HCS-DFC: A Diffusion Classifier for Mode of Action Prediction Using Morphological Profiles

Anonymous CVPR submission

Paper ID 3

Abstract

Phenotypic-driven drug discovery is gaining popularity due to the advances in high-content imaging and machine learning, particularly for predicting compound Mode of Action (MoA) and properties. However, reliance on biochemical assays for label acquisition introduces noise and sparsity, complicating reliability estimation in traditional discriminative models.

In this work, we propose a High Content Screening Diffusion Classifier (HCS-DFC), reformulating prediction as a conditional generation task to inherently model label distributions and co-dependencies without requiring calibration datasets. By leveraging diffusion models' ability to capture complex data distributions, HCS-DFC outperforms conformal prediction methods in reliability estimation and achieves state-of-the-art accuracy on synthetic (MNIST-based multi-task classification) and real-world cell painting datasets.

1. Introduction

High content imaging (HCS) is a powerful technique that allows to screen the huge amounts of compounds and genetic perturbations by analyzing morphological changes in cells [21, 25]. An example of such an approach is cell painting protocol [7, 12, 28], which generates numerous cell images representing cellular morphology after treatment. By studying changes in cell phenotype, researchers can draw conclusions about particular treatments, such as their toxicity or ability to induce specific cellular state changes (e.g., tubulin inhibition) [1, 10, 15, 19]. These insights are crucial for successful drug development programs.

With the advancement of imaging techniques and analytical approaches, phenotypic-driven drug discovery has gained popularity, leading to the development of multiple methods for analyzing HCS data [5, 6, 9, 17, 23, 27, 28]. One task being modeled based on these data is the prediction of mode of action (MoA) and properties for specific sets of compounds [6]. However, the image labels (e.g., specific MoA) are often derived from experiments that do not consider cell morphology. To showcase that phenomenon, let us assume that we want to model the response of U2OS cells to compound X. When we design this experiment, we do not know if this X will be reflected in U2OS morphology. Moreover, we know that X has a given MoA α but may also have MoA β, which has not yet been discovered. As a result, we do not know if morphological change is the result only of α or β or both of them. This results in noisy and incomplete labels, posing challenges for machine learning models [6].

To address this issue, methods estimating prediction reliability, such as conformal prediction [13, 29], are widely used. However, these methods have significant limitations, as they require a calibration set for reliability estimation while still operating on noisy and sparse labels. Properly modeling the conformity score would require different calibration sets for each task (single MoA) or expensive and time-consuming biological experiments to confirm and input missing labels in the calibration set.

We propose a novel approach to this problem by exploiting a diffusion-based classifier [11, 30] instead of an MLP-based one. This approach transforms the problem formulation from classical classification to generating a multi-task label vector from noise conditioned by the cellular image representation. The key difference is that diffusion models can model the distribution of data, in this case, the distribution of labels and their co-dependencies.

To demonstrate the effectiveness of this approach, we first define a toy task using the MNIST dataset [18], where we concatenate two digits and perform multi-task classification. This results in better reliability estimation than conformal prediction in terms of accuracy versus data coverage, as shown in Figure 2. We then adapt this method to the Bray et al. dataset [8], which is widely used to benchmark predictions of MoA and properties for small molecules [6], where we introduce the HCS Diffusion Classifier (HCS-DFC). Experimental results show that the diffusion model better handles noisy label characteristics and not only provides better

metrics than conformal prediction but also achieves better accuracy than MLP using the same image representation type obtained via CellProfiler [22]. Our contributions can be summarized as follows:

- We redefine the MoA and property prediction problem for HCS data as a conditional generation.
- We introduce the HCS-DFC model for effective and accurate modeling of morphological responses.
- We conduct validation on a toy dataset and the challenging Bray et al. dataset to showcase the effectiveness of the HCS-DFC model.

2. Preliminaries

2.1. Diffusion Process

Diffusion models have gained significant recognition due to their ability to generate highly detailed content like images [26] or videos [4]. These models operate through an iterative noising and denoising procedure. The core mechanism involves two key processes: a fixed forward process that systematically adds noise to data and a learned backward process that attempts to recover the original structure, optionally conditioning on variable c [16].

