SNM Radiation Signature Classification Using Semi-Supervised Machine Learning in Shadow

Jordan Stomps, Paul P.H. Wilson
University of Wisconsin-Madison
stomps@wisc.edu

Background

- The timely detection of special nuclear material (SNM) transfers is important for nuclear nonproliferation monitoring
- It can be costly to characterize large and unlabeled nuclear datasets
- Semi-supervised machine learning allows a model to utilize well-characterized labeled samples and plentiful unlabeled data for understanding a data distribution

Goal: Using MINOS data, demonstrate that semi-supervised machine learning (SSL) can outperform supervised models for binary classification of anomalous nuclear measurements.

MINOS Data Preparation

- Anomalous measurements are identified using a hypothesis testing algorithm (see previous ETI presentations)
  - One-minute measurement frequency over 5 months
  - 1,000 binned feature spectra with background subtracted
  - Background is estimated from lowest gross count-rate in previous 20 minutes
- Since some data must be labeled, a basic heuristic is used to label samples
  - This could be alleviated with domain expert labeling
- Example classifications:
  - SNM transfers exhibit a low energy Compton continuum with a shielded photopeak due to shielded containers
  - Ar-41 photopeak is present when HFIR is active
  - Rainy weather washes out Rn signals
- Classes are collapsed into binary anomalies: SNM transfer or other

$$\boldsymbol{D} = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{|\boldsymbol{L}|}, y^{|\boldsymbol{L}|}), x^{|\boldsymbol{L}|+1}, \ldots, x^{|\boldsymbol{L}|+|\boldsymbol{U}|}\}$$

- Each x is one spectrum sample with its label y and **U** is used in SSL

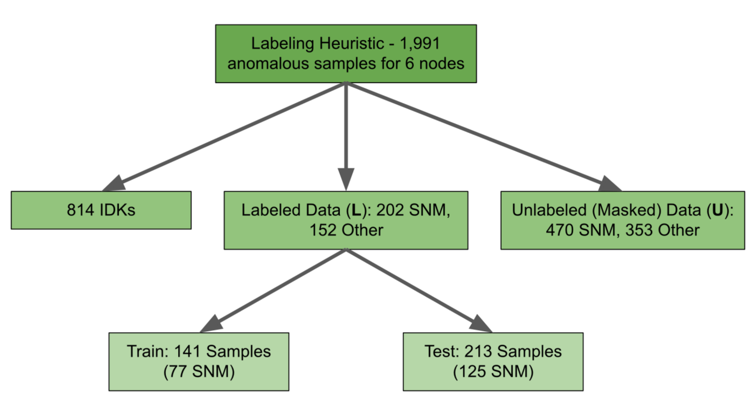

Figure 1: Breakdown of identified anomalies from hypothesis testing, including the training data regime.

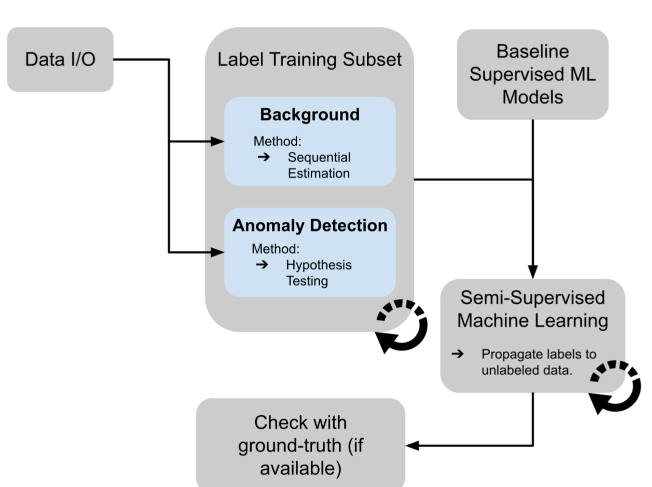

Figure 2: Algorithmic workflow for SNM classification.

Baseline Supervised Model

- The baseline supervised model is Logistic Regression from Scikit-learn
- Can only utilize information from the labeled dataset **L**
- Minimizes loss function:

$$\min_{w,c} \frac{1}{2} w^T w + C \sum_{i=1}^{n} log\left(exp\left(-y_i\left(X_i^T w + c\right)\right) + 1\right)$$

