Handling High Level of Censoring for Endovascular Aortic Repair Risk Prediction

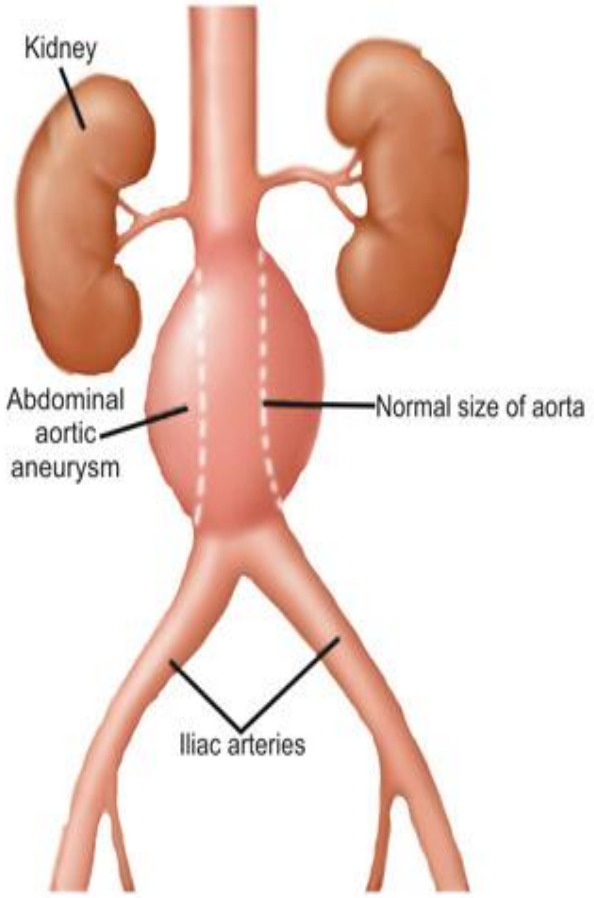
Presented by: Omneya Attallah

Agenda

- Aortic Aneurysm and Endovascular Aortic Aneurysm Repair.
- External Validation and EVAR Datasets.
- Survival Analysis and Censoring.
- Handling Censoring.
- Ensemble Classifiers.
- Survival Feature Selection (FS).
- Proposed Approach to Perform FS for Survival Data.
  - Uncensoring Step.
  - Hybrid Feature Selection.
- Results.
- Conclusions.

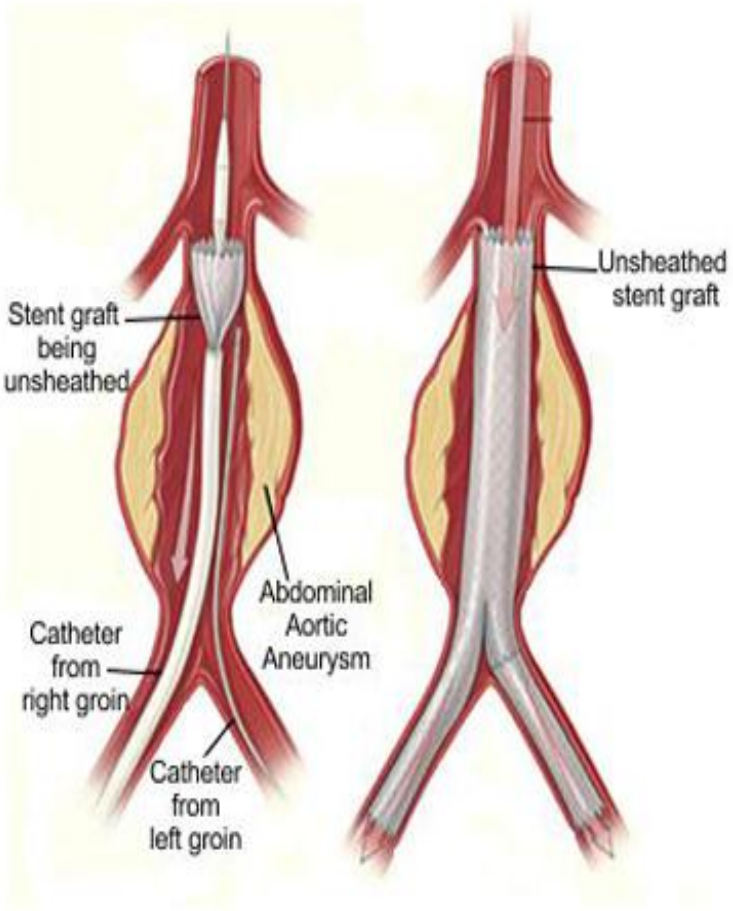
Aortic Aneurysm

- Aorta is the biggest blood vessel in the body.
- Sometimes, part of aorta may balloon and form an aneurysm.
- It increases the risk of aorta breach which could lead to death due to internal bleeding.
- Endovascular aortic aneurysm repair and open surgery are the common solutions.
- EVAR carries significantly lower preoperative morbidity and mortality risks than open repair surgery.
- EVAR enables shorter length of hospital stay and faster patient's recovery rates than open repair surgery.



