JAMA Network Open

Original Investigation | Health Informatics

Diagnostic Codes in AI Prediction Models and Label Leakage of Same-Admission Clinical Outcomes

Bashar Ramadan, MBBS; Ming-Chieh Liu, MS; Michael C. Burkhart, PhD; William F. Parker, MD, PhD; Brett K. Beaulieu-Jones, PhD

Abstract

IMPORTANCE Artificial intelligence models that predict same-admission outcomes for hospitalized patients, such as inpatient mortality, often rely on *International Classification of Diseases* (*ICD*) diagnostic codes, even when these codes are not finalized until after discharge.

OBJECTIVE To investigate the extent to which the inclusion of *ICD* codes as features in predictive models are associated with inflated performance metrics via label leakage (eg, including the code for cardiac arrest into an inpatient mortality prediction model) and assess the prevalence and implications of this practice in existing literature.

DESIGN, SETTING, AND PARTICIPANTS This prognostic study examined publicly available, deidentified inpatient electronic health record data from the Medical Information Mart for Intensive Care IV (MIMIC-IV) database. Patients admitted to an intensive care unit or emergency department at Beth Israel Deaconess Medical Center between January 1, 2008, and December 31, 2019, were included. These data were analyzed between December 18, 2024, and January 14, 2025. A targeted literature review of same-admission prediction models using MIMIC with *ICD* codes as features was performed between November 20 and 27, 2024.

MAIN OUTCOME AND MEASURES Using a standard training-validation-test split procedure, prediction models were developed for inpatient mortality (logistic regression, random forest, and XGBoost) using only *ICD* codes as features. Performance in the test set was analyzed using areas under the receiver operating curve and variable importance. Frequencies of studies using same-admission prediction models using MIMIC with *ICD* codes were calculated from the targeted literature review.

RESULTS The study cohort consisted of 180 640 patients (mean [SD] age at admission, 58.7 [19.2] years; 53.0% female), of whom 8573 (4.7%) died during the admission. The models using *ICD* codes predicted in-hospital mortality with high performance in the test dataset, with areas under the receiver operating curve of 0.976 (95% CI, 0.973-0.980) (logistic regression), 0.971 (95% CI, 0.967-0.974) (random forest), and 0.973 (95% CI, 0.968-0.977) (XGBoost). The most important *ICD* codes were subdural hemorrhage (OR, 389.99; 95% CI, 28.79-5283.59), cardiac arrest (OR, 219.58; 95% CI, 159.61-302.08), brain death (OR, 112.78; 95% CI, 13.42-947.70), and encounter for palliative care (OR, 98.04; 95% CI, 83.16-115.58). The literature review found that 37 of 92 studies (40.2%) using MIMIC to predict same-admission outcomes included *ICD* codes as features, even though both MIMIC publications and documentation clearly state that *ICD* codes are derived after discharge.

CONCLUSIONS AND RELEVANCE This prognostic study of the MIMIC-IV database suggests that using *ICD* codes as features in same-admission prediction models may be a severe methodological flaw associated with inflated performance metrics, rendering models incapable of clinically useful

(continued)

Key Points

Question Are *International Classification of Diseases* (*ICD*) diagnostic codes, which are only finalized after hospital discharge, associated with inflated performance of artificial intelligence (AI) health care prediction models?

Findings In this prognostic study of 180 640 patients, 40.2% of published AI models trained to predict same-admission outcomes used *ICD* codes as features. Prediction models for inpatient mortality trained on *ICD* codes predicted in-hospital mortality with high accuracy, with the most important codes (eg, brain death, encounter for palliative care) not available in time for clinically useful mortality prediction.

Meaning These findings suggest that to ensure that AI prediction models are both reliable and clinically deployable, greater diligence is needed in identifying and preventing label leakage.

+ Invited Commentary

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

Abstract (continued)

predictions. The literature review found that the practice is common. Addressing this challenge is essential for advancing trustworthy artificial intelligence in health care.

JAMA Network Open. 2025;8(12):e2550454. doi:10.1001/jamanetworkopen.2025.50454

Introduction

Artificial intelligence (AI) and machine learning models have shown impressive performance in predicting critical same-admission outcomes, such as in-hospital mortality.[1-3] Some models use *International Classification of Diseases* (*ICD*) diagnostic billing codes as input features. Since *ICD* codes are entered in the electronic health record (EHR) after a clinical event, can be revised over the course of an admission, and are finalized only after discharge, their inclusion introduces data leakage, in which information unavailable in deployed clinical settings is improperly used during model training and evaluation.

There are many published examples of machine learning models in health care achieving unrealistic performance by relying on unintended features, a phenomenon termed shortcut learning.[4-8] In this work, we specifically examined the issue of temporal label leakage, as described by Davis et al,[9] in which model inputs are used before they are actually available. For example, imagine a patient admitted with unspecified abdominal pain. After further evaluation, the patient is diagnosed with appendicitis, develops septic shock, and experiences cardiac arrest several days later before dying. Early in the patient's admission, only unspecified abdominal pain would be available. However, if a model incorporates all *ICD* codes subsequently assigned after the end of a hospital stay, it unfairly leverages hindsight information to predict mortality, achieving deceptively high accuracy.

This work aimed to illustrate how seemingly accurate same-admission prediction models may be driven by leakage and to quantify how frequently such leakage appears in the literature on machine learning for health care. To examine outcomes associated with this problem, we performed 2 analyses. First, we use *ICD* codes in models predicting inpatient mortality, one of the most common same-admission prediction tasks. Second, we performed a targeted literature review of studies that have built AI models to predict inpatient outcomes and identified the percentage of those that included *ICD* codes from the same admission as input features.

Methods

Data Source and Study Population

This prognostic study used the Medical Information Mart for Intensive Care IV database (MIMIC-IV), version 2.2,[10] a publicly available, deidentified, EHR database of patients admitted to an intensive care unit (ICU) or emergency department at Beth Israel Deaconess Medical Center between January 1, 2008, and December 31, 2019. The MIMIC-IV database is a large, freely accessible EHR resource released in deidentified form, with dates shifted and other deidentification safeguards applied per Health Insurance Portability and Accountability Act deidentification standards. Because the research used only deidentified data and involved no interaction with individuals, no access to identifiable private information, and no intervention, it did not constitute human participants research under the Common Rule and, therefore, did not require institutional review board review or informed consent. Access to the MIMIC data followed standard credentialing requirements and data use agreement. This study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis and AI (TRIPOD+AI) and Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines.

All admissions with *ICD* codes were included in our study, with less than 1% excluded. The MIMIC-IV dataset categorizes race and ethnicity data of admitted patients as Asian, Black, Hispanic, White, other, or unknown, which are reported herein for descriptive purposes.[10] We partitioned the dataset by the date of admission into train (70%), validation (10%), and test (20%) sets per TRIPOD+AI guidelines,[11] excluding patients from the validation and test sets who also had admissions in the training set. Because our cohort spans the US transition from The *International Classification of Diseases, Ninth Revision, Clinical Modification* (*ICD-9-CM*) to the *International Statistical Classification of Diseases, Tenth Revision, Clinical Modification* (*ICD-10-CM*) (October 1, 2015), we mapped all *ICD-10-CM* diagnoses to *ICD-9-CM* using the Centers for Medicare & Medicaid Services General Equivalence Mappings to harmonize the code space across years; the additional granularity of *ICD-10-CM* was not required for our aims focused on leakage. We also removed *ICD* codes that had low variance (<0.0001) or high covariance (>0.8) with other *ICD* codes.

