CRISPR Off-Target Risk Prediction and Uncertainty

Juehao Pan*

College of Letters & Science, University of Wisconsin–Madison, Madison, Wisconsin 53706, United States

Abstract. Due to their high efficiency and programmability, CRISPR/Cas systems are commonly used in gene function experiments and gene therapy, but off-target effects are still a big problem-Cas proteins may cut non-target sites. Most machine learning and deep learning methods for predicting the off-target activity of gRNA–DNA pairs are point predictions, which are hard to quantify and weaken the validity of the assessment and amplify the risk of decision-making in high-risk scenarios. Therefore, uncertainty quantification approaches have been increasingly applied into the research of this field. This paper reviews the related works on applying uncertainty modeling in this field. Specifically, the concepts of data uncertainty and model uncertainty and their modeling techniques will be reviewed, including heteroscedastic regression, probabilistic predictive models, Deep Ensembles and MC Dropout. Furthermore, this paper will introduce some key metrics for the evaluation of the reliability of uncertainty models, including predictive accuracy, confidence interval calibration and performance on out-of-distribution (OOD) data. By reviewing the related studies, this paper concludes that it is very important to use uncertainty-aware models to improve the reliability of CRISPR predictions in this field.

1 Introduction

In recent years, besides its wide application in molecular biology research, CRISPR/Cas system has gradually developed into an indispensible tool in studying gene function. Due to its high programmability, superior efficiency, and relatively simple design, CRISPR-Cas9 has been widely used in research of gene function, building disease models, and studying potential gene therapies. By using gRNAs to guide Cas nucleases to target DNA sequences, researchers can introduce targeted modifications into the genome. This system has pushed the development of basic genetics and provided new approaches for precision medicine.

Despite its many advantages, CRISPR technology still meets a big challenge: off-target effects. Researchers have found that CRIPSR/Cas9 can induce cleavage at other, non-target sites across the entire genome and thus threaten the safety and efficiency of gene editing [1][2]. Off-target effects are the phenomenon that Cas proteins accidentally cut the genome at some other locations which are similar (but not identical) to the target sequence. If the Cas proteins cleave the DNA at these locations, they will cause unknown mutations on the genome. Previous experiments have confirmed that these erroneous cleavage events may really happen in human cells and even the whole genome [3]. In an experiment setting, these unintended edits may sometimes just be regarded as noises. However, in clinical applications, their impacts may be very serious. If it happens in gene therapy, even one gene editing mistake may cause permanent genetic editing in the patient and even some unknown biological effects.

Given the possible risks exists, accurately predicting CRISPR off-target activity has become a increasingly become a significant research problem in recent years. Recently, researchers began to use the machine learning(ML) and deep learning(DL) models to predict potential off-target cleavage activity based on gRNA–DNA sequence pairs. For example, the powerful prediction performance of models like Deep CRISPR in this task has already been showed in ref [4]. In addition, recent review papers summarized the wide application of deep learning methods in CRISPR off-target prediction [5] [6].

These models usually predict the cleavage probability of candidate sequences by learning some patterns from experimental data. However, most of the existing methods focus on improving the prediction accuracy, and these methods usually just give a "point prediction", which means that these models just output one value to represent the predicted cleavage activity.

However, in high-risk application scenarios like gene editing, it's usually insufficient to rely on only one value output by a model. For researchers, the question they usually want to know is not only what the predicted value is, but more importantly, is this value reliable? When sequences which are quite different from training data come, the model can still output some "definitive" predictions; The existing research has identified this situation is due to the lack of enough model generalization ability [6]—even if the model lacks enough information about such data. However, this

* Corresponding author: panjuehao1030@gmail.com

"overconfident prediction" is very dangerous in biomedical applications.

To address this problem, researchers began to introduce the concept of uncertainty quantification(UQ) into machine learning models. The predictive uncertainty in general can be divided into two types. The first is aleatoric uncertainty, which is caused by the random noise existing in the data itself, such as experimental errors and measurement fluctuations. This kind of uncertainty is usually hard to remove in practice. The second is epistemic uncertainty, which is caused by the model's lack of understanding for the data distribution, for example, limited training data or incompleteness of training data distribution. Unlike aleatoric uncertainty, epistemic uncertainty can be reduced in principle with more data available.

In the field of CRISPR off-target prediction, modeling uncertainty is very important. When the input sequence comes from a distribution different from the training data, a reliable model should output a higher level of predictive uncertainty, which can warn the researcher of the possible risk brought by current prediction. Therefore, introducing uncertainty estimation into deep learning models has become an important step to build more reliable predictive systems.

This paper mainly introduces the role of uncertainty modeling in CRISPR off-target prediction, discuss about the methods to model data uncertainties and model uncertainties, and explores how the uncertainty-aware models can improve the reliability of predictions in high-risk biomedical applications.

