Advances in Artificial Intelligence and Machine Learning; Research 5 (1) 3565-3604 Received 30-10-2024; Accepted 23-03-2025; Published 30-03-2025

Predicting COVID-19 Outcomes Among Albertans With Diabetes and COVID-19: A Machine Learning Approach

Dean T Eurich deurich@ualberta.ca
School of Public Health, University of Alberta
Edmonton, Alberta, Canada

Darren Lau darren.lau@ualberta.ca
Department of Medicine, Faculty of Medicine & Dentistry,
University of Alberta
Edmonton, Alberta, Canada

Weiting Li
OKAKI
Calgary, Alberta, Canada

Olivia R Weaver
School of Public Health, University of Alberta
Edmonton, Alberta, Canada

Tanya Joon
OKAKI
Calgary, Alberta, Canada

Ming Ye
School of Public Health, University of Alberta
Edmonton, Alberta, Canada

Finlay A McAlister
Department of Medicine, Faculty of Medicine & Dentistry, University of Alberta
Edmonton, Alberta, Canada

Padma Kaul
Department of Medicine, Faculty of Medicine & Dentistry, University of Alberta
Edmonton, Alberta, Canada

Salim Samanani
OKAKI
Calgary, Alberta, Canada

Corresponding Author: Dean T. Eurich PhD

Citation: Dean T Eurich, et al. Predicting COVID-19 Outcomes Among Albertans With Diabetes and COVID-19: A Machine Learning Approach. Advances in Artificial Intelligence and Machine Learning. 2025;5(1):204.

Abstract

Background: Certain patients with diabetes and COVID-19 are at high risk of severe outcomes. Identification of risk factors among this group is required to risk-stratify those who may benefit from further surveillance. We aimed to develop machine learning (ML) models predicting severe outcomes among individuals with diabetes and COVID-19 in Alberta, Canada.

Methods: Patients with diabetes and COVID-19 determined by PCR test administered in community and/or emergency department (ED) settings (March 2020-March 2021) were included. Outcomes were ED visit, hospitalization or death for those tested in the community ("Community cohort") and hospitalization or death for those tested in ED ("ED cohort"), and in the combined cohorts ("Community+ED cohort"). Outcomes and features (sociodemographics, drug/healthcare utilization, health history) were identified using healthcare administrative data (2008-2021). Calibration plots, areas under the receiver operating curve, precision-recall curves (AUC, AUPRC), and threshold analyses were used to assess the models.

Results: The Community cohort included 11,247 individuals (1,665 ED visits; 756 hospitalizations; 421 deaths). AUCs for models predicting ED/hospitalization/death were 0.65/0.70/0.93. The AUCs for predicting death in ED (1,495 individuals; 169 deaths) and Community+ED (12,410 individuals; 582 deaths) cohorts were 0.82 and 0.93. Models predicting hospitalization in these cohorts performed poorly and are not reported. Of all models, that predicting death from the Community performed best (sensitivity 0.77, specificity 0.91, positive predictive value 0.26, negative predictive value 0.99), and improved the prediction of death at a 10% risk threshold (compared to the pre-test probability, positive likelihood ratio 9.06 and negative likelihood ratio 0.25).

Conclusion: Identifying diabetes patients at the highest risk of the worst outcomes would assist in triaging patients to ensure appropriate resource use in times of high demand. Overall, the model predicting death among patients with diabetes and COVID-19 in the community could be useful in identifying who requires additional care.

Keywords: COVID-19, Diabetes, Machine learning

Abbreviations:

ATC: Anatomical Therapeutic Classification
AUC: Area Under the Receiver Operating Characteristic Curve
AUPRC: Area Under the Precision Recall Curve
CatBoost: Categorical Boosting
DM: Diabetes Mellitus
ED: Emergency Department
ICD: International Statistical Classification of Diseases and Related Health Problems
LGBoost: Light Gradient Boost
LR-: Negative Likelihood Ratio
LR+: Positive Likelihood Ratio

ML: Machine Learning
NACRS: National Ambulatory Care Reporting System
NPV: Negative Predictive Value
PIN: Pharmaceutical Information Network
PPV: Positive Predictive Value
SHAP: Shapley Additive Explanation
XGBoost: Extreme Gradient Boosting

1. INTRODUCTION

The COVID-19 pandemic placed a considerable burden on the Alberta healthcare system [1]. Even in the current endemic state, COVID-19 remains an important respiratory virus in Canada [2] and Alberta, a province in Canada [3]. There is a continued need to triage high-risk patients presenting with COVID-19 and efficiently allocate limited healthcare resources. One such high-risk group includes individuals with diabetes mellitus who are at greater risk of severe outcomes related to COVID-19 (e.g. hospitalization or death) compared to individuals without diabetes [4]. Individuals with diabetes accounted for 45% of the COVID-19 related deaths in Alberta as of May 2021 [5]. However, what characteristics of diabetes patients put them at high risk is uncertain as not all diabetes patients are seen at risk of severe COVID-19 outcomes. Indeed, simply knowing someone with COVID-19 has diabetes is insufficient. There were 370,535 individuals with prevalent diabetes in Alberta as of May 2021 (~8% of the total population) [6], yet only a subset experienced severe COVID-19 outcomes. Additional approaches are required to risk-stratify diabetes patients who may be at elevated risk and, therefore, may benefit from further surveillance or acute care referral upon testing positive for COVID-19.

Supervised machine learning (ML) uses a subset of outcome-labelled data on which to train models (i.e., generate algorithms that predict a specified outcome by uncovering patterns in the data) [7–9]. In order to avoid over-fitting, the generated algorithms of interest are then tested by predicting the outcome in the remaining portion of the study data and assessing the performance [7, 8]. ML techniques have the capacity to incorporate large numbers of predictors ("features") into prediction algorithms, such as many of those that are routinely collected by Alberta Health [7, 8]. Models such as these have the potential to be published for use by physicians to aid in triaging patients and resources [10, 11].

ML is becoming more common in health sciences research, including COVID-19 and diabetes research. For instance, several studies have predicted the risk of developing COVID-19 [12], diabetes [13] and diabetes-related complications [14, 15]. We conducted a review of relevant literature predicting COVID-19 outcomes (see Supplemental Appendix 1). Of the 30 studies identified, there was only one based on Canadian (Ontario) data, and two specific to diabetes. More than half of the studies were conducted in COVID-19 patients who would likely be considered high-risk and already having a relevant outcome on interest (i.e. already admitted to hospital or ED). A wide range of models (including but not limited to XGBoost, Random Forest, Gradient Boosting Models, etc) were used and resulted in a wide range of AUC's (0.50-1.00) and positive predictive values (0.03-0.99) (See supplemental Appendix 1). Although other studies have used ML to predict adverse COVID-19 outcomes in populations with and without diabetes who have already contracted COVID-19, none have been within the context of Alberta's universal access healthcare system without user fees

at the point of service, or within the context of the routine administrative health data collection specific to Alberta Health [12, 16, 17]. Therefore, the objective of this study was to generate ML models to predict the risk of severe health outcomes in Albertans with diabetes who tested positive for COVID-19.

2. MATERIALS AND METHODS

2.1 Design, Setting and Participants

Data from January 1, 2008 to March 31, 2021 were used to generate a classifier identifying the risk of severe negative COVID-19 related outcomes among Albertans with Diabetes Mellitus (DM) and a positive SARS-CoV-2 PCR test, using supervised ML. Participants residing in Alberta were included if they were over 18 years of age, had a DM diagnosis (identified by a modified version of the National Diabetes Surveillance System definition including any drug dispense for anti-hyperglycemic medication from Alberta's Pharmaceutical Information Network (PIN) dataset, Anatomical Therapeutic Classification (ATC) code A10), and tested positive for SARS-CoV-2 by PCR test between March 1, 2020 and March 1, 2021 [18–20]. This time period, the first year of the COVID-19 pandemic in Alberta, was chosen to avoid secular changes/heterogeneity in outcomes risk, since the COVID-19 vaccine was not yet available, population testing was offered universally and there were no home testing kits, and almost all COVID-19 cases in Alberta were with one (Wuhan wild-type) strain. In the second year of the pandemic, at-home rapid tests for COVID-19 were introduced, likely leading to a substantial number of missed COVID-19 cases and thus, the data was restricted to the first year of the pandemic.

The study population was stratified into three sub-cohorts based on where COVID-19 tests were administered (community vs Emergency Department (ED)) as these settings are clinically relevant use-case scenarios for the derived models. Tests performed in the community setting ("Community cohort") were defined as any tests not performed in ED or hospital settings, where a test performed in the ED or hospital was defined as a COVID-19 test occurring between the start and end date of an ED or hospitalization record. Tests performed in the ED formed the "ED cohort", and the third cohort ("Community + ED cohort") was a combination of the previous two. The Community cohort and ED cohort are both important with regards to identifying patients to refer or admit to hospital. The Community+ED cohort is additionally important since, particularly in rural areas, the ED can act as a secondary source of primary care. Of note, cohorts were not mutually exclusive, and repeat positive COVID-19 tests in the same individual were treated as separate cases if they occurred 3 months or more apart in accordance with Alberta Health's case definition of re-infection.

2.2 Outcome Variables

The outcomes of interest in this study for the Community cohort were COVID-19 related ED visits, hospitalization or death. For Community + ED and ED cohorts, the outcomes of interest were COVID-19 related hospitalization or death. In each cohort setting, the prediction of these outcomes could potentially allow derived models to be used, in addition to clinical judgement, to identify individuals in need of closer observation (e.g.: via referral to a virtual hospital or remote telemonitoring

program) or acute care referral. COVID-19 related ED, hospitalization and death outcomes were identified as ED (National Ambulatory Care Reporting System [NACRS]), discharge abstracts, and vital statistics records within 5 days prior or 30 days after a COVID-19 positive test in any of the three cohorts of interest, bearing an International Statistical Classification of Diseases and Related Health Problems (ICD)-10 diagnostic code of U071 or U072 [21, 22]. We accepted ED visits or hospital admissions for COVID-19 within 5 days prior to the COVID-19 positive test since the confirmatory diagnostic lab testing may be delayed until part-way through an individual's acute care stay; in any event, all outcomes were ultimately diagnosed as COVID-19 as per validated ICD-10-based case definitions, as above.

With respect to the prediction of hospitalization from the ED and Community + ED cohorts, all models performed poorly and would not be beneficial in any clinical scenarios and are therefore not further reported. Outcomes such as ICU admission or need for ventilation were also not included, as initial modelling performed poorly in all cohorts due to insufficient data (e.g.: vital signs).

2.3 Feature Selection

Based on previous literature, we selected 345 features (SUPPLEMENTAL TABLE S1 and SUPPLEMENTAL FIGURE S1) to include in our models, which included social-demographic information, drug and healthcare utilization history and health history to account for both clinical and socio-demographic factors that could contribute to COVID-19 outcomes [10, 11, 15, 16, 23]. SUPPLEMENTAL TABLE S2 contains additional feature definitions, including administrative data definitions for health history.

Socio-demographic features were extracted from the Alberta Health Population Registry (April 1, 2019-March 31, 2021) and included census-derived neighborhood income levels, age, sex, and geographical location (urban vs rural). Additionally, the Pampalon Deprivation Index was used to derive estimated social and material factors which summarize high-level aspects of individuals' living arrangements and means of livelihood [24].

Physician claims, ED and hospital data, which record each interaction with a physician, ED or hospital, respectively, were used to extract features such as the number/type of healthcare visits and the number of healthcare providers 1 year before the positive COVID-19 test date using ICD-9 and ICD-10 codes. Similarly, flags for specific drugs and total drugs dispensed 6 months before the positive COVID-19 test date were identified in the PIN data, using level 4 ATC codes to identify drug classes, except anti-hyperglycemic agents, which used ATC level 5 codes. Alberta benefits from the capture of all drug dispensations at the point-of-sale, regardless of insurance status or age, in its Pharmaceutical Information Network (PIN) dataset.

Biomarkers consistent with COVID-19 and diabetes-related clinical history were also selected as features. Laboratory values were extracted from Alberta Precision Laboratories data between March 1, 2019 and February 29 2020, which maintains records of biological specimen test results in both community and acute settings. Only individuals with more severe health conditions are likely to have extensive lab tests as they progress through the health system. Therefore, laboratory data was only considered up to February 29 2020 in order to avoid selection bias of individuals with the most severe (potentially COVID-19 related) health status. In the case of multiple identical tests, values

closest to the COVID-19 test date were used. Prior occurrence of diabetes events (from physician, ED or hospital data), comorbidities (including Elixhauser conditions [25] and frailty [26] indicators) and ambulatory care sensitive conditions (ACSC) were extracted up to 5 years before an individual's positive COVID-19 test date. *Data Pre-Processing*

Categorical features were one-hot encoded to transform them into a suitable format for machine learning. Missing values in the income data from the Alberta Health Population Registry were imputed using the average income, providing a reasonable estimate for incomplete records. Additionally, lab data was reviewed by clinicians and related lab test categories (such as WBC EST, Instr WBC and WBC) were grouped into a single feature after clinicians reviewed statistical measures like mean and standard deviation This merging of tests eliminated redundant features, enhancing the dataset's clarity and reliability.

2.4 Machine Learning Methods

The likelihood of ED visit, hospitalization, and death in the respective cohorts of interest was predicted using a ML approach as described by Sharma et al, 2021 [10] (scikit-learn package, version 1.2.2) [27]. We utilized the PyCaret library (version 3.0.4) [28] to build, train and compare models, and select the best at the end of this pipeline. The training sets for the PyCaret experiments consisted of 80% of the cohort, with 20% held for validation (FIGURE 1).

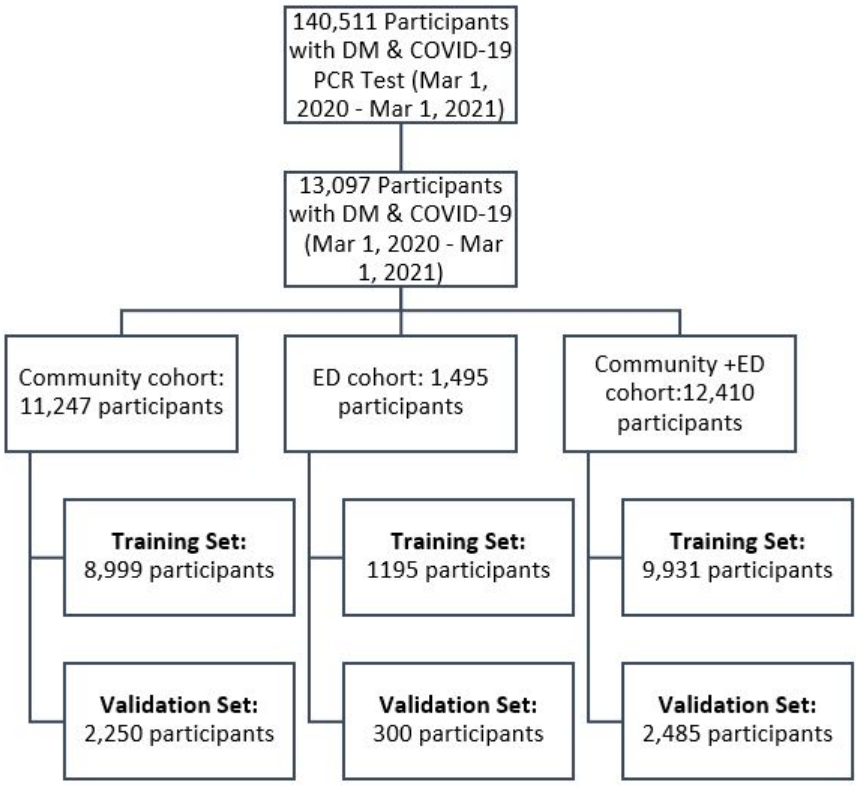


Figure 1: Participant selection and analysis.

