NUREG/CR-7294
INL/EXT-21-61117

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Exploring Advanced Computational Tools and Techniques with Artificial Intelligence and Machine Learning in Operating Nuclear Plants

Office of Nuclear Regulatory Research

AVAILABILITY OF REFERENCE MATERIALS IN NRC PUBLICATIONS

NRC Reference Material

As of November 1999, you may electronically access NUREG-series publications and other NRC records at the NRC's Library at www.nrc.gov/reading-rm.html. Publicly released records include, to name a few, NUREG-series publications; *Federal Register* notices; applicant, licensee, and vendor documents and correspondence; NRC correspondence and internal memoranda; bulletins and information notices; inspection and investigative reports; licensee event reports; and Commission papers and their attachments.

NRC publications in the NUREG series, NRC regulations, and Title 10, "Energy," in the *Code of Federal Regulations* may also be purchased from one of these two sources:

1. **The Superintendent of Documents**
   U.S. Government Publishing Office
   Washington, DC 20402-0001
   Internet: www.bookstore.gpo.gov
   Telephone: (202) 512-1800
   Fax: (202) 512-2104

2. **The National Technical Information Service**
   5301 Shawnee Road
   Alexandria, VA 22312-0002
   Internet: www.ntis.gov
   1-800-553-6847 or, locally, (703) 605-6000

A single copy of each NRC draft report for comment is available free, to the extent of supply, upon written request as follows:

Address: **U.S. Nuclear Regulatory Commission**
Office of Administration
Digital Communications and Administrative Services Branch
Washington, DC 20555-0001
E-mail: distribution.resource@nrc.gov
Facsimile: (301) 415-2289

Some publications in the NUREG series that are posted at the NRC's Web site address www.nrc.gov/reading-rm/doc-collections/nuregs are updated periodically and may differ from the last printed version. Although references to material found on a Web site bear the date the material was accessed, the material available on the date cited may subsequently be removed from the site.

Non-NRC Reference Material

Documents available from public and special technical libraries include all open literature items, such as books, journal articles, transactions, *Federal Register* notices, Federal and State legislation, and congressional reports. Such documents as theses, dissertations, foreign reports and translations, and non-NRC conference proceedings may be purchased from their sponsoring organization.

Copies of industry codes and standards used in a substantive manner in the NRC regulatory process are maintained at—

The NRC Technical Library
Two White Flint North
11545 Rockville Pike
Rockville, MD 20852-2738

These standards are available in the library for reference use by the public. Codes and standards are usually copyrighted and may be purchased from the originating organization or, if they are American National Standards, from—

American National Standards Institute
11 West 42nd Street
New York, NY 10036-8002
Internet: www.ansi.org
(212) 642-4900

The NUREG series comprises (1) technical and administrative reports and books prepared by the staff (NUREG–XXXX) or agency contractors (NUREG/CR–XXXX), (2) proceedings of conferences (NUREG/CP–XXXX), (3) reports resulting from international agreements (NUREG/IA–XXXX),(4) brochures (NUREG/BR–XXXX), and (5) compilations of legal decisions and orders of the Commission and the Atomic and Safety Licensing Boards and of Directors' decisions under Section 2.206 of the NRC's regulations (NUREG–0750).

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

NUREG/CR-7294
INL/EXT-21-61117

Exploring Advanced Computational Tools and Techniques with Artificial Intelligence and Machine Learning in Operating Nuclear Plants

Manuscript Completed: September 2021
Date Published: February 2022

Prepared by:
Z. Ma[1]
H. Bao[1]
S. Zhang[1]
M. Xian[2]
A. Mack[1]

[1]Idaho National Laboratory
Idaho Falls, Idaho 83415

[2]University of Idaho, Idaho Falls
Idaho Falls, ID 83402

John Lane, NRC Project Manager

Office of Nuclear Regulatory Research

ABSTRACT

This report presents the project Idaho National Laboratory conducted for the Nuclear Regulatory Commission (NRC) to explore the advanced computational tools and techniques, such as artificial intelligence (AI) and machine learning (ML), for operating nuclear plants. The report reviews the nuclear data sources, with the focus on operating experience data, that could be applied by advanced computational tools and techniques. Plant-specific and generic (national and international) data from different sources are described. The report describes the relationships between statistics and AI/ML and then introduces the most widely used AI/ML algorithms in both supervised and unsupervised learning. The report reviews the recent applications of advanced computational tools and techniques in various fields of nuclear industry, such as reactor system design and analysis, plant operation and maintenance, and nuclear safety and risk analysis. The report presents the insights from the project on the potential applicability of AI/ML techniques in improving advanced computational capabilities, how the advanced tools and techniques could contribute to the understanding of safety and risk, and what information would be needed to provide meaningful insights to decision makers.

The report also documents an NRC survey on the current state of commercial nuclear power operations relative to the use of AI and ML tools as well as the role of AI/ML tools in nuclear power operations, which was published by the NRC as in the Federal Register Notice NRC-2021-0048 in April 2021. A summary of the survey, including the survey questions, survey participants, survey responses, and the conclusions and insights derived from the survey, is provided in the report.

Finally, the report investigates potential applications of using AI/ML in operating nuclear power plants and advanced reactors (both advanced light-water reactors and advanced non-light-water reactors) to improve nuclear plant safety and efficiency. Three main application fields are considered: plant safety and security assessments; plant degradation modeling, fault and accident diagnosis and prognosis; and plant operation and maintenance efficiency improvement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

ADAMS	agencywide documents access and management system
ADDM	accident detection, diagnosis and mitigation
AF	application fields
AI	artificial intelligence
ANN	artificial neural network
AP	affinity propagation
BMU	best matching unit or neuron
BN	Bayesian network
BNN	Bayesian neural network
BWR	boiling water reactor
CAP	corrective action program
CART	classification and regression tree
CCF	common-cause failure
CNN	convolutional neural network
DNN	deep neural network
DOE	Department of Energy
DT	decision tree
EPRI	Electric Power Research Institute
FDDP	fault detection, diagnosis and prognosis
FNN	feedforward neural network
FPL	Florida Power & Light Company
FRN	Federal Register Notice
GA	generic algorithm
GP	Gaussian process
HC	hierarchical clustering
HRA	human reliability analysis
I&C	instrumentation and control
IAEA	International Atomic Energy Agency
IE	initiating event
IEI	Insight Enterprises, Inc.
INL	Idaho National Laboratory
INPO	Institute of Nuclear Power Operations
IRIS	Industry Reliability and Information System
JRC	European Committee Joint Research Centre
LER	licensee event report
LLE	local linear embedding
LOOP	loss-of-offsite-power
LSTM	Long Short-Term Memory
LWR	light water reactor

ML	machine learning
NDE	non-destructive examination
NEA	Nuclear Energy Agency
NEI	Nuclear Energy Institute
NLP	natural language processing
NLWR	non-light-water reactor
NN	neural network
NPP	nuclear power plant
NRC	Nuclear Regulatory Commission
OECD	Organisation for Economic Co-operation and Development
OpE	operating experience
PCA	principal component analysis
PRA	probabilistic risk assessment
PSA	probabilistic safety assessment
PWR	pressurized-water reactor
RAVEN	Risk Analysis and Virtual Environment
RCA	root cause analysis
RF	random forest
RNN	recurrent neural network
RUL	remaining useful life
SRI	Southern Research Institute
SSC	structures, systems and components
SVM	support vector machine
WANO	World Association of Nuclear Operators
WEC	Westinghouse Electric Company LLC

1 INTRODUCTION

1.1 Background

Idaho National Laboratory (INL) has provided technical assistance to the Nuclear Regulatory Commission (NRC) Division of Risk Analysis in the Office of Nuclear Regulatory Research in the areas of reliability and risk analysis since the 1980s. INL developed an integrated coding system that captures the necessary information from the nuclear industry operating experience (OpE) to update and maintain industry and plant-specific system and component reliabilities, initiating event frequencies, common-cause failure (CCF) parameter estimates and to conduct component and system trending analysis. The periodic analysis, calculation, and updates to these parameters used in the standardized plant analysis risk models are based upon the statistical methods developed in NUREG/CR-6823 (Atwood et al. 2003) and NUREG/CR-6928 (Eide et al. 2007), which included a portion of all available statistics methods, were first published in the early 2000s.

In the meantime, artificial intelligence (AI), machine learning (ML), Big Data, content analytics, and other advanced computational tools and techniques have shown promise as being beneficial across several disciplines and for a variety of applications in both the private and public sectors. The development and use of these information technologies is becoming more widespread among industry organizations, academic institutions, and federal agencies to help improve their efficiency, effectiveness, and decision-making. The nuclear industry has also been investigating the adoption of such technologies to improve operational efficiency at nuclear power plants (NPPs).

Some NPP licensees are in the process of demonstrating new approaches (e.g., NEI 18-10 (Nuclear Energy Institute 2018)) for meeting regulatory requirements in Title 10 of the Code of Federal Regulations (10 CFR) 50.65, “Requirements for Monitoring the Effectiveness of Maintenance at Nuclear Power Plants.” The new approach in NEI 18-10 is a departure from the current preventative maintenance assessment paradigm (e.g., establishing structures, systems and components [SSC] performance criteria) and is intended to allow for a more dynamic assessment of maintenance effectiveness based on the use of data and risk trending analytics. As a result, however, the licensees have also opted to discontinue use of the NRC-endorsed approach in NUMARC 93-01 (Nuclear Energy Institute 2011) for meeting requirements in 10 CFR 50.65. As such, NRC resident inspectors are tasked with understanding the underlying technologies employed in these new approaches (e.g., AI, ML, and data analytical tools) to ensure the adequate inspection of the licensee’s ability to meet the requirements in 10 CFR 50.65.

A new project was conducted by INL for the NRC to explore the advanced computational tools and techniques, such as AI, ML, and other analytics, in operating NPPs and developing advanced computational predictive capabilities in nuclear OpE. The project has three major tasks:

- Task 1: Explore the advanced computational tools and techniques for operating nuclear plants.
- Task 2: Assess the use of advanced computational tools and techniques in the commercial nuclear industry.
- Task 3: Explore the potential applications and impact of advanced computational tools and techniques on operating NPPs and advanced reactors.

The purpose of Task 1 is to perform an assessment of the advanced computational tools and techniques to address the following questions: What types of advanced computational tools and techniques may be employed, how would they work, and how effective would they be expected to be? What aspects of the advanced tools and techniques could contribute to our increased understanding of safety and risk? What types and quantities of information would be needed, in concert with the new tools and techniques, to generate safety and risk implications? The purpose of Task 2 is to perform a survey assessment of the state of practice and future trends related to the advanced computational tools and techniques in advancing the state-of-the-art in predictive reliability and predictive safety assessments in the commercial nuclear industry. The purpose of Task 3 is to investigate the potential applications of the advanced computational tools and techniques, including AI, ML, big data, and content analytics to operating NPPs and advanced reactors, including advanced light water reactors (LWRs) and advanced non-light-water reactors (NLWRs).

This report documents the results for all three project tasks by assessing the advanced computational tools and techniques that may be applied to nuclear OpE (Task 1), conducting a survey to request public comments on the current state of commercial nuclear power operations relative to the use of AI and ML tools (Task 2), and investigating potential applications of using AI/ML in operating NPPs, advanced LWRs, and advanced NLWRs (Task 3).

1.2 Outline

Since one of the major factors for a successful application of advanced computational tools and techniques is the data availability and quality, this report first looks at the nuclear data that may be available and could be used in advanced computational tools and techniques. Section 2 presents a categorization of nuclear data sources and focuses on different types of OpE data that may be applied through advanced computational tools and techniques. Section 3 presents an overview of advanced computational tools and techniques. It first describes the relationships between statistics and AI/ML and then introduces the most widely used AI/ML algorithms in both supervised and unsupervised learning. Section 4 reviews the existing applications of advanced computational tools and techniques, including AI/ML in various nuclear industry fields, such as reactor system design and analysis, plant operation and maintenance, and nuclear safety and risk analysis. Section 5 provides insights for the three questions under Task 1. Section 6 presents the survey on the role of AI tools in U.S. commercial nuclear power operations responses, including the survey questions, survey participants, survey responses, and the conclusions and insights derived from the survey. Section 7 investigates the potential applicability of the new computational tools and techniques with AI/ML to inform and simplify the regulatory process on the operating NPPs and advanced reactors while simultaneously improving plant safety and efficiency and enhancing regulatory oversight. This report provides details of these three main technological application fields (AFs) in Sections 7.1, 7.2, and 7.3, respectively. Section 8 provides the report's conclusion. Section 9 lists the references cited in the report. Appendix A provides a list of recent applications of advanced computational tools and techniques in the nuclear industry.

2 NUCLEAR DATA OVERVIEW

This section provides an overview of data sources in NPPs. The scope is focused on data sources for commercial NPPs but does not exclude data sources in other nuclear installations or nonnuclear industries, which are potentially relevant to or useful for building up advanced computational capabilities for NPPs. Section 2.1 presents a categorization of NPP data sources. Section 2.2 further categorizes the NPP OpE data. Section 2.3 introduces the characteristics of NPP OpE data sources. Section 2.4 discusses the relevancies of NPP OpE data to probabilistic risk assessment (PRA).

2.1 Nuclear Data Sources

Nuclear data sources can be categorized in different ways. For example, (Atwood et al. 2003) utilizes two types of data sources, plant-specific and generic, to produce various parameter estimates that are needed in a PRA, and (Al Rashdan et al. 2019, Al Rashdan and St. Germain 2019) categorize fifteen typical NPP data sources based on their data-collection methods. In this report, nuclear data sources are categorized into observed data and synthetic data, while observed data includes OpE data and experimental data, and synthetic data includes simulated data (see Figure 1). Both observed data and synthetic data can be anonymized such as by removing sensitive information to protect data source privacy and confidentiality.

1. **OpE data** observed and harvested as NPPs operate (including during the maintenance).
2. **Experimental data** produced by lab or field experiments. Experimental data and OpE data could overlap if an experiment is conducted as part of plant operations, such as surveillance testing.
3. **Synthetic/simulated data**, which are artificially generated from running computational models to simulate processes or systems using computer programs (such as physics simulation codes) or digital twins (such as plant simulators).

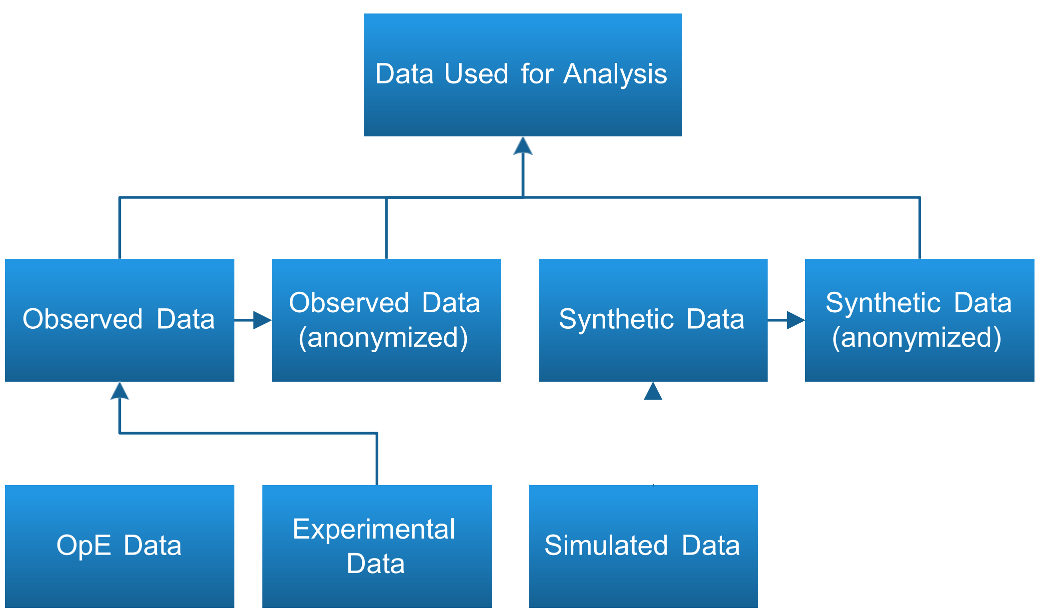

Figure 1 Relationships Between Observed Data and Synthetic/simulated Data

It should be noted that "data" and "information" are different concepts in a strict sense, as shown in the following definitions from Merriam-Webster (Merriam-Webster) and Kelly and Smith (Kelly and Smith 2011). "Data" is the basic, unrefined, and generally observable information, while "information" is the processed, more refined, and often inferred data.

