



Drug Discovery with few-Shot Learning

Group 2
Case Study Representation

109136502 陳宗聖

109033573 吳育葦

109033804 張軼峯

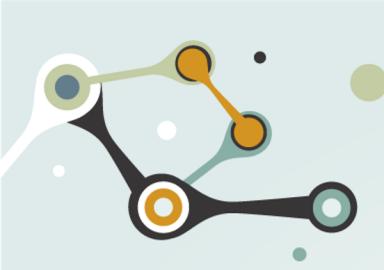


Outline

- Topic – **Drug Discovery**
- Challenges - **Why few-shot learning?**
- Domain Knowledge in drug discovery
- Benchmark Dataset
- State-of-the-art Approaches
- Conclusion & Future Directions
- Reference – papers & websites

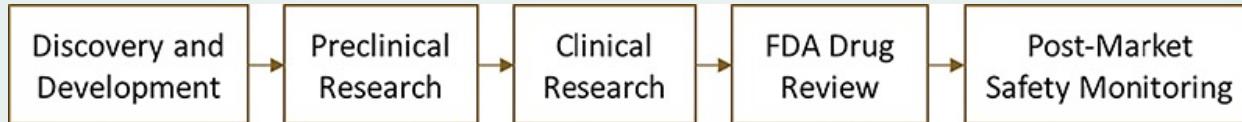


Drug Discovery & Why Few-shot Learning?

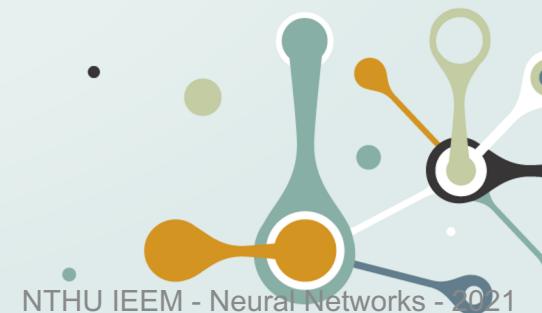


Topic – Drug Discovery

- Drug discovery and development is an **expensive** and **time-consuming** process.



- Possibility to reduce costs by **computational approaches**.
 - Activity prediction, toxicity prediction, molecular property prediction
- Machine learning and deep learning is powerful tool for **computer-aided drug discovery (CADD)**.





Topic – Drug Discovery

- **Small-molecule based drug-discovery:**

Optimize the candidate molecule by discovering analogue molecules with

- 就医 Greater pharmacological activity & Optimal therapeutic effects
- 骷髅 Reduced risks for the organism & Less toxicity
- 试管 Better solubility and selectivity

Novel Approaches of deep learning were demonstrated :

- 显示屏 Molecular property prediction
- 星星 Processing and extracting features from molecular structures



Challenges – Low data problem

- The lead optimization step of drug discovery is a **low-data problem**.
The biological labelled data is often limited in size and is very expensive to obtain.
 - » The scarcity of experimental data.
 - » Costs in data collection and processing.
 - » Highly imbalanced classification.
- Recent works use **pre-training** and **multitask learning** to leverage data from multiple sources.
- The problem of learning in low-data domain has been tackled vehemently by
 - »» **Few Shot Learning**

Domain Knowledge in Drug Discovery

SMILES, Lead Optimization, QSAR,
In Silico Virtual Screening

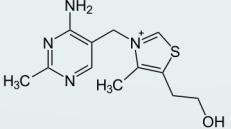


Representation - SMILES

The **simplified molecular input line entry specification (SMILES)** is a specification for unambiguously describing the structure of chemical molecules using short **ASCII** strings.

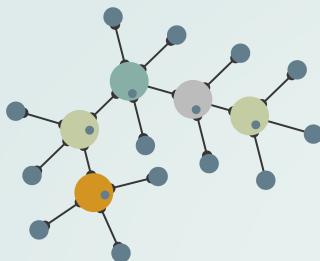
SMILES strings can be imported by most molecule editors for conversion back into **2D drawings** or **3D models** of the molecules.