Fifteen ML algorithms were evaluated, including tree classifiers, gradient boost machine learning models such as CatBoost classifier, Light Gradient Boosting Model, AdaBoost classifier, Extreme Gradient Boost model, Random Forest, etc. Each model was trained and cross validated on 5 random splits of the training data. Out of the 5 random subsets, 4 of them were used during training while the last was saved for cross-validation, which delivers more robust results than if this process was performed on a single combination of training and tuning sets. Hyperparameters (e.g. learning rate, maximum depth, number of estimators, etc) were optimized through an iterative process of tuning. PyCaret's "tune_model" function and "Randomized Search" function from the scikit_learn package were utilized to iteratively search over a predefined grid of potential values, progressively adjusting hyperparameters to identify the best configuration for each model.

Top performing models were selected based on the value of the area under the receiver operating characteristic curve (AUC) (SUPPLEMENTAL TABLE S3). AUC was selected as it provides a robust measure of overall model performance. For predicting ED, hospitalization and death outcomes in the Community cohort, these included Extreme Gradient Boosting (XGBoost) [29], Light Gradient Boost (LGBoost) [30], and Categorical Boosting (CatBoost) [31, 32] models respectively. As a sensitivity analysis, the base XGBoost and LGBoost models for predicting ED and hospitalization were calibrated using Spline [33] and Sigmoid calibration techniques in an attempt to improve performance. Results were not greatly affected despite improved calibration, therefore only results from base models are presented. LGBM and CatBoost were chosen for predicting death in ED and Community + ED cohorts, respectively.

2.5 Statistical Analysis

Participant characteristics were reported descriptively as means (SD) and counts (%) and compared between training and validation sets using t-tests and two-sample tests of proportion for continuous and categorical variables, respectively (STATA/IC, version 15.1). ML models were assessed using AUC and area under the precision recall curve (AUPRC). Shapley Additive Explanation (SHAP) values represent each feature's responsibility for a given predicted risk, compared to a baseline level of that feature, and can be used to characterize the importance of a given feature for a given prediction. The sum of SHAP values is equal to the model-predicted risk [34]. SHAP values were visualized, in aggregate, using plots of absolute values, and beeswarm plots. Model calibration was assessed visually through plots of actual vs predicted risk, and the clinical utility of the most discriminative models was assessed using a threshold analysis for different risk levels of the specified outcome. For the threshold analysis, we chose relatively low thresholds (e.g.: individuals with model-predicted outcome risk >= 10% as "test positive") to assess the clinical utility of models, to ensure we classified as many people as possible with a severe outcome as "positive".

3. RESULTS

3.1 Participant Characteristics

Of 372,055 individuals with DM, 140,511 received at least one PCR test between March 1, 2020 and March 1, 2021. The total study population included 13,097 unique individuals among these who

had a positive COVID-19 test result, and were therefore eligible for inclusion in one or more of the three cohorts (FIGURE 1). In all cohorts, individuals in the training and validation sets generally had similar characteristics, as expected. Small differences were observed in proportions of several Elixhauser comorbidities in each of the cohorts, and in the average minimum drug dispense quantity, for the ED cohort (SUPPLEMENTAL TABLE S4a and SUPPLEMENTAL TABLE S4b).

3.2 Community Cohort Model Performances

The Community cohort consisted of 11,255 positive COVID-19 cases (11,247 individuals) leading to 1,665 ED visits (mean 10.09 ± 7.63 days from testing), 756 hospitalizations (mean 9.33 ± 6.66 days from testing) and 421 deaths (mean 12.10 ± 7.12 days from testing) (FIGURE 1).

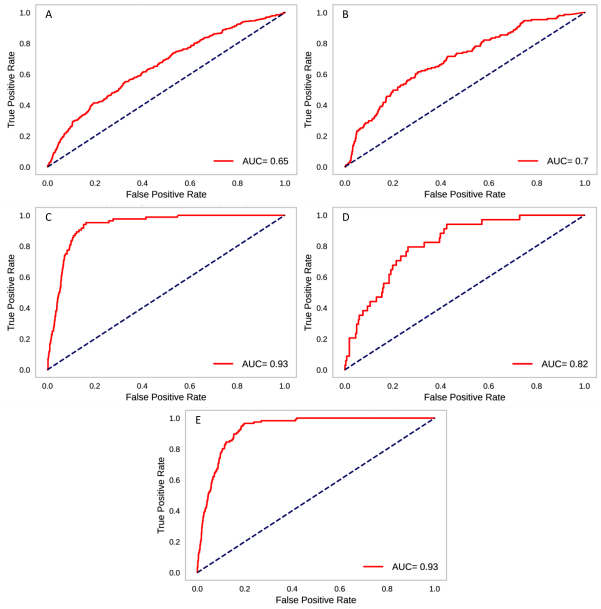


Figure 2: Test receiver operating characteristic curves for all models. A) Predicting ED visits in the Community cohort B) Predicting hospitalization in the Community cohort; C) Predicting death in the Community cohort; D) Predicting death in the ED cohort; E) Predicting death in the Community+ED cohort.

The model predicting ED visits had only moderate discriminatory ability (AUC 0.65) (FIGURE 2a). The precision (positive predictive value) of the model at the 0.1 risk threshold remained relatively constant as recall (sensitivity) increased, leading to a low AUPRC value (0.25) (SUPPLEMENTAL FIGURE S2a). This is expected in datasets with imbalanced data due to rare outcomes like ours, highlighting the trade-off between sensitivity and specificity. However, the model was not well-calibrated based on visual inspection (SUPPLEMENTAL FIGURE S3a) and did not merit further threshold or feature importance analysis.

The model predicting hospitalization also had moderate discriminatory ability (AUC 0.70) (FIGURE 2b) and small AUPRC (0.14, again due to the small number of outcomes) (SUPPLEMENTAL

FIGURE S2b). Based on visual inspection, the model was moderately calibrated to a risk threshold of about 0.2 (SUPPLEMENTAL FIGURE S3b), indicating that for most individuals in this cohort the models were moderately accurate, since the risk of COVID-related hospitalization in community cases was below this 0.2 threshold in almost all cases. At the 0.1 risk threshold, the model had low sensitivity (0.27, indicating accurate prediction of only 27% of all hospitalizations that occurred) and positive predictive value (PPV 0.19; only 19% of hospitalizations predicted actually occurred). The model also had high specificity (0.92; accurate prediction of 92% of all those who were not hospitalized) and negative predictive value (NPV 0.95; 95% of those predicted to not be hospitalized turned out to be true). The post-test probability was minimally changed as indicated by small likelihood ratios (LR+ 3.35; LR- 0.79) (TABLE 1) and it therefore provides minimal prognostic utility. Age, historical number of drug dispenses and average drug dispense quantity had the greatest impact on model performance (SUPPLEMENTAL FIGURE S4a). Higher values were associated with higher risk of hospitalization except for historical number of drug dispenses where very high or low values were associated with increased risk (SUPPLEMENTAL FIGURE S5a).

FN, false negative; FNR, false negative rate; FP, false positive; FPR, false positive rate; inf, infinite; LR+, positive likelihood ratio; LR-, negative likelihood ratio; NaN, undefined result; NPV, negative predictive value; PPV, positive predictive value; TN, true negative; TNR, true negative rate; TP, true positive; TPR, true positive rate.

The AUC for the model predicting death indicated good discriminatory ability (0.93) (FIGURE 2c), and, similar to other models, a small AUPRC (0.28) (SUPPLEMENTAL FIGURE S2c). The model showed good calibration for risk levels of death up to approximately 0.4-0.5, while generally overestimating risk at higher thresholds (SUPPLEMENTAL FIGURE S3c). This means that for most individuals in this cohort, the models were accurate, since the risk of COVID-related death in community cases was generally below 0.2. The sensitivity of the model at the 0.1 risk threshold was moderate (0.77) and the specificity was high (0.91). The model had high PPV (0.26) and NPV (0.99) and positive likelihood ratio (LR+, 9.06). The LR+ indicates that the post-test probability was substantially increased (from 3.74% to approximately 84%) while the small LR- (0.25) indicates no substantial decrease in post-test probability (TABLE 1) when tests are negative. Estimated material factor was the most important feature, followed by age and historical number of drug dispenses (SUPPLEMENTAL FIGURE S4b). Greater values of age and lower values of estimated material factor (i.e. less deprived) and historical drug dispenses were generally associated with a greater likelihood of death (SUPPLEMENTAL FIGURE S5b).

3.3 ED Cohort Model Performance

The ED cohort consisted of 1,497 COVID-19 cases (1,495 individuals), and 169 deaths (FIGURE 1) occurred (mean 11.50 ± 7.48 days from testing). The model exhibited good discriminatory ability for predicting death (AUC 0.82) (FIGURE 2d), low AUPRC (0.36) (SUPPLEMENTAL FIGURE S2d), and moderate performance. The model was well-calibrated to a risk threshold of approximately 0.4, indicating accurate predictions for the majority of the study cohort which had risk levels below this threshold (SUPPLEMENTAL FIGURE S3d). Sensitivity (0.74) and specificity (0.76) were lower in this model at the 0.1 threshold than in the model predicting death in the Community cohort, and the PPV and NPV were similar (0.28 and 0.96 respectively). The LR+ and LR- showed minimal changes in post-test probabilities, with LR+ at 3.06 and LR- at 0.35 (TABLE 1), and it therefore

Table 1: Threshold analysis for all models

Threshold	TP	FN	FP	TN	TPR	TNR	FNR	FPR	PPV	NPV	Sensitivity	Specificity	LR+	LR-
Predicting hospitalization in Community cohort														
0	151	0	2100	0	1.00	0.00	0.00	1.00	0.07	NaN	1.00	0.00	1.00	NaN
0.05	128	23	1358	742	0.85	0.35	0.15	0.65	0.09	0.97	0.85	0.35	1.31	0.43
0.1	41	110	170	1930	0.27	0.92	0.73	0.08	0.19	0.95	0.27	0.92	3.35	0.79
0.15	4	147	37	2063	0.03	0.98	0.97	0.02	0.10	0.93	0.03	0.98	1.50	0.99
0.2	0	151	1	2099	0.00	1.00	1.00	0.00	0.00	0.93	0.00	1.00	0.00	1.00
0.3	0	151	0	2100	0.00	1.00	1.00	0.00	NaN	0.93	0.00	1.00	NaN	1.00
0.4	0	151	0	2100	0.00	1.00	1.00	0.00	NaN	0.93	0.00	1.00	NaN	1.00
0.5	0	151	0	2100	0.00	1.00	1.00	0.00	NaN	0.93	0.00	1.00	NaN	1.00
0.6	0	151	0	2100	0.00	1.00	1.00	0.00	NaN	0.93	0.00	1.00	NaN	1.00
0.7	0	151	0	2100	0.00	1.00	1.00	0.00	NaN	0.93	0.00	1.00	NaN	1.00
0.8	0	151	0	2100	0.00	1.00	1.00	0.00	NaN	0.93	0.00	1.00	NaN	1.00
0.9	0	151	0	2100	0.00	1.00	1.00	0.00	NaN	0.93	0.00	1.00	NaN	1.00
0.99	0	151	0	2100	0.00	1.00	1.00	0.00	NaN	0.93	0.00	1.00	NaN	1.00
Predicting death in Community cohort														
0	84	0	2167	0	1.00	0.00	0.00	1.00	0.04	NaN	1.00	0.00	1.00	NaN
0.05	74	10	269	1898	0.88	0.88	0.12	0.12	0.22	0.99	0.88	0.88	7.10	0.14
0.1	65	19	185	1982	0.77	0.91	0.23	0.09	0.26	0.99	0.77	0.91	9.06	0.25
0.15	55	29	143	2024	0.65	0.93	0.35	0.07	0.28	0.99	0.65	0.93	9.92	0.37
0.2	43	41	110	2057	0.51	0.95	0.49	0.05	0.28	0.98	0.51	0.95	10.08	0.51
0.3	25	59	61	2106	0.30	0.97	0.70	0.03	0.29	0.97	0.30	0.97	10.57	0.72
0.4	14	70	27	2140	0.17	0.99	0.83	0.01	0.34	0.97	0.17	0.99	13.38	0.84
0.5	6	78	9	2158	0.07	1.00	0.93	0.00	0.40	0.97	0.07	1.00	17.20	0.93
0.6	1	83	5	2162	0.01	1.00	0.99	0.00	0.17	0.96	0.01	1.00	5.16	0.99
0.7	0	84	3	2164	0.00	1.00	1.00	0.00	0.00	0.96	0.00	1.00	0.00	1.00
0.8	0	84	0	2167	0.00	1.00	1.00	0	NaN	0.96	0.00	1.00	NaN	1.00
0.9	0	84	0	2167	0.00	1.00	1.00	0	NaN	0.96	0.00	1.00	NaN	1.00
0.99	0	84	0	2167	0.00	1.00	1.00	0	NaN	0.96	0.00	1.00	NaN	1.00
Predicting death in ED cohort														
0	34	0	266	0	1.00	0.00	0.00	1.00	0.11	NaN	1.00	0.00	1.00	NaN
0.05	27	7	85	181	0.79	0.68	0.21	0.32	0.24	0.96	0.79	0.68	2.49	0.30
0.1	25	9	64	202	0.74	0.76	0.26	0.24	0.28	0.96	0.74	0.76	3.06	0.35
0.15	19	15	47	219	0.56	0.82	0.44	0.18	0.29	0.94	0.56	0.82	3.16	0.54
0.2	17	17	42	224	0.50	0.84	0.50	0.16	0.29	0.93	0.50	0.84	3.17	0.59
0.3	14	20	26	240	0.41	0.90	0.59	0.10	0.35	0.92	0.41	0.90	4.21	0.65
0.4	11	23	15	251	0.32	0.94	0.68	0.06	0.42	0.92	0.32	0.94	5.74	0.72
0.5	7	27	9	257	0.21	0.97	0.79	0.03	0.44	0.90	0.21	0.97	6.08	0.82
0.6	5	29	5	261	0.15	0.98	0.85	0.02	0.50	0.90	0.15	0.98	7.82	0.87
0.7	2	32	2	264	0.06	0.99	0.94	0.01	0.50	0.89	0.06	0.99	7.82	0.95
0.8	1	33	1	265	0.03	1.00	0.97	0.00	0.50	0.89	0.03	1.00	7.82	0.97
0.9	0	34	0	266	0.00	1.00	1.00	0.00	NaN	0.89	0.00	1.00	NaN	1.00
0.99	0	34	0	266	0.00	1.00	1.00	0.00	NaN	0.89	0.00	1.00	NaN	1.00

continued..

provides minimal clinical utility. Age, sex and average drug dispense quantity were the primary features contributing to model performance, with higher values of age and drug dispense quantity,

Table 1: Continued..

Threshold	TP	FN	FP	TN	TPR	TNR	FNR	FPR	PPV	NPV	Sensitivity	Specificity	LR+	LR-
Predicting death in Community + ED cohort														
0	116	0	2369	0	1.00	0.00	0.00	1.00	0.05	NaN	1.00	0.00	1.00	NaN
0.05	100	16	351	2018	0.86	0.85	0.14	0.15	0.22	0.99	0.86	0.85	5.82	0.16
0.1	92	24	247	2122	0.79	0.90	0.21	0.10	0.27	0.99	0.79	0.90	7.61	0.23
0.15	76	40	184	2185	0.66	0.92	0.34	0.08	0.29	0.98	0.66	0.92	8.44	0.37
0.2	66	50	139	2230	0.57	0.94	0.43	0.06	0.32	0.98	0.57	0.94	9.70	0.46
0.3	46	70	75	2294	0.40	0.97	0.60	0.03	0.38	0.97	0.40	0.97	12.53	0.62
0.4	24	92	34	2335	0.21	0.99	0.79	0.01	0.41	0.96	0.21	0.99	14.42	0.80
0.5	14	102	12	2357	0.12	0.99	0.88	0.01	0.54	0.96	0.12	0.99	23.83	0.88
0.6	6	110	4	2365	0.05	1.00	0.95	0.00	0.60	0.96	0.05	1.00	30.63	0.95
0.7	3	113	0	2369	0.03	1.00	0.97	0.00	1.00	0.95	0.03	1.00	inf	0.97
0.8	2	114	0	2369	0.02	1.00	0.98	0.00	1.00	0.95	0.02	1.00	inf	0.98
0.9	0	116	0	2369	0.00	1.00	1.00	0.00	NaN	0.95	0.00	1.00	NaN	1.00
0.99	0	116	0	2369	0.00	1.00	1.00	0.00	NaN	0.95	0.00	1.00	NaN	1.00

and female sex, being associated with greater risk of death (SUPPLEMENTAL FIGURE S4c and SUPPLEMENTAL FIGURE S5c).

