

Predictive Modeling of TBI Outcomes in Rwanda through Machine Learning

by

Shantanu Srivatsa
Duke Global Health Institute
Duke University

Date: _____

Approved:

Joao Vissoci, Advisor

Catherine Staton

Lawrence Park

Thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science in the Duke Global Health Institute
in the Graduate School of Duke University

2020

ABSTRACT

Predictive Modeling of TBI Outcomes in Rwanda through Machine Learning

by

Shantanu Srivatsa
Duke Global Health Institute
Duke University

Date: _____

Approved:

Joao Vissoci, Advisor

Catherine Staton

Lawrence Park

An abstract of a thesis submitted in partial
fulfillment of the requirements for the degree
of Master of Science in the Duke Global Health Institute
in the Graduate School of Duke University

2020

Copyright by
Shantanu Srivatsa
2020

Abstract

Background: Globally, many low-income settings lack diagnostic tools to handle prognosis of TBI patients. In such settings, development of generalizable predictive models which indicate likelihood of patient outcomes may help improve decision-making for physicians and health care providers.

Methods: An analysis of a Rwanda TBI registry (n=682) was conducted to determine key predictors of TBI mortality. A previously developed prognostic model of a Tanzania TBI registry (n=3209) was subsequently implemented in the Rwanda TBI registry for external validation. 8 different machine learning models were implemented in the Rwanda dataset. Subsequently, 6 Tanzania models and a combined model aggregating the Tanzania model predictions were used to compare and predict Rwanda patient outcomes.

Results: The predictive models developed in Rwanda had satisfactory predictive ability, with the best performing model, Ridge Regression having an AUC of 90% (CI: 89.3%-90.7%). The models developed in Tanzania and used to predict outcomes within the Rwanda dataset showed similar predictive ability, the best performing Random Forest model, AUC 91.3% (CI: 88.0%-94.6%) and the combined machine learning model, AUC 91.9% (CI: 88.7%-95.1%).

Conclusions: The results from the Tanzania models indicate satisfactory predictive ability. The ability of the models to hold similar predictive power in an external dataset, with the use of indicators collectable at triage suggests potential applicability in other low-resource settings.

Contents

Abstract	iv
List of Tables.....	viii
List of Figures.....	ix
Acknowledgements.....	x
1. Introduction.....	1
2. Methods.....	4
2.1 Study Design.....	4
2.2 Setting	4
2.3 Measures.....	6
2.4 Analysis	6
Processing of Data	6
Machine learning models	7
3. Results	12
3.1 Demographic characteristics	12
3.2 Model Results	13
3.2.1 Rwanda machine learning models	13
3.2.2 External Validation of Tanzania developed machine learning models in Rwanda.....	17
4. Discussion.....	20
4.1 Study strengths and limitations	24
5. Conclusion	26

Appendix A	27
References	29

List of Tables

Table 1: Demographic characteristics of Rwanda Dataset.....	12
Table 2: Performance Metrics of Rwanda Machine Learning models	14
Table 3: Performance Metrics of Tanzania predictive models externally validated in Rwanda	18

List of Figures

Figure 1: Analysis flow: Development and implementation of machine learning models	8
Figure 2: Rwanda Machine Learning model ROC curves	15
Figure 3: Variable importance for Rwanda TBI Random Forest and Ridge Regression models	16
Figure 4: Tanzania in Rwanda model ROC curves.....	18
Figure 5: Calibration Curves for best performing Tanzania models in Rwanda	20
Figure S1: Visualization of AUC performance as model ensembling increases for individual models, simple averaging (WOC), and optimal weighted averaging (SUMMA).....	27
Figure S2: ROC/Calibration curves for algorithm-based Weighted Ensemble Model	28

Acknowledgements

I would like to acknowledge the support and guidance of Dr. Joao Vissoci, Dr. Catherine Staton, and Dr. Larry Park as committee members and mentors. Furthermore, I would like to acknowledge Armand Zimmerman for initial guidance and support with machine learning models.

1. Introduction

Traumatic Brain Injury (TBI) is a neurological condition caused by an impact or injury to brain tissue. It has a widely heterogeneous presentation, ranging from anything between a mild concussion to injury leading to deficits, disability, or death. The economic and social impacts of TBI are enormous, including hospitalization costs, lost productivity, cognitive deficits, and disability.^{1,2}

Globally, TBI is responsible for more than 4.7 million deaths and 50 million disabilities annually.^{5,6} In 2020, researchers predict that TBI will become a leading cause of global injury-related mortality and morbidity. Unfortunately, the burden of TBI is disproportionately shared, with low and middle-income countries (LMICs) bearing 90% of TBI injuries.^{7,8} Of patients with these injuries, odds of mortality are twice as high as a patient with a similar injury in a high income country (HIC).³¹ One factor contributing to rising rates of TBI injury in LMICs is rapid urbanization, which has led to increased traffic infrastructure without necessary safety precautions. In addition, nascent health-care systems which haven't fully developed may struggle to integrate diagnostic and chronic care models. Among global regions, TBI related injury in Sub-Saharan Africa is a concern in particular. Rates of road traffic injuries have exceeded both the global and regional rates and seem poised to continue rising.² In Rwanda, TBI rates match that of the region, largely due to road traffic injuries.² A study at the University Teaching Hospital of Kigali (UTHK) found that TBI related mortality

was 58%, with severe TBI patient mortality at 89%.⁹ Furthermore, diagnostic and treatment delays have severe implications for poor outcomes. Although many severely injured patients may receive quick treatment decisions, patients with moderate TBI injuries who experience diagnostic or treatment delays may deteriorate and have increased risk of a poor outcome.^{17,22}

Thus, timely diagnostic tools are critical when dealing with neurological trauma, since even small treatment delays have implications for poor health outcomes. Unfortunately, many low resource settings lack diagnostic tools such as validated clinical guidelines, available CT scans, and an adequate number of providers to handle high volumes of TBI patients.³²⁻³⁴ However, in such settings where diagnostic tools are lacking, prognostic models which indicate the likelihood of patient outcomes may help improve decision-making capabilities for physicians and health care providers. These prognostic models assess patient demographic and clinical data gathered during triage, and can assist physicians in clinical decision making by predicting the likelihood of poor health outcomes with inputted information. Accurate prognosticating is an acquired skill which requires a robust health system and experienced physicians, components which are often lacking in LMIC settings. As a majority of patients in emergency care see junior practitioners, models can help junior practitioners develop this skill more quickly.

Predictive modeling using machine learning is one such method of assessing TBI prognostically and analyzing indicators that contribute to specific outcomes such as

death or GOS (Glasgow Outcome Score). Although these models have shown promising success and in some cases superior decision-making capacity to clinicians, little work has been done in low and middle income countries (LMIC) with location specific data.^{11,27} This is important, given the fact that models cannot simply be implemented in various locations; models developed in high income countries function very poorly when data from LMICs are applied without adjustment.³⁰ Furthermore, there are very few examples of models with generalizable predictive ability for TBI outcomes across other low-resource settings.³⁷ The aims of the present study are to model TBI mortality with an array of machine learning methods (1) using data from a Rwanda TBI registry and (2) using models trained with data from a Tanzania TBI registry and examining their predictive value in Rwanda.

