Expanding P-NET, a multi-purpose biologically informed deep learning framework

Marc Glettig [1 2] **Chenzhang Zhou** [1 2 3] **Giuseppe Tarantino** [1 2 3] **Tyler Aprati** [1 2] **Eliezer M. Van Allen** [1 2 3] **David Liu** [1 2 3] **Haitham Elmarakeby** [1 2 3]

Abstract

We present expanded P-NET, a versatile framework for deep learning in computational biology based on the original P-NET, leveraging biological pathways for interpretable predictions. Our framework achieves competitive performance in genomic & transcriptomic prediction tasks. We demonstrate its stability and interpretability compared to traditional machine learning models. P-NET incorporates gene and pathway information, providing valuable insights into complex biological processes. The framework is publicly available, enabling its application in various computational biology tasks.

1. Introduction

In recent years, deep learning has emerged as a powerful tool for analyzing complex biological data and has revolutionized various domains of computational biology. The integration of deep learning techniques with biological knowledge has led to the development of specialized neural network architectures tailored to capture the intricate relationships within biological systems. One such architecture, the Pathway Neural Network (P-NET) (Elmarakeby et al., 2021), draws inspiration from the organizational principles of biological pathways and has shown promising results in biological applications (Hartman et al., 2023).

In this work, we present an extended and versatile implementation of P-NET as a multi-purpose framework for deep learning applications in computational biology. The framework leverages the inherent structure and organization of biological pathways to limit the modeling capabilities of neural networks to meaningful connections. By incorporating domain-specific knowledge into the architecture, our framework is guided by the interpretability-by-design principle. Through this re-implementation, researchers can seamlessly apply deep learning techniques to a wide range of computational biology tasks, eliminating the need for task-specific network architectures. The flexibility and adaptability of our framework not only save valuable time and computational resources but also facilitate knowledge transfer across different biological applications.

To validate the effectiveness of our framework we first apply it to the original task of predicting metastatic vs. non-metastatic prostate cancer. Then we demonstrate the framework versatility and effectiveness by applying it to the clinically relevant task of predicting whole genome doubling (WGD) in cancer samples. Through extensive experiments on public datasets, we showcase the superior performance of our framework compared to regular deep learning architectures. Moreover, we provide an in-depth analysis of the learned representations and highlight the interpretability of our model, enabling biologists to gain valuable insights into complex biological processes.

2. Methods

2.1. P-NET architecture

Our implementation of P-NET adopts a sparse feed-forward neural network architecture, Figure 1. The sparsity is achieved through masked linear layers, where the masks are generated based on the adjacency matrix of the Reactome hierarchical pathway database (Gillespie et al., 2022). The Reactome database provides a comprehensive collection of hierarchical pathway relations and gene-to-pathway associations.

To handle different gene-level input modalities, we introduced a Gene Input Layer (GIL) that aggregates the input information for each gene. The GIL then establishes sparse connections between genes and the pathways they are associated with. The pathways, organized hierarchically, encompass various fine-grained biological sub-processes. By adding the pathway representations to the gene representations, we capture the collective influence of genes within

[1]Dana-Farber Cancer Institute, Boston, MA, USA [2]Broad Institute of MIT and Harvard, Cambridge, MA, USA [3]Harvard Medical School, Boston, MA, USA. Correspondence to: Marc Glettig <marc_glettig@dfci.harvard.edu>, Haitham Elmarakeby <haithama_elmarakeby@dfci.harvard.edu>.

The 2023 ICML Workshop on Computational Biology. Baltimore, Maryland, USA, 2023.