ICD Code Prediction Model Development and Evaluation

We trained classification models (logistic regression,[12] random forest,[12] and XGBoost[13]) using only *ICD-9* codes as features, tuning hyperparameters in the validation set. We chose these models because they are some of the commonly used classifiers, achieve strong performance with tabular data, and offer approaches to interpret models. Other predictive features, such as vital signs, laboratory values, and medications, were intentionally excluded to examine only the potential for *ICD* code–driven label leakage. The trained models were then evaluated on the held-out test set, with performance assessed using the area under the receiver operating characteristic curve (AUROC) and balanced accuracy.

Targeted Literature Review

To assess the pervasiveness of this issue, we performed a targeted literature review of studies that used either MIMIC-III or MIMIC-IV. To do so, we used Google Scholar between November 20 and 27, 2024, with 2 search queries (case insensitive): (1) *prediction model machine learning mimic-IV* OR *mimic IV* OR *mimic 4* OR *mimic-4* and (2) *prediction model machine learning mimic-III* OR *mimic III* OR *mimic 3* OR *mimic-3*. We sorted results by citations per year to avoid bias against recently published studies and screened them sequentially until we identified 100 prediction modeling studies (50 each for MIMIC-III[14] and MIMIC-IV[10]). We then performed a manual review of the articles to (1) categorize whether the studies predicted clinical events during the same admission and (2) investigate whether *ICD* codes were used as input features to predict an outcome during that same admission.

Statistical Analysis

We calculated odds ratios (ORs) and *P* values for *ICD* codes in the logistic regression model and applied the Benjamini-Hochberg procedure to control for false discovery rate, with a threshold of $P < .05$. For the random forest and XGBoost models, we assessed feature importance with each library's respective default criterion, namely Gini importance and gain, to identify which *ICD* codes were considered important for the prediction task. The analysis was performed between December 18, 2024, and January 14, 2025, using Python, version 3.10 (Python Software Foundation) with the packages numpy, version 2.0.2; pandas, version 2.2.2; scikit-learn, version 1.4.2; scipy, version 1.13.0; shap, version 0.46.0; statsmodels, version 0.14.2, and xgboost, version 2.0.3. The full source code is available on Github.[15]

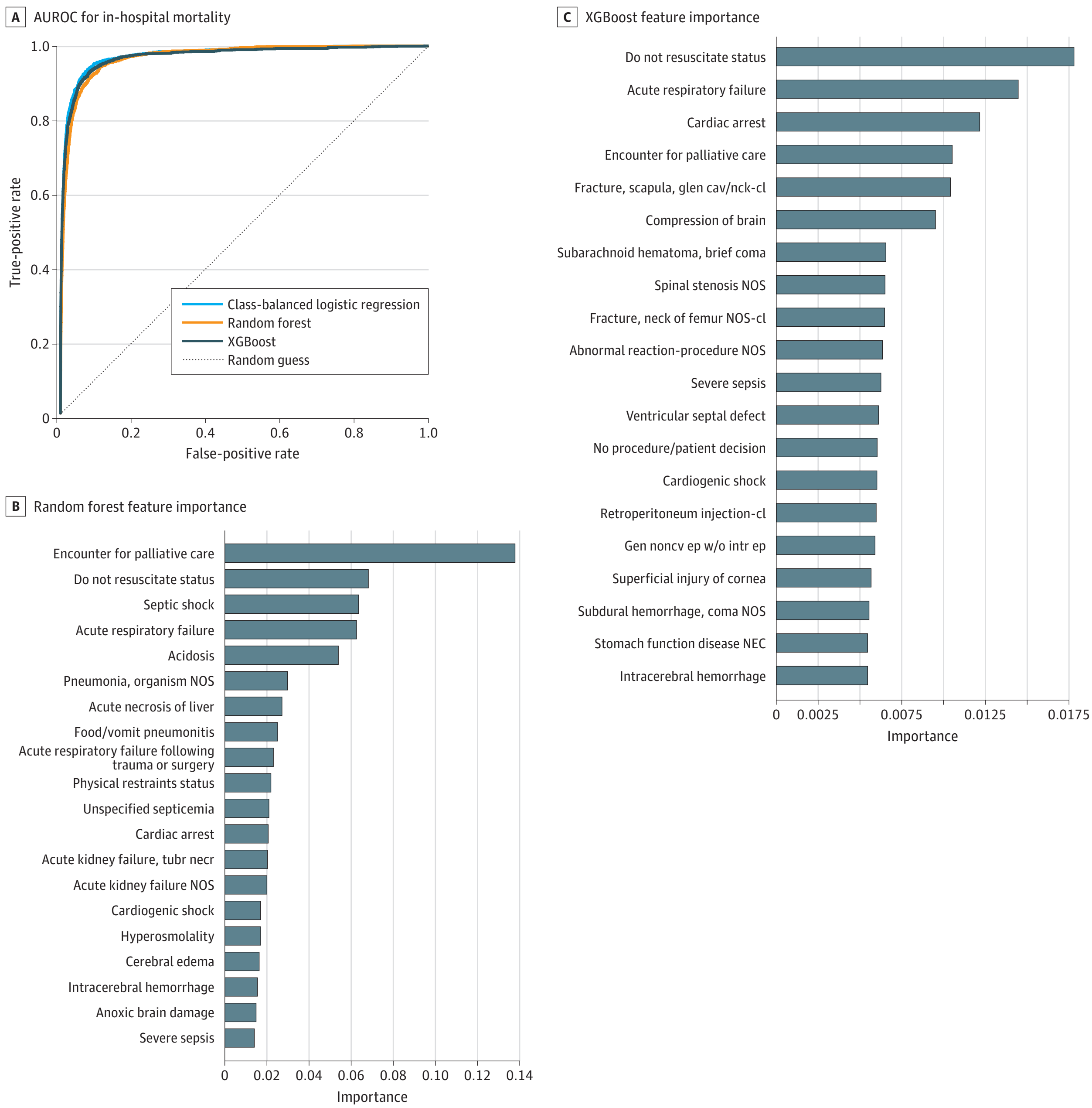
Results

ICD Code–Based Prediction Models

The study cohort included 422 534 hospital admissions from 180 640 unique patients (mean [SD] age at admission, 58.7 [19.2] years; 53.0% female and 47.0% male; 3.5% categorized in MIMIC-IV as Asian, 16.2% as Black, 5.9% as Hispanic, 67.2% as White, 4.1% as other, and 3.3% as unknown race and ethnicity). In-hospital mortality occurred in 8417 admissions (2.0%). In the held-out test set, all 3 models achieved high predictive performance, with AUROCs of 0.976 (95% CI, 0.973-0.980) (logistic regression), 0.971 (95% CI, 0.967-0.974) (random forest), and 0.973 (95% CI, 0.968-0.977) (XGBoost) (**Figure 1**A; eFigure in Supplement 1). These results are even better than published models trained on the same data that also included many additional predictive features from the rest of the EHR.[1,2]

The **Table** highlights the 20 diagnostic codes with the highest ORs used by the logistic regression model. Complete logistic regression feature results are available in eTable 2 in Supplement 2. All codes were statistically significant after the Benjamini-Hochberg procedure ($P <$ .05). Acute diagnoses typically arose during hospitalization and dominated the list, such as subdural

Figure 1. Model Predictive Performance and Feature Importance for Predicting In-Hospital Mortality



A, Shading indicates the 95% CI. AUROC indicates area under the receiver operating characteristic curve; cl, closed; gen noncv ep w/o intr ep, generalized nonconvulsive epilepsy without mention of intractable epilepsy; glen cav/nck, glenoid cavity and the scapular neck; NEC, necrotizing enterocolitis; NOS, not otherwise specified; tubr necr, tubular necrosis.

hematoma, deep coma (OR, 389.99; 95% CI, 28.79-5283.59); cardiac arrest (OR, 219.58; 95% CI, 159.61-302.08); brain death (OR, 112.78; 95% CI, 13.42-947.70); and encounter for palliative care (OR, 98.04; 95% CI, 83.16-115.58), all of which carry an obvious high risk of mortality. The additional features included rare diagnoses and symptoms that occurred in cases more often than controls within this dataset.

