EMPIRICAL EVALUATION OF QNTR MODELS BUILD ON PHYSICO-CHEMICAL CHARACTERISTICS

ANDĚL Michael[1, 2], KLÉMA Jiří[1, 2], TOPINKA Jan[1]

[1]*Institute of Experimental Medicine AS CR, Prague, Czech Republic, EU*
[2]*Czech Technical University in Prague, Faculty of Electrical Engineering, Prague, Czech Republic, EU*

Abstract

To model the quantitative relationship of the nanoparticle toxicity we can use theoretical molecular descriptors or physico-chemical characteristics. The former provide an auspicious interpretation of the toxicity mechanisms, however their computation may be very demanding, namely in the nanoscale. The latter are on the other hand fully observable, yet scarcely available for all the toxicity-assessed particles. Currently, there are large initiatives generating data for QNTR, including their toxicity and physico-chemical features. Resulting data are naturally very heterogeneous because of multiple subjects involved in the project. In this study, we investigate whether the data generated from such large projects are sufficient to induce well-generalizing models. We used the data generated by MODENA-COST, consisting of the toxicity measurements and physico-chemical characteristics of 192 nanoparticles. We build several machine-learning based models and focused on their statistical validity. The internal evaluation of these models (i.e. protocol using the same data set, such as cross-validation) suggests quite good validity of these models. Then we employed a rigorous validation protocol and external data set of our own measurements related to 10 standardized MeOx nanoparticles. Hence, the result were not so optimistic at all. Instead, they seem valid only for a well-defined set of experimental conditions. This research is supported by the Czech Ministry of Education, Youth and Sports (Grants No. LD 14002 and LO1508).

Keywords: Nano-QSAR, machine learning, nanoparticle characteristics

1. INTRODUCTION

Transferring the QSAR paradigm for nanoparticles is still a challenging task. The QSAR (quantitative structure-activity relationship) methods predict activity of a class of compounds, mainly the organic ones, based on their common molecular structure. The nano-QSAR predicts the activity of entire particles, namely the nanoparticles. The activity most researched in nano-QSAR is the particles *toxicity*, as the model-based toxicity predictions could facilitate the complicated controlling process of industrial nanoparticles.

However, as the conventional QSAR approaches profit from structural diversity of the organic molecules, the nanoparticles have quite simple chemical composition, but an immense variety of physical-chemical properties, such as surface structure, size-distribution, porosity, electrical potential, etc. These physical-chemical characteristics influence the events crucial for the particles kinetics, such as the *agglomeration* or a*ggregation* of the particles, their *sedimentation* or *uptake* by a cell, and thus influence their biological interaction and consequently their toxicity effect. The toxicity itself is then induced by quantum properties of the particle, often specific to the nanosize. The physical-chemical characteristics are affected by *design*, e.g. by its primary properties like shape, size, or by surface modification, but they actually origin from the quantum properties related to each compound and its higher crystalline or other macromolecular structure.

Obviously, it is difficult to properly model these properties and relations. Moreover, all these properties are induced by particle's interaction with the environment. And the environment differ, the experiment by

experiment. There are some relatively plausible models [1, 2], of nanoparticles toxicity, but their applicability is limited due to the variety of particles characteristics and heterogeneous experimental conditions that the toxicity assays are performed in. Moreover, the toxicity per se is actually a *hidden variable*, whose *observable* image is the *toxicity response*, which is measured by the means of toxicology assays. The response may further be affected by the *treatment time* (time of exposition), concentration of the particles or by the defence mechanisms of cell.

Generally, there are two fundamental approaches to model the nanotoxicity. The first [1, 2] tries to model the causal mechanisms of toxicity while uses mostly the quantum chemical descriptors. This computationally demanding approach can provide interpretable results, yet valid only in quite narrowly defined conditions, i.e. for the particles that were assessed under homogeneous experimental conditions. Such an approach does not fit for large pooled data sets, collected in the inter laboratory projects, such as MODENA-COST. The latter approach [3] employs specifically the *observable* physical-chemical characteristics. This approach nonetheless does not explain the causal mechanisms, and therefore could not be employed in order to manufacture the particles *safe by* design. Moreover, the physical-chemical characterization can be even more costly than the toxicology experiments. However, if there was properly described the relationship of particles toxicity and their physical-chemical properties, and on the other hand there was a mechanistic model of these properties such as [4, 5], we could further employ it in some composite model which would alternately use these approaches, learning one from the other.

