# Supplemental Digital Content 1

## Comparison of Algorithm Performance

We investigated several machine learning algorithms, including logistic regression, random forest, CatBoost, and stacked models combining the results of the three using voting. CatBoost generally yielded the best AUC, precision-recall, and calibration results, as can be seen in Figure E2, Figure E3, and Table E5, and therefore it was used in the final models.

## References

1. Hur S, Ko RE, Yoo J, Ha J, Cha WC, Chung CR: A machine learning-based algorithm for the prediction of intensive care unit delirium (PRIDE): Retrospective study. JMIR Med Informatics 2021; 9:1–20.
2. Boogaard M Van Den, Pickkers P, Slooter AJC, Kuiper MA, Spronk PE, Voort PHJ Van Der, Hoeven JG Van Der, Donders R, Achterberg T Van, Schoonhoven L: Development and validation of PRE-DELIRIC (PREdiction of DELIRium in ICu patients) delirium prediction model for intensive care patients: Observational multicentre study. BMJ 2012; 344:17
3. Chen X, Lao Y, Zhang Y, Qiao L, Zhuang Y: Risk predictive models for delirium in the intensive care unit: A systematic review and meta-analysis. Ann Palliat Med 2021; 10:1467–79.
4. Marra A, Pandharipande PP, Shotwell MS, Chandrasekhar R, Girard TD, Shintani AK, Peelen LM, Moons KGM, Dittus RS, Ely EW, Vasilevskis EE: Acute Brain Dysfunction: Development and Validation of a Daily Prediction Model. Chest 2018; 154:293–301.
5. Moon KJ, Jin Y, Jin T, Lee SM: Development and validation of an automated delirium risk assessment system (Auto-DelRAS) implemented in the electronic health record system. Int J Nurs Stud 2018; 77:46–53.
6. Fan H, Ji M, Huang J, Yue P, Yang X, Wang C, Ying W: Development and validation of a dynamic delirium prediction rule in patients admitted to the Intensive Care Units (DYNAMIC-ICU): A prospective cohort study. Int J Nurs Stud 2019; 93:64–73.

## Table E1: Completed Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis (TRIPOD) checklist.

|  |  |  |  |
| --- | --- | --- | --- |
| Section/Topic | Item | Checklist item | Completed? |
| Title | 1 | Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted. | Yes |
| Abstract | 2 | Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions. | Yes |
| Background | 3a | Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references toexisting models. | Yes |
| Objectives | 3b | Specify the objectives, including whether the study describes the development or validation of the model, or both. | Yes |
| Source of Data | 4a | Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable. | Yes |
|  | 4b | Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up. | Yes |
| Participants | 5a | Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres. | Yes |
|  | 5b | Describe eligibility criteria for participants. | Yes |
|  | 5c | Give details of treatments received, if relevant. | N/A because retrospective study.  |
| Outcome | 6a | Clearly define the outcome that is predicted by the prediction model, including how and when assessed. | Yes |
|  | 6b | Report any actions to blind assessment of the outcome to be predicted. | N/A because retrospective study. |
| Predictors | 7a | Clearly define all predictors used in developing the multivariable prediction model, including how and when they were measured. | Yes |
|  | 7b | Report any actions to blind assessment of predictors for the outcome and other predictors. | Yes |
| Sample Size | 8 | Explain how the study size was arrived at. | Yes |
| Missing Data | 9 | Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method. | Yes |
| Statistical Analysis Methods | 10a | Describe how predictors were handled in the analyses. | Yes |
|  | 10b | Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation. | Yes |
|  | 10c | For validation, describe how the predictions were calculated. | Yes |
|  | 10d | Specify all measures used to assess model performance and, if relevant, to compare multiple models. | Yes |
|  | 10e | Describe any model updating (e.g., recalibration) arising from the validation, if done. | N/A because no model updating. |
| Risk Groups | 11 | Provide details on how risk groups were created, if done. | N/A, not done. |
| Development vs. validation | 12 | For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors. | Yes |
| Participants | 13a | Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram maybe helpful. | Yes |
|  | 13b | Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome. | Yes |
|  | 13c | For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors, and outcome). | Yes |
| Model Development | 14a | Specify the number of participants and outcome events in each analysis. | Yes |
|  | 14b | If done, report the unadjusted association between each candidate predictor and outcome. | Yes |
| Model Specification | 15a | Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point). | Yes |
|  | 15b | Explain how to use the prediction model. | Yes |
| Model Performance | 16 | Report performance measures (with CIs) for the prediction model. | Yes |
| Model Updating | 17 | If done, report the results from any model updating (i.e., model specification, model performance). | Yes |
| Limitations | 18 | Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data). | Yes |
| Interpretation | 19a | For validation, discuss the results with reference to performance in the development data, and any other validation data. | Yes |
|  | 19b | Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence. | Yes |
| Implications | 20 | Discuss the potential clinical use of the model and implications for future research. | Yes |
| Supplementary Information | 21 | Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets. | Yes |
| Funding | 22 | Give the source of funding and the role of the funders for the present study. | Yes |

