



Machine learning model in predicting the minimal endoscopy services requirement for gastrointestinal cancer diagnosis in Hong Kong during various phases of COVID-19

Thomas KL Lui¹, Wai K Leung¹

¹Department of Medicine, University of Hong Kong, Hong Kong, China

Introduction

- The ongoing COVID-19 outbreak has severely affected the health care service delivery, including routine endoscopy services.
- We have previously shown that there were 37-46% drop in gastrointestinal cancers diagnosed during the initial phase of COVID-19 in Hong Kong.
- As the situation would unlikely to resolve soon, we have to develop a new strategy to monitor and predict our endoscopic service requirement, in order to provide a robust guidance on optimal endoscopy volume to minimize the delay in gastrointestinal cancers diagnosis.

Aims

- To build a machine learning model to forecast the optimal number of upper and lower endoscopy volume in the next 4 weeks in order to avoid delay in diagnosis of gastrointestinal cancers

Methodology

- CDARS of the Hospital Authority: the number of patients who had upper and lower endoscopy and the number of new cases of gastrointestinal cancers diagnosed
- The School of Public Health dashboard: the real time reproductive number of local COVID-19 case and the number of new COVID-19 cases.
- The number of patients with potential delay in cancer diagnosis was estimated with the ARIMA model.
- Four different phases of COVID-19 in the preliminary training data was previously defined as follow. Phase 1: preparedness and imported infection from mainland China (31 December 2019 to 3 February 2020); Phase 2: small scale local transmission (4 February to 3 March 2020); Phase 3: imported infection from overseas countries associated with local transmission (4 March to 19 April 2020); Phase four—controlled imported infection with limited local transmission (20 April to 30 June 2020)
- Six different machine learning models: traditional linear regression (LR), random forest (RF), support vector machine (SVM), stochastic gradient boosting (SGB), neural network (NN) and extreme gradient boosting (XGBoost) were used to estimate the requirement of endoscopy service for gastrointestinal cancer diagnosis during different phases of COVID-19 in Hong Kong. (Figure 3)
- Our target outcome is to predict the minimal number of upper and lower endoscopy to be performed per week in the coming 4-week without delay in cancer diagnosis
- Variables used to build machine learning models will include:
 - One-week moving average of new COVID-19 cases,
 - Mean number of new COVID-19 cases,
 - Number of weeks from the first COVID-19 case on a weekly basis,
 - Weekly average real-time effective reproductive number for local case (Rt),
 - Number of new patients with diagnosis of gastric or colorectal cancer (per week)
- The data set will be separated into training and validation set in a proportion of 90% and 10%. The mean absolute percentage error (MAPE) of each model will be calculated and compared

Figure 1. Upper endoscopy volume and percentage reduction in gastric cancer diagnosis during different phases of COVID-19 in Hong Kong as compared to historical cohort 2017-2019

