Artificial Intelligence in the Life Sciences 4 (2023) 100087

Contents lists available at ScienceDirect

Artificial Intelligence in the Life Sciences

journal homepage: www.elsevier.com/locate/ailsci

Experimental Uncertainty in Training Data for Protein-Ligand Binding Affinity Prediction Models

Carlos A. Hernández-Garrido [a], Norberto Sánchez-Cruz [a,b,*]

[a] Instituto de Química, Unidad Mérida, Universidad Nacional Autónoma de México, Carretera Mérida-Tetiz Km. 4.5, 97357, Ucú, Yucatán, Mexico
[b] Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas Unidad Mérida, Universidad Nacional Autónoma de México, Sierra Papacál, 97302, Mérida, Yucatán, Mexico

ARTICLE INFO

Keywords:
Binding affinity prediction
uncertainty estimation
machine-learning

ABSTRACT

The accuracy of machine learning models for protein-ligand binding affinity prediction depends on the quality of the experimental data they are trained on. Most of these models are trained and tested on different subsets of the PDBbind database, which is the main source of protein-ligand complexes with annotated binding affinity in the public domain. However, estimating its experimental uncertainty is not straightforward because just a few protein-ligand complexes have more than one measurement associated. In this work, we analyze bioactivity data from ChEMBL to estimate the experimental uncertainty associated with the three binding affinity measures included in the PDBbind (K_i, K_d, and IC_{50}), as well as the effect of combining them. The experimental uncertainty of combining these three affinity measures was characterized by a mean absolute error of 0.78 logarithmic units, a root mean square error of 1.04 and a Pearson correlation coefficient of 0.76. These estimations were contrasted with the performances obtained by state-of-the-art machine learning models for binding affinity prediction, showing that these models tend to be overoptimistic when evaluated on the core set from PDBbind.

1. Introduction

One of the main tools in structure-based drug design is molecular docking [1–3], whose main goals are the identification of compounds able to bind to a macromolecular target as well as their binding modes [3]. For such a task, molecular docking software rely on two main components, a search strategy capable of exploring the space of binding conformations of a protein-ligand complex, and a scoring function capable of estimating the binding affinity of such conformations, its scoring function.

Scoring functions can be classified into two main groups: classical and machine-learning based scoring functions [4,5]. The first group assumes a functional form to establish the relationships between features characterizing a protein-ligand complex and its binding affinity (force fields, empirical scoring functions, statistical potentials). In contrast, machine-learning based scoring functions learn the connection between the features describing the system and its binding affinity through a machine-learning algorithm. Machine-learning scoring functions have shown to outperform classical scoring functions regarding their capability of accurately predicting the binding affinity of protein-ligand benchmark sets, such as those from the Comparative Assessment of Scoring Functions (CASF) [6].

Numerous papers have explored the use of different machine-learning algorithms for the development of scoring functions, going from algorithms based on decision trees such as random forest or gradient boosting trees (GBT), to artificial neural networks with different architectures, such as multilayer perceptrons (MLP), convolutional neural networks (CNN) or graph neural networks (GNN), which have been reviewed elsewhere [5,7,8].

The performance of these models is compared by calculating the Pearson correlation coefficient (R_p) and the root mean squared error (RMSE) between their predictions and the binding affinity of 285 protein-ligand complexes extracted from the PDBbind database [9], the named "core set". On the other hand, for training purposes these models employ either the 'general set', containing all protein–ligand structures in the database, or the 'refined set', containing a subset of high-quality structures from the "general set".

Although the PDBbind database represents the major source of

Abbreviations: CASF, Comparative Assesment of Scoring Functions; CNN, Convolutional neural networks; GBT, Gradient boosting trees; GNN, Graph neural networks; MAE, Mean absolute error; MLP, Multilayer perceptron; RMSE, Root mean square error; R_p, Pearson correlation coefficient.

* Corresponding author.
E-mail address: norberto.sanchez@iquimica.unam.mx (N. Sánchez-Cruz).

https://doi.org/10.1016/j.ailsci.2023.100087
Received 30 June 2023; Received in revised form 30 September 2023; Accepted 4 October 2023
Available online 4 October 2023

protein-ligand complexes with annotated binding affinity in the public domain, there are a couple of problems that cannot be overlooked. On the one hand is its documented bias towards preferred crystallographic targets [10], that will diminish only as more data become available. On the other hand, is the fact that the binding affinity data for distinct protein-ligand complexes comes from different sources (assays, laboratories, experimental conditions), leading to an inherent experimental uncertainty, which in principle delimits the achievable performance of the machine-learning models trained on such data.

