INL/CON-23-72154-Revision-0

A New Approach to Identify and Characterize Low Power Shutdown Initiating Events Using Machine Learning Techniques

October 2023

Zhegang Ma, Fei Xu, Sai Zhang, Min Xian

Changing the World's Energy Future

Idaho National Laboratory

INL is a U.S. Department of Energy National Laboratory operated by Battelle Energy Alliance, LLC

INL/CON-23-72154-Revision-0

A New Approach to Identify and Characterize Low Power Shutdown Initiating Events Using Machine Learning Techniques

Zhegang Ma, Fei Xu, Sai Zhang, Min Xian

October 2023

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

October 23–25, 2023

Zhegang Ma, Ph.D., P.E.

INL/CON-23-72154

Applying AI/ML Techniques to U.S. Nuclear Operating Experience Program

Zhegang Ma[1], Min Xian[2], Fei Xu[1], Sai Zhang[1]

[1]Idaho National Laboratory, USA

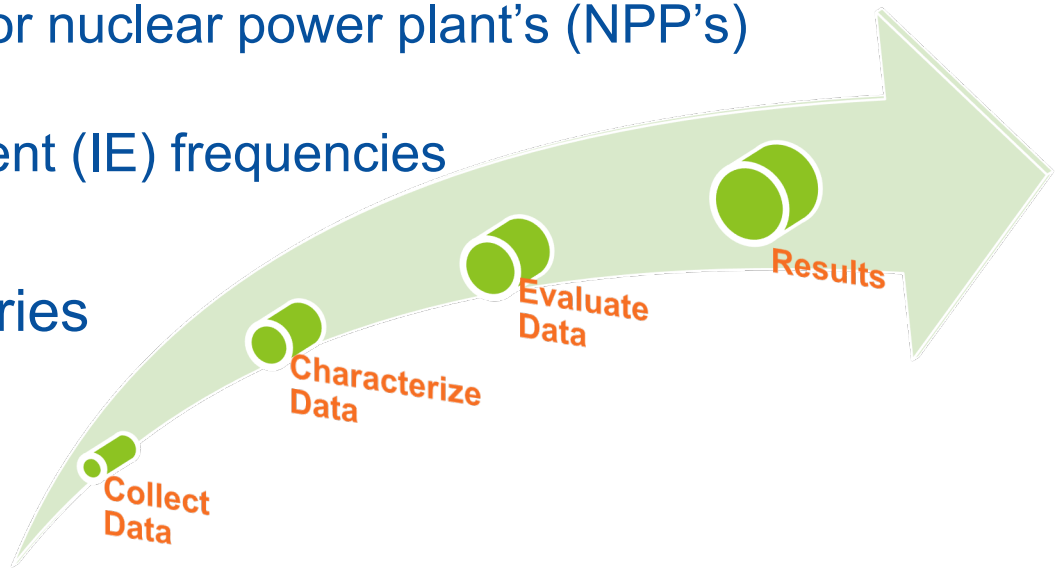
[2]University of Idaho, USA

Zhegang.Ma@inl.gov, mxian@uidaho.edu, Fei.Xu@inl.gov, Sai.Zhang@inl.gov

Idaho National Laboratory

U.S. Nuclear OpE Program: Overview

- Idaho National Laboratory (INL) has provided technical assistance to the Nuclear Regulatory Commission (NRC) in reliability and risk analysis including the operating experience (OpE) program since the 1980s
- The U.S. nuclear OpE program provides input parameters to the NRC's Standardized Plant Analysis Risk models
  - Also used by nuclear industry as generic data for nuclear power plant's (NPP's) probabilistic risk assessment (PRA) models
  - Equipment reliability, unavailability, initiating event (IE) frequencies
  - Common cause failure (CCF) probabilities
- The nuclear OpE program includes two categories
  - Data collection and characterization
  - Data computations and presentation



U.S. Nuclear OpE Program: Main Data Sources

- **Institute for Nuclear Power Operations (INPO) Database**
  - INPO is the U.S. industry organization that promotes safety and reliability in NPPs
  - Plants report the equipment failure/unavailability data as well as demand and run time data to INPO
  - INPO provides the above proprietary data to the NRC/INL quarterly for data analysis
- **Licensee Event Reports (LERs)**
  - Plants are required to submit LERs to the NRC within 60 days of a reportable event
  - Example reportable events: plant manual or automatic reactor scram; safety barriers degraded; inadvertent actuation of emergency core cooling system
- **Event Notification (EN) Reports**
  - Immediate notification reports from plants to the NRC
- **Monthly Operating Reports (MORs)**
  - Each plant's online hours, critical hours, and shutdown information

