XAI in Genomics with LLM Generated Explanations in Oncology *

Isabelle Bichindaritz[1,*,†], Shazveen Saleem[1,†], and Leszek Kotula[3,†]

[1] *Biomedical and Health Informatics, State University of New York at Oswego, NY, USA*
[2] *Upstate Medical University, State University pf New York, Syracuse, NY, USA*

Abstract

The amount of genetic data available for analysis is rapidly growing, and the importance of analyzing this type of data is paramount to the field of precision medicine, in which diagnostic and prognostic tasks rely on genetic data analysis in many fields, including in oncology. Given the high dimensionality of genomic data, AI or machine learning in genomics requires more sophisticated methods to help clinicians interpret their results, as gene-level explanations may not be sufficient to make sense of these results. Given the growing importance of XAI, a rapidly growing number of AI papers in genomics refer to XAI methods. In this paper, we present general methods for XAI in genomics and focus on how LLMs can contribute to explanations at the case level in the domain of oncology. The preliminary results obtained in this study are encouraging about the role LLM chatbots can play in the user interface to help explain machine learning results to users at the case level in conversational or visual interaction modalities.

Keywords

case-based reasoning, XAI, LLMs, GenAI, survival analysis, oncology

1. Introduction

Precision medicine proposes to diagnose and treat diseases by adjusting diagnosis and treatment approaches to each patient, taking into account their specificities to tailor care to each individual (Chen et al., 2022). Instance-based machine learning and their case-specific explanations represent a promising avenue for precision medicine decision support (Toussaint et al., 2024).

Decision support systems for precision medicine often resort to larger, more precise sets of data for each patient, and often involve genetic data, and even multiomics data, to screen each patient more thoroughly as the genetic information may provide a more precise and fine-grained description of a patient's ultimate bodily functioning (Toussaint et al., 2024). Due to the size and complexity of this data, machine learning models have recently been poised to provide decision support tools for precision medicine (Hsu et al., 2026), in addition to classical biostatistical analytical methods.

If the predictive performance of machine learning algorithms has progressed tremendously, many recent works have focused on improving explainability (Lundberg et al., 2017). Machine learning researchers typically allude to a "black- box" effect (Budhkar et al., 2025) when input and output are understood, but the processing that occurs in-between is obscure. Methods leading to a greater level of explainability, such as models interpretable by design, are necessary in medicine since users are ultimately responsible for their clinical decisions and thus need to make informed decisions (Hsu et al., 2026).

However, when genetic data is involved like in precision medicine, XAI methods are particularly important as the feature set may be very large, and not straightforward to interpret by medical users (Toussaint et al., 2024). This article explores some of the roles LLM chatbots can play to tackle the task of explaining machine learning models on evolving genetic data, and presents some possible use cases in the task of survival analysis in oncology as a case study.

*ICCBR Workshop on Memory-Based Reasoning for Responsible GenAI at ICCBR2026, August 15, 2026, Bremen, Germany
* *Corresponding author.*
† *These authors contributed equally.*
ibichind@oswego.edu (I. Bichindaritz)
0000-0003-1712-490X (I. Bichindaritz); 1111-0002-7977-7124 (L. Kotula)

2. XAI Background

Rudin (2019) stresses the importance of machine learning models that are interpretable by design, and not only explainable after the fact (post-hoc). Further, Rudin et al. (2022) highlights ten challenges in XAI, with a focus on preferring interpretable models for high stakes decisions such as medical decisions, and methods to address them. Actually, most XAI tools recently applied in Genomics (Toussaint et al., 2024) to explain blackbox models are considered as model-agnostic, post-hoc, tools by Rudin et al. (2022). Rudin et al. (2022) refer to the following list of methods for interpreting machine learning Models. We also highlight counterfactuals as promising explanatory artifacts (Toussaint et al., 2024).

I. XAI models interpretable by design (Rudin et al., 2022): case-based reasoning, decision trees/lists/rules, scoring systems, generalized linear/additive models (GLMs/GAMs), physics/causal modeling, and disentangled neural networks (Chen et al., 2019) are all represented in this category.

II. SHAP: The SHapley Additive exPlanations (SHAP) model is a post-hoc agnostic, model-dependent, XAI model, based on feature attribution able to generate explanations for tabular and image data (Lundberg and Lee, 2017).

