# Supplement

## Supplemental Figures and Tables

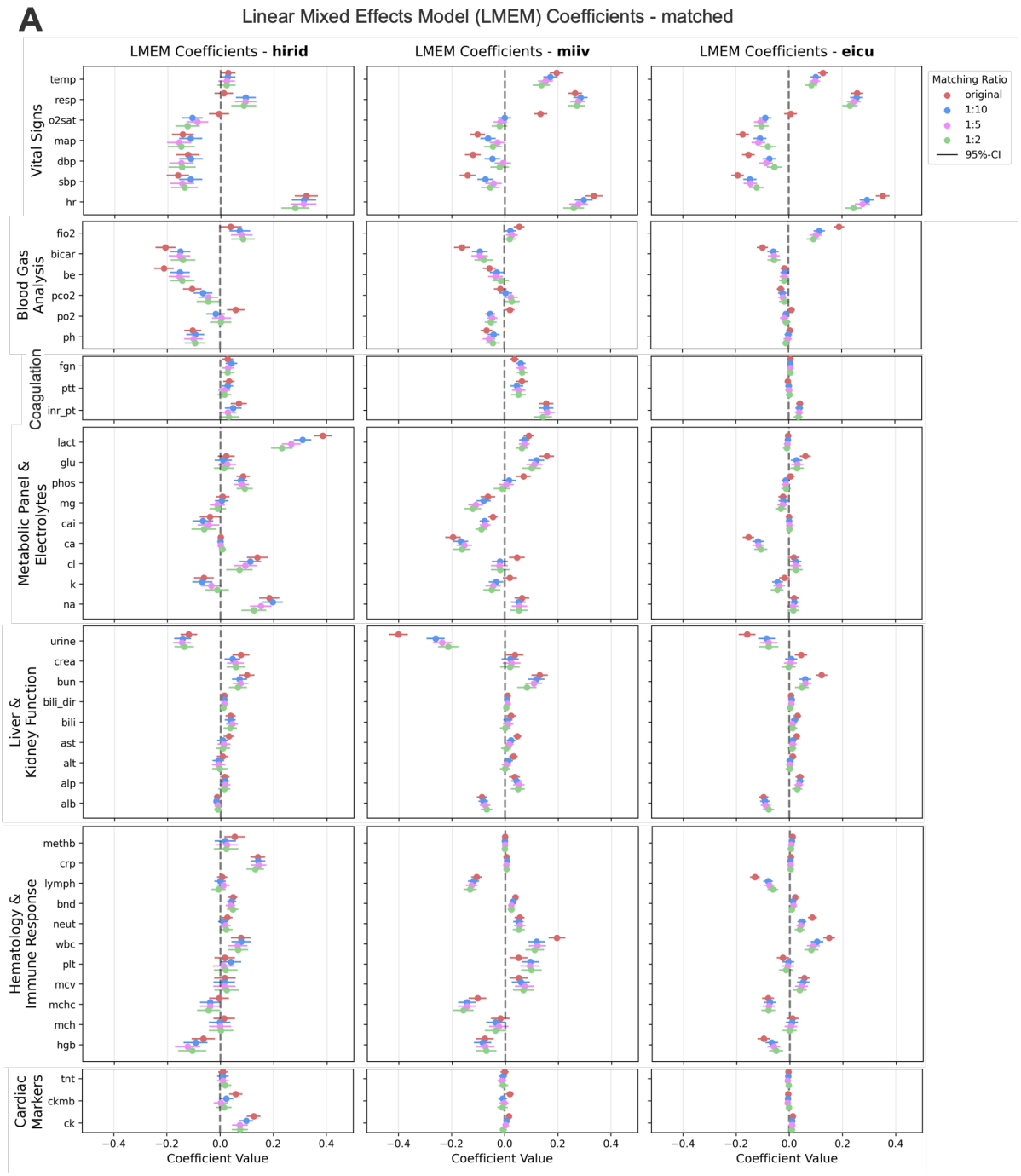


**Figure 1. Absolute LMEM coefficients across datasets (original vs. matched). A)** LMEM coefficients exhibit consistent patterns for dynamic features across all three datasets with higher magnitudes in original versus matched training cohorts, indicating attenuated discriminative capacity between sepsis and non-sepsis groups. Coefficients are plotted on the same scale for all grouped dynamic features and for all matched training cohort matching ratios. Positive coefficients indicate elevated feature values in the sepsis group relative to the non-sepsis group, whereas negative coefficients signify reduced values in sepsis cases. A coefficient of 1.0 denotes that the sepsis group exhibits, on average, a one-unit increase in the feature value compared to controls, after accounting for within-subject correlations via the mixed-effects model. All coefficients are presented with their respective 95% confidence intervals.

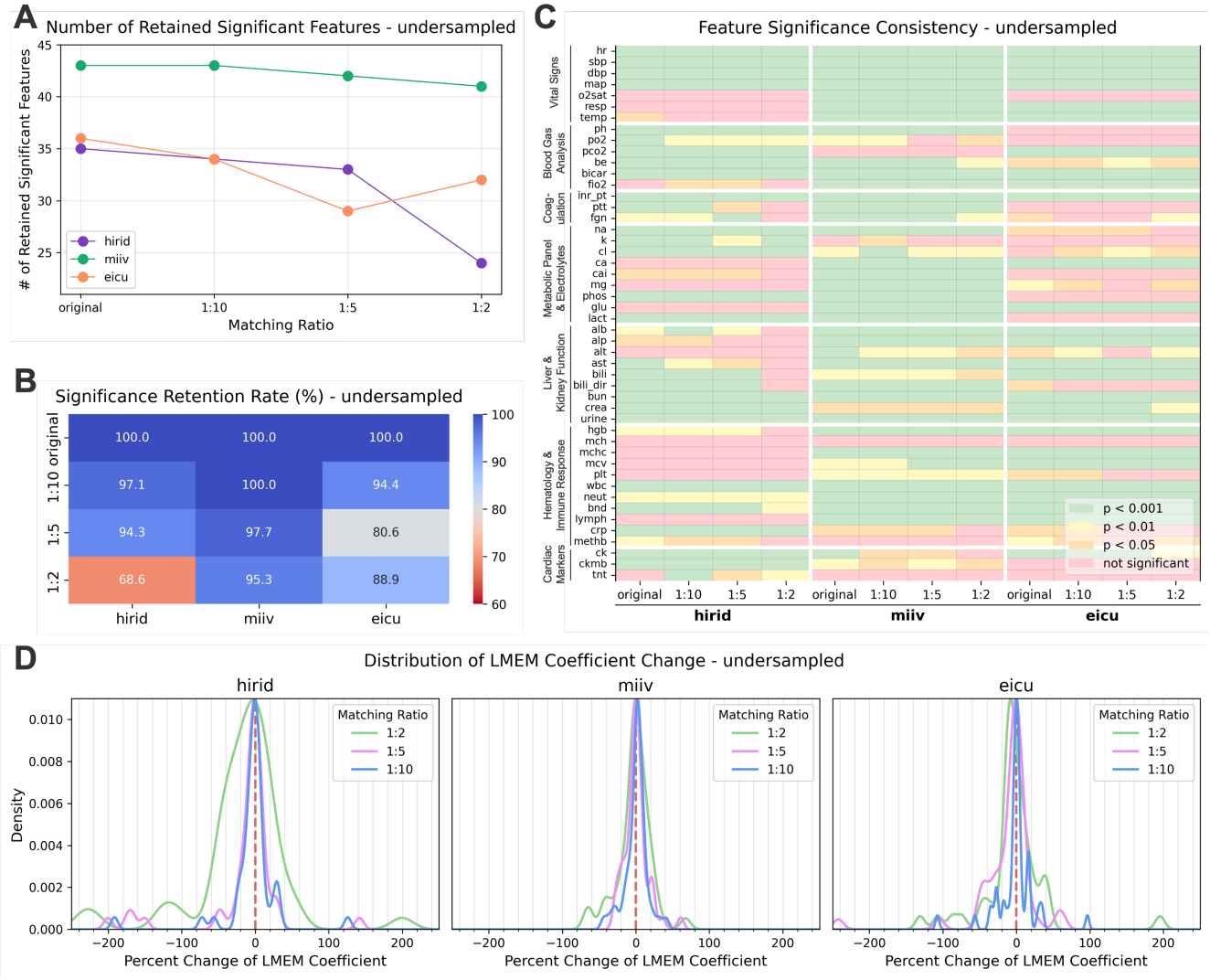


**Figure 2. Evaluation of LMEM coefficient alterations and their statistical significance (original vs. undersampled). A)** The absolute count of retained significant features in original versus undersampled training cohorts exhibits minimal reduction at 1:10 and 1:5 matching ratios. At 1:2 matching ratios, dataset-specific patterns emerge: MIMIC-IV and, to a slightly lesser extent, eICU preserve nearly all significant features, whereas HiRID demonstrates a substantial decrease. **B)** The relative significance retention rate similarly demonstrates dataset-specific patterns, with MIMIC-IV maintaining high preservation and HiRID exhibiting diminished retention at the 1:2 matching ratio. **C)** Analysis of significance levels reveals distinct heterogeneity in feature significance profiles across datasets. **D)** The percent change distribution in LMEM coefficients clusters around the original baseline with predominantly minor deviations, suggesting preserved discriminative capacity of collective dynamic features for sepsis versus non-sepsis cohort differentiation.

