Lucyde: A Demonstrator for Explainable Artificial Intelligence and Interactive Machine Learning

Eda Ismail-Tsaous[1] , **Ute Schmid**[1,2]
[1]Bavarian Research Institute for Digital Transformation
[2]University of Bamberg
eda.ismail-tsaous@bidt.digital

Abstract

As AI-based decision-support systems become increasingly widespread, methods aimed at improving the performance of human-AI teams are gaining attention. In recent years, explainable artificial intelligence (XAI) has received growing interest, as it provides methods to make the behavior of machine learning models more transparent and can help to identify errors and flaws, which is particularly important in safety critical domains such as medicine and law. However, explanations themselves can be misleading, inconsistent, or incorrect, which makes it essential to raise awareness of the possibilities and limitations of these methods. We introduce Lucyde, a web-based demonstrator designed to help users explore, compare, and better understand XAI methods across different datasets, models, and configurations. Lucyde provides a curated collection of explanation techniques, enables side-by-side comparison of methods, and offers easy-to-understand supplementary information for different user groups. It also illustrates interactive machine learning workflows by allowing users to correct model outputs or explanations. Lucyde thereby fosters informed and reflective engagement with AI systems and their explanations.

1 Introduction

As AI-based decision-support systems become increasingly integrated into real-world workflows, users are required to understand, interpret and evaluate their outcomes. Explainable Artificial Intelligence (XAI) can help make black-box machine learning (ML) models more transparent by revealing model errors and providing insight into how a system arrived at its outcome. In recent years, an extensive range of explanation techniques has emerged [Ali *et al.*, 2023; Schwalbe and Finzel, 2024] with the broader expectation that provided explanations will foster appropriate trust and support decision making, but the current body of research presents a highly heterogeneous picture, with findings spanning positive, null, and even negative effects [Rong *et al.*, 2024; Vaccaro *et al.*, 2024]. It is evident that the effectiveness and impact of explanations are highly context-dependent [Ismail-Tsaous *et al.*, 2025], and that individuals have different informational needs depending on their role, task and preferences [Schmid and Wrede, 2022]. Explanations describe the behavior of a model, not the underlying reality, and results for the same data point can vary substantially depending on the chosen model, parameters, and XAI technique.

Moreover, not only can system outputs be incorrect, but explanations themselves can also be inconsistent, misleading or wrong [Teso and Kersting, 2019; Molnar, 2025]. It is therefore essential to raise awareness of these possibilities and limitations, not only among machine learning experts, but also among lay users who increasingly encounter XAI methods in applications where explanations are presented as justifications.

In this work, we introduce Lucyde, a web demonstrator for XAI methods. The main idea behind this application is that seeing and comparing different explanations in different configurations across multiple implementations helps gaining a solid understanding of the strengths, weaknesses, and appropriate application contexts of ML models and XAI methods. With Lucyde, we aim to give experts and non-experts alike a structured overview of different types of explanations, easy-to-understand information, and a means to compare different XAI methods side by side. Our focus, however, is primarily on domain experts, especially in safety-critical areas such as medicine, as they stand to benefit the most from AI systems equipped with XAI capabilities. In addition, Lucyde provides insights into interactive machine learning approaches by illustrating how human-in-the-loop workflows can be realized by correcting wrong model outputs or explanations, storing the information and using it to retrain and refine a model at a later stage. To ensure that it can be accessed easily by anyone without installation overhead, this application is available online[1] and built with modern web technologies, featuring a React.js frontend and Django REST Framework backend.