III. LIME: The Local Interpretable Model-agnostic Explanations model is a post-hoc agnostic XAI model, based on feature attribution able to generate explanations for tabular and image data (Ribeiro et al., 2016).

IV. Grad-CAM: The Gradient-weighted Class Activation Mapping model is a post-hoc agnostic XAI model, based on feature attribution able to provide visual explanations for image data based on the gradient values of the last convolutional layer from convolutional neural networks (CNNs) (Selvaraju et al., 2017).

V. Twins: ANN-CBR twins are a post-hoc agnostic XAI model, which proposes to explain deep learning using examples (Kenny and Keane, 2021). The explanation takes the form of a set of examples.

VI. Counterfactuals: Among explanation methods, grounded in cognitive psychology, counterfactual explanations stand out for their power of persuasion by generating contrastive scenarios based on such questions as "What would happen if?" (Byrne, 2019) to show what should be changed in current features to shift the current classification or risk into a positive outcome.

Because genomics programs are often written in R scripting language, a large number of XAI tools are available for R, such as *DALEX* (Biecek, 2018) and *iml* (Hothorn and Kook, 2025), as well as in Python libraries.

3. XAI in Genomics

We conducted a survey of the biomedical literature in AI in Genomics by searching in *Pubmed* using MESH global terms (query = "Artificial Intelligence"[Mesh]) AND ("Genomics"[Mesh] OR "Multiomics"[Mesh]) AND explainable), we obtained 224 papers, with a sharp increase in 2025 (see Fig.1), which we analyzed for XAI methods and evaluations. We also performed specific searches at the intersection of AI, Genomics and the XAI methods listed.

Toussaint et al. (2024) developed a classification map for XAI methods in genomics and multiomics including 10 facets on 405 papers. Of particular interest, they noted that, if the main XAI methods used in genomics are feature relevance, simplification, local explanation, and visual explanation, other methods highlighted in the XAI literature were promising and too

underrepresented, namely explanation by example (Papernot and McDaniel, 2018) and text explanations (Feldhus et al., 2023). Further they propose to investigate how explanation by example can improve post-hoc analysis of this data. Several other papers stress the importance for the future of example-based and counterfactual and example-based explanations (Gates et al., 2023; Bhudkar et al., 2025; Hsu et al, 2026).

Based on this literature review, it is clear that the main quality searched for in terms of explainability in this domain is feature relevance (Toussaint et al., 2024). All machine learning models in this domain select features to simplify their models (feature selection), while they often also practice feature attribution for explainability purposes. Both processes often work synergistically for biomarker discovery. Hence all papers reviewed were interested in biomarker discovery (Toussaint et al., 2024), and discussed both the confirmation of known biomarkers of certain diseases, and the discovery of potentially interesting novel biomarkers, which could then lead to research experiments to confirm their clinical relevance. In that sense, this major quality searched for relates as well to model sparsity since one of its effects is to reduce the input set of the highly dimensional data.

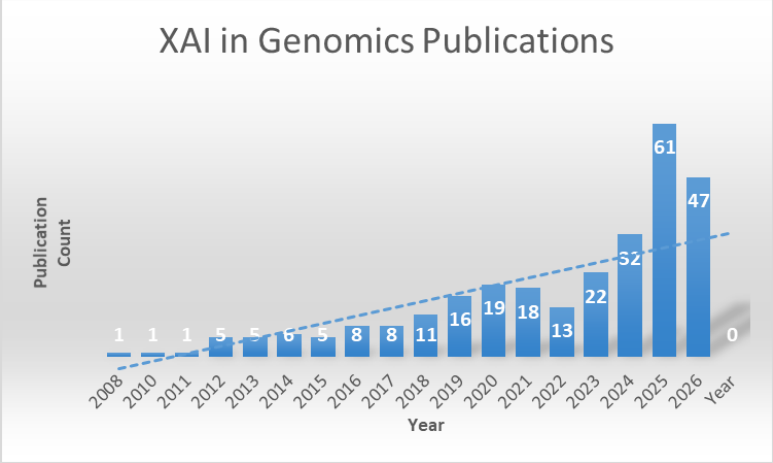


Figure 1. The increase in XAI in Genomics publications

4. LLM Generated Explanations Use Cases

This paper explores how textual explanations for case-based machine learning models can be generated through GenAI prompting with a use case in XAI in Genomics for Oncology.

