Predicting Post-Liver Transplant Outcomes in Patients with Acute-on-Chronic Liver Failure using Expert-Augmented Machine Learning

」
(
1

Jin Ge, MD, MBA⁻; Jean C. Digitale, MPH, KN⁻; Cynthia Fenton, MD⁻; Charles E. McCulloch, PhD⁻; Jennifer
C. Lai, MD, MBA¹; Mark J. Pletcher, MD, MPH², Efstathios D. Gennatas, MBBS, AlCSM, PhD²
¹ Division of C. Lai, MD, MBA ; Mark J. Pletcher, MD, MPH , Erstathios D. Gennatas, MBBS, AlCSM, PhD
¹ Division of Gastroenterology and Hepatology, Department of Medicine, University of Calif
Francisco, San Francisco, CA
² Departmen .
r $\begin{array}{c} 1 \\ 2 \\ 3 \end{array}$

Parancisco, San Francisco, CA
Department of Epidemiology and Biostatistics, University of California – San Francisco, San Francisco,
CA
Division of Hospital Medicine, Department of Medicine, University of California – San

Prancisco, San Francisco, San

² Department of Epidemiolog

² Division of Hospital Medicir
Francisco, CA CA
Division of Hospital Medicine, Department of Medicine, University of California – San Francisco, San
Francisco, CA
Francisco, CA ² Di
Fra
Fina

Financial/Grant Support:

Francisco, CA
Financial/Grant Support:
The authors of this study were supported in part by the KL2TR001870 (National Center for Advancing Francial/Gra
Financial/Gra
The authors o
Translational $\frac{1}{2}$ Translational Sciences, Ge and McCulloch), AASLD Anna S. Lok Advanced/Transplant Hepatology Awar
AHL21-104606 (AASLD Foundation, Ge), American Society of Transplantation LICOP Grant Award CA-
0182782 (American Society of T AHL21-104606 (AASLD Foundation, Ge), American Society of Transplantation LICOP Grant Award CA-
0182782 (American Society of Transplantation, Ge), P30DK026743 (UCSF Liver Center Grant, Ge and Lai),
F31HL156498 (National Hea official views of the National Institutes of Health or any other funding agencies. The funding agencies 1922 (Mathematic Center, 2018) and Blood Institute, Digitale), UL1TR001872 (National Center for
Advancing Translational Sciences, Pletcher), and R01AG059183/K24AG080021 (National Institute on
Aging, Lai). The content is so F31HL15764 (Mathematically 2003) and 2012 and R01AG059183/K24AG080021 (National Institute on
Aging, Lai). The content is solely the responsibility of the authors and does not necessarily represent to
official views of the Aging, Lai). The content is solely the responsibility of the authors and does not necessarily represent
official views of the National Institutes of Health or any other funding agencies. The funding agencie
played no role official views of the National Institutes of Health or any other funding agencies. The funding agencies
played no role in the analysis of the data or the preparation of this manuscript.
Abbreviations:

Abbreviations:

played no role in the analysis of the data or the preparation of this manuscript.
Abbreviations:
ACLF, acute-on-chronic liver failure; APASL ACLF, Asian Pacific Association for the Study of the Liver AC
Research Consective played no role in the analysis of the data or the preparation of this manuscript.
Abbreviations:
Research Consortium; AUROC, area-under-the-receiver-operating characteristic
Data Driven Insights and Innovation: CDS, clinic $\frac{1}{2}$ Research Consortium; AUROC, area-under-the-receiver-operating characteristic curve; CDI2, Center for
Data-Driven Insights and Innovation; CDS, clinical decision support; CI, confidence interval; CORDS, UC
COVID Research Da Data-Driven Insights and Innovation; CDS, clinical decision support; CI, confidence interval; CORDS, UC
COVID Research Data Set; CPT4, Current Procedural Terminology version 4; EAML, Expert-Augmented
Machine Learning; EF-C Data-Driven Insights and Intervals, 2007, and an additional pepperty by conductive intervaly clieby and
COVID Research Data Set; CPT4, Current Procedural Terminology version 4; EAML, Expert-Augmented
Machine Learning; EF-C Machine Learning; EF-CLIF, European Association for the Study of the Liver-Chronic Liver Failure
Consortium; EHR, electronic health record; ESLD, end-stage liver disease; FiO2, fraction of inspired
oxygen; GBM, Gradient Bo Consortium; EHR, electronic health record; ESLD, end-stage liver disease; FiO2, fraction of inspire
oxygen; GBM, Gradient Boosting Machine; GLMNET, Elastic-Net Regularized Generalized Linear I
NOTE: This preprint reports n Consorting; CONSTRIGUTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
NOTE: This preprint reports new research that has not been certified

ICD-10-CM, INTERNATION OF THE TERRORY CHAMPETERS, THAMPETERS, THAMPETERS, THE CONDITIONS.

National COVID Cohort Collaborative; NACSELD, North American Consortium for the Study of End-Stag.

Liver Disease; OMOP, Observatio Mational COVID Cohort Collaborative; NACSELD, North American Consortium for the Study of End
Liver Disease; OMOP, Observational Medical Outcomes Partnership; PaO2, arterial partial pressur
oxygen; RF, Random Forest; SNOMED Liver Disease; OMOP, Observational Medical Outcomes Partnership; PaO2, arterial partial pressure of
oxygen; RF, Random Forest; SNOMED, Standard Nomenclature of Medicine; SpO2, partial oxygen
saturation; TAM, Transplantatio Oxygen; RF, Random Forest; SNOMED, Standard Nomenclature of Medicine; SpO2, partial oxygen
saturation; TAM, Transplantation for ACLF-3 Model; UCH, University of California Health; UCHDW,
University of California Health Dat saturation; TAM, Transplantation for ACLF-3 Model; UCH, University of California Health; UCHDW,
University of California Health Data Warehouse; VHACDW, Veterans Health Administration Corpor
Data Warehouse University of California Health Data Warehouse; VHACDW, Veterans Health Administration Corporate University of California Health Data Warehouse; VHACDW, VEterans Health Administration Corporation
Correspondence:
F13 Democrate Administration Corporation Corporation Corporation Corporation Corporation Corporation Corpor

Correspondence:

Jin Ge, MD, MBA
513 Parnassus Avenue, S-357
San Francisco, CA 94143 an Francisco, CA
513 Parnassus Av
San Francisco, CA
E-mail: jin.ge@uc
Fax: 415-476-065 52 Parmassus Avenue, S-357
San Francisco, CA 94143
E-mail: jin.ge@ucsf.edu
Fax: 415-476-0659
Disclosures E-mail: jin.ge@ucsf.edu
Fax: 415-476-0659
Disclosures:
Disclosures: Fax: 415-476-0659
Fax: 415-476-0659
Disclosures:
Dr. Jin Ge receives resea

Disclosures:

**Disclosures:
Dr. Jin Ge receives
Writing Accistones** |
|
|

Dr. Jin Ge receives research support from Merchands
Writing Assistance:
Author Contributions: \
|
| Writing Assistance:

Author Contributions:

Autho
Author
*Ge: S*tu |
|
|
| Ge: Study concept and design; data extraction; analysis and i
manuscript; critical revision of the manuscript for important
obtained funding; study supervision
Digitale: Analysis and interpretation of data; drafting of man Ge: Study concept and design; data extraction; analysis and interpretation of data; drafting or
manuscript; critical revision of the manuscript for important intellectual content; statistical an
Digitale: Analysis and in

obtained funding; study supervision

Digitale: Analysis and interpretation of data; drafting of manuscript; critical revision of the manuscript

for important intellectual content

Fenton: Data extraction; critical revisio for important intellectual content

Bightare: Analysis and interpretation of data; drafting of manuscript; critical revision of the manuscript
for important intellectual content
Fenton: Data extraction; critical revision of the manuscript for important int Fenton: Data extraction; critical revision of the manuscript for important intellectual content *McCulloch:* Interpretation of data; critical revision of the manuscript for important intellectual content
McCulloch: Interpretation of data; critical revision of the manuscript for important intellectual
 McCulloch: Interpretation of data; critical revision of the manuscript for important intellectual content

Lai: Study concept and design; interpretation of data; critical revision of the manuscript for important
intellectual content; study supervision
Pletcher: Interpretation of data; critical revision of the manuscript for imp

manuscript for important intellectual content Gennatas: Study concept and design; analysis and interpretation of data; critical revision of the

Data Acknowledgement:

Gennatas: Study concept and design; analysis and interpretation of data; critical revision of the
manuscript for important intellectual content
Data Acknowledgement:
The authors thank the Center for Data-driven Insights an manuscript for important international contents
Data Acknowledgement:
The authors thank the Center for Data-driven
California Health (UCH),(1) for its analytical an |
|
|
| California Health (UCH),(1) for its analytical and technical support related to the use of the UC Health
Data-Warehouse (UCHDW) and related data-assets, including the UC COVID Research Data Set (CORD
Word Count:
3,592 (wit Data Warehouse (UCHDW) and related data assets, including the UC COVID Research Data Set (CORDS).

