Machine Learning in **Medicine**

Muhammad Owais Bawany
Felix Singerman

Machine Learning in Medicine

Author of this research paper is:

Rahul C. Deo, MD, PHD

Agenda

1. Machine Learning: Supervised vs Unsupervised Learning
2. The Learning problem
3. Illustrative Examples of Machine Learning (Supervised)
   - Supervised Learning - learning from Forests & Trees
   - C-Path
   - Attractor metagenes in Cancer and bake-offs in ML
4. Illustrative Examples of Machine Learning (Supervised)
   - in HFpEF - Towards Precision Medicine?
5. Discussion - summary

Machine Learning

- Scientific discipline that focuses on how computers learn from data

- Subclassified into categories such as supervised and unsupervised learning
  - Tasks physicians can already do well vs learning those where physicians have had limited success

Supervised vs Unsupervised Learning

		Supervised	Unsupervised	Comments
1	Predicting known output or target	✓	✗	• Unsupervised Learning is used for Analysis
2	Classification and regression	✓	✗	• Unsupervised Learning - Clustering, density estimation and dimensionality reduction
3	Input data is labeled	✓	✗	
4	Uses Training dataset	✓	✗	• Uses just input dataset

THE LEARNING PROBLEM

Machine learning algorithms have had limited presence in clinical practice

Supervised Learning Problem

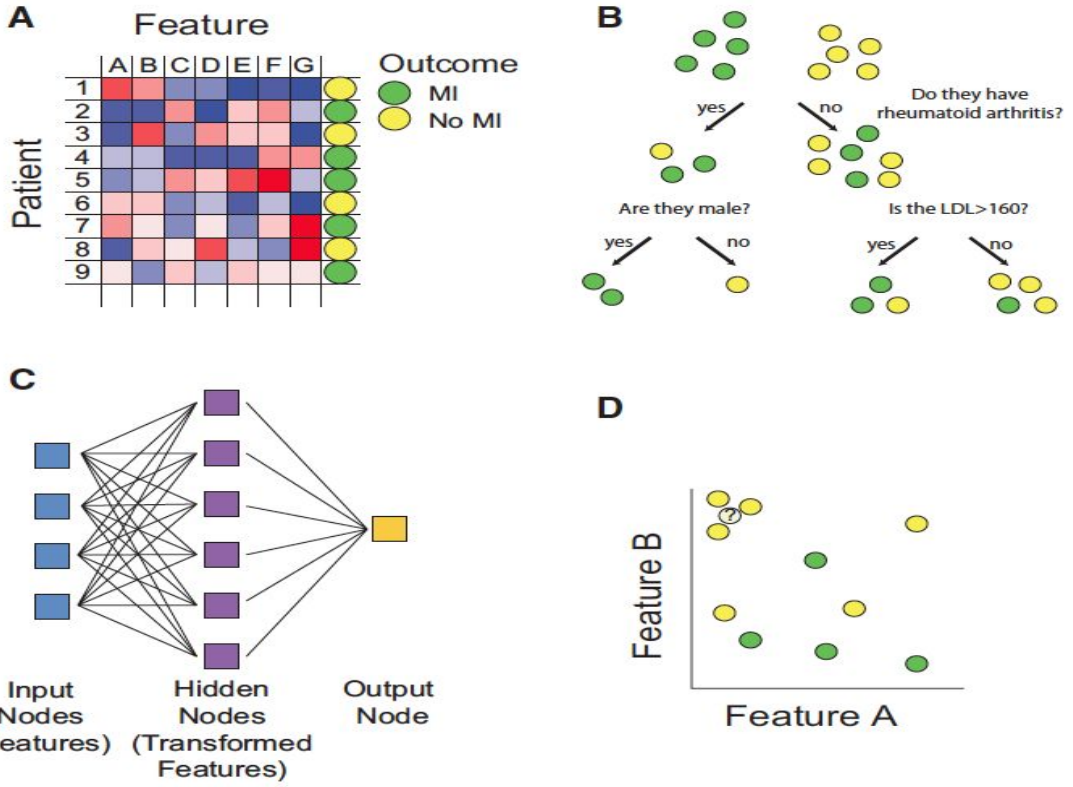
- Want to build an accurate model to ***discriminate*** between classes.
  - ➔ Finding predictors or features (Feature Selection)
    - ◆ Issues and Solutions
  - ➔ Find function that relates values of a feature to its prediction
    - ◆ Which function to work with?