Different rules for atoms and bonds, branching, aromaticity, stereochemistry, isotopes...

| Molecule | Structure | SMILES formula |
|--|--|---|
| Thiamine (vitamin B ₁ , C ₁₂ H ₁₇ N ₄ OS ⁺) |  | OC(Cc1c(N)nc2c(N)cc(C)c2)C([n+](cs1)Cc2cnc(C)nc2N)O |

Reference [1]: https://en.wikipedia.org/wiki/Simplified_molecular_input_line-entry_system

Lead Optimization & In Silico Virtual Screening



Lead Optimization

藥物的研究開發過程，大約每八萬個化合物才能獲得一個先導化合物 (lead compound)，必須經過繁雜的化學修飾來進行先導化合物優化 (lead optimization)，以增強其活性或降低毒性等工作。

需一定數量以上之化合物方能建立起結構與活性的構效關係 (structure-activity relationship)，作為下次合成與修飾之依據。

In Silico Virtual Screening

近年來電腦模擬已經被廣泛地運用在各個領域，且有相當不錯的成果，電腦虛擬篩選 (in silico virtual screening) 因此也成為開發先導化合物中非常重要的一環。

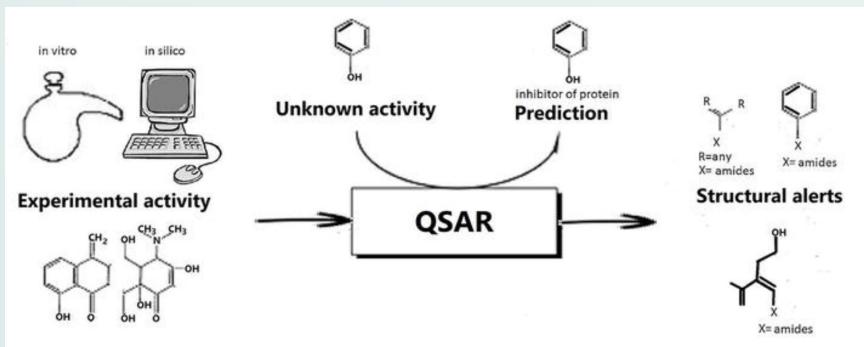
拉丁片語 **in vivo** (in life) 強調為體內與 **in vitro** (in 297 glass) 則為體外，「**In silico**」為因應而生的現代用語，**in silico** (in silicon ; on a computer chip) 強調使用電腦作試驗。



Quantitative Structure-Activity Relationship

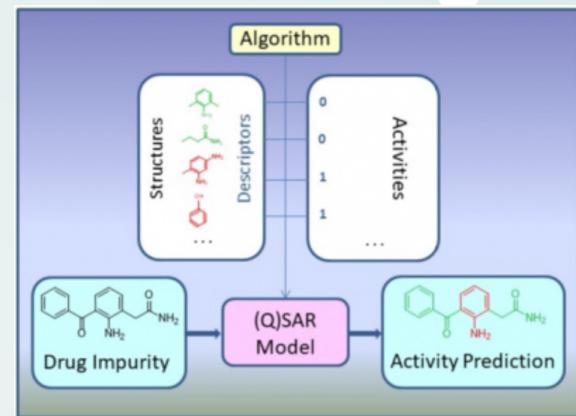
Quantitative structure-activity relationships (**QSARs**) models summarize and attempt to correlate the relationship between the **chemical structural features** of a compound and its **physicochemical properties or biological activities**.

The QSAR models are useful for various purposes including the **prediction of activities of untested chemicals**.



Reference:

<https://www.intechopen.com/books/bayesian-networks-advances-and-novel-applications/quantitative-structure-activity-relationship-modeling-and-bayesian-networks-optimality-of-naive-baye>



Reference:

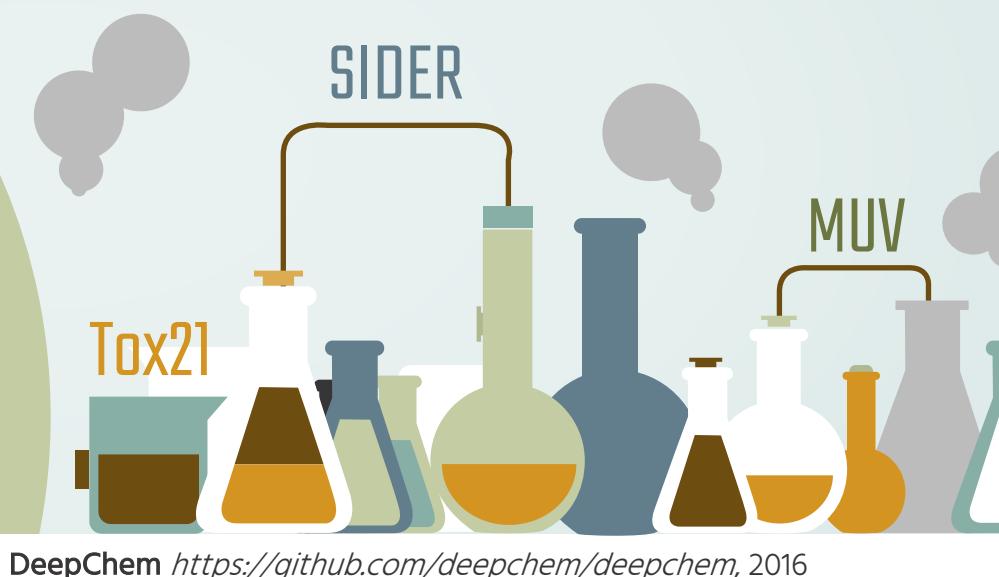
<https://www.fda.gov/drugs/regulatory-science-action/impact-story-cder-assessment-drug-impurity-mutagenicity-quantitative-structure-activity-relationship>

