HIGH SCHOOL EDITION
Journal of Student Research

Volume 12 Issue 3 (2023)

Predictability of Off-Targets in CRISPR-Cas9 Gene Editing Systems using Convolutional Neural Networks

Parsh Verma, Jothsna Kethar[#] and Dr. Rajagopal Appavu, Ph.D.[#]

[#]Advisor

ABSTRACT

CRISPR-Cas systems have catalyzed a quantum leap in genomics, providing a remarkable degree of precision in genomic sequence alterations. Despite these advancements, one notable conundrum persists - the prediction and mitigation of off-target effects. These unintended effects occur when the CRISPR apparatus interacts with homologous, albeit non-identical sequences, which could potentially induce unanticipated mutations. This investigation employs a comparative approach, juxtaposing machine-learning models that harness sophisticated Convolutional Neural Networks with different architectures. The standard CNN model utilizes a stratified architecture that meticulously filters pertinent genetic patterns, and identifies salient genetic features using a convolutional layer along with a batch normalization layer and a max pooling layer. This architecture prevents overfitting by applying a dropout layer, and 2 dense layers to categorize the two outputs. In contrast, the AttnToMismatch_CNN model marries the attention mechanism with the convolutional paradigm to encode sgRNA and DNA sequences into vector representations by using embedding and transformer layers. Which is then passed into a convolution layer and then to a dense layer to result in two outputs. Performance appraisal of the models, through the Area Under the Curve (AUC) of the Receiver Operating Characteristic Curve (AUC-ROC) score, indicated the standard CNN model's superior predictive accuracy. This research accentuates the untapped potential of Convolutional Neural Networks in augmenting the predictability of off-target effects in CRISPR-Cas systems, thereby fostering safer and more efficacious applications of this transformative gene editing tool.

Introduction

Gene Editing is a technique that enables precise and efficient modification of the genetic material of living organisms. It has been revolutionary to the biotechnology field and is very promising for advancing human health and addressing some of the most pressing challenges yet to be resolved. The clustered regularly interspaced short palindromic repeats(CRISPR)-associated(Cas) protein 9 system has been one of the most widely used techniques for precise gene editing due to its ease of use, accuracy, and high efficiency. The CRISPR-Cas9 system, utilizing single-guide RNA and Cas9 protein to carry out insertions, deletions, or replacements of genetic fragments, has made significant progress in gene editing by offering promising solutions for treating various diseases, such as genetic disorders, cancer, viral infections, autoimmune disease, etc. However, despite the high efficiency of the CRISPR-Cas9 system, it still has several limitations, such as, including off-target effects, low specificity in certain cases, and occasionally resulting in incomplete editing. To overcome these challenges, researchers have begun exploring the application of Artificial Intelligence(AI) models to enhance the precision, accuracy, and predictiveness of the CRISPR-Cas9 system. More specifically, AI algorithms can analyze vast amounts of genomic data and predict the most effective guide RNA for a given target. Machine Learning models can learn from previous CRISPR-Cas9 experiments and build upon them to optimize future results.

Overview: CRISPR-Cas9

What is CRISPR?

Bacteria are frequently exposed to viral infections and have limited time to identify and neutralize them. To combat this challenge, many bacteria possess an adaptive immune system known as CRISPR, which enables them to detect and destroy viral DNA. One crucial component of the CRISPR system is a protein called Cas9, which can locate and cleave viral DNA. When a virus infects a bacterial cell, it injects its DNA, and the CRISPR system enables the extraction of this DNA from the virus and its integration into the bacterium's chromosome. The clustered regularly interspaced short palindromic repeats (CRISPR) mechanism allows cells to record the viruses they encounter over time, and this information is passed down through generations via the CRISPR locus. As a result, cells are protected over multiple generations, as they possess a record of past infections. Following the insertion of viral DNA bits into the bacterial chromosome, the cell synthesizes a copy of RNA. RNA interacts with DNA molecules that possess matching sequences. RNA associates with the Cas9 protein, which searches the DNA for matching RNA sites. When it locates a site, the complex attaches to the DNA, enabling the Cas9 cleaver to make a highly precise cut in the viral DNA.