Endovascular Aortic Aneurysm Repair (EVAR)

- The surveillance after EVAR is obligatory, but is considered to be expensive, varied, and poorly-calibrated.
- Patients may be exposed to radiations and contrast nephropathy as a result of frequent surveillance.
- By specifying which patients are more likely to require the EVAR re-intervention after 5 years (high risk patients) and which are less likely(low risk patients), a cost-effective and risk-stratified surveillance system could be achieved.



External validation

- In order to produce a clinically validated model to be used in medical prediction, it should achieve acceptable performance on a dataset not used in the training process.
- However external validation is essential.
- It tests the capability of the predictive model to establish a corresponding performance on other patients from another center who have dissimilar population from the one used to construct the model.
- The opportunity to cross check the model prediction upon different centers strengthens the results and study outcomes.
- Therefore, in this study two EVAR datasets were collected from two vascular centers in the UK (St George and Leicester vascular institutes respectively)

EVAR Datasets

- Both centers have 45 variables.
- Patients containing missing values were removed.
- The number of patients after removal is 457 and 286 for center 1 and 2 respectively.
- Patients that actually experienced the EVAR during the observations are 42 and 26 for center 1 and 2 respectively.
- The main problem of these datasets is high level of censoring.

Survival Analysis and Censoring

- Generally, medical survival data has a time variable which can indicate either the time when a certain event of interest occurs such as mortality or the time when a disease recurs.
- It may also describe the time till last follow up for a patients who did not complete observation (censored patients) due to various reasons.
- Survival analysis techniques analyze this unique data without ignoring any instance in the dataset even if the event had not occurred (censored).

Survival Analysis Methods

- They can be classified into:
  - Standard statistical survival analysis.
    - Such as: Kaplan Meier (KM), cox's proportional hazard model, and Accelerated failure time (AFT)
  - ML survival analysis techniques which are preferred due to:
    - Their capability to deal with complex and non linear relations between covariates.
    - Considering the variable information in constructing the predictive model not like nonparametric method (KM).
    - Do not assume a hazard distribution to construct the predictive model like parametric methods (AFT) (hassle of choosing the best hazard distribution to fit the model is avoided).
- However, ML techniques cannot be used directly with censored survival data.

Handling Censoring

- Deleting censored patients or considering that the event of interest will definitely not happen to all of them (zero target output).
- Using only uncensored patients to build the predictive model.
- Use Kaplan Meier (KM) survival method to weigh the output probability estimates of censored patients.
- Splitting observation time interval and repeating censored patients to all intervals.

Ensemble Classifiers

- Ensembles classifiers is preferred over individual classifiers, as it merges the outputs of multiple classifiers which usually improves prediction results.
- In health studies, it corresponds to asking several doctors diagnosing opinions in order to come up to a more confident and accurate last decision

Survival Feature Selection (FS)

- FS is important in medical field.
- It process becomes complicated with the presence of censored survival data.
- Most survival FS techniques are based on Cox's model
- Most of the work done using ML techniques (MLT) for censored data did not perform FS.
- Fewer work was done for survival FS using MLT.
- Limitation of these methods: The way they handled censoring by: deleting, ignoring, repeating, using only uncensored patients, or considering censored patients as event free (can't be used with high censoring)