$\frac{1}{2}$ Word Count:

Paradonal Count:
1999 (Without references) and related data assets, including the UC COVID Research Data Set (CORDS). $\frac{3}{5}$

Abstract:

Background: Liver transplantation (LT) is a treatment for acute-on-chronic liver failure (ACLT) but up to
40% mortality post-LT has been reported. Existing post-LT models in ACLF have been limited by small
samples. In this samples. In this study, we developed a novel Expert-Augmented Machine Learning (EAML) model to predict post-LT outcomes. samples. In this study, we developed a novel Expertiting, and the Learning (EAML) model to
predict post-LT outcomes.
<u>Methods:</u> We identified ACLF patients in the University of California Health Data Warehouse (UCHDV

Methods: We identified A
We used EAML, which uses $\frac{1}{1}$ Methods: We identified ACLF patients in the University of California Health Data Warehouse (UCHDW).
We used EAML, which uses the RuleFit machine learning (ML) algorithm to extract rules from decision-
trees that are then e trees that are then evaluated by human experts, to predict post-LT outcomes. We compared
EAML/RuleFit's performances versus other popular models.
Results: We identified 1,384 ACLF patients. For death at one-year: areas-und

Easults: We identified 1,384 ACLF patients. For death at one
Characteristic curve (AUROCs) were 0.707 (Confidence Interv $\frac{1}{2}$ Results: We identified 1,384 ACLF patients. For death at one-year: areas-under-the-receiver-operating
characteristic curve (AUROCs) were 0.707 (Confidence Interval [CI] 0.625-0.793) for EAML and 0.719 (C
0.640-0.800) for R 0.640-0.800) for RuleFit. For death at 90-days: AUROCs were 0.678 (CI 0.581-0.776) for EAML and 0.707
(CI 0.615-0.800) for RuleFit. In pairwise comparisons, EAML/RuleFit models outperformed crosssectional models. Divergences between experts and ML in rankings revealed biases and artifacts in the
underlying data. sectional moderlying data.
Sectional models. Dignificant discrepancies between experts and ML in the Fit outperformed cross-sectional models. Significant discrepancies between

underlying data.
Conclusions: EAM
experts and ML c $rac{6}{1}$ experts and ML occurred in rankings of biomarkers used in clinical practice. EAML may serve as a
method for ML-guided hypothesis generation in further ACLF research. experts and ML occurred in rankings of a community assume incompetence is and MA₂ contract.

The method for ML-guided hypothesis generation in further ACLF research. method for ML-guided hypothesis generation in further ACLF research.

Introduction:

Acute-on-chronic liver failure (ACLF) is commonly defined as acute decompensation of endstage liver disease (ESLD) with extra-hepatic organ failure and is associated with high short-term
mortality (2–7) Liver transplantation (LT) is a well-established treatment for patients with ACLF who are refractory to supportive care and treatment for the underlying precipitant. Due to critical illness,
however, LT is estimated to be feasible in only 25% of ACLF patients (8) Moreover, there have beer
conflicting post-LT ou however, LT is estimated to be feasible in only 25% of ACLF patients. (8) Moreover, there have been conflicting post-LT outcomes reported for ACLF patients with some sub-populations having up 40%
three-month mortality.(9,10) There is an unmet need for tools to predict post-LT outcomes for ACLF
patients in the pre-LT sett confliction three-month mortality.(9,10) There is an unmet need for tools to predict post-LT outcomes for ACL
patients in the pre-LT setting (and without intra-operative or post-LT data) to ensure utility.(11,12)
Multiple

patients in the pre-LT setting (and without intra-operative or post-LT data) to ensure utility.(11,12)
Multiple international research consortia, such as the North American Consortium for the Stu
of End-Stage Liver Disease patients in the pre-LT setting (and without in the periodic or post-LT data) to entitle utility.(11)(11)
Multiple international research consortia, such as the North American Consortium for the S
of End-Stage Liver Disease Multiple international research international research consortium (EF-CLIF),(4) and the Asian Pacific Association for the Study of the Liver ACLF
Iure Consortium (EF-CLIF),(4) and the Asian Pacific Association for the Stud Liver Failure Consortium (EF-CLIF),(4) and the Asian Pacific Association for the Study of the Liver ACLF Research Consortium (APASL ACLF);(13) have developed scoring systems to predict pre-LT outcomes. None of these models, however, specifically evaluates for post-LT outcomes. One of the few models that specifically evaluate for post-LT outcomes is the Transplantation for ACLF-3 Model (TAM) score, which was trained on a cohort of 76 patients with EF-CLIF grade-3 (severe) ACLF at a single French
center and validated in 76 patients in four other centers.(14) Despite its potential utility, the TAM
model has not been st center and validated in 76 patients in four other centers.(14) Despite its potential utility, the TAM
model has not been studied in non-European settings outside of limited cohorts.
In addition, one of the major barriers t

that ACLF is a heterogeneous and dynamic clinical syndrome, as evidenced by diverging definitions in different geographies.(3,4,6,8,13) Existing prediction models do not utilize vast numbers of data In a sevidenced by diverging definitions in
In addition, original syndrome, as evidenced by diverging definitions in
In addition, one of the majority of the major barriers of data
In available in electronic health records different geographies. (3,4,6,8,13) Existing prediction models do not utilize vast numbers of data
features available in electronic health records (EHRs) to better define dynamic clinical trajectories see
in patients with diffeatures available in electronic health records (EHRs) to better define dynamic clinical trajectories.
In patients with ACLF. Our group had previously demonstrated an informatics approach to extract
data that yielded a in patients with ACLF. Our group had previously demonstrated an informatics approach to extract EHR
data that yielded a median of 454 features per admission to more accurately represent ACLF patients' data that yielded a median of 454 features per admission to more accurately represent ACLF patients'
data that yielded a median of 454 features per admission to more accurately represent ACLF patients' data that yielded a median of 454 features per admission to more accurately represent ACLF patients'

clinical courses.(15) Machine Learning (ML) is well-suited for analyzing such data, but can be misleading
when taken out of context of biological or clinical mechanisms.(16,17)
Expert-Augmented Machine Learning (EAML) is a Expert-Augmented Machine Learning (EAML) is an emerging te
limitation of ML by extracting rules from decision-tree ML models for h
has two potential benefits: 1) To create combined models that incorpo Expert Augmented Machine Learning (EAML) is an emerging terminal and Carl interest and
n of ML by extracting rules from decision-tree ML models for human expert feedback. EAM
potential benefits: 1) To create combined model has two potential benefits: 1) To create combined models that incorporate the best of human and ML
knowledge, and 2) To evaluate for differences between humans and ML. These differences could knowledge, and 2) To evaluate for differences between humans and ML. These differences could
represent human biases (e.g., experts ignoring important variables identified by ML) or artifacts in the
underlying data (e.g., e knowledge, and 2) To evaluate for anti-time and human and ML. These anti-times could
represent human biases (e.g., experts ignoring important variables identified by ML) or artifacts in
underlying data (e.g., experts are i underlying data (e.g., experts are identifying the important variables but there is over-representation of
other clinical characteristics in this population not seen elsewhere).
In this study, we utilized a novel multi-cen

underlying data (e.g., experts are identifying the important variables but there is the inportant variables but
other clinical characteristics in this population not seen elsewhere).
In this study, we utilized a novel mult In this study, we utilized a novel multi-center EHR database,
Data Warehouse (UCHDW), to construct an EAML model to predict p
ACLF. In the study, we understand a study, we have the UCHDW), to construct an EAML model to predict post-LT outcomes in patients with $\frac{1}{2}$ PACLF.
Data Warehouse (UCHDW), to construct an EAML model to predict post-LT outcomes in patients with patients with
A Methods:

Metho
Metho
<u>The Ui</u> Methods:

The University of California Health Data Warehouse (UCHDW)

 $\frac{1}{2}$ University of California Health (UCH) Medical Centers (Davis, Irvine, Los Angeles, San Diego, and San
Francisco) and managed by the Center for Data-Driven Insights and Innovation (CDI2).(1) UCHDW hold
data on 6.2+ million Francisco) and managed by the Center for Data-Driven Insights and Innovation (CDI2).(1) UCHDW ho
data on 6.2+ million well-characterized patients seen at UCH since 2012. All data in UCHDW are
harmonized in the Observationa Francisco)
Francisco data on 6.2+ million well-characterized patients seen at UCH since 2012. All data in UCHDW are
harmonized in the Observational Medical Outcomes Partnership (OMOP) common data model, version
5.3.1.(18) harmonized in the Observational Medical Outcomes Partnership (OMOP) common data model, v
5.3.1.(18) All data elements in UCHDW are de-identified prior to the receipt by end-users with n
clinical notes or imaging. UCHDW has harmonized in the Observational Medical Outcomes Partnership (OMOP) common data model, version
5.3.1.(18) All data elements in UCHDW are de-identified prior to the receipt by end-users with no
clinical notes or imaging. UC 5.3.1.(18) All data elements in UCHDW are de-identified prior to the receipt by end-users with no content in the previously between UCH health systems and amongst individual providers within each health system. (19) For all
analyses, we utilized UCHDW, versioned as of September 22, 2022 and accessed on October 20, 2022 between UCH health systems and amongst individual providers within each health system.(19) For all
analyses, we utilized UCHDW, versioned as of September 22, 2022 and accessed on October 20, 2022. analyses, we utilized UCHDW, versioned as of September 22, 2022 and accessed on October 20, 2022.

Study Population

 $\frac{1}{\sqrt{2}}$ We isolated all adults (>= 18 years) who underwent an orthotopic liver transplantation
procedure, as defined by the OMOP concept identifiers 2109321 (CPT4 code) or 4067458 (SNOMED code), based on the ATHENA OMOP vocabulary dictionary, (20) in UCHDW between January 1, 2013 through December 31, 2021. We included patients who underwent multi-organ (such as simultaneous
liver-kidney transplant) and re-transplant procedures as they may have been in ACLF prior to transplant
Consistent with prior liver-kidney transplant) and re-transplant procedures as they may have been in ACLF prior to transplant.
Consistent with prior informatics approaches for detecting ACLF admissions, we excluded all patients
who underwent tr Liver-kidney consistent with prior informatics approaches for detecting ACLF admissions, we excluded all patients

who underwent transplant within 48 hours of admission as they were likely admitted electively.

