**SPECIFIC AIMS**

Machine learning (ML) and artificial intelligence (AI) are expected to significantly influence medical research and practice,1 facilitated by the ongoing expansion of electronic health record (EHR) systems since 2009.2 However, outside of image classification, few ML algorithms have overcome the ‘AI chasm,’3 *i.e.*, the gap between developing a scientifically sound algorithm and its beneficial use in real-world settings. The principal investigator of this proposal, Dr. Byron Jaeger, is an early-stage investigator who has previously developed a ML algorithm, oblique random survival forests (ORSF),4 and disseminated it with an open-source software package (*obliqueRSF)*5 on the Comprehensive R archive network.6 ORSF predicts risk for censored time-to-event outcomes using recursive partitioning with oblique splits, i.e., splitting data into subsets based on linear combinations of predictors rather than a single predictor.7–10 We have applied ORSF to predict risk for incident heart failure (HF), a form of cardiovascular disease (CVD) with growing burden in the United States.11 In this work, recently published in *Circulation*,12 ORSF obtained better discrimination (i.e., less false positive and less false negative cases) in external data versus competing ML algorithms and previously published HF risk prediction equations. In addition to obtaining higher discrimination, ORSF’s predictions highlighted differences in the drivers of predicted risk for HF across racial groups. While natriuretic peptide levels were the strongest predictor overall, HF risk was driven by prevalent CVD and traditional CVD risk factors for white adults, with diabetes/glycemia and socioeconomic factors being more relevant for black adults.

With the work above demonstrating tremendous promise in research settings, ORSF is poised to bridge the AI chasm in clinical settings. However, ORSF’s high computational overhead limits its application to larger datasets collected in clinical settings and its potential to provide interpretable predictions. This is a critical barrier given the scale of current clinical populations, the need for interpretable predictions from “black-box” ML algorithms, and the need for continual learning and model updating to avoid decreases in performance through temporal mechanisms such as calibration drift.13 In addition, while the ORSF model we previously developed identifies patients at high risk for incident HF and is available in a public web application, it does not support patients with prevalent HF. Decision support is critical for patients diagnosed with HF, as two-thirds of patients with HF experience a rehospitalization within the first-year hospital discharge in the US and recent studies have found roughly 10–50% of patients with HF were prescribed at least one potentially harmful medication in ambulatory care and inpatient health care settings. Therefore, we propose to demonstrate feasibility of ORSF in a dedicated clinical decision support system by training it on EHR data and medication records to identify medication related problems such as drug-drug interactions or drug-disease interactions that increase risk for HF-related hospitalization. Our findings may prevent medication related treatment failure and medical problems caused by medication, which cost the US approximately $528.4 billion annually. To accomplish these objectives, we propose following aims:

**Specific Aim 1**. Increase the computational efficiency of the *obliqueRSF* software package. ObliqueRSF relies on external R packages (*e.g.*, *glmnet* and *survival*) to run the ORSF algorithm, leading to unnecessary copying of data and computations. We will develop customized routines using the highly optimized Armadillo library for matrix algebra in C++,14 which seamlessly transfers data to and from the R computing environment,15 and benchmark the accelerated ORSF (AORSF) algorithm against its predecessor using simulated data. Our preliminary findings suggest the computational time required for ORSF can be reduced by a factor of 100 or more, i.e., reducing one hour of computation to 36 seconds.

**Specific Aim 2.** Test the feasibility of using AORSF to guide decisions in clinical settings related to HF. HF is the leading cause of hospitalization among older adults, and Medicare beneficiaries with HF have the highest readmission rate of any condition.16,17 We will develop and evaluate an EHR-adapted model for predicting incident HF risk using patients (free of a HF diagnosis) with a Wake Forest affiliated primacy care provider part of an affiliated Accountable Care Organization from 2017-2019 (expected N>50,000 patients).

