MPAFib: A Mortality Predictor for Emergency Department Patients with Atrial Fibrillation

LIM Su May *Student, Singapore Management University*
TAN Kar Way *Associate Professor, Singapore Management University*
Marcus ONG Eng Hock *Professor, Duke-NUS Medical School and Singapore General Hospital*
LIU Nan *Associate Professor, Duke-NUS Medical School*

SMU Singapore Management University — School of Computing and Information Systems
Duke-NUS Medical School
Singapore General Hospital — SingHealth

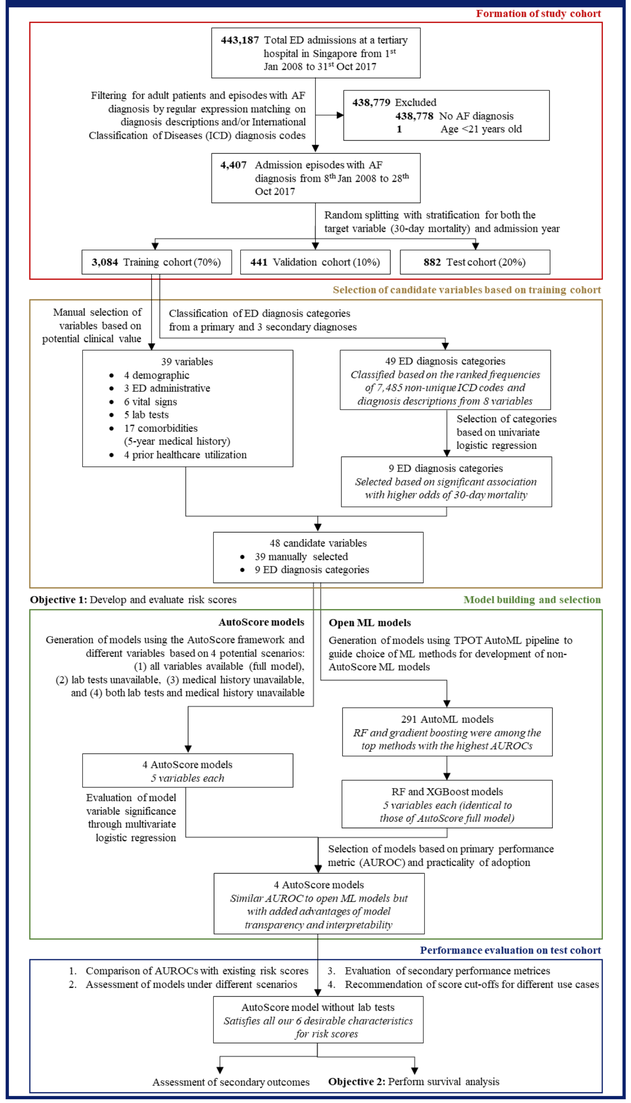
Background and Objectives

Atrial fibrillation (AF) is the most common cardiac arrhythmia with a pooled relative risk of 1.46 for all-cause mortality. Despite higher mortality risk, not all patients presenting with AF at the emergency department (ED) require inpatient care. In Singapore, more ED patients with AF are admitted to hospital compared to other developed countries, suggesting room for improving risk assessment and ED disposition. Existing risk stratification scores for AF patients are not ideal for use in the fast-paced ED environment as they often comprise numerous variables, some of which are not readily obtainable or objectively derived.

Objective 1: Develop and evaluate pragmatic 30-day mortality (primary outcome) risk scores using machine learning (ML) methods to identify low-risk AF patients for safe discharge while maintaining acceptable miss rates (e.g., <3%)

Objective 2: Perform survival analysis using to visualize the survival probabilities of low- and high-risk patients over 30 days

Overview of Work

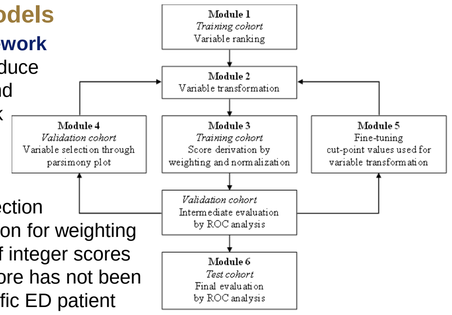


Model Building and Selection

AutoScore Models

AutoScore framework

- Designed to produce parsimonious and interpretable risk scores
- Random forest for variable ranking and selection
- Logistic regression for weighting and derivation of integer scores
- To date, AutoScore has not been applied on specific ED patient subsets of interest