included pa Consideration with prior information approaches for detecting their distributed and patients and patients
who underwent transplant within 48 hours of admission as they were likely admitted electively.(15) \
included patien included patients who had evidence of ACLF prior to the time of LT through a previously published
informatics-driven approach.(15) Briefly, this involves identifying any patient who meets ACLF
diagnostic criteria based on informatics-driven approach.(15) Briefly, this involves identifying any patient who meets ACLF
diagnostic criteria based on the NACSELD or EF-CLIF definitions prior to LT. We did not use the APA
ACLF diagnostic criteria du diagnostic criteria based on the NACSELD or EF-CLIF definitions prior to LT. We did not use the
ACLF diagnostic criteria due to bacterial infection being the most common precipitant of ACLF i
patients with cirrhosis in the diagnostic criteria due to bacterial infection being the most common precipitant of ACLF in patients with cirrhosis in the United States.(21,22) patients with cirrhosis in the United States.(21,22)
Measurements

patients
Measurements
We extracted all structured clinical informa We extra

We extr

Baseline charac

identity of the l We extracted all structured clinical information associated with the admission of interest.
Baseline characteristics included age, sex, race/ethnicity, height, weight, body mass index, and censored
identity of the UCH faci Baseline characteristy of the UCH facility (defined as "UC-1," "UC-2," and "UC-3"). Laboratory measurements, liver
disease etiologies, complications of cirrhosis, comorbid medical conditions, dialysis state, ventilation
pa disease etiologies, complications of cirrhosis, comorbid medical conditions, dialysis state, ventilation
parameters, and vasopressor administration were extracted based on previously defined OMOP conc
identifiers (20,23) disease etiologies, complications of climinary permembentum community, analysis state, ventilation
parameters, and vasopressor administration were extracted based on previously defined OMOP condidentifiers.
(20,23)
As pati

parameters. (20,23)
As patients may have different lengths of stay before LT, we focused only on data values from
the day of admission and the day before LT. We dropped measurements from other time points from $\frac{1}{2}$ As patient:
the day of admissive of admission and the day before LT. We dropped measurements from other time points from

Solution of admission and the day before LT. We dropped measurements from other time points from the day of admission and the day before LT. We dropped measurements from other time points from

between admission and transplant based on the differences between data features between admission
and day before LT.(24–26) All intra-operative data values and values after transplant were excluded
from our analyses as our and day before LT.(24–26) All intra-operative data values and values after transplant were excluded
from our analyses as our intent was to develop a predictive model utilizing only pre-transplant data.
Missing data feature and any server LT.(24–22) All intra-operative data values and values after transplant data.

From our analyses as our intent was to develop a predictive model utilizing only pre-transplant data.

Missing data features and Missing data features and variables underwent single imputation with chained random forests, which
has been shown to produce low errors and good performances in previous studies utilizing EHR
data.(27–29) Mas been shown to produce low errors and good performances in previous studies utilizing EHR
data (27–29)
data (27–29) has been shown to produce low entities in good performances in previous studies utilizing EHR
data.(27–29)
Outcomes

Outcomes

outcomes
1
1 The p $rac{1}{\sqrt{2}}$ The primary outcome was all-cause mortality at one-year after LT. The secondary outcomes
included: 1. All-cause readmissions within 90-days, and 2. All-cause mortality within 90-days after LT.
Death was ascertained based o included: 1. All-causes the California Death Registry.
Included: 1. All-causes reading Seath Registry (1)
Model Development and Expert-Augmented Machine Learning (EAML)

Model Development and Expert-Augmented Machine Learning (EAML)

Death was assessed on synchronized as a state with the California Death Registry.(2)

Death Development and Expert-Augmented Machine Learning (EAML)

The sample of ACLF patients isolated from UCHDW was split by random samp $\frac{1}{\sqrt{2}}$ The sample of ACLF patients isolated from UCHDW was split by random sampling into training,
validation, and test sets in a 60:20:20 ratio.(30–32) The training set was used to fit the model, the
validation set was utilized validation set was utilized to tune hyperparameters, and the test set was held-out for independent
testing. RuleFit training and testing plots are shown in Supplemental Figures 1, 2, and 3 for our three
outcomes. We then u ML model for each of our primary and secondary outcomes of interest (total of three models). (33) outcomes. We then utilized EAML, as implemented in the rtemis R package, version 0.91, to train one
ML model for each of our primary and secondary outcomes of interest (total of three models).(33)
rtemis is a platform for out comes of interest (total of three models). (33)
Then utilizations is a platform for advanced ML research and applications, which incorporates several algorithms,
Including EAML. (34) ME MERENET CRIMITY ON PRIMARY AND SECONDARY OF MICHAEL (TOTAL OF MICE MODELS).
The micrositions is a platform for advanced ML research and applications, which incorporates several algorith
Including EAML.(34)
As described

rtemis including EAML.

As described above, EAML is an ensemble ML algorithm that incorporates human knowledge by

converting high-dimensional training data into Likert-scale questions. (33) EAML first trains a predictive including Education
As described
converting high-dime As an engineer above, the state above, the state above, the ML algorithm that incorporates human constants and the Muslim and the Muslim of the Muslim of the Muslim that include the Muslim of the Muslim corporates human co converting high-dimensional training data into Likert-scale questions.(33) EAML first trains a predictive

model using the RuleFit algorithm,(35) which is a combination of a Gradient Boosting Machine (1211),
decision-tree model (trained on the data to generate rules), and a Least Absolute Shrinkage and
Selection Operator (LASSO Selection Operator (LASSO) model (used to select rules generated by the GBM model).(35) The I
model training outputs include the detailed rules, model coefficients (represents the change in r
associated with the rule), and Selection operator (LASSO) model constraints generated by the Cambridge processes as
model training outputs include the detailed rules, model coefficients (represents the change in respons
associated with the rule), and em associated with the rule), and empirical risk (rating of the rule importance by the machine).
Utilizing the rules selected by RuleFit, we then created an online survey on the Qualtrics

Utilizing the rules selected by RuleFit, we then created an online survey on the Qual
platform (example question in Figure 1) that was sent to 15 hepatologists throughout the w
conduct clinical care and research in ACLF re Untertaing the rules selected by RuleFit, we then created an entire survey on the Quality of the creation in Figure 1) that was sent to 15 hepatologists throughout the world is clinical care and research in ACLF recruited conduct clinical care and research in ACLF recruited from a convenience sample. These experts were
asked to rate rules on a 5-point Likert-scale based on perceived associations with the outcomes of
interest. We calculated Exhault clinical care and research in ACLF research in ACLF remembers cample. These experts and as
asked to rate rules on a 5-point Likert-scale based on perceived associations with the outcomes of
interest. We calculated interest. We calculated expert rankings based on the averages of these ratings. We then took the differences in rankings between the experts and those generated by the RuleFit model to calculate penalties. These penalties differences in rankings between the experts and those generated by the RuleFit model to calculate
penalties. These penalties were then incorporated into the RuleFit models by eliminating the top
quartile of the most discre penalties. These penalties were then incorporated into the RuleFit models by eliminating the top
quartile of the most discrepant rules (highest fourths of absolute rank differences between RuleFit
expert rankings) to creat penalties. These penalties are then incorporated into the RuleFit models of the RuleFit expert rankings) to create the EAML models for each of the three outcomes. (33) expert rankings) to create the EAML models for each of the three outcomes.(33)
Statistical Analyses and Model Performance Evaluation

Statistical Analyses and Model Performance Evaluation

expert ranks.

Statistical Analyses and Model Performance Evaluation

Clinical characteristics and laboratory data were summarized by medians $\frac{4}{1}$ CLINICAL CONTINUOUS VARIACTERISTY, AND MELT CRIMINATIONS, MEDIANS AND MELT PARTICLE.
CLINICAL CONTINUOUS VARIACTERISTY OF DETAILS AND INTERFERIENCY OF CATABORATIONS.
ISONS between the training, validation, and test sets we ranges (IQR) for continuous variables or numbers and percentages (%) for categorical variables.
Comparisons between the training, validation, and test sets were performed using chi-square and
Kruskal-Wallis tests where app

models through area-under-the-receiver-operating characteristic curve (AUROC), which has been used We evaluated the performances
models through area-under-the-receive
previously to evaluate ML models in trai We evaluated the performances of
Show to evaluate ML models in transplant hepatology. (36–39) To compare the performances of
IL and RuleFit models versus cross-sectional models (MELDNa, NACELD-ACLF, CLIF-C-ACLF, TAM) models through area-under-the-receiver-operating characteristic curve (Netterly), which has been used
previously to evaluate ML models in transplant hepatology.(36–39) To compare the performances of
the EAML and RuleFit mo the EAML and RuleFit models versus cross-sectional models (MELDNa, NACELD-ACLF, CLIF-C-ACLF, TAM) the EAML and RuleFit models versus cross-sectional models (MELDNa, NACELD-ACLF, CLIF-C-ACLF, TAM)

and other ML algorithms (Random Forest [RF], Gen., and Elastic Resingmenties Centrumes Entrom Forest
Model [GLMNET]), we calculated AUROC differences between each pair of models (e.g., AUROC
differences between EAML and NA differences between EAML and NACSELD) and their confidence intervals using bootstrapping wit
iterations per pair-wise comparison.(40,41) We calculated MELDNa, NACSELD-ACLF, CLIF-C-ACLI
TAM scores per previously published l differences between ERML and NACSELD, and MAC standards MELDNa, NACSELD-ACLF, CLIF-C-ACLF, and
TAM scores per previously published literature. (3,4,14,42) We used rtemis implementations of RF,
GBM, and GLMNET to generate c TAM scores per previously published literature.(3,4,14,42) We used rtemis implementations of RF,
GBM, and GLMNET to generate comparison ML models.
All data queries, extractions, and transformations of OMOP concept identifi

conducted using the Microsoft Azure implementations of Spark, version 2.12. All statistical analyses All data queries, extractions, and transformation
conducted using the Microsoft Azure implementations of
were performed utilizing Spark-R, version 4.1.3 "One Pus ed using the Microsoft Azure implementations of Spark, version 2.12. All statistical analyses
rformed utilizing Spark-R, version 4.1.3 "One Push-Up" (R Core Team, Vienna, Austria), and R
s previously noted and documented i conducted and documented using the Microsoft Azure in the Microsoft Azure in the Superior Azustria), and R
packages previously noted and documented in the supplemental materials. (43) Two-sided p-values
<0.05 were consider performed until α performed until α performed utilizing the Push-Up" (R Core Team), Theore, Austria), and R
packages previously noted and documented in the supplemental materials. (43) Two-sided p-values
<0.05 were c end onto the supplemental scheme of UCHDW data for this study
authorized by the Institutional Review Board at the University of California, San Francisco under #20
32717 for model generation and #22-37555 for expert input. authorized by the Institutional Review Board at the University of California, San Francisco under #20-
32717 for model generation and #22-37555 for expert input. authorized by the Institutional Review Board at the University of California, San Francisco under the University
32717 for model generation and #22-37555 for expert input.
Results:

Results:

32717 For Mesults:
32717 A total of 1,384 patients with ACLF were identified from the A \overline{a} er 31, 2021. Of the 1,384: 611 (44.1%) were women, 576 (41.6%) Hispanic, 472 (34.1%) non-
Nite, 138 (10.0%) Asian, 60 (4.3%) Black, and 122 (8.8%) of Unknown/Other race/ethnicity.
Nition of patients by University of Califo Distribution of patients by University of California sites were 410 (29.6%) at UC-1, 173 (12.5%) at UC-2
and 801 (57.9%) at UC-3.
The patients were randomly divided based on a 60:20:20 ratio with 841 patients in the traini

and 801 (57.9%) at UC-3.

