Modeling Drug Toxicity Using Machine Learning

Expanding & Empowering Scientists With Neural Networks

TKS x BenchSci

The Impact & Mechanism of Drug Interactions

Failed Clinical Drugs

90% of clinical trials fail. **30%** of the trial consists of drugs being overly toxic. **40% to 50%** lacked its desired effect meanwhile another portion had unexpected harmful drug properties that were unable to continue. **Billions** of dollars are funded for these drugs and at the end, most of the drug candidates fail the trials.

Unknown Drug Interactions

With **unknown** mechanisms, **many** drug interactions with the body are still unavailable and we are not able to determine the **negative** effects of a drug, such as **toxicity**. Thus, many drug pathways are still a mystery, but provide the fundamental question of to how it reacts to the human body.

Ambiguity

Although drugs work, its mechanism could cause several **adverse** health impacts that are still unknown. Positive and negative effects are weighed, but there aren't any models that are **flexible** enough to screen a drug and provide a specific interaction result that can be measured.

Introducing

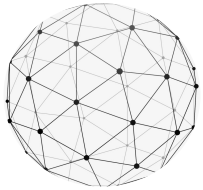
S.M.A.R.T.

Small Molecular Analysis Reporting Tool

Understanding drugs that cause cellular death - our starting point

Extracting Data From The Tox21 Dataset

We use the RD-Kit library to extract **smiles** (chemical structures) and biomarker test data to convert them into a **CSV** format. This dataset allows us to **effectively** parse the data for our model.



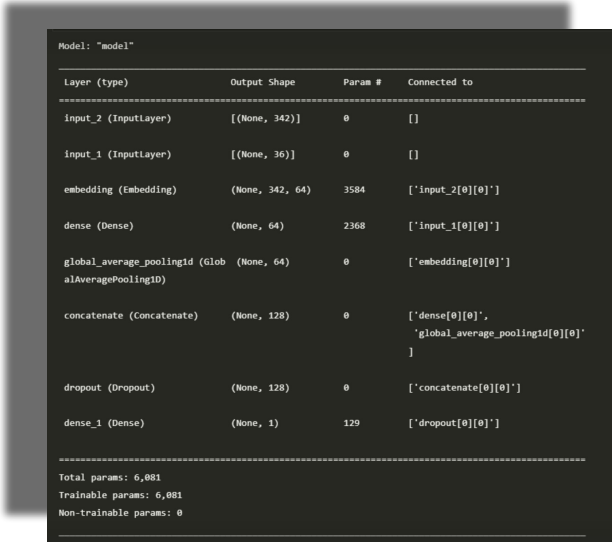
Developing A Tokenization And Embedding A Neural Network Model

The data collected from the dataset is used to create a neural network model with a **smile tokenization** layer. The tokenized smiles run through an embedded layer with features (chemical tests) through dense **TF layers**. The output is activated using **sigmoid** and determines if the drug is **toxic**.

Training The Dataset

With the neural network established, **accuracy** is ensured by **training** the model to achieve a model of **99%+** accuracy to allow for **widespread** use by the scientific community.

Implementing A Machine Learning Neural Network

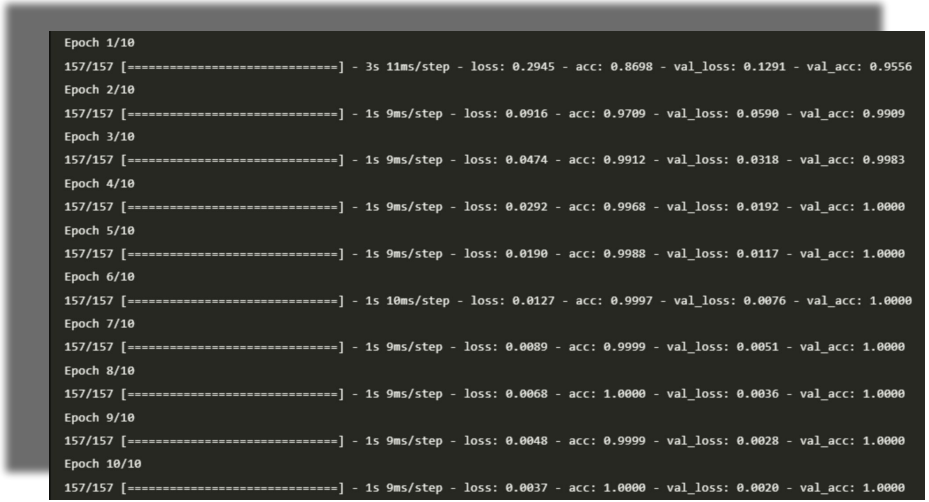


```
Model: "model"
__________________________________________________________________________________________________
 Layer (type)                Output Shape              Param #   Connected to
==================================================================================================
 input_2 (InputLayer)        [(None, 342)]             0         []

 input_1 (InputLayer)        [(None, 36)]              0         []

 embedding (Embedding)       (None, 342, 64)           3584      ['input_2[0][0]']

 dense (Dense)               (None, 64)                2368      ['input_1[0][0]']

 global_average_pooling1d (Glob  (None, 64)            0         ['embedding[0][0]']
 alAveragePooling1D)

 concatenate (Concatenate)   (None, 128)               0         ['dense[0][0]',
                                                                  'global_average_pooling1d[0][0]'
                                                                 ]

 dropout (Dropout)           (None, 128)               0         ['concatenate[0][0]']

 dense_1 (Dense)             (None, 1)                 129       ['dropout[0][0]']

==================================================================================================
Total params: 6,081
Trainable params: 6,081
Non-trainable params: 0
__________________________________________________________________________________________________
```

Neural Network

S.M.A.R.T uses a complex neural network to ensure accuracy. Our AI takes the **features** (SMILES and biomarker assays) and passes them to their own paths through **hidden layers** for calculation. Our system then concatenates the outputs, passes it through a final layer. This will **output** the drugs toxicity..