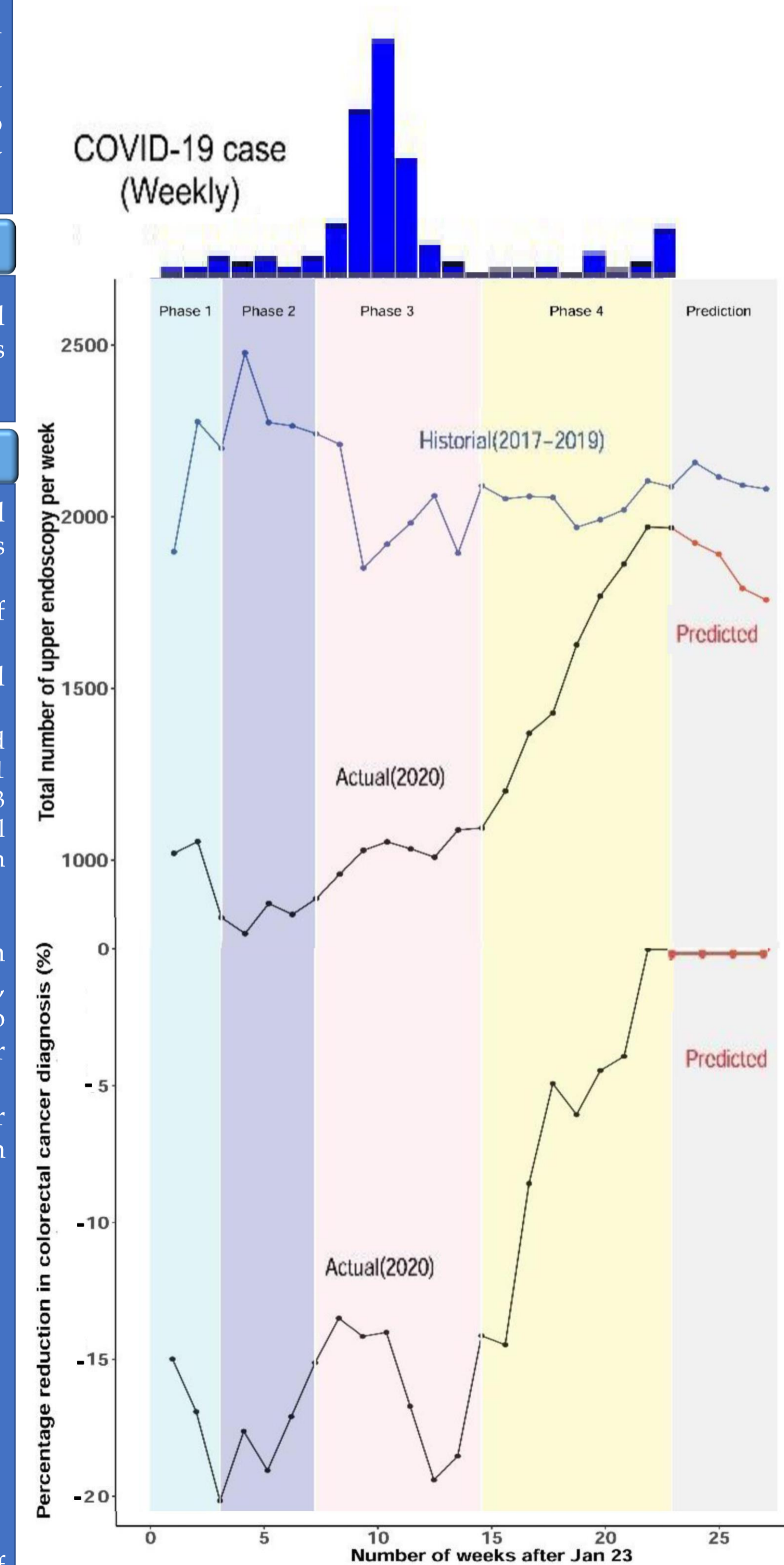


Figure 2. Lower endoscopy volume and percentage reduction in colorectal cancer diagnosis during different phases of COVID-19 in Hong Kong as compared to historical cohort 2017-2019

