

Comparing the Kattan-type nomogram to random forest models to predict post-prostatectomy pathology

John Kang, MD, PhD¹

Christopher Doucette, MD¹

Issam El Naqa, PhD²

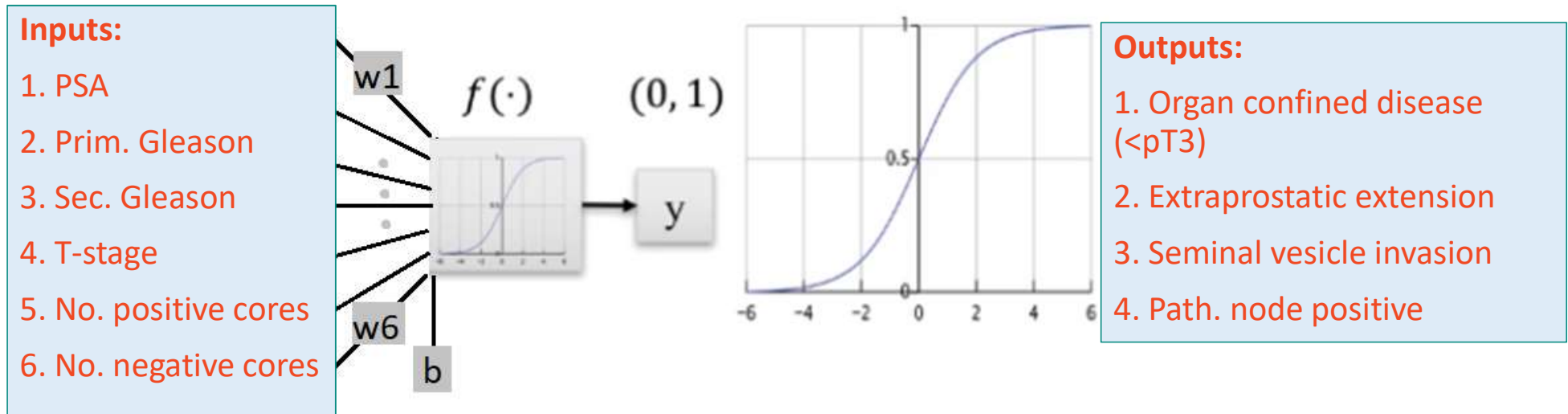
Hong Zhang, MD, PhD¹

¹University of Rochester Medical Center; ²University of Michigan

Purpose

- **Provide more upfront information to patients with prostate cancer regarding treatment options**
- Compare random forests to the Kattan-type model in prediction of adverse pathology in clinically intermediate risk prostate cancer
 - n=1560 radical prostatectomy patients from 2005-2015
 - Pathology report outcomes
 - Organ confined disease (<pT3): 62.6%
 - Extraprostatic extension: 37.1%
 - Seminal vesicle invasion 6.9%
 - Pathologic node positive 4.2%

Kattan-type nomogram uses logistic regression on human-selected variables



$$\hat{y} = \text{logistic}(\hat{b} + \hat{w}_1 \cdot x_1 + \cdots \hat{w}_n \cdot x_n)$$
$$= \frac{1}{1 + \exp[-(\hat{b} + \hat{w}_1 \cdot x_1 + \cdots \hat{w}_n \cdot x_n)]}$$