## Table E2: All signals/features extracted and examined in the study, and which were ultimately used in final predictive models.

|  |  |  |  |
| --- | --- | --- | --- |
| Category or Variable | Specific Processing and Features Extracted | First 24 hour Model | Dynamic 1 hour Model |
| Demographics |  |  |  |
| Age | Ages greater than 90 were converted to 90 due to dataset limitations | X | X |
| Sex | Male or female, one-hot encoded | X | X |
| Ethnicity | One-hot encoded 6 categories: Caucasian, African American, Hispanic, other, Asian, Native American |  |  |
| Admission Information |  |  |  |
| Admission Diagnosis | Used in Elixhauser calculation | X | X |
| Unit Admit Source | One-hot encoded, options varied by dataset | X | X |
| ICU Type | One-hot encoded, options varied by dataset | X | X |
| Hospital Admit Time | Calculated number of seconds since midnight | X | X |
| Unit Admit Time | Calculated number of seconds since midnight |  | X |
| Height | On admission | X | X |
| Weight | First value after ICU admission |  |  |
| Urgent Admission | Binary flag, determined using certain diagnoses and elective flags in data | X | X |
| Hospital Information |  |  |  |
| Number of Beds | Categorical variable with 4 options: <100, 100-250, 250-499, or 500+ beds | X | X |
| Teaching Status | Binary flag | X | X |
| Physiological Scoring |  |  |  |
| Glasgow Coma Score (total, verbal, motor, eyes) | Mean, Last recorded value | X | X |
| RASS | Mean, Last recorded value | X | X |
| APACHE IV | Extracted from first 24 hours after ICU admission | X | X |
| Q-SOFA (and sub scores) | Calculated over whole ICU stay up to prediction time | X | X |
| SOFA (and sub scores) | Calculated over whole ICU stay up to prediction time | X | X |
| Medical History |  |  |  |
| Automatic Implantable Cardioverter Defibrillator | Binary flag | X | X |
| Angina | Binary flag | X | X |
| Aplastic Anemia | Binary flag | X | X |
| Arrhythmia | Binary flag | X | X |
| Asthma | Binary flag | X |  |
| Cancer | Binary flag |  | X |
| Chemotherapy | Binary flag |  |  |
| Chronic Obstructive Pulmonary Disease | Binary flag | X | X |
| Cirrhosis | Binary flag |  |  |
| Clotting Disorder | Binary flag | X |  |
| Congestive Heart Failure | Binary flag |  |  |
| Coronary Artery Bypass Graft Surgery | Binary flag | X | X |
| Cushing’s Disease | Binary flag |  |  |
| Dementia | Binary flag | X | X |
| Diabetes | Binary flag | X |  |
| Heart Transplant | Binary flag | X | X |
| Hemolytic Anemia | Binary flag | X |  |
| Hypercalcemia | Binary flag | X | X |
| Hypercoagulable Condition | Binary flag | X |  |
| Hypersplenism | Binary flag | X |  |
| Hypertension | Binary flag |  | X |
| Hyperthyroid Disease | Binary flag | X | X |
| Hypothyroid Disease | Binary flag |  | X |
| Immune Suppression | Binary flag |  | X |
| Intracranial Mass | Binary flag | X |  |
| Liver Transplant | Binary flag | X | X |
| Lung Transplant | Binary flag | X | X |
| Myocardial Infarction | Binary flag | X |  |
| Myeloproliferative Disease | Binary flag | X | X |
| Neurogenic Bladder | Binary flag |  | X |
| Neuromuscular Disease | Binary flag | X |  |
| Oncology | Binary flag | X | X |
| Pacemaker | Binary flag | X | X |
| Percutaneous Coronary Intervention | Binary flag | X |  |
| Peptic Ulcer Disease | Binary flag | X | X |
| Pulmonary Embolism | Binary flag | X | X |
| Peripheral Vascular Disease | Binary flag |  |  |
| Radiation Therapy | Binary flag | X | X |
| Renal Failure | Binary flag | X |  |
| Renal Insufficiency | Binary flag |  |  |
| Renal Transplant | Binary flag |  | X |
| Respiratory Failure | Binary flag | X | X |
| Restrictive Disease | Binary flag | X | X |
| Rheumatic Disease | Binary flag |  | X |
| Renal Tubular Acidosis | Binary flag | X |  |
| Sarcoidosis | Binary flag |  |  |
| Seizures | Binary flag |  | X |
| Sickle Cell Disease | Binary flag | X | X |
| Steroid Use | Binary flag | X | X |