The patients were randomly divided based on a 60:20:20 ratio with 841 patients in the training

set, 255 in the validation set, and 288 in the test set. The three sets were broadly similar across The patients were
set, 255 in the validation :
demographic and clinical In the validation set, and 288 in the test set. The three sets were broadly similar across multiple
aphic and clinical characteristics (e.g. age, race/ethnicity, liver disease etiologies, comorbid
ns, and distribution betw demographic and clinical characteristics (e.g. age, race/ethnicity, liver disease etiologies, comorbid conditions, and distribution between UCH facilities). Of note, the median MELDNa scores at admission conditions, and distribution between UCH facilities). Of note, the median MELDNa scores at admission

were 34 (interquartile range [IQR] 29-39), 34 (IQR 30-38), and 34 (IQR 30-38) for the training, validation,
and test sets, respectively. Detailed patient characteristics at time of admission are reported in Table 1.

Primary and Secondary Outcomes

and test sets sets sets sets sets sets and test sets of the test
in the total sample of 1,384 patients: 149 (10.8%) met the primary outcome of death at one-
in the total sample of 1,384 patients: 149 (10.8%) met the primar $\frac{1}{2}$ year, 97 (7%) met the secondary outcome of death at 90-days, and 621 (44.9%) met the secondary
outcome of readmission within 90-days. Distributions and prevalence of the primary and secondary
outcomes were similar between outcomes were similar between the training, validation, and test sets; and are reported in Table 2.

RuleFit and Expert Augmentation

RuleFit and Expert Augmentation

RuleFit and Expert Augmentation
After identification and division of the ACLF patient population as above, we then applied t $\frac{1}{2}$ After identification and division of the ACLF patient population as above, we then applied the
RuleFit algorithm. RuleFit generated 20 rules for the primary outcome of death at one-year (Table 4), 18
rules for the secondar Rules for the secondary outcome of death within 90-days (Table 5), and 6 rules for the secondary
outcome of readmission within 90-days (Table 6). The rules generated by RuleFit for each of the
outcomes were then distribute outcome of readmission within 90-days (Table 6). The rules generated by RuleFit for each of the
outcomes were then distributed to 15 hepatologists throughout the world who conduct clinical c
research in ACLF who rated the outcomes were then distributed to 15 hepatologists throughout the world who conduct clinical c
research in ACLF who rated the importance of rules based on a 5-point Likert scale. The aggregat
physician rankings along with research in ACLF who rated the importance of rules based on a 5-point Likert scale. The aggregated
physician rankings along with rank differences between RuleFit and experts are also reported in Tabl
4, 5, and 6 for each o physician rankings along with rank differences between RuleFit and experts are also reported in Tables physician rankings along manufalm and along the three manufalm differences between RuleFit and
4, 5, and 6 for each of the three outcomes. Of note, the greatest discrepancies between RuleFit and
human experts occurred in t 4, 4, 4, and 6 for each of the three states of the three outcomes in the states of biomarkers more commonly utilized in clinical practice, sure
as age and MELDNa score.
The rand Fit and MELDNa score. human experts of the rankings of an angle in the rankings of an angle and MELDNa score.
EAML Model Performance Versus Cross-Sectional and Other ML Models as age and MELDNa score.
1998 - EAML Model Performance Versus Cross-Sectional and Other ML Models
1997 - For the primary outcome of death at one-year: AUROCs were 0.707 (Confidence Interval [CI]

 $\frac{l}{2}$ EAML Model Performance Versus Cross-Sectional and Other ML Models
For the primary outcome of death at one-year: AUROCs were 0.
0.625-0.793) for the EAML and 0.719 (CI 0.640-0.800) for the RuleFit mo F o.625-0.793) for the EAML and 0.719 (CI 0.640-0.800) for the RuleFit models. For the secondary 0.625-0.793) for the EAML and 0.719 (CI 0.640-0.800) for the RuleFit models. For the secondary

outcome of death at one-year and in Figure 3 for the secondary outcome of death at 90-days
primary outcome of death at one-year and in Figure 3 for the secondary outcome of death at 90-days Pairwise AUROC differ
Pairwise AUROC differ
primary outcome of death at c
In general, for the outcomes o primary outcome of death at one-year and in Figure 3 for the secondary outcome of death at 90-days.
In general, for the outcomes of death at one-year and death at 90-days, AUROC differences between
EAML and RuleFit models In general, for the outcomes of death at one-year and death at 90-days, AUROC differences between
EAML and RuleFit models showed that RuleFit outperformed EAML but this was not significant: Δ (Rule
- EAML) was 0.013 (CI EAML and RuleFit models showed that RuleFit outperformed EAML but this was not significant: Δ (RuleFit
- EAML) was 0.013 (CI -0.027-0.052) for death at one-year and Δ (RuleFit - EAML) was 0.030 (CI -0.100-
0.071) for EAML) was 0.013 (CI -0.027-0.052) for death at one-year and Δ(RuleFit - EAML) was 0.030 (CI -0.100-
0.071) for death at 90-days. Moreover, AUROC differences between the EAML/RuleFit models and
GBM, and those between the E - The D.071) for death at 90-days. Moreover, AUROC differences between the EAML/RuleFit models and
GBM, and those between the EAML/RuleFit models and GLMNET were also not significant. In contras
for the outcomes of death a GBM, and those between the EAML/RuleFit models and GLMNET were also not significant. In contr
for the outcomes of death at one-year and death at 90-days, the EAML/RuleFit models consistently
outperformed cross-sectional mo For the outcomes of death at one-year and death at 90-days, the EAML/RuleFit models consistently
outperformed cross-sectional models (MELDNa, NACSELD, CLIF-ACLF, and TAM).
For the secondary outcome of readmission at 90-day

outperformed cross-sectional models (MELDNa, NACSELD, CLIF-ACLF, and TAM).
For the secondary outcome of readmission at 90-days: AUROCs were 0.557 (CI 0.493-0.623)
the EAML and 0.564 (CI 0.498-0.629) for the RuleFit models. For the secondary outcome of readmission at 90-days: AUROCs were 0.5!
the EAML and 0.564 (Cl 0.498-0.629) for the RuleFit models. Pairwise AUROC diff
confidence intervals are reported in detail in Figure 4 for the secondar For the secondary outcome of readministration at 90-days: AUROC differences and
IL and 0.564 (CI 0.498-0.629) for the RuleFit models. Pairwise AUROC differences and
note intervals are reported in detail in Figure 4 for the confidence intervals are reported in detail in Figure 4 for the secondary outcome of readmission at 90days. In general, the EAML and RuleFit models did not show significant differences in predictive abilitie
versus each other and versus other ML models. Moreover, while EAML/RuleFit showed significant
differences in AUROC v versus each other and versus other ML models. Moreover, while EAML/RuleFit showed significant
differences in AUROC versus some of the cross-sectional models (MELDNa, NACSELD, and CLIF-ACL
overall predictive abilities of al differences in AUROC versus some of the cross-sectional models (MELDNa, NACSELD, and CLIF-ACLF) differences in AUROC versus some of the cross-some of the cross-sectional models (MELDNA, NACSELD, and CLIF-ACLF) –
overall predictive abilities of all models evaluated were poor.
Discussion:

Discussion:

This study is one of the first to explicitly combine human expert knowledge with ML to create an l
i interpretable ML model for a clinical problem within transplantation. In this study, we generated two
models (EAML, which incorporates human expert content, and RuleFit, which does not incorporate interpretable ML models (EAML, which incorporates human expert content, and RuleFit, which does not incorporate
human input) for each of the three outcomes (post-transplant mortality at one-year, post-transplant
duman inpu human input) for each of the three outcomes (post-transplant mortality at one-year, post-transplant human input) for each of the three outcomes (post-transplant mortality at one-year, post-transplant

mortality at 90-days, and readmission after transplant at 90-days). Our ML models (EAML and RuleFit)
significantly outperformed existing cross-sectional models with mean AUROCs clustering around 0.700
for the outcomes of p

significant of post-transplant mortality at one-year and mortality at 90-day.
In our pairwise comparisons of models utilizing AUROC differences, we found that while there
were no significant differences between EAML and Ru for the outcomes of post-transplant mortality at one-year and mortality at 90-0 any.
In our pairwise comparisons of models utilizing AUROC differences, we fou
were no significant differences between EAML and RuleFit, and b In our pairwise comparisons of models unitality of the same there, we compare more displained into the significant differences between EAML and RuleFit, and between EAML/RuleFit and other
ML algorithms, such as GBM and GLM popular ML algorithms, such as GBM and GLMNET. Moreover, while these were not statistically
significant, but the EAML models consistently had lower AUROCs versus the RuleFit models. The r
likely explanation in this situati popular ML algorithms, such a CLM and CLMNET. MOTOLS, while these is the RuleFit models. The likely explanation in this situation is due to residual artifacts in study population as the training, validation, and test sets likely explanation in this situation is due to residual artifacts in study population as the training,
validation, and test sets are all derived from the same database. In this circumstance, the process of
incorporating ex Interty experiment in this situation is due to residual artifacts in study, population in the training,
validation, and test sets are all derived from the same database. In this circumstance, the proce
incorporating expert incorporating expert input with EAML is not expected to improve the performance of the model since
the test set have similar distributions of demographic and clinical characteristics as the training sets.
The purpose of EA

