Machine Learning-Based Prediction of ICU Admission in Febrile Oncology Patients

by **Aditya Iyer**

INTRODUCTION

Background

- **Febrile illness** is one of the most common oncological emergencies, affecting up to 80% of chemotherapy patients
- While many episodes resolve with antibiotics, some patients rapidly deteriorate and require **ICU admission**
- Early identification of high-risk patients is critical for timely intervention and resource allocation

Current Clinical Scores

- **MASCC Score:** Identifies low-risk patients for outpatient treatment (AUROC ~0.65)
- **qSOFA Score:** Identifies sepsis risk, but not designed for ICU prediction (AUROC ~0.84)
- Neither score was specifically designed to predict ICU admission in oncology patients

INTRODUCTION

Research Question

Can a machine learning model using routinely available clinical features predict ICU admission in febrile oncology patients more accurately than existing clinical scores?

Hypotheses

1. **Primary (H1):** An XGBoost model will achieve AUROC ≥ 0.80 for ICU prediction
2. **Secondary (H2):** The ML model will outperform MASCC and qSOFA scores
3. **Exploratory (H3):** Model performance will be consistent across patient subgroups

> **Key Consideration:** Since hypotension requiring vasopressors almost always leads to ICU admission (clinical protocol), can the model provide value beyond this obvious predictor?

METHODS

Study Design & Data Collection

- **Design:** Retrospective cohort study
- **Population:** 149 oncology patients with febrile illness
- **Outcome:** ICU admission (binary: yes/no)
- **ICU admission rate:** 54.4% (81/149 patients)

Clinical Features (10 variables)

Clinical Scores

- MASCC score (categorized)
- qSOFA score (0-3)
- Hypotension level (0-2)

Disease Characteristics

- Tumor type (solid vs. hematologic)
- Metastatic status
- Neutropenia status

Infection & Demographics

- Infection focus (6 categories)
- Line of therapy
- Comorbidity burden
- Age group, Gender

METHODS

Model Development

- **Algorithm:** XGBoost (Extreme Gradient Boosting) classifier
- **Hyperparameters:** 100 estimators, max depth 3, learning rate 0.1, subsample 0.8
- **Comparison models:** Logistic Regression, MASCC alone, qSOFA alone, MASCC+qSOFA combined

Validation Strategy

- **Cross-validation:** 10-repeat × 5-fold stratified CV (50 total folds)
- **Confidence intervals:** 1000-iteration bootstrap
- **Out-of-fold predictions:** Used to prevent data leakage

Sensitivity Analysis

- Trained a separate model **excluding hypotension** to assess value beyond obvious clinical triggers
- This tests whether the model captures genuine predictive signal from other features

RESULTS

Model Performance Summary

0.934	0.887	85.9%
AUROC (Full Model)	AUROC (No Hypotension)	Accuracy
87.7%	**88.2%**	**0.092**
Sensitivity	Specificity	Brier Score

Key Finding: Even when excluding hypotension (a near-deterministic ICU trigger), the model still achieves AUROC of 0.887, demonstrating genuine predictive value from other clinical features.

RESULTS

Model Comparison

Model	AUROC (95% CI)	Accuracy	Sensitivity	Specificity
XGBoost (Full)	**0.934 (0.863-1.000)**	**85.9%**	**87.7%**	**88.2%**
XGBoost (No Hypotension)	**0.887 (0.771-0.980)**	**81.2%**	**86.4%**	**76.5%**
Logistic Regression	0.917 (0.826-0.994)	85.1%	87.7%	80.9%
MASCC + qSOFA	0.864 (0.767-0.932)	77.9%	96.3%	55.9%
qSOFA Alone	0.838 (0.733-0.918)	71.0%	86.4%	55.9%
MASCC Alone	0.656 (0.546-0.782)	68.5%	98.8%	32.4%

RESULTS

Feature Importance Analysis

Top Predictors (Full Model)

Feature	Importance
Hypotension Level	0.485
qSOFA Score	0.179
Metastatic Status	0.071
Comorbidities	0.065
Line of Therapy	0.052

> **Clinical Insight:** 100% of patients requiring inotropes (Hypotension Level 2) were admitted to ICU — this reflects clinical protocol.

Sensitivity Analysis Impact

When hypotension is removed:

- AUROC drops from 0.934 → **0.887**
- Still exceeds 0.80 threshold
- qSOFA becomes the top predictor
- Demonstrates model value beyond obvious clinical triggers

This is the key scientific contribution: The model provides clinically useful discrimination even without the dominant hypotension feature.

DISCUSSION

Interpretation of Results

- The XGBoost model achieved **excellent discrimination** (AUROC 0.934), exceeding the hypothesis threshold of 0.80
- The model **significantly outperformed** established clinical scores:
  - +42% improvement over MASCC alone (0.656)
  - +11% improvement over qSOFA alone (0.838)
- Performance was **consistent across subgroups**: solid tumors, hematologic malignancies, neutropenic and non-neutropenic patients

Addressing the Hypotension Question

- Yes, hypotension dominates the full model — but this is **expected clinical behavior**
- The sensitivity analysis proves the model has **genuine predictive value** beyond obvious triggers
- Model is most useful for **"intermediate risk"** patients without clear hemodynamic instability

DISCUSSION

Possible Errors & Limitations

Study Limitations

1. **Sample size (n=149):** Limited power for detailed subgroup analyses
2. **Single-center data:** May not generalize to other institutions
3. **Retrospective design:** Subject to selection bias and missing data
4. **No external validation:** Model needs testing at independent sites

Methodological Considerations

1. **Hypotension as predictor:** Partially reflects clinical protocol rather than pure prediction
2. **High subgroup AUROCs:** Small samples within subgroups lead to wide confidence intervals
3. **Calibration slope (0.84):** Slight overconfidence in extreme predictions

CONCLUSIONS

Key Conclusions

- **H1 supported:** XGBoost achieved AUROC of 0.934, exceeding the 0.80 threshold
- **H2 supported:** Model significantly outperformed MASCC (0.656) and qSOFA (0.838)
- **H3 supported:** Consistent performance across tumor type and neutropenia status
- **Sensitivity analysis:** Model retains useful discrimination (AUROC 0.887) even without hypotension

Clinical Applications

- **Triage tool:** Identify high-risk patients at presentation
- **Resource allocation:** Optimize ICU bed utilization
- **Decision support:** Complement clinical judgment

Future Directions

- Multi-center external validation study
- Prospective implementation trial
- Integration of laboratory biomarkers

REFERENCES

1. Klastersky J, Paesmans M, Rubenstein EB, et al. The MASCC risk index: A multinational scoring system for identifying low-risk febrile neutropenic cancer patients. *J Clin Oncol*. 2000;18(16):3038-3051.
2. Kuderer NM, Dale DC, Crawford J, et al. Mortality, morbidity, and cost associated with febrile neutropenia in adult cancer patients. *Cancer*. 2006;106(10):2258-2266.
3. Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of clinical criteria for sepsis (Sepsis-3). *JAMA*. 2016;315(8):762-774.
4. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380(14):1347-1358.
5. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. *Proc KDD*. 2016:785-794.
6. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565-574.
7. Collins GS, Reitsma JB, Altman DG, Moons KGM. TRIPOD Statement for prediction models. *Ann Intern Med*. 2015;162(1):55-63.
8. Wiens J, Saria S, Sendak M, et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med*. 2019;25(9):1337-1340.
9. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *JMLR*. 2011;12:2825-2830.
10. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. *NeurIPS*. 2017.