

1 Forecasting local surges in COVID-19 hospitalizations 2 through adaptive decision tree classifiers

3 Rachel E. Murray-Watson, PhD,^{1,2,*} Alyssa Bilinski, PhD, ³ Reza Yaesoubi, PhD^{1,2}

4 1. Department of Health Policy and Management, Yale School of Public Health, New
5 Haven, Connecticut, USA

6 2. Public Health Modeling Unit, Yale School of Public Health, New Haven, Connecticut,
7 USA

8 3. Department of Health Services, Policy and Practice, Brown University School of Public
9 Health, Providence, Rhode Island, USA

10 * Corresponding author (rachel.murray-watson@yale.edu).

11 **Abstract**

12 During the COVID-19 pandemic, many communities across the US experienced surges in
13 hospitalizations, which strained the local hospital capacity and affected the overall quality
14 of care. Even when effective vaccines became available, many communities remained at high
15 risk of surges in COVID-19-related hospitalizations due to waning immunity, low uptake of
16 booster vaccinations, and the continual emergence of new variations of SARS-CoV-2. Some
17 risk metrics, such as the CDC’s Community Levels, were developed to predict the impact
18 of COVID-19 on the community-level healthcare system based on routine surveillance data.
19 However, they had limited utility as they were not routinely updated based on accumulating
20 data and were not directly linked to specific outcomes, such as surges in COVID-19 hospital-
21 izations beyond local capacities. Regression models could resolve these limitations, but they
22 have limited interpretability and do not convey the reasoning behind their predictions. In
23 this paper, we evaluated decision tree classifiers that were developed in “real-time” to predict
24 surges in local hospitalizations due to COVID-19 between July 2020 and November 2022.
25 These classifiers would have provided visually intuitive and interpretable decision rules for
26 local decision-makers to understand and act upon, and by being updated weekly, would have
27 responded to changes in the epidemic. We showed that these classifiers exhibit reasonable
28 predictive ability with the area under the receiver operating characteristic curve (auROC)
29 $> 80\%$. These classifiers maintained their performance temporally (i.e., over the duration of
30 the pandemic) and spatially (i.e., across US counties). We also showed that these classifiers
31 outperformed the CDC’s Community Levels for predicting high hospital occupancy.

32 **Significance Statement**

33 A major concern during the COVID-19 pandemic was the risk of exceeding local healthcare
34 capacity due to COVID-19-related hospitalizations. To assess this risk and inform mitigating
35 strategies, several risk assessment tools were developed during the pandemic. Many of these
36 tools, however, did not predict local outcomes, were not updated as the pandemic progressed,
37 and/or were not interpretable by decision-makers. We propose an adaptive framework of de-
38 cision tree classifiers to predict whether COVID-19-related hospital occupancy would exceed
39 a given capacity threshold. This framework would provide interpretable classification rules to
40 predict surges in local hospitalizations, and maintained its performance over time and across
41 US counties, and outperformed the CDC’s Community Level tool.

42 **Keywords**

43 COVID-19, decision tree classifier, risk prediction, surveillance

44 1 Introduction

45 When COVID-19 emerged in the United States in 2020, it proved an immediate threat to
46 healthcare systems. Rapid surges in hospitalizations threatened to overwhelm the healthcare
47 capacity and compromise the standard of care and patient health outcomes [1]. Even after
48 vaccinations were introduced in late 2020, many communities remained at risk of intermittent
49 surges in COVID-19 hospitalizations due to low rates of vaccination, waning infection- and
50 vaccine-induced immunity, and seasonal changes in transmission dynamics [2, 3].

51 To assess the risk of surges in hospitalizations, significant efforts were made to predict the
52 trajectory of the pandemic [4], including the COVID-19 Forecast Hub [5], the Scenario
53 Modeling Hub [6], and the IHME COVID-19 Forecast Model [7]. However, the majority of
54 these predictions are made available at the state or national levels [4], reducing their utility
55 in more local settings, where the pandemic trajectory could bear little resemblance to that
56 of the nation or state. Hence, there was a need among local policymakers for tools that will
57 convert the data collected by surveillance systems into meaningful predictions for their area.

58 One attempt at such a tool was the CDC’s COVID-19 Community Levels [8] (replaced with
59 the COVID-19 Hospital Admission Rate [9] and the COVID-19 County Check [10] in 2023),
60 which was designed to indicate when there may be an upcoming strain on healthcare systems.
61 The COVID-19 Community Levels were based on the number of new weekly COVID-19
62 cases, weekly hospital admissions due to COVID-19, and weekly inpatient beds occupied by
63 COVID-19 patients and predicted whether the Community Level would be low, medium, or
64 high. While these metrics were chosen for their correlation with future high mortality and
65 hospital occupancy [8], they did not directly predict any specific outcome of interest [11].
66 Additionally, the thresholds for the Community Levels were chosen once and not updated
67 as the pandemic progressed or new coronavirus strains became dominant. This limited the
68 overall utility of the Community Levels tool in the long term. Previous work has attempted
69 to address these issues by using regression models that are continuously updated to predict
70 concrete outcomes (e.g., mortality level) [12]. The main limitation of regression models is
71 that they are not easily interpretable and do not convey the reasoning behind their assessment
72 of risk for some adverse outcomes (e.g., surges in COVID-19 hospitalizations).

73 In this paper, we investigate whether accurate decision tree classifiers could be developed
74 to predict local surges in COVID-19 hospitalizations. Decision tree classifiers are machine
75 learning models that provide simple and interpretable classification rules to make predictions
76 [13, 14]. They resemble the threshold-based, flowchart-like structure of the CDC Community
77 Levels, which makes them easy to use and interpret by local policymakers. Using decision
78 tree classifiers, however, allows us to explicitly link the surveillance data to the outcome of
79 interest, which here is to predict whether local hospital occupancy due to COVID-19 patients
80 would exceed a set threshold. To develop these decision tree classifiers, we use data collected
81 from US county surveillance systems and evaluate their performance during different waves
82 of the pandemic between July 2020, before which data was not routinely collected from each
83 county in the US, and November 2022. We also compare the performance of these decision
84 tree classifiers with that of CDC Community Levels.

85 2 Method

86 2.1 Overview

87 From July 2020, several COVID-19 indicators (e.g., deaths, cases, hospital admissions) were
88 collected through surveillance systems to monitor and predict trends in the pandemic [15, 16].
89 Our goal is to use these indicators and metrics derived from these indicators as features in
90 decision tree classifiers to predict whether the local hospital capacity is expected to exceed
91 in the short term due to surges in COVID-19 hospitalizations.

92 We consider four groups of classifiers that differ based on the predictors (features) they use to
93 predict whether local hospital capacity would be surpassed in a three weeks from the current
94 week. These models are described in detail below. To develop and evaluate these models, we
95 used data collected between July 15th, 2020, and November 7th, 2022, a 123-week period.
96 Before July 2020, data was not routinely collected and reported. In December 2022, there
97 was a change in hospitalization reporting guidelines, and data was reported to the Centers
98 for Disease Control and Prevention’s (CDC) National Health Safety Network (NHSN) rather
99 than to the Department of Health and Human Services (HHS). After this period, there were
100 changes in the quality of the data being reported. Hence, we focused our analysis over the
101 period July 15th, 2020 to November 7th, 2022.

102 For a given week t , we train decision tree classifiers using data collected between week 1
103 through week $t - 1$. We then use the data collected in week t to predict the outcome in
104 week $t + 3$. For example, at the beginning of week $t = 10$, we use the data collected through
105 week 9 to develop models to predict whether hospital capacity would be exceeded in week
106 13. To evaluate how the performance of these classifiers changes throughout the pandemic
107 (especially during the phases where novel variants emerge), we repeat this procedure for
108 every week $t \in \{2, 3, 4, \dots, 118\}$.