- Data—"factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation"
- Information—"knowledge obtained from investigation, study, or instruction."

To simplify the terms, this report uses "data" in a more general sense in that it includes "information," such as the Licensee Event Report (LER) submitted by licensees to the NRC or analysis report.

Simulated data and experimental data are widely used in nuclear industry and academic research in fields, such as reactor system design and analysis and plant operations (see Section 4). These data sources are not introduced further in this section. This section is focused on introducing characteristics of OpE data.

2.2 OpE Data

Nuclear OpE data can be divided into plant-specific, generic (national), and generic (international) data, which can be further categorized according to data collection scopes, relevant activities, or collecting countries and organizations (see Figure 2). Each subcategory of OpE data is described in the following subsections.

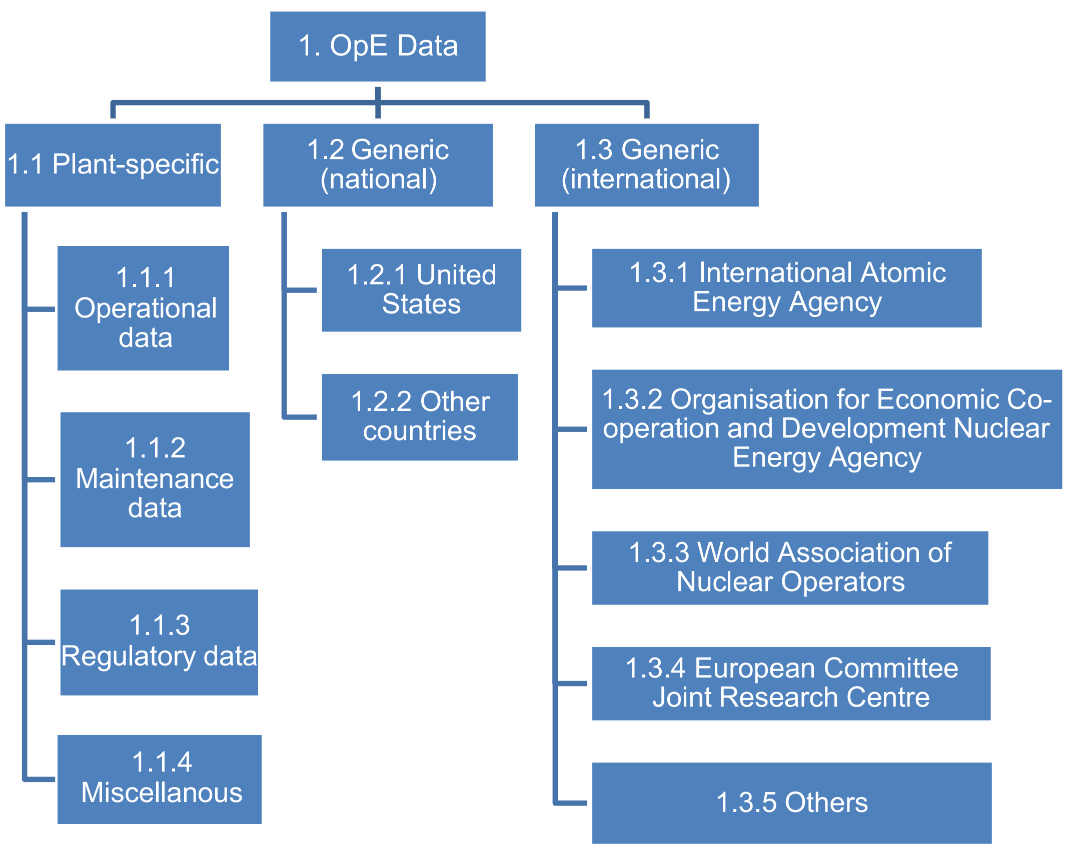

Figure 2 Subcategories of NPP OpE data

2.2.1 Plant-Specific OpE Data (Category 1.1)

The OpE data collected on a plant basis can be broken down according to relevant activities, including operational data (**Category 1.1.1** in Table 1), maintenance data (**Category 1.1.2**), regulatory data (**Category 1.1.3**), and miscellaneous data (**Category 1.1.4**).

Category 1.1.1: Operational data are defined in this report as the data accumulated as a plant operates and are usually proprietary to the plant. Examples of this category include process instrumentation and control (I&C) data, plant logs, and internal plant failure reports.

- Process I&C data are the data collected in real-time from plant-wide sensors for process measurements and monitoring; such data have diverse data formats corresponding to the monitored process variables (including neutron flux, reactor pressure, coolant temperature, steam generator water level, radiation dose, etc.) and are usually quantitative and structured.
- Plant logs refer to all types of operational logs maintained to record all important events in the plant. Examples include control room logs, operator round notes, and engineer notebooks. Different plants could have different types and formats of plant logs; regarding the data types, plant logs usually contain numerical data, categorical data, and narrative data. Plant logs can either be handwritten or electronic and are usually routinely maintained.
- Internal plant failure reports refer to the documentation of equipment failures or human errors for a plant's own use rather than for a regulatory purpose. A single failure report is usually in the form of condition report and then entered into the plant's corrective action program. Like plant logs, failure reports usually contain numerical data, categorical data, and narrative data and can be either handwritten or electronic. Rather than routinely maintained, the preparation of a failure report is conditioned on failure occurrence.

Category 1.1.2: Maintenance data refer to the records (including plans, actions, results, and all relevant documents) of plant maintenance activities. This report adopts "maintenance" to refer to a broad scope of activities—it is not confined to preventive and corrective maintenance, but includes replacement, inspection, calibration, and surveillance testing. Maintenance data could contain numerical, categorical, narrative, or graphical data and could be either handwritten or electronic. The velocity of maintenance data is dependent on the frequency of corresponding activity, either routinely coming in or conditioned on event occurrence (such as corrective maintenance).

Category 1.1.3: Regulatory data are defined in this report as the data that a plant prepares or receives to comply with regulatory requirements. Different countries have different regulations, leading to differences in regulatory data. The characteristics of regulatory data described in this report are based on the NPPs in the United States. Unlike other propriety plant-specific OpE data, regulatory data are usually made publicly available and published on the NRC website. Regulatory data are either submitted by plant licensees to the NRC (such as monthly operating reports and LERs) or issued by the NRC to licensees (such as inspection reports and notices of enforcement discretion). The data velocity depends on the type of submission, either routine submission (daily, quarterly, annually, etc.) or special submission conditioned on event occurrence. Most regulatory data contain a mix of numerical, categorical, checkbox, and narrative data. Some regulatory data are further visualized in graphical forms, such as the reactor oversight process performance indicators presented online in color codes. Regulatory data are mostly electronic reports, while some data are extracted and built into a specific database, such as the NRC LERSearch platform to search for LERs and inspection reports.

Category 1.1.4: Miscellaneous data refer to all other plant-specific OpE data that are not captured by the previous three subcategories. Examples include plant design and license-related documents, plant operating guidance documents (such as technical specifications, procedures, and guidelines), and plant business data (such as data associated with enterprise asset management, procurement and logistics, project scheduling and management, human resources, finance, etc.). These data are usually proprietary to the plant and could have diverse data formats and structures. The data may be stored in maturely-designed commercial databases (like enterprise resource and human resource data) or in the form of digital files or hard copies.

2.2.2 Generic (National) OpE Data (Category 1.2)

Besides on a plant basis, the OpE data can be collected throughout the nuclear power industry. Some OpE data sources, based on country-specific industrywide data collection, are introduced below.

Category 1.2.1: United States has several domestic OpE data sources established by the NRC and other organizations, such as Nuclear Energy Institute (NEI), Institute of Nuclear Power Operations (INPO), Electric Power Research Institute (EPRI), Department of Energy (DOE), and Energy Information Administration. These data sources provide data with a variety of focuses, including OpE feedback, component performance data, human performance data, and general statistical analyses. Most data, although stored in databases, are semi-structured, containing a mix of numerical, categorical, narrative, check box, and graphical data. General statistical analyses are publicly available and updated on a regular basis. The other data (OpE feedback, component performance data, and human performance data) are conditioned on event occurrences.

- OpE feedback includes the publicly available NRC LERSearch that could be used to search LERs.
- Component performance data include the proprietary INPO Industry Reliability and Information System (IRIS) database, the NRC Integrated Data Collection and Coding System database, NRC Reactor Operating Experience Data web app for nuclear event searching, NRC Reliability and Availability Data System web app for PRA data calculations, NRC initiating event (IE) database, NRC loss of offsite power (LOOP) database, and NRC CCF database. The raw data and other details of the above sources are proprietary, while the generic results of the data analysis are part of the general statistical analyses and are publicly available on the NRC websites.
  Component performance data also include general statistical analyses (i.e., “information” rather than “data”) such as (Eide, 2003) and its updates on industry-average performance for components and initiating events, the NRC annual LOOP analysis and IE analysis, the NRC component performance studies, the NRC system reliability studies, the DOE generic component failure data base for light water and liquid-sodium reactor PRAs, and EPRI reports on pipe rupture frequencies, components, and shutdown accident events.
- Human performance data are publicly available on the NRC website and via NUREG reports.

Category 1.2.2: Other countries have established their domestic databases for OpE feedback and have made them publicly available (at least the generic results). Examples include the Nuclear Power Plant Event Reporting run by the Canadian Nuclear Safety Administration, the Experience Feedback Platform run by Chinese National Nuclear Safety Administration, the Nuclear Event Evaluation Database Incident Reporting System run by the Korean Institute of Nuclear Safety, and the Nuclear Events Databases curated by the Swiss ETH Zurich. The data in these OpE feedback databases are semi-structured data, which are a mix of numerical, categorical, and narrative data. Besides OpE feedback, there are nationwide databases for component performance data, such as the Swedish Reliability Data of Components in Nordic Nuclear Power Plants (the T-book), that are not publicly available but open for purchase.

2.2.3 Generic (International) OpE Data (Category 1.3)

Nuclear OpE data are also collected by international organizations, such as the International Atomic Energy Agency (IAEA), Organisation for Economic Co-operation and Development (OECD) Nuclear Energy Agency (NEA), World Association of Nuclear Operators (WANO), and European Committee Joint Research Centre (JRC). Some OpE data sources maintained by different organizations are introduced below.

It should be noted that, although these data sources are collected by international organizations, a country's participation is on a voluntarily basis. The scopes of data collection are determined by the participants of certain organizations (such as the IAEA and the OECD NEA) or countries in certain regions (such as Nordic countries).

The data sources are focused on providing OpE feedback, component performance, and plant performance data. OpE feedback data sources have different focuses on commercial nuclear power plants, research reactors, and fuels. Component performance data sources include data for generic NPP components as well as data with specific topics, such as aging, CCF, cable, and piping comments. Most of these data sources are available to participants only.

Table 1 shows the categories of the nuclear OpE data sources as well as the examples in different categories. The examples are focused on active data sources. Inactive data sources with no update in 10 years or longer, such as the Nuclear Plant Reliability Data System, Nuclear Computerized Library for Assessing Reactor Reliability, and IAEA Component Reliability Data for Use in Probabilistic Safety Assessment (PSA), are not listed.

Table 1 NPP OpE Data Source Subcategories

NPP OpE Data Source Subcategory		
Data Collection Scope	Activity/Country/Organization	Example of Data Sources
1.1 Plant-specific data	1.1.1 Operational data	Process I&C data Plant logs Internal plant failure reports (for licensee’s own use) Condition reports Correction action programs
	1.1.2 Maintenance data	Maintenance and replacement records Inspection, calibration, and surveillance test records
	1.1.3 Regulatory data (applicable to United States NPPs)	Submitted by licensees to NRC and other regulatory bodies Routine submissions Daily power reactor status reports Monthly operating reports Quarterly operating reports Quarterly reactor oversight process performance indicators Annual radioactive effluent reports Annual environmental reports Special submissions Event notification reports LERs Issued by regulatory bodies to licensees Routine issuances Quarterly inspection findings and reports Special issuances Special inspection findings and reports Preliminary notification reports Significant enforcement actions Notices of enforcement discretion

Table 1 NPP OpE Data Source Subcategories (continued)

NPP OpE Data Source Subcategory		
Data Collection Scope	Activity/Country/Organization	Example of Data Sources
	1.1.4 Miscellaneous	Plant design and license-related documents Plant operating guidance documents Technical specifications Procedures and guidelines Plant business data
1.2 Generic (national) data	1.2.1 United States	OpE feedback NRC LERSearch Component performance data Raw data INPO IRIS database (proprietary) NRC IDCCS database NRC NROD web app for nuclear event searching NRC RADS web app for PRA data calculations NRC IE database NRC LOOP database NRC CCF database General statistical analyses NUREG/CR-6928 (Eide et al. 2007) and its updates on industry-average performance for components and initiating events NRC annual LOOP analysis and IE analysis NRC component performance studies NRC system reliability studies DOE generic component-failure data base for light water and liquid-sodium reactor PRAs EPRI reports on pipe rupture frequencies, components, shutdown accident events (proprietary) Human performance data NRC human factors information system

Table 1 NPP OpE Data Source Subcategories (continued)

NPP OpE Data Source Subcategory		
Data Collection Scope	Activity/Country/Organization	Example of Data Sources
	1.2.2 Other countries	OpE feedback Canadian Nuclear Safety Administration nuclear power plant event reporting Chinese National Nuclear Safety Administration experience feedback platform Korean nuclear event evaluation database incident reporting system Swiss ETH Zurich-curated nuclear events database Component performance data Swedish Reliability data of components in Nordic nuclear power plants (T-Book)
Generic (international) data	1.3.1 International Atomic Energy Agency (IAEA)	OpE feedback International nuclear and radiological event scale Incident Reporting System for research reactors Component performance data Reliability data for research reactor probabilistic safety assessment
	1.3.2 Organisation for Economic Co-operation and Development/Nuclear Energy Agency (OECD NEA)	OpE feedback Fire incidents records exchange project Component performance data International common-cause data exchange project Component operational experience, degradation & aging program Cable aging data and knowledge project
	1.3.3 World Association of Nuclear Operators (WANO)	Plant performance data Performance analysis program
	1.3.4 European Committee Joint Research Centre (JRC)	OpE feedback Clearinghouse on operating experience feedback database

Table 1 NPP OpE Data Source Subcategories (continued)

NPP OpE Data Source Subcategory		
Data Collection Scope	Activity/Country/Organization	Example of Data Sources
	1.3.5 Others	OpE feedback International reporting system for operating experience jointly managed by IAEA and OECD NEA Fuel incident notification and analysis system jointly managed by IAEA and OECD NEA Component performance data Centralized reliability and events database, containing events collected from nuclear power plants in Germany, Netherlands, and Switzerland Nordic reliability data of piping components in Nordic nuclear power plants (R-book) Nordic/German CCF reliability book (C-book)

2.3 Characteristics of NPP OpE Data Sources

This subsection introduces characteristics of each NPP OpE data source. Selected characteristics include:

- **Data format,** such as numerical (discrete and continuous), categorical (including binary data as a special case), check box, narrative (or free text), image, symbol, audio, and video.
- **File format,** such as digital files and written notes. This report lists file format and data format separately since both could affect the selection of applicable AI/ML techniques. For instance, there are two files containing the same set of tabular numerical data; however, one file is handwritten, and the other is its digitized version. Processing the handwritten file, although containing structured data, will require more advanced techniques with handwriting recognition functions when compared to the digitized file.
- **Data structure,** including structured and unstructured. Structured data is organized, formatted, and easily searchable, while unstructured data has no predefined format and is much more difficult to collect, process, and analyze. It is also possible that a data source contains both structured data and unstructured data; in this case, this data source will be deemed semi-structured in this report.
- **Data velocity,** which specifies how frequent the data are comes in. Data could arrive in real-time like from sensors, routinely (daily or monthly) or conditioned on the occurrences of certain events. It should be noted that some data sources take in and process raw data; under these cases, this report separates the data velocity as a (raw) data-sampling velocity and data-processing velocity.
- **Data accessibility,** which specifies who have access to the data. A data source can be publicly available or proprietary and only accessible for authorized users.
- **Relevancy to PRA,** which specifies if the data could be used to support PRA. Formulated in 1970s, PRA is a well-established technique to systematically develop accident scenarios and generate probabilistic, system- and plant-level risk estimates. This report defines four levels of relevancy to PRA, including (1) direct relevancy, (2) indirect relevancy, (3) potential relevancy, and (4) no relevancy.
    - Direct relevancy is defined as providing the information needed to construct a PRA model (such as system design) or providing a PRA model parameter estimates (such as component-failure probabilities).
    - Indirect relevancy is defined as providing raw data for estimating PRA model parameters (such as component-failure events occurred in individual plants).
    - Potential relevancy is defined as not directly or indirectly supporting the current practice of PRAs but having a possibility to be connected to PRAs in the future as modeling techniques advance. One potential case is that the current PRA practice includes only component-failure events, but if PRA modeling can be expanded to include component degradation events, a lot more data sources could then be utilized.
    - No relevancy is defined as not directly or indirectly supporting the current practice of PRAs and is not projected to be connected to PRAs in the future.