Training A Machine Learning Neural Network



Validation Overview

Our AI goes through several training periods, called **epochs**, to improve accuracy. Here, our AI's validation accuracy improves to more than **99%**.

Future results

Our plan

A Solid Framework

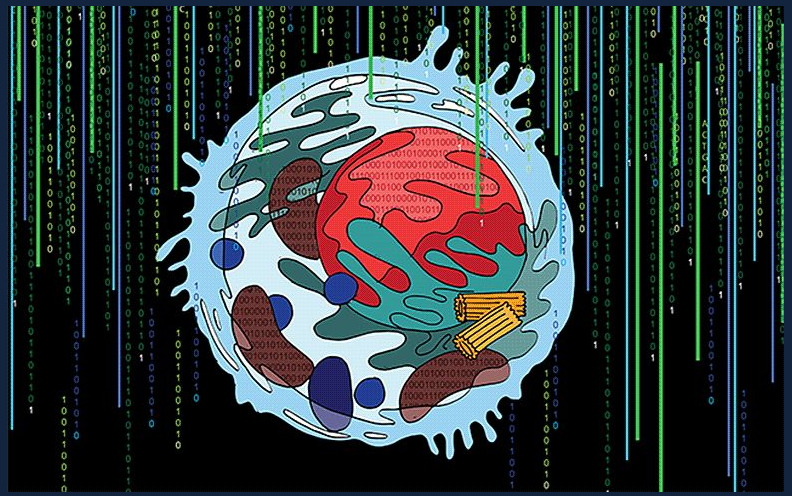
By tokenizing SMILES and then training a separate embedding layer in our net, we have developed an incredibly flexible model that can take any kind of biomarker assay test as features. This allows us to repurpose the embedding layer for understanding and predicting for other drug pathway interactions

Consumer Tailored Model

An example of our model's versatility is its capability to guide and tailor it to any research program depending on their biological assay data). By taking their previous SMILES and biomarker assay results, training a model on those results while using our pretrained SMILES embedding layer.

Top Chemical Assay Candidate

Because this model screens for any drug, biological assays performed will bring a specific result in which based off of this result, scientists would be able to determine what the drug is best suited for.



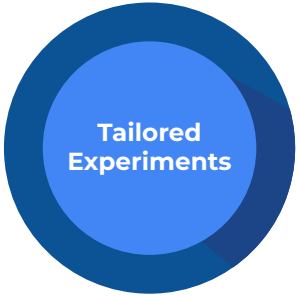
The Benefits

Cost-Efficient

By using our neural network technology, many drug research and clinical trial efforts can **prevent wasted funding** by investing in a drug that is harmful, such as high toxicity, or causes increased risks for other diseases, and more.

Applicable

Our neural network can **screen** various drugs and would be able to perform **different biomarker** assays, making it a widespread and applicable tool for the **entirety** of the scientific community.



Customization

Researchers would be able to **tailor** their experiments for drugs that perform at a certain interaction, **eliminating** drugs that can negatively affect patients through health-induced risks. This can advance society as a whole, much more rapidly by **targeting** specific drug cures.

Optimization

By tailoring experiments to specific drugs, researchers may **eliminate** unwanted drugs known to be toxic, through **microscopic** and **biological** properties. It can also screen for multiple markers, increasing efficiency for optimal effect

Solution Feasibility

Irina Sementchoukova

Research & Development Scientist
@ **BenchSci**

"**SMART** uses **machine learning** to extract meaningful **biological** information for any compound, using only the **chemical structure.** This technology has the **potential** to **expedite** early screening in any drug **discovery** research."

Darien Schettler

Staff Machine Learning Engineer
@ **VMware**

"The **embedding** layer for this project could possibly be used for **more** than just gauging **toxicity effectively**. This model could be **capable** of retrieving information from the **SMILES** alone.**"**

To The Future

Next Steps For S.M.A.R.T

Our Platform:

At its very core, our promise behind S.M.A.R.T is to deliver a reliable and accurate model to empower the scientific community with the necessary resources to accelerate drug discovery by understanding its feasibility through toxicity properties.

Our Dream For The Future:

Here at S.M.A.R.T, our mission is to play a role in creating a knowledge empowering environment to accelerate the discoveries of biological and drug inventions. We hope to continue to expand our database and platform, fully implement S.M.A.R.T, and ensure a widespread and impactful use for all scientists, accelerating drug discovery one at a time.

Caleb Lo

Abhinav Menon

Muhammed Tameem Mughal

Dave Nguyen

Amory Cheng

Thank you, BenchSci!

Being able to work with such a society advancing organization has been a true learning experience for us. Throughout the process, we were able to expand our perspectives and horizons to take apply real world experience and innovation to genuinely be able to make the world a better place with technology.

With scientific research and healthcare being such a significant issue, we were intrigued with curiosity and an innovative mindset to learn more and be able to find a solution to expand and empower scientists around the world with knowledge.

We can confidently say that we had an amazing opportunity to grow, learn, and think both critically while creatively outside the box. We strongly hope that this project and idea will be taken under consideration to be executed and implemented to solve the scientific community's drug related problems.

If you have any questions, feel free to reach out to us and we would love to answer them!