

Phylogenetic dynamics and early variant identification analysis of respiratory viruses genome

PhD Candidate – Aung Pone Myint

Supervisors: Prof. Marta Nunes (Centre International de la Recherche en Infectiologie (CIRI))
Prof. Giacomo Cacciapaglia (Institut de Physique des Deux Infinis de Lyon (IP2I))

Background

	COVID-19	Seasonal Influenza	RSV infection
Virus	SARS-CoV-2	Influenza viruses (A/H1N1, A/H3N2, B/Victoria, and B/Yamagata)	Respiratory syncytial virus (RSV-A and RSV-B)
Cases	770m ¹ (confirmed, until Sept 2023) 3.4b ² (estimated, until 2021 Nov)	54.5m ³ (2017)	33m ⁴ (U5, 2019) 1.5m* ⁵ (65+, 2015)
Deaths	6m ¹ (confirmed, until Sept 2023) 15.1m ² (estimated, until 2021 Nov)	145k ³ (2017)	101k ⁴ (U5, 2019) 14k ⁵ (65+, 2015)
Mutation rate	0.81–1.08 × 10 ⁻³ substitutions/site/year ⁶	3.87-4.84 × 10 ⁻³ substitutions/site/year ⁶	10 ⁻³ to 10 ⁻⁴ substitutions/site/year ⁷

