Supplement to "Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model"

Benjamin Letham, Operations Research Center, Massachusetts Institute of Technology. bletham@mit.edu

Cynthia Rudin, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology. rudin@mit.edu

Tyler H. McCormick, Department of Statistics, Department of Sociology, Center for Statistics and the Social Sciences, University of Washington. tylermc@u.washington.edu

David Madigan, Department of Statistics, Columbia University. madigan@stat.columbia.edu

This supplement provides the Bayesian Rule Lists (BRL) point estimates (BRL-point) recovered from the five folds of cross-validation on the full stroke-prediction experiment in Figs 1-5. Figs 6 and 7 give BRL-point models from the female-only and male-only experiments, respectively.

if hemiplegia **and** age>60 **then** *stroke risk* 58.9% (53.8% - 63.8%)
else if cerebrovascular disorder **then** *stroke risk* 47.8% (44.8% - 50.7%)
else if transient ischaemic attack **then** *stroke risk* 23.8% (19.5% - 28.4%)
else if occlusion and stenosis of carotid artery without infarction **then** *stroke risk* 15.8% (12.2% - 19.6%)
else if altered state of consciousness **and** age>60 **then** *stroke risk* 16.0% (12.2% - 20.2%)
else if age≤70 **then** *stroke risk* 4.6% (3.9% - 5.4%)
else *stroke risk* 8.7% (7.9% - 9.6%)

Figure 1: Decision list for determining 1-year stroke risk following diagnosis of atrial fibrillation from patient medical history. The risk given is the mean of the posterior consequent, and in parentheses is the 95% credible interval. Obtained from the first of five folds of cross-validation.

if hemiplegia **and** cerebrovascular disorder **then** *stroke risk* 64.7% (59.6% - 69.6%)
else if cerebrovascular disorder **then** *stroke risk* 44.5% (41.6% - 47.5%)
else if hemiplegia **then** *stroke risk* 32.7% (23.8% - 42.2%)
else if congestive cardiac failure **and** hydrocodone **then** *stroke risk* 9.9% (8.4% - 11.5%)
else if transient ischaemic attack **then** *stroke risk* 30.5% (25.1% - 36.2%)
else if age>70 **then** *stroke risk* 9.1% (8.3% - 10.0%)
else *stroke risk* 4.0% (3.3% - 4.8%)

Figure 2: Stroke prediction decision list obtained from the second fold of cross-validation.

if hemiplegia **and** cerebrovascular disorder **then** *stroke risk* 61.7% (56.5% - 66.9%)
else if cerebrovascular disorder **then** *stroke risk* 44.8% (41.8% - 47.8%)
else if transient ischaemic attack **then** *stroke risk* 26.1% (21.7% - 30.7%)
else if occlusion and stenosis of carotid artery without infarction **then** *stroke risk* 15.2% (11.8% - 18.9%)
else if hemiplegia **then** *stroke risk* 37.8% (27.7% - 48.5%)
else if age≤60 **then** *stroke risk* 3.5% (2.8% - 4.4%)
else *stroke risk* 8.1% (7.4% - 8.8%)

Figure 3: Stroke prediction decision list obtained from the third fold of cross-validation.

if hemiplegia **and** cerebrovascular disorder **then** *stroke risk* 61.3% (56.2% - 66.3%)
else if cerebrovascular disorder **then** *stroke risk* 44.5% (41.5% - 47.5%)
else if sodium chloride **and** chronic obstructive pulmonary disease **then** *stroke risk* 10.6% (8.3% - 13.1%)
else if transient ischaemic attack **then** *stroke risk* 27.6% (22.8% - 32.7%)
else if hemiplegia **then** *stroke risk* 41.6% (30.9% - 52.7%)
else if age≤60 **then** *stroke risk* 3.2% (2.4% - 4.1%)
else *stroke risk* 8.2% (7.5% - 8.9%)

Figure 4: Stroke prediction decision list obtained from the fourth fold of cross-validation.

if hemiplegia **and** cerebrovascular disorder **then** *stroke risk* 64.5% (59.3% - 69.5%)
else if cerebrovascular disorder **then** *stroke risk* 44.2% (41.2% - 47.2%)
else if chronic obstructive pulmonary disease **and** chest pain **then** *stroke risk* 8.4% (7.2% - 9.8%)
else if transient ischaemic attack **then** *stroke risk* 30.2% (24.8% - 35.8%)
else if age≤60 **then** *stroke risk* 3.1% (2.3% - 4.0%)
else if hemiplegia **then** *stroke risk* 44.6% (32.8% - 56.7%)
else *stroke risk* 8.9% (8.1% - 9.7%)

Figure 5: Stroke prediction decision list obtained from the fifth fold of cross-validation.

if hemiplegia **then** *stroke risk* 59.0% (53.4% - 64.6%)
else if cerebrovascular disorder **then** *stroke risk* 44.7% (41.2% - 48.3%)
else if hypovolaemia **and** chest pain **then** *stroke risk* 14.6% (11.6% - 17.9%)
else if transient ischaemic attack **then** *stroke risk* 29.9% (24.0% - 36.2%)
else if age$\leq$70 **then** *stroke risk* 4.5% (3.6% - 5.5%)
else *stroke risk* 9.0% (8.0% - 10.0%)

Figure 6: Stroke prediction decision list obtained from the first fold of cross-validation on the females-only dataset.

if hemiplegia **and** age>70 **then** *stroke risk* 57.6% (47.8% - 67.1%)
else if transient ischaemic attack **and** chest pain **then** *stroke risk* 39.1% (31.6% - 46.9%)
else if occlusion and stenosis of carotid artery without infarction **and** coronary artery arteriosclerosis **then** *stroke risk* 21.1% (14.9% - 28.0%)
else if cerebrovascular disorder **then** *stroke risk* 49.6% (43.8% - 55.5%)
else *stroke risk* 6.8% (5.8% - 7.7%)

Figure 7: Stroke prediction decision list obtained from the first fold of cross-validation on the males-only dataset.