- ChEMBL[5] – an open large-scale bioactivity database

Reference: <https://www.ebi.ac.uk/chembl>

Dataset: MoleculeNet [4] – a large scale Benchmark

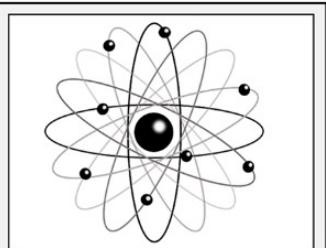
MoleculeNet curates **multiple** public datasets, establishes **metrics for evaluation**, and offers high quality open-source implementations of multiple previously proposed molecular featurization and learning algorithms.





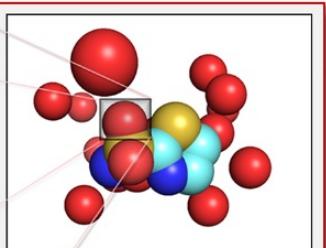
MoleculeNet [4] – a large scale benchmark

MoleculeNet datasets collection cover **various levels** of molecular properties:



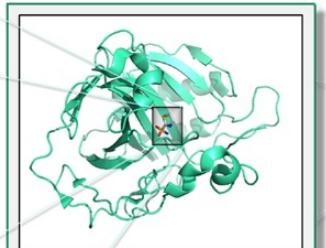
Quantum Mechanics

- QM7
- QM8
- QM7b
- QM9



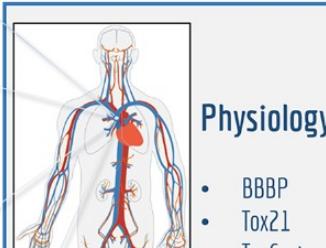
Physical Chemistry

- ESOL
- Lipophilicity
- FreeSolv



Biophysics

- HIV
- PCBA
- PDBbind
- MUV
- BACE



Physiology

- BBBP
- Tox21
- ToxCast
- SIDER
- ClinTox

Ranging from molecular-level properties to macroscopic influences on human body.

State-of-the-art Approaches



SOTA Approaches in Drug Discovery via Deep Learning

Graph Convolutions

Graph-convolutional architectures

Variational Autoencoder

Encoder and Decoder:
discrete representation

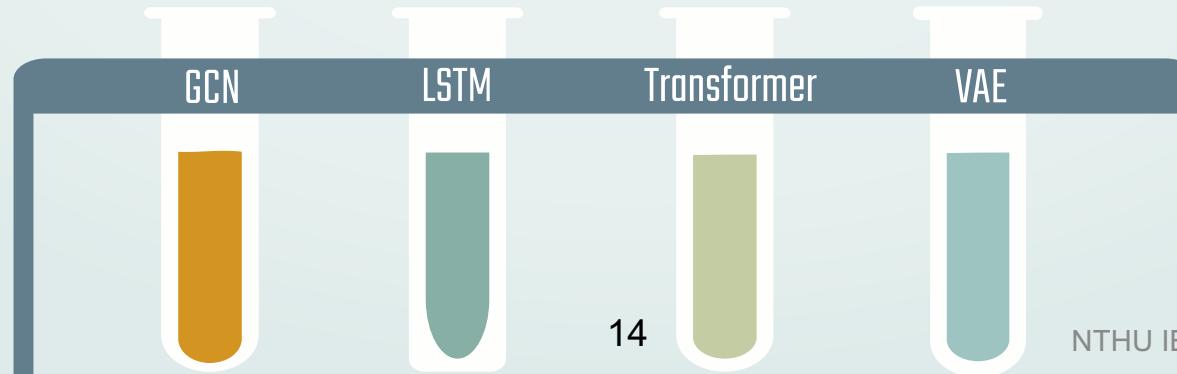
continuous vectors (latent space).

SMILES Transformer

No recurrent connections,
more stable and faster to converge.
Better featurization performance.