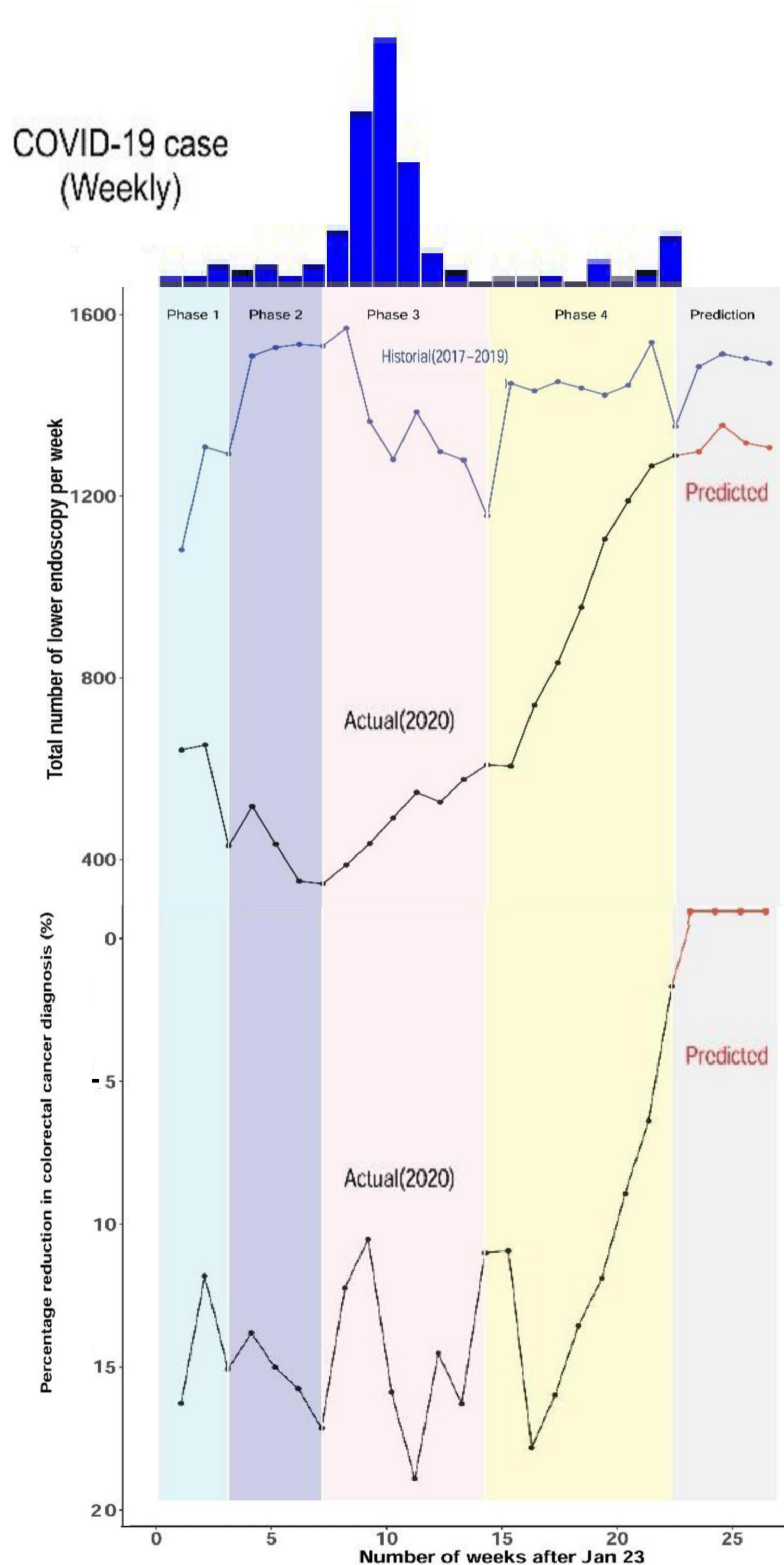
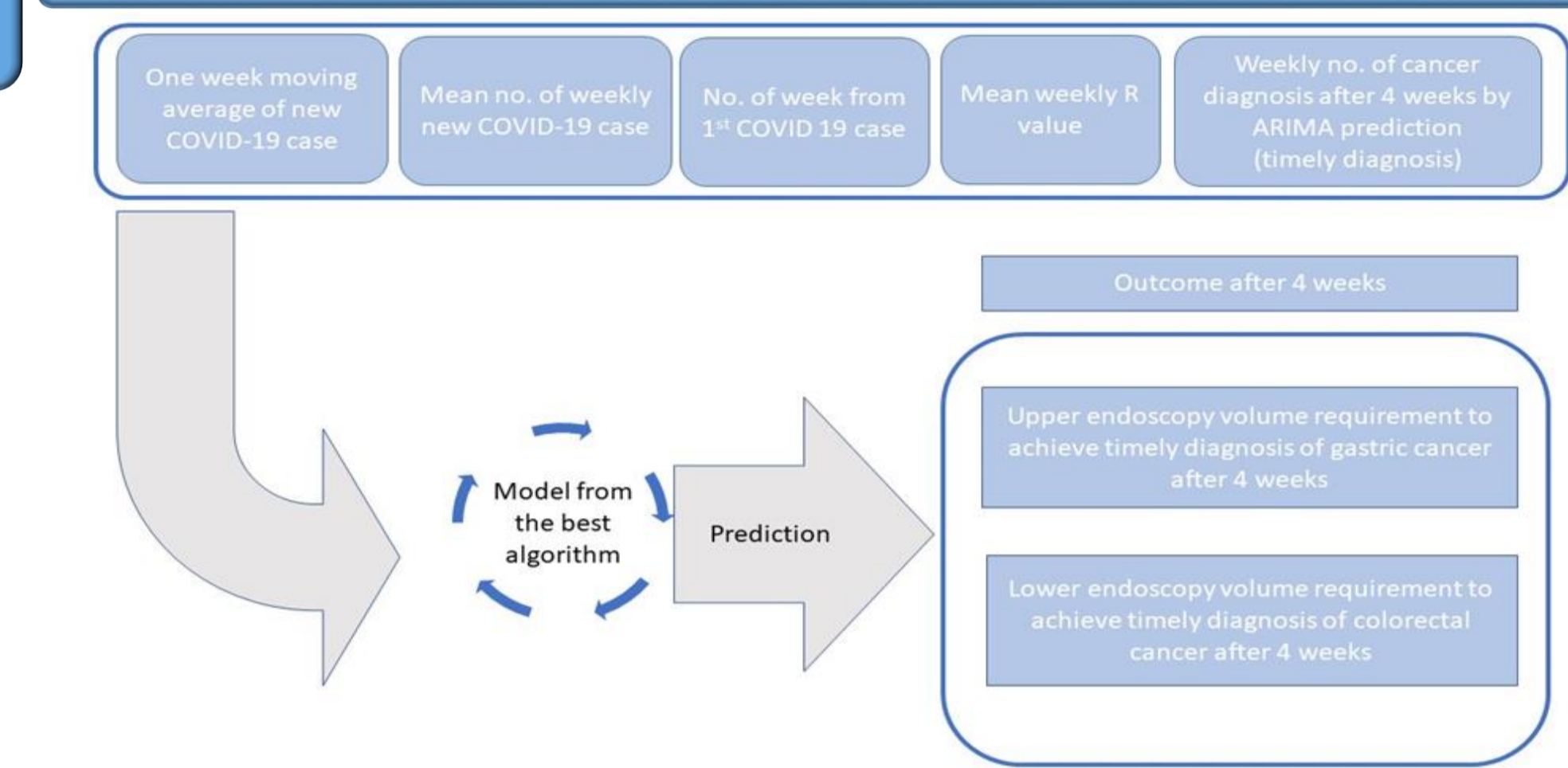