| Stone Disease | Binary flag |  | X |
| Stroke | Binary flag | X | X |
| Transient Ischemic Attack | Binary flag | X | X |
| Valve Disorder | Binary flag | X |  |
| Venous Thrombosis | Binary flag | X |  |
| Labs |  |  |  |
| Bicarbonate | Mean, minimum, maximum | X |  |
| Blood urea nitrogen | Mean, minimum, maximum | X |  |
| Calcium | Mean, minimum, maximum | X |  |
| Chloride | Mean, minimum, maximum | X |  |
| Creatinine | Mean, minimum, maximum |  |  |
| Glucose | Mean, minimum, maximum | X |  |
| Hematocrit | Mean, minimum, maximum | X |  |
| Hemoglobin | Mean, minimum, maximum | X |  |
| Mean corpuscular hemoglobin concentration | Mean, minimum, maximum | X |  |
| Mean corpuscular volume | Mean, minimum, maximum | X |  |
| Platelets | Mean, minimum, maximum | X |  |
| Potassium | Mean, minimum, maximum | X |  |
| Red blood cells | Mean, minimum, maximum |  |  |
| Red cell distribution width | Mean, minimum, maximum | X |  |
| Sodium | Mean, minimum, maximum | X |  |
| White blood cells | Mean, minimum, maximum | X |  |
| Medications |  |  |  |
| Acetaminophen | Binary indicator of administration |  |  |
| Adrenergic Bronchodilators | Binary indicator of administration |  | X |
| Aminoglycosides | Binary indicator of administration |  | X |
| Anticholinergic Bronchodilators | Binary indicator of administration | X | X |
| Anticholinergics | Binary indicator of administration |  | X |
| Anticoagulants | Binary indicator of administration |  |  |
| Antidiarrheals | Binary indicator of administration |  | X |
| Antiemetics | Binary indicator of administration | X | X |
| Antihistamines | Binary indicator of administration |  | X |
| Barbiturates | Binary indicator of administration | X | X |
| Benzodiazepines | Binary indicator of administration | X | X |
| Beta Blockers | Binary indicator of administration |  | X |
| Calcium Channel Blockers | Binary indicator of administration |  |  |
| Carbapenems | Binary indicator of administration | X | X |
| Cephalosporins | Binary indicator of administration | X | X |
| Class V Antiarrhythmics | Binary indicator of administration |  |  |
| Colloid fluids | Binary indicator of administration |  |  |
| Crystalloid fluids | Binary indicator of administration | X | X |
| Diuretics | Binary indicator of administration | X |  |
| General Anesthetics | Binary indicator of administration | X | X |
| Glucocorticoids | Binary indicator of administration |  |  |
| Glucose Elevating Drugs | Binary indicator of administration | X | X |
| Glycopeptides | Binary indicator of administration | X | X |
| H2 Receptor Blockers | Binary indicator of administration | X | X |
| Insulin | Binary indicator of administration | X |  |
| Laxatives | Binary indicator of administration |  |  |
| Lincomycins | Binary indicator of administration | X | X |
| Macrolides | Binary indicator of administration | X | X |
| Monoamine oxidase inhibitors Antidepressants | Binary indicator of administration |  | X |
| Methylxanthines | Binary indicator of administration |  |  |
| Neuromuscular Blockers | Binary indicator of administration | X | X |
| NSAIDs | Binary indicator of administration |  |  |
| Opioids | Binary indicator of administration | X | X |
| Other antidepressants | Binary indicator of administration | X | X |
| Beta-Lactams | Binary indicator of administration | X | X |
| Phenylpiperazine Antidepressants | Binary indicator of administration | X | X |
| Potassium Channel Blockers | Binary indicator of administration | X |  |
| Proton Pump Inhibitor | Binary indicator of administration |  | X |
| Quinolones | Binary indicator of administration | X | X |
| serotonin and norepinephrine reuptake inhibitor Antidepressants | Binary indicator of administration | X | X |
| Sodium Channel Blockers | Binary indicator of administration |  |  |
| Somatostatin | Binary indicator of administration | X |  |
| SSRI Antidepressants | Binary indicator of administration | X |  |