Table 2 presents the characteristics of plant-specific OpE data sources. Table 3 shows the characteristics of generic (national) OpE data sources. Table 4 presents the characteristics of generic (international) OpE data sources.

Table 2 Characteristics of Plant-Specific OpE Data Sources

Data Source	File Format	Data Format	Data Structure	Data Velocity	Data Accessibility	Relevancy to PRA
Process I&C data	Digital files	Numerical	Structured	Real-time	Proprietary	Potential relevancy
Plant logs	Digital files Written notes	Numerical Categorical Narrative	Unstructured	Routine	Proprietary	Indirect relevancy (providing raw data for PRA parameter estimation)
Internal plant failure reports	Digital files	Numerical Categorical Narrative	Unstructured	Conditioned on failure occurrence	Proprietary	Indirect relevancy (providing raw data for PRA parameter estimation)
Maintenance and replacement records	Digital files Written notes	Numerical Categorical Narrative	Unstructured	Conditioned on activity frequencies or event occurrences	Proprietary	Indirect relevancy (providing raw data for PRA parameter estimation)
Inspection, calibration, and surveillance test records	Digital files Written notes	Sensor data Numerical Categorical Narrative Graphical	Unstructured	Conditioned on activity frequencies	Proprietary	Indirect relevancy (providing raw data for PRA parameter estimation)
Regulatory data submitted by licensees to regulatory bodies	Database Digital files	Numerical Categorical Narrative Graphical	Unstructured	Routine or conditioned on event occurrence	Publicly available	Indirect relevancy (providing raw data for PRA parameter estimation)
Regulatory data issued by regulatory bodies to licensees	Database Digital files	Numerical Categorical Narrative Graphical	Unstructured	Routine or conditioned on event occurrence	Publicly available	Indirect relevancy (providing raw data for PRA parameter estimation)
Plant design and license-related documents	Digital files Hard copies	Unknown	Unstructured	Updated as needed	Proprietary	Direct relevancy (providing information for PRA model development)

Table 2 Characteristics of Plant-Specific OpE Data Sources (continued)

Data Source Process I&C data	**File Format** Digital files	**Data Format** Numerical	**Data Structure** Structured	**Data Velocity** Real-time	**Data Accessibility** Proprietary	**Relevancy to PRA** Potential relevancy
Plant operating guidance documents	Digital files Hard copies	Narrative	Unstructured	Updated as needed	Proprietary	Direct relevancy (providing information for PRA model development)
Plant business data	Database Digital files	Numerical Narrative	Structured or Unstructured	Updated as needed	Proprietary	No relevancy

Table 3 Characteristics of Generic (National) OpE Data Sources

Data Source	File Format	Data Format	Data Structure	Data Velocity	Data Accessibility	Relevancy to PRA
NRC LERSearch	Database Digital files	Numerical Categorical Checkbox Narrative	Unstructured	Conditioned on event occurrence	Publicly available	Indirect relevancy (providing raw data for PRA parameter estimation)
U.S. component performance database (e.g., INPO IRIS, NRC IDCCS, NROD, RADS) and general statistical analyses	Database	Numerical Categorical Narrative	Structured (for database) or semi-structured (for analysis results)	Routinely updated	Generic results publicly available; raw data proprietary	Direct relevancy (providing raw data and PRA parameter estimates) IE frequencies Component unreliability Component unavailability CCF parameters Component studies System studies
NRC human factors information system	Database Digital files	Numerical Categorical	Structured	Conditioned on event occurrence	Publicly available	Indirect relevancy (providing raw data for PRA parameter estimation)
Canadian Nuclear Safety Administration nuclear power plant event reporting	Database	Numerical Categorical Narrative	Semi-structured	Conditioned on event occurrence	Publicly available	Indirect relevancy (providing raw data for PRA parameter estimation)
Chinese National Nuclear Safety Administration experience feedback platform	Database	Numerical Categorical Narrative	Semi-structured	Conditioned on event occurrence	Generic results publicly available; raw data proprietary	Indirect relevancy (providing raw data for PRA parameter estimation)
Korean nuclear event evaluation database incident reporting system	Database	Numerical Categorical Narrative Graphical	Semi-structured	Conditioned on event occurrence	Publicly available	Indirect relevancy (providing raw data for PRA parameter estimation)

Table 3 Characteristics of Generic (National) OpE Data Sources (continued)

Data Source	File Format	Data Format	Data Structure	Data Velocity	Data Accessibility	Relevancy to PRA
Swiss ETH Zurich-curated nuclear events database	Database	Numerical Categorical Narrative	Semi-structured	Conditioned on event occurrence	Publicly available	Indirect relevancy (providing raw data for PRA parameter estimation)
Swedish T-book	Handbook	Numerical Categorical Narrative	Semi-structured	Raw data conditioned on event occurrence	Available for purchase	Direct relevancy (providing PRA parameter estimates) Component unreliability

Table 4 Characteristics of Generic (International) OpE Data Sources

Data Source	File Format	Data Format	Data Structure	Data Velocity	Data Accessibility	Relevancy to PRA
IAEA international nuclear and radiological event scale	Database	Numerical Categorical Narrative	Semi-structured	Raw data collection conditioned on event occurrence	Database available to participants; events rated at Level 2 posted publicly	Indirect relevancy to PRA (providing raw data for PRA parameter estimation)
IAEA Incident Reporting System for research reactors	Database	Numerical Categorical Narrative	Semi-structured	Ra data collection conditioned on event occurrence	Available to participants	Indirect relevancy to PRA (providing raw data for PRA parameter estimation)
IAEA reliability data for research reactor probabilistic safety assessment	Digital file	Numerical Categorical Narrative	Semi-structured	Raw data collection conditioned on event occurrence	Publicly available	Direct relevancy to PRA (providing PRA parameter estimates) Component failure probabilities Component failure rates
OECD NEA fire incidents records exchange project	Database	Numerical Categorical Narrative Graphical	Semi-structured	Raw data collection conditioned on event occurrence; database updated annually	Available to participants	Direct relevancy to PRA (providing PRA parameter estimates) Fire event frequencies
OECD NEA international common- cause data exchange project	Database	Numerical Categorical Narrative	Semi-structured	Raw data collection conditioned on event occurrence	Database available to participants; topical reports publicly available	Direct relevancy to PRA (providing PRA parameter estimates) CCF parameters
OECD NEA component operational experience, degradation & ageing programme	Database	Unknown	Unknown	Raw data collection conditioned on event occurrence	Available to participants	Indirect relevancy to PRA (providing raw data for PRA parameter estimation)

Table 4 Characteristics of Generic (International) OpE Data Sources (continued)

Data Source	File Format	Data Format	Data Structure	Data Velocity	Data Accessibility	Relevancy to PRA
OECD NEA cable ageing data and knowledge project	Database	Numerical Categorical Narrative	Semi-structured	Raw data collection conditioned on event occurrence	Database available to participants; topical reports publicly available	Potential relevancy to PRA
WANO performance analysis programme	Digital files	Unknown	Unknown	Raw data collection conditioned on event occurrence	Available to WANO members	Indirect relevancy to PRA (providing raw data for PRA parameter estimation) Event reports Performance insight reports Performance indicator (safety system performance indicator) Potential relevancy to PRA Performance indicators (fuel reliability) No relevancy to PRA Performance indicators (e.g., forced loss rate; collective radiation exposure; industrial safety accident rate; unplanned total scrams per 7,000 hours critical)

Table 4 Characteristics of Generic (International) OpE Data Sources (continued)

Data Source	File Format	Data Format	Data Structure	Data Velocity	Data Accessibility	Relevancy to PRA
European Committee JRC clearinghouse on operating experience feedback database	Digital files	Numerical Categorical Narrative	Semi-structured	Raw data collection conditioned on event occurrence; newsletters published quarterly	Publicly available	Indirect relevancy to PRA (providing raw data for PRA parameter estimation)
IAEA international reporting system for operating experience	Database	Numerical Categorical Narrative	Semi-structured	Raw data collection conditioned on event occurrence; summary report published every three years	Database available to participants; summary reports publicly available	Indirect relevancy to PRA (providing raw data for PRA parameter estimation)
IAEA fuel incident notification and analysis system	Database	Numerical Categorical Narrative	Semi-structured	Raw data collection conditioned on event occurrence	Available to participants	Indirect relevancy to PRA (providing raw data for PRA parameter estimation)
Centralized Reliability and Events Database containing events collected from nuclear power plants in Germany, Netherlands, and Switzerland	Database	Numerical Categorical Narrative	Semi-structured	Raw data collection conditioned on event occurrence	Generic results available for purchase; raw data proprietary	Direct relevancy to PRA (providing PRA parameter estimates) Component failure rates Component failure probabilities

Table 4 Characteristics of Generic (International) OpE Data Sources (continued)

Data Source	File Format	Data Format	Data Structure	Data Velocity	Data Accessibility	Relevancy to PRA
Nordic R-book	Handbook	Numerical Categorical Narrative	Semi-structured	Raw data conditioned on event occurrence	Access possible via the Nordic PSA Group	Direct relevancy (providing PRA parameter estimates) Piping reliability parameters
Nordic/German C-book	Handbook	Numerical Categorical Narrative	Semi-structured	Raw data conditioned on event occurrence	Access possible via the Nordic PSA Group	Direct relevancy (providing PRA parameter estimates) CCF parameters

2.4 Relevancies of OpE Data to PRA

This subsection discusses how the OpE data sources relate to PRA. As introduced in Section 2.3, this report defines four levels of relevancy to PRA, including (1) direct relevancy, (2) indirect relevancy, (3) potential relevancy, and (4) no relevancy.

The data sources with direct relevancy to PRA either provide information for PRA model development, such as plant design and license-related documents and plant operating guidance documents, or provide estimates of PRA parameters, such as IE frequencies (provided by the NRC IDCCS and RADS), component reliability and unavailable data (provided by the NRC IDCCS and RADS, the Swedish T-book, the IAEA reliability data for research reactors, the Centralized Reliability and Events Databases run by Germany/Netherlands/Switzerland, and the Nordic R-book), CCF parameters (provided by the NRC IDCCS and RADS, the OECD NEA international CCF data exchange project, and the Nordic/German C-book), and hazard occurrence frequencies (provided by the OECD NEA fire incidents records exchange project).

The data sources with indirect relevancy to PRA share the feature of providing raw data for PRA parameter estimation from plant logs, failure reports, maintenance records, regulatory data, and the national and international OpE feedback platforms.

The data sources with potential relevancy to PRA refer to those not directly or indirectly supporting the current practice of PRA but having a possibility to be connected to PRA in the future as modeling techniques advance. Three data sources are characterized as having a potential relevancy to PRAs, including the process I&C data, the OECD NEA cable aging data and knowledge project, and the WANO performance analysis program (the fuel reliability part). There are two angles of PRA advancement to facilitate the potential incorporation of these three data sources. On one hand, the PRA modeling scope in current practice only includes component-failure events; if the PRA modeling scope could be expanded to include events representing component degradation (or “unhealthy”) states, more data sources might be utilized, such as the process I&C data and the data from the OECD NEA cable ageing data and knowledge project. On the other hand, the current PRA modeling scope includes limited components; if the scope could be expanded to include more “micro-level” components, more data sources might be adopted, such as the fuel reliability data provided by the WANO performance analysis program.

3 AN OVERVIEW OF ADVANCED COMPUTATIONAL TOOLS AND TECHNIQUES - AI/ML

This section presents an overview of advanced computational tools and techniques which couldinclude advanced statistical algorithms, AI/ML algorithms, and relevant hybrid applications suchas physics-informed machine learning. Section 3.1 describes the relationship between statisticsand AI/ML. Section 3.2 introduces the most widely used AI/ML algorithms in both supervised learning and unsupervised learning. Section 3.3 presents how AI/ML approaches could be applied for advanced computational capabilities in nuclear industry provides a list of example use cases and AI/ML algorithms.

3.1 Statistics and AI/ML

The relationship between statistics and ML is a topic that interests many people. Currently the NRC and INL are using the statistical methods in NUREG/CR-6823, which was published in 2003 and includes a minute portion of all available statistical methods, for nuclear OpE analysisto estimate industry-average and plant-specific system and component reliabilities, initiating event frequencies, CCF parameters and to conduct component and system trending analysis. Abrief look at the relationship between statistics and AI/ML would help us to understand why we are exploring advanced computational predictive capabilities using AI/ML in nuclear OpE.

Statistics and ML are closely related in terms of methodological principles but are different in their primary goals: ML concentrates on prediction to identify the best course of actions with no or limited understanding of the underlying mechanism, while statistics have a focus on inferenceby modeling the data generation process to formalize understanding (although statistics can perform predictions as well) (Bzdok et al. 2018). Statistics is a subfield of mathematics while MLis a subfield of computer science and grew out of AI to focus on learning from data. It follows that ML and AI developed with the advancement of computing power. Statistical methods have traditionally been used on smaller data sets, in cases where the entire population of data is not known. Advanced AI methods require much more data than the traditional statistics methods buthave the ability to predict when relationships are more complex. Early ML had emphasis on symbolic representation and knowledge-based learning, such as decision trees and logic formulate. ML started to flourish as a separate field in the 1990s and changed the focus to methods borrowed from statistics and probability theory (Langley 2011). Some methods from ML were adopted and led to a combined field called statistical learning (James et al. 2013).
Some of the AI/ML algorithms described in Section 3.2, e.g., Bayesian Network and GaussianProcess, are also popular approaches in statistics.

Statistical learning methods have traditionally been used and achieved success on smaller datasets, in cases where the entire population of data is not known. Modern AI/ML approachestake advantage of high-performance computing and large datasets and have pushed the learning capacity of models to the next level to solve extremely complex problems, e.g., self- driving, cancer early detection, and smart agriculture. However, many machine learning approaches, e.g., deep learning, sacrifice some degrees of interpretability for predictive power.

3.2 AI/ML Algorithms Overview

AI/ML algorithms can be categorized into supervised learning and unsupervised learning, depending on whether the dataset is labeled and whether training samples are involved. Learning can occur in two ways. First, in supervised learning, the true output values are known. The algorithm then learns the relationship between input variables (features) and the known output (response) variable. In unsupervised learning, the response variable is not known. The algorithm learns patterns in the data to discover groupings, or clusters of data. The two categories of AI/ML algorithms are further broken down in the following sections.

