
Deep Learning in Genomics

Visualizing and interpreting deep learning methods for on-target prediction in CRISPR/Cas9

Klimentina Krstevska¹, Xilin Wang¹, Zhuoyang Lyu¹ and Hyegi Bang¹

¹ Department of Computer Science, Brown University, Providence, 02912, USA.

*To whom correspondence should be addressed.

Associate Editor: Ritambhara Singh

Received on 17 December 2023;

Abstract

Motivation: CRISPR/Cas 9 has become the most efficient technology for target gene editing and understanding the creation of highly efficient single-guide RNA (sgRNA) is of the highest importance for the scientific advancement of this technology. Currently there exist many deep-learning models that perform well when predicting the on-target activity of sgRNA, however, many of them face the black-box problem and lack any interpretability. In this project we attempt to bridge this gap and apply several visualization methods to different CRISPR on-target prediction models in order to gain interpretation into their predictions and link this to existing biological knowledge.

Results: We apply different visualization techniques including feature importance maps, gradient-based saliency maps and SHAP (Shapley Additive Explanations) for interpreting three different deep learning architectures (CNN-autoencoder, CNN-SVR, CNN-RNN) that perform on-target predictions. Our results show that feature importance maps give the least insights into the models while gradient-based saliency maps are able to show some meaningful interpretation. However, due to the zero-gradient problem with gradient-based saliency maps, SHAP is the technique which offers the most insights into the understanding of efficient sgRNA by recovering important and biologically relevant features.

Availability: Code and results available at github.com/Robins0000onLyu/CSCI2952G-group-project
Contact:

Supplementary information: Supplementary data and information are available at the end of this paper.

1 Introduction

The Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated protein 9, commonly referred to as CRISPR/Cas 9 is the most accurate and versatile technique for targeted gene editing. By precisely manipulating, inserting and deleting sequence fragments in the DNA the potential applications for the CRISPR/Cas 9 technology include developing highly-effective drugs, gene therapy, pathogen detection, biofuel production and plant biotechnology. This system comprises two components: the Cas9 protein, which possesses nuclease activity, and a single-guide RNA (sgRNA) responsible for programming specificity. The sgRNA directs the protein to precise genomic targets. When choosing the optimal

target site, an ideal sgRNA should be capable of maximizing on-target activity (guide efficiency) while simultaneously minimizing unfavorable off-target effects (guide specificity). Nowadays deep learning models have dominated the field with their accurate prediction of on-target activity for a sgRNA. However, the black-box problem of deep learning models persists and in this paper we perform a comparative analysis of different models used for on-target prediction with the aim to see which one performs the best on a given dataset and understand how each one makes its predictions. With this our aim is to close an existing research gap in the field and utilize our visualization methods to better understand efficient sgRNA.

2 Related Work

A recent review by Sherkatghanad et al.[1] summarizes existing deep learning models applied to the prediction of on-target efficacy and off-target profile of a given sgRNAs. However, this extensive summary lacks a detailed look into the explainability and interpretability of the listed deep learning methods. Therefore, we have selected three different architectures for which we offer prediction explanation and model interpretation when comparing the models' performance on the same dataset. The first model is DeepCRISPR [2], an autoencoder based on deep convolutional denoising neural network (DNCNN), which was the first deep learning model applied to the sgRNA design and novel in its automated feature selection. The model's convolutional neural network (CNN) predicts the sgRNA efficiency on the CRISPR/Cas sgRNA sequence representation extracted by the autoencoder. The second model is a hybrid CNN-SVR [3] in which the CNN is trained to extract resilient features from the sgRNA, and then a support vector regression (SVR) classifier is employed to make predictions about the sgRNA's efficacy. The third model is C-RNNCrispr [4], a combination of CNN and bidirectional gated recurrent unit (GRU), where the former is used to extract features and the latter models the abstract sgRNA features' sequential dependencies. Another drawback of Sherkatghanad et al. [1] is the lack to compare the models on the same data (where applicable) and show the performance of each model on a dataset different from the one that they were trained on, which we address in our experiments and determine how much each of the three model is able to generalize on unseen data. Finally we compare the model's performance across the same metrics for all of them instead of just reporting the ones that have been used individually for each model.