| Sulfonamides | Binary indicator of administration | X |  |
| Tetracyclic Antidepressants | Binary indicator of administration |  | X |
| Tetracyclines | Binary indicator of administration |  | X |
| Thrombolytics | Binary indicator of administration | X | X |
| Tricyclic Antidepressants | Binary indicator of administration |  |  |
| Vasodilators | Binary indicator of administration |  |  |
| Vasopressors | Binary indicator of administration | X | X |
| Physiological Measurements |  |  |  |
| Temperature | Mean, minimum, maximum | X |  |
| Blood pressure | tsfresh package features |  |  |
| SaO2 | tsfresh package features |  |  |
| Respiratory Rate | tsfresh package features |  |  |
| Heart rate | tsfresh package features |  |  |
| Urine output | Volume in mL over last 24 hours |  |  |
| Treatments |  |  |  |
| Dialysis | Binary indicator |  |  |
| Mechanical Ventilation | Binary indicator, total duration in minutes | X | X |
| Blood product transfusions (Red blood cells, plasma, platelets, other) | Binary indicator | X | X |
| Surgery | Binary indicator |  |  |
| Comorbidities |  |  |  |
| Acute Kidney Injury | Binary indicator from diagnoses | X | X |
| AIDS/HIV | Binary indicator from Elixhauser ICD codes | X |  |
| Alcohol Abuse | Binary indicator from Elixhauser ICD codes | X | X |
| Blood Loss Anemia | Binary indicator from Elixhauser ICD codes | X |  |
| Cardiac Arrhythmias | Binary indicator from Elixhauser ICD codes | X | X |
| Chronic Pulmonary Disease | Binary indicator from Elixhauser ICD codes | X |  |
| Coagulopathy | Binary indicator from Elixhauser ICD codes | X | X |
| Coma | Binary indicator of Richmond Agitation Sedation Scale <= -4 | X |  |
| Congestive Heart Failure | Binary indicator from Elixhauser ICD codes | X |  |
| Deficiency Anemia | Binary indicator from Elixhauser ICD codes | X | X |
| Depression | Binary indicator from Elixhauser ICD codes | X | X |
| Diabetes, Complicated | Binary indicator from Elixhauser ICD codes |  |  |
| Diabetes, Uncomplicated | Binary indicator from Elixhauser ICD codes | X | X |
| Drug Abuse | Binary indicator from Elixhauser ICD codes | X |  |
| Fluid and Electrolyte Disorders | Binary indicator from Elixhauser ICD codes | X |  |
| Hypertension, Complicated | Binary indicator from Elixhauser ICD codes | X | X |
| Hypertension, Uncomplicated | Binary indicator from Elixhauser ICD codes |  | X |
| Hypothyroidism | Binary indicator from Elixhauser ICD codes | X |  |
| Infection | Binary indicator from ICD codes | X | X |
| Liver Disease | Binary indicator from Elixhauser ICD codes |  |  |
| Lymphoma | Binary indicator from Elixhauser ICD codes |  | X |
| Metabolic Acidosis | Binary indicator from lab data |  | X |
| Metastatic Cancer | Binary indicator from Elixhauser ICD codes | X | X |
| Obesity | Binary indicator from Elixhauser ICD codes |  | X |
| Other Neurological Disorders | Binary indicator from Elixhauser ICD codes | X | X |
| Paralysis | Binary indicator from Elixhauser ICD codes | X | X |
| Peptic Ulcer Disease | Binary indicator from Elixhauser ICD codes | X |  |
| Peripheral Vascular Disorders | Binary indicator from Elixhauser ICD codes | X | X |
| Pulmonary Circulation Disorders | Binary indicator from Elixhauser ICD codes | X |  |
| Renal Failure | Binary indicator from Elixhauser ICD codes | X | X |
| Rheumatoid Arthritis/Collagen Vascular Diseases | Binary indicator from Elixhauser ICD codes | X |  |
| Sepsis | Binary indicator from calculated SOFA scores and infection flag | X | X |
| Solid Tumor Without Metastasis | Binary indicator from Elixhauser ICD codes | X |  |
| Valvular Disease | Binary indicator from Elixhauser ICD codes | X | X |
| Weight Loss | Binary indicator from Elixhauser ICD codes |  | X |
| Miscellaneous |  |  |  |
| Current Length of Stay | Minutes since ICU admission at prediction time |  | X |
| Current Time of Day | Seconds since midnight at prediction time |  | X |