109 2.2 Data

110 We obtained COVID-19 hospital admissions, occupancy, and ICU occupancy data from the
111 Department of Health and Human Services [16], and the data on cases and deaths from
112 the New York Times [15]. Only cases confirmed by a reverse-transcriptase polymerase chain
113 reaction (RT-PCR) test were included in the definition of “case”. We followed the procedures
114 outlined in previous studies for data pre-processing, including the aggregation of weekly
115 observations and the imputation of missing values [11, 12].

116 To account for patients leaving their county of residence to access healthcare, we aggregated
117 data by Health Services Areas (HSAs)[17] consistent with the CDC’s Community Level
118 calculations. We compiled data at the midpoint of each week. A total of 804 HSAs were
119 included in the analysis; for each of the classifiers we developed, $< 1.5\%$ of the health service
120 areas were omitted from the data set due to missing data.

121 2.3 Outcomes

122 The outcome of interest was whether the COVID-19-caused hospital occupancy would ex-
123 ceed 15 per 100,000 population in exactly three weeks' time. The capacity threshold of 15
124 per 100,000 population is calculated in prior studies and falls in the middle of the CDC Com-
125 munity Level's "Medium" risk assessment for the hospital admissions indicator [18, 19, 20]
126 (Figure S1 in SI). We chose the three-week period for consistency with the CDC Community
127 Levels [8, 12]. However, we note that knowledge of hospital capacity in interceding weeks
128 is also of use to policymakers. Consequently, we considered whether this outcome occurs
129 any week during the interval $[t + 1, t + 3]$ as a sensitivity analysis. This time period skips
130 the immediate next week but maintains a relatively short period between the training and
131 target weeks.

132 In addition to the initial threshold of 15 cases per 100,000 population over a three-week
133 period, we performed a sensitivity analysis using alternative threshold values and outcome
134 periods.

135 2.4 Features

136 The prediction models we consider utilize all or a subset of the following features:

- 137 1. Number of COVID-19 cases per 100,000 population in the last week,
- 138 2. Number of COVID-19 deaths per 100,000 population in the last week,
- 139 3. Number of COVID-19-related hospital admissions per 100,000 population in the last
140 week,
- 141 4. Number of hospital beds occupied by COVID-19 patients per 100,000 population in
142 the last week,
- 143 5. Number number of ICU beds occupied by COVID-19 patients per 100,000 population
144 in the last week,
- 145 6. Portion of hospital beds occupied by COVID-19 patients in the last week,
- 146 7. Change in each of the aforementioned metrics in the last week, and
- 147 8. Whether the hospital occupancy by COVID-19 patients exceeded the selected capacity
148 threshold in the last week.

149 2.5 Decision tree classifiers

150 We consider four broad categories of decision tree classifiers, which differ in the features they
151 include, to predict whether hospital capacity will be exceeded:

- 152 1. "Naive" classifier that uses only a binary variable of whether the current hospital
153 capacity threshold is exceeded;

- 154 2. “CDC Optimized” classifier that uses the same features included in the CDC Commu-
155 nity Levels (i.e., new weekly COVID-19 cases, hospital admissions, and the percent of
156 inpatient beds used by COVID-19 patients);
- 157 3. “Reduced” classifier, which uses features related to hospital admissions, hospital and
158 ICU occupancy, but not cases or deaths (because as of mid-2023, the county-level case
159 and death data was no longer routinely reported [15]);
- 160 4. “Full” classifier that uses all features listed in the previous section.

161 We compare the performance of these classifiers with that of the CDC’s Community Levels
162 (Figure S1 of SI), which were developed using all data collected between March 1st, 2021
163 and January 24th, 2022 [8]. To measure the performance of Community Levels, we evaluate
164 it on the period between March 3rd, 2022 and November 20th, 2022, during which its use
165 was encouraged by the CDC.

166 We also note that the Community Levels designate an area as “low”, “medium”, or “high”
167 risk; hence, to evaluate its ability to predict surges in COVID-19 hospitalizations, we counted
168 all weeks in which the set hospitalization threshold was exceeded as equivalent to a “high”
169 community level, and weeks where it was below as equivalent to “medium/low”.

170 **2.6 Model development**

171 For each week t between July 15th, 2020 and November 7th, 2022, we use the data collected
172 from all HSAs from week 1 to week $t - 1$ as a single training set to develop our classification
173 trees. We used 10-fold spatial and temporal cross-validation both to optimize the model
174 hyperparameters (see §S2 in SI for details). To ensure that resulting decision trees were easy
175 to interpret, we restricted the depth of trees to less than five layers and then found optimal
176 depth through hyperparameter tuning.

177 Since the data from larger populations are expected to be less noisy compared to data from
178 smaller populations, we included instance weights in the model fitting procedure, based on
179 HSA population. Additionally, to account for class imbalance in the outcome, we trained
180 the model using “balanced” class weights, with higher weights assigned to the minority class
181 in the decision tree classifier’s function [21, 22].

182 Though we used 123 weeks of data, due to the 3-week prediction task, the need for one week
183 of test data, and one week for initial training, we could only train 117 decision tree classifiers.

184 We developed our decision tree classifiers using the scikit-learn package in Python [23].

185 **2.7 Model evaluation**

186 In line with Transparent Reporting of a multivariable prediction model for Individual Prog-
187 nosis or Diagnosis (TRIPOD) recommendations [24], model validation was carried out using
188 temporal validation, partitioning our dataset into training and testing subsets based on time.
189 To evaluate the performance of a decision tree classifier, we calculate the area under the re-
190 ceiver operating curve (auROC) based on data from all HSAs collected during the projection

191 period. A classifier that provide 100% correct predictions has an auROC of 1 and a classifier
192 that randomly guess that outcomes has an auROC of 0.5.

193 To investigate the regional variation in the performance of our classifiers, we also present the
194 auROC scores based on the out-of-sample predictions by HSA. These auROC scores were
195 calculated based on the predictions made by each weekly classifier for a given HSA, which
196 was then matched with the corresponding HSA.

197 The auROC metric is agnostic to the selected classification threshold and presume that
198 true/false positives and true/false negatives are equally desirable outcomes. In reality, a
199 predicted surge in hospital capacity will cause different responses among policymakers, each
200 with an associated cost. Under this assumption, the value of true and false predictions may
201 differ. To allow for any differences in preferences policymakers may have, we additionally
202 calculate the net benefit [25], a metric that allows for differences in the weighting of true
203 and false positives, facilitated by an “exchange” parameter, ω . The net benefit function
204 incorporating true positive and false positive rates is given by:

$$NB_P(\omega, p) = TP(p) - \omega FP(p), \quad (1)$$

where $TP(p)$ is the true positive rate and $FP(p)$ is the false positive rate when the classification threshold p is selected (a classifier with classification threshold p predicts the hospital capacity would be exceeded if the estimated the probability of exceeding capacity is greater than p). A true positive result means that the model has correctly predicted the surge in hospital capacity; hence, any action undertaken by the policymaker to avoid such a surge was justified and could have prevented loss in population health. In contrast, a false positive result could waste resources if it leads to using mitigating strategies. Equation 1 allows policymakers to weigh the economic costs of unnecessary action ($\omega FP(p)$) with the health benefits of justified action ($TP(p)$). These outcomes, however, depend on the classification threshold p . To find the optimal classification threshold for a given ω , we solve:

$$NB_P^*(\omega) = \max_p [TP(p) - \omega FP(p)],$$

205 using a grid search over values of p .