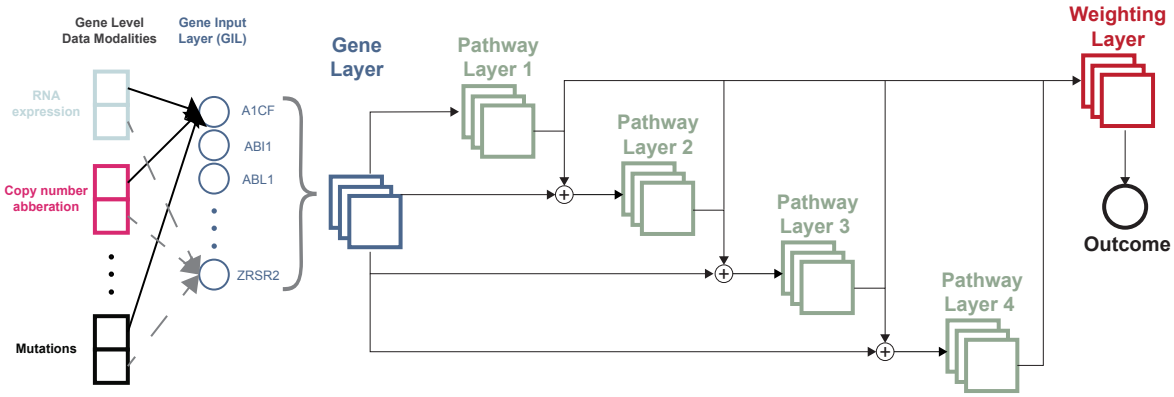

Figure 1. P-NET Model architecture, in the left part we demonstrate how different gene level input modalities are sparsely connected to the Gene Input Layer (GIL). A single representation of each gene in the GIL is then connected to it's respective biological pathways in the neural network layers. After each layer an outcome prediction is made and outcome predictions are then combined by a weighted average to form the final model prediction.

higher-level pathways. It is important to note that each gene connects only to the lowest pathway level in which it is present.

Within each pathway layer, we incorporate a prediction head to generate predictions specific to that pathway layer. These pathway-level predictions are then combined using a weighted average approach, with the weights learned in a dedicated weighting layer. Finally, the overall outcome prediction is obtained as the weighted average of the predictions from different pathway layers.

2.2. Model performance

To assess the performance of our model, we compared it against several other models, including a normal feed-forward neural network, a sparse neural network, and traditional machine learning models such as random forests (RF) and support vector machines (SVM). The sparse neural network follows the same model architecture as P-NET, utilizing the same number of connections. However, the sparse neural network randomly selects the connections. This approach allows us to investigate the impact of structured connections derived from biological pathways versus randomly chosen connections in the model's performance. These models were implemented and trained using established frameworks and libraries (Pedregosa et al., 2011).

For measuring the performance of the model, we utilized the area under the receiver operating characteristic curve (AUC-ROC). Additionally, to gain insights into feature and neuron importance scores, we employed the Integrated Gradients method (Sundararajan et al., 2017) and its extension (Shrikumar et al., 2018). These methods enabled us to attribute scores to individual features and neurons, highlighting their significance in the model's decision-making process.

Furthermore, to assess the stability of feature importance scores, we introduced the concept of rank variance. Specifically, we measured the variance in the importance rank of the 50 most important genes within the GIL. A lower rank variance means the model interpretations are more stable and can be considered better. This allowed us to evaluate the consistency and robustness of feature importance rankings across different runs or iterations.

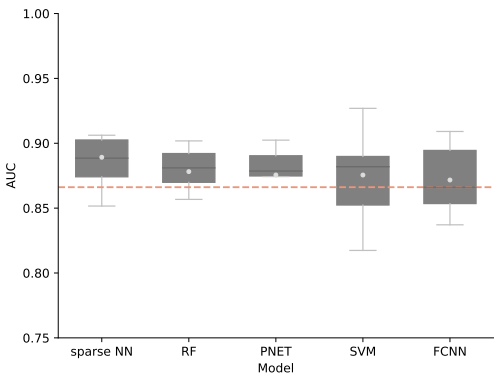

Figure 2. AUC (10-fold CV) for predicting metastatic vs. non-metastatic prostate cancer in the original cohort (Armenia et al., 2018). The white dot represents the mean AUC. Models are displayed in descending performance order. Sparse deep learning models (PNET and sparse NN) achieve similar performance to the traditional ML models (RF and SVM). Notably, the fully connected neural network exhibits slightly lower performance compared to the other models.