3.4 Community + ED Cohort Model Performance

Lastly, there were 12,421 positive COVID-19 cases (12,410 individuals) in the Community + ED cohort. The cohort experienced 582 COVID-related deaths (mean 11.80 ± 7.21 days from testing) (FIGURE 1). The AUC for this model predicting death showed strong discriminatory ability at 0.93 and had a low AUPRC (0.37), similar to the other models (FIGURE 2e and SUPPLEMENTAL FIGURE S2e). Similar to the model predicting death in the Community cohort, this model also had good calibration to a risk level of 0.4 (SUPPLEMENTAL FIGURE S3e) and performed well, exhibiting moderate sensitivity at 0.79 and high specificity at 0.90, with similar PPV and NPV (PPV 0.27; NPV 0.99). The LR+ showed favorable shifts in post-test probabilities (7.61) while the LR- indicated no change (0.23) (TABLE 1). Notably, these values derived from the Community + ED cohort were likely driven by the proportionally larger Community cohort. Age, sex, and number of physician visits contributed the most to this model, with higher values of age and physician visits, and male sex, contributing to higher risk of death (SUPPLEMENTAL FIGURE S4d and SUPPLEMENTAL FIGURE S5d).

4. DISCUSSION

We trained and validated a suite of ML models to predict the risk of severe outcomes (ED, hospitalization and death) in adults with diabetes testing positive for COVID-19 in the community or in the ED. Most models showed good discrimination and were moderately predictive with the exception of the prediction of hospitalization in the ED and Community + ED cohorts where models performed poorly and were not further reported. In this case, it is likely that by the time someone with COVID-19 was sick enough to present to ED, the decision to admit to hospital was largely

based on hypoxemia and other vital signs or functional / clinical status indicators, for which we did not have data. Our models are better oriented towards understanding an individual's underlying risk of poor clinical outcomes based on their pre-morbid characteristics or laboratory parameters.

In contrast, the model for predicting death following a community case of COVID-19 was highly predictive. At a 10% risk threshold, this model's positive likelihood ratio was substantial, and it correctly predicted over 3 quarters of death-cases (sensitivity 77%), with high negative predictive value (99%). With a positive predictive value of 26%, this model can identify (i.e.: "rule in") a group of individuals in which *a quarter will die in hospital without additional intervention*, based solely on pre-morbid characteristics. The high negative predictive value is re-assuring, but, given the dire consequences of missing an individual at high risk of dying (false negatives in 19 of 2,001 individuals), our suggested use of this model is not to rule out severe illness.

The predictive ability of models in the Community cohort was much lower for ED visits and hospitalization outcomes, than for death. Unlike death, ED visits and hospital admission are ultimately determined by the availability of hospital services and capacity, and the behaviour and decision-making of patients and healthcare professionals. This decision-making process may be more subjective and rely heavily on features unavailable to us, e.g.: hypoxemia in the ED, or patient/family concerns about COVID-19. Alternatively, due to the disruption to health services during the pandemic, it is possible that rapidly shifting care practices translated to features with too inconsistent a relationship with ED visits and hospital admissions for these models to be accurate.

Previously, two studies developed ML models predicting deaths among patients with COVID-19 and diabetes who were admitted to hospital [16, 35]. All models assessed had high accuracy (80-87%) and sensitivity (75-88%), and moderate-high specificity (55-95%). These models used features from the time of hospital admission and/or duration of the hospital stay which are not available at the time of community or ED assessments., Relatively small sample sizes and lack of assessment of model calibration additionally makes it difficult to discern how these models would perform in a real-world clinical setting. While these models are intended for patients already admitted to hospital, ours have the benefit of catching patients at an earlierpoint in the COVID-19 trajectory, where anticipatory interventions may make a bigger difference on outcomes. Among studies of non-hospitalized COVID-19 patients (none specific to adults with diabetes), AUC's range from 0.62 to 0.84 for predicting hospitalization (compared to 0.70 in our own study) and from 0.61 to 0.91 for predicting death (compared to 0.93 in the present study) (supplemental appendix 1).

We specifically envision implementing our model predicting death among community adults with diabetes and COVID-19 positivity at community practitioners' offices, urgent care centres, rapid screening sites, and in rural or remote centres. Positive COVID results in these settings could be accompanied by an automated report of model-predicted risk where a sufficiently integrated data environment exists. Alternatively, in less integrative data environments, providers may refer to an online repository of ML models and enter data manually for the salient features, just as most prediction algorithms are accessed presently (e.g.: https://www.mdcalc.com). If predicted risk is >= 10%, the following recommendations would be provided: refer the individual to the ED if in a non-acute community setting; transfer the individual to a tertiary care hospital if in a rural or remote acute or urgent care environment; consider hospital admission and close observation contingent on clinical parameters for individuals transferred to or being seen in a tertiary acute care ED; prioritize this patient for receipt of COVID-19-targeted therapies. For those predicted to be at low (< 10%) risk, routine clinical judgement and follow-up should apply.

The challenges of implementing such a model primarily relate to the integration of prediction models in busy clinical workflows. An integrative data environment, where running the prediction model can be automated, would be ideal. Prediction models would need to be updated continuously and rapidly, to maintain value in the face of viral evolution. Finally, the practical points of rolling out decision support tools such as this, in diverse environments, would need to be considered, including processes for enlisting providers and evaluating their responses to the predictions and recommendations. While we have shown that good predictions are technically possible and may have a marked impact for adults with diabetes and incident COVID-19 positivity, the challenge for health systems in the digital era will be to translate complex and powerful machine learning models to actual practice change.Our application of a more proactive approach to prognostication is one of the strengths of the present study, along with the use of routinely collected administrative data which does not impose any additional burden of data collection on physicians to use these models. This study does have limitations. There was a small number of participants and outcomes in the ED cohort relative to the Community cohort which likely biased results to be more favorable for the combined cohort. We were also limited to data that is routinely collected by Alberta Health, and while extensive and generally accurate, may not include relevant clinical or functional / frailty-related variables, such as vital signs in the ED. Results from this study may not be generalizable to other countries or Canadian provinces with different healthcare systems, and are specific to adults with diabetes. Other health systems may collect different routine administrative health data than the variables used in our models. Models for other jurisdictions would need to be validated in other populations. Finally, we only had data for Wuhan virulence COVID-19. Other strains may have different risks of severe outcomes and the model therefore may not translate into other time periods or be applicable to the endemic phase of predominantly Omicron variants, therefore also impacting the generalizability of the model and requiring further validation.

5. CONCLUSION

In conclusion, our ML model predicting COVID-19 related death among individuals testing positive for COVID-19 in the community setting appears to be clinically useful, as a supplement to clinical assessment, for identifying a subset of individuals with diabetes in whom over a quarter will die without intervention. Such individuals would benefit from acute care referral, prioritization for prophylactic COVID-19-targetted therapies, and other enhanced observation / preventive measures to mitigate their high mortality risk. These models can also be used on an anticipatory basis, to assist with risk counselling, e.g.: regarding COVID-19 vaccinations. Future research could expand upon cohorts tested in other settings such as in-hospital, and validating the current models in other populations and with other strains of COVID-19 variants. While good predictions are technically possible and may have a marked impact for adults with diabetes and COVID-19, the translation of complex and powerful machine learning models to actual practice change remains a health system challenge.

6. ACKNOWLEDGEMENT

Ethics Approval

This study received ethics approval from the University of Alberta Research Ethics Board (Pro00097641).

Funding:

This study was supported by the University Hospital Foundation (UHF7064).

Author's Contributions:

DTE, SS, FAM and PK contributed to the conception and design. DTE, SS contributed to data acquisition and analysis. DL and MY contributed to conception, design, analysis and interpretation of results. WL and TJ analyzed the data. ORW and WL wrote the first draft of the manuscript. All authors contributed to reviewing, interpretation of results, editing and approving the final version of the manuscript.

References

[1] https://www.cbc.ca/news/canada/calgary/respiratory-viruses-surge-alberta-hospitals-overwhelmed-1.7056500

[2] https://health-infobase.canada.ca/covid-19/current-situation.html#a4

[3] https://www.alberta.ca/stats/dashboard/respiratory-virus-dashboard.htm?data=historical-data#historical-data

[4] Apicella M, Campopiano MC, Mantuano M, Mazoni L, Coppelli A, et al. COVID-19 in People With Diabetes: Understanding the Reasons for Worse Outcomes. Lancet Diabetes Endocrinol. Sep 2020;8:782-792.

[5] https://mcc-covid.crc.pitt.edu/COVID19_official_websites/Canada/alberta_case_count/index_2021-05-12_082304.html#comorbidities

[6] https://health-infobase.canada.ca/ccdss/data-tool/

[7] Baştanlar Y, Ozuysal M. Introduction to Machine Learning. In: Yousef M, Allmer J, editors. miRNomics: MicroRNA Biology and Computational Analysis. Methods in Molecular Biology. Springer. 2014;1107:105-128.

[8] Yusuf M, Atal I, Li J, Smith P, Ravaud P, et al. Reporting Quality of Studies Using Machine Learning Models for Medical Diagnosis: A Systematic Review. BMJ Open. 2020;10:e034568.

[9] Liu Y, Chen PC, Krause J, Peng L. How to Read Articles That Use Machine Learning: User's Guides to the Medical Literature. JAMA. 2019;322:1806-1816.

[10] Sharma V, Kulkarni V, Eurich DT, Kumar L, Samanani S. Safe Opioid Prescribing: A Prognostic Machine Learning Approach to Predicting 30-Day Risk After an Opioid Dispensation in Alberta Canada. BMJ Open. 2021;11:e043964.

[11] Sharma V, Kulkarni V, Jess E, Gilani F, Eurich D, et al. Development and Validation of a Machine Learning Model to Estimate Risk of Adverse Outcomes Within 30 Days of Opioid Dispensation. JAMA Netw Open. 2022;5:e2248559.

[12] Willette AA, Willette SA, Wang Q, Pappas C, Klinedinst BS, et al. Using Machine Learning to Predict COVID-19 Infection and Severity Risk Among 4510 Aged Adults: A UK Biobank Cohort Study. Sci Rep. 2022;12:7736.

[13] Ravaut M, Harish V, Sadeghi H, Leung KK, Volkovs M, et al. Development and Validation of a Machine Learning Model Using Administrative Health Data to Predict Onset of Type 2 Diabetes. JAMA Netw Open. 2021;4:e2111315.

[14] Nicolucci A, Romeo L, Bernardini M, Vespasiani M, Rossi MC, et al. Prediction of Complications of Type 2 Diabetes: A Machine Learning Approach. Diabetes Res Clin Pract. 2022;190:110013.

[15] Ravaut M, Sadeghi H, Leung KK, Volkovs M, Kornas K, et al. Predicting Adverse Outcomes Due to Diabetes Complications With Machine Learning Using Administrative Health Data. NPJ Digit Med. 2021;4:24.

[16] Khadem H, Nemat H, Eissa MR, Elliott J, Benaissa M. COVID-19 Mortality Risk Assessments for Individuals With and Without Diabetes Mellitus: Machine Learning Models Integrated With Interpretation Framework. Comput Biol Med. 2022;144:105361.

[17] Gao Y, Cai GY, Fang W, Li HY, Wang SY, et al. Machine Learning Based Early Warning System Enables Accurate Mortality Risk Prediction for COVID-19. Nat Commun. 2020;11:5033.

[18] LeBlanc AG, Jun Gao YJ, McRae L, Pelletier C. At-A-Glance – Twenty Years of Diabetes Surveillance Using the Canadian Chronic Disease Surveillance System. Health Promot Chronic Dis Prev Can. 2019;39:306-309.

[19] Lipscombe LL, Hwee J, Webster L, Shah BR, Booth GL, et al. Identifying Diabetes Cases From Administrative Data: A Population-Based Validation Study. BMC Health Serv Res. 2018;18:316.

[20] Mathe N, Ryan A, Cook A, Sargious P, Senior P, et al. Enhancing Diabetes Surveillance Across Alberta by Adding Laboratory and Pharmacy Data to the National Diabetes Surveillance System Methods. Can J Diabetes. 2022;46:375-380.

[21] Kadri SS, Gundrum J, Warner S, Cao Z, Babiker A, et al. Uptake and Accuracy of the Diagnosis Code for COVID-19 Among Us Hospitalizations. JAMA. 2020;324:2553-2554.

[22] Wu G, D'Souza AG, Quan H, Southern DA, Youngson E, et al. Validity of Icd-10 Codes for COVID-19 Patients With Hospital Admissions or Ed Visits in Canada: A Retrospective Cohort Study. BMJ Open. 2022;12:e057838.

[23] Sharma V, Kulkarni V, Joon T, Eurich DT, Simpson SH, et al. Predicting Falls-Related Admissions in Older Adults in Alberta Canada: A Machine-Learning Falls Prevention Tool Developed Using Population Administrative Health Data. BMJ Open. 2023;13:e071321.

[24] https://www.arcgis.com/home/item.html?id=29198abead1b42b18ae5e5d156dafa4a

[25] Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, et al. Coding Algorithms for Defining Comorbidities in Icd-9-CM and Icd-10 Administrative Data. Med Care. 2005;43:1130-1139.

[26] Urquhart R, Giguere AM, Lawson B, Kendell C, Holroyd-Leduc JM, et al. Rules to Identify Persons With Frailty in Administrative Health Databases. Can J Aging. 2017;36:514-521.

[27] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, et al. Scikit-Learn: Machine Learning in Python. J Mach Learn Res. 2011;12:2825-2830.

[28] https://pycaret.org/

[29] Chen T, Guestrin C. XGBOOST: A Scalable Tree Boosting System. Computing Research Repository. 2016. Arxiv preprint: https://arxiv.org/pdf/1603.02754v1

[30] Ke G, Meng Q, Finley T, Wang T, Chen W, et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Advances in neural information processing systems. 2017;30.

[31] Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. Catboost: Unbiased Boosting With Categorical Features. Computing Research Repository. 2019. Arxiv preprint: https://arxiv.org/pdf/1706.09516

[32] Dorogush AV, Ershov V, Gulin A. CatBOOST: Gradient Boosting With Categorical Features Support. Computing Research Repository. 2018. Arxiv preprint: https://arxiv.org/pdf/1810.11363.

[33] Lucena B. Spline-Based Probability Calibration. 2018. Arxiv preprint: https://arxiv.org/pdf/1809.07751

[34] Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. 2017. Arxiv preprint: https://arxiv.org/pdf/1705.07874v1

[35] Khodabakhsh P, Asadnia A, Moghaddam AS, Khademi M, Shakiba M, et al. Prediction of In-Hospital Mortality Rate in COVID-19 Patients With Diabetes Mellitus Using Machine Learning Methods. J Diabetes Metab Disord. 2023;22:1-14.

