CRISPR application predictor

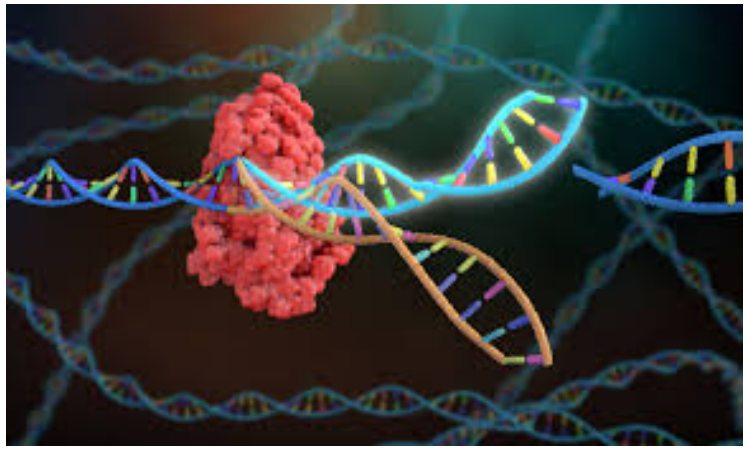
Emilia Grzesiak
DSSI Summer 2020
Mentor: Nisha Mulakken (Bioinformatics)

LLNL-PRES-812938
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Lawrence Livermore National Laboratory

Motivation

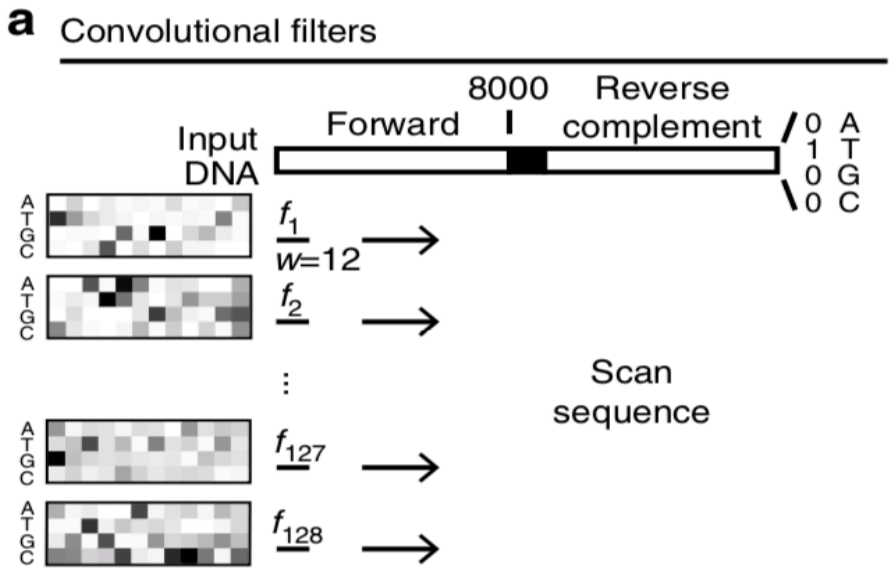
- CRISPR, a genome editing tool, has become increasingly popular and cost-effective
  - Not everyone will use this technology ethically, some may use it nefariously
- Can we identify who bioengineered DNA and what their goals were?
  - Proactive project

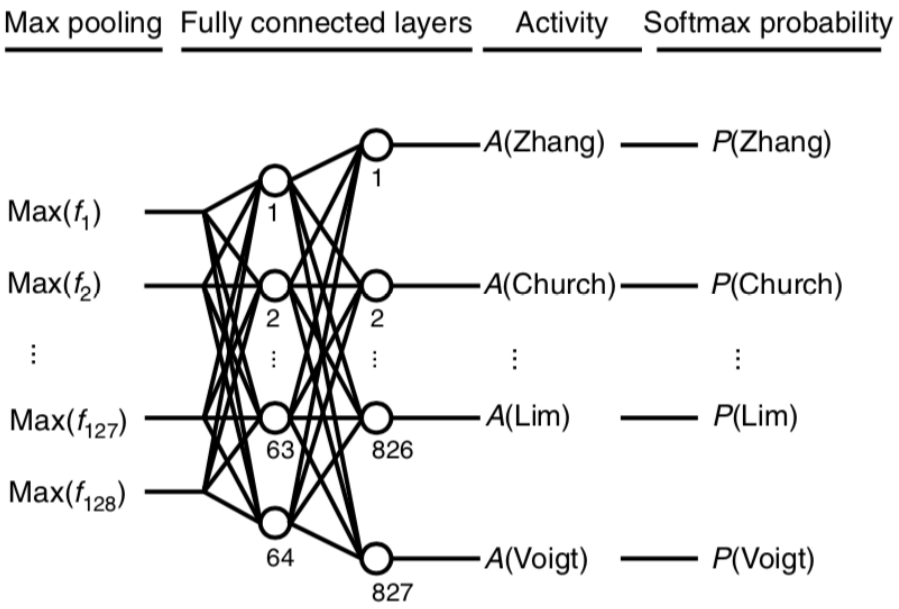


LLNL-PRES-812938

Approach

- Bioengineers make small design choices that add up as their “DNA signature”, but impossible to notice with human eye
- Database: Addgene (popular service that delivers plasmids for research purposes)
- Used 1-D Convolutional Neural Network to classify depositor per plasmid
- Model based off published journal paper [1]





[1] Nielsen, A.A.K., Voigt, C.A. Deep learning to predict the lab-of-origin of engineered DNA. Nat Commun 9, 3135 (2018). https://doi.org/10.1038/s41467-018-05378-z

Results & Conclusion

- Achieved validation and cross validation accuracy of 66%
  - Had 66,000+ plasmids and 1,400 depositors
- Original paper achieved 48% validation and cross validation accuracy
  - Had 36,764 sequences and 827 depositors
- Next steps:
  - Evaluate performance of RNN models
  - Scan sequences to determine “DNA signature” of depositor