The test set have similar distributions of demographic and clinical characteristics as the training sets.
The purpose of EAML, therefore, in this situation is to reveal key insights from the discrepancie
between human expe The purpose of EAML, therefore, in this situation is to reveal key insights from the discrepand
between human expert and ML rankings of rules. These reveal residual biases and areas for future
research. For instance, in th The purpose of EAML, therefore, in this situation is to reveal residual biases and areas for future
The purpose of EAML insights from the BAML model for post-transplant mortality at one-year, rule #18 (MELDNa
The of transp research. For instance, in the EAML model for post-transplant mortality at one-year, rule #18 (MEL
at the time of transplant being > 32.47) was ranked as the most important by experts, but only tent
most important by RuleF at the time of transplant being > 32.47) was ranked as the most important by experts, but only tenth most important by RuleFit. This difference in rank by nine positions indicated that experts may have
biases favoring of a well-known and established clinical scoring system – whereas the RuleFit algorithm
determined it to biases favoring of a well-known and established clinical scoring system – whereas the RuleFit algorithm
determined it to be not as important. In general, across the three outcomes, ACLF experts were more
likely to over-ran age, and white blood cell count. In contrast, RuleFit was more likely to elevate the importance of dikely to over-rank the importance of commonly used physiologic and clinical makers, such as MELDNa,
age, and white blood cell count. In contrast, RuleFit was more likely to elevate the importance of
electrolytes and hemat age, and white blood cell count. In contrast, RuleFit was more likely to elevate the importance of
electrolytes and hematological parameters, such as ionized calcium, sodium, and lactate dehydrogenase
as important data fea age, and then extracted count in counter, matrix matrix and the important incompetitive or
electrolytes and hematological parameters, such as ionized calcium, sodium, and lactate dehydrog
as important data features. These as important data features. These results imply additional avenues for further research in the clinical care of patients with ACLF (Figure 5). Moreover, this study demonstrates that EAML's use may not be care of patients with ACLF (Figure 5). Moreover, this study demonstrates that EAML's use may not be

minted to predictive modeling, but also as an artificial intelligence-guide method for hypothesis
generation.
database, and the first to derive insights on transplant patients. UCHDW is based on the OMOP g
Fina
database, ar
common da Finally, the state second study to fully study study study study study study study is and the first to derive insights on transplant patients. UCHDW is based on the OMOP and a model, which is also utilized in several other common data model, which is also utilized in several other big data multi-center EHR databases, such as
the National COVID Cohort Collaborative (N3C), (44) All of Us, (45) and the Veterans Health the National COVID Cohort Collaborative (N3C),(44) All of Us,(45) and the Veterans Health
Administration Corporate Data Warehouse (VHACDW).(46,47) While patients with ACLF and LT patients
have been extensively studied in t Administration Corporate Data Warehouse (VHACDW).(46,47) While patients with ACLF and
have been extensively studied in the VHACDW, the VHACDW is not broadly representative
general population. While patients with cirrhosis have been extensively studied in the VHACDW, the VHACDW is not broadly representative of the
general population. While patients with cirrhosis have been studied in N3C, the current purviews of N3C
limits research topics to have been emergined, While patients with cirrhosis have been studied in N3C, the current purviews
limits research topics to those related to the novel coronavirus pandemic. It is our hope that our
analytical approach of ut general population. Then patients with circuit interesting and a studied in N3C, the current purviews of the curre
limits research topics to those related to the novel coronavirus pandemic. It is our hope that our
analytic analytical approach of utilizing OMOP will become more common as increasing numbers of institutions of interval
have or are in the processing of harmonizing their EHR data to the OMOP common data model.
There are several l

analytical approach of harmonizing their EHR data to the OMOP common data model.
There are several limitations to this study due to its retrospective nature, its use of a novel
database, and its analytical processes. First There are several limitations to this study due to its retrospective nature, its use of a not
database, and its analytical processes. First, there is selection bias – we had only included patient
ACLF who had successfully There are selection bias – we had only included patients
There is selection bias – we had only included patients
There are study due to its retrospective nature, its use of a normal
Therefore is retrospective nature, its u ACLF who had successfully undergone LT, and not those who were listed to undergo LT but then
subsequently died or recovered and not those who were never listed for LT. This means that the
patients with ACLF who ultimately ACLF who had successfully analogene LT, and the three three there is that the subsequently died or recovered and not those who were never listed for LT. This means that the patients with ACLF who ultimately made it LT suff patients with ACLF who ultimately made it LT suffered from a survivorship bias and are unlikely to be
representative of the entire ACLF population. While it is feasible to pull data for all patients with ACLI
who did not u representative of the entire ACLF population. While it is feasible to pull data for all patients with ACLF
who did not undergo LT, we have no visibility into whether these patients were listed for LT and we
would not be ab who did not undergo LT, we have no visibility into whether these patients were listed for LT and we
would not be able to evaluate for the post-transplant outcomes of interest.
Second, we do not have intraoperative or donor

would not be able to evaluate for the post-transplant outcomes of interest.
Second, we do not have intraoperative or donor derived data for the patients in our cohort.
This, however, is not necessarily a significant limita Second, we do not have intraoperative or donor derived data for the
This, however, is not necessarily a significant limitation in our study as our ir
pre-LT model to predict post-LT outcomes. The ultimate clinical decision Wever, is not necessarily a significant limitation in our study as our intended goal was to derivated to predict post-LT outcomes. The ultimate clinical decision that this model would help
o whether to proceed to LT for an The LT model to predict post-LT outcomes. The ultimate clinical decision that this model would help
with is to whether to proceed to LT for an ACLF patient. Third, to take advantage of the high-
with is to whether to proce with is to whether to proceed to LT for an ACLF patient. Third, to take advantage of the high-
with is to whether to proceed to LT for an ACLF patient. Third, to take advantage of the highdimensional nature of UCHDW, we only sourced data from three transplant centers within UCH. In
addition, all three UCH facilities included are in the state of California. While this population is
demographically diverse, C demographically diverse, California has some of the highest MELDNa scores at the time of trans
The models and their results, therefore, may not be generalizable to other settings. In addition
etiologies may be variable acr The models and their results, therefore, may not be generalizable to other settings. In addition, as AC
etiologies may be variable across geographies, our models and conclusions may not be generalizable t
populations outsi The models and their results, the setting, may not to generalizable to estinalize to such a same in a sumplemen
etiologies may be variable across geographies, our models and conclusions may not be generalizable to
populati populations outside the United States. These geographic-based differences may be a contributor to why
the TAM model based on French ACLF patients performed poorly in our populations. External validation
should be undertake the TAM model based on French ACLF patients performed poorly in our populations. External validation
should be undertaken for these model prior to their potential deployment in clinical practice.
Finally, the analysis code

should be undertaken for these model prior to their potential deployment in clinical practice.
Finally, the analysis codes utilized to derive the data from UCHDW were written for this specific
(UCHDW) implementation of the should be undertaken for these model prior of their potential depth yielding annual practice.
Finally, the analysis codes utilized to derive the data from UCHDW were written for the
(UCHDW) implementation of the OMOP commo Finally, the analysis codes annual to derive the data from UCHDW were untilected this specific
V) implementation of the OMOP common data model. While OMOP is a common data model
ws for generalization of analyses across dif (Unch) that allows for generalization of analyses across different datasets, there may be minor variations and
differences in data structures, semantics, and coding. The OMOP-based extraction methods and
algorithms for the differences in data structures, semantics, and coding. The OMOP-based extraction methods and
algorithms for these analyses have not been tested on other OMOP-based data sources – further
research is required to evaluate fo differences in data structures, semantics, and coding. The CMOP-based data sources – further
algorithms for these analyses have not been tested on other OMOP-based data sources – further
research is required to evaluate fo

research is required to evaluate for true "out-of-the-box" interoperability.
Despite these limitations, this study represents "proof of concept" for several key concep
developments for health services research in transplan research in transplantation: 1. Use of hurto-boxed by-
developments for health services research in transplantation: 1. Use of hur
ML modeling, 2. Generation of multiple ML models that outperforms tradit Despite these intrinsicting time critical process processes.

Despite these formulations in the study represents in the study represents the support of conceptual in

Deling, 2. Generation of multiple ML models that outpe ML modeling, 2. Generation of multiple ML models that outperforms traditional cross-sectional models
for predicting post-transplant outcomes in ACLF, and 3. Utilizing of a novel data source and common
data model in transpl this study could be refined and evaluated in an iterative manner in clinical decision support (CDS) systems to actively guide clinical decision-making. In such a CDS-based implementation, prospective this study could be refined and evaluated in an iterative manner in clinical decision support (CDS)
systems to actively guide clinical decision-making. In such a CDS-based implementation, prospective
surveillance of outcom this study could be refined and evaluation in an iterative manner in clinical decision positions and systems to actively guide clinical decision-making. In such a CDS-based implementation, prospect surveillance of outcomes surveillance of outcomes would then allow for active feedback to further improve these models.
Surveillance of outcomes would then allow for active feedback to further improve these models. surveillance of outcomes would then allow for active feedback to further improve these models.