2. Methods

2.1 Study Design

A retrospective analysis of a Rwanda hospital TBI registry was conducted to determine key predictors of TBI mortality. In addition, a previously developed prognostic model of a Tanzania TBI registry was subsequently implemented in the Rwanda TBI registry for external validation (Rocha).

2.2 Setting

Data were sourced from UTHK (University Teaching Hospital), located in Kigali, Rwanda. UTHK is a public referral center centrally located in Kigali which holds approximately 500 beds.¹⁶ UTHK A&E holds 30 patient beds, a 5-bed intensive care unit with ventilators, a 64-slice CT scanner.¹⁶ During this study, there was 1 neurosurgeon on staff along with 2 general practitioners, who were trained in neurosurgery service.¹⁶ These providers would accept consultations while they were in Kigali and not occupied, and practical availability was limited by the size of the service.¹⁶ Furthermore, critical patients could not always be assessed immediately.¹⁶ The dataset included 682 de-identified patients whose data were collected between October 2013, and April 2014. Inclusion criteria included presentation with TBI (headache, reported trauma to the head, visible trauma to the head, or any alteration in consciousness in the context of an

injury), ≥ 10 years of age, and acute injury (< 48 hours), and alive upon arrival. The registry does not include information after discharge due to lack of patient follow-up.

The Tanzania dataset used to train models for external validation in Rwanda included 3209 patients admitted for TBI at Kilimanjaro Christian Medical Centre (KCMC) hospital in Moshi, Tanzania from 2013 to 2017. KCMC is a 630 bed, tertiary level hospital serving a population of over 15 million people in Tanzania.²² TBI patients at KCMC are evaluated and triaged in the emergency department, which is staffed with one trained emergency medicine physician and general surgery interns.²² The general surgery interns, residents and physicians are trained in basic neurosurgical procedures and rarely craniotomies and craniectomies.²² There are no formally trained neurosurgeons at KCMC and initial assessment of injury severity and risk of poor outcome is largely based on the healthcare provider's judgement.²² The use of resource intensive diagnostics such as CT is limited by cost, which must be paid prior to obtaining the imaging, as well as availability and functioning of the equipment.²² Data in the KCMC registry included information on demographics, vital signs, injury type, treatment, and outcome. The registry also did not include information after discharge due to lack of patient follow-up.

2.3 Measures

The variables selected were: Age, Gender, Mechanism of Injury, presence of alcohol, clinical data collected in the ER (Temperature, Heart Rate, Respiratory Rate, Systolic Blood Pressure, Diastolic Blood Pressure, Pulse Oxygenation), Glasgow Coma Score, and if the patient had surgery for TBI. The choice of indicators was decided with the consultation of emergency medicine and neurosurgery clinicians, as well as through comparison with variables mutually present in the previously developed prognostic model from Tanzania.²² Only indicators practical for use in a low-resource setting were considered, emphasizing variables that could be collected during triage (not post-surgical care or follow-up data). The outcome variable analyzed was mortality (alive or dead). Mortality was generated from existing GOS outcome data. The GOS is a 5 point scale which indicates the severity of the patient outcome ranging from 1 (death) to 5 (excellent recovery). For this case study, the GOS was dichotomized as either dead (GOS 1) or alive (GOS 2-5).

2.4 Analysis

Processing of Data

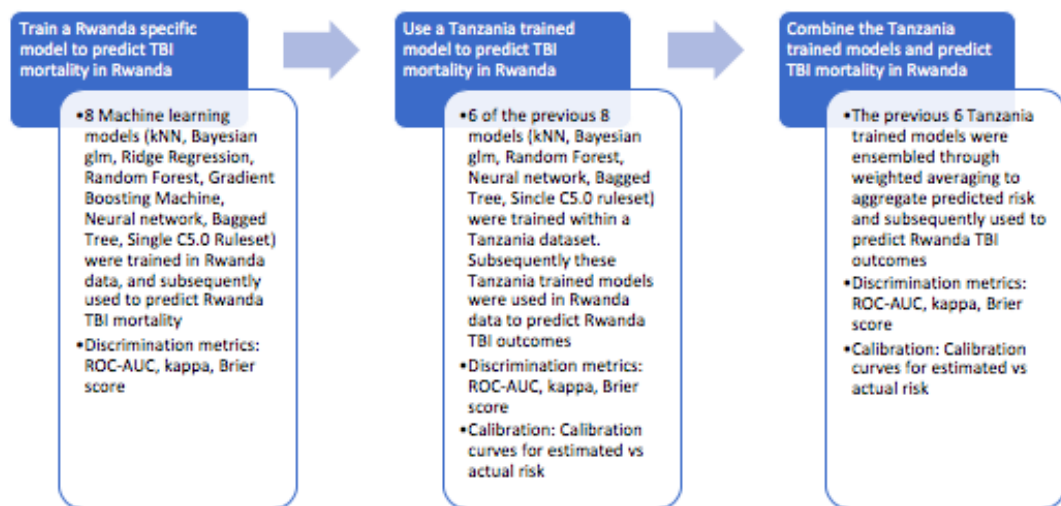
Age, GCS, and clinical data were treated as continuous predictors. Gender, presence of alcohol, TBI surgery, and mechanism of injury were coded as categorical predictors with gender, alcohol, and TBI surgery being dichotomous, and mechanism of

injury including the categories of road traffic injury, assault, fall, and other. Of the 682 patients in the Rwanda registry, 332 patients left or were discharged without recorded GOS outcome data. Despite their lack of specific GOS outcome data, we included these individuals within the cohort of alive patients due to presence of indicator data and knowledge of their mortality status upon departure. Next, missing data was handled and multiple imputation using chained equations was done to impute missing values and the resulting dataset was obtained after 10 iterations (all variables with less than 20% missing values prior to imputation). Numeric variables were imputed using predictive mean matching whereas categorical variables were imputed using polytomous regression imputation for unordered categorical data. A high correlation analysis was performed following the imputation steps to identify high correlation variables but no variables were dropped. After, the categorical variables were expanded to include all the factor levels through dummy variable conversion. The data was investigated for the presence of near zero variance, high correlation and linear combinations. Ultimately, the Rwanda dataset included 682 patients and 17 variables after all processing steps, and the Tanzania dataset included 3209 patients and 17 variables.

Machine learning models

8 different machine learning models were tested in Rwanda, including Bayesian generalized linear model, K-nearest neighbor, Ridge Regression, Random Forest, Neural

Network, Bagged Tree, Single C5.0 Ruleset, and Gradient Boosting Machine. Of the aforementioned modelling techniques, all but Ridge Regression and Gradient Boosting Machine, were trained in Tanzania data and subsequently examined for predictive ability in the Rwanda registry. Finally, a combined machine learning model was implemented, combining the predictive ability of the 6 Tanzania models, as discussed further later. An analysis flow depicting the use of these models is also shown below.



Analysis Flow: Development and implementation of machine learning models

Figure 1: Analysis flow: Development and implementation of machine learning models

Modeling was done using the R packages Caret, pROC, arm, scoring, SuperLearner, SUMMA, and obliqueR which contained the necessary tools to implement the training and prediction of the modeling techniques as well as AUC, metric, and calibration visualizations. The data was internally validated and split using repeated cross validation with 10 fold partitioning and five repetitions. Since the outcome was imbalanced due to only a small fraction of patients dying (9.4% mortality), a normalization procedure called Synthetic Minority Over-sampling Technique (SMOTE) was used in order to account for this imbalance. This helps account for a model in which the algorithm predicts everyone died, but still has good accuracy due to the imbalance in outcome distribution.