206 The above definition of net benefit does not consider the health and economic consequences
207 of true and false negatives. True negatives mean that policymakers can avoid implementing
208 costly interventions and false negatives may risk overwhelming hospital capacity. To account
209 for these factors, we also evaluate our classifiers using the following version of net benefit:

$$NB_{P,N}(\omega, p) = [TP(p) - FN(p)] - \omega [FP(p) - TN(p)]. \quad (2)$$

210 For a given ω , we use a grid search to find the optimal classification threshold that maximizes
211 the above function.

212 We also evaluate the performance of different classifiers using maximum regret (MR). MR is
213 the difference between the performance of the best-performing model and that of the model
214 under consideration and gives a metric of loss or gain in difference in performance by using

215 one model over another. Mathematically, for a given metric q (e.g., auROC), the MR of
216 model m is defined as [12]:

$$MR_m = \max_{m \in M}(q_m) - q_m, \quad (3)$$

217 where M is the set of all models under consideration and q_m is the performance (e.g., auROC)
218 of model m .

219 Finally, we used SHAP (Shapley Additive exPLanations) values to evaluate the contribution
220 of each feature to predictions [26]. SHAP values provide a way to allocate the contribution
221 of each feature to a model’s prediction by considering all possible combinations of feature
222 values across instances [27]. It provides insight into changes in model performance that may
223 occur if some features are not included. We use the shap package in Python to calculate the
224 SHAP values and present the SHAP summary plot [27]. In this plot, we present the SHAP
225 values for each prediction period across all 117 Reduced classifiers.

226 3 Results

227 Between July 15th, 2020 and November 7th, 2022, there was a substantial variation in the
228 burden of COVID-19 across HSAs (Figure 1). Throughout this 123-week period, on average,
229 78% of HSA-weeks surpassed the 15 COVID-19 patients per 100,000 hospitalized threshold
230 (dashed curve in Figure 1C).

231 COVID-19 cases, new hospital admissions, hospital beds and ICU beds occupied by COVID-
232 19 patients, and the percentage of hospital beds occupied by COVID-19 patients were posi-
233 tively correlated with surpassing the hospitalization threshold (Table 1).

234 Across all classifiers, the auROC was always above 0.5, including for the Naive classifiers
235 which used just one feature. The Full classifiers generally had the highest auROC, though
236 the Reduced classifiers were competitive despite not using features related to cases or deaths.
237 The CDC Optimized, Reduced and Full classifiers had an auROC of > 0.8 across all weeks
238 (Figure 2A). The performance of these classifiers was relatively consistent across different
239 waves of the pandemic. However, the regret for auROC scores was best controlled by the
240 Full and Reduced classifiers (Figure 2B).

241 Across all models, around week 62, there was a decrease in the auROC score, which coincided
242 with when the omicron strain became the predominant strain in the US. There is additionally
243 a decrease in performance when there was a peak in cases during the omicron wave (around
244 week 90). Conversely, the highest auROC scores were achieved when a high proportion of
245 HSAs exceeded the 15 per 100,000 hospitalization threshold (Figure 2A). As county-level
246 case and death data ceased being available from mid-2023, we focused the remainder of the
247 analysis on the Reduced classifiers.

248 For 76% of the counties, the auROC that could be calculated using the Reduced classifiers
249 over the entire study period exceeded 0.80 (Figure 3). However, for around 23% of counties,
250 an auROC could not be calculated as across each 117 outcome weeks, the hospital beds
251 occupied by COVID-19 patients always exceeded the 15 per 100,000 threshold, or never did.
252 Thus, there was no “true negative” with which to calculate the auROC.

253 The added benefit of the Reduced classifiers compared with the Naive classifiers varied over
254 the pandemic weeks and depended on the selected penalty value (ω in Eq. 1) to model the
255 trade-off between the true positive and false positive rates (Figure 4A). For smaller values of
256 ω , which represent scenarios where the false positive is minimally penalized, the added benefit
257 of using the Reduced classifiers is unimportant. Reduced classifiers provide greater benefit
258 when false positives are penalized more substantially and when the proportion of HSAs where
259 hospital capacity exceeds 15 per 100,000 is large (Figure 4A). However, the Naive classifiers
260 provide additional net benefit over the Reduced classifiers when the proportion of HSAs
261 exceeding the hospital occupancy threshold was low. The Full classifier presented higher net
262 benefit compared to the Naive classifier over each penalty value and weekly classifier (Figure S8
263 in the SI).

264 We observe similar behavior when the net benefit function $NB_{P,N}()$, which accounts for true
265 and false positive rates and true and false negative rates, is used (Figure 4B). When this net
266 benefit function is used, the Reduced classifiers outperform the Naive classifiers for a larger
267 number of weeks and penalty values.

268 The change in the number of COVID-19 hospital admissions in the previous week, the
269 number of beds occupied by COVID-19 patients, and the number of COVID-19 hospital
270 admissions, were selected as important features in more than 50% of 117 Reduced classifiers
271 between July 15th, 2020 and November 7th, 2022 (Figure 5). Whether the current hospital
272 capacity exceeded the 15 per 100,000 threshold was not included as an important feature
273 in any classifier, despite the good performance of the Naive classifier, which only used this
274 feature. However, this binary feature is highly correlated with other features such as the
275 number of hospital beds by week and the percentage of beds occupied by COVID-19 patients
276 (Table 1), which do appear as important features (Figure 5).

277 The SHAP values also indicate that COVID-19 admissions and ICU beds occupied by
278 COVID-19 patients have a large influence on model predictions (Figure 6). The SHAP
279 values broadly follow a trend where high numbers of admissions and occupied hospital beds
280 (indicated by the pink dots in Figure 6) increase the log odds of a positive prediction, whilst
281 lower admissions decrease the log odds.

282 One major advantage of and the main motivation for the use of decision tree classifiers is that
283 they are interpretable. To demonstrate, we present three decision tree classifiers developed
284 for three different stages of the pandemic (Figure 7): the week of July 14th, 2021, when
285 the delta variant was circulating in the population but was not yet the dominant strain
286 (Panel A), the week of August 4th, 2021, when the delta variant was dominant (Panel B),
287 and the final week in our study time period, i.e., the week of November 20, 2022 (Panel
288 C). To illustrate how these decision tree classifiers could be used in practice by a local
289 policymaker, we consider the scenario observed in an HSA in Maryland for July 14th, 2021
290 (Table 2). Since hospital beds occupied by COVID-19 patients = 22.82, which is less than
291 29.0, the condition of the first decision node is satisfied. Hence, we check the condition of
292 the second node, whether hospital beds occupied by COVID-19 patients ≤ 10.0 , which is
293 satisfied. Therefore, this classification decision tree predicts that the hospital capacity of 15
294 per 100,000 population is not expected to be exceeded in 3 weeks' time.

295 We note that while these classification trees are developed for different phases of the pan-
296 demic, they all use the same features: 1) the number of hospital beds occupied by COVID-19
297 patients, 2) the number of ICU beds occupied by COVID-19 patients, and 3) the number
298 of hospital admissions of COVID-19 patients. What is different between these classification
299 trees is the classification thresholds and the depth of the tree; for example, the classification
300 threshold at the first node of these trees are 29.0 (Panel A), 30.0 (Panel B), and 27.0 (Panel
301 C).

302 4 Discussion

303 We presented an adaptive framework to generate simple classification rules that predict
304 whether COVID-19 hospitalizations would have exceeded local capacity. These classification
305 rules are easy to communicate and use the latest data from surveillance systems that are
306 available at the local level. To develop these classification rules during the weeks of July,
307 2020 to November, 2022, we trained decision tree models on “expanding” datasets, where
308 all available data up to the target week of interest are used in the training procedure. This
309 allowed for the model training procedure to account for changes in pandemic trajectories as
310 population-level factors not directly accounted for as features in the model (e.g., vaccina-
311 tion rates, waning immunity, and easing non-pharmacological interventions). Our analysis
312 suggests that classification rules that are adaptively updated could maintain their predictive
313 ability temporally (i.e., over the duration of the pandemic; Figure 2) and spatially (i.e.,
314 across US counties, Figure 3). Additionally, these classifiers outperform other classifiers
315 trained only on more recent datasets (see SI S4).