*in industrialized countries

Viral evolution

Studies

Impact

Phylogenetic tree,
antigenic cartograph,
modelling

Machine learning

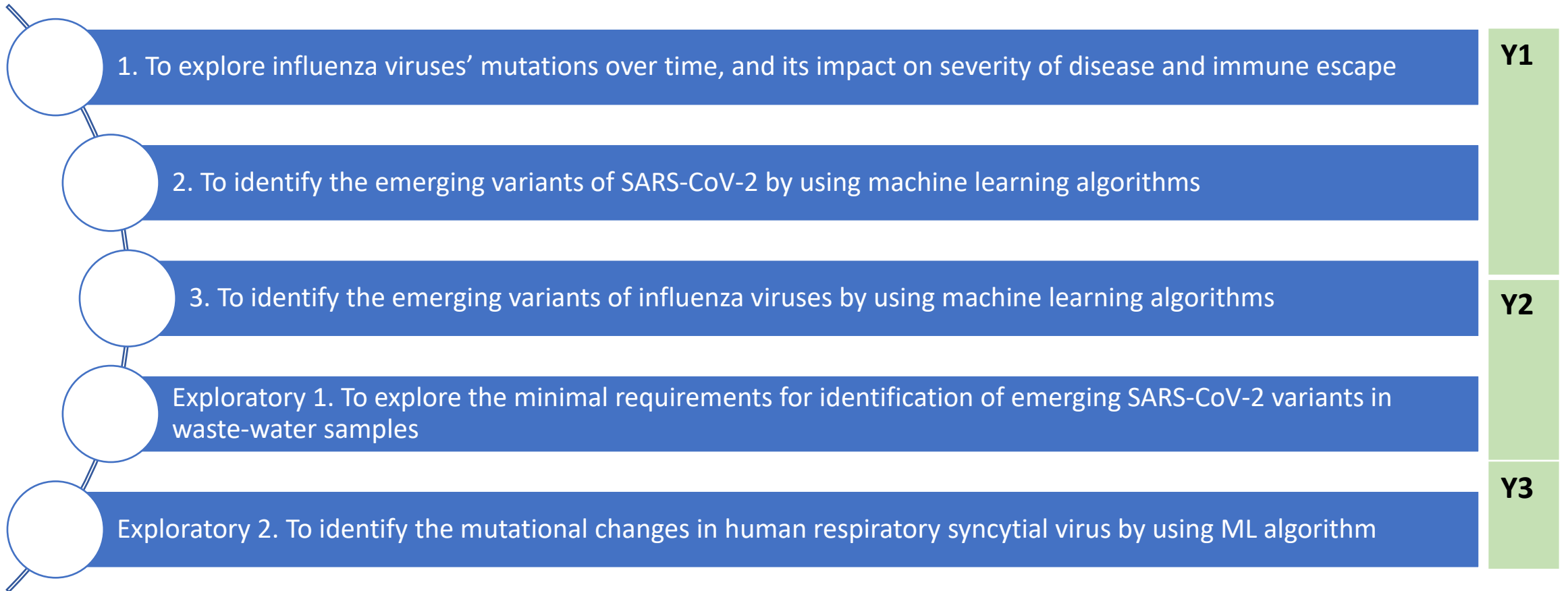
