Proposal submitted to the National Nuclear Security Administration, Defense Nuclear Nonproliferation Research and Development

Transferability in Multisource Data Analytics for Nuclear Operations Scenarios

webPMIS assigned proposal #	LB20-Nuc Ops Transferability-PD3SS
Office/Team/Objective/Requirement	PD3SS
Submitting Lab	Lawrence Berkeley National Laboratory (LBNL)
Classified Annex	N/A
Principal Investigator	Lee Bernstein labernstein@lbl.gov (510) 486-4951
Lab Manager	John Valentine jdvalentine@lbl.gov (510) 486-4920
Partner Organizations/POC	University of California, Berkeley: Bethany Goldblum

Fiscal Year	Total Budget Request	Fiscal Year	Goal TRL (at EOY)
2020		2020	TRL-3

Date	Revision Number	Description
2019-Sept-14	1	Original

BERKELEY LAB Lawrence Berkeley National Laboratory

Abstract

Machine learning models crafted using data obtained in a specific setting often produce incorrect outputs when applied to data collected in a different setting—even for the same domain and classification task. Transferability, the application of models generated at one facility to other contexts or settings, is critical to the application of multi-modal informatics in proliferation detection scenarios, as it is the *transferability* of the learned representations obtained at known, instrumented nuclear facilities (not the representations themselves) that bring value in characterizing or monitoring unknown facilities of interest. Transferability for proliferation detection applications is a non-trivial problem as differences in nuclear facility layout, equipment, operations cycles, personnel management, etc., will greatly impact the data recorded, but methods are emerging to address this challenge. This work targets data collection at two different nuclear reactor facilities with the goal of developing transferable multi-source machine learning methods to classify nuclear operations—such as reactor operational states, fuel delivery, and reactor refueling—at previously unseen facilities. This represents a presently unmet need in the proliferation detection community for generalizing the applicability of multisource data analytics capabilities.

1 Mission Relevance

1.1 Mission Relevance, Goals and Impact

In order for multimodal analytic methods to provide value in nuclear security applications, we need to demonstrate the ability to transfer proliferation signatures obtained in testbed settings to other settings of interest. The Multi-Informatics for Nuclear Operations Scenarios (MINOS) multi-lab venture funded by DNN R&D uses sensors placed outside the Oak Ridge National Laboratory High Flux Isotope Reactor (HFIR) and the Radiochemical Engineering Development Center (REDC) reprocessing facility to monitor facility operations. This testbed facilitates the identification of measurable phenomena from nuclear facilities and the venture makes available a variety of data products, along with the relevant ground truth, to assess operational questions related to nuclear reactors and target/material processing facilities. A key part of the value proposition, however, is to develop methods and tools that can be applied to a broad range of nuclear facilities. In the case of machine learning methods for data fusion, conventional approaches have traditionally been designed to operate with a feature-space distribution representative of a single setting. The goal of this work is to identify and demonstrate methods to transfer machine learning models trained in known settings to previously unseen facilities of interest.

The expected outcome is a pathway for generalizing the applicability of multisource machine learning classification for a range of facilities and potential proliferation scenarios. Successful demonstration of this capability will benefit the field of data science for proliferation detection and has broader applicability for stakeholders in defense, intelligence, diplomacy, and homeland security.[1]

1.2 Context

1.2.1 Current State of the Art/Practice

Multisource machine learning classification has high potential for mission applications in proliferation detection. In this approach, measurements of various data modalities are obtained and then sorted into a set of predefined classes reflective of the phenomena of interest. For example, data on radionuclide, seismic, thermal, and magnetic emissions are obtained at a nuclear reactor and then evaluated in concert to classify reactor startup/shutdown operations and power level changes that may be reflective of the production of weapons-grade plutonium.