316 Our classifiers were trained on a range of COVID-19 indicators that were routinely reported
317 between July 2020 and November 2022, including hospital admissions, and hospital occu-
318 pancy. However, the source of case and death data used in this study for the “Full” classifiers
319 ceased being updated in mid-2023 [15], and other data sources are only updated at the state
320 level [28] if at all [29, 30]. Though this limits the training data on which the classifiers can be
321 trained, we showed omitting death and case data does not meaningfully impact the perfor-
322 mance of the characterized classification rules (Figure 2). Indeed, the CDC Optimized, Full,
323 and Reduced presented reasonable predicted accuracy, with an auROC of > 0.80 for the ma-
324 jority of classifiers; even classifiers based solely on current occupancy can achieve relatively
325 high auROC, including when a new strain becomes dominant in the population (Figure 2).
326 In the Reduced classifiers, the admissions and change in admissions, and occupied ICU beds
327 were most frequently considered to be important in the decision tree classifiers (Figure 5).
328 Similarly, these features had some of the largest explanatory effects on the auROC (Figure
329 6).

330 We focused on providing predictions at the level of HSAs as this would have enabled local
331 policymakers to detect potential surges in COVID-19 hospitalizations and to respond accord-
332 ingly before the local hospital capacity is overwhelmed. One such previous attempt was the
333 CDC’s Community Levels framework. This framework, however, lacked a direct relationship
334 with a specific outcome of interest [18] and was never updated after its development. Other
335 studies have provided more systematic classification rules predicting specific outcomes (such

336 as hospital capacity [20] or high mortality [12]). We extended this work by demonstrating
337 that decision tree classifiers trained on surveillance data could predict whether COVID-19
338 hospitalizations may exceed the local capacity level with reasonable accuracy. These classi-
339 fication rules provided by these models are easy to use and interpret in practice (Figure S9
340 in SI).

341 The spread of SARS-CoV-2 varied substantially across different communities and geographic
342 regions, depending on the implementation of non-pharmaceutical interventions [31], local
343 rates of vaccination [32], and the emergence of novel strains, amongst other factors [33, 34,
344 35]. Despite the substantial heterogeneity in data reported by HSAs, our analysis suggests
345 that our Reduced classifiers maintained their performance across US counties (Figure 3).
346 We note that the spatial performance of classifiers was not dramatically improved when case
347 and death data were used to develop classification rules (Figure S7).

348 By using the net benefit framework, our approach allows for incorporating policymakers'
349 preferences between different prediction outcomes (i.e. false negative and positive and true
350 negative and positive). For example, predicting that there will be a surge in hospital occu-
351 pancy when there will not actually be one (false positive) or predicting no surge when one is
352 going to occur (false negative) have distinct health and economic costs; the former could have
353 high economic costs and the latter could lead to high adverse health outcomes. Our analysis
354 shows that the classification rules identified by our approach present positive net benefit
355 values, especially during the phases of the pandemic when the COVID-19 hospitalizations
356 were rising or declining (Figure 4).

357 Two of the classifiers considered here (i.e., “CDC Optimized” and “Full”) use case and death
358 data. The available case data undercounted the actual number of infections, as only cases
359 confirmed by a molecular laboratory test were included [15]. Limited testing availability [36],
360 particularly at the beginning of the pandemic, and asymptomatic transmission contributed to
361 this undercount [37]. Similarly, death counts relied on the patient to be diagnosed according
362 to guidelines specified by state and federal bodies [38], which could have led to both the
363 mis- and underdiagnosis of patients. However, our analysis shows that even without these
364 COVID-19 indicators (as in our “Reduced” classifier), reasonable predictive power can still
365 be achieved (Figure 2).

366 An important limitation of predictive models that are only trained on historical data is that
367 the historical data may not sufficiently capture the possible changes in pandemic trajectories
368 due to future events. This is particularly true if novel strains emerge with characteristics that
369 are substantially different from what is manifested in historical data. The use of simulated
370 trajectories, which can be parameterized to allow for such variations, may help to make the
371 predictions of the classifier more robust [20].

372 The proposed approach is flexible to facilitate a number of extensions. While we trained our
373 models to predict a binary outcome, classification decision trees could also be developed to
374 predict multiple outcomes, such as “low”, “medium” and “high” levels of COVID-19 hos-
375 pital capacity, if thresholds to define these outcomes are available. Furthermore, we only
376 included indicators related to the COVID-19 pandemic to predict surges in hospital occu-
377 pancy. However, there is evidence of cocirculation of influenza, COVID-19, and respiratory

378 syncytial virus infections during recent winters [39, 40], which threatens to overwhelm the
379 healthcare system. Data from surveillance systems related to these infectious diseases could
380 be incorporated to provide more robust predictions. Finally, while we focused on predicting
381 surges in hospitalizations, other metrics of interest, such as ICU capacity or mortality, could
382 also be considered depending on the availability and the quality of data related to these
383 outcomes.

384 5 Figures

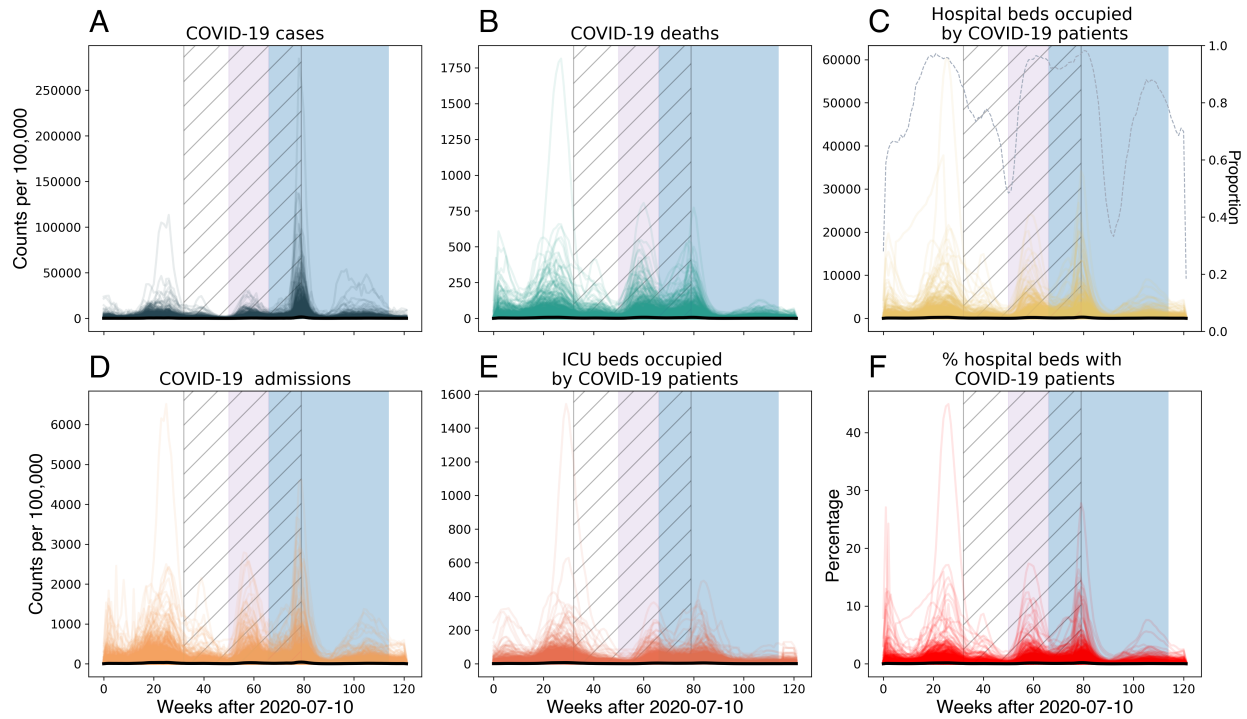


Figure 1: **Weekly COVID-19 indicators between July 15th, 2020 and November 7th, 2022 reported by HSAs.** The purple and blue boxes shows the period when the delta and omicron variants were dominant, respectively. The hatching shows the data used to develop the CDC Community Levels. The black curve shows the mean across all HSAs for the given indicator.

