Building comprehensive searches through a Machine Learning approach for systematic reviews

Corrado Lanera[1] (corrado.lanera@unipd.it), Clara Minto[1] (clara.minto@unipd.it), Abhinav Sharma[2] (abhishar@iitk.ac.in), Dario Gregori[1] (dario.gregori@unipd.it), Paola Berchialla[3] (paola.berchialla@unito.it), Ileana Baldi[1] (ileana.baldi@unipd.it).

[1] Unit of Biostatistics Epidemiology and Public Health (UBESPH) at the Department of Cardiac, Thoracic and Vascular Sciences (DCTV) at University of Padova (ITALY)
[2] Biological Sciences and Bioengineering (BSBE) Department at IIT Kanpur (India)
[3] Department of Clinical and Biological Science (DSCTB) at University of Torino (ITALY)

UNIVERSITÀ DEGLI STUDI DI PADOVA

UBSEPH

Systematic review on clinical trials

Searching not only through standard databases leads to

- Increase of patients from 10% to 50%
- Change in statistics from 0% to 29%

Impact of searching clinical trial registries in systematic reviews of pharmaceutical treatments: methodological systematic review and reanalysis of meta-analyses

BMJ 2017 ; 356 doi: https://doi.org/10.1136/bmj.j448 (Published 17 February 2017)
Cite this as: *BMJ* 2017;356:j448

Depite their relevant role,clinical trial registries are under-utilized

- No hierarchical branching structure
- Text search is based on few fields
- Cannot use queries' combination

Syst Rev. 2014 Oct 27;3:126. doi: 10.1186/2046-4053-3-126.

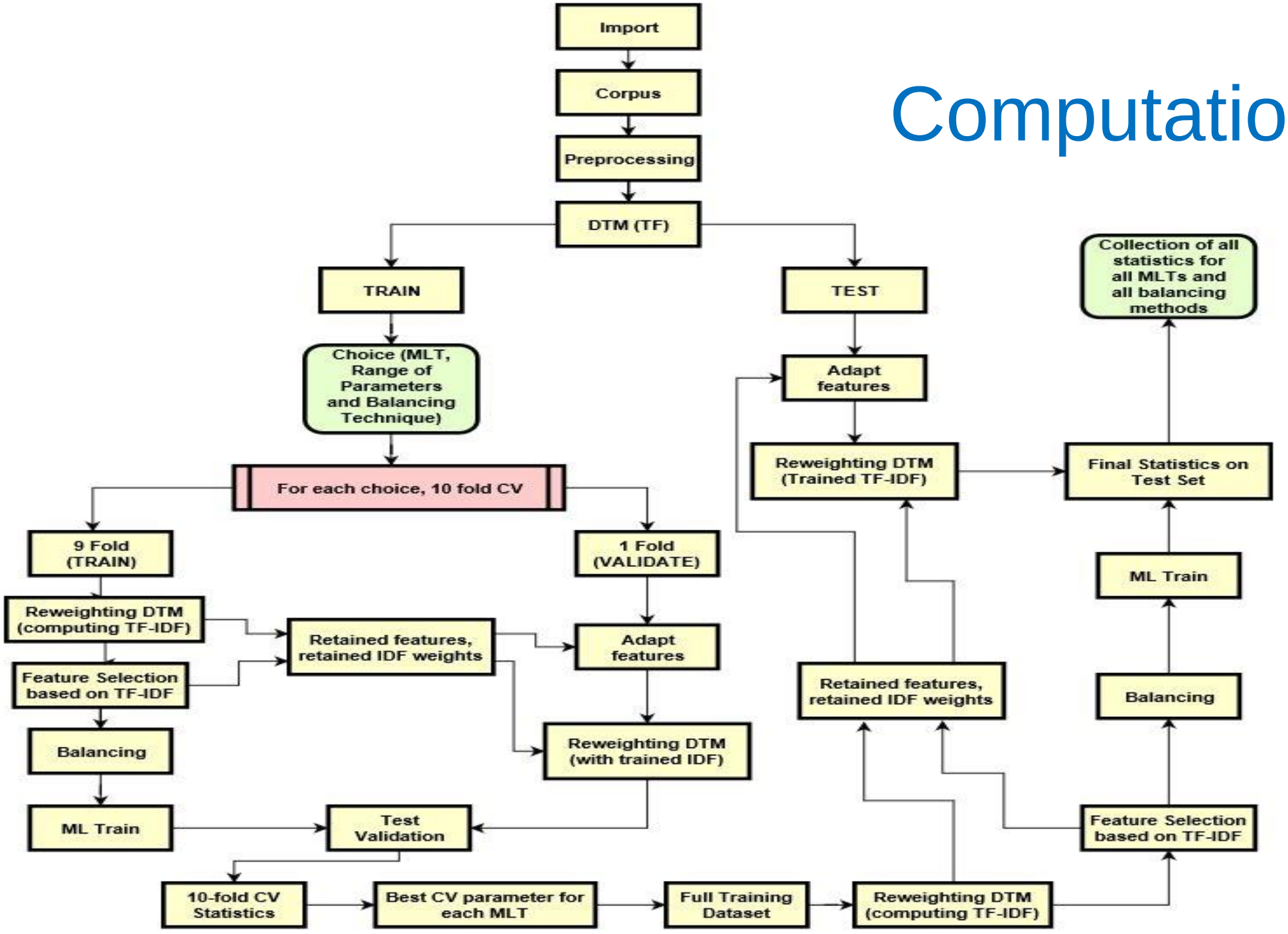
Clinical trials registries are under-utilized in the conduct of systematic reviews: a cross-sectional analysis.