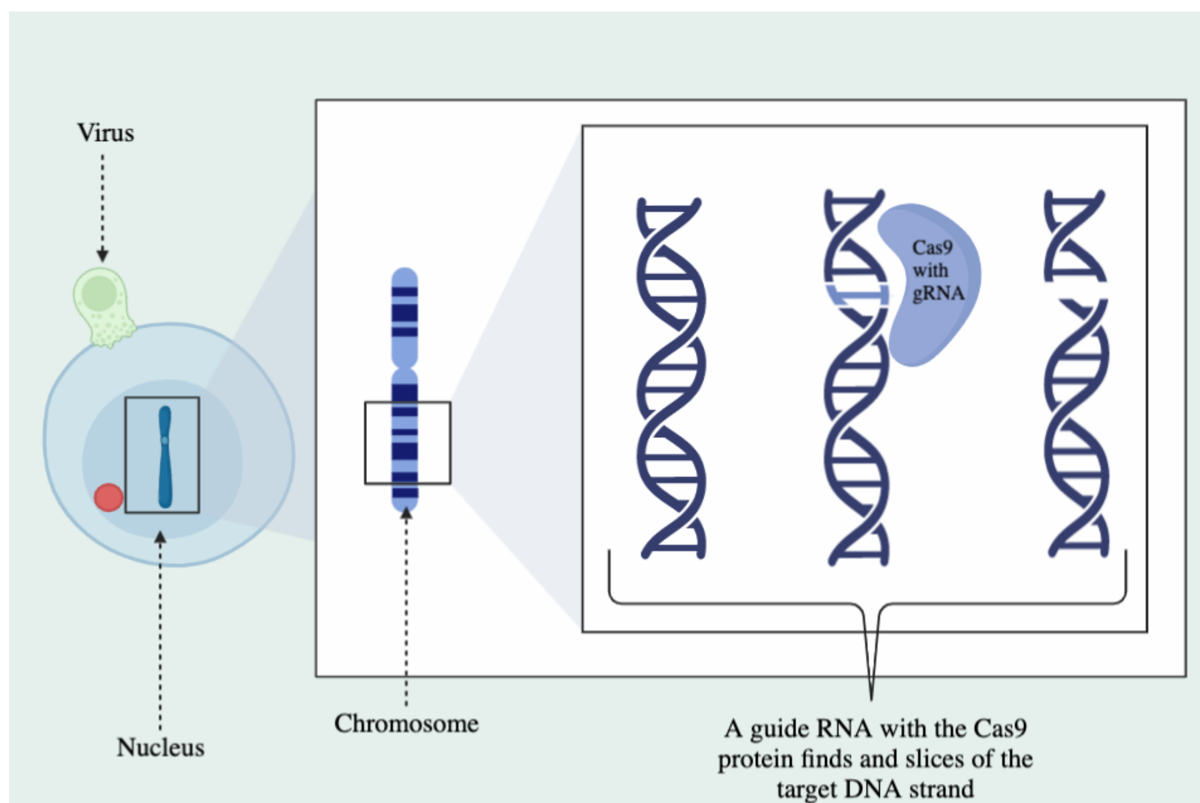

Figure 1: The diagram depicts the CRISPR system's functionality within a cell after a virus has been attached. The Cas9 protein is responsible for recording the viral DNA, locating it, and cleaving it from the strand. The first strand in the image represents a typical double-helix DNA strand. However, with the help of guide RNA, the Cas9 protein traverses the strand to identify the targeted site in the DNA. Once the target site is located, the enzyme precisely cleaves the specific sequence. Following the cleavage, the cell has two options for repairing the cut ends - it can either join the broken ends, leading to sequence modification, or introduce new DNA to the cut site, restoring the original sequence. This technology in the cell was later harnessed and replicated with human control, opening a vast ocean of opportunities for genetic engineering.

Genetic Engineering Technique

Researchers have found that this complex is programmable, allowing it to recognize specific DNA sequences and cleave DNA at those sites, thereby enabling the harnessing of this system for genetic engineering purposes. In instances of DNA breakage, cells can either connect the broken ends with sequence modifications or repair the break by incorporating a new DNA piece into the cut site. One of the methods is known as non-homologous end joining (NHEJ), which is a process that involves directly ligating the two broken ends of DNA together. NHEJ is often error-prone and can lead to mutations at the site of the break. Non-homologous end joining is a DNA repair pathway that repairs double-strand breaks in DNA. It is the predominant pathway used by mammalian cells to repair DNA double-strand breaks. The NHEJ pathway functions by directly joining two broken ends of DNA together without using a homologous template to guide repair, making it a relatively quick and efficient method of repair. NHEJ can result in small insertions or deletions at the site of the break due to imprecise repair, which can lead to mutations or changes in the DNA sequence. This can be useful for creating targeted mutations in genes for genetic engineering purposes, such as knockout or knock-in experiments. The other method is Homology-directed repair (HDR), which is a more precise repair mechanism that involves using a template DNA sequence to guide the repair process. HDR is often used with CRISPR-Cas9 to introduce specific genetic changes into the DNA sequence. By providing a template DNA sequence with the desired changes, researchers can guide the repair process to incorporate those changes at the site of the break. This has revolutionized the field of genetic engineering and has enabled researchers to study the function of genes and their role in disease in a much more targeted and precise manner. This capability of being able to make genetic changes can enable the correction of various mutations responsible for diseases such as Cystic Fibrosis, Sickle Cell Anemia, and more.