3 Methods

3.1 Data

For running our experiments we used the public experimentally-validated gRNA datasets packed by Chuai et al [2]. Every entry in these datasets contains the specified 23-nt gRNA sequence, binary cleavage efficacy and numerical cleavage efficacy. For each entry there are also four different kinds of epigenetic features which have been obtained from ENCODE[6]. These are DNA methylation information from the RRBS assay, chromatin accessibility information from the DNase-Seq assay, and CTCF binding and H3K4me3 information from the ChIP-Seq assay. Every epigenetic feature is represented by a sequence of "A"s and "N"s with length 23, where "A" means that the epigenetic feature is present at the specific base position and "N" means that it is absent. Binary classification of gRNA cleavage efficacy is established with a value of 1 denoting high efficiency and 0 indicating low efficiency.

3.2 Visualization Methods:

3.2.1 Feature Importance Map

For the on-target prediction task we visualize the feature importance for each model with the method presented by Xie et al., in which each nucleotide's importance is determined directly by feeding it into the network [14].

Specifically, for each possible nucleotide (A, T, C, or G) at each position (1-23), one sequence is constructed by assigning a 1 to this target of interest, and 0 to all other nucleotides and positions. Corresponding to this constructed sequence, the four epigenetic sequences are created similarly. Then, this sequence along with its epigenetic features are fed into the neural network, and the output indicates the probability of such sequence being an efficient sgRNA. In other words, this score represents the importance score of the nucleotide at that position along with its epigenetic features. Eventually, a heat map is generated where each entry is the importance score for the corresponding nucleotide and position in the sequence.

3.2.2 Gradient-Based Saliency Map

For gradient-based saliency maps, we visualize the influence of the nucleotides (A, T, C and G) at each position in the sequence on the final output score with the method presented in Simonyan et al. [16]. Here the derivative is with regard to the input sequence and it is easy to compute because it simply is one step of backpropagation in the models. The derivatives' values for the nucleotides tell us what smallest nucleotide changes in the sequence have the largest effect on the output score.

3.2.3 SHAP (Shapley Additive Explanations) Value (Kernel SHAP)

SHAP value is an interpretation approach for machine learning models using the ideas of Shapley values from game theory [15]. It calculates the contribution score for each input feature to the model's prediction. In particular, SHAP value summarizes the currently available additive feature attribution methods, such as LIME and DeepLIFT, leveraging insights from these approaches to effectively calculate the importance scores for every input feature [15].

We present here the Kernel SHAP, which utilizes the idea of linear explanation in Linear LIME to approximate SHAP values [15]. Specifically, Linear LIME attempts to interpret models by locally approximating the model using the linear function:

$$g(z) = \phi_0 + \sum_{i=1}^M \phi_i z_i,$$

where M is the dimension of input features, and z represents the encoding for a M-dimensional input in a binary form, such that $z \in \{0, 1\}^M$. Each ϕ_i is the weight for the corresponding dimension in the input feature. This linear approximation is learnt through minimizing the loss:

$$\argmin_{g \in G} L(f, g, \pi_x) + \Omega(g)$$

In this equation, f is the original model, L is the loss function, π_x is the weighting kernel on the input space, and Ω is a regularization function on g. Kernel SHAP uses the specific combination of L, π_x , and Ω such that the learned linear approximation recovers the Shapley values from game theory.

Note that Kernel SHAP is a model-agnostic estimation of SHAP values as it only relies on the linear regression function g .

The use of Kernel SHAP requires the specification of a reference value. This reference value, or background data, represents a neutral and uninformative input such that prediction on this reference value reflects the expectation of the model outcome. In Kernel SHAP, the importance score for a feature is calculated by setting the feature value to the reference value and examining changes in model prediction. We implement Kernel SHAP using the KernelExplainer from the Python shap library.

The choice of an appropriate reference value plays a crucial role in accurately assessing the importance of features. When comparing different reference values, such as a zero input versus context-specific references like the frequency of nucleotides for DNA or the mean of input values for epigenetic data, the aim is to identify the reference that best captures the model's baseline expectation. The zero input reference, where all feature values are set to zero, serves as a neutral and uninformative baseline, representing a scenario where no specific input information is considered. This approach is particularly suitable when the model's behavior is expected to be unbiased and symmetrical around zero. On the other hand, context-specific references, such as nucleotide frequencies or mean epigenetic values, reflect domain-specific expectations. The choice between these approaches should consider the nature of the data and the model's sensitivity to different input features. In instances where the model is agnostic to specific patterns in the input data, the zero input reference may be preferred as it provides a generalizable and neutral baseline for feature importance calculations.