2 Related Work

There are various tools and applications that pursue similar goals or approaches as Lucyde. However, many of them are targeted at machine learning experts. In the following, we

[1]www.lucyde.info

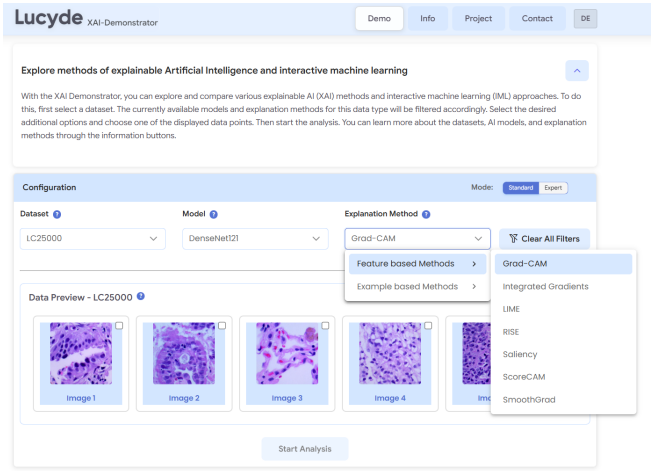

Figure 1: The selection interface allows to choose and combine a dataset, model, and explanation method. The available explanations are filtered depending on the chosen dataset and model.

present several examples and outline how they vary from our approach.

2.1 Comparison of Models and XAI Methods

Several systems focus on the comparative analysis of multiple models to facilitate diagnosis and selection. For instance, ModelWise [Meng *et al.*, 2022] provides a visual analytics workflow for comparing multi-class classifiers on tabular data, but uses only SHAP values [Lundberg and Lee, 2017] to align feature importance across different models. FACEE [Rezaei *et al.*, 2023] focuses on the evaluation of CNNs and provides a score to rank pairs of CNN models and saliency-based explainers. Visagreement [Silva *et al.*, 2025], also designed for experts, compares explanations, but focuses on the problem of explanatory multiplicity, i.e., when different XAI methods provide inconsistent or even contradicting explanations for the same model decision. The tool explAIner [Spinner *et al.*, 2019] allows to simultaneously display various XAI methods, but its main purpose is to compare different models using explanations as diagnostic tools.

2.2 Interactive Machine Learning

In recent years, the number of systems employing interactive approaches for model improvement has increased substantially [Liu *et al.*, 2025]. The already mentioned explAIner system [Spinner *et al.*, 2019] provides such a pipeline for model diagnosis and refinement. CorrectMe [Liu *et al.*, 2025] introduces an iterative workflow for object detection in which users can identify failure patterns and correct system outputs to incrementally retrain the model. However, although there are an increasing number of approaches aimed at improving explanations themselves [Cappuccio *et al.*, 2025; Slany *et al.*, 2022], we are not aware of any applications that implement this in a substantial way.

Lucyde is a multi-modal comparison tool aimed at both non-experts and experts, integrating the concepts outlined above for informational purposes.

Figure 2: The complete results page in comparison mode. At the top, the classification result and information about the data point are displayed. Below, the explanations generated by SmoothGrad and Score-CAM for this data point and this model (DenseNet121) are shown side by side in comparison mode.

3 Overview: Lucyde

In this section we present the key concepts of Lucyde. Note that this application is work in progress and is continuously being expanded and enhanced. It targets both lay users and machine learning experts, with different focal points depending on the audience: For laypeople and users with less technical expertise, the goal is to inform about AI-based classification systems, raise awareness about their limitations and show to what extent XAI methods can help to uncover such errors. For experts, we aim to develop a continuously growing collection and comparison tool for XAI methods complemented by in-depth information such as fidelity and robustness metrics, references and empirical evaluations.

3.1 Curated Collection of XAI Methods

One of the main objectives is to provide a solid overview of different explanation approaches using meaningful and illustrative examples, without claiming completeness. At present, Lucyde primarily covers local and feature-based explanation methods for various tabular and image datasets, such as Grad-CAM [Selvaraju *et al.*, 2017], SHAP [Lundberg and Lee, 2017], SmoothGrad [Smilkov *et al.*, 2017],

