JOURNAL OF MULTIDISCIPLINARY RESEARCH AND INTEGRATED SCIENCES

Volume 2, Issue 1, 2022 | **Article ID:** JMRIS-2022-001 | **Pages:** 1-13 | **Article Type:** Systematic Review
Received: 12-12-2021 | **Revised:** 25-01-2022 | **Accepted:** 15-02-2022 | **Published:** 15-03-2022

Machine-Learning Prediction of Intensive Care Unit Admission Among Adults Presenting With Community-Acquired Pneumonia Using Routinely Collected Clinical Variables: A Systematic Evidence Synthesis and Model-Development Framework

Jonathan Reeves[1*]

[1] *University of Toronto, Toronto, Canada*

***Corresponding author:** Jonathan Reeves, King's College London, London, United Kingdom | Jonreeves124125@gmail.com

ABSTRACT

Background: Community-acquired pneumonia (CAP) ranges from mild disease manageable outside hospital to rapidly progressive respiratory or circulatory failure requiring intensive care. Accurate identification of patients likely to deteriorate remains difficult. Conventional scores such as the Pneumonia Severity Index (PSI), CURB-65, SMART-COP and Risk of Early Admission to Intensive Care Unit (REA-ICU) provide structured risk assessment, but fixed weighting and limited interaction modelling may restrict performance. Routinely collected electronic clinical data provide an opportunity for machine-learning prediction. **Objective:** To synthesize high-quality evidence available through 2021 regarding prediction of intensive care unit (ICU) requirement in adults with CAP and to define a reproducible machine-learning framework using routinely collected clinical variables. **Methods:** A structured evidence synthesis was conducted using studies published no later than 31 December 2021 in first-quartile journals relevant to respiratory medicine, infectious diseases, critical care, clinical epidemiology and medical informatics. CAP severity scores, ICU-prediction studies and methodological literature on clinical prediction modelling were examined. Candidate predictors were restricted to variables realistically available at or shortly after emergency-department presentation. **Results:** Existing prediction systems consistently identified respiratory rate, blood pressure, oxygenation, mental status, urea, age, comorbidity, radiographic extent and selected laboratory parameters as markers of severe disease. SMART-COP was designed specifically to identify intensive respiratory or vasopressor support and demonstrated high sensitivity in its derivation and validation cohorts. REA-ICU stratified risk of ICU admission within the first three hospital days from approximately 0.7% to 31%. More recent weighted approaches achieved discrimination approaching an AUC of 0.91. Electronic-health-record studies demonstrated that automated multivariable prediction using routinely captured variables was feasible. However, methodological literature showed that machine learning does not inherently outperform logistic regression and that calibration, external validation and clinical utility are essential. **Conclusion:** Machine learning offers a plausible method for improving early CAP triage when applied to clinically available variables and evaluated against established severity scores. A robust model should prioritize early availability, prevent information leakage, address missing data within validation folds, compare interpretable and nonlinear algorithms, and report discrimination, calibration and decision-analytic utility. External validation is required before clinical implementation.

Keywords: community-acquired pneumonia; machine learning; intensive care unit; clinical prediction; risk stratification; electronic health records; severe pneumonia

1. Introduction

Community-acquired pneumonia remains a clinically heterogeneous disease. Some adults recover with outpatient antimicrobial therapy, whereas others develop hypoxaemic respiratory failure, shock, multiorgan dysfunction or death.

A central challenge at emergency-department presentation is therefore determining the appropriate site and intensity of care.

Traditional prognostic tools initially focused primarily on mortality.

Fine et al. developed the Pneumonia Severity Index (PSI) using demographic characteristics, comorbid conditions, examination findings and laboratory variables to identify patients at low mortality risk [1]. The PSI remains one of the best-established CAP risk tools but contains numerous variables and was not primarily designed to determine the requirement for intensive care.

CURB-65 subsequently provided a more parsimonious alternative based on confusion, blood urea nitrogen, respiratory rate, blood pressure and age ≥65 years [2].

Although mortality prediction is clinically important, the prediction of ICU-level deterioration represents a different problem.

A patient may require mechanical ventilation or vasopressor support and survive. Conversely, an older patient with substantial comorbidity may have a high predicted mortality risk without necessarily being an appropriate or imminent ICU candidate.

This distinction motivated development of ICU-oriented tools.