3.2.1 Supervised Learning

Supervised learning implies a training data set contains the observed values of the variable of interest. The observed values can be either categorical (labels), discrete, or continuous. Supervised learning implies the availability of a labeled training dataset that consists of a set of training samples. In its most common form, each data sample pair has an input feature vector and a desired output value. A supervised learning algorithm learns the underlying model (inferred function) between the input and the output using the training set, and the requirement is that the model should be able to generalize from the training set to unseen data samples. A wide collection of supervised learning algorithms is available, each with its strengths and weaknesses. In this section, we will review the most widely used learning algorithms including artificial neural networks (ANNs), Gaussian processes (GPs), Bayesian networks (BNs), support vector machine (SVM), decision trees (DTs), and random forests (RFs).

3.2.1.1 Artificial Neural Networks

ANNs are the most well-known methods in supervised learning and have the capabilities to be applied in broad areas, including regression analysis, classification, data preprocessing, and robotics. ANNs have architectures for both supervised learning and unsupervised learning (e.g., autoencoders), and we discuss supervised ANNs approaches in this section and discuss unsupervised neural networks (NNs) in Section 3.2.2.

An ANN is composed of three types of layers: input, hidden, and output layers. Each layer consists of a set of nodes called neurons. A typical ANN has one input layer, one output layer, and multiple hidden layers. The connections between nodes in different layers are associated with the weights that define the connection strength and are adjusted as learning proceeds. The number of input nodes is decided by the dimensionality of the data samples, and the number of hidden nodes determines the complexity of the model. ANNs are powerful nonlinear function approximators, and the universal approximation theorem (Baldi and Hornik 1989) states that any function may be approximated by a sufficiently large ANN. An ANN with more than three hidden layers is called a deep NN. Recent progress on deep learning has demonstrated that deep NNs can achieve impressive performance for many tasks, such as object recognition (Cireşan et al. 2012), image classification (Ciregan et al. 2012), semantic segmentation (Long et al. 2015), medical applications (Cheng et al. 2016), facial expression recognition (Glauner 2015), and speech recognition (Deng et al. 2013).

Feedforward Networks (Figure 3) represent the most common ANN architectures. Feedforward networks are composed of layers of nodes, where a weighted output from one layer is the net input to the next layer. Nonlinear activation functions of nodes in the hidden layers enable a NN to fit nonlinear relationships between features and output variables. Backpropagation (Rumelhart et al. 1986) based gradient decent approaches are used to learn the network weights to minimize the error between the prediction and true label of the input data. Unsupervised pre-training (Hinton et al. 2006) and increased computing power from graphics processing units allowed the use of deep NNs, which became known as deep learning (Goodfellow et al. 2016).

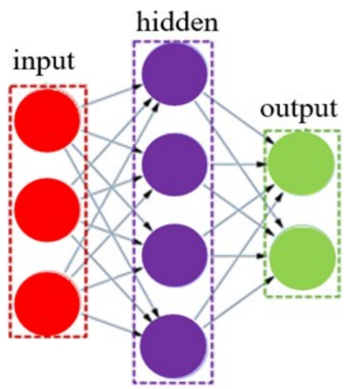

Figure 3 A three-layer feedforward ANN

There are three main breakthroughs in ANNs history: (1) Convolutional Neural Networks (CNNs) that have proven successful in processing visual and other two-dimensional data (LeCun et al. 1989); (2) Long Short-Term Memory (LSTM)-based Recurrent Neural Networks (RNNs) avoid the vanishing gradient problem (Goodfellow et al. 2014); and (3) competitive networks, such as Generative Adversarial Networks (GANs) (Hubel and Wiesel 1968), in which multiple networks compete with each other.

CNNs are feedforward networks and are based on a shared-weight architecture using a number of small kernels and filters. They were inspired by the organization of the animal visual cortex (Hubel and Wiesel 1968, Matsugu et al. 2003) in that the biological neurons respond to stimuli only in a restricted region of the visual field. CNNs have achieved great success in image and video recognition, image classification, medical image analysis, natural language processing (NLP), and time series analysis. Popular CNN architectures include AlexNet (Krizhevsky et al. 2017), VGG-16 (Simonyan and Zisserman 2014), FCN (Long et al. 2015), GoogLeNet (Szegedy et al. 2015), ResNet (He et al. 2016), and DenseNet (Huang et al. 2017). One of the major advantages is that CNNs are more independent from prior knowledge and human effort in feature design.

RNNs are a class of ANNs that allow modeling of the dynamic behavior of time series data. RNN weights are also learned using backpropagation-based optimization algorithms. RNNs have been quite successful for NLP and speech recognition. NLP has been used in applications such as email/web autocorrect grammar, language translation, aircraft maintenance by synthesizing information from large manuals, and to identify motives in actions based on speech. Applications in speech recognition include voice to text, voice control of computer-based technology, and personal identification based on voice, among others.

The effectiveness of early RNNs has been hindered by the problems of gradient vanishing or exploding. The development of the LSTM algorithms (Hochreiter and Schmidhuber 1996) renewed interest in RNNs. An LSTM unit consists of a cell, an input gate, an output gate, and a forget gate. The cell remembers values over arbitrary time intervals, and the three gates regulate the flow of information into and out of the cell. LSTM alleviates the problems with gradients and the transmission of long-term information from which standard RNNs suffer.

3.2.1.2 Gaussian Processes

A GP is a probabilistic method that can be applied to both regression and classification tasks. GPs aim to find a probability distribution over all possible functions between the features and responses. The most important advantage of GPs is the incorporation of the confidence of the prediction into the result, and one can decide based on the confidence intervals if the refitting is needed for some region of interest. However, GP models use the whole dataset to perform prediction and often scale poorly as the amount of data increases.

Figure 4 shows a GP example. The blue squares are eight training samples from a sine function. The red dashed curve shows the mean output/predictions of the test data from -5 to 5. The pink shaded region shows the confidence for the predictions.

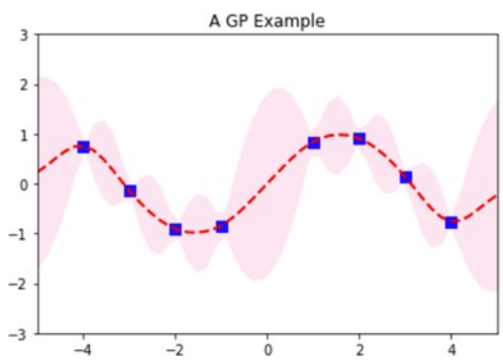

Figure 4 A Gaussian process example

3.2.1.3 Bayesian Networks

A BN or Bayesian belief network (BNN) is a graphical model that captures the known probabilistic relationship using a directed acyclic graph. BNs are ideal for predicting the likelihood of any possible cause of an event that occurred. For example, a BN could represent the joint distribution between features and class labels. Given features, the BN can be used to compute the probabilities of possible class labels (classification). BN classifiers are special BNs designed for classification problems and offer the benefit of explainability. Naïve Bayes (NB) is a special BN with strong independence assumptions between features (see Figure 5). In NB classifiers, the class label variable will be the parent of all feature variables, and the joint distribution is

$$p(l, x) = p(l) \prod_{i=1}^{n} p(x_i|l)$$

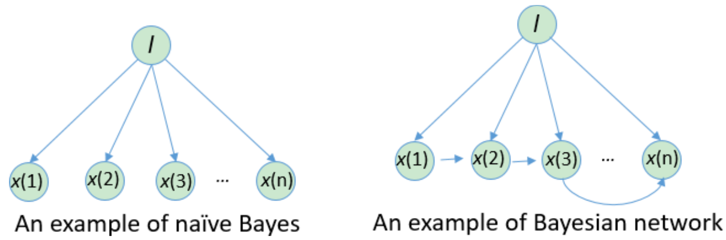

Figure 5 Examples of Bayesian Networks

BNs have two major advantages. First, because a BN encodes dependencies among all random variables, it can handle the missing data problem. Second, a BN can learn causal relationships, and can be used to gain understanding about the data and a problem.

3.2.1.4 Decision Trees and Random Forests

Decision Trees aim to create a tree model that predicts the target value by learning simple decision rules inferred from the data features. In the tree structures, leaves represent target values, internal nodes are labeled with features, and branches (from root to a leaf) represent the conjunction of features that lead to the target values. The most common strategy used to build DTs from data is a top-down greedy approach (Quinlan, 1986) which recursively splits the source dataset into subsets by choosing a feature at each step that give best partitions. Different algorithms use different quantitative metrics to measure the best partition, similar to the loss function choice. Common metrics includes the Gini impurity [Z], information gain [Y, AA], and variance reduction (Breiman et al. 1984). They can be used for both regression and classification tasks. Notable algorithms are the Classification and Regression Tree (CART) (Breiman et al. 1984), Iterative Dichotomiser 3 (ID3) (Quinlan 1986), C4.5 (Quinlan 1993), Chi-square automatic interaction detection (CHAID) (Kass 1980), and MARS (Friedman 1991).

DTs are among the most popular ML algorithms and are also a common statistical approach. DTs are popular because (1) they are simple and can visually and explicitly represent decisions and decision-making processes; (2) they can handle both numerical and categorical data; (3) they make no assumptions of the training data (e.g., distributional and model assumptions); and (4) the hierarchy of features in a DT reflects the importance of features. The features on top are the most informative. However, DTs can create over-complex trees that lead to the overfitting problem.

Random Forests (Tin Kam 1998, Breiman 2001) were proposed to overcome the weakness of the overfitting problem of DTs. Overfitting occurs when an algorithm performs well on the data set used to build the model but performs poorly when applied to new data sets. They construct a number of DTs and output the class label for classification or average prediction values of the individual trees for regression. RFs build each tree using a randomly drawn subset (with a replacement) from the training set. When splitting each node during a tree construction, the best split is found from a random subset of features. The injection of two sources of randomness was designed to reduce the variance of the DTs and help RFs outperform DTs.

3.2.1.5 Support Vector Machines

SVMs (Cortes and Vapnik 1995) aim to find a hyperplane or set of hyperplanes to separate data samples, which can be used for classification and regression tasks. SVMs follow the intuition that a good hyperplane should have the largest distance (maximal margin) to the nearest training samples of any class because, the larger the margin, the lower the generalization error of the classifier. The data samples on the margin are support vectors. The original SVM algorithm is called a linear SVM, which can only apply to linearly separable data and perform binary classifications. Let $\{(x_i, y_i)\}_{i=1}^{n}$ be the training set, where x_i is the feature vector and $y_i \in \{-1, 1\}$

is the target value. The hard-margin (-1 or 1) linear SVM can be formulated as

$$minimize \; \|w\|$$
$$s.t. \; y_i(w^T x_i - b) \geq 1 \; \forall \; i = 1, 2, \cdots, n$$

where w is the normal vector to the hyperplane ($w^T x_i - b = 0$). The class label of a new data sample x is sgn($w^T x_i - b$), where sgn() is the sign function. To deal with data that are not linearly separable, the soft-margin method was proposed, which is defined by

$$minimize \; \lambda \|w\|^2 + \frac{1}{n} \sum_{i=1}^{n} \max\left(0, 1 - y_i(w^T x_i - b)\right)$$

where λ determines the tradeoff between increasing the margin size and ensuring that the data samples lie on the correct side of the margin. The original SVMs also have been extended to use the kernel trick to create nonlinear classifiers (Boser et al. 1992). SVMs have two advantages: (1) they work effectively on high-dimensional data and (2) they are memory efficient because the decision function is determined only by the support vectors.

3.2.2 Unsupervised Learning

In unsupervised learning, all one has is a set of data samples without being told their expected labels (ground truths) for categorical variables, nor the true numeric values for continuous variables. Unsupervised learning methods are promising in many applications due to three major reasons. First, labeling a large dataset can be surprisingly expensive and time consuming. If a method can be trained and run with a small to no amount of human supervision, researchers can save massive amounts of time and trouble. Second, we can use unsupervised learning to find features that can best represent the data and will be useful for future prediction tasks. Third, in the early stages of a research project, unsupervised learning methods are valuable tools used to gain insights into the structure of the data (i.e., the understanding of the probability density and subgroups can help influence the design for data classification and regression applications). Two central applications in unsupervised learning are clustering analysis and dimensionality reduction.

3.2.2.1 Clustering Analysis

Clustering Analysis is used to identify data subgroups and clusters in such a way that data samples from the same cluster are more similar to each other than to those from different clusters. There are many clustering algorithms because the notion of a 'cluster' cannot be clearly defined (Estivill-Castro 2002). The most appropriate algorithm for a particular task needs to be chosen experimentally. Popular clustering approaches are k-means (Lloyd 1982), spectral clustering (Ng et al. 2002), hierarchical clustering (Ward Jr 1963), DBSCAN (Ester et al. 1996), OPTICS (Ankerst et al. 1999), and Affinity propagation (Frey and Dueck 2007). A comparison of those approaches is shown in Table 5.

Table 5 Clustering Algorithms

Algorithms	Parameters	Scalability	Metrics Used	Use Case
K-means	K	Large N Medium K	Euclidean distance between points	Even cluster size, flat geometry
Spectral Clustering	K	Medium N Small K	Graph distance	Even cluster size, non-flat geometry
DBSCAN	maximum distance, number of neighbors	Large N Medium K	Distances between nearest points	Uneven cluster sizes, non-flat geometry
OPTICS	maximum distance (optional), number of neighbors	Large N Large K	Any distances between points	Uneven cluster sizes, non-flat geometry, variable cluster density
Hierarchical Clustering	K or linkage distance threshold	Large N Large K	Any distances between points	Many clusters possibly connectivity constraints
Affinity Propagation	Damping and sample preference	Small or Medium N	Graph distance	Many clusters, uneven cluster size, non-flat geometry

*K: number of clusters, and N: number of data samples.

K-means clustering is formulated as an optimization problem. It aims to find k cluster centers that minimize the sum of square distances within clusters and assigned data samples to the nearest cluster center. The optimization problem is known as NP-hard, and we often refer to the approximation method, Lloyd's algorithm (Lloyd 1982), as the k-means algorithm. Variations of k-means include k-medoids, k-medians, k-means++, and fuzzy c-means. K-means scales well to large datasets and has been used in a wild range of applications. The standard formulation of the k-means clustering is defined by

$$\arg\min_{C} \sum_{i=0}^{N} \min_{\mu_j \in C} \left\| x_i - \mu_j \right\|^2$$

where $C = \{\mu_j\}_{j=1}^{k}$ is a set cluster center and x_i is the ith data sample. When applying the k-means-type algorithms, the following issues should be considered:

- They require users to specify the number of clusters (K) in advance, which is considered one of the major drawbacks. A searching processing or domain knowledge can help find the most appropriate k.
- The performance highly depends on the initialization of the k centers, which is often alleviated by running the computation several times or choosing the initial centers to be distant from each other (k-means++).
- The Euclidean-distance-based clustering criteria assumes that clusters are convex and isotropic, which responds poorly to clusters with elongated or irregular shapes.
- When dealing with high-dimensional data, running dimensionality deduction approaches prior to clustering can mitigate the 'cure-of-dimensionality' problem and speed up the computation (e.g., spectral clustering).

Spectral Clustering (Ng et al. 2002) performs dimensionality reductions using the eigenvalues and eigenvectors of the similarity matrix of the data before clustering. The similarity matrix quantitatively measures the relative similarity of each pair of data samples. It is computationally efficient if the similarity matrix is sparse. Spectral clustering requires the number of clusters to be specified in advance. It works well for medium size datasets and a small number of clusters.

