suitable methods to incorporate them during machine learning. Recognizing this gap, we have developed extensions to several established machine learning models, enabling them to effectively learn from censored labels while providing robust and reliable uncertainty quantification. Our results showed a particular advantage of including the censored labels for ensemble-based and Bayesian models but also enhanced Gaussian models in cases where >35% of the data contained censoring.

Through a comprehensive temporal study using internal pharmaceutical datasets, we demonstrated the importance of these extended models in accurately predicting key affinity scores, and side effects, of potential drug compounds, and their associated uncertainties. Our model comparison included approaches that estimate both aleatoric and epistemic uncertainties, providing a more complete understanding of prediction confidence. Specifically, we found that a straightforward ensemble of individually trained neural networks achieves generally high predictive accuracy. However, when also accounting for the calibration of uncertainty and computational cost, we recommend using Bayes by Backprop.

Thanks to the temporal evaluation, we were able to detect key differences between the distributions of target-based assays versus ADME-T assays. The ADME-T assays are more diverse in terms of the chemical space and thus exhibit less of a shift throughout time compared to the target-based assays where distinct differences could be observed for different time points. In light of these trends, we showed that the best-performing model in terms of predictive accuracy and calibration of aleatoric uncertainty are typically robust throughout time for ADME-T assays but not always for target-based assays. As such, an evaluation of the best model for ADME-T assays can be trusted to hold without reevaluation whereas target-based assays may need to be reassessed occasionally.

Finally, our case study showed how the uncertainty estimates from the most effective models can be practically applied to inform and guide ongoing drug development efforts, offering valuable insights for risk management and decision-making. Herein, we found that the epistemic uncertainty estimates correlate well with theoretical assumptions, but that the aleatoric uncertainty estimates require further analysis to understand their relation to the underlying, inherent noise in the data.

Acknowledgements

We thank our colleagues for their valuable feedback. This study was partially funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Actions grant agreement "Advanced machine learning for Innovative Drug Discovery (AIDD)" No. 956832. HRF, and AA are affiliated to Leuven.AI and received funding from the Flemish Government (AI Research Program).

Data Availability Statement

This work was conducted on internal data that cannot be disclosed beyond the information provided in Table 1. Despite this, we believe the work brings great value to the community by showcasing crucial aspects of the data that could not otherwise be analyzed. First, the temporal evaluation would not be possible to the same extent on public data, which led to conclusions about how the models compare over time. Second, while censoring in experimental labels are naturally occurring in internal data, it is less common in public data. Therefore, the key contribution of adapting and evaluating current uncertainty quantification approaches to censored labels, would not be possible without the proprietary data. All methodology will however be made available in our code, soon to be published on GitHub under the MIT license. Instructions to prepare the programming environment, as well as how to run the training, inference, and evaluation procedures on similar public data from Therapeutics Data Commons [60] can also be found there. Throughout this work, all experiments were run on a cluster of servers with diverse Nvidia GPUs using Python 3.11 with PyTorch 2.0.1 [46].

Disclosure of Interests

The authors have no competing interests to declare that are relevant to the content of this article.

References

[1] Lewis H Mervin, Simon Johansson, Elizaveta Semenova, Kathryn A Giblin, and Ola Engkvist. Uncertainty Quantification in Drug Design. *Drug Discovery Today*, 26(2):474–489, 2021.