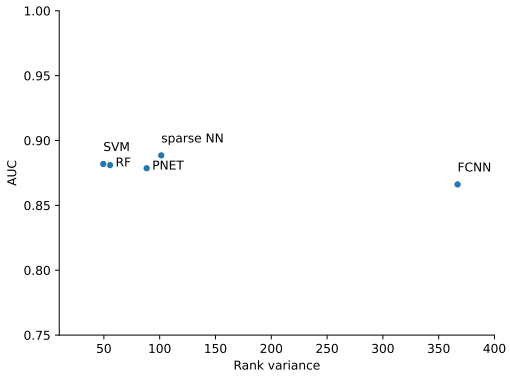

Figure 3. Comparison of variance in gene importance rank for the 50 most important genes. The rank variance serves as a stability measure for model interpretability, with lower scores indicating higher stability. RF & SVM, which capture simple interactions in the data, demonstrate robust interpretability. Among the deep learning models, our P-NET framework stands out as the most stable, with the lowest rank variance. The superior stability of P-NET highlights its ability to reliably attribute importance to specific genes & pathways.

3. Results

3.1. Reproduction of Original Model

In this experiment, we aimed to validate the performance of our new P-NET framework by reproducing the outcomes from the original study (Elmarakeby et al., 2021). We compared the results obtained with our framework against various machine learning models. The evaluation was conducted on the original task of binary classification between metastatic and non-metastatic prostate cancer samples.

Our results demonstrated that all models performed reasonably well in capturing the metastatic vs. non-metastatic prediction task, with mean AUCs across 10 folds exceeding 0.85. Specifically, our P-NET framework achieved similar AUC-ROC scores compared to the other models, as shown in Figure 2. P-NET achieved the most consistent results with the smallest variance among the tested models showing that P-NET predictions are more stable over the used 10 folds.

Furthermore, the interpretability analysis revealed that our P-NET implementation exhibited more stable feature scores compared to other deep models, such as the sparse neural network and the fully connected neural network, as shown in Figure 3. Demonstrating that the embedding biological information in the computational model may help enhance the predictions stability and feature ranking consistency. As expected, we observed even more stable features in classical machine learning models like RF and SVM that are known to be generally stable compared to deep learning methods.

3.2. Prediction of Ploidy in Skin Cutaneous Melanoma (SKCM)

In this experiment, we applied the P-NET framework to predict ploidy in skin cutaneous melanoma (SKCM) samples from TCGA (Hoadley et al., 2018). High ploidy and WGD are of clinical interest as predictors of Immune Checkpoint Blockade (ICB) therapy response. Our objective was to investigate the relationship between high ploidy and genomic and transcriptomic mechanisms.

By incorporating gene mutations and normalized RNAseq data as inputs to the P-NET framework, we achieved accurate prediction of WGD in SKCM samples. The performance of our model was evaluated using the AUC-ROC, which demonstrated competitive performance in distinguishing between samples with and without WGD (mean test AUC of 0.72). Additionally, we evaluated the generalization ability of our model on an external validation dataset (Liu et al., 2019), where it exhibited promising performance with a mean validation AUC of 0.69, demonstrating its efficacy in handling unseen data.

The interpretability analysis allowed us to gain insights into the contributions of individual features and neurons in the prediction process. By leveraging the design of P-NET, it becomes feasible to link the importance scores to corresponding biological pathways and genes. Figure 4 showcases a Sankey diagram highlighting the highest importance genes and pathways. Notably, genes known to be involved in DNA repair, such as RPA1 and BRCA1, were found to have significant involvement in the model's prediction. Moreover, the DNA double-strand repair pathway emerged as an important feature for high ploidy predictions. These findings demonstrate the P-NET framework's ability to establish connections between biological processes and network neurons, with model learning guided by established biological knowledge.

4. Discussion

We successfully reproduced the outcomes from the original study using our new P-NET framework. The performance evaluation showcased competitive results, with our framework achieving similar performance to state-of-the-art models in the task of identifying advanced prostate cancer patients based on their genomic profiles. Notably, our P-NET implementation exhibited superior stability of feature scores compared to other deep models. This indicates that the incorporation of biological pathways in P-NET enables more reliable and interpretable predictions.