4 Experiments and Results:

We apply effective and comparable visualization approaches on the three models to analyze important sgRNA features that potentially lead to a better sgRNA design. The similarity and differences in the models' visualizations directly reflect the model's strengths. For more details on the models' different architectures, the definitions and calculations of evaluation metrics as well as additional results, please refer to the supplementary section at the end of this paper.

4.1 Feature importance map is insufficient

The feature importance map used by Zhang et al. [4] fails to consider interactions among the sequence locations. Indeed, this importance map is created by zeroing out all remaining features, focusing solely on one nucleotide at a specific location. This independent consideration is insufficient to reveal the complex mechanism of sgRNA in on-target cleavage, as can be seen from the heatmap in Figure 1 which does not show much biological relevance, except for the general positions more adjacent to the PAM region of the sequence being generally more important than the PAM-distal regions for the identification of an efficient sgRNA [17].

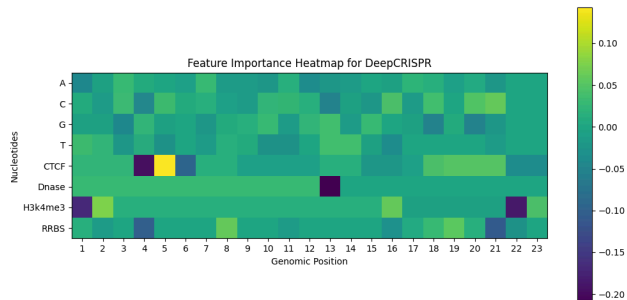


Fig. 1. Feature Importance Map for DeepCRISPR model

4.2 Gradient-based saliency maps show some meaningful interpretation

To overcome the shortcomings of feature importance maps, we have incorporated a gradient-based saliency map as an additional interpretation approach, and compared the results to the existing feature importance map.

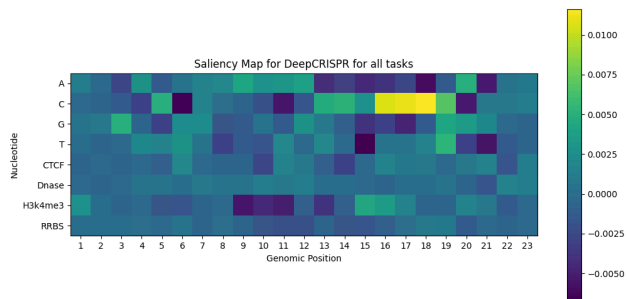


Fig. 2. Gradient-Based Saliency Map for DeepCRISPR model

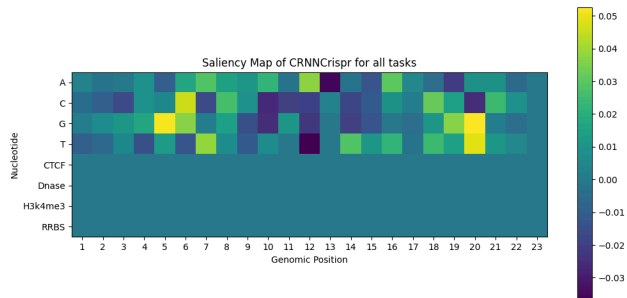


Fig. 3. Gradient-Based Saliency Map for C-RNNCrispr model

We see an alignment between the visualization maps and previous findings. In DeepCRISPR, cytosine is preferred at position 18 which is the DNA cleavage site for the CRISPR system, consistent with in vitro and in vivo studies [9, 10]. Furthermore, the location of PAM NGG exhibits a low preference for thymine [9, 10]. The PAM-adjacent regions are more crucial in on-target prediction than the PAM-distal region in the sgRNA [17].

Interestingly, for C-RNNCrispr we see a preference for thymine at position 20 consistent with previous studies that adenine or thymine at this position increases the indel [18]. However, the saliency maps for the four different kinds of epigenetic features in this model suffer from the zero-gradient problem, exhibiting the shortcomings of relying solely on this interpretation method.

4.3 SHAP value recovers important and biologically relevant features

Given the zero-gradient issue in gradient-based saliency map, we apply Kernel SHAP to visualize the contribution of each input feature to the C-RNNCrispr and CNN-SVR models prediction. The SHAP-based saliency map has the same configuration as the feature importance map and the gradient-based saliency map for direct comparison. For the genomic sequence component of the inputs, we also visualize them through figures in which the base letter at each position is present along with its height indicating the importance score. An upright letter indicates positive contribution, and an inverted letter means negative contribution. We apply the Kernel SHAP visualization on the selected maximum-frequency and minimum-frequency data samples. We observed that the SHAP visualizations using zero-input as a reference value showed better results for both models than the visualizations using nucleotide frequencies and mean epigenetic values, presented in supplementary information.

