Predicting Side Effects of Brain Drugs Using Machine Learning

Alex Whelan, Giovanna-Maria Dimitri, Pietro Lio, Max Garagnani

MSc Computational Cognitive Neuroscience

awhel001@gold.ac.uk

Goldsmiths UNIVERSITY OF LONDON

Background

Drug development costs around **$2.6 billion** (Tufts Centre for the Study of Drug Development, 2016) and few drugs reach phase III trials. Reducing this cost means a wider range of cheaper drugs. One way to accomplish this is to analyse likely side effects of drugs from known chemical features, which has been done in previous research:

- Using **chemical structure** and an ensemble classification model approach. 888 drugs, rare side effects predicted (*Jahid and Ruan, 2013*)
- Or with **targeted proteins**, correlational analyses, 658 drugs (*Mizutani et al., 2012*)
- Or with chemical and biological properties, plus **phenotypic information**. SVMs, 832 drugs. (*Liu et al., 2012*)
- DrugClust: Clustering algorithms used for side effect prediction using protein interactions (*Dimitri and Lio, 2017*)

But what about the brain? It's a highly complex system including off-target effects. Recreational psychoactives add to the need for safe and efficient drug analysis.

- Predicting BBB permeability using chemical structure is helpful to some degree, 153 compounds in rats (*Suenderhauf et al., 2012*)
- SVMs using side effect information also applied (*Gao et al., 2016*)
- Reversing this, so BBB permeability predicts side effects, has also been effective (*Fan et al., 2018*)

Overall there is a need for further work in predicting side effects of drugs that affect the brain.

Methods and Data

Chemical features database

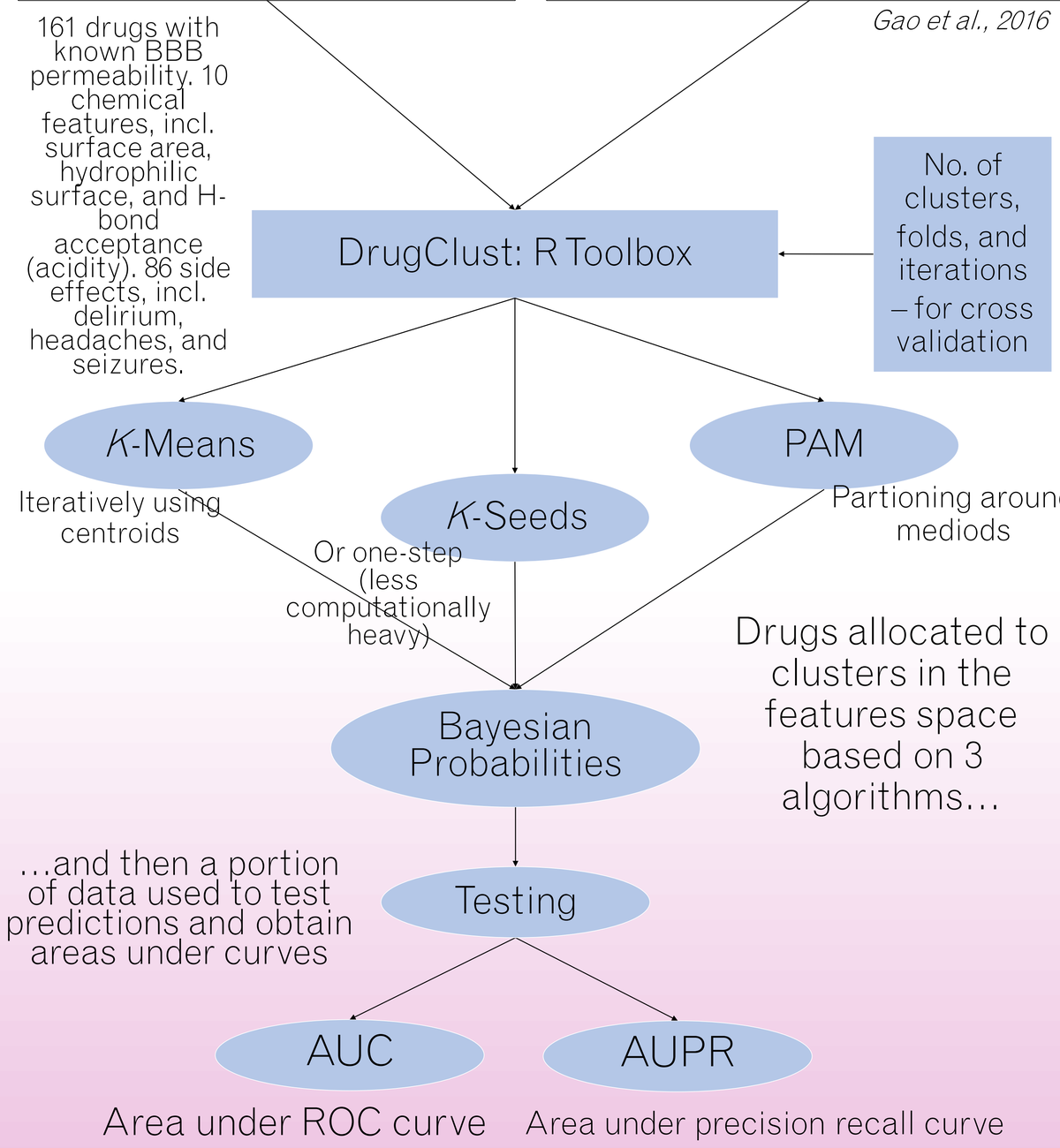
Drug ID	Drug Name	BBB +/-	MW	Volume	surface area	%hydrophillic surf.	logP
CID100000214	Alprostadil	N	354.486	205.543	27.66	27.1819	1.1
CID100000298	chloramphenicol	N	323.132	180.846	24.57	98.2414	-2
CID100000338	salicylic acid	N	138.123	71.5378	9.124	55.8614	1.
CID100000444	bupropion	Y	239.744	137.071	18.01	35.1429	2.615
CID100000450	estradiol	N	272.386	208.974	27.34	15.1308	4.3
CID100000453	mannitol	N	182.173	96.213	13.85	100	-
CID100000681	dopamine	N	153.18	88.0365	11.47	50.9332	0.0
CID100000698	estrone	N	270.371	203.892	26.43	14.2347	4.062
CID100000942	nicotine	Y	162.234	101.521	12.77	26.2634	1.4
CID100001206	methamphetamine	Y	149.235	96.1664	12.51	24.7682	-
[illegible]	doxo[illegible]		543	277.528	33.99	77.5749	-3.95
[illegible]			2[illegible]	[illegible]9	[illegible]	[illegible]9	1

