**Supplemental Methods:**

*The Observational Medical Outcomes Partnership (OMOP) Data Model*

OMOP is an open-community and common data model to enable standard analyses of observational databases. In the OMOP common data model, classification vocabularies, such as International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM), Current Procedural Terminology version 4 (CPT4), or Standard Nomenclature of Medicine (SNOMED); are mapped to standard OMOP concepts based on semantic and clinical relationships.(1) Vocabulary classification and mapping of various ontologies to the OMOP standard vocabulary is maintained by OHDSI and publicly available on ATHENA (http://athena.ohdsi.org/), which is a web-based vocabulary repository.(2)

*Definitions of ACLF*

The NACSELD ACLF diagnostic criterion is based on two or more organ failures: shock, West-Haven grade III/IV hepatic encephalopathy, dialysis, and/or mechanical ventilation.(3) The EF-CLIF ACLF diagnostic criterion is based on gradations of organ failures based on laboratory measurements and/or events across six organ systems.(4) As the ratio of arterial partial pressure of oxygen (PaO2) to fraction of inspired oxygen (FiO2) was not always available, we utilized the threshold of partial oxygen saturation (SpO2) to FiO2 ratio of ≤ 214 as an equivalent of PaO2/FiO2 ratio of ≤ 200 as per custom in the CLIF-C-ACLF model.(4)

*Missingness and Imputation*

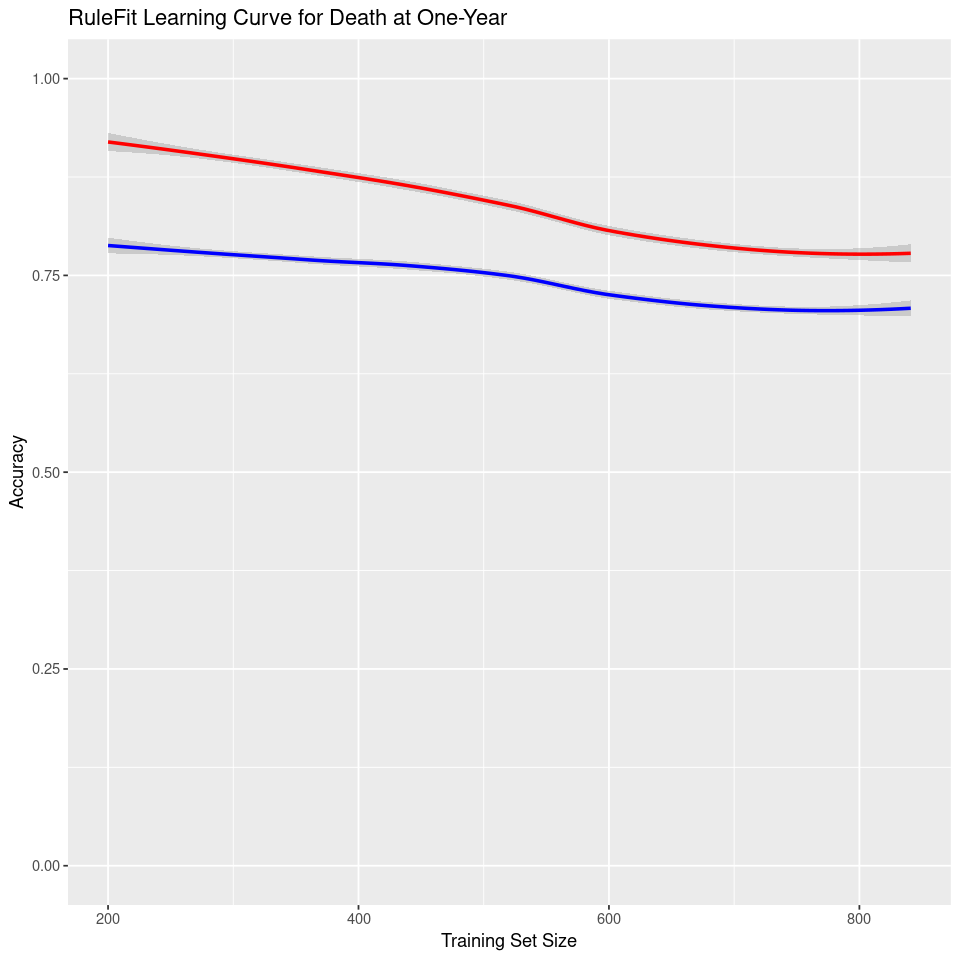
Implausible values, as defined as those greater than three standard deviations from the mean values were removed as per previous data processing in the generation of ML models.(5) Data features and variables with greater than or equal to 25% missingness were excluded from analyses, similar to previous ML analyses of transplant hepatology patients.(5) Data features with less than 25% missingness were imputed with single imputation with chained random forests as implemented in the *missRanger*, version 2.1.3, R package.(6,7)