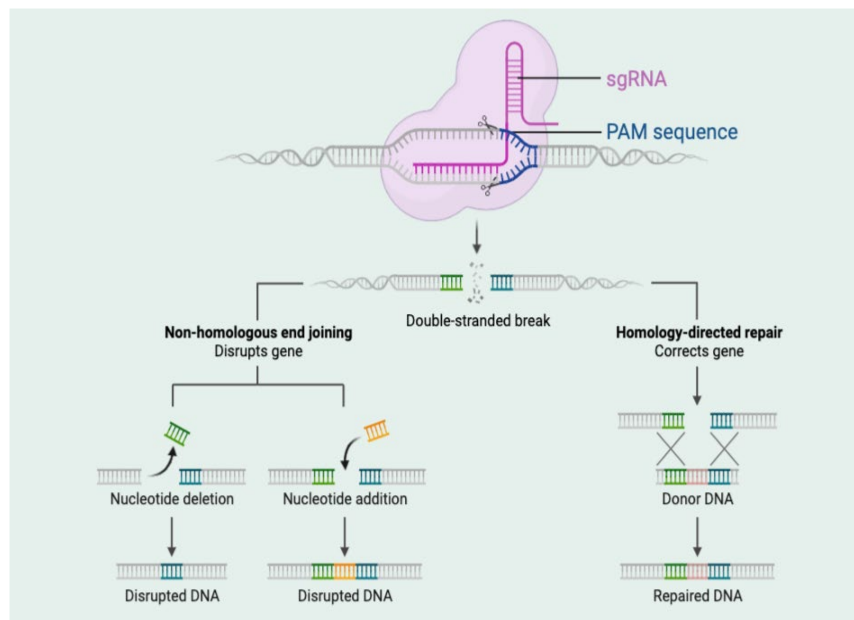

Figure 2: This diagram depicts the process of breaking the double-stranded DNA and the two types of processes that can be done after the break has been made using CRIPSR-Cas9. The first process is Non-Homologous End Joining(NHEJ), which is an error-prone process that can introduce small deletions or insertions at the break site. This can result in gene disruptions or knockouts. NHEJ is the preferred pathway when the goal is to knock out a gene. The second process shown above, Homology-Directed Repair(HDR), is a precise repair mechanism that uses a template DNA sequence, also known as the donor DNA, to repair the break. HDR can introduce specific changes or modifications at the break site.

Challenges of CRISPR-Cas9

Despite the great potential for this gene engineering technique, there are multiple concerns that need to be addressed, such as off-target effects, PAM dependence, gRNA production precision, and more.

Off-Targets:

Off-target effects occur when the CRISPR-Cas9 system inadvertently targets and modifies unintended genes. This can result in genetic changes that can have detrimental effects on an organism, including changes in gene expression, disrupting cellular function, and even the development of tumors. Several factors can contribute to off-target effects. One of the primary factors is the nature of the CRISPR-Cas9 system itself. The Cas9 protein, which is responsible for cleaving the DNA at the target site, can sometimes recognize and bind to regions of DNA that are similar to the target sequence but not an exact match. This can lead to unintended cleavage and modification of other genes. Another contributing factor is the design of the guide RNA (gRNA), which is responsible for directing the Cas9 enzyme to the target site. While gRNA design has improved significantly over the years, it is still possible for gRNAs to have off-target effects if they are not designed optimally. For example, gRNAs that are too long or contain certain nucleotide sequences can increase the likelihood of off-target effects. Off-target effects can also be influenced by the type of cells being targeted. Off-target mutations may “lead to genomic instability and disturb the normal gene function”(Lin et. al, 2018). Different cell types have unique chromatin structures and gene expression profiles, which can affect the accessibility of the target site and the specificity of the CRISPR-Cas9 system. Despite these challenges, researchers have made significant progress in minimizing the off-target effects of CRISPR-Cas9. One approach is to use bioinformatics tools to predict potential off-target sites and avoid designing gRNAs that target these sites. Researchers can also use modified versions of the Cas9 enzyme that are less likely to have off-target effects, such as the high-fidelity Cas9 variant. While progress has been made in reducing off-target effects, there is still much that researchers do not understand about the mechanisms that contribute to these effects. Further research is needed to better understand how the CRISPR-Cas9 system interacts with DNA and to develop more effective methods for predicting and minimizing off-target effects. One promising avenue of research is the use of machine learning algorithms to predict off-target effects. These algorithms can be trained on previous experiments, along with other genomic data, and then be used to predict and identify any potential off-target effects. This approach has shown promise in predicting off-target effects in a wide range of cell types and organisms.