Trying to estimate this experimental uncertainty in the PDBbind database is not straightforward because just a few protein-ligand complexes have more than one measurement associated and the bias towards preferred crystallographic targets is also present. An alternative is to estimate the experimental uncertainty by analyzing data from larger bioactivity databases where no structural information of the protein-ligand complexes is available, such as ChEMBL [11]. In [12], for instance, the authors analyze the experimental uncertainty on K_i measurements from ChEMBL version 12. However, data from the PDBbind database includes not only K_i, but also K_d and IC_{50} measurements, and most machine-learning models derived from such data treat them as the same target variable / endpoint. Although this is a reasonable approximation, combining different types of binding affinity estimation would only increase the uncertainty of estimations.

In this work, we analyze bioactivity data from ChEMBL version 33 to estimate the experimental uncertainty associated with the three binding affinity measures included in the PDBbind (K_i, K_d and IC_{50}), as well as the effect of combining them. These approximations are suggested to be applicable for cases where multiple independent measurements are not available, such as the PDBbind database.

2. Materials and methods

2.1. Data

Bioactivity data from ChEMBL version 33 was downloaded as a tab-separated value file from the European Bioinformatics Institute website using the following criteria: 1) Only measurements on single proteins and with BioAssay Ontology labeled as "single protein format" were included, 2) Only entries with bioactivity data (Standard Type) annotated as K_i, K_d and IC_{50} were kept. All further processing was performed using in-house Python scripting.

2.1. Data curation

Different sources of systematic errors have been identified when dealing with bioactivity data from ChEMBL [12]. In order to avoid as much as possible these sources of error, different preprocessing steps were needed as detailed below. The number of bioactivity data points preserved at each step is shown in Table 1.

Assay selection. All bioactivity data from ChEMBL is annotated with a target (Target ChEMBL ID), a small molecule (Molecule ChEMBL ID) and an assay (Assay ChEMBL ID). As protein mutants are associated with the same Target ID, the first step consisted in the elimination of assays performed on such mutants. For that, all assays including the words "recombinant" or "mutant" in either its assay parameters or descriptions were discarded, as well as those with an assay variant accession or an assay variant mutation annotated. To avoid data coming from high throughput screening campaigns, whose quality is not straightforward to guarantee, all assays including "HTS" or "screening" in their assay parameters, as well as those with more than 100 data points associated were also discarded.

Table 1
Data curation of ChEMBL data.

Step	Number of data points
Crude data	964,038
Assay selection	672,655
High quality data	477,962
Single measures of binding affinity	pK_i = 10,249
	pK_d = 3,549
	pIC_{50} = 35,000
Multiple measures of binding affinity	pK_i-pK_d = 1,200
	pK_i-pIC_{50} = 9,080
	pK_d-pIC_{50} = 4,012
	pK_i-pK_d-pIC_{50} = 13,685

High quality data. In order to select the binding affinity data to be included in the study, distinct criteria regarding data quality were included. Data points with no pChEMBL Value associated (pK_i, pK_d or pIC_{50}) or whose associated values were lower or equal than zero were discarded, this step removed data points with reported units distinct to nM. Data points associated with the Target ID CHEMBL612545 (ChEMBL ID for unchecked targets) were also removed. Only bioactivity data unequivocally assigned were preserved (measurements with standard relation other than "=" were discarded). If a protein-ligand pair had multiple measurements reported for the same assay, only the highest value was preserved (assuming it to be the one with optimal conditions). This partially curated dataset was the starting point for the construction of smaller subsets focused on single activity types (pK_i, pK_d or pIC_{50}) or the possible combinations between two or the three of them (pK_i-pK_d, pK_i-pIC_{50}, pK_d-pIC_{50} and pK_i-pK_d-pIC_{50}).

Single measures of binding affinity. The partially curated dataset was split into three subsets according to the binding affinity measure included (pK_i, pK_d or pIC_{50}). In the context of each subset, to exclude data coming from citations of previously reported values, if a protein-ligand pair had the same bioactivity value reported multiple times, only one was preserved. Finally, only protein-ligand pairs with at least two independent measures were considered for further analysis.