This pilot project would lay the groundwork for numerous future directions. The most immediate direction for the proposed study team would be prediction of incident frailty**.** There are several ongoing efforts leveraging the eFI to target interventions for frail older adults, including targeting outreach of community health workers, optimizing pre-operative screening, and deprescribing for patients with type II diabetes. However, these efforts focus on patients that are believed to already be frail (eFI>0.21). Another interventional target could be identifying patients likely to accumulate more age-related deficits, and thus transition from pre-frailty to frailty. Using AORSF to predict risk of transition from pre-frailty to frailty using the same cohort as in Aim 2 (though including patients with HF diagnoses) could create an opportunity to intervene before the transition.

**RESEARCH PLAN**

**Background and Significance**

Machine learning and artificial intelligence are expected to significantly influence medical research and practice,1 facilitated by the ongoing expansion of electronic health record (EHR) systems since 2009.2 However, few machine learning algorithms have overcome the ‘artificial intelligence chasm,’ *i.e.*, the gap between developing a scientifically sound algorithm and its beneficial use in real-world applications.3 A data-driven learning healthcare system that bridges the artificial intelligence chasm requires prediction models with accuracy, interpretability, and robustness in the presence of noisy and irregular data, subject to imperfect ascertainment and documentation.

Neural networks are at the forefront of artificial intelligence for classification based on images, text, or audio data, but they have had limited success with structured clinical data.18 In a recent analysis of prediction models for 30-day readmission following hospitalization for heart failure, neural networks obtained concordance (C-) statistics ranging from 0.628 to 0.643, which was no better that predictions obtained from standard logistic regression (C-statistic = 0.644).19 In addition, the interpretation of predictions from neural networks is complicated by their highly parameterized design, with indistinguishable perturbations of input data leading to different, sometimes conflicting interpretations of predictions from the network.20

Random forests are a machine learning technique that develop an ensemble of decision trees and forms ensemble predictions by aggregating predictions from all the trees.21 Randomness is injected into trees by growing each one with a random sample of the training data, often sampled with replacement, and by using random subsets of predictor variables as candidates to create new nodes in the tree. A benchmark experiment based on 243 tabular datasets, 67 of which were from biomedical settings, estimated a mean C-statistic of 0.867 (95% confidence interval [CI] 0.847, 0.884) for random forest classifiers versus 0.826 (95% CI 0.807, 0.844) for logistic regression.22 Random forests also have a rich and growing set of methods to facilitate interpretability,23–25 and their ensemble design has robustness to noisy or mis-classified data,26,27 making them an ideal method to assist decisions in the clinical setting.

A core principle of personalized medicine is disease prevention, which can only be (correctly) studied in a ‘time to event with censoring’ framework. However, ML algorithms that can engage with censored time to event data are less common than algorithms for classification and regression. In 2008, Ishwaran et al. developed the random survival forest, a formulation of the random forests that can engage with right censored time to event data,9 and later proved that random survival forests were an asymptotically consistent estimator of a survival function conditional on having a discrete feature space comprising all relevant predictor variables.28The principal investigator of this proposal and early stage investigator, Dr. Byron Jaeger, built on Ishwaran et al’s methodology by developing the oblique random survival forest (ORSF),4 which allows the method to further explore high dimensional feature spaces by performing oblique splits (see Section xx.x), thereby improving the chance that relevant features can be leveraged for prediction and a consistent estimator of a survival function can be formed.

[Seems like we need one more paragraph here in background. Probably something related to use of the HER to help facilitate clinical decision support, alluding to eFI work?]