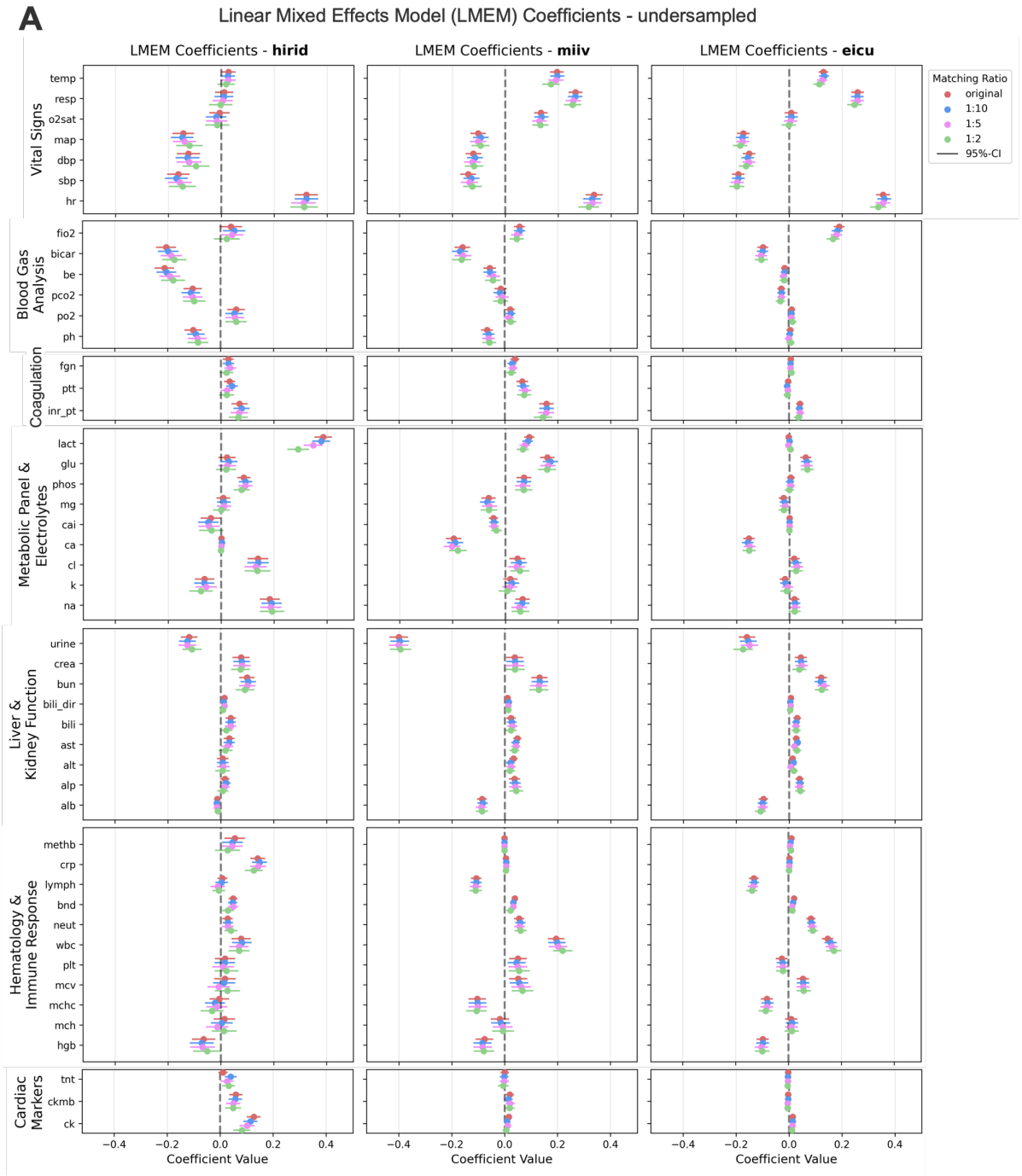


**Figure 3. Absolute LMEM coefficients across datasets (original vs. undersampled). A)** LMEM coefficients demonstrate negligible variation between original and undersampled training cohorts. Coefficients are plotted on the same scale for all grouped dynamic features and for all undersampled training cohort matching ratios. Positive coefficients indicate elevated feature values in the sepsis group relative to the non-sepsis group, whereas negative coefficients signify reduced values in sepsis cases. A coefficient of 1.0 denotes that the sepsis group exhibits, on average, a one-unit increase in the feature value compared to controls, after accounting for within-subject correlations via the mixed-effects model. All coefficients are presented with their respective 95% confidence intervals.

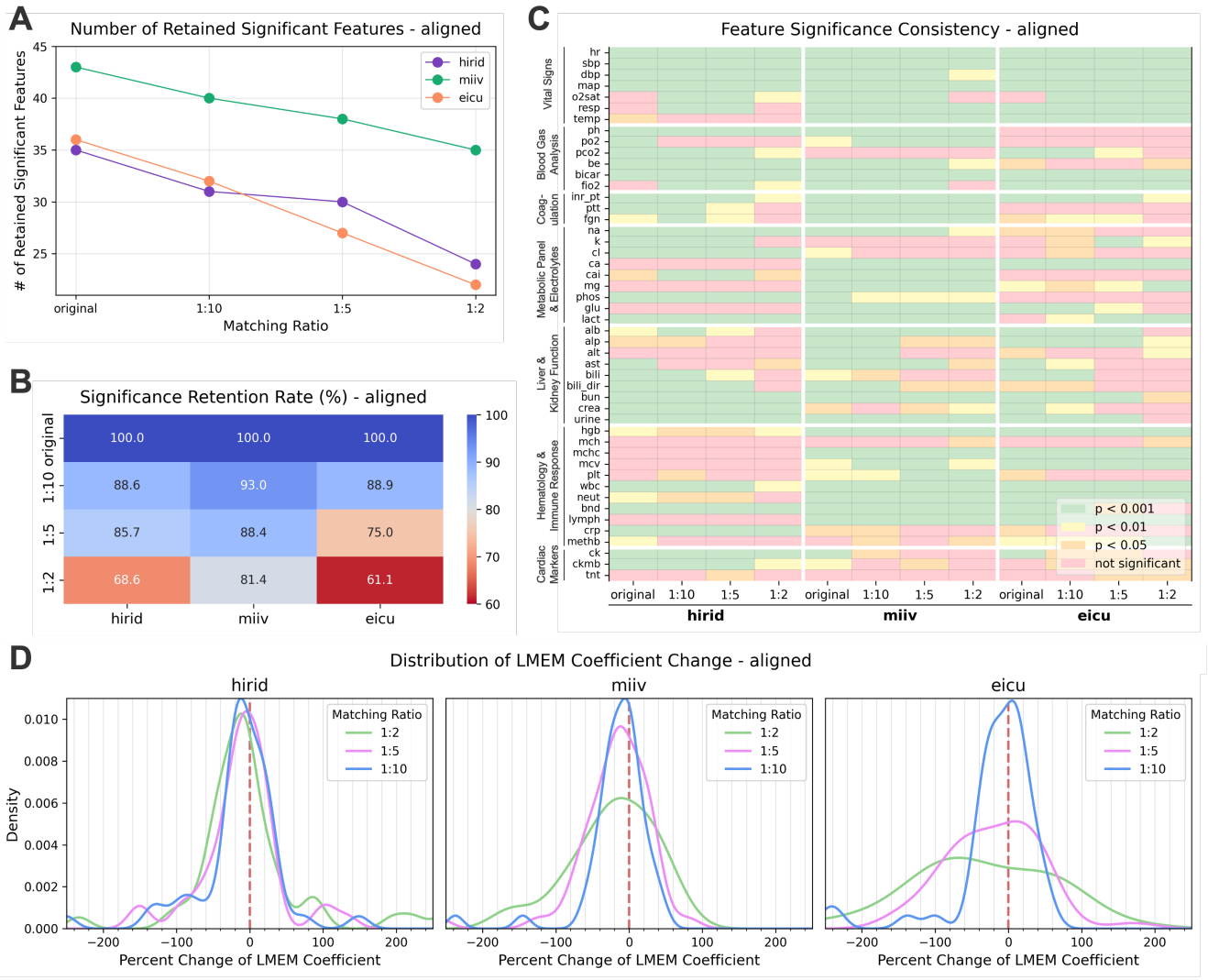


**Figure 4. Evaluation of LMEM coefficient alterations and their statistical significance (original vs. aligned). A)** The absolute number of retained significant features in aligned training cohorts compared to original training sets demonstrates an inverse relationship with increasing matching ratio. This relationship is comparable to results from matched training cohorts. **B)** The relative significance retention rate demonstrates attenuated decrements with increasing matching ratios in MIMIC-IV compared to HiRID and eICU. **C)** Analysis of significance levels reveals distinct heterogeneity in feature significance profiles across datasets. **D)** The distribution shift toward negative percent changes in LMEM coefficients indicates reduced discriminative capacity of the collective dynamic features for differentiating between sepsis and non-sepsis cohorts. Percent change values exhibited greater dispersion for eICU matching ratios of 1:5 and 1:2.

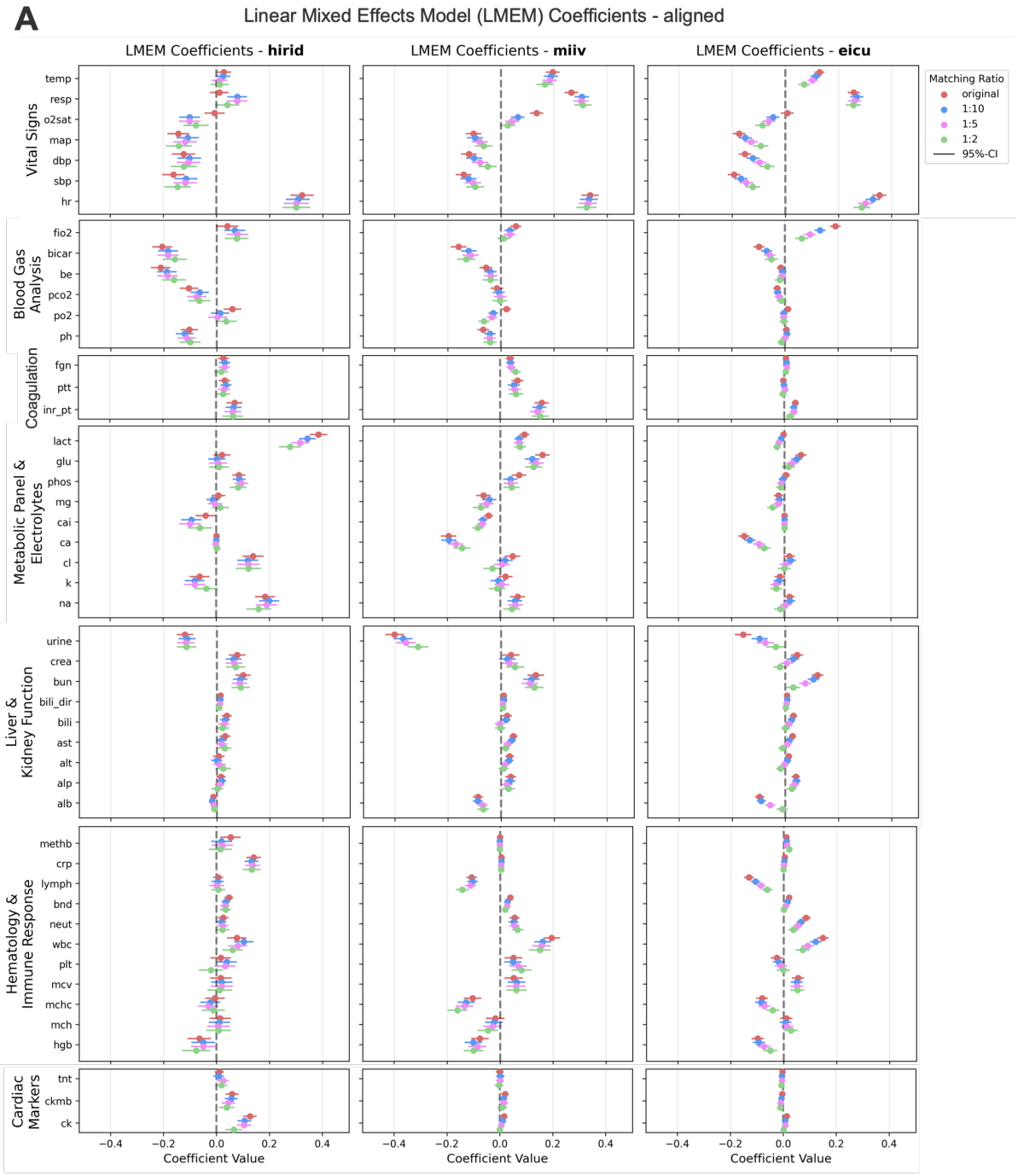


**Figure 5. Absolute LMEM coefficients across datasets (original vs. aligned). A)** LMEM coefficients parallel those derived from LMEM analysis of matched training cohorts. Coefficients are plotted on the same scale for all grouped dynamic features and for all aligned training cohort matching ratios. Positive coefficients indicate elevated feature values in the sepsis group relative to the non-sepsis group, whereas negative coefficients signify reduced values in sepsis cases. A coefficient of 1.0 denotes that the sepsis group exhibits, on average, a one-unit increase in the feature value compared to controls, after accounting for within-subject correlations via the mixed-effects model. All coefficients are presented with their respective 95% confidence intervals.