Feature Selection - choice of functional class

Type of functions

1. Logistic regression model - generalized linear model
2. Decision trees - mutually exclusive causes?
3. Neural networks
4. Support Vector Machines
5. Prototype methods - k-nearest neighbours

Feature Selection



The Learning Problem

- Free Parameters
- Separation of tasks
- Estimating training error
- Collecting requisite inputs
- Test data vs training data
- Complexity vs generalizability
- Best solution?
- How much data?

Supervised Learning — Learning From Forests and Trees

- Provide sufficient flexibility to minimize training error
- Allow generalization to new data sets
  AND
- In a computationally efficient wav

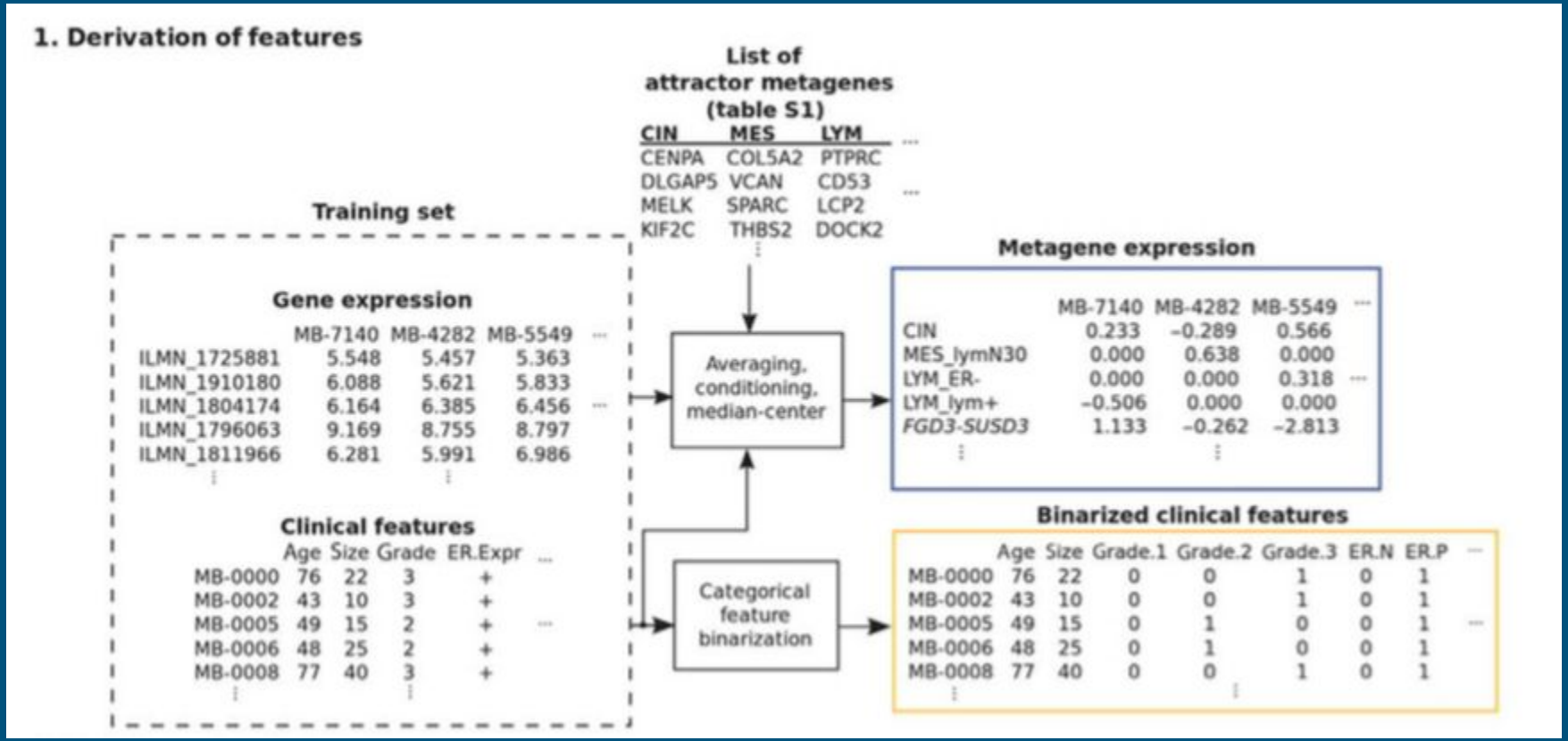
Innovative and highly effective algorithm - constructed from trees - regularization - bagging - subset of features at each node