The 2007 Infectious Diseases Society of America/American Thoracic Society (IDSA/ATS) criteria defined severe CAP using

major and minor criteria [3]. Charles et al. subsequently developed SMART-COP specifically to predict intensive respiratory or vasopressor support [4].

Renaud et al. developed the REA-ICU index to predict ICU admission during the first three hospital days among patients who had no obvious requirement for immediate intensive care [5].

The increasing availability of structured electronic health records provides an opportunity to extend these approaches.

Machine-learning algorithms can accommodate nonlinear relationships and interactions without requiring the clinician to calculate a fixed point score manually. However, complexity does not guarantee better prediction. Clinical machine-learning models must be evaluated against simpler statistical models and established clinical rules [11-16].

The present study therefore aimed to synthesize pre-2022 evidence concerning prediction of ICU-level deterioration in adults with CAP; identify routinely available candidate predictors; establish an appropriate outcome definition for machine-learning development; compare conventional and machine-learning approaches; and propose a reproducible framework for model development, validation and eventual clinical implementation.

2. Methods

2.1 Evidence-Synthesis Design

A structured evidence synthesis was performed using literature available through 31 December 2021.

Search concepts included:

community-acquired pneumonia, severe pneumonia, intensive care, ICU admission, mechanical ventilation, vasopressor support, clinical prediction, severity score, machine learning, electronic health record, risk model, SMART-COP, REA-ICU, PSI and CURB-65.

Evidence was restricted to peer-reviewed Q1 journals within relevant clinical and methodological categories.

2.2 Eligibility Criteria

Included studies were required to examine at least one of the following areas: CAP severity, ICU admission, mechanical ventilation, vasopressor requirement, early clinical deterioration, electronic CAP prediction, or methodological standards for clinical prediction modelling.

Variables unavailable until after the ICU decision were considered inappropriate for an admission-time prediction model.

2.3 Intended Prediction Question

The proposed clinical question was whether routinely available clinical variables among adults presenting with community-acquired pneumonia can predict requirement for ICU-level care during the early course of hospitalization.

The intended time zero is the patient's initial emergency-department or hospital assessment.

The preferred outcome is unplanned ICU admission or initiation of intensive respiratory or vasopressor support within 72 hours.

Patients requiring immediate mechanical ventilation or vasopressors at initial presentation can either be excluded from model development or classified separately, because no prediction model is required to recognize an already established ICU indication.

3. Existing CAP Prediction Systems

Major pre-2022 CAP severity and ICU-risk approaches are summarised in Table 1.

3.1 Pneumonia Severity Index and CURB-65

The PSI established that combinations of routinely collected variables can provide clinically meaningful risk stratification [1].

However, age contributes substantially to the PSI score. This is appropriate for mortality prediction but may reduce its usefulness for predicting acute physiological deterioration in younger patients with severe pneumonia.

CURB-65 is substantially simpler [2].

Its five components permit rapid bedside use, but simplification comes at the cost of reduced physiological granularity. Oxygenation, radiographic extent, pH and several laboratory abnormalities associated with critical illness are not explicitly represented.

Thus, PSI and CURB-65 are valuable comparators for any machine-learning model, but the ML outcome should not simply replicate their mortality endpoint.

3.2 IDSA/ATS Severe-Pneumonia Criteria

The IDSA/ATS framework identifies severe CAP using major and minor criteria [3].

Major criteria include invasive mechanical ventilation and septic shock requiring vasopressors.

These are essentially established manifestations of critical illness.

Minor criteria include respiratory rate ≥30/min, low PaO_2/FiO_2, multilobar infiltrates, confusion, uraemia, leukopenia, thrombocytopenia, hypothermia and hypotension requiring aggressive fluid resuscitation.

Subsequent validation studies supported their relationship with severe outcomes [6,7].

From a machine-learning perspective, these criteria are informative because they provide an evidence-based initial feature set.

However, conventional scoring assigns predetermined thresholds. Machine learning can instead retain continuous values and investigate nonlinear relationships.

A respiratory rate of 29 and 30 breaths/minute, for example, need not be treated as fundamentally different observations.

3.3 SMART-COP

SMART-COP was explicitly designed to identify patients likely to require intensive respiratory or vasopressor support [4].