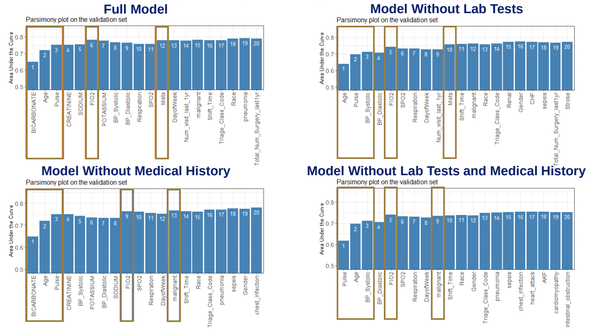


4 scenarios (1 model per scenario)

1. Lab test results and medical history are available (full model)
2. Only lab test results are unavailable (model without lab tests)
3. Only medical history is unavailable (model without medical history)
4. Both lab test results and medical history are unavailable (model without lab tests and medical history)

Variable ranking and selection

- 5 variables selected for each of the 4 AutoScore models based on incremental improvement to the area under the receiver operating characteristic (ROC) curve (AUROC) of the validation set



Scoring table

- Higher values correspond to higher predicted mortality risk
- Values are summed over all 5 variables to obtain patient's risk score
- All variables are objectively derived (i.e., not based on subjective judgement or manual assessment)

Variable	Full model	Without lab tests	Without medical history	Without lab tests and med history
Age (years)				
<65	0	0	0	0
65–74	7	7	6	7
75–84	13	12	12	13
85–94	22	21	22	20
≥95	24	24	25	24
Pulse (/min)				
<60	2	2	3	2
60–89	0	0	0	0
90–109	2	3	3	4
110–129	11	12	12	13
≥130	18	19	19	20
Systolic BP (mm Hg)				
<90	N/A	16	N/A	15
90–119	N/A	5	N/A	7
120–149	N/A	2	N/A	2
150–179	N/A	0	N/A	0
≥180	N/A	12	N/A	11
Oxygen supplementation				
No	0	0	0	0
Yes	16	14	16	15
Bicarbonate (mmol/L)				
<16	13	N/A	12	N/A
16–17	2	N/A	3	N/A
18–19	4	N/A	6	N/A
≥20	0	N/A	0	N/A
Metastatic solid tumor (5-year medical history)				
Absent	0	0	N/A	N/A
Present	29	28	N/A	N/A
Malignant neoplasm (ED diagnosis)				
Absent	N/A	N/A	0	0
Present	N/A	N/A	28	26

Mean scores (95% CI)

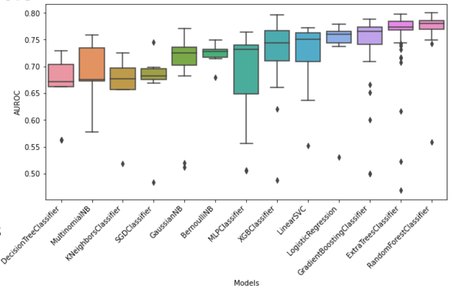
- Scores of those who died were higher than those who survived

Outcome	Full model	Without lab tests	Without medical history	Without lab tests and medical history
Without 30-day mortality	20.6 (19.6–21.6)	21.7 (20.8–22.6)	20.9 (19.9–21.9)	23.4 (22.5–24.4)
With 30-day mortality	39.9 (36.8–43.1)	39.0 (36.0–42.0)	39.8 (36.5–43.2)	50.0 (37.8–44.1)

Open ML Models

TPOT AutoML

- To guide the choice of other ML methods
- TPOT pipeline produced 291 models
- Random forests and gradient boosting models had the highest AUROCs



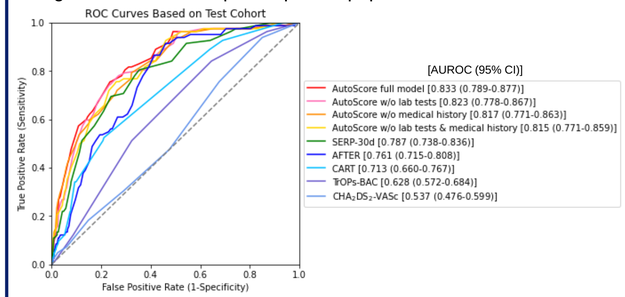
Model performance [AUROC (95% CI)]

- Random forest model: 0.834 (0.779–0.889)
- XGBoost model: 0.836 (0.781–0.891)
- Comparable performance to AutoScore models; hence, did not evaluate open ML models further since they do not offer additional advantages in terms of inherent transparency and interpretability

Performance Evaluation

Comparing AutoScore models and existing scores

- **AutoScore** models have similar ROCs with AUROCs of over 0.8
- AUROCs are not significantly different across AutoScore models, supporting the case for selecting a single model for implementation
- **SERP-30d**: AutoScore model trained with general ED population, same ED dataset, same target
- **AFTER** and **TrOPs-BAC**: same target, some variables not available
- **CART** and **CHA2DS2-VASc**: different target, all variables available
- Demonstrates value of developing risk scores that are tailored for target outcomes and specific patient populations of interest