| Category | | Feature | ricu | *unit* | Reference Range | | Thinkable Range* | |
|---|---|---|---|---|---|---|---|---|
| | | | | | lower | upper | lower | upper |
| Static Features | | sex | sex | *M/F* | | | | |
| | | age | age | *years* | | | | |
| | | height | height | *cm* | | | 135 | 225 |
| | | weight | weight | *kg* | | | 40 | 250 |
| Dynamic Features | Vital Signs | Heart Rate | hr | *bpm* | 60 | 100 | 20 | 320 |
| | | Systolic Blood Pressure | sbp | *mmHg* | 90 | 120 | 30 | 300 |
| | | Diastolic Blood Pressure | dbp | *mmHg* | 60 | 80 | 10 | 200 |
| | | Mean Arterial Pressure (MAP) | map | *mmHg* | 65 | 100 | 20 | 250 |
| | | Oxygen Saturation | o2sat | *%* | 95 | 100 | 50 | 100 |
| | | Respiratory Rate | resp | */min* | 12 | 20 | 4 | 80 |
| | | Temperature | temp | *°C* | 36.5 | 37.5 | 30 | 42 |
| | Blood Gas Analysis | pH Level | ph | – | 7.35 | 7.45 | 6.7 | 8 |
| | | Partial Pressure of Oxygen (PaO2) | po2 | *mmHg* | 75 | 100 | 40 | 600 |
| | | Partial Pressure of Carbon Dioxide (PaCO2) | pco2 | *mmHg* | 35 | 45 | 10 | 150 |
| | | Base Excess | be | *mmol/L* | -2 | 2 | -25 | 25 |
| | | Bicarbonate | bicar | *mmol/L* | 22 | 29 | 5 | 50 |
| | | Fraction of Inspired Oxygen (FiO2) | fio2 | *%* | 21 | 100 | 21 | 100 |
| | Coagulation | International Normalized Ratio (INR) | inr_pt | – | 0.8 | 1.2 | 0.5 | 20 |
| | | Partial Thromboplastin Time (PTT) | ptt | *sec* | 25 | 35 | 10 | 250 |
| | | Fibrinogen | fgn | *mg/dL* | 200 | 400 | 30 | 1100 |
| | Metabolic Panel & Electrolytes | Sodium | na | *mmol/L* | 135 | 145 | 90 | 170 |
| | | Potassium | k | *mmol/L* | 3.5 | 5 | 1 | 9 |
| | | Chloride | cl | *mmol/L* | 96 | 106 | 70 | 140 |
| | | Calcium | ca | *mg/dL* | 8.5 | 10.5 | 4 | 20 |
| | | Ionized Calcium | cai | *mmol/L* | 1.1 | 1.3 | 0.4 | 2.2 |
| | | Magnesium | mg | *mg/dL* | 1.7 | 2.2 | 0.5 | 5 |
| | | Phosphate | phos | *mg/dL* | 2.5 | 4.5 | 0.5 | 15 |
| | | Glucose | glu | *mg/dL* | 70 | 140 | 25 | 1000 |
| | | Lactate | lact | *mmol/L* | 0.5 | 2 | 0.1 | 20 |
| | Liver & Kidney Function | Albumin | alb | *g/dL* | 3.5 | 5 | 0.5 | 6 |
| | | Alkaline Phosphatase | alp | *U/L* | 44 | 147 | 10 | 1200 |
| | | Alanine Aminotransferase (ALT) | alt | *U/L* | 7 | 56 | 10 | 5000 |
| | | Aspartate Aminotransferase (AST) | ast | *U/L* | 10 | 40 | 10 | 8000 |
| | | Total Bilirubin | bili | *mg/dL* | 0.1 | 1.2 | 0.1 | 50 |
| | | Direct Bilirubin | bili_di | *mg/dL* | 0 | 0.3 | 0 | 30 |
| | | Blood Urea Nitrogen (BUN) | bun | *mg/dL* | 7 | 20 | 1 | 180 |
| | | Creatinine | crea | *mg/dL* | 0.6 | 1.3 | 0.1 | 20 |
| | Hematology & Immune Response | Hemoglobin | hgb | *g/dL* | 13.5 | 17.5 | 3 | 20 |
| | | Mean Corpuscular Hemoglobin (MCH) | mch | *pg* | 27 | 33 | 15 | 45 |
| | | Mean Corpuscular Hemoglobin Concentration (MCHC) | mchc | *g/dL* | 32 | 36 | 20 | 45 |
| | | Mean Corpuscular Volume (MCV) | mcv | *fL* | 80 | 100 | 50 | 130 |
| | | Platelets | plt | $10^3/\mu L$ | 150 | 450 | 10 | 1500 |
| | | White Blood Cell Count (WBC) | wbc | $10^3/\mu L$ | 4 | 11 | 0.1 | 500 |
| | | Neutrophils | neut | *%* | 55 | 70 | 0 | 100 |
| | | Band Neutrophils | bnd | *%* | 0 | 6 | 0 | 50 |
| | | Lymphocytes | lymph | *%* | 20 | 40 | 0 | 90 |
| | | C-Reactive Protein (CRP) | crp | *mg/L* | 0 | 10 | 0 | 500 |
| | | Methemoglobin | methb | *%* | 0 | 2 | 0 | 60 |
| | Cardiac Markers | Creatine Kinase (CK) | ck | *U/L* | 30 | 200 | 10 | 100000 |
| | | Creatine Kinase-MB (CK-MB) | ckmb | *ng/mL* | 0 | 5 | 0 | 500 |
| | | Troponin T | tnt | *ng/mL* | 0 | 14 | 0 | 1000 |

**Table 1. Overview of features in harmonized datasets.** Clinical concepts were extracted using `ricu`. Reference and thinkable ranges were defined after harmonization to facilitate data interpretation and cleaning. *Ranges were defined with generous boundaries and with respect to the circumstances of ICU stays. Outliers that were under or over the limits by more than 5% of the thinkable range were replaced with NaNs.

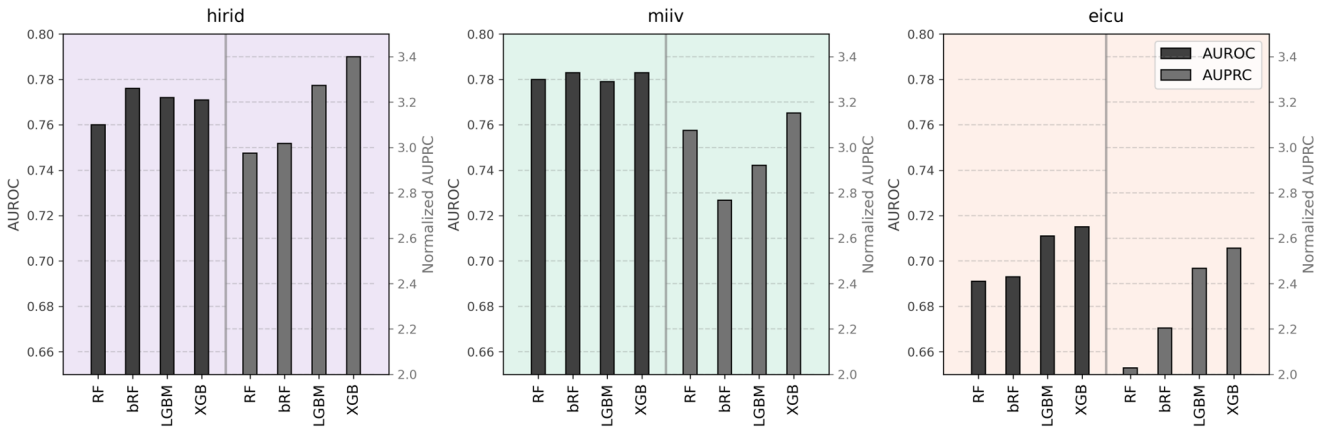


**Figure 6.** AUROC and nAUPRC metrics of machine learning models following Bayesian hyperparameter optimization, trained and evaluated on original datasets.

| | | | Models Analyzed (Σ 108) | | Metric Change | |
|---|---|---|---|---|---|---|
| | | | **Count** | **Percentage** | **Mean** | **Standard Deviation** |
| Differences in Model Performance (Matched vs. Aligned) | AUROC | ↑/= | 58 | 54% | 0,019 | 0,017 |
| | | ↓ | 50 | 46% | -0,017 | 0,014 |
| | nAUPRC | ↑/= | 54 | 50% | 0,025 | 0,030 |
| | | ↓ | 54 | 50% | -0,040 | 0,027 |
| | AUPRC | ↑/= | 63 | 58% | 0,002 | 0,002 |
| | | ↓ | 45 | 42% | -0,003 | 0,003 |

**Table 2. Differences in model performance metrics between matched and aligned cohorts.** Arrows indicate performance improvement (↑), no change (=), or degradation (↓) of models trained on matched training sets (across all matching ratios) when compared to models trained on aligned training sets. A comprehensive evaluation was conducted across 108 distinct models, comprising a factorial design of 3 datasets, 4 model algorithms, 3 data-prediction-window configurations, and 3 matching ratios. The comparative analysis of matched versus aligned sampling methods reveals that the incorporation of additional demographic matching criteria produces minimal impact on model performance. The metrics AUROC, nAUPRC, and AUPRC demonstrate comparable magnitudes of change, with approximately half of the models showing slight improvements and half showing slight decrements in performance across each metric.