Table 1 – Baseline Clinical and Demographic Characteristics of the Training, Validation, and Test Set

Table 2 – Outcomes of the Training, Validation, and Test Set Populations

Table 3 – RuleFit and Expert Rankings for the Primary Outcome of Mortality at One-Year

Differences are defined between values on the day prior to transplant versus those on the day of admission

Table 4 – RuleFit and Expert Rankings for the Secondary Outcome of Mortality at 90-Days

Differences are defined between values on the day prior to transplant versus those on the day of admission

Table 5 – RuleFit and Expert Rankings for the Secondary Outcome of Readmissions at 90-Days

Figure Legends:

Figure 1 - Example Survey Question Utilized to Obtain Expert Input

Figure 2 – AUROC Differences and Confidence Intervals for EAML/RuleFit versus Other Models for the
Outcome of Death at One-Year
Figure 3 – AUROC Differences and Confidence Intervals for EAML/RuleFit versus Other Models for $\frac{1}{2}$

Figure 3 – AUROC Differences and Confidence Intervals for EAML/RuleFit versus Other Models for the
Dutcome of Death at 90-Days
Figure 4 – AUROC Differences and Confidence Intervals for EAML/RuleFit versus Other Models for

Figure 3 – AUROC Differences a
Outcome of Death at 90-Days
Figure 4 – AUROC Differences a
Outcome of Readmissions at 90 $\frac{1}{2}$ Figure 4 – AUROC Differences and Confidence Intervals for EAML/RuleFit versus Other Models for the
Dutcome of Readmissions at 90-Days
Figure 5 – Disagreements Between Experts and RuleFit May Reflect Biases, Artifacts, and Eigure 4 – AUROC Differences
Outcome of Readmissions at 9
Figure 5 – Disagreements Betv
Further Research $\frac{1}{2}$

Figure 5 – Disagreements Between Experts and RuleFit May Reflect Biases, Artifacts, and Areas for Further Research
Figure 5 – Disagreements Between Experts and RuleFit May Reflect Biases, Artifacts, and Areas for
Further R **Figure 5 – Disagreements Between Ex.**
Further Research
Versions at 90-00-Days $\frac{1}{2}$ Further Research
Further Research $\frac{1}{\sqrt{2}}$

- References:
1. Center for Data-driven Insights and Innovations (CDI2) | UCOP [Internet]. [cited 2021 Feb
- 17];Available from: https://www.ucop.edu/uc-health/functions/center-for-data-driven-in:

and-innovations-cdi2.html

2. Bajaj JS, O'Leary JG, Reddy KR, Wong F, Biggins SW, Patton H, et al. Survival in infection-re

acute-on Bajaj JS, O'Leary JG, Reddy
acute-on-chronic liver failu
256.
O'Leary JG, Reddy KR, Garc
chronic liver failure (NACSE
cirrhosis. Hepatology. 2018
Jalan R, Saliba F, Pavesi M,
- cirrhosis. Hepatology. 2018;67:2367-2374. 2. Bajaj JS, O'Leary JG, Reddy KR, Garcia-Tsao G, Biggins SW, Wong F, Fallon MB, et al. NACSELD acute-on-
2. Bajaj JG, Reddy KR, Garcia-Tsao G, Biggins SW, Wong F, Fallon MB, et al. NACSELD acute-on-
2. Battonic liver fail
- 256.

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. 20 256.

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. cirrhosis. Hepatology. 2018;67:2367–2374.
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 failu Jalan R, Saliba F, Pavesi M, Amoros A, More
prognostic score to predict mortality in pati
2014;61:1038–1047.
Gustot T, Moreau R. Acute-on-chronic liver
cirrhosis. J. Hepatol. 2018;69:1384–1393.
Hernaez R, Solà E, **Moreau R** 1. January 2. January 2.
- cirrhosis. J. Hepatol. 2018;69:1384-1393.
-
- prognostic score to prognostic score to provide a distinct more to provide to the compensation of
cirrhosis. J. Hepatol. 2018;69:1384–1393.
Hernaez R, Solà E, **Moreau R, Ginès P.** Acute-on-chronic liver failure: an update. Gustot T, Moreau R. J
cirrhosis. J. Hepatol. 2
Hernaez R, Solà E, Mo
2017;66:541–553.
Moreau R, Jalan R, Gi
distinct syndrome tha
Gastroenterology. 20 cirrhosis. J. Hepatol. 2018;69:1384–1393.

6. Hernaez R, Solà E, **Moreau R, Ginès P.** Acute-on-chronic liver failure: an update. Gut.

2017;66:541–553.

7. Moreau R, Jalan R, Gines P, Pavesi M, Angeli P, Cordoba J, et al. ministry of protonally of the names.
Hernaez R, Solà E, Moreau R, Ginès P. Act
2017;66:541–553.
Moreau R, Jalan R, Gines P, Pavesi M, Ang
distinct syndrome that develops in patient
Gastroenterology. 2013;144:1426–37, 143
S Fremaez R, Sola E, Moreau R, Sines P. Acute-on-chronic liver failure: an update. Gut.

2017;66:541–553.

Moreau R, Jalan R, Gines P, Pavesi M, Angeli P, Cordoba J, et al. Acute-on-chronic live

distinct syndrome that devel Moreau R, Jalan R,
distinct syndrome
Gastroenterology
Sarin SK, Choudhuı
Nat. Rev. Gastroen
Levesque E, **Winte**
liver failure on 90distinct syndrome that develops in patients with acute decompensation of cirrhosis.

Gastroenterology. 2013;144:1426–37, 1437.e1.

8. Sarin SK, Choudhury A. Acute-on-chronic liver failure: terminology, mechanisms and manag
-
-
- Gastroenterology. 2013;144:1426–37, 1437.e1.
Sarin SK, Choudhury A. Acute-on-chronic liver failure: terminology, mechanisms and
Nat. Rev. Gastroenterol. Hepatol. 2016;13:131–149.
Levesque E, Winter A, Noorah Z, Daurès J-P, Sarin SK, Choudhury A. Acute-on-chronic liver fa
Nat. Rev. Gastroenterol. Hepatol. 2016;13:131-
Levesque E, **Winter A, Noorah Z,** Daurès J-P, Lar
liver failure on 90-day mortality following a first
Umgelter A, Lange K, Kor Nat. Rev. Gastroenterol. Hepatol. 2016;13:131–149.

9. Levesque E, Winter A, Noorah Z, Daurès J-P, Landais P, Feray C, et al. Impact of acute-on-chronic

liver failure on 90-day mortality following a first liver transplant Levesque E, **Winter A, Noorah Z,** Daurès J-P, Landais
liver failure on 90-day mortality following a first liver
Umgelter A, Lange K, Kornberg A, Büchler P, Friess H,
in critically ill cirrhotic patients with multi-organ fa 9. Levesque E, Winter A, Nooran Z, Daures J-1, Editable 1, Feray C, et al. Impact of acute on emonic
liver failure on 90-day mortality following a first liver transplantation. Liver Int. 2017;37:684–693.
10. Umgelter A, La Umgelter A, Lange K, Kornberg A, Büchler P, Friess H, Schmid RM. Orthotopic liver transplantation
in critically ill cirrhotic patients with multi-organ failure: a single-center experience. Transplant.
Proc. 2011;43:3762–37 11. Umgelter A, Hernaldy, J. Lange A, Marin Lange A, Lange A, Lange A, Lange Conter A, Lange-Center experience. Transplant.

Proc. 2011;43:3762–3768.

11. Bajaj JS, Verna EC. What role should ACLF play in liver transplant
- Proc. 2011;43:3762–3768.
Bajaj JS, Verna EC. What role should ACLF play in liver transplant prioritization? survey of us-bas
transplant providers. Liver Transpl. 2020;
Wu T, Sundaram V. Transplantation for Acute-on-Chronic
-
- Proc. 2011;43:3762–3768.
Bajaj JS, Verna EC. What role should ACLF play in liver transplant prioritization? survey of us-based
transplant providers. Liver Transpl. 2020;
Wu T, Sundaram V. Transplantation for Acute-on-Chron 12. Wu T, Sundaram V. Transplantation for Acute-on-Chronic Liver Failure. Clin Liver Dis (Hoboken).

12. Wu T, Sundaram V. Transplantation for Acute-on-Chronic Liver Failure. Clin Liver Dis (Hoboken).

13. Sarin SK, Kedari Wu T, Sundaram V. Transplantation for Accords, 2019;14:152–155.
Sarin SK, Kedarisetty CK, Abbas Z, Amarap
failure: consensus recommendations of the
(APASL) 2014. Hepatol Int. 2014;8:453–47
Artzner T, Michard B, Weiss E, Ba
- 2019;14:152–155.

13. Sarin SK, Kedarisetty CK, Abbas Z, Amarapurkar D, Bihari C, Chan AC, et al. Acute-on-chronic liver

failure: consensus recommendations of the Asian Pacific Association for the Study of the Liver

(APA Sarin SK, Kedariset
Failure: consensus
failure: consensus
(APASL) 2014. Hep
Artzner T, Michard
critically ill cirrhotic
2020;20:2437–244
Ge J, Najafi N, Zhac 13. failure: consensus recommendations of the Asian Pacific Association for the Study of the Liver

13. (APASL) 2014. Hepatol Int. 2014;8:453–471.

