Material design assisted by machine learning

M. Druchok[a,b], D. Yarish[a], O. Gurbych[a], and M. Maksymenko[a]

[a] *Softserve Inc., 2D Sadova Str., 79021 Lviv, Ukraine*

[b] *Institute for Condensed Matter Physics of NAS of Ukraine, 1 Svientsitskii Str., 79011 Lviv, Ukraine*

Efficient design and screening of the novel molecules is a major challenge in drug and material design. In this contribution we present a multi-stage pipeline in which several deep neural networks are used to generate and validate novel molecular structures with the desired properties. Here the Autoencoder network is trained on existing structures to convert discrete molecular representations to continuous vector representation and reconstruct back the structure for a given vector in that space. An Attention-based Sequence to Sequence model "spell-checks" errors in the generated structures, while a fully connected Regressor type network is trained to predict desired molecular descriptors. In addition, we extend the scheme by adding few steps assessing the quality of the generated molecules. To this end, we use oversampling techniques in the continuous space to generate candidate structures and compute Synthetic accessibility score to assess the likeliness of the molecule synthesis.