Supplemental Data

Supplemental Table S1: List of features

Feature Category	Description
Demographics	Age, sex, income, estimated material factor, estimated social factor
Drug utilization	Polypharmacy, number of dispenses, total quantity of historical dispenses, maximum quantity of historical dispenses, minimum quantity of historical dispenses, average quantity of historical dispenses, historical number of dispenses, individual drugs including: N02B, A10BA, A10B, M05B, C08C, C09A, A11C, H03A, N05C, A02B, C03B, C10A, N06A, A10AE, A10A, A10AB, C03C, C08D, R03B, A10BB, R03A, C09C, B03B, N03A, L04A, J07B, N05B, R06A, R01A, G03C, A10BK, M03A, A10BJ, M01A, J05A, M04A, G04C, S01E, S01B, A02A, N02A, A10BD, N05A, A12A, A03B, S01F, G04B, C09B, C07A, J01X, J01D, J01A, J01M, D06B, A10AD, D08A, B03A, A01A, D07A, S02C, D01A, D06A, A06A, C09D, A10AC, A11D, A05A, J01C, J01F, L02A, L02B, N04B, A10BH, J02A, A04A, A03A, J01E, N07C, D10A, P01A, C03A, D03B, G03A, N07B, C01C, V03A, A12C, G03B, C03D, M03B, C05A, R03D, S01X, H02B, D02A, M02A, A10BX, H02A, N02C, P03A, D01B, S01A, P01B, H05A, H01C, A10BG, C03E, R05D, D11A, A09A, A07D, N06D, C01D, N07A, A07A, J07A, R05F, N01B, B05C, L01X, G03D, A07E, N06B, G01A, A12B, C01A, H03B, H04A, A03F, G02B, D07C, R05C, S01C, C02C, C02A, L01B, D05B, D05A, S01G, B02A, B03X, A11G, G03F, D04A, S03A, C04A, C02D, S01, L01C, G03H, H01B, L03A, N04A, P02C, D02B, G02C, R05X, V04C, C02K, N01A, C01B, A08A, B05X, S02D, S02B, S01L, G03G, J04B, J04A, L01A, L01D, A16A, H05B, C01E, A07B, G03X, A11H, A07C, D10B, A11J, A10BF, A03C, S03C, V07A, R03C, D07X, R01B, A11E, R02A, A11A, N07X, H01A, C07C, M09A, V06D, B01A
Healthcare utilization	Number of physician visits, number of physicians, number of emergency department visits, number of hospitalizations, number of days in acute care
Historic diabetes events	Hypoglycemia, cardiovascular diseases, retinopathy, amputation, debridement, advanced foot infection, mild foot infection, tissue infection
Ambulatory Care Sensitive Conditions	Hypertension, diabetes, angina, chronic obstructive pulmonary disease, asthma, heart failure/pulmonary edema, grand mal status and other epileptic convulsions,
Elixhauser comorbidities	Neurologic disorders, hypertension, diabetes, hypothyroidism, obesity, drug abuse, arrhythmia, chronic pulmonary disease, rheumatoid arthritis, deficiency anemia, alcohol abuse, depression, psychoses, renal failure, peptic ulcer disease, fluid and electrolyte disorders, solid tumor cancer, liver disease, peripheral vascular disorders, valvular disease, weight loss, coagulopathy, blood loss anemia, pulmonary circulation disorders, paralysis, HIV/AIDS, lymphoma, congestive heart failure
Frailty indicators	Falls, malaise & fatigue/debility, delirium, dementia, stroke, senility without mention of psychosis, pressure ulcer, vascular dementia, dementia in other diseases classified elsewhere, urinary incontinence, other cerebral degenerations incl. Alzheimer's, abnormal weight loss, gait abnormality, dementia in Alzheimer's, muscle weakness, muscular wasting and disuse atrophy, difficulty walking, fecal incontinence, cerebral generations usually manifest in childhood
Clinical history	eGFR, creatinine, Hemoglobin A1C, lymphocyte, alanine transferase, hematocrit, white blood count, platelet count, neutrophil, red blood count, hemoglobin, potassium, sodium, high-density lipoprotein cholesterol, thyroid stimulating hormone, urate, calcium, low-density lipoprotein cholesterol, fasting glucose, alkaline phosphatase, albumin:creatinine, total cholesterol, albumin, total bilirubin, gamma-glutamyl transferase, triglycerides, aspartate aminotransferase, non-high-density lipoprotein cholesterol, chloride, triiodothyronine, thyroxine, prothrombin time international normalized ratio, urea, activated partial thromboplastin time, lactate dehydrogenase, anion gap, bicarbon, magnesium, reticulocytes, ferritin, total iron-binding capacity, B-type natriuretic peptide, creatine kinase, vitamin B12, iron, polymerase chain reaction, C-reactive protein, N-terminal prohormone of brain natriuretic peptide, D-dimer, arterial partial pressure of O2, partial pressure of CO2, ionized calcium, arterial blood pH, fibrinogen, lactate, folate, erythrocyte sedimentation rate, PaO2:FiO2, arterial saturation of O2
Other	Hyperlipidemia, smoking, prior ischemic heart disease, injury, poison,

Supplemental Table S2: Administrative codes for feature extraction

Features	Source	ICD 9, ICD 10, CCI or HSC codes
Historic Diabetes Events		
Hypoglycemia	CLAIMS, ED, HOSP, VS	**ICD-9:** 250.3, 250.8, 251, 270.3, 775, 962.3; **ICD 10:** E100-E101, E110-E111, E130-E131, E140-E141, E160-E162, E0865, E0801, E08641, E13641, E1165, E1065, E11641, E10641
Retinopathy		**ICD-9:** 362.01-362.07; **ICD-10:** E1031-E1036, E1131-E1136, E1331-E1336, E08311, E08319, E0836, E0839
Cardiovascular diseases		**ICD-9:** 410-414, 428, 433, 435-436, 362.3 **ICD-10:** I20-I22, I61, I63-I64, I50, G450, G453, G458-G459
Amputation		**ICD-9:** 84.11-84.17, V49.71-V49.77; **CCI:** 1.VC.93, 1.VG.93, 1.VQ.93, 1.WA.93, 1.WE.93, 1.WI.93, 1.WJ.93, 1.WK.93, 1.WL.93, 1.WM.93, 1.WN.93; **HSC:** 96.1; **ICD-10:** 1VC, 1VG, IVQ, 1WA, 1WE, 1WI, 1WJ, 1WK, 1WL, 1WM, 1WN
Advanced foot infection		**ICD-9:** 040, 785.4, 250.7, 440.2-440.24, 730.07, 730.17, 730.27, 730.97, 707.14-707.15, 707.1, 680.7, 682.7, 681.1; **ICD-10:** I96, I70.26, I70.36, I70.46, I70.56, I70.66, I70.76, A48.0, E08.51, E09.51, E10.51, E11.51, E13.51, E08.71, E09.71, E10.71, E11.71, E13.71, M86.8X6, M86.8X7, M72.5, M72.6, M86.09, M86.19, M86.29, M86.39, M86.49, M86.59, M86.69, M86.89, I70.23-I70.24, I70.33-I70.34, I70.43-I70.44, I70.53-I70.54, I70.63-I70.64, I70.73-I70.74, L97.2-L97.5, L97.8-L97.9, E08.70, E09.70, E10.70, E11.70, E13.70
Tissue infection	ED, HOSP, VS	**ICD-10:** L00-L05, L08, M725-M726, A480, E1051, E1151, E1351, E1451, R02, E1061, E1161, E1361, E1461, E1070, E1171, E1371, E1471, E08620-E08622, E08628, E09620, E09622, E09628
Mild foot infection		**ICD-10:** L03.01-L03.12, M14.2, M14.6
Debridement	CLAIMS	**CCI:** 1.VQ.87, 1.VS.52.LA, 1.VX.59, 1.VZ.70.LA, 1.WA.52.WJ, 1.WA.52.WK, 1.WA.87, 1.WE.52.WJ, 1.WE.52.WK, 1.WE.87, 1.WI.52.WJ, 1.WI.52.WK, 1.WI.87, 1.WJ.52.WJ, 1.WJ.52.JK, 1.WJ.87, 1.WM.52.WJ, 1.WM.52.WK, 1.WN.52.WJ, 1.WN.52.WK, 1.WV.52.LA, 1.WV.59; **HSC:** 98.11
ACSC Conditions		
Angina	NACRS	**ICD-10:** I20, I2382, I240, I248-I249
Chronic obstructive pulmonary disease		**ICD-10:** J41-J44, J47
Asthma		**ICD-10:** J45
Hypertension		**ICD-10:** I100, I101, I11
Diabetes		**ICD-10:** E100-E101, E1063, E109-E111, E1163, E119, E130-E131, E1363, E139-E141, E1463, E149

continued..

Supplemental Table S2: Continued..

Features	Source	ICD 9, ICD 10, CCI or HSC codes
Grand mal status and other epileptic convulsions		**ICD-10:** G40, G41
Heart failure/pulmonary edema		**ICD-10:** I50 and J81
Elixhauser conditions		
Alcohol abuse	DAD, CLAIMS, NACRS	**ICD-9:** 265, 291, 303, 305, 357, 425, 535, 571, 265.2, V11.3, 980; **ICD-10:** G621, I426, K292, K700, K703, K709, Z502, Z714, Z721, F10, E52, T51
Arrhythmia		**ICD-9**: 426-427, 785, 996, V45.0, V53.3; **ICD-10:** I441-I443, I456, I459, R000-R001, R008, T821, Z450, Z950, I47-I49
Blood loss anemia		**ICD-9:** 280; **ICD-10:** D500
Congestive heart failure		**ICD-9:** 398, 402, 404, 425, 398.9, 428; **ICD-10:** I099, I110, I130, I132, I255, I420, I425-I429, P290, I43, I50
Coagulopathy		**ICD-9:** 286-287; **ICD-10:** D691, D693-D696, D65-D68
Deficiency anemia		**ICD-9:** 280-281
Depression		**ICD-9:** 300, 296.2-296.3, 296.5 309, 311; **ICD-10:** F204, F313-F315, F341, F412, F432, F32-F33
Diabetes		**ICD-9:** 250; **ICD-10:** E10-E14
Drug abuse		**ICD-9:** 304-305, V65.4, 292; **ICD-10:** Z715, Z722, F11-F16, F18, F19
Fluid & electrolyte disorders		**ICD-9:** 253; **ICD-10:** E222, E86-E87
Hypertension		**ICD-9:** 401-405; **ICD-10:** I10-I13, I15
Hypothyroidism		**ICD-9:** 240, 246, 243-244; **ICD-10:** E890, E00-E03
Liver disease		**ICD-9:** 070, 456, 572-573, V42.7, 570-571; **ICD-10:** I864, I982, K711, K713-K715, K717, K760, K762-K769, Z944, B18, I85, K70, K72-K74
Lymphoma		**ICD-9:** 203, 238, 200-202; **ICD-10:** C900, C902, C81-C85, C88, C96
Neurologic disorders		**ICD-9:** 331-336, 340-341, 345, 348, 780, 784, 348.1, 348.3; **ICD-10:** G254-G255, G312, G318, G319, G931, G934, R470, G10-G13, G20-G22, G32, G35-G37, G40-G41, R56
Obesity		**ICD-9:** 278; **ICD-10:** E66
Psychoses		**ICD-9:** 293, 295-298; **ICD-10:** F302, F312, F315, F20, F22-F25, F28-F29

continued..

Supplemental Table S2: Continued..

Features	Source	ICD 9, ICD 10, CCI or HSC codes
Renal failure		**ICD-9:** 403-404, 588, V42.0, V45.1, 585-586, V56; **ICD-10:** I120, I131, N250, Z490-Z492, Z940, Z992, N18-N19
Rheumatoid arthritis		**ICD-9:** 701, 710-711, 719, 728-729, 446, 714, 720, 725; **ICD-10:** L940, L941, L943, M120, M123, M310-M313, M461, M468-M469, M05-M06, M08, M30, M32-M35, M45
Solid tumor cancer		**ICD-9:** 140-172, 174-199; **ICD-10:** C00-C26, C30-C34, C37-C41, C43, C45-C58, C60-C80, C97
HIV/AIDS		**ICD-9:** 042-044; **ICD-10:** B20-B22, B24
Paralysis		**ICD-9:** 334, 342-344; **ICD-10:** G041, G114, G801-G802, G830-G834, G839, G81-G82
Peptic ulcer disease		**ICD-9:** 531-534; **ICD-10:** K257, K259, K267, K269, K277, K279, K287, K289
Pulmonary circulation disorders		**ICD-9:** 415-417; **ICD-10:** I280, I288-I289, I26-I27
Chronic pulmonary disease		**ICD-9:** 416, 506, 508, 499-505, 490-498; **ICD-10:** I278-I279, J684, J701, J703, J40-J47, J60-J63
Peripheral vascular disorders		**ICD-9:** 093, 437, 443, 447, 557, V43.4, 440-441; **ICD-10:** I731, I738-I739, I771, I790, I792, K551, K558-K559, Z958-Z959, I70-I71
Valvular disease		**ICD-9:** 093, 746, V42.2, V43.3, 394-397, 424; **ICD-10:** A520, I091, I098, Q230-Q233, Z952-Z954, I05-I08, I34-I39
Weight loss		**ICD-9:** 783, 799, 260-263; **ICD-10:** R634, E40-E46, R64
Frailty conditions		
Other cerebral degenerations including Alzheimer's	DAD, CLAIMS, NACRS	**ICD-9:** 336.0-336.2, 331-335; **ICD-10:** G31-G32
Cerebral generations usually manifest in childhood		**ICD-9:** 330; **ICD-10:** G94
Delirium		**ICD-9:** 293; **ICD-10:** F05
Dementia		**ICD-9:** 290; **ICD-10:** F03.90
Difficulty walking		**ICD-9:** 719.7; **ICD-10:** R26.2
Falls		**ICD-9:** E91.77-E91.78, E92.93; **ICD-10:** W01, W05-W19
Fecal incontinence		**ICD-9:** 787.6; **ICD-10:** R15
Gait abnormality		**ICD-9:** 781.2; **ICD-10:** R26
Senility without mention of psychosis		**ICD-9:** 797; **ICD-10:** R54
Urinary incontinence		**ICD-9:** 788.3; **ICD-10:** R32

continued..

Supplemental Table S2: Continued..

Features	Source	ICD 9, ICD 10, CCI or HSC codes
Vascular dementia		**ICD-9:** 290.4; **ICD-10:** F01
Pressure ulcer		**ICD-9:** 707; **ICD-10:** L89
Dementia in other diseases classified elsewhere		**ICD-9:** 294.1; **ICD-10:** F02, F03
Malaise & fatigue/debility		**ICD-9:** 780.7, 799.3; **ICD-10:** G93.3, R53
Dementia in Alzheimer's		**ICD-9:** 331; **ICD-10:** G30.0
Abnormal weight loss		**ICD-9:** 783.0, 783.21-783.22, 783.3, 783.9; **ICD-10:** R63.0, R63.3-R63.4, R63.6, R63.8
Muscle weakness		**ICD-9:** 728.87; **ICD-10:** M62.81
Muscular wasting and disuse atrophy		**ICD-9:** 728.2; **ICD-10:** M62.50
Sroke		**ICD-9:** 430-434; **ICD-10:** I60-I64
Other		
Injury	DAD, CLAIMS, NACRS	**ICD-9:** 800 <= icd of length 3 ≤ 999; **ICD-10:** start with either S or T
Poison		**ICD-9:** 960 ≤ icd of length 3 ≤ 989; **ICD-10:** 36 ≤ icd[1 : 3] ≤ 65
Smoking		**ICD-9:** 989.84, V15.82, 305.1, 649.00-649.04; **ICD-10:** F17, Z71.6, Z72.0, O99.33, T65.2, Z86.43
Hyperlipidemia		**ICD-9:** 272; **ICD-10:** E78
Prior ischemic heart disease		**ICD-9:** 410-414; **ICD-10:** I20-I25

CCI, Canadian Classification of Health Interventions; CLAIMS, physician claims; DAD, discharge abstract data; ED, emergency department; HOSP, hospital; HSC, Health Service Codes; ICD, International Statistical Classification of Diseases and Related Health Problems; NACRS, National Ambulatory Care Reporting System; VS, vital statistics.