Its components comprise systolic blood pressure, multilobar chest-radiograph involvement, albumin, respiratory rate, tachycardia, confusion, oxygenation and arterial pH.

In the original derivation cohort, 10.3% of patients required intensive respiratory or vasopressor support.

A SMART-COP score ≥3 identified approximately 92% of patients who subsequently required such support [4].

The score therefore represents a strong benchmark for machine-learning ICU prediction.

3.4 REA-ICU

REA-ICU was developed specifically for patients without obvious criteria for immediate ICU admission [5].

The score incorporated 11 variables including sex, age, comorbidity, respiratory rate, heart rate, radiographic disease, leukocyte count, hypoxaemia, urea, arterial pH and sodium.

Risk classes corresponded to early ICU-admission probabilities ranging from approximately 0.7% to 31%, and overall discrimination was approximately AUC 0.81 [5].

This design is particularly useful conceptually because it asks the clinically relevant question:

Which apparently non-critical patient is likely to deteriorate shortly after admission?

That should also be the objective of a machine-learning model.

4. Candidate Predictors for Machine Learning

The predictor set should satisfy two requirements: demonstrated clinical relevance and realistic availability at prediction time.

The routinely collected clinical feature architecture proposed for early ICU-admission prediction is presented in Figure 1.

Recommended routinely available candidate predictors are summarised in Table 2.

A parsimonious primary model should avoid exotic biomarkers not routinely available across emergency departments.

The principal advantage of machine learning in this context is not necessarily collecting more information.

It is extracting more information from variables already available.

5. Why Machine Learning May Add Value

Traditional clinical scores generally impose three constraints: predictor values are categorized, each category receives a predetermined weight, and interactions are rarely modeled.

Machine learning can potentially overcome these limitations.

For example, the effect of respiratory rate may differ according to oxygen saturation, age and blood pressure.

A gradient-boosting model can represent such interactions without explicitly defining each one beforehand.

However, evidence from clinical prediction research indicates that machine-learning models frequently fail to outperform properly developed logistic-regression models [12].

Complexity should therefore be treated as a hypothesis rather than an advantage.

Candidate algorithms for CAP ICU prediction are compared in Table 3.

Christodoulou et al. systematically compared logistic regression with machine-learning approaches in clinical prediction and found no consistent performance advantage for machine learning [12].

The proper comparison is therefore not machine learning versus no model, but machine learning versus strong statistical and clinical benchmarks.

6. Model-Development Framework

The recommended model-development and validation pipeline is presented in Figure 2.

6.1 Cohort Definition

Eligible participants should include adults aged ≥18 years presenting with clinically and radiographically confirmed CAP.

Patients with hospital-acquired or ventilator-associated pneumonia should be excluded.

A separate decision is required regarding patients with immediate major severity criteria.

For a deterioration model, they should preferably be excluded because their ICU need is already apparent.

6.2 Missing Data

Missingness is unavoidable in routine clinical datasets.

Complete-case analysis can discard substantial information and create selection bias.

Multiple imputation or model-specific imputation strategies are preferable when assumptions are reasonable.

Crucially, imputation must be estimated within training data.

Calculating imputation values using the complete dataset before train-test separation introduces information from the validation sample into model development.

6.3 Continuous Variables

Continuous predictors should not routinely be dichotomized using conventional clinical thresholds.

For example, urea, respiratory rate, sodium and oxygen saturation should initially be preserved as continuous measurements.

Machine-learning algorithms can learn nonlinear effects directly.

Logistic-regression comparators can use restricted cubic splines where appropriate.

6.4 Class Imbalance

ICU admission is expected to occur in a minority of CAP presentations.

Therefore, accuracy alone is an inappropriate performance measure.

A model that predicts “no ICU admission” for every patient may achieve deceptively high accuracy in a low-event population.

Evaluation should include AUROC, area under the precision-recall curve, sensitivity, specificity, positive predictive value, negative predictive value and calibration.

7. Validation Strategy

A simple random 70:30 development-test split is convenient but statistically inefficient, particularly when ICU events are limited.

Prediction-model methodology favors resampling approaches for internal validation [13-16].

The recommended validation strategy is summarised in Table 4.

External validation is particularly important because ICU-admission decisions depend partly on bed availability, institutional policy, clinician behaviour, patient preferences and local thresholds for respiratory support.

A model trained in one institution may therefore learn both disease severity and local practice patterns.