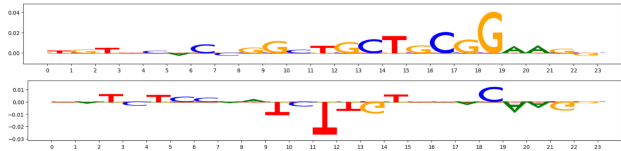


Fig. 4. Input sequences for minimum and maximum data points

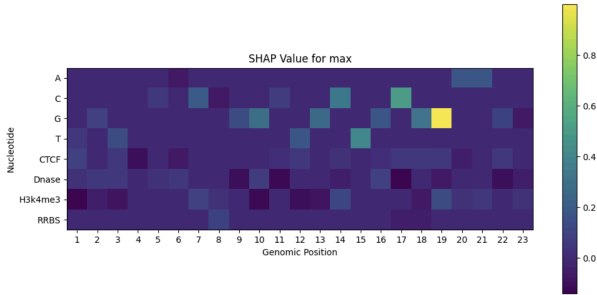


Fig. 5. SHAP Values Map for C-RNNCrispr model for maximum data point

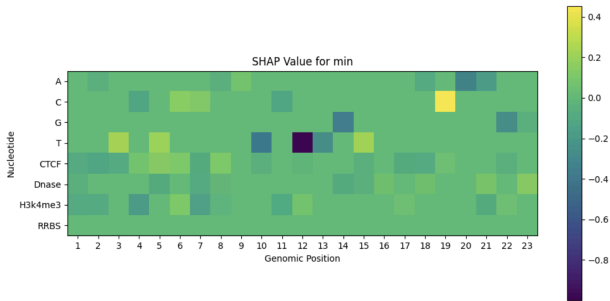


Fig. 6. SHAP Values Map for C-RNNCrispr model for minimum data point

The following results can be observed in Figures 6-8: (i) The positions more adjacent to the PAM region of the sequence are generally more important than the PAM-distal regions for the identification of an efficient sgRNA [17] (ii) Homopolymers (a consecutive sequence of two or more identical nucleotides) are preferred at position 17-19. This is consistent with previous studies that homopolymers increase the proportion of deletions at the cut site (3bp upstream of PAM) [18] (iii) Cytosine is also

favorable at position 17, as indicated from a previous study that points to the possibility of the cleavage being 3bp or more upstream of the PAM [19] (iv) DNA methylation (H3K4me3) contributes negatively to the model prediction as indicated from the SHAP-based saliency map for the minimum-frequency sample. This avoidance of DNA methylation was found by Chuai et al [2]. In general, an efficient sgRNA is signified by specific nucleotides at particular location and a general avoidance of DNA methylation as indicated by H3K4me3 features. We also found observations that were not present in any of the examined model's papers: (i) nucleotide at the second position upstream to the PAM is important [20], and both models have a high preference of guanine consistent with Hiranniramol et al. [21] (ii) Both CNN-SVR and C-RNN show disfavor of thymine, which is consistent with a previous finding that thymine is depleted for most of the positions of a functional sgRNA [21].



Fig. 7. SHAP Values Map for CNN-SVR model for maximum data point

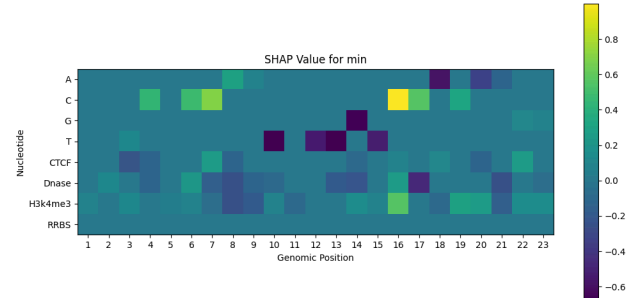


Fig. 8. SHAP Values Map for CNN-SVR model for minimum data point

5 Discussion

Given that DL methods have emerged as state-of-the art approaches for predicting on-target activity in CRISPR/Gas9, we work on overcoming the black-box problem by comparing different architectures' performances and extracting meaningful features from the models to better understand how each one makes its predictions. This comparative analysis offers interesting insights into the optimization of sgRNA designing and the benefits and disadvantages of using different visualization techniques on different DL architectures for the CRISPR/Gas9 on-target prediction task. The visualizations show the underlying biological concepts that these models are learning and which have previously been well established. We have also highlighted the biological relevance of these prediction models which have not been previously linked to them.