**Approach**

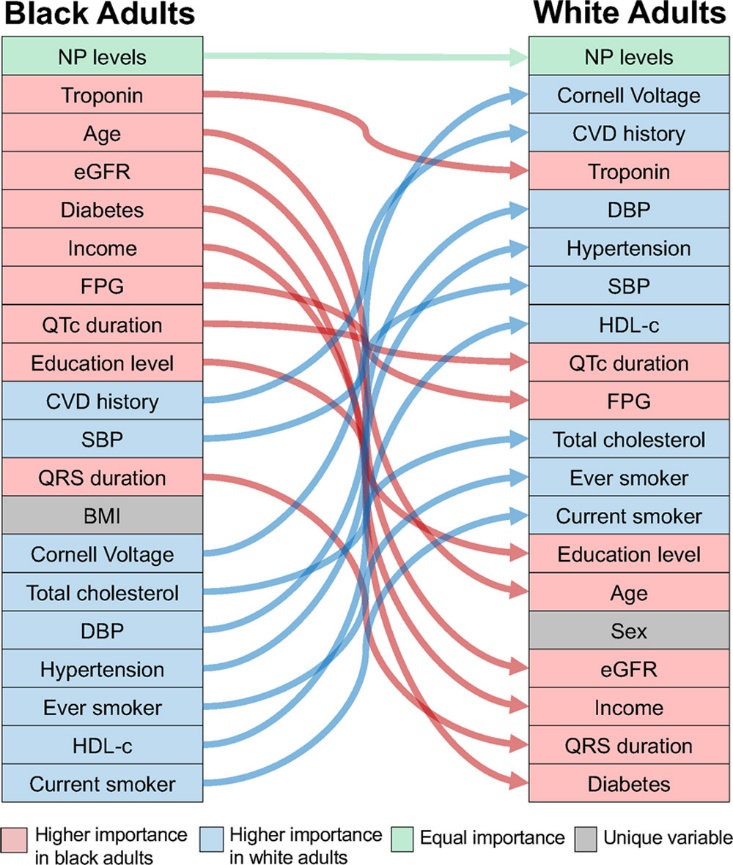
Research Relevant to the Planned Proposal

In 2019, Dr. Jaeger developed the oblique random survival forest (ORSF).4 ORSF forms prediction rules for censored time-to-event outcomes, a ubiquitous and critical data structure in medical research. <ADD paragraph on how oblique splitting works and include figure from seminar?> ORSF has obtained improved or similar predictive performance versus state-of-the-art algorithms, including random survival forests,9 regularized regression (*i.e.*,ridge regression and the LASSO),7,8 and xgboost.10 We recently applied ORSF in the context of risk stratification for incident heart failure (HF), a form of cardiovascular disease (CVD) with growing burden in the United States.11 In this work, recently published in *Circulation*,12 we applied ORSF to develop race-specific risk prediction models for black and white adults using data from several epidemiological cohorts. In external validation, ORSF obtained adequate calibration and better discrimination (i.e., less false positive and less false negative cases) compared to all competing ML algorithms and compared to previously published HF risk prediction equations. ORSF also highlighted different drivers of risk for HF in black and white adults (**Figure 1**). While natriuretic peptide levels were the strongest predictor in both races, HF risk was largely driven by prevalent CVD and traditional CVD risk factors (hypertension, cholesterol, etc.) for white adults, with diabetes/glycemia and socio-economic factors being more relevant for black adults. However, the models developed in this analysis are limited in their scope of application as several predictors are not widely available in clinical practice (*e.g.*, high-sensitivity cardiac troponin, details concerning social determinants of health).

Figure 1: Alluvial plot of the 20 most important predictors for heart failure risk prediction identified by the variable importance metric in the oblique random survival forest.

Overview of Study Design

Study Population

For analyses of incident heart failure and frailty, we will include patients 65 years or older that are part of an affiliated Accountable Care Organization (ACO) and also have a Wake Forest affiliated primary care provider (PCP). We will use an index date of October 1st, 2017, using the previous two years of data in order to characterize patient characteristics, and using data through January 1st, 2022 to examine incident heart failure and frailty. We chose to focus on this population for several reasons. First, ACO patients with a Wake Forest-PCP will tend to have more complete ascertainment of their healthcare utilization within the EHR. Second, heart failure is the leading cause of hospitalization among older adults, and Medicare beneficiaries with heart failure have the highest readmission rate of any condition.16,17 Third, the prevalence of frailty is generally low in adults less than 65 years of age (<5 to 10%), with a prevalence exceeding 20% in adults 75 years or older. As part of several ongoing studies, Dr. Pajewski and members of the Center for Health Care Innovation have developed automated SQL queries that extract and quality control relevant patient characteristics from Clarity (relational database underlying Epic) including demographics, diagnosis codes, medications, vital signs, laboratory data, healthcare utilization, screening data from Medicare Annual Wellness Visits, and other relevant data points using a nightly extract, transform, and load (ETL) process. Therefore, we do not expect any barriers in constructing the necessary analytic datasets needed for our two planned application of the optimized *obliqueRSF* software package.