2.2. Diffusion Classifier

Diffusion classifier [20] is a novel approach for using pretrained diffusion models to perform classification tasks without requiring additional training. The method can transform any generative diffusion model into a discriminative classifier by exploiting its conditional density estimation capabilities. Given an input image $\mathbf{x}$ and a conditioning $\mathbf{c}$, the diffusion model can be used to select the class that best fits the image.

The diffusion classifier works by using Bayes theorem on the model predictions and the prior $p(\mathbf{c})$ over labels $\mathbf{c}_i$:

$$p_\theta(\mathbf{c}_i \mid \mathbf{x}) = \frac{p(\mathbf{c}_i)\, p_\theta(\mathbf{x} \mid \mathbf{c}_i)}{\sum_j p(\mathbf{c}_j)\, p_\theta(\mathbf{x} \mid \mathbf{c}_j)} \tag{1}$$

Assuming uniform prior over $\mathbf{c}_i = \frac{1}{n}$ (where n corresponds to the number of labels), the classification task reduces to finding the conditioning c_i that maximizes $p_\theta(\mathbf{x}|\mathbf{c}_i)$. For diffusion models, computing $p_\theta(\mathbf{x} \mid \mathbf{c})$ directly is intractable, therefore diffusion classifier utilizes the Evidence Lower BOund (ELBO) approximation:

$$\text{ELBO} \approx -\mathbb{E}_{t,\epsilon}\|\epsilon - \epsilon_\theta(\mathbf{x}_t, \mathbf{c}_i)\|^2 \tag{2}$$

to obtain posterior distribution over $\{\mathbf{c}_i\}_{i=1}^n$:

$$p_\theta(\mathbf{c}_i \mid \mathbf{x}) \approx \frac{\exp\{-\mathbb{E}_{t,\epsilon}\|\epsilon - \epsilon_\theta(\mathbf{x}_t, \mathbf{c}_i)\|^2\}}{\sum_j \exp\{-\mathbb{E}_{t,\epsilon}\|\epsilon - \epsilon_\theta(\mathbf{x}_t, \mathbf{c}_j)\|^2\}} \tag{3}$$

which can be estimated by sampling (t_i, ϵ_i) pairs, where $t_i \sim [1, 1000]$ and $\epsilon \sim \mathcal{N}(0, I)$ and than computing:

$$\frac{1}{N}\sum_{i=1}^{N} \left\| \epsilon_i - \epsilon_\theta(\sqrt{\bar{\alpha}_{t_i}}\mathbf{x} + \sqrt{1-\bar{\alpha}_{t_i}}\epsilon_i, \mathbf{c}_j) \right\|^2 \tag{4}$$

2.3. Conformal Prediction

Conformal prediction is a model agnostic method of generating prediction sets for any desired level of confidence [3]. Conformal prediction works by first sampling a small subset $(x_1, y_1), ..., (x_m, t_m)$ of the training set known as the calibration set. By using a small subset of the training set, an equal distribution of classes can be ensured. Using a calibration set and a fitted model $\hat{f}$, a conformal score s_i can be obtained. From this distribution we can compute $1-\alpha$ quantile $\hat{q}$, with α being the maximum permissible error rate. The ceiling function of such quantile is thus a threshold of minimum probability value for a prediction to be accepted as confident. When a prediction set $C(\mathbf{x}_{\text{test}})$ is formed using this threshold, it satisfies the following guarantee:

$$1 - \alpha \leq P(y_{\text{test}} \in C(\mathbf{x}_{\text{test}})) \leq 1 - \alpha + \frac{1}{m+1} \tag{5}$$

where n is size of $\mathbf{x}_{\text{test}}$. A prediction set $C(\mathbf{x}_{\text{test}})$ can then be formed by including all classes whose softmax outputs exceed $1-\hat{q}$. The true classes will be included in the prediction set with probability at least $1-\alpha$, regardless of model accuracy or data distribution.

3. Methods

In this section, we first describe HCS-DFC, our approach to reliability estimation with diffusion-based methods. Then, we discuss the implementation of conformal prediction in the multi-task prediction setting as a comparable baseline.

3.1. High Content Screening Diffusion Classifier

We propose High Content Screening Diffusion Classifier (HCS-DFC) as an adaptation of the Diffusion Classifier [20] for Mechanism of Action (MoA) prediction from High-Content Screening (HCS) images. We utilize image conditioning obtained using Cell Profiler features, which are used to condition the model at every timestep $t_i \sim [1, 1000]$. Through this process, we aim to generate accurate representations of MoAs observed in the input data.

