Applicability domain and confidence in predictions
# Towards a more formal framework

Thierry Hanser, Chris Barber, Sébastien Guenes, Jean-François Marchaland, Stéphane Werner

*Lhasa Limited Granary Wharf House, 2 Canal Wharf, Leeds, LS11 5PS*

Lhasa Limited

■ Abstract

Applicability domain and confidence in an individual prediction are important aspects in QSAR modelling and key concerns when it comes to human safety assessment. Many approaches have been elaborated in order to estimate the validity of and the confidence in a prediction; however, there is no clearly unified definition of the concept of applicability domain and even less so a common framework for assessing the confidence in a prediction. Here, we will explore a new methodology that aims to clarify and formalise the assessment of a prediction. This new framework relies on the separation of concerns into three independent and staged aspects of the concept of confidence: applicability domain, reliability domain and decision domain. This new staged definition will provide the end user with clear and easy to interpret information describing the confidence in a prediction and we hope it will initiate a convergence towards a common framework within the QSAR community.
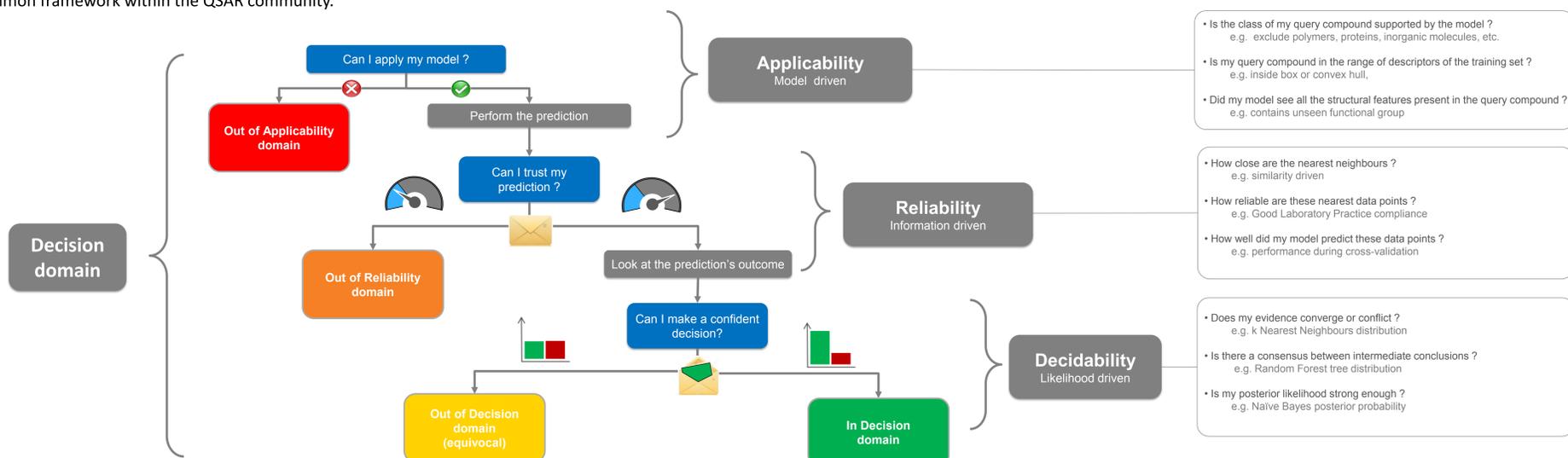
■ Why we need to be able to estimate the confidence in individual predictions ?

Before even addressing the problem of Applicability Domain we need to understand the very important concept of confidence in an individual prediction. If we had built a good model providing 83% accuracy on a challenging toxicity prediction task and would use this model in the context of screening thousands of compounds, we probably feel very confident since the numbers are in our favour. In fact it feels like going to the casino and playing with strongly biased dice; we know that in the long run we will win a lot of money. However, in the context of risk assessment, we need to focus on one individual prediction with potentially life threatening consequences. In this case, 83% is exactly a 5 out of 6 ratio with one in 6 chance of a lethal outcome. In the first case we are interested in the global accuracy of the model and in the second we want to estimate the accuracy for a very specific individual prediction. The question we ask in this latter case is "Can we trust this specific individual prediction?" and this is a very challenging question.
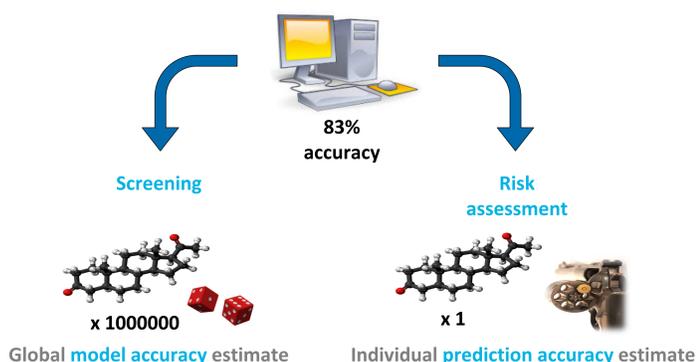