DBSCAN (Ester et al. 1996) **and OPTICS** (Ankerst et al. 1999) are two density-based clustering approaches, in which clusters are viewed as areas of high density and are separated by sparse areas of outliers or noise samples. The central concept in DBSCAN is the core samples, which are samples in high density areas. There are two parameters to the algorithm, the radius of a neighborhood $\varepsilon\varepsilon$ and the number of data sample minPts in a neighborhood to be considered core points. DBSCAN has three major steps: (1) selecting core points. A data point p is labeled as a core point if at least minPts points are within distance $\varepsilon\varepsilon$ of it; (2) finding directly reachable points. A point q is directly reachable from a core point p if q is within distance ε of p. A point q is reachable from p if there is a path that starts at p and ends at q. Any two adjacent points should be directly reachable along the path; and (3) a cluster is formed by any core point and all its (directly) reachable points. All points not reachable from any core points are labeled as outliers or noise.

The major advantages of DBSCAN are threefold: (1) it does not require the number of clusters to be specified, as opposed to k-means and spectral clustering; (2) it can find clusters with arbitrary shapes, whether they are convex or nonconvex; and (3) it is robust to outliers. Note that if the dataset has large differences in density, it can be difficult to determine a meaningful ε for DBSCAN. OPTICS aims to solve this weakness by removing the need to choose the distance threshold ε. The data points in the dataset are (linearly) ordered such that the spatially closest points become neighbors in the ordering. Strictly speaking, the distance threshold ε is not needed in OPTICS, but one can set it to speed up the algorithm.

Hierarchical Clustering (HC) builds clusters by recursively partitioning data samples using merging or splitting strategies (Rokach and Maimon 2005). In the merging strategy, each data sample starts in its own cluster, and cluster pairs are recursively merged. The splitting strategy starts putting all data samples in one cluster and performs splits recursively. HC introduced the linkage criterion to decide which cluster pairs should be merged or where a cluster should be split. The linkage criterion defines the dissimilarity and distance between sets of data samples. Popular linkage criteria are Ward's (Ward Jr 1963), complete linkage, single linkage, and average linkage. HC requires one to specify the number of clusters to find the linkage distance. The recursive merging or splitting can terminate if enough clusters have been produced or all between-set distances are larger than the threshold.

The distinct advantage of HC is that any distance metrics (e.g., Euclidean distance, squared Euclidean distance, Manhattan distance, and Hamming distance) can be used in the linkage distance, which broadens the applications of HC to data with different formats. For numeric data, the most common metrics are the Euclidean distance and squared Euclidean distance, while the Hamming distance is more appropriate for text or other non-numeric data. In addition, HC perform all operations using the distance matrix, and the original data are not required in the clustering process. The standard HC algorithms calculated all distances between points from two different sets, which is computationally expensive and hinders the application of HC algorithms to big datasets. One can solve this issue by specifying a connectivity matrix to define the neighborhood structures.

Affinity propagation (AP) (Frey and Dueck 2007) aims to identify the exemplars that are representative of other data samples by using two message-passing steps between data points. Clusters are constructed by finding data samples that share the same exemplar. The first category of message is the responsibility $r(i, k)$, which accumulates evidence that data sample k should be the exemplar for sample i. The second message is the availability $a(i, k)$, which accumulates the evidence that data sample i should choose sample k to be its exemplar. $r(i, k)$ and $a(i, k)$ are updated iteratively until there are either no changes over a number of iterations or some predefined number of iterations is reached. The exemplars are chosen as those whose "responsibility + availability" for themselves is positive.

AP does not require the number of clusters to be specified or estimated in advance. However, two parameters, the damping factor (λ) and preference, should be determined. The damping factor is used in the message updating steps to avoid numerical oscillations, and the preference value specifies the preference of data sample to be chosen as an exemplar. The final number of clusters will be influenced by the preference. The main disadvantage of AP is its complexity. It has a time complexity and memory complexity of the order O(N2T) and of the order of O(N2), respectively, where N is the number of data samples and T is the number of iterations. It is most appropriate for small to medium sized datasets.

3.2.2.2 Dimensionality Reduction

Dimensionality Reduction transforms high-dimensional data to low-dimensional representation that preserves meaningful properties of the original data. It is often used as an intermediate step to remove redundant features and noisy data to enable a better performance of other data analysis tasks (e.g., data classification and visualization). There are two major benefits to dimensionality reduction: (1) it helps some algorithms to work more efficiently and improve performance after the redundant, irrelevant, and noisy data are removed; and (2) it allows us to visualize patterns of high-dimensional data. Dimensionality reduction approaches are commonly divided into two categories: linear and nonlinear approaches (Van Der Maaten et al. 2009). Table 6 presents several popular Dimensionality Reduction algorithms.

Table 6 Dimensionality Reduction Algorithms

Algorithms	Linear/Non-linear	Parameters	Computation	Memory
PCA	Linear	none	$O(D^3)$	$O(D^3)$
Kernel PCA	Non-linear	kernel function	$O(N^3)$	$O(N^3)$
Isomap	Non-linear	k	$O(N^3)$	$O(N^3)$
Local linear embedding (LLE)	Non-linear	k	$O(pN^2)$	$O(pN^2)$
Self-organizing map	Non-linear	number of neurons, weight vectors, iteration limit, correction step	/	/
Autoencoders	Non-linear	Network architecture and weights Learning rate	/	/

*k is the number of nearest neighbors, D is the dimensionality of the input data, and N is number of data samples. '/' denotes not applicable.

Principle components analysis (PCA) and Kernel PCA: PCA transforms the data to a lower-dimensional space so that the variance of the data is maximized. PCA attempts to find a linear mapping M that maximizes the following cost function

$$trace(\boldsymbol{M}^T S \boldsymbol{M})$$

where S is the scatter matrix (zero-mean) or the covariance matrix of the data samples. Let D be the dimensionality of the original data samples. M is constructed by using the S's d principal eigenvectors that correspond to the largest d (d < D) eigenvalues. The low-dimensional data representation $y = (y_1, y_1, \cdots, y_d)^T$ of a data sample $x = (x_1, x_1, \cdots, x_D)^T$ is computed by

$$y = x^T \boldsymbol{M}$$

PCA is still the most popular linear approach for dimensionality reduction. However, when we apply PCA, several practical issues should be considered: (1) PCA is sensitive to the scaling of the features, and we need to scale each feature by its standard deviation; (2) PCA captures linear correlations between features, but it may fail if the potential correlation is nonlinear; and (3) the size of the covariance matrix ($D \times D$) is proportional to the dimensionality of the data samples (D). Consequently, it might be not be feasible to compute the eigenvectors for very high-dimensional data if the first few components do not explain a large proportion of the total variability in the data. **Kernel PCA** (Schölkopf et al. 1998) extends PCA to achieve non-linear dimensionality reduction by using the "kernel trick." It constructs the kernel matrix of data points using a kernel function $\kappa(x_i, x_j)$ and computes the eigenvectors of the kernel matrix, rather than those of the scatter matrix. The size of the kernel matrix is $N \times N,$ where N is the number of data samples. Therefore, a large data set will lead to high memory complexity. One way to solve this is to perform clustering on the dataset first and calculate the kernel matrix using the cluster centers.

Isomap (Tenenbaum et al. 2000) nonlinear dimensionality reduction is an approach that aims to preserve pairwise geodesic distances between data samples. The geodesic distances are computed by constructing a neighborhood graph G, in which every data point xi connects its k-nearest neighbors in the dataset. The shortest path of two points in the graph is used to estimate the geodesic distance. All geodesic distances between every data pair are used to build a geodesic distance matrix. The low-dimensional representations of the data points are computed by applying multidimensional scaling algorithm. The Isomap performance is sensitive to the chosen k. If k is large, Isomap will be vulnerable to the short-circuit error because the geodesic distance calculation propagates the error to the whole distance matrix. On the other hand, if k is too small, G will be too sparse to calculate the geodesic distance accurately.

Local linear embedding (Roweis and Saul 2000) also constructs graphical representations of data samples, while aiming to retain only the local properties of the data. LLE algorithms includes three major steps: (1) it finds a set of the nearest neighbors $\{x_i\}_{i=1}^k$of each point x; (2) it computes a set of weights $\{w_i\}_{i=1}^k$that best describes a data sample as a linear combination of its neighbors $x = \sum_{i=1}^k w_i x_i$; and (3) it uses an eigenvector-based optimization to find the low-dimensional embedding y, such that each data sample is still described with the same linear combination $y = \sum_{i=1}^k w_i y_i$. In contrast with Isomap, LLE includes a faster optimization when implemented to take advantage of sparse matrix algorithms and is less sensitive to the short-circuit error. However, it handles non-uniform sample densities poorly because the weights may drift drastically as regions differ in sample densities.

Self-organizing map (Kohonen 1982, Kohonen and Honkela 2007) is an artificial NN-based approach that maps high-dimensional input data to a finite 2D map space. The map space contains neurons than are arranged in a regular hexagonal or rectangular grid. Each neuron is assigned a weight vector that has the same dimension as input data samples. The training process updates weight vectors toward the input data without spoiling the topology. The self-organizing map training uses competitive learning and includes three steps: (1) it calculates the Euclidean distance between the input data and all weight vectors; (2) it finds the best matching unit or neuron (BMU) whose weight vector is most similar to the input; and (3) it updates the weight vectors of the neurons in the neighborhood of the BMU using

$$w^{t+1}(i) = w^t(i) + h(u,i,t) \cdot \alpha(s) \cdot (x(t) - w^t(i))$$

where t is the iteration index, u is the index of the BMU, and i is the index of a neuron in BMU's neighborhood. $\alpha(s)$ is a scalar factor that defines the size of the correction, and its values decreases with the step index t. $w^t(i)$ is the weight vector of the ith neuron at iteration t. $h(u,i,t)$ is the neighborhood function. It is equal to 1 when i=u, and its value decreases when the distance between the neurons I and u increases. The above three steps will be repeated until it reaches the iteration limit. After the training, we can map input data samples to a 2D coordinates of the BMUs.

Autoencoders are feedforward neural networks that aim to learn a low-dimensional representation (encoding) for a dataset by training the network to ignore signal "noise" or redundancies. It includes two main parts (see Figure 6): an encoder that maps the input into the low-dimensional feature vector (code) and a decoder that reconstructs the original input. Autoencoders are often trained with only a single layer encoder and a single layer decoder, but using deep encoders and decoders can reduce the computation cost and yield better compression (Goodfellow et al. 2016). Figure 6 shows the architecture of an autoencoder with seven hidden layers. The rectangles illustrate the feature maps generated in the hidden layers. The code y is the output of the most internal layers and is the low-dimensional representation of x. $\hat{x}$ is reconstructed data of x using the code y.

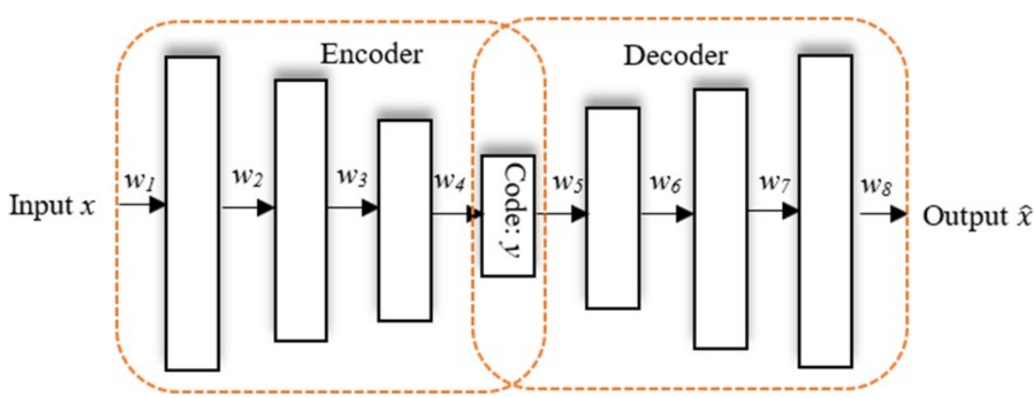

Figure 6 Architecture of an Autoencoder with Seven Hidden Layers

3.3 AI/ML Algorithms for Computational Predictive Capabilities

Section 3.2 introduces the most popular AI/ML algorithms in nuclear science and engineering. This section focuses on AI/ML algorithms that are proper for developing advanced computational predictive capabilities. The algorithm selection for a successful nuclear application depends on two major factors: (1) the nature and objectives of the task, e.g., classification or clustering analysis; and (2) data availability and quality. For instance, if the application is to detect cracks of reactors using surface images, the task can be viewed as a classification problem that classifies each image pixel into two categories: crack and non-crack. Therefore, we can select classification algorithms from the supervised learning category, e.g., CNNs. If the task requires the explainability of the algorithm to demonstrate the reasoning process, DTs, RFs and BNs might be better choices. In practice, the two above factors can be used to narrow down the searching range; however, we do not have a generic principle to determine the perfect algorithm for a specific task. The best strategy is to evaluate and compare different algorithms using extensive experiments along with specific physical phenomenon considerations, and the final algorithm(s) should be determined by using values of quantitative metrics, e.g., accuracy, precision, and recall rate, on new datasets.

In nuclear science and engineering, AI/ML approaches have been widely applied to enhance equipment reliability, reduce radiation exposure to personnel, assist with decision making and optimize maintenance schedule in in three major areas: (1) nuclear power plant health and management, (2) nuclear operations and controls, and (3) radiation protection. A list of example use cases and algorithms are shown in Table 7.

Table 7 Example Use Cases and Algorithms

Application Area	Use Case	Algorithms
Plant health and management	System behavior prediction	BNs, NB, ANNs, SVM
	Severe accident classifications	ANNs, DTs, BNs
	Functional failure of systems	ANNs, Clustering algorithms, e.g., K-means
	Crack detection	CNNs
	Equipment monitoring	CNNs, ANNS, BNs,
Nuclear operations and controls	Anomalous event detection	AEs, SVM, ANN, DTs
	Unattended operations	DTs, BNs
	Detection and response to degraded or failure conditions	CNNs, ANNs
	Radwaste management	CNNs
Radiation protection	Radionuclide identification	ANNs, SVM
	Special nuclear material identification	ANNs, GPs, NB, Clustering algorithms

3.4 AI/ML Languages and Tools

Python, C++, and R are among the most popular ML programming languages. Python is the fastest-growing programming language in recent years with its readability and good structure. It is a general purpose language but has high-quality ML and data analysis libraries and is suitable for ML model development. C++ is a flexible, object-oriented, middle-level language based on the C programming language. It can directly interact with hardware under real-time constraints and can be used for parallel computing. R is a top choice for many data scientists as a language and environment for statistics, visualization, and data analysis. R libraries provide numerous statistical and graphical techniques and can also be extended with R machine learning packages.

There are extensive and ever-evolving ML tools, platforms, and software for data analytics and visualization, e.g., Python Pandas, NumPy, KNIME, TensorFlow, Pytorch, Accord.net, Google cloud AutoML, and Jupyter notebook. There are also commercial off-the-shelf software such as SAS and MATLAB that can be used to apply ML in data analytics. These off-the-shelf software include the tools and functions that can handle big data and make ML accessible with prebuilt functions, extensive toolboxes, and specialized apps for classification, regression, and clustering. And the results from these software are generally trusted in the numerical analysis research community.