AI-Assisted CRISPR-Cas9 Gene Editing

To overcome the challenges of the CRISPR-Cas9 system and improve its accuracy, researchers have begun exploring the application of Artificial Intelligence algorithms. AI-assisted CRISPR-Cas9 gene editing involves the use of Machine Learnings models, which are trained on large amounts of genomic data to possibly identify the most effective guide RNA for a given target, predict the off-target effects and increase the accuracy of gene editing.

Methods

Method 1:

Standard Convolutional Neural Network model

In the 2018 study by Jiencong Lin and Ka-Chun Wong, the standard CNN architecture applies convolutional filters of various sizes to the input sequence, followed by batch normalization, global max-pooling, and fully connected dense layers. This design enables the network to learn and capture important features in the sgRNA-DNA sequence.

To apply a neural network, numerical data is required. To achieve this, the study encodes each of the 4 bases of the sgRNA and target DNA into one hot vector:

A: [1,0,0,0]
G:[0,1,0,0]
C:[0,0,1,0]
T:[0,0,0,1]

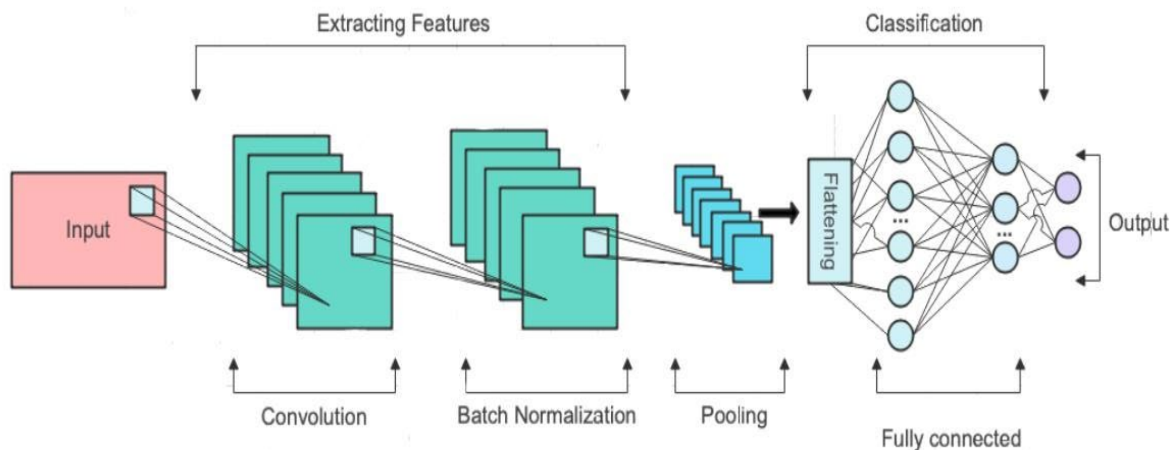

Figure 3: The diagram above is a visual representation of the standard convolution neural network used in this study. It displays the specific layers and what their main purpose is: For Extracting Features, the convolutional and batch normalization layers are used; For Classification, two fully connected dense layers are used. This results in a 2-neuron result for the prediction of the off-target.
Created and copyrighted by Parsh Verma.

Convolutional Layer:
The first layer of the model was a convolutional layer "designed for extracting sgRNA-DNA"(Lin et al., 2018). This layer applied 40 different filters, the filters in the layer scan over the one hot vectors(input data) looking for patterns that may be relevant for the task, which in this case is extracting matching sgRNA- target DNA. This layer utilizes filters of multiple sizes which allows for filtration at different levels of specificity. This layer "gives a 1 x 23 x 40 feature map"(Lin et al., 2018). This feature map contains matching information of the encoded sgRNA-DNA sequence, detected by the first layer's 40 filters. The output of this can be inputted into subsequent layers in the network.

Batch Normalization Layer:
The second layer is a Batch Normalization Layer, "designed for reducing internal covariate shift in the neural network"(Sergey et al., 2015). Internal Covariate Shift refers to the phenomenon where the distribution of the inputs of the layer of neutral network changes during training, which makes the learning of the network less accurate. The BN layer is responsible for normalizing the inputs by subtracting the mean and dividing the standard deviation of the batch, which reduces the shift. The BN layer also prevents smaller changes to some parameters, allowing for more efficient learning. This is a crucial layer for genetic analysis as genetic data often have high dimensionality, which makes generalizing the model for new data.