**AUPRC**

| | | | original | undersampled | | | temporal alignment | | | demographics matching | | | demographics + temporal | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 1:2 | 1:5 | 1:10 | 1:2 | 1:5 | 1:10 | 1:2 | 1:5 | 1:10 | 1:2 | 1:5 | 1:10 |
| 1h data / 0h pred | hirid | RF | 0.061 | 0.881 | 0.899 | 0.950 | 0.037 | 0.025 | 0.020 | 0.030 | 0.021 | 0.020 | 0.028 | 0.018 | 0.018 |
| | | bRF | 0.070 | 0.883 | 0.881 | 0.916 | 0.034 | 0.024 | 0.018 | 0.032 | 0.023 | 0.021 | 0.028 | 0.020 | 0.019 |
| | | LGBM | 0.060 | 0.147 | 0.202 | 0.319 | 0.024 | 0.021 | 0.020 | 0.026 | 0.022 | 0.023 | 0.020 | 0.019 | 0.019 |
| | | XGB | 0.062 | 0.188 | 0.338 | 0.416 | 0.026 | 0.023 | 0.018 | 0.024 | 0.021 | 0.021 | 0.021 | 0.020 | 0.019 |
| | | LR | 0.055 | 0.054 | 0.057 | 0.057 | 0.021 | 0.025 | 0.020 | 0.019 | 0.026 | 0.022 | 0.021 | 0.024 | 0.026 |
| | | CNN | 0.072 | 0.064 | 0.084 | 0.097 | 0.028 | 0.028 | 0.028 | 0.030 | 0.035 | 0.028 | 0.030 | 0.026 | 0.027 |
| | miiv | RF | 0.029 | 0.865 | 0.948 | 0.939 | 0.019 | 0.010 | 0.010 | 0.027 | 0.012 | 0.010 | 0.021 | 0.015 | 0.010 |
| | | bRF | 0.035 | 0.826 | 0.919 | 0.909 | 0.021 | 0.012 | 0.011 | 0.024 | 0.014 | 0.012 | 0.017 | 0.015 | 0.011 |
| | | LGBM | 0.034 | 0.073 | 0.098 | 0.148 | 0.012 | 0.011 | 0.011 | 0.012 | 0.012 | 0.010 | 0.012 | 0.011 | 0.010 |
| | | XGB | 0.034 | 0.099 | 0.170 | 0.269 | 0.011 | 0.011 | 0.011 | 0.015 | 0.011 | 0.010 | 0.012 | 0.012 | 0.011 |
| | | LR | 0.033 | 0.034 | 0.034 | 0.034 | 0.012 | 0.012 | 0.013 | 0.012 | 0.013 | 0.013 | 0.012 | 0.011 | 0.013 |
| | | CNN | 0.038 | 0.043 | 0.051 | 0.059 | 0.016 | 0.015 | 0.013 | 0.017 | 0.017 | 0.013 | 0.016 | 0.014 | 0.013 |
| | eicu | RF | 0.017 | 0.891 | 0.885 | 0.906 | 0.016 | 0.011 | 0.011 | 0.014 | 0.010 | 0.011 | 0.024 | 0.012 | 0.009 |
| | | bRF | 0.025 | 0.865 | 0.851 | 0.878 | 0.018 | 0.015 | 0.011 | 0.016 | 0.012 | 0.013 | 0.026 | 0.013 | 0.011 |
| | | LGBM | 0.025 | 0.047 | 0.059 | 0.085 | 0.011 | 0.011 | 0.011 | 0.010 | 0.010 | 0.011 | 0.011 | 0.010 | 0.011 |
| | | XGB | 0.023 | 0.057 | 0.106 | 0.119 | 0.011 | 0.011 | 0.012 | 0.010 | 0.011 | 0.010 | 0.011 | 0.011 | 0.011 |
| | | LR | 0.020 | 0.020 | 0.020 | 0.020 | 0.012 | 0.012 | 0.012 | 0.011 | 0.011 | 0.013 | 0.011 | 0.011 | 0.011 |
| | | CNN | 0.027 | 0.029 | 0.034 | 0.038 | 0.012 | 0.012 | 0.014 | 0.011 | 0.012 | 0.012 | 0.012 | 0.011 | 0.011 |
| 6h data / 0h pred | hirid | RF | 0.069 | 0.871 | 0.895 | 0.945 | 0.040 | 0.024 | 0.023 | 0.031 | 0.022 | 0.023 | 0.032 | 0.021 | 0.019 |
| | | bRF | 0.078 | 0.812 | 0.862 | 0.921 | 0.036 | 0.024 | 0.021 | 0.033 | 0.027 | 0.024 | 0.036 | 0.023 | 0.021 |
| | | LGBM | 0.086 | 0.204 | 0.287 | 0.399 | 0.025 | 0.025 | 0.021 | 0.022 | 0.024 | 0.025 | 0.022 | 0.023 | 0.023 |
| | | XGB | 0.086 | 0.322 | 0.415 | 0.498 | 0.027 | 0.020 | 0.019 | 0.022 | 0.024 | 0.023 | 0.022 | 0.021 | 0.022 |
| | | LR | 0.077 | 0.075 | 0.080 | 0.081 | 0.025 | 0.025 | 0.026 | 0.022 | 0.029 | 0.022 | 0.022 | 0.029 | 0.027 |
| | | CNN | 0.090 | 0.141 | 0.214 | 0.296 | 0.029 | 0.028 | 0.022 | 0.025 | 0.030 | 0.027 | 0.026 | 0.023 | 0.021 |
| | miiv | RF | 0.043 | 0.847 | 0.944 | 0.942 | 0.020 | 0.010 | 0.010 | 0.027 | 0.013 | 0.011 | 0.021 | 0.015 | 0.010 |
| | | bRF | 0.049 | 0.689 | 0.852 | 0.888 | 0.021 | 0.012 | 0.012 | 0.026 | 0.015 | 0.012 | 0.018 | 0.015 | 0.012 |
| | | LGBM | 0.048 | 0.098 | 0.143 | 0.190 | 0.011 | 0.011 | 0.011 | 0.013 | 0.012 | 0.010 | 0.012 | 0.011 | 0.010 |
| | | XGB | 0.052 | 0.148 | 0.232 | 0.318 | 0.011 | 0.010 | 0.011 | 0.013 | 0.012 | 0.010 | 0.012 | 0.012 | 0.011 |
| | | LR | 0.046 | 0.049 | 0.048 | 0.049 | 0.013 | 0.011 | 0.017 | 0.013 | 0.015 | 0.019 | 0.012 | 0.013 | 0.013 |
| | | CNN | 0.065 | 0.102 | 0.165 | 0.235 | 0.013 | 0.011 | 0.012 | 0.017 | 0.013 | 0.012 | 0.013 | 0.013 | 0.010 |
| | eicu | RF | 0.022 | 0.882 | 0.884 | 0.905 | 0.017 | 0.012 | 0.010 | 0.013 | 0.010 | 0.011 | 0.024 | 0.014 | 0.010 |
| | | bRF | 0.033 | 0.734 | 0.780 | 0.835 | 0.020 | 0.016 | 0.014 | 0.015 | 0.013 | 0.014 | 0.027 | 0.016 | 0.013 |
| | | LGBM | 0.037 | 0.067 | 0.072 | 0.062 | 0.011 | 0.012 | 0.012 | 0.010 | 0.012 | 0.011 | 0.012 | 0.011 | 0.011 |
| | | XGB | 0.034 | 0.095 | 0.123 | 0.174 | 0.012 | 0.012 | 0.012 | 0.010 | 0.011 | 0.012 | 0.012 | 0.011 | 0.011 |
| | | LR | 0.026 | 0.027 | 0.028 | 0.027 | 0.012 | 0.013 | 0.014 | 0.012 | 0.014 | 0.015 | 0.013 | 0.012 | 0.012 |
| | | CNN | 0.043 | 0.069 | 0.095 | 0.139 | 0.012 | 0.013 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.011 | 0.012 |
| 6h data / 4h pred | hirid | RF | 0.064 | 0.872 | 0.895 | 0.941 | 0.044 | 0.020 | 0.022 | 0.031 | 0.021 | 0.023 | 0.030 | 0.021 | 0.019 |
| | | bRF | 0.075 | 0.722 | 0.828 | 0.911 | 0.037 | 0.022 | 0.023 | 0.035 | 0.026 | 0.025 | 0.033 | 0.026 | 0.021 |
| | | LGBM | 0.073 | 0.206 | 0.271 | 0.383 | 0.021 | 0.025 | 0.019 | 0.024 | 0.021 | 0.022 | 0.023 | 0.020 | 0.022 |
| | | XGB | 0.085 | 0.312 | 0.402 | 0.586 | 0.031 | 0.023 | 0.019 | 0.024 | 0.024 | 0.020 | 0.027 | 0.022 | 0.020 |
| | | LR | 0.070 | 0.069 | 0.074 | 0.074 | 0.022 | 0.023 | 0.027 | 0.022 | 0.030 | 0.021 | 0.021 | 0.028 | 0.028 |
| | | CNN | 0.086 | 0.140 | 0.200 | 0.313 | 0.025 | 0.029 | 0.020 | 0.025 | 0.029 | 0.025 | 0.025 | 0.021 | 0.021 |
| | miiv | RF | 0.037 | 0.839 | 0.937 | 0.940 | 0.020 | 0.011 | 0.009 | 0.019 | 0.010 | 0.009 | 0.018 | 0.013 | 0.009 |
| | | bRF | 0.042 | 0.519 | 0.776 | 0.856 | 0.022 | 0.011 | 0.012 | 0.021 | 0.015 | 0.011 | 0.017 | 0.014 | 0.012 |
| | | LGBM | 0.037 | 0.092 | 0.130 | 0.172 | 0.010 | 0.013 | 0.010 | 0.011 | 0.011 | 0.009 | 0.011 | 0.010 | 0.009 |
| | | XGB | 0.039 | 0.112 | 0.257 | 0.338 | 0.011 | 0.011 | 0.010 | 0.010 | 0.011 | 0.009 | 0.011 | 0.010 | 0.010 |
| | | LR | 0.035 | 0.038 | 0.037 | 0.038 | 0.011 | 0.010 | 0.012 | 0.012 | 0.013 | 0.013 | 0.011 | 0.012 | 0.011 |
| | | CNN | 0.055 | 0.092 | 0.150 | 0.228 | 0.012 | 0.010 | 0.010 | 0.013 | 0.012 | 0.010 | 0.011 | 0.013 | 0.010 |
| | eicu | RF | 0.021 | 0.882 | 0.886 | 0.906 | 0.016 | 0.012 | 0.010 | 0.012 | 0.010 | 0.011 | 0.023 | 0.012 | 0.010 |
| | | bRF | 0.031 | 0.606 | 0.722 | 0.813 | 0.021 | 0.017 | 0.014 | 0.015 | 0.014 | 0.016 | 0.027 | 0.017 | 0.013 |
| | | LGBM | 0.029 | 0.061 | 0.072 | 0.083 | 0.011 | 0.011 | 0.012 | 0.010 | 0.011 | 0.011 | 0.011 | 0.011 | 0.012 |
| | | XGB | 0.030 | 0.093 | 0.132 | 0.169 | 0.011 | 0.013 | 0.011 | 0.010 | 0.011 | 0.011 | 0.011 | 0.012 | 0.011 |
| | | LR | 0.024 | 0.025 | 0.025 | 0.025 | 0.011 | 0.012 | 0.013 | 0.011 | 0.013 | 0.013 | 0.012 | 0.011 | 0.011 |
| | | CNN | 0.040 | 0.072 | 0.086 | 0.113 | 0.011 | 0.012 | 0.011 | 0.012 | 0.011 | 0.012 | 0.011 | 0.011 | 0.011 |

**Table 3. AUPRC Performance of machine learning models.** Comparative analysis of non-optimized (vanilla) machine learning models trained on specified training datasets (original, demographics-excluded, as well as undersampled, aligned, and matched at 1:10/1:5/1:2 ratios each) and evaluated on the original test cohort. This experimental framework was implemented across three distinct temporal configurations: 1-hour data window with 0-hour prediction window, 6-hour data window with 0-hour prediction window, and 6-hour data window with 4-hour prediction window. Models trained on undersampled cohorts demonstrate substantial performance metric improvements, while those trained using temporal alignment, demographics matching, or both, exhibit performance metrics below baseline levels.