LSTM

Attention LSTM based
iterative refinement LSTM

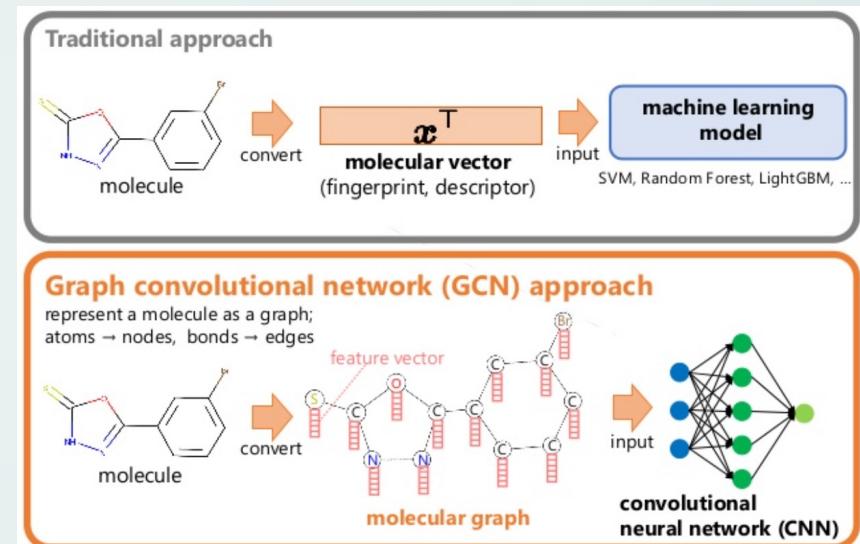


Graph Convolutional Network (GCN)

Standard convolutional architectures use a **fixed computational graph**, making them difficult to apply to objects of **varying size** or structure, such as molecules.

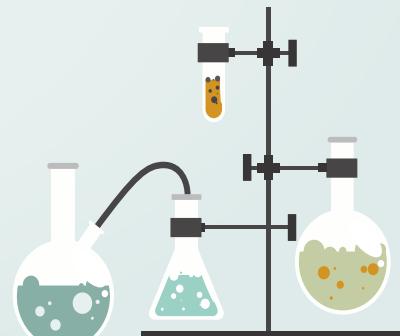
Compare with CNN, GCN can apply **more completely information of structure** to train the neural network.

Recently, GCN is widely used in deep learning application such as drug discovery.



Training Strategy – Few Shot Learning

- Few-Shot Learning (FSL) is a type of machine learning problems, where our data only contains a limited number of examples with supervised information for the target.
- Few-shot classification is a task in which a classifier must be adapted to **accommodate new classes not seen in training**, given only a few examples of each of these classes.
- And it learns a classifier which predicts label y_i for each input x_i .
 - one considers the N -way- K -shot classification



Training strategy - N -way- K -shot classification

- Episodic training strategy : D_{train} contains $I = KN$ examples from N classes and each with K examples in one episode.
- The labeled set of examples (the support set) to **predict classes for the unlabeled points** (the query set).
- In each tasks, the training classes, which are selected to be support set, are randomly sampled from training set.

Training task 1

Support set



Query set



Training task 2

Support set



Query set



Test task 1

Support set



Query set



Reference:

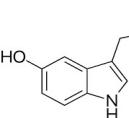
<https://medium.com/ai%2B3-theory-practice-business/what-is-few-shot-learning-4b2842646b47>

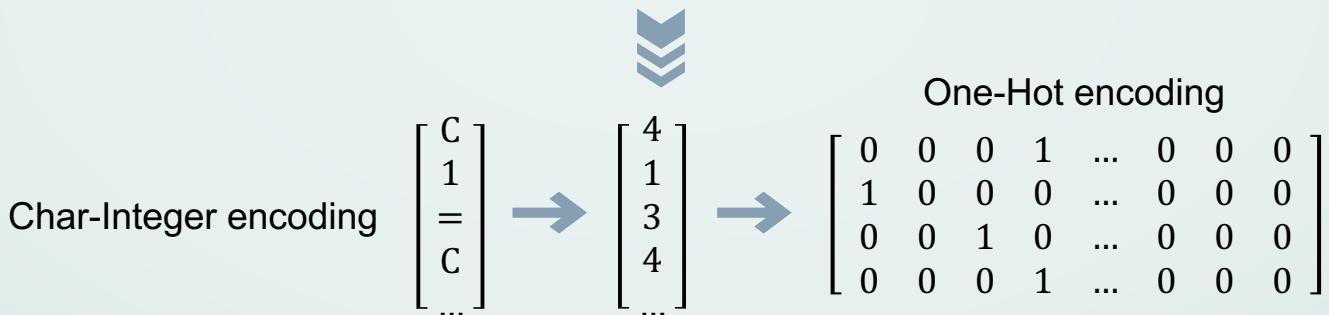


Siamese Neural Network [3]