Global Max-Pooling Layer:
The third layer is a global max-pooling layer that takes the output of the BN layer and reduces the dimensionality of the feature map. This layer performs a max-pooling operation on the feature maps from the previous layer. Max-Pooling is used to reduce the dimensionality of the feature maps while retaining the important information of genetic patterns and information. In the context of genomics, global max-pooling can be used to identify the most significant genomic features, such as binding sites, motifs, or mutations, and capture their presence in the input sequence.

Dense Layers:
In the 2018 study by Jiencong Lin and Ka-Chun Wong, they included two layers at the end that is "fully connected dense layers with the sizes of 100 and 23"(Lin and Wong, 2018). The first dense layer has 100 neurons, and the second has 23 neurons. These dense layers allow the network to learn more complex genetic features of the sequences from the feature map extracted from the previous layer. To prevent overfitting a dropout layer was applied to the last dense layer(Lin and Wong, 2018). The dropout layer randomly masks some of the output values, which forces the network to learn more robust features. This ensures that the network does not become too reliant on specific genetic features that may not be representative of the larger population.

Output Layer:
The output layer consists of 2 neurons corresponding to the two classification results. The two neurons go through the softmax function, which converts the output of each neuron into a probability. The two neurons represent the probability that the input sequence is off-target or on-target.

Method 2:

"AttnToMismatch_CNN" Model for Off-Target Classification

To improve off-target prediction accuracy, Qiao Liu, Di He, Lei, and Lei Xie conducted a study to develop a deep neural network called AttnToMismatch_CNN, which is comprised of four components: an embedding layer, a transformer layer, a convolutional neural network layer, and a fully connected layer. The embedding layer encodes sgRNA and DNA sequences into vector representations, which are concatenated together to form a matrix. The transformer layer, which has shown superior performance in sequential analysis, has an encoder and decoder parts that both have multiple multi-head scaled dot product-based attention modules. The output of the transformer layer flows into a convolutional neural network layer, which includes two Conv2d layers and two Maxpooling layers. Finally, the output from the convolutional neural network layer is flattened and flows into a fully connected layer with a softmax function that predicts the probability of a sgRNA being a positive or negative sample. The AttnToMismatch_CNN method was evaluated using data from a CRISPR-Cas9 library screen and was found to have significantly higher prediction accuracy compared to other existing off-target prediction tools. This method has the potential to improve the safety and efficacy of CRISPR-Cas9 gene editing for future gene therapy applications.

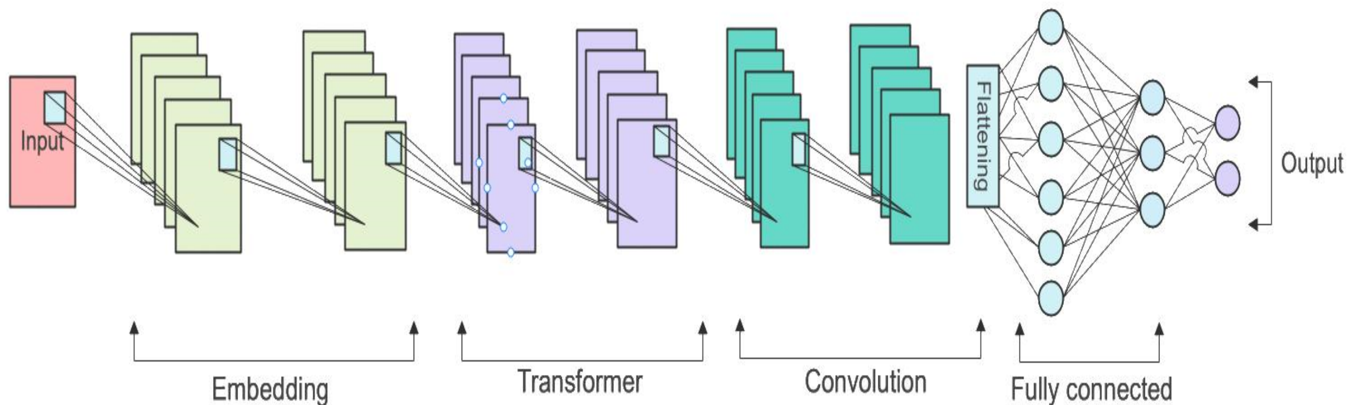

Figure 3: The diagram above is a visual representation of the AttnToMismatch_CNN model used in this study. It displays the specific layers and the flow of the data through the model. Contrastingly, this model utilizes embedding and transformer layer which are responsible for the vectorization of data and establishment of relationships between the data. Then that data is passed over to the Convolutional layer, similar to Method 1, and then a fully connected dense layer is applied to output the result.
Created and copyrighted by Parsh Verma.