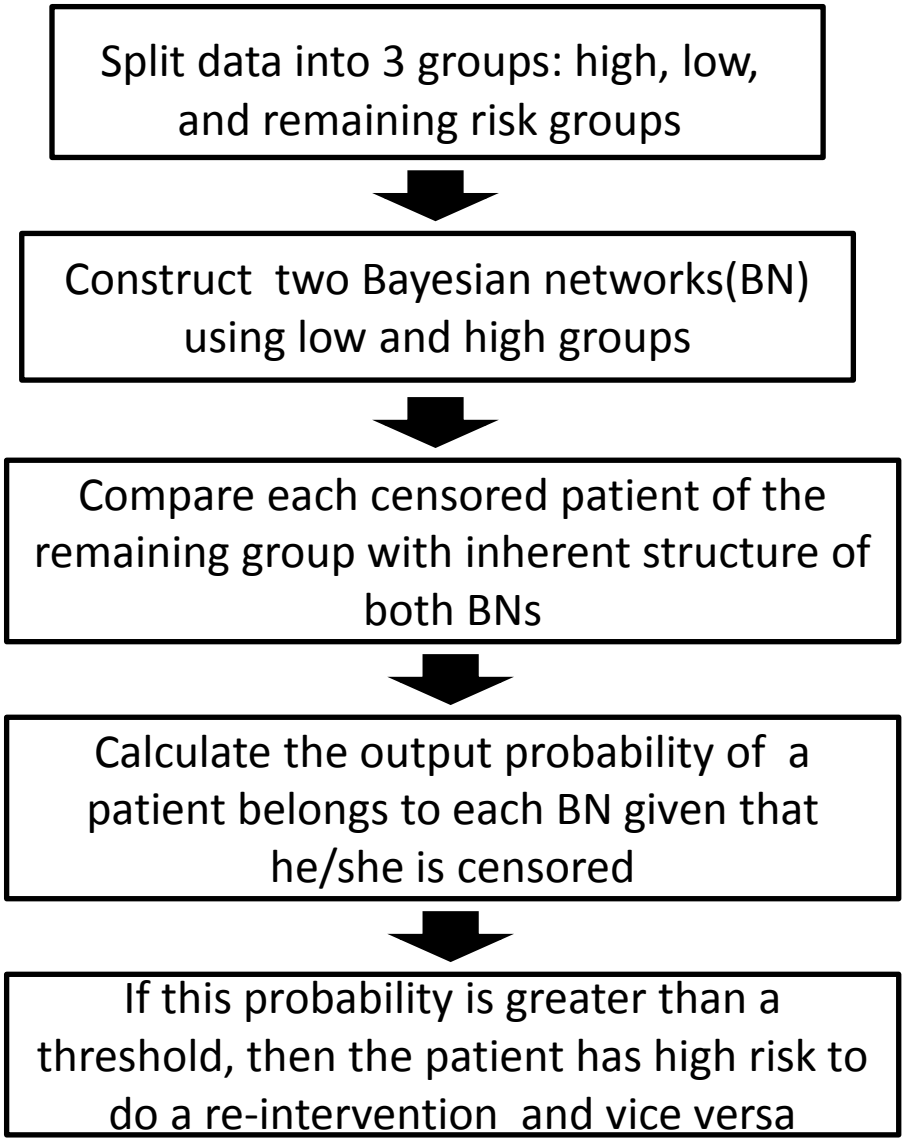
Proposed Approach to Perform FS for Survival Data

- The proposed algorithm consists of two steps which are the : Uncensoring and Hybrid FS steps.
- It is an extension to the our previous work made to predict the risk of intervention after 5 years of EVAR using a single Neural networks.
- However in this paper, ensemble ANNs was used instead and its performance was compared with individual ANN.
- Moreover, it was merged with hybrid FS process in the second step in order to select a reduced number of variables that improve prediction results

Uncensoring Step

- This step presents a new solution to high censoring .
- It uses Bayesian network in order to deal with high censoring problem located in EVAR datasets.
- It does not depend on deleting, ignoring, repeating, weighting , or considering censored patients as event free.