8. Performance Assessment

Model discrimination is only one component of performance.

An AUROC describes ranking: whether higher-risk patients tend to receive higher predictions.

It does not establish that predicted probabilities are numerically accurate.

Van Calster et al. emphasized calibration as a central requirement for prediction models [15].

A model predicting a 30% risk of ICU admission should ideally observe approximately 30 ICU events per 100 similar predictions.

The minimum model-performance reporting set is summarised in Table 5.

Clinical utility should be evaluated across realistic risk thresholds.

For example, an emergency physician may tolerate considerable over-triage if the alternative is failing to identify a patient who will require ventilation within hours.

The optimal threshold therefore depends on consequences rather than statistical performance alone.

9. Evidence Supporting Automated Prediction

By 2021, computerized CAP prediction had become feasible at substantial scale.

Jones et al. evaluated electronic severity assessment among adults presenting with CAP across 117 Veterans Affairs Medical Centers [10].

The study compared a computerized PSI-type approach with additional logistic, nonlinear and machine-learning models using routinely captured electronic variables.

This demonstrates an important practical principle:

CAP severity prediction can be automated directly from electronic clinical data.

Automation offers several potential advantages.

First, the model can operate without requiring manual calculation.

Second, risk can be recalculated as new clinical information becomes available.

Third, predictions can be integrated into emergency-department workflows.

However, dynamic prediction raises a further methodological issue.

A model designed for initial triage must not accidentally incorporate variables measured because clinicians already suspected deterioration.

Otherwise, the algorithm predicts clinical decision-making rather than underlying disease progression.

10. Proposed Primary Model

A clinically useful initial model should prioritize approximately 15–25 routinely available variables.

A recommended development set would include age, sex, major chronic comorbidities, mental status, respiratory rate, heart rate, systolic blood pressure, temperature, oxygen saturation, supplemental oxygen requirement, urea, sodium, glucose, white blood-cell count, haematocrit, platelets, pH when available, multilobar radiographic involvement and pleural effusion.

Three core algorithms should then be compared: penalized logistic regression, random forest and gradient boosting.

PSI, CURB-65 and SMART-COP should serve as clinical benchmarks.

The objective should not be selecting the algorithm with the highest third decimal place of AUROC.

Model selection should consider discrimination, calibration, parsimony, interpretability, reproducibility, robustness to missingness and clinical utility.

11. Interpretation and Explainability

Prediction and explanation are different objectives.

A variable with high predictive importance does not necessarily cause ICU admission.

For example, supplemental oxygen use may strongly predict ICU transfer because it reflects existing respiratory severity.

Feature-importance techniques should therefore be described as predictive contributions, not causal effects.

Global interpretation can determine which predictors influence the model overall.

Patient-level interpretation can help clinicians understand why an individual prediction is high.

However, explanations should complement rather than replace rigorous validation.

An interpretable poorly calibrated model remains clinically unsafe.

12. Discussion

The evidence available through 2021 provides a strong basis for machine-learning prediction of ICU requirement in CAP.

Importantly, the clinical problem is already well characterized.

Respiratory rate, hypoxaemia, haemodynamic instability, mental-status change, renal/metabolic abnormalities and radiographic extent repeatedly appear across established severity systems [1-7].

Machine learning therefore does not require discovering an entirely new set of biomarkers.

Its potential contribution is improved integration of familiar information.

The most important methodological decision is the endpoint.

Mortality, hospital admission and ICU admission should not be considered interchangeable outcomes.

For early triage, ICU admission or intensive respiratory/vasopressor support is more closely aligned with the decision being supported.

SMART-COP and REA-ICU provide particularly relevant benchmarks because they were developed around critical-care outcomes [4,5].

The second major consideration is timing.

Prediction models should use only information available before the decision they aim to predict.

Including downstream interventions can create severe information leakage and artificially inflate model performance.

Third, machine learning should not be presumed superior.

Christodoulou et al. demonstrated that apparent superiority of ML over logistic regression is not reliably supported across clinical-prediction studies [12].

A well-specified logistic model therefore remains an essential comparator.

Fourth, calibration must receive equivalent attention to discrimination.

Two models can have similar AUROC values while producing substantially different individual probabilities.

Because ICU triage decisions will often use probability thresholds, poorly calibrated probabilities can directly influence resource allocation.

