activity showed clear peaks during infomercials and other major media mentions. To distinguish the users that were truly suspecting a COVID-19 exposure from the users that were visiting ©Omaolo just out of curiosity, a question about this matter was added in the questionnaire in the form of a simple tick-box on March 28, 2020. After this update, it was found that approximately 40% of the

responses had the out-of-curiosity option checked. We then investigated whether it was possible to distinguish the two response profiles (out-of-curiosity or not) and which questions were the best predictors of this behaviour. We used a naive Bayes classifier [20], logistic regression [16] and XGBoost binary classifier [17]. We chose naive Bayes and logistic regression because they are widely used in medicine and elsewhere. XGBoost was included since it is based on a different approach (an ensemble of trees) and thus provides an interesting comparison to the two established methods.

All three models were tested with 5-fold random cross-validation and the sensitivity and specificity for each fold were computed and finally averaged over all folds. The size of the majority set (not out-of-curiosity) was balanced by under-sampling to the size of the minority prior to cross-validation.

All analyses, both the predictions and profiling, were conducted using Python version 3.6 [21] with the feature selectors, classifiers and regressors from Scikit-learn module version 1.0 [22].

3 Results

3.1 Predicting the daily use of healthcare resources

Women used the web-based symptom checker more often than men; about two-thirds of the filled-in forms were filled-in by women (Table 1). People of working age were also using the symptom checker more than other age groups. Most of the questionnaires were filled-in in Southern Finland, around half of them. Cough was the most common symptom followed by sore throat, fever, headache, and difficulty breathing.

Table 1. Distribution of age, employment status, symptoms, and region of the filled-in ©Omaolo questionnaires.

	Category	Total	Men	Women
Age				
	0-9	8652 (2%)	4291	4361
	10-19	30,805 (7%)	10,689	20,116
	20-29	93,511 (23%)	30,241	63,270
	30-39	98,041 (24%)	34,727	63,314
	40-49	77,869 (19%)	29,010	48,859
	50-59	59,432 (14%)	22,788	36,644
	60-69	31,935 (8%)	13,392	18,543
	70-79	12,824 (3%)	6519	6305
	80 +	2453 (1%)	1286	1167
	Total	415,522	152,943	262,579
Employment status				
	Not currently working	133,199 (33%)	47,674	85,525
	Healthcare worker	65,379 (16%)	7953	57,426
	Cannot avoid contact (service worker)	102,544 (25%)	41,948	60,596
	Can avoid contact	107,558 (26%)	51,869	55,689
	Total	408,680	149,444	259,236
Symptoms				

	Cough	140,661 (35%)	53,364	87,297
	Trouble breathing	31,143 (8%)	11,312	19,831
	Sore throat	86,985 (21%)	24,505	62,480
	Headache	35,886 (9%)	12,233	23,653
	Myalgia	10,213 (3%)	4552	5661
	Vomiting	7406 (2%)	2640	4766
	Fever	59,041 (15%)	23,842	35,199
	Loss of smell	433 (0%)	182	251
	Dysphagia	7971 (2%)	2946	5025
	Trismus	6469 (2%)	1938	4531
	Trouble speaking	6392 (2%)	2971	3421
	Other	13,952 (3%)	5589	8363
	Total	406,552	146,074	260,478
Region				
	Western central Finland (City of Tampere)	42,759 (14%)	15,370	27,389
	Western coastal Finland (City of Turku)	36,321 (12%)	13,155	23,166
	Northern Finland (City of Oulu)	27,419 (9%)	9894	17,525
	Southern Finland (City of Helsinki)	149,145 (51%)	54,447	94,698
	Eastern Finland (City of Kuopio)	39,340 (13%)	13,819	25,521
	Total	294,984	106,685	188,299