**Specific Aim 1**: Increase the computational efficiency of the *obliqueRSF* software package. The three areas we will target for improving computational speed are (1) obtaining linear combinations of predictors, (2) selecting optimal cut-points of linear combinations via the log-rank test, and (3) computing predicted risk on new data. To show feasibility, we completed (1) and present computation time of our proposed algorithm to obtain linear combinations of predictors versus algorithms used by *obliqueRSF* (*i.e.*, glmnet with and without 10-fold cross-validation; **Table 1**). Further details and code to reproduce these preliminary findings are available on GitHub.29

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 1:** Average time required to fit modeling techniques  that provide linear combination of inputs | | | | |
| Modeling technique | Milliseconds | | Ratio | |
| Mean | Median | Mean | Median |
| Primary Biliary Cholangitis data (N = 276) | | | | |
| glmnet, 10-fold CV | 8.999 | 8.828 | 240.0 | 236.8 |
| glmnet, no CV | 0.4713 | 0.4638 | 12.57 | 12.45 |
| survival | 0.0922 | 0.0884 | 2.460 | 2.371 |
| AORSF | 0.0375 | 0.0372 | 1 (ref) | 1 (ref) |
| Assay of Serum Free Light Chain data (N = 6,524) | | | | |
| glmnet, 10-fold CV | 90.97 | 90.22 | 252.3 | 252.0 |
| glmnet, no CV | 5.381 | 5.348 | 14.93 | 14.93 |
| survival | 0.8787 | 0.7923 | 2.439 | 2.214 |
| AORSF | 0.3615 | 0.3588 | 1 (ref) | 1 (ref) |
| Computations were repeated 500 times, CV = cross-validation | | | | |

To obtain linear combinations of predictors up to 250 times faster than *obliqueRSF* routines, we created an optimized Newton Raphson algorithm using the Armadillo library in C++. Our algorithm maximizes the partial likelihood of outcomes in training data iteratively with respect to a set of regression coefficients. We tested our algorithm by verifying it gives the exact same parameter estimates as the proportional hazard regression function in the *survival* R package and can run over twice as fast. Code to select optimal cut-points of linear combinations will make use of fast sorting algorithms in the Armadillo library to identify potential cut-points in the vector derived by taking the dot product of a set of predictor values with estimated regression coefficients. In addition, we will leverage computational shortcuts that can be taken when only two groups are compared by the log-rank test, which is always the case when growing twonew nodes in a decision tree. The computation of predicted risk for new data is anticipated to receive the greatest increase in efficiency by computing predictions for leaves in the tree rather than for individual observations. Thus, instead of computing predicted risk for each observation, we only need to identify which leaf the observation is mapped to.

**Specific Aim 2.** Apply the AORSF algorithm to predict incident risk for HF. In our *Circulation* paper, ORSF was the best method for developing HF risk equations, but the models developed were limited in their scope of application as they require several predictors that are not widely available in clinical practice. Therefore, we will use AORSF to develop an EHR-adapted risk prediction model for HF. The EHR-adapted model will account for patient history of diagnoses, procedures, and medications, as well as derived features such as the eFI, healthcare utilization, medication use, and geographic indicators of socioeconomic status like the Area Deprivation Index on or prior to the index date: January 1st, 2017.30

Outcomes

I have questions that we may want to discuss and bring up with teammates before writing this section:

-How will we identify incident HF from 1/1/2017 to 1/1/2022?

-Concerns about under-reporting after 3/12/2020? Move index date to 1/1/2015 to avoid pandemic?