In supervised classification approaches, a training data set is obtained and labeled based on the phenomena of interest, e.g., reactor operational state. The machine learning model is then applied to unlabeled data to sort these sets of signals into classes reflective of different operational states of the reactor—thus fusing the data to enable breakthrough capabilities in reactor monitoring. Supervised classification is a powerful technique that has been shown to yield high classification accuracy for a wide range of tasks.[2] However, it requires a training set of labeled data, "ground truth" that includes detailed operational logs or on-the-ground observations, which may be difficult to obtain for some nuclear facilities of interest. Unsupervised classification obviates the requirement for "ground truth," but classification tasks can be much more error prone in comparison supervised approaches.[3] Testbed facilities allow for both the development and validation of supervised machine learning methods, providing a sandbox for new multimodal analytics and illuminating the art of the possible. But the challenge remains to apply these classifiers in new settings—to enable the monitoring of plutonium production at a variety of reactors and nuclear facilities.

Transferability in machine learning classification is the process of applying a model developed for one task or setting to another task or setting.[4,5] Traditionally, inductive approaches have been used, where a base model is trained on the dataset obtained in a given setting, say the testbed facility, and then applied to the data obtained in Setting 2, the nuclear facility of interest. Such an approach makes an implicit assumption that the feature-space distribution is the same (or sufficiently similar) in both settings. However, individual nuclear facilities differ significantly in terms of their peak power production, fuel inventory, coolant type, reactor vessel and core characteristics, reactivity control, coolant and containment systems, and so on. And these changes will be manifest as changes in the underlying feature-space distribution.

Transductive transfer learning methods avoid this assumption and instead use information embedded in the dataset obtained in Setting 2 to develop a modified model targeting the new setting. By leveraging knowledge, e.g., network structure, features, weights, etc., from models previously trained on testbed data, reasoning from training instances can be applied in a setting where no labeled data are available.[6] Transductive transfer learning has shown success in real-world applications including action recognition,[7] computer vision,[8] and text classification,[9] but little to no work has been performed to assess its applicability and performance for multimodal supervised classification for proliferation detection scenarios.

1.2.2 Relevant Prior Work

The research team has experience in multisource data collection, analysis, and methods development for characterization of nuclear facility operations, most recently via the DNN-supported Interdependent Networks project. At the 88-Inch Cyclotron at LBNL, our team conducted four experimental campaigns, each involving the deployment of an array of multisensor devices. We obtained ground truth from the facility using operator logs, labeled the data, and then developed software to achieve the following tasks:

- Predict ON/OFF status of the beam, given indirect physical sensor data from multisensors deployed in a variety of configurations at various standoffs from the beam generation apparatus, and in different seasons,
- Algorithmically reject data which correspond to experimenter- or staff-induced perturbation of the sensor array,

- Identify the sensing modalities which had the greatest contributions to classification accuracy, assess performance on subsets of the input modalities to avoid spurious correlations, and
- Assess transferability of classifiers by predicting beam ON/OFF on data from experimental campaigns other than the one used for training.

The cyclotron operational status was classified via supervised learning using a multilayer feed-forward neural network and a suite of sensing modalities as inputs. Using multisensor networks deployed inside the facility, the classifier consistently scored a Matthews Correlation Coefficient (MCC) > 0.90 corresponding to classification accuracies greater than 95% when trained and evaluated on data from the same experimental campaign. This work was recently published as a conference proceeding and is currently under development for submission to a peer-reviewed journal.[10]

As a part of this work, our team also developed and released a software package, MIMOSAS (Multisource Input Model Output Security Analysis Suite), for classification of multimodal data to inform nuclear security and proliferation detection scenarios.[11] MIMOSAS provides an end-to-end data processing workflow, from ingestion and pre-processing to model training and test set classification. The pipeline is specified via an input deck, making workflow customization easy, and the framework is modular to allow for the straightforward addition of new learning algorithms. In the current build, the user selects from decision tree, random forest, and feed-forward neural network classifiers to train customizable models with built-in cross validation methods for hyperparameter optimization. With MIMOSAS, the machine learning aspects are fully managed, so the analyst can focus on the mission-relevant aspects of the problem while maintaining full access to the model internals. Trained model outputs are stored with the associated metadata for rapid deployment. These can be applied in supervised classification to assess previously unseen data or for further training as new observations are added to the existing data set. MIMOSAS provides the capability to fuse a wide range of data sources (e.g., radiation, environmental, acoustic, seismic, imagery, etc.) to make, confirm, and correlate machine learning predictions for nuclear security applications.