Supplemental Table S3: Table of AUC's prior to model selection

Model	Outcomes in Community cohort			Outcomes in ED cohort	Outcomes in Community + ED cohort
	Hospitalization	ED visit	Death	Death	Death
CatBoost Classifier	0.6777	0.6471	0.9309	0.8553	0.918
Extreme Gradient Boosting	0.6836	0.6557	0.9271	0.8541	0.915
Ada Boost Classifier	0.6685	0.6407	0.9263	0.8406	0.9123
Light Gradient Boosting Machine	0.6898	0.6521	0.9253	0.8572	0.9139
Gradient Boosting Classifier	0.685	0.6551	0.9238	0.845	0.916
Random Forest Classifier	0.6741	0.6504	0.9226	0.8354	0.9021
Extra Trees Classifier	0.6767	0.6503	0.9216	0.8322	0.9069
Logistic Regression	0.6689	0.6342	0.9214	0.824	0.9056
Decision Tree Classifier	0.6282	0.6063	0.9088	0.7555	0.8814

Supplemental Table S4: Categorical Variables

(a) Participant Comparisons

	Community (n=11,247)					ED (n=1,495)					Community + ED (n=12,410)				
	Train (n=8,999)		Validation (n=2,250)		p-value	Train (n=1,195)		Validation (n=300)		p-value	Train (n=9,931)		Validation (2,485)		p-value
	n	percent	n	percent		n	percent	n	percent		n	percent	n	percent	
Sex															
Males	4086	45.41	1002	44.53	0.45	535	44.77	146	48.67	0.23	4578	46.10	1118	44.99	0.32
Females	4906	54.52	1244	55.29	0.51	656	54.90	154	51.33	0.27	5341	53.78	1366	54.97	0.29
Elixhauser comorbidities															
Cardiac arrhythmias	2834	31.49	693	30.80	0.53	458	38.33	116	38.67	0.91	3194	32.16	762	30.66	0.15
Diabetes	7353	81.71	1849	82.18	0.61	1042	87.20	252	84.00	0.15	8141	81.98	2060	82.90	0.28
Hypertension	5789	64.33	1476	65.60	0.26	899	75.23	217	72.33	0.3	6480	65.25	1624	65.35	0.93
Other Neurological disorders	5790	64.34	1441	64.04	0.79	818	68.45	233	77.67	**0.002**	6441	64.86	1614	64.95	0.93
Liver Disease	812	9.02	204	9.07	0.94	134	11.21	37	12.33	0.59	920	9.26	217	8.73	0.41
Rheumatoid arthritis	2118	23.54	549	24.40	0.39	361	30.21	95	31.67	0.62	2396	24.13	625	25.15	0.29
Fluid & electrolyte disorders	1421	15.79	391	17.38	0.07	387	32.38	108	36.00	0.23	1736	17.48	435	17.51	0.97
Deficiency anemia	1317	14.63	336	14.93	0.72	198	16.57	60	20.00	0.16	1504	15.14	351	14.12	0.20
Renal Failure	861	9.57	206	9.16	0.55	221	18.49	48	16.00	0.32	966	9.73	284	11.43	**0.01**
Congestive heart failure	829	9.21	234	10.40	0.08	201	16.82	60	20.00	0.19	988	9.95	268	10.78	0.22
Depression	3356	37.29	869	38.62	0.24	523	43.77	124	41.33	0.45	3769	37.95	958	38.55	0.58
Hypothyroidism	1486	16.51	327	14.53	**0.02**	198	16.57	43	14.33	0.35	1609	16.20	393	15.81	0.64
Peripheral vascular disorders	518	5.76	149	6.62	0.12	135	11.30	31	10.33	0.63	630	6.34	163	6.56	0.69
Chronic pulmonary disease	2014	22.38	503	22.36	0.98	400	33.47	113	37.67	0.17	2351	23.67	567	22.82	0.37
Obesity	2768	30.76	691	30.71	0.96	382	31.97	100	33.33	0.65	3062	30.83	786	31.63	0.44
Psychoses	904	10.05	230	10.22	0.81	152	12.72	53	17.67	**0.03**	1030	10.37	252	10.14	0.74
Alcohol abuse	1313	14.59	310	13.78	0.33	256	21.42	61	20.33	0.68	1502	15.12	368	14.81	0.70
Drug abuse	1707	18.97	446	19.82	0.36	270	22.59	64	21.33	0.64	1928	19.41	469	18.87	0.54
Solid Tumor Cancer	942	10.47	237	10.53	0.93	158	13.22	55	18.33	**0.02**	1083	10.91	250	10.06	0.22
Valvular disease	234	2.60	77	3.42	**0.03**	58	4.85	10	3.33	0.26	287	2.89	79	3.18	0.44
Pulmonary circulation disorders	293	3.26	64	2.84	0.31	68	5.69	20	6.67	0.52	331	3.33	91	3.66	0.42
Coagulopathy	232	2.58	59	2.62	0.91	57	4.77	9	3.00	0.18	276	2.78	64	2.58	0.58
Weight loss	623	6.92	179	7.96	0.09	121	10.13	39	13.00	0.15	744	7.49	171	6.88	0.30
Peptic Ulcer Disease*	212	2.36	48	2.13	0.52	55	4.60	9	3.00	0.22	223	2.25	79	3.18	**0.007**
Paralysis	141	1.57	38	1.69	0.68	38	3.18	3	1.00	**0.04**	158	1.59	44	1.77	0.53
Lymphoma	105	1.17	24	1.07	0.69	30	2.51	7	2.33	0.86	124	1.25	31	1.25	1.00
HIV/AIDS	30	0.33	3	0.13	0.11	6	0.50	1	0.33	0.7	29	0.29	9	0.36	0.57
Blood loss anemia	45	0.50	6	0.27	0.15	12	1.00	4	1.33	0.62	54	0.54	5	0.20	**0.03**

*excluding bleeding

Supplemental Table S4: Continued..

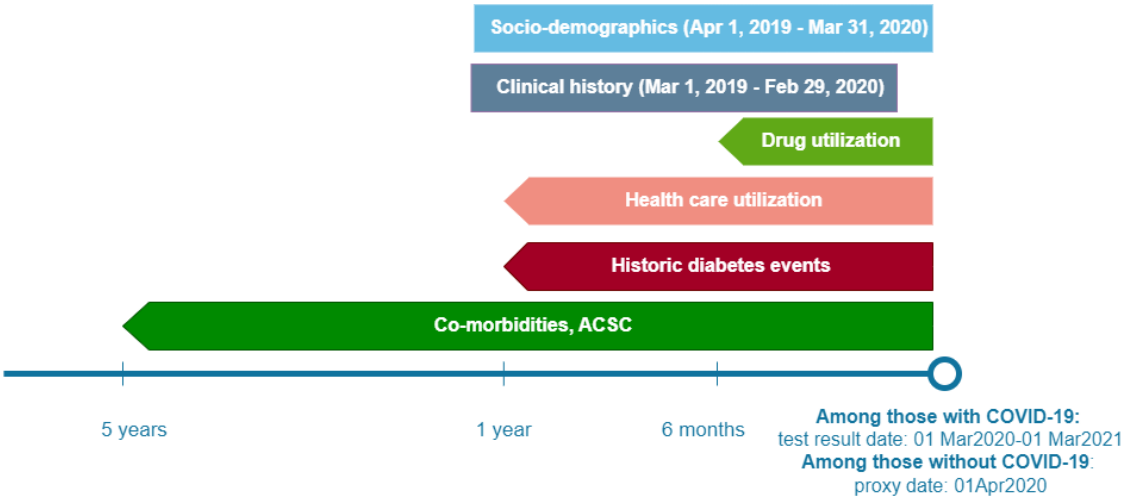
(b) Continuous variables

	Community (n=11,247)					ED (n=1,495)					Community + ED (n=12,410)				
	Train (n=8,999)		Validation (n=2,250)		p-value	Train (n=1,195)		Validation (n=300)		p-value	Train (n=9,931)		Validation (2,485)		p-value
	Mean	SD	Mean	SD		Mean	SD	Mean	SD		Mean	SD	Mean	SD	
Socio-demographics															
Age	55.99	16.83	56.12	17.13	0.74	61.56	16.90	60.65	17.65	0.41	56.37	17.02	56.31	16.84	0.87
Income	93752.11	23135.72	94226.29	24169.52	0.40	88995.30	25304.88	87311.76	23274.28	0.30	93461.64	23534.42	93161.64	23459.56	0.57
EMF	0.01	0.03	0.01	0.03	1.00	0.01	0.03	0.01	0.03	1.00	0.01	0.03	0.01	0.03	1.00
ESF	0.00	0.02	0.00	0.02	1.00	0.01	0.02	0.01	0.02	1.00	0.00	0.02	0.00	0.02	1.00
Healthcare utilization															
# physician visits	28.35	40.56	29.92	42.60	0.11	38.48	48.28	40.09	48.49	0.61	29.35	40.53	29.26	44.27	0.93
# physicians	6.53	6.49	6.80	7.29	0.11	9.40	10.11	10.17	10.83	0.25	6.82	6.83	6.79	7.45	0.86
# hospital visits	0.18	0.61	0.20	0.66	0.19	0.47	1.28	0.58	1.53	0.25	0.20	0.66	0.21	0.75	0.54
# days in hospital acute care	1.68	9.84	1.83	9.68	0.52	3.60	12.34	4.69	16.37	0.28	1.84	9.93	1.73	9.51	0.61
# ED visits	2.27	9.81	2.54	10.33	0.26	4.93	13.10	4.29	11.48	0.40	2.54	9.96	2.59	10.77	0.83
Drug utilization															
Total drug dispense qty	1353.74	1784.39	1350.04	1517.66	0.92	2127.44	4341.24	2131.19	2089.51	0.98	1415.07	1880.40	1379.99	1600.02	0.35
Max drug dispense qty	197.33	268.98	192.36	238.94	0.39	276.72	836.60	267.29	387.20	0.77	204.29	302.73	195.01	241.19	0.10
Min drug dispense qty	17.11	31.35	18.74	88.83	0.39	15.24	63.67	10.50	16.72	**0.02**	17.16	53.78	17.24	30.78	0.92
Average drug dispense qty	63.42	52.05	64.38	95.31	0.64	66.08	90.00	64.10	57.24	0.64	63.83	68.58	63.58	48.42	0.83
Historical # drug dispenses	32.27	59.20	33.91	60.46	0.24	46.07	83.45	50.30	107.40	0.53	33.95	63.49	32.30	58.55	0.22

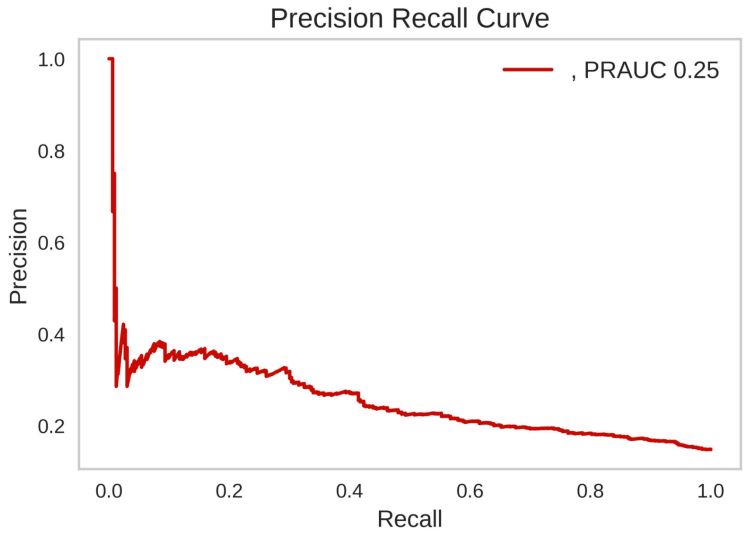
EMF, Estimated Material Factor; ESF, Estimated Social Factor, qty; quantity

Supplemental Table S5: Model Specification

CatBoost classifier for predicting death in the community cohort:
CatBoostClassifier(nan_mode='Min', gpu_ram_part=0.95, eval_metric= 'Logloss' iterations= 130, leaf_estimation_method= 'Newton', grow_policy= 'SymmetricTree', boosting_type= 'Plain', feature_border_type= 'GreedyLogSum', devices= '-1', l2_leaf_reg= 3, random_strength= 0.10000000149011612, rsm= 1, boost_from_average= False, model_size_reg= 0.5, use_best_model= False, class_names= [0, 1], random_seed= 123, depth= 3, border_count= 32, data_partition= 'DocParallel', bagging_temperature= 1, classes_count= 0, auto_class_weights= None, leaf_estimation_backtracking= 'AnyImprovement', best_model_min_trees= 1, min_data_in_leaf= 1, loss_function= 'Logloss', learning_rate= 0.10000000149011612, score_function= 'Cosine', task_type= 'GPU', leaf_estimation_iterations= 10, bootstrap_type= 'Bayesian', max_leaves= 8)
CatBoost classifier for predicting death in the community + ER cohort:
CatBoostClassifier(nan_mode=None, gpu_ram_part=0.95, eval_metric= 'Logloss',iterations= 130, leaf_estimation_method= 'Newton', grow_policy= 'SymmetricTree',boosting_type= 'Plain', feature_border_type= 'GreedyLogSum', devices= '-1', l2_leaf_reg= 3,random_strength= 0.1000000015, rsm= 1, boost_from_average= False, model_size_reg= 0.5, use_best_model= False, class_names= [0, 1], random_seed= 123, depth= 3, border_count= 32, data_partition= 'DocParallel', bagging_temperature= 1, classes_count= 0,auto_class_weights= None, leaf_estimation_backtracking= 'AnyImprovement', best_model_min_trees= 1, min_data_in_leaf= 1, loss_function= 'Logloss', learning_rate= 0.1000000015, score_function= 'Cosine', task_type= 'GPU', leaf_estimation_iterations= 10, bootstrap_type= 'Bayesian', max_leaves= 8)
LGBM model for predicting hospitalization in the ER cohort:
LGBMClassifier(bagging_fraction=0.6, bagging_freq=2, boosting_type='gbdt', class_weight=None, colsample_bytree=1.0, feature_fraction=0.4, importance_type='split', learning_rate=0.1, max_depth=-1, min_child_samples=41, min_child_weight=0.001, min_split_gain=0.9, n_estimators=260, n_jobs=10, num_leaves=70, objective=None, random_state=123, reg_alpha=2, reg_lambda=3, silent='warn', subsample=1.0, subsample_for_bin=200000, subsample_freq=0)
LGBM model for predicting hospitalization in the community cohort:
LGBMClassifier(bagging_fraction=1.0, bagging_freq=7, boosting_type='gbdt', class_weight=None, colsample_bytree=1.0, feature_fraction=0.7, importance_type='split', learning_rate=0.01, max_depth=-1, min_child_samples=11, min_child_weight=0.001, min_split_gain=0.2, n_estimators=80, n_jobs=10, num_leaves=60, objective=None, random_state=123, reg_alpha=0.0005, reg_lambda=0.4, silent='warn', subsample=1.0, subsample_for_bin=200000, subsample_freq=0)
XGBM model for predicting ER in the community cohort:
XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1, colsample_bynode=1, colsample_bytree=1, enable_categorical=False, gamma=0, gpu_id=0, importance_type=None, interaction_constraints='', learning_rate=0.0005, max_delta_step=0, max_depth=10, min_child_weight=1, missing=nan, monotone_constraints='()', n_estimators=300, n_jobs=10, num_parallel_tree=1, objective='binary:logistic', predictor='auto', random_state=123, reg_alpha=0.05, reg_lambda=5, scale_pos_weight=2.1, subsample=0.5, tree_method='gpu_hist', use_label_encoder=True, validate_parameters=1, verbosity=0)



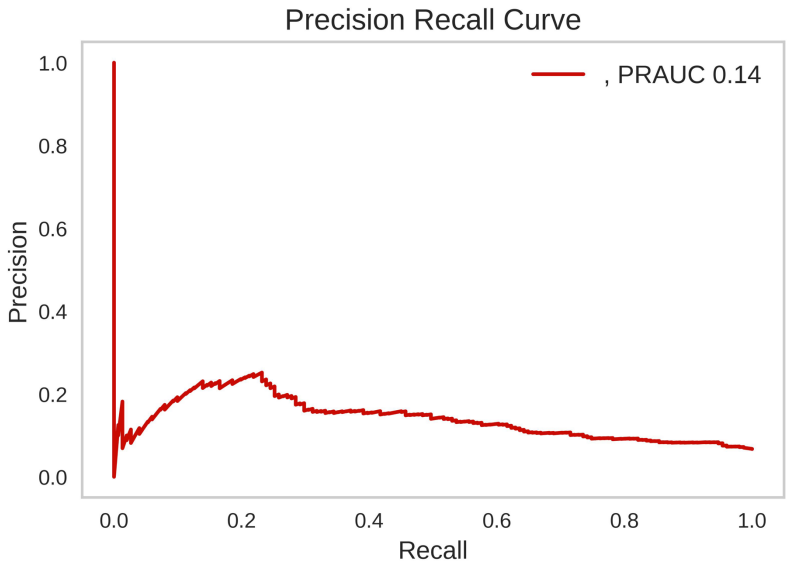
Supplemental Figure S1: Feature collection timeline. ACSC, ambulatory care sensitive conditions.