The combined machine learning model integrated the predictive ability of the Tanzania models using weighted averaging of the predicted probabilities to create an ensemble model. Higher weights were given to models which showed better predictive ability in the Rwanda dataset.³⁹ The rationale behind model ensembles is based on the theory that aggregating several model outputs can average out biases and reduce total model variance.^{36,38} Furthermore, if each model varies in its specific error and misclassification, averaging out these models will effectively “cancel” these differing errors.^{35,36,39} Ensemble methods tend to show less improvement over individual models in cases where model predictions greatly overlap.³⁵⁻³⁹ A visualization of how weighted ensembling increases predictive power over simple averaging and individual modeling

is shown in the Appendix in Figure S1. For the Tanzania model ensemble, rather than simple averaging, a weighted average was employed, providing greater weight to the predictions with greater accuracy in the Rwanda dataset. These averages were calculated through trial and error, eventually employing the Random Forest, Neural network, and Bayesian models to maximize AUC. In addition, weighted ensembling was performed through the use of the SUMMA package, which requires no knowledge of actual values to predict optimal model weights for binary classification modeling.³⁹ Although the weighted averages calculated through trial and error are provided in the Results, a secondary ROC curve with weights calculated through outcome blinded algorithm optimization is provided in the Appendix in Figure S2.

Model metrics

Discrimination

The ROC-AUC, kappa statistic and the Brier Score were the metrics used to assess the discriminatory ability of the prognostic models. The Brier Score is a score function ranging from 0 to 1 which measures the accuracy of probabilistic predictions, with values closer to 0 indicating higher accuracy. It measures the differences between predicted probabilities of the outcome(0%-100%) and the outcome (0 or 1) across the dataset. The metrics for comparison among the models were chosen based on confusion matrix statistics, including area under the ROC curve (AUC), sensitivity, specificity, positive predictive value, and negative predictive value.

Calibration

Calibration curves which show the estimated vs. actual risk were created for all external validation models implemented in Rwanda. Calibration is an important metric for machine learning models because although a model may discriminate well between two classes (distinguish between dead and alive), the underlying probabilities may not reflect accurate risk, potentially affecting ability to aid clinical decision-making.

3. Results

3.1 Demographic characteristics

Demographic characteristics of patients in the Rwanda registry are shown in Table 1. Most patients (81%) were male, had a road traffic injury (62%), and did not have surgery (95%). Mean age was 31 (S.D. 12) and median GCS was 13 (IQR 11-13).

Table 1: Demographic characteristics of Rwanda Dataset

Basic descriptive statistics from the Rwanda dataset				
	Alive No. 618	Dead No. 64	Total No. 682	P-value
Male				0.044
Female	123 (20%)	6 (9%)	129 (19%)	
Male	492 (80%)	58 (91%)	550 (81%)	
Missing	3 (0%)	0 (0%)	3 (0%)	
Age				0.022
Mean (SD)	31 (\pm 12)	35 (\pm 13)	31 (\pm 12)	
Missing	1 (0%)	2 (3%)	3 (0%)	
Respiration Rate				0.21
Mean (SD)	19 (\pm 2)	20 (\pm 4)	19 (\pm 3)	
Missing	36 (6%)	8 (12%)	44 (6%)	
Heart Rate				0.030
Mean (SD)	86 (\pm 17)	94 (\pm 30)	87 (\pm 18)	
Missing	28 (5%)	7 (11%)	35 (5%)	
Pulse ox				< 0.0001
Mean (SD)	97 (\pm 3)	91 (\pm 12)	96 (\pm 5)	
Missing	90 (15%)	10 (16%)	100 (15%)	
Mode of Injury				0.011
Assault	153 (25%)	13 (20%)	166 (24%)	
Fall	73 (12%)	6 (9%)	79 (12%)	
Other	10 (2%)	6 (9%)	16 (2%)	
Road Traffic Injury	382 (62%)	39 (61%)	421 (62%)	
TBI surgery				0.046
No	592 (96%)	58 (91%)	650 (95%)	
Yes	23 (4%)	6 (9%)	29 (4%)	
Missing	3 (0%)	0 (0%)	3 (0%)	
GCS (Mean/SD)	14 (\pm 2)	8 (\pm 4)	13 (\pm 3)	< 0.0001
GCS (Median/IQR)	13 (12 - 13)	5 (3 - 9)	13 (11 - 13)	< 0.0001

3.2 Model Results

3.2.1 Rwanda machine learning models

Results from the 8 different models implemented in the Rwanda TBI registry are shown in Table 2. The AUC of the model with the highest performance was the Ridge Regression model with an AUC of 90% (CI: 89.3%-90.7%, Brier Score 0.246), and the model with the worst predictive ability was the K-nearest neighbor model with an AUC of 73.2% (CI: 72.6%-73.7%, Brier Score 0.309). A comparison of the model ROC curves is shown in Figure 2.

Variable importance for the two best performing models in Rwanda was calculated and is shown in Figure 3. In the Random Forest model, the most important indicators for predicting TBI mortality were GCS score, pulse oximetry, heart rate, respiratory rate, age, temperature, systolic blood pressure, and diastolic blood pressure. In the Ridge Regression model, the most important indicators for predicting TBI mortality were GCS score, pulse oximetry, heart rate, age, respiratory rate, male gender, and absence of alcohol.

Table 2: Performance Metrics of Rwanda Machine Learning models

Performance Metrics of Rwanda Machine Learning models

	Model	AUC	95% Confidence Interval	Kappa Statistic	Brier Score
1	Bayesian GLM	89	84.4%-93.6%	0.476	0.193
2	Ridge Regression	90	89.3%-90.7%	0.441	0.246
3	Random Forest	89.9	88.3%-91.6%	0.466	0.182
4	Bagged Tree	89.5	87.9%-91.2%	0.438	0.188
5	Gradient Boosting Machine	89.2	88.8%-89;5%	0.486	0.181
6	Neural Network	77.1	74.3%-79.9%	0.338	0.314
7	K-nearest neighbor	73.2	72.6%-73.7%	0.331	0.309
8	Single C5.0 ruleset	81.8	79.3%-84.3%	0.395	0.243