Approach. We base our methodology on the Diffusion Classifier principles [20], introducing several architectural modifications to enhance computational efficiency and adapt the model to the MoA prediction task. Specifically, we replace the pretrained diffusion transformer architecture with a simpler autoencoder, which we train from

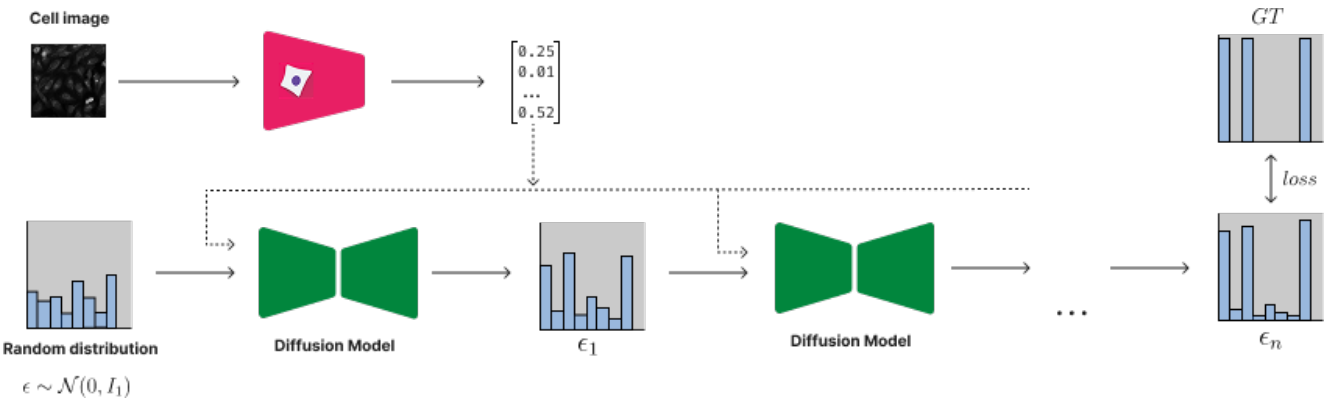

Figure 1. To predict MoA activities for a given cell image, we first use Cell Profiler to obtain a representation vector, which then serves as class conditioning for the diffusion model. We condition the diffusion model to generate MoA probabilities for a given image from n (number of classes) shaped noise vector. We train our model using masked BCE loss, calculating the loss only for known targets.

scratch. The optimal network depth was determined empirically through extensive experimentation. This architectural simplification yields satisfactory results for the given tasks while significantly reducing computational requirements.

Classification Framework. In contrast to the classifier proposed by [20], we modify the classification objective. Instead of $\mathbf{x}$ being an image and $\mathbf{c}$ a class description, we feed the model with an n_{classes}-dimensional vector sampled from a random distribution and use an image representation extracted via CellProfiler [22] as conditioning $\mathbf{c}$, as illustrated in Figure 1. This approach allows us to generate predictions relying solely on features extracted from morphological image analysis.

3.2. Multi Label Classification with Conformal Prediction

We benchmark our method against conformal prediction, which is a commonly used technique for uncertainty estimation in the drug discovery process [2]. We first process image representation using an MLP classifier and then apply conformal prediction to the results.

While conformal prediction was primarily to work with single-label classification [3], it can also be adapted to work in the multi-label setting [24, 31]. To accurately gauge the model's confidence, we calculate separate distributions of conformal scores s for each class during the calibration step. We then calculate $1-\alpha$ quantile for each of those distributions, thus creating an individual threshold $(t_1, ..., t_n)$ for each class. For $i = 1, ..., n$, we classify predictions $y_i > t_i$ as positive, $y_i < 1 - t_i$ as negative, and other as uncertain.

4. MNIST-based Toy Dataset

4.1. Experiment design

We evaluate two models built from an equal number of equally sized layers, one utilizing diffusion and the other being regular MLP combined with conformal prediction. We train the models using a custom dataset, created by stitching together random digits coming from MNIST dataset [18], ensuring even distribution of all possible digit combinations. Each example is labeled by two classes corresponding to the digits present in the image. In this scenario, instead of CellProfiler, we extract image representation using CNN pretrained on the same dataset and then classify them using both models. To better replicate the HCS setting, Gaussian noise is added to each example during inference.