Embedding Layer:
The embedding layer encodes each base from sgRNA and its corresponding base in DNA into a vector representation. "Then all positional vector representations are concatenated together to output a matrix for the aligned sequence."(Liu et al., 2019). In addition, the embedding layer also includes a positional embedding component, which encodes the position of each base pair into a vector representation. This positional information is important for the subsequent layers to capture the contextual relationship between bases in the sequences. By combining the base-pair and positional matrices using elementwise addition, the embedding layer ensures that the base pairs at different positions are encoded into distinct and informative vector representations. This embedding method enables the model to capture the subtle differences and variations in the sgRNA and DNA sequences, which is crucial for accurately predicting off-target sites.

Transformer Layer:
The Transformer layer is a deep learning component that uses a self-attention mechanism to learn contextual relationships between sequential data. It has been shown to have superior performance in sequential analysis, which fits this problem of off-target specificity prediction, as it requires capturing the contextual relationships between the sgRNA and DNA sequences. The Transformer layer in the AttnToMismatch_CNN model has both an encoder part and a decoder part, each with "multiple multi-head scaled dot product-based attention modules" (Liu et al.,2019), allowing it to effectively encode the input sequence and decode the output sequence for accurate prediction.

Convolutional Neural Network Layer:
The purpose of the Convolutional Neural Network (CNN) layer in this context is to capture local features and patterns in the encoded sequence matrix output from the Transformer layer. This is achieved through a series of two Conv2d layers and two Maxpooling layers interleaved with each other, as described by Liu et al. (2019): "The CNN comprises two Conv2d layers and two Maxpooling layers interleaved with each other, which captures the local features and patterns in the embedded sequence matrix." The output of the CNN layer is then flattened and fed into the fully

connected layer for classification. By incorporating a CNN layer in the AttnToMismatch_CNN model, it allows for capturing more complex patterns in the data, which can lead to improved classification accuracy for off-target prediction in CRISPR-Cas gene editing.

Fully Connected Layer:
The output from the CNN layer is flattened and then passed to a fully connected layer, which predicts the probability of a sgRNA to be a positive or negative sample using a softmax function: "The output from CNN is flattened and flows into the fully connected layer, which includes a softmax function to predict the probability of a sgRNA to be positive samples or negative samples."(Liu et al., 2019). This is a common approach in classification tasks, where the final output is a probability distribution over the possible classes. In the context of this study, the fully connected layer serves as the final decision layer that maps the learned features from the previous layers to the prediction task of classifying sgRNA sequences as either positive or negative samples. The softmax function is used to produce a probability distribution over the possible classes, which facilitates decision-making for downstream applications such as CRISPR-Cas gene editing.

Results/Comparison

The results of the study showcased the performance of both the Convolutional Neural Network (CNN) and the AttnToMismatch_CNN methods in predicting off-target sequences.

The Convolutional Neural Network based model demonstrated a significant ability to predict off-target sequences effectively. With its multi-layered approach, this model was able to analyze and classify sequences in an organized manner. The first layer, the convolutional layer, was effective in filtering out the significant genetic patterns, providing a 1 x 23 x 40 feature map. The subsequent batch normalization layer reduced the internal covariate shift, ensuring accurate learning of the network. The global max-pooling layer and dense layers effectively reduced dimensionality while ensuring important genetic features are recognized. The dropout layer reduced overfitting, while the output layer offered a clear classification.

The AttnToMismatch_CNN model, on the other hand, provided promising results with higher prediction accuracy than most models. It effectively encoded sgRNA and DNA sequences into vector representations and contextualized relationships between bases in the sequences through the transformer layer. The CNN layer in this model captured local features and patterns in the gene sequence matrix, while the fully connected layer effectively predicted the sgRNA sequences as either positive or negative samples.

The Convolutional Neural Network method outperformed the AttnToMismatch_CNN model in terms of prediction accuracy, demonstrating the potential for better off-target prediction. This conclusion can be reached by looking at the Area Under the Curve(AUC), in this case, the Reciever Operating Characteristic Curve(AUC-ROC), score to determine the predictiveness of each model.

As presented in the graph the Standard CNN, explained in Method 1, outperforms the other models. However, it should be noted that both methods have their strengths and can be chosen depending on the specific requirements and constraints of the task at hand. A major factor that may be affecting these techniques is the data that these models were trained and tested on. The Standard CNN(Method 1) was trained on the CRISPOR dataset and tested on the GUIDE-seq dataset. While the AttnToMismatch_CNN is trained and tested “with Crispr-Cas9 dataset in A) K562 cell line B) A549 cell line and C) NB4 cell line”(Liu et al., 2019).

