

# Building a Healthier Tomorrow

## Designing More Effective Influenza Vaccines

Gianna August, Pushkar Chougule, Amy Femal, Carlos Palomo, Jonathan Renstrom



### Introduction

The current COVID-19 outbreak generated a curiosity in our group to know about another major disease – Influenza. From the information available about the impact of Flu, WHO estimates that around 290,000 to 650,000 deaths are caused by the Flu globally, every year. According to CDC, in United States alone, there are 12,000 to 61,000 deaths associated with Flu each year. Studying the data available for the effectiveness of existing flu vaccines, the rates varies between 20% - 60% over the last 10 years, with majority of times around 40% mark. The cell receptors of Influenza A undergo a high rate of mutation that can compromise the adaptive immune systems acquired immunity. This is one of the major factors for the low effectiveness rate of vaccines. Vaccines that target the proper viral epitopes can prevent infection. Unfortunately, Influenza vaccines are educated guesses based on prior viral strains. As a part of this study, we aim at using the power of Data Science to predict the viral mutations and in turn help improve the accuracy of Influenza vaccines.

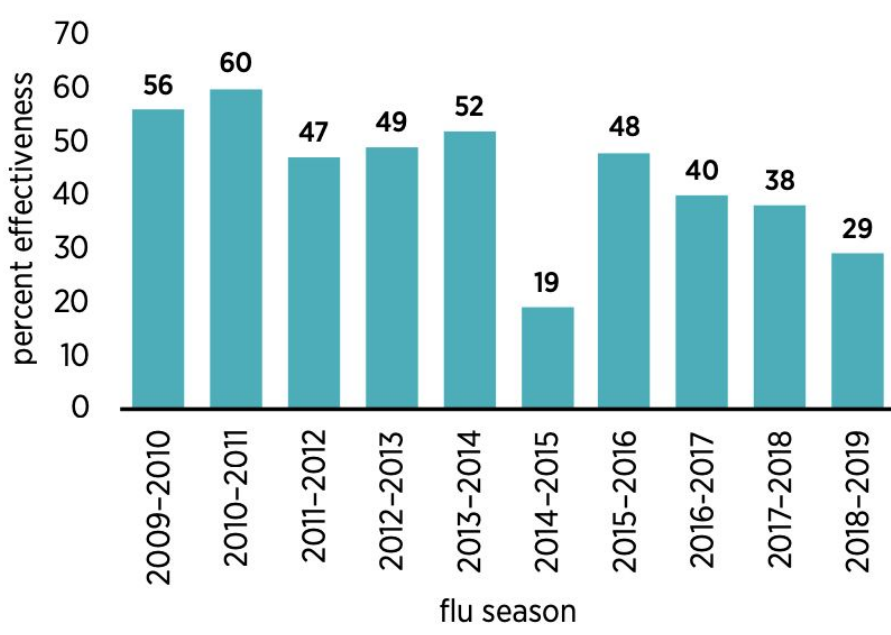
### Objectives

We will use various machine learning techniques to generalize data, perform classification analysis, and make predictions of the antigenic structure of glycoprotein hemagglutinin and antigenic variants.

### Why is this Data Science?

As a part of this research work, we have to collect Flu data for the past several years, study the trends in the mutations in the virus structures (hemagglutinin - H, neuraminidase - N) proteins through different flu cycles as well as geographic regions and variance among different age groups. To be able to handle and clean such a huge amount of data and deriving the predictions in the virus mutations or future virus strains, would certainly require a collaboration between human brain and processing powers of modern computers, to be able to create meaningful predictions, which will help make informed decisions for the vaccine creation. Hence, we think this study is good fit for Data Science.