4 AN OVERVIEW OF APPLICATIONS OF ADVANCED COMPUTATIONAL TOOLS AND TECHNIQUES IN NUCLEAR INDUSTRY

Advanced computational capabilities have been required, developed, and deployed in various fields of the nuclear industry, such as reactor system design and analysis, plant operation and maintenance, and nuclear safety and risk analysis. In the most-recent decade, research in nuclear engineering has produced a large amount of experimental and numerical data, which, if measured in petabytes, is perhaps several orders of magnitude more than all accumulated data in the previous history of the industry. However, there is still a severe lack of data to validate the multiphysics, multiscale capability because very little experimental and plant data are directly relevant to validate high-fidelity mechanistic models and codes, particularly when advanced reactor designs are involved. The enormous progress in science, technology, and engineering in past decades brings opportunities to help deal with these proposed issues. Some capabilities that are enabled in this environment have been summarized in (Dinh et al. 2013) which are still instructive and informative today:

- "Increasing affordability of advanced experimental and diagnostic techniques for the experimentation under some high-temperature high-pressure conditions of interest to reactor applications." It provides a technical basis for generating necessary but extreme experimental data for validation and uncertainty quantification purpose.
- "Advancement of data science, including statistical analysis methods and tools for processing of multi-field, multi-dimensional heterogeneous datasets, data mining, pattern recognition, data aggregation, and data assimilation."
- "Methods and tools for sensitivity analysis, uncertainty quantification, model calibration and validation, and design of experiments to maximize the data's informative value."
- "Advanced methods in computational physics that enable effective and accurate solutions for complex non-linear multi-scale problems."
- "Advancement in computer science and software engineering that provides methods and tools to accommodate increasingly and necessarily sophisticated software architectural requirements in a new modeling framework (e.g., flexible data-model integration)."
- "Affordable data storage and computational power needed for data processing, sensitivity and uncertainty analysis, model calibration and time- and space- resolved high-fidelity simulations."
- "Community-wide experience, shared best practice, standards development and accumulative knowledge base from using, innovating, and pushing existing methods and tools in nuclear industry to the limit, particularly driven by common goals in nuclear reactor safety."

This section presents an overview of AI/ML applications in the nuclear industry and academic research in these fields, evaluates the potential applicability of AI/ML techniques in improving advanced computational capabilities, and provides insights on how to utilize AI/ML techniques in simultaneously improving plant safety and enhancing regulatory oversight.

4.1 AI/ML in Reactor System Design and Analysis

AI/ML techniques have been widely introduced and applied in the design and analysis of reactor systems, including reactor thermal hydraulics, reactor physics, and reactor system performance. Part of these efforts after 2010 are listed and reviewed in Table A-1 of Appendix A. The main data types are structured experimental data or numerical data generated by simulation codes or simulators. These datasets are good supplements to plant operating experience as key parameters and variables in important reactor systems, and components can be simulated using different simulation codes or observed on different tests via full-scaled or scaled-down facilities.

For reactor system design and analysis, a majority of AI/ML applications are focused on the model and code uncertainty analysis and closure model development. ML-based methods have emerged as a valuable approach to aid in the development and application of thermal-hydraulic or neutronic methods. ML provides new avenues for dimensionality reduction and reduced-order modeling in fluid mechanics or neutronics by providing a concise framework that complements and extends existing methodologies. In past decades, many simulation codes have been developed based on various empirical correlations and numerical algorithms for the physics and phenomena existing in nuclear power plants. Different supervised and unsupervised ML/AI algorithms, such as ANN (Bao et al. 2021), deep learning (Chang and Dinh 2019), SVM (Trontl et al. 2008), GP (Pastore et al. 2017), DT (Ling and Templeton 2015), RF, Bayesian neural network (BNN) (Utama et al. 2016), Cascade Fuzzy NNs (Choi et al. 2016), and Kernel Regression (Tracey et al. 2013), have been widely applied for different research and application objectives in reactor system design and analysis, as listed in Table A-1 (most of them have also been described in Section 3).

These ML/AI algorithms are suitable for processing and analyzing structured data; however, the main technical challenge comes from their “black box” nature, which brings in a new uncertainty source and makes it difficult to explain and trust AI/ML techniques when applied to nuclear research or in the industry. Also, the “superpower” of AI/ML techniques to capture the features of training data may lead to overfitting for the predictions.

4.2 AI/ML in Plant Operation and Maintenance

In recent decades, AI/ML techniques have also been investigated in supporting and optimizing nuclear power plant operations and maintenance. Some of recent efforts are summarized in Table A-2 of Appendix A. (Lin et al. 2021a) developed an autonomous management and control system for advanced reactors to mitigate plant anomalies and accidents by using a feedforward neural network (FNN) and RNN. Similarly, other ML/AI techniques or advanced statistical methods, such as Bayesian network (Cetiner and Ramuhalli 2019), Answer Set Programming (Hanna et al. 2020a), and LSTM (Lee et al. 2018a), were also demonstrated and applied to the development of advanced control systems to support the operators’ decision making. Owing to their accurate, real-time predictions, these advanced AI-guided systems are able to help operators understand the current plant status and make optimal mitigation planning for specific plant anomalies and hazardous events. Besides, AI/ML techniques, particularly ANN in recent years, are frequently deployed for various (semi)automatic operation and controls for different purposes, such as a load following operation (Khajavi et al. 2002), smart core controller (Boroushaki et al. 2003), alarm processing system (Park and Seong 2002), symptom-based diagnostic system (Vinod et al. 2003), real-time nuclear power plant monitoring (Nabeshima et al. 2012), plant abnormality identification (Ayo-Imoru and Cilliers 2018), and component detection (Gao et al. 2020).

Besides, researchers developed various AI/ML-aided methods for cybersecurity studies. (Zhang et al. 2020) developed a cybersecurity solution platform using an AI/ML hub, including k-nearest neighbor, DT, bootstrap aggregating, RT, auto-associative kernel regression, and PCA. (Gawand et al. 2017) developed and tested a cyber-physical system via least squares approximation. (Poolsappasit et al. 2012) and (Shin et al. 2017) utilized Bayesian methods for dynamic security risk management and evaluation. (Lee and Huh 2019) applied unsupervised ML (classification), reinforcement learning for a plant security measure.

Most of these AI/ML applications are performed based on simulated data since plant operating data is rarely available; however, these robust, accurate, and fast computational capabilities enabled by AI/ML techniques are very instructive and informative for realizing autonomous plant control and management for reducing cost and improving reactor resilience. However, most of these AI/ML-aided techniques are developed for the safety-significant or safety-related I&C of nuclear power plants; there are strict regulatory requirements for their licensing process. The trustworthiness, transparency, and robustness of these AI/ML-aided techniques should be identified, analyzed, and evaluated in future research.

4.3 AI/ML in Nuclear Safety and Risk Analysis

Table A-3 of Appendix A lists recent studies that developed or applied AI/ML approaches for nuclear safety and risk analysis, primarily for the PRA of nuclear power plants. Unlike the AI/ML applications in reactor design & systems analysis and plant operation & maintenance, AI/ML applications in nuclear safety and risk analysis are performed for both structured data and unstructured free-text data. (Siu et al. 2013) discussed the role that content analytics and text analytics plays in supporting regulatory decision making, and the NRC's plan to initiate scoping studies to explore the application of advanced data analytics techniques to support PRA activities. To recognize free-text data and extract implied information inside ML algorithms such as NLP, supervised and unsupervised ML is applied. (Zhao et al. 2019) utilized NLP to extract the causal relationships among failure-contributing factors from free-text reports. (Moura et al. 2017) applied unsupervised ML (clustering) to validate risk studies using information from past major accidents. (Mandelli et al. 2018) developed a data-driven method for cost risk analysis via supervised ML (classification).

Another difference between these ML/AI applications in nuclear safety and risk analysis is that AI/ML techniques are not only directly applied for model development or uncertainty quantification but are embedded in complicated frameworks for different purposes. (Christian et al. 2020) developed a data-driven framework for the estimation of pressurized-water reactor (PWR) coping time, wherein the GP, SVM, k-nearest-neighbor classifier and regressor, Shepard's method, and the spline interpolation method can be selected and applied. (Kim et al. 2020a) introduced dynamic Bayesian network and clustering methods for the risk assessment of dynamic systems. (Park et al. 2017) extracted the relative importance of performance shaping factors for human reliability analysis using CART. (Zou et al. 2018) developed a data mining framework, combining three statistical approaches (i.e., correlation analysis, cluster analysis and association rule mining) to identify intrinsic correlations among human factors. (Maljovec et al. 2015) introduced unsupervised ML (clustering) to analyze simulation-based PRA data. (Di Maio et al. 2016b) applied semi-supervised ML to post-process the multi-valued dynamic scenarios.

In 2020, (Pence et al. 2020) reviewed existing studies that developed and/or applied ML approaches for the PRA of nuclear power plants, which highlighted following results:

- “There are a limited number of studies using machine learning to quantify PRA model elements, and none of the studies included organizational factors.......The application of machine learning approaches for PRA primarily analyzed physical phenomena, where machine learning was used to cluster the simulation outcomes. In these studies, the data are not historical events but instead are the results of simulation codes; therefore, the main challenge is dealing with large volume of data rather than processing heterogeneous data.”
- “Several studies leveraged the Risk Analysis and Virtual Environment (RAVEN) computational platform to operationalize machine learning for time-dependent data resulted from simulations that were equipped with sampling and uncertainty analysis (e.g., ADAPT/RELAP/RAVEN (Mandelli et al. 2013a)).”
- Among the PRA-oriented AI/ML studies, most of these efforts used historical event data rather than results of simulation codes, such as(Young et al. 2004, Maljovec et al. 2015, Siu et al. 2016, Christian et al. 2020, Ham and Park 2020). “There are limited studies using text mining approaches for PRA. Additional research is needed to compare the performance evaluation of machine learning techniques for unstructured data to justify the best selection for PRA.”

These highlights are instructive for the NRC to guide and initiate future applications of advanced computational tools and techniques including AI/ML in nuclear safety and risk analysis, particularly in PRA. ML-based methods have recently emerged as a valuable approach to aid the development and application of methods for solving different technical issues in the nuclear industry. These applications have constructed very diverse and solid technical bases for improving the use of AI/ML in dealing with numerous technical issues. The rapid progress of the AI/ML techniques in other industrial fields also provides very valuable lessons to similar problems. However, due to AI/ML uncertainty, the insufficiency of data quality and quantity, and lack of cognition about how to efficiently incorporate knowledge and data, challenges of adapting AI/ML techniques still exist. New perspectives and advanced frameworks should be proposed for different purposes in nuclear engineering. Particularly, the “black box” nature of ML/AI brings challenges with respect to the trustworthiness and transparency of the results in nuclear industry. This challenge makes the deployment of ML/AI-guided applications difficult to satisfy the regulatory requirements of NRC.

5 INSIGHTS ON TASK 1 QUESTIONS

This section provides the insights for the three questions under the purpose of Task 1.

Question 1 for Task 1: What types of advanced computational tools and techniques may be employed, how would they work, and how effective would they be expected to be?

Advanced computational tools and techniques include advanced statistical algorithms (e.g., Bayesian methods), AI/ML algorithms (e.g., ANN, SVM, and RF), and relevant hybrid applications (e.g., physics-informed machine learning). All of them are promising to be deployed in the nuclear industry. These advanced computational tools and techniques, particularly ML/AI, have been used in a wide variety of applications, such as self-driving cars and computer vision, where it is difficult or unfeasible to develop conventional algorithms using human knowledge to perform the needed tasks. With excellent performance in various fields, different types of advanced computational tools and techniques reviewed in Section 3 can be applied in nuclear industry for different scopes and purposes. While Table 7 in Section 3.3 provides a list of example use cases and advanced computational techniques in different areas, there is no generic principle to determine the perfect technique for a specific task. Instead, different techniques should be evaluated and compared with extensive experiments and quantitative metrics such as accuracy, precision, and recall rate. A successful nuclear application of advanced computational tools and techniques will depend on two major factors: (1) the nature and objectives of the task, e.g., classification or clustering analysis, CNNs or RFs, BNs; and (2) data availability and quality. In practice, these two factors can be used to narrow the selection of advanced techniques for a specific task.

On the other hand, there are some basic principles to remember when applying the advanced computational tools and techniques in nuclear OpE in order to ensure the trustworthiness, transparency, and explainability of the predictions from the tools and techniques. For instance, although ML can provide satisfactory predictive capabilities, machine-learning programs sometimes fail to deliver expected results. Potential reasons include lack of suitable data, unsuitable process of training data, insufficient training and selection of unsuitable algorithms. Uncertainty always exists in ML predictions for unknown values or unfamiliar problems. This uncertainty may lead to unexpected results which can lead to an inappropriate suggestion to plant operator for diagnosis and prognosis of plant anomalies. Particularly, the “black box” nature of AI/ML brings challenges to the trustworthiness and transparency of applications in nuclear OpE. These challenges make the deployment of ML/AI-guided applications difficult to satisfy the regulatory requirements. Accordingly, before the advanced computational tools and techniques being deployed in nuclear industry, “use cases” of ML/AI applications in nuclear industry must be carefully developed with the results being validated (e.g., compare the results to those from traditional methods when proper) in order to improve regulator’s confidence in new tools/techniques and meet various regulatory requirements.

Question 2 for Task 1: What aspects of the advanced tools and techniques could contribute to our increased understanding of safety and risk?

AI/ML approaches have been applied in the nuclear industry to enhance equipment reliability, reduce radiation exposure to personnel, assist with decision making and optimize maintenance schedule. There are many aspects where advanced computational tools and techniques could contribute to our increased understanding of safety and risk. For example,

1. Advanced computational tools and techniques are capable of recognizing and processing both structured data and unstructured data (e.g., free-text data). They can extract both explicit and implied information from unstructured data. Such capability could facilitate the usage of more data sources and provide a larger quantity of raw data for PRA parameter estimation.
2. Advanced computational tools and techniques are capable of training/enhancing data-driven models and reflecting the relationships between model inputs and outputs, even without knowing the underlying physics. Such capabilities could make it feasible to measure the impacts of fluctuations of potential influencing factors on PRA parameters (such as testing the impact of room temperature on component unreliability) or PRA outputs, which could facilitate uncovering previously unknown (or not explainable using physics) risk-contributing factors.
3. Advanced computational tools and techniques are capable of developing predictive models for key physics in NPPs or developing surrogate models for computationally-expensive simulations of system physics. Such capabilities could help satisfy PRA needs, such as, simulating accident progression in dynamic PRA, simulating component failure mechanism in classical PRA, or examining the impacts of uncertainties in deeper-level physical parameters on PRA outputs, associated with running a large number of physics simulations within an acceptable time period.
4. Advanced computational tools and techniques can be applied in supporting and optimizing nuclear power plant operation and maintenance, where there are knowledge gaps and technical issues. They are capable of handling high-volume, high-frequency data such as sensor data. Combining this capability and the above-mentioned model-training capability, they could facilitate construction of advanced diagnostic (such as detecting failure cause) and prognostic models (such as predicting remaining useful life), solving these models based on high-volume, high-frequency data, and generating real-time diagnostic and prognostic results. With these capabilities, they could also facilitate integrating micro-level prognostic models with PRA (such as constructing a model reflecting the relationship between component remaining useful life and component failure probability) and updating PRA outputs in real time.
5. Besides the above-mentioned capabilities that could help expand the modeling scope and enhancing the modeling resolution of PRA, advanced computational tools and techniques are capable of automating manually-conducted analyses. Such capability could help improve the efficiency of model development (such as better visualizing raw data or easing event tree and fault tree construction) and parameter estimation (such as accelerating raw data processing) of PRA.

Another thought is whether there needs to be a fundamental shift away from “failures” and more on “success” in PRA. In other words, more focus on reliability than P(failure) as there is much more information on working operation and much less on failures (and why things fail). With this potential shift, advanced computational tools and techniques could play a significant role to deal with enormous “success” data as well as compare the results from the “success” data to those from the “failure” data.

Question 3 for Task 1: What types and quantities of information would be needed, in concert with the new tools and techniques, to generate safety and risk implications?

To ensure accurate and acceptable predictive capabilities for key plant OpE parameters and behaviors, the following information are needed to reduce the uncertainty of the prediction results and provide meaningful insights to decision makers when assessing safety and risk:

1. Clear description and sufficient understanding of the task, including the expected outputs and the key metrics of safety and risk significance.
2. Suitable and sufficient data for the training of AI/ML models.
3. Suitable methods for training data processing.
4. Sufficient knowledge repository for target (e.g., physics, system) of interest and the approaches to validate the results.