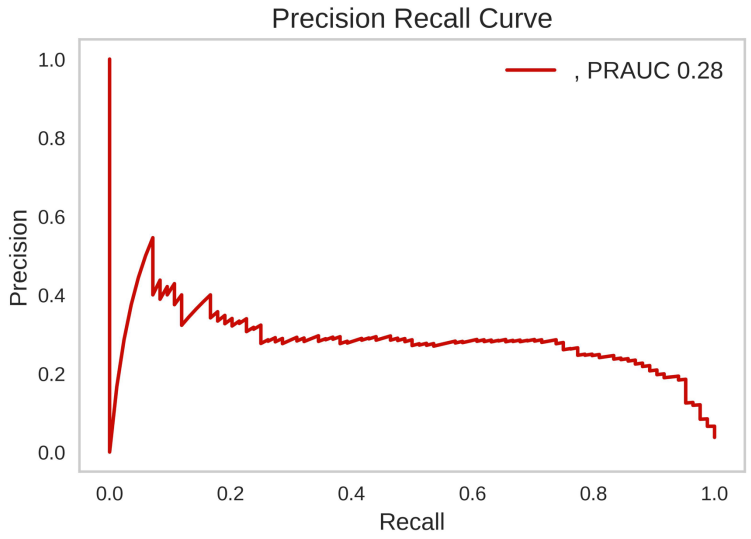
Supplemental Figure S2: Area under the precision-recall curve for all models. A) Predicting ED visit in Community cohort

SHAP bar plots present the overall rank of each features. Each bar in the mean summary plot of SHAP values illustrates the extent to which the corresponding feature influences the predicted primary outcome (death) among 3 subcohorts. The plot begins with features possessing the highest SHAP values and continues to delineate features with lower influence on the overall predictions. For example, according to A) the features with the most significant marginal effect on the highest importance outcome predictions in the Community cohort were age, followed by number of physician visits and history of dementia.

Beeswarm plots displays each individual instance of a feature value. With each dot representing an instance in the dataset, the distribution of dots offers insight into how a feature's value contributes positively or negatively to the prediction. Red dots signify higher feature values, and blue dots

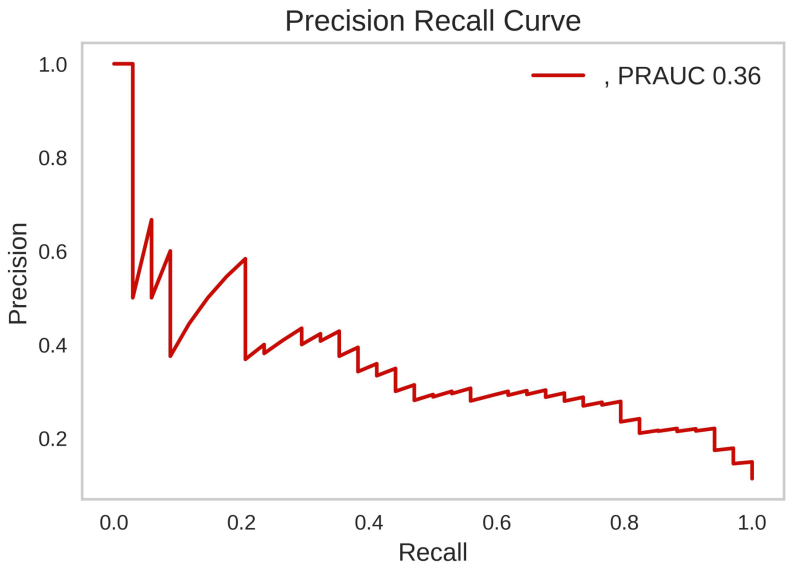


Supplemental Figure S2: B) Predicting hospitalization in the Community cohort

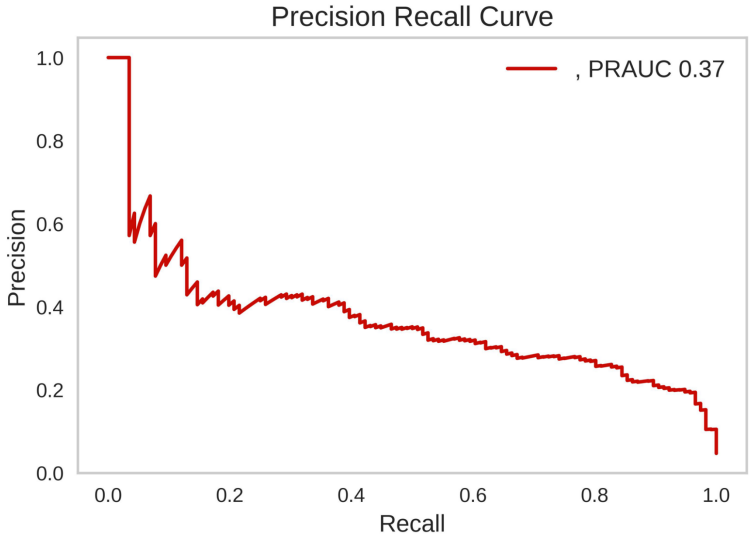


Supplemental Figure S2: C) Predicting death in the Community cohort

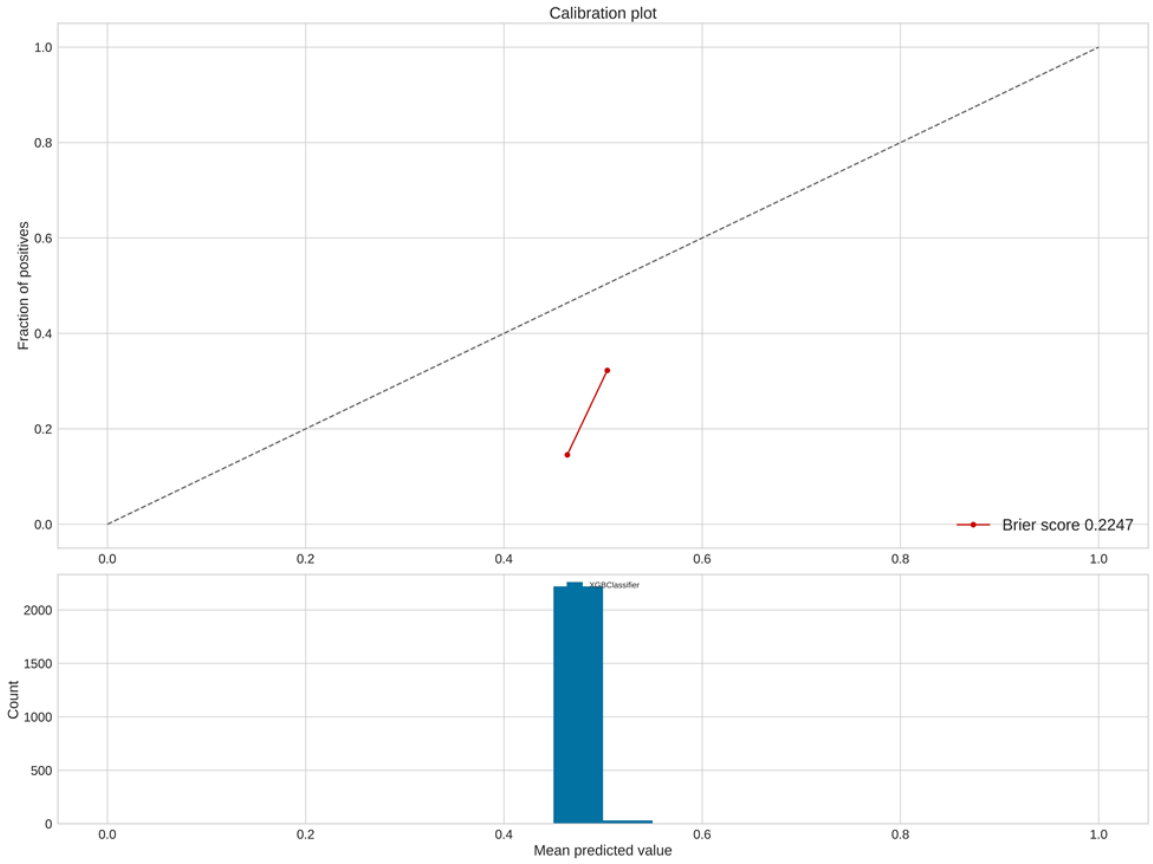
denote lower values. For example, in A) the beeswarm plot indicates that older age, larger numbers of physician visits and larger values for history of dementia contribute to a more likely death outcome in the Community cohort.



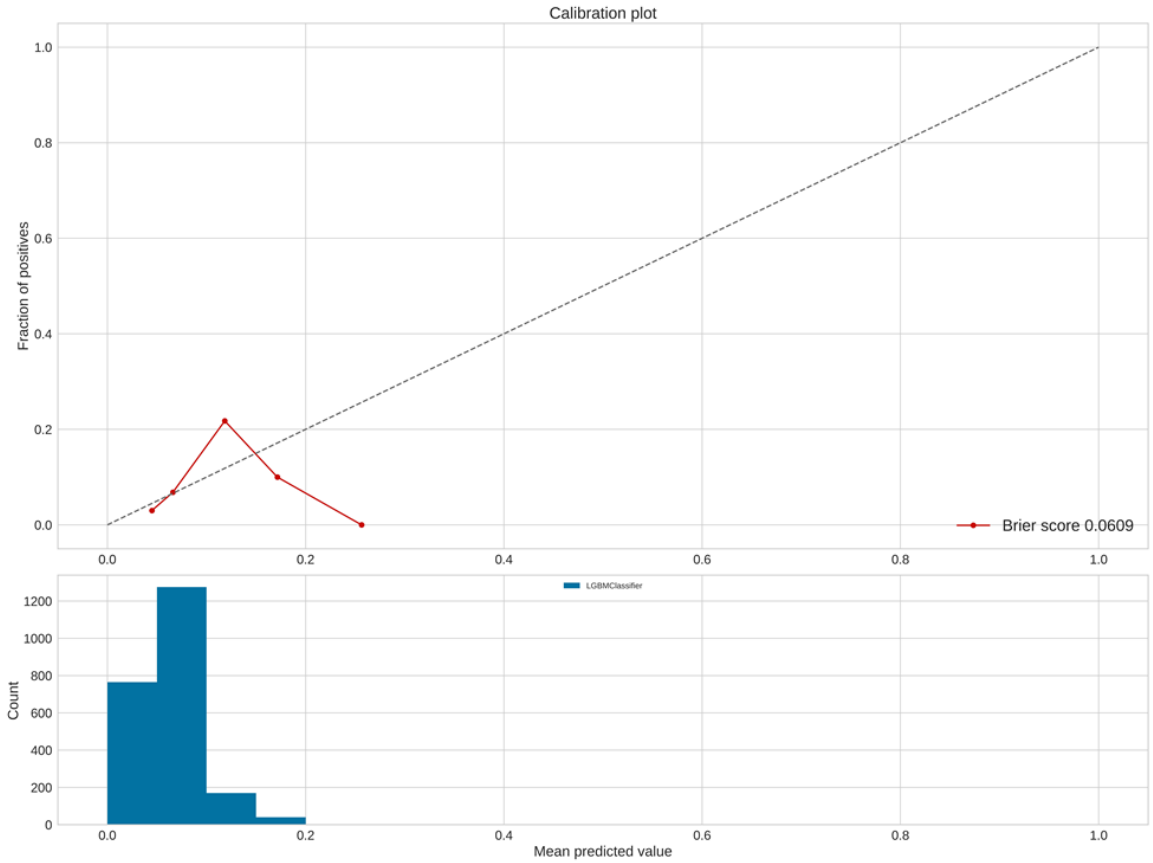
Supplemental Figure S2: D) Predicting death in ED cohort



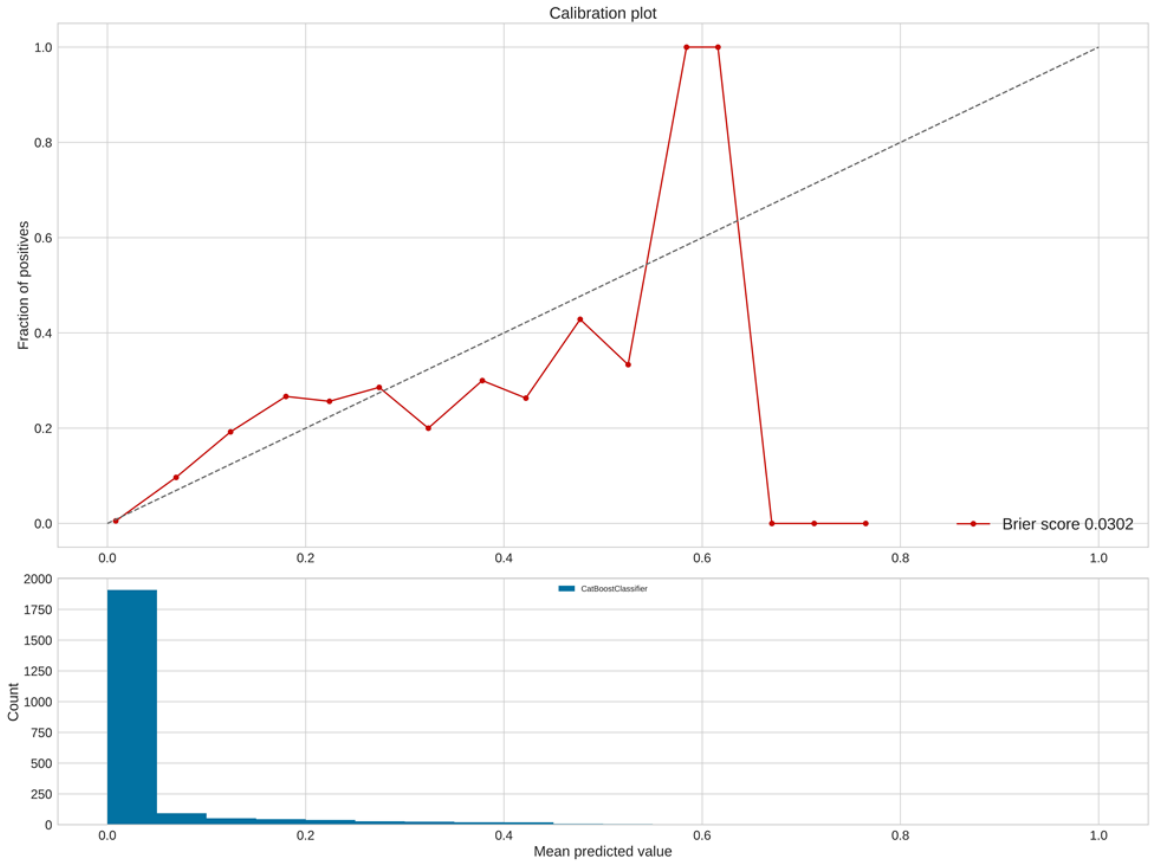
Supplemental Figure S2: E) Predicting death in the Community+ED cohort. PRAUC, Area under the precision-recall curve