Multiple measures of binding affinity. To study the effect of combining binding affinity measures in the experimental uncertainty, subsets containing pairwise combinations between binding affinity measures were generated (pK_i-pK_d, pK_i-pIC_{50}, pK_d-pIC_{50}). For consistency with the previous methodology, if a protein-ligand pair contained more than one measurement for an affinity measure, only the highest affinity was preserved, and only protein pairs with at least one measurement of each affinity measure were considered for further analysis. Finally, a dataset containing data from the three binding affinity measures was built (pK_i-pK_d-pIC_{50}), which represents the worst-case scenario where all possible affinity measures are combined. In this last case, protein pairs with measurements for at least two affinity measures were considered for further analysis.

2.2. Measures of uncertainty

For each subset of binding affinity measurements, the differences for the possible combinations between pairs of measurements for each protein-ligand pair were obtained. To exclude rounded citation errors of previously published values, pairs with an absolute difference lower than 0.05 units were discarded. To exclude unit-transcription errors, data with an absolute difference equal to 3 or 6 units were also removed. The mean absolute error (MAE), the root mean square error (RMSE), as well as the Pearson's correlation coefficient (R_p) between the pairs of measurements were employed to estimate the uncertainty of the data. These metrics were calculated as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|M_{i,1} - M_{i,2}\right|$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(M_{i,1} - M_{i,2}\right)^2}$$

$$R_p = \frac{\sum_{i=1}^{n}(M_{i,1} - \overline{M}_1)(M_{i,2} - \overline{M}_2)}{\sqrt{\sum_{i=1}^{n}(M_{i,1} - \overline{M}_1)^2}\sqrt{\sum_{i=1}^{n}(M_{i,2} - \overline{M}_2)^2}}$$

Where $M_{i,1}$ and $M_{i,2}$ correspond to two different measurements for the same protein-ligand pair, where all n distinct protein-ligand pairs are considered for the calculations, while $\overline{M}_1$ and $\overline{M}_2$ are the mean affinity value found on each affinity dataset. These measures of experimental uncertainty can be seen as the best achievable performance of a model built on such heterogeneous data.

Datasets combining binding affinity measures were processed using the same procedure described for the datasets for single binding activity measures. In this case, the removal of data with an absolute difference lower than 0.05 or an absolute difference equal to 3 or 6 units was performed to keep consistency and not induce an artificial bias in the results, and the same metrics of uncertainty estimation were calculated.

2.3. Comparison with predictions from state-of-the-art models

Uncertainty metrics for all the datasets of binding affinity measures were contrasted with those obtained for the performance of machine-learning models. As models are tested on different benchmark sets and subsets of them, models selected for comparison include only those tested on the 285 protein-ligand complexes included in the "core set" from the PDBbind database showing state-of-the-art performance, defined as an R_p of at least 0.83. This selection included models using different machine learning algorithms and architectures. Brief descriptions of these models are provided below.

- AGL-Score [13] employs multiscale weight-colored subgraphs to describe molecular interactions where nodes represent atoms and its spatial position, and edges representing noncovalent interactions between atoms. A set of statistics is produced from the eigenvalues of the adjacency matrix and is used to train a GBT algorithm for the prediction of the binding affinity.
- ECIF [14] is another model employing a GBT algorithm for the prediction of binding affinity, but in this case, the input consisted of a vector with 1540 atom-pair counts found in protein-ligand complexes.
- AEScore [15] employs atomic environment vectors derived by combining radial and angular atom-centered symmetry functions in a one-dimensional vector that is further employed to feed multiple MLP, whose outputs are summed together to obtain the binding affinity estimation.
- OnionNet-2 [16] describes the protein-ligand interactions as a number of contacts between protein residues and ligand atoms in multiple distance shells which are converted into a 2D image that is processed by a CNN to perform the binding affinity prediction.
- graphDelta [17] employs a graph representation of the ligand and incorporates information about the target as node features in the form of atom-centered symmetry functions, and such representation is further processed by a GCN in order to perform the binding affinity prediction.
- PointTransformer [18] represents the protein-ligand system as 3D point clouds which are processed by different 1D CNN layers combined with multiple self-attention layers (ATT) to yield the final binding affinity prediction.