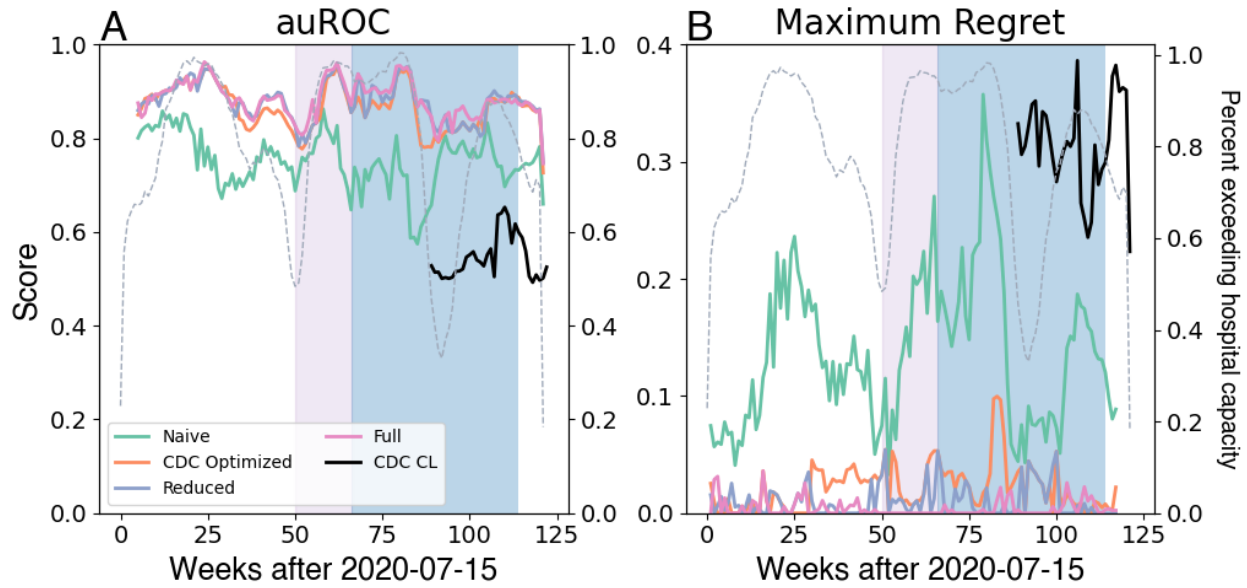


Figure 2: Performance of decision tree classifiers to predict whether the COVID-19 hospital occupancy is expected to exceed 15 per 100,000 in exactly 3 weeks. (A) auROC, (B) maximum regret. The purple and blue boxes show the period when, respectively, the delta and omicron variants were dominant. The gray dashed line shows the proportion of HSAs that exceeded the hospitalization threshold of 15 per 100,000 population for a given week.

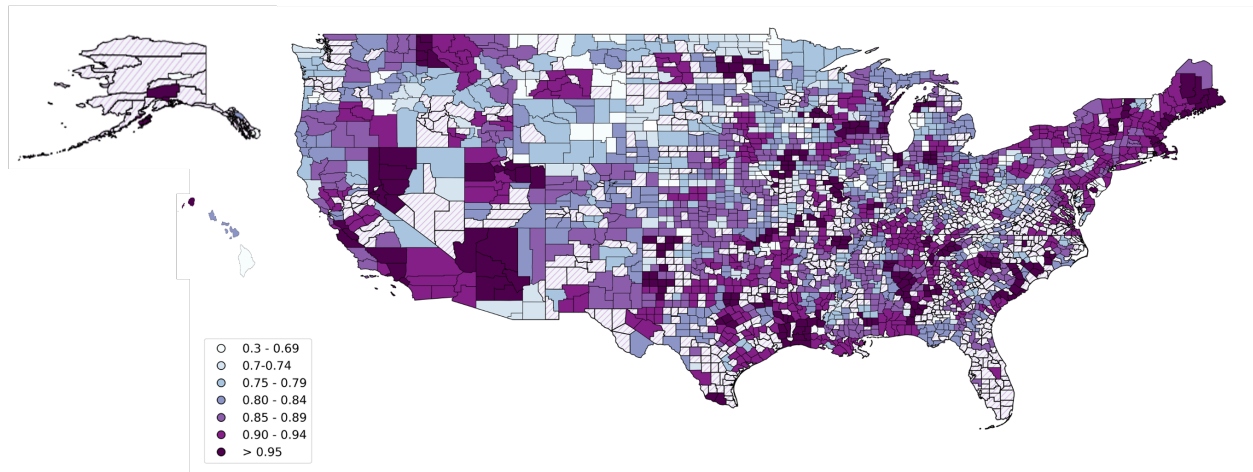


Figure 3: Performance of Reduced classifiers across all counties. The auROC was calculated by HSA using the predictions from all 117 Reduced classifiers. The hatching indicates where there were no true negative instances with which to calculate the auROC, and the auROC is recorded as NA. See Figure S7 for the spatial performance of classifiers that includes case and death data.

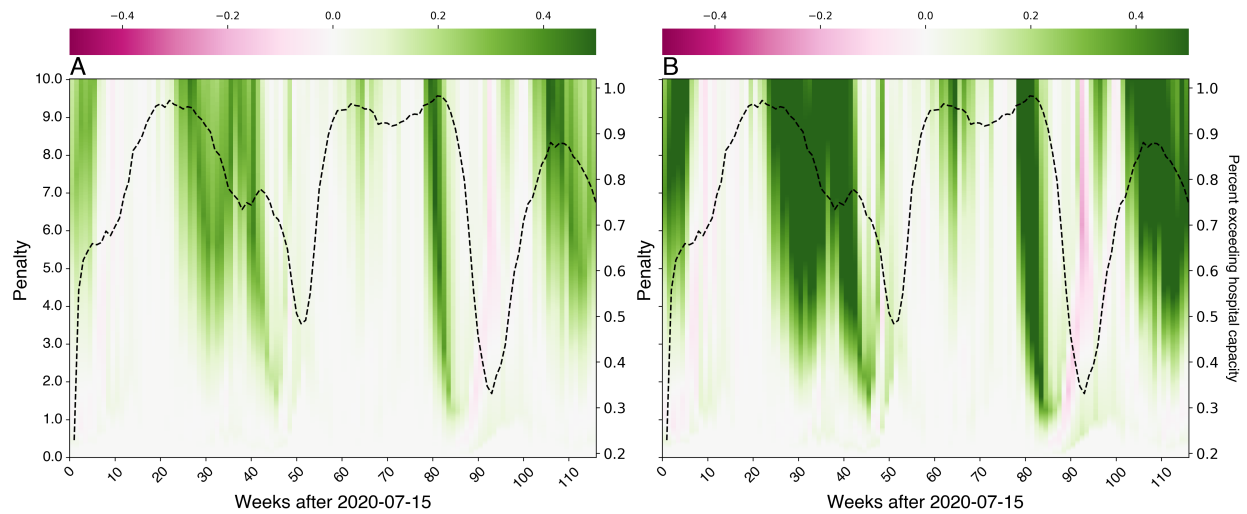


Figure 4: **The net benefit of the Reduced classifiers related to the Naive classifiers.** (A) Using the net benefit function $NB_P()$, which accounts for false positive and true positive rates and (B) Using the net benefit function $NB_{P,N}()$, which accounts for true and false positive rates and true and false negative rates. In areas shaded green, the Reduced classifiers outperform the Naive classifier, while areas shaded pink indicate where the Naive classifier perform better. The gray dashed line is the proportion of HSAs that exceeded the 15 per 100,000 hospital capacity for a given week.