For the analyses, we chose two regressors, linear regression and XGBoost, and three feature pre-selection strategies for each and compared their performance. To predict the COVID-19-related admissions for each day, seven days ahead of the prediction point, the features given to the model were extracted from the responses and the Hilmo register on the 7 and 14 days prior (lag variables). The use of lag variables essentially means that two sets of the time-dependent features was formed. The first was delayed 7 days and the second for 14 days. This was to ensure that no data from any of the sources, Omaolo or Hilmo, is leaked past the point of prediction during feature selection, model training or testing. The regressors were first trained with 5 expert-selected features: how many of the questionnaires were filled-in by people over 60 years old, how many reported lengths of symptoms greater than 10 days, and how many were assigned an urgency code P1 (the most urgent) in the care recommendations, in addition to the number of COVID-19-related admissions. The feature pre-selectors F-score and mutual information were added later.

The predictions made by both models for data gathered between March 16, 2020 and June 15, 2020 were compared to the true admission count (Figure 1). The first four weeks were reserved for training the models before predictions were made in the week starting on April 16, 2020. For the consecutive weeks, the models were retrained with the data from the previous weeks. Both models make predictions with a similar error (linear regression: mean absolute error (MAE) 138, mean absolute percentage error (MAPE) 31; XGBoost regression: MAE 135, MAPE 31). The predictions become more accurate towards the end of the study period, since the models have more data to learn from, which reduces the sampling bias. May 22, 2020 was a public holiday in Finland, explaining a similar drop in the true admission count as on weekends.

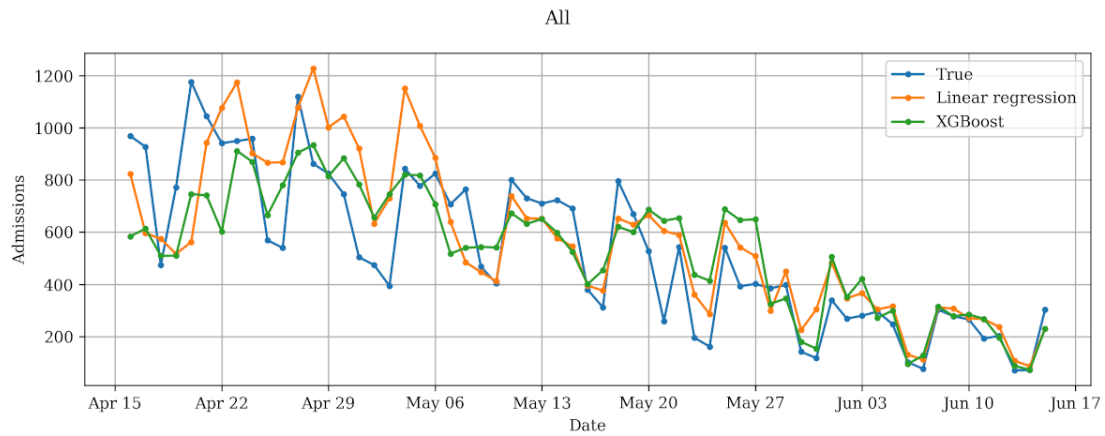


Figure 1. COVID-19-related admissions predicted by linear regression and XGBoost regression models, together with the true admission count during the first wave of the pandemic in 2020.

To check that the ©Omaolo questionnaire data are relevant for the predictions, we compared the error of both models with and without the questionnaire data, and only with the questionnaire data (Table 2). Both models achieved the lowest error when the registry and the questionnaire data were combined, indicating that the questionnaire data are relevant for making accurate predictions compared to the registry data alone.

Table 2. Comparison of the effect of ©Omaolo and Hilmo data on the error of the models on expert-selected features

	Features	Mean absolute error (MAE)	Mean absolute percentage error (MAPE)
Linear regression			
	©Omaolo + Hilmo	137.79	31.33
	©Omaolo	175.38	44.60
	Hilmo	184.37	34.98
XGBoost regression			
	©Omaolo + Hilmo	139.11	31.78
	©Omaolo	165.18	46.94
	Hilmo	178.92	46.63

In addition to expert-selected features, we tried automated feature selection methods that select the top eight features based on the F-score or mutual information of the feature with the predicted variable (number of admissions on a given day) (Table 3). The F-score and mutual information are measures of dependence between the feature values and the values of the predicted variable in the historical data. Different feature selection strategies worked better for the two models: linear regression was the most accurate with the mutual information criterion, whereas XGBoost with the expert-selected features.