Furthermore, we applied the P-NET framework to predict

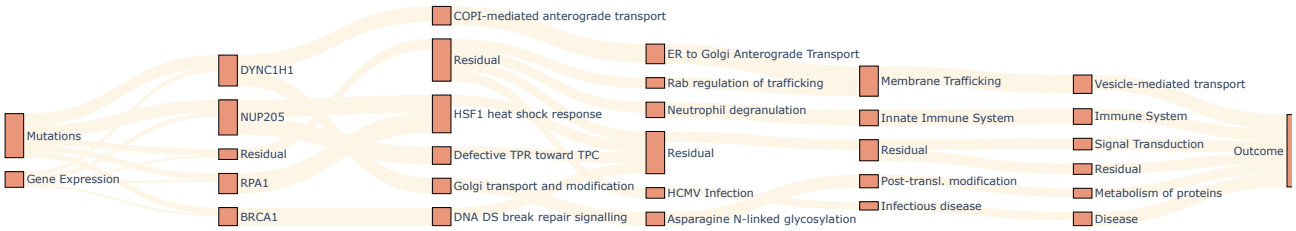

Figure 4. Sankey diagram illustrating the flow of feature importances through P-NET in the prediction of WGD in SKCM. The diagram shows the most important genes and pathways per model layer. Notably, genes such as RPA1 and BRCA, which are known to be involved in DNA repair, as well as the pathway for DNA double-strand (DS) break repair emerge as important drivers of the model's prediction.

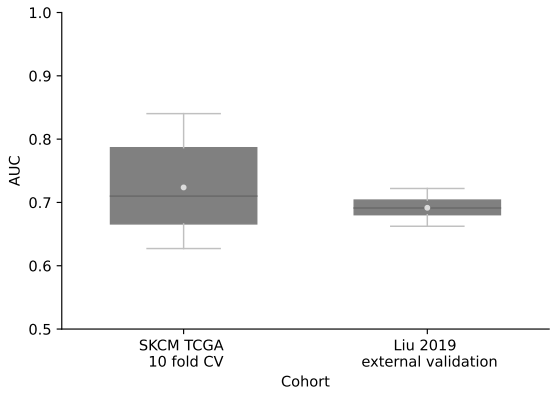

Figure 5. AUC (10-fold CV) for predicting WGD in SKCM cancer samples. The white dot represents the mean AUC. The training and testing were conducted on the TCGA cohort. P-NET demonstrates strong predictive performance in discerning WGD events in SKCM samples (mean test AUC$= 0.72$. Furthermore, we show the generalization capability of our model on an external validation cohort (Liu et al., 2019). P-NET showcases consistent performance across different datasets (mean validation AUC$= 0.69$.

WGD in skin cutaneous melanoma (SKCM) samples, aiming to explore the relationship between high ploidy and genomic and transcriptomic mechanisms. By integrating gene mutations and normalized RNAseq data as inputs, our model accurately predicted WGD in SKCM samples. The evaluation results, demonstrated competitive performance in distinguishing between samples with and without WGD. The interpretability analysis provided valuable insights into the contributions of individual features and neurons, linking them to specific biological pathways and genes. Notably, genes involved in DNA repair, such as RPA1 and BRCA1, emerged as important predictors, emphasizing the relevance of these biological processes in determining ploidy status.

Our findings highlight the effectiveness of the P-NET framework in deep learning applications within computational biology. By integrating biological pathways into the model architecture, we provide a powerful tool for gaining interpretable insights into complex biological processes. The stability of feature scores and the ability to link them to biological knowledge make P-NET a valuable asset in understanding the underlying mechanisms.

5. Conclusion

Our study showcases the versatility and effectiveness of the P-NET framework in deep learning applications within computational biology. By combining domain knowledge with advanced neural network techniques, our framework offers a unified solution to tackle various challenges in biological data analysis. Our framework provides a powerful tool for integrating multiple data modalities including genomic and transcriptomic profiles to predict biological and clinical outcomes. Moreover, our framework provides interpretable predictions and valuable insights into complex biological processes. This work contributes to the growing field of biologically-informed deep learning in computational biology, opening new avenues for leveraging biological knowledge to guide clinical prediction and biological discovery.

The P-NET comprehensive multi-purpose framework stands out for its user-friendly implementation, requiring straightforward data integration. To facilitate its adoption, we have made the framework readily accessible at the following GitHub repository:

`https://github.com/vanallenlab/pnet`

Looking ahead, we aim to expand the scope of P-NET by exploring its applicability to unsupervised tasks and transfer-learning approaches in various cancer-related use cases. Additionally, we see potential in leveraging the framework for the analysis of single-cell data, further extending its utility and impact in computational biology.