Feature importance analyses from the random forest and XGBoost models (Figure 1B) found *ICD* codes for do not resuscitate status (random forest rank, 2nd; XGBoost rank, 1st), acute respiratory failure (random forest rank, 4th; XGBoost rank, 2nd), and encounter for palliative care (random forest rank, 1st; XGBoost rank, 4th) to be powerful predictors of mortality. In addition to *ICD* codes that obviously represent label leakage (eg, brain death), the diagnosis superficial injury to the cornea was the 17th most important feature to the XGBoost model, which stood out as it is not an acute diagnosis. This anomaly may be associated with the model's ability to detect a clinician's focus on documenting less severe conditions, signaling relative patient stability and, therefore, low mortality risk.

Literature Review

Figure 2 outlines our study-screening process. We reviewed 100 studies that built a prediction model from an initial set of the 140 citing MIMIC and sorted them in descending order by the mean number of citations per year (the full list is provided in eTable 1 in Supplement 1).[16-115] Of these articles, 92 (92.0%) reported building predictive models that targeted outcomes within the same admission,[17-19, 21-60, 62-65, 67-79, 82-92, 94-107, 109-112] and among those, 37 (40.2%) used *ICD* diagnostic codes as input features.[17, 21, 22, 26, 30, 34, 35, 38, 40-43, 47, 49, 50, 58-60, 68, 70, 71, 75, 78, 83, 84, 87, 90, 92, 97-99, 102, 103, 107, 109, 111, 115]

Table. Top 20 Features in the Logistic Regression Model

Feature	OR (95% CI)	Adjusted *P* value[a]
Subdural hemorrhage, deep coma	389.99 (28.79-5283.59)	<.001
Cardiac arrest	219.58 (159.61-302.08)	<.001
Brain death	112.78 (13.42-947.70)	<.001
Encounter for palliative care	98.04 (83.16-115.58)	<.001
Transient visual loss	96.12 (45.58-202.69)	<.001
Kidney sclerosis, unspecified	69.83 (43.92-111.03)	<.001
Unspecified intracranial hemorrhage	59.52 (32.96-107.48)	<.001
Acute maxillary sinusitis	37.24 (12.36-112.17)	<.001
Chronic glomerulonephritis with unspecified pathologic lesion in kidney	36.30 (15.77-83.56)	<.001
Subarachnoid hemorrhage following injury without mention of open intracranial wound, with prolonged (>24 h) loss of consciousness without return to preexisting conscious level	32.08 (2.23-461.70)	.04
Abdominal aneurysm, ruptured	30.90 (14.75-64.72)	<.001
Postoperative shock, cardiogenic	28.64 (15.67-52.37)	<.001
Influenza due to identified avian influenza virus with other respiratory manifestations	26.00 (11.72-57.69)	<.001
Other abnormality of urination	25.95 (13.06-51.54)	<.001
Intracerebral hemorrhage	25.85 (21.83-30.60)	<.001
Nonpressure chronic ulcer of other part of right foot with other specified severity	25.70 (6.97-94.78)	<.001
Ulcer of thigh	22.75 (9.76-53.02)	<.001
Acute myeloid leukemia, in relapse	21.79 (13.21-35.94)	<.001
Viral hepatitis B with hepatic coma, acute or unspecified, without mention of hepatitis delta	21.68 (7.21-65.18)	<.001
Unspecified drug dependence, unspecified	20.96 (7.81-56.27)	<.001

Abbreviation: OR, odds ratio.

[a] Benjamini-Hochberg correction.

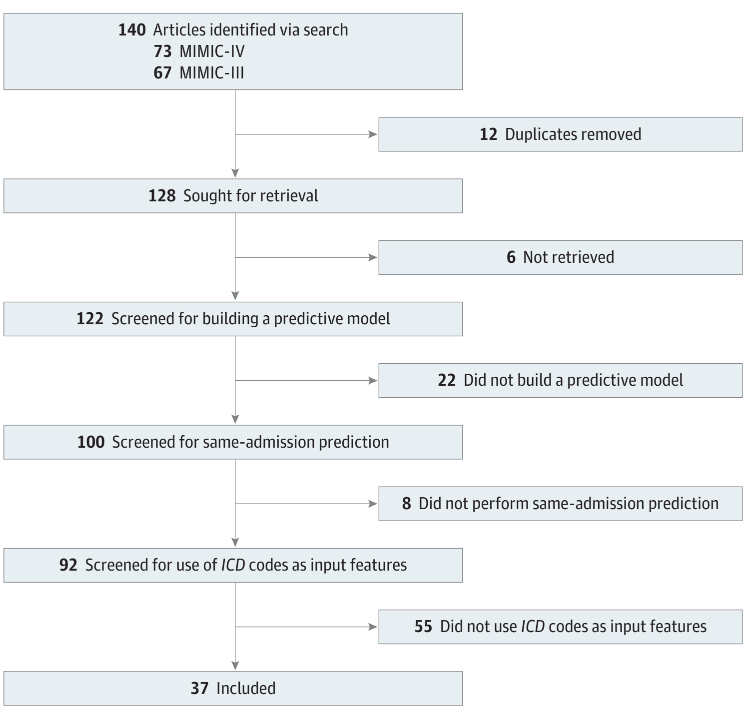
Discussion

This prognostic study found that a specific problem within the machine learning health care literature may be the presence of data leakage in same-admission prediction models associated with the inclusion of diagnostic codes as input features. These codes, finalized only after discharge, provide models with hindsight information that would not be available at the time of prediction. This practice causes 2 distinct and serious problems. First, codes that clinicians document in the EHR after a clinical encounter cannot be used to guide real-time clinical decision making during that encounter. Second, a subset of these codes (eg, brain death for inpatient mortality) document highly correlated events with the outcome being predicted. This issue underscores a broader concern that machine learning models trained with retrospective data risk misrepresenting their value in actual clinical care. If these models do not account for the realities of real-time clinical workflows, their success in research will not translate into meaningful improvements in patient outcomes.

Both MIMIC-III and MIMIC-IV carry explicit warnings against using an admission's *ICD* codes to predict outcomes from that same admission. In MIMIC-III, *ICD-9* codes arise from patient discharges,[14] while MIMIC-IV clarifies that diagnoses are determined by trained professionals after reviewing signed patient notes.[10] These datasets do not provide an audit log of changes or updates to *ICD* codes but, instead, provide only the final set of *ICD* diagnoses. Given the prevalence of *ICD* code use in MIMIC-based studies despite this direct guidance, we suspect that publications on private institutional data, especially those that do not share source code, could potentially be even more likely to be compromised by label leakage.