3. Results and discussion

3.1. Measures of uncertainty

Overall, ChEMBL version 33 includes binding affinity data measurements (defined as K_i, K_d or IC_{50}) for 793,823 different protein-ligand pairs. Multiple measurements of the same affinity measure are available only for a small fraction of them, being 1433 for K_d, 4200 for K_i, and 13,203 for IC_{50}, with an average number of measurements per protein-ligand pair of 2.44, 2.48, and 2.65 respectively (Fig. 1). It might be thought that for the datasets combining binding affinity measures, the upper limit on the number of protein-ligand pairs would be defined by the lower number of protein-ligand pairs within the datasets of single measures. However, this is not the case, as for the datasets combining affinity measures, protein-ligand pairs with single measurements for the corresponding affinity measures were also included, which explains the increased number of protein ligand pairs and protein targets, particularly for the datasets covering IC_{50}. Summary statistics and uncertainty estimations for all datasets are shown in Table 2.

Differences between affinity measurements in all datasets were found to be lower than 2.5 logarithmic units in more than 95% of the data (data not shown). This is in agreement with the empirical threshold proposed in [12] as the maximum tolerable agreement between different measurements. However, for this study, even data above that value was considered for the calculation of the measures of uncertainty. Distribution of affinity values for the datasets of single affinity measures and plots for the pairs of affinity measurements are shown in Fig. 2. Distributions of affinity values for the three affinity measures employed are slightly skewed to the left due to the natural behavior of scientific literature where high affinity compounds tend to be reported more often. The statistical parameters calculated to explore the experimental uncertainty showed that the dataset with less experimental uncertainty is the one associated with pK_d measurements, obtaining the lowest MAE and RMSE (0.69 and 1.03, respectively), as well as the highest R_p (0.76) among the three datasets of single measures. In terms of MAE it was followed by the pIC_{50} dataset (0.76), and in terms of R_p by the pK_i dataset (0.72).

Regarding the datasets combining affinity measures, in most cases they showed MAE, RMSE and R_p in the same range or higher than those for the single affinity measure datasets. The only exception is the pK_i-pIC_{50} dataset, where all metrics improved. This implies that although different activity measures show the same binding affinity tendency (high R_p), the individual differences in the affinity measures tend to be higher between measurements. In terms of MAE, RMSE and R_p, the pK_d-pIC_{50} dataset is the one with more experimental uncertainty (0.87, 1.12 and 0.69, respectively), as even the dataset combining the three affinity measures obtained better statistics. Pairs of affinity measurements for the datasets combining binding affinity measures are shown in Fig. 3.

3.2. Comparison with predictions from state-of-the-art models

The performance of a selection of machine learning models for binding affinity prediction tested in the "core set" of the PDBbind database is shown in Table 3. The R_p and RMSE values reported for these models can be contrasted with the R_p and RMSE reported herein. As the training data for these models comes from the PDBbind database, which combines pK_i, pK_d, and pIC_{50} as affinity measures, the closest estimations of their achievable performance are in principle the statistics calculated for the pK_i-pK_d-pIC_{50} dataset. The R_p and the RMSE values reported for these models are higher than those for the dataset herein built, especially for R_p. These findings indicate that machine learning models tend to be overoptimistic when assessing the general behavior of binding affinity values for protein ligand pairs (high R_p) but the individual predictions are already very close to the experimental uncertainty observed for the pK_i-pK_d-pIC_{50} dataset. It is worth noting that the comparison between these models and the herein obtained statistics is not straightforward, as the size and composition of the dataset where they are computed is different. So, considering the size of the core set from PDBbind and its known biases, it is not surprising to obtain such statistics.

Empirically, the R_p and RMSE of the pK_i-pK_d-pIC_{50} dataset would establish a limit of the achievable R_p and RMSE for any model trained on such heterogeneous data. Although this is observed to be true for the RMSE, that is not the case for R_p. This discrepancy on R_p values could