Immune escape

Severity of disease

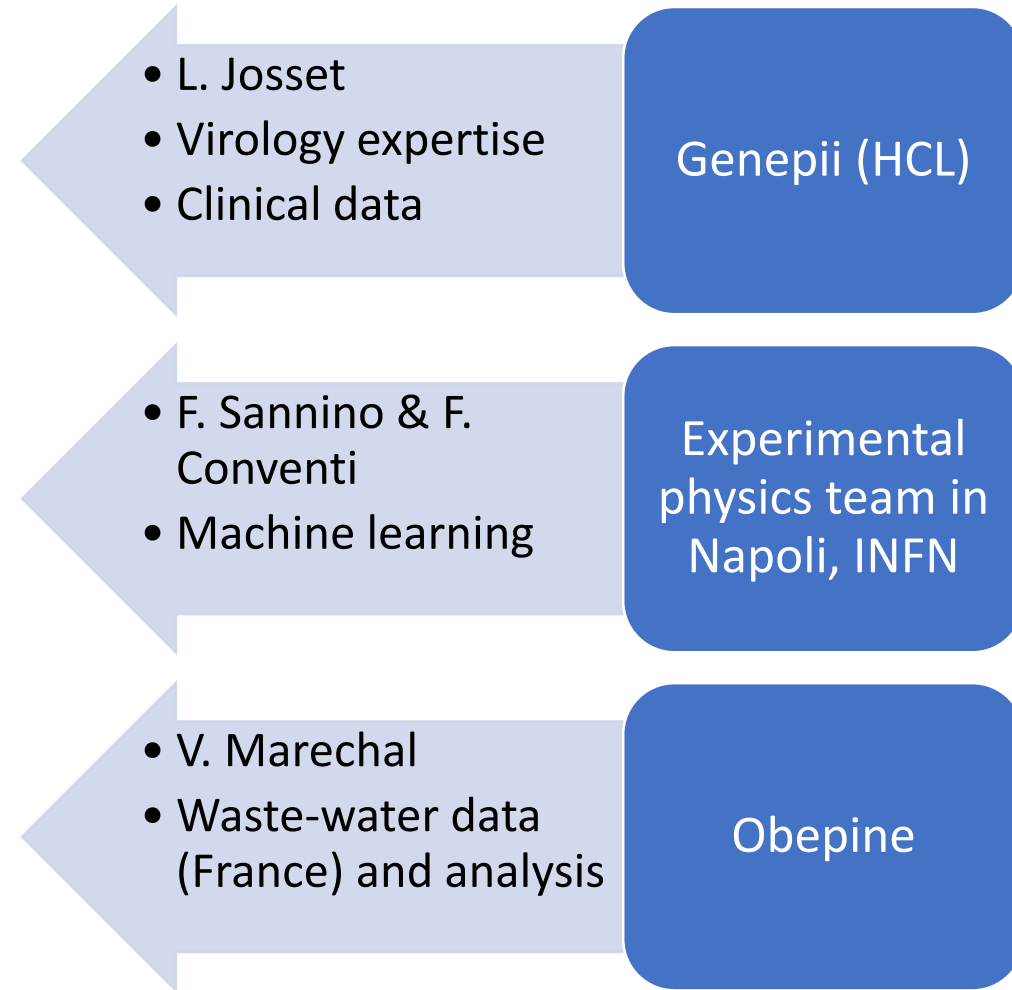
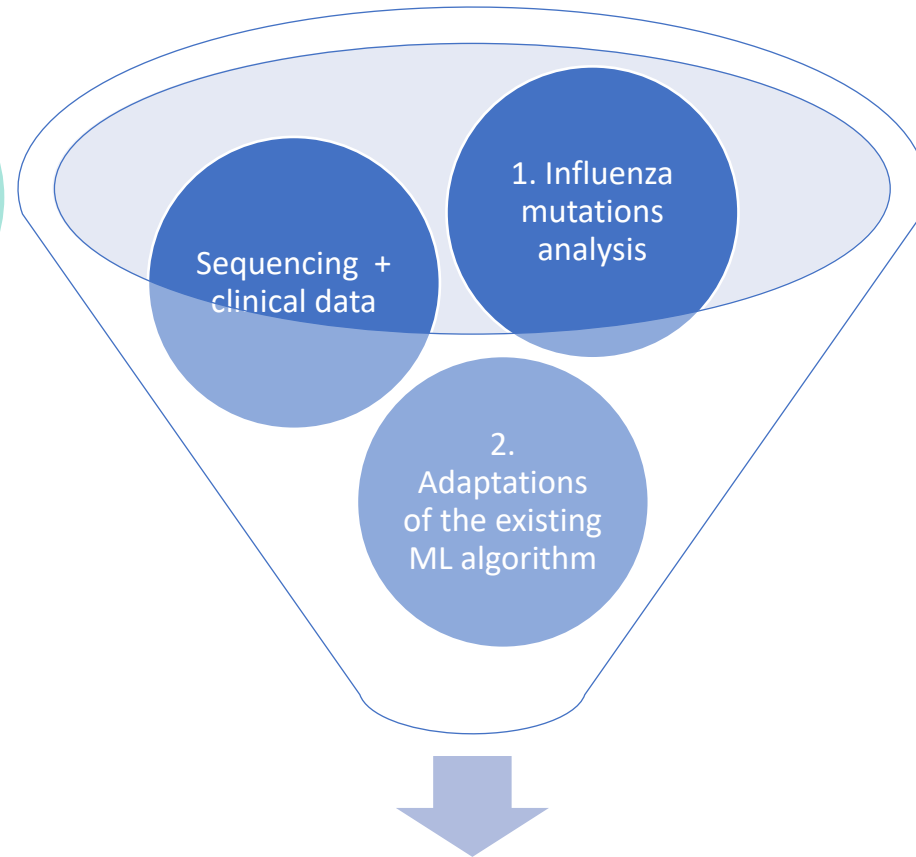
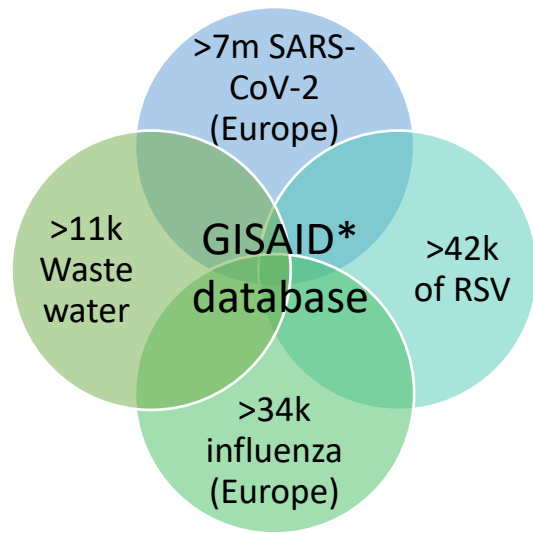
Increased burden of
disease