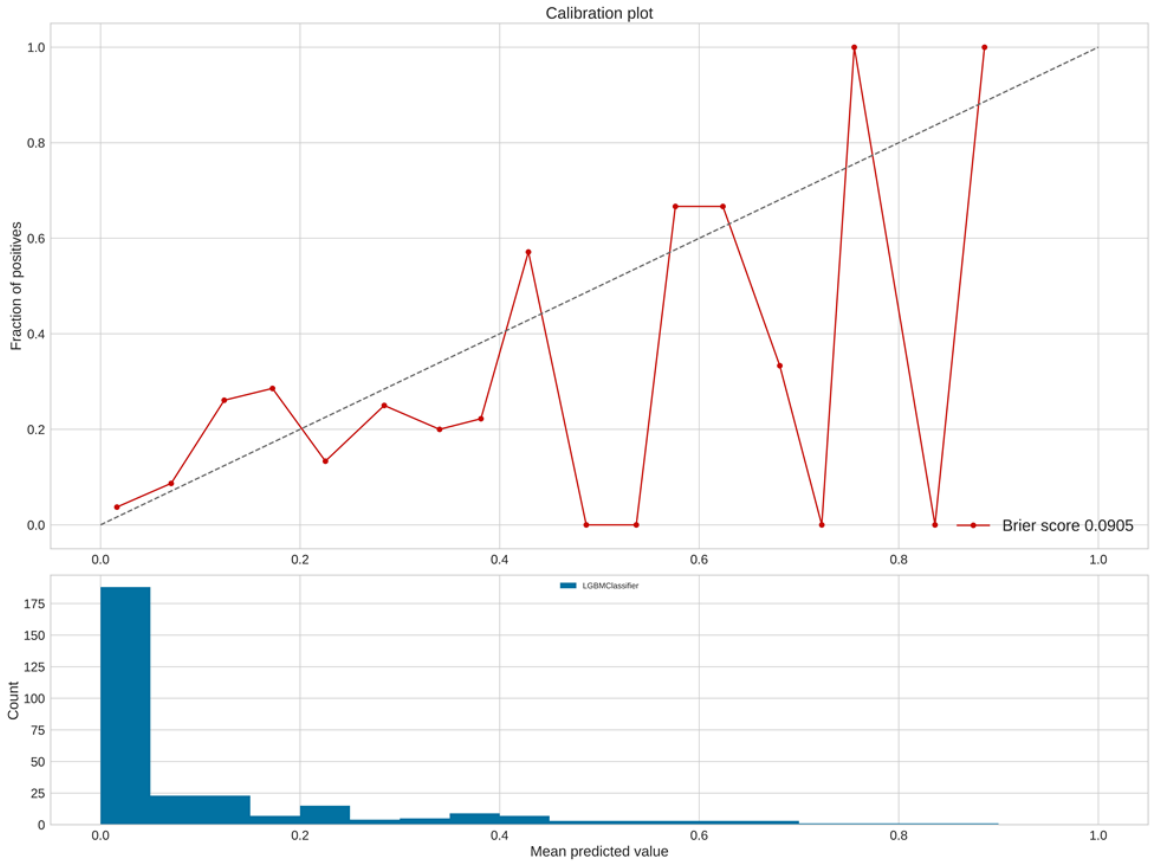
Supplemental Figure S3: Calibration plots for all models. A) Predicting ED visit in Community cohort



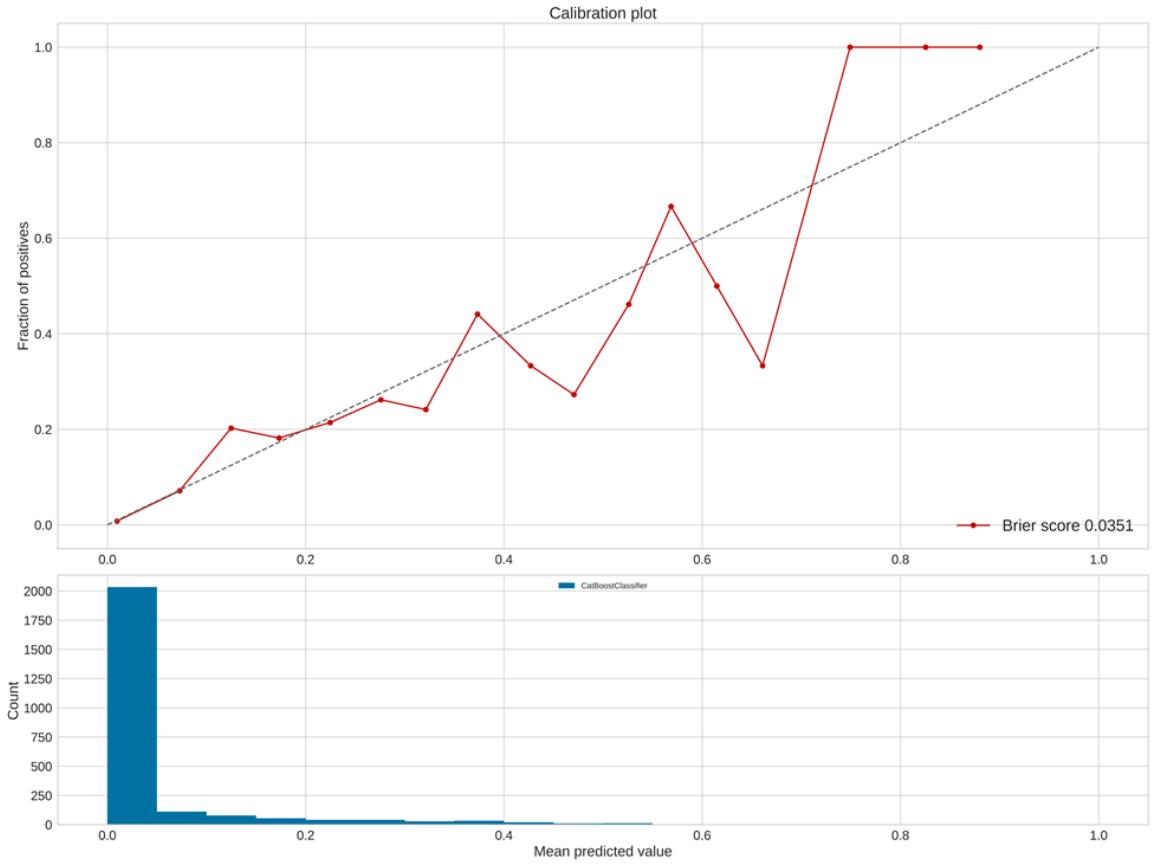
Supplemental Figure S3: B) Predicting hospitalization in Community cohort



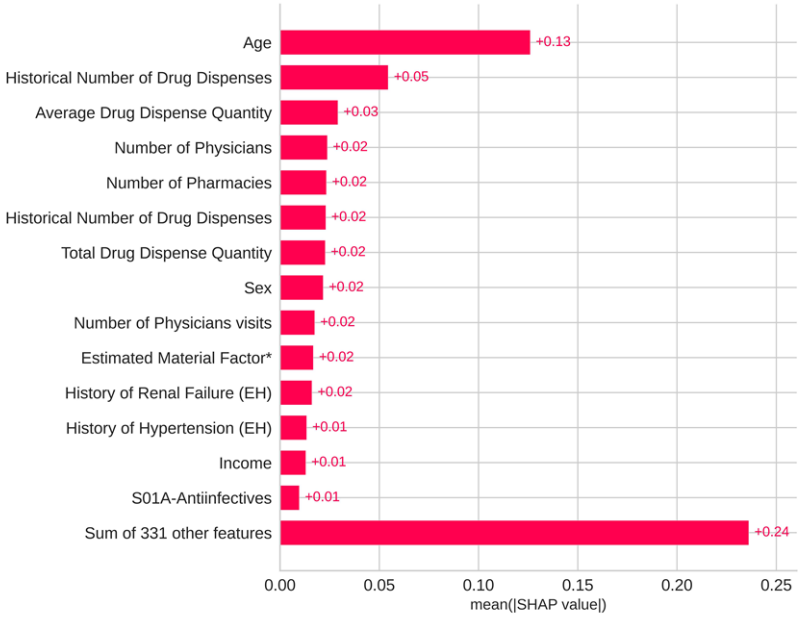
Supplemental Figure S3: C) Predicting death in Community cohort



Supplemental Figure S3: D) Predicting death in ED cohort

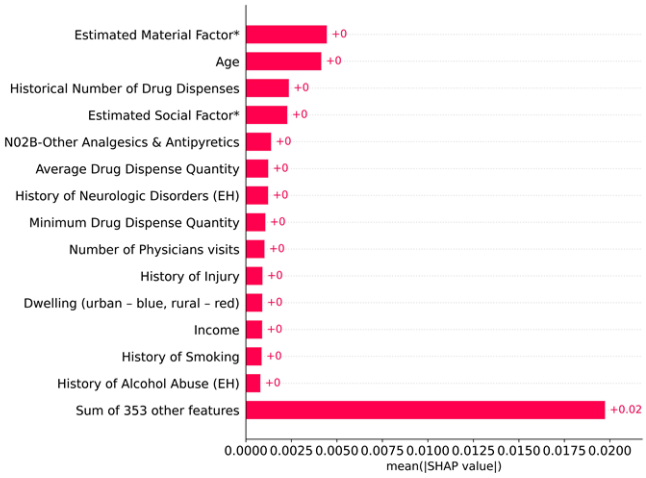


Supplemental Figure S3: E) Predicting death in Community+ED cohort



*Estimated Material Factor (Education, income, employment factors using the neighborhood-based Pampalon Deprivation Index)

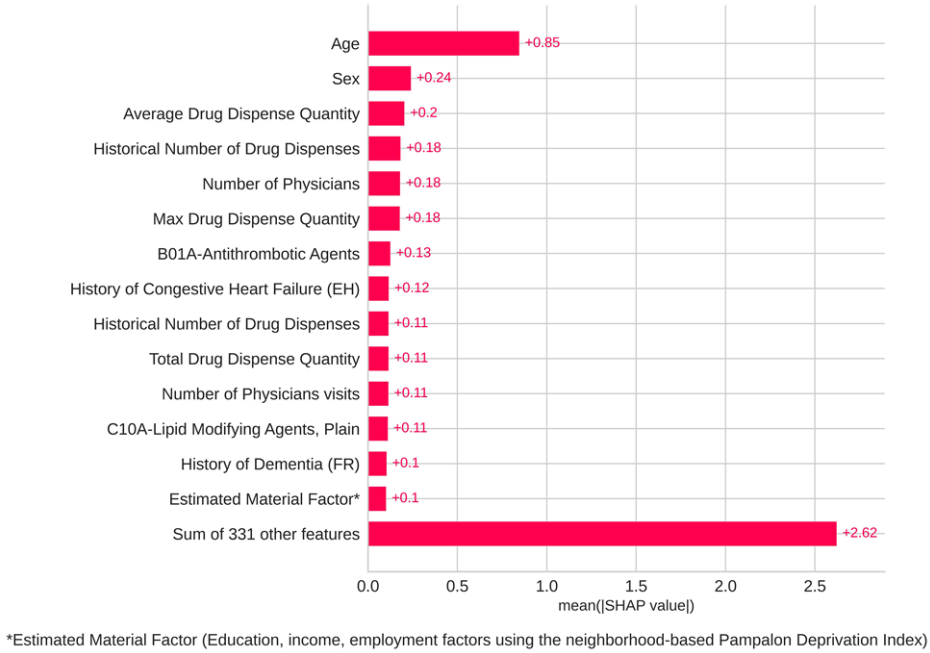
Supplemental Figure S4: SHAP values for models of interest. A) Absolute SHAP values for model predicting hospitalization in Community cohort



*Estimated Material Factor (Education, income, employment factors using the neighborhood-based Pampalon Deprivation Index)

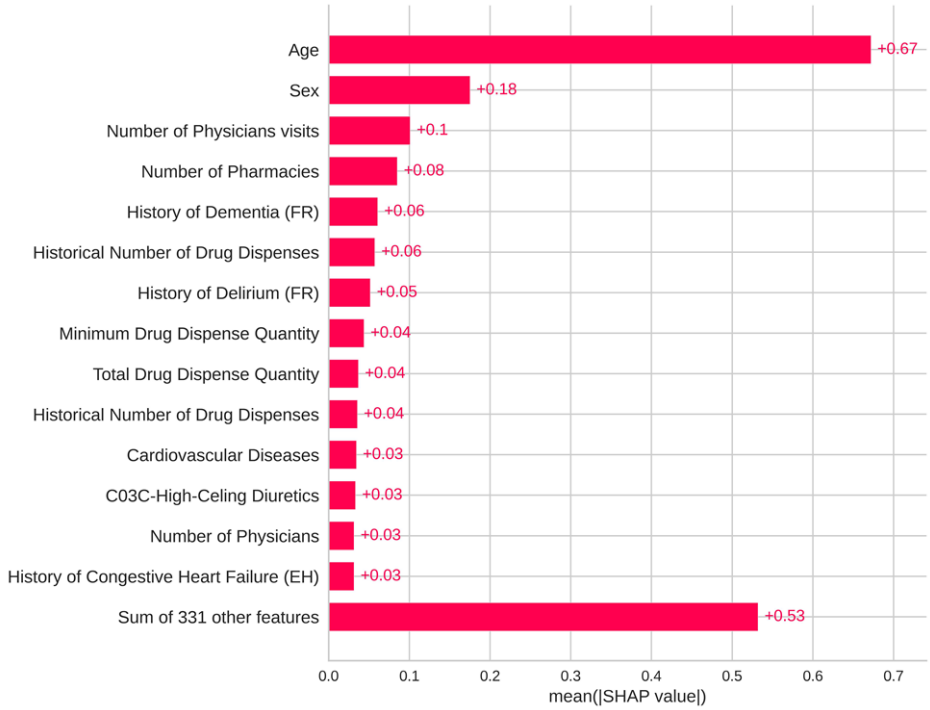
*Estimated Social Factor (Marital status, parenting arrangements and dwelling arranges using the neighborhood-based Pampalon Deprivation Index)

Supplemental Figure S4: B) Absolute SHAP values for model predicting death in Community cohort

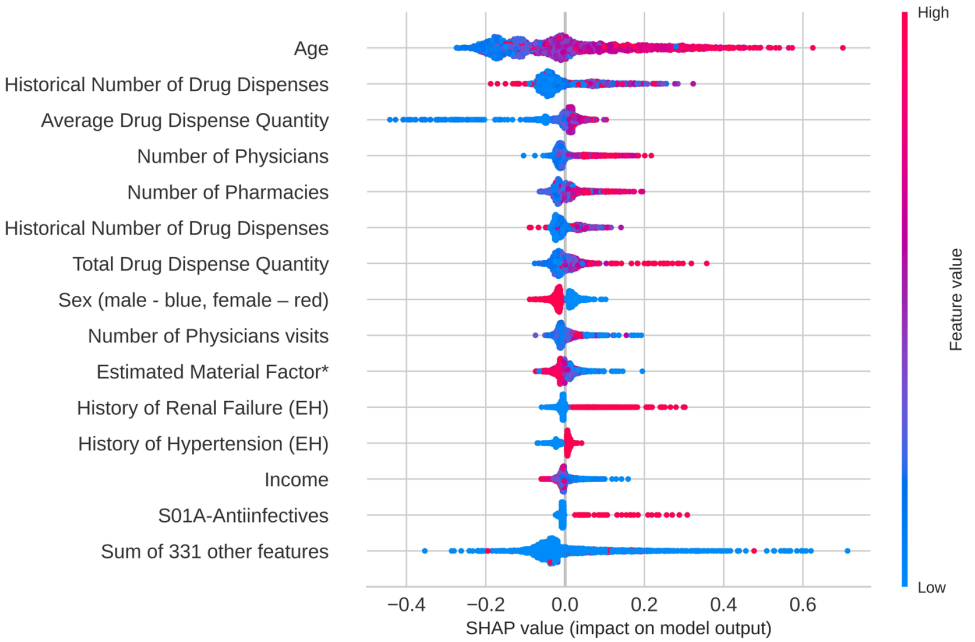


*Estimated Material Factor (Education, income, employment factors using the neighborhood-based Pampalon Deprivation Index)

Supplemental Figure S4: C) Absolute SHAP values for model predicting death in ED cohort