Acknowledgements

We would like to extend our sincere gratitude to the funding agencies, including the Department of Defense/LC220330, NIH K08CA234458, Doris Duke Clinical Scientist Development Award, PCF-Movember Challenge Award, NIH U01CA233100, and NIH R01CA227388. Their generous financial support has been instrumental in enabling this research. We deeply appreciate their belief in our work and their investment in scientific endeavors.

Furthermore, we would like to express our appreciation to Dr. Jackson Nyman, Pasha Trukhanov, Gwen Miller, and Andrew Zhou for their valuable contributions, insightful discussions, and constructive feedback throughout the development of this research project. Their expertise and collaboration have greatly enriched our work.

References

Armenia, J., Wankowicz, S. A., Liu, D., Gao, J., Kundra, R., Reznik, E., Chatila, W. K., Chakravarty, D., Han, G. C., Coleman, I., et al. The long tail of oncogenic drivers in prostate cancer. *Nature genetics*, 50(5):645–651, 2018.

Carter, S. L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P. W., Onofrio, R. C., Winckler, W., Weir, B. A., Beroukhim, R., Pellman, D., Levine, D. A., Lander, E. S., Meyerson, M., and Getz, G. Absolute quantification of somatic dna alterations in human cancer. *Nature Biotechnology*, 30(55):413–421, May 2012. ISSN 1546-1696. doi: 10.1038/nbt.2203.

Elmarakeby, H. A., Hwang, J., Arafeh, R., Crowdis, J., Gang, S., Liu, D., AlDubayan, S. H., Salari, K., Kregel, S., Richter, C., Arnoff, T. E., Park, J., Hahn, W. C., and Van Allen, E. M. Biologically informed deep neural network for prostate cancer discovery. *Nature*, 598(7880): 348–352, Oct 2021. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-021-03922-4.

Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., Griss, J., Sevilla, C., Matthews, L., Gong, C., Deng, C., Varusai, T., Ragueneau, E., Haider, Y., May, B., Shamovsky, V., Weiser, J., Brunson, T., Sanati, N., Beckman, L., Shao, X., Fabregat, A., Sidiropoulos, K., Murillo, J., Viteri, G., Cook, J., Shorser, S., Bader, G., Demir, E., Sander, C., Haw, R., Wu, G., Stein, L., Hermjakob, H., and D'Eustachio, P. The reactome pathway knowledgebase 2022. *Nucleic Acids Research*, 50(D1):D687–D692, Jan 2022. ISSN 0305-1048. doi: 10.1093/nar/gkab1028.

Hartman, E., Scott, A., Malmström, L., and Malmström, J. Interpreting biologically informed neural networks for enhanced biomarker discovery and pathway analysis. pp. 2023.02.16.528807, Feb 2023. doi: 10.1101/2023.02.16.528807. URL https://www.biorxiv.org/content/10.1101/2023.02.16.528807v1.

He, K., Zhang, X., Ren, S., and Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. (arXiv:1502.01852), Feb 2015. doi: 10.48550/arXiv.1502.01852. URL http://arxiv.org/abs/1502.01852. arXiv:1502.01852 [cs].

Hoadley, K. A., Yau, C., Hinoue, T., Wolf, D. M., Lazar, A. J., Drill, E., Shen, R., Taylor, A. M., Cherniack, A. D., Thorsson, V., Akbani, R., Bowlby, R., Wong, C. K., Wiznerowicz, M., Sanchez-Vega, F., Robertson, A. G., Schneider, B. G., Lawrence, M. S., Noushmehr, H., Malta, T. M., Network, C. G. A., Stuart, J. M., Benz, C. C., and Laird, P. W. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell*, 173(2):291–304.e6, Apr 2018. ISSN 1097-4172. doi: 10.1016/j.cell.2018.03.022.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. (arXiv:1412.6980), Jan 2017. doi: 10.48550/arXiv.1412.6980. URL http://arxiv.org/abs/1412.6980. arXiv:1412.6980 [cs].