## Table E3: Final specified hyper parameters of predictive models.

|  |
| --- |
| First 24-Hour Model Hyperparameters |
| Logistic Regression (sklearn) | penalty: l1C: 0.02class\_weight: balancedmax\_iter: 100solver: liblinear |
| Random Forest (sklearn) | n\_estimators: 750max\_depth: 50ccp\_alpha: 0.00013488757502313955max\_features: 0.06307313831160079max\_samples: 0.8211980594548445min\_samples\_leaf: 5min\_samples\_split: 3 |
| CatBoost | n\_estimators: 3000 |
| 12-Hour Lead Dynamic Model Hyperparameters |
| Logistic Regression (sklearn) | penalty: l1C: 0.02class\_weight: balancedmax\_iter: 100solver: liblinear |
| Random Forest (sklearn) | n\_estimators: 750max\_depth: 60ccp\_alpha: 0.0004713116875479681max\_samples: 0.9515116343018055min\_samples\_leaf: 4min\_samples\_split: 7 |
| CatBoost | n\_estimators: 3000 |

## Table E4: Modifications made to reference model features.

|  |  |  |
| --- | --- | --- |
| **Original Feature** | **Modified Feature** | **Reason** |
| Age | No modification | N/A |
| APACHE-II Score | APACHE-IV scores | Ease of calculation and likely improvement in using more modern version |
| Non-coma, drug-induced coma, miscellaneous coma, or combination coma. | Simplified non-coma, or coma. | Lack of documentation clarifying between different forms of coma |
| Surgical vs. medical vs. trauma vs. neurological patients | No modification | N/A |
| Infection | No modification | N/A |
| Metabolic acidosis | No modification | N/A |
| Morphine use at specified levels | Binary flag for morphine or other opioid use | Lack of precise documentation on dosage |
| Use of sedatives | No modification | N/A |
| Urea concentration | No modification | N/A |
| Urgent admission | No modification | N/A |

## Table E5: Evaluation metrics from the highest performing models on the development dataset.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model  | Algorithm | AUC (95% CI) | Mean Precision (95% CI) | Brier Score (95% CI) |
| First 24-Hour Model | CatBoost | 0.791 (0.775, 0.806) | 0.397 (0.369, 0.425) | 0.103 (0.097, 0.108) |
| Logistic Regression | 0.775 (0.749, 0.802) | 0.363 (0.333, 0.392) | 0.102 (0.095, 0.108) |
| Random Forest | 0.778 (0.757, 0.800) | 0.376 (0.345, 0.407) | 0.104 (0.098, 0.110) |
| Stacked | 0.786 (0.767, 0.805) | 0.389 (0.360, 0.418) | 0.103 (0.098, 0.107) |
| Dynamic (1 Hour Lead Time) | CatBoost | 0.899 (0.890, 0.907) | 0.665 (0.635, 0.695) | 0.091 (0.087, 0.095) |
| Logistic Regression | 0.836 (0.822, 0.850) | 0.493 (0.460, 0.527) | 0.111 (0.106, 0.115) |
| Random Forest | 0.881 (0.873, 0.888) | 0.610 (0.586, 633) | 0.096 (0.093, 0.099) |
| Stacked | 0.889 (0.879, 0.900) | 0.641 (0.605, 0.678) | 0.093 (0.088, 0.098) |
| Dynamic (3 Hour Lead Time) | CatBoost | 0.882 (0.875, 0.889) | 0.638 (0.612, 0.665) | 0.096 (0.092, 0.100) |
| Logistic Regression | 0.831 (0.821, 0.842) | 0.489 (0.459, 0.518) | 0.112 (0.107, 0.117) |
| Random Forest | 0.849 (0.828, 0.870) | 0.575 (0.533, 0.617) | 0.108 (0.101, 0.114) |
| Stacked | \* | \* | \* |
| Dynamic (6 Hour Lead Time) | CatBoost | 0.879 (0.871, 0.887) | 0.638 (0.612, 0.664) | 0.099 (0.096, 0.102) |
| Logistic Regression | 0.823 (0.813, 0.834) | 0.481 (0.462, 0.501) | 0.116 (0.113, 0.119) |
| Random Forest | \* | \* | \* |
| Stacked | \* | \* | \* |
| Dynamic (12 Hour Lead Time) | CatBoost | 0.845 (0.831, 0.859) | 0.590 (0.566, 0.613) | 0.111 (0.106, 0.116) |
| Logistic Regression | 0.807 (0.787, 0.827) | 0.478 (0.455, 0.501) | 0.123 (0.118, 0.129) |
| Random Forest | 0.828 (0.816, 0.839) | 0.548 (0.525, 0.572) | 0.118 (0.112, 0.123) |
| Stacked | \* | \* | \* |

\* Based on predictive performance from the first 24-hour models and short lead time dynamic models, and high computational needs of training stacked and random forest models, we did not use these algorithms for certain models.