6 A SURVEY ON THE ROLE OF ARTIFICIAL INTELLIGENCE TOOLS IN U.S. COMMERCIAL NUCLEAR POWER OPERATIONS

On April 21, 2021, the NRC published a Federal Register Notice (FRN) NRC-2021-0048 (U.S. Nuclear Regulatory Commission 2021) containing an 11-question survey to request public comments on the current state of commercial nuclear power operations relative to the use of AI and ML tools as well as the role of AI/ML tools in U.S. commercial nuclear power operations. The survey results were planned for use in enhancing the NRC's understanding of the short- and long-term applications of AI and ML in nuclear power industry operations and management, as well as potential pitfalls and challenges associated with their applications. The survey had been open for one month until May 21, 2021. Twelve participants (individuals or organizations) responded to this survey and submitted their written-form responses.

This section provides a summary of the survey responses as well as the conclusions and insights derived from the survey. Section 6.1 presents the 11 survey questions. Section 6.2 lists the 12 survey participants that provided responses to the NRC. Section 6.3 provides a survey question-response matrix for the explicit responses of each survey question by the participants and the detailed responses to each survey question. Section 6.4 provides the insights obtained from the survey responses.

6.1 Survey Questions

For reference, the 11 survey questions in FRN NRC-2021-0048 are listed below:

1. What is status of the commercial nuclear power industry development or use of AI/ML tools to improve aspects of nuclear plant design, operations or maintenance or decommissioning? What tools are being used or developed? When are the tools currently under development expected to be put into use?
2. What areas of commercial nuclear reactor operation and management will benefit the most, and the least, from the implementation of AI/ML? Possible examples include, but are not limited to, inspection support, incident response, power generation, cybersecurity, predictive maintenance, safety/risk assessment, system and component performance monitoring, operational/maintenance efficiency, and shutdown management.
3. What are the potential benefits to commercial nuclear power operations of incorporating AI/ML in terms of (a) design or operational automation, (b) preventive maintenance trending, and (c) improved reactor operations staff productivity?
4. What AI/ML methods are either currently being used or will be in the near future in commercial nuclear plant management and operations? Example of possible AI/ML methods include, but are not limited to, artificial neural networks (ANN), decision trees, random forests, support vector machines, clustering algorithms, dimensionality reduction algorithms, data mining and content analytics tools, gaussian processes, Bayesian methods, natural language processing (NLP), and image digitization.
5. What are the advantages or disadvantages of a high-level, top-down strategic goal for developing and implementing AI/ML across a wide spectrum of general applications versus an ad-hoc, case-by-case targeted approach?
6. With respect to AI/ML, what phase of technology adoption is the commercial nuclear power industry currently experiencing and why? The current technology adoption model characterizes phases into categories such as: the innovator phase, the early adopter phase, the early majority phase, the late majority phase, and the laggard phase.

7. What challenges are involved in balancing the costs associated with the development and application of AI/ML, against plant operational and engineering benefits when integrating AI/ML applications into operational decision-making and workflow management?
8. What is the general level of AI/ML expertise in the commercial nuclear power industry (e.g., expert, well-versed/skilled, or beginner)?
9. How will AI/ML effect the commercial nuclear power industry in terms of efficiency, costs, and competitive positioning in comparison to other power generation sources?
10. Does AI/ML have the potential to improve the efficiency and/or effectiveness of nuclear regulatory oversight or otherwise affect regulatory costs associated with safety oversight? If so, in what ways?
11. AI/ML typically necessitates the creation, transfer and evaluation of very large amounts of data. What concerns, if any, exist regarding data security in relation to proprietary nuclear plant operating experience and design information that may be stored in remote, offsite networks?

6.2 Survey Participants

A total of 12 participants responded to FRN NRC-2021-0048, including Florida Power & Light Company (FPL), Xcel Energy (Xcel), EPRI, NEI, Westinghouse Electric Company LLC (WEC), Framatome Inc., X-energy, Blue Wave AI Labs, and a few other consulting companies or anonymous participants. Table 8 lists the survey participants as well as the NRC Agencywide Documents Access and Management System (ADAMS) access numbers for the comments they provided.

Table 8 A List of Survey Participants

No.	Participant	Response Accession Number
1	Anonymous	ML21113A083
2	Southern Research Institute (SRI)	ML21126A011
3	FPL	ML21139A103
4	EPRI	ML21141A184
5	Xcel	ML21141A185
6	ForHumanity	ML21145A363
7	Blue Wave AI Labs (Blue Wave)	ML21145A364
8	X-energy	ML21145A366
9	Insight Enterprises, Inc. (IEI)	ML21145A367
10	NEI	ML21145A369
11	Framatome Inc. (Framatome)	ML21153A056
12	WEC	ML21202A180

6.3 Survey Responses

This section provides the detailed survey responses to each survey question by participants. Table 9 shows a matrix of survey question and response by participants.

Table 9 Survey Question and Response Matrix

No.	Participants	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Notes
1	Anonymous	N	N	N	N	N	N	N	N	N	N	N	Referenced two publications (Kortelainen et al. 2020, Suman 2020)
2	SRI	N	N	Y	N	Y	N	N	N	N	N	Y	
3	FPL	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
4	EPRI	N	N	N	N	N	N	N	N	N	N	N	Referenced four EPRI-authored publications (Electric Power Research Institute 2020b, Electric Power Research Institute 2020a, Electric Power Research Institute 2021b, Electric Power Research Institute 2021a)
5	Xcel	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
6	ForHumanity	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Included an introduction of AI-related work conducted by the ForHumanity
7	Blue Wave	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
8	X-energy	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
9	IEI	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
10	NEI	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
11	Framatome	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
12	WEC	Y	Y	Y	Y	N	Y	N	N	N	N	Y	

* N = not answered, Y = answered

Survey Question 1

Survey Question 1 asks the status, tools, and the expected launch timeline of the commercial nuclear power industry development or use of AI/ML tools. Nine of the 12 participants responded to this question, and their responses are summarized in Table 10.

Table 10 Summaries of Survey Responses to Survey Question 1

Participants	Summary of Responses to Survey Question 1
FPL	Currently using AI/ML to improve several aspects, including work management, the corrective action program (CAP), and equipment reliability. Several ongoing projects in various stages of development; some applications have been in place for over a year.
Xcel	Beginning to leverage cloud-based technology to create AI/ML applications. The first application currently under development is the CAP Intelligent Advisor. The first deployment is anticipated in late 2021 on the CAP support.
ForHumanity	Not directly relevant to the question.
Blue Wave	Believes the nuclear industry is in the early adopter phase. Recommends several areas for AI/ML applications in both existing reactor fleet and next generator reactor designs. Currently working with utilities on using AI for fuel cycle management and predictive maintenance; this work has been going on for the last three years. Two tools have been developed, including the MCO.ai to predict and manage moisture carryover and the Eigenvalue.ai to predict boiling water reactor (BWR) eigenvalue evolution for future fuel cycles. Both tools have been utilized in BWRs and led to significant cost savings over the last two years using these tools.
X-energy	Has identified and is actively seeking the use of AI/ML in a wide variety of applications for the advanced reactors. Several applications under development, while the others in theory exploration. Has been using Python to build custom AI/ML models; many data science code packages are used, with the most notable being TensorFlow. The expected release and use for the Xe-100 will be when the first unit is commissioned under the Advanced Reactor Demonstration Program; tentatively between 2025–2027.
IEI	AI/ML in the early stages of development and use in the nuclear power industry.
NEI	Industry use of AI/ML tools varies from company to company but also within each company from organization to organization, with some in the mature stage being actively used in the organization, while others are still under development. Use varies across organizations, with some using internally developed solutions and others relying upon external vendors (most are collaborating with EPRI). A variety of areas have been identified for AI/ML use, including textual report analysis, condition-based maintenance, work order planning, fuel performance prediction, and reactor core design optimization.

Table 10 Summaries of Survey Responses to Survey Question 1 (continued)

Participants	Summary of Responses to Survey Question 1
	One licensee is using AI in a limited business process of classifying condition reports using IBM Watson to make the initial classification. The licensee does not have any immediate plans or approved projects for other AI applications.
Framatome	Focused on several aspects, including root cause analysis (RCA) support, non-destructive examination (NDE) support, and reactor operation and control. For RCA support, an AI/ML tool, Metroscope, is currently deployed on more than 40 reactor units; research is ongoing to enhance the technical capability and generalize the application fields of the Metroscope. For NDE support, AI/ML tools are actively being developed; some tools are already developed and being beta tested. For reactor operation and control, AI/ML tools are being developed; an example is the Operator Assistance Predictive System, employing the Artificial Narrow Intelligence software.
WEC	Has identified and initiated a variety of AI/ML applications related to NPP design, operations, maintenance, and decommissioning. One application (use of surrogate modeling to streamline design analysis process) has been partially used, while most of the applications are ongoing. Broadly groups AI/ML tools into two categories, including anomaly detection and ML (teaching the software/human to find a pattern). WEC has recently developed a tool that evaluates over 10 regression-based AI/ML algorithms to find trends in the data and then select the optimal algorithm based on data-driven modeling validation metrics. Clarifies that data does not have to be numeric; one of the largest demonstrated benefits of AI/ML in the nuclear power industry is the use of AI/ML coupled with text recognition.

Survey Question 2

Survey Question 2 asks what areas of NPP operation and management will benefit the most, and the least, from AI/ML implementation. Nine of the 12 participants responded to this question, and their responses are summarized in Table 11.

Table 11 Summaries of Survey Responses to Survey Question 2

Participants	Summary of Response to Survey Question 2
FPL	Not specifying the most/least areas but providing NextEra’s main target areas to date, including operational/maintenance efficiency, system and component performance monitoring, and work management improvements.
Xcel	Suggesting that the areas that require significant repetitive manual input as the focus areas for future implementation. An example is the enhanced identification and communication of equipment conditions.
ForHumanity	Not directly relevant to the question.

Table 11 Summaries of Survey Responses to Survey Question 2 (continued)

Participants	Summary of Response to Survey Question 2
Blue Wave	Acknowledging most areas listed in the question description can benefit significantly. Specifying that fuel cycle planning and risk management have already benefitted from AI/ML uses, and there is an ongoing DOE-sponsored program on predictive maintenance. Adding one area, next generation reactor design, which is not included in the question description.
X-energy	Rating areas on a scale of 1–5 stars; one star for least beneficial and five stars for most: 5-star: cybersecurity, predictive maintenance, system and component performance monitoring; 4-star: physical security, inspection support; 3-star: power generation, safety/risk assessment; 2-star: incident response, operational/maintenance efficiency.
IEI	Largest benefits to: design, fuel management, outage reduction (both in number and duration). Other areas will have less impact either due to small return-of-investment or long development cycles due to risk and regulation.
NEI	Significant benefits to: system and component performance monitoring. Also beneficial to: predictive maintenance, reducing required man hours, aiding inspection analysis of videos and photographs, and cybersecurity.
Framatome	Significant benefits to: predictive maintenance, system and component monitoring, NDE inspections. May also have benefits to: achieving safe and efficient operation and allowing for semi-autonomous or autonomous operation.
WEC	Most beneficial: digital twins (validity and computational efficiency).

Survey Question 3

Survey Question 3 asks the potential benefits of incorporating AI/ML to commercial nuclear power operations in terms of three specific areas: design or operational automation, preventive maintenance trending, and improved reactor operations staff productivity. Ten of the 12 participants responded to this question, and their responses are summarized in Table 12.

Table 12 Summaries of Survey Responses to Survey Question 3

Participants	Summary of Response to Survey Question 3
SRI	Allow easier design changes through performing retraining Increase staff productivity through automating labor-intensive work and/or replacing workers with robots.
FPL	Preventive maintenance trending is a current project in progress. Equipment monitoring and early diagnostics is also in progress. Other projects support staff productivity efficiencies.

Table 12 Summaries of Survey Responses to Survey Question 3 (continued)

Participants	Summary of Response to Survey Question 3
Xcel	Efficiently identify and address the most important issues through automated data collection, computer-supported trending, and enhanced predictability and planning. Gain greater insights and understanding from the nuclear data and information produced.
ForHumanity	Not directly relevant to the question.
Blue Wave	Allow condition-based maintenance through developing remaining useful life models. Allow the creation of virtual sensors and virtual calibration tools.
X-energy	For design/operational automation: optimize designs, identify patterns unnoticed by humans, significantly save time of running simulations, suggest control strategies (including novel approaches not thought of beforehand), simultaneously tuning multiple controllers in control loops. For staff productivity improvement: reduce human workload through the uses of autonomous/automated control systems and smarter alarm systems.
IEI	Design/operation optimization will possibly reduce overall operational costs. Maintenance optimization might improve plant productivity. Staff productivity will also have an impact, but not as much, since labor cost in NPPs are dwarfed by big ticket items, such as fuel purchasing or outages.
NEI	Preventive maintenance trending allows for optimizing resource allocation. Use of AI/ML can improve preventive maintenance trending by incorporating many sources of information. Use of AI/ML can increase staff productivity by handling massive sensor-based/textual data, which would otherwise require many man hours.
Framatome	Reduce scheduled tasks and waive programmatic requirements through accurate monitoring for known root causes. Allow for planning intrusive maintenance activities outside of an emergent basis.
WEC	Reduce time for a design cycle and potentially reduce the testing workload. Appropriately scope predictive maintenance activities and potentially achieve financial benefit. Identify the most cost-effective strategies and optimize resource allocations Reduce or eliminate human tasks performed during daily plant procedures.

Survey Question 4

Survey Question 4 asks what AI/ML methods are either currently being used or will be in the near future. Nine of the 12 participants responded to this question, and their responses are summarized in Table 13.

Table 13 Summaries of Survey Responses to Survey Question 4

Participants	Summary of Response to Survey Question 4
FPL	The following methods are currently being used: NLP – for CAP trending; ANN – for searching information related to plant equipment and procedures; Clustering algorithms – for optimizing preventive maintenance scope and frequency.
Xcel	All the listed methods are being considered for use.
ForHumanity	Not directly relevant to the question.
Blue Wave	All the listed methods are currently being used. Several methods have been proven to be valuable – CNN (yielded breakthrough progress in BWR applications), clustering algorithms (such as k-means and DBSCAN), Pearson's, transfer learning
X-energy	Not specifying which methods are being used or to be used. Specifying that advanced reactor companies (like X-energy) can incorporate AI/ML easier since they are not restrained by in-place data infrastructure and could design and build a new data infrastructure using state-of-the-art technologies; this could allow them to incorporate AI/ML into just about any tool where it is deemed beneficial and appropriate.
IEI	All the listed methods are currently being used if considering the entire nuclear power industry. Adding one AI/ML area which is not listed – explainable AI.
NEI	Almost all the listed methods are currently being used with specifications for the following methods: Optical character recognition – for understanding handwritten or text in images; IBM Watson virtual assistant – for condition report classification process.
Framatome	The following methods are currently being used: Clustering algorithms – for anomaly detection; Gaussian approaches – for correcting instrument error; Bayesian method – employed by the Metroscope to find root causes; Autocorrelation methods (not listed) – being explored for NDE applications.
WEC	A lot of listed methods are already or currently being used. Applications of two methods, decision trees and random forests, are not known to the authors of the response without further search.

Survey Question 5

Survey Question 5 asks what the advantages or disadvantages are of a high-level, top-down strategic goal for developing and implementing AI/ML across a wide spectrum of general applications versus an ad-hoc, case-by-case targeted approach. Eight of the 12 participants responded to this question, and their responses are summarized in Table 14.

