Permeability Estimation of the Drug Through The Blood-Brain Barrier

Osman Onur KUZUCU
CE Hacettepe University, n21131025@cs.hacettepe.edu.tr

Assoc. Prof. Tunca DOĞAN
CE Hacettepe University, tuncadogan@hacettepe.edu.tr

***Abstract* - Drug design is a very popular topic these days. A lot of researchers want to cure patients by designing drugs. Drug design is a very complicated area and hard to make predictions. This is a very time-consuming process. In this paper, proposed methods are built to help drug designers on blood-brain barrier permeability.**

Introduction

The blood-brain barrier is a very selective cell that blocks harmful molecules to pass through the neuron side. This structure has different endothelial cells. These are astrocytes and pericytes. Astrocytes are endothelial cells that control the biochemical molecules. Pericytes are endothelial cells that help to maintain the brain's homeostatic and hemostatic functions. Detailed structure is presented in Figure 1.

Pathogens, solutes, big or hydrophilic molecules are blocked by the blood-brain barrier. Blood-brain barrier allows hydrophobic molecules such as O_2, CO_2 and hormones to pass through the neuron side. Metabolic products such as glucose are transferred by transport proteins. In addition, the blood-brain barrier does not allow antibodies, immune cells to cross the barrier.

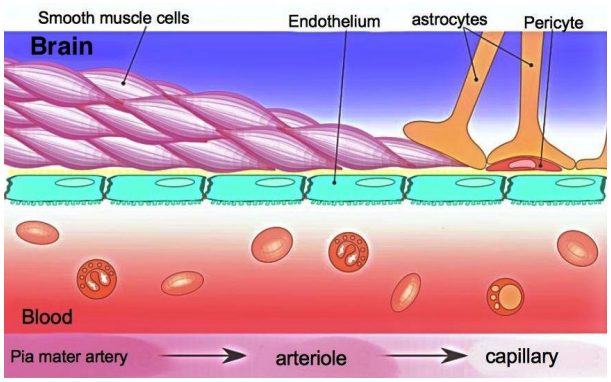

Figure 1: Structure of blood-brain barrier

In the blood-brain barrier, brain capillary endothelium blocks %100 of the big-molecule neurotherapeutics and %98 of small-molecule drugs. Some peptides are able to cross the barrier. Antibodies and therapeutic molecules are designed to cure disease in the brain. However, some useful molecules can not cross the barrier effectively. The main challenge to cure brain disorders is prediction of which molecule can cross the barrier.

Design development is a very time-consuming process. In this process, designers design drugs to cure diseases. Then, some tests are applied to this designed drug. If the drug can pass the test, the drug is applied on animals. After animal tests, if the drug has good results, the drug is applied to some people in increasing amounts. After all tests, the results are shown to the government. This process takes too much time. In some step in the process, it is hard to estimate that designed drugs can pass all the steps or which step drug can not pass. We want to help designers to solve these problems.

The design process and test process is difficult. Because of that we want to build a model that helps designers. The model makes a prediction for a designed drug that can cross the blood-brain barrier.

In this paper, our focus point is prediction that drugs cross the blood-brain barrier. In Section 2 we explain the dataset. In Section 3, related works are explained. In Section 4, the proposed methods are explained. In Section 5, results are shown. Our comments explain the results.

Section 2
Dataset

In this research, a blood-brain barrier dataset was used. In this dataset, there are some drugs that are able to penetrate the blood-brain barrier. The other drugs are not able to penetrate the blood-brain barrier. These informations are labeled in this dataset. Drug's chemical structure information is stored with the SMILES string.

Drug ids, drug's SMILES strings and labels are stored. Dataset has 3 parts. They are train, test and valid. There is a total of 2030 drug information in the dataset. In the train part 1421, in validation part 203 and in the test part 406 drugs are stored.

Section 3
Related Works

In "A Bayesian approach to in silico blood-brain barrier penetration modeling", researchers want to build a robust model to help real world drug research. In the research, Bayesian statistic, state of art machine learning methods were used. Random Forest, Support Vector Machines and 5-fold cross-validation processes are tested. Proposed model's accuracy is %83 that is for positive drugs. %96 accuracy is the result of negative drugs.

In "MoleculeNet: a benchmark for molecular machine learning" , researchers want to build a benchmark based on molecules. In this research, many datasets are used. One of the dataset is the blood-brain barrier dataset. The SMILES strings are used for featurization. ECFP, Coulomb matrix, frid featurizer, symmetry function, graph convolution and weave techniques are used to generate more features. Logistic regression, support vector machine, kernel ridge regression, random forest, gradient boosting, multitasking/singletask network, bypass multitask network, influence relevance voting, graph convolutional models, Weave models, directed tensor neural networks , ANI-1 and message passing neural networks are used. Proposed methods' results on the blood-brain barrier dataset are in Figure 2. Best conventional method is KernelSVM that performs %72.9 test accuracy. Best graph-based method is GC that performs %69 test accuracy.