■ Separation of concepts

When we think of applicability domain we tend to mix several aspects into a single monolithic concept. We often attempt to merge the notion of descriptor range, information density and likelihood of the prediction's outcome (e.g. probability of the prediction to be correct) into one global metric. This is often a source of confusion and leads to multiple ways to combine these aspects. We propose disentangling these aspects into three separate concepts: Applicability, Reliability and Decidability. These three concepts form a cascade of requirements that gradually refine the confidence in a prediction. The resulting approach provides an intuitive, robust and transparent assessment of the confidence that can be assigned to an individual prediction.


■ Applicability

Applicability tells us whether the model can be applied or not for a given query. Typically a model can only be used for some type of queries (categories of molecules) based on the limitation of the model's algorithm or training history. It is important not to base the applicability of a model on the result of a prediction made with this model. Indeed it would be self contradictory to have to use the model to decide whether a model should be used or not. Unfortunately this is common practice. In this method the assessment of the applicability is conducted prior to any prediction. The nature of the query compound and the specifications of the model are the only criteria involved and lead to a binary answer "Yes I can apply the model" or "No I can't apply the model". In the latter case the query is said to be "Out of the applicability domain".
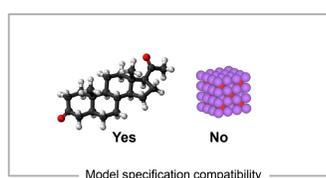
Applicability

Can the model be used to make a prediction for the specific query. The answer is driven by the specifications of the model and must be assessed prior to any prediction.

■ Reliability

Once it is established that the model is applicable, it becomes meaningful to perform the actual prediction and assess if this prediction is reliable. Reliability depends on how much information the algorithm has access to (quantity and quality of supporting evidence). This information is not dependent on the result of the prediction. We can think of the outcome of the algorithm as a closed envelope where the reliability captures how much we trust in the content of the envelope prior to opening it. The desired level of reliability is use case dependent and can be chosen by the end user. The reliability metric must be calibrated and normalised. Methods based on information density can be used for this purpose. If a prediction's reliability falls below the desired threshold, the prediction is said to be "Outside the reliability domain".
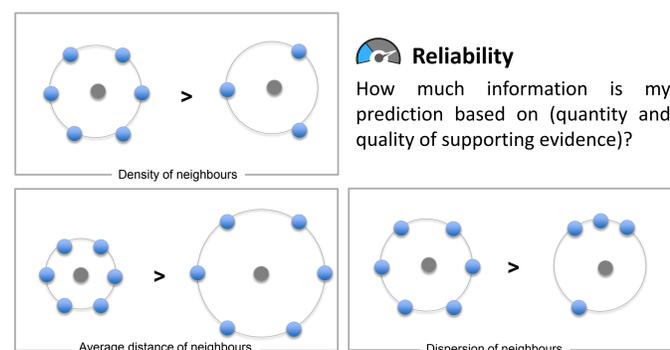

Reliability

How much information is my prediction based on (quantity and quality of supporting evidence)?

■ Decidability

Once we know that we have made a reliable prediction, we can look into the outcome of the algorithm. Decidability is telling us how assertive the model is (usually based on the likelihood provided by the algorithm). If the outcome is equivocal, the prediction can't be the base of a confident decision and is said to be "Outside the decision domain". Like reliability, decidability needs to be calibrated; the mathematical framework of conformal predictors can be used for that purpose.
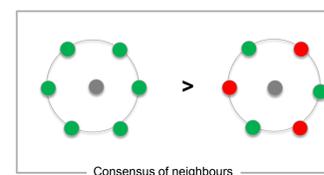
Decidability

How assertive is my model for the given prediction (likelihood provided by the model).

■ Conclusion

The cascading approach defines a rational chronology where the confidence in a prediction is refined to ensure that a confident decision can be made based on the prediction; this domain is called the Decision Domain. *"The Decision Domain is the scope in which it is possible to apply the model to make a non equivocal decision based on a reliable prediction".*

The proposed method offers the basis for an intuitive, rigorous and transparent framework to help the QSAR community adopt a common methodology that will provide the end user with consistent decision support across models and applications.

shared **knowledge** ● shared **progress**