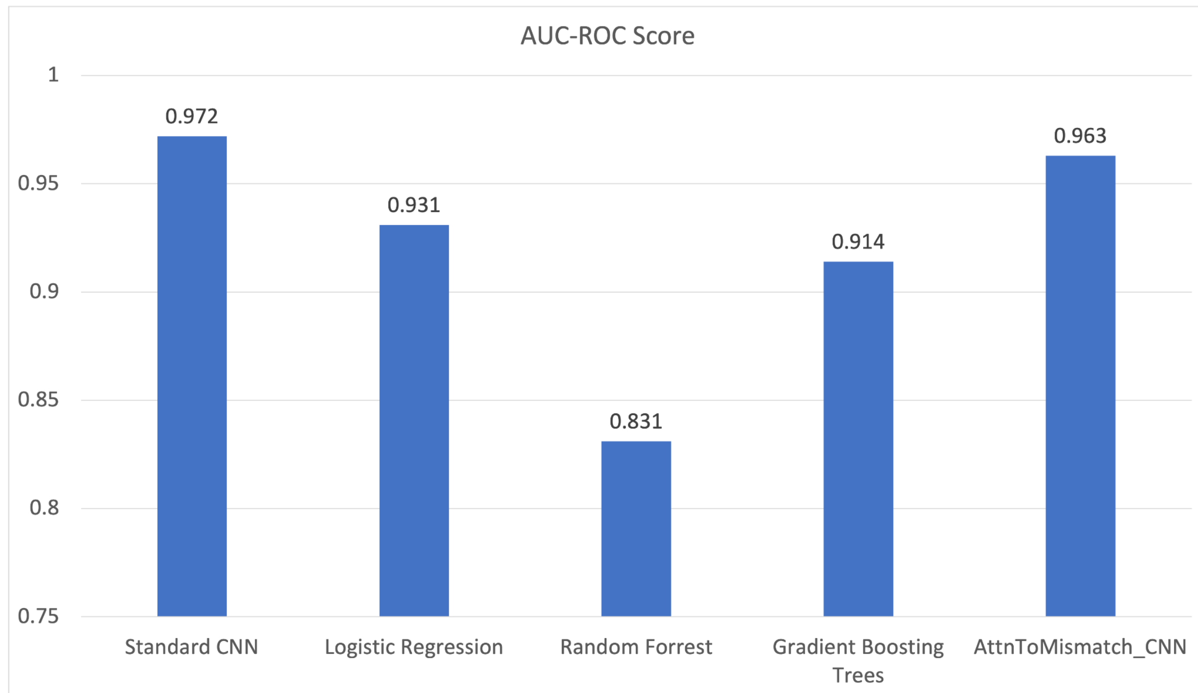

Figure 3: According to Lin et al. 2018, the scores achieved by Standard CNN: 0.972; Logistic Regression: 0.931; Random Forrest: .831; Gradient Boosting Trees: 0.914. The AttnToMismatch_CNN score can be calculated using the average of the two scores given in the research conducted by Liu et al. in 2019. The average of those two scores results in a score of 0.963.

Conclusion

CRISPR-Cas9 gene editing holds great promise for precise genetic modifications, but it also presents challenges, including off-target effects. Researchers have turned to artificial intelligence to enhance the accuracy and predictability of the CRISPR-Cas9 system. AI-based off-target prediction methods, particularly Convolutional Neural Networks, have shown significant potential in improving the safety and efficacy of gene editing.

The standard CNN-based model effectively predicts off-target sequences by filtering genetic patterns and accurately classifying sequences. It leverages its multi-layered architecture, including convolutional, batch normalization, global max-pooling, and dense layers, to reduce dimensionality and improve the recognition of important genetic features. Additionally, the AttnToMismatch_CNN model encodes sgRNA and DNA sequences into vector representations, contextualizing relationships between bases through the transformer layer, and making precise predictions with a fully connected layer. The main difference between the architectures is the method of data vectorization: The standard CNN had 4 different encoded values for the 4 bases (A, C, T, G), while the AttnToMismatch_CNN used an embedding layer to vectorize the data. While the standard CNN-based model outperforms the AttnToMismatch_CNN model in terms of prediction accuracy, both methods have their strengths and can be chosen based on specific requirements for the most effective data vectorization. Factors such as the training and testing datasets and the constraints of the task at hand should also be considered when selecting the appropriate method.