Uncensoring Step



Results of Uncensoring using Individual ANN and Ensemble ANNs

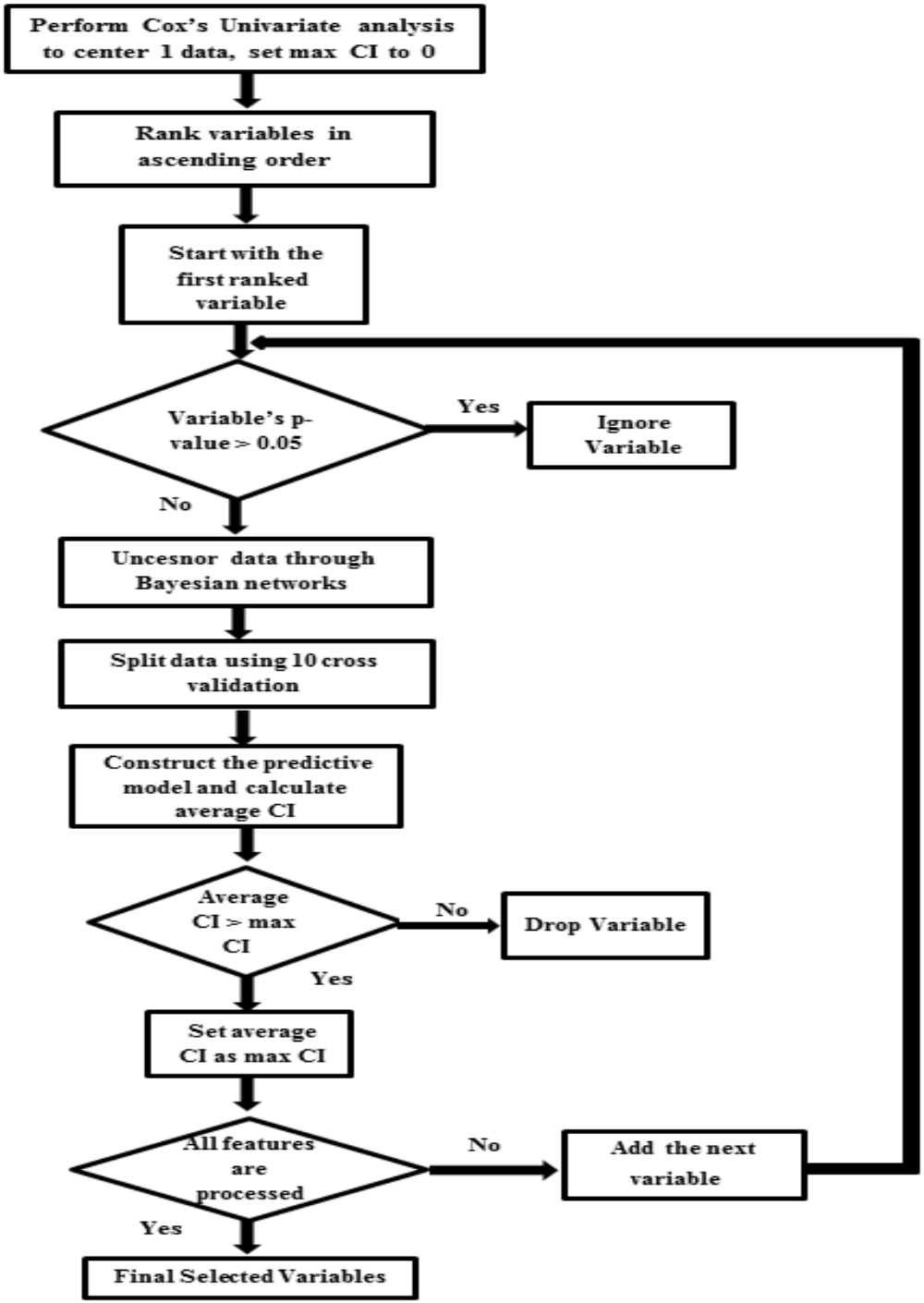
Center 1

(a) Before uncensoring (Center 1)			
Class	True	False	ROC
High risk (positive)	0.095	0.905	0.543
Low risk (negative)	0.929	0.071	
(b) After uncensoring using individual ANN			
High risk (positive)	0.654	0.346	0.832
Low risk(negative)	0.859	0.141	
(c) After uncensoring using Ensemble ANNs			
High risk (positive)	0.667	0.333	0.857
Low risk (negative)	0.889	0.111	

Center 2

(a) Before uncensoring (Center 2)			
Class	True	False	ROC
High risk (positive)	0.038	0.962	0.431
Low risk (negative)	0.95	0.095	
(b) After uncensoring using individual ANN			
High risk (positive)	0.63	0.37	0.764
Low risk(negative)	0.786	0.214	
(c) After uncensoring using Ensemble ANNs			
High risk (positive)	0.65	0.35	0.786
Low risk (negative)	0.805	0.195	

Hybrid Feature Selection



Results of the Proposed Hybrid FS

Center 1

Classifier	FS method	Model Size	CI(SD)	Sensitivity
ANN	Hybrid FS	7	0.7 (0.0487)	0.8852
Ensemble ANNs	Hybrid FS	7	0.6912 (0.0496)	0.8798
Cox Regression	Lasso FS	6	0.7191 (0.0448)	0.6905
	AIC FS	4	0.7043 (0.0484)	0.6905

Center 2

Classifier	FS method	Model Size	CI (SD)	Sensitivity
ANN	Hybrid FS	7	0.6473 (0.0545)	0.6923
Ensemble ANNs	Hybrid FS	7	0.6507 (0.0643)	0.6923
Cox Regression	LASSO FS	6	0.6134 (0.0850)	0.5385
	AIC FS	4	0.6070 (0.0845)	0.5385

Conclusions

Conclusions

- A Hybrid feature selection technique was proposed to build and validate a predictive reduced model for EVAR survival data.
- It was able to deal with the high level of censoring.
- The results indicated that the machine learning techniques performed better than statistical survival models such as Cox's models.
- It also showed that the ensemble neural networks can outperform the single neural network model in predicting the risk of endovascular aortic re-intervention after 5 years of surgery.
- The reduced predictive model will enable clinicians to put a future follow up and surveillance plan for those having higher chance of re-intervention as they need more regular monitoring schedule.

Thank you