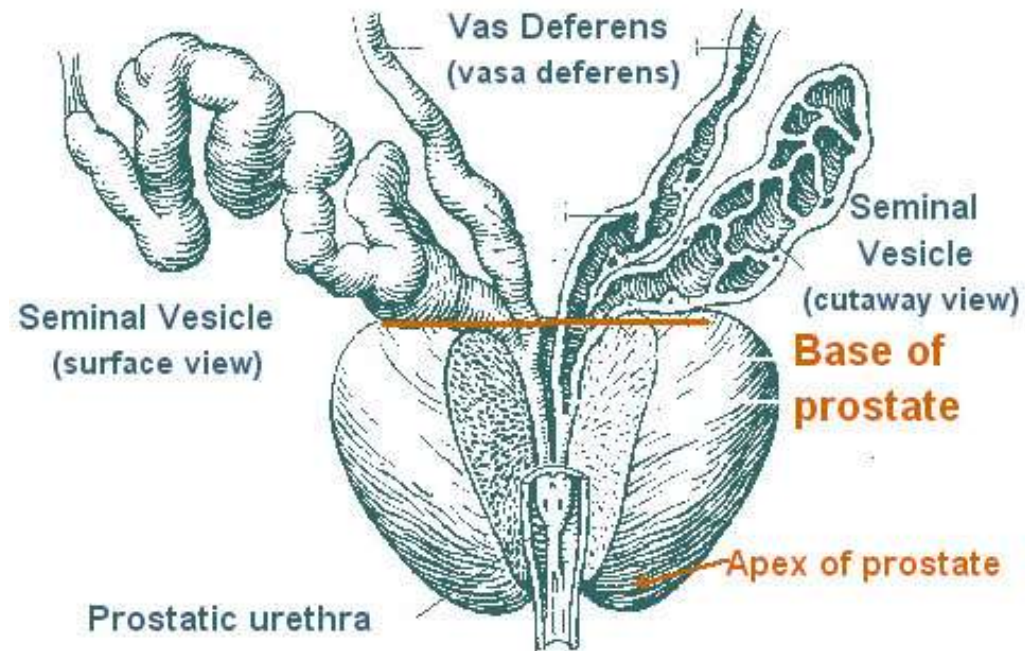
Kattan-type nomogram: https://www.mskcc.org/nomograms/prostate/pre_op

Kattan et al., "A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer." JNCI 1998.

Fig: "Applied Machine Learning in Python". Course by University of Michigan. <https://www.coursera.org/lecture/python-machine-learning/logistic-regression-bEtYh>