In this paper, we present an extensive study documenting whether it is possible to learn (model) the toxicity response from the physical-chemical characteristics of the particles pooled by the MODENA-COST project. We propose a robust validation protocol to assess plausibility of the model.

2. DATA DESCRIPTION

We used the data collected during MODENA-COST project. The data set contains 189 measurements of two toxicity endpoints, i.e. the effective lethal concentrations EC25 [µg ml^{-1}] and EC50 [µg ml^{-1}], which were performed on 11 cell types, using altogether 4 cytotoxicity assays WST-1, MTT, ATP and LDH. The cells were altogether exposed to 46 nanoparticles of 12 different core-materials, altogether with 18 different coating (or without) for 2 - 114 hours. The experiments were done under 4 dispersion protocols with increasing energy input denoted as stirring, vortexing, bath sonication, cup horn sonication, and tip sonication. The data set contains also the following physical-chemical characteristics: 1) the shape of particles; 2) primary size of the particles, measured in two dimensions mostly by TEM; 3) specific surface, either calculated or measured by BET; 4) particles size in situ (DLS, NTA); and 5) zeta potential.

To sum up, each *record* (row) in the data set represents a toxicity assessment (expressed by EC25 and EC50) of a *particle,* performed on certain *cell line* under given *experimental conditions*, by the means of certain cytotoxicity *assay*. Each of these records represents a potential *example* for the learning (modelling) algorithm.

3. EXPERIMENTAL PROTOCOL

The most popular procedure for testing the machine-learning models is *cross-validation*. The aim of the cross-validation is to assess the accuracy of learning algorithm which was employed for building the model. Standardly, it splits the data *examples* into *k* disjoint subsets, leaves one of the subsets out for testing the learner which will have subsequently been learnt on the rest of the subsets. The predictions on the all left-out subsets are then aggregated into one accuracy measure. This aggregated accuracy measure is *estimation* of the model's general performance, i.e. of its *future* performance on unseen examples.

However, in our case the data examples are the toxicity experiments and related particles with their characteristics. It means that the data examples are *not mutually independent*, as there is standardly more particles made of he same material. Henceforth, when trying to extract the endpoint-relevant physical-chemical characteristics from the model, we cannot be certain whether they seem important for their *true meaning for the toxicity*, or because they are merely the features of several participating particles which are all made of the same toxic material. It also implies the uncertainty about the model's accuracy as follows. Having a seemingly accurate model, we are not able to distinguish whether the model is accurate for its generalization power which must hold for diverse types of materials, or simply because it have accurately classified only the particles of *prevailing material*.

Henceforth, we designed a robust validation protocol which mimics the real-life application of a model. The protocol is based on so-called *leave-one-label-out* cross-validation (LOLO) [6], where the examples are divided into the disjoint subsets consistently with some *third* labelling. Here, we therefore divided the data set according to the core-materials of respective particles. Then, for each material presented in the data set, we *left out* the related examples, where the particles were made of that material, for testing, and learned a model on the rest. The learned model was thereafter applied on the *left-out* examples to predict their endpoint. This way, the predictions made for each left-out material were deposited together with the true endpoint values. Finally, the *correlation coefficient* (ρ) between all the predictions and related true endpoint values was calculated. Thus, we obtained an unbiased estimation of the model performance and correctly assessed the real ability of the model for *inter-material generalizatio*n.

We used the Pearson correlation coefficient as the accuracy measure, because its intuitive interpretation, i.e. range between 0 and 1, where the values close to zero suggest random result. To distinguish which results are "close to zero", namely to determine whether the model's correlation coefficient implies merely the *correlation by chance*, we designed a *permutation test* and realized *a null distributio*n of the model. We randomly permuted the endpoint values of the examples inside the same material label (preserving the distribution inside of particles inside the material categories). Then, for each of the permutation, particularly we made 50 ones, we run the LOLO validation protocol as described a paragraph above. Hence, we obtained the accuracy measures (correlation coefficients) of 50 random models. The best of them was used as a threshold to determine whether the true (non-permuted) models stands above the null distribution. Finally, we filtered each model (of the true models, learned for each split of the data set), whose accuracy measure was *below* the accuracy of the *best performing* random model which had been learned on related random permutation.