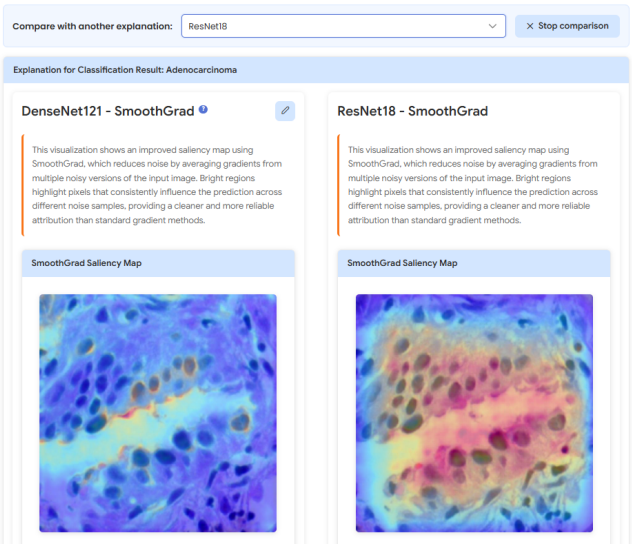

Figure 3: The explanation section for the same data point in comparison mode. The explanations are generated both with the method SmoothGrad using two different models: DenseNet121 and ResNet18.

RISE [Petsiuk *et al.*, 2018], and others. In addition, some example-based methods, such as "Near Hit & Near Miss" examples [Kim *et al.*, 2016; Finzel *et al.*, 2025], are implemented as well.

What makes Lucyde special is that it does not only show isolated examples of explanations, but provides examples of the same explanation method applied to different instances across different datasets and, additionally, to different classification models. As can be seen in Figure 1, users can select a combination of a specific data point from a dataset, a model and an explanation method. Based on this selection, the tool provides the model's classification result, a pre-generated explanation, and additional information about the chosen method. Currently, several correctly and incorrectly classified data points of tabular and image datasets from different domains are available to choose from, e.g., the Palmers Archipelago Penguin Dataset [Horst *et al.*, 2020], Lung and Colon Cancer data set (LC25000) [Borkowski *et al.*, 2019], Felix [Krügel *et al.*, 2025], etc. The available models are currently different CNN architectures such as EfficientNetB4 [Tan and Le, 2019], DenseNet121 [Huang *et al.*, 2017], ResNet18 [He *et al.*, 2015], or tree-based classifiers such as random forest [Breiman, 2001] or XGBoost [Chen and Guestrin, 2016].

For each dataset, model, and explanation method, additional information and references are provided, which can be accessed via the question-mark icons. These resources will be continuously expanded to include not only the original sources but also relevant extensions or empirical findings.

3.2 Comparison of Explanations

The core idea of Lucyde is that the strengths and weaknesses of different explanation methods become more apparent when they are evaluated side by side. For this reason, its central feature is the comparative presentation of explanations. Currently, there are two ways to conduct such comparisons: Figure 2 illustrates how for the selected data point and model, the result of an additional explanation method is displayed side by side. The second variant is shown in Figure 3: For the selected data point and explanation method, the result of an additional model is displayed side by side.

3.3 Work in Progress

In this section, we briefly outline two features that are still under development: the expert mode and the components for correcting system outputs and explanations.

Standard and expert mode. Information and explanation needs can be very different across possible target audiences [Schmid and Wrede, 2022]. To lower the entry barrier for users who may feel overwhelmed or discouraged by too much information or technical detail, in standard mode, concepts are explained as simply as possible and the number of options in the selection interface is reduced. However, more detailed information about datasets, models, and methods can still be opened on demand in the side panel.

The expert mode is aimed at machine learning practitioners, who are typically familiar with model architectures and parameters but may lack an overview of the broader spectrum of XAI methods beyond widely known approaches such as LIME [Ribeiro *et al.*, 2016] or SHAP [Lundberg and Lee, 2017]. When fully implemented, it will provide technical details, quantitative metrics to evaluate the explanations, and additional options to compare the results of specific methods with different parameter settings.

