

Forecasting PEDv in the US through the use of machine learning

K. VanderWaal, ¹ PhD; I. Paploski, ¹ DVM, MSc, DPH; R. Bhojwani, ¹ MDS; C. A. Corzo, ¹ DVM, MS, PhD

¹Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, Saint Paul, Minnesota

Key Points:

- Machine learning is still underutilized in veterinary medicine, but offers great potential for the field
- An ongoing project uses machine learning to predict PEDv cases in certain MSHMP participant systems
- The predictions generated provide production systems and veterinarians with data based tools to aid in disease management

Machine learning is the use of algorithms and statistical models to recognize nuanced patterns in large and complex datasets. In agricultural sciences, machine learning approaches have been applied to a myriad of problems. Despite thousands of publications utilizing machine learning approaches, machine learning has not yet been widely used in the veterinary sciences (Figure 1). Examples of machine learning usage in the veterinary sciences range from algorithms that predict success of conception given insemination in cows, to the prediction of the occurrence of swine shipments to better predict disease occurrence. Based on the potential this new technology has, we decided to use these innovative data science methods to generate new knowledge about the processes that contribute to viral spread, while allowing for the generation of actionable information. Our goals are to further understand the determinants for disease occurrence and how the industry can act to diminish associated risks, while producing real time information that can assist production systems in disease management, and creating a tool that can be applied to more diseases and more systems as we progress.

Since 2017, we have been using machine learning approaches to try and predict the occurrence of porcine epidemic diarrhea virus (PEDV) in sow farms in the United States (U.S.). Utilizing a subset of data from the Morrison Swine Health Monitoring Project (MSHMP) containing weekly PEDV breeding herd status, we built predictive machine-learning models that forecast the probability of a PEDV outbreak two weeks from the day of the assessment.

By using data on weekly PEDV status in breeding pig farms, pig movements, geolocations of farms, environmental, and weather factors we can predict the probability that a sow farm will become infected two weeks from the day of the assessment. Our best model for PEDV prediction so far had a sensitivity (probability of predicting an outbreak when it indeed occurred) of approximately 20%, while the specificity (probability of not predicting an outbreak when none occurred) was >99%. The positive predictive value (probability of an outbreak occurring given that our pipeline predicted one) was approximately 70%, while the negative predictive value (probability of no outbreak occurring given that our pipeline did not predict one) was >99%. This contributes to an overall accuracy of our predictive model of 99.3%. However, this must be interpreted with caution due to the highly unbalanced nature of our data.

We are finalizing the implementation of this tool in a server-run platform. The server communicates directly to the MSHMP database in order to access weekly outbreak data, and has the capability to easily incorporate new production systems that join the project. The platform can also be readily applied to other diseases being tracked by the MSHMP, such as PRRS or a foreign animal disease. However, predictive performance will vary for different diseases according to disease epidemiology. This platform is scalable and was built with the ultimate goal of creating a sustainable tool that improves the industry's ability to anticipate outbreaks and manage disease risk.

Our farm-level forecasts of disease occurrence are unique in that they account for the behavior of (e.g., movements received by) neighboring farms and changes in infection status of neighboring farms. Participating producers have access to their own weekly farm-level forecasts and are able to use these data-based predictions when making decisions about disease risk management. Ultimately, the near real-time estimation of forecasts will provide tools for data-informed actions by producers and practitioners to control outbreaks.

More information on this project, with further details on its methods and results will be presented at the 2020 American Association of Swine Veterinarians, in Atlanta, GA.

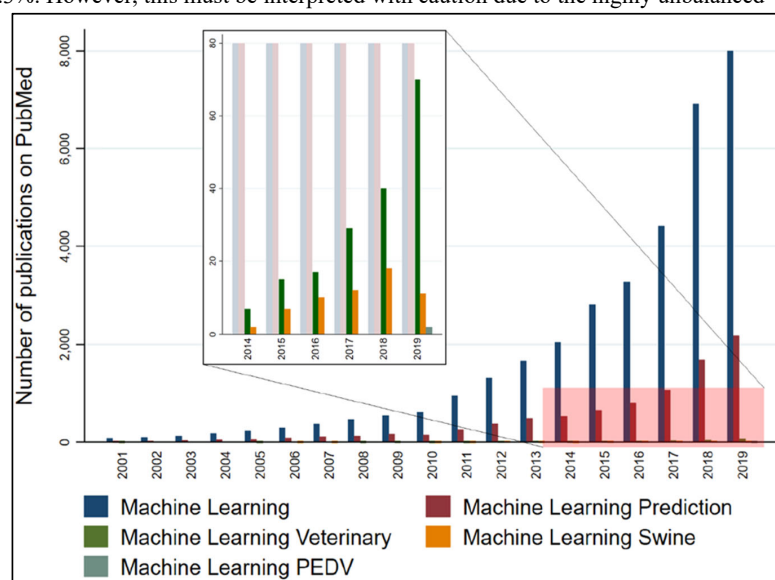


Figure 1. Number of papers published yearly according to different search terms in PubMed. The inset zooms in 2014-2019 to display number of papers in the veterinary science, including swine and PEDV (note difference in scale).

The authors would like to acknowledge the industry partners (production systems and veterinarians) who contributed to data for this analysis and to SHIC, as without their contributions this project would have not been possible.