161 drugs with known BBB permeability. 10 chemical features, incl. surface area, hydrophilic surface, and H-bond acceptance (acidity). 86 side effects, incl. delirium, headaches, and seizures.

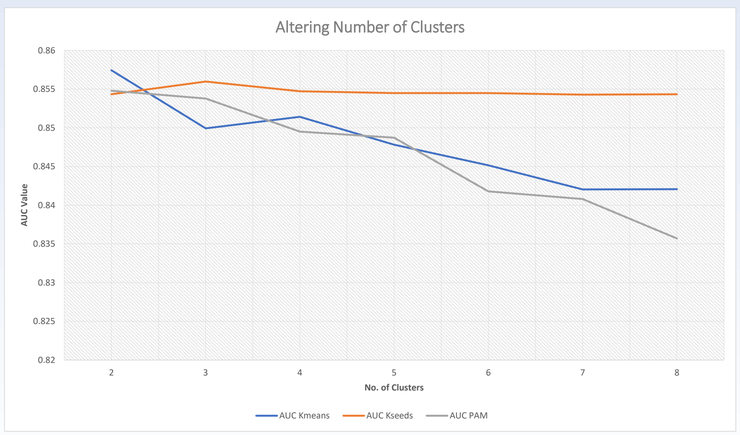
Side effects database

drug	cid	Eating disorders	Impulse control	Manic and bipolar	Mood disorders	Personality disord
chloramphenicol	CID100000298	0	0	0	1	
salicylic acid	CID100000338	0	0	0	0	
bupropion	CID100000444	0	0	0	1	
Alprostadil	CID100000214	0	1	1	1	
estradiol	CID100000450	0	0	0	1	
mannitol	CID100000453	0	0	0	0	
dopamine	CID100000681	0	0	0	0	
estrone	CID100000698	0	0	0	1	
nicotine	CID100000942	0	0	0	1	
methamphetamine	CID100001206	0	0	0	1	
[illegible]rubicin	CID100001690	0	0	[illegible]	1	
	[illegible]00017	0			1	

Gao et al., 2016



Results



Varying number of clusters from 2 to 8 had a slight negative effect on AUC, with *K*-Seeds as an exception.

Mean AUC:

K-Means – 0.848

_K_-Seeds – 0.855

PAM – 0.846

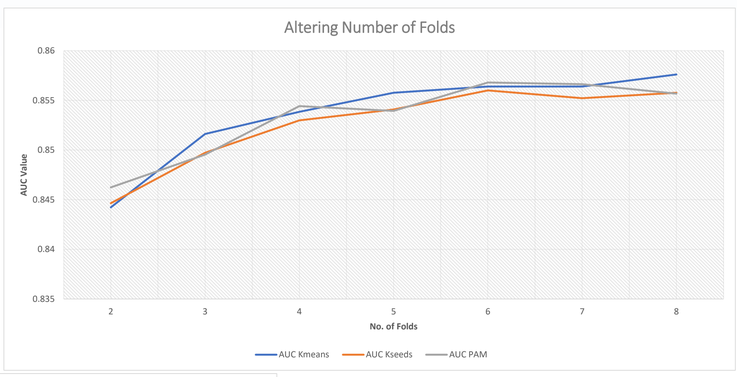
Varying number of folds from 2 to 8 was positively correlated with AUC.

