# System Requirements

This software requires Matlab 2020b or above running under Windows 7 or above. The statistics and optimization toolboxes for Matlab are required and the Matlab report generator is optional.

The software should theoretically run in Matlab on Linux or other operating systems except that there may be some hardcoded, Windows-style file separators in the code. It may also run on older versions of Matlab but has not been tested on them.

This software has been tested on Matlab 2021b running under Windows 10. RStudio 1.4.1717 with MICE 3.13.0 was used to impute missing data values.

There is no required non-standard hardware.

# Installation Guide

Unzip the zip file containing the code and obtain the data from the Synapse [dataset at](javascript:;)  <https://doi.org/10.7303/syn39792658.1>

There is no other installation for this software.

# Demo

## Multiple Imputation

This step takes the raw data, COVID\_reduced\_Spring\_2022.csv, as input and produces imputed datasets: COVID\_reduced\_Spring\_2022\_imp1.csv, COVID\_reduced\_Spring\_2022\_imp2.csv, …, COVID\_reduced\_Spring\_2022\_impM.csv.

Set the filename on line 13 of impute\_COVID\_reduced.r to the full path and filename of the input csv file (e.g. C:\COVID\_reduced\_Spring\_2022.csv). Also specify the full path and filename of the imputed output files on line 121. Specify the categorical variables on line 26, the vitals on line 32, and the labs on line 33. Also set the exclude list on line 55. Then run the file within R-studio. Then open each of the output files and shift the header row to the right one column and delete the first column.

## Creating the mat file

This step takes the raw data and the imputed datasets and produces a mat file that is formatted for the simulation.

Run the following in Matlab:  
create\_mat\_file(ds, arrayfun(@(x) sprintf('COVID\_reduced\_Spring\_2022\_imp%d.csv', x), 1:5, 'UniformOutput', false), [28 7]\*24);

This will create X\_May\_2022.mat. Change the output filename on line 111 if desired.

## Main results

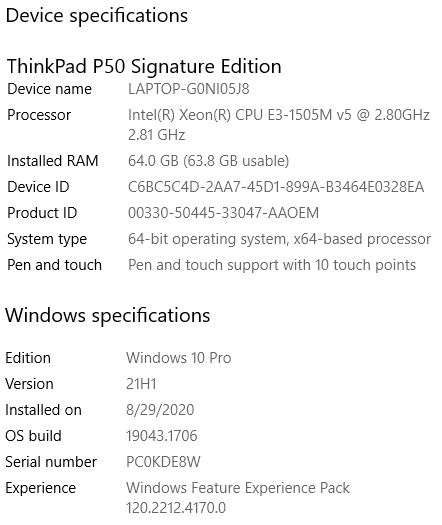
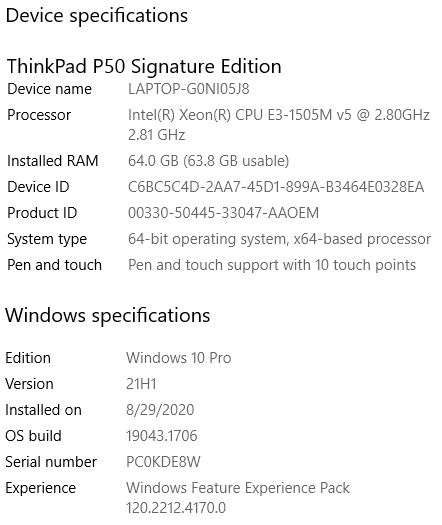
To run the software, run the script sliding\_window\_update\_batch.m in Matlab. This program loops over the 28-day and 7-day outcome time horizons and the NOCOS, logistic regression, and gradient boosted decision tree classifiers. The random number generator is set to default at each iteration to ensure consistency of the results.

The output is a set of figures in the figs folder. The figures are organized into the subfolders 7day and 28 day, and inside each folder is a nocos, LR, and xgboost folder. If generateReport is set to true on line 25 of sliding\_window\_update.m and there is a Matlab Report Generator license, then a PowerPoint file containing the figures is also created.