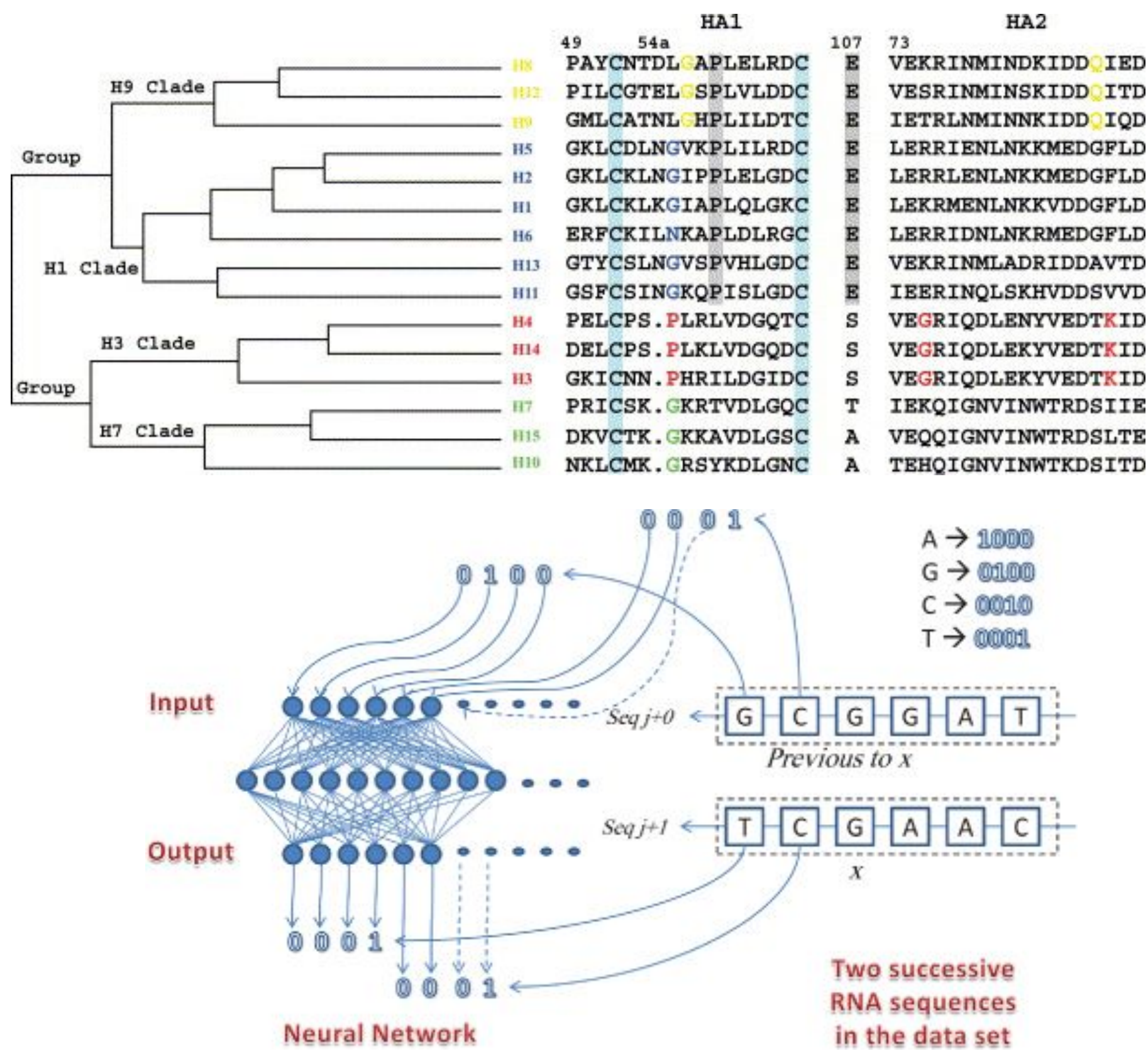
Figure 1. Seasonal Flu Vaccine Effectiveness



Source: "CDC Seasonal Flu Vaccine Effectiveness Studies." Centers for Disease Control and Prevention, last updated February 21, 2020, <https://www.cdc.gov/flu/vaccines-work/effectiveness-studies.htm>.

### Project Deliverables

- A decision tree to illustrate the closeness in the variations of the Influenza subtypes based off the classification analysis performed.
- A neural network that can predict Influenza subtypes.



### Conclusions

29% IS NOT GOOD ENOUGH

Being able to predict the upcoming influenza virus strains and preparing the appropriate vaccines would certainly help prevent the loss of human lives reflected in current numbers on an annual basis and an opportunity to create healthy future for the society using the potential power of data science.

### Acknowledgements

We would like to thank the brave health care workers that are currently risking their lives to fight off the Covid-19 virus.