- i.e. the model trains on blue and orange samples from Fig. 3

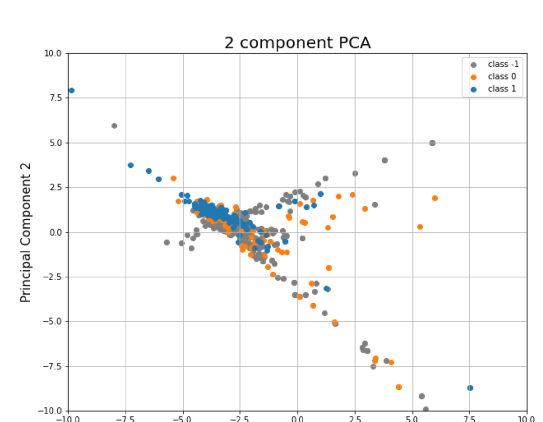

*Figure 3: Principal component analysis for data samples in **L** and **U**. Class 0 (SNM transfers) and class 1 (Other) are used in supervised learning.*

Semi-supervised Models

- Co-training with Logistic Regression
  - Given **L** and **U**, split **L** between two logistic regression models (h_1 and h_2)
  - For each iteration
    - Train h_1 and h_2
    - Have h_1 predict u_1 and h_2 predict u_2 from **U**
    - Include u_1 in the labeled dataset for h_2 and u_2 in the labeled dataset for h_1
  - Evaluate on test set

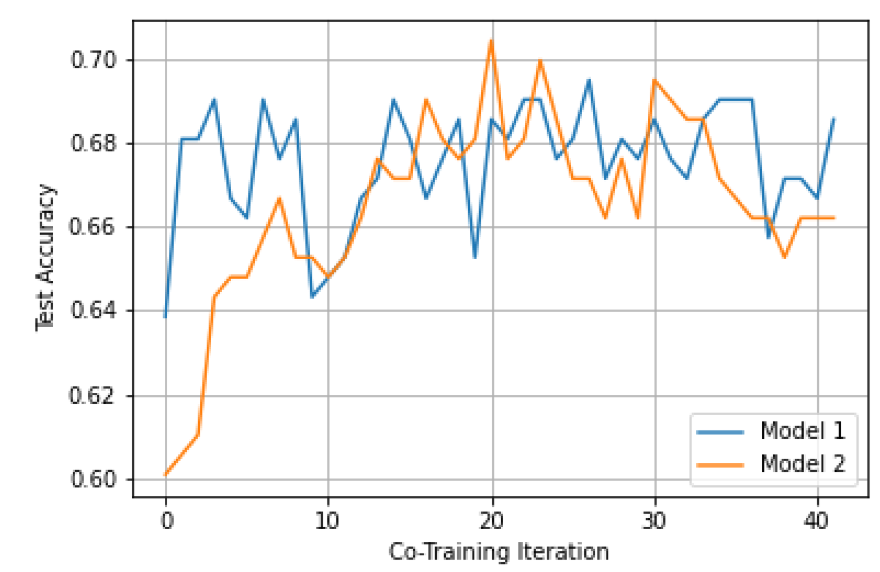

Figure 4: Evaluation curves on test dataset for each model over co-training iterations.

- Label Propagation
  - One of only two SSL models implemented in Scitkit-learn
  - Effectively a semi-unsupervised approach
  - Iterates weighted nearest neighbors from labeled data to unlabeled data to "propagate" labels
- Shadow Python Package
  - Combines loss on labeled and unlabeled data with consistency regularization (with model f_θ)

$$\mathcal{L}(f_\theta(x_l), y_l) + \alpha g(f_\theta, x)$$

  - Exponential Averaging Adversarial Training extends Mean Teacher and Virtual Adversarial Training perturbing samples in a teacher-student framework:

$$g(f_\theta, x) = d(f_\theta(x + r_{adv}), f'_\theta(x))$$

  - Implemented using a Convolutional Neural Network in Pytorch
    - Convolves multiple samples in the network (i.e. treat each spectrum as a 1D image)