Textual and/or conversational explanations have been proposed and successfully evaluated for machine learning models (Feldhus et al., 2023; slack et al., 2023), and this type of explanation is being more and more proposed to users since the large-scale availability of GenAI tools. However, researchers highlight the limitations of these tools regarding the possibility of generating inaccurate information and hallucinations (Palod et al., 2026). Solutions have been proposed to counter this limitation, for example by Palod et al. (2026), who enhance the effectiveness of GenAI explanations by presenting arguments for and against the AI's answer.

In this section, we present a case study of GenAI explanations for a case-based reasoning survival analysis model in oncology as a proof of concept. Given the range of questions users may want to ask from machine learning models for explanation purposes, interfaces to LLM APIs have been made available for bioinformatics analytical pipelines. LLMs can be resorted to for a number of questions, by transmitting prompts to GPTs and displaying results to users (Ex: *mergen* R package).

We selected a number of use cases to explain case-level recommendations by a case-based survival analysis model, *CBRinR* (Bichindaritz et al, 2025) with the goal of evaluating the feasibility of using prompts to generate sophisticated genomics explanations to users through conversations and visualizations.

More specifically, the dataset used contains breast cancer gene expressions from the Illumina HiSeq RNA sequencing data collected from TCGA. This dataset contains 17,002 features with values ranging between -1 and 1. After preprocessing, differential expression analysis, and reducing samples to just the primary cancer tissue, 1151 features and 761 samples were retained. The LLM chatbot tested is an enterprise-level university-managed Google Gemini Education chatbot.

CBRinR is a CBR framework for survival analysis, CBRinR, is described in (Bichindaritz et al., 2025). Briefly, it proceeds first through learning prognostic risk groups based on differentially expressed genes determined by the Cox proportional hazards model, which can be visualized through Kaplan-Meier survival curves. This helps the interpretability as we can visualize where a query case fits among the risk groups. Then a query case is matched to a varying number of retrieved cases from the memory, before being adapted based on the differences between the query and retrieved cases. Our framework also takes into account confounding factors in its retrieval step, since the underlying Cox model can handle them. It is interesting to note that almost all ML survival models rely on the Cox model, whether random survival forests, or deep learning models such as LSTM or transformers.

In this case study, we use a single retrieved case ('TCGA-C8-A273') and the corresponding query case ('TCGA-D8-A1JC') were entered as attached files into the system. No other information than the presented prompts was entered into the system, which appears preloaded with biological and genetic knowledge bases. The following use cases were selected and tested on the query case and the corresponding retrieved case for the task of survival analysis:

P1. Please tell me the features in common between the retrieved cases and the query case.

P2. Please tell me the main differences between the retrieved cases and the query case.

P3. Can you visualize these commonalities and differences between the retrieved cases and the query case?

P4. Can you map these common features at the biological pathway level or the functional level?

P5. Same question for the differences?

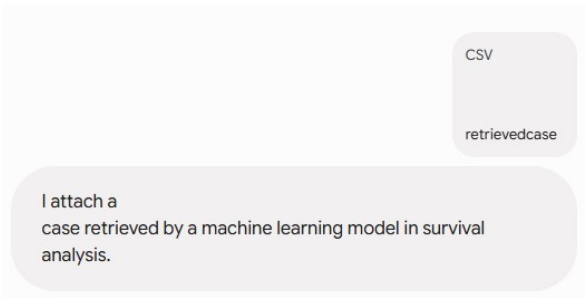
P6. Can you provide a visual representation at the pathway level?

P7. Can you summarize for an oncologist what you have found about the relationship between the retrieved cases and the query case?

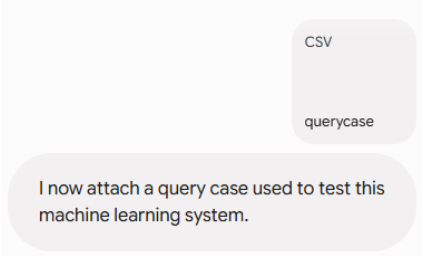
P8. Same question for a patient.

P9. Can you generate a counterfactual for this query case?

4.1 Initialization



The intent was to ask the chatbot to determine the main commonalities and differences between the query case and the retrieved case (prompts P1, P2, P3). Interestingly, the chatbot performed this analysis automatically before prompting, and also automatically detected that the machine learning model is based on case-based reasoning.