This imputation method handles both continuous and categorical data features within the distribution of the original data, thereby avoiding outliers and recovers the natural data variability. In previous performance evaluations of imputation methods in a large electronic health record dataset, *missRanger* algorithms produced the most reliable results with the lowest average standard errors. (8–10)

*Confidence Interval Calculations for AUROC Differences*

To estimate point estimates and confidence intervals for AUROC differences between pair-wise comparisons of models (e.g. EAML versus MELDNa), we utilized bootstrapping over 2,000 iterations for each difference as implemented in the *boot*, version 1.3-28, R package.(11–13)

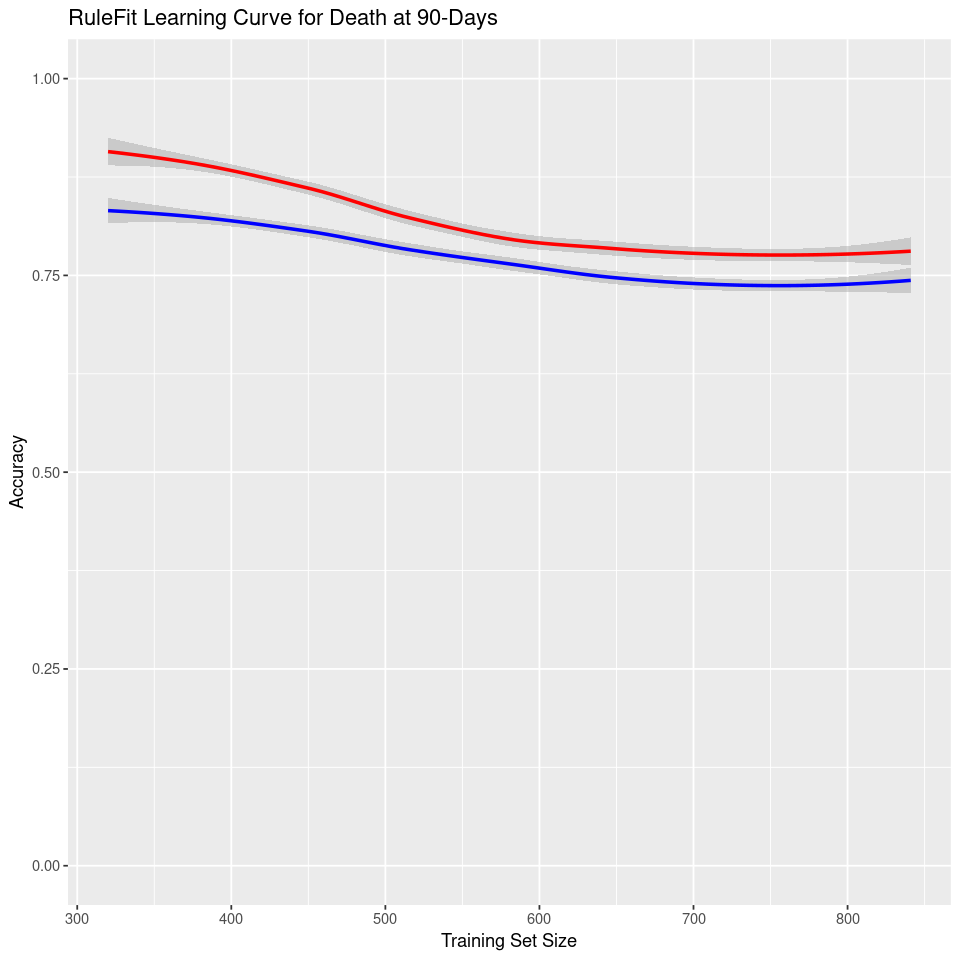
**Supplemental Figure 1 – RuleFit Training and Testing Plot for Outcome of Death at One-Year**