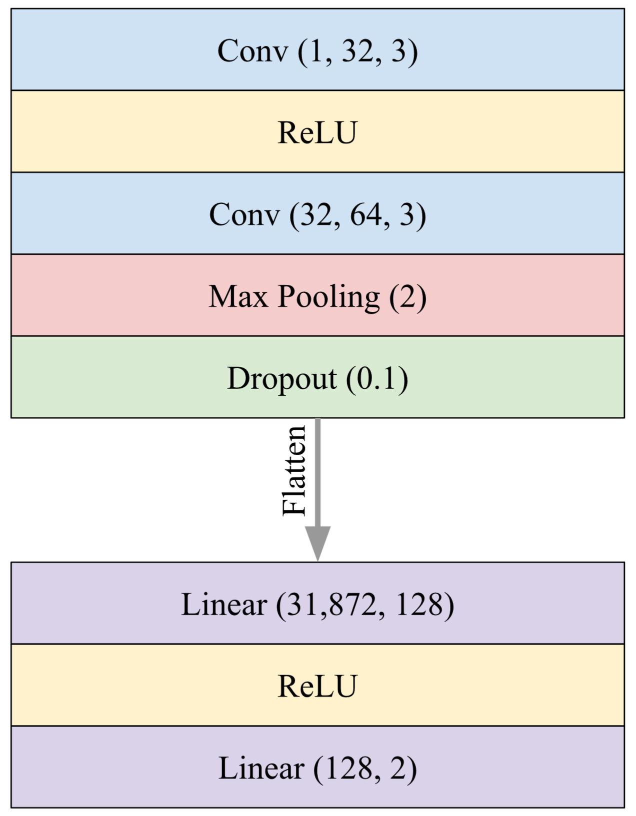

Figure 5: CNN architecture for Shadow EAAT. Reminiscent of basic architectures used for MNIST examples but simplified for 1D images.

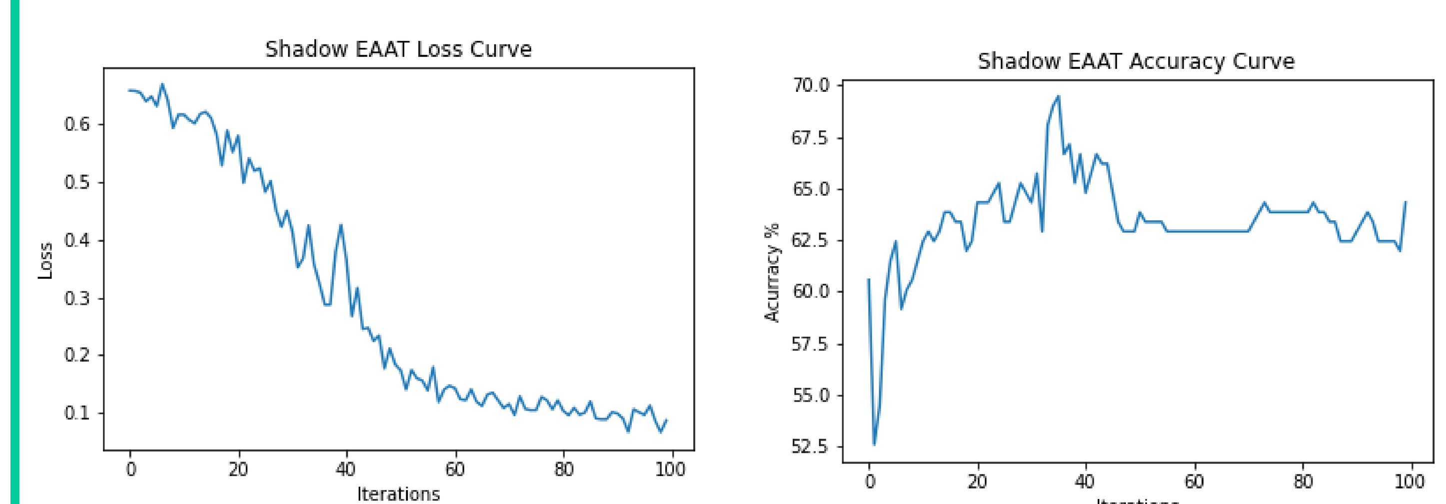

Figure 6: Loss minimization curve and evaluation accuracy curve on test dataset over training iterations for CNN.