3 use cases for recommending score cut-offs

1. "Rule out" for low-risk patients (false negative rate / miss rate <3%)
2. Optimal tradeoff (upper-left corner of ROC curve)
3. "Rule in" for high-risk patients that are likely to be admitted to ICU in situations with limited beds (specificity close to 95%)

Applying rule out use case

- Around at third of "low-risk" episodes could potentially have been consideration for safe discharge

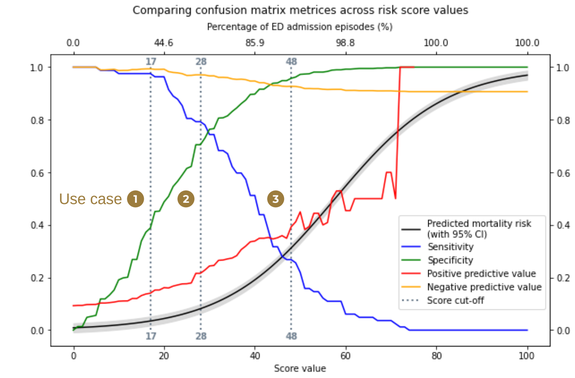
AutoScore model	"Low-risk" episodes under rule out use case
Full model	32.8%
Without lab tests	**35.5%**
Without medical history	31.7%
Without lab tests and medical history	35.3%

Recommended model (model without lab tests)

- Models without lab tests are inclusive (i.e., can be used for most patients) and can be applied earlier in the ED patient's journey
- If medical history is not available, patients/caregivers can simply be questioned to obtain relevant information under most circumstances
- Taken together, the model without lab tests can be realistically applied under various scenarios and is recommended
- Named the recommended model **MPAFib** (**M**ortality **P**redictor for patients with **A**trial **F**ibrillation), which also serves as a mnemonic for the model's variables (**M**etastatic tumour, **P**ulse, **A**ge, **Fi**O2, Systolic **b**lood pressure)

Recommended MPAFib score cut-offs for the 3 use cases and corresponding confusion matrix metrices

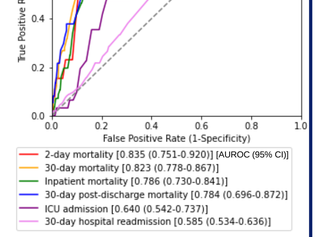
- The 3 use cases cover a range of values, allowing the user to determine which score cut-off to apply according to present need



Use case	Score cut-off	Predicted risk	% episodes above cut-off	Sensitivity	Specificity	Positive predictive value	Negative predictive value
1. Rule out	≥17	≥3.5%	64.5%	97.6% (93.9-100%)	38.9% (35.5-42.1%)	14.1% (13.3-14.9%)	99.4% (98.4-100%)
2. Optimal tradeoff	≥28	≥8.1%	34.1%	79.3% (70.7-87.8%)	70.5% (67.3-73.8%)	21.6% (19.1-24.2%)	97.1% (95.8-98.3%)
3. Rule in	≥48	≥31.2%	6.3%	26.8% (17.1-36.6%)	95.8% (94.3-97.1%)	39.6% (27.9-51.9%)	92.7% (91.9-93.7%)

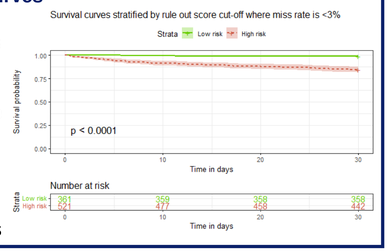
Secondary outcomes

- MPAFib performed well on all mortality-related secondary outcomes
- Although the model was trained to predict 30-day mortality, it may reliably predict other short-term mortality outcomes
- Lower AUROCs of ICU admission and 30-day hospital readmission may be due to inherently different patient characteristics since these outcomes are not mortality-related



Kaplan-Meier survival curves

- All 3 score cut-offs separate survival curves of "low-" and "high-risk" cases over 30 days (figure showing only rule out use case)
- The two risk classes were significantly different in terms of survival for all use cases



Conclusions and Contributions

- Demonstrated that the AutoScore framework can be effectively used to generate risk scores for specific patient populations
- Compared to existing risk scores, AUROCs of AutoScore models were among the highest reported while maintaining parsimony
- MPAFib can potentially flag around 35% of cases for considerations for safe discharge while maintaining a low miss rate of <3%
- MPAFib is a practical ED tool as it is simple, explainable, inclusive, uses only readily-obtainable variables that are objectively derived, and can potentially reduce unnecessary admissions for AF patients

Acknowledgements: I would like to thank my supervisors, Profs Tan, Ong, and Liu, and Xie Feng for their mentorship, help, and strong support