4.2. Results

We evaluate both models, aiming to compare the models' performance for a given confidence level as well as the number of labels that were rejected in order to obtain a confident prediction set. As illustrated in Figure 2, our HCS-DFC model achieves higher accuracy with greater data coverage compared to conformal prediction.

5. Mechanism of Action Prediction

5.1. Experiment design

To evaluate models for the HCS, we use publicly available features derived from Bray et al. dataset [8] using CellProfiler. We follow the steps proposed by [6] by first pooling MoA labels taken from the ChEMBL database and then selecting those with at least 25 active and 25 inactive labels in the dataset. Following [6], we also train both models using masked BCE loss, calculating the loss only for known targets. We use the weighted loss for the conformal model to

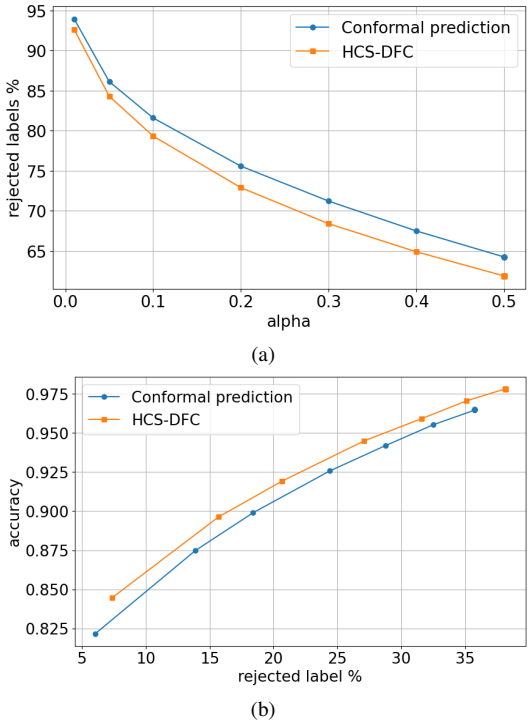

Figure 2. Percent of rejected labels depending on error rate (a) and accuracy depending on rejected labels (b) for MNIST-based toy dataset. Our HCS-DFC model achieves higher accuracy with greater data coverage compared to conformal prediction.

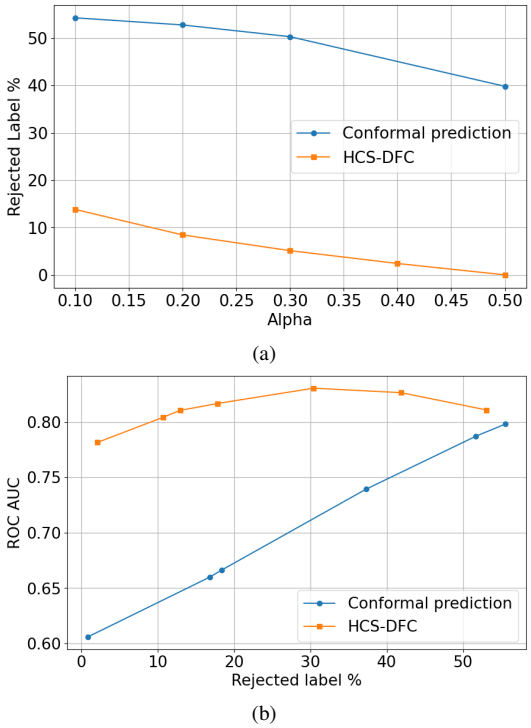

Figure 3. Percent of rejected labels depending on error rate (a) and accuracy depending on rejected labels (b) for MoA prediction. The diffusion-based approach achieves substantially greater coverage relative to the conformal-based approach.

ensure a fair comparison.

5.2. Results

As demonstrated in Figure 3, our HCS-DFC model exhibits superior overall performance compared to the conformal baseline. Notably, the diffusion-based approach achieves substantially greater coverage relative to the conformal model, suggesting that HCS-DFC represents a viable approach for reliability estimation in drug discovery applications.

6. Conclusions

In this work, we reformulate the problem of Mechanism of Action (MoA) prediction as a conditional generation process using our novel HCS-DFC model. We demonstrate the effectiveness of this approach on a synthetic dataset and a MoA prediction with the Bray et al. dataset. Our model achieves superior accuracy and coverage compared to the baseline method based on conformal prediction while maintaining the same desired reliability score.