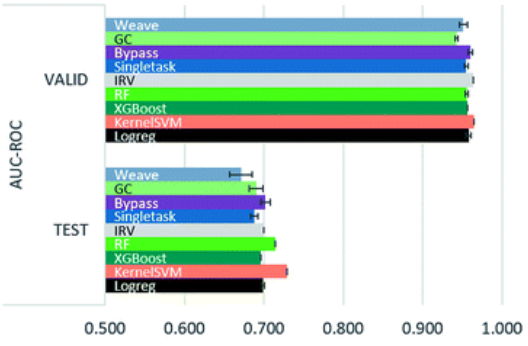

Figure 2: Results of proposed model based on blood-brain barrier dataset.

Section 4
Proposed methods

In our research, we want to build a robust method that will help drug designers. Proposed methods are worked with a blood-brain barrier dataset.In these methods, different algorithms are used.

In the research, all convolutional methods are listed in Table 1. In this research, many methodologies were designed and tested.

ID	Classifier	Note
1	Logistic regression	with Count vectorizer
2	Random Forest Classifier	with Count vectorizer
3	Tf-idf transformer	with Count vectorizer
4	Linear SVC	with Count vectorizer
5	Multinomial Naive Bayes	with Count vectorizer
6	Neural network(10,1)	with Count vectorizer
7	Logistic regression	with Hashing Vectorizer
8	Neural network(10,1)	with Hashing Vectorizer
9	Logistic regression	with Tf-Idf vectorizer
10	Graph Convolutional Network	4 GCN layers, 1 GR layer and 2 Keras dense layers(16,1)

Table 1: All methods are mentioned

Section 5
Results

In this research, we have worked on a blood-brain barrier dataset. Many different methodologies were used. All results are stated in Table 2.

Methods	Accuracy Result
Logistic regression with count vectorizer	%83.5
Random Forest with count vectorizer	%81.5
Tf-Idf transformer with count vectorizer	%81.7
Linear SVC with count vectorizer	%82.55
Multinomial Naive Bayes with count vectorizer	%82.75
Neural network with count vectorizer(2 Layers:10,1)	%80.3
Logistic regression with hashing vectorizer	%83.48

Neural Network with hashing vectorizer	%80.3
Logistic regression with tfıdf vectorizer	%80.59
Graph Convolutional Network (adam,mean squared error,10 epochs)	%74.5
Graph Convolutional Network (adam,mean squared error,20 epochs)	%74.63
Graph Convolutional Network (adam,binary cross entropy,20 epochs)	%74.87
Graph Convolutional Network (adam,binary cross entropy,20 epochs,learning rate: 0.01)	%74.89
Graph Convolutional Network (adam,categorical cross entropy,20 epochs,learning rate: 0.01)	%74.87

All experiments done in this research, the best method is logistic regression with count vectorizer. The second best method is logistic regression with a hashing vectorizer. The third best method is multinomial naive bayes with a counter vectorizer. Our prediction is that graph convolutional networks perform best. However the results surprised us.

Section 6
Conclusion

In this research, we wanted to build a method that helps drug designers while designing drugs. In this research, a blood-brain barrier dataset is used to train and test the models. Many models are designed to find the best methods. Our expectation is that graph convolutional networks can perform best. Because graphs store the relation between nodes. In our SMILES string store the relationship between molecules. Therefore, graph based methods can perform better than other methods. However, the results are surprising. The text based classification algorithms have performed better than graph based algorithms.

In future works, graph based methods can be developed to perform better. Some text based classification methods can be implemented. All methods' parameters can be adjusted. New vectorizer methods can be added.

References

[1] https://en.wikipedia.org/wiki/Blood%E2%80%93brain_barrier

[2] https://tdcommons.ai/single_pred_tasks/adme/#bbb-blood-brain-barrier-martins-et-al

[3] https://pubs.rsc.org/en/content/articlelanding/2018/sc/c7sc02664a

[4] https://pubs.rsc.org/en/content/articlelanding/2018/sc/c7sc02664a

[5] https://qbi.uq.edu.au/brain/brain-anatomy/what-blood-brain-barrier

[6] https://www.ncbi.nlm.nih.gov/books/NBK519556/#:~:text=The%20blood%2Dbrain%20barrier%20is,%2C%20i.e.%2C%20projections%20of%20astrocytes.

[7] https://fluidsbarrierscns.biomedcentral.com/articles/10.1186/s12987-020-00230-3