14. Artzner T, Michard B, Weiss E, Barbier L, Noorah Z, Merle J-C, et al. (APASL) 2014. Hepatol Int. 2014;8:453–471.
Artzner T, Michard B, Weiss E, Barbier L, Noorah Z, Merle J-C, et al. Liver transplantation for
critically ill cirrhotic patients: Stratifying utility based on pretransplant facto Artzner T, Michard B, Weiss E, Barbier L, Noc
critically ill cirrhotic patients: Stratifying utili!
2020;20:2437–2448.
Ge J, Najafi N, Zhao W, Somsouk M, Fang M,
Updated Acute-On-Chronic Liver Failure Prog
Data. Hepatol. C
- 14. Artically ill cirrhotic patients: Stratifying utility based on pretransplant factors. Am. J. Transp
2020;20:2437–2448.
15. Ge J, Najafi N, Zhao W, Somsouk M, Fang M, Lai JC. A Methodology to Generate Longitudina
Update 2020;20:2437–2448.
Ge J, Najafi N, Zhao W, Somsouk M, Fang M, Lai JC. A Methodology to Generate Longitudinally
Updated Acute-On-Chronic Liver Failure Prognostication Scores From Electronic Health Record
Data. Hepatol. Comm 2020;20:24:25
Ge J, Najafi N, Zhao V
Updated Acute-On-C
Data. Hepatol. Comm
Danziger J, Zimolzak A
analysis of electronic
Hu C, Anjur V, Saboo
and death using mach 15. Ge J, Najafi N, Zhao W, Somsouk M, Fang M, Lai JC. A Methodology to Generate Longitudinally

Updated Acute-On-Chronic Liver Failure Prognostication Scores From Electronic Health Record

Data. Hepatol. Commun. 2021;5:10
- Data. Hepatol. Commun. 2021;5:1069–1080.
Danziger J, Zimolzak AJ. Residual confounding lurking in big data: A source of error. In: Seconda
analysis of electronic health records. Cham: Springer International Publishing; 201 Danziger J, Zimolzak AJ. Residual confounding
analysis of electronic health records. Cham: S
Hu C, Anjur V, Saboo K, Reddy KR, O'Leary J, T
and death using machine learning in cirrhosis
OMOP Common Data Model – OHDSI [Inte
- analysis of electronic health records. Cham: Springer International Publishing; 2016. p. 71–78.
17. Hu C, Anjur V, Saboo K, Reddy KR, O'Leary J, Tandon P, et al. Low predictability of readmissions
and death using machine l analysis of electronic health records. Chain is the computation of the computations and death using machine learning in cirrhosis. Am. J. Gastroenterol. 2021;116:336–346.
OMOP Common Data Model – OHDSI [Internet]. [cited 2
- and death using machine learning in cirrhosis. Am. J. Gastroenterol. 2021;116:336–346.
18. OMOP Common Data Model OHDSI [Internet]. [cited 2021 Feb 17];Available from:
https://www.ohdsi.org/data-standardization/the-commo and defining machine learning in comparison of the community of the death of the death of the death of the sol
OMOP Common Data Model – OHDSI [Internet]. [cited 2021 Feb 17];Available from:
https://www.ohdsi.org/data-stand https://www.ohdsi.org/data-standardization/the-common-data-model/
https://www.ohdsi.org/data-standardization/the-common-data-model/
https://www.ohdsi.org/data-standardization/the-common-data-model/ https://www.ohdsi.org/data-standardization/the-common-data-model/

- 19. Peterson Taylor Controlling Controlling Controlling Controlling Controlling Controlling Controlling Adventure Controlling And SCS Care.

2021;44:908-914.

20. Pintus R, Yang Y, Rushmeier H. ATHENA. J. Comput. Cult. Her
-
- For type 2 diabetes and the caliform of the 2021;44:908-914.
Fintus R, Yang Y, Rushmeier H. ATHENA. J. Comput. Cult. Herit. 2015;8:1-25.
Singh H, Pai CG. Defining acute-on-chronic liver failure: East, West or Middle ground
-
- Pintus R, Yang Y, R
Singh H, Pai CG. De
Hepatol. 2015;7:25
Zaccherini G, Weis
principles of treatn
Ge J, Pletcher MJ, I
Chronic Liver Disea 21. Singh H, Pai CG. Defining acute-on-chronic liver failure: East, West or Middle
Hepatol. 2015;7:2571–2577.
22. **Zaccherini G, Weiss E,** Moreau R. Acute-on-chronic liver failure: Definitions, principles of treatment. JHE 22. **Zaccherini G, Weiss E,** Moreau R. Acute-on-chronic liver failure: Definitions, pathophysiology a
principles of treatment. JHEP Rep. 2021;3:100176.
23. Ge J, Pletcher MJ, Lai JC, N3C Consortium. Outcomes of SARS-CoV-2 Zaccherini G, Weiss E, Morea
principles of treatment. JHEP
Ge J, Pletcher MJ, Lai JC, N3C
Chronic Liver Disease and Cir
Gastroenterology. 2021;161:
Wu J, Roy J, Stewart WF. Prec
comparison of machine learn 22. Zaccherini G, weiss E, Moreau R. Acute-on-chronic liver failure: Definitions, pathophysiology and
principles of treatment. JHEP Rep. 2021;3:100176.
23. Ge J, Pletcher MJ, Lai JC, N3C Consortium. Outcomes of SARS-CoV-2 Formula Consortium. Outcome

Ge J, Pletcher MJ, Lai JC, N3C Consortium. Outcome

Chronic Liver Disease and Cirrhosis: A National COV

Gastroenterology. 2021;161:1487–1501.e5.

Wu J, Roy J, Stewart WF. Prediction modeling u
-
- Chronic Liver Disease and Cirrhosis: A National COVID Cohort Collaborative Study.

Gastroenterology. 2021;161:1487–1501.e5.

24. Wu J, Roy J, Stewart WF. Prediction modeling using EHR data: challenges, strategies, and a

c Chronic Liver Disease and Constrainers and Constroenterology. 2021;161:1487–1501.e5.

Wu J, Roy J, Stewart WF. Prediction modeling using EHR data: challenges, strategie

comparison of machine learning approaches. Med. Care Factor Christmas, Prediction modelin
Wu J, Roy J, Stewart WF. Prediction modelin
comparison of machine learning approaches
Perotte A, Ranganath R, Hirsch JS, Blei D, Elh
progression using heterogeneous electronic
Med. Info
- 24. Wurst, 2017, 2014 and he learning approaches. Med. Care. 2010;48:S106-13.
25. Perotte A, Ranganath R, Hirsch JS, Blei D, Elhadad N. Risk prediction for chronic kidney dise
25. Perotte A, Ranganath R, Hirsch JS, Blei D, Perotte A, Ranganath R, Hirsch JS, Blei D, Elhadad N. Risk prediction for chro
progression using heterogeneous electronic health record data and time ser
Med. Inform. Assoc. 2015;22:872–880.
Singh A, Nadkarni G, Gottesman progression using heterogeneous electronic health record data and time series analysis. J. Am.

Med. Inform. Assoc. 2015;22:872–880.

26. Singh A, Nadkarni G, Gottesman O, Ellis SB, Bottinger EP, Guttag JV. Incorporating t
- progression using the anti-matematic means and time series analysis. The Med. Inform. Assoc. 2015;22:872–880.
Singh A, Nadkarni G, Gottesman O, Ellis SB, Bottinger EP, Guttag JV. Incorporating temporal EHI
data in predicti Med. Information of the Unit Singh A, Nadkarni G, Gottesman O, Ellis
data in predictive models for risk stratif
2015;53:220–228.
Wong KC-Y, Xiang Y, Yin L, So H-C. Unco
Cases Using UK Biobank Data: Machine
2021;7:e29544.
R 26. Singh A, Nadiation of the Statification of renal function deterioration. J. Biomed. Inform.
2015;53:220–228.
27. Wong KC-Y, Xiang Y, Yin L, So H-C. Uncovering Clinical Risk Factors and Predicting Severe COVID-
2021;7:e 2015;53:220–228.
Wong KC-Y, Xiang Y, Yin L, So H-C. Uncovering Clinical Risk Factors and Predicting Severe COVID-1:
Cases Using UK Biobank Data: Machine Learning Approach. JMIR Public Health Surveill.
2021;7:e29544.
Rios R Wong KC-Y, Xiang
Cases Using UK Bio
2021;7:e29544.
Rios R, Miller RJH,
machine learning t
SPECT registry. Cor
Liu D, Oberman HI,
- 27. Wong KC-Y, Xiang Y, Yin L, So H-C. Uncorrecting Control River Conses Using UK Biobank Data: Machine Learning Approach. JMIR Public Health Surveill.
2021;7:e29544.
28. Rios R, Miller RJH, Manral N, Sharir T, Einstein AJ 2021;7:e29544.
2021;7:e29544.
Rios R, Miller RJH, Manral N, Sharir T, Einstein AJ, Fish MB, et al. Handling missing value
machine learning to predict patient-specific risk of adverse cardiac events: Insights fron
SPECT reg 2021;
Rios R, Miller RJ
machine learnin
SPECT registry. (
Liu D, Oberman
arXiv. 2021;
Razavian N, Maji
-
- 22. Rios R, Miller R, Miller, Miller R, Miller, Specific risk of adverse cardiac events: Insights from RE
28. Liu D, Oberman HI, Muñoz J, Hoogland J, Debray TPA. Quality control, data cleaning, imputa
28. Liu D, Oberman HI machine learning to predict patient-specific risk of adverse cardiac events: Insights from REFINE
SPECT registry. Comput Biol Med. 2022;145:105449.
Liu D, Oberman HI, Muñoz J, Hoogland J, Debray TPA. Quality control, data Liu D, Oberman HI, Muñoz J, Hoogland J, Debray TPA
arXiv. 2021;
Razavian N, Major VJ, Sudarshan M, Burk-Rafel J, Stel
time prediction model for favorable outcomes in hos
2020;3:130.
Ayala Solares JR, Diletta Raimondi FE, Z 29. Particular N, Major VJ, Sudarshan M, Burk-Rafel J, Stella P, Randhawa H, et al. A validated, real-
2020;3:130.
2020;3:130. Ayala Solares JR, Diletta Raimondi FE, Zhu Y, Rahimian F, Canoy D, Tran J, et al. Deep learning Razavian N, I
time predict
2020;3:130.
Ayala Solare
electronic he
Inform. 2020
- 2020;3:130.
Ayala Solares JR, Diletta Raimondi FE, Zhu Y, Rahimian F, Canoy D, Tran J, et al. Deep learning for
electronic health records: A comparative review of multiple deep neural architectures. J. Biomed.
Inform. 2020 Ayala Solare:
electronic he
Inform. 2020
Juhn Y, Liu H
based clinica
Gennatas ED
augmented r Experiment Controller and Markon and The Universe Present and Theorem 2020;101:103337.