4.2 Cases comparison

Output:

Thank you for providing the query case. Analyzing `querycase.csv` alongside the previously provided `retrievedcase.csv` provides clear insights into why your machine learning survival analysis model paired these two specific profiles together.

In case-based or retrieval-based survival models (such as deep k-NN survival models, attention-based matching networks, or prototype frameworks), the system evaluates a patient test case (the "Query") and searches the training registry for the most genomically and clinically similar patient (the "Retrieved Case") to guide its survival prediction or generate explanations.

1. Patient Profile Alignment

The clinical and metadata alignment between the query case and the retrieved case is highly concordant:

Case Type	Patient ID	Survival Time (Years)	Survival Status
Query Case	TCGA-D8-A1JC	~ 1.32$	0 (Right-Censored / Alive)
Retrieved Case	TCGA-C8-A273	~ 1.41$	0 (Right-Censored / Alive)

The machine learning system has successfully matched patients who share nearly identical mid-term, event-free clinical outcomes (≈ 1.32 vs. ≈ 1.41 years).

It is interesting to note that this summary table of comparison between the query case and the test case has selected the survival time and survival status automatically as the most important features for this survival analysis model, without even explaining the model type.

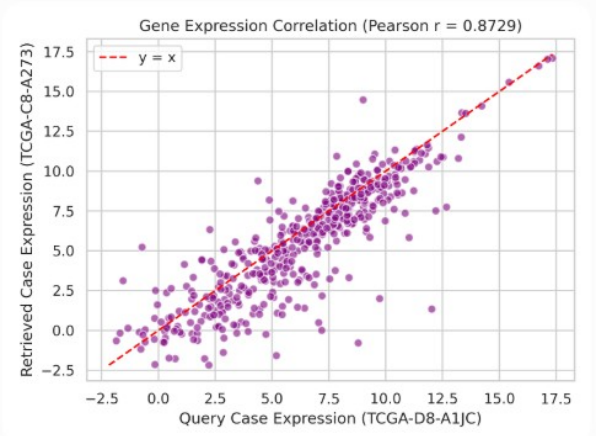
2. High Genomic Profile Concordance

Evaluating the expression levels across all 500 genes reveals a remarkably robust relationship:

- **Pearson Correlation Coefficient (r):** ≈ 0.8729
- **Euclidean Distance:** 41.66

A correlation coefficient of 0.8729 indicates that the model has identified a high-fidelity genomic "twin." The overall distribution tightly tracks a $y = x$ reference line, confirming that the network is leveraging global transcriptomic patterns rather than relying on an isolated set of sparse markers.

A visualization of this cross-case correlation has been generated and saved as `query_vs_retrieved_scatter.png` at the top of your panel.



The proposed visualization is a Pearson correlation between the genes of the query case and the retrieved case. The user could request a different type of visualization as needed.

3. Shared Biological Fingerprints (The Extracellular Matrix)

Looking at the highest-expressed genes in the query case confirms that both profiles share the exact same biological signature:

- **Query Top Expressed Genes: COL1A1** (17.33), **FN1** (17.12), **COL3A1** (16.73), **GAPDH** (15.43), and **BGN** (14.22).
- **Overlapping Signature:** Out of the top 15 highest-expressed genes in both patients, **8 major genes are identical**: `COL1A1`, `FN1`, `COL3A1`, `GAPDH`, `BGN`, `COL5A1`, `COL5A2`, and `DCN`.

This heavy overlap emphasizes that the machine learning system sees **stromal matrix activation** or a **mesenchymal-like tumor microenvironment** (characterized by dense collagen deposition, fibronectin expression, and matrix-remodeling proteoglycans) as the defining clinical-genomic axis for this patient cohort. In breast invasive carcinoma, this signature often points to highly reactive tissue architecture or Cancer-Associated Fibroblast (CAF) infiltration, which directly maps to specific survival risks or therapeutic sensitivities.

4. Key Expressions Divergences

While their global expression vectors are highly correlated, the model must navigate minor deviations. The genes with the largest absolute expression differences between the query and retrieved case include:

1. **CRABP1** (Cellular Retinoic Acid Binding Protein 1): Highly expressed in the query case (12.03) but low in the retrieved case (1.34).
2. **ANO3** (Anoctamin 3): Expressed in the query case (8.79) but completely suppressed in the retrieved case (−0.78).
3. **AKR1C2** (Aldo-Keto Reductase Family 1 Member C2): Highly active in the query case (9.73) compared to the retrieved case (2.00).