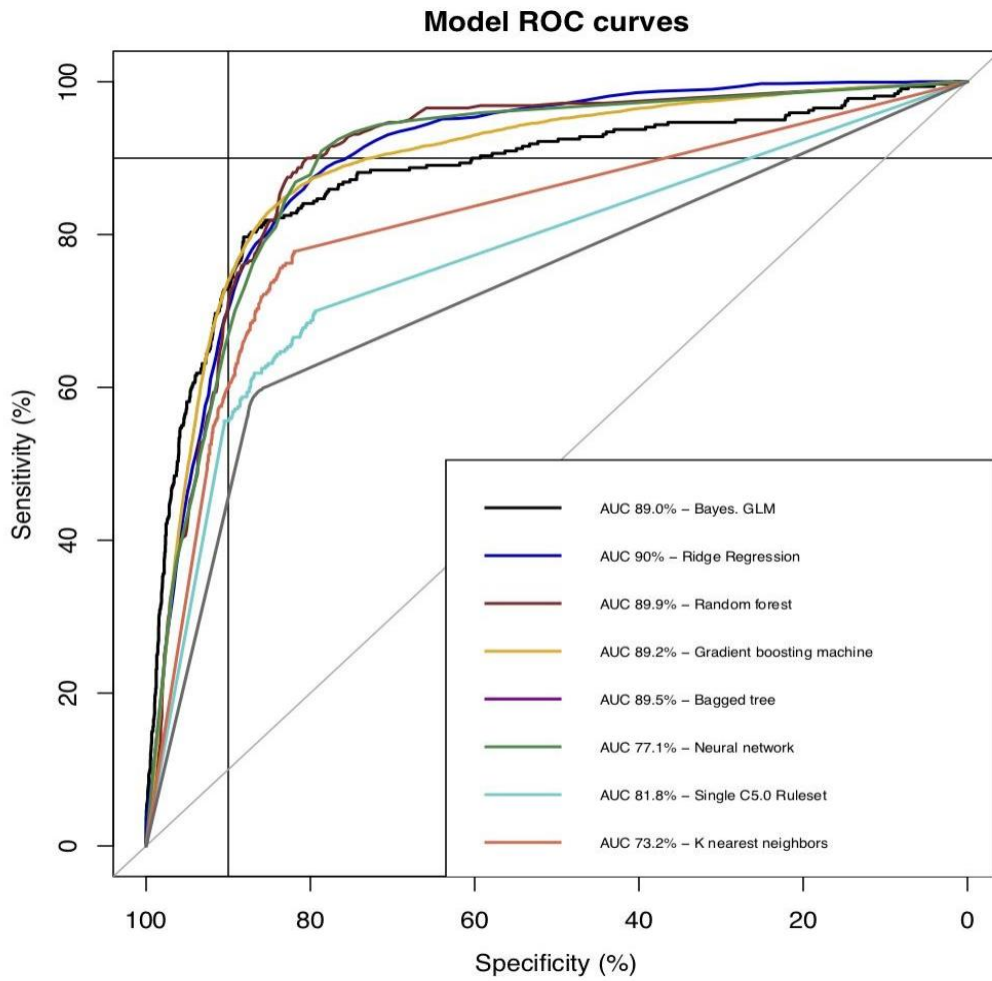
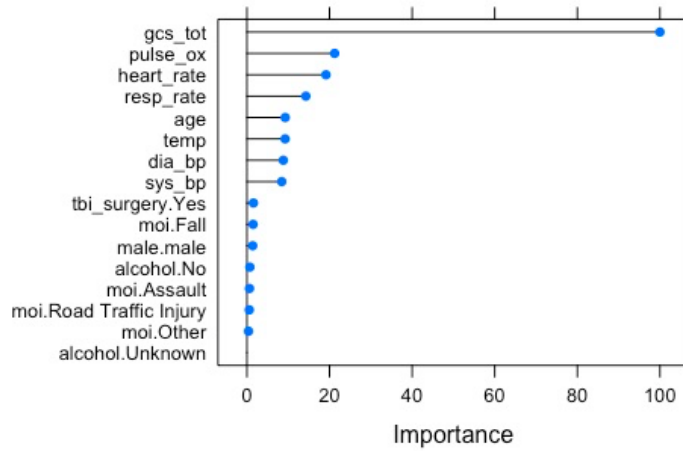


Figure 2: Rwanda Machine Learning model ROC curves

Importance of all Variables for Random Forest Model



Importance of all Variables for Ridge Regression Model

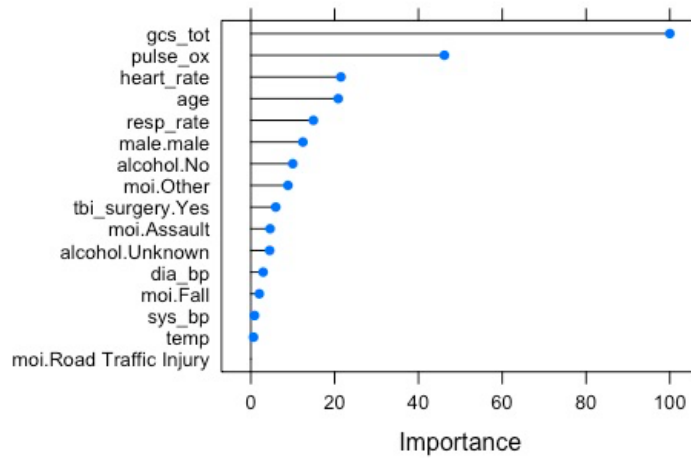


Figure 3: Variable importance for Rwanda TBI Random Forest and Ridge Regression models

3.2.2 External Validation of Tanzania developed machine learning models in Rwanda

A Tanzania TBI dataset was used to train and develop predictive models using 6 of the eight aforementioned algorithms (K-nearest neighbor, Random Forest, Neural Network, Bagged Tree, and Single C5.0 Ruleset). The models were subsequently examined for generalizable predictive value by external validation within the Rwanda TBI registry. Performance metrics for the externally validated models are presented in Table 3 and ROC curves for the models are shown in Figure 4 below. Calibration curves of the best performing models are provided in Figure 5. Variable importance for the best performing Tanzania developed models are previously reported in the literature (Rocha et al).

Finally, the Tanzania machine learning model predictions were combined into a single model prediction using the (1) mode of the predictions and (2) weighted averaging of the predictions, with larger weight given to the models with higher predictive ability. The Performance metrics for this combined model is shown in Table 4. The calibration and ROC curves for this model are shown in Figure 5.

Table 3: Performance Metrics of Tanzania predictive models externally validated in Rwanda

Performance Metrics of Tanzania Machine Learning models in Rwanda					
	Model	AUC	95 % Confidence Interval	Kappa Statistic	Brier Score
1	Bayesian GLM	89	84.4%-93.6%	0.516	0.161
2	Random Forest	91.3	88.0%-94.6%	0.465	0.154
3	Bagged Tree	87.6	83.1%-92.1%	0.437	0.176
4	Neural Network	89.9	85.6%-94.2%	0.523	0.161
5	K-nearest neighbor	69.3	64.6%-77.1%	0.344	0.316
6	Single C5.0 ruleset	79.3	73.1%-85.4%	0.432	0.17
7	Combined models	91.9	88.7%-95.1%	0.542	0.126