1.2.3 Technology Readiness Level

The technology readiness level (TRL) at the beginning of this project is TRL-2. The basic principles of transfer learning have been observed, reported, and the concept has been demonstrated for image classification techniques. The application of transfer learning for the characterization of nuclear facility operations is speculative and experimental studies have not yet been conducted to physically validate the formulated application.

2 Scientific and Technical Plan

2.1 Scientific Basis

To transfer supervised classification models obtained in one setting to another in which "ground truth" is not available, transductive transfer learning methods leverage the unlabeled dataset during training.[12] That is, for a given set of *n* multimodel input features obtained at a testbed facility, $x_i = \{x_1, \dots, x_n\}$, these have a label associated with them, y_i, derived from "ground truth" obtained from operational logs and local observations. For the new setting of interest, multimodal input features, $x_i' = \{x_1', \dots, x_n'\}$ are obtained through measurements, but the associated labels in the new setting, y_i', are not known. Transductive transfer learning uses the measurements obtained in the new setting to refine base models for robust classification performance.

There exists a range of approaches for incorporating the new information, $x_i' = \{x_1', \dots, x_n'\}$, into the base model in an attempt to make use of both the unlabeled data, along with labeled data, in training. Here, we explore two: 1) semi-supervised learning under clustering/manifold assumptions, and 2) weak supervision using labeling functions derived from domain expertise. Semi-supervised learning typically refers to a case in which a single dataset has both labeled and unlabeled inputs; In its application for transfer learning, we assume that the base and target datasets have similar underlying features. These underlying distributions are identified, e.g., through clustering or manifold assumptions, and then used to assign proxy labels to the target dataset. For example, using a cluster-then-label approach to semi-supervised learning, both labeled and unlabeled data are clustered to identify the underlying structure of the distributions and this structure is used to illuminate proxy labels.[13] Consider the case of binary classification (on/off) of the reactor operational state. Clustering algorithms may be performed on the testbed dataset in an attempt to identify a natural grouping of the multisource inputs that are largely distinguished based on reactor operational state. The same clustering algorithm is then applied to the target dataset and proxy labels are assigned based on the structure observed in the testbed set. In the case of manifold assumptions, the data are assumed to be distributed relative to the labels in a low-dimensional manifold. This manifold is learned for the testbed dataset and then applied to the target dataset to generate proxy labels. Weak supervision also employs proxy labels, but these are generated through domain heuristics in combination with higher-level supervised classification.[14] Such an approach uses a Label Model to optimize labeling functions to generate improved proxy labels.

Once the proxy labels, y_i', are generated, some or all of the data from the target set, $x_i' = \{x_1', \dots, x_n'\}$, are used to refine the base model. This is done either through continued training of the base model with the new dataset or by "learning without forgetting," i.e., continued training while freeing only the weights of the last few layers of the base network.[15] The expected output is a tailored classifier that takes advantage of the information obtained from the testbed setting, but honed for the target setting, thus providing improved classification accuracy.

In addition, a comprehensive literature review will be performed to assess state-of-the-art transductive transfer learning approaches. The most promising of this selection will also be implemented in toy models for performance assessment and, if time allows, a select set will be incorporated into the codebase for evaluation of the target dataset.

2.2 Proposed Statement of Work

Multisource data will be collected using an array of MERLYN multisensor platforms at two different nuclear reactor facilities: the High Flux Isotope Reactor (HFIR) at ORNL and an alternate university-operated nuclear reactor. The HFIR dataset will be used to develop a base model and provide a benchmark for performance assessment. Transfer learning methods will be developed and implemented to optimize classification performance on data obtained at the alternate nuclear reactor. Transfer learning performance will be evaluated by comparing the benchmark classification accuracy to that obtained for under different transfer learning paradigms.

2.2.1 Task 1: Data Collection at HFIR

An array of 12 MERLYN multisensor platforms are already deployed at HFIR/REDC and have been persistently collecting data since FY19Q3. The MERLYN instrument houses several types of sensors, collects data from these sensors, stores the data locally, and then automatically transmits this data to a central computer via an Ethernet connection. The sensors contained on this board are a magnetometer, ambient temperature sensor, color photo sensor, accelerometer, proximity sensor, and atmospheric pressure sensor, along with GPS used for setting the system time of the device.