## Table E6: Characteristics of cases and controls from validation dataset 1 for the first 24-hour model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | No Delirium | Delirium | Total | p-Value |
| Patient Characteristics |  |  |
| Sex |  |  |  | 0.683 |
| Male | 2,486 (54.9%) | 437 (56.9%) | 2,923 (55.2%) |  |
| Female | 2,045 (45.1%) | 331 (43.1%) | 2,376 (44.8%) |  |
| Median Age [Interquartile Range] (Years) | 65.4 [53.8-77.4] | 67.25 [55.3-78.5] | 65.68 [54.0-77.7] | 0.017 |
| Ethnicity |  |  |  | 0.395 |
| Caucasian | 3,122 (68.9%) | 502 (65.4%) | 3,624 (68.4%) |  |
| Other/Unknown | 537 (11.9%) | 145 (18.9%) | 682 (12.9%) |  |
| African American | 531 (11.7%) | 79 (10.3%) | 610 (11.5%) |  |
| Hispanic | 206 (4.6%) | 23 (3.0%) | 229 (4.3%) |  |
| Asian | 133 (2.9%) | 19 (2.5%) | 152 (2.9%) |  |
| Native American | 2 (0.04%) | 0 (0%) | 2 (0.04%) |  |
| Median First 24 Hour APACHE IV Score [Interquartile Range] | 62.0 [48.0-77.0] | 86.0 [70.0-105.0] | 65.0 [50.0-81.0] | <0.001 |
| Patient History |  |  |  |
| Dementia |  |  |  | 0.799 |
| No | 4,337 (96.6%) | 725 (96.2%) | 5,062 (96.6%) |  |
| Yes | 152 (3.4%) | 29 (3.9%) | 181 (3.5%) |  |
| Alcohol Abuse |  |  | 0.002 |
| No | 4,132 (91.2%) | 633 (82.4%) | 4,765 (89.9%) |  |
| Yes | 399 (8.8%) | 135 (17.6%) | 534 (10.1%) |  |
| Drug Abuse |  |  | 0.274 |
| No | 4,327 (95.5%) | 716 (93.2%) | 5,043 (95.2%) |  |
| Yes | 204 (4.5%) | 52 (6.8%) | 256 (4.8%) |  |
| Treatments in First 24hrs |  |  |
| Dialysis |  |  |  | 0.117 |
| No | 4,276 (94.4%) | 697 (90.8%) | 4,973 (93.9%) |  |
| Yes | 255 (5.6%) | 71 (9.2%) | 326 (6.1%) |  |
| Mechanical Ventilation |  |  | <0.001 |
| No | 2,895 (63.9%) | 254 (33.1%) | 3,149 (59.4%) |  |
| Yes | 1,636 (36.1%) | 514 (66.9%) | 2,150 (40.6%) |  |
| Primary Diagnostic Grouping |  |  | 0.476 |
| General Medicine | 2,349 (51.9%) | 392 (51.0%) | 2,741 (51.8%) |  |
| Cardiovascular | 1,112 (24.6%) | 149 (19.4%) | 1,261 (23.8%) |  |
| Neurologic | 541 (12.0%) | 119 (15.5%) | 660 (12.5%) |  |
| Musculoskeletal/Skin | 216 (4.8%) | 33 (4.3%) | 249 (4.7%) |  |
| Trauma | 155 (3.4%) | 57 (7.4%) | 212 (4.0%) |  |
| Thoracic | 105 (2.3%) | 15 (2.0%) | 120 (2.3%) |  |
| Genitourinary | 32 (0.7%) | 2 (0.3%) | 34 (0.6%) |  |
| Ear Nose Throat | 14 (0.3%) |  0 (0%) | 14 (0.3%) |  |
| Obstetrics | 3 (0.07%) | 1 (0.1%) | 4 (0.08%) |  |
| Outcomes |  |  |  |
| Median ICU Length of Stay [Interquartile Range] (Hours) | 50.4 [33.1-83.4] | 172.4 [101-291] | 56.9 [36.4-106.4] | <0.001 |
| Mortality |  |  |  | <0.001 |
| No | 4,196 (92.6%) | 636 (82.8%) | 4,832 (91.2%) |  |
| Yes | 335 (7.4%) | 132 (17.2%) | 467 (8.8%) |  |