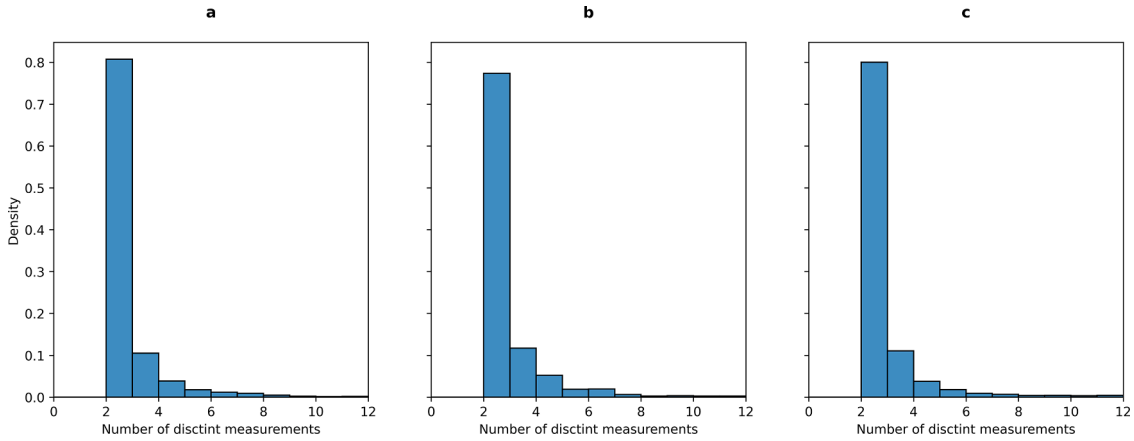

Fig. 1. Distribution of affinity measurements per protein-ligand pair. a: pK_i, b: pK_d, and c: pIC_{50}. For visualization purposes, the number of measurements was truncated to 12 as it covers 99% of the data. The maximum number of measurements was 27, 11 and 78 for pK_i, pK_d, and pIC_{50}, respectively.

Table 2
Summary statistics of experimental uncertainty for protein-ligand binding affinity.

Dataset	Number of unique targets	Number of unique protein-ligand pairs	MAE	RMSE	R_p
pK_i	548	4200	0.78	1.11	0.72
pK_d	327	1433	0.69	1.03	0.76
pIC_{50}	1235	13203	0.76	1.03	0.69
pK_i-pK_d	211	600	0.81	1.07	0.78
pK_i-pIC_{50}	717	4540	0.73	1.00	0.78
pK_d-pIC_{50}	577	2006	0.87	1.12	0.69
pK_i-pK_d-pIC_{50}	1092	6719	0.78	1.04	0.76

imply that machine learning models trained on structures from the PDBbind database are learning only the general tendency on binding affinity values on such biased set and not the intermolecular interactions captured in the structures of protein-ligand complexes, something that has been questioned previously [19,20], and that would limit their applicability domain.

4. Conclusions

The experimental uncertainty of binding affinity measures from different sources was estimated. Our results suggested that combining pK_d data from distinct sources results in the lowest experimental uncertainty, with a MAE of 0.69 logarithmic units, a RMSE of 1.03, and an R_p of 0.76. Combination of distinct binding affinity measures brings about an increase in uncertainty, where the lowest uncertainty was associated to the combination of pK_i and pIC_{50} measures, with a MAE of

