***Intro***

In recent years huge DNA, RNA and protein sequences have been available, for efficiently analyzing them, it is required to convert the symbolic sequence to a numerical sequence which is essential to genome comparison, compression and encryption

By merging the idea of chaotic dynamics with the sequence bases (4 letters) of DNA or RNA,

rather than yielding in a random structure as in using only four corners (no consistent result each run), the randomness of the structure is taken out by the sequence governing, the resulted image is the genetic structure of the entered seq

this technique is known as Chaos game representation (CGR) which is a numerical mapping of sequences that is represented in a graph, the last edited version of it is ICGR (integer Chaos game representation) which allows to encode and decode the sequence into 3 integer numbers

The ICGR is one-to-one lossless (the k-th point in image corresponds to the k-th nucleotide), but is limited by not exceeding 1024 base of the sequence in order to be able to succeed in encoding and decoding

***Academic motivation***

Based on the paper published in 2018 (Encoding DNA sequences by integer chaos game representation)[2](https://arxiv.org/pdf/1712.04546v2.pdf)

As DNA sequences are fundamental for encoding genetic information and converting the sequence into numerical values is essential for comparison, compression and encryption,

the goal is to add the ability of increasing the sequence length over 1024 base, succeeding in encoding and decoding it back, and making further analysis on the resulted images using machine learning

***Impact/users***

The targeted users are researchers in a related biological field, bioinformaticians

***Scope***

The range of working project is on DNA and RNA sequences

***Algorithm***

* CGR algorithm
* Machine learning algorithms

***Benefits***

* Converting the sequence into an image
* Using machine learning for further analysis on the images
* Encoding the sequence into 3 integer numbers that can be decoded back to the right sequence
* Increasing the size of the input sequence that could be more than 1024 nucleotides