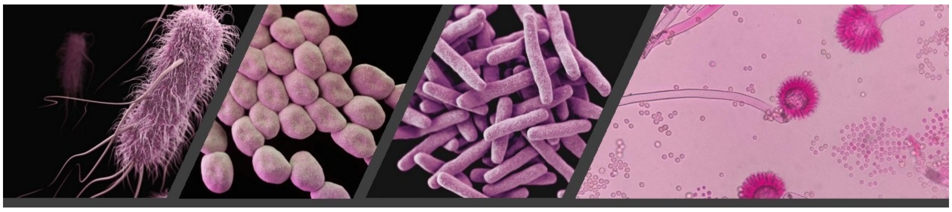

Abstract ID : **65**

Application of Machine Learning to Predict Antigenic Distance Between Newcastle Disease Virus Strains from Sequence Data

Content

Newcastle disease virus (NDV) remains a major vaccination challenge due to its rapid evolution and the emergence of new variants.. Although molecular and sequence data are now quickly and inexpensively produced, genetic distance rarely are a good proxy for cross-protection, while experimental studies to assess antigenic differences are time-consuming and resource-intensive. In the present study several machine learning (ML) methods were developed and compared to predict the antigenic distance between NDV strains as determined by haemagglutination-inhibition (HI) assays based on F and HN gene sequences -and corresponding amino acid features - analysis. Among the models evaluated, the random forest (RF) approach outperformed traditional linear models, achieving a predictive accuracy with an R^2 value of 0.723 compared to only 0.051 for linear models based on genetic distance alone. This significant improvement demonstrates the usefulness of applying flexible ML approaches as a rapid and reliable tool for vaccine selection, minimizing the need for labour-intensive experimental trials. Moreover, the flexibility of this ML framework allows the application of comparable approaches to other infectious diseases in both animals and humans, particularly in scenarios where prompt response and ethical constraints limit conventional experimental approaches.

Keywords

NDV; machine learning; sequencing; cross-protection; hemagglutination inhibition;
antigenic cartography

Registration ID

120

Professional Status of the submitter, who is also the speaker

Professor

Track Classification: Veterinary Bacteriology, Mycology and Virology

Contribution Type: Oral presentation