Master's Thesis Information Showcase
Spring Semester-2026
Year 2026

Department: Computer Science & Information Technology
Programme: Masters of Science
Specialization: Computer Science & Information Technology

1	**Title of the Thesis** AI-Driven Off-Target Prediction in Gene Editing
2	**Abstract** (min 300 to 500 words) Gene editing has provided a major technical breakthrough in medicine and agriculture but there is a critical safety limitation hindering its growth. The CRISPR-Cas9 gene editing tool,has an off-target problem. It does not always cut where it is directed to. Its guide mechanism tolerates minor sequence mismatches. This means the Cas9 protein can bind and sever DNA at unintended genomic locations that just resemble the true target. These off-target mutations have critical biological consequences. These can include silencing essential genes to triggering large-scale chromosomal damage. Detecting every possible erroneous cut through laboratory analysis alone is extremely expensive and time consuming. This thesis addresses that exact bottleneck by developing a hybrid deep learning model specifically designed to predict off-target cleavage sites computationally. The proposed architecture combines 3D one-hot encoded sequence representations with a small set of biologically grounded numerical features. This enables the system to flag high-risk genomic locations before any physical experiment is carried out. Biological datasets normally have severe classification imbalances. To counter this statistical skew, the proposed model uses a 5-model deep ensemble architecture. Dual-input 1D-Convolutional Neural Networks operate alongside parallel Dense branches. The entire pipeline is optimized by a Binary Focal Loss function. Empirical evaluation on an unseen hold-out dataset yielded a Precision-Recall Area Under the Curve (PRAUC) of 0.7380. The ensemble achieved an overall precision of 86.67%. These metrics show that the model makes confident predictions and actually understands the data. It was observed that combining computationally inexpensive metrics improves performance. Thermodynamic stability indicators like GC content alongside count of mismatches within the sensitive PAM-proximal seed region were used. These metrics combined with the 3D one-hot encoded sequences create a very precise model. The system requires fewer than 600,000 total ensemble parameters. With a very small computational cost, the design is a compromise between very resource-intensive physical simulations and sequence based deep learning approaches. The resulting model is very precise and scalable for predicting CRISPR-Cas9 off-targets.
3	**Impact on SDG# 3 Good Health and Well-being"** (min 500 words) Sustainable Development Goal 3 focuses on ensuring healthy lives and promoting well-being for all. Gene editing is a major frontier for this objective. It holds the massive promise of fixing inherited diseases. But there is a serious roadblock. The primary tool we use, known as CRISPR, is highly effective but prone to making biological mistakes. If we want

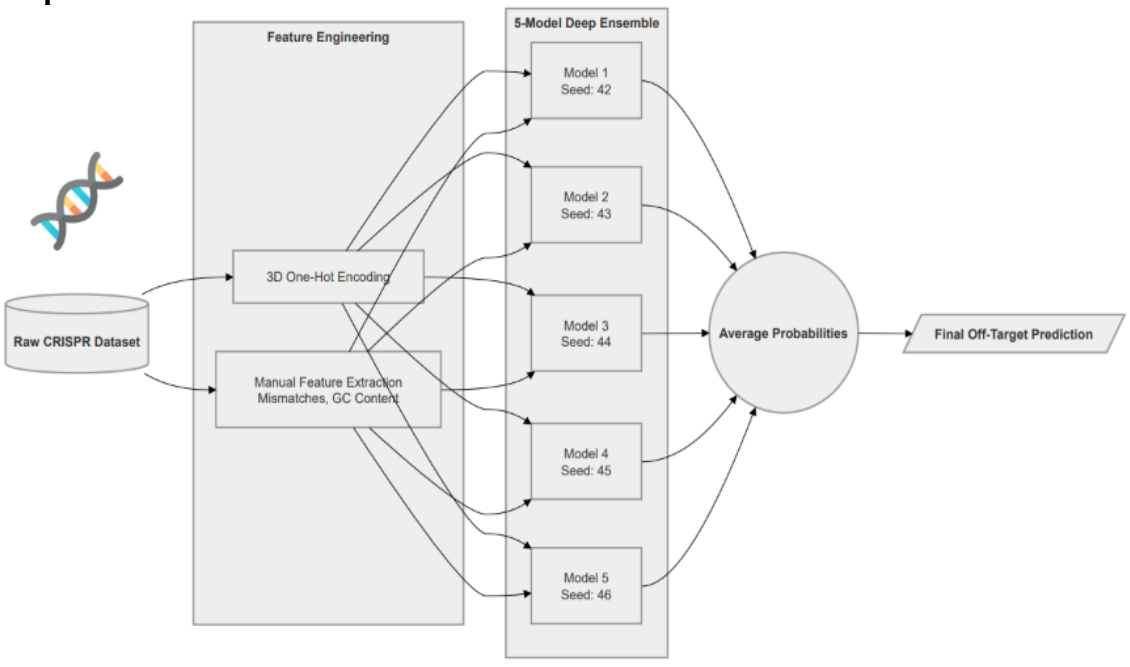
to cure diseases without accidentally causing new ones, we have to solve this underlying safety problem. This thesis tackles that exact problem, bringing us a vital step closer to making genetic medicine a safe reality for patients worldwide.