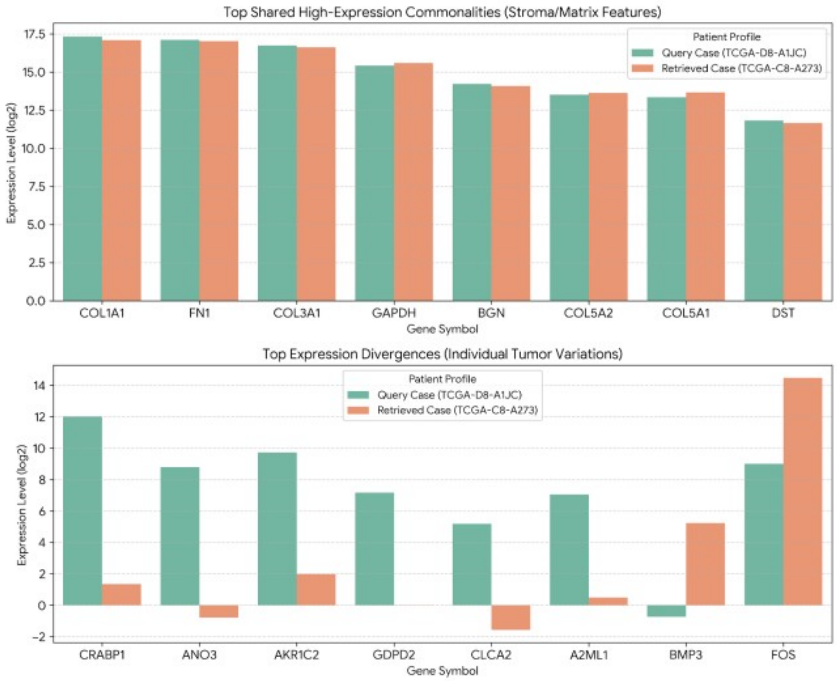
Conclusion

Your survival analysis machine learning model is operating with high biological and clinical fidelity. It has retrieved a training case that matches the query case not only in its overall genomic landscape ($r = 0.8729$) and its specific hyper-activated stromal phenotype, but also in its clinical trajectory—producing a highly stable, context-aware match for survival evaluation.

The proposed explanation resorts to genetic knowledge for this detailed interpretation, and likely genetic pathway-level interpretation of the genes in common. We can further ask the system to provide a visual representation of the comparison between cases.

Can you
visualize these commonalities and differences between the retrieved cases and
the query case?

Output:



Overall, the chatbot was able to detect, quantify, visualize, and interpret reasonably well the similarity and differences between cases.

4.3 Genetic pathway explanations

Can you map
these common features at the biological pathway level or the functional level?

Output:

To map the shared high-expression signature between the query case and the retrieved case at the functional and biological pathway level, we can look at the overrepresented annotation networks in standard databases like **Reactome**, **KEGG**, and **Gene Ontology (GO)**.

Gemini generated several pages of interpretations at the pathway level and concludes by providing a precise explanation of results at the pathway level.

3. Clinical & Survival Implications for the Machine Learning Model

Why did the machine learning survival model pick this specific functional module as the primary axis to pair these two patients?

1. **The Desmoplastic/CAF Signature:** In breast invasive carcinoma, this exact network of genes is the hallmark of **desmoplasia**—the development of a dense, fibrous stroma heavily populated by active **Cancer-Associated Fibroblasts (CAFs)**. The model has recognized that both patients share an identical stroma-rich tumor microenvironment phenotype.
2. **Mechanotransduction & Tumor Stiffness:** A high concentration of cross-linked collagens and fibronectin increases interstitial tissue stiffness. This physical rigidity triggers mechanotransduction loops inside tumor cells, promoting epithelial-to-mesenchymal transition (EMT), increasing stemness, and facilitating local invasion.
3. **Physical Barrier to Therapy:** High expression of this matrisome network creates a dense physical matrix and elevates interstitial fluid pressure within the tumor. This acts as a barrier that hinders the efficient diffusion of chemotherapeutic agents and restricts the infiltration of cytotoxic T-lymphocytes.