2 Methods to predict uncertainty

In CRISPR off-target prediction tasks, traditional deep models only provide one type of predictive output, i.e., cleavage activity scores for a certain gRNA–DNA sequence pair. However, in many application scenarios involving high safety requirements, such as gene editing, it is hard to make decisions based on single point predictions, and researchers need to know not only the predictions themselves, but also the reliability of these predictions. Therefore, it has become an important trend in recent research to introduce uncertainty modeling into CRISPR off-target prediction models.

Predictive uncertainty can be divided into two types: aleatoric uncertainty and epistemic uncertainty. Aleatoric uncertainty originates from the random noise existing in the experimental data itself, i.e., random sequencing errors, different experimental conditions and measurement instruments. In many CRISPR experiments, even for the same gRNA–DNA sequence pair, different experimental conditions will lead to significantly different measurement results ; this phenomenon has been widely verified in many CRPSR experiments [1][2]. Since this type of uncertainty exists in the randomly measured data itself, it will be hard to remove even if the model is sufficiently complex. Epistemic uncertainty occurs when the model's knowledge about the data distribution is incomplete, e.g., limited training data or incomplete training data distribution [7].

2.1 Data uncertainty characterization

A common method to characterize data uncertainty is heteroscedastic regression. In this method, the neural network outputs not only the mean of the target variable, but also the variance of the predicted value, and thus can characterize the noise level of different input samples. When the input sample shows high observational noise, the model outputs a larger predicted variance, and thereby reflects the uncertainty of that sample. In this way, the range of model output expands from a single predicted value to a certain probability distribution, and the predictions accommodate the random fluctuations of experimental data.

Another way to characterize data uncertainty is probabilistic output modeling. For example, in the CRISPR-DBA method, the model uses probabilistic modeling method to output predictive distribution instead of a single point estimate, and thereby characterizes the predictive uncertainty of off-target activity [8]. Experimental results show that this method not only maintains the prediction accuracy, but also outputs the variance information of each prediction, and thereby more accurately reflects the stochasticity existing in the experimental data.

2.2 Model uncertainty

One of the most widely used methods to deal with model uncertainty is Deep Ensembles. By training multiple independent neural networks and aggregating the output of these models, one can approximately estimate how uncertain the model is for certain predictions. Since the parameters learned by different models during the training process may differ significantly, the variance among the individual predictions of these models provides an estimate of the model's uncertainty with regards to the input data. Previous work has shown that compared to other methods, Deep Ensembles provide relatively stable estimates of uncertainty across a range of machine learning tasks, and also exhibit improved generalization in certain data-rich settings[7].

However, when the models are applied to input samples that are different from the training data distribution, the predictions made by the model may become unreliable. Ovadia et al. have shown that in cases where there is a change in the data distribution (dataset shift), uncertainty estimates may break even if the model generalizes well to the original training data[9]. Therefore, in practical applications, it is very important to include estimation of uncertainty methods to assess the model performance on the out-of-distribution data(OOD).

Another commonly used method is MC Dropout. Gal and Ghahramani proposed that Dropout can also be viewed as a type of Bayesian approximation method[10]. During the testing phase, Dropout layers are left on, and multiple forward passes are run on the same input data. Since the neural network's structure changes slightly each time a forward pass is made, the model outputs a set of slightly different predictions. By statistically analyzing these predictions, the uncertainty involved in the model's output can be further estimated. In practice,

during practical applications, the Dropout rate is usually set between 0.2 and 0.5, and 20 to 50 forward passes are run on the same input and a stable predictive distribution is obtained[10]. By computing the predicted mean and variance, the model's output is no longer a single point estimate, and instead a predictive distribution providing information regarding uncertainty. Denoting the T forward passes made on the same input, and getting predictions $\boldsymbol{f_t(x)}$, the predictive mean can be expressed as:

$$\mu = \frac{1}{T}\sum_{t=1}^{T} f_t\,(x) \tag{1}$$

The prediction variance can be expressed as:

$$\sigma^2 = \frac{1}{T}\sum_{t=1}^{T} (\,f_t(x) - \mu)^2 \tag{2}$$

By computing the predicted mean and variance, the model's output is no longer a single point estimate, and instead a predictive distribution providing information regarding uncertainty. For example, in practical applications, if the predicted mean for the same input data is 0.62 and the predicted variance is 0.04, the model can output a prediction interval and provide information regarding the reliability of that prediction result.

The experimental results of Gal and Ghahramani [10] show that compared to other methods, MC Dropout can effectively estimate model uncertainty while maintaining good prediction accuracy, and does so at relatively low computational cost, and is therefore commonly used in uncertainty quantification problems in deep learning.

After the introduction of uncertainty modeling methods, it is of importance to verify further if the model really increases the reliability of its predictions. First of all, the model should really provide reasonable predictions. For example, for the case of CRISPR-DBA, the Pearson correlation coefficient between the models predictions on the public datasets and the respective experimentally measured values is around 0.78 [8]. This value is comparable to the performance levels achieved by existing deep learning models for the CRISPR off-target prediction task [4]. If uncertainty modeling would really strongly degrade the predictions accuracy, the practical usage of such a method would be quite limited.