Table 3. Comparison the effect of different feature selection methods on the error of the models

	Feature selection	Mean absolute error (MAE)	Mean absolute percentage error (MAPE)
Linear regression			
	Expert-selected	137.79	31.33
	F-score	141.40	30.46
	Mutual information	112.16	24.23
XGBoost regression			
	Expert-selected	139.11	31.78
	F-score	150.07	36.40
	Mutual information	146.12	33.33

Older age groups, who are more likely to have a severe form of COVID-19 and hence be admitted to hospital, are under-represented in the ©Omaolo questionnaire data. This

could affect the performance of the models, as it is more difficult for the models to learn from imbalanced training data. To assess the problem, oversampling of the under-represented age groups was conducted to see if it would decrease the error of the models (Table 4). Resampling gave linear regression model slightly smaller error, whereas the XGBoost regressor performed worse with resampling. The results indicate that it is not essential when minimizing the prediction error to oversample the questionnaire answers of the under-represented age groups to match the age distribution of the population.

Table 4. The effect of oversampling on the error of the models

	Resampling	Feature selection	Mean absolute error (MAE)	Mean absolute percentage error (MAPE)
Linear regression				
	No resampling	Expert-selected	137.79	31.33
	Oversampling	Expert-selected	131.13	30.47
XGBoost regression				
	No resampling	Expert-selected	139.11	31.78
	Oversampling	Expert-selected	153.35	35.22

3.2 Profiling the motives of the @Omaolo COVID-19 symptom check users

The three different models produced the following sensitivity and specificity for detecting users that answered using the service out of curiosity: the naive Bayes classifier with

sensitivity of 0.622 and specificity of 0.367, logistic regression of 0.665 and 0.332 and XGBoost binary classifier with 0.607 and 0.388, respectively. These results were acquired by maximising the number of correct classifications.

4 Discussion

4.1 Principal results

In this study, we examined whether it is possible to predict the national epidemic progression and the burden of healthcare using machine learning methods with real-life data on symptoms and usage of healthcare. The main finding of this study is that it is possible to predict national healthcare admissions related to COVID-19 using a symptom checker combined with register data by using machine learning methods with considerable accuracy (small MAPE error). These methods were tested in a scenario where the predictions were made one week ahead, once per week. The best model was achieved using the symptom checker data combined with register data (MAPE 24.23%). This result was reached by using linear regression with mutual information as the feature pre-selector. All tested models and combinations of feature pre-selectors and models were able to produce predictions that followed the true epidemic progression (Figure 1). Overall, linear regression was better than XGBoost, though only marginally. This suggests that in our research scenario there was no benefit in using a model that has many trainable hyperparameters (XGBoost regression) over a simple one (linear regression). All tested models seemed to improve towards the end of our study period as more data were available for training. Additionally, the differences in accuracy between the models are more visible at the start of the period and seem to diminish towards the end. Based on the results, the F-score and especially mutual information appear to improve the results for linear regression. Feature pre-selection may improve the predictions by, for example, reducing the risk of overfitting. This is relevant in our dataset since the feature set used in the classification was rather large and likely suffers from multicollinearity. Using the pre-selectors did not improve XGBoost regression. This suggests that we were not able to find a suitable pre-selection strategy for the method.

Finally, predictions that can follow the progression can be made using either ©Omaolo symptom checks or historic Hilmo counts separately. However, the best results were

achieved by combining both. Adding ©Omaolo to Hilmo counts reduces the MAPE of linear regression from 34.98% to 31.33% and XGBoost from 46.63% to 31.78%. The results suggest that ©Omaolo contains information of the pandemic progression that is not present in Hilmo alone.