Correction of system outcomes and explanations. The limitations of both the machine learning models and their explanations make corrective mechanisms necessary. Lucyde is intended to illustrate different approaches to correcting both system outputs and generated explanations without introducing unintended side effects on the underlying model. We therefore provide example components that show how user feedback and corrections can be collected and managed.

4 Conclusion

Lucyde aims to promote informed, critical, and reflective use of AI systems by making the strengths, weaknesses, and appropriate applications of models and XAI methods more transparent and accessible. It enables users to explore multiple explanation methods across different models, datasets, and configurations, highlighting the variability, possibilities, and limitations inherent in both machine learning outcomes and XAI techniques. Through user corrections of model outputs and explanations, Lucyde illustrates how human oversight can refine model behavior in interactive machine learning settings. While other tools and applications are targeted at ML practitioners, this web-based application seeks to engage all user groups equally by enabling them to explore methods autonomously and adapt the level of detail to their own needs.

Acknowledgments

This work was funded by Bundesministerium für Forschung, Technologie and Raumfahrt (BMFTR) grant 01IS24067B.

References

[Ali *et al.*, 2023] Sajid Ali, Tamer Abuhmed, Shaker El-Sappagh, Khan Muhammad, Jose M. Alonso-Moral, Roberto Confalonieri, Riccardo Guidotti, Javier Del Ser, Natalia Díaz-Rodríguez, and Francisco Herrera. Explainable Artificial Intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence. *Information Fusion*, 99:101805, 2023.

[Borkowski *et al.*, 2019] Andrew A. Borkowski, Marilyn M. Bui, L. Brannon Thomas, Catherine P. Wilson, Lauren A. DeLand, and Stephen M. Mastorides. Lung and colon cancer histopathological image dataset (LC25000), 2019.

[Breiman, 2001] Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, October 2001.

[Cappuccio *et al.*, 2025] Eleonora Cappuccio, Bahavathy Kathirgamanathan, Salvatore Rinzivillo, Gennady Andrienko, and Natalia Andrienko. Integrating human knowledge for explainable AI. *Machine Learning*, 114(11):250, October 2025.

[Chen and Guestrin, 2016] Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 785–794, New York, NY, USA, 2016. Association for Computing Machinery.

[Finzel *et al.*, 2025] Bettina Finzel, Judith Knoblach, Anna Thaler, and Ute Schmid. Near hit and near miss example explanations for model revision in binary image classification. In Vicente Julian, David Camacho, Hujun Yin, Juan M. Alberola, Vitor Beires Nogueira, Paulo Novais, and Antonio Tallón-Ballesteros, editors, *Intelligent Data Engineering and Automated Learning – IDEAL 2024*, pages 260–271, Cham, 2025. Springer Nature Switzerland.

[He *et al.*, 2015] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[Horst *et al.*, 2020] Allison Marie Horst, Alison Presmanes Hill, and Kristen B Gorman. *palmerpenguins: Palmer Archipelago (Antarctica) penguin data*, 2020. R package version 0.1.0.

[Huang *et al.*, 2017] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017.

[Ismail-Tsaous *et al.*, 2025] Eda Ismail-Tsaous, Matthias Uhl, Celine Spannagl, Sebastian Krügel, and Ute Schmid. Context-dependent effects of explanations on multiple layers of trust. In *Proc. of ContEx 2025: Contextualizing Explanations, Bielefeld*, 2025.

[Kim *et al.*, 2016] Been Kim, Rajiv Khanna, and Oluwasanmi Koyejo. Examples are not enough, learn to criticize! criticism for interpretability. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, page 2288–2296, Red Hook, NY, USA, 2016. Curran Associates Inc.

[Krügel *et al.*, 2025] Sebastian Krügel, Jonas Ammeling, Emely Rosbach, Matthias Uhl, Christof Bertram, and Marc Aubreville. A validated task for studying human-AI interaction in multiclass image classification. Technical Report 5193553, Rochester, NY, March 2025.