Secondly, it is of importance to verify if the prediction intervals are reasonably calibrated. For example, if a model outputs a 90% predictive confidence interval, one would expect that around 90 % of the actually observed values lay within that interval. If this rate drops significantly below the expected value, the model shows overconfidence; whereas if this rate significantly exceeds the theoretical value, the prediction intervals are too wide and the model failed to make full use of the information within the data. As pointed out by Ovadia et al. [9], the degree of calibration of the prediction intervals is one of the metrics to measure the reliability of uncertainty models.

Additionally, on the side of OOD data, the uncertainty of the models predictions should strongly increase. If input samples are not part of the data distribution during training, a reliable model should output more variance in the predictions to inform the researchers that the respective prediction might pose a high-risk. If a model continues to output low-variance predictions during OOD scenarios, this would indicate that the model is not able to detect unfamiliar data correctly and might lead to false high-confidence predictions [9].

3 Current challenges and future work

Although providing a suitable path for uncertainty quantification, current approaches still face several challenges. First, dataset shift leads to invalid model uncertainty estimates. If the training data differs from the actual application scenario, the output variance which the model is able to provide does not reflect the real underlying risk anymore. Second, existing models are mainly based on sequential information for predictions and therefore do not take full advantage of the more biological contextual information, e.g. chromatin, into account which leads to an information deficit regarding systematic errors and possible future research could add more biological structural information to enhance the cross domain generalization capability of the model as well as embed the OOD detection into the uncertainty estimation process to build a more robust risk assessment framework.

4 Conclusions

CRISPR/Cas system is a widely used and indispensable part of molecular biology research, and its safe application is limited by the off-target effects. In high-risk application scenarios, such as gene therapy, accurately predicting the off-target activity is not just a prediction problem, but a decision-making problem involving risk assessment. Most traditional deep learning models output only one predictive value. This "point prediction" method cannot reflect the model's confidence in different data situations, and may cause overconfident prediction.

This article reviews the latest progress in modeling uncertainty in CRISPR off-target prediction. The article starts by introducing the reader to two kinds of predictive uncertainty, namely aleatoric uncertainty and epistemic uncertainty. Afterwards, the article discusses several commonly used methods to model uncertainty. These methods include heteroscedastic regression and probabilistic predictive models (used to model data noise) as well as Deep Ensembles and MC Dropout methods (used to estimate model uncertainty). By using these methods, the model is no longer restricted to outputting one single estimate. Instead, it can output a predictive distribution. Thus, the researchers can use this to better understand the reliability of the prediction results.

Subsequently, this paper discusses the methods for evaluating uncertainty models. These include the metric of predictive accuracy, the metric of confidence interval calibration and the metric of model performance on OOD data. These metrics can help the researchers to determine whether the estimation of uncertainty enhances the credibility of the model.

The results of the analysis show that when evaluating the uncertainty model, one should not rely on only one factor but should consider all of the factors, including the metric of predictive accuracy, the metric of the calibration of confidence intervals and the metric of model performance on OOD data. Only by providing accurate uncertainty estimates while maintaining predictive performance can the model output more reliable results in high-stakes application scenarios, such as CRISPR gene editing.

References

1. S.Q. Tsai, J.K. Joung, Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. Nat. Rev. Genet. 17, 300–312 (2016)
2. X.-H. Zhang, L.Y. Tee, X.-G. Wang, et al., Off-target effects in CRISPR/Cas9-mediated genome engineering. Mol. Ther. Nucleic Acids 4, e264 (2015)
3. D. Kim, S. Bae, J. Park, et al., Digenome-seq: Genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nat. Methods 12, 237–243 (2015)
4. G. Chuai, et al., DeepCRISPR: Optimized CRISPR guide RNA design by deep learning. Genome Biol. 19, 80 (2018)
5. M. Cao, A. Brennan, C.M. Lee, et al., Deep learning based models for CRISPR/Cas off-target prediction. Small Methods 9, 2500122 (2025)
6. W. Du, et al., A versatile CRISPR/Cas9 system off-target prediction tool using language model. Commun. Biol. 8 (2025)
7. B. Lakshminarayanan, A. Pritzel, C. Blundell, Simple and scalable predictive uncertainty estimation using deep ensembles. Adv. Neural Inf. Process. Syst. 30 (2017)
8. X. Cao, P. Minary, CRISPR-DBA: A deep learning framework for uncertainty quantification of CRISPR off-target activities. In: Proc. IEEE Int. Conf. Bioinformatics and Biomedicine (BIBM) (2024)
9. Y. Ovadia, et al., Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift. Adv. Neural Inf. Process. Syst. 32 (2019)
10. Y. Gal, Z. Ghahramani, Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In: Proc. 33rd Int. Conf. Mach. Learn. (ICML), PMLR, 48, 1050–1059 (2016).