Predictive and Explanatory Uncertainties in Graph Neural Networks: A Case Study in Molecular Property Prediction

Anonymous Full Paper
Submission 39

Abstract

Accurate molecular property prediction is a key challenge in fields such as drug discovery and materials science, where deep learning models offer promising solutions. However, the widespread use of these models is hindered by their lack of transparency and the difficulty in assessing the reliability of their predictions. In this study, we address these issues by integrating uncertainty quantification and explainable AI techniques to enhance the trustworthiness of graph neural networks for molecular property prediction. We focus on predicting two distinct properties: aqueous solubility and mutagenicity.

By deriving explanations in the form of substructure masks, we obtain interpretable explanations in the form of chemically meaningful substructures that influence the model's predictions. Additionally, we incorporate uncertainty quantification to evaluate the confidence of both the predictions and their explanations. Our results demonstrate that (1) predictive uncertainty scores correlate with the accuracy of the predictions for both tasks, (2) uncertainties in the explanations also correlate with prediction correctness, and (3) there is a weak to moderate correlation between the uncertainties in the predictions and those in the explanations. These findings highlight the potential of combining uncertainty quantification and explainability to improve the trustworthiness of molecular property prediction models. [1]

1 Introduction

Molecular property prediction is a critical task in computational chemistry, material science, and drug discovery, where understanding the relationships between molecular structures and their properties can guide the discovery of new materials and therapeutics [2, 3]. Machine learning (ML), and particularly deep learning (DL) methods have revolutionized this field, enabling models to learn complex, high-dimensional representations of molecular data and provide accurate predictions for various molecular properties [4].

However, despite the promising performance of DL models in molecular property prediction, concerns about their reliability remain significant barriers to their widespread adoption, particularly in high-stakes domains like drug discovery or material design. These models often lack transparency in how they arrive at predictions, which can be problematic in safety-critical applications. The absence of interpretability and reliability can make it difficult for chemists to trust model outputs and make informed decisions. Thus, ensuring the trustworthiness of predictions is a critical step toward advancing the utility of these models in real-world applications.

Explainable AI (XAI) techniques have emerged as a solution to address some of these challenges. By providing interpretable explanations for model predictions, XAI allows users to gain insights into the underlying decision-making process, fostering confidence in the predictions [5]. In the context of molecular property prediction, XAI can reveal how specific molecular substructures contribute to the model's output, providing valuable insights for researchers and guiding further experimental investigations. Additionally, uncertainty quantification (UQ) has become an essential tool in assessing the reliability of model predictions [6]. By quantifying the uncertainty associated with a prediction, UQ helps identify regions where the model is less confident and may be prone to errors, allowing for more reliable and cautious decision-making.

In this work, we aim to bridge these critical aspects of deep learning models in molecular property prediction: uncertainty quantification, explainability, and their interplay. Specifically, we (1) show how uncertainty quantification can be applied to molecular property predictions to assess their trustworthiness, (2) show how substructure-mask-explanations can be used to interpret these predictions, and propose and compare several ways to determine what role uncertainties play in these explanations, (3) show that there is a correlation of uncertainty scores and correctness of predictions for both predictive uncertainties and explanation uncertainties, and (4) analyze the relationship between these different uncertainties.

Through these investigations, we aim to contribute to the development of more trustworthy and interpretable deep learning models for molecular property prediction.

[1] The code is available on https://github.com/anonymous-user3/NLDL2025-project, and is based on [1].

2 Background

In recent years, ML and DL techniques have emerged as powerful tools for molecular property prediction [4]. These approaches, particularly neural networks, can capture complex patterns in molecular data and provide accurate predictions across a wide range of tasks, including predicting solubility, toxicity, bioactivity, and mutagenicity. Graph neural networks (GNNs), in particular, have gained prominence due to their ability to directly operate on molecular graphs, which naturally represent atoms and bonds, thus preserving the inherent structure of molecules [7]. Despite the significant advances in predictive performance, challenges remain regarding the interpretability and reliability of deep learning models. These DL models offer little insight into the decision-making process, making it difficult to understand how certain molecular features contribute to the predicted properties.

In the context of molecular property prediction, XAI methods can help researchers understand which molecular characteristics, such as specific substructures, functional groups, or atom-level interactions, drive the predictions of a model. This interpretability is crucial for validating model predictions, especially in high-stakes domains such as drug discovery, where understanding the rationale behind a prediction can help researchers make more informed decisions and avoid potentially harmful or costly errors.