RSF performance was actually inferior compared to Framingham Risk Score, despite fewer variables & a more complex model for RSF

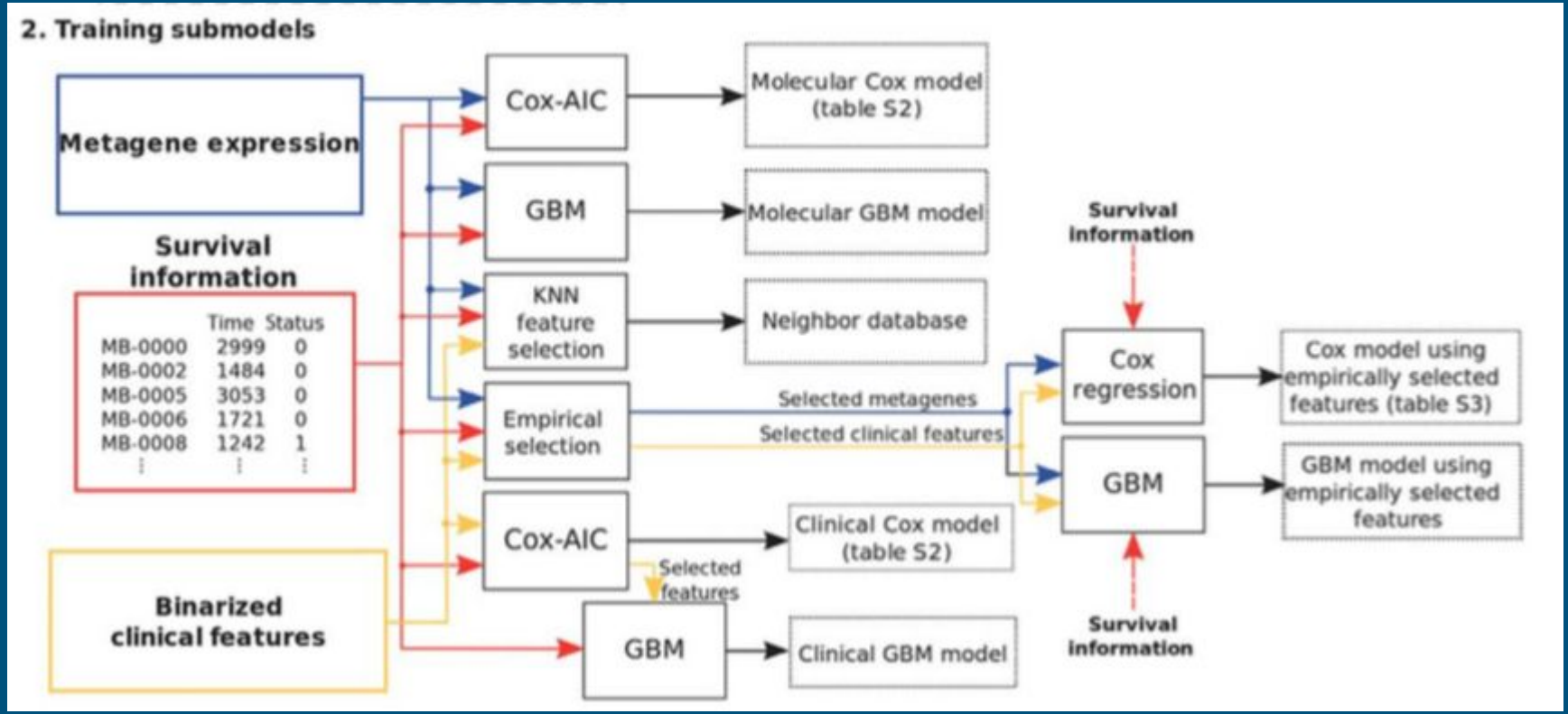
Attractor Metagenes in Cancer and Bake-Offs in Machine Learning

- Netflix $1,000,000 bake-off
- Rare in medicine
- Sage Bionetworks-DREAM Breast Cancer Prognosis Challenge
  - Lessons Learned
- Attractor Metagenes