-Should we make separate models for HF with PEF versus HF with REF?

|  |  |
| --- | --- |
| **Hospitalization for Heart Failure in the past 12 months, defined as:** | |
| Diagnosis code (ICD-10s provided below) for HF, and encounter = inpatient in the 12 months prior to query date | |
|  |  |
| I11.0 | Hypertensive heart disease with heart failure |
| I13.0 | Hypertensive heart and chronic kidney disease with heart failure and stage 1 through stage 4 chronic kidney disease, or unspecified chronic kidney disease |
| I13.2 | Hypertensive heart and chronic kidney disease with heart failure and with stage 5 chronic kidney disease, or end stage renal disease |
| I50.1 | Left ventricular failure, unspecified |
| I50.20 | Unspecified systolic (congestive) heart failure |
| I50.21 | Acute systolic (congestive) heart failure |
| I50.22 | Chronic systolic (congestive) heart failure |
| I50.23 | Acute on chronic systolic (congestive) heart failure |
| I50.30 | Unspecified diastolic (congestive) heart failure |
| I50.31 | Acute diastolic (congestive) heart failure |
| I50.32 | Chronic diastolic (congestive) heart failure |
| I50.33 | Acute on chronic diastolic (congestive) heart failure |
| I50.40 | Unspecified combined systolic (congestive) and diastolic (congestive) heart failure |
| I50.41 | Acute combined systolic (congestive) and diastolic (congestive) heart failure |
| I50.42 | Chronic combined systolic (congestive) and diastolic (congestive) heart failure |
| I50.43 | Acute on chronic combined systolic (congestive) and diastolic (congestive) heart failure |
| I50.9 | Heart failure, unspecified |

Model Development

Before taking any steps in model development, we will set aside 30% of our data for validation. The remaining 70% of our data will be used for derivation of a 5-year risk prediction model for incident HF.

Data pre-processing steps: We will derive predictors based on a 2-year lookback period preceding the index date. As the number of unique medical diagnosis, procedure, and medication codes is large, we will use standard medical concept groupers to group codes documented during the lookback period into overarching concepts. ICD-9 diagnosis codes can be grouped by the Clinical Classification Software diagnosis grouper into 283 concepts,31 medication codes can be grouped by the Generic Product Identifier into 96 concepts,32 and Current Procedural Terminology codes can be grouped by the Clinical Classification Software procedure grouper into 244 concepts.33 We will count the number of diagnoses, procedures, and medications in each concept for each patient during the lookback period. For each numeric laboratory value, we will estimate the mean and slope of the value during the lookback period, yielding 2 predictors for each laboratory measurement (e.g., mean and slope of systolic blood pressure during the lookback period), with a missing slope value when one or fewer measurements were present. The same strategy will be applied to numeric variables pertaining to census and income information, with a slope of 0 if there is no documented change during the 2-year lookback period. Missing values for continuous and categorical variables will be imputed to the mean and mode, respectively, as this method has been shown to produce consistent prediction models when missing values are not informative.34 Last, a subset of the 20 most important predictor variables will be selected using permutation importance.

Tuning: Tunable parameters for AORSF include the minimal number of events and observations required in decision tree leaves, the number of input variables included in linear combinations, and the minimal log-rank statistic required to create two new nodes in the tree. These parameters will be tuned with 10-fold cross-validation using the derivation data. We will divide the derivation data into 10 non-overlapping subsets of approximately equal size, and label them ‘fold 1’ through ‘fold 10’. Then, one by one, we will set 1 fold aside as testing data, and use the remaining 9 folds as training data. To ensure that cross-validation will not be optimistically biased by allowing information from the testing data to inform pre-processing of training data, the pre-processing steps outlined above will be carried out in the 9 folds reserved for training, and the held-out fold will undergo pre-processing based on the values computed in the training data. We will compute the integrated time-dependent Brier score of each AORSF model in the held-out fold of data before repeating the procedure with a new held-out fold.35 The Brier score is an overall goodness of fit metric based on both discrimination and calibration of a prediction model. After each fold has been held out once, we will aggregate discrimination and calibration performance statistics across all 10 folds using the median.