Oversampling the data to balance the regional differences between the ©Omaolo users and general population seems to produce conflicting results: marginal gain with linear regression and but a loss of accuracy with XGBoost. Using oversampling leads to added complexity in the analysis pipeline, and without a clear benefit its use is hard to justify.

4.2 Profiling the motives of the ©Omaolo COVID-19 symptom check users

The answer profiles of users using the COVID-19 symptom check out of curiosity were very similar to the rest, and no reliable classification between the groups could be made by any of the tested models. Neither of the groups reported longer or more serious symptoms over the other. The only striking difference between the groups was that there appeared to be more out-of-curiosity responses during high service utilisation, for example after TV infomercials.

4.3 Consideration of other sources

We also considered using the daily number of phone calls received at the 116117 Medical Helpline service. The 116117 Medical Helpline provides professional assistance on healthcare-related topics in urgent, but non-emergency cases to over 4 million Finns in extended business hours. However, since it was not possible to extract the topic of the call, whether they were COVID-19-related or not, and since the calls could be localised with much poorer resolution than with the rest of the sources, this dataset was eventually dropped from the analyses.

The Google Trends [23] of popular COVID-19-related search terms was another potential data source considered for analyses. However, it was found that the publicly available Finnish trends only covered major cities. Additionally, the overlap of trending search terms between cities was found to be small, making the data very sparse. For these reasons, this dataset was not used in the analyses.

4.4 Strengths and limitations

This study had several potential limitations. A considerable number of responses (132,951) were not saved during the study period. Nearly all of these were submitted during the first two weeks of the study, complicating the analysis for the first month. This may have contributed to the relatively poor prediction accuracy for the related weeks by delaying the convergence of the regressors to the true admission count.

The true admission count showed a strong diminishing trend towards the end of our study period. During the last weeks of the period, there were days when only a few dozen new admissions were recorded nationally. Because our error metric MAPE is relative to the true values, them being small will result in error values that appear high even though the absolute error between true and predicted counts remains low. Despite this, we decided to use MAPE for its intuitiveness, wide use, and easy comparability of the error between the days, weeks, and methods.

At the beginning, the survey did not include an item about the motive to fill in the survey, i.e., whether it was due to actual symptoms or out of curiosity. This adds some additional forms to our data that do not reflect the situation at hand. The proportion of responses filled-in out of curiosity remained remarkably stable at around 40% throughout the study period. Moreover, the results of trying to separate the responses filled-in out of curiosity from the rest with binary classifiers (naive Bayes, logistic regression, XGBoost classifier) failed to reveal any meaningful differences between the answer profiles of these groups. Thus, we did not find a justification to remove these responses from the analyses or handle them differently than the rest of responses (not filled-in out of curiosity).

The data available for this study contained COVID-19-related admissions data with a steady downward trend, and it would have been interesting to see if the models could predict a reversal of the trend before it occurs. However, there is weekly variability in the admissions and the models learned this pattern well. An interesting deviation from the weekly pattern was Ascension Day on Thursday May 21, which was a public holiday in Finland. The models did not have enough training data on admissions on public holidays

on a weekday to predict a similar dip in the admissions as on weekends, but longer datasets could allow the models to learn this pattern as well.

Much of the prediction errors took place during days that showed sharp peaks of increased or decreased demand that were not immediately explainable with the data available. Some of these errors may be due to technical reasons, for example a major care provider suffering an error on one day and reporting higher counts on the following day. Naturally, these kinds of special events cannot be learned from the admission count data alone. On a positive note, the developed models appear robust and thus not susceptible to these kinds of anomalies.