How can we improve on Kattan-type model?

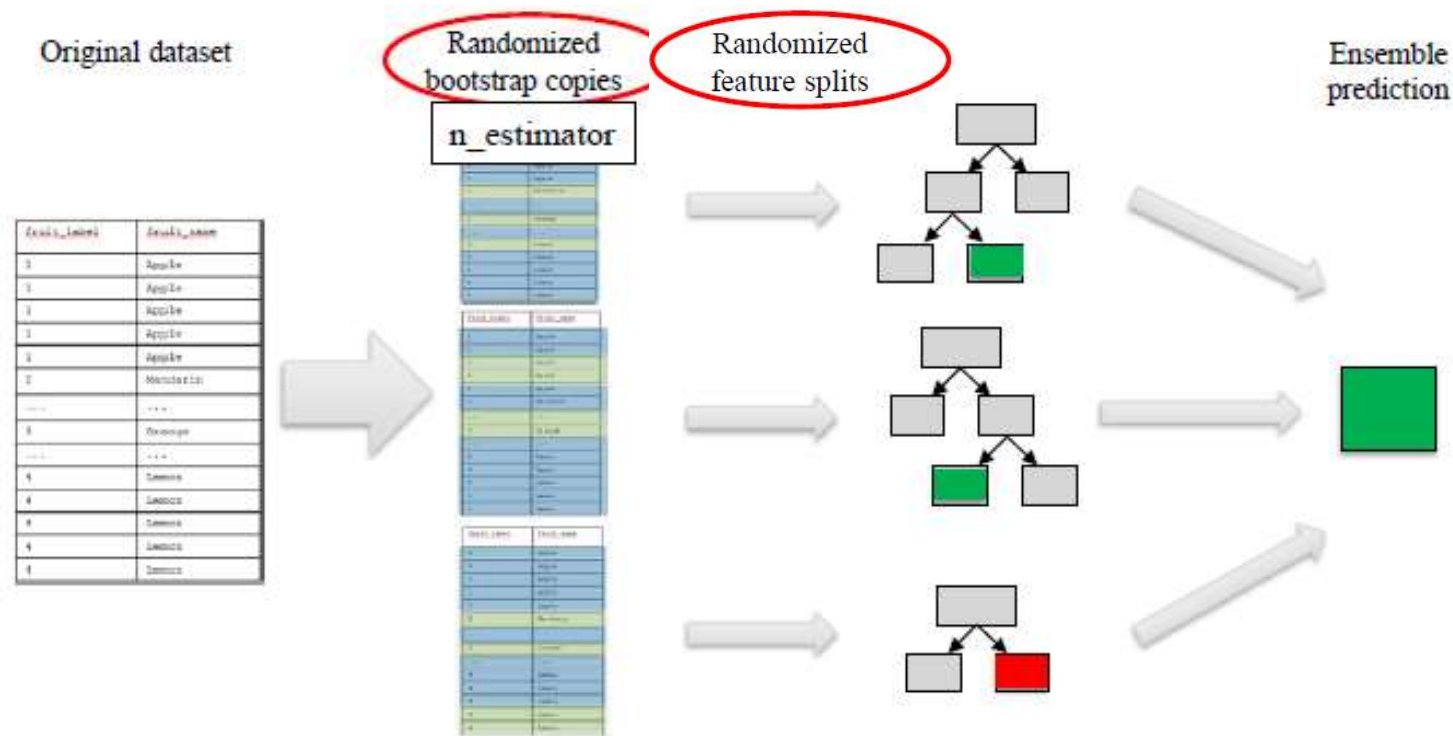
- Adverse pathology may be linked to location-specific biopsy data
 - Prostate “capsule” absent at base and poorly defined at apex¹
- Ensemble tree models may have superior performance