DNA can be as a considered a massive molecular instruction manual with billions of letters that defines how the human body functions. The CRISPR tool acts as a pair of programmable molecular scissors. Scientists use it to cut problematic parts of DNA. The issue is that the CRISPR method can cut wrong parts of the DNA that look similar to the actual target. In medical treatment these accidental cuts are catastrophic. They can trigger massive chromosomal damage or even turn off the specific genes that protect the human body from cancer. Finding every single possible accidental cut by running physical experiments in a lab is incredibly slow and very expensive. This thesis builds a permanent bridge over that bottleneck by using artificial intelligence.

Instead of relying entirely on trial and error in physical test tubes, this research introduces a specialized artificial intelligence model. It acts as an advanced digital warning system. By feeding the AI model information about the physical properties of the DNA structure, the model learns to spot the danger zones. It predicts exactly where the wrong cuts can occur. To improve the prediction accuracy, architecture uses five differently configured models working simultaneously. This approach ensures the predictions are exceptionally reliable. The system catches nearly 87 percent of these dangerous locations before a single physical experiment even begins.

This technological leap directly supports global health equity. A major part of the Sustainable Development Goals is making sure lifesaving technology is shared globally. Physical lab testing requires millions of dollars in continuous funding. This creates a massive financial barrier to entry for new research. The AI model developed is very cost-effective as it requires very little computing power to run. This democratizes the field of gene editing. An institution in a developing nation with limited resources can use AI based prediction to screen potential genetic treatments just as effectively as a massive pharmaceutical company.

Before any new genetic treatment can be given to a human patient, regulatory agencies demand absolute proof that it will not cause unintended harm. AI based methods provide a computational safety check in addition to existing physical testing methods. It gives regulators the hard data they need to approve clinical trials with confidence, significantly speeding up the timeline from laboratory discovery to medical treatment.

The flexibility of a computational model also amplifies its global impact. While the main target of this thesis is SDG#3 human health, the exact same AI pipeline can be used to improve global food security. This connects directly to SDG#2, which aims to end hunger. Agricultural scientists use gene editing to create crops that can survive severe climate change and resist local diseases. Just like in humans, agricultural modifications need to be completely accurate to avoid ruining crop yields or nutritional value. By making the design process safer and cheaper across the board, this technology accelerates progress in multiple areas of human survival.

	We cannot achieve universal good health by relying exclusively on manual laboratory testing techniques. The transition from physical trial and error to intelligent digital prediction is mandatory for the future of science. This thesis delivers a scalable, highly accurate software tool that directly neutralizes a major safety hazard in modern genetic engineering.
4	**Scholar Name** (along with contact details) Name: Humza Ahmed Email: humza.ahmed235@gmail.com Phone: +92 3111649254
5	**Supervisor & Co-supervisor Name** (along with contact details) Name: Prof. Dr. Shariq Mahmood Khan Email: shariq@neduet.edu.pk
6	**Graphical Abstract**