Author’s response to reviews

Title: Development, Implementation, and Prospective Validation of a Model to Predict 60-day End-of-Life in Hospitalized Adults upon Admission at Three Sites

Authors:
Vincent J Major (vincent.major@nyulangone.org)
Yindalon Aphinyanaphongs (yin.a@nyulangone.org)

Version: 2 Date: 13 Aug 2020

Author’s response to reviews:

See attached response pdf document for formatted version.

Development, Implementation, and Prospective Validation of a Model to Predict 60-day End-of-Life in Hospitalized Adults upon Admission at Three Sites

Authors:
Vincent J Major, ME, MPhil
Yindalon Aphinyanaphongs, MD PhD

General response:
We thank the reviewers and editor for their time and consideration of our revised manuscript. We appreciate the benefit of adding a comparison of several alternative algorithms and have repeated experiments to add these results to Table 2. The best model after cross-validation remains the random forest and the remainder of the manuscript is unchanged.

Reviewer 1:
The authors have address reviewers' comments sufficiently
Author Response:
We thank the reviewer for their time and consideration.

Reviewer 2:
I thank the authors for their efforts on addressing the comments. Most of my comments were well addressed. However, as an informatics journal, I feel like a comparison of several algorithms are still necessary to justify the use of random forest in the study. In the response, the authors mentioned that they did some experiments comparing logistic regression, gradient boosting (XgBoost). It would be nice if the authors could add these results in the manuscript.
Author Response:
We thank the reviewer for their second read. We have adapted our prior experiments to the same cross-validation folds and added AUROC and AUPRC results to Table 2.
Changes
Methods > Model Development

… An algorithm is needed that can learn which features—and which values within those features—are prognostic. Three classifiers were considered (eMethods): logistic regression with lasso regularization (15), XGBoost (16), and random forest (17). The later two are tree-based algorithms, known for their consistent performance on a variety of datasets (15) and within similar mortality work (10,16). …

Retrospective Modeling > Performance Within the Training Cohort

… AUROC and AUPRC within cross-validation from each model is reported in Table 2. The random forest classifier with 100 trees and a maximum depth of 1000 (eResults) outperformed the lasso regression model with marginal improvement over the XGBoost model and was selected as the final model (Table 2). …

Retrospective Modeling > Performance Within the Testing Cohort

Table 2. Model performance within cross-validation, applied to the testing set, and stratified by site.

<table>
<thead>
<tr>
<th>Model</th>
<th>Measure</th>
<th>Cohort</th>
<th>AUROC</th>
<th>AUPRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lasso regression</td>
<td>Training (Cross-validation)</td>
<td>Mean [min, max]</td>
<td>78.8 [78.0, 80.2]</td>
<td>21.0 [18.3, 22.0]</td>
</tr>
<tr>
<td>XGBoost</td>
<td>Training (Cross-validation)</td>
<td>Mean [min, max]</td>
<td>84.6 [83.8, 86.0]</td>
<td>25.7 [21.2, 27.4]</td>
</tr>
<tr>
<td>Random forest</td>
<td>Training (Cross-validation)</td>
<td>Mean [min, max]</td>
<td>86.9 [85.3, 87.7]</td>
<td>26.4</td>
</tr>
</tbody>
</table>
Testing (Bootstrapped)
Median
[95% CI]
87.2
[86.1, 88.2]
28.0
[25.0, 31.0]

Brooklyn
Median
[95% CI]
83.8
[81.9, 85.6]
26.6
[22.5, 31.0]

Non-Brooklyn
Median
[95% CI]
88.9
[87.5, 90.2]
30.1
[26.4, 33.7]

Supplementary Material > eMethods > Model Development
Three alternative models were considered: 1. logistic regression with lasso regularization implemented with the glmnet package in R (1), 2. XGBoost with a logistic objective implemented with the xgboost package in R (2), and 3. random forest implemented in R using the fest program (3). Empirical testing of model parameters was conducted within 5-fold cross-validation within the training cohort where patients (4) are partitioned into five groups and five models are learnt, each leaving out a different fifth for validation. Different parameters are compared by computing the areas under the receiver operating characteristic (AUROC) and precision-recall curves (AUPRC) within each cross-validation fold and the mean AUROC and AUPRC across folds.