U.S. Nuclear OpE Program: Current Practice

- **Data Characterization**
  - Analysts review OpE data source documents such as LERs and INPO data and code the events into different studies
    - **Component Failure**: component type, failure mode, cause, degradation value, detection, recovery…
    - **Initiating Event**: IE category
    - **Shutdown IE (SDIE)**: SDIE category, cause, duration, temperature change…
    - **Loss of Offsite Power (LOOP)**: event category, timeline, cause, unit status, initiator, consequential…
    - **CCF**: component type, failure mode, cause, group size, coupling factor/strength, time delay factor, defense mechanism, shock type…
- **Data Computations**
  - Use number of failures and valid demands to estimate failure probability
  - Use number of failures and run time to estimate failure rate or IE frequency
  - Use number of CCF events and total failure events to estimate CCF alpha factor

U.S. Nuclear OpE Program: Challenges

- Manually review source documents and add coding to studies and events
- Established coding system and data fields in existing database and studies
  - What about new studies or expanding existing studies?
    - Digital system
    - Shutdown IE
- Decreasing number of CCF events
- Classic statistical approaches to estimate failure probabilities/rates
  - How good is previous performance to predict current or future performance?

U.S. Nuclear OpE Program: Opportunities with AI/ML

- Advanced computational tools and techniques with artificial intelligence (AI) and machine learning (ML)[1]
  - Recognize and process both structured data and unstructured data
    - Data classification for new or expanded data fields, studies, or source documents
    - Causal learning, causal inference
  - Train and enhance data-driven models and reflect the relationships between model inputs and outputs
  - Develop predictive models for parameter estimations
  - Other applications in supporting and optimizing NPP operation and maintenance including advanced diagnostic (such as detecting failure cause) and prognostic models (such as predicting the remaining useful life)

[1] Ma et al. (2022). *Exploring Advanced Computational Tools and Techniques with Artificial Intelligence and Machine Learning in Operating Nuclear Plants.* NUREG/CR-7294, INL/EXT-21-61117. U.S. Nuclear Regulatory Commission.

AI/ML Applications in Nuclear OpE Program

- Low Power Shutdown (LPSD) Initiating Event Study
  - Using free text data (event descriptions)
  - To automate identify and characterize a LPSD initiating event
- Causal Learning Study [2,3]
  - Using free text data (event descriptions)
  - To automatically learn causal relationships and integrate them into a causal network
- Causal Inference Study
  - Based on the causal network from the causal learning study
  - To estimate failure probabilities conditioned on degraded conditions

[2.] Zhang, S., Xu, F., Ma, Z., & Xian, M. (2023). *Natural Language Processing-Enhanced Nuclear Industry Operating Experience Data Analysis: Aggregation and Interpretation of Multi-Report Analysis Results*. 2023 International Topical Meeting on Probabilistic Safety Assessment and Analysis (PSA 2023). Knoxville, TN

[3.] Zhang, S., Xu, F., & Ma, Z. (2022). *Natural Language Processing-Enhanced Nuclear Industry Operating Experience Data Analysis to Support Risk Model Parameter Estimations*. 16th Probabilistic Safety Assessment & Management Conference. Honolulu, HI

LPSD Initiating Event Study

- One issue in developing LPSD PRA is the reasonable estimation of LPSD IE frequencies
  - NRC has a SDIE database based on LER review
    - However, the shutdown IE database does not include all LPSD IEs as many LPSD IEs may not be reportable events and thus not included in LER
    - The NRC shutdown IE database is only a subset of the LPSD IEs in nuclear plants
  - Manually reviewing OpE data is resource intensive

LPSD Initiating Event Study (cont.)

- Electrical Power Research Institute (EPRI) did conduct LPSD IE study using data from the INPO industry database
  - Keyword search plus manual review
- We are investigating a new approach to identify and characterize LPSD IEs from the INPO database using AI/ML techniques
  - Independent from the EPRI study
  - Use and benchmark with the NRC SDIE database and the EPRI study