Researchers aim to harness available knowledge to the greatest extent possible when training models, and there is a reasonable expectation that some diagnoses are known to clinicians shortly after admission (eg, broken limbs, burns). Some information could potentially be gleaned from patient notes or physician problem lists that may be available during a patient's stay. Often, codes are carried over from previous visits, eg, chronic conditions or comorbidities such as diabetes and hypertension, and these can safely be assumed as known. However, diagnoses in the form of *ICD* codes for a given admission in MIMIC are explicitly derived after discharge. In other datasets, it may

Figure 2. Overview of Targeted Literature Screening and Review



The screening involved searching and filtering studies citing Medical Information Mart for Intensive Care III (MIMIC-III) or MIMIC-IV, developing a prediction model, performing same-admission predictions, and using *International Classification of Diseases* (*ICD*) codes as input features.

be possible to use *ICD* codes without label leakage if these codes are time-stamped and derived from problem lists. However, there are still substantial limitations given that these codes are used for billing purposes and represent clinical thinking as opposed to patient state.[3]

Both analyses in this study have a scope limited to the MIMIC dataset. However, thousands of studies have used data from the MIMIC database[10,14] for a wide variety of tasks, including the portion incorporating *ICD* codes for same-admission prediction tasks identified in this study. While it is not possible to quantify this issue for private or institutional datasets, we suspect that similar issues may be at least as prevalent in analyses on less transparent and thoroughly documented datasets. The MIMIC database is well described, with detailed publications, well-developed documentation, and example code for analyses. Institutional and private datasets generally have less transparency and do not allow for reproducibility, reflecting a broader challenge in health care machine learning research.[9] That label leakage occurs this often in a well-defined dataset that explicitly describes the nature of *ICD* codes should raise questions when evaluating research using less transparent datasets and methods.

A solution to the problem of temporal label leakage is to diligently examine the input features to ensure that these features are truly available at the time of prediction, which could be a challenging problem in health care due to the complicated nature of data generation. For example, present-on-admission flags seem like an easy way to decide whether an *ICD* code could be used in same-admission prediction. In reality, the Centers for Medicare & Medicaid Services states that "subsequent to the assignment of the *ICD-10-CM* codes, the [present-on-admission] indicator should then be assigned to those conditions that have been coded."[116] There are many examples of apparent timestamps, which are actually imperfect proxies for when information is known because of the way documentation lags clinical reality. Accordingly, our recommendation is to ensure that model developers are only using data based on the EHR storage time as opposed to either making assumptions about availability or using other timing information. Model developers could visualize the passage of time with patient timelines based on the EHR storage time to emulate the clinical deployment of prediction models. It is critical for research teams to work with clinical domain experts, as well as information technologists and informaticians, to understand the meaning of different timestamps in clinical data. We advise defining the prediction time point a priori and, for any candidate variable, establishing whether it is truly known by that moment through provenance review and clinician or domain-expert input. We also recommend that articles include a brief variable availability statement that names the source and timing assumptions for each variable class and explains how those assumptions align with the intended clinical use.

The utility of *ICD* codes geared at billing for deployable prediction models is debatable, but at a minimum, researchers need to be careful to ensure that the codes are available prior to the time a prediction needs to be made. Ensuring codes would be available may require only using codes from prior admissions, which still requires ensuring that they are not edited during any adjudication processes with payers or deriving these diagnoses from a time-stamped problem list. The MIMIC database, however, does not include either timestamps or codes from the problem list. The frequency of this error suggests a need for researchers to more closely read the documentation of third-party datasets. While it is not possible to estimate how frequently this issue occurs in private, institutional datasets, we believe that the frequency also suggests a need for greater engagement of prediction model developers with experts covering the full data generation (clinicians) and preparation (eg, informaticians and data warehousing teams) process.

Limitations

This study had some limitations. The scope was limited to studies that used the MIMIC-III and MIMIC-IV datasets. Our findings suggest a clear problem within this subset of the literature but did not provide direct evidence of whether or how frequently this issue occurs in studies that used private or other institutional datasets. Furthermore, this analysis did not account for potential differences between MIMIC and private institutional data, which may have different coding practices, data structures, or documentation. While we suspect that similar or greater challenges may exist in less transparent datasets because of their less transparent nature, it is not possible to empirically test this. Thus, this study includes no findings to support that assertion. Any generalization of our findings beyond the MIMIC-based literature would require further investigation.

Conclusions

This prognostic study of patient data in the MIMIC-IV database found that using *ICD* codes as features in same-admission prediction models may be a severe methodological flaw that inflates performance metrics and renders models incapable of making clinically useful predictions in real time. Our literature review found that the practice is common. Addressing this challenge is essential for advancing trustworthy AI in health care.

ARTICLE INFORMATION

Accepted for Publication: October 15, 2025.

Published: December 26, 2025. doi:10.1001/jamanetworkopen.2025.50454

Corresponding Author: Brett K. Beaulieu-Jones, PhD, Division of the Biological Sciences, Department of Biomedical Informatics, The University of Chicago, 5841 S Maryland Ave, M607, Chicago, IL 60637 (beaulieujones@uchicago.edu).

Author Affiliations: Center for Computational Medicine and Clinical Artificial Intelligence, Department of Medicine, University of Chicago, Chicago, Illinois (Ramadan, Liu, Burkhart, Parker, Beaulieu-Jones); MacLean Center for Clinical Medical Ethics, University of Chicago, Chicago, Illinois (Parker).

Author Contributions: Drs Ramadan and Beaulieu-Jones had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Dr Ramadan and Mr Liu contributed equally as co–first authors.

Concept and design: Ramadan, Liu, Parker, Beaulieu-Jones.

Acquisition, analysis, or interpretation of data: Liu, Burkhart, Beaulieu-Jones.

Drafting of the manuscript: Ramadan, Liu, Beaulieu-Jones.

Critical review of the manuscript for important intellectual content: Liu, Burkhart, Parker, Beaulieu-Jones.

Statistical analysis: Liu, Burkhart, Parker, Beaulieu-Jones.

Obtained funding: Beaulieu-Jones.

Administrative, technical, or material support: Liu, Burkhart, Beaulieu-Jones.

Supervision: Ramadan, Parker, Beaulieu-Jones.

Conflict of Interest Disclosures: Dr Parker reported receiving grants from the National Institutes of Health during the conduct of the study. No other disclosures were reported.

Funding/Support: This work was funded, in part, by grant R00NS114850 from the National Institute of Neurological Disorders and Stroke of the National Institutes of Health (Dr Beaulieu-Jones).

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data Sharing Statement: See Supplement 3.

