Potential Application of Machine Learning in Drug Development and Regulation

Qi Liu, Ph.D., M.Stat.
Office of Clinical Pharmacology
OTS/CDER/FDA

ASQ 509 SIG meeting
29-August-2019

The views expressed are those of the author and do not reflect official policy of the FDA.
Outline

• Background
• Examples of ML submissions to the FDA
• Examples of ML use at the FDA
• Challenges and future directions
What is Machine Learning (ML)?

Two definitions of Machine Learning are offered. Arthur Samuel described it as: "the field of study that gives computers the ability to learn without being explicitly programmed." This is an older, informal definition.

Tom Mitchell provides a more modern definition: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

https://www.coursera.org/learn/machine-learning/home/welcome
The field of computer science dedicated to solving cognitive problems commonly associated with human intelligence, such as learning, problem solving, and pattern recognition.

https://aws.amazon.com/machine-learning/what-is-ai/
Machine Learning and Related terminology

Artificial Intelligence

Machine Learning

Artificial Neural Network

Deep Neural Network ≈ Deep Learning

Diagram showing a neural network with input layer, hidden layers, and output layer.
Machine Learning and Related terminology

Artificial Intelligence

Machine Learning

Artificial Neural Network

Deep Neural Network

Deep Learning

≈
Types of Machine Learning

- Machine Learning
  - Supervised Learning
    - Classification
    - Regression
  - Semi-supervised Learning
  - Unsupervised Learning
    - Clustering
  - Reinforcement Learning
    - Reward based learning on how to act in certain environment
The Potential Use of ML in Drug Development

Aggregating data, synthesizing information, seeking patterns and optimizing decisions

- Understanding disease/targets
- Generating/evaluating drug candidate(s) and combinations
- Improving trial design
- Advancing precision medicine by improving diagnosis and treatment
Examples of ML Submissions to the FDA

- Diagnostic tools
- Pharmacometrics analyses
- Disease models as a drug development tool (randomization aid, patient enrichment, virtual control)
- Prediction of drug response based on baseline factors (e.g., genomic/proteomic data) or imaging
- Identification of predictive biomarkers
- Application to claims and electronic medical records to identify drug abuse related problems in post marketing setting
Examples of Use of Machine Learning at the FDA

• Resource allocation
  – Prediction of workflow

• Review and research
  – Machine-learning-based quantitative structure-activity relationship
  – Investigation of AI/ML in the interpretation of adverse event reports
  – Pharmacometric/statistical analyses
    • Covariate identification
    • Exposure-Response analyses
  – Imaging data and precision medicine
Prediction of the first ANDA Submission for New Chemical Entities (NCEs) Utilizing ML Methodology

- Objectives: Prioritize research efforts, product-specific guidance (PSG) development and resource allocation
- Formulate the prediction question:
  \[
  \text{Time to first ANDA submission} \sim f(\text{product, regulatory, pharmacoeconomic})
  \]
- Methods of analysis:
  - Cox regression model
  - Machine-learning based method: Random Survival Forest (RSF)
- Results and Conclusions:
  - The RSF model outperforms the Cox regression model in prediction.
  - This approach can be expanded to other prediction tasks, e.g., predicting the number of ANDAs submitted.

To investigate the ability of NLP/ML techniques to assess safety reports received in FAERS and VAERS

To ensure efficient triage to FDA reviewers, potentially leading to better safety surveillance and improved protection of public health.

Identifying Predictors for All-Cause Mortality in Diabetic Patients in the ACCORD Trial Using RSF

• **Background:** The ACCORD trial was terminated early due to an unexpected increase in mortality with intensive vs. standard glucose-lowering treatment.

• **Methods:**
  – A total of 240 variables including demographic, clinical and laboratory data, and their change from baseline during follow-up, were analyzed as potential predictors of mortality, using RSF.
  – The top 20 predictors, identified by RSF, were included in a Cox proportional hazards model with stepwise selection to validate the results.

• **Conclusion:**
  – Some markers (e.g., urinary biomarkers, loop diuretic use and age) emerged as important predictors of mortality.
  – RSF is a rapid and flexible approach to identify potential outcome predictors among a large number of variables.

https://www.ahajournals.org/doi/abs/10.1161/circ.136.suppl_1.18043
Exposure Response Analysis using Machine Learning

A simulation study

Slide courtesy of Dr. Chao Liu
Question: How do we know if further efficacy could be achieved by increasing dose given data from only one dose level were available?