4. RESULTS

First of all, we observed that the estimation of model accuracy by standard cross-validation procedure were indeed *over-estimated* as we supposed (see **Figure 1**). The accuracy estimation yielded with the standard cross-validated protocol (**Figure 1**, left) appears quite good (ρ = 0.93), and therefore we can suppose the related model will generalize well. But, in fact when employing our robust LOLO-based validation protocol we observe the results fundamentally different and worse (**Figure 1**, right). It implies that the model actually does not generalize, namely it does not generalize to the other types of material.

Henceforth, we were using our designed LOLO-based protocol together with the null-distribution filtering as described in Chapter 3. Finally, we observed only two experimental conditions consistently appearing in the significant results, namely the WST-1 assays where the respective models yielded the maximal accuracy ρ = 0.72 (see **Figure 2**), and the ATP assays, but only at the SK-OV-3 bone-marrow cell line, with the maximal accuracy ρ = 0.62. In **Figure 2**, left, left, we demonstrate the predictions for particles of certain material. We can observe that still for some of TiO2 and ZnO particles the model fails. In **Figure 2**, right, we can observe

the location of the former model (the red point) in the space of random models (the blue points). The location of true model is evidently outlying, which suggests significance of that result.

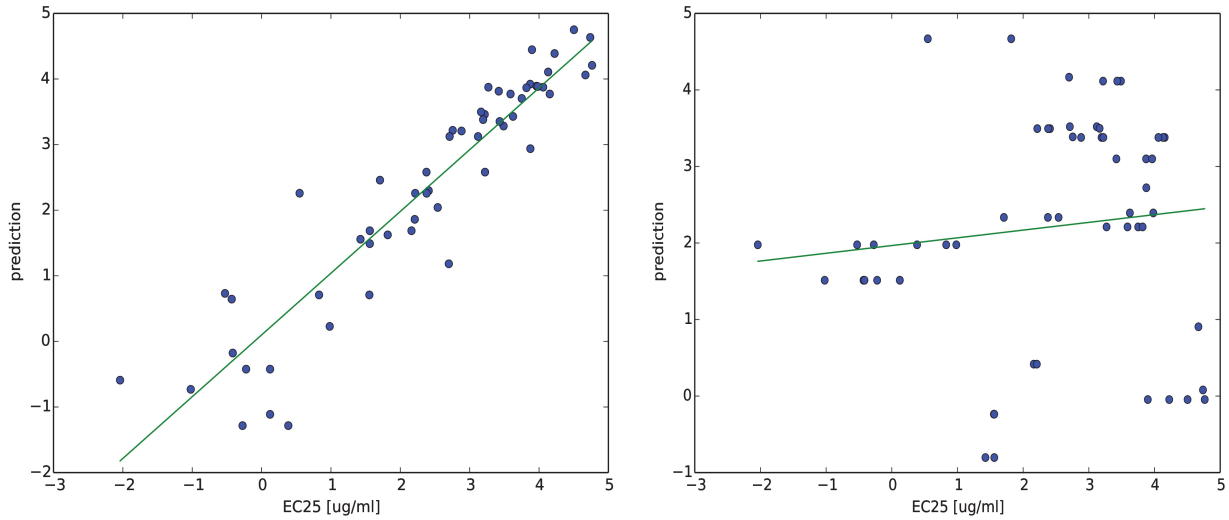

Figure 1 The predictions of the model learned on all the examples where the EC25 had been assessed with the ATP assay. The predictions yielded with the standard cross-validation protocol (left) are fundamentally worse than those yielded with our robust LOLO protocol (right)