Jones CW[1], Keil LG, Weaver MA, Platts-Mills TF.

Statistical challenges

- Making text data amenable to analysis
  - Train and Predict on different sources
  - Dimensionality of data is huge
- Dealing with unbalanced dataset
  - Most of the classifiers are biased towards majority class and perform poorly on minority ones.

Computation plan



Case Studies (method)

Impact of searching clinical trial registries in systematic reviews of pharmaceutical treatments: methodological systematic review and reanalysis of meta-analyses

Marie Baudard,[1,2] Amélie Yavchitz,[1,2,3] Philippe Ravaud,[1,2,3,4,5] Elodie Perrodeau,[1,2,3,4] Isabelle Boutron[1,2,3,4]

14 Systematic reviews and Meta-analyses re-analyzed by Baudard et al.

Training (PubMED**)**:

- positive : trials originally included in the systematic review
- negative : off topic trials randomly choosed (1/20)

Test (ClinicalTrial.gov**)**:

- positive: trials by Baudard et al. study
- Negative: off topic trials randomly choosed (up to a set of 100)

MLT:

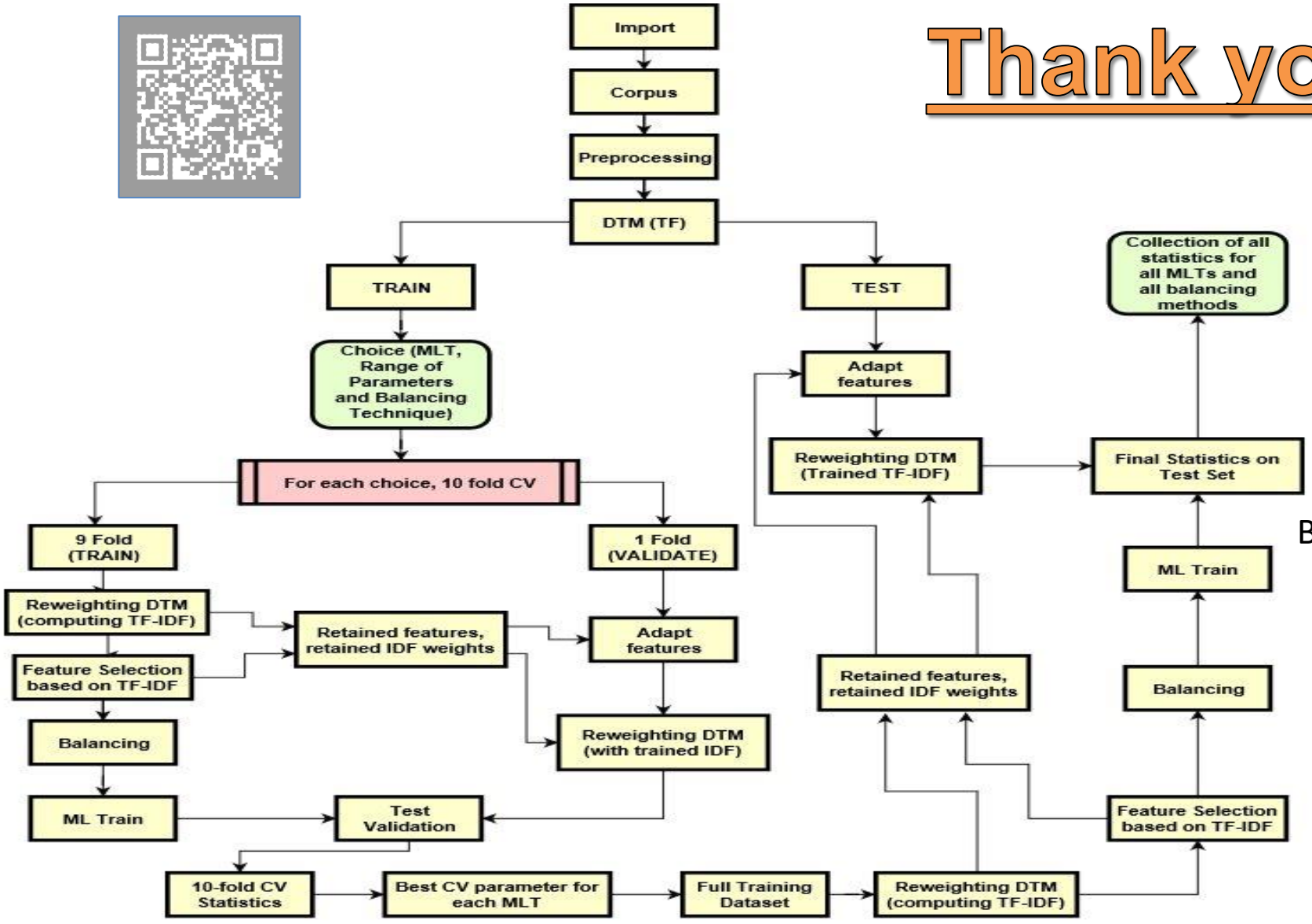
- Random Forest (10-fold CV, Random Oversampling Sampling)

Median AUC	0.798
Median SENS & RECALL	1
Median SPEC	0.596

	sr	train_pos	train_neg	test_pos	test_neg	AUC	PPV	PREV	SENS	SPEC
Yang		4	200	1	99	NA	NA	NA	NA	NA
Meng		9	200	1	99	0.677	0.015	0.01	1	0.354
Segelov		13	400	2	98	0.770	0.043	0.02	1	0.541
Li		6	200	1	99	0.859	0.034	0.01	1	0.717
Lv		12	400	1	99	0.798	0.024	0.01	1	0.596
Wang		32	800	1	99	0.909	0.053	0.01	1	0.818
Zhou		9	200	1	99	0.803	0.025	0.01	1	0.606
Liu		23	600	21	79	0.918	0.618	0.21	1	0.835
Douxfils		13	400	1	99	0.606	0.013	0.01	1	0.212
Kourbeti		75	1600	4	96	0.854	0.125	0.04	1	0.708
Li		9	200	2	98	0.592	0.024	0.02	1	0.184
Cavender		6	200	1	99	NA	NA	NA	NA	NA
Chatterjee		18	400	1	99	0.515	0.010	0.01	1	0.030
Funakoshi		9	200	2	98	NA	NA	NA	NA	NA

Thank you

corrado.Lanera@unipd.it



POSTER #17:
Session Number **214592** on **8/2/2017** beginning at **10:30:31 AM** in the Baltimore Convention Center, CC-Halls A&B.

QUESTIONS ?

UNIVERSITÀ DEGLI STUDI DI PADOVA
UBSEPH