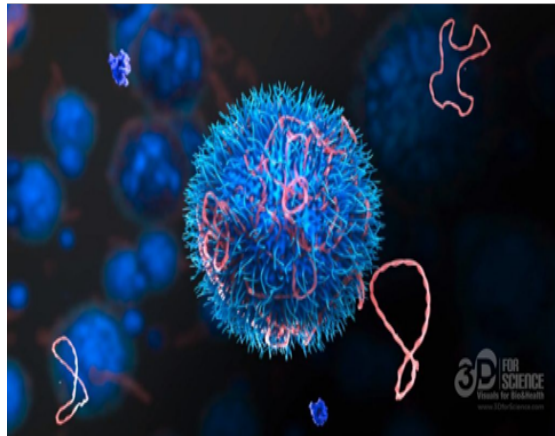# Machine Learning in Medicine: MoA Prediction

Challenges in Medicine

In the past, scientists derived drugs from natural products or were inspired by traditional remedies. Many common drugs were put into widely used decades before their biological mechanisms were fully understood, which could be very dangerous. Lack of understand makes inventing new drugs more and more difficult as well.

MoA Matters

MoA stands for Mechanism of Actions. Nowadays, when studying for a certain disease, scientists often identify a protein target associated with that disease and wish to develop a certain molecule that can modulate that protein target. As a shorthand to describe the biological activity of a given molecure, scientists assign a label referred to as MoA.

Our Solution

We treat sample of human cells with the drug and then analyze the cellular responses, searching for similarity to known patterns in large genomic databases. We apply machine learning to train a model using these data to predict what MoA of a drug will have with certain genomic data such as gene expression, cell viability patterns and so on.

Why Our Solution

- Discover hidden correlation pattern that is impossible to see by human eyes
- Dig out known information from unknown drugs
- Satisfying prediction accuracy
- Low cost (Computational Cost only)
- Free model update if more experimental data is available

AlphaUnFold
Technology

Technical Details

We use 4 datasets: Train features, Test features, Train targets scored and Train targets unscored.

Features include 5 features: treatment time, treatment dose, treatment method, gene expression, cell viability. (876 entries)

Basically, we train our model using train features and test our model using test features. We compare our prediction with train target scored data. Train targets unscored data is ancillary data set that could potentially provide more training options.
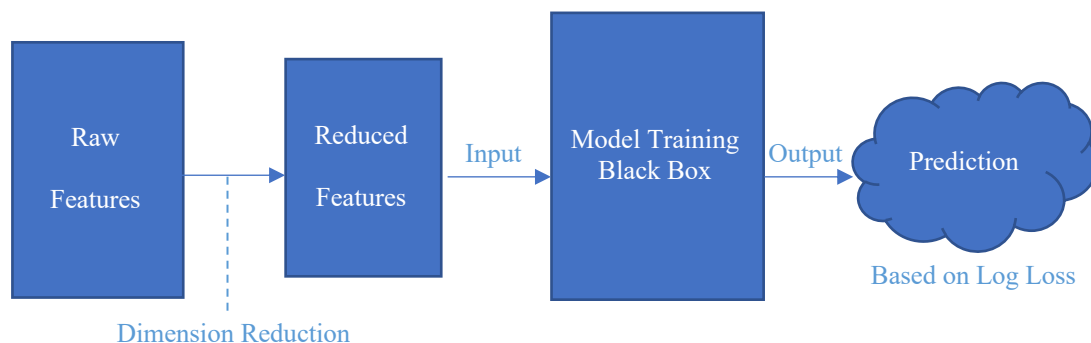
Model Optimization

We apply several model optimizations during the training process. For example, for basic model, we use Batch normalization and scheduler.

Data Preprocessing: The most important step to do in data preprocessing is dimension reduction. (or feature reduction) Several common techniques such as PCA analysis, K-means clustering, RankGauss are used.
Model selection: This is another crucial step. Since this is a multi-features, multi-classes classification problem and there are so many features, we choose to use Neural Networks, including basic model, TabNet model and deep neural network.
Model Interference: We assign weights to three different models and calculate integrated results, improving the overall prediction accuracy.

Flow Chart Demo

Raw Features → Reduced Features —Input→ Model Training Black Box —Output→ Prediction

Based on Log Loss

Dimension Reduction

Future Work

As data getting more, Cloud computing can be applied

amazon web services

Microsoft Azure

AlphaUnFold
Technology

Founded by:

Haoyu Guo, Xinbo Xia, Rocha Zhou