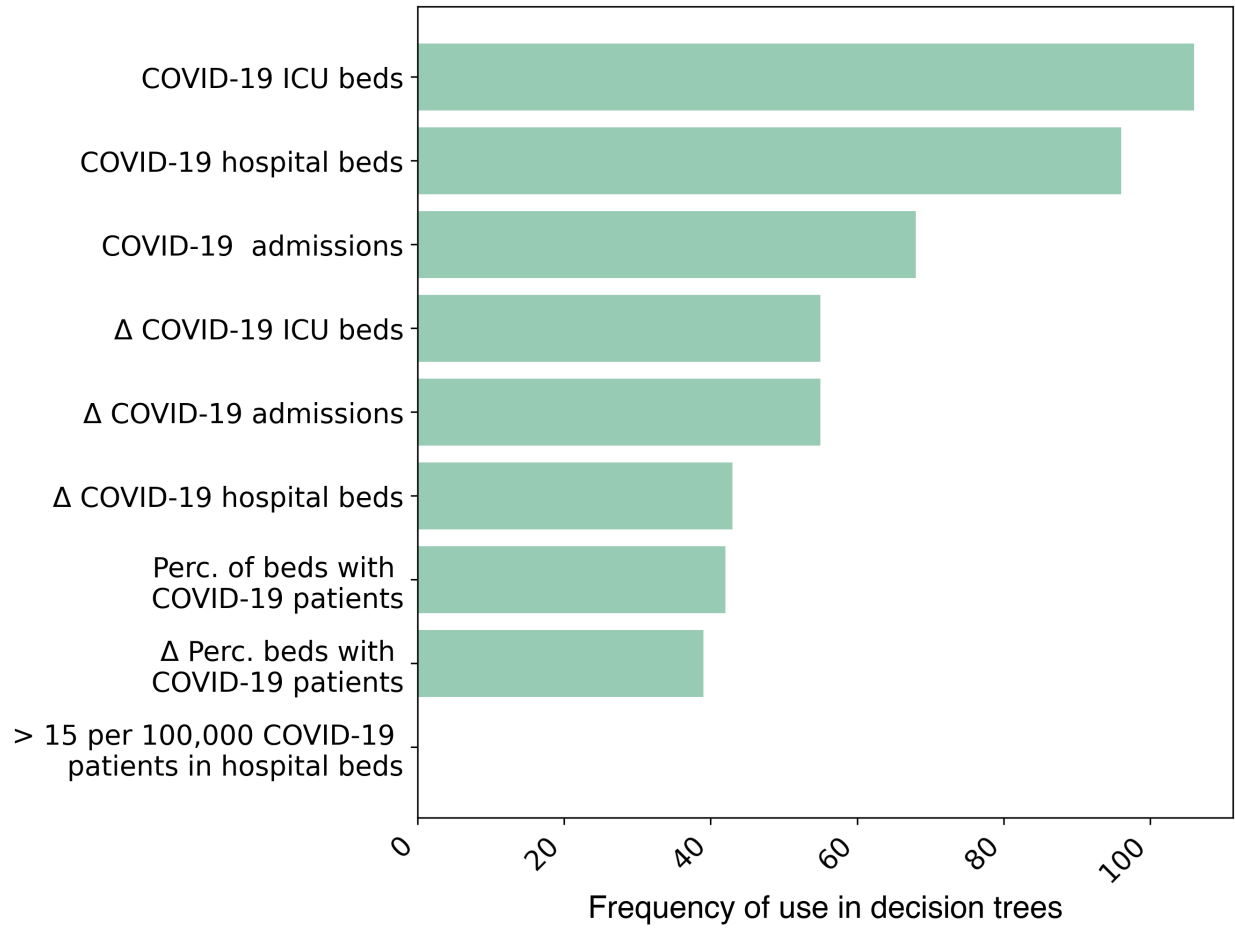


Figure 5: Frequency at which a feature was identified as important in 117 Reduced classifiers created between July 15th, 2020 and November 7th, 2022.

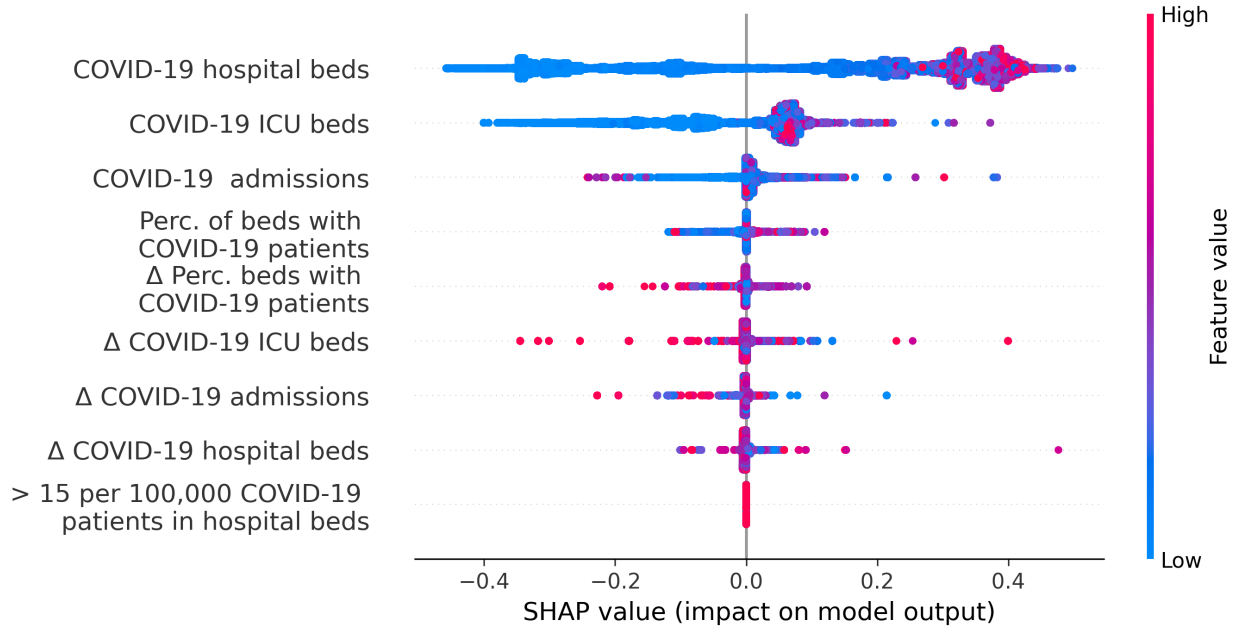


Figure 6: **SHAP summary plot.** Each point represents the SHAP value for a feature and an instance (observations from a single HSA over a single week). The color of each point represents the value of the feature from low to high. Overlapping points are jittered vertically to display the distribution of SHAP values per feature.