<https://training.seer.cancer.gov/prostate/anatomy/>

¹ <https://www.auanet.org/education/auauniversity/education-products-and-resources/pathology-for-urologists/normal-histology-and-important-histo-anatomic-structures/prostatic-capsule-and-periprostatic-tissues>

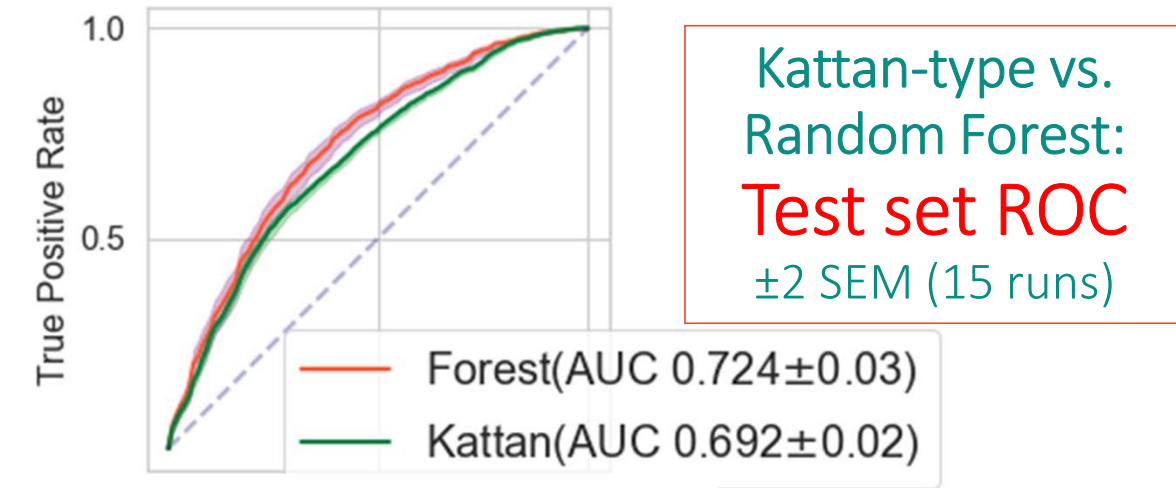
Random Forest: two points of randomness



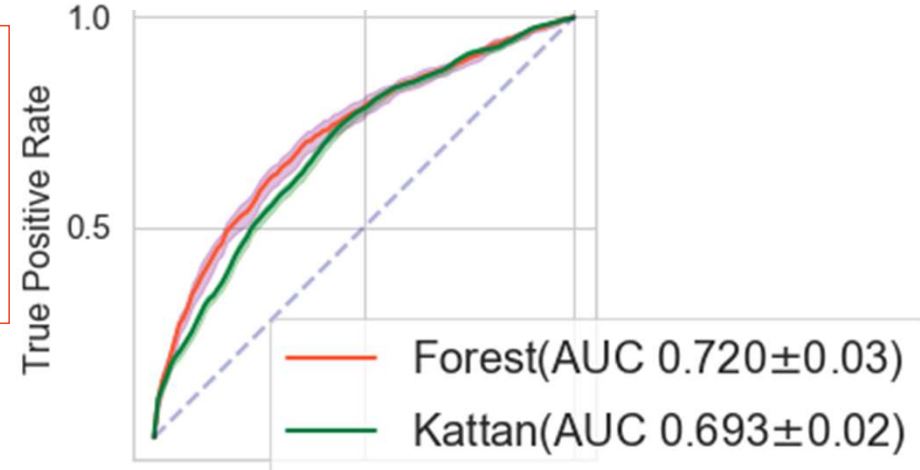
Breiman. Random Forests. Machine Learning 2001.

Fig. "Applied Machine Learning in Python". Course by University of Michigan <https://www.coursera.org/lecture/python-machine-learning/random-forests-IF9QN>

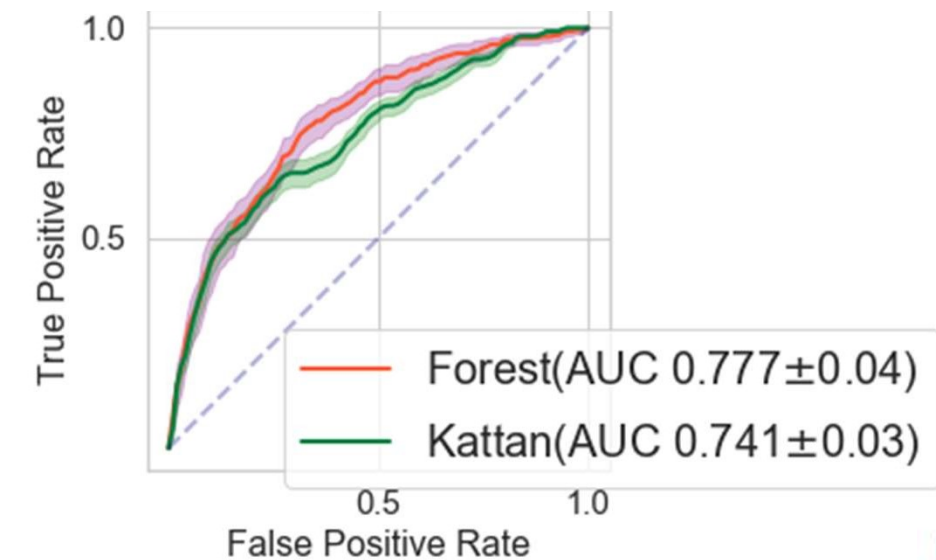
Organ confined disease (<pT3)