The outputs include:

* auc\_ici\_vs\_{patients-date}\_{model}\_{horizon}.fig: area under the ROC curve vs the cumulative number of patients for each model/horizon iteration (Figure 3 and Supplemental Figure 2).
* coefs\_{model}\_{horizon}\_{updateMethod}.fig: model coefficient for each model/horizon/update method iteration (Figure 5 as a line graph, Supplemental Figure 4, Supplemental Figure 5 as a line graph)
* The legends of the following figures are used to fill Table 2 and Supplemental Table 1
  + cal\_{model}\_{horizon}\_{updateMethod}.fig: calibration plot for each model/horizon/update method iteration (Figure 2 B-C, Figure 4 B, Supplemental Figure 1 B-C, Supplemental Figure 3 B).
  + pr\_{model}\_{horizon}.fig and roc\_{model}\_{horizon}.fig: precision-recall curves for each model/horizon/update method iteration (Figure 2 A, Figure 4 A, Supplemental Figure 1 A, Supplemental Figure 3 A)
* nb\_{model}\_{horizon}.fig: decision curve analyses for each model/horizon iteration (Figure 2 D, Figure 4 C, Supplemental Figure 1 D, Supplemental Figure 3 C)

The following is a table of the run times achieved for each iteration on a computer with the specifications listed below.



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Iteration | 28-day NOCOS | 28-day LR | 28-day XGBoost | 7-day NOCOS | 7-day LR | 7-day XGBoost |
| Run Time (s) | 132 | 1035 | 105 | 93 | 1276 | 97 |

## Additional Results

In addition to the main results, several other analyses were performed by modifying options in the code.

* To replicate Supplemental Table 4, copy any existing figures to another location to prevent them from being overwritten, replace the cell array that is set to {‘LIJ’} with the length 12 cell array of all hospitals that is commented out on line 5 of sliding\_window\_update\_batch.m, and rerun the script.
* To replicate Figure 6, Supplemental Figure 7, Supplemental Table 2, and Supplemental Table 3, set plotVariantGenderRaceEth to true on line 51 of sliding\_window\_update.m, copy any existing figures to another location to prevent them from being overwritten, and rerun the script.
* Table 1 can be deduced manually from the data.
* Figure 1 was created using the raw data. The patient lengths of stays are needed to reproduce this figure. This is not a quantitative result.
* Supplemental Figure 6 can be reproduced by rerunning the analysis after changing various parameters in the code one-at-a-time.
  + minWindowLengths can be changed on line 34 of sliding\_window\_update.m
  + causal can be changed on line 17 of sliding\_window\_update.m
  + eavgThresh can be changed on line 36 of sliding\_window\_update.m
  + numImputations can be changed to a constant on line 86 of sliding\_window\_update.m

# Instructions for Use

To run the software on arbitrary data, set the mat file on line 62 of sliding\_window\_update.m to the name of the mat file containing the data. The mat file must contain the following variables:

* Xmi: length numImputations cell array of tables where each table is numPatients x numPredictors and each predictor is numeric. Each table is a different randomly imputed version of the raw data.
* admitDtm: length numPatients datetime array of the time of admission for each patient
* cvguid: length numPatients numeric array of unique identifiers for each patient. This is used to align data for performing the pairwise comparisons.
* finalHospital: length numPatients categorical array of the final hospital for each patient. This is used for separating the development cohort and the retrospective validation cohort.
* outcome7 - length numPatients numeric array of 0’s and 1’s. 1 indicates 7-day survival and 0 indicates 7-day mortality.
* outcome28 - length numPatients numeric array of 0’s and 1’s. 1 indicates 28-day survival and 0 indicates 28-day mortality.

To create Xmi from unimputed data, run MICE using R-studio. See impute\_COVID\_reduced.r for an example.