The integration of CNN models for off-target prediction methods enhances the safety and efficacy of CRISPR-Cas9 gene editing. Further research is needed to refine and optimize these AI-based methods, ensuring the precise and safe application of this technology, as it can be extremely dangerous if not applied precisely. Ethical considerations and thorough assessment of risks associated with AI-assisted gene editing are crucial for the responsible and safe utilization of this technology. The integration of AI algorithms in gene editing raises concerns regarding

accuracy, biases, and transparency. By addressing these considerations and establishing comprehensive ethical guidelines and robust regulatory frameworks, we can navigate the complexities of gene editing responsibly.

In summary, the integration of CRISPR-Cas9 gene editing with AI-based off-target prediction methods represents a transformative frontier in genetic engineering. By effectively addressing the challenges associated with off-target effects, researchers can unlock the full potential of this technology, advancing precision medicine and offering new possibilities for the treatment of genetic disorders and other diseases. The comparative analysis of CNN-based models showcases the significant potential of Convolutional Neural Networks in accurately predicting off-target sequences and improving the precision of gene editing procedures. This advancement not only contributes to the field of genetic engineering but also holds great promise for enhancing human health outcomes and revolutionizing the therapeutic landscape.

References

Almagro Armenteros, J. J., Sønderby, C. K., Sønderby, S. K., Nielsen, H., & Winther, O. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics*, *33*(21), 3387–3395. https://doi.org/10.1093/bioinformatics/btx431

Kang, S.-H., Lee, W., An, J.-H., Lee, J.-H., Kim, Y.-H., Kim, H., Oh, Y., Park, Y.-H., Jin, Y. B., Jun, B.-H., Hur, J. K., Kim, S.-U., & Lee, S. H. (2021). Author Correction: Prediction-based highly sensitive CRISPR off-target validation using target-specific DNA enrichment. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-020-20559-5

Lin, J., & Wong, K.-C. (2018). Off-target predictions in CRISPR-Cas9 gene editing using deep learning. *Bioinformatics*, *34*(17), i656–i663. https://doi.org/10.1093/bioinformatics/bty554

Lino, C. A., Harper, J. C., Carney, J. P., & Timlin, J. A. (2018). Delivering CRISPR: a review of the challenges and approaches. *Drug Delivery*, *25*(1), 1234–1257. https://doi.org/10.1080/10717544.2018.1474964

Liu, Q., Cheng, X., Liu, G., Li, B., & Liu, X. (2020). Deep learning improves the ability of sgRNA off-target propensity prediction. *BMC bioinformatics*, *21*(1), 51. https://doi.org/10.1186/s12859-020-3395-z

Liu, Q., He, D., & Xie, L. (2019). Prediction of off-target specificity and cell-specific fitness of CRISPR-Cas System using attention boosted deep learning and network-based gene feature. PLoS computational biology, 15(10), e1007480. https://doi.org/10.1371/journal.pcbi.1007480

Ma, Y., Zhang, L., & Huang, X. (2014). Genome modification by CRISPR-Cas9. *FEBS Journal*, *281*(23), 5186–5193. https://doi.org/10.1111/febs.13110

Manghwar, H., Lindsey, K., Zhang, X., & Jin, S. (2019). CRISPR-Cas System: Recent Advances and Future Prospects for Genome Editing. *Trends in Plant Science*, *24*(12), 1102–1125. https://doi.org/10.1016/j.tplants.2019.09.006

Ran, F. Ann, Hsu, Patrick D., Lin, C.-Y., Gootenberg, Jonathan S., Konermann, S., Trevino, A. E., Scott, David A., Inoue, A., Matoba, S., Zhang, Y., & Zhang, F. (2013). Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell*, *154*(6), 1380–1389. https://doi.org/10.1016/j.cell.2013.08.021

Redman, M., King, A., Watson, C., & King, D. (2016). What is CRISPR-Cas9? *Archives of Disease in Childhood - Education & Practice Edition*, *101*(4), 213–215. https://doi.org/10.1136/archdischild-2016-310459

Zhang, F., Wen, Y., & Guo, X. (2014). CRISPR-Cas9 for genome editing: progress, implications and challenges. *Human Molecular Genetics*, *23*(R1), R40–R46. https://doi.org/10.1093/hmg/ddu125

Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S., & Yang, S.-H. (2015). Off-target Effects in CRISPR-Cas9-mediated Genome Engineering. *Molecular Therapy - Nucleic Acids*, *4*(1), e264. https://doi.org/10.1038/mtna.2015.37

Zhu, L. J. (2015). Overview of guide RNA design tools for CRISPR-Cas9 genome editing technology. *Frontiers in Biology*, *10*(4), 289–296. https://doi.org/10.1007/s11515-015-1366-y