A Platform for Drug-Drug Interaction Prediction
Our Team

Mai La
Data Scientist
Data Engineer | MLOps

Grace Lee
Data Scientist
Product Manager | UI

Radhika Mardikar
Data Scientist
ML Engineer | UI
Drug - Drug interactions (DDI) can have severe effects

52% of adverse drug effects are preventable¹

Account for 18.3% of all adverse drug effects²

22.4% of adults in the US use 5+ medications³

¹ Source: Hakkarainen et al., 2012  
² Source: Jiang et al., 2022  
³ Source: Hales et al., 2019
Drug Interaction Testing is Expensive

$60 Billion
2023 Global Spending

60% increase
Drug approvals 2010-2019

$133 Billion
2032 Global Spending

4-5 Years
For testing full panel and reporting

Source: Precedence Research
Source: CBO
Source: User Interviews
User Interviews

**Pharmacist - Private**
- Manually find DDI

**Supplement Co CEO**
- Report for FDA
- Lots of testing

**Org Chem PhD**
- Chemical Formula
- Mechanism of Action

**Pharmacist - CVS**
- Have built in database for searching
Optimize DDI testing and assist drug companies to bring new drugs to market

Input:
- SMILES (Simplified Molecular-Input Line-Entry System) Drug structure
  \[ COC1=CC2=C(C=C1)N=C(N2)=O)CC1=NC=C(C)(OC)=C1C \]
- Mechanism of Action Pathway
  This drug stops gastric acid secretion by selective inhibition of the H+/K+ ATPase enzyme system...

Output:
- Prediction Probability of Positive Interactions
- Breakdown of supplement vs medication
- Downloadable table with drug list and probabilities
02

Product

Demo
Impressive that AI could extract DDI for new drug development.

I was amazed by how fast this tool can help define possible interactions between a new drug with hundreds of available drugs at a very high level of accuracy. This definitely help drug discovery much faster and more efficient with a lower cost. Traditionally, the process of evaluating drug interactions might take months or years of experiments in vitro or in vivo. With DD.ai, it’s less than 1 minute. Simply amazing!!!

Good start for interaction checker. It would bring greater values to the users by serving a high accurate AI model which focus on a certain drug family like antibiotics or cancer treatment drugs.
Approach
Data Sources

DrugBank (Features):
- Drug SMILES
- Action Pathway

NIH (Outcome):
- Drug Interactions

2,979 Drugs
Key Considerations

- **850+ Interaction Types**
  - Need to reduce scope

- **SMILES Chemical structures**
  - Non-traditional text encodings

- **512 Tokens**
  - Limitation for NLP BERT encodings

- **Drugbank data**
  - Limited to pairs of drugs
Scope and Encodings

**Scope**

- 850+ Class Prediction
  - Unfeasible
- Binary Classification

**Text-based features**

- **Morgan Similarity**
  - Encode SMILES chemical structures into vectors
- **Google Pegasus PubMed**
  - Summarize large text blocks of drug information
- **BioClinical BERT**
  - Encode above features for enhanced differentiability.
Architecture

Structure (SMILES)

Morgan Similarity Encoding

Action Pathway

DRUG 1
Potassium-transporting ATPase alpha chain 1 Inhibitor...

DRUG 2
Sodium-dependent serotonin transporter Inhibitor...

Pegasus PubMed

Summary 1

Summary 2

BioClinical BERT

Pooler H1 H2 H3

SSE1 SSE2

Output

Concat

H4 DDI
DrugA = Omeprazole

DrugB = Citalopram

SMILES (Simplified Molecular-Input Line-Entry System)

COC1=CC=CC(=C1)N=N(C2)=C=S(=O)C

C1=NC=C(C)C(OC)=C1C

Morgan Fingerprints Encoding
(1024 bits)

00000010000001000010010000
000000010000100010000100...

00000000010000000100000010
000000000001000010000100...

Structure Similarity (SS)

<table>
<thead>
<tr>
<th>Drug - Basis</th>
<th>Cosine Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SS_{AC1}$</td>
<td>0.161</td>
</tr>
</tbody>
</table>

Structure Similarity Encoding (SSE) - 300 centroids from KMeans

$SSE_A = [SS_{AC1}, SS_{AC2}, SS_{AC3}, ...]$  

$SSE_B = [SS_{BC1}, SS_{BC2}, SS_{BC3}, ...]$  

Key Takeaways:

- Encode SMILES with Morgan Similarity
- Create 300 clusters of encodings
- Take cosine similarity of new SMILES with 300 clusters to reduce dimensionality
Text Encodings

Google Pegasus PubMed
Summarize large text blocks of drug information

BioClinical BERT
Encode features using BERT model pre-trained on biomedical texts

Image Source: https://www.weak-learner.com/blog/2020/06/13/pegasus/
04

Results
Model Registry

**Baselines**
- Logistic Regression
  - F2: 70%  AUPRC: 72%
- XGBoost
  - F2: 70%  AUPRC: 74%

**Improvement**
- Morgan Similarity Neural Net
  - F2: 73%  AUPRC: 83%
- BioClinicalBERT SMILES Only
  - F2: 81%  AUPRC: 86%

**Final**
- Morgan Similarity + BioClinicalBERT SMILES Only
  - F2: 83%  AUPRC: 81%
- Morgan Similarity + BioClinicalBERT SMILES & Pathway
  - F2: 83%  AUPRC: 83%
Evaluation

Morgan Similarity + BioClinicalBERT on SMILES & Pathway

- F2: 83%
- AUPRC: 83%
- Recall: 86%
- Precision: 76%
Evaluation

Further hypertuning in progress!
Future Developments

**Product Focus**
Improve model predictions on specific drugs groups (i.e. antibiotics)

**Encoding**
Train BioBERT with drugs target & action pathway texts

**Modeling**
Graph base GNN
Additional features: target proteins

**Latency**
Inference with GPU instances
Accomplishments

**New Drug Development**
Inform potential DDI with drug formula

**Scalability**
Autoscaling with AKS

**NLP Model Serving**
Morgan Similarity + BioClinical BERT

**Interpretable Results**
Graphs and downloadable CSV
Mission

Reimagine drug development by transforming drug-drug interaction detection
Thank you!

dd-ai-predict.streamlit.app
Acknowledgements

Professors:
- Joyce Shen
- Cornelia Ilin
- James Winegar

SME, Advisors:
- Roop Raich
- Tim Roth
- Dr. Han Dang, PharmD
- Mai Do, RPh., MBA
- Dr. Tuan La, PharmD
References

1. Preventable ADE
2. Percent of ADE
3. Adults using multiple medications
4. Drug Discovery Market - Global Industry Analysis, Size, Share, Growth, Trends, Regional Outlook, and Forecast 2023 – 2032
5. Drug Approvals
7. Drugbank: a knowledgebase for drugs, drug actions and drug targets
8. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules
10. Google/Pegasus-Pubmed
11. Publicly Available Clinical BERT Embeddings
Thanks

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and images by Freepik