Various XAI techniques have been explored in molecular property prediction. These techniques include SHAP (SHapley Additive exPlanations) e.g. in [8], MolSHAP [9], LIME (Local Interpretable Model-agnostic Explanations), e.g. [10], and substructure-based explanation methods, e.g. [1]. In particular, substructure-mask-explanations have gained attention as they allow for the identification of chemically meaningful substructures that influence predictions. By highlighting these key substructures, researchers can not only gain insight into the behavior of the model but also identify potential areas for further experimental validation or molecular optimization.

While XAI helps to understand model behavior, UQ provides a complementary approach to assess the reliability of predictions. UQ focuses on measuring the confidence or uncertainty in a model's outputs, offering valuable information on the regions where a model might be less certain or prone to error [6]. In molecular property prediction, the incorporation of uncertainty quantification can help identify predictions that are likely to be incorrect, thus improving the overall trustworthiness of the model, and can help determine which molecules should be selected for further experimental testing [11]. A variety of UQ methods exists that be can used for this task, including ensemble-based methods and distance-based methods.

Table 1. Datasets and corresponding details used in this study. Size refers to the total number of molecules in a dataset before splitting into training, test and validation set. Metric refers to the evaluation metric used for assessing the performance.

Dataset	Task	Size	Metric
ESOL	regression	1111	MSE
Mutagenicity	classification	7672	AUC

Recent research has begun to explore the integration of XAI and UQ to provide a more comprehensive understanding of model reliability [12]. By combining UQ with XAI, not only can the certainty of the prediction being correct and the most likely explanation for the model's prediction be assessed, but also the confidence in the explanations themselves. This integrated approach has the potential to enhance the interpretability and trustworthiness of molecular property prediction models, offering a deeper understanding of the factors driving model decisions and their associated uncertainties. There remains a gap in understanding how the uncertainties in molecular property predictions relate to those in their corresponding explanations, which is essential for advancing reliable, interpretable, and actionable models.

3 Methods

3.1 Data

Two datasets are used in this study, one for predicting aqueous solubility (ESOL) and one for predicting mutagenicity. Each dataset was randomly split into training, test and validation data with a 80-10-10 split. Details about the data are shown in Table 1.

3.2 Model Construction

The molecular prediction model is constructed as a neural network ensemble [13] consisting of 10 relational graph convolution network (RGCN) models, as suggested in [1]. After three RCGN layers, attention pooling (weighted sum along the feature dimension over all nodes) is used to create a molecular embedding followed by three fully connected layers. Each model is initialized with a different random seed, leading to network diversification. The final prediction $F(x)$ for an input x is calculated as the average of the predictions of the 10 ensemble members:

$$F(x) = \frac{\sum_{i=1}^{m} f_i(x)}{m} \quad (1)$$

where m is the number of ensemble members and $f_i(x)$ is the prediction of ensemble member i.

3.3 Predictive Uncertainty Quantification

Two different methods for quantifying predictive uncertainties are used in this study: variance-based total predictive uncertainty (regression and classification) and softmax score (only classification).

3.3.1 Variance-based total predictive uncertainty

The predictive uncertainty can be measured by the variance of the individual predictions of the ensemble members for an input. If the ensemble is certain that a prediction is correct, then all ensemble members should align in their predictions, while a lot of variation means that the members disagree, and hence the overall ensemble is not sure. Formally, the predictive uncertainty is measured by

$$U_{pred}(x) = \frac{\sum_{i=1}^{m}(F(x) - f_i(x))^2}{m} \tag{2}$$

where m is the number of ensemble members, $F(x)$ is the ensemble prediction for input x and $f_i(x)$ is the prediction of ensemble member i for input x.

3.3.2 Softmax Score

For a classification task, the predictive uncertainty can be measured by using the values output from the softmax activation function after the last layer of a neural network for the predicted class, where a high softmax score means low uncertainty and a low softmax score means high uncertainty. The softmax values are negated to represent uncertainties instead of certainties to be comparable to other methods.