LPSD Initiating Event Study: SDIE Categories

Shutdown IE Category	Description	Details
ISOL	Trip or Isolation of Shutdown Cooling Loop	Primary isolation, does not include low level trip due to LOCA
LOAC	Loss of Safety or Vital Bus for Shutdown Cooling Equipment	Loss of vital bus due to LOOP or local fault
FLOW	Diversion or Loss of Cooling Water Flow	Blockage or diversion of primary coolant or service/closed cooling water flow path such that heat removal is no longer accomplished, does not include primary isolations or losses of primary coolant from the primary system
LOCA	Loss of Coolant Accident (LOCA)	Includes inadvertent drain-down of primary system where sufficient coolant is no longer available for the normal decay heat removal process
LOOP	Loss of Offsite Power	
SPF	Spent Fuel Pool	

LPSD Initiating Event Study: New AI/ML Approach

- The main process in the AI/ML approach is to
  - Find out the relationship between keywords in the event descriptions and the LPSD IE categories
  - Apply the relationship to the INPO database records
- The new approach can be used as a supplemental method to peer review the EPRI study results
- If the new approach is proved to be accurate and efficient, it could be applied to other NRC OpE data characterization processes for LOOP events, IEs, CCFs, and multi-unit CCFs

LPSD Initiating Event Study: 2022 Study

- Three popular classification models were used in our earlier study in 2022
  - Support Vector Machine (SVM), with two different feature selecting strategies
  - Naïve Bayes
  - Random Forest
- Data
  - Training/Testing Dataset: ~ 150 shutdown IEs and > 500 non-shutdown IEs
  - Target Dataset: more than 12,000 records from INPO (both at-power and shutdown events)

Categories information	Category #	Total # of records
'ISOL'	1	24
'LOAC'	2	64
'FLOW'	3	17
'LOCA'	4	3
'LOOP'	5	35
'SPF'	6	2
'NONSDIE'	7	542
Total of the records		687

LPSD Initiating Event Study: 2022 Study (cont.)

- Input: event description field of records
- Output: one of seven event categories
- Process:
  1) Remove stop words from the event description
  2) Create term frequency-inverse document frequency features
  3) Separate the training and testing dataset (e.g., 70% of the data is randomly selected as a training set with all remaining as a testing set)
  4) Apply training and testing process to the models and obtain the overall performances
  5) Calculate the accuracy of each category

Accuracy = Sum of Predicted Correctly Events (from Cat. 1 to Cat. 7)
/ Sum of Known Events (from Cat. 1 to Cat. 7)

LPSD Initiating Event Study: 2022 Study (cont.)

- Performance comparisons of the four models

Models	Average Accuracy	Category Accuracy for Test Dataset							Category Accuracy for Training Dataset						
		1	2	3	4	5	6	7	1	2	3	4	5	6	7
		5	16	4	0	13	0	169	19	48	13	3	22	2	373
SVM	**0.8647**	0	11	0	0	2	0	167	19	45	13	3	21	2	373
Naive Bayes	0.8164	0	0	0	0	0	0	169	0	2	0	0	0	0	373
Random Forest	0.8454	2	5	0	0	0	0	168	19	45	13	3	21	2	373
SVM with Chi-Square Feature	**0.8599**	1	9	1	0	9	0	138	17	39	9	2	18	1	373

→ Category

→ "Known" Events

ML Predicted Events

LPSD Initiating Event Study: 2022 Study (cont.)

- Applying the AI/ML models to the Target Dataset

	Certainty Degree	ISOL	LOAC	FLOW	LOCA	LOOP	SFP	NonSDIE	Total
High Certainty	1	1	287	1	0	1	0	542	832
	0.83	6	182	8	0	5	1	0	202
Low Certainty	0.66	190	276	16	0	1	0	0	483
	0.5	645	266	39	2	3	1	0	956
	0.33	187	30	0	0	1	1	0	219
Total		1029	1041	64	2	11	3	542	2692
Certainty of SDIE = 1 or 0.83		7	469	9	0	6	1	0	492
Known Events		24	64	17	3	35	2	542	687