Model Fitting and Interpretation: Whichever AORSF parameter specification obtains the best Brier score will be used to develop a final AORSF model using the entire derivation dataset. Similar to our previous work in *Circulation*, we will compute partial dependence estimates of predicted risk with respect to each predictor in the model, giving a depiction of how each variable contributes to the model’s predicted risk function. In addition, we will compute SHapley Additive exPlanation (SHAP) values,36 which provide a measure of variable importance and can also be used to show the relationship between a model’s prediction function and a predictor variable.23

Model Validation and Fairness

Using the validation set, we will assess discrimination of the final AORSF model for 5-year HF risk prediction with a time-dependent C-statistic. Calibration will be assessed using the Greenwood Nam D’Agostino test and a calibration slope curve. These metrics will be assessed overall and in subgroups determined by race, sex, age, and geographic region. We will also run standardized checks (e.g., accuracy equality, equal opportunity, predictive equality, and predictive and statistical parity ratio) to assess model fairness for these subgroups. If unfairness is detected for any subgroup, we will identify group-specific risk thresholds for risk stratification to mitigate unfairness.

Research Environment, Resources, and Dissemination

This study will leverage existing open-source tools for development and dissemination of statistical software. For development, we will use the R Statistical Computing Environment, the leading programming language for academic research in data science. We will leverage Rcpp and RcppArmadillo to seamlessly transfer data between R and C++, a computationally efficient low-level programming language. For dissemination, we will develop a new R package, *aorsf*, and publish this package on the Comprehensive R Archive Network, as Dr. Jaeger has previously done for three R packages he maintains. Publication on this open-source platform will allow any R user with internet connection access to the software. Dr. Jaeger and Dr. Pajewski will manage the development of *aorsf* using git, the industry-standard platform for collaborative code development. As with previous R packages developed by Dr. Jaeger, source code will be disseminated through GitHub under the MIT license. Resources for package users, including documentation and hands-on examples using ORSF, will be hosted via GitHub Pages as Dr. Jaeger has done for previous software dissemination.37

Limitations

Our EHR-adapted HF risk prediction model may not be able to leverage all the biomarkers that were used in the HF risk prediction model we published in *Circulation*, as some of these measurements are rarely taken in routine clinical practice. However, the EHR-adapted model will leverage more data pertaining to conditions in the environments where patients live, learn, work, play, worship, and age, *i.e.,* social determinants of health. Our previous work showed that social determinants of health (*e.g.*, income and education) were among the most important risk factors for HF among black adults. Therefore, the proposed HF risk prediction model may improve risk stratification for black adults with adverse social determinants of health.

**STUDY MILESTONES: Anticipated Outcomes and Dissemination Plan**

This study proposes to adapt previously written code for a novel ML algorithm into a software package that can operate at scale with massive databases extracted from the EHR. The proposed study will also apply this software package to develop risk prediction equations for two highly relevant outcomes for older adults, incident heart failure and frailty. **Table 2** presents our proposed timeline of milestones and actions to disseminate the proposed research. We expect to spend the first four months of the project developing core features of *aorsf*, *i.e.*, writing the functions that will grow decision trees and generate risk predictions. Future grant applications will build additional features on top of the core ones, *e.g.*, conducting statistical tests to determine whether an individual variable contributes to the model’s prediction function, accounting for competing risks and recurrent events such as hospitalizations, and incorporating missing data as a feature. We will also initialize the data extraction from WakeOne during this period by conducting initial data extractions and evaluating the validity of derived variables such as the eFI. During the second four months of the project, we plan to complete data extraction, data cleaning and quality control, and perform robust testing of the core features in *aorsf*. Given the high informatics and analytic burden of this project, the study team will meet at least bi-monthly in order to review progress. Scholarly products from this project will include a manuscript describing our development, validation, and fairness assessment of the HF risk prediction model. We intend to apply for R21 funding to conduct an external validation of the HF risk prediction model within the network infrastructure provided by Atrium Health’s merger with WakeOne. Finally, we expect to develop an R01 application based on this work, for example using the AORSF algorithm to predict the onset of frailty. There are several ongoing efforts leveraging the eFI to target interventions for frail older adults, including targeting outreach of community health workers, optimizing pre-operative screening, and deprescribing for patients with type II diabetes. However, these efforts focus on patients that are believed to already be frail (eFI>0.21). Another interventional target could be identifying patients likely to accumulate more age-related deficits, and thus transition from pre-frailty to frailty. Our planned R01 application will apply the *aorsf* Rpackage to examine predictors of the transition from pre-frailty to frailty using the same cohort as derived in this pilot application (though including patients with HF diagnoses).