Contribution

Zhuoyang Lyu proposed the idea of working with the CRISPR/Cas9 system and worked on the C-RNNCrispr model maps generation.

Xilin Wang worked on all the maps generation for CNN-SVR and the SHAP implementation for both CNN-SVR and C-RNNCrispr.

Hyegi Bang worked on all the map generation for the DeepCRISPR model including all the necessary Docker setup to run Tensorflow version 1.3.

Klimentina Krstevska gathered the data, the minimum and maximum points for testing and all the biological information to be used for interpretation of our results as well as doing the organizing and writing of the majority of the paper.

All authors contributed to the paper writing and presentations preparation.

References

- [1] Zeinab Sherkatghanad, Moloud Abdar, Jeremy Charlier, Vladimir Makarenkov, Using traditional machine learning and deep learning methods for on- and off-target prediction in CRISPR/Cas9: a review, *Briefings in Bioinformatics*, Volume 24, Issue 3, May 2023, bbad131, <https://doi.org/10.1093/bib/bbad131>
- [2] Chuai, G., Ma, H., Yan, J. *et al.* DeepCRISPR: optimized CRISPR guide RNA design by deep learning. *Genome Biol* 19, 80 (2018). <https://doi.org/10.1186/s13059-018-1459-4>
- [3] Zhang, G., Dai, Z., & Dai, X. (2020). A Novel Hybrid CNN-SVR for CRISPR/Cas9 Guide RNA Activity Prediction. *Frontiers in genetics*, 10, 1303. <https://doi.org/10.3389/fgene.2019.01303>
- [4] Zhang G, Dai Z, Dai X. C-rnn-crispr: prediction of crispr/cas9 sgRNA activity using convolutional and recurrent neural networks. *Comput Struct Biotechnol J* 2020; 18:344–54.
- [5] Badaro, G., Hajj, H., El-Hajj, W., Nachman, L. (2013). “A hybrid approach with collaborative filtering for recommender systems,” in: 2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC) (Publisher: IEEE), pp. 349–354. doi: 10.1109/IWCMC.2013.6583584
- [6] Consortium, E. P. (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 306 (5696), 636–640. doi: 10.1126/science.1105136
- [7] Kim, D., Kim, J., Hur, J. K., Been, K. W., Yoon, S. H., Kim, J. S. (2016). Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* 34 (8), 863–868. doi: 10.1038/nbt3609
- [8] Kleinstiver, B. P., Tsai, S. Q., Prew, M. S., Nguyen, N. T., Welch, M. M., Lopez, J. M., et al. (2016). Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat. Biotechnol.* 34 (8), 869–874. doi: 10.1038/nbt3620
- [9] Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., et al. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* 32 (12), 1262. doi: 10.1038/nbt3026
- [10] Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34 (2), 184. doi: 10.1038/nbt3437
- [11] Hart, T., Chandrasekhar, M., Aregger, M., Steinhart, Z., Brown, K. R., MacLeod, G., et al. (2015). High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* 163 (6), 1515–1526. doi: 10.1016/j.cell.2015.11.015
- [12] Wang, T., Wei, J. J., Sabatini, D. M., Lander, E. S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343 (6166), 80–84. doi: 10.1126/science.1246981
- [13] Simonyan K, Vedaldi A, Zisserman A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv:1312.6034*
- [14] Xie B, Jankovic BR, Bajic VB, Song L, Gao X. Poly(A) motif prediction using spectral latent features from human DNA sequences. (2013). *Bioinformatics*. 29(13):i316-25. doi: 10.1093/bioinformatics/btt218. PMID: 23813000; PMCID: PMC3694652.
- [15] Lundberg, S.M., & Lee, S. (2017). A Unified Approach to Interpreting Model Predictions. *Neural Information Processing Systems*.
- [16] Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*.
- [17] Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology*, 31(3), 233–239. <https://doi.org/10.1038/nbt.2508>
- [18] Leenay, R. T., Aghazadeh, A., Hiatt, J., Tse, D., Roth, T. L., Apathy, R., Shifrut, E., Hultquist, J. F., Krogan, N., Wu, Z., Cirolia, G., Canaj, H., Leonetti, M. D., Marson, A., May, A. P., & Zou, J. (2019). Large dataset enables prediction of repair after CRISPR-Cas9 editing in primary T cells. *Nature biotechnology*, 37(9), 1034–1037. <https://doi.org/10.1038/s41587-019-0203-2>
- [19] Shou, J., Li, J., Liu, Y., & Wu, Q. (2018). Precise and Predictable CRISPR Chromosomal Rearrangements Reveal Principles of Cas9-Mediated Nucleotide Insertion. *Molecular cell*, 71(4), 498–509.e4. <https://doi.org/10.1016/j.molcel.2018.06.021>
- [20] Abadi, S., Yan, W. X., Amar, D., & Mayrose, I. (2017). A machine learning approach for predicting CRISPR-Cas9 cleavage efficiencies and patterns underlying its mechanism of action. *PLoS computational biology*, 13(10), e1005807. <https://doi.org/10.1371/journal.pcbi.1005807>
- [21] Hiranniramol, K., Chen, Y., Liu, W., & Wang, X. (2020). Generalizable sgRNA design for improved CRISPR/Cas9 editing efficiency. *Bioinformatics (Oxford, England)*, 36(9), 2684–2689. <https://doi.org/10.1093/bioinformatics/btaa041>