3.4 Substructure-Mask-Explanations

We base our analysis on three different ways of breaking molecules into chemically meaningful substructures: BRICS [14] uses a library of retrosynthetically feasible fragments, Murcko [15] splits molecules into a central core with side chains, and functional groups identify small local features linked to reactivity. In order to measure how good of an explanation a specific chemical substructure is, a prediction is calculated once for the whole molecule, and then a second time for the molecule when *masking* out all atoms that are part of this substructure. This masking is only applied after the message passing layers, i.e. during the attention pooling. For each atom that is part of the mask, the weight of its corresponding node will be set to 0 when creating the molecular embedding. Based on these two different predictions, we can measure the impact or *attribution* of this substructure on the prediction. The attribution score A is defined as the difference of all predictions for input x before applying the mask, i.e. the full molecule, and after applying the mask (x_{sub}), i.e. the molecule without the respective substructure.

$$A_{sub}(x) = \sum_{i=1}^{m} f_i(x) - \sum_{i=1}^{m} f_i(x_{sub}) \tag{3}$$

A value close to 0 should indicate that the masked substructure has little to no impact on the molecular property that is predicted. In contrast, a large absolute value should indicate that this substructure is of high relevance.

3.5 Explanation Uncertainty Quantification

The uncertainty of an explanation can be measured by calculating each ensemble member's attribution score separately and taking the variance of this.

$$U_{sub}(x) = \frac{\sum_{i=1}^{m}\Big(\big(f_i(x) - f_i(x_{sub})\big) - A_{sub}(x)\Big)^2}{m} \tag{4}$$

The overall uncertainty of the explanations can be defined in different ways. Here, we are specifically looking at four different option: Only taking into account the uncertainty of the most likely explanation, taking all uncertainties into account, the latter with additional weighing of importance of each uncertainty, and the same with a scaling method to ensure comparability between molecules.

$$sub^* = \arg\max_{sub} |A_{sub}(x)| \tag{5}$$

$$UX_{highest}(x) = U_{sub^*}(x) \tag{6}$$

$$UX_{all}(x) = \sum_{sub} U_{sub}(x) \tag{7}$$

$$UX_{weighted}(x) = \sum_{sub} U_{sub}(x) \cdot |A_{sub}(x)| \tag{8}$$

For the last method, the sum of the uncertainties of all possible explanations is weighted by the absolute attribution score for each explanation, meaning that large uncertainties get penalized more when the explanation is more likely. In order for this sum to be more comparable between different molecules and to be invariant to the number of substructure masks, we can scale this value first.

$$\alpha_{sub}(x) = \frac{|A_{sub}(x)|}{\sum_{sub} |A_{sub}(x)|} \tag{9}$$

Table 2. Performance on the test datasets for predicting solubility and mutagenicty.

Dataset	Metric	Result
ESOL	MSE	0.350
Mutagenictiy	AUC	0.896

$$UX_{scaled}(x) = \sum_{sub} U_{sub}(x) \cdot \alpha_{sub}(x) \qquad (10)$$

3.6 Evaluation

The network performance for the task of molecular property prediction is measured the mean squared error (MSE) for the predicting aqueous solubility on the ESOL dataset and by the area under the receiver operating characteristic curve (AUC) for predicting mutagenicity on the respective dataset.

To evaluate the preditive uncertainies, the relation between these uncertainties and the performance at molecular property prediction is studied. This is done by dividing the data into different subsets based on uncertainty scores and looking at the molecular property prediction for each of the subsets. Additionally, the performance on the test dataset is evaluated when excluding predictions with the highest uncertainty scores.

To evaluate the uncertainties of the explanations, the data is again divided into different subsets based on the explanation uncertainty, and the performance on each subset is calculated separately. This division into subsets is done separately for each explanation uncertainty method and for each of the three methods to determine substructure mask explanations.

The relation between the predictive uncertainties and the explanation uncertainties is evaluated by calculating the correlation coefficients for each combination of explanation uncertainty, predictive uncertainty, SME method and dataset.

4 Results

4.1 Molecular Property Prediction

The performance of the RGCN ensemble for the task of predicting aqueous solubility and mutagenicity is shown in Table 2.

4.2 Predictive Uncertainty

For both datasets a correlation between predictive uncertainties and correctness of the prediction could be observed.

When dividing the ESOL test data into three equally sized subsets based on uncertainties, the subset with the lowest uncertainties also has the lowest MSE (see Fig 1).