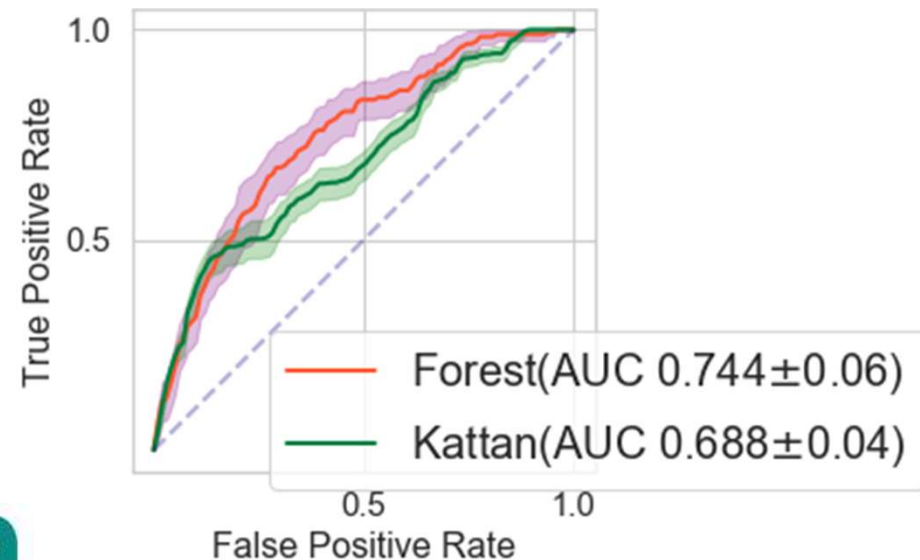
Extraprostatic extension



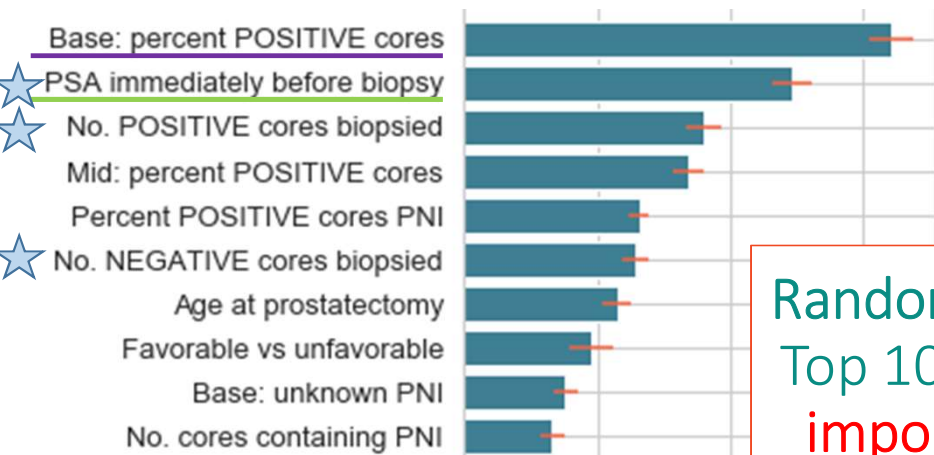
Seminal vesicle invasion



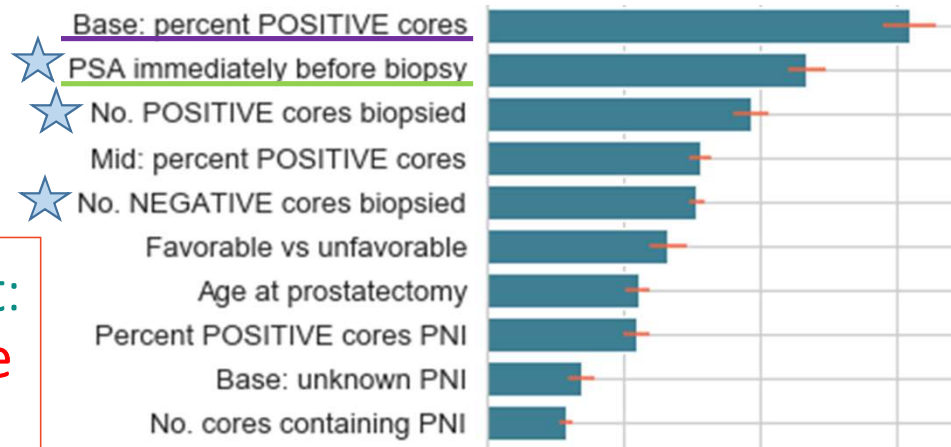
Pathologic node positive (pN+)