Supplementary Information

6 Metrics

6.1 AUROC

Area Under Receiver Operating Characteristic Curve (AUROC) is a widely used metric in deep learning that evaluates performance of a classification model. One major advantage of using AUROC compared to vanilla accuracy is that it considers every threshold and therefore evaluates the model independent of the choice of the threshold. A major problem arises from the fact that AUROC is in its essence a measurement of how good a model is at ranking the data by the predicted probability. As evidence, if we multiply every single prediction by a constant, the AUROC will not change. Therefore, AUROC might not truthfully represent a model's performance when the dataset suffers from label imbalance due to the use of FN. AUROC tends to give an overly optimistic evaluation to the model when there is a significant class imbalance in the data.

6.2 Spearman Correlation

Spearman Correlation measures the strength and direction of two ranked variables. It compares the rank of the predicted value against the rank of true values. Spearman Correlation measures only the monotonic association between two variables, so it is suitable when we only care about the rank of the prediction rather than the actual predicted value. Just like AUROC, it measures the model's performance or ranking data independent of the actual values of the predictions and the ground truth labels. It is robust to outliers and appropriate for ordinal and non-normally distributed data. A downside to this metric is that it ignores Non-Monotonic Relationships between the variables and does not capture other types of relationships, such as curvilinear or nonlinear associations between the variables.

7 Models:

7.1 DeepCRISPR

DeepCRISPR [2] represents a comprehensive computational platform that harnesses deep learning to forecast the effectiveness of single guide RNA (sgRNA) for knocking out target genes. It streamlines the process of identifying epigenetic attributes and sequences which can impact the efficiency of sgRNA knockout through data-driven methods. This autoencoder-based deep convolutional denoising neural network (DNCNN) utilizes the convolutional neural network (CNN) to predict the sgRNA efficiency on the CRISPR/Cas sgRNA sequence representation extracted by the autoencoder. The architecture of DeepCRISPR employs deep learning models for both classification (for on-target prediction) and regression (for sgRNA knockout efficacy prediction) tasks..

7.2 CNN-SVR

CNN-SVR [3] is a unique combination of two essential components: a convolutional neural network (CNN) tasked with gRNA feature extraction and a support vector regression (SVR) system responsible for the prediction of the gRNA on-target efficiency. The CNN-SVR system is composed of a front-end merged CNN for gRNA feature extraction and a back-end SVR for regression and prediction tasks.

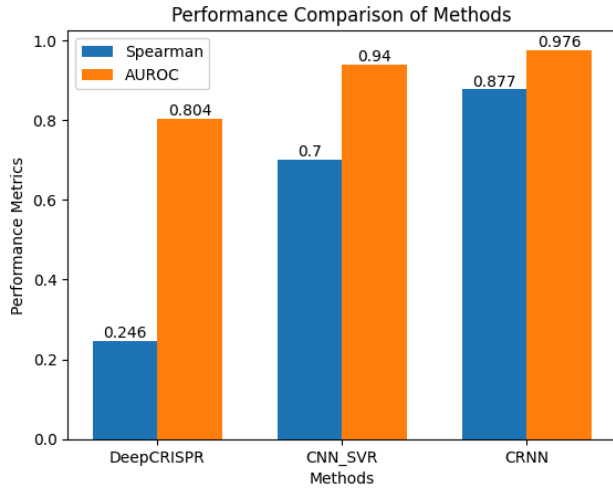
The CNN-SVR model that's been proposed consists of two main flows which are the gRNA stream and the epigenetic stream. Both of the branches share an identical structure, which encompasses two one-dimensional (1D) convolution layers, followed by two average-pooling layers, followed by four fully connected layers. The convolution layers (conv1 and conv2) and average pooling layers (pool1 and pool2) are instrumental in extracting features from the input data. The output of these layers is then flattened and subsequently fed through four fully connected layers (fc1, fc2, fc3, and fc4), with their sizes progressively decreasing to 256, 128, 64, and 40, respectively. To maintain model robustness and mitigate overfitting, dropout is applied as a regularization technique.