LBNL/UCB owns the MERLYN sensor array and will serve as stewards for platform maintenance, data management, and quality control. Each quarter, MERLYN data will be evaluated, cleaned, and transferred to the MINOS database.

2.2.2 Task 2: Sensor Fabrication, Deployment, and Data Collection at Alternate Reactor

The McClellan Nuclear Research Center (MNRC) reactor at UC Davis provides nuclear reactor services and associated support to the UC community for educational and research purposes. The facility hosts a TRIGA (Training, Research, and Isotope Production General Atomics) reactor rated at 2 MW in steady state. LBNL/UCB will procure, assemble, and deploy an array of 5 MERLYN multisensors at the MNRC reactor to collect data related to reactor operations and reactor fuel activities. Ground truth will be obtained through operator and facility logs and data will be labeled based on relevant nuclear facility operations macrophenomena. If the MNRC reactor experiences outages or is otherwise unavailable during this period, the TRIGA reactor at the UC Irvine Nuclear Reactor Facility or an alternate suitable facility will be pursued.

2.2.3 Task 3: Base Network Training

Leveraging MIMOSAS (Multisource Input Model Output Security Analysis Suite) developed under the DNN-supported Interdependent Networks,[16] LBNL/UCB will train supervised learning models using neural and temporal convolutional networks[17] with input data obtained from MERLYN multisensor platforms deployed at HFIR to classify nuclear facility operations. Classification will focus on one or more of the following phenomena:

- reactor fuel activities, including delivery, refueling/defueling, and shipments
- reactor operations, including operational states, outages, and power levels

2.2.4 Task 4: Baseline Transfer Performance Evaluation

LBNL/UCB will apply the models trained at HFIR to data collected from the MNRC reactor to evaluate baseline classification performance i.e., classification accuracy, using MNRC ground truth.

Task 2 must be completed before beginning Task 4.

2.2.5 Task 5: Transfer Learning Method Development

LBNL/UCB will research transfer learning methods with an emphasis on transductive approaches, including but not limited to semi-supervised learning under clustering/manifold assumptions and weak supervision using labeling functions derived from domain expertise. An extensive literature review on these and other methods for machine learning transferability will be conducted. Toy models will be developed, and performance evaluation will be conducted using data previously obtained at the 88-Inch Cyclotron at LBNL under the DNN-supported Interdependent Networks with the goal of optimizing the transferability of multisource machine learning methods. Software will be developed for promising approaches and integrated into the modular MIMOSAS code base.

2.2.6 Task 6: Transfer Learning Performance Evaluation

Using classification performance as a metric and the previously-trained HFIR models as a base, transfer learning approaches will be applied on the Alternate Reactor dataset and performance evaluation will be assessed.

Tasks 2, 3, and 4 must be completed before beginning Task 6. The outcome of Task 5 will inform model selection for Task 6.

2.2.7 Task 7: Final Report and Deep Dive at HQ

A final report will be provided upon completion of the project describing the technical underpinnings of our approach, the success of our approach as demonstrated on this use case, the specific conclusions/contributions that we can make about our approach as relevant for proliferation detection applications, and the potential for transition and next steps.

A Deep Dive will be conducted at NNSA HQ to address the motivations for transfer learning, specific architecture of the models, and utility of these methods for proliferation detection mission applications. A walk-through of the approach from data collection, base network training, transfer learning methods, and performance evaluation will be provided along with a real-time demonstration of transferability in action.

2.3 Multi-year Work Plan

2.3.1 Graphical format

A multi-year work plan in graphical format is shown below.