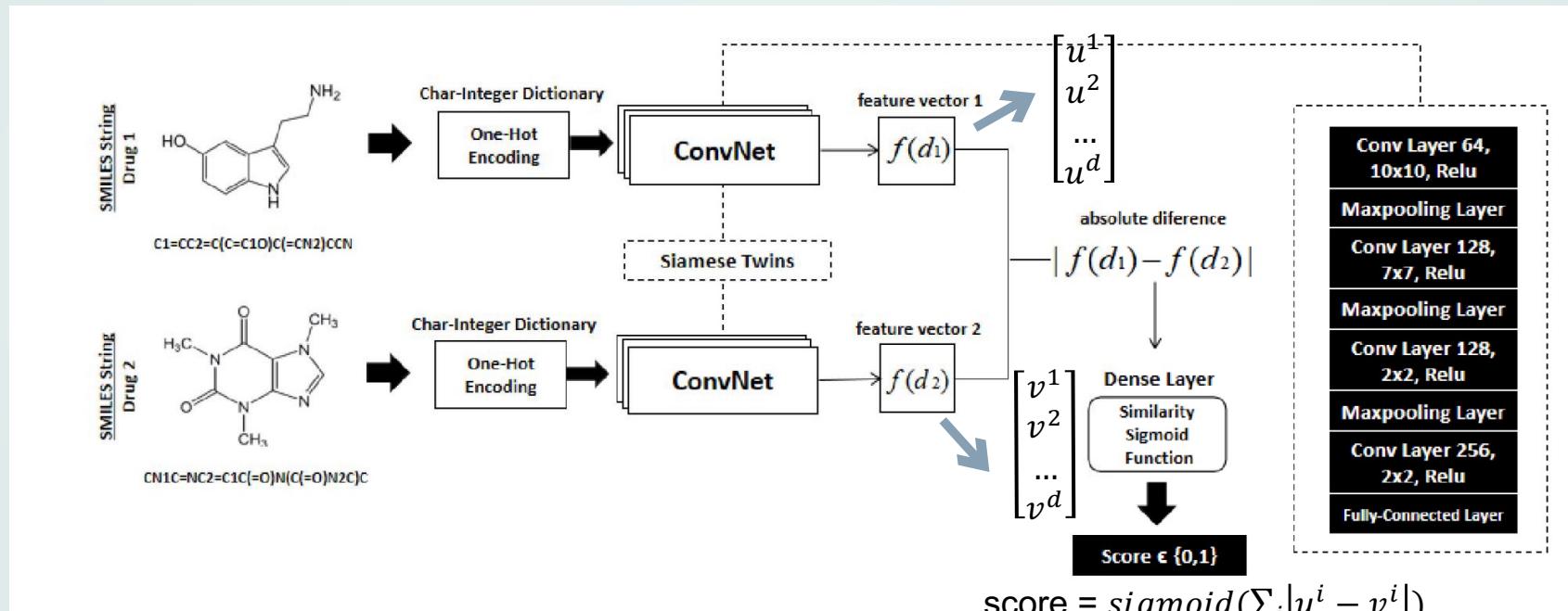
- Objective: The normal neural network models is the **need for a large number of training examples** per class, which is not always feasible in drug discovery applications.
The main objective of the study is to optimize the discovery of novel compounds based on a reduced set of candidate drugs.
- Proposed architecture : **A Siamese neural network** architecture for one-shot classification, based on **Convolutional Neural Networks**, that learns from a similarity score between two input molecules according to a given similarity function.

Data Representation

| Molecule | Structure | SMILES formula |
|--|--|---------------------------|
| Serotonin (血清素, C ₁₀ H ₁₂ N ₂ O) |  The diagram shows the chemical structure of serotonin (5-hydroxytryptamine). It consists of a tryptamine ring system where the nitrogen atom is substituted with an amino group (-NH ₂) and the adjacent carbon is connected to a methylene group (-CH ₂ -). This methylene group is further substituted with a hydroxyl group (-OH). | C1=CC2=C(C=C1O)C(=CN2)CCN |



Model architecture





How to update the parameters ?

- The two **output feature vectors** is corresponding to \mathbf{u} and \mathbf{v} . The output layer learns a similarity function between two feature vectors by applying a distance metric to the learned feature map. And the score serves as output vectors **passing a sigmoid**.