7.3 C-RNNCrispr

C-RNNCrispr [4] represents a hybrid framework that skillfully blends two powerful neural network components: convolutional neural networks (CNNs) and bi-directional gate recurrent unit networks (BGRU) for the purpose of forecasting the on-target activity of CRISPR/Cas9 sgRNAs. The architectural design of C-RNNCrispr comprises two distinct branches, namely sgRNA and epigenetic. In the sgRNA branch, the model takes the binary matrix representation of the sgRNA sequence as input, while the epigenetic branch leverages four key epigenetic features as its input. These two branches collaborate to generate a regression score as the final output.

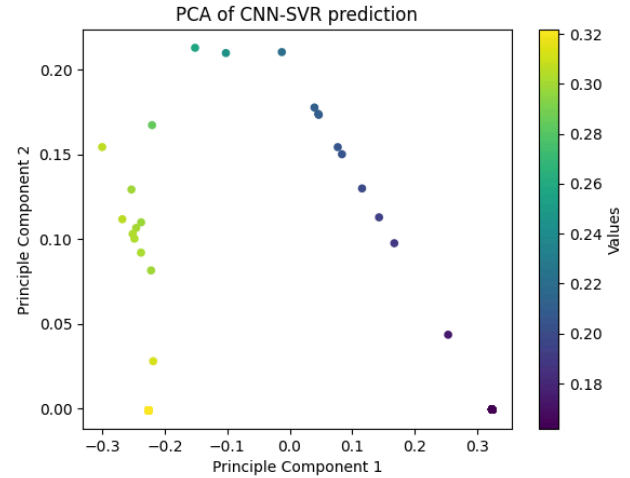
7.4 Model Performance Comparison:

DeepCRISPR has the lowest Spearman value, indicating a weaker monotonic relationship compared to CNN-SVR and C-RNNCrispr. C-RNNCrispr has the highest AUROC, suggesting superior discriminatory ability compared to DeepCRISPR and CNN-SVR



Sup. Fig. 1 AUROC and Spearman Performance Comparison of the three models

which the features fall. We were not able to extract any relevant biological features from this graph.

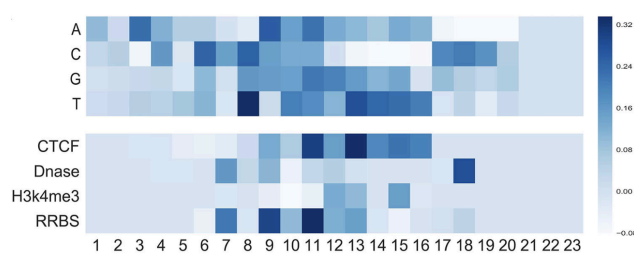


Sup. Fig. 4. Principal Component Analysis (PCA) of CNN-SVR model

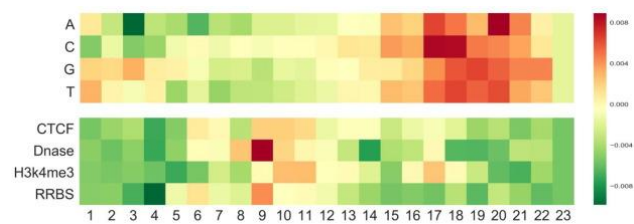
8 Experiments and Results:

8.1 Feature importance map

Below are the feature maps for the CNN-SVR and C-RNNCrispr models. We can see from Figure 1 and Supplementary Figures 2 and 3 that the heatmap for the C-RNNCrispr model exhibits the well-known pattern of general positions more adjacent to the PAM region of the sequence being more important than the PAM-distal regions for the identification of an efficient sgRNA [17].