Task	Year 1			
	Q1	Q2	Q3	Q4
Task 1: Data Collection at HFIR				
Task 2: Sensor Fabrication, Deployment, and Data Collection at Alternate Reactor				
Task 3: Base Network Training				
Task 4: Baseline Transfer Performance Evaluation				
Task 5: Transfer Learning Method Development				
Task 6: Transfer Learning Performance Evaluation				
Task 7: Final Report and Deep Dive at HQ				

2.3.2 Technology Readiness Level

Project Year	TRL Goal	Explanatory Notes
2020	TRL-3	Proof-of-concept demonstration

2.4 Impact

If successful, we will demonstrate the capability for transferring knowledge obtained from multisource data in one proliferation-detection-relevant setting to another. Following the proof-of-concept demonstration, the goal is to expand this R&D into a multi-year effort to plug a key gap in the proliferation detection portfolio by linking testbed learning to real-world proliferation signatures.

2.5 Technical Risks/Issues

The primary technical risk is related to nuclear reactor access and operations. This work requires access to an external nuclear reactor facility and the associated operations logs.

3 Management Plan & Budget

3.1 Management Plan

This is a small project team, with LBNL and UC Berkeley working in joint coordination under the leadership and technical direction of Goldblum (UCB/LBNL). The work will be supported by postdoctoral scholar Dr. Christopher Stewart and a junior research specialist at UCB. The

Nuclear Science and Security Consortium will provide support in the form of an undergraduate research assistant at UCB.

Brief biographies for the expected major contributors are provided in Appendix 1.

3.2 Risk Mitigation

This project is intended as a high-risk, high-reward research and development effort. Our goal is to push beyond the state-of-the-art in transfer learning for proliferation detection applications in machine learning to achieve the best results for overcoming the isolated learning paradigm.

3.3 Collaboration/Work for Others

We anticipate this project to be integrated to some degree with the MINOS venture. Since the tasking and deliverables resulting from this integration are not yet clear, they have not been described in this LCP in any detail. We acknowledge that tasking and focus may change to some degree as this collaboration is fleshed out. Additionally, the University of California, Berkeley will leverage NNSA-sponsored consortium funding via the Nuclear Science and Security Consortium (NSSC) whenever possible to support students and facilitate other interactions.

3.4 Schedule, Milestones, and Deliverables

TABLE OF DELIVERABLES

Title (as will be entered in webPMIS)	Description (as will be entered in webPMIS)	Associated Task(s)	Due Date
Quarterly Report	Per HQ format	All	w/in 10 days EOQ
Annual TRL Readiness Assessment	Per HQ format	All	w/in 30 days EOY
Final Report	Per HQ format	All	w/in 30 days EOP

TABLE OF MILESTONES

Description	Associated Task(s)	Due Date
Upload MERLYN data for transfer to MINOS database	1	w/in 30 days EOQ
MERLYN sensors assembled and ready for deployment	2	12/31/2019
Transfer learning literature review complete	5	12/31/2019
Data collection at alternate reactor completed	2	3/31/2020
Base network trained on HFIR dataset	3	3/31/2020
Toy transfer methods implemented and evaluated	5	3/31/2020
Baseline transferability performance assessed	4	4/30/2020
Transfer learning performance evaluated	6	9/1/2020

Note that deliverables and milestones may be updated as priorities are established. The LCP may be updated at a later date with these deliverables/milestones and any corresponding changes in task structure.

3.5 Budget

Fiscal Year	Budget Request

2020	

Appendix 1 – Research Team

Lee Bernstein, Lawrence Berkeley National Laboratory, labernstein@lbl.gov, 510-486-4951
Bernstein leads the Bay Area Nuclear Data (BAND) group at Lawrence Berkeley National Laboratory. He has served as PI on several NA-22 efforts, including the Interdependent Networks predecessor. Bernstein will be the principal investigator for this project and will manage administrative aspects at LBNL.

Bethany Goldblum, UC Berkeley, bethany@nuc.berkeley.edu, 510-643-2065
Goldblum is a Research Engineer in the Department of Nuclear Engineering at the University of California, Berkeley and Affiliate at Lawrence Berkeley National Laboratory. She is the Executive Director for the Nuclear Science and Security Consortium and leads research efforts in neutron detection and multisource data analytics. Goldblum has a Ph.D. in Nuclear Engineering from UC Berkeley and is the author of more than 30 peer-reviewed publications. Goldblum will be principal investigator for the UC Berkeley subcontract and serve as lead on the effort in her role as LBNL Affiliate.