## Positive Predictive Value (Precision)

| | | | original | undersampled | | | temporal alignment | | | demographics matching | | | demographics + temporal | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 1:2 | 1:5 | 1:10 | 1:2 | 1:5 | 1:10 | 1:2 | 1:5 | 1:10 | 1:2 | 1:5 | 1:10 |
| 1h data / 0h pred | hirid | RF | 0.000 | 0.791 | 0.990 | 0.998 | 0.128 | 0.085 | 0.012 | 0.116 | 0.072 | 0.028 | 0.105 | 0.047 | 0.022 |
| | | bRF | 0.092 | 0.338 | 0.451 | 0.481 | 0.110 | 0.066 | 0.014 | 0.098 | 0.061 | 0.027 | 0.083 | 0.050 | 0.019 |
| | | LGBM | 0.072 | 0.226 | 0.313 | 0.581 | 0.069 | 0.067 | 0.000 | 0.131 | 0.000 | 0.019 | 0.000 | 0.053 | 0.000 |
| | | XGB | 0.148 | 0.217 | 0.474 | 0.674 | 0.075 | 0.081 | 0.000 | 0.089 | 0.020 | 0.023 | 0.026 | 0.041 | 0.019 |
| | | LR | 0.000 | 0.090 | 0.105 | 0.043 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | CNN | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | miiv | RF | 0.000 | 0.904 | 0.998 | 0.999 | 0.103 | 0.039 | 0.022 | 0.123 | 0.053 | 0.023 | 0.116 | 0.098 | 0.026 |
| | | bRF | 0.046 | 0.143 | 0.189 | 0.192 | 0.074 | 0.028 | 0.014 | 0.069 | 0.030 | 0.008 | 0.063 | 0.046 | 0.015 |
| | | LGBM | 0.032 | 0.141 | 0.159 | 0.300 | 0.024 | 0.000 | 0.020 | 0.150 | 0.019 | 0.006 | 0.000 | 0.000 | 0.000 |
| | | XGB | 0.069 | 0.142 | 0.300 | 0.572 | 0.016 | 0.024 | 0.032 | 0.068 | 0.034 | 0.000 | 0.040 | 0.042 | 0.004 |
| | | LR | 0.231 | 0.071 | 0.088 | 0.178 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | CNN | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | eicu | RF | 0.000 | 0.979 | 1.000 | 1.000 | 0.108 | 0.058 | 0.033 | 0.072 | 0.046 | 0.036 | 0.147 | 0.076 | 0.023 |
| | | bRF | 0.041 | 0.249 | 0.285 | 0.291 | 0.083 | 0.053 | 0.021 | 0.064 | 0.033 | 0.024 | 0.099 | 0.042 | 0.018 |
| | | LGBM | 0.018 | 0.073 | 0.098 | 0.220 | 0.000 | 0.000 | 0.002 | 0.000 | 0.000 | 0.018 | 0.033 | 0.000 | 0.049 |
| | | XGB | 0.020 | 0.109 | 0.288 | 0.293 | 0.009 | 0.007 | 0.012 | 0.018 | 0.004 | 0.012 | 0.017 | 0.015 | 0.029 |
| | | LR | 0.000 | 0.020 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | CNN | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 6h data / 0h pred | hirid | RF | 0.000 | 0.663 | 0.958 | 1.000 | 0.139 | 0.083 | 0.015 | 0.117 | 0.081 | 0.034 | 0.121 | 0.051 | 0.016 |
| | | bRF | 0.106 | 0.240 | 0.334 | 0.365 | 0.111 | 0.060 | 0.016 | 0.092 | 0.071 | 0.031 | 0.095 | 0.055 | 0.017 |
| | | LGBM | 0.179 | 0.262 | 0.502 | 0.668 | 0.058 | 0.321 | 0.000 | 0.057 | 0.007 | 0.138 | 0.019 | 0.143 | 0.000 |
| | | XGB | 0.179 | 0.294 | 0.496 | 0.713 | 0.076 | 0.066 | 0.000 | 0.034 | 0.033 | 0.067 | 0.037 | 0.030 | 0.018 |
| | | LR | 0.159 | 0.135 | 0.127 | 0.144 | 0.500 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | CNN | 0.000 | 0.244 | 0.402 | 0.518 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | miiv | RF | 0.000 | 0.814 | 0.996 | 0.999 | 0.105 | 0.041 | 0.025 | 0.117 | 0.044 | 0.019 | 0.096 | 0.109 | 0.030 |
| | | bRF | 0.055 | 0.116 | 0.146 | 0.147 | 0.081 | 0.029 | 0.017 | 0.068 | 0.032 | 0.009 | 0.070 | 0.051 | 0.016 |
| | | LGBM | 0.028 | 0.166 | 0.281 | 0.376 | 0.000 | 0.000 | 0.023 | 0.057 | 0.059 | 0.003 | 0.058 | 0.033 | 0.014 |
| | | XGB | 0.140 | 0.165 | 0.350 | 0.523 | 0.012 | 0.016 | 0.010 | 0.046 | 0.030 | 0.024 | 0.052 | 0.039 | 0.029 |
| | | LR | 0.189 | 0.100 | 0.122 | 0.186 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | CNN | 0.046 | 0.159 | 0.390 | 0.738 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | eicu | RF | 0.000 | 0.941 | 0.998 | 0.999 | 0.107 | 0.064 | 0.033 | 0.075 | 0.043 | 0.035 | 0.148 | 0.075 | 0.021 |
| | | bRF | 0.047 | 0.159 | 0.180 | 0.190 | 0.089 | 0.057 | 0.023 | 0.062 | 0.033 | 0.025 | 0.103 | 0.045 | 0.020 |
| | | LGBM | 0.033 | 0.164 | 0.148 | 0.043 | 0.000 | 0.000 | 0.000 | 0.000 | 0.009 | 0.000 | 0.000 | 0.009 | 0.000 |
| | | XGB | 0.003 | 0.160 | 0.273 | 0.472 | 0.012 | 0.009 | 0.020 | 0.015 | 0.015 | 0.043 | 0.011 | 0.010 | 0.018 |
| | | LR | 0.000 | 0.058 | 0.053 | 0.067 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | CNN | 0.000 | 0.141 | 0.833 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 6h data / 4h pred | hirid | RF | 0.000 | 0.703 | 0.975 | 1.000 | 0.131 | 0.068 | 0.018 | 0.115 | 0.072 | 0.037 | 0.115 | 0.055 | 0.015 |
| | | bRF | 0.104 | 0.249 | 0.312 | 0.342 | 0.103 | 0.051 | 0.017 | 0.083 | 0.067 | 0.030 | 0.082 | 0.057 | 0.022 |
| | | LGBM | 0.095 | 0.278 | 0.380 | 0.572 | 0.081 | 0.302 | 0.000 | 0.018 | 0.143 | 0.052 | 0.035 | 0.000 | 0.000 |
| | | XGB | 0.247 | 0.275 | 0.476 | 0.815 | 0.096 | 0.071 | 0.018 | 0.053 | 0.023 | 0.007 | 0.051 | 0.035 | 0.000 |
| | | LR | 0.191 | 0.123 | 0.125 | 0.167 | 0.500 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | CNN | 0.000 | 0.273 | 0.380 | 0.644 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | miiv | RF | 0.000 | 0.823 | 0.995 | 0.999 | 0.103 | 0.036 | 0.022 | 0.101 | 0.041 | 0.018 | 0.100 | 0.078 | 0.029 |
| | | bRF | 0.055 | 0.141 | 0.177 | 0.176 | 0.086 | 0.027 | 0.018 | 0.064 | 0.032 | 0.007 | 0.070 | 0.047 | 0.017 |
| | | LGBM | 0.013 | 0.143 | 0.251 | 0.278 | 0.000 | 0.211 | 0.030 | 0.017 | 0.000 | 0.016 | 0.060 | 0.000 | 0.000 |
| | | XGB | 0.137 | 0.145 | 0.387 | 0.593 | 0.023 | 0.031 | 0.011 | 0.029 | 0.037 | 0.005 | 0.027 | 0.034 | 0.000 |
| | | LR | 0.273 | 0.088 | 0.103 | 0.143 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | CNN | 0.241 | 0.161 | 0.335 | 0.778 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | eicu | RF | 0.000 | 0.911 | 0.993 | 0.998 | 0.110 | 0.065 | 0.030 | 0.069 | 0.042 | 0.034 | 0.145 | 0.075 | 0.019 |
| | | bRF | 0.041 | 0.141 | 0.154 | 0.165 | 0.085 | 0.058 | 0.022 | 0.056 | 0.036 | 0.024 | 0.102 | 0.044 | 0.023 |
| | | LGBM | 0.017 | 0.129 | 0.130 | 0.176 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.007 | 0.000 | 0.005 | 0.085 |
| | | XGB | 0.033 | 0.158 | 0.282 | 0.431 | 0.012 | 0.033 | 0.022 | 0.014 | 0.011 | 0.012 | 0.028 | 0.016 | 0.017 |
| | | LR | 0.000 | 0.054 | 0.040 | 0.059 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | CNN | 0.000 | 0.134 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

**Table 4. Positive Predictive Value (PPV) performance of machine learning models.** This metric is equivalent to **precision**. Comparative analysis of non-optimized (vanilla) machine learning models trained on specified training datasets, case control ratios, and data/prediction window combinations. Models trained on undersampled cohorts demonstrate substantial performance metric improvements, while those trained using temporal alignment, demographics matching, or both, exhibit performance metrics below baseline levels.

## Recall/Sensitivity

| | | | original | undersampled | | | temporal alignment | | | demographics matching | | | demographics + temporal | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 1:2 | 1:5 | 1:10 | 1:2 | 1:5 | 1:10 | 1:2 | 1:5 | 1:10 | 1:2 | 1:5 | 1:10 |
| 1h data / 0h pred | hirid | RF | 0.000 | 0.838 | 0.838 | 0.900 | 0.108 | 0.067 | 0.010 | 0.096 | 0.056 | 0.027 | 0.076 | 0.033 | 0.015 |
| | | bRF | 0.127 | 0.917 | 0.889 | 0.928 | 0.137 | 0.092 | 0.021 | 0.121 | 0.082 | 0.043 | 0.112 | 0.067 | 0.026 |
| | | LGBM | 0.014 | 0.167 | 0.069 | 0.049 | 0.002 | 0.001 | 0.000 | 0.004 | 0.000 | 0.001 | 0.000 | 0.001 | 0.000 |
| | | XGB | 0.014 | 0.401 | 0.312 | 0.203 | 0.028 | 0.014 | 0.000 | 0.026 | 0.002 | 0.002 | 0.009 | 0.007 | 0.002 |
| | | LR | 0.000 | 0.029 | 0.010 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | CNN | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | miiv | RF | 0.000 | 0.806 | 0.924 | 0.906 | 0.076 | 0.033 | 0.022 | 0.064 | 0.029 | 0.010 | 0.050 | 0.028 | 0.014 |
| | | bRF | 0.213 | 0.922 | 0.954 | 0.944 | 0.120 | 0.053 | 0.030 | 0.108 | 0.052 | 0.015 | 0.087 | 0.063 | 0.029 |
| | | LGBM | 0.011 | 0.067 | 0.029 | 0.048 | 0.001 | 0.000 | 0.001 | 0.004 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 |
| | | XGB | 0.005 | 0.257 | 0.181 | 0.140 | 0.005 | 0.004 | 0.004 | 0.010 | 0.003 | 0.000 | 0.011 | 0.004 | 0.000 |
| | | LR | 0.001 | 0.014 | 0.003 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | CNN | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | eicu | RF | 0.000 | 0.814 | 0.809 | 0.868 | 0.033 | 0.028 | 0.026 | 0.049 | 0.032 | 0.026 | 0.065 | 0.033 | 0.009 |
| | | bRF | 0.092 | 0.909 | 0.893 | 0.904 | 0.081 | 0.076 | 0.037 | 0.089 | 0.053 | 0.044 | 0.123 | 0.064 | 0.029 |
| | | LGBM | 0.006 | 0.009 | 0.007 | 0.028 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.001 |
| | | XGB | 0.001 | 0.129 | 0.079 | 0.050 | 0.002 | 0.001 | 0.001 | 0.009 | 0.000 | 0.001 | 0.004 | 0.002 | 0.002 |
| | | LR | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | CNN | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 6h data / 0h pred | hirid | RF | 0.000 | 0.859 | 0.842 | 0.889 | 0.110 | 0.059 | 0.010 | 0.097 | 0.058 | 0.030 | 0.084 | 0.036 | 0.009 |
| | | bRF | 0.230 | 0.924 | 0.894 | 0.932 | 0.123 | 0.071 | 0.020 | 0.115 | 0.082 | 0.044 | 0.109 | 0.063 | 0.020 |
| | | LGBM | 0.025 | 0.283 | 0.139 | 0.094 | 0.004 | 0.008 | 0.000 | 0.003 | 0.001 | 0.004 | 0.002 | 0.004 | 0.000 |
| | | XGB | 0.018 | 0.557 | 0.395 | 0.305 | 0.031 | 0.013 | 0.000 | 0.015 | 0.005 | 0.011 | 0.016 | 0.007 | 0.002 |
| | | LR | 0.005 | 0.081 | 0.020 | 0.008 | 0.003 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | CNN | 0.000 | 0.106 | 0.115 | 0.021 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | miiv | RF | 0.000 | 0.802 | 0.903 | 0.884 | 0.078 | 0.037 | 0.024 | 0.067 | 0.025 | 0.009 | 0.045 | 0.031 | 0.017 |
| | | bRF | 0.359 | 0.925 | 0.960 | 0.958 | 0.109 | 0.048 | 0.032 | 0.103 | 0.050 | 0.015 | 0.086 | 0.057 | 0.027 |
| | | LGBM | 0.005 | 0.132 | 0.064 | 0.058 | 0.000 | 0.000 | 0.001 | 0.002 | 0.002 | 0.000 | 0.001 | 0.001 | 0.001 |
| | | XGB | 0.011 | 0.348 | 0.253 | 0.205 | 0.004 | 0.003 | 0.002 | 0.010 | 0.003 | 0.002 | 0.019 | 0.005 | 0.003 |
| | | LR | 0.002 | 0.053 | 0.016 | 0.006 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | CNN | 0.001 | 0.175 | 0.032 | 0.009 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | eicu | RF | 0.000 | 0.824 | 0.797 | 0.854 | 0.032 | 0.030 | 0.024 | 0.052 | 0.032 | 0.027 | 0.061 | 0.035 | 0.009 |
| | | bRF | 0.188 | 0.920 | 0.906 | 0.912 | 0.080 | 0.070 | 0.034 | 0.077 | 0.048 | 0.043 | 0.116 | 0.062 | 0.029 |
| | | LGBM | 0.009 | 0.044 | 0.017 | 0.036 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | XGB | 0.000 | 0.169 | 0.103 | 0.081 | 0.003 | 0.002 | 0.003 | 0.006 | 0.003 | 0.003 | 0.003 | 0.002 | 0.002 |
| | | LR | 0.000 | 0.011 | 0.002 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | CNN | 0.000 | 0.039 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 6h data / 4h pred | hirid | RF | 0.000 | 0.871 | 0.841 | 0.886 | 0.104 | 0.052 | 0.013 | 0.103 | 0.052 | 0.033 | 0.080 | 0.039 | 0.008 |
| | | bRF | 0.237 | 0.928 | 0.889 | 0.934 | 0.119 | 0.062 | 0.022 | 0.117 | 0.077 | 0.045 | 0.103 | 0.068 | 0.027 |
| | | LGBM | 0.013 | 0.314 | 0.134 | 0.121 | 0.004 | 0.009 | 0.000 | 0.001 | 0.003 | 0.003 | 0.002 | 0.000 | 0.000 |
| | | XGB | 0.026 | 0.541 | 0.425 | 0.311 | 0.039 | 0.015 | 0.003 | 0.022 | 0.005 | 0.001 | 0.022 | 0.007 | 0.000 |
| | | LR | 0.005 | 0.066 | 0.015 | 0.006 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | CNN | 0.000 | 0.121 | 0.089 | 0.017 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | miiv | RF | 0.000 | 0.791 | 0.897 | 0.895 | 0.082 | 0.035 | 0.024 | 0.059 | 0.025 | 0.009 | 0.045 | 0.022 | 0.017 |
| | | bRF | 0.263 | 0.907 | 0.947 | 0.952 | 0.121 | 0.051 | 0.036 | 0.104 | 0.055 | 0.014 | 0.089 | 0.059 | 0.031 |
| | | LGBM | 0.004 | 0.118 | 0.064 | 0.078 | 0.000 | 0.004 | 0.001 | 0.000 | 0.000 | 0.001 | 0.002 | 0.000 | 0.000 |
| | | XGB | 0.011 | 0.306 | 0.257 | 0.217 | 0.011 | 0.011 | 0.002 | 0.009 | 0.004 | 0.000 | 0.013 | 0.006 | 0.000 |
| | | LR | 0.001 | 0.039 | 0.009 | 0.003 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | CNN | 0.002 | 0.145 | 0.025 | 0.007 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | eicu | RF | 0.000 | 0.822 | 0.797 | 0.861 | 0.035 | 0.031 | 0.022 | 0.048 | 0.033 | 0.028 | 0.062 | 0.035 | 0.008 |
| | | bRF | 0.193 | 0.916 | 0.905 | 0.911 | 0.082 | 0.073 | 0.035 | 0.073 | 0.054 | 0.041 | 0.116 | 0.060 | 0.034 |
| | | LGBM | 0.006 | 0.035 | 0.018 | 0.044 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.002 |
| | | XGB | 0.002 | 0.173 | 0.106 | 0.081 | 0.004 | 0.004 | 0.004 | 0.004 | 0.002 | 0.001 | 0.009 | 0.003 | 0.001 |
| | | LR | 0.000 | 0.008 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | CNN | 0.000 | 0.043 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