Estimate the causal effects of exposure on response in a highly nonlinear system

Option A: Combining ML prediction & causal inference tools

Option B: Apply ML toolbox for approximating the ground truth
Option A: Combining ML Prediction & Causal Inference Tools

*Inverse-Probability Weighting + Marginal Structural Model*

**Propensity Score Estimation**
- *Prediction problem*
- *Nonlinear with interactions*

**Generate Weighted Population**
- *Association between confounders and exposure removed in the weighted samples*

**Inference on Causality**
- *Average treatment effects (ATE) estimated using marginal structural model (MSM)*
Option B: Apply ML to Approximate the Ground Truth

**Neural network: A universal function approximator**

![Neural network diagram with input layers and output layers](image)
Simulation Study Conclusions

• In a strong nonlinear system, linear model may provide biased estimate on E-R relationship, even all confounders were adjusted

• Estimation of E-R relationship could be more robust with causal inference toolbox if machine learning was applied in its prediction steps

• Neural Network has the potential to recover the heterogeneous E-R relationships by approximating the ground truth
Use of Machine Learning for Imaging Data to Advance Precision Medicine

• Machine learning/computer vision is a great tool to unlock the power of imaging data
  – Detection, characterization, staging of the disease
  – Prediction and assessment of treatment response
• Progress has been made in the fields in radiomics and radiogenomics
• Opportunities:
  – Incremental improvements to current practice
  – New paradigms that did not exist before
• Examples:
  – Novel metrics for response assessment
  – Predictive imaging biomarkers for drug response
A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study

Roger Sun, MD  Elaine Johanna Limkin, MD  Maria Yakalopoulou, PhD  Laurent Dercle, MD  Stéphane Champiat, MD  Shan Rong Han, MD  et al.  Show all authors  Show footnotes

Published: August 14, 2018  DOI: https://doi.org/10.1016/S1470-2045(18)30413-3

Interpretation

The radiomic signature of CD8 cells was validated in three independent cohorts. This imaging predictor provided a promising way to predict the immune phenotype of tumours and to infer clinical outcomes for patients with cancer who had been treated with anti-PD-1 and PD-L1. Our imaging biomarker could be useful in estimating CD8 cell count and predicting clinical outcomes of patients treated with immunotherapy, when validated by further prospective randomised trials.
A Proposal of Using Machine Learning and Imaging Data to Facilitate Precision Treatment

- Establish a platform/procedure for imaging data submission/storage at the FDA
- Combine imaging data with other data (e.g., clinical data, liquid biopsy data)
- Develop algorithm(s) that can predict each patient’s response to treatment and support treatment decisions
- Seek industry and academia collaborators
Challenges and Future Directions in the Application of ML in Drug Development

• The expectation needs to be fit-for-purpose and risk-based
• Validation of the algorithms
• Generalizability
  – Issues with biased training data/underrepresented population (especially with high-dimensional data)
  – Training data should be unbiased and diverse/inclusive
  – Methods need to be developed for performance guarantee
• Transparency/interpretability
  – Black box nature of some algorithms
  – Methods are being developed to improve interpretability
• Deep learning is data hungry
  – ML+ Real world data (e.g., electronic medical records, patient-generated data/wearables)
  – Data sharing/precompetitive collaboration
FDA Machine Learning Fellowship

- FDA-Stanford/UCSF CERSI Machine Learning Fellowship

- Harvard-FDA INFORMED Post-Doctoral Fellowship in Artificial Intelligence and Machine Learning
  https://spark.adobe.com/page/5ehro94b1FpGv/
Acknowledgement

– Chao Liu
– Hao Zhu
– Yaning Wang
– Xiangyu Liu
– Lin Lin (NJIT)
– Daphney Jean
– Shiew-Mei Huang
– Issam Zineh
– Peter Lee
– Brian Booth
– Atiqr NAM Rahman
– Paul Howard
– Jeffry Florian
– Kyle Myers
– Nicholas Petrick
– Berkman Sahiner
– Yutao Gong
– Sean Khozin
– Julia Beaver
– Julie Schneider
– Marc Theorect
– Gideon Blumenthal
– Richard Pazdur
– Meng Hu
– Liang Zhao
– Scott Proestel
– Robert Ball
– Keith Burkhart
– David Strass