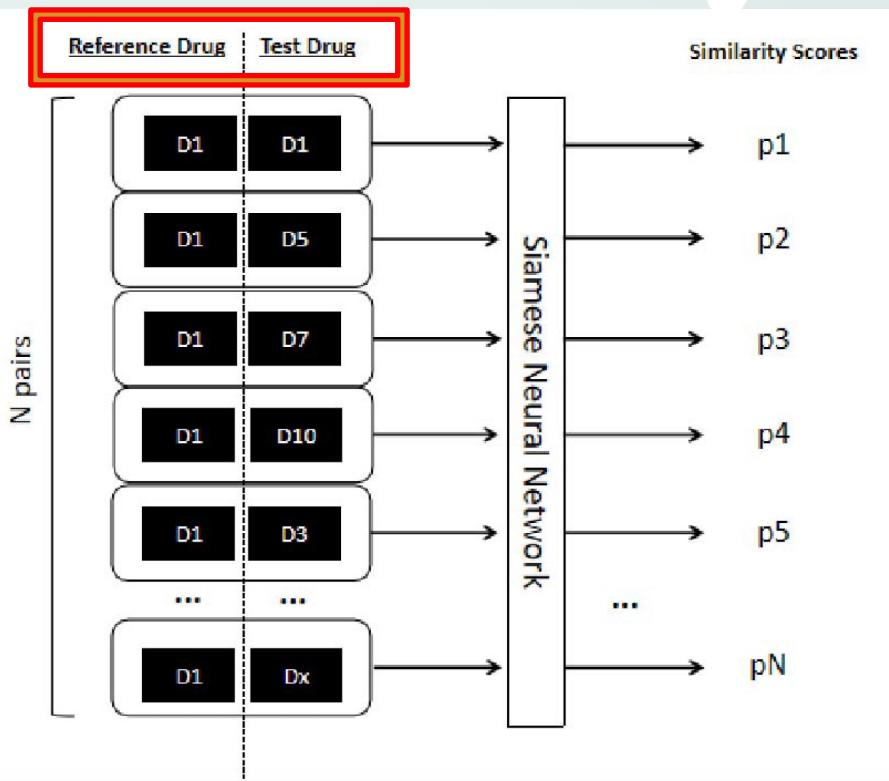
$$\text{score} = \text{sigmoid}(\sum_i |u^i - v^i|)$$

- This similarity measure (score) is a probability, assuming a value between 0 and 1.
- The loss function corresponds to a **binary cross entropy with regularization term** between the target and the prediction.

$$L(\mathbf{x}_1, \mathbf{x}_2, t) = t \cdot \log(p) + (1 - t) \cdot \log(1-p) + \lambda \|w\|^2$$

One shot learning

- Knowing that only one instance in our support set corresponds to that same class
- If the pair of compounds of the same class gets the maximum score, the model prediction is correct.



Performance between different algorithm

TABLE IV
N-WAY ONE-SHOT LEARNING ACCURACY RESULTS.

| | N | | | | | |
|--|------------|------------|------------|------------|------------|------------|
| | 2 | 3 | 4 | 5 | 7 | 10 |
| Siamese Neural Network (validation) | 94% | 90% | 84% | 78% | 70% | 65% |
| Siamese Neural Network (training) | 95% | 92% | 86% | 84% | 72% | 70% |
| KNN | 70% | 55% | 49% | 43% | 36% | 30% |
| Naïve Model | 61% | 43% | 34% | 31% | 22% | 19% |
| SVM | 56% | 42% | 30% | 24% | 16% | 12% |
| Random Forest | 71% | 58% | 60% | 44% | 34% | 20% |
| Multi-Layer Perceptron | 76% | 60% | 36% | 34% | 22% | 13% |
| Convolutional Neural Network | 81% | 70% | 58% | 46% | 41% | 39% |

Performance between different algorithm

Table 1. ROC-AUC Scores of Models on Median Held-out Task for Each Model on Tox21^a

| Tox21 | RF (100 trees) | Graph Conv | Siamese | AttnLSTM | IterRefLSTM |
|---------|-------------------|-------------------|-------------------|-------------------|-------------------|
| 10+/10- | 0.586 ± 0.056 | 0.648 ± 0.029 | 0.820 ± 0.003 | 0.801 ± 0.001 | 0.823 ± 0.002 |
| 5+/10- | 0.573 ± 0.060 | 0.637 ± 0.061 | 0.823 ± 0.004 | 0.753 ± 0.173 | 0.830 ± 0.001 |
| 1+/10- | 0.551 ± 0.067 | 0.541 ± 0.093 | 0.726 ± 0.173 | 0.549 ± 0.088 | 0.724 ± 0.008 |
| 1+/5- | 0.559 ± 0.063 | 0.595 ± 0.086 | 0.687 ± 0.210 | 0.593 ± 0.153 | 0.795 ± 0.005 |
| 1+/1- | 0.535 ± 0.056 | 0.589 ± 0.068 | 0.657 ± 0.222 | 0.507 ± 0.079 | 0.827 ± 0.001 |

^aNumbers reported are means and standard deviations. Randomness is over the choice of support set; experiment is repeated with 20 support sets. The [Appendix](#) contains results for all held-out Tox21 tasks. The result with highest mean in each row is highlighted. The notation 10+/10- indicates supports with 10 positive examples and 10 negative examples.