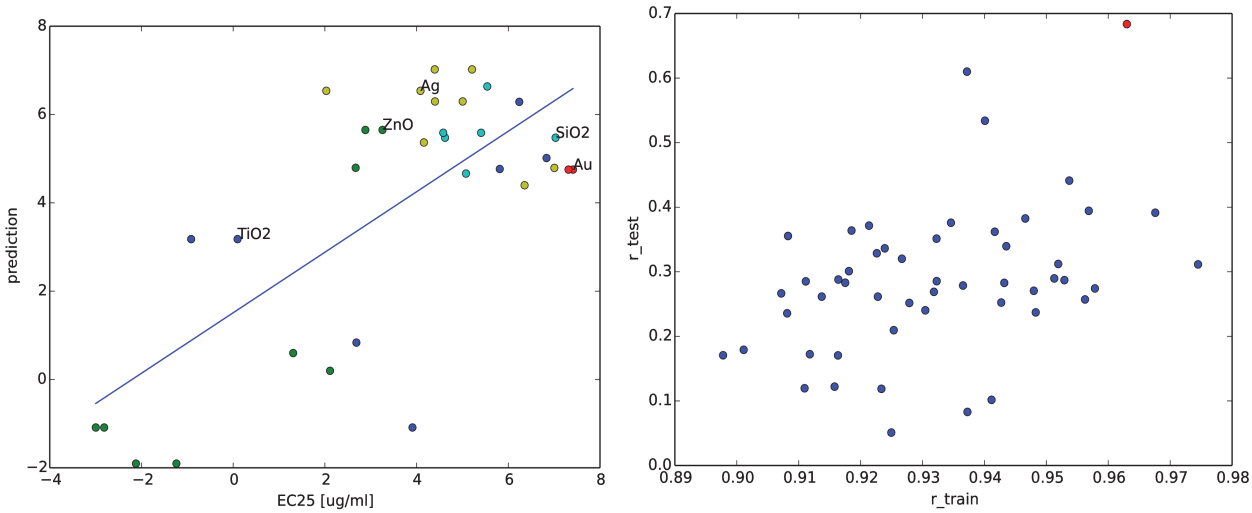

Figure 2 The unbiased predictions of a model learned on the examples where the endpoint (EC25) had been assessed with the WST-1 toxicology assay, with bath sonication. The location of that model in the space of random models depicts the red point in the right plot. The coordinates *x*, y represent the training and validation accuracy respectively

The structure of the model (from the **Figure 2**) itself is reported in **Figure 3**. We can observe that the physico-chemical characteristics are placed in the upside part of the tree. It suggests the primary importance of these characteristics. On the other hand, the conditions determining the particular cell lines are just in front of the leaf nodes. It suggests that the physical-chemical characteristics have primal meaning for the general toxicity, while the model eventually fits for the particular cell line. However, having more data for particular cell line could probably result in more accurate model, as it would be less fitted particular conditions (cell lines).