REFERENCES

1. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning for electronic health records. *arXiv*. Preprint posted online January 24, 2018. doi:10.1038/s41746-018-0029-1

2. Renc P, Jia Y, Samir AE, et al. Zero shot health trajectory prediction using transformer. *NPJ Digit Med*. 2024;7(1):256. doi:10.1038/s41746-024-01235-0

3. Beaulieu-Jones BK, Yuan W, Brat GA, et al. Machine learning for patient risk stratification: standing on, or looking over, the shoulders of clinicians? *NPJ Digit Med*. 2021;4(1):62. doi:10.1038/s41746-021-00426-3

4. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS Med*. 2018;15(11): e1002683. doi:10.1371/journal.pmed.1002683

5. Banerjee I, Bhattacharjee K, Burns JL, et al. "Shortcuts" causing bias in radiology artificial intelligence: causes, evaluation, and mitigation. *J Am Coll Radiol*. 2023;20(9):842-851. doi:10.1016/j.jacr.2023.06.025

6. Seah J, Tang C, Buchlak QD, et al. Do comprehensive deep learning algorithms suffer from hidden stratification? a retrospective study on pneumothorax detection in chest radiography. *BMJ Open*. 2021;11(12):e053024. doi:10.1136/bmjopen-2021-053024

7. Nauta M, Walsh R, Dubowski A, Seifert C. Uncovering and correcting shortcut learning in machine learning models for skin cancer diagnosis. *Diagnostics (Basel)*. 2021;12(1):40. doi:10.3390/diagnostics12010040

8. Ong Ly C, Unnikrishnan B, Tadic T, et al. Shortcut learning in medical AI hinders generalization: method for estimating AI model generalization without external data. *NPJ Digit Med*. 2024;7(1):124. doi:10.1038/s41746-024-01118-4

9. Davis SE, Matheny ME, Balu S, Sendak MP. A framework for understanding label leakage in machine learning for health care. *J Am Med Inform Assoc*. 2023;31(1):274-280. doi:10.1093/jamia/ocad178

10. Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data*. 2023;10(1):1. doi:10.1038/s41597-022-01899-x

11. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378. doi:10.1136/bmj-2023-078378

12. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011; 12(85):2825-2830.

13. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *arXiv*. Preprint posted online March 9, 2016. doi:10.1145/2939672.2939785

14. Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016;3(1): 160035. doi:10.1038/sdata.2016.35

15. Liu MC. Data leakage. Github. Accessed November 17, 2025. https://github.com/bbj-lab/data-leakage

16. Hou N, Li M, He L, et al. Predicting 30-days mortality for MIMIC-III patients with Sepsis-3: a machine learning approach using XGBoost. *J Transl Med*. 2020;18(1):462. doi:10.1186/s12967-020-02620-5

17. Li F, Xin H, Zhang J, Fu M, Zhou J, Lian Z. Prediction model of in-hospital mortality in intensive care unit patients with heart failure: machine learning-based, retrospective analysis of the MIMIC-III database. *BMJ Open*. 2021;11 (7):e044779. doi:10.1136/bmjopen-2020-044779

18. Scherpf M, Gräßer F, Malberg H, Zaunseder S. Predicting sepsis with a recurrent neural network using the MIMIC III database. *Comput Biol Med*. 2019;113:103395. doi:10.1016/j.compbiomed.2019.103395

19. Bao C, Deng F, Zhao S. Machine-learning models for prediction of sepsis patients mortality. *Med Intensiva (Engl Ed)*. 2023;47(6):315-325. doi:10.1016/j.medin.2022.06.004

20. Zhong Z, Yuan X, Liu S, Yang Y, Liu F. Machine learning prediction models for prognosis of critically ill patients after open-heart surgery. *Sci Rep*. 2021;11(1):3384. doi:10.1038/s41598-021-83020-7

21. Lei M, Han Z, Wang S, et al. A machine learning-based prediction model for in-hospital mortality among critically ill patients with hip fracture: an internal and external validated study. *Injury*. 2023;54(2):636-644. doi:10.1016/j.injury.2022.11.031

22. Zhu Y, Zhang J, Wang G, et al. Machine learning prediction models for mechanically ventilated patients: analyses of the MIMIC-III database. *Front Med (Lausanne)*. 2021;8:662340. doi:10.3389/fmed.2021.662340

23. McWilliams CJ, Lawson DJ, Santos-Rodriguez R, et al. Towards a decision support tool for intensive care discharge: machine learning algorithm development using electronic healthcare data from MIMIC-III and Bristol, UK. *BMJ Open*. 2019;9(3):e025925. doi:10.1136/bmjopen-2018-025925

24. Zhao QY, Wang H, Luo JC, et al. Development and validation of a machine-learning model for prediction of extubation failure in intensive care units. *Front Med (Lausanne)*. 2021;8:676343. doi:10.3389/fmed.2021.676343

25. Xie F, Zhou J, Lee JW, et al. Benchmarking emergency department prediction models with machine learning and public electronic health records. *Sci Data*. 2022;9(1):658. doi:10.1038/s41597-022-01782-9

26. Hu C, Tan Q, Zhang Q, et al. Application of interpretable machine learning for early prediction of prognosis in acute kidney injury. *Comput Struct Biotechnol J*. 2022;20:2861-2870. doi:10.1016/j.csbj.2022.06.003

27. Wang Z, Zhang L, Huang T, et al. Developing an explainable machine learning model to predict the mechanical ventilation duration of patients with ARDS in intensive care units. *Heart Lung*. 2023;58:74-81. doi:10.1016/j.hrtlng.2022.11.005

28. Nistal-Nuño B. Developing machine learning models for prediction of mortality in the medical intensive care unit. *Comput Methods Programs Biomed*. 2022;216:106663. doi:10.1016/j.cmpb.2022.106663

29. Bendavid I, Statlender L, Shvartser L, et al. A novel machine learning model to predict respiratory failure and invasive mechanical ventilation in critically ill patients suffering from COVID-19. *Sci Rep*. 2022;12(1):10573. doi:10.1038/s41598-022-14758-x

30. Gentimis T, Alnaser AJ, Durante A, Cook K, Steele R. Predicting hospital length of stay using neural networks on MIMIC III data. In: *Proceedings of the 2017 IEEE 15th International Conference on Dependable, Autonomic and Secure Computing, 15th International Conference on Pervasive Intelligence and Computing, 3rd International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*. Institute of Electrical and Electronics Engineers; 2017:1194-1201.

31. Peng X, Li L, Wang X, Zhang H. A machine learning-based prediction model for acute kidney injury in patients with congestive heart failure. *Front Cardiovasc Med*. 2022;9:842873. doi:10.3389/fcvm.2022.842873

32. Hu C, Li L, Li Y, Wang F, Hu B, Peng Z. Explainable machine-learning model for prediction of in-hospital mortality in septic patients requiring intensive care unit readmission. *Infect Dis Ther*. 2022;11(4):1695-1713. doi:10.1007/s40121-022-00671-3

33. Bashar SK, Hossain MB, Ding E, Walkey AJ, McManus DD, Chon KH. Atrial fibrillation detection during sepsis: study on MIMIC III ICU data. *IEEE J Biomed Health Inform*. 2020;24(11):3124-3135. doi:10.1109/JBHI.2020.2995139

34. Zhao QY, Liu LP, Luo JC, et al. A machine-learning approach for dynamic prediction of sepsis-induced coagulopathy in critically ill patients with sepsis. *Front Med (Lausanne)*. 2021;7:637434. doi:10.3389/fmed.2020.637434

35. Hempel L, Sadeghi S, Kirsten T. Prediction of intensive care unit length of stay in the MIMIC-IV dataset. *Appl Sci (Basel)*. 2023;13(12):6930. doi:10.3390/app13126930

36. Liang Y, Zhu C, Tian C, et al. Early prediction of ventilator-associated pneumonia in critical care patients: a machine learning model. *BMC Pulm Med*. 2022;22(1):250. doi:10.1186/s12890-022-02031-w

37. Sayed M, Riaño D, Villar J. Predicting duration of mechanical ventilation in acute respiratory distress syndrome using supervised machine learning. *J Clin Med*. 2021;10(17):3824. doi:10.3390/jcm10173824

38. Hasan MN, Hamdan S, Poudel S, Vargas J, Poudel K. Prediction of length-of-stay at intensive care unit (ICU) using machine learning based on MIMIC-III database. In: *Proceedings of the 2023 IEEE Conference on Artificial Intelligence (CAI)*. Institute of Electrical and Electronics Engineers; 2023:321-323.