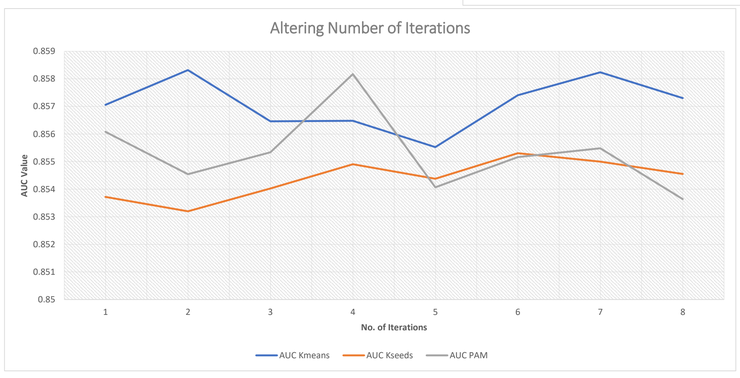
Mean AUC:

_K_-Means – 0.854

K-Seeds – 0.853

PAM – 0.853





Varying number of iterations from 1 to 8 had no clear relationship.

Mean AUC:

_K_-Means – 0.857

K-Seeds – 0.854

PAM – 0.855

Varying number of clusters and iterations was done to see if a clearer relationship emerged. The slight negative correlation remained (again with the exception of the K-Seeds algorithm).

Mean AUC (top):

K-Means – 0.845

_K_-Seeds – 0.854

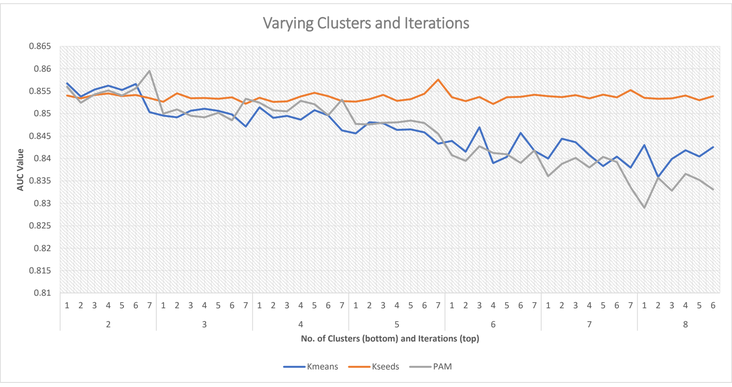
PAM – 0.846

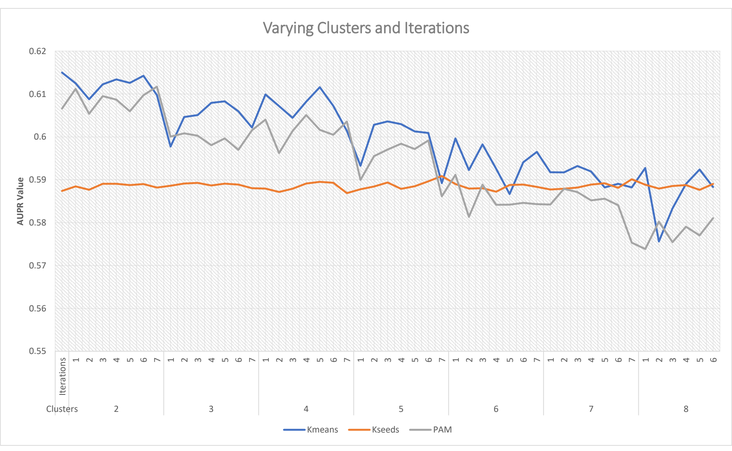
Mean AUPR (bottom):

_K_-Means – 0.600

K-Seeds – 0.589

PAM – 0.594





Discussion

Results compare favourably both with original DrugClust results (Liu dataset AUC – ~0.89, AUPR ~0.40, Mizutani dataset AUC ~ 0.89, AUPR ~ 0.40), Liu et al.'s results (AUC ~ 0.77), and Mizutani et al's results (AUC ~ 0.88, AUPR ~ 0.37 – 0.41).

Clusters having a slight negative correlation is expected, due to a larger choice when assigning drugs to clusters. Increasing folds is likely to have a positive relationship with AUC as it increases validation of the algorithm. The relationship between iterations and AUC/AUPR is less clear and still requires elucidating.

Next steps:

- Perform **further analyses**, significance testing, ANOVAs etc. to further explore relationships between folds, clusters, iterations, and AUC/AUPR
- Find more drug information and combine with current database to form a **novel database** for more robust predictions
- Remove features from database and **re-train** to see what features are strongest for predictions

Dimitri, G. M., Lio, P. (2017). DrugClust: A machine learning approach for drugs side effects prediction. *Computational Biology and Chemistry, 68,* 204-210.

Gao, Z., Chen, Y., Cai, X. (2016). Predict drug permeability to blood-brain barrier from clinical phenotypes: drug side effects and drug indications. *Bioinformatics, 33(6),* 901-908.