Future research will focus on extensively evaluating the usefulness of this approach across various phenotypic-driven drug discovery tasks, such as phenotypic virtual screening [27]. Additionally, we aim to benchmark our model in other drug discovery-related tasks in which labels can be characterized as sparse and noisy. Finally, we plan to explore the integration of a conditioned diffusion-based classifier with deep image representation models, such as SubCell [14], to enhance performance further.

References

[1] Mohammad Akbarzadeh, Ilka Deipenwisch, Beate Schoelermann, Axel Pahl, Sonja Sievers, Slava Ziegler, and Herbert Waldmann. Morphological profiling by means of the cell painting assay enables identification of tubulin-targeting compounds. *Cell Chemical Biology*, 29(6):1053–1064, 2022. 1

[2] Jonathan Alvarsson, Staffan Arvidsson McShane, Ulf Norinder, and Ola Spjuth. Predicting with confidence: Using conformal prediction in drug discovery. *Journal of Pharmaceutical Sciences*, 110(1):42–49, 2021. 3

[3] Anastasios N. Angelopoulos and Stephen Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *CoRR*, abs/2107.07511, 2021. 2, 3

[4] Andreas Blattmann, Tim Dockhorn, Sumith Kulal, Daniel Mendelevitch, Maciej Kilian, Dominik Lorenz, Yam Levi, Zion English, Vikram Voleti, Adam Letts, Varun Jampani,

and Robin Rombach. Stable video diffusion: Scaling latent video diffusion models to large datasets, 2023. 2

[5] Adriana Borowa, Szczepan Kruczek, Jacek Tabor, and Bartosz Zieliski. Weakly-supervised cell classification for effective high content screening. In *International Conference on Computational Science*, pages 318–330. Springer, 2022. 1

[6] Adriana Borowa, Dawid Rymarczyk, Marek Żyła, Maciej Kańduła, Ana Sánchez-Fernández, Krzysztof Rataj, Łukasz Struski, Jacek Tabor, and Bartosz Zieliński. Decoding phenotypic screening: A comparative analysis of image representations. *Computational and Structural Biotechnology Journal*, 23:1181–1188, 2024. 1, 3

[7] Mark-Anthony Bray, Shantanu Singh, Han Han, Chadwick T Davis, Blake Borgeson, Cathy Hartland, Maria Kost-Alimova, Sigrun M Gustafsdottir, Christopher C Gibson, and Anne E Carpenter. Cell painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nature protocols*, 11(9):1757–1774, 2016. 1

[8] Mark-Anthony Bray, Sigrun M Gustafsdottir, Mohammad H Rohban, Shantanu Singh, Vebjorn Ljosa, Katherine L Sokolnicki, Joshua A Bittker, Nicole E Bodycombe, Vlado Dančík, Thomas P Hasaka, et al. A dataset of images and morphological profiles of 30 000 small-molecule treatments using the cell painting assay. *Gigascience*, 6(12):giw014, 2017. 1, 3

[9] Juan C Caicedo, Claire McQuin, Allen Goodman, Shantanu Singh, and Anne E Carpenter. Weakly supervised learning of single-cell feature embeddings. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9309–9318, 2018. 1

[10] Juan C Caicedo, John Arevalo, Federica Piccioni, Mark-Anthony Bray, Cathy L Hartland, Xiaoyun Wu, Angela N Brooks, Alice H Berger, Jesse S Boehm, Anne E Carpenter, et al. Cell painting predicts impact of lung cancer variants. *Molecular biology of the cell*, 33(6):ar49, 2022. 1

[11] Huanran Chen, Yinpeng Dong, Shitong Shao, Hao Zhongkai, Xiao Yang, Hang Su, and Jun Zhu. Diffusion models are certifiably robust classifiers. *Advances in Neural Information Processing Systems*, 37:50062–50097, 2024. 1

[12] Beth A Cimini, Srinivas Niranj Chandrasekaran, Maria Kost-Alimova, Lisa Miller, Amy Goodale, Briana Fritchman, Patrick Byrne, Sakshi Garg, Nasim Jamali, David J Logan, et al. Optimizing the cell painting assay for image-based profiling. *Nature protocols*, 18(7):1981–2013, 2023. 1