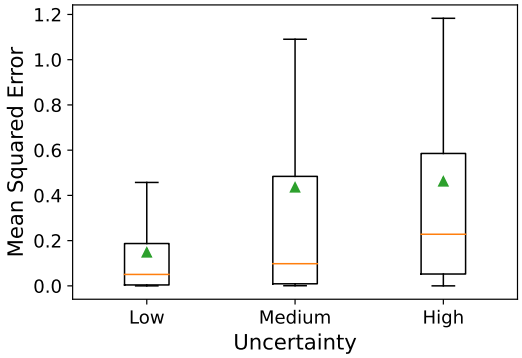

Figure 1. Mean squared error for predictions with low, medium and high predictive uncertainties for the ESOL test dataset. All subsets were chosen to be of equal size.

When using uncertainty thresholds to determine when a prediction can be trusted and when not, the overall performance improves (i.e., lower MSE) the lower the uncertainty threshold would be picked (see Fig. 2).

The results for different uncertainty thresholds for the mutagenicity prediction task are shown in Fig. 3 (performance when excluding data with high uncertainties) and Fig. 4 (likelihood of prediction being correct given the uncertainty score), showing that there is a relation between uncertainty scores and the correctness of the prediction.

4.3 Substructure Mask Explanations

An example of several substructure mask explanations for a molecule from the Mutagenicity dataset is shown in Fig. 5, using all three methods for dividing the molecule into meaningful chemical substructures (BRICS, murcko scaffolds and functional groups).

4.4 Explanation Uncertainty

The performance on data with low, low-medium, medium-high and high explanation uncertainties for the ESOL test dataset is shown in Fig. 6 and for the mutagenicity test dataset is shown in Fig. 7. For both datasets, the highest molecular property prediction performance is achieved on data with the lowest explanation uncertainties for all four explanation uncertainty metrics across all types of chemical substructures (BRICS, murcko scaffolds, functional groups), with the exception of functional groups for ESOL, where no significant relation between the explanation uncertainies and the MSE of the prediction could be observed.

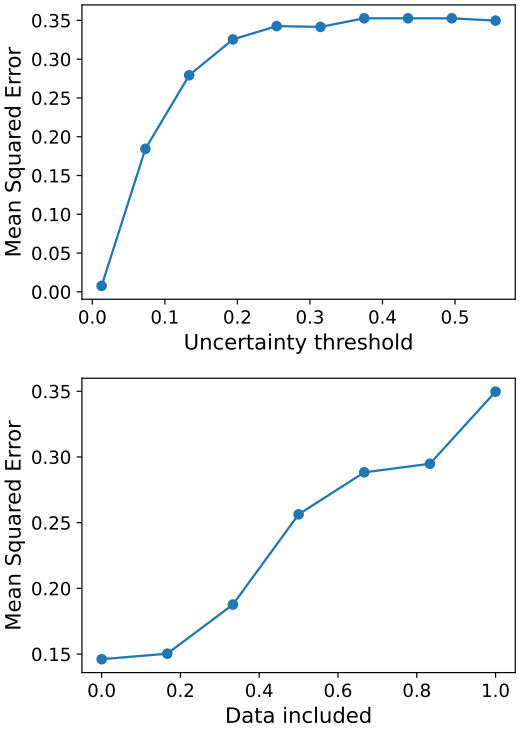

Figure 2. Mean squared error when only including predictions with the lowest predictive uncertainties on the ESOL test data. The upper row shows the performance based on the uncertainty threshold values, while lower row shows the performance based on how much of the data samples will still be included.

Table 3. Correlations between the predictive uncertainty and the different types of explanation uncertainties for the ESOL test dataset.

	brics	murcko	fg
$UX_{highest}$	**0.45**	**0.37**	**0.17**
UX_{all}	0.14	0.34	0.07
$UX_{weighted}$	0.06	0.29	-0.05
UX_{scaled}	0.35	0.44	0.15

4.5 Relation of Predictive Uncertainty and Explanation Uncertainty

On the ESOL dataset, weak to moderate correlations were found between the predictive uncertainties and the different types of explanation uncertainties (Table 3). The murcko substructures explanation uncertainties showed the strongest correlation overall with the predictive uncertainties. The highest correlations were found when using Eqs. 6 and 10 for calculating the explanation uncertainty.

The correlation between explanation uncertainties and predictive uncertainties for the mutagenicity dataset were found to be moderate to strong (see Table 4 and 5 for the results when using the ensemble variance and the softmax score respectively as the predictive UQ method). Similar to ESOL, the murcko substructure explanation uncertainties correlate the a lot with the predictive uncertainties, as well as the BRICS explanation uncertainties, however here, the highest correlations are found when using Eq. 7 as the explanation UQ metric.