Results

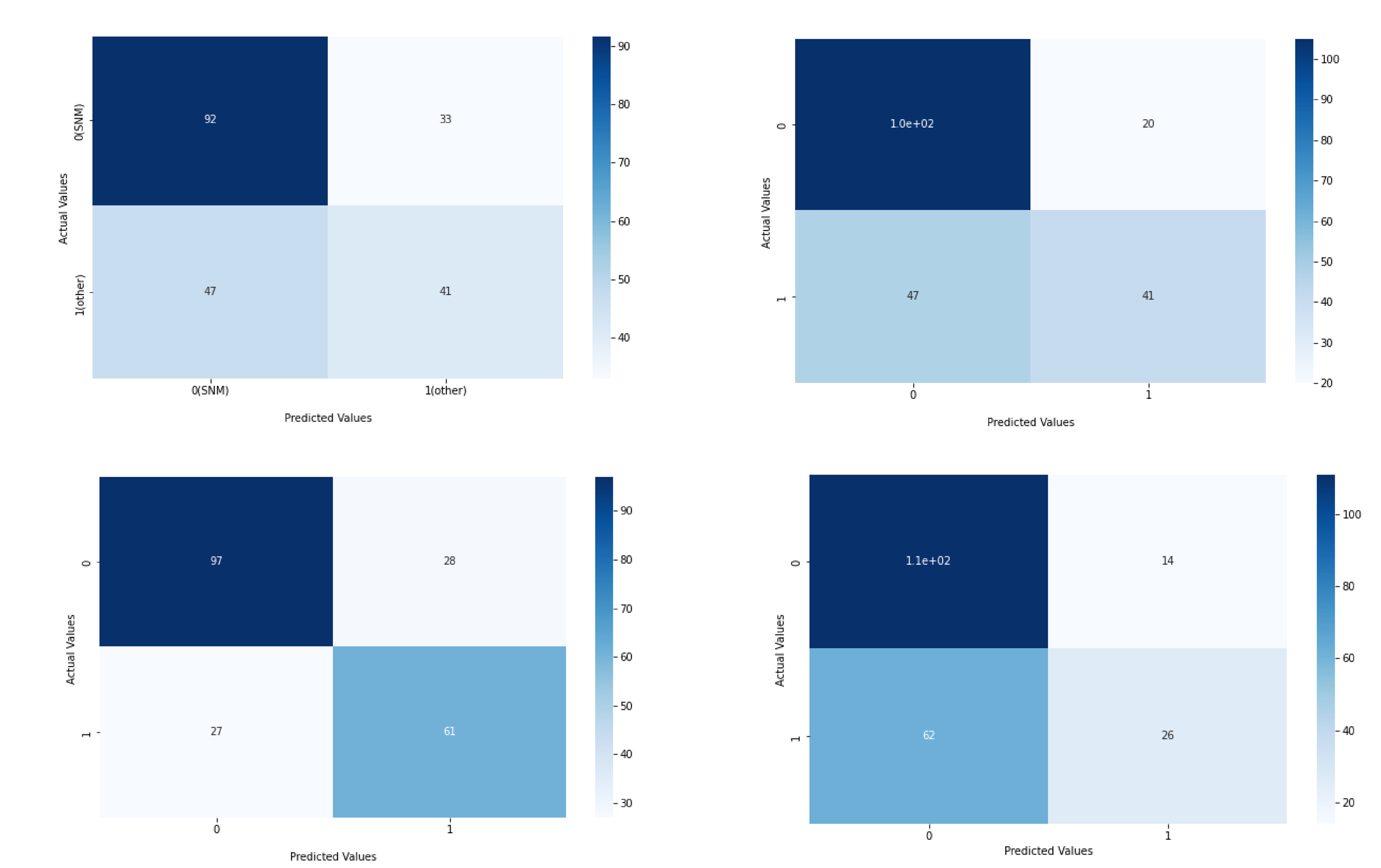

Figure 7: Confusion matrices on test dataset for each machine learning model. (Top Left) Logistic Regression: 62.4% (Top Right) Co-training with Logistic Regression: 66.2-68.5% (Bottom Left) Label Propagation: 74.2% (Bottom Right) Shadow EAAT CNN: 64.3%

Conclusion

- Initial performances show that SSL performs better than supervised methods
- Hyperparameter and architecture optimization should improve performance further
  - The existing algorithm is limited by its labeling heuristic
- Further SSL models should be compared and characterized
- The binary classifier should be extended to multiclass

Acknowledgements & References

We acknowledge and are thankful for the help of the MINOS collaboration at Oak Ridge National Laboratory: Daniel E. Archer, Michael J. Willis, James M. Ghawaly, and Andrew D. Nicholson. They have collected, organized, and shared the data used in this project. Thank you to our collaborators on this research: Ken Dayman, Rob Nowak, and Danica Fliss. Thank you also to the ETI consortium and the NNSA for their support.

[1] A. D. Nicholson, D. E. Archer, I. Garishvili, I. R. Stewart, and M. J. Willis, "Characterization of gamma-ray background outside of the High Flux Isotope Reactor," *J Radioanal Nucl Chem*, vol. 318, no. 1, pp. 361–367, Oct. 2018, doi: 10.1007/s10967-018-6097-5.
[2] A. Blum and T. Mitchell, "Combining Labeled and Unlabeled Data with Co-Training y," p. 10.
[3] X. Zhu and Z. Ghahramani, "Learning from Labeled and Unlabeled Data with Label Propagation," p. 8.
[4] L. Linville, D. Anderson, J. Michalenko, J. Galasso, and T. Draelos, "Semisupervised Learning for Seismic Monitoring Applications," *Seismological Research Letters*, vol. 92, no. 1, pp. 388–395, Oct. 2020, doi: 10.1785/0220200195.

This material is based upon work supported by the Department of Energy / National Nuclear Security Administration under Award Number(s) DE-NA0003921.