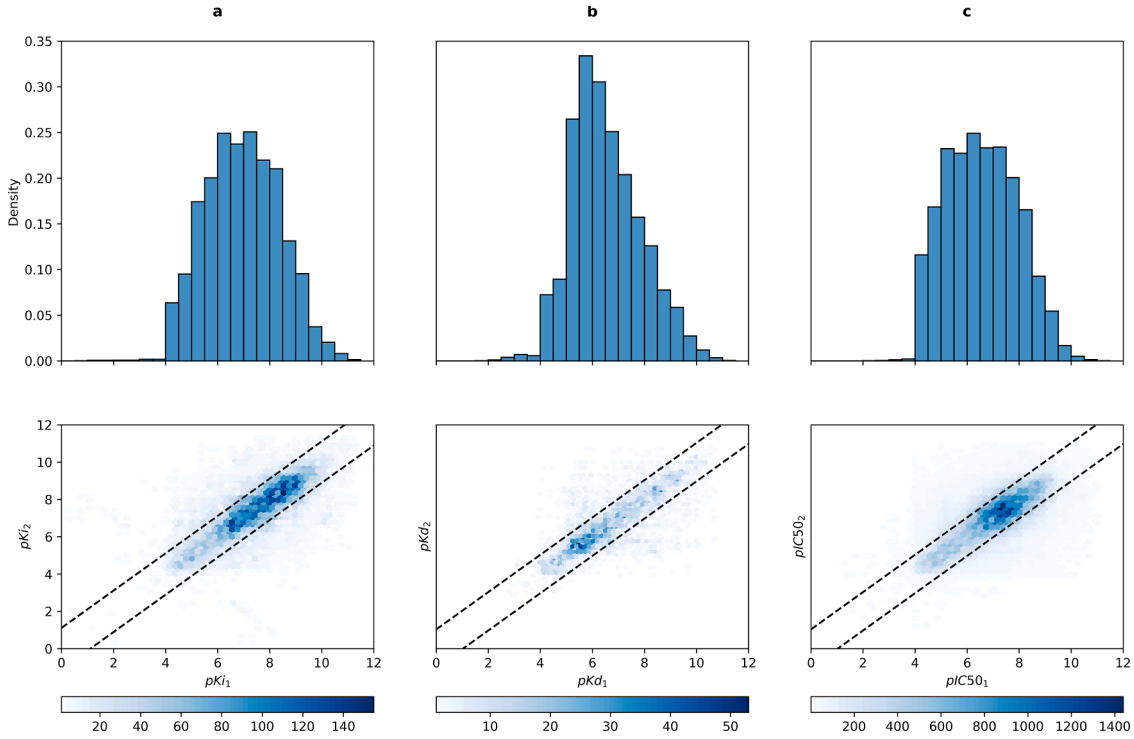

Fig. 2. Distribution of affinity values and pairs of affinity measurements for datasets of single affinity measures. a: pK_i, b: pK_d, and c: pIC_{50}. Dashed lines in the lower panel indicate the values covered by the RMSE.