By mapping the query case to a retrieved training case with this precise functional signature, the machine learning system is asserting that the **biophysical microenvironment** of the tumor is a dominant driver governing the patient's survival probability and clinical trajectory.

Same
question for the differences?

Output excerpt

Gemini generated several pages of interpretations of differences at the pathway level and concludes by providing a precise explanation of results at the pathway level.

Summary: Why the Machine Learning Model Accepted These Differences

In machine learning algorithms designed for survival retrieval, matching cases is a trade-off between **global genomic structure** and **local genetic anomalies**.

The model chose to pair these two patients because the macro-environmental features—such as the massive structural matrisome shared between them—exert a dominant influence over long-term clinical survival trajectories. The system effectively treated the divergences highlighted above as localized metabolic variations (e.g., how they handle vitamin A or lipid storage) that do not overwrite the fundamental biophysical reality that both tumors are encased in a dense, protective, and highly reflective stromal matrix.

Gemini also generated a concise visual representation of both commonalities and differences at the pathway level (not shown in this article).

4.4 Summaries for specific users

For the other tasks, Gemini provided detailed explanations to oncologist users, and with a different language, to patient users. For the clinicians, it proposed some treatment ideas to explore. The system was also able to generate some basic counterfactuals.

Output excerpt for an oncologist

Clinical Recommendations & Therapeutic Insights

1. **Stroma-Targeting Rationale:** Because the model identifies the dense matrisome as the primary prognostic driver, therapeutic strategies aimed at disrupting matrix cross-linking or CAF activation (e.g., targeting the TGF-β pathway, FAK inhibitors, or hyaluronidase-based co-therapies) could be considered to lower interstitial pressure and improve systemic drug delivery.
2. **Mitigating Query-Specific Resistance:** The high $CRABP1$ and $AKR1C$ footprint in your query patient suggests potential resistance to conventional oxidative stress-inducing therapies. This profile makes downstream MEK/ERK inhibition a biological vulnerability worth monitoring if the disease progresses.

In summary, the machine learning system suggests that **the structural "soil" (the desmoplastic microenvironment) is exerting a more uniform influence on survival outcomes than the specific metabolic "seed" (individual signaling mutations)**, making the stroma a critical focus for this patient's clinical management.

Output excerpt for a patient

Understanding Your Tumor's Match: A Guide to What the Computer Found

When doctors look at a tumor, they don't just look at the cancer cells under a microscope; they can also look at a massive "genetic fingerprint" consisting of hundreds of genes. A advanced computer program looked at your tumor's specific genetic fingerprint (the **Query Case**) and searched a large database to find a patient from the past whose tumor had a matching pattern (the **Retrieved Case**).

The program found a very strong match. In fact, your tumor and the matched patient's tumor share a **87% global genetic match**. Because their patterns are so similar, looking at how the past patient responded to treatment gives your medical team a highly accurate roadmap for your own care.

5. Discussion and Conclusion

In this paper we discussed the state-of-the-art in XAI in genomics and experimentally tested available R packages for LLM explanations for case-based survival analysis in the oncology domain. The tested use cases provided detailed responses to a number of a user's questions about the CBR system recommendations, which demonstrated that the chatbot had at its disposal deep knowledge about genetics, allowing it to map genes to genetic pathways, in particular. Although these results are preliminary, they at least provide a proof-of-concept that LLMs can quickly generate genetic explanations to CBR recommendations in this domain, which can be used to create explanation prototypes for user feedback when designing CBR systems in this domain. This type of LLM-generated explanations can be used for rapid prototyping of CBR explanations to collect user feedback on preferences. It would be possible to mine users' interaction with the explanation system to learn their preferences and questions. However, before considering using LLM-based explanations for CBR recommendations in the clinic, more thorough validations of the accuracy of the explanations needs to be undertaken (Weber et al., 2024), as well as steps to ensure the safety of these explanations since the research on safety-critical systems has demonstrated that accuracy is not sufficient to ensure safety (Leveson and Turner, 1993). Hence controlling the existence of inaccurate or hallucinated results as proposed by (Palod et al., 2026) would be a valuable step for clinical settings, however would also require safety-critical design considerations. We also plan to develop many other types of questions that users may find interesting to discuss with the CBR system, as prompt patterns for XAI in Genomics (White et al., 2023).

Acknowledgements

We want to thank the National Cancer Institute of the National Institutes of Health for their support of this research under award Number R15CA297754, and the reviewers for their detailed feedback.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools for writing this paper, except for generating machine learning models explanations as this is the topic of this research.