Table 2. ROC-AUC Scores of Models on Median Held-out Task for Each Model on SIDER^a

| SIDER | RF (100 trees) | Graph Conv | Siamese | AttnLSTM | IterRefLSTM |
|---------|-------------------|-------------------|-------------------|-------------------|-------------------|
| 10+/10- | 0.535 ± 0.036 | 0.483 ± 0.026 | 0.687 ± 0.089 | 0.553 ± 0.058 | 0.669 ± 0.007 |
| 5+/10- | 0.533 ± 0.030 | 0.473 ± 0.029 | 0.648 ± 0.070 | 0.534 ± 0.053 | 0.704 ± 0.002 |
| 1+/10- | 0.540 ± 0.034 | 0.447 ± 0.016 | 0.544 ± 0.056 | 0.506 ± 0.016 | 0.556 ± 0.011 |
| 1+/5- | 0.529 ± 0.028 | 0.457 ± 0.029 | 0.530 ± 0.050 | 0.505 ± 0.022 | 0.644 ± 0.012 |
| 1+/1- | 0.506 ± 0.039 | 0.468 ± 0.045 | 0.510 ± 0.016 | 0.501 ± 0.022 | 0.697 ± 0.002 |

^aNumbers reported are means and standard deviations. Randomness is over the choice of support set; experiment is repeated with 20 support sets. The [Appendix](#) contains results for all held-out SIDER tasks. The result with highest mean in each row is highlighted. The notation 10+/10- indicates supports with 10 positive examples and 10 negative examples.



05

Conclusion & Future Direction



Conclusion & Future Direction

- According to Matching Network, if using **N-way K-shot** training method, it prevents from overfitting even having few data.
- The state-of-the-art architecture provides an accurate and reliable prediction of novel compounds considering the lack of biological data available, and it also give the another direction about few-shot learning in drug discovery tasks.
- **GCN** is a class of computational techniques aiming at **extracting features from general graphs** through graph convolutions.
- **Interpretability** is very important when assessing a GCN model and for better understanding the underlying mechanisms.



Reference Papers

1. Low Data Drug Discovery with One-Shot Learning

Han Altae-Tran, Bharath Ramsundar, Aneesh S. Pappu, and Vijay Pande, ACS Central Science 2017 3 (4), 283-293,
DOI: 10.1021/acscentsci.6b00367

2. From machine learning to deep learning: progress in machine intelligence for rational drug discovery

Lu Zhang, Jianjun Tan, Dan Han and Hao Zhu, Drug Discov Today (2017), <http://dx.doi.org/10.1016/j.drudis.2017.08.010>

3. Exploring a Siamese Neural Network Architecture for One-Shot Drug Discovery

L.Torres, N. Monteiro, J. Oliveira, J. Arrais and B. Ribeiro,
2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), 2020, pp. 168-175,
DOI: 10.1109/BIBE50027.2020.00035.

4. MoleculeNet: A Benchmark for Molecular Machine Learning

Zhenqin Wu, Bharath Ramsundar, Evan N. Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S. Pappu, Karl Leswing, Vijay Pande
<https://arxiv.org/abs/1703.00564>

5. The ChEMBL database in 2017.

Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, Mutowo P, Atkinson F, Bellis LJ, Cibrián-Uhalte E, Davies M, Dedman N, Karlsson A, Magariños MP, Overington JP, Papadatos G, Smit I, Leach AR.
Nucleic Acids Res. 2017 Jan 4;45(D1):D945-D954. doi: 10.1093/nar/gkw1074. Epub 2016 Nov 28.
PMID: 27899562; PMCID: PMC5210557.