## Table E7: Characteristics of cases and controls from validation dataset 2 for the first 24-hour model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | No Delirium | Delirium | Total | p-Value |
| Patient Characteristics |
| Sex |  |  |  | 0.84 |
| Male | 17,036 (56.3%) | 3,411 (57.3%) | 20,447 (56.5%) |  |
| Female | 13,208 (43.7%) | 2,539 (42.7%) | 15,747 (43.5%) |  |
| Median Age [Interquartile Range] (Years) | 64 [53-75] | 67 [55-77] | 65 [53-75] | <0.001 |
| Ethnicity |  |  |  | 0.661 |
| Caucasian | 20,547 (67.9%) | 3,727 (62.6%) | 24,274 (67.1%) |  |
| Other/Unknown | 4,331 (14.3%) | 1,207 (20.3%) | 5,538 (15.3%) |  |
| African American | 3,130 (10.3%) | 634 (10.7%) | 3,764 (10.4%) |  |
| Hispanic | 1,161 (3.8%) | 206 (3.5%) | 1,367 (3.8%) |  |
| Asian | 1,017 (3.4%) | 155 (2.6%) | 1,172 (3.2%) |  |
| Native American | 58 (0.2%) | 21 (0.4%) | 79 (0.2%) |  |
| Median First 24 Hour APACHE IV Score [Interquartile Range] | 57 [43-72] | 86 [68-107] | 61 [46-78] | <0.001 |
| Patient History |
| Dementia (Not in dataset) |  |  |  | N/A |
| N/A |  |  |  |  |
| Alcohol Abuse |  |  | 0.02 |
| No | 27,669 (91.5%) | 5,056 (85.0%) | 32,725 (90.4%) |  |
| Yes | 2,575 (8.5%) | 894 (15.0%) | 3,469 (9.6%) |  |
| Drug Abuse |  |  | 0.351 |
| No | 28,954 (95.7%) | 5,584 (93.8%) | 34,538 (95.4%) |  |
| Yes | 1,290 (4.3%) | 366 (6.2%) | 1,656 (4.6%) |  |
| Treatments in First 24hrs |
| Dialysis |  |  |  | 0.071 |
| No | 28,834 (95.3%) | 5,446 (91.5%) | 34,280 (94.7%) |  |
| Yes | 1,410 (4.7%) | 504 (8.5%) | 1,914 (5.3%) |  |
| Mechanical Ventilation | <0.001 |
| No | 20,140 (66.6%) | 2,256 (37.9%) | 22,396 (61.9%) |  |
| Yes | 10,104 (33.4%) | 3,694 (62.1%) | 13,798 (38.1%) |  |
| Primary Diagnostic Grouping |  |  | 0.621 |
| General Medicine | 12,980 (42.9%) | 2,993 (50.3%) | 15,973 (44.1%) |  |
| Cardiovascular | 9,739 (32.2%) | 1,306 (21.9%) | 11,045 (30.5%) |  |
| Neurologic | 4,583 (15.2%) | 973 (16.4%) | 5,556 (15.4%) |  |
| Musculoskeletal/Skin | 1,090 (3.6%) | 203 (3.4%) | 1,293 (3.6%) |  |
| Trauma | 824 (2.7%) | 306 (5.1%) | 1,130 (3.1%) |  |
| Thoracic | 597 (2.0%) | 125 (2.1%) | 722 (2.0%) |  |
| Genitourinary | 221 (0.7%) | 27 (0.5%) | 248 (0.7%) |  |
| Ear Nose Throat | 125 (0.4%) | 13 (0.2%) | 138 (0.4%) |  |
| Obstetrics | 84 (0.3%) | 4 (0.1%) | 88 (0.2%) |  |
| Outcomes |
| Median ICU Length of Stay [Interquartile Range] (Hours) | 48.7 [33.1-75.5] | 166.3 [97.4-294.2] | 54.7 [36.1-98.2] | <0.001 |
| Mortality |  |  |  | <0.001 |
| No | 28,570 (94.5%) | 4,730 (79.5%) | 33,300 (92.0%) |  |
| Yes | 1,674 (5.5%) | 1,220 (20.5%) | 2,894 (8.0%) |  |

## Table E8: Mean (95% CI) area under the receiver operating curve when using nested cross validation, across various lead times, comparing performance with and without high-frequency physiological time series features.