**Table 5. Recall (sensitivity) performance of machine learning models.** Comparative analysis of non-optimized (vanilla) machine learning models trained on specified training datasets, case control ratios, and data/prediction window combinations. Models trained on undersampled cohorts demonstrate substantial performance metric improvements, while those trained using temporal alignment, demographics matching, or both, exhibit performance metrics below baseline levels.

## Specificity

| | | | original | undersampled | | | temporal alignment | | | demographics matching | | | demographics + temporal | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 1:2 | 1:5 | 1:10 | 1:2 | 1:5 | 1:10 | 1:2 | 1:5 | 1:10 | 1:2 | 1:5 | 1:10 |
| 1h data / 0h pred | hirid | RF | 1.000 | 0.995 | 1.000 | 1.000 | 0.984 | 0.984 | 0.982 | 0.984 | 0.984 | 0.980 | 0.986 | 0.985 | 0.986 |
| | | bRF | 0.972 | 0.960 | 0.976 | 0.978 | 0.975 | 0.971 | 0.968 | 0.975 | 0.972 | 0.966 | 0.973 | 0.972 | 0.970 |
| | | LGBM | 0.996 | 0.987 | 0.997 | 0.999 | 0.999 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 |
| | | XGB | 0.998 | 0.968 | 0.992 | 0.998 | 0.992 | 0.997 | 0.997 | 0.994 | 0.998 | 0.998 | 0.993 | 0.997 | 0.998 |
| | | LR | 1.000 | 0.994 | 0.998 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | CNN | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | miiv | RF | 1.000 | 0.999 | 1.000 | 1.000 | 0.992 | 0.991 | 0.989 | 0.995 | 0.994 | 0.995 | 0.996 | 0.997 | 0.994 |
| | | bRF | 0.950 | 0.937 | 0.953 | 0.955 | 0.983 | 0.979 | 0.976 | 0.984 | 0.981 | 0.979 | 0.985 | 0.985 | 0.979 |
| | | LGBM | 0.996 | 0.995 | 0.998 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 |
| | | XGB | 0.999 | 0.982 | 0.995 | 0.999 | 0.997 | 0.998 | 0.999 | 0.999 | 0.999 | 0.999 | 0.997 | 0.999 | 0.999 |
| | | LR | 1.000 | 0.998 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | CNN | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | eicu | RF | 1.000 | 1.000 | 1.000 | 1.000 | 0.997 | 0.995 | 0.992 | 0.993 | 0.993 | 0.992 | 0.996 | 0.996 | 0.996 |
| | | bRF | 0.976 | 0.969 | 0.975 | 0.975 | 0.990 | 0.985 | 0.981 | 0.985 | 0.983 | 0.980 | 0.988 | 0.984 | 0.982 |
| | | LGBM | 0.996 | 0.999 | 0.999 | 0.999 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 |
| | | XGB | 0.999 | 0.988 | 0.998 | 0.999 | 0.997 | 0.999 | 0.999 | 0.995 | 0.999 | 0.999 | 0.997 | 0.999 | 0.999 |
| | | LR | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | CNN | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 6h data / 0h pred | hirid | RF | 1.000 | 0.990 | 0.999 | 1.000 | 0.984 | 0.985 | 0.984 | 0.983 | 0.984 | 0.980 | 0.986 | 0.984 | 0.987 |
| | | bRF | 0.954 | 0.931 | 0.958 | 0.962 | 0.977 | 0.974 | 0.971 | 0.973 | 0.975 | 0.968 | 0.975 | 0.974 | 0.973 |
| | | LGBM | 0.997 | 0.981 | 0.997 | 0.999 | 0.999 | 1.000 | 0.999 | 0.999 | 0.998 | 0.999 | 0.998 | 1.000 | 0.999 |
| | | XGB | 0.998 | 0.968 | 0.991 | 0.997 | 0.991 | 0.996 | 0.996 | 0.990 | 0.997 | 0.996 | 0.990 | 0.995 | 0.998 |
| | | LR | 0.999 | 0.988 | 0.997 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | CNN | 1.000 | 0.992 | 0.996 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | miiv | RF | 1.000 | 0.998 | 1.000 | 1.000 | 0.993 | 0.990 | 0.989 | 0.994 | 0.994 | 0.995 | 0.995 | 0.997 | 0.994 |
| | | bRF | 0.930 | 0.921 | 0.937 | 0.937 | 0.986 | 0.982 | 0.979 | 0.984 | 0.983 | 0.982 | 0.987 | 0.988 | 0.981 |
| | | LGBM | 0.998 | 0.993 | 0.998 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 |
| | | XGB | 0.999 | 0.980 | 0.995 | 0.998 | 0.996 | 0.998 | 0.998 | 0.998 | 0.999 | 0.999 | 0.996 | 0.999 | 0.999 |
| | | LR | 1.000 | 0.995 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | CNN | 1.000 | 0.990 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | eicu | RF | 1.000 | 0.999 | 1.000 | 1.000 | 0.997 | 0.995 | 0.992 | 0.993 | 0.992 | 0.991 | 0.996 | 0.995 | 0.995 |
| | | bRF | 0.956 | 0.943 | 0.952 | 0.955 | 0.991 | 0.987 | 0.983 | 0.986 | 0.984 | 0.981 | 0.988 | 0.985 | 0.983 |
| | | LGBM | 0.997 | 0.997 | 0.999 | 0.991 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 |
| | | XGB | 0.999 | 0.990 | 0.997 | 0.999 | 0.997 | 0.998 | 0.999 | 0.996 | 0.998 | 0.999 | 0.997 | 0.998 | 0.999 |
| | | LR | 1.000 | 0.998 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | CNN | 1.000 | 0.997 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 6h data / 4h pred | hirid | RF | 1.000 | 0.992 | 1.000 | 1.000 | 0.984 | 0.984 | 0.984 | 0.982 | 0.985 | 0.980 | 0.986 | 0.985 | 0.988 |
| | | bRF | 0.953 | 0.936 | 0.955 | 0.959 | 0.976 | 0.973 | 0.972 | 0.971 | 0.975 | 0.967 | 0.973 | 0.974 | 0.972 |
| | | LGBM | 0.997 | 0.981 | 0.995 | 0.998 | 0.999 | 1.000 | 0.999 | 0.999 | 1.000 | 0.999 | 0.999 | 0.999 | 0.999 |
| | | XGB | 0.998 | 0.967 | 0.989 | 0.998 | 0.992 | 0.996 | 0.996 | 0.991 | 0.996 | 0.997 | 0.991 | 0.996 | 0.996 |
| | | LR | 1.000 | 0.989 | 0.998 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | CNN | 1.000 | 0.993 | 0.997 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | miiv | RF | 1.000 | 0.998 | 1.000 | 1.000 | 0.993 | 0.990 | 0.989 | 0.995 | 0.994 | 0.995 | 0.996 | 0.997 | 0.994 |
| | | bRF | 0.954 | 0.944 | 0.955 | 0.955 | 0.987 | 0.982 | 0.980 | 0.985 | 0.983 | 0.981 | 0.988 | 0.988 | 0.982 |
| | | LGBM | 0.997 | 0.993 | 0.998 | 0.998 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 |
| | | XGB | 0.999 | 0.982 | 0.996 | 0.999 | 0.995 | 0.997 | 0.998 | 0.997 | 0.999 | 0.999 | 0.995 | 0.998 | 0.998 |
| | | LR | 1.000 | 0.996 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | CNN | 1.000 | 0.992 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | eicu | RF | 1.000 | 0.999 | 1.000 | 1.000 | 0.997 | 0.995 | 0.992 | 0.993 | 0.992 | 0.991 | 0.996 | 0.995 | 0.995 |
| | | bRF | 0.950 | 0.938 | 0.945 | 0.949 | 0.990 | 0.987 | 0.983 | 0.986 | 0.984 | 0.981 | 0.989 | 0.985 | 0.984 |
| | | LGBM | 0.996 | 0.997 | 0.999 | 0.998 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 |
| | | XGB | 1.000 | 0.990 | 0.997 | 0.999 | 0.997 | 0.999 | 0.998 | 0.997 | 0.998 | 0.999 | 0.996 | 0.998 | 0.999 |
| | | LR | 1.000 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | CNN | 1.000 | 0.997 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

**Table 6. Specificity performance of machine learning models.** Comparative analysis of non-optimized (vanilla) machine learning models trained on specified training datasets, case control ratios, and data/prediction window combinations. While the undersampled models seem to underperform on this metric, taken in conjunction with the other reported metrics, these models still continue to outperform. This metric may also rely on particular threshold selection depending on the model, where results shown are using the default classification threshold of 0.5.