Christopher Stewart, UC Berkeley, clstewart@berkeley.edu, (501) 352-2046
Chris Stewart is a postdoctoral scholar at UC Berkeley. He holds a Ph.D. in Nuclear Engineering from the Georgia Institute of Technology.

Appendix 2: REFERENCES

1. "Artificial Intelligence and National Security," Congressional Research Service Report R45178, January 30, 2019.

2. S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques," Informatica **31** (2007) 249-268.

3. "Learning Module 6.1: Image Classification," in GSP 216: Introduction to Remote Sensing, Geospatial Science, Humboldt State University, 2019.
http://gsp.humboldt.edu/OLM/Courses/GSP_216_Online/lesson6-1/supervised.html

4. S. J. Pan and Q. Yang, "A Survey on Transfer Learning," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345-1359, Oct. 2010. doi: 10.1109/TKDE.2009.191

5. Kouw, W.M., Loog, M. 2018. An introduction to domain adaptation and transfer learning. arXiv e-prints arXiv:1812.11806.

6. A. Arnold, R. Nallapati, and W.W. Cohen, "A Comparative Study of Methods for Transductive Transfer Learning," *Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007), Omaha, 2007*.

7. N. FarajiDavar, T. de Campos, J. Kittler and F. Yan, "Transductive transfer learning for action recognition in tennis games," *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, Barcelona, 2011, pp. 1548-1553.
doi: 10.1109/ICCVW.2011.6130434

8. N. FarajiDavar, "Transductive Transfer Learning for Computer Vision," University of Surrey, PhD Thesis, 2015.

9. J. Pan, "Feature-based Transfer Learning with Real-World Applications," The Hong Kong University of Science and Technology, PhD Thesis, 2010.

10. C.L. Stewart, B.L. Goldblum, Yi-Ting Alicia Tsai, S. Chockkalingam, S. Padhy, and A. Wright, "Multimodal Data Analytics for Nuclear Facility Monitoring," in *Proceedings of the Institute of Nuclear Materials Management 60th Annual Meeting, Palm Desert, 2019*, (Institute of Nuclear Materials Management, 2019).

11. Jared Zhao, Bethany L. Goldblum, Christopher Stewart, Alicia Ying-Ti Tsai, Shruthi Chockkalingam, and Pedro Vicente Valdez. (2019). MIMOSAS: Multimodal Input Model Output Security Analysis Suite (Version 1.0) [Computer software]. Berkeley, CA: University of California, Berkeley. https://complexity.berkeley.edu/mimosas/

12. M. Rohrbach, S. Ebert, and B. Schiele, "Transfer learning in a transductive setting," in Advances in Neural Information Processing Systems, eds. C. J. C. Burges, L. Bottou, M. Welling, Z. Gharhramani, and K. Q. Weinberger, 2013, 46-54. http://papers.nips.cc/paper/5209-transfer-learning-in-a-transductive-setting.pdf

13. M. Peikari, S. Salama, S. Nofech-Mozes, and A. L. Martel, "A Cluster-the-label Semi-supervised learning approach for pathology image classification," Nature Scientific Reports, 8 (2018), 7193. https://www.nature.com/articles/s41598-018-24876-0

14. Ratner, Alexander et al. "Snorkel: Rapid Training Data Creation with Weak Supervision." *Proceedings of the VLDB Endowment. International Conference on Very Large Data Bases* vol. 11,3 (2017): 269-282. doi:10.14778/3157794.3157797

15. Z. Li and D. Hoiem, "Learning without Forgetting," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 12, pp. 2935-2947, 1 Dec. 2018. doi: 10.1109/TPAMI.2017.2773081

16. Jared Zhao, Bethany L. Goldblum, Christopher Stewart, Alicia Ying-Ti Tsai, Shruthi Chockkalingam, and Pedro Vicente Valdez. (2019). MIMOSAS: Multimodal Input Model Output Security Analysis Suite (Version 1.0) [Computer software]. Berkeley, CA: University of California, Berkeley. https://complexity.berkeley.edu/mimosas/

17. Bai, S., Zico Kolter, J., Koltun, V. 2018. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. arXiv e-prints arXiv:1803.01271.