One could also question if biological tests make other monitoring redundant. While in many countries, biological tests are done to follow the pandemic, it is important to note that the online symptom checking does not replace the need for biological testing but provides a different and valuable perspective. In many countries, including Finland, the testing services are saturated, and only some population groups get tested, making our picture of the pandemic progression biased. The biological testing further comes with an immense cost, particularly with an exponential rise in cases.

The symptom checker, in turn, can be filled in anytime and by large numbers. Furthermore, it can be continuously updated to include the most relevant questions. It is possible that health care burden is going to change even quite rapidly, and we can see that using ©Omaolo surveys even without an indication from the time series data. Additionally, the machine learning models are trained every week, allowing adaptability to changes in the statistical relationship between the predictors and the predicted variable as the pandemic progresses.

Furthermore, while our data only cover the first wave of the pandemic, it remains, however, important in the aspect of the early decision making against a new threat, and overcoming a challenge on modelling a novel phenomenon from the start with no history. Additionally, we do not only validate the data, but our approach is on the prediction and modeling of the pandemic. Emergence of vaccinations, new variants, new policies and restrictions will all affect the progression of the pandemic. These changes have also affected ©Omaolo that has been continuously updated, e.g., including questions about

vaccination status. Our prediction models are also upgradeable continuously, thus the drift in data and concept can be mitigated on the go and features that will drop or rise in importance can be monitored along with the actual result of the prediction.

Finally, the lack of reported anosmia could suggest a lack of specificity. In our data we see that anosmia is a rarely reported symptom. It is also a known problem that without a specific test COVID-19 is notoriously hard to distinguish from a common flu. Despite of the relatively small number of reported anosmia cases, we have other features that have been shown to be highly significant in predicting ICD-10 and ICPC-2 coded admissions. These include features such as how many of the questionnaires were filled-in by people over 60 years old, how many reported lengths of symptoms greater than 10 days, and how many were assigned an urgency code P1 (the most urgent) in the care recommendations. Alternatively, if anosmia is important for specificity, it is likely that we have more false positives and our final models are less accurate. In other words, with the key symptoms, the precision of the model is improved and the error is smaller. The consequences of such a lack depends on how the results are used. If the burden of healthcare were solely estimated based on these models, we might underestimate the need for health care. However, it is unlikely that the model would be used to define any absolute health care need, whereas it can provide an indication to prepare for an increased health care burden.

The study also has several strengths. We had access to a nationwide online symptom checker data source, ©Omaolo, which is a CE-marked medical device complying with the Medical Device Directive, used by healthcare and social service professionals (Instructions: [24]). We also had data on all hospital admission records from public or private hospitals in Finland on a weekly basis. Such data are rarely available anywhere in the world and provide unique opportunities to produce new information about the possibilities of using such real-life data in predicting subsequent healthcare burden. Similar symptom checkers could be adopted for use in many other countries, and they could provide an opportunity to collect data on symptom development very rapidly, and at a relatively low cost at a national level. These symptom checkers and the findings are not restricted or solely applicable to the current pandemic or its first wave, but could be applied in any other future epidemic or pandemic – or for collection of other types of symptoms as well.

In addition, a clear advantage of machine learning methods is that both the model selection and fit are automatized. This means that the prediction method adapts in any new data and attempts to make as accurate predictions as possible, with the data in use at any given time. In case the phenomenon in question changes notably, changes in data use and sources and machine learning are also required.”

4.5 Conclusions

Our study shows that COVID-19-related healthcare admissions in the short-term can be predicted with considerable accuracy using symptom checker data combined with register data using machine learning methods. This kind of approach could help healthcare providers better assess the burden of the healthcare system in advance, which would make resource allocation more predictable. Furthermore, we think that this kind of approach could also be implemented in different stages of the pandemic and in future pandemics as well.