- Genomic analysis
- Benefited from three developments⁸:
 1. Increasing availability and quality of viral genome sequences;
 2. Growth in computer processing power; and
 3. Development of sophisticated statistical methods.

Objectives



Methodology



GISAID: Global Initiative on Sharing All Influenza Data
ML: Machine learning

3. Using ML algorithm for detection of variants

Mutation analysis of influenza

Mutations overtime

34,885
sequences
from GISAID
for Europe
region

Fluserver tool
to identified
the mutation
sites

Clinical data from HCL Lyon

Hospitalization
status

Vaccination
status

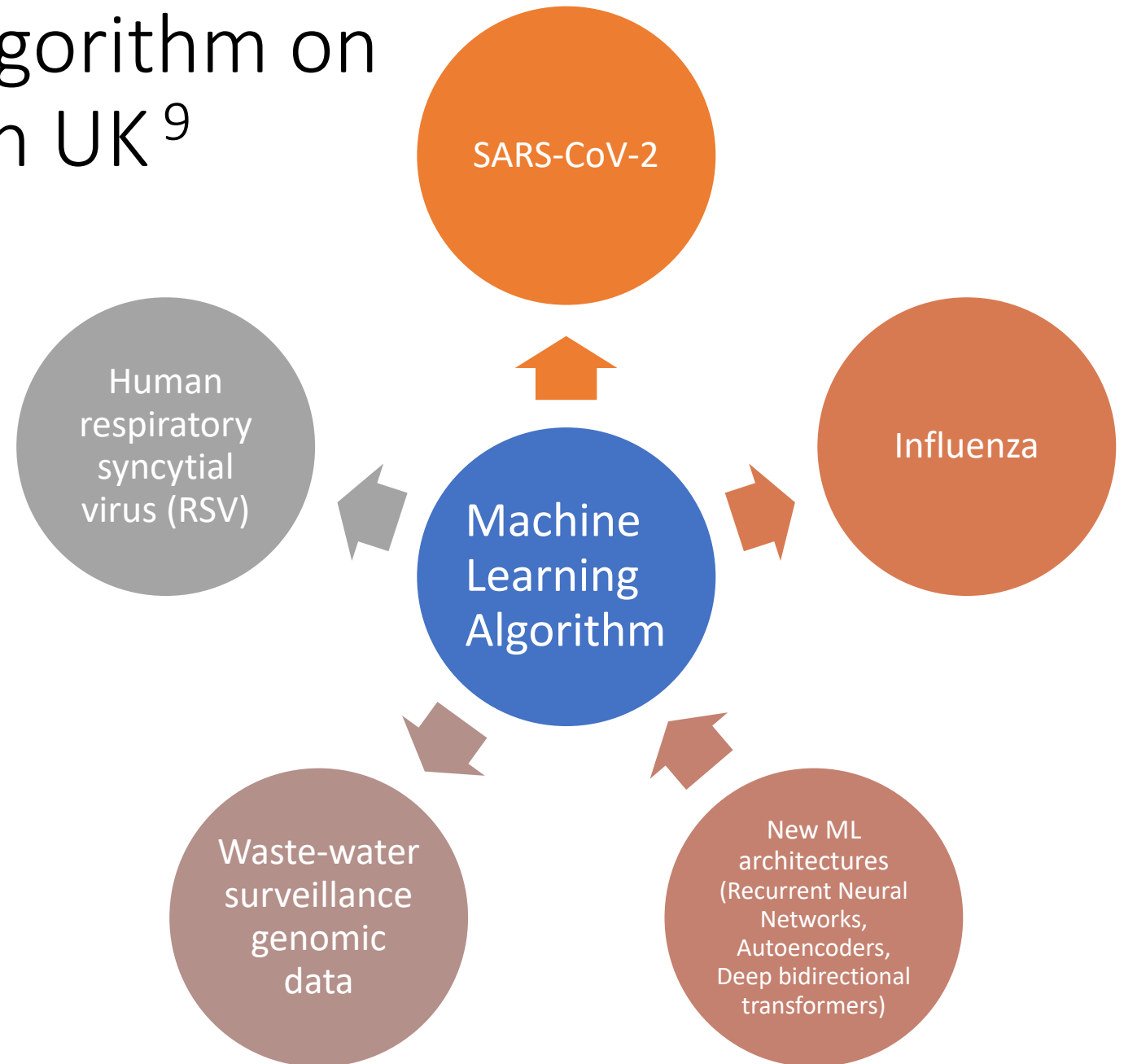
Mutations associated with

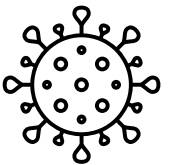
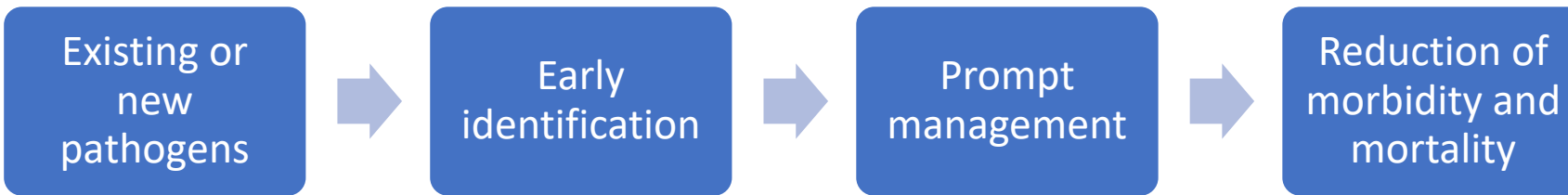
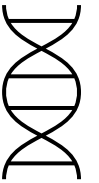
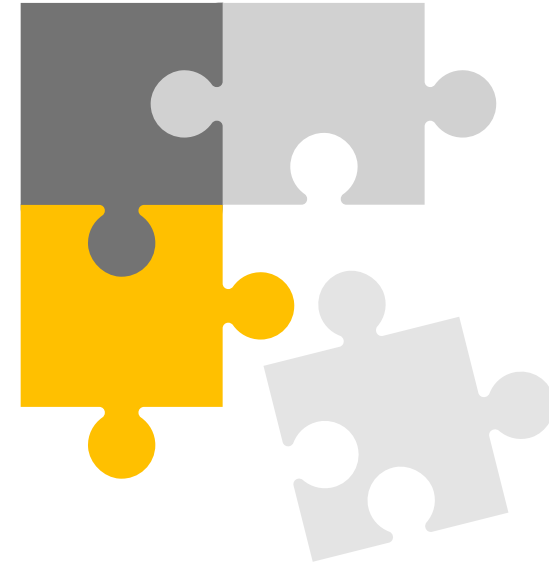
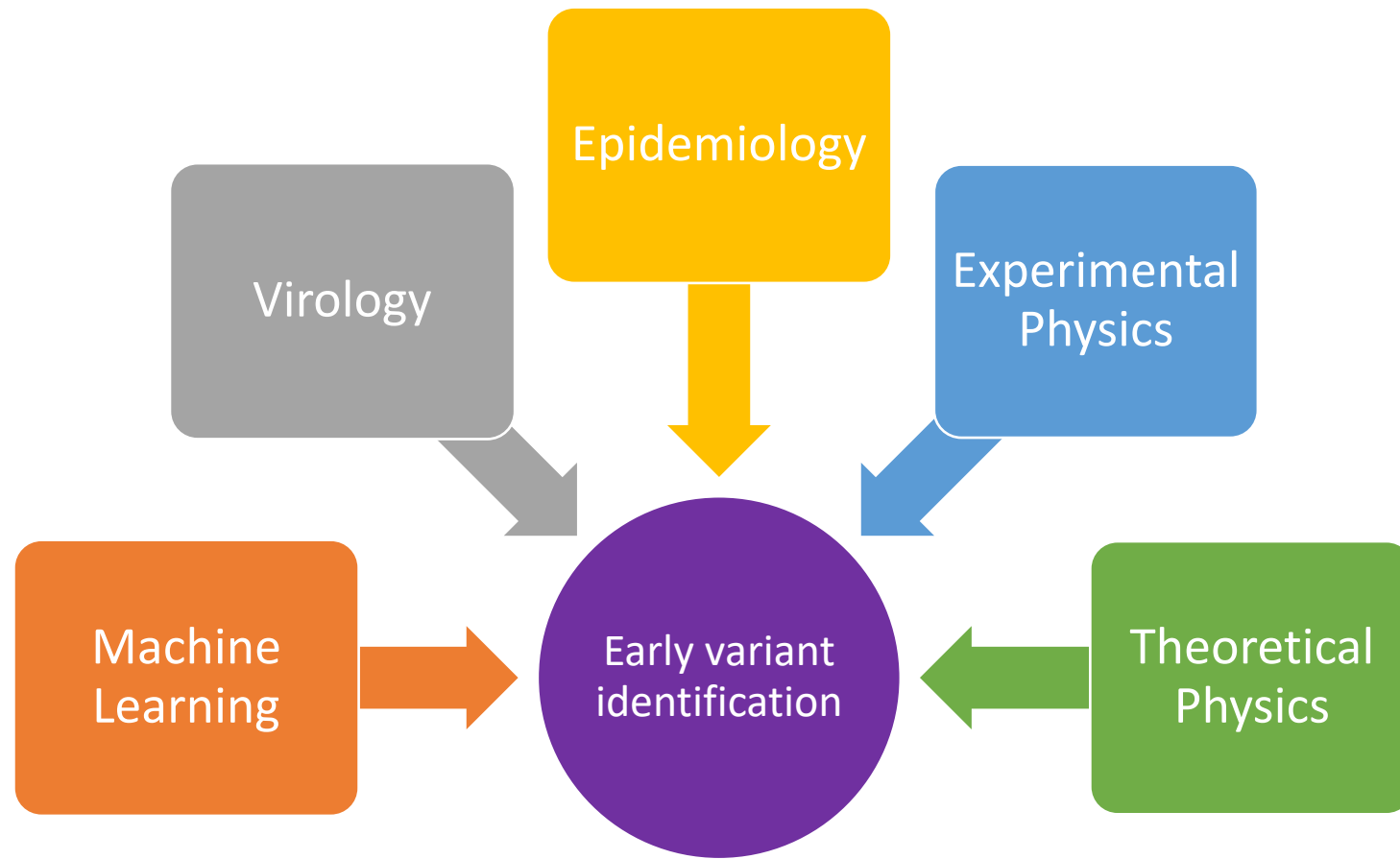
Severity of disease