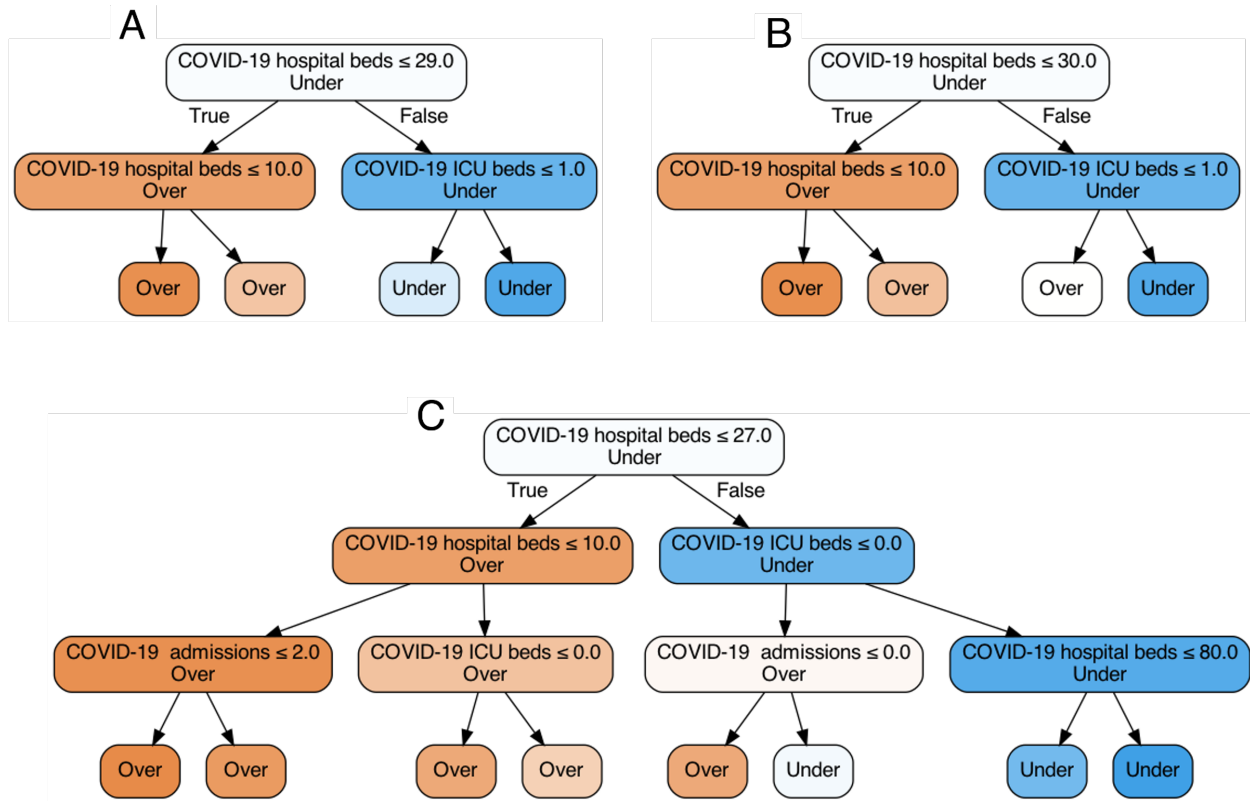


Figure 7: **Decision tree classifiers developed at three different points during the pandemic.** The week of July 14th, 2021, when the delta variant was circulating in the population but was not yet the dominant strain (Panel A), the week of August 4th, 2021, when the delta variant was dominant (Panel B), and the final week in our study time period, i.e., the week of November 20, 2022 (Panel C).

385 6 Tables

Table 1: Mean and standard deviation of weekly COVID-19 observations stratified by whether the hospital occupancy is surpassed in 3 weeks' time or not.

Feature	Surpassed		Not surpassed	
	Mean	Std Dev	Mean	Std Dev
Cases	143.63	290.4	53.4	122.9
Deaths	1.8	3.2	1.4	3.0
Hospital beds	37.9	87.3	1.4	4.3
Admissions	9.4	21.6	2.3	9.1
ICU beds	1.75	3.5	0.08	0.44
Percent beds occupied by COVID-19 patients	4.3	9.4	0.023	0.073
Change in cases	-1.4	169.0	2.5	88.1
Change in deaths	0.059	1.1	-0.16	1.3
Change in hospital beds	-0.098	41.6	0.016	4.3
Change in admissions	-0.19	14.7	0.36	9.1
Change in ICU beds	-0.0057	1.64	0.0076	0.38
Change in percent beds occupied by COVID-19 patients	-0.00022	0.071	0.000024	0.0064

Table 2: Data for Allegany/Garrett HSA in Maryland on July 14, 2021. This HSA contains Allegany and Garrett counties in Maryland and Grant, Hardy, and Mineral Counties in West Virginia.

Value	
Hospital Admissions	1.35
ICU beds	0.52
Hospital beds	22.82
Percent beds occupied by COVID-19 patients	1.5
Change in admissions	-1.81
Change in ICU beds	0.34
Change in hospital beds	5.65
Change in percent beds occupied by COVID-19 patients	0.38
Currently exceed threshold capacity	Yes
Exceed capacity three weeks later	Yes

References

- 386
- 387 [1] Sameer S Kadri, Junfeng Sun, Alexander Lawandi, Jeffrey R Strich, Lindsay M Busch,
388 Michael Keller, Ahmed Babiker, Christina Yek, Seidu Malik, Janell Krack, et al. As-
389 sociation between caseload surge and covid-19 survival in 558 us hospitals, march to
390 august 2020. *Annals of Internal Medicine*, 174(9):1240–1251, 2021.
- 391 [2] Christopher JL Murray and Peter Piot. The potential future of the COVID-19 pandemic:
392 will sars-cov-2 become a recurrent seasonal infection? *Jama*, 325(13):1249–1250, 2021.
- 393 [3] Chadi M Saad-Roy, Caroline E Wagner, Rachel E Baker, Sinead E Morris, Jeremy
394 Farrar, Andrea L Graham, Simon A Levin, Michael J Mina, C Jessica E Metcalf, and
395 Bryan T Grenfell. Immune life history, vaccination, and the dynamics of sars-cov-2 over
396 the next 5 years. *Science*, 370(6518):811–818, 2020.
- 397 [4] Kristen Nixon, Sonia Jindal, Felix Parker, Nicholas G Reich, Kimia Ghobadi, Eliza-
398 beth C Lee, Shaun Truelove, and Lauren Gardner. An evaluation of prospective COVID-
399 19 modelling studies in the usa: From data to science translation. *The Lancet Digital*
400 *Health*, 4(10):e738–e747, 2022.
- 401 [5] Estee Y Cramer, Yuxin Huang, Yijin Wang, Evan L Ray, Matthew Cornell, Johannes
402 Bracher, Andrea Brennen, Alvaro J Castro Rivadeneira, Aaron Gerding, Katie House,
403 et al. The United States COVID-19 forecast hub dataset. *Scientific data*, 9(1):462,
404 2022.
- 405 [6] Joseph C Lemaitre, Kyra H Grantz, Joshua Kaminsky, Hannah R Meredith, Shaun A
406 Truelove, Stephen A Lauer, Lindsay T Keegan, Sam Shah, Josh Wills, Kathryn Kamin-
407 sky, et al. A scenario modeling pipeline for COVID-19 emergency planning. *Scientific*
408 *reports*, 11(1):7534, 2021.
- 409 [7] IHME COVID-19 Forecasting Team. Modeling COVID-19 scenarios for the United
410 States. *Nature medicine*, 27(1):94–105, 2021.
- 411 [8] Centers for Disease Control and Prevention. Science brief: Indicators for monitoring
412 COVID-19 Community Levels and making public health recommendations. In *CDC*
413 *COVID-19 Science Briefs [Internet]*. Centers for Disease Control and Prevention (US),
414 2022.
- 415 [9] Heather M Scobie, Mark Panaggio, Alison M Binder, Molly E Gallagher, William M
416 Duck, Philip Graff, and Benjamin J Silk. Correlations and timeliness of COVID-19
417 surveillance data sources and indicators—United States , october 1, 2020–march 22,
418 2023. *Morbidity and Mortality Weekly Report*, 72(19):529, 2023.
- 419 [10] Centers for Disease Control and Prevention. COVID-19 County Check, Year Accessed:
420 2024. Accessed: 4th March 2024.
- 421 [11] Joshua A Salomon and Alyssa Bilinski. Evaluating the performance of Centers for
422 Disease Control and prevention COVID-19 Community Levels as leading indicators of
423 COVID-19 mortality. *Annals of Internal Medicine*, 175(9):1240–1249, 2022.