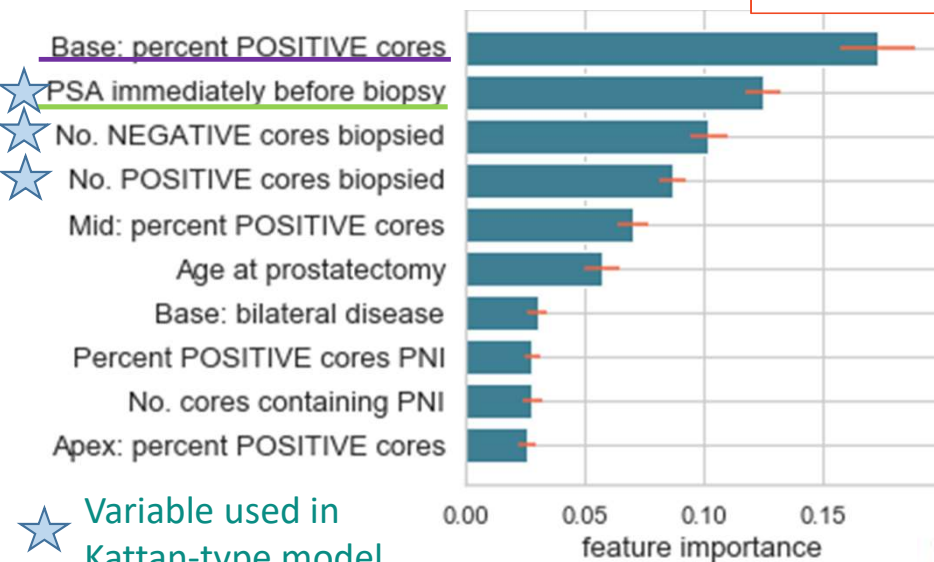
Organ confined disease (<pT3)



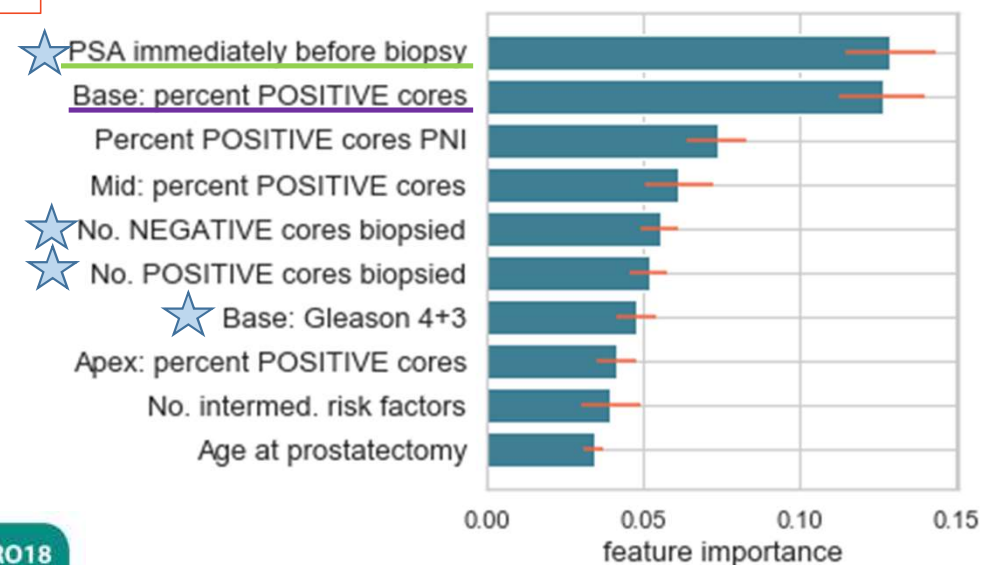
Extraprostatic extension



Seminal vesicle invasion



Pathologic node positive (pN+)



Random forest:
Top 10 feature
importances
±2 SEM (15 runs)

Variable used in
Kattan-type model

Conclusions

- Including location specific biopsy data in a random forest may improve prediction of pathologic upstaging vs. Kattan-type model
 - Burden of disease at the base seems to be of similar importance as PSA
- Low event rates for seminal vesicle invasion and pathologic node positivity are a limitation
- Future work
 - See if results generalize to other methods
 - See if results hold up with a more sparse feature set
 - Toxicity prediction using genomics

Research team



Christopher
Doucette, MD



Hong Zhang,
MD, PhD



Issam El Naqa, PhD