Table 13 Summaries of Survey Responses to Survey Question 5

Participants	Summary of Response to Survey Question 5
SRI	If a high-level, top-down approach is taken from the beginning, it can have more flexibility than targeted approaches.
FPL	Has seen advantages of a top-down approach for targeted cost savings and business efficiency.
Xcel	Currently on a case-by-case basis. This approach provides agility to review current, critical needs while evaluating the benefits in small, measured improvements. In the future, would be open to considering use of a common standard. This might create open, efficient lines of communication to streamline alignment within the regulatory and oversight process.
Blue Wave	There needs to be a top-down approach in support and funding for AI-based work in the nuclear power arena. The specific programs are naturally more purpose and problem specific. Our approach has been to work with utilities to identify the absolutely most critical problems.
X-energy	Top-down company strategy or top-down guidance from NRC? If from NRC, a similar example is the inclusion of risk-informed analysis criteria for all revised standards, which has experienced some challenges because there is no practical way to evaluate risk for some systems. Adding an AI/ML requirement in this manner would be a similar challenge. The case-by-case basis is more natural and targets application to the systems that are most conducive to AI/ML methods. If from individual companies, the top-down approach will make more sense. If applying a top-down approach, the data infrastructure must be built accordingly. This could be a challenge for established companies with the data infrastructure already in place but might not be an issue for advanced reactor companies who don't currently reply on existing infrastructures.
IEI	Both approaches are required for AI/ML success. The top-down approach ensures that AI/ML successes are repeatable and fundamental infrastructures and architectures are reusable; but the targeted approach allows the industry to align on immediate value delivery. The best practice is for organizations to create a Center of Excellence, focusing on a series of high-value point solutions and accumulating success experiences.
NEI	The advantages of a top-down approach within a company include: enables a holistic approach to choose one product that meets the needs of all possible AI use cases; creates a standardized data and solution architecture; has an ability to easily share knowledge and utilized learned experience; etc. The disadvantages of a top-down approach within a company include: time-consuming, framework limiting to a constantly changing technology landscape, and potential loss of employee insights on how to innovate and where the true value propositions exist.

Table 13 Summaries of Survey Response to Survey Question 5 (continued)

Participants	Summary of Response to Survey Question 5
Framatome	The advantage of high-level strategic goals is that many methods may be generalized to other areas with proper research and incremental development; the disadvantage is that real benefits in operation and maintenance relies heavily on expert knowledge to build the tools, and knowledge-based AI relies on a physics model or digital twin rather than pure data-driven methods. Having a top-down approach has strengths in providing a roadmap to synergize various data streams and generate holistic insights; however, an ad-hoc approach also has its own advantage in allowing for phased integration of AI/ML that supports the development of trust in the system.

Survey Question 6

In general, there are five groups of personality traits or phases in how people/industry accept an innovative technology: (1) innovators, who are willing to take risks and are the first ones to adopt an innovation; (2) early adopters, who adopt an innovation slower than innovators but quicker than other groups; (3) early majority, who adopt an innovation significantly after the innovators and early adopters but are still at or above average overall; (4) late majority, who adopt an innovation after the average time; and (5) laggards, who are the last to adopt an innovation.

Survey Question 6 asks which phase of AI/ML adoption the commercial nuclear power industry is currently experiencing. Nine of the 12 participants responded to this question, and their responses are summarized in Table 15.

Table 15 Summaries of Survey Responses to Survey Question 6

Participants	Summary of Response to Survey Question 6
FPL	Consider itself in the early adopter phase.
Xcel	Consider itself in the early adopter phase.
ForHumanity	Not directly relevant to the question.
Blue Wave	Consider the nuclear industry in the early adopter phase although an argument could made for the innovator phase.
X-energy	Consider the nuclear industry in either the late majority or laggard phase. Consider itself to be in the innovator phase.
IEI	Consider the nuclear industry in the innovator phase.
NEI	Consider the nuclear industry in the early adopter phase.
Framatome	Consider the nuclear industry ranging from early majority to late majority.
WEC	Not specify at which phase. For nuclear vendors, AI/ML is just beginning to be adopted at a large scale, with focuses on preventive maintenance and digital twins. For nuclear power industry, AI/ML is being readily incorporated by the larger nuclear power utilities (some have been investing in AI/ML for more than a decade); smaller nuclear utilities are just beginning to apply AI/ML on a limited basis.

Survey Question 7

Survey Question 7 asks the challenges in balancing the costs associated with AI/ML development and application, against plant operational and engineering benefits. Eight of the 12 participants responded to this question, and their responses are summarized in Table 16.

Table 16 Summaries of Survey Responses to Survey Question 7

Participants	Summary of Response to Survey Question 7
FPL	The high costs and regulatory requirements could hamper further development. However, NextEra has been able to develop business cases that balance the development costs against expected efficiency improvements.
Xcel	For each possible case, a business case is created to determine potential future value. AI/ML integration into nuclear processes are still at a very early stage; the current challenges surrounds data quality and availability.
ForHumanity	Good data governance and compliance-by-design can increase the development costs of software; however, the downside risks associated with defective, weak, or easy-to-fail software will likely result in harms that far outweigh the upfront costs.
Blue Wave	According to their experience, the benefits have been much higher compared to the cost of developing and maintaining AI-based tools.
X-energy	The existing nuclear fleet would have high costs of developing AI/ML tools; but for advanced reactor companies, such costs are negligible compared to design engineering and capital costs for building an NPP.
IEI	The high levels of risk-aversion and regulation add additional restrictions and thus add to the cost of developing AI solutions.
NEI	Implementing AI does have a considerable upfront cost, but all uses of AI/ML do not inherently result in timely benefits. Success is not guaranteed for the innovation projects like the use of AI/ML The readiness and culture of the organization to accept and adopt these innovative tools after development is another challenge.
Framatome	One challenge is capturing accurate and credible operations and maintenance cost data that supports the benefit evaluation in a return-on-investment study. Acceptance, qualification(s), and integration into existing NDE procedures can also be costly.

Survey Question 8

Survey Question 8 asks the general level of AI/ML expertise in the commercial nuclear power industry. Eight of the 12 participants responded to this question, and their responses are summarized in Table 17.

Table 17   Summaries of Survey Responses to Survey Question 8

Participants	Summary of Response to Survey Question 8
FPL	Has a team of data scientist that are experts in AI/ML. Work with several vendors that have expert data scientists.
Xcel	Utilizes external support through vendors, national labs, and universities. Internal talent is being hired and developed.
ForHumanity	Not directly relevant to the question.
Blue Wave	For the most part, the commercial nuclear power industry is at the beginner stage; rely on external organizations with AI expertise.
X-energy	Has an internal team to develop digital twin concepts, including AI/ML applications. Work with external partners on digital twin concepts; there are experts on the team with mixed nuclear engineering and data science backgrounds tackling the AI/ML applications. In the nuclear power industry, the interest in AI/ML applications is growing rapidly; relevant graduate programs are starting around computational analysis and AI-assisted design optimization; however, the expertise on human factors and autonomous control systems is less common.
IEI	In the nuclear power industry, AI/ML expertise is overall well-versed/skilled Specifically, the AI/ML and data-understanding capabilities of data scientist in the nuclear power industry are expert. But the needed adjacent skills of business analysis, MLOps, and organizational change management are much less mature.
NEI	Varied across the industry. In some utilities, AI/ML has been a focus with staffing aligned. Other utilities may have employees with AI/ML expertise but not yet assigned to these type of activities. In some cases, rely on outside entities (vendors and industry resources).
Framatome	The industry in terms of predictive maintenance. Well-versed in some areas, like cluster analysis for monitored anomalies. May be characterized as beginner when considering other tools. Framatome in terms of NDE. Well-versed/skilled. Also partners with vendors that are expert level in various AI/ML technologies.

Survey Question 9

Survey Question 9 asks how AI/ML will affect the commercial nuclear power industry in terms of efficiency, costs, and competitive positioning in comparison to other power generation sources. Eight of the 12 participants responded to this question, and their responses are summarized in Table 18.

Table 18 Summaries of Survey Responses to Survey Question 9

Participants	Summary of Response to Survey Question 9
FPL	With the use of AI/ML, expect improvements in reliability and efficiency, which will aid nuclear power to remain competitive and cost effective.
Xcel	Reduction of compliance cost through workload automation will allow for gaining cost efficiencies and improved performance.
ForHumanity	Reminding that AI/ML may be a double-edged sword; and the special feature of nuclear power industry could amplify the potential technology risks.
Blue Wave	Aggressive use of ML techniques has the potential to lower operating costs by 20–30%; to achieve these cost reductions, AI needs to be applied to operation and fuel programs in a great capacity. The use of drones and robots for inspection and repair in hazardous parts of the facility could reduce human labor costs significantly.
X-energy	Applying AI/ML could help nuclear power better survive in U.S. energy market.
IEI	AI/ML will have a similar competitive impact on the commercial nuclear power industry compared to other power generation sources. For other power generation sources, savings is generally lower impact and more distributed across more time; in the nuclear power industry, savings are more concentrated on fewer, higher impact events, such as fuel optimization and crud maintenance to drive better fuel re-use and fewer fuel purchases.
NEI	AI/ML will support staff in operation and management at increasingly reliable levels, resulting in early detection of incipient failures, optimizing resources and the timing of maintenance. This will improve efficiency, lower costs, and position nuclear power more favorably with competing carbon-free generation sources.
Framatome	The opportunity is significant in comparison with fossil generation due to typical operating and maintenance costs that may be reduceable with robust monitoring and diagnostics.

Survey Question 10

Survey Question 10 asks whether AI/ML has the potential to improve regulatory efficiency and/or effectiveness or otherwise affect regulatory costs associated with safety oversight. Eight of the 12 participants responded to this question, and their responses are summarized in Table 19.

Table 19 Summaries of Survey Responses to Survey Question 10

Participants	Summary of Response to Survey Question 10
FPL	The NRC may benefit from utilizing AI to review plant documentation to identify trends in performance or more efficiently analyze plant performance and issues. The added efficiency of NRC would also benefit the nuclear industry by reducing inspection burden.
Xcel	Provide the ability to assess compliance on a continual basis.
ForHumanity	At the outset, regulatory governance and oversight, if accomplished by independent audit of AI systems would result in decreased efficiency and increased cost; but this is likely a temporary state and could have tremendous gains in multiple aspects; compliance cost will stabilize over time, if not decline. The value for the NRC is anticipated to be enormous; compliance-in-a-box solutions could create a systemic funnel of normalized and automated compliance resulting in tremendous leverage for the NRC.
Blue Wave	AI-trained monitoring software could replace some portions of human surveillance; these systems could monitor data and written reports and detect problems/compliance issues before they occur.
X-energy	Use NLP to make the NRC ADAMS database more searchable and user friendly. As part of regulatory process, NRC should examine the model submitted by the applicants, run simulations, and evaluate if the simulation results are acceptable; surrogate model could be developed to save computational cost and facilitate running significantly more simulations of a proposed model, which might give the NRC more confidence in their decision-making.
IEI	While the potential exists for AI/ML to improve regulatory oversight, it is likely many years away from becoming a reality.
NEI	AI/ML is expected to improve the effectiveness associated with safety oversight based on improved equipment operation, fewer plant events, and improved performance indicators.
Framatome	Yes, if diagnostics can reasonably expand in coordination with risk-informed categorization, oversight might be improved by simplification of inspections and standardized rationale for maintenance deferral.

Survey Question 11

Survey Question 11 asks the concerns regarding data security. Ten of the 12 participants responded to this question, and their responses are summarized in Table 20.

Table 20 Summaries of Survey Responses to Survey Question 11

Participants	Summary of Response to Survey Question 11
SRI	The training environment, data, and model must all be protected. Since the model is the most important, it is likely important to store the model. The proprietary information in the training environment/data need to be protected.
FPL	Cybersecurity is critically important – NextEra has relevant policies and infrastructure controls in place to maintain data security.
Xcel	Follows both company and nuclear requirements pertaining to data security regulations in their first use of offsite networks. Overall security around data and information continues to strengthen as new tools and capabilities become available (i.e., encryption of data at rest and transit).
ForHumanity	The size and turnover of data is a new security vector. Data labeling attacks, model inversion, membership inference, and other data entry point attacks can render models useless or adversarial to the nuclear facility safety function. Large sums of data or source code present tremendous cover for malicious entry, highlighting a protocol concern about segmentation and separation of AI/ML/autonomous systems.
Blue Wave	Utilities have very sophisticated risk management programs for data/software. Data is encrypted both at rest and in transit; require frequented penetration tests and remediation efforts for any vulnerabilities uncovered by the tests Unlikely that significant insight could be gleaned from the purloined data, unless the hackers had access to the application software that generated the data.
X-energy	Cybersecurity and data security is a major concern. Plans to apply appropriate security controls to prevent unauthorized access of data.
IEI	Compared to other industries, nuclear industry has a smaller number of data, which can impact the ability to generate meaningful intelligence Attempts at accessing plant-specific data and aggregating data are also difficult.
NEI	Concerns include cybersecurity, proprietary information concerns, export control information, data curation challenges, and WiFi connectivity in the power plants.
Framatome	Concerns include cyber-intrusion, loss of export control, and enterprise risk from public and shareholder perception. Cloud systems now exist with cyber-controls that make data as secure as private intranet networks. Novel NDE AI/ML techniques are exploring AI platforms that consist of hardware that is collocated with inspection systems and are prepared to process data locally without the need to transfer.
WEC	Care must be taken to not allow co-mingling of data between various data sources. Security, privacy, export control compliance, and access needs are also key areas.

6.4 <u>Insights from Survey Responses</u>

This section provides the insights obtained from the survey responses. It should be noted that the insights and conclusions in this section are developed based on the collected responses from a limited survey scope and might not fully reflect the practice of the commercial nuclear power industry.

1. A broad range of potential AI/ML applications have been identified across the nuclear power industry to improve different plant aspects, and the development statuses of these applications are varied. According to the responses to Survey Question 1, most of the identified AI/ML applications are under concept exploration or strategic consideration, a small portion are under development, and a few are already developed and in place for plant use. The current areas of AI/ML development or use include textual report analysis, predictive maintenance, work management, fuel cycle management, reactor operation and control, surrogate model development, and supports to CAP, RCA, and NDE.
2. Several AI/ML tools have been developed and put in use for plant improvements. According to the responses to Survey Question 1, examples of the developed tools include:
    a. A tool to predict moisture carryover in BWRs (i.e., MCO.ai developed by the Blue Wave AI Labs).
    b. A tool to predict the BWR eigenvalue evolution for future fuel cycles (i.e., Eigenvalue.ai developed by the Blue Wave AI Labs).
    c. A tool to determine root causes derived from symptoms (i.e., Metroscope developed by the Électricité de France).
    d. A WEC-developed tool to evaluate multiple regression-based AI/ML algorithms to find trends in the data and select the optimal algorithm.

Besides these customized tools, commercial off-the-shelve software tools are also being used, such as the IBM Watson. Several survey participants also mentioned that some of their tools are not yet developed but are well underway, such as Xcel Energy's CAP Intelligence Advisor (targeting late 2021 for the first deployment) and X-Energy's Xe-100 Digital Twin (targeting 2025–2027 for the first deployment).

3. The survey participants held diverse views for the areas that could benefit the most and least from AI/ML applications. According to the responses to Survey Question 2, nine areas are deemed by the survey participants to be the most beneficial, including.
    a. System and component monitoring (mentioned by three participants, referred to as three votes).
    b. Predictive maintenance (two votes).
    c. Digital twins (one vote).
    d. NDE inspections (one vote).
    e. Automating human labor (one vote).
    f. Cybersecurity (one vote).
    g. Design support (one vote).
    h. Fuel management (one vote).
    i. Outage reduction (one vote).

System and component monitoring and predictive maintenance turn out to be the most-voted, most-beneficial areas. Most survey participants believed all the example areas listed in the question description could benefit from AI/ML applications to different extents and did not specify which areas are expected to benefit the least.

4. The survey participants expected benefits to the nuclear power industry of incorporating AI/ML in design or operational automation, preventive maintenance trending, and improved staff productivity; clear pathways can be envisioned to achieve these benefits. According to the responses to Survey Question 3, potential benefits through each of the three areas were extensively discussed, and the benefits through different areas considerably overlapped. The most mentioned expected benefits from these areas include:
    a. Increasing design-process efficiency.
    b. Enabling data collection and analysis at a larger scope and faster speed.
    c. Identifying patterns unnoticed by humans.
    d. Suggesting control strategies not necessarily thought of beforehand.
    e. Automating labor-intensive work.
    f. Optimizing resource allocation.
    g. Streamlining maintenance scheduling.
5. The commercial nuclear power industry has conducted or is currently conducting the tryouts for most AI/ML methods. According to the responses to Survey