Reference Papers

6. **Matching networks for one shot learning,**
O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra et al., NeurIPS, 2016.
7. **Generalizing from a few examples: A survey on few-shot learning,**
Wang, Yaqing and Yao, Quanming and Kwok, James T and Ni, Lionel M. ACM Computing Surveys, 2020.
8. **Convolutional Networks on Graphs for Learning Molecular Fingerprints**
David Duvenaud, Dougal Maclaurin, Jorge Aguilera-Iparraguirre, Rafael Gómez-Bombarelli, Timothy Hirzel, Alán Aspuru-Guzik, Ryan P. Adams, NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2, December 2015, Pages 2224–2232
9. **Molecular Graph Convolutions: Moving Beyond Fingerprints**
Steven Kearnes, Kevin McCloskey, Marc Berndl, Vijay Pande, Patrick Riley, J Comput Aided Mol Des (2016), arXiv:1603.00856, DOI: 10.1007/s10822-016-9938-8
10. **Graph convolutional networks for computational drug development and discovery**
Mengying Sun, Sendong Zhao, Coryandar Gilvary, Olivier Elemento, Jiayu Zhou, Fei Wang,
Briefings in Bioinformatics, Volume 21, Issue 3, May 2020, Pages 919–935, <https://doi.org/10.1093/bib/bbz042>
11. **Graph Convolutional Neural Networks for Predicting Drug-Target Interactions**
Wen Torng, Russ B. Altman, J. Chem. Inf. Model. 2019, 59, 4131–4149, <https://doi.org/10.1021/acs.jcim.9b00628>
12. **SMILES Transformer: Pre-trained Molecular Fingerprint for Low Data Drug Discovery**
Shion Honda, Shoi Shi, Hiroki R. Ueda, [arXiv:1911.04738 \[cs.LG\]](https://arxiv.org/abs/1911.04738)



Reference Websites

1. https://en.wikipedia.org/wiki/Simplified_molecular-input_line-entry_system
2. https://archive.epa.gov/med/med_archive_03/web/html/smiles.html
3. https://bc.imb.sinica.edu.tw/images/biotech/20_23.pdf
4. <https://www.intechopen.com/books/bayesian-networks-advances-and-novel-applications/quantitative-structure-activity-relationship-modeling-and-bayesian-networks-optimality-of-naive-baye>
5. <https://www.fda.gov/drugs/regulatory-science-action/impact-story-cder-assessment-drug-impurity-mutagenicity-quantitative-structure-activity-relationship>
6. <https://www.slideshare.net/tonets/pdpta19ohue>
7. <https://qrlearning.github.io/slides/zitnik.pdf>

The background features a light gray surface with a network of colored nodes and connections. Nodes are represented by circles in various colors: orange, teal, blue, and green. Some nodes are filled, while others are outlines. Lines connect the nodes, forming a web-like structure. The overall aesthetic is clean and modern, suggesting a digital or scientific theme.

Thanks for your listening.

問題回覆

1. 投影片架構清楚，設計不錯
2. 想了解組員背景是否為藥物開發相關，若為相關背景，建議也可新增相關背景說明並提出個人專業看法等評論
非相關背景。
3. Low data problem的用詞較少見，想確認是否為該領域常見用詞？若是，則請說明出處(e.g., 參考文獻中的ACS Central Science?) 或是有其它更精確或合適的用詞(e.g., few labeled data) **請參考 reference [1]**。
4. 第14頁中有關不同方法類別的試管圖示，裡面的高低值是否有其涵義？例如是否為被採用的程度比較或是有效程度？
無特殊涵義，並修正以避免誤導。
5. 第18頁的paper title與出處未在同一頁中說明。另外想確認該會議是否為該領域的top conference或知名會議？
為該領域最新發表的論文，並已附上連結。
6. 第19頁的char-integer encoding建議可補充相關網頁連結 **此作法為作者原創，並已附上reference**。
7. 第20頁有提到「同樣結構，但不同architecture」，這部分容易困惑，需解釋清楚。另外，實際上是同一個CNN架構嗎？同架構的話有parameter sharing嗎？還是不同parameters? **CNN的架構一致，architecture並無不同。並且沒有 parameter sharing，即使架構與parameter initialization一致，但在back propagation更新kernel的parameter後，仍會使裡面的parameter不相同。**
8. 第22頁的說明不是很懂，建議可實際舉例說明會比較清楚
9. 採用Siamese Neural Network是為了學similarity function，看第20頁中的drug 1與drug 2是不同的化學式，那麼這裡正確答案的similarity (或label)是怎麼定義的？或是具備什麼化學意義？同樣的治療效果？
8.與9.一併回答:由於採用N-way one-shot method，故假設我們每個episode隨機採樣N個class出來，總共僅有N個examples。故當query example與每個examples做比較，勢必只會與其中的一個example的class相同，故在計算similarity時，由於最後面會加上sigmoid，使similarity介於0到1之間，並希望預測出契合的class與ground truth一致，且similarity能夠越高越好。
10. 好奇GCN在結論的部分看起來在該領域似乎很重要，但中間講述方法的部分反而是用CNN的方法？
本篇報告著重在 few-shot learning應用在drug discovery的實例，而前面說到GCN和CNN的比較是為了提供大家一個較為完整的背景知識。