Acknowledgments

ITEA / Business Finland grant ITEA-2019-18022-IVVES

Business Finland grant 32258/31/2020/AHMED

Conflicts of interest

None declared

Abbreviations

COVID-19: Coronavirus disease 19

CE: Conformité européenne

THL: Finnish Institute for Health and Welfare

XGBoost: Extreme gradient boosting

SARS-CoV-2: severe acute respiratory syndrome coronavirus two

MAE: Mean absolute error

MAPE: Mean absolute percentage error

References

1. Tiirinki H, Tynkkynen L, Sovala M, et al. COVID-19 pandemic in Finland – Preliminary analysis on health system response and economic consequences. *Health Policy and Technology* 2020; 9(4):649-662. PMID:32874860, doi: 10.1016/j.hlpt.2020.08.005
2. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet* 2020; 395(10229):1054-1062. PMID:32171076, doi: 10.1016/S0140-6736(20)30566-3
3. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 2020; 382(8):727-733. PMID:31978945, doi: 10.1056/NEJMoa2001017
4. World Health Organisation. Rolling updates on coronavirus disease (COVID-19) <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen> [accessed 24-10-2021]
5. Jormanainen V, Kaikkonen R, Isosomppi S, et al. The Finns quickly found the coronavirus symptom assessment online service: those who responded to the Omaolo coronavirus symptom assessment from 16 March to 15 May, 2020. Finnish Institute for Health and Welfare (THL) 2020, ISBN 978-952-343-523-0 (online publication). (Article in Finnish, abstract in English).
6. Judson TJ, Odisho AY, Neinstein AB, et al. Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for COVID-19. *J Am Med Inform Assoc* 2020 Jun 1; 27(6):860-866. PMID:32267928, doi: 10.1093/jamia/ocaa051
7. Galmiche S, Rahbe E, Fontanet A, et al. Implementation of a Self-Triage Web Application for Suspected COVID-19 and Its Impact on Emergency Call Centers: Observational Study. *J Med Internet Res* 2020; 22(11):e22924. PMID:33147165, doi: 10.2196/22924
8. Timmers T, Janssen L, Stohr J, et al. Using eHealth to Support COVID-19 Education, Self-Assessment, and Symptom Monitoring in the Netherlands: Observational Study. *JMIR Mhealth Uhealth* 2020; 8(6):e19822. PMID:32516750, doi: 10.2196/19822
9. Bastiani L, Fortunato L, Pieroni S, et al. Rapid COVID-19 Screening Based on Self-Reported Symptoms: Psychometric Assessment and Validation of the EPICOV19 Short Diagnostic Scale. *J Med Internet Res* 2021; 23(1):e23897. PMID:33320825, doi: 10.2196/23897

10. Rossman H, Keshet A, Shilo S, et al. A framework for identifying regional outbreak and spread of COVID-19 from one-minute population-wide surveys. *Nat Med* 2020; 26(5):634-638. PMID:32273611, doi: 10.1038/s41591-020-0857-9
11. Spector T, Wolf J, Willett WC, et al. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science* 2020; 368(6497):1362-1367. PMID:32371477, doi: 10.1126/science.abc0473
12. Jormanainen V, Soininen L Use and Users of the Web-Based Omaolo Covid-19 Symptom Self-Assesment Tool in Finland Since March 16, 2020 2021; 281 PMID:34042674, doi: 10.3233/SHTI210270
13. Budd J, Miller BS, Manning EM, et al. Digital technologies in the public-health response to COVID-19. *Nat Med* 2020; 26(8):1183-1192. PMID:32770165, doi: 10.1038/s41591-020-1011-4
14. Sund R Quality of the Finnish Hospital Discharge Register: A systematic review. *Scand J Public Health* 2012; 40(6):505-515. PMID:22899561, doi: 10.1177/1403494812456637
15. Statistics Finland. Paavo database https://www.stat.fi/tup/paavo/index_en.html [accessed 30-10-2021]
16. James Gareth, Witten Daniela, Hastie Trevor, et al. , *An Introduction to Statistical Learning with Applications in R*. New York: Springer; 2013
17. Chen T, Guestrin C. , *XGBoost: A Scalable Tree Boosting System*. New York, NY, USA: Association for Computing Machinery; 2016
18. Kraskov A, Stögbauer H, Grassberger P Estimating mutual information. *Phys Rev E* 2004; 69(6):066138. doi: 10.1103/PhysRevE.69.066138
19. Courtney Cochrane. Time series nested cross validation <https://towardsdatascience.com/time-series-nested-cross-validation-76adba623eb9> [accessed 31-10-2021]
20. McCallum Andrew, Nigam Kamal A comparison of event models for Naive Bayes text classification. *AAAI* 1998
21. Python <https://www.python.org> [accessed 31-10-2021]