Across all preditive UQ methods, all four explanation UQ methods and both datasets, the functional groups explanation uncertainties showed the lowest correlation scores. An example scatter plot of the relation between predictive uncertainties and explanation uncertainties is shown in Fig. 8.

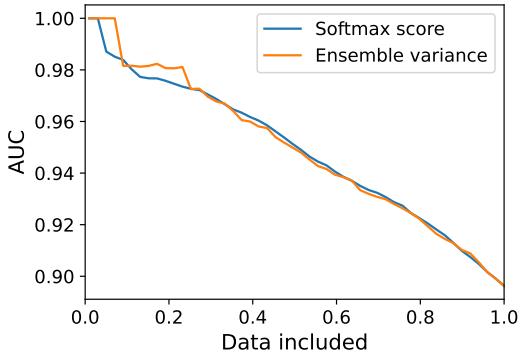

Figure 3. AUC when only including predictions with the lowest predictive uncertainty scores for the Mutagenicity test dataset.

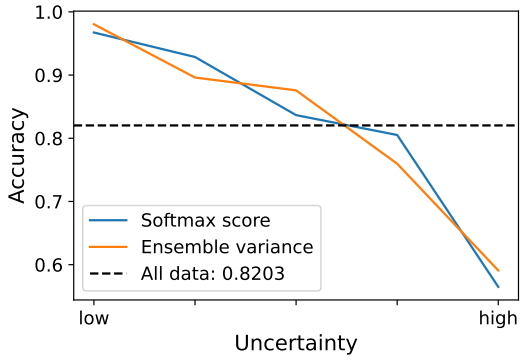

Figure 4. Likelihood of a prediction being correct given its predictive uncertainty score. The likelihood is calculated as the accuracy, i.e. the fraction of correct predictions over all predictions within a small range of uncertainty scores.

Table 4. Correlations between the predictive uncertainty (ensemble variance) and the different types of explanation uncertainties for the Mutagenicity test dataset.

	brics	murcko	fg
$UX_{highest}$	0.47	0.59	0.29
UX_{all}	**0.76**	**0.66**	**0.40**
$UX_{weighted}$	0.5	0.43	0.18
UX_{scaled}	0.24	0.52	0.25

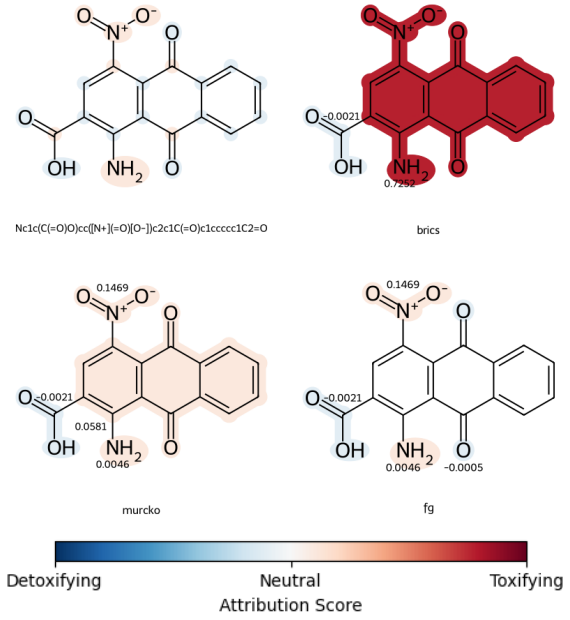

Figure 5. Substructure mask explanations for an example molecule from the Mutagenicity dataset. Red color refers to a positive impact of a substructure to the final prediction, and blue color refers to a negative impact. The intensity of the color describes how high this effect is. The attribution values of each substructure is written next to it.