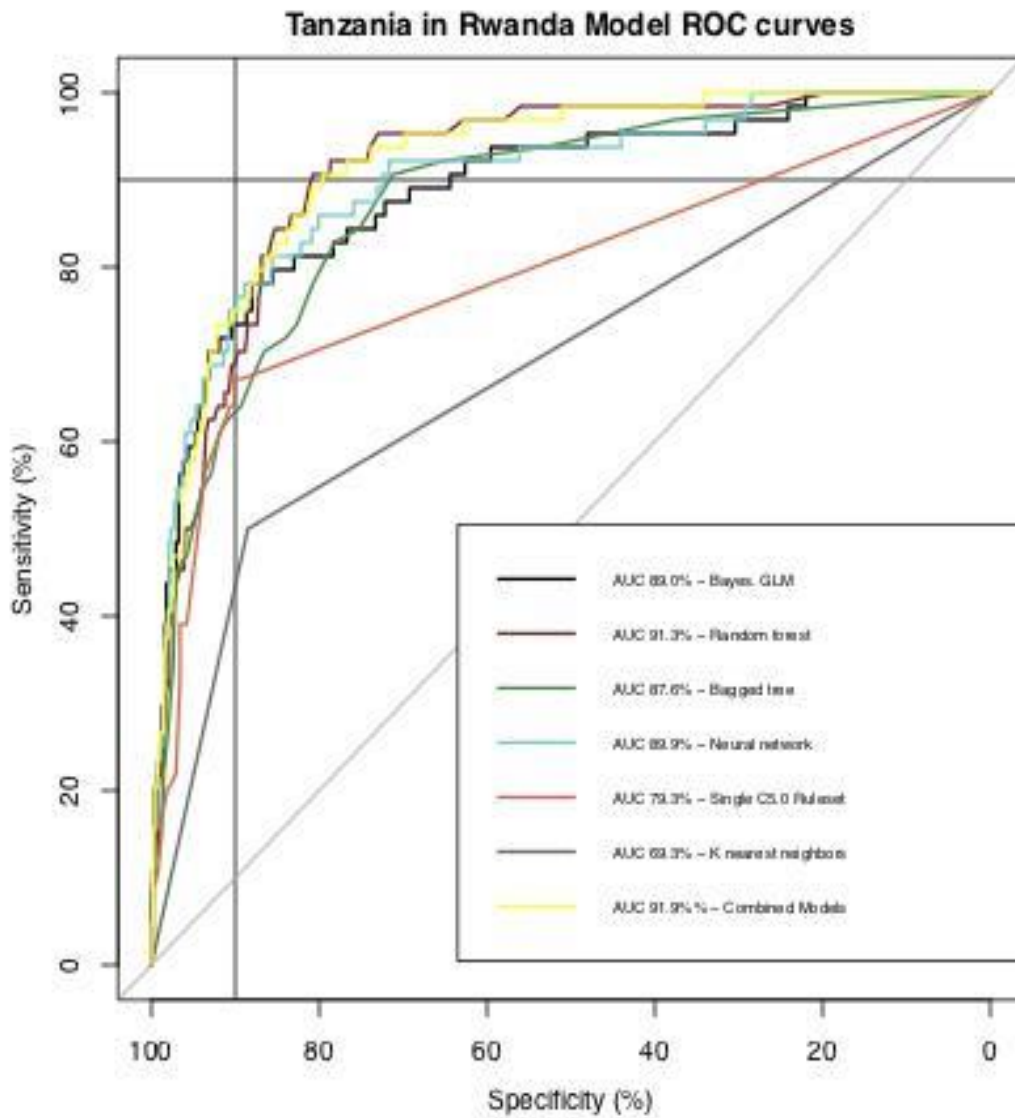


Figure 4: Tanzania in Rwanda model ROC curves

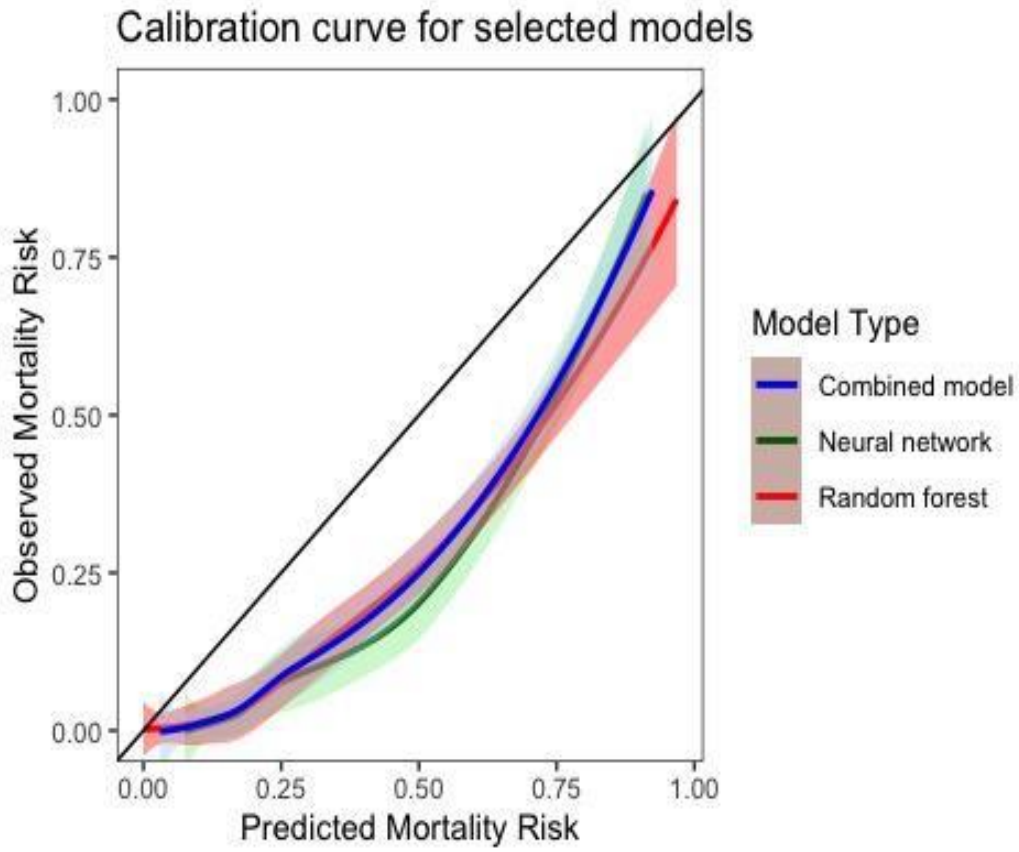


Figure 5: Calibration Curves for best performing Tanzania models in Rwanda

4. Discussion

Prognostic TBI models have been developed in many High Income Countries (HIC), with satisfactory generalizability across other high income settings.⁴⁰⁻⁴² However, these models cannot be simply implemented across widely diverse settings, especially low-resource and low-income settings. This is due to factors such as access to diagnostic tools, differences in availability of clinical indicators collected, number of clinical specialists, and timely hospital admission, among others.³²⁻³⁴ Due to the high burden of TBI mortality within LMICs, there is a necessity for relevant diagnostic tools to aid in early assessment of TBI and predict patient outcomes based on clinical data. Although previous studies have been done on TBI mortality in LMIC settings, there is still a gap in the literature regarding generalizability of predictive models across low resource settings.^{26,37} To our knowledge, this is the first study utilizing ensemble methods to model TBI outcomes in a LMIC setting.

Machine learning models

Model discrimination

The predictive models trained and implemented in Rwanda had satisfactory predictive ability, with the best performing model, Ridge Regression having an AUC of 90% (CI: 89.3%-90.7%). However, the models developed in Tanzania and subsequently used to predict outcomes within the Rwanda dataset showed similar predictive ability, with no significant difference between the AUCs of the best performing Random Forest

model, AUC 91.3 % (CI: 88.0%-94.6%) and the Rwanda Ridge Regression, AUC 90% (CI: 89.3%-90.7%). This demonstrates potential for generalizability of the Tanzania model in the Rwanda setting, as it had similar predictive power in an external dataset.

Furthermore, the results of the model which combines the Tanzania machine learning models strengthen the cross-setting predictive ability, with AUC 91.9% (CI: 88.7%-95.1%). Since it is unknown which specific machine learning models will perform better in a given dataset, a model which can combine the predictive ability of several algorithms lends itself highly to external use. The predictive ability of the combined model was similar to both the best performing Rwanda and Tanzania in Rwanda models, indicating it had satisfactory predictive ability and generalizability.