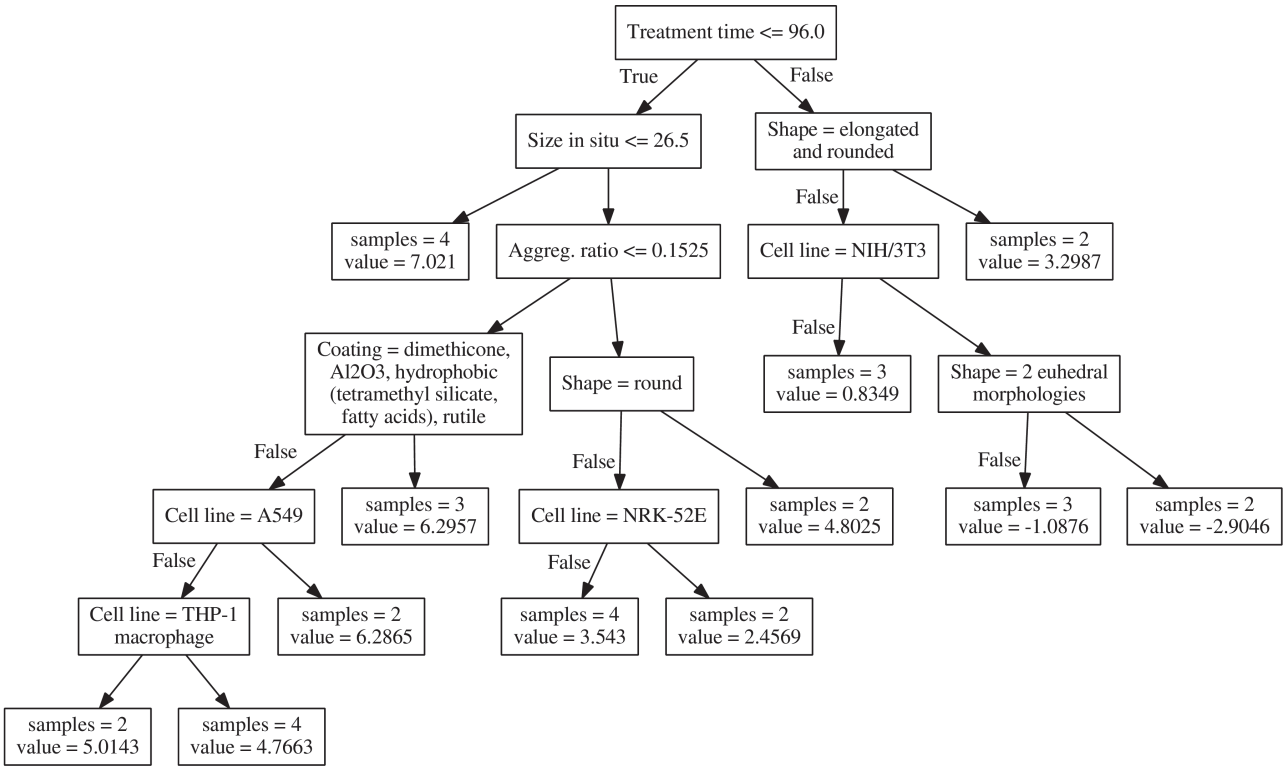

Figure 3 The decision tree of the model learned on the examples where the endpoint (EC25) had been assessed with the WST-1 toxicology assay, with bath sonication. The nodes stand for the features, the leaves for predicted endpoint values

5. CONCLUSION

From the results observed, we can deduce that it is possible to learn from pooled data sets of nanoparticle characteristics and related toxicology endpoints. The resulting models can potentially generalize to other unseen types of materials. Nonetheless, when validating the models it is necessary to use a robust statistical protocol which mimics the model application in the real life.

The aim of this work was nonetheless not to provide a general model of nanotoxicity. The reported models have still limited validity. The used data set is still too small, particularly regarding the fact that the data must be split somehow to deliver quite homogeneous experimental conditions. Moreover, there would be advisable to expand also the feature space of the data set, namely by measuring more physico-chemical characteristics of respective particles, or even to employ some quantum mechanical descriptors.

ACKNOWLEDGEMENTS

This research is supported by the Czech Ministry of Education, Youth and Sports (Grants No. LD 14002 and LO1508). More over we would like to acknowledge the COST action TD1204 MODENA for providing the data and establishing the network which this work has benefited from.

REFERENCES

[1] PUZYN, T., RASULEV, B., GAJEWICZ, A., et al. Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nature nanotechnology*, 2011, vol. 6, no. 3, pp. 175-178.

[2] GAJEWICZ, A., SCHAEUBLIN, N., RASULEV, B., et al. Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: Hints from nano-QSAR studies. *Nanotoxicology*, 2015, vol. 9, no. 3, pp. 13-325.

[3] SAYES, C., IVANOV, I. Comparative Study of Predictive Computational Models for Nanoparticle-Induced Cytotoxicity. *Risk Analysis*, 2010, vol. 30, no. 11, pp. 1723-1734.

[4] HOTZE, E. M., PHENRAT, T., LOWRY, G. V. Nanoparticle aggregation: challenges to understanding transport and reactivity in the environment. *Journal of environmental quality*, 2010, vol. 39, no. 6, pp. 1909-1924.

[5] MIKOLAJCZYK, A., GAJEWICZ, A., RASULEV, B., et. al. Zeta potential for metal oxide nanoparticles: a predictive model developed by a nano-quantitative structure-property relationship approach. *Chemistry of Materials*, 2015, vol. 27, no. 7, pp. 2400-2407.

[6] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., et a. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 2011, vol. 12, no. Oct, pp. 2825-2830.