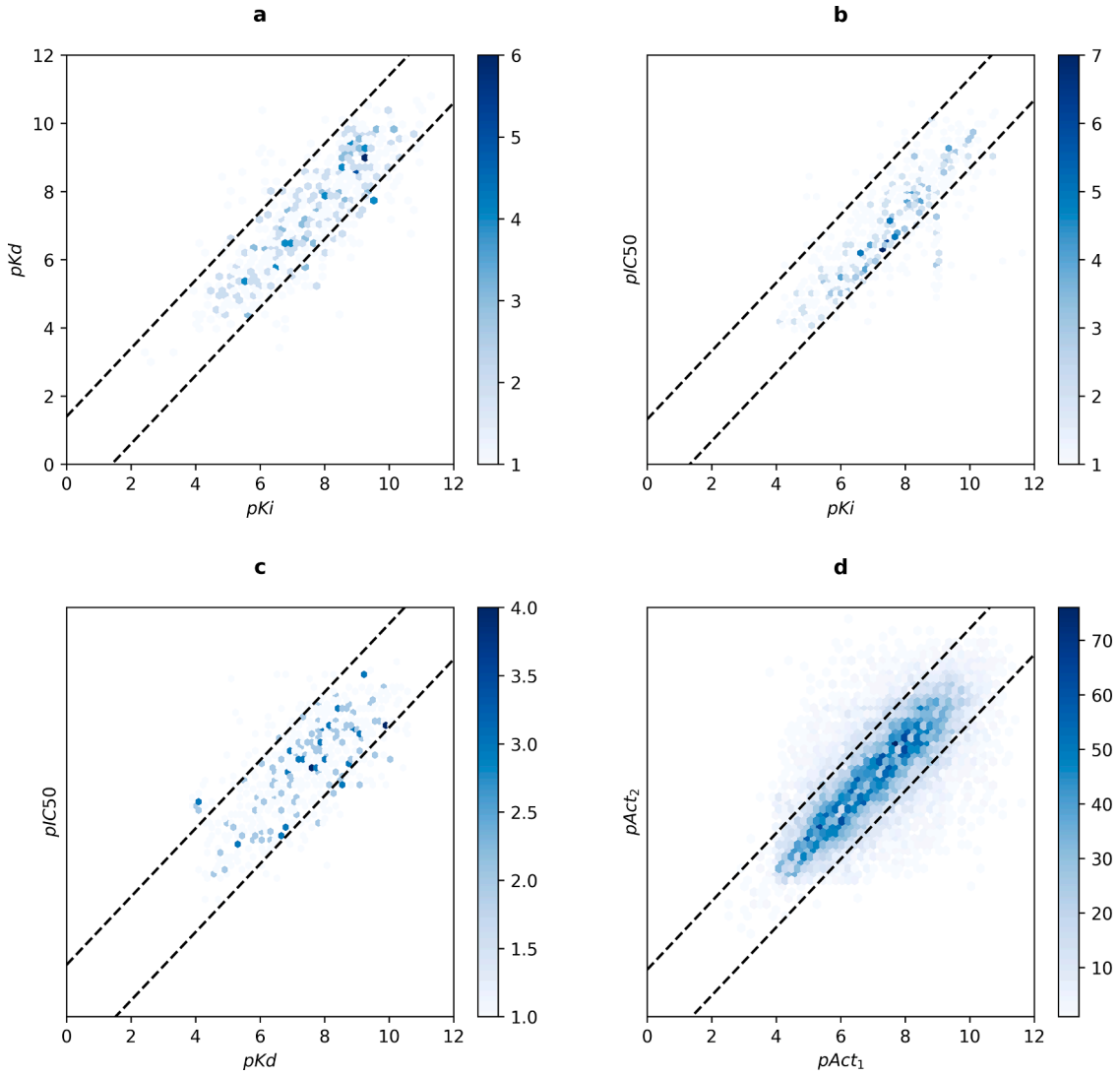

Fig. 3. Pairs of affinity measurements for datasets combining affinity measures. **a**: pK_i-pK_d, **b**: pK_i-pIC_{50}, **c**: pK_d-pIC_{50}, and **d**: pK_i-pK_d-pIC_{50}. Dashed lines in the lower panel indicate the values covered by the RMSE.

Table 3
Non-exhaustive selection of machine learning models for protein-ligand binding affinity prediction.

Model	Algorithm	R_p	RMSE	Reference
AGL-Score	GBT	0.83	1.27	[13]
ECIF	GBT	0.87	1.17	[14]
AEScore	MLP	0.83	1.22	[15]
OnionNet-2	CNN	0.86	1.16	[16]
graphDelta	GNN	0.87	1.05	[17]
PointTransformer	CNN+ATT	0.85	1.19	[18]

0.73, RMSE of 1.00, and R_p of 0.78. We also showed that even in the worst-case scenario (combining pK_i, pK_d and pIC_{50} measures), the uncertainty remains comparable, with MAE of 0.78, RMSE of 1.04, and R_p of 0.76.

By comparing these measures of uncertainty with those related to the performance of machine learning models for protein-ligand binding affinity prediction, we provide insights on how the assessment of performance of these models on a biased set such as the core set of the PDBbind using R_p and RMSE tend to be overoptimistic, suggesting that these metrics are not enough to assess the performance of models trained on such heterogeneous data, giving the wrong perception that we are achieving predictions with almost experimental precision. This agrees with the results obtained by modeling different levels of noise on synthetic numerical data and comparing how well these levels of noise are detected by distinct metrics. Such study suggests that R_p is not sensitive enough to the error present in the data, as modelled data with up to 50% of noise ratio achieve R_p of up to 0.80. According to that, metrics such as the Fractional Gross Error, the Variance Accounted For, and the Mean Normalized Bias seem as promising choices for estimating the performance of predictive models, as they showed to be more sensitive to the noise present in the data [21]. We hope this brief study will contribute to the ongoing discussion within the scientific community about the limits and applicability of machine learning models for binding affinity prediction trained on heterogeneous binding affinity data.

Data and Code Availability

All analysis were performed on public data. The raw data and code to reproduce all the results presented in this work can be found at https://doi.org/10.6084/m9.figshare.24031836.v2.

Declaration of Competing Interest

The authors declare that they have no known competing financial

interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work was funded by the Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) IA210023.

