

**ACCESSIBLE AND RELIABLE MACHINE LEARNING MODELS FOR THE BLOOD-BRAIN BARRIER PERMEABILITY ASSESSMENT**

**Isakova A.M.** (ITMO University), **Shkil D.O.** (Syntelly LLC), **Steshin I.S.** (Syntelly LLC)  
**Scientific supervisor – Prof., MD, PhD, DSc Sergey Shityakov** (ITMO University)

**Introduction.** The blood-brain barrier (BBB) is a semi-permeable barrier that protects our nervous system from unwanted substances. The ability to penetrate the BBB is one of the most important parameters in pharmacology, especially when it comes to drugs for neurodegenerative diseases [1]. However, it is very difficult to measure permeability experimentally and scientists have high hopes for machine learning. To date, all high-quality machine learning models for this problem are either not publicly available [2], making it impossible to reproduce their results, or they have been trained on small data sets, so these models are not reliable enough [3].

**Main part.** The aim of our project was to develop more reliable publicly available machine learning models:

- 1) A binary classification model that determines whether the molecule passes through BBB.
- 2) A regression model that predicts the concrete logBB value: the higher the logBB value is, the easier the molecule overcomes BBB and enters the nervous system.

Different machine learning algorithms (CatBoost, XGBoost, LGBM, Random Forest, artificial neural networks) and various sets of input molecular features (Avalon, MACCS, ECFP6, etc.) were investigated. The key point is that we created our own set of fragment-based molecular descriptors specific to our task. We also introduced the idea of using a classification label as an additional feature for the regression task.

**Conclusions.** Publicly available machine learning models for classification and regression tasks were created. Fragments proved to be the optimal set of input data for all machine learning algorithms in the case of the regression task and for most algorithms in the case of the classification task. The use of the classification label as a feature can substantially increase the quality of regression models.

**References:**

1. Di, L., Rong, H., & Feng, B. (2013). Demystifying brain penetration in central nervous system drug discovery. //Journal of Medicinal Chemistry — 2013. —V.56 (1).
2. Mazumdar, B., Deva Sarma, P. K., Mahanta, H. J., & Sastry, G.N. Machine learning based dynamic consensus model for predicting blood-brain barrier permeability. // Computers in Biology and Medicine — 2023. —V.160.
3. Ciura, K., Ulenberg, S., Kapica, H., Kawczak, P., Belka, M., & Bączek, T. Assessment of blood-brain barrier permeability using micellar electrokinetic chromatography and P\_VSA-like descriptors. //Microchemical Journal — 2020. —V.158.