22. Scikit-learn <https://scikit-learn.org/stable/> [accessed 31-10-2021]
23. Google. Google trends <http://www.google.com/trends/> [accessed 30-10-2021]
24. Omaolo Instructions <https://www.omaolo.fi/kayttoohjeet/omaolo-instructions.pdf> [accessed 31-10-2021]

Multimedia Appendix 1 : ©Omaolo questionnaire

©Omaolo symptom check questions July 2020

- I just want to try or test the symptom checker

- What is your postcode (zip code)?

- Your age (in years)

- Gender (answer required for further questions)
 - Male
 - Female

- Are you pregnant?
 - Yes
 - No
 - Not sure

- Why are you filling out this form?
 - I have symptoms and want to know if they are caused by the coronavirus.
 - What symptoms do you have?
 - Cough
 - Difficulty of breathing
 - Sore throat
 - Runny nose or stuffiness (does not mean chronic or allergic symptoms)
 - Headache
 - Muscle pain
 - Diarrhea
 - Vomiting (throwing up)
 - I have lost my sense of smell or taste
 - Other symptoms
 - How many days have your current symptoms lasted? Give the answer as a whole number (for example 4)

- Do you have fever? (Above 37,5 degrees Celsius?)
 - Yes
 - What has been the highest fever reading you have had in the last 24 hours? Give the answer with a decimal point (for example 37,5 degrees Celsius).
 - No
- What is your general condition?
 - I am well enough to be out of bed and can complete regular household tasks
 - I am well enough to be out of bed, but I can only complete the most essential household tasks
 - I feel very weak and can only lie in bed (at most I can visit the bathroom)
- What is your weight (kg)?
- What is your height (cm)?
- Do you smoke on a daily basis?
- Do you have factors that increase your risk of thrombosis (blood clot)? You can select more than one option.
 - A history of deep venous thrombosis (blood clot) or pulmonary embolism
 - Increased hereditary risk for thrombosis (blood clot) that has been verified by a blood test.
 - Active cancer with ongoing treatment
 - Blood or bone marrow disease such as myeloma, polycythemia vera, thrombocytosis, PNH
 - Severe inflammatory disease such as rheumatoid disease, inflammatory bowel disease or vasculitis
 - Major surgical operation, such as knee or hip replacement within 2 months
 - Cast in the lower extremity
 - Delivery (childbirth) within 6 weeks
 - Poorly controlled diabetes
 - Severe renal insufficiency
 - None of the above
- Please answer the following questions to identify serious symptoms not relating to coronavirus
 - Do you have difficulty of swallowing food or drink because it's so painful?
 - Yes
 - No
 - Do you have difficulty of opening your mouth (lockjaw)?
 - Yes
 - No
 - Have your speech become difficult for others to understand?
 - Yes
 - No
- I do not have any symptoms but would like to know if I have caught coronavirus.

- Yes
 - No

- If medical professional recommend that you should be examined, how would you be able to travel?
 - I would be able to travel alone or with my family member living with me (my own car or other transportation) and so that I would not infect others.
 - I would have to use public transportation.
 - I would have to use a taxi.
 - I am so unwell or it is so difficult for me to walk that I would need an ambulance, medical taxi or a home visit by a nurse.