The implications of this finding are important in the context of low-resource settings. If TBI registries or tools such as CT scans are unavailable in the specific settings that diagnostic ability is desired, then the need for a model which can aid clinical monitoring and decision-making is crucial. These models were developed largely using indicators that were easily obtainable at triage or early in clinical evaluation, aiding in decision making by providing risk stratification based on inputted data. Although the capacity sites for both the Rwanda and Tanzania settings are both low resource settings, they differ in several ways. Although the Tanzania dataset had more robust data (larger sample size, more indicators collected), the Tanzania emergency department only had one physician who himself was not trained in neurosurgery. Furthermore, there was

only one CT for which access was limited at best. The Rwanda setting in contrast, had more healthcare providers including a neurosurgeon supported by two general practitioners. However, although the setting had access to more resources than the Tanzania site, the data was less robust (fewer patients and indicators collected). Although these settings differ, this strengthens the generalizability of the Tanzania predictive models. The machine learning models h similar predictive value in two LMIC setting which are largely dissimilar. Thus, based on the initial model results presented here, these models as prognostic tools seem adaptable within differing LMIC sites. If data collection is occurring in a site with similar capacity or collecting similar clinical data, these models may be able to predict patient outcome with a degree of certainty.

Calibration

The calibration curves of the Rwanda and Tanzania in Rwanda machine learning models show general overestimation of predicted mortality risk by the models, indicating conservative predictions. The combined machine learning model had a slightly better calibration curve, as seen in Figure 5, than the previously mentioned models, although similarly overestimated mortality risk. The overestimation of the models is most pronounced for patients in a moderate to moderate/severe risk category (50-75%). Since patients in a moderate risk range may deteriorate quickly and have an increased likelihood for a poor outcome, this overestimation may indicate the model can be used to monitor patients in this category (50-75% estimated risk).^{17,22} The implications

for clinical decision making indicate that patients who have between 50%-75% estimated risk of poor outcome should still be monitored closely as to avoid deterioration into a severe risk category. One potential reason that mis-calibration was most pronounced for moderate risk patients is due to the wide disparity in risks of patients who present. Most patients can be accurately identified upon arrival with indicators such as GCS as being either high risk or low risk; however, it is hard to ascertain what moderate risk signifies for TBI outcomes. Patients who present with low GCS and high GCS can be determined to have high likelihood and low likelihood of mortality, respectively. However, due to the nature of the GCS score and this general ability to distinguish between more extreme levels of risk, our model has higher levels of miscalibration at intermediate levels of risk.

4.1 Study strengths and limitations

Strengths of this study include access to a large TBI dataset and the use of indicators that can be easily collected at triage, a benefit for predictive ability as early as possible following patient arrival. Furthermore, the Tanzania models developed in this study, both individually, and the ensemble were able to be externally validated within the Rwanda dataset. Although the models showed satisfactory predictive power, a limitation of the Rwanda dataset is the lack of specific indicators that could aid in TBI prognosis. Perhaps the use of more robust data with access to indicators such as pupil reactivity and delays in time to care would increase both predictive ability and calibration of the models. The use of repeated clinical measures to monitor risk

prediction over time would additionally strengthen the model as well as improve clinical assistance to providers by analyzing longitudinal risk for TBI mortality.

However, further studies should be done evaluating further generalizability as well as feasibility studies exploring implementation of an app in an in-patient setting.

5. Conclusion

Thus, this study demonstrates the generalizability of a predictive model between two different LMIC settings. The similar predictive ability and calibration between the models developed in Rwanda and Tanzania indicates the potential use of such models, especially the ensemble, within mobile or web-based applications in low-resource settings or settings with a lack of electronic medical records. The use of indicators easily implementable at triage further strengthens the applicability of these predictive models, since they aid in clinical decision-making early in patient arrival. As literature and research increases on prognostic TBI modelling in LMICs, models with better predictive power can aid health care providers in reducing the disproportionate burden of TBI mortality.

Appendix A

Figure S1: Visualization of AUC performance as model ensembling increases for individual models, simple averaging (WOC), and optimal weighted averaging (SUMMA)

