

168 Supplemental Material - Methods

169 Training data

170 Emergency department patients >18 years old who had a WHO-based SARS-CoV-2 PCR test
171 and a complete blood count (CBC) test ordered within 24 hours from 3/1/2020 to 3/20/2020 at
172 Stanford Health Care were included in the training cohort. All data collection was conducted with
173 approval of the IRB at Stanford University and with approval of waiver of consent. If multiple
174 SARS-CoV-2 PCR test results were available for a patient, the first dated result was used, and
175 the subsequent results excluded. If multiple CBC results were available for a patient within 24
176 hours of a SARS-CoV-2 test result, the CBC result that was ordered closest in time to the
177 SARS-CoV-2 PCR order was used and the others excluded.

178 Feature and model selection

179 CBC features were selected based on a univariate analysis of CBC components and the SARS-
180 CoV-2 PCR result within the training set and an analysis of correlation between individual CBC
181 components. Absolute neutrophil count (ANC), absolute lymphocyte count (ALC), and
182 hematocrit were selected based on this analysis. The combination of this manual feature
183 selection and an L2-regularized logistic regression model used for the final model was
184 compared to model-based feature selection methods (e.g. recursive feature elimination, L1/L2-
185 regularized logistic regression) using cross-validation within the training set and demonstrated
186 similar performance.

187 Validation data

188 Model performance was tested on data unseen during training of the model. Data from Stanford
189 Health Care for validation was collected from 3/21/2020 to 4/7/2020 using the same criteria as
190 training. Data collection for validation at University of Washington, Northwestern University,
191 University of Ulsan College of Medicine, and Bundang Jesaeng General Hospital was