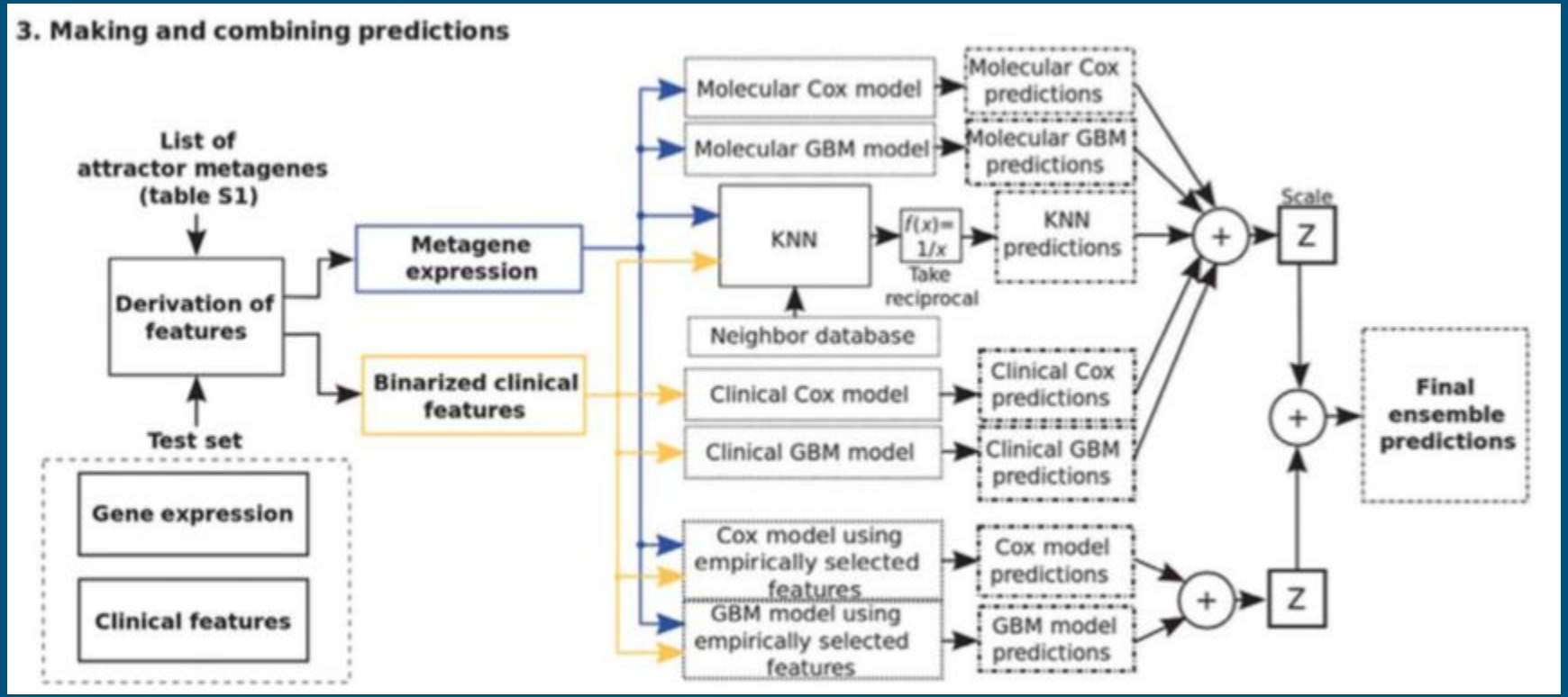
Attractor Metagenes in Cancer



Attractor Metagenes in Cancer



Attractor Metagenes in Cancer

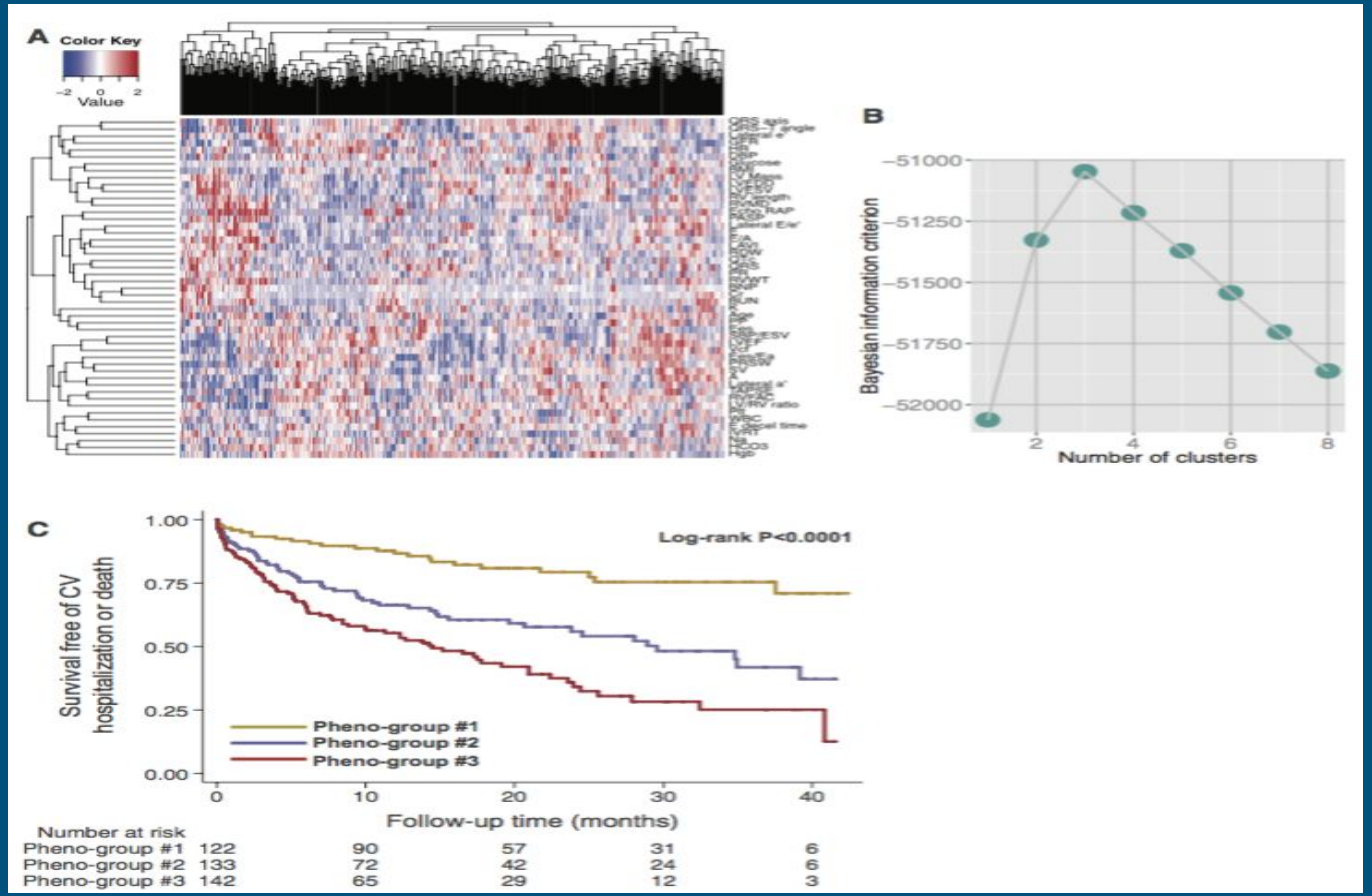


Unsupervised Learning in HFpEF

- Classifying HFpEF patients - seeking to find ***internal structure*** in the data.
- Instances (patients) characterized by a feature vector & value to their attributes (age, sex,..)
- Matrix Representation - to find instances (patients) similar to one another
- Agglomerative hierarchical clustering/K-medoids clustering

➔ Sparse Coding

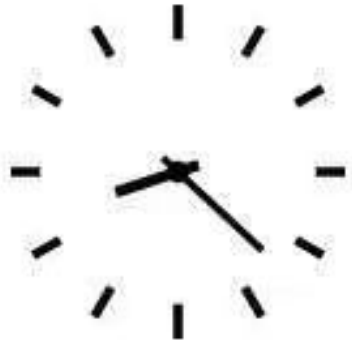
Unsupervised Learning in HFpEF



Conclusion

- Application of ML to clinical datasets - robust risk models/redefined patient classes
- Predicting outcome from diverse features/finding recurring patterns
- Limited clinical footprint of ML
  - Reluctance to completely entrust a Machine
  - Reimbursement and liability
  - Black-box nature of automated systems
- Reimbursement model - integrated man-and-machine approach

QUESTIONS??



Q & A time


Thank You