Table 5. Correlations between the predictive uncertainty (softmax score) and the different types of explanation uncertainties for the Mutagenicity test dataset.

	brics	murcko	fg
$UX_{highest}$	0.24	0.36	0.13
UX_{all}	**0.57**	**0.46**	**0.24**
$UX_{weighted}$	0.24	0.2	0.03
UX_{scaled}	-0.03	0.27	0.10

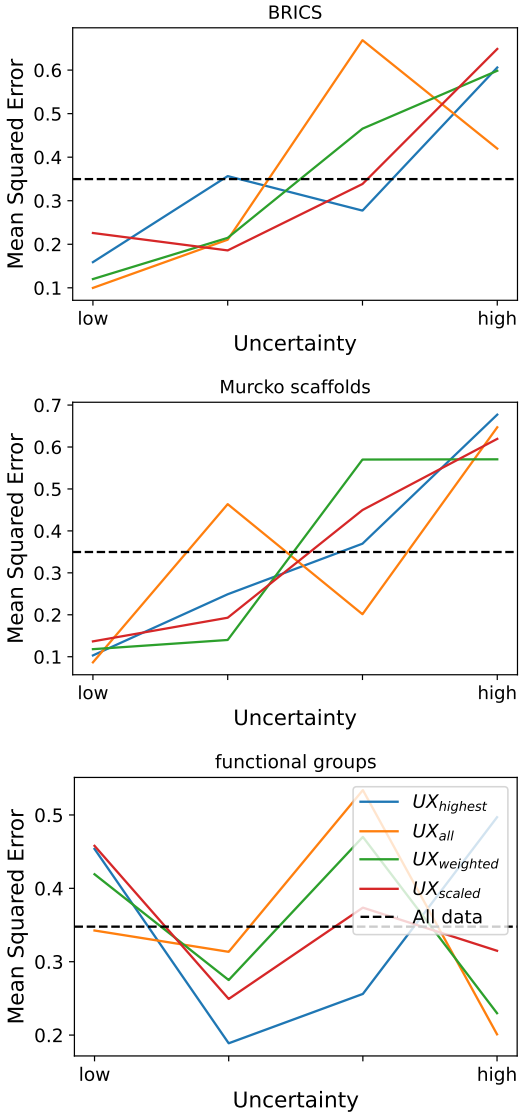

Figure 6. Mean squared error for predictions with low to high explanation uncertainties on the ESOL test dataset.

5 Discussion

One key observation of this study is that predictive uncertainty scores correlate with the accuracy of predictions, supporting the hypothesis that uncertainty estimation can be a reliable indicator of prediction correctness. As shown in Fig. 2 for ESOL and in Fig. 3 for Mutagenicity, samples with high predictive uncertainties can be excluded, which will lead to an improved overall performance of the remaining data. Although using uncertainty thresholds like this comes at the expense of not being able to make predictions for some molecules, it also has the benefit that the remaining predictions are more trustworthy and reliable, which can be crucial in high-stake applications such as drug design.

By dividing the data into different subsets based on their predictive uncertainties, this observation could be further confirmed. For ESOL, the MSE for low-uncertainty predictions was close to zero with a median of 0.05 (Fig. 1). For Mutagenicity, the likelihood of a low-uncertainty prediction being correct was close to 1.0, while the likelihood of high-uncertainty prediction was just above chance level (Fig. 4). This is a strong indicator for the argument that predictions with high uncertainties should not be blindly trusted, as there is a high chance that they are incorrect.

Furthermore, explanation uncertainties were shown to have a correlation with prediction correctness, emphasizing that they are not only of importance for increasing model interpretability, but that when the explanation is unsure or unclear, it is often connected to an unreliable prediction. Although this relation is clear, it is less strong than the relationship between the predictive uncertainty