[Liu *et al.*, 2025] Yinuo Liu, Zhiyuan Wu, Xiaoju Dong, Shaoxiong Jiang, and Weijie Li. CorrectMe: An interactive framework for human-in-the-loop correction and explanation of object detection models. *Proc. ACM Hum.-Comput. Interact.*, 9(8), November 2025.

[Lundberg and Lee, 2017] Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 4768–4777, Red Hook, NY, USA, 2017. Curran Associates Inc.

[Meng *et al.*, 2022] Linhao Meng, Stef van den Elzen, and Anna Vilanova. ModelWise: Interactive Model Comparison for Model Diagnosis, Improvement and Selection. *Computer Graphics Forum*, 41(3):97–108, 2022.

[Molnar, 2025] Christoph Molnar. *Interpretable Machine Learning*. 3 edition, 2025.

[Petsiuk *et al.*, 2018] Vitali Petsiuk, Abir Das, and Kate Saenko. Rise: Randomized input sampling for explanation of black-box models. In *British Machine Vision Conference (BMVC)*, 2018.

[Rezaei *et al.*, 2023] Ahmad Rezaei, Johannes Nau, Johannes Richter, Detlef Streitferdt, and Jörg Schambach. FACEE: Framework for Automating CNN Explainability Evaluation. pages 67–78, June 2023. ISSN: 0730-3157.

[Ribeiro *et al.*, 2016] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "[W]hy Should I Trust You?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 1135–1144, New York, NY, USA, 2016. Association for Computing Machinery.

[Rong *et al.*, 2024] Yao Rong, Tobias Leemann, Thai-Trang Nguyen, Lisa Fiedler, Peizhu Qian, Vaibhav Unhelkar, Tina Seidel, Gjergji Kasneci, and Enkelejda Kasneci. Towards human-centered explainable ai: A survey of user studies for model explanations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(4):2104–2122, 2024.

[Schmid and Wrede, 2022] Ute Schmid and Britta Wrede. What is missing in XAI so far?: An interdisciplinary perspective. *KI - Künstliche Intelligenz*, 36(3–4):303–315, 2022.

[Schwalbe and Finzel, 2024] Gesina Schwalbe and Bettina Finzel. A comprehensive taxonomy for explainable artificial intelligence: a systematic survey of surveys on methods and concepts. *Data Min. Knowl. Discov.*, 38(5):3043–3101, January 2024.

[Selvaraju *et al.*, 2017] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 618–626, 2017.

[Silva *et al.*, 2025] Priscylla Silva, Vitoria Guardieiro, Brian Barr, Claudio Silva, and Luis Gustavo Nonato. Visagreement: Visualizing and exploring explanations (dis)agreement. *IEEE Transactions on Visualization and Computer Graphics*, 31(10):7862–7875, 2025.

[Slany *et al.*, 2022] Emanuel Slany, Yannik Ott, Stephan Scheele, Jan Paulus, and Ute Schmid. *CAIPI in Practice: Towards Explainable Interactive Medical Image Classification*, pages 389–400. Springer International Publishing, 2022.

[Smilkov *et al.*, 2017] Daniel Smilkov, Nikhil Thorat, Been Kim, Fernanda Viégas, and Martin Wattenberg. SmoothGrad: removing noise by adding noise, 2017.

[Spinner *et al.*, 2019] Thilo Spinner, Udo Schlegel, Hanna Schäfer, and Mennatallah El-Assady. explAIner: A visual analytics framework for interactive and explainable machine learning. *IEEE transactions on visualization and computer graphics*, 26(1):1064–1074, 2019.

[Tan and Le, 2019] Mingxing Tan and Quoc Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6105–6114. PMLR, 09–15 Jun 2019.

[Teso and Kersting, 2019] Stefano Teso and Kristian Kersting. Explanatory interactive machine learning. In *Proc. of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, page 239–245, 2019.

[Vaccaro *et al.*, 2024] Michelle Vaccaro, Abdullah Almaatouq, and Thomas Malone. When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, 8(12):2293–2303, October 2024.