32. Juhn Y, Liu H. Artificial intelligence approaches using natural language processing to advance EHR

33. Gennatas ED, Friedman JH,
- based clinical research. J. Allergy Clin. Immunol. 2020;145:463-469. Inform. 2020;101:103337.
Juhn Y, Liu H. Artificial intelligence approaches using natural language processing to advance EHR-
based clinical research. J. Allergy Clin. Immunol. 2020;145:463–469.
Gennatas ED, Friedman JH, Un Inform. 2021
Juhn Y, Liu H. Artificial inte
based clinical research. J. A
Gennatas ED, Friedman JH,
augmented machine learni
rtemis ML [Internet]. [cited
Friedman JH, Popescu BE. I
Lau L, Kankanige Y, Rubinst
-
-
-
- 313. Algorithms Predict Graft Failure After Liver Transplantation. Transplantation. 2017;101:e125 Gennatas ED, Friedman JH, Ungar LH, Pirracchio R, Eaton E, Reichmann LG, et al. Expert-
augmented machine learning. Proc. Natl. Acad. Sci. USA. 2020;117:4571–4577.
rtemis ML [Internet]. [cited 2023 Jan 2];Available from: h rtemis ML [Internet]. [cited 2023 Jan 2];Available from: https://rtemis.lambdamd.org/
Friedman JH, Popescu BE. Predictive learning via rule ensembles. Ann Appl Stat. 2008;2
Lau L, Kankanige Y, Rubinstein B, Jones R, Christ 35. Friedman JH, Popescu BE. Predictive learning via rule ensembles. Ann Appl Stat. 2008;2
36. Lau L, Kankanige Y, Rubinstein B, Jones R, Christophi C, Muralidharan V, et al. Machine-
Algorithms Predict Graft Failure After 36. Lau L, Kankanige Y, Rubinstein B, Jones R, Christophi C, Muralidharan V, et al. Machine-Learning
Algorithms Predict Graft Failure After Liver Transplantation. Transplantation. 2017;101:e125–
e132.
Ferrarese A, Sartori Algorithms Predict Graft Failure After Liver Transplantation. Transplantation. 2017;101:e125–

132. Ferrarese A, Sartori G, Orrù G, Frigo AC, Pelizzaro F, Burra P, et al. Machine learning in liver

138. Spann A, Yasodhara
- 37. Ferrarese A, Sartori G, Orrù G, Frigo AC, Pelizzaro F, Burra P, et al. Machine learning in liver transplantation: a tool for some unsolved questions? Transpl Int. 2021;34:398–411.
38. Spann A, Yasodhara A, Kang J, Watt e132.
Ferrarese A, Sartori G, Orrù G, Frigo AC, Pelizzaro F, Burra P, et al. Machine learning in liver
transplantation: a tool for some unsolved questions? Transpl Int. 2021;34:398–411.
Spann A, Yasodhara A, Kang J, Watt K
- 17. Ferrares A, Spann A, Yasodhara A, Kang J, Watt K, Wang B, Goldenberg A, et al. Applying machine learn

liver disease and transplantation: A comprehensive review. Hepatology. 2020;71:1093-110

liver disease and transpla transplantation: a tool for some unit. The problems with problems paints of spann A, Yasodhara A, Kang J, Watt K, Wang B, Goldenberg A, et al. Applying machine liver disease and transplantation: A comprehensive review. Hep Iiver disease and transplantation: A comprehensive review. Hepatology. 2020;71:1093-1105. liver disease and transplantation: A comprehensive review. Hepatology. 2020;71:1093–1105.

- 39. In the controll of the mortality prediction models using national liver transplantation registries are feasible but have

limited utility across countries. Am. J. Transplant. 2022;

39. DiCiccio TJ, Efron B, Hall P, Ma
-
-
- limited utility across countries. Am. J. Transplant. 2022;
DiCiccio TJ, Efron B, Hall P, Martin MA, Canty AJ, Davison AC, et al. Bootstrap confidence
intervalsCommentCommentCommentCommentRejoinder. Stat Sci. 1996;11:189–22 limited utility across countries. Am. J. Transplant. 2022;
DiCiccio TJ, Efron B, Hall P, Martin MA, Canty AJ, Davison AC, et al. Bootstrap confidence
intervalsCommentCommentCommentCommentRejoinder. Stat Sci. 1996;11:189–22 intervalsCommentCommentCommentCommentRejoinder. Stat Sci. 1996;11:189–228.
41. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guid
medical statisticians. Stat. Med. 2000;
42. Kim WR, intervalser J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical g
medical statisticians. Stat. Med. 2000;
Kim WR, Biggins SW, Kremers WK, Wiesner RH, Kamath PS, Benson JT, et al. Hyponatre
mortalit 11. Carpenter J, Bithell Statisticians. Stat. Med. 2000;

42. Kim WR, Biggins SW, Kremers WK, Wiesner RH, Kamath PS, Benson JT, et al. Hyponatremia and

mortality among patients on the liver-transplant waiting list. N. Eng Kim WR, Biggins SW, Kremers WK, Wie
mortality among patients on the liver-1
1026.
Team RC. R: A language and environme
Haendel MA, Chute CG, Bennett TD, Ei
Cohort Collaborative (N3C): Rationale,
Inform. Assoc. 2021;28:427–
-
- mortality among patients on the liver-transplant waiting list. N. Engl. J. Med. 2008;359:1018–1026.
Team RC. R: A language and environment for statistical computing. 2013;
Haendel MA, Chute CG, Bennett TD, Eichmann DA, Gui Team RC. R: A language and environment for statistical computing. 2013;
Haendel MA, Chute CG, Bennett TD, Eichmann DA, Guinney J, Kibbe WA, et al. The National COVID
Cohort Collaborative (N3C): Rationale, design, infrastru ----
Team
Haend
Cohor
Inform
Klann
Progra
2019;1 Haendel MA, Chute CG, Bennett TD, Eichmann DA, Guinney J, Kibbe WA, 6
Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployn
Inform. Assoc. 2021;28:427–443.
Klann JG, Joss MAH, Embree K, Murphy SN. Data
- Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment. J. Am. Med.

Inform. Assoc. 2021;28:427–443.

45. Klann JG, Joss MAH, Embree K, Murphy SN. Data model harmonization for the All Of Us Research
 Inform. Assoc. 2021;28:427–443.
Klann JG, Joss MAH, Embree K, Murphy SN. Data model harmonization for the All Of Us Rese
Program: Transforming i2b2 data into the OMOP common data model. PLoS One.
2019;14:e0212463.
Wang H, Klann JG, Joss MAH, Embree K, M
Program: Transforming i2b2 data
2019;14:e0212463.
Wang H, Belitskaya-Levy I, Wu F, I
method for longitudinal observati
VA million veteran program. BMC
Viernes B, Lynch KE, South B, Coro
- Frame: Transforming i2b2 data into the OMOP common data model. PLoS One.

2019;14:e0212463.

Wang H, Belitskaya-Levy I, Wu F, Lee JS, Shih M-C, Tsao PS, et al. A statistical quality assessment

method for longitudinal obse Program: Transforming the Unit of the OMOP, 14:e0212463.

Wang H, Belitskaya-Levy I, Wu F, Lee JS, Shih M-C, Tsao PS, et al. A statistical quali

method for longitudinal observations in electronic health record data with a 2019;

Wang H, Belitskaya-

method for longitud

VA million veteran p

Viernes B, Lynch KE,

Common Data Mod method for longitudinal observations in electronic health record data with an application to the
VA million veteran program. BMC Med Inform Decis Mak. 2021;21:289.
47. Viernes B, Lynch KE, South B, Coronado G, DuVall SL. C
- VA million veteran program. BMC Med Inform Decis Mak. 2021;21:289.
Viernes B, Lynch KE, South B, Coronado G, DuVall SL. Characterizing VA Users with the OMOP
Common Data Model. Stud. Health Technol. Inform. 2019;264:1614–1 VA million veteran program. BMC Med Inform Decis Mak. 2021;21:289.
Viernes B, Lynch KE, South B, Coronado G, DuVall SL. Characterizing VA Users with the OMOP Common Data Model. Stud. Health Technol. Inform. 2019;264:1614–1 47. Viernes B, Lynch MB, Lynch Lynch Lynch Common Data Model. Stud. Health Technol. Inform. 2019;264:1614–1615. Common Data Model. Stud. Health Technol. Inform. 2019;264:1614–1615.

This ACLF patient has:

- Glucose at the time of transplant <= 136.9, and
- INR at the time of transplant > 2.4 , and
- Serum sodium at the time of transplant <= 142.5

This combination of co-existing clinical conditions is associated with an increased likelihood of post-transplant mortality at 1 year:

- \Box Strong disagree
- \Box Somewhat disagree
- \Box Neither agree nor disagree
- \Box Somewhat agree
- \Box Strongly agree

A. EAML vs Other Models - AUROC Differences and 95% CIs for Death at One-Year

B. RuleFit vs Other Models - AUROC Differences and 95% CIs for Death at One-Year

A. EAML vs Other Models - AUROC Differences and 95% CIs for Death at 90-Days

B. RuleFit vs Other Models - AUROC Differences and 95% CIs for Death at 90-Days

A. EAML vs Other Models - AUROC Differences and 95% CIs for Readmission at 90-Days

B. RuleFit vs Other Models - AUROC Differences and 95% CIs for Readmission at 90-Days

Expert versus RuleFit Rankings for the Outcome of Death at One-Year

Rule #	Rule	Expert Rank	RuleFit Rank
18	MELDNa at Transplant > 32.47		10
14	Serum Alkaline Phosphate at Admission > 289 AND Age > 51.79		19
16	$\left \right $ lonized Calcium at Admission <= 0.98 AND WBC Count at Transplant > 11.89		20
8	SpO2 at Transplant > 98.29 AND Difference in Lactate Dehydrogenase between		17
	Admission and Transplant > 6		
	Serum Glucose at Transplant <= 136.92 AND INR at Transplant > 2.41 AND Serum !		
	Sodium at Transplant <= 142.50		
	Age \le 51.79	18	

Potential Biases from Human Experts

ACLF human experts may be biased/fixated on the importance of commonly utilized biomarkers and scoring systems

Potential Areas for Further Research

Decision-tree algorithms may be identifying previously unrecognized data features or biomarkers that may be more important for ACLF prognostication