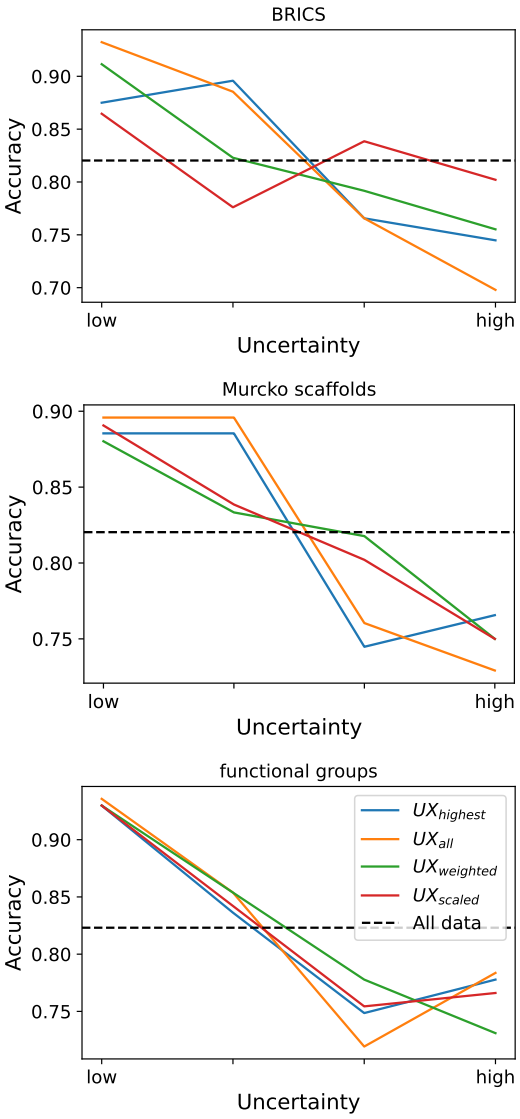

Figure 7. Likelihood of a prediction being correct given the uncertainty of the explanation for the Mutagenicity test dataset.

and the performance, which can be seen in Fig. 6 and 7. For mutagenicity, the accuracy of predictions for samples with low explanation uncertainties was around 0.87-0.95, and around 0.70-0.80 for samples with high explanation uncertainties, which is still a large performance difference, suggesting that predictions should not be trusted when the explanation uncertainty is high.

The findings indicate that predictive uncertainty and explanation uncertainty are interrelated, but the strength of this relationship varies depending on the dataset and uncertainty estimation method used. In the ESOL dataset, weak to moderate correlations were observed, with Murcko scaffold explanations showing the strongest alignment with predictive uncertainty (Table 3). In contrast, the mutagenicity dataset exhibited moderate to strong correlations, particularly when using ensemble variance as the

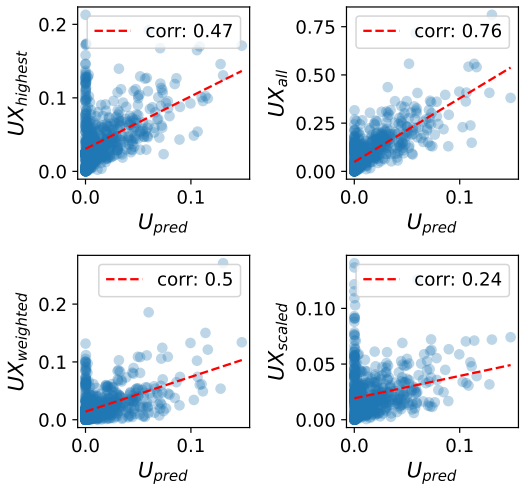

Figure 8. Correlation between predictive uncertainty and explanation uncertainty from the BRICS substructure explanations for the Mutagenicity test datset. Here, the predictive uncertainty was measured as the ensemble variance.

predictive UQ method (Tables 5 and 4).

The strongest correlations were observed when summing the attribution scores of all possible explanations within a single SME method. While BRICS and Murcko scaffold explanations demonstrated the highest correlation with predictive uncertainty, functional group explanations consistently showed lower correlation scores.

When comparing the different UQ methods used in this study, a similar performance is achieved between the variance-based ensemble uncertainty and the softmax score for measuring predictive uncertainty scores (Fig. 3 and Fig. 4). The four different explanation uncertainty methods all exhibited similar performance at detecting untrustworthy predictions, but they differ largely in their correlations with the predictive uncertainties. All UQ methods seem to offer valuable insights and no clear winning method could be determined.

Since it was shown that both the predictive uncertainties and the explanation uncertainties have a strong relation of their scores to the correctness of a prediction, but the correlations between these uncertainties are mostly only moderate (with a few being high or low), these results suggest that predictive uncertainty and explanation uncertainty provide complementary perspectives on model trustworthiness. The correlations, varying in strength, indicate that combining both approaches may yield a more comprehensive measure of reliability.

6 Conclusion

Using UQ methods can increase the trustworthiness of a model by excluding predictions that are

likely not correct with large predictive uncertainties. The interpretability of a model can be improved by including explainability methods.

For both datasets, a clear correlation between predictive uncertainties and correctness of the prediction could be observed. A similar behavior was found between the explanation uncertainties and the correctness of the prediction. When comparing the predictive uncertainties and the explanation uncertainties, positive correlations were found, with the strongest one when using the sum of the attribution scores of all possible explanations for a molecule within one SME method as the measure for explanation uncertainty. While some correlations are only weak, the results suggest that the different methods find different data samples untrustworthy, implying that a combination of all methods should be used to decide whether to trust a prediction or not.