39. Beaulieu-Jones BK, Orzechowski P, Moore JH. Mapping patient trajectories using longitudinal extraction and deep learning in the MIMIC-III critical care database. *Pac Symp Biocomput*. 2018;23:123-132.

40. Xie W, Li Y, Meng X, Zhao M. Machine learning prediction models and nomogram to predict the risk of in-hospital death for severe DKA: a clinical study based on MIMIC-IV, eICU databases, and a college hospital ICU. *Int J Med Inform*. 2023;174:105049. doi:10.1016/j.ijmedinf.2023.105049

41. Zhang Y, Hu J, Hua T, Zhang J, Zhang Z, Yang M. Development of a machine learning-based prediction model for sepsis-associated delirium in the intensive care unit. *Sci Rep*. 2023;13(1):12697. doi:10.1038/s41598-023-38650-4

42. Wang B, Li Y, Tian Y, Ju C, Xu X, Pei S. Novel pneumonia score based on a machine learning model for predicting mortality in pneumonia patients on admission to the intensive care unit. *Respir Med*. 2023;217:107363. doi:10.1016/j.rmed.2023.107363

43. Tang F, Xiao C, Wang F, Zhou J. Predictive modeling in urgent care: a comparative study of machine learning approaches. *JAMIA Open*. 2018;1(1):87-98. doi:10.1093/jamiaopen/ooy011

44. Lu Z, Zhang J, Hong J, et al. Development of a nomogram to predict 28-day mortality of patients with sepsis-induced coagulopathy: an analysis of the MIMIC-III database. *Front Med (Lausanne)*. 2021;8:661710. doi:10.3389/fmed.2021.661710

45. Huang B, Liang D, Zou R, et al. Mortality prediction for patients with acute respiratory distress syndrome based on machine learning: a population-based study. *Ann Transl Med*. 2021;9(9):794. doi:10.21037/atm-20-6624

46. Camacho-Cogollo JE, Bonet I, Gil B, Iadanza E. Machine learning models for early prediction of sepsis on large healthcare datasets. *Electronics (Basel)*. 2022;11(9):1507. doi:10.3390/electronics11091507

47. Danilatou V, Nikolakakis S, Antonakaki D, et al. Outcome prediction in critically-ill patients with venous thromboembolism and/or cancer using machine learning algorithms: external validation and comparison with scoring systems. *Int J Mol Sci*. 2022;23(13):7132. doi:10.3390/ijms23137132

48. Zhao Y, Zhang R, Zhong Y, et al. Statistical analysis and machine learning prediction of disease outcomes for COVID-19 and pneumonia patients. *Front Cell Infect Microbiol*. 2022;12:838749. doi:10.3389/fcimb.2022.838749

49. Sun Y, He Z, Ren J, Wu Y. Prediction model of in-hospital mortality in intensive care unit patients with cardiac arrest: a retrospective analysis of MIMIC-IV database based on machine learning. *BMC Anesthesiol*. 2023;23(1):178. doi:10.1186/s12871-023-02138-5

50. Shu T, Huang J, Deng J, et al. Development and assessment of scoring model for ICU stay and mortality prediction after emergency admissions in ischemic heart disease: a retrospective study of MIMIC-IV databases. *Intern Emerg Med*. 2023;18(2):487-497. doi:10.1007/s11739-023-03199-7

51. Ning YL, Sun C, Xu XH, et al. Tendency of dynamic vasoactive and inotropic medications data as a robust predictor of mortality in patients with septic shock: an analysis of the MIMIC-IV database. *Front Cardiovasc Med*. 2023;10:1126888. doi:10.3389/fcvm.2023.1126888

52. Budrionis A, Miara M, Miara P, Wilk S, Bellika JG. Benchmarking PySyft federated learning framework on MIMIC-III dataset. *IEEE Access*. 2021;9:116869-116878. doi:10.1109/ACCESS.2021.3105929

53. Tang H, Jin Z, Deng J, et al. Development and validation of a deep learning model to predict the survival of patients in ICU. *J Am Med Inform Assoc*. 2022;29(9):1567-1576. doi:10.1093/jamia/ocac098

54. Pang K, Li L, Ouyang W, Liu X, Tang Y. Establishment of ICU mortality risk prediction models with machine learning algorithm using MIMIC-IV database. *Diagnostics (Basel)*. 2022;12(5):1068. doi:10.3390/diagnostics12051068

55. Zeng Z, Yao S, Zheng J, Gong X. Development and validation of a novel blending machine learning model for hospital mortality prediction in ICU patients with sepsis. *BioData Min*. 2021;14(1):40. doi:10.1186/s13040-021-00276-5

56. Wang R, Cai L, Liu Y, Zhang J, Ou X, Xu J. Machine learning algorithms for prediction of ventilator associated pneumonia in traumatic brain injury patients from the MIMIC-III database. *Heart Lung*. 2023;62:225-232. doi:10.1016/j.hrtlng.2023.08.002

57. Liu F, Yao J, Liu C, Shou S. Construction and validation of machine learning models for sepsis prediction in patients with acute pancreatitis. *BMC Surg*. 2023;23(1):267. doi:10.1186/s12893-023-02151-y

58. 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 Inform*. 2021;9(7):e23401. doi:10.2196/23401

59. Huang AA, Huang SY. Dendrogram of transparent feature importance machine learning statistics to classify associations for heart failure: a reanalysis of a retrospective cohort study of the Medical Information Mart for Intensive Care III (MIMIC-III) database. *PLoS One*. 2023;18(7):e0288819. doi:10.1371/journal.pone.0288819

60. Yang W, Zou H, Wang M, Zhang Q, Li S, Liang H. Mortality prediction among ICU inpatients based on MIMIC-III database results from the conditional medical generative adversarial network. *Heliyon*. 2023;9(2):e13200. doi:10.1016/j.heliyon.2023.e13200

61. Zhang X, Fei N, Zhang X, Wang Q, Fang Z. Machine learning prediction models for postoperative stroke in elderly patients: analyses of the MIMIC database. *Front Aging Neurosci*. 2022;14:897611. doi:10.3389/fnagi.2022.897611

62. Su Y, Guo C, Zhou S, Li C, Ding N. Early predicting 30-day mortality in sepsis in MIMIC-III by an artificial neural networks model. *Eur J Med Res*. 2022;27(1):294. doi:10.1186/s40001-022-00925-3

63. Chang HH, Chiang JH, Wang CS, et al. Predicting mortality using machine learning algorithms in patients who require renal replacement therapy in the critical care unit. *J Clin Med*. 2022;11(18):5289. doi:10.3390/jcm11185289

64. Liu W, Tao G, Zhang Y, et al. A simple weaning model based on interpretable machine learning algorithm for patients with sepsis: a research of MIMIC-IV and eICU databases. *Front Med (Lausanne)*. 2022;8:814566. doi:10.3389/fmed.2021.814566

65. Nowroozilarki Z, Pakbin A, Royalty J, Lee DKK, Mortazavi BJ. Real-time mortality prediction using MIMIC-IV ICU data via boosted nonparametric hazards. In: *Proceedings of the 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI).* Institute of Electrical and Electronics Engineers; 2021:1-4.