[13] Adam Corrigan, Philip Hopcroft, Ana Narvaez, and Claus Bendtsen. Batch mode active learning for mitotic phenotypes using conformal prediction. In *Conformal and Probabilistic Prediction and Applications*, pages 229–243. PMLR, 2020. 1

[14] Ankit Gupta, Zoe Wefers, Konstantin Kahnert, Jan N Hansen, Will Leineweber, Anthony Cesnik, Dan Lu, Ulrika Axelsson, Frederic Ballllosera Navarro, Theofanis Karaletsos, et al. Subcell: Vision foundation models for microscopy capture single-cell biology. *bioRxiv*, pages 2024–12, 2024. 4

[15] Dorota Herman, Maciej M Kanduła, Lorena GA Freitas, Caressa van Dongen, Thanh Le Van, Natalie Mesens, Steffen Jaensch, Emmanuel Gustin, Liesbeth Micholt, Charles-Hugues Lardeau, et al. Leveraging cell painting images to expand the applicability domain and actively improve deep learning quantitative structure–activity relationship models. *Chemical Research in Toxicology*, 36(7):1028–1036, 2023. 1

[16] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020. 2

[17] Oren Kraus, Kian Kenyon-Dean, Saber Saberian, Maryam Fallah, Peter McLean, Jess Leung, Vasudev Sharma, Ayla Khan, Jia Balakrishnan, Safiye Celik, et al. Masked autoencoders for microscopy are scalable learners of cellular biology. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11757–11768, 2024. 1

[18] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. 1, 3

[19] Laurent Lejeune, Morgane Roussin, Bruno Leggio, and Aurelia Vernay. An interpretable framework to characterize compound treatments on filamentous fungi using cell painting and deep metric learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 495–504, 2023. 1

[20] Alexander C. Li, Mihir Prabhudesai, Shivam Duggal, Ellis Brown, and Deepak Pathak. Your diffusion model is secretly a zero-shot classifier. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2206–2217, 2023. 2, 3

[21] Sean Lin, Kenji Schorpp, Ina Rothenaigner, and Kamyar Hadian. Image-based high-content screening in drug discovery. *Drug discovery today*, 25(8):1348–1361, 2020. 1

[22] Claire McQuin, Allen Goodman, Vasiliy Chernyshev, Lee Kamentsky, Beth A Cimini, Kyle W Karhohs, Minh Doan, Liya Ding, Susanne M Rafelski, Derek Thirstrup, et al. Cellprofiler 3.0: Next-generation image processing for biology. *PLoS biology*, 16(7):e2005970, 2018. 2, 3

[23] Nikita Moshkov, Michael Bornholdt, Santiago Benoit, Matthew Smith, Claire McQuin, Allen Goodman, Rebecca A Senft, Yu Han, Mehrtash Babadi, Peter Horvath, et al. Learning representations for image-based profiling of perturbations. *Nature communications*, 15(1):1594, 2024. 1

[24] Harris Papadopoulos. A cross-conformal predictor for multi-label classification. In *Artificial Intelligence Applications and Innovations: AIAI 2014 Workshops: CoPA, MHDW, IIVC, and MT4BD, Rhodes, Greece, September 19-21, 2014. Proceedings 10*, pages 241–250. Springer, 2014. 3

[25] James Pilling and Neil Carragher. *High Content Imaging in Drug Discovery*. Royal Society of Chemistry, 2025. 1

[26] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. *CoRR*, abs/2112.10752, 2021. 2

[27] Ana Sanchez-Fernandez, Elisabeth Rumetshofer, Sepp Hochreiter, and Günter Klambauer. Cloome: contrastive learning unlocks bioimaging databases for queries with

chemical structures. *Nature Communications*, 14(1):7339, 2023. 1, 4

[28] Srijit Seal, Maria-Anna Trapotsi, Ola Spjuth, Shantanu Singh, Jordi Carreras-Puigvert, Nigel Greene, Andreas Bender, and Anne E Carpenter. Cell painting: a decade of discovery and innovation in cellular imaging. *Nature methods*, pages 1–15, 2024. 1

[29] Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(3), 2008. 1

[30] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. pmlr, 2015. 1

[31] Chhavi Tyagi and Wenge Guo. Multi-label classification under uncertainty: a tree-based conformal prediction approach. In *Conformal and Probabilistic Prediction with Applications*, pages 488–512. PMLR, 2023. 3