References

[1] Budhkar, A., Song, Q., Su, J., & Zhang, X. (2025). Demystifying the black box: A survey on explainable artificial intelligence (XAI) in bioinformatics. Computational and Structural Biotechnology Journal, 27, 346-359.

[2] Bichindaritz, I., Berrios, L., & Desai, M. (2025, June). CBRinR Multitask Multiomics Case-Based Reasoning in Bioinformatics. In International Conference on Case-Based Reasoning (pp. 363-377). Cham: Springer Nature Switzerland.

[3] Biecek, P. (2018). DALEX: Explainers for complex predictive models in R. Journal of Machine Learning Research, 19(84), 1-5.

[4] Byrne, R. M. (2019, August). Counterfactuals in explainable artificial intelligence (XAI): Evidence from human reasoning. In IJCAI (pp. 6276-6282).

[5] Chen RJ, Lu MY, Williamson DFK, Chen TY, Lipkova J, Noor Z, Shaban M, Shady M, Williams M, Joo B, Mahmood F. (2022) Pan-cancer integrative histology-genomic analysis via multimodal deep learning. Cancer Cell. 2022 Aug 8;40(8):865-878.e6.

[6] Feldhus, N., Wang, Q., Anikina, T., Chopra, S., Oguz, C., & Möller, S. (2023, December). InterroLang: Exploring NLP models and datasets through dialogue-based explanations. In Findings of the Association for Computational Linguistics: EMNLP 2023 (pp. 5399-5421).

[7] Gates, L., Leake, D., & Wilkerson, K. (2023, July). Cases are king: a user study of case presentation to explain CBR decisions. In International Conference on Case-Based Reason-ing (pp. 153-168). Cham: Springer Nature Switzerland.

[8] Hothorn, T., Frick, H., & Kook, L. (2025). CRAN task view: Machine learning & statistical learning.

[9] Hsu, C. Y., Askar, S., Alshkarchy, S. S., Nayak, P. P., Attabi, K. A., Khan, M. A., ... & Soleimani Samarkhazan, H. (2026). AI-driven multi-omics integration in precision oncolo-gy: bridging the data deluge to clinical decisions. Clinical and experimental medicine, 26(1), 29.

[10] Kenny, E. M., & Keane, M. T. (2021). Explaining Deep Learning using examples: Optimal feature weighting methods for twin systems using post-hoc, explanation-by-example in XAI. Knowledge-Based Systems, 233, 107530.

[11] Leveson, N. G., & Turner, C. S. (1993). An investigation of the Therac-25 accidents. *Computer*, *26*(7), 18-41.

[12] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predic-tions. Advances in neural information processing systems, 30.

[13] Palod, V., Biswas, U., & Kambhampati, S. (2026). Evaluating the False Trust Engendered by LLM Explanations. arXiv preprint arXiv:2605.10930.

[14] Papernot, N., & McDaniel, P. (2018). Deep k-nearest neighbors: Towards confident, inter-pretable and robust deep learning. arXiv preprint arXiv:1803.04765.

[15] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Ex-plaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD inter-national conference on knowledge discovery and data mining (pp. 1135-1144).

[16] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes deci-sions and use interpretable models instead. Nature machine intelligence, 1(5), 206-215.

[17] Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., & Zhong, C. (2022). Interpretable machine learning: Fundamental principles and 10 grand challenges. Statistic Surveys, 16, 1-85.

[18] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In Pro-ceedings of the IEEE international conference on computer vision (pp. 618-626).

[19] Slack, D., Krishna, S., Lakkaraju, H., & Singh, S. (2023). Explaining machine learning models with interactive natural language conversations using TalkToModel. Nature Machine Intelligence, 5(8), 873-883.

[20] Toussaint PA, Leiser F, Thiebes S, Schlesner M, Brors B. (2023) Sunyaev A. Explainable artificial intelligence for omics data: a systematic mapping study. Brief Bioinform. 2023 Nov 22;25(1):bbad453. doi: 10.1093/bib/bbad453. PMID:38113073; PMCID: PMC10729786.

[21] Weber, R. O., Johs, A. J., Goel, P., & Silva, J. M. (2024). XAI is in trouble. AI Maga-zine, 45(3), 300-316.

[22] White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., ... & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382*.