References

[1] Anderson AC. The Process of Structure-Based Drug Design. Chem Biol 2003;10:787–97. https://doi.org/10.1016/j.chembiol.2003.09.002.
[2] Batool M, Ahmad B, Choi S. A Structure-Based Drug Discovery Paradigm. Int J Mol Sci 2019;20:2783. https://doi.org/10.3390/ijms20112783.
[3] Lyu J, Wang S, Balius TE, Singh I, Levit A, Moroz YS, et al. Ultra-large library docking for discovering new chemotypes. Nature 2019;566:224–9. https://doi.org/10.1038/s41586-019-0917-9.
[4] Liu J, Wang R. Classification of Current Scoring Functions. J Chem Inf Model 2015;55:475–82. https://doi.org/10.1021/ci500731a.
[5] Ain QU, Aleksandrova A, Roessler FD, Ballester PJ. Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. Wiley Interdiscip Rev Comput Mol Sci 2015;5:405–24. https://doi.org/10.1002/wcms.1225.
[6] Su M, Yang Q, Du Y, Feng G, Liu Z, Li Y, et al. Comparative Assessment of Scoring Functions: The CASF-2016 Update. J Chem Inf Model 2019;59:895–913. https://doi.org/10.1021/acs.jcim.8b00545.
[7] Meli R, Morris GM, Biggin PC. Scoring Functions for Protein-Ligand Binding Affinity Prediction Using Structure-based Deep Learning: A Review. Frontiers in Bioinformatics 2022;2:885983. https://doi.org/10.3389/FBINF.2022.885983.
[8] Sánchez-Cruz N. Deep graph learning in molecular docking: Advances and opportunities. Artificial Intelligence in the Life Sciences 2023;3:100062. https://doi.org/10.1016/j.ailsci.2023.100062.
[9] Liu Z, Su M, Han L, Liu J, Yang Q, Li Y, et al. Forging the Basis for Developing Protein–Ligand Interaction Scoring Functions. Acc Chem Res 2017;50:302–9. https://doi.org/10.1021/acs.accounts.6b00491.
[10] Yang J, Shen C, Huang N. Predicting or Pretending: Artificial Intelligence for Protein-Ligand Interactions Lack of Sufficiently Large and Unbiased Datasets. Front Pharmacol 2020;11. https://doi.org/10.3389/fphar.2020.00069.
[11] Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, et al. ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res 2019;47:D930–40. https://doi.org/10.1093/nar/gky1075.
[12] Kramer C, Kalliokoski T, Gedeck P, Vulpetti A. The experimental uncertainty of heterogeneous public K i data. J Med Chem 2012;55:5165–73. https://doi.org/10.1021/JM300131X/SUPPL_FILE/JM300131X_SI_001.ZIP.
[13] Nguyen DD, Wei GW. AGL-Score: Algebraic Graph Learning Score for Protein-Ligand Binding Scoring, Ranking, Docking, and Screening. J Chem Inf Model 2019;59:3291–304. https://doi.org/10.1021/acs.jcim.9b00334.
[14] Sánchez-Cruz N, Medina-Franco JL, Mestres J, Barril X. Extended connectivity interaction features: improving binding affinity prediction through chemical description. Bioinformatics 2021;37:1376–82. https://doi.org/10.1093/bioinformatics/btaa982.
[15] Meli R, Anighoro A, Bodkin MJ, Morris GM, Biggin PC. Learning protein-ligand binding affinity with atomic environment vectors. J Cheminform 2021;13:1–19. https://doi.org/10.1186/S13321-021-00536-W/FIGURES/11.
[16] Wang Z, Zheng L, Liu Y, Qu Y, Li YQ, Zhao M, et al. OnionNet-2: A Convolutional Neural Network Model for Predicting Protein-Ligand Binding Affinity Based on Residue-Atom Contacting Shells. Front Chem 2021;9:753002. https://doi.org/10.3389/FCHEM.2021.753002/BIBTEX.
[17] Karlov DS, Sosnin S, Fedorov MV, Popov P. GraphDelta: MPNN Scoring Function for the Affinity Prediction of Protein-Ligand Complexes. ACS Omega 2020;5:5150–9. https://doi.org/10.1021/acsomega.9b04162.
[18] Wang Y, Wu S, Duan Y, Huang Y. A point cloud-based deep learning strategy for protein–ligand binding affinity prediction. Brief Bioinform 2022;23. https://doi.org/10.1093/BIB/BBAB474.
[19] Bajorath J. Deep learning of protein–ligand interactions—Remembering the actors. Artificial Intelligence in the Life Sciences 2022;2:100037. https://doi.org/10.1016/j.ailsci.2022.100037.
[20] Volkov M, Turk J-A, Drizard N, Martin N, Hoffmann B, Gaston-Mathé Y, et al. On the Frustration to Predict Binding Affinities from Protein–Ligand Structures with Deep Neural Networks. J Med Chem 2022;65:7946–58. https://doi.org/10.1021/acs.jmedchem.2c00487.
[21] Plevris V, Solorzano G, Bakas N, Ben Seghier M. Investigation of performance metrics in regression analysis and machine learning-based prediction models. In 8th European Congress on Computational Methods in Applied Sciences and Engineering; CIMNE 2022. https://doi.org/10.23967/eccomas.2022.155.