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2**: Timeline and Milestones for Program Activities. | | | | | | | | | | | | |
| **Month** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Development Phase | | | | | | | | | | | | |
| Twice monthly project meetings | X | X | X | X | X | X | X | X | X | X | X | X |
| Initial data queries to evaluate data quality | X | X | X | X |  |  |  |  |  |  |  |  |
| Develop core features in aorsf package | X | X | X | X |  |  |  |  |  |  |  |  |
| Evaluate availability of Medicare claims | X | X | X | X |  |  |  |  |  |  |  |  |
| Implementation Phase | | | | | | | | | | | | |
| Test validity of aorsf core features |  |  |  | X | X |  |  |  |  |  |  |  |
| Final data extraction from WakeOne |  |  |  | X | X | X |  |  |  |  |  |  |
| Link to Medicare Claims data |  |  |  | X | X | X |  |  |  |  |  |  |
| Data cleaning |  |  |  |  | X | X | X |  |  |  |  |  |
| Data analysis |  |  |  |  |  |  | X | X | X | X |  |  |
| Scholarly Products | | | | | | | | | | | | |
| Pilot Manuscript |  |  |  |  |  |  |  |  |  | X | X | X |
| Publish aorsf on CRAN |  |  |  |  |  |  |  |  |  | X | X | X |
| Further grant application submission |  |  |  |  |  |  |  |  |  | X | X | X |

CRAN: Comprehensive R Archive Network.

**REFERENCES**

1. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. *npj Digital Medicine*. 2020;3:126.

2. Evans RS. Electronic Health Records: Then, Now, and in the Future. *Yearbook of Medical Informatics*. 2016;:S48.

3. Keane PA, Topol EJ. With an eye to AI and autonomous diagnosis. *npj Digital Medicine*. 2018;1:40.

4. Jaeger BC, Long DL, Long DM, Sims M, Szychowski JM, Min Y-I, Mcclure LA, Howard G, Simon N. Oblique random survival forests. *The Annals of Applied Statistics*. 2019;13. doi:10.1214/19-AOAS1261.

5. Jaeger BC. obliqueRSF. Available at https://cran.r-project.org/package=obliqueRSF. Accessed October 26, 2021.

6. Hornik K. The Comprehensive R Archive Network. *Wiley Interdisciplinary Reviews: Computational Statistics*. 2012;4:394–398.

7. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for Cox’s proportional hazards model via coordinate descent. *Journal of Statistical Software*. 2011;39:1–13.

8. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*. 2010;33:1–22.

9. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Annals of Applied Statistics*. 2008;2:841–860.

10. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016;13-17-Augu:785–794.

11. Urbich M, Globe G, Pantiri K, Heisen M, Bennison C, Wirtz HS, Di Tanna GL. A Systematic Review of Medical Costs Associated with Heart Failure in the USA (2014–2020). *PharmacoEconomics*. 2020;38:1219–1236.

12. Segar MW, Jaeger BC, Patel K V, Nambi V, Ndumele CE, Correa A, Butler J, Chandra A, Ayers C, Rao S, Lewis AA, Raffield LM, Rodriguez CJ, Michos ED, Ballantyne CM, Hall ME, Mentz RJ, de Lemos JA, Pandey A. Development and Validation of Machine Learning-Based Race-Specific Models to Predict 10-Year Risk of Heart Failure: A Multicohort Analysis. *Circulation*. 2021;143:2370–2383.