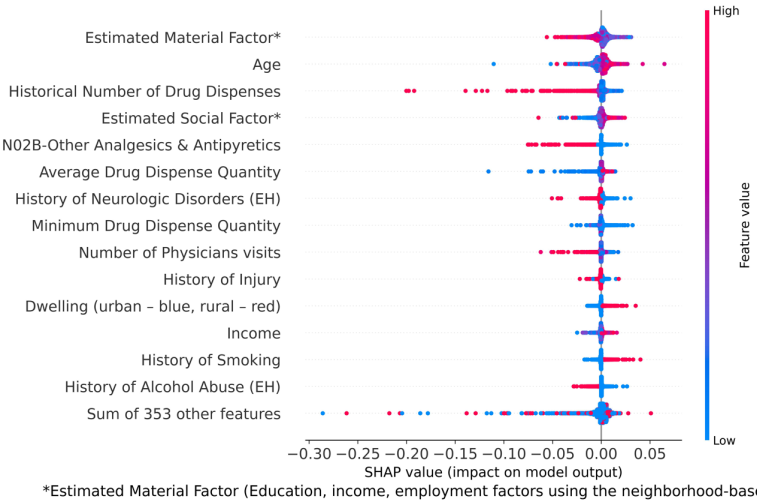


Supplemental Figure S4: D) Absolute SHAP values for model predicting death in Community + ED cohort



*Estimated Material Factor (Education, income, employment factors using the neighborhood-based Pampalon Deprivation Index)

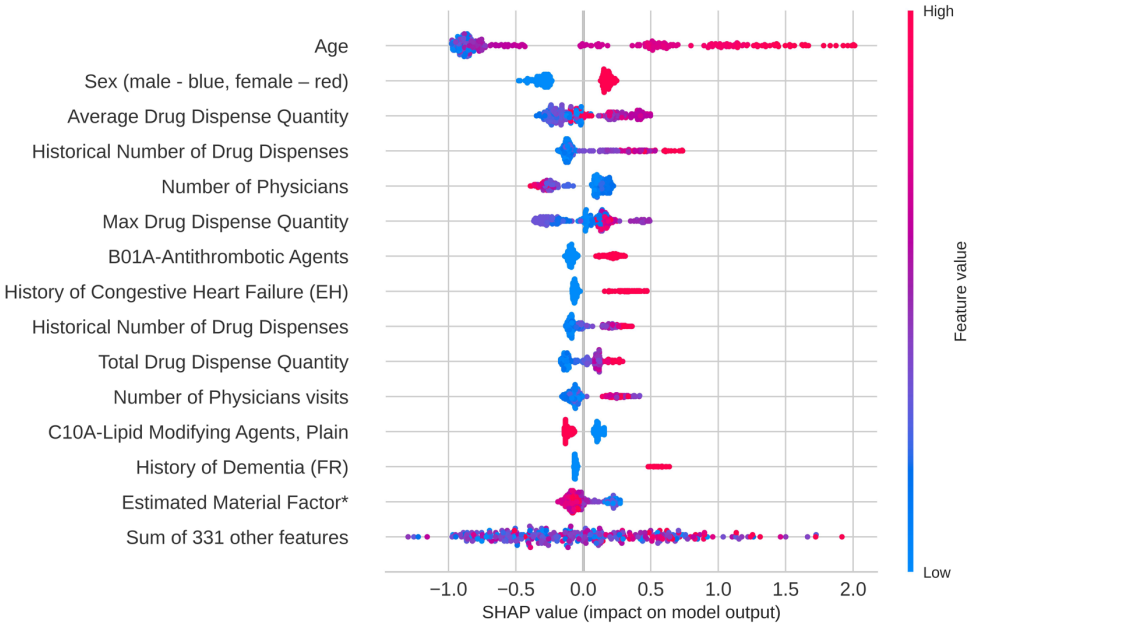
Supplemental Figure S5: Beeswarm plots of SHAP values for models of interest. A) For model predicting hospitalization in Community cohort



*Estimated Material Factor (Education, income, employment factors using the neighborhood-based Pampalon Deprivation Index)

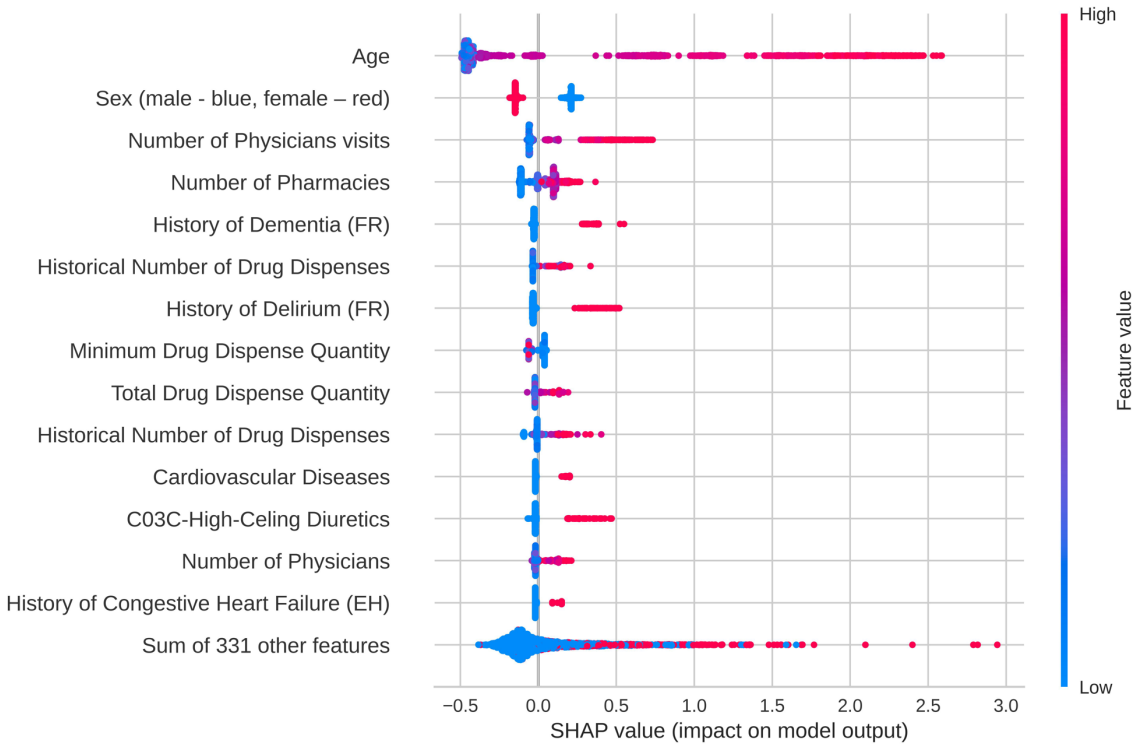
*Estimated Social Factor (Marital status, parenting arrangements and dwelling arranges using the neighborhood-based Pampalon Deprivation Index)

Supplemental Figure S5: B) For model predicting death in Community cohort



*Estimated Material Factor (Education, income, employment factors using the neighborhood-based Pampalon Deprivation Index)

Supplemental Figure S5: C) For model predicting death in ED cohort



Supplemental Figure S5: D) For model predicting death in Community + ED cohort

Appendix 1.

Methods: We conducted a search in Medline to review relevant literature predicting adverse COVID-19 outcomes: [exp *Machine Learning] AND [*COVID-19/ or *SARS-CoV-2/] AND [Hospitalization/ or "hospital visit".mp. or "emergency department visit".mp. or Emergency Room Visits/ or Death/ or Mortality/]. After review, this returned 49 results, and 23 were excluded due to ML models not being used for predictive modelling, or were based on populations, outcomes or features that were not of interest to us. Four additional studies were identified by review of references and were also included resulting in 30 included articles.

Results: Review of relevant literature predicting COVID-19 outcomes.

Author (Year)	Population of COVID-19 patients at BL	Outcome of interest	ML model	Test AUC (external validation AUC)	Model PPV/ precision
Ramalho-Pinto et al (2025)[1]	Hospitalized or non-hospitalized, Brazil	Hospitalization	Decision tree	N/A	N/A
Nirmalarajah et al (2025)[2]	Hospitalized or non-hospitalized, Canada	Hospitalization	Logistic regression	0.83	0.86
			Stacking (GBoost + Logistic regression-LASSO)	0.80	0.82
			GBoost	0.78	0.80
			Logistic regression-LASSO	0.77	0.89
			XGBoost	0.80	0.84
Delgado et al (2024)[3]	Hospitalized, Spain	Death	Multi-thresholding meta-algorithm	N/A	N/A
Mesinovic et al (2024)[4]	Hospitalized, global	Death	Logistic regression	0.73	N/A
			Linear discriminant analysis	0.73	N/A
			Naïve Bayes	0.71	N/A
			Random forest	0.74	N/A
			Stacking Ensemble	0.74	N/A
			XGBoost	0.74	N/A
			Ensemble (AdaBoost)	0.73	N/A
			Ensemble (XGBoost)	0.74	N/A
Benny et al (2024)[5]	Non-hospitalized, Italy	Hospitalization	AdaGrad	0.62-0.80*	0.62-0.85*
Govindan et al (2024)[6]	Hospitalized, USA	Death	XGBoost	0.88 (0.86)	0.03-0.35** (0.03-0.23)**
Alie et al (2024)[7]	Hospitalized, Ethiopia	Death	J48 Decision tree	0.50	0.95
			Random forest	0.55	0.95
			K-nearest neighbor	0.97	0.93
			Multi-layer perceptron	0.76	0.99
			Naïve Bayes	0.75	0.91
			XGBoost	0.71	0.95
			Logistic regression	0.58	0.91
Wei et al (2024)[8]	Non-hospitalized, USA	Combined hospitalization/ death/ ventilation	Logistic regression	0.70-0.71*	N/A
			XGBoost	0.78	N/A
			AdaBoost	0.69-0.71*	N/A
			K-nearest neighbor	0.64-0.65*	N/A
			Support vector machine	0.70	N/A
		Combined ventilation/ death	Logistic regression	0.78-0.81*	N/A
			XGBoost	0.86-0.88*	N/A
			AdaBoost	0.77-0.79*	N/A
			K-nearest neighbor	0.61-0.70*	N/A
			Support vector machine	0.70-0.78*	N/A
		Death	Logistic regression	0.79-0.84*	N/A
			XGBoost	0.89-0.91*	N/A
			AdaBoost	0.80-0.84*	N/A
			K-nearest neighbor	0.63-0.72*	N/A
			Support vector machine	0.73-0.81*	N/A

continued..

Author (Year)	Population of COVID-19 patients at BL	Outcome of interest	ML model	Test AUC (external validation AUC)	Model PPV/ precision
Cogill et al (2024)[9]	Non-hospitalized, USA	Hospitalization Death	GBoost GBoost	0.82 0.90	0.74 0.83
Dipaola et al (2023)[10]	Emergency department, Italy	Death	TensorFlow Fastai XGBoost	0.87-0.88* 0.84 0.84	0.49-0.53* 0.43-0.46* 0.43
Yu et al (2022)[11]	Hospitalized, Europe	Death	Cox regression Logistic regression Random forest	0.64 0.72 0.78	N/A N/A N/A
Jung et al (2022)[12]	Hospitalized, France	Combined death/ ICU/ adult respiratory distress syndrome/ unspecified coma	XGBoost	0.62-0.66*	N/A
Kamran et al (2022)[13]	Hospitalized, USA	Combined death/ ventilation/ nasal cannula/ vasopressors	Ensemble (logistic regression)	0.80(0.77-0.84)*	N/AN/A
Prieto (2022)[14]	Hospitalized or non-hospitalized, Mexico	Hospitalization, death	Logistic regression Decision tree K-Neighbors Bayes (Bernoulli) XGBoost Random forest	N/A N/A N/A N/A N/A N/A	N/A N/A N/A N/A N/A N/A
Marcos et al (2021)[15]	Hospitalized, Spain	Death	Random forest XGBoost Logistic regression	0.85 (0.81) 0.84 (0.82) 0.83 (0.83)	N/A N/A 0.55-0.78**
Yan et al (2021)[16]	Non-hospitalized, USA	Hospitalization	CatBoost ExtraTreesClassifier Logistic regression Ensemble models	0.79-0.80* 0.73-0.74* 0.67-0.70* 0.74	N/A N/A N/A N/A
Baqui et al (2021)[17]	Hospitalized, Brazil	Death	XGBoost Logistic regression K-nearest neighbor Neural Network Random forest Support vector machine	0.80-0.81* 0.76-0.77* 0.75-0.76* 0.78-0.80* 0.78-0.80* 0.76-0.77*	N/A N/A N/A N/A N/A N/A
Pyrros et al (2021)[18]	Non-hospitalized, USA	Hospitalization	Ensemble XGBoost	0.84	N/A
Vaid et al (2021)[19]	Hospitalized, USA	Combined death/ dialysis	Logistic regression LASSO Random forest XGBoost XGBoost + imputation	0.84-0.87* (0.71-0.81)* 0.92-0.95* (0.73-0.86)* 0.92-0.94* (0.83-0.84)* 0.93-0.98* (0.85-0.87)* 0.92-0.96* (0.82-0.86)*	0.07-0.24* (0.04-0.33)* 0.19-0.40* (0.05-0.38)* 0.16-0.36* (0.10-0.31)* 0.37-0.45* (0.10-0.37)* 0.19-0.40* (0.06-0.34)*

continued..

Author (Year)	Population of COVID-19 patients at BL	Outcome of interest	ML model	Test AUC (external validation AUC)	Model PPV/ precision
Estiri et al (2021)[20]	Non-hospitalized, USA	Hospitalization	Gboost, XGBoost DART	0.74-0.80*	N/A
		Death	Gboost, XGBoost DART	0.87-0.91*	N/A
Heldt et al (2021)[21]	Hospitalized, England	Death	Logistic regression	0.70	N/A
			Random forest	0.77	N/A
			XGBoost	0.76	N/A
Jimenez-Solem et al (2021)[22]	Hospitalized or non-hospitalized, Europe	Hospitalization	Random forest	0.82 (0.65-0.66)*	N/A
		Death	Random forest	0.72-0.91* (0.62-0.74)*	N/A
Sang et al (2021)[23]	Hospitalized, USA	Death	XGBoost	0.93	0.29
Guan et al (2021)[24]	Hospitalized, China	Death	Multi-tree XGBoost	0.98 (0.98)	N/A
			Simple-tree XGBoost	1.00 (1.00)	1.00 (1.00)
			Logistic regression	0.88-0.94* (0.94-1.00)*	N/A
Hu et al (2021)[25]	Hospitalized, China	Death	Logistic regression	0.90 (0.88)	N/A
			Random forest	0.92	N/A
			Bagged flexible discriminant analysis	0.90	N/A
			Partial least squares	0.89	N/A
			Elastic net	0.88	N/A
Rechtman et al (2020)[26]	Hospitalized or non-hospitalized, USA	Death	XGBoost	0.86	N/A
Willette et al (2022)[27]	Hospitalized or non-hospitalized, England	Hospitalization	Linear discriminant analysis	0.52-0.93*	N/A
Khadem et al (2022)[28]	Hospitalized, UK, with diabetes	Death	Random Forest	0.80	N/A
	Hospitalized, UK, without diabetes	Death	Random Forest	0.84	N/A
Gao et al (2020)[29]	Hospitalized, China	Death	Logistic regression	0.96 (0.92-0.97)*	0.83 (0.70-0.87)*
			Support vector machine	0.96 (0.91-0.98)*	0.82 (0.69-0.88)*
			K-nearest neighbor	0.91 (0.82-0.90)*	0.92 (0.56-0.81)*
			Random Forest	0.91 (0.79-0.91)*	0.82 (0.71-0.82)*
			GBoost decision tree	0.95 (0.90-0.95)*	0.75 (0.73-0.86)*
			Neural network	0.96 (0.92-0.98)*	0.89 (0.75-0.90)*
			Ensemble (logistic regression, support vector machine, GBoost decision tree, neural network)	0.96 (0.92-0.98)*	0.85 (0.73-0.90)*
Khodabakhsh et al (2023)[30]	Hospitalized, Iran, with diabetes	Death	Decision Tree	0.87	0.96
			Logistic Regression	0.83	N/A

AdaBoost, Adaptive Boostig; AdaGrad, adaptive gradient; GBoost, gradient boosting machine; LASSO, least absolute shrinkage and selection operator; XGBoost, eXtreme gradient boosting; N/A, not applicable ie not reported; *varies by strata (i.e.age, sex, time period, feature type, cohort); **depends on threshold used