- 424 [12] Alyssa M Bilinski, Joshua A Salomon, and Laura A Hatfield. Adaptive metrics for
425 an evolving pandemic: A dynamic approach to area-level COVID-19 risk designations.
426 *Proceedings of the National Academy of Sciences*, 120(32):e2302528120, 2023.
- 427 [13] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The ele-*
428 *ments of statistical learning: data mining, inference, and prediction*, volume 2. Springer,
429 2009.
- 430 [14] Andriy Burkov. *The hundred-page machine learning book*, volume 1. Andriy Burkov
431 Quebec City, QC, Canada, 2019.
- 432 [15] The New York Times. Coronavirus (Covid-19) Data in the United States . <https://github.com/nytimes/covid-19-data>, 2021. Retrieved: [Insert Date Here].
433
- 434 [16] U.S. Department of Health and Human Services. COVID-19 Reported Patient
435 Impact and Hospital Capacity Statistics. [https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/anag-cw7u/about_](https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/anag-cw7u/about_data)
436 [data](https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/anag-cw7u/about_data), 2023. Accessed: August 7, 2023.
437
- 438 [17] Diane M Makuc, Bengt Haglund, Deborah D Ingram, Joel C Kleinman, and Jacob J
439 Feldman. Health service areas for the United States . *Vital and Health Statistics. Series*
440 *2, Data Evaluation and Methods Research*, (112):1–102, 1991.
- 441 [18] Centers for Disease Control, Prevention, et al. Indicators for monitoring COVID-19
442 Community Levels and COVID-19 and implementing COVID-19 prevention strategies.
443 *PowerPoint presentation, February, 25, 2022*.
- 444 [19] Seyed M Moghadas, Affan Shoukat, Meagan C Fitzpatrick, Chad R Wells, Pratha Sah,
445 Abhishek Pandey, Jeffrey D Sachs, Zheng Wang, Lauren A Meyers, Burton H Singer,
446 et al. Projecting hospital utilization during the COVID-19 outbreaks in the United
447 States. *Proceedings of the National Academy of Sciences*, 117(16):9122–9126, 2020.
- 448 [20] Reza Yaesoubi, Shiyong You, Qin Xi, Nicolas A Menzies, Ashleigh Tuite, Yonatan H
449 Grad, and Joshua A Salomon. Generating simple classification rules to predict local
450 surges in COVID-19 hospitalizations. *Health Care Management Science*, pages 1–12,
451 2023.
- 452 [21] Gary King and Langche Zeng. Logistic regression in rare events data. *Political analysis*,
453 9(2):137–163, 2001.
- 454 [22] Scikit-learn Developers. Scikit-learn: Machine Learning in Python. https://scikit-learn.org/stable/modules/generated/sklearn.utils.class_weight.compute_class_weight.html, 2021. Accessed: March 18, 2024.
455
456
- 457 [23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel,
458 P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau,
459 M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python.
460 *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- 461 [24] Karel GM Moons, Douglas G Altman, Johannes B Reitsma, John PA Ioannidis, Petra
462 Macaskill, Ewout W Steyerberg, Andrew J Vickers, David F Ransohoff, and Gary S

- 463 Collins. Transparent Reporting of a multivariable prediction model for Individual
464 Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of Internal
465 Medicine*, 162(1):W1–W73, 2015.
- 466 [25] Andrew J Vickers, Ben Van Calster, and Ewout W Steyerberg. Net benefit approaches
467 to the evaluation of prediction models, molecular markers, and diagnostic tests. *bmj*,
468 352, 2016.
- 469 [26] Scott M Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M Prutkin, Bala
470 Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. From local
471 explanations to global understanding with explainable ai for trees. *Nature machine
472 intelligence*, 2(1):56–67, 2020.
- 473 [27] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions.
474 In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and
475 R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–
476 4774. Curran Associates, Inc., 2017.
- 477 [28] Centers for Disease Control and Prevention. COVID Data Tracker. Online; accessed 02
478 November 2023.
- 479 [29] Johns Hopkins University . Johns Hopkins University and Medicine Coronavirus Re-
480 source Center. Online; accessed 02 November 2023.
- 481 [30] USA Facts. US COVID-19 cases and deaths by state . Online; accessed 02 November
482 2023.
- 483 [31] Surya Singh, Mujahed Shaikh, Katharina Hauck, and Marisa Miraldo. Impacts of
484 introducing and lifting nonpharmaceutical interventions on covid-19 daily growth rate
485 and compliance in the united states. *Proceedings of the National Academy of Sciences*,
486 118(12):e2021359118, 2021.
- 487 [32] John M McLaughlin, Farid Khan, Sarah Pugh, David L Swerdlow, and Luis Jodar.
488 County-level vaccination coverage and rates of covid-19 cases and deaths in the united
489 states: An ecological analysis. *The Lancet Regional Health–Americas*, 9, 2022.
- 490 [33] Sylvain Gandon, Margaret J Mackinnon, Sean Nee, and Andrew F Read. Imperfect
491 vaccines and the evolution of pathogen virulence. *Nature*, 414(6865):751–756, 2001.
- 492 [34] Michael Ganslmeier, Davide Furceri, and Jonathan D Ostry. The impact of weather on
493 covid-19 pandemic. *Scientific reports*, 11(1):22027, 2021.
- 494 [35] Lauren M Andersen, Stella R Harden, Margaret M Sugg, Jennifer D Runkle, and Tay-
495 lor E Lundquist. Analyzing the spatial determinants of local covid-19 transmission in
496 the united states. *Science of the Total Environment*, 754:142396, 2021.
- 497 [36] Jacob M Souch and Jeralynn S Cossman. A commentary on rural-urban disparities
498 in covid-19 testing rates per 100,000 and risk factors. *The Journal of Rural Health*,
499 37(1):188, 2021.

- 500 [37] Asma Rashki, Amirhoshang Hoseinpour Dehkordi, Reza Nemati, and Pouya Tavousi.
501 What does covid-19 testing results really say? the real statistics concealed behind the
502 accessible data. *Journal of Medical Virology*, 93(10):5977–5987, 2021.
- 503 [38] Q Nguyen, J Rivera, K Schechtman, R Glassman, and M Mart. Methods for counting
504 covid-19 deaths in us states and territories. *International Journal of Infectious Diseases*,
505 116:S48, 2022.
- 506 [39] Tej A Patel, Bhav Jain, and Julia Raifman. Revamping public health systems: Lessons
507 learned from the tripledemic. *American Journal of Preventive Medicine*, 66(1):185–188,
508 2024.
- 509 [40] Bryant Furlow. Triple-demic overwhelms paediatric units in us hospitals. *The Lancet*
510 *Child & Adolescent Health*, 7(2):86, 2023.