2022 LPSD IE Study Summary

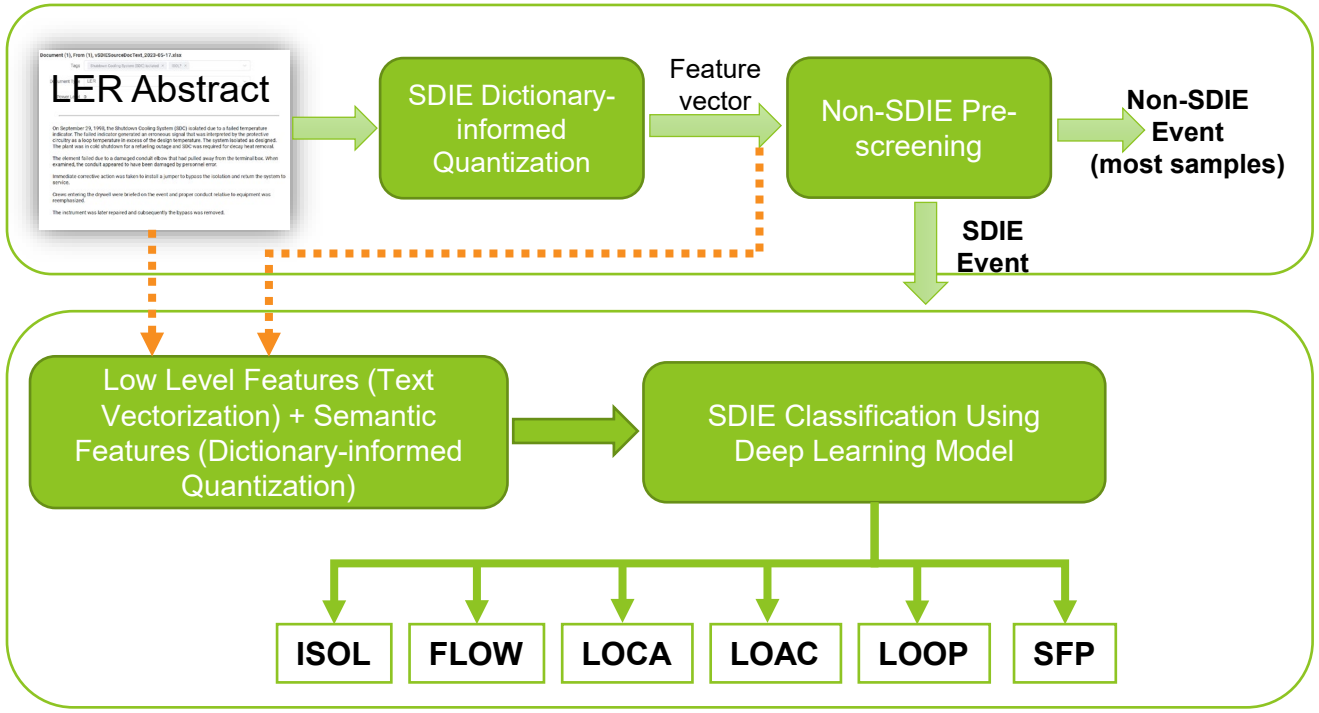
- Preliminary models were developed and applied to the nuclear OpE data
- From a total of about 2,700 shutdown events (IE or not), the models characterize
  - About 500 shutdown IEs with high certainty
  - About 2,100 shutdown IEs with either high or low certainty
  - Still need extensive manual review to confirm the results from the models
  - The models should be improved further
- Issues identified from the 2022 study
  - Training/testing datasets are too small
  - Imbalance data points
    - Much more non-SDIE events than SDIEs
    - The number of SDIEs vary significantly among the six shutdown IE categories
    - Event descriptions that were used as inputs for AI/ML models are input by plant engineers and are in “free style”

LPSD Initiating Event Study: 2023 Study

- Preliminary models and performance in the 2022 Study were biased because of the imbalance dataset, i.e., >80% of samples are non-SDIEs.
- In the 2023 study
  - Built a SDIE dictionary and used it to extract features from event description
  - Proposed a two-stage pipeline that can handle the imbalance issue
  - Built a web-based NLP tool to accelerate SDIE annotation
  - Models can be improved further using high-quality training set

LPSD Initiating Event Study: 2023 Study (cont.)

- Proposed, implemented and evaluated a dictionary-informed deep learning approach
  - Develop a **two-stage pipeline** that solves the issue of imbalance dataset
  - Extract **semantic features using a SDIE dictionary**
  - Integrate low-level text features
  - Use a deep neural network for SDIE classification



Pipeline of the newly proposed dictionary-informed deep learning approach for SDIE classification

LPSD Initiating Event Study: 2023 Study (cont.)

- Issues identified from the ongoing 2023 study
  - Training/testing datasets are still small
  - Some FLOW and ISOL events are incorrectly annotated
  - More FLOW, ISOL and SFP events are needed
  - Event descriptions and terms in the records are not standardized
- More results will be provided after the study is concluded

Idaho National Laboratory

Battelle Energy Alliance manages INL for the U.S. Department of Energy's Office of Nuclear Energy. INL is the nation's center for nuclear energy research and development, and also performs research in each of DOE's strategic goal areas: energy, national security, science and the environment.