Amalgamation of automated Machine Learning and Ensemble approaches to achieve state-of-the-art post-Heart Failure Survival predictions

Ali Haider Bangash (1)(2)§, Ali Haider Shah (1)§, Arshiya Fatima (1), Saiqa Zehra (1), Syed Mohammad Mehmood Abbas (1), Hashir Fahim Khawaja (1), Muhammad Ashraf (3), Adil Baloch (1)

§ Ali Haider Bangash and Ali Haider Shah contributed equally.
(1) STMU Shifa College of Medicine, Islamabad, Pakistan; (2) Member, Working Group 3, COST Action Evidence-Based RESearch (EVBRES), Western Norway University of Applied Sciences, Bergen, Norway; (3) Rawalpindi Medical College, Rawalpindi Medical University, Rawalpindi, Pakistan

Background

Heart Failure (HF) is the underlying cause of a significant proportion of deaths attributed to Cardiovascular diseases with diabetes mellitus as its universally-recognized risk factor. Studies have attested to the significance of early diagnosis and subsequent initiation of treatment as a predictor of post-HF mortality (PHM). With this context, the current state-of-the-art (SOTA) for automated Machine Learning (aML) was explored to develop models that accurately, precisely and instantaneously predict PHM.

OBJECTIVE

Adoption of Automated Machine Learning to develop predictive models for PHM

RATIONALE

Easily deployed, robust machine learning predictive models that have the ability to provide instantaneous predictions can optimize risk stratification for optimal allocation of resources in a bid to decrease associated morbidity and mortality.

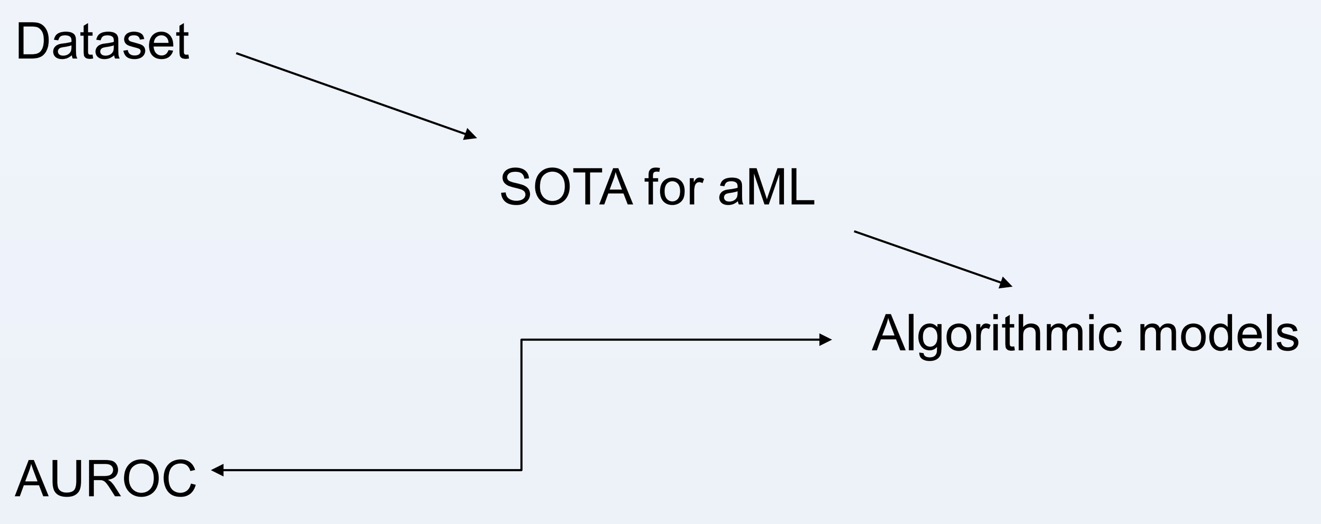
Training and validation dataset

Data was adopted as shared by Chicco D et al.(1) which was originally collected by Ahmad T et al.(2)

299 heart failure patients (35% women); age range: 40-95 years; All patients: Left ventricular systolic dysfunction and previous heart failures [New York Heart Association (NYHA) classification of the stages of heart failure: Classes III or IV].

Variables include CPK levels, Ejection fraction, serum creatinine and Hx of DM.

Protocol for models' development, training and assessment- Flow chart



RESULTS

- Upon training without the follow-up time (FUT) (scenario 1), an ensemble of RF and NN algorithms achieved the highest AUROC of 89% while predicting PHM.
- Upon training with the FUT (scenario 2), an ensemble of RF and Llr algorithms exhibited the highest AUROC of 88%.
- Upon training only with serum creatinine and ejection fraction variables without FUT (scenario 3), an ensemble of RF and Xgboost algorithms achieved an ACC of 80%.
- The presented models outperformed those developed by Chicco D et al. (1) in all aspects with the exception of F1 score in scenario 3.

Algorithm	MCC	F1 score	Accuracy	AUROC
Decision Tree	+0.542	0.7	0.785	0.768
Linear Logistic Regression	+0.556	0.703	0.803	0.813
Xgboost	+0.56	0.711	0.786	0.848
Neural Network	**+0.604**	**0.737**	**0.821**	0.822
Random Forest	+0.585	0.727	0.803	0.82
Best ensemble : Random Forest + Neural Network	+0.578	0.711	**0.821**	**0.89**

Table 1 PHM predictions without FUT: Characteristics of the algorithmic models

Algorithm	MCC	F1 score	Accuracy	AUROC
Decision Tree	+0.509	0.682	0.786	0.783
Linear Logistic Regression	+0.578	0.686	0.821	0.864
Xgboost	**+0.684**	0.789	**0.857**	0.83
Neural Network	+0.636	0.667	0.839	0.792
Random Forest	+0.621	**0.75**	0.839	0.879
Best ensemble: Random Forest + Linear Logistic Regression	+0.64	0.762	0.839	**0.881**

Table 2 PHM predictions with FUT: Characteristics of the algorithmic models

CONCLUSIONS

Adoption of the current SOTA for aML with subsequent incorporation of Ensemble approach (EA) yielded SOTA PHM predictive models. Such models have the potential to be incorporated into HF management protocols to achieve optimal and instantaneous risk stratification that would potentially translate into a significant decrease in the associated morbidity and mortality.

TAKE-HOME MESSAGE

By the incorporation of aML and EA, optimized and instantaneous PHM prediction can be achieved. This also underpins the superior statistical capabilities embedded into Artificial intelligence in general and Machine Learning in particular.

REFERENCES

1. Chicco D, Jurman G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Med Inform Decis Mak. 2020;20(1):16. Published 2020 Feb 3. doi:10.1186/s12911-020-1023-5
2. Ahmad T, Munir A, Bhatti SH, Aftab M, Raza MA. Survival analysis of heart failure patients: A case study. PLoS One. 2017;12(7):e0181001. Published 2017 Jul 20. doi:10.1371/journal.pone.0181001