Figure 3: Prediction of outcomes by the machine learning models



Results

- There were a total of 337,903 upper endoscopies and 254,588 lower endoscopies performed between Oct 2016 and June 2020. (Figure 1 and 2)
- The model with the best performance in terms of prediction of minimal upper and lower endoscopy volume without delay in diagnosis of gastric and colorectal cancer is the XGBoost (MAPE±95%CI: 3.0±2.6 and 2.4±0.01, respectively).
- The minimal average weekly number of upper endoscopy to be performed in the subsequent month is 1781±149 (mean ±95%CI), which is still significantly lower than usual upper endoscopy volume in the pre-COVID19 period (1974±818, p<0.01).
- Accordingly, the minimal average weekly number of lower endoscopy required is 1115±386 as compared to usual lower endoscopy volume of 1330±456 (p<0.01).

Table 1: Mean absolute percentage error of various machine learning models in the prediction of the minimal weekly upper and lower endoscopy in the subsequent 4-week to achieve a timely diagnosis of gastric and colorectal cancer

Model	Upper endoscopy	Lower endoscopy
	Mean Percentage error(%) +/-95%CI	Mean Percentage error(%) +/-95%CI
Traditional linear regression	3.8+/-2.3	3.6+/-2.7
Random forest	3.5+/-2.1	3.0+/-2.1
Support vector machine	4.9+/-3.7	3.4+/-1.3
Neural network	9.0+/-3.0	3.4+/-0.6
Stochastic gradient boosting	3.9+/-1.9	2.6+/-2.0
Extreme gradient boosting	3.0+/-2.6	2.4+/-0.01

Conclusion

- Machine learning model, particularly the XGBoost, can be applied in the prediction of minimal number of endoscopy service required to minimize delay in gastrointestinal cancer diagnosis during the ongoing COVID-19 outbreak in Hong Kong.