Limitations and future work In the future, it should be explored how the different uncertainty measures could be combined into one meaningful, potentially more powerful, measure of trustworthiness. This could be done by training a small neural network that takes all the uncertainty scores and the prediction as input, and learns to predict whether this prediction was correct.

Furthermore, this study has the limitation of only investigating two datasets. This should be extended to more different datasets, with different molecular property prediction tasks.

Additionally, more uncertainty quantification methods for measuring the predictive uncertainties should be taken into account. Here, it might be interesting to also study methods which make it possible to decompose the total preditive uncertainy into its epistemic and aleatoric parts.

More advanced graph neural networks that take into account the geometry of the molecules could also be explored in the future.

References

[1] Z. Wu, J. Wang, H. Du, D. Jiang, Y. Kang, D. Li, P. Pan, Y. Deng, D. Cao, C.-Y. Hsieh, et al. “Chemistry-intuitive explanation of graph neural networks for molecular property prediction with substructure masking”. In: *Nature Communications* 14.1 (2023), p. 2585.

[2] W. P. Walters and R. Barzilay. “Applications of deep learning in molecule generation and molecular property prediction”. In: *Accounts of chemical research* 54.2 (2020), pp. 263–270.

[3] O. Wieder, S. Kohlbacher, M. Kuenemann, A. Garon, P. Ducrot, T. Seidel, and T. Langer. “A compact review of molecular property prediction with graph neural networks”. In: *Drug Discovery Today: Technologies* 37 (2020), pp. 1–12.

[4] Z. Li, M. Jiang, S. Wang, and S. Zhang. “Deep learning methods for molecular representation and property prediction”. In: *Drug Discovery Today* 27.12 (2022), p. 103373.

[5] D. Minh, H. X. Wang, Y. F. Li, and T. N. Nguyen. “Explainable artificial intelligence: a comprehensive review”. In: *Artificial Intelligence Review* (2022), pp. 1–66.

[6] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, et al. “A review of uncertainty quantification in deep learning: Techniques, applications and challenges”. In: *Information fusion* 76 (2021), pp. 243–297.

[7] P. Reiser, M. Neubert, A. Eberhard, L. Torresi, C. Zhou, C. Shao, H. Metni, C. van Hoesel, H. Schopmans, T. Sommer, et al. “Graph neural networks for materials science and chemistry”. In: *Communications Materials* 3.1 (2022), p. 93.

[8] D. Shi, F. Zhou, W. Mu, C. Ling, T. Mu, G. Yu, and R. Li. “Deep insights into the viscosity of deep eutectic solvents by an XGBoost-based model plus SHapley Additive exPlanation”. In: *Physical Chemistry Chemical Physics* 24.42 (2022), pp. 26029–26036.

[9] T. Tian, S. Li, M. Fang, D. Zhao, and J. Zeng. “Molshap: Interpreting quantitative structure–activity relationships using shapley values of r-groups”. In: *Journal of Chemical Information and Modeling* 64.7 (2023), pp. 2236–2249.

[10] C. M. C. Nascimento, P. G. Moura, and A. S. Pimentel. “Generating structural alerts from toxicology datasets using the local interpretable model-agnostic explanations method”. In: *Digital Discovery* 2.5 (2023), pp. 1311–1325.

[11] L. Hirschfeld, K. Swanson, K. Yang, R. Barzilay, and C. W. Coley. “Uncertainty quantification using neural networks for molecular property prediction”. In: *Journal of Chemical Information and Modeling* 60.8 (2020), pp. 3770–3780.

[12] C.-I. Yang and Y.-P. Li. “Explainable uncertainty quantifications for deep learning-based molecular property prediction”. In: *Journal of Cheminformatics* 15.1 (2023), p. 13.

[13] L. K. Hansen and P. Salamon. "Neural network ensembles". In: *IEEE transactions on pattern analysis and machine intelligence* 12.10 (1990), pp. 993–1001.

[14] J. Degen, C. Wegscheid-Gerlach, A. Zaliani, and M. Rarey. "On the art of compiling and using'drug-like'chemical fragment spaces". In: *ChemMedChem* 3.10 (2008), p. 1503.

[15] G. W. Bemis and M. A. Murcko. "The properties of known drugs. 1. Molecular frameworks". In: *Journal of medicinal chemistry* 39.15 (1996), pp. 2887–2893.