## Negative Predictive Value (NPV)

| | | | original | undersampled | | | temporal alignment | | | demographics matching | | | demographics + temporal | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 1:2 | 1:5 | 1:10 | 1:2 | 1:5 | 1:10 | 1:2 | 1:5 | 1:10 | 1:2 | 1:5 | 1:10 |
| 1h data / 0h pred | hirid | RF | 0.978 | 0.996 | 0.996 | 0.998 | 0.980 | 0.980 | 0.978 | 0.980 | 0.979 | 0.979 | 0.980 | 0.979 | 0.978 |
| | | bRF | 0.981 | 0.998 | 0.998 | 0.998 | 0.981 | 0.980 | 0.978 | 0.981 | 0.980 | 0.979 | 0.980 | 0.979 | 0.978 |
| | | LGBM | 0.979 | 0.982 | 0.980 | 0.979 | 0.979 | 0.978 | 0.978 | 0.979 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 |
| | | XGB | 0.979 | 0.987 | 0.985 | 0.983 | 0.979 | 0.979 | 0.978 | 0.979 | 0.978 | 0.978 | 0.979 | 0.979 | 0.978 |
| | | LR | 0.978 | 0.979 | 0.979 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 |
| | | CNN | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 |
| | miiv | RF | 0.989 | 0.998 | 0.999 | 0.999 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| | | bRF | 0.991 | 0.999 | 0.999 | 0.999 | 0.990 | 0.989 | 0.989 | 0.990 | 0.989 | 0.989 | 0.990 | 0.989 | 0.989 |
| | | LGBM | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| | | XGB | 0.989 | 0.991 | 0.991 | 0.990 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| | | LR | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| | | CNN | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| | eicu | RF | 0.989 | 0.998 | 0.998 | 0.999 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.990 | 0.989 | 0.989 |
| | | bRF | 0.990 | 0.999 | 0.999 | 0.999 | 0.990 | 0.990 | 0.989 | 0.990 | 0.989 | 0.989 | 0.990 | 0.989 | 0.989 |
| | | LGBM | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| | | XGB | 0.989 | 0.990 | 0.990 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| | | LR | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| | | CNN | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| 6h data / 0h pred | hirid | RF | 0.977 | 0.997 | 0.996 | 0.997 | 0.979 | 0.978 | 0.977 | 0.979 | 0.978 | 0.977 | 0.979 | 0.977 | 0.977 |
| | | bRF | 0.981 | 0.998 | 0.997 | 0.998 | 0.979 | 0.978 | 0.977 | 0.979 | 0.978 | 0.977 | 0.979 | 0.978 | 0.977 |
| | | LGBM | 0.977 | 0.983 | 0.980 | 0.979 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 |
| | | XGB | 0.977 | 0.989 | 0.986 | 0.984 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 |
| | | LR | 0.977 | 0.979 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 |
| | | CNN | 0.977 | 0.979 | 0.979 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 | 0.977 |
| | miiv | RF | 0.989 | 0.998 | 0.999 | 0.999 | 0.990 | 0.989 | 0.989 | 0.990 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| | | bRF | 0.992 | 0.999 | 1.000 | 0.999 | 0.990 | 0.989 | 0.989 | 0.990 | 0.989 | 0.989 | 0.990 | 0.989 | 0.989 |
| | | LGBM | 0.989 | 0.990 | 0.990 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| | | XGB | 0.989 | 0.993 | 0.992 | 0.991 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| | | LR | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| | | CNN | 0.989 | 0.991 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| | eicu | RF | 0.989 | 0.998 | 0.998 | 0.998 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| | | bRF | 0.990 | 0.999 | 0.999 | 0.999 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.990 | 0.989 | 0.989 |
| | | LGBM | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.988 | 0.989 | 0.989 | 0.989 |
| | | XGB | 0.989 | 0.990 | 0.990 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| | | LR | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| | | CNN | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| 6h data / 4h pred | hirid | RF | 0.978 | 0.997 | 0.996 | 0.997 | 0.980 | 0.978 | 0.978 | 0.979 | 0.978 | 0.978 | 0.979 | 0.978 | 0.977 |
| | | bRF | 0.982 | 0.998 | 0.997 | 0.998 | 0.980 | 0.978 | 0.977 | 0.980 | 0.979 | 0.978 | 0.979 | 0.979 | 0.978 |
| | | LGBM | 0.978 | 0.984 | 0.980 | 0.980 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 |
| | | XGB | 0.978 | 0.989 | 0.987 | 0.984 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 |
| | | LR | 0.978 | 0.979 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 |
| | | CNN | 0.978 | 0.980 | 0.979 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 | 0.978 |
| | miiv | RF | 0.990 | 0.998 | 0.999 | 0.999 | 0.991 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 |
| | | bRF | 0.992 | 0.999 | 0.999 | 0.999 | 0.991 | 0.990 | 0.990 | 0.991 | 0.990 | 0.990 | 0.991 | 0.990 | 0.990 |
| | | LGBM | 0.990 | 0.991 | 0.991 | 0.991 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 |
| | | XGB | 0.990 | 0.993 | 0.992 | 0.992 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 |
| | | LR | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 |
| | | CNN | 0.990 | 0.991 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 |
| | eicu | RF | 0.989 | 0.998 | 0.998 | 0.998 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.990 | 0.989 | 0.989 |
| | | bRF | 0.991 | 0.999 | 0.999 | 0.999 | 0.990 | 0.990 | 0.989 | 0.990 | 0.989 | 0.989 | 0.990 | 0.989 | 0.989 |
| | | LGBM | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| | | XGB | 0.989 | 0.991 | 0.990 | 0.990 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| | | LR | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |
| | | CNN | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 | 0.989 |

**Table 7. Negative Predictive Value (NPV) performance of machine learning models.** Comparative analysis of non-optimized (vanilla) machine learning models trained on specified training datasets, case control ratios, and data/prediction window combinations. Models trained on undersampled cohorts demonstrate substantial performance metric improvements, while those trained using temporal alignment, demographics matching, or both, exhibit performance metrics below baseline levels.

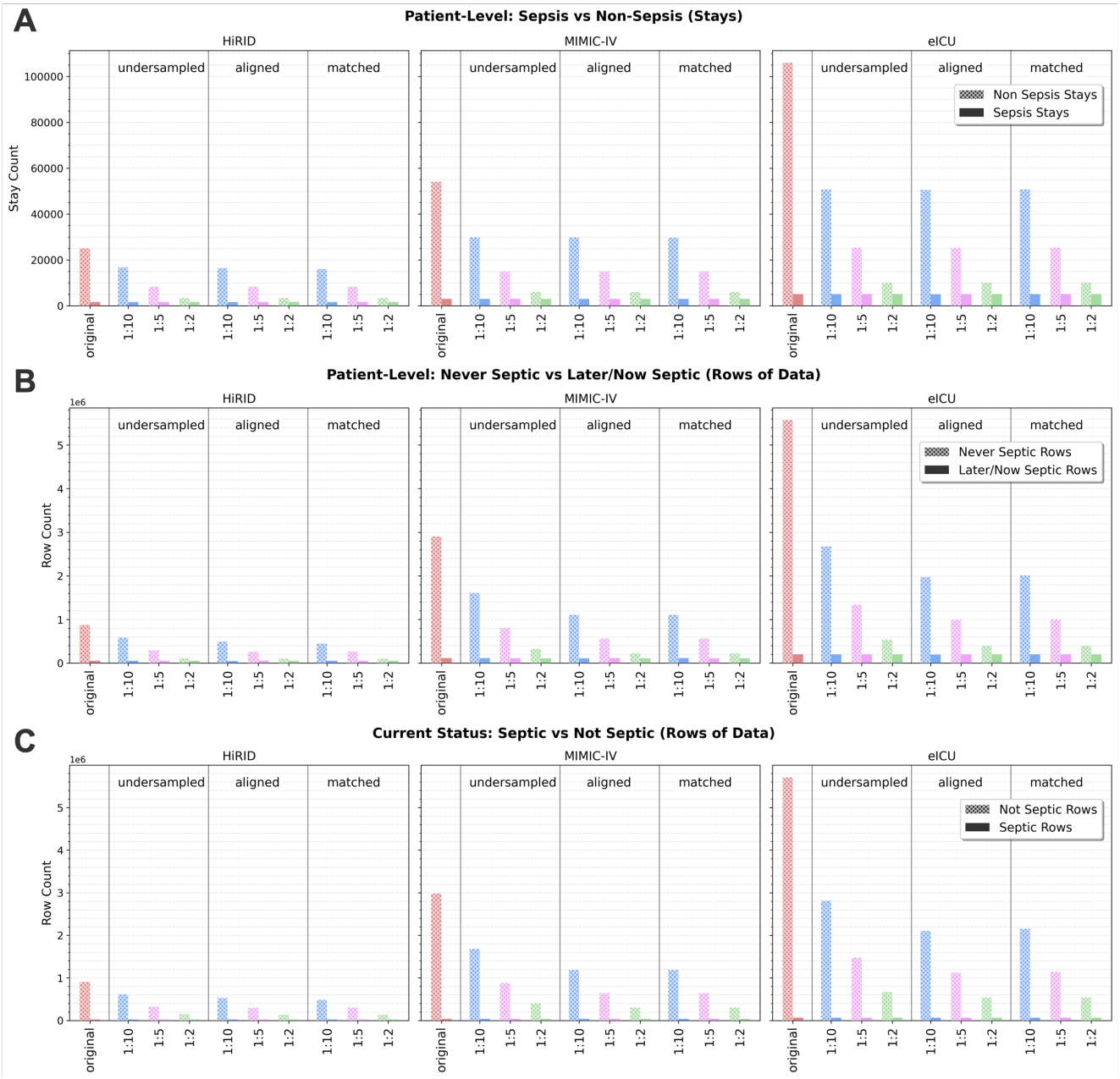


**Figure 7. Overview of training set sizes and sepsis case and label prevalences.** Implementation of three case-control matching ratios (1:10, 1:5, and 1:2) yields progressively reduced training cohorts with inversely increasing sepsis prevalence. Numerical data for these distributions are presented in Table 8. **A)** Number of stays in the sepsis and non-sepsis groups across different datasets (HiRID, MIMIC-IV, eICU), matching methods (undersampled, aligned, matched) and matching ratios (1:10, 1:5, 1:2). **B)** Combined number of rows (i.e. hours of data) of sepsis (later or now septic) and non-sepsis (never septic) stays. **B)** Number of rows (i.e. hours of data) with positive sepsis label (currently septic) and negative sepsis label (currently not septic).