Immune escape

Proof-of-concept ML algorithm on SARS-CoV-2 evolution in UK⁹

- Early identification of Alpha, Delta, and Delta plus
- Levenshtein distance on spike protein to detect dominant variants, and persistency to detect emerging variants





References

1. WHO (2023, September 29). *COVID-19 situation report*. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>
2. Barber, R. M., Sorensen, R. J. D., Pigott, D. M., Bisignano, C., Carter, A., Amlag, J. O., Collins, J. K., Abbafati, C., Adolph, C., Allorant, A., Aravkin, A. Y., Bang-Jensen, B., Castro, E., Chakrabarti, S., Cogen, R. M., Combs, E., Comfort, H., Cooperrider, K., Dai, X., . . . Murray, C. J. L. (2022, June 1). *Estimating global, regional, and national daily and cumulative infections with SARS-CoV-2 through Nov 14, 2021: a statistical analysis*. *The Lancet*; Elsevier BV. [https://doi.org/10.1016/s0140-6736\(22\)00484-6](https://doi.org/10.1016/s0140-6736(22)00484-6)
3. Troeger, C., Blacker, B. F., Khalil, I., Zimsen, S. R. M., Albertson, S. B., Abate, D., Abdela, J., Adhikari, T. B., Aghayan, S. A., Agrawal, S., Ahmadi, A., Aichour, A. N., Aichour, I., Aichour, M. T. E., Al-Eyadhy, A., Al-Raddadi, R., Alahdab, F., Alene, K. A., Aljunid, S. M., . . . Reiner, R. C. (2019, January 1). *Mortality, morbidity, and hospitalisations due to influenza lower respiratory tract infections, 2017: an analysis for the Global Burden of Disease Study 2017*. *The Lancet Respiratory Medicine*; Elsevier BV. [https://doi.org/10.1016/s2213-2600\(18\)30496-x](https://doi.org/10.1016/s2213-2600(18)30496-x)
4. Li, Y., Wang, X., Blau, D. M., Caballero, M. T., Feikin, D. R., Gill, C., Madhi, S., Omer, S. B., Simões, E. A. F., Campbell, H., Pariente, A. B., Bardach, D., Bassat, Q., Casalegno, J., Chakhunashvili, G., Crawford, N., Danilenko, D., Ha, L. A., Echavarría, M., . . . Walker, G. J. (2022, May 1). *Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in children younger than 5 years in 2019: a systematic analysis*. *The Lancet*; Elsevier BV. [https://doi.org/10.1016/s0140-6736\(22\)00478-0](https://doi.org/10.1016/s0140-6736(22)00478-0)
5. Shi, T., Denouel, A., Tietjen, A. K., Campbell, I. H., Moran, E., Li, X., Campbell, H., Demont, C., Nyawanda, B., Chu, H. Y., Stoszek, S. K., Krishnan, A., Openshaw, P., Falsey, A. R., Nair, H., & Investigators, R. (2019, March 18). *Global Disease Burden Estimates of Respiratory Syncytial Virus–Associated Acute Respiratory Infection in Older Adults in 2015: A Systematic Review and Meta-Analysis*. *The Journal of Infectious Diseases*; Oxford University Press. <https://doi.org/10.1093/infdis/jiz059>
6. Kawasaki, Y., Abe, H., & Yasuda, J. (2023, August 11). *Comparison of genome replication fidelity between SARS-CoV-2 and influenza A virus in cell culture*. *Scientific Reports*; Nature Portfolio. <https://doi.org/10.1038/s41598-023-40463-4>
7. Pandya, M. C., Callahan, S. M., Savchenko, K., & Stobart, C. C. (2019, May 21). *A Contemporary View of Respiratory Syncytial Virus (RSV) Biology and Strain-Specific Differences*. *Pathogens*; Multidisciplinary Digital Publishing Institute. <https://doi.org/10.3390/pathogens8020067>
8. Pybus, O. G., & Rambaut, A. (2009, August 1). *Evolutionary analysis of the dynamics of viral infectious disease*. *Nature Reviews Genetics*; Nature Portfolio. <https://doi.org/10.1038/nrg2583>
9. De Hoffer, A., Vatani, S., Cot, C., Cacciapaglia, G., Chiusano, M. L., Cimarelli, A., Conventi, F., Giannini, A., Hohenegger, S., & Sannino, F. (2022, June 3). *Variant-driven early warning via unsupervised machine learning analysis of spike protein mutations for COVID-19*. *Scientific Reports*; Nature Portfolio. <https://doi.org/10.1038/s41598-022-12442-8>