Liu, D., Schilling, B., Liu, D., Sucker, A., Livingstone, E., Jerby-Arnon, L., Zimmer, L., Gutzmer, R., Satzger, I., Loquai, C., Grabbe, S., Vokes, N., Margolis, C. A., Conway, J., He, M. X., Elmarakeby, H., Dietlein, F., Miao, D., Tracy, A., Gogas, H., Goldinger, S. M., Utikal, J., Blank, C. U., Rauschenberg, R., von Bubnoff, D., Krackhardt, A., Weide, B., Haferkamp, S., Kiecker, F., Izar, B., Garraway, L., Regev, A., Flaherty, K., Paschen, A., Van Allen, E. M., and Schadendorf, D. Integrative molecular and clinical modeling of clinical outcomes to pd1 blockade in patients with metastatic melanoma. *Nature Medicine*, 25(1212):1916–1927, Dec 2019. ISSN 1546-170X. doi: 10.1038/s41591-019-0654-5.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Shrikumar, A., Su, J., and Kundaje, A. Computationally efficient measures of internal neuron importance.

(arXiv:1807.09946), Jul 2018. doi: 10.48550/arXiv.1807.09946. URL http://arxiv.org/abs/1807.09946. arXiv:1807.09946 [cs, stat].

Sundararajan, M., Taly, A., and Yan, Q. Axiomatic attribution for deep networks. (arXiv:1703.01365), Jun 2017. doi: 10.48550/arXiv.1703.01365. URL http://arxiv.org/abs/1703.01365. arXiv:1703.01365 [cs].

A. Methods

A.1. Dataset preparation

A.1.1. Metastatic Prostate Cancer data

For the validation of the original task, we followed the dataset preparation process described in (Elmarakeby et al., 2021). We utilized pre-processed data, which included mutation counts and copy number aberrations (CNA). The mutation data was transformed into a binary input format, while the CNA data was split into separate binary datasets representing amplification (AMP) and deletion (DEL). To ensure consistent evaluation, we employed the same data splits for 10-fold cross-validation, dividing the samples into training and test datasets. The target variable for this task involved binary classification between metastatic and non-metastatic prostate cancer samples.

A.1.2. Whole genome doubling in Melanoma

In our example use case of predicting whole genome doubling (WGD) in skin cutaneous melanoma (SKCM), we utilized the publicly available TCGA PanCancer cohort (Hoadley et al., 2018). Specifically, we focused on the subset of SKCM samples within this cohort. The determination of tumor ploidy for these samples was performed using the ABSOLUTE pipeline (Carter et al., 2012). A Gaussian mixture model (GMM) was employed to establish a threshold for calling WGD based on ploidy values. Samples with ploidy ≥ 2.23 were considered to exhibit whole genome doubling. In the additional external validation dataset, we used the same methods to determine ploidy as well as WGD. To link high ploidy with genomic and transcriptomic mechanisms, we utilized gene mutations and normalized RNAseq data to predict WGD using the P-NET framework. As additional external validation dataset, we used the (Liu et al., 2019) cohort.

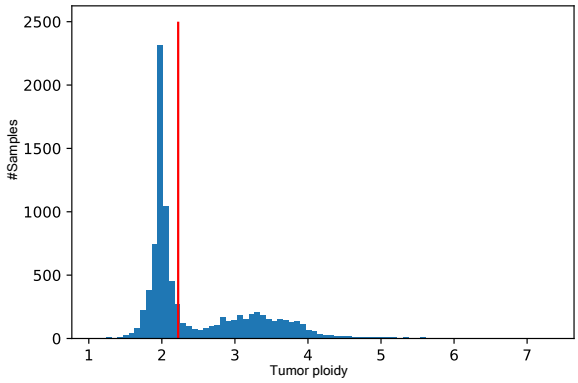

Figure 6. Histogram of ploidy values shown. The bimodal distribution shows many diploid samples with a clear peak around 2. The genome doubled samples have a much wider distribution. Our GMM model determined the cutoff between the two distributions to be at 2.23.

A.2. Model training

The model was implemented using the PyTorch framework (Paszke et al., 2019). We initialized the weights using the He initialization method (He et al., 2015) and employed the Adam optimizer (Kingma & Ba, 2017) for training. A learning rate of 10^{-4} was utilized, and the training process spanned 300 epochs with the option of early stopping. Since all tasks involved binary prediction, Binary Cross Entropy was chosen as the loss function. To promote model generalization, we introduced L2 regularization.