| Dataset | Matching Method | Variant | Set | Total Stays | Non-Sepsis Cases | Sepsis Cases (Prevalence) |
|---|---|---|---|---|---|---|
| HiRID | original | | train | 26,728 | 25,040 (93.7%) | 1,688 (6.3%) |
| | original | | test | 2,970 | 2,800 (94.3%) | 170 (5.7%) |
| | undersampled | 10 | train | 18,394 | 16,726 (90.9%) | 1,668 (9.1%) |
| | undersampled | 5 | train | 10,023 | 8,354 (83.3%) | 1,679 (16.7%) |
| | undersampled | 2 | train | 5,016 | 3,343 (66.6%) | 1,673 (33.4%) |
| | aligned | 10 | train | 18,394 | 16,726 (90.9%) | 1,668 (9.1%) |
| | aligned | 5 | train | 10,011 | 8,350 (83.2%) | 1,668 (16.8%) |
| | aligned | 2 | train | 5,016 | 3,343 (66.6%) | 1,673 (33.4%) |
| | matched | 10 | train | 17,681 | 16,013 (90.6%) | 1,668 (9.4%) |
| | matched | 5 | train | 10,028 | 8,354 (83.3%) | 1,674 (16.7%) |
| | matched | 2 | train | 5,069 | 3,395 (67.0%) | 1,674 (33.0%) |
| MIMIC-IV | original | | train | 57,082 | 54,058 (94.7%) | 3,024 (5.3%) |
| | original | | test | 6,343 | 6,046 (95.3%) | 297 (4.7%) |
| | undersampled | 10 | train | 32,877 | 29,874 (90.9%) | 3,003 (9.1%) |
| | undersampled | 5 | train | 17,933 | 14,933 (83.3%) | 3,000 (16.7%) |
| | undersampled | 2 | train | 8,966 | 5,969 (66.6%) | 2,997 (33.4%) |
| | aligned | 10 | train | 32,751 | 29,765 (90.9%) | 2,986 (9.1%) |
| | aligned | 5 | train | 17,933 | 14,934 (83.3%) | 2,999 (16.7%) |
| | aligned | 2 | train | 8,965 | 5,969 (66.6%) | 2,996 (33.4%) |
| | matched | 10 | train | 32,734 | 29,737 (90.8%) | 2,998 (9.2%) |
| | matched | 5 | train | 17,933 | 14,946 (83.3%) | 2,987 (16.7%) |
| | matched | 2 | train | 8,966 | 5,969 (66.6%) | 2,997 (33.4%) |
| eICU | original | | train | 111,071 | 105,988 (95.4%) | 5,083 (4.6%) |
| | original | | test | 12,342 | 11,787 (95.5%) | 555 (4.5%) |
| | undersampled | 10 | train | 55,816 | 50,739 (90.9%) | 5,077 (9.1%) |
| | undersampled | 5 | train | 30,445 | 25,363 (83.3%) | 5,082 (16.7%) |
| | undersampled | 2 | train | 15,222 | 10,143 (66.6%) | 5,079 (33.4%) |
| | aligned | 10 | train | 55,539 | 50,522 (91.0%) | 5,017 (9.0%) |
| | aligned | 5 | train | 30,522 | 25,341 (83.0%) | 5,181 (17.0%) |
| | aligned | 2 | train | 15,222 | 10,141 (66.6%) | 5,081 (33.4%) |
| | matched | 10 | train | 55,816 | 50,734 (90.9%) | 5,082 (9.1%) |
| | matched | 5 | train | 30,445 | 25,363 (83.3%) | 5,082 (16.7%) |
| | matched | 2 | train | 15,222 | 10,155 (66.7%) | 5,067 (33.3%) |

| Dataset | Matching Method | Variant | Set | Total Rows | Group | | Label | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Never-Septic | Later-/Now-Septic | Not Septic | Septic |
| HiRID | original | | train | 930,608 | 877,612 (94.3%) | 52,996 (5.7%) | 908,712 (97.6%) | 21,896 (2.4%) |
| | original | | test | 102,473 | 96,175 (93.9%) | 6,298 (6.1%) | 100,267 (97.8%) | 2,206 (2.2%) |
| | undersampled | 10 | train | 504,536 | 451,320 (89.5%) | 53,220 (10.5%) | 482,883 (95.7%) | 21,754 (4.3%) |
| | undersampled | 5 | train | 319,754 | 266,046 (83.2%) | 53,708 (16.8%) | 297,983 (93.2%) | 21,771 (6.8%) |
| | undersampled | 2 | train | 159,190 | 105,281 (66.1%) | 53,909 (33.9%) | 137,402 (86.3%) | 21,788 (13.7%) |
| | aligned | 10 | train | 519,289 | 466,172 (89.8%) | 53,126 (10.2%) | 497,492 (95.8%) | 21,798 (4.2%) |
| | aligned | 5 | train | 328,286 | 266,859 (81.3%) | 61,432 (18.7%) | 306,455 (93.3%) | 21,837 (6.7%) |
| | aligned | 2 | train | 159,076 | 105,229 (66.1%) | 53,848 (33.9%) | 137,284 (86.3%) | 21,802 (13.7%) |
| | matched | 10 | train | 504,519 | 451,264 (89.4%) | 53,255 (10.6%) | 482,883 (95.7%) | 21,636 (4.3%) |
| | matched | 5 | train | 319,754 | 266,046 (83.2%) | 53,708 (16.8%) | 298,047 (93.2%) | 21,714 (6.8%) |
| | matched | 2 | train | 159,404 | 105,563 (66.2%) | 53,891 (33.8%) | 137,598 (86.3%) | 21,836 (13.7%) |
| MIMIC-IV | original | | train | 3,017,646 | 2,903,317 (96.2%) | 114,329 (3.8%) | 2,978,396 (98.7%) | 39,250 (1.3%) |
| | original | | test | 341,958 | 330,615 (96.7%) | 11,343 (3.3%) | 338,099 (98.9%) | 3,859 (1.1%) |
| | undersampled | 10 | train | 1,722,288 | 1,608,255 (93.4%) | 114,037 (6.6%) | 1,683,395 (97.7%) | 38,896 (2.3%) |
| | undersampled | 5 | train | 978,358 | 865,528 (88.5%) | 112,832 (11.5%) | 939,470 (96.0%) | 38,893 (4.0%) |
| | undersampled | 2 | train | 489,048 | 376,253 (76.9%) | 112,795 (23.1%) | 450,127 (92.0%) | 38,942 (8.0%) |
| | aligned | 10 | train | 1,272,709 | 1,152,570 (90.6%) | 119,530 (9.4%) | 1,234,222 (96.9%) | 38,528 (3.1%) |
| | aligned | 5 | train | 697,878 | 584,053 (83.6%) | 114,440 (16.4%) | 659,128 (94.3%) | 39,570 (5.7%) |
| | aligned | 2 | train | 337,942 | 225,240 (66.6%) | 113,034 (33.4%) | 299,211 (88.5%) | 38,795 (11.5%) |
| | matched | 10 | train | 1,223,350 | 1,109,025 (90.7%) | 114,324 (9.3%) | 1,184,435 (96.8%) | 38,915 (3.2%) |
| | matched | 5 | train | 678,271 | 565,436 (83.4%) | 112,889 (16.6%) | 639,377 (94.3%) | 38,910 (5.7%) |
| | matched | 2 | train | 336,483 | 224,231 (66.6%) | 112,252 (33.4%) | 297,583 (88.4%) | 38,900 (11.6%) |
| eICU | original | | train | 5,772,659 | 5,571,761 (96.5%) | 200,898 (3.5%) | 5,707,176 (98.9%) | 65,483 (1.1%) |
| | original | | test | 644,384 | 621,458 (96.4%) | 22,926 (3.6%) | 637,231 (98.9%) | 7,153 (1.1%) |
| | undersampled | 10 | train | 2,575,364 | 2,370,363 (92.0%) | 205,001 (8.0%) | 2,512,803 (97.6%) | 62,561 (2.4%) |
| | undersampled | 5 | train | 1,530,332 | 1,330,332 (86.9%) | 200,001 (13.1%) | 1,467,619 (95.9%) | 62,713 (4.1%) |
| | undersampled | 2 | train | 737,419 | 535,312 (72.6%) | 202,204 (27.4%) | 671,967 (91.1%) | 65,432 (8.9%) |
| | aligned | 10 | train | 1,609,570 | 1,375,572 (85.5%) | 198,653 (14.5%) | 1,514,923 (94.8%) | 84,647 (5.2%) |
| | aligned | 5 | train | 920,273 | 721,472 (78.4%) | 200,503 (21.6%) | 838,823 (91.0%) | 83,064 (9.0%) |
| | aligned | 2 | train | 602,885 | 400,605 (66.5%) | 202,349 (33.5%) | 537,348 (90.0%) | 65,484 (10.0%) |
| | matched | 10 | train | 2,214,452 | 2,013,255 (90.9%) | 201,197 (9.1%) | 2,148,990 (97.0%) | 65,462 (3.0%) |
| | matched | 5 | train | 1,207,335 | 1,004,836 (83.2%) | 202,499 (16.8%) | 1,141,852 (94.6%) | 65,483 (5.4%) |
| | matched | 2 | train | 603,836 | 402,276 (66.6%) | 201,560 (33.4%) | 538,555 (89.2%) | 65,281 (10.8%) |

**Table 8. Summary of training and testing set characteristics across different matching methods and ratios.** Machine learning analyses were conducted using the original dataset versions for both model training and evaluation, whereas the undersampled, aligned and matched dataset variants were exclusively utilized for model training. All models, regardless of training dataset, were subsequently evaluated on the original test sets to maintain external validity and enable assessment under clinically representative conditions. The total number of sepsis cases can vary slightly because of randomized splitting into training and testing sets for all variants. Visual representations of this data are depicted in Figure 7.

| Model | Hyperparameter | Range | Best Parameters | | |
|---|---|---|---|---|---|
| | | | HiRID | MIMIC-IV | eICU |
| RF | n_estimators | 50-500 | 486 | 477 | 467 |
| | max_depth | 3-30 | 19 | 16 | 13 |
| | min_samples_split | 2-10 | 3 | 7 | 2 |
| | min_samples_leaf | 1-4 | 2 | 3 | 2 |
| bRF | n_estimators | 50-500 | 409 | 415 | 400 |
| | max_depth | 3-30 | 28 | 29 | 29 |
| | min_samples_split | 2-10 | 9 | 9 | 9 |
| LGBM | n_estimators | 50-500 | 81 | 76 | 489 |
| | learning_rate | 0.01-0.3 | 0.054 | 0.055 | 0.012 |
| | max_depth | 3-30 | 6 | 7 | 15 |
| | num_leaves | 20-100 | 92 | 89 | 62 |
| XGB | n_estimators | 50-500 | 188 | 179 | 478 |
| | learning_rate | 0.01-0.3 | 0.053 | 0.092 | 0.028 |
| | max_depth | 3-7 | 6 | 4 | 5 |
| | min_child_weight | 1-5 | 4 | 3 | 2 |

**Table 9. Bayesian optimization hyperparameter spaces and best parameters.** Hyperparameter ranges were defined for each machine learning model, with optimal parameters determined independently for each model-dataset pairing.

## Code Availability

Scripts used to generate the results presented in this study are publicly available in a GitLab repository.
The following Python scripts are included:

1. **Data Preparation**

   - `preprocessing.py`
     This script utilizes harmonized datasets generated via the YAIB workflow (https://github.com/rvandewater/YAIB). It provides a flexible pipeline for specifying datasets (HiRID, MIIV, eICU), matching methods (original, undersampled, aligned, matched), matching ratios (1:2, 1:5, 1:10), and tolerances for strict demographics-based matching.
   - `windowing.py`
     This script processes the preprocessed training and testing sets (output from `preprocessing.py`) to generate windowed datasets. Users can select preprocessed datasets by specifying the dataset (HiRID, MIIV, eICU), matching method (original, undersampled, aligned, matched), and matching ratio (1:2, 1:5, 1:10). Data and prediction window sizes are configurable (default combinations: 1/0, 6/0, 6/4).

2. **Linear Mixed Effects Models**

   - `lmem.py`
     This script analyzes unwindowed preprocessed training sets (output from `preprocessing.py`) and trains univariate linear mixed effects models for each dynamic feature. Users can specify datasets (HiRID, MIIV, eICU) and matching methods (undersampled, aligned, matched), while original training sets are always included as a comparison. Output is provided as a text file report.
   - `lmem_visualizations.py`
     This script converts the text file report (output from `lmem.py`) into a pandas DataFrame and generates visualizations.

3. **Machine Learning Models**

   - `ml_vanilla.py`
     This script operates on preprocessed unwindowed (1-hour data window, 0-hour prediction window) or windowed training and testing sets (output from `preprocessing.py` and `windowing.py`). It offers a flexible framework for specifying datasets (HiRID, MIIV, eICU), machine learning algorithms (RF, bRF, LGBM, XGB), window configurations (default data-prediction-window combinations: 1/0, 6/0, 6/4), matching methods (original, undersampled, aligned, matched), and matching ratios (1:2, 1:5, 1:10). The script trains models and evaluates their performance.
   - `ml_bayesianopt.py`
     This script utilizes preprocessed unwindowed (1-hour data window, 0-hour prediction window) training and testing sets (output from `preprocessing.py`). Users can specify datasets (HiRID, MIIV, eICU) and machine learning algorithms (RF, bRF, LGBM, XGB). The script applies Bayesian optimization to machine learning models, identifies optimal hyperparameters, and evaluates the resulting model performance.