### References

1. Anon. 2019. Predicting correct serotypes using machine-learning models based on codon usage patterns of influenza A viruses. (January 2019). Retrieved April 23, 2020 from <https://www.biorxiv.org/content/10.1101/528083v1.abstract>
2. Anon. 2020. CDC Seasonal Flu Vaccine Effectiveness Studies. (February 2020). Retrieved April 23, 2020 from <https://www.cdc.gov/flu/vaccines-work/effectiveness-studies.htm>
3. Anon. Classification of Host Origin in Influenza A virus by ... Retrieved April 23, 2020 from <http://www.naun.org/main/NAUN/bio/2017/a182010-049.pdf>
4. Claire Gillespie. This Is How Many People Die From the Flu Each Year. Retrieved April 23, 2020 from <https://www.health.com/condition/cold-flu-sinus/how-many-people-die-of-the-flu-every-year>
5. Edward A. Belongia et al. 2009. Effectiveness of Inactivated Influenza Vaccines Varied Substantially with Antigenic Match from the 2004–2005 Season to the 2006–2007 Season. The Journal of Infectious Diseases 199, 2 (2009), 159–167. DOI:<http://dx.doi.org/10.1086/595861>
6. Eng CLP et al. Predicting host tropism of influenza A virus proteins using random forest. BMC Medical Genomics. 2014;7:1-11. <https://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=135645438&site=eds-live>.
7. J.J. Treanor et al. 2012. Effectiveness of Seasonal Influenza Vaccines in the United States During a Season With Circulation of All Three Vaccine Strains. Clinical Infectious Diseases 55, 7 (2012), 951–959. DOI:<http://dx.doi.org/10.1093/cid/cis574>
8. Jia Wang et al. 2013. Using Amino Acid Factor Scores to Predict Avian-to-Human Transmission of Avian Influenza Viruses: A Machine Learning Study. Protein & Peptide Letters 20, 10 (January 2013), 1115–1121. DOI: <http://dx.doi.org/10.2174/0929866511320100005>
9. Li Yan et al. Prediction of criticality in patients with severe COVID-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan. DOI: <https://www.medrxiv.org/content/10.1101/2020.02.27.20028027v>
10. Lin Tang. 2013. Machine Learning Methods for Annual Influenza Vaccine Update. Masters Thesis. Waterloo University, Ontario, Canada.
11. Mostafa A. Salama et al. 2016. The prediction of virus mutation using neural networks and rough set techniques. EURASIP Journal on Bioinformatics and Systems Biology 2016, 1 (2016). DOI:<http://dx.doi.org/10.1186/s13637-016-0042-0>
12. Nermeen Shaltout et al. 2014. Information Gain as a Feature Selection Method for the Efficient Classification of Influenza Based on Viral Hosts. (2014). Retrieved April 22, 2020 from
13. Pavan Attaluri et al. 2009. Applying machine learning techniques to classify H1N1 viral strains occurring in 2009 flu pandemic. (2009).
14. Pavan Attaluri et al. Applying neural networks to classify influenza virus antigenic types and hosts. Retrieved April 23, 2020 from <https://ieeexplore.ieee.org/abstract/document/5510726>
15. Pavan Attaluri. 2010. Classifying influenza subtypes and hosts using machine learning techniques. (May 2010). Retrieved April 22, 2020 from <http://ezproxy.bellevue.edu/login?url=https://search-proquest-com.ezproxy.bellevue.edu/docview/219921882?accountid=28125>
16. Pavan K. Attaluri, Zhengxin Chen, and Guoqing Lu. 2010. Applying neural networks to classify influenza virus antigenic types and hosts. 2010 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (2010). DOI: <http://dx.doi.org/10.1109/cibcb.2010.5510726>
17. Pavan K. Attaluri, Zhengxin Chen, Aruna M. Weerakoon, and Guoqing Lu. 2009. Integrating Decision Tree and Hidden Markov Model (HMM) for Subtype Prediction of Human Influenza A Virus. (June 2009). Retrieved April 23, 2020 from [https://link.springer.com/chapter/10.1007/978-3-642-02298-2\\_8](https://link.springer.com/chapter/10.1007/978-3-642-02298-2_8)
18. R. Parida, M.s. Shaila, S. Mukherjee, N.r. Chandra, and R. Nayak. 2007. Computational analysis of proteome of H5N1 avian influenza virus to define T cell epitopes with vaccine potential. Vaccine 25, 43 (2007), 7530–7539. DOI:<http://dx.doi.org/10.1016/j.vaccine.2007.08.044>
19. Reatha Sandie and Stéphane Aris-Brosou. 2013. Predicting the Emergence of H3N2 Influenza Viruses Reveals Contrasted Modes of Evolution of HA and NA Antigens. Journal of Molecular Evolution 78, 1 (2013), 1–12. DOI:<http://dx.doi.org/10.1007/s00239-013-9608-6>
20. Rui Yin, Yu Zhang, Xinrui Zhou, and Chee Keong Kwoh. 2020. Time series computational prediction of vaccines for influenza A H3N2 with recurrent neural networks. Journal of Bioinformatics and Computational Biology (2020). DOI: <http://dx.doi.org/10.1142/s0219720020400028>
21. Shaomin Yan and Guang Wu. 2010. Prediction of Mutation Positions in H5N1 Neuraminidases From Influenza A Virus by Means of Neural Network. Annals of Biomedical Engineering 38, 3 (2010), 984–992. DOI:<http://dx.doi.org/10.1007/s10439-010-9907-7>
22. WHO. Coronavirus disease (covid-19) outbreak situation. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
23. Xiaowei Xu et al. Deep learning system to screen coronavirus disease 2019 pneumonia. DOI: <https://arxiv.org/ftp/arxiv/papers/2002/2002.09334.pdf>