

Data Science

Cytel data scientists apply advanced statistical techniques including predictive modeling of biological processes and drug interactions to unlock the potential of big data. Our team supports biomarker discovery and diagnostic test development based on biomedical signals and images, and real world evidence analysis.



The Cytel Difference

Cytel is the world's largest biometrics CRO and recognized leader in biostatistics and trial design. Our Data Science group brings together a multi-disciplinary team of biostatisticians, biomedical engineers, bioinformaticians and computer scientists to tackle a range of complex biological problems. The team brings a critical perspective to issues of false-discovery control, reproducibility and generalizability.

Our Services



Biomarker Discovery

Developing reliable biomarkers can guide drug development decision-making. The application of statistical and machine learning techniques to large datasets from varied sources can assist the process of biomarker identification. Once identified, these biomarkers can be applied to population enrichment strategies and precision medicine.

Techniques

- Hierarchical and model based clustering
- Supervised and unsupervised regression trees
- Semi-supervised models
- Support Vector Machines
- Random Forest
- Self-organizing Maps
- Deep Learning
- ROC analysis
- Genetic Algorithm

Signal Processing and Classification

Application to medical diagnostics based on biomedical signals and images.

Techniques

- Time-Frequency Analysis of Signals (Fourier Transforms)
- Multi-Resolution Analysis (MRA) using Wavelets
- Feature Extraction
- Feature Selection
 - *Genetic Algorithm for Feature Selection*
- Model building and validation
- Adaptive Design of validation studies for diagnostic devices
- Strong experience in regulatory interaction with medical device regulatory agencies.

Data Mining

Data mining uses algorithms and techniques from machine learning and statistics to allow us to extract information from large datasets and identify patterns and trends.

Techniques

- Data exploration using graphical tools
- Logistic regression models
- Survival models
- Partial least squares models
- Machine learning tools (RF, SVM, Deep learning)
- Functional data analysis

Programming Languages

Cytel data scientists use high performance computing methods and software engineering techniques in a range of programming languages to create customized solutions including:



R



C++



R-Shiny

*used for
customized software*



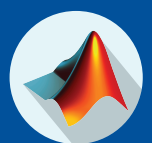
Python



Perl



Julia



Matlab

Case Examples

- Identification of biomarkers in Parkinson's disease that discriminate between treatment responders and non-responders. A range of techniques were used including Hierarchical and Random Forest Clustering, semi-supervised methods and self-organizing maps (SOM). These biomarkers are now being used as enrichment strategies for designing Phase 2 trials.
- Data mining using predictive modelling approaches to ascertain the influence of treatment on imbalance of cardiovascular events in a large vaccine trial. Cytel data scientists found evidence that patient level variables may explain the imbalance rather than the vaccination.
- Signal processing and classification in an ophthalmology indication using support vector machine algorithms and wavelets. The sponsor gained regulatory approval based on this work.
- Signal and image processing and classification in the area of neuro-sciences using wavelets and machine learning algorithms including Bayesian algorithms. A pivotal trial (also designed by Cytel statisticians) to validate this classifier algorithm is ongoing. The 2/3-stage flexible trial design for this pivotal trial has been accepted by the CDRH and the FDA.