66. Khope SR, Elias S. Critical correlation of predictors for an efficient risk prediction framework of ICU patient using correlation and transformation of MIMIC-III dataset. *Data Sci Eng*. 2022;7(1):71-86. doi:10.1007/s41019-022-00176-6

67. Hirano Y, Shinmoto K, Okada Y, et al. Machine learning approach to predict positive screening of methicillin-resistant *Staphylococcus aureus* during mechanical ventilation using synthetic dataset from MIMIC-IV database. *Front Med (Lausanne)*. 2021;8:694520. doi:10.3389/fmed.2021.694520

68. Gao T, Nong Z, Luo Y, et al. Machine learning-based prediction of in-hospital mortality for critically ill patients with sepsis-associated acute kidney injury. *Ren Fail*. 2024;46(1):2316267. doi:10.1080/0886022X.2024.2316267

69. Jiang M, Pan CQ, Li J, Xu LG, Li CL. Explainable machine learning model for predicting furosemide responsiveness in patients with oliguric acute kidney injury. *Ren Fail*. 2023;45(1):2151468. doi:10.1080/0886022X.2022.2151468

70. Zhou S, Lu Z, Liu Y, et al. Interpretable machine learning model for early prediction of 28-day mortality in ICU patients with sepsis-induced coagulopathy: development and validation. *Eur J Med Res*. 2024;29(1):14. doi:10.1186/s40001-023-01593-7

71. Wang G, Xu J, Lin X, et al. Machine learning-based models for predicting mortality and acute kidney injury in critical pulmonary embolism. *BMC Cardiovasc Disord*. 2023;23(1):385. doi:10.1186/s12872-023-03363-z

72. Tsiklidis EJ, Sinno T, Diamond SL. Predicting risk for trauma patients using static and dynamic information from the MIMIC III database. *PLoS One*. 2022;17(1):e0262523. doi:10.1371/journal.pone.0262523

73. Ko RE, Cho J, Shin MK, et al. Machine learning-based mortality prediction model for critically ill cancer patients admitted to the intensive care unit (CanICU). *Cancers (Basel)*. 2023;15(3):569. doi:10.3390/cancers15030569

74. Tian J, Cui R, Song H, Zhao Y, Zhou T. Prediction of acute kidney injury in patients with liver cirrhosis using machine learning models: evidence from the MIMIC-III and MIMIC-IV. *Int Urol Nephrol*. 2024;56(1):237-247. doi:10.1007/s11255-023-03646-6

75. Wei S, Zhang Y, Dong H, et al. Machine learning-based prediction model of acute kidney injury in patients with acute respiratory distress syndrome. *BMC Pulm Med*. 2023;23(1):370. doi:10.1186/s12890-023-02663-6

76. Liu C, Yao Z, Liu P, et al. Early prediction of MODS interventions in the intensive care unit using machine learning. *J Big Data*. 2023;10(1):55. doi:10.1186/s40537-023-00719-2

77. Ren W, Zou K, Huang S, et al. Prediction of in-hospital mortality of intensive care unit patients with acute pancreatitis based on an explainable machine learning algorithm. *J Clin Gastroenterol*. 2024;58(6):619-626. doi:10.1097/MCG.0000000000001910

78. Zhang J, Li H, Ashrafi N, Yu Z, Placencia G, Pishgar M. Prediction of in-hospital mortality for ICU patients with heart failure. *medRxiv*. Preprint posted online June 25, 2024. doi:10.1101/2024.06.25.24309448

79. Pettinati MJ, Chen G, Rajput KS, Selvaraj N. Practical machine learning-based sepsis prediction. *Annu Int Conf IEEE Eng Med Biol Soc*. 2020;2020:4986-4991.

80. Assaf R, Jayousi R. 30-Day hospital readmission prediction using MIMIC data. In: *Proceedings of the 2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT)*. Institute of Electrical and Electronics Engineers; 2020:1-6.

81. Yang S, Cao L, Zhou Y, Hu C. A retrospective cohort study: predicting 90-day mortality for ICU trauma patients with a machine learning algorithm using XGBoost using MIMIC-III database. *J Multidiscip Healthc*. 2023;16:2625-2640. doi:10.2147/JMDH.S416943

82. Hu F, Zhu J, Zhang S, et al. A predictive model for the risk of sepsis within 30 days of admission in patients with traumatic brain injury in the intensive care unit: a retrospective analysis based on MIMIC-IV database. *Eur J Med Res*. 2023;28(1):290. doi:10.1186/s40001-023-01255-8

83. Wang W, Jin X. Prostate cancer prediction model: a retrospective analysis based on machine learning using the MIMIC-IV database. *Intelligent Pharmacy*. 2023;1(4):268-273. doi:10.1016/j.ipha.2023.04.010

84. Lin S, Lu W, Wang T, et al. Predictive model of acute kidney injury in critically ill patients with acute pancreatitis: a machine learning approach using the MIMIC-IV database. *Ren Fail*. 2024;46(1):2303395. doi:10.1080/0886022X.2024.2303395

85. Tsoni R, Kaldis V, Kapogianni I, Sakagianni A, Feretzakis G, Verykios VS. A machine learning pipeline using KNIME to predict hospital admission in the MIMIC-IV database. In: *Proceedings of the 2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA)*. Institute of Electrical and Electronics Engineers; 2023:1-6.

86. Pattalung TN, Chaichulee S. Comparison of machine learning algorithms for mortality prediction in intensive care patients on multi-center critical care databases. *IOP Conf Ser Mater Sci Eng*. 2021;1163(1):012027. doi:10.1088/1757-899X/1163/1/012027

87. Liu R, Liu H, Li L, Wang Z, Li Y. Predicting in-hospital mortality for MIMIC-III patients: a nomogram combined with SOFA score. *Medicine (Baltimore)*. 2022;101(42):e31251. doi:10.1097/MD.0000000000031251

88. Kang S, Park C, Lee J, Yoon D. Machine learning model for the prediction of hemorrhage in intensive care units. *Healthc Inform Res*. 2022;28(4):364-375. doi:10.4258/hir.2022.28.4.364

89. Yu Z, Ashrafi N, Li H, Alaei K, Pishgar M. Prediction of 30-day mortality for ICU patients with Sepsis-3. *BMC Med Inform Decis Mak*. 2024;24(1):223. doi:10.1186/s12911-024-02629-6

90. Xia M, Jin C, Cao S, et al. Development and validation of a machine-learning model for prediction of hypoxemia after extubation in intensive care units. *Ann Transl Med*. 2022;10(10):577. doi:10.21037/atm-22-2118

91. Langenberger B. Machine learning as a tool to identify inpatients who are not at risk of adverse drug events in a large dataset of a tertiary care hospital in the USA. *Br J Clin Pharmacol*. 2023;89(12):3523-3538. doi:10.1111/bcp.15846

92. Dong L, Liu P, Qi Z, Lin J, Duan M. Development and validation of a machine-learning model for predicting the risk of death in sepsis patients with acute kidney injury. *Heliyon*. 2024;10(9):e29985. doi:10.1016/j.heliyon.2024.e29985

93. Xia Z, Xu P, Xiong Y, Lai Y, Huang Z. Survival prediction in patients with hypertensive chronic kidney disease in intensive care unit: a retrospective analysis based on the MIMIC-III database. *J Immunol Res*. 2022;2022:3377030. doi:10.1155/2022/3377030

94. Medina M, Sala P. On the early detection of sepsis in MIMIC-III. In: *Proceedings of the 2021 IEEE 9th International Conference on Healthcare Informatics (ICHI)*. Institute of Electrical and Electronics Engineers; 2021: 171-180.