AUC Performance vs. Number of models in various ensemble methods

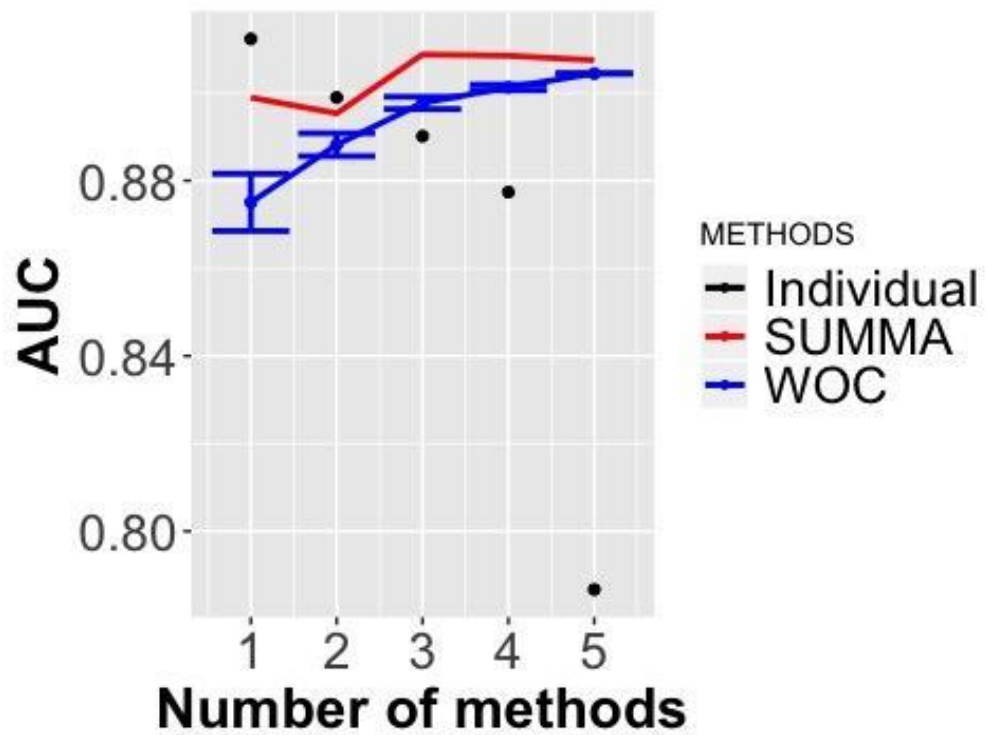
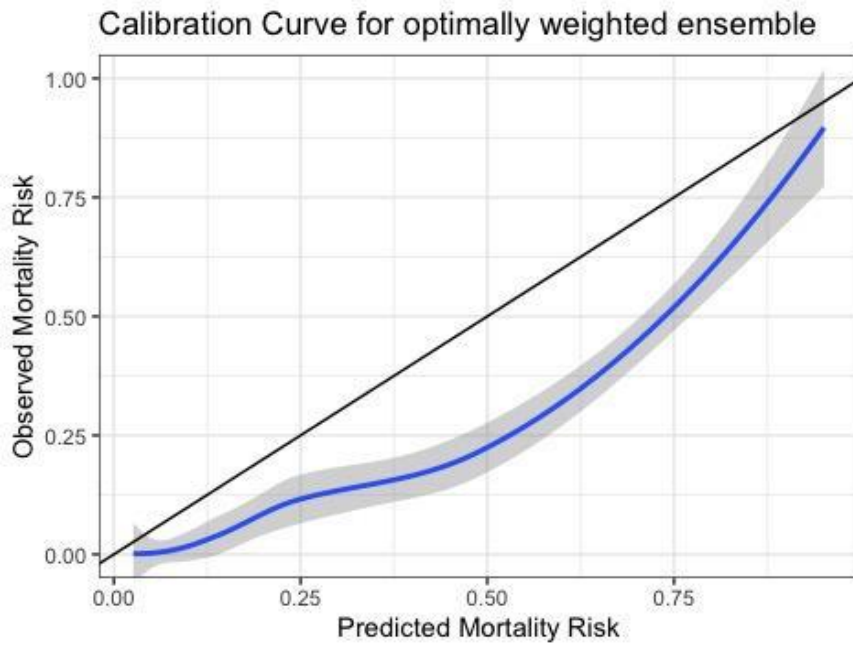
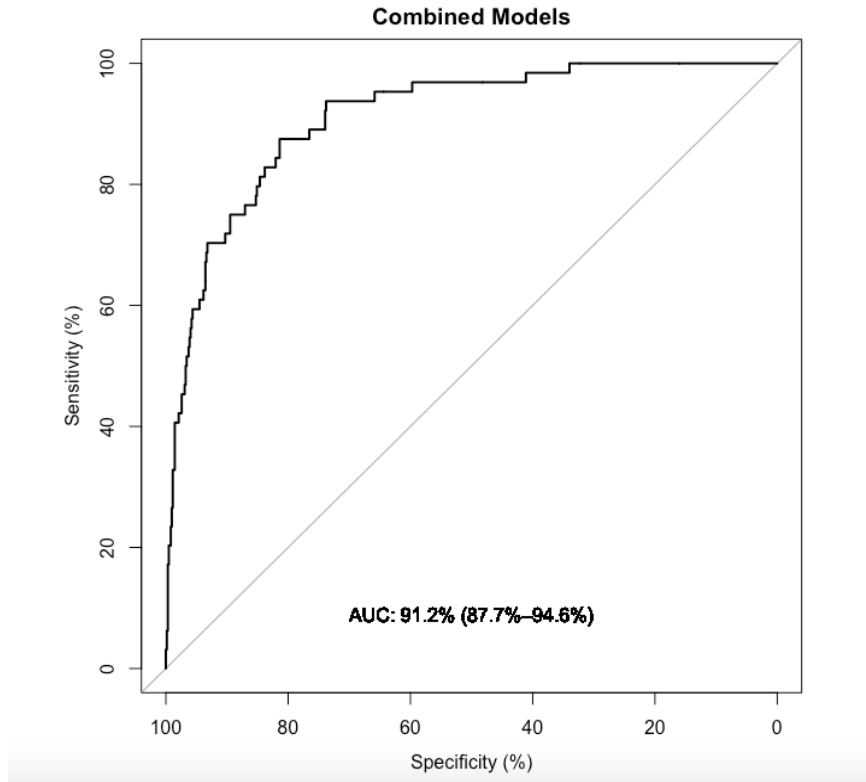


Figure S2: ROC/Calibration curves for algorithm-based Weighted Ensemble Model



References

1. De Silva MJ, Roberts I, Perel P, et al. Patient outcome after traumatic brain injury in high-, middle- and low-income countries: analysis of data on 8927 patients in 46 countries. *Int J Epidemiol.* 2009;38(2):452-458. doi:10.1093/ije/dyn189
2. Desautels T, Calvert J, Hoffman J, et al. Using Transfer Learning for Improved Mortality Prediction in a Data-Scarce Hospital Setting. *Biomed Inform Insights.* 2017;9. doi:10.1177/1178222617712994
3. Kodliwadmth HB, Koppad SN, Desai M, Badiger SP. Correlation of Glasgow outcome score to Glasgow coma score assessed at admission. *International Surgery Journal.* 2016;3(4):1959-1963. doi:10.18203/2349-2902.isj20163172
4. Twagirayezu E, Teteli R, Bonane A, Rugwizangoga E. Road traffic injuries at Kigali University Central Teaching Hospital, Rwanda. *East and Central African Journal of Surgery.* 2008;13(1):73-76-76.
5. Chaganti S, Plassard AJ, Wilson L, Smith MA, Patel MB, Landman BA. A Bayesian Framework for Early Risk Prediction in Traumatic Brain Injury. *Proc SPIE Int Soc Opt Eng.* 2016;9784. doi:10.1117/12.2217306
6. Chong S-L, Liu N, Barbier S, Ong MEH. Predictive modeling in pediatric traumatic brain injury using machine learning. *BMC Med Res Methodol.* 2015;15:22. doi:10.1186/s12874-015-0015-0
7. Dewan MC, Rattani A, Gupta S, et al. Estimating the global incidence of traumatic brain injury. *J Neurosurg.* April 2018:1-18. doi:10.3171/2017.10.JNS17352
8. Eaton J, Hanif AB, Grudziak J, Charles A. Epidemiology, Management, and Functional Outcomes of Traumatic Brain Injury in Sub-Saharan Africa. *World Neurosurg.* 2017;108:650-655. doi:10.1016/j.wneu.2017.09.084
9. Galgano M, Toshkezi G, Qiu X, Russell T, Chin L, Zhao L-R. Traumatic Brain Injury: Current Treatment Strategies and Future Endeavors. *Cell Transplant.* 2017;26(7):1118-1130. doi:10.1177/0963689717714102
10. De Silva MJ, Roberts I, Perel P, et al. Patient outcome after traumatic brain injury in high-, middle- and low-income countries: analysis of data on 8927 patients in 46 countries. *Int J Epidemiol.* 2009;38(2):452-458. doi:10.1093/ije/dyn189

11. Hale AT, Stonko DP, Brown A, et al. Machine-learning analysis outperforms conventional statistical models and CT classification systems in predicting 6-month outcomes in pediatric patients sustaining traumatic brain injury. *Neurosurg Focus*. 2018;45(5):E2. doi:10.3171/2018.8.FOCUS17773
12. Hyder AA, Wunderlich CA, Puvanachandra P, Gururaj G, Kobusingye OC. The impact of traumatic brain injuries: a global perspective. *NeuroRehabilitation*. 2007;22(5):341-353.
13. Iaccarino C, Carretta A, Nicolosi F, Morselli C. Epidemiology of severe traumatic brain injury. *J Neurosurg Sci*. 2018;62(5):535-541. doi:10.23736/S0390-5616.18.04532-0
14. Khan A, Prince M, Brayne C, Prina AM. Lifetime Prevalence and Factors Associated with Head Injury among Older People in Low and Middle Income Countries: A 10/66 Study. *PLoS One*. 2015;10(7):e0132229. doi:10.1371/journal.pone.0132229
15. Khellaf A, Khan DZ, Helmy A. Recent advances in traumatic brain injury. *J Neurol*. September 2019. doi:10.1007/s00415-019-09541-4
16. Krebs E, Gerardo CJ, Park LP, et al. Mortality-Associated Characteristics of Patients with Traumatic Brain Injury at the University Teaching Hospital of Kigali, Rwanda. *World Neurosurg*. 2017;102:571-582. doi:10.1016/j.wneu.2017.03.001
17. Kuo BJ, Vaca SD, Vissoci JRN, et al. A prospective neurosurgical registry evaluating the clinical care of traumatic brain injury patients presenting to Mulago National Referral Hospital in Uganda. *PLoS One*. 2017;12(10):e0182285. doi:10.1371/journal.pone.0182285
18. Menon DK, Zahed C. Prediction of outcome in severe traumatic brain injury. *Curr Opin Crit Care*. 2009;15(5):437-441. doi:10.1097/MCC.0b013e3283307a26
19. Patel A, Vieira MMC, Abraham J, et al. Quality of the Development of Traumatic Brain Injury Clinical Practice Guidelines: A Systematic Review. *PLoS One*. 2016;11(9):e0161554. doi:10.1371/journal.pone.0161554
20. Rudolfson N, Dewan MC, Park KB, Shrimel MG, Meara JG, Alkire BC. The economic consequences of neurosurgical disease in low- and middle-income countries. *J Neurosurg*. May 2018:1-8. doi:10.3171/2017.12.JNS17281