Finally, external validation is indispensable.

CAP epidemiology, admission practices and critical-care resources vary geographically and temporally.

A model developed in one healthcare system should not be deployed elsewhere solely because internal discrimination appears high.

13. Clinical Implementation

A future validated model could be embedded within an electronic health record.

After routine observations and laboratory results become available, the system could calculate ICU-deterioration probability automatically.

A clinically sensible interface might present three categories: low risk, for which standard management is appropriate unless clinical judgment indicates otherwise; intermediate risk, prompting enhanced observation, repeat assessment and consideration of higher-acuity placement; and high risk, prompting urgent senior and critical-care review.

The model should never automatically determine ICU admission.

ICU care depends on factors beyond statistical risk, including treatment limitations, frailty, patient preferences, reversibility, resource availability and clinician assessment.

The model should therefore function as decision support, not decision replacement.

14. Strengths and Limitations

This synthesis deliberately focuses on prediction of ICU-level deterioration rather than mortality alone.

It also emphasizes routinely collected variables, increasing potential clinical feasibility.

The principal limitation is that machine-learning studies directly predicting CAP ICU admission were still comparatively limited by the end of 2021.

Consequently, the framework integrates evidence from established CAP risk scores, computerized prognosis research and broader prediction-model methodology.

Second, ICU admission is an imperfect outcome because it reflects both physiological severity and healthcare-system practices.

A combined endpoint incorporating mechanical ventilation or vasopressor support may reduce this limitation.

Third, no novel patient-level model was trained in the present study.

This was intentional. Reporting algorithm performance without an actual accessible cohort would create fabricated precision.

The paper therefore establishes the analytical framework required for a subsequent reproducible modelling study.

15. Future Research

A high-quality next-stage study should use a multicentre CAP cohort with sufficient ICU events.

Sample size should be determined from prediction-model principles rather than a simplistic fixed events-per-variable rule [16].

Development should follow TRIPOD reporting recommendations [13], and risk of bias should be assessed using PROBAST principles [14].

A particularly strong study design would include geographically separate development and validation sites, a prespecified 72-hour deterioration endpoint, comparison against PSI, CURB-65 and SMART-COP, penalized logistic regression, random forest and gradient boosting, nested hyperparameter tuning, multiple imputation within resampling, calibration assessment, decision-curve analysis and model availability for independent validation.

Only after this process should prospective clinical-impact testing be considered.

16. Conclusion

Early identification of adults with community-acquired pneumonia who will require intensive care remains a clinically important prediction problem.

Existing tools demonstrate that routinely collected information already contains substantial prognostic signal.

SMART-COP and REA-ICU are particularly relevant because they target critical-care outcomes rather than mortality alone.

Machine learning may improve upon fixed clinical scores by preserving continuous variables, modeling nonlinear relationships and incorporating interactions automatically.

However, complexity itself provides no guarantee of superior performance.

A clinically credible CAP machine-learning model should therefore be built around five principles:

appropriate outcome definition, temporally valid predictors, rigorous internal validation, external validation, and complete assessment of calibration and clinical utility.

The most useful future system is unlikely to be the most complex algorithm.

It will be the model that provides accurate, calibrated, interpretable and transportable early risk estimates using information already available to clinicians at the bedside.

Declarations

Author Contributions

The authors contributed to conceptualisation, evidence synthesis, methodological interpretation, manuscript drafting and critical revision.

Funding

No external funding was received.

Ethical Approval

Not applicable. This study synthesised previously published evidence and did not involve direct patient-level data collection.

Consent to Participate

Not applicable.