Training Plot

Test Plot

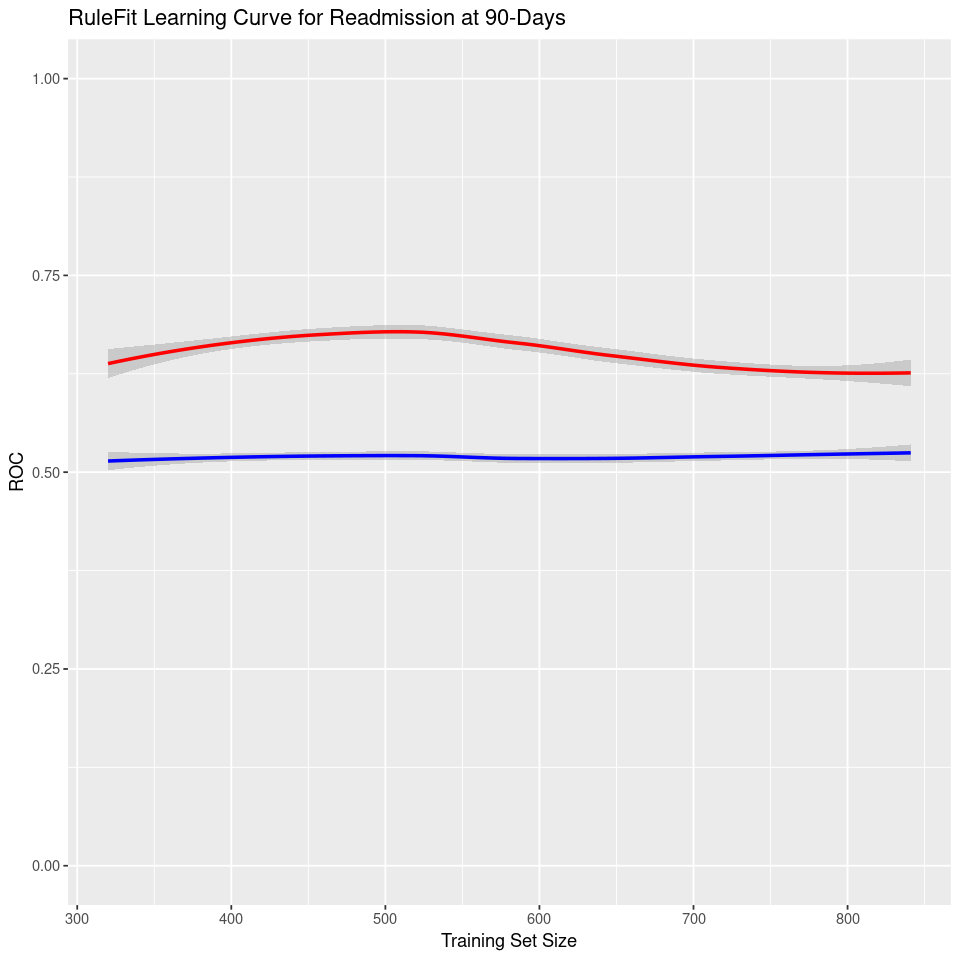
**Supplemental Figure 2 – RuleFit Training and Testing Plot for Outcome of Death at 90-Days**



Training Plot

Test Plot

**Supplemental Figure 3 – RuleFit Training and Testing Plot for Outcome of Readmissions at 90-Days**



Training Plot

Test Plot

**Supplemental Materials References:**

1. documentation:vocabulary:mapping [Observational Health Data Sciences and Informatics] [Internet]. [cited 2021 Apr 15];Available from: https://www.ohdsi.org/web/wiki/doku.php?id=documentation:vocabulary:mapping

2. Pintus R, Yang Y, Rushmeier H. ATHENA. J. Comput. Cult. Herit. 2015;8:1–25.

3. O’Leary JG, Reddy KR, Garcia-Tsao G, Biggins SW, Wong F, Fallon MB, et al. NACSELD acute-on-chronic liver failure (NACSELD-ACLF) score predicts 30-day survival in hospitalized patients with cirrhosis. Hepatology. 2018;67:2367–2374.

4. Jalan R, Saliba F, Pavesi M, Amoros A, Moreau R, Ginès P, et al. Development and validation of a prognostic score to predict mortality in patients with acute-on-chronic liver failure. J. Hepatol. 2014;61:1038–1047.

5. Ivanics T, So D, Claasen MPAW, Wallace D, Patel MS, Gravely A, et al. Machine learning–based mortality prediction models using national liver transplantation registries are feasible but have limited utility across countries. Am. J. Transplant. 2022;

6. Using missRanger [Internet]. [cited 2023 Jan 6];Available from: https://cran.microsoft.com/snapshot/2020-04-16/web/packages/missRanger/vignettes/vignette\_missRanger.html

7. Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;28:112–118.

8. Wong KC-Y, Xiang Y, Yin L, So H-C. Uncovering Clinical Risk Factors and Predicting Severe COVID-19 Cases Using UK Biobank Data: Machine Learning Approach. JMIR Public Health Surveill. 2021;7:e29544.

9. Rios R, Miller RJH, Manral N, Sharir T, Einstein AJ, Fish MB, et al. Handling missing values in machine learning to predict patient-specific risk of adverse cardiac events: Insights from REFINE SPECT registry. Comput Biol Med. 2022;145:105449.

10. Liu D, Oberman HI, Muñoz J, Hoogland J, Debray TPA. Quality control, data cleaning, imputation. arXiv. 2021;

11. Bootstrap Functions (Originally by Angelo Canty for S) [R package boot version 1.3-28.1] [Internet]. 2022 [cited 2023 Jan 6];Available from: https://cran.r-project.org/web/packages/boot/index.html

12. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. Stat. Med. 2000;

13. DiCiccio TJ, Efron B, Hall P, Martin MA, Canty AJ, Davison AC, et al. Bootstrap confidence intervalsCommentCommentCommentCommentRejoinder. Stat Sci. 1996;11:189–228.