13. Davis SE, Lasko TA, Chen G, Matheny ME. Calibration Drift Among Regression and Machine Learning Models for Hospital Mortality. *AMIA . Annual Symposium proceedings AMIA Symposium*. 2017;2017:625–634.

14. Sanderson C. Armadillo: An open source C++ linear algebra library for fast prototyping and computationally intensive experiments. 2010.

15. Eddelbuettel D, Analysis CS-CS& D. RcppArmadillo: Accelerating R with high-performance C++ linear algebra. *Elsevier*. 2014.

16. Jencks SF, Williams M V., Coleman EA. Rehospitalizations among Patients in the Medicare Fee-for-Service Program. *http://dx.doi.org/101056/NEJMsa0803563*. 2009;360:1418–1428.

17. Jackson SL, Tong X, King RJ, Loustalot F, Hong Y, Ritchey MD. National Burden of Heart Failure Events in the United States, 2006 to 2014. *Circulation Heart failure*. 2018;11:e004873.

18. Arık S, arXiv TP-, 2020 undefined. Tabnet: Attentive interpretable tabular learning. *aaai.org*.

19. Allam A, Nagy M, Thoma G, Krauthammer M. Neural networks versus Logistic regression for 30 days all-cause readmission prediction. *Scientific Reports 2019 9:1*. 2019;9:1–11.

20. Ghorbani A, Abid A, Zou J. Interpretation of neural networks is fragile. *33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019*. 2019;:3681–3688.

21. Breiman L. Random forests. *Machine Learning*. 2001;45:5–32.

22. Couronné R, Probst P, Boulesteix A-L. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics 2018 19:1*. 2018;19:1–14.

23. Lundberg SM, Erion GG, Lee S-I. Consistent Individualized Feature Attribution for Tree Ensembles. 2018.

24. Hatwell J, Gaber MM, Azad RMA. CHIRPS: Explaining random forest classification. Springer Netherlands; 2020: 1-5747–5788. doi:10.1007/s10462-020-09833-6.

25. Bénard C, Biau G, da Veiga S, Scornet E. Interpretable Random Forests via Rule Extraction. 2020.

26. Dietterich TG. Experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. *Machine Learning*. 2000;40:139–157.

27. Grandvalet Y. Bagging equalizes influence. *Machine Learning*. 2004;55:251–270.

28. Ishwaran H, Kogalur UB. Consistency of Random Survival Forests. *Statistics & probability letters*. 2010;80:1056.

29. Jaeger BC. Fast linear combinations of predictors with RcppArmadillo; Preliminary results for CTSI Pilot application. 2021. doi:10.5281/zenodo.5655329. Available at https://bcjaeger.github.io/coxph\_bench/index.

30. Kind AJH, Buckingham WR. Making Neighborhood-Disadvantage Metrics Accessible - The Neighborhood Atlas. *The New England journal of medicine*. 2018;378:2456–2458.

31. Clinical classifications software (CCS) for ICD-9-CM. 2018. Available at https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp. Accessed November 9, 2021.

32. Medi-Span Electronic Drug File (MED-File) v2. Wolters Kluwer Clinical Drug Information. Available at https://www.wolterskluwer.com/en/solutions/medi-span/medi-span/drug-data/core-drug-data-and-picklists. Accessed November 9, 2021.

33. HCUP-US Tools and Software Page CCS-Services and Procedures. Agency for Healthcare Research and Quality. Available at https://www.hcup-us.ahrq.gov/toolssoftware/ccs\_svcsproc/ccssvcproc.jsp. Accessed November 9, 2021.

34. Josse J, Prost N, Scornet E, Varoquaux G. On the consistency of supervised learning with missing values. *arXiv preprint arXiv:190206931*. 2019.

35. Mogensen UB, Ishwaran H, Gerds TA. Evaluating random forests for survival analysis using prediction error curves. *Journal of statistical software*. 2012;50:1.

36. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems. ; 2017: 4765–4774.

37. Jaeger B. Make and Apply Customized Rounding Specifications for Tables: table.glue. Available at https://bcjaeger.github.io/table.glue/index.html. Accessed November 10, 2021.