21. Staton CA, Msilanga D, Kiwango G, et al. A prospective registry evaluating the epidemiology and clinical care of traumatic brain injury patients presenting to a regional referral hospital in Moshi, Tanzania: challenges and the way forward. *Int J Inj Contr Saf Promot.* 2017;24(1):69-77. doi:10.1080/17457300.2015.1061562
21. Subaiya S, Roberts I, Komolafe E, Perel P. Predicting intracranial hemorrhage after traumatic brain injury in low and middle-income countries: a prognostic model based on a large, multi-center, international cohort. *BMC Emerg Med.* 2012;12:17. doi:10.1186/1471-227X-12-17
22. Hernandez Rocha TA, Elahi C, Cristina da Silva N, et al. A traumatic brain injury prognostic model to support in-hospital triage in a low-income country: a machine learning-based approach. *J Neurosurg.* May 2019:1-9. doi:10.3171/2019.2.JNS182098
23. Matsuo K, Aihara H, Nakai T, Morishita A, Tohma Y, Kohmura E. Machine Learning to Predict In-Hospital Morbidity and Mortality after Traumatic Brain Injury. *J Neurotrauma.* September 2019. doi:10.1089/neu.2018.6276
24. Molaei S, Korley FK, Reza Soroushmehr SM, et al. A machine learning based approach for identifying traumatic brain injury patients for whom a head CT scan can be avoided. *Conf Proc IEEE Eng Med Biol Soc.* 2016;2016:2258-2261. doi:10.1109/EMBC.2016.7591179
25. Rau C-S, Kuo P-J, Chien P-C, Huang C-Y, Hsieh H-Y, Hsieh C-H. Mortality prediction in patients with isolated moderate and severe traumatic brain injury using machine learning models. *PLoS One.* 2018;13(11):e0207192. doi:10.1371/journal.pone.0207192
26. van der Ploeg T, Nieboer D, Steyerberg EW. Modern modeling techniques had limited external validity in predicting mortality from traumatic brain injury. *J Clin Epidemiol.* 2016;78:83-89. doi:10.1016/j.jclinepi.2016.03.002
27. Perel P, Edwards P, Wentz R, Roberts I. Systematic review of prognostic models in traumatic brain injury. *BMC Med Inform Decis Mak.* 2006;6:38. doi:10.1186/1472-6947-6-38
28. Nguyen R, Fiest KM, McChesney J, et al. The International Incidence of Traumatic Brain Injury: A Systematic Review and Meta-Analysis. *Can J Neurol Sci.* 2016;43(6):774-785. doi:10.1017/cjn.2016.290

29. Systematic review of prognostic models in traumatic brain injury. - PubMed - NCBI. <https://www.ncbi.nlm.nih.gov/pubmed/17105661>. Accessed February 2, 2020.
30. Steyerberg EW, Mushkudiani N, Perel P, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. *PLoS Med.* 2008;5(8):e165; discussion e165. doi:10.1371/journal.pmed.0050165
31. Samanamalee S, Sigera PC, De Silva AP, et al. Traumatic brain injury (TBI) outcomes in an LMIC tertiary care centre and performance of trauma scores. *BMC Anesthesiol.* 2018;18. doi:10.1186/s12871-017-0463-7
32. Appenteng R, Nelp T, Abdelgadir J, et al. A systematic review and quality analysis of pediatric traumatic brain injury clinical practice guidelines. *PLOS ONE.* 2018;13(8):e0201550. doi:10.1371/journal.pone.0201550
33. Rubiano AM, Carney N, Chesnut R, Puyana JC. Global neurotrauma research challenges and opportunities. *Nature.* 2015;527(7578):S193-S197. doi:10.1038/nature16035
34. Kinyanjui B. Traumatic Brain Injury in Kenya: A Preliminary Review of the Literature. *SAGE Open.* 2016;6(1):2158244016638392. doi:10.1177/2158244016638392
35. Shahhosseini M, Hu G, Pham H. Optimizing Ensemble Weights and Hyperparameters of Machine Learning Models for Regression Problems. arXiv:190805287 [cs, stat]. January 2020. <http://arxiv.org/abs/1908.05287>. Accessed March 2, 2020.
36. Caruana R. Ensemble selection from libraries of models | Proceedings of the twenty-first international conference on Machine learning. https://dl.acm.org/doi/abs/10.1145/1015330.1015432?casa_token=RztSsYxbW34AAAAA%3AdhXgVViojFQaGBP-tepmY6-BeqQ7Q3xCfVH6EZZwlfzEAeqlOP3adJ9boiSI9IO7pZeZoHKS4N7Q2w. Accessed March 2, 2020.
37. Amorim RL, Oliveira LM, Malbouisson LM, et al. Prediction of Early TBI Mortality Using a Machine Learning Approach in a LMIC Population. *Front Neurol.* 2020;10. doi:10.3389/fneur.2019.01366
38. Dietterich TG. Ensemble Methods in Machine Learning. In: Multiple Classifier Systems. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2000:1-15. doi:10.1007/3-540-45014-9_1

39. Ahsen ME, Vogel R, Stolovitzky GA. R/PY-SUMMA: An R/Python Package for Unsupervised Ensemble Learning for Binary Classification Problems in Bioinformatics. *Journal of Computational Biology*. January 2020. doi:10.1089/cmb.2019.0348

40. Lingsma HF, Roozenbeek B, Steyerberg EW, Murray GD, Maas AIR. Early prognosis in traumatic brain injury: from prophecies to predictions. *Lancet Neurol*. 2010;9(5):543-554. doi:10.1016/S1474-4422(10)70065-X

41. Castaño-Leon AM, Lora D, Munarriz PM, et al. Predicting Outcomes after Severe and Moderate Traumatic Brain Injury: An External Validation of Impact and Crash Prognostic Models in a Large Spanish Cohort. *J Neurotrauma*. 2016;33(17):1598-1606. doi:10.1089/neu.2015.4182

42. Maeda Y, Ichikawa R, Misawa J, et al. External validation of the TRISS, CRASH, and IMPACT prognostic models in severe traumatic brain injury in Japan. *PLoS ONE*. 2019;14(8):e0221791. doi:10.1371/journal.pone.0221791