Sup Fig. 2. Feature Importance Map for CNN-SVR model [3]



Sup. Fig. 3. Feature Importance Map for C-RNNCrispr model [4]

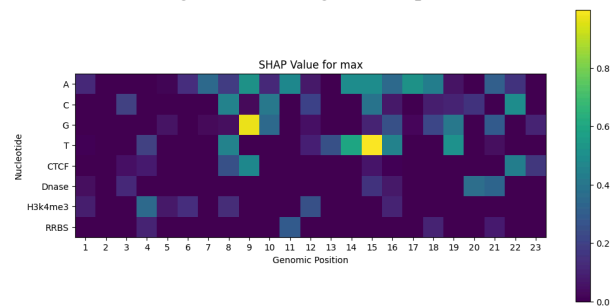
8.2 Gradient-based saliency maps

For CNN-SVR, because of the SVR classifier that is sitting on top of the CNN, it is not possible to generate a gradient-based saliency map on the entire model's architecture, so instead we performed a Principal Component Analysis (PCA) for this model and displayed the visualization in Supplementary Figure 4 from which we can observe two general clusters in

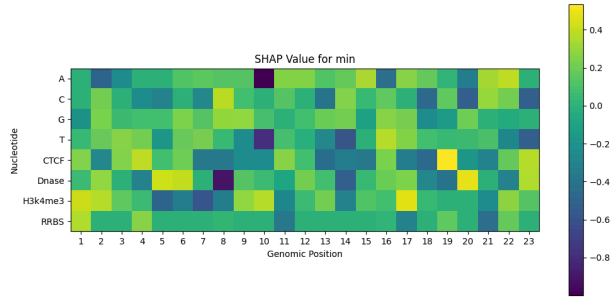
8.3 SHAP

In the main part of the paper, we presented the SHAP value maps for which the reference value was the zero input because we observed that these SHAP visualizations showed better results for both models which we assume is because the zero-input reference provides a generalizable and neutral baseline for feature importance calculations.

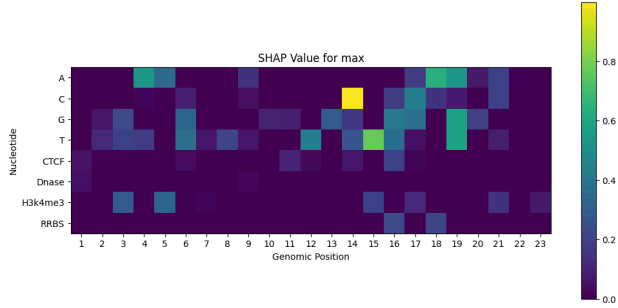
Here we also provide the Kernel SHAP visualization on the selected maximum-frequency and minimum-frequency data samples with the nucleotide frequencies and mean epigenetic values being used as the reference points in the calculations. Interestingly, the SHAP values maps for the minimum data point for both the CNN-SVR and C-RNNCrispr models have high variation between neighboring squares and it is impossible to gather any relevant biological interpretation from them. The SHAP values maps for the maximum data point for both models are more concise and slightly exhibit the general rule that the positions more adjacent to the PAM region of the sequence are generally more important than the PAM-distal regions for the identification of an efficient sgRNA [17], however, besides that there are no other significant biological interpretations.



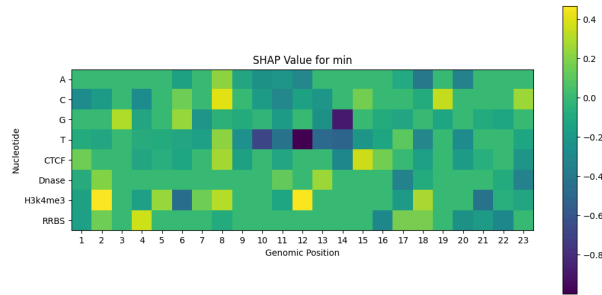
Sup. Fig. 5. SHAP Values Map for CNN-SVR model for maximum data point



Sup. Fig. 6. SHAP Values Map for CNN-SVR model for minimum data point



Sup. Fig. 7. SHAP Values Map for C-RNNCrispr model for maximum data point



Sup. Fig. 8. SHAP Values Map for C-RNNCrispr model for minimum data point

We were not be able to apply the Kernel SHAP visualization for analyzing the DeepCRISPR model because the Python shap library installation requires at least TensorFlow version 2 and DeepCRISPR is written in TensorFlow version 1 and does not run in environments with later TensorFlow versions.

9 Future Work

In the future we would like to extend this analysis to other models which perform off-target and both on-target and off-target predictions for CRISPR/Gas9.