95. Jiang X, Dai W, Cai Y. Comparison of machine learning algorithms to SAPS II in predicting in-hospital mortality of fractures of the pelvis and acetabulum: analyzes based on MIMIC-III database. *All Life*. 2022;15(1):1000-1012. doi:10.1080/26895293.2022.2125448

96. Mu S, Yan D, Tang J, Zheng Z. Predicting mortality in sepsis-associated acute respiratory distress syndrome: a machine learning approach using the MIMIC-III database. *J Intensive Care Med*. 2025;40(3):294-302. doi:10.1177/08850666241281060

97. Sanii J, Chan WY. Explainable machine learning models for pneumonia mortality risk prediction using MIMIC-III data. In: *Proceedings of the 2022 9th International Conference on Soft Computing & Machine Intelligence (ISCMI)*. Institute of Electrical and Electronics Engineers; 2022:68-73.

98. Khope SR, Elias S. Simplified & novel predictive model using feature engineering over MIMIC-III dataset. *Procedia Comput Sci*. 2023;218:1968-1976. doi:10.1016/j.procs.2023.01.173

99. Yuan W, Xiao M, Wang R, Liu G, Wu J, Wang X. XGBoost in the prediction of 28-day mortality in critical elderly patients with hip fracture: a MIMIC-IV cohort study. *Altern Ther Health Med*. 2024;30(9):432-436.

100. Reponen M. Predicting prescribed medications from the MIMIC-IV event and measurement data. University of Eastern Finland. September 2023. Accessed November 17, 2025. https://dspace.uef.fi/bitstream/handle/123456789/30580/urn_nbn_fi_uef-20231201.pdf?sequence=1&isAllowed=y

101. Shi J, Chen F, Zheng K, et al. Clinical nomogram prediction model to assess the risk of prolonged ICU length of stay in patients with diabetic ketoacidosis: a retrospective analysis based on the MIMIC-IV database. *BMC Anesthesiol*. 2024;24(1):86. doi:10.1186/s12871-024-02467-z

102. Huang S, Teng Y, Du J, Zhou X, Duan F, Feng C. Internal and external validation of machine learning-assisted prediction models for mechanical ventilation-associated severe acute kidney injury. *Aust Crit Care*. 2023;36(4): 604-612. doi:10.1016/j.aucc.2022.06.001

103. Tu Y, Zhang J, Zhao M, He F. Nomogram establishment for short-term survival prediction in ICU patients with aplastic anemia based on the MIMIC-IV database. *Hematology*. 2024;29(1):2339778. doi:10.1080/16078454.2024.2339778

104. Henriksson F, Svensson P. Predicting patient outcome from clinical journals and biomedical articles: using the MIMIC-IV database, multiple in-hospital mortality prediction models are created, to which improvements are attempted through the use of word embeddings trained on scientific biomedical literature. Chalmers University of Technology. 2022. Accessed November 17, 2025. https://odr.chalmers.se/items/a104afb1-3220-494c-b149-0a85b9d18117

105. Royalty JP. Machine learning time-to-event mortality prediction in MIMIC-IV critical care database. Texas A&M. July 24, 2021. Accessed November 17, 2025. https://hdl.handle.net/1969.1/194429

106. Lin X, Pan X, Yang Y, et al. Machine learning models to predict 30-day mortality for critical patients with myocardial infarction: a retrospective analysis from MIMIC-IV database. *Front Cardiovasc Med*. 2024;11:1368022. doi:10.3389/fcvm.2024.1368022

107. Lin S, Lu W, Wang T, et al. ML-Based AKI prediction in acute pancreatitis: innovative models from MIMIC-IV database. *Research Square*. Preprint posted online September 25, 2023. doi:10.21203/rs.3.rs-3347996/v1

108. Jung J, Kim D, Hwang I. Exploring predictive factors for heart failure progression in hypertensive patients based on medical diagnosis data from the MIMIC-IV database. *Bioengineering (Basel)*. 2024;11(6):531. doi:10.3390/bioengineering11060531

109. Chen Y, Zong C, Zou L, et al. A novel clinical prediction model for in-hospital mortality in sepsis patients complicated by ARDS: a MIMIC IV database and external validation study. *Heliyon*. 2024;10(13):e33337. doi:10.1016/j.heliyon.2024.e33337

110. Sheng S, Li A, Liu X, et al. Factors and machine learning models for predicting successful discontinuation of continuous renal replacement therapy in critically ill patients with acute kidney injury: a retrospective cohort study based on MIMIC-IV database. *BMC Nephrol*. 2024;25(1):407. doi:10.1186/s12882-024-03844-z

111. Li J, Sun Y, Ren J, Wu Y, He Z. Machine learning for in-hospital mortality prediction in critically ill patients with acute heart failure: a retrospective analysis based on MIMIC-IV databases. *Research Square*. Preprint posted online January 8, 2024. doi:10.21203/rs.3.rs-3834698/v1

112. Kakadiaris A. Evaluating the fairness of the MIMIC-IV dataset and a baseline algorithm: application to the ICU length of stay prediction. *arXiv*. Preprint posted online December 31, 2023. doi:10.48550/arXiv.2401.00902

113. Lin MY, Chi HY, Chao WC. Multitask learning to predict successful weaning in critically ill ventilated patients: a retrospective analysis of the MIMIC-IV database. *Digit Health*. Published online October 8, 2024. doi:10.1177/20552076241289732

114. Yang M, Hu W, Yan J. Development of machine learning models for predicting acute respiratory distress syndrome: evidence from the MIMIC-III and MIMIC-IV. *Research Square*. Preprint posted online September 1, 2023. doi:10.21203/rs.3.rs-3221576/v1

115. Liu Y, Mo W, Wang H, Shao Z, Zeng Y, Bi J. Feature selection and risk prediction for diabetic patients with ketoacidosis based on MIMIC-IV. *Front Endocrinol (Lausanne)*. 2024;15:1344277. doi:10.3389/fendo.2024.1344277

116. *ICD-10-CM* official guidelines for coding and reporting. Centers for Medicare & Medicaid Services. Updated April 1, 2024. Accessed October 6, 2025. https://www.cms.gov/files/document/fy-2024-icd-10-cm-coding-guidelines-updated-02/01/2024.pdf

SUPPLEMENT 1.

eTable 1. Studies Citing MIMIC That Built a Predictive Model
eFigure. Calibration Curves and Predicted Distributions for Trained Classifiers
eReferences.

SUPPLEMENT 2.

eTable 2. Full List of Features' Odds Ratios for the Logistic Regression Model

SUPPLEMENT 3.

Data Sharing Statement