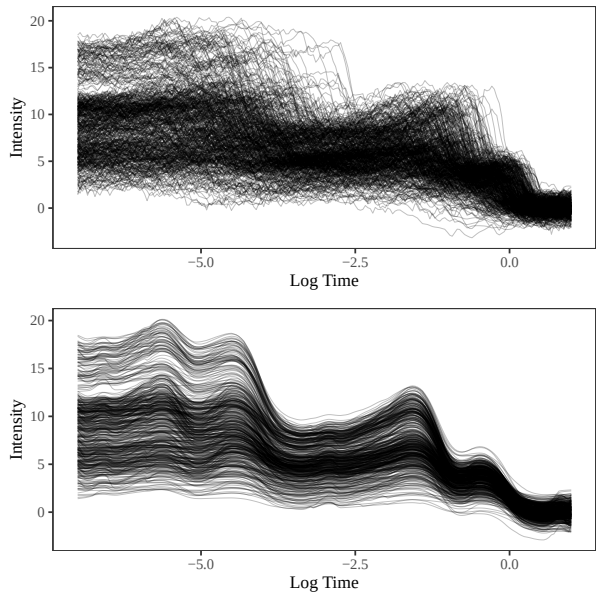

Figure 7: (Top) Examples of simulated optical spectral-temporal signatures from explosions with more variability. (Bottom) The signatures from the plot on the left after applying smoothing and alignment (using box filtering and time warping, respectively) from the fdasrvf R package (version 1.9.3; **? ?**).

5 Discussion

The prediction of explosive device characteristics using optical spectral-temporal signatures from explosions is an example in national security where machine learning applied to functional data could improve performance in practice. However, this is also an example of a machine learning application where not only is high predictive accuracy important, but it is also imperative that it is possible to explain how the model makes predictions. In this paper, we propose a procedure for training and explaining machine learning models with functional data inputs that accounts for the functional nature of the data. We implement our procedure to provide explainable predictions for neural networks trained to predict explosive device characteristics. In particular, the transformation of the optical spectral-temporal signatures using fPCA permits the identification of fPCs important to prediction in a neural network for an explosive device characteristic using PFI, and visualizations for interpreting the variability captured by the important fPCs allows for the determination of the aspects of the signatures that are important for prediction. The validation from an SME of the meaningfulness of the fPCs identified by PFI allows us to be confident that the neural networks are using trustworthy aspects of the signatures to make predictions. A limitation of this method is that the ability to explain a prediction made by the neural network is dependent on the ability to interpret the fPCs. In our example, PFI identifies the first four fPCs as important for predicting at least one of the characteristics, and it is possible to determine meaningful variation captured by these fPCs. However, if PFI identifies fPCs that are not able to be interpreted, it would not be possible to explain the aspects of a function that are important to the neural network for prediction. While the earlier fPCs explain larger amounts of the variability in a data set, it is not necessarily true that the earlier fPCs will be the best for discrimination of response characteristics. If PFI identifies a higher numbered fPC as important, it is likely to be more difficult to interpret.

Another aspect not considered in this paper is that fPCA accounts for amplitude variability (vertical variability) but does not account for phase variability (horizontal variability) in the functions. Joint functional principal component analysis (joint fPCA) is a method that can be applied after smoothing and aligning functional data that accounts for both amplitude and phase variability (**?**; **?**). The procedure in this paper could be adjusted by substituting fPCA with smoothing, aligning, and applying joint fPCA to the signatures (Figure 7). With noisier signatures, accounting for phase variability is important to capture the signals in the data. The focus of this paper is the explanation of the predictions. However, another aspect important to applications connected to high consequence decisions is the ability to report uncertainty quantification to gauge the model's confidence in a prediction. Bayesian neural networks (BNNs) are an example of a machine learning method that returns uncertainty quantification. BNNs produce a distribution for prediction as opposed to a single prediction. If a BNN is used as the machine learning model in our procedure, it would be possible to develop a method that adjusts the computation of PFI to account for the distribution of predictions.

In this paper, we present a method for explaining predictions from a machine learning model trained using functional data. We demonstrate the method on a national security application in which model authentication is crucial. While all statistical methods used within our procedure have been developed previously, we join these techniques in a new way that provides insight to a black-box model. Additionally, our approach accounts for the functional nature of the data, which is an aspect that has been overlooked in the explainable machine learning literature.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. SAND2020-11247 C

Acknowledgments

The authors would like to thank J. Derek Tucker for his advice on fPCA.

Itaque pariatur quibusdam repellendus labore beatae, corrupti sed quia illum numquam fuga quasi at, accusamus quis

cum perferendis quam veniam ullam