|  |  |  |
| --- | --- | --- |
| Lead Time (hours) | With physiological time series-based features | Without physiological time series-based features |
| 1 | 0.856 (0.84, 0.871) | 0.853 (0.836, 0.870) |
| 3 | 0.849 (0.828, 0.87) | 0.848 (0.836, 0.860) |
| 6 | 0.841 (0.833, 0.848) | 0.84 (0.828, 0.851) |

## Table E9: Comparison of development characteristics from current model with prediction models from recent literature.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Prediction Model | Years of Study | Primary Outcome | Outcome Incidence | Sample Size | Patient Description | Internal AUROC | External AUROC | Prediction Time Point | Hospitals in Study |
| Current Model | 2014-2015 | Confusion Assessment Matrix-ICU or Intensive Care Delirium Screening Checklist | 17.0% | 22,234 | Age >= 18, in ICU | up to 0.899 | up to 0.912 | 1-12 hours before delirium | 111 |
| PRIDE E1 | 2016-2019 | Confusion Assessment Matrix-ICU | 30.8% | 2,061 | Age >= 18, Medical ICU or Surgical ICU | 0.919 | 0.697 | First 24 hours of ICU stay | 1 |
| PRE-DELIRIC E2, E3 | 2008-2009 | Confusion Assessment Matrix-ICU,Diagnostic and statistical Manual of Mental Disorders -IV | 25.5% | 4,033 | Age >= 18, in ICU | 0.844 | N/A (meta-analysis) | First 24 hours of ICU Stay | 5 |
| ABD-pm E4 | 2007-2010 | Confusion Assessment Matrix-ICU | 75% | 810 | Age >= 18, Medical ICU or Surgical ICU, had respiratory failure or shock | 0.79 | None | Daily | 2 |
| Auto-DelRAS E5 | 2009-2012 | Confusion Assessment Matrix-ICU | 21.1% | 4,304 | Age >= 18,Medical ICU or Surgical ICU | 0.89 | 0.72 | Daily at midnight | 2 |
| DYNAMIC-ICU E6 | 2009-2010 | Modified Confusion Assessment Matrix-ICU | 20.2% | 560 | Age >= 18, no neurological disease, no sedation | 0.9 | None | Daily | 1 |

## Supplemental Digital Content Figure Legends

Figure E1: Model development process. 66

Figure E2: AUC plots comparing the performance of different algorithms for the first 24-hour model. 67

Figure E3: AUC plots comparing the performance of different algorithms for the dynamic model at 1-hour lead time. 68

Figure E4: AUC performance of the dynamic model at 1-hour lead time. 69

Figure E5: Precision recall curves of the dynamic model at 1-hour lead time. 70

Figure E6: Calibration curve and Brier score of the dynamic model at 1-hour lead time. 71

Figure E7: Feature analysis of the first 24-hour model**.** Shapley summary plot of the top 20 features for the dynamic model, at 1-hour lead time for prediction of delirium. Each dot represents the Shapley value of one sample for that feature. A feature’s Shapley value represents the association of that feature to the risk score, with positive values indicating an association with a higher risk of delirium, and negative values indicating an association with a lower risk of delirium. The location of the dot on the x-axis represents its Shapley value, while its color represents the feature’s absolute value. For example, low age is associated with low Shapley values, while high age is associated with high Shapley values, indicating that elderly patients are at higher risk for delirium. 72

Figure E8: Risk scores over time for one positive patient, obtained using the output of the 1-hour lead time dynamic model, showing minor fluctuations and a major one shortly before delirium onset. 73



Figure E1: Model development process.



Figure E2: AUC plots comparing the performance of different algorithms for the first 24-hour model.

 

Figure E3: AUC plots comparing the performance of different algorithms for the dynamic model at 1-hour lead time.



Figure E4: AUC performance of the dynamic model at 1-hour lead time.



Figure E5: Precision recall curves of the dynamic model at 1-hour lead time.



Figure E6: Calibration curve and Brier score of the dynamic model at 1-hour lead time.



Figure E7: Feature analysis of the first 24-hour model**.** Shapley summary plot of the top 20 features for the dynamic model, at 1-hour lead time for prediction of delirium. Each dot represents the Shapley value of one sample for that feature. A feature’s Shapley value represents the association of that feature to the risk score, with positive values indicating an association with a higher risk of delirium, and negative values indicating an association with a lower risk of delirium. The location of the dot on the x-axis represents its Shapley value, while its color represents the feature’s absolute value. For example, low age is associated with low Shapley values, while high age is associated with high Shapley values, indicating that elderly patients are at higher risk for delirium.



Figure E8: Risk scores over time for one positive patient, obtained using the output of the 1-hour lead time dynamic model, showing minor fluctuations and a major one shortly before delirium onset.