Data Availability

No novel patient-level dataset was generated or analysed. All evidence was obtained from previously published research.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Fine MJ, Auble TE, Yealy DM, Hanusa BH, Weissfeld LA, Singer DE, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. N Engl J Med. 1997;336(4):243-250. doi:10.1056/NEJM199701233360402.
2. Lim WS, van der Eerden MM, Laing R, Boersma WG, Karalus N, Town GI, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax. 2003;58(5):377-382. doi:10.1136/thorax.58.5.377.
3. Mandell LA, Wunderink RG, Anzueto A, Bartlett JG, Campbell GD, Dean NC, et al. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. Clin Infect Dis. 2007;44 Suppl 2:S27-S72. doi:10.1086/511159.
4. Charles PGP, Wolfe R, Whitby M, Fine MJ, Fuller AJ, Stirling R, et al. SMART-COP: a tool for predicting the need for intensive respiratory or vasopressor support in community-acquired pneumonia. Clin Infect Dis. 2008;47(3):375-384. doi:10.1086/589754.
5. Renaud B, Labarère J, Coma E, Santin A, Hayon J, Gurgui M, et al. Risk stratification of early admission to the intensive care unit of patients with no major criteria of severe community-acquired pneumonia: development of an international prediction rule. Crit Care. 2009;13(2):R54. doi:10.1186/cc7781.
6. Phua J, See KC, Chan YH, Widjaja LS, Aung NW, Ngerng WJ, et al. Validation and clinical implications of the IDSA/ATS minor criteria for severe community-acquired pneumonia. Thorax. 2009;64(7):598-603.
7. Chalmers JD, Taylor JK, Mandal P, Choudhury G, Singanayagam A, Akram AR, et al. Validation of the Infectious Diseases Society of America/American Thoracic Society minor criteria for intensive care unit admission in community-acquired pneumonia patients without major criteria or contraindications to intensive care unit care. Clin Infect Dis. 2011;53(6):503-511. doi:10.1093/cid/cir463.
8. Renaud B, Schuetz P, Claessens YE, Labarère J, Albrich W, Mueller B. Proadrenomedullin improves Risk of Early Admission to ICU score for predicting early severe community-acquired pneumonia. Chest. 2012;142(6):1447-1454. doi:10.1378/chest.11-2574.
9. Gearhart AM, Furmanek S, English C, Ramirez J, Cavallazzi R. Predicting the need for ICU admission in community-acquired pneumonia. Respir Med. 2019;155:61-65. doi:10.1016/j.rmed.2019.07.007.
10. Jones BE, Ying J, Nevers M, Alba PR, He T, Patterson OV, et al. Computerized mortality prediction for community-acquired pneumonia at 117 Veterans Affairs Medical Centers. Ann Am Thorac Soc. 2021;18(7):1175-1184. doi:10.1513/AnnalsATS.202011-1372OC.
11. Metlay JP, Waterer GW, Long AC, Anzueto A, Brozek J, Crothers K, et al. Diagnosis and treatment of adults with community-acquired pneumonia: an official clinical practice guideline of the American Thoracic Society and Infectious Diseases Society of America. Am J Respir Crit Care Med. 2019;200(7):e45-e67. doi:10.1164/rccm.201908-1581ST.
12. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction

models. J Clin Epidemiol. 2019;110:12-22. doi:10.1016/j.jclinepi.2019.02.004.

13. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162(1):55-63. doi:10.7326/M14-0697.
14. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51-58. doi:10.7326/M18-1376.
15. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17:230. doi:10.1186/s12916-019-1466-7.
16. Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. BMJ. 2020;368:m441. doi:10.1136/bmj.m441.

Table 1. Major pre-2022 CAP severity and ICU-risk approaches

Tool/model	Primary purpose	Major predictors	Principal strength	Principal limitation
PSI [1]	Mortality/site-of-care stratification	Age, demographics, comorbidity, examination, laboratory and radiographic findings	Extensively validated	Complex; mortality rather than ICU need
CURB-65 [2]	Mortality severity	Confusion, urea, RR, BP, age	Simple and rapid	Limited physiological detail
IDSA/ATS severe-CAP criteria [3]	Identify severe pneumonia/ICU candidates	Major respiratory/shock criteria plus nine minor criteria	Clinically intuitive	Equal weighting of heterogeneous minor criteria
SMART-COP [4]	Predict intensive respiratory or vasopressor support	BP, multilobar disease, albumin, RR, HR, confusion, oxygenation, pH	ICU-oriented outcome	Requires multiple variables
REA-ICU [5]	Predict ICU admission days 1–3	Age, sex, comorbidity, vital signs, imaging and laboratory variables	Directly targets early ICU transfer	More complex than CURB-65
Weighted IDSA/ATS models [9]	Predict ICU admission	Differently weighted severity variables	Allows unequal predictor importance	Requires additional validation
Computerized EHR models [10]	Automated CAP prognosis	Routinely captured physiologic/clinical variables	Scalable and automatable	Dependent on EHR quality and transportability

Table 2. Recommended candidate predictors

Domain	Candidate predictors	Rationale
Demographics	Age, sex	Established CAP risk modifiers
Comorbidity	Heart disease, chronic lung disease, renal disease, liver disease, malignancy, cerebrovascular disease	Represents baseline physiological reserve
Mental status	Confusion/altered consciousness	Included in multiple CAP scores
Respiratory physiology	Respiratory rate, SpO_2, PaO_2, oxygen requirement	Directly reflects respiratory compromise
Cardiovascular physiology	Systolic/diastolic BP, heart rate	Identifies haemodynamic instability
Temperature	Initial temperature	Captures inflammatory/hypothermic response
Renal/metabolic	Urea/BUN, sodium, glucose	Components of established severity systems
Haematology	WBC, haematocrit, platelets	Reflects inflammatory and systemic response
Acid-base	Arterial pH	Important SMART-COP predictor
Imaging	Multilobar involvement, pleural effusion	Reflects anatomical disease burden
Derived values	Shock index, oxygenation ratios	May capture interactions efficiently
Existing scores	PSI, CURB-65, SMART-COP	Useful comparison or secondary predictors

Table 3. Candidate algorithms for CAP ICU prediction

Model	Strength	Limitation	Recommended role
Logistic regression	Interpretable; probabilistic; familiar	Requires appropriate functional-form specification	Mandatory baseline
Penalized logistic regression	Reduces overfitting	Still largely additive unless interactions added	Strong primary comparator
Random forest	Captures nonlinear effects/interactions	Probabilities may require calibration	Secondary ML model
Gradient boosting	High predictive flexibility	Hyperparameter-sensitive; less transparent	Primary nonlinear candidate
Support-vector machine	Handles complex boundaries	Less intuitive probability interpretation	Exploratory
Neural network	Highly flexible	Usually unnecessary for modest tabular clinical datasets	Exploratory only
Simple clinical score	Easy implementation	Lower flexibility	Clinical benchmark

Table 4. Recommended validation strategy

Stage	Recommended method	Purpose
Model development	Training dataset	Estimate model parameters
Hyperparameter selection	Nested cross-validation	Prevent optimistic tuning
Internal validation	Bootstrap or repeated cross-validation	Estimate optimism
Temporal validation	Later patient cohort	Test stability over time
Geographic validation	Different hospital/system	Test transportability
Recalibration	Update intercept/slope when required	Adapt risk estimates
Impact study	Prospective implementation	Determine clinical benefit

Table 5. Minimum model-performance reporting set

Metric	Question answered	Clinical importance
AUROC	Can the model rank high-risk above low-risk patients?	Discrimination
AUPRC	How well does the model identify events when events are uncommon?	Class-imbalance sensitive
Sensitivity	What proportion of future ICU patients are identified?	Critical for screening
Specificity	What proportion of non-ICU patients are correctly classified?	Limits over-triage
PPV	How many high-risk predictions truly deteriorate?	Resource planning
NPV	How reassuring is a low-risk prediction?	Ward/outpatient safety
Calibration intercept	Is risk systematically over/underestimated?	Probability accuracy
Calibration slope	Are predictions too extreme or too conservative?	Overfitting assessment
Brier score	How accurate are probabilities overall?	Combined calibration/discrimination
Decision-curve analysis	Does model-guided triage improve net clinical benefit?	Implementation relevance

Figure 1. Routinely collected clinical feature architecture for early ICU-admission prediction in community-acquired pneumonia

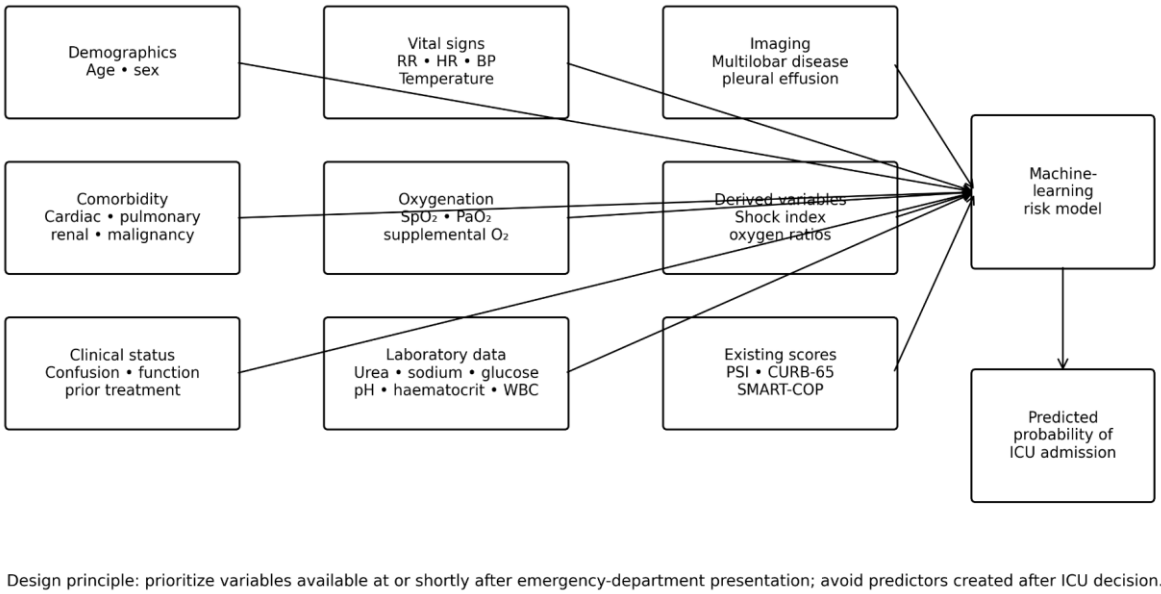

Figure 1. Routinely collected clinical feature architecture for early ICU-admission prediction in community-acquired pneumonia.

Proposed feature architecture for an early machine-learning ICU-admission model. Variables should be restricted to information available at or shortly after emergency-department presentation to avoid information leakage.

Figure 2. Recommended development and validation pipeline for a CAP ICU-admission machine-learning model

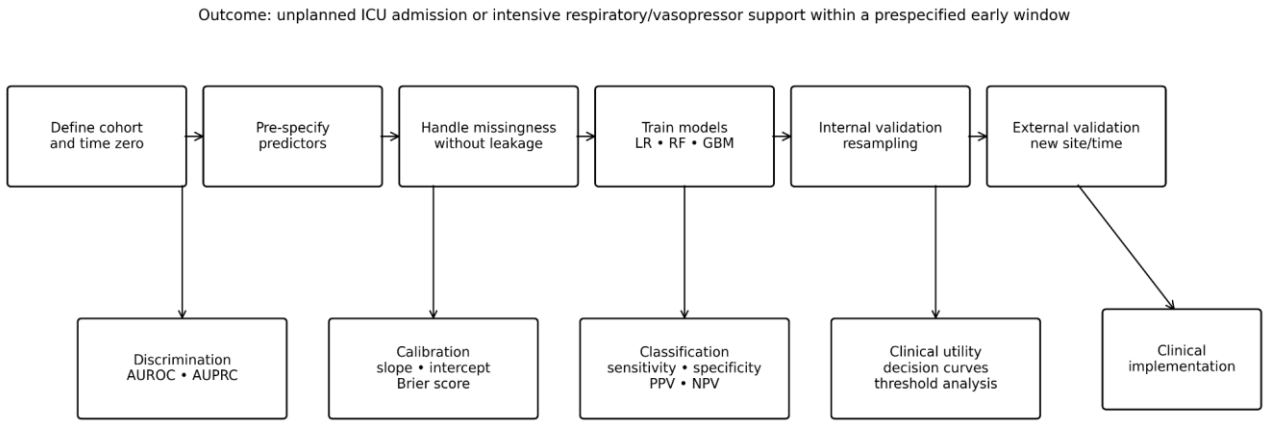

Figure 2. Recommended development and validation pipeline for a CAP ICU-admission machine-learning model.

Recommended analytical pipeline. Data preprocessing, imputation, feature engineering and hyperparameter tuning should occur within resampling procedures to prevent information leakage. External validation should precede clinical implementation.

How to cite: Reeves J, Turner M, Martínez E. (2022). Machine-Learning Prediction of Intensive Care Unit Admission Among Adults Presenting With Community-Acquired Pneumonia Using Routinely Collected Clinical Variables: A Systematic Evidence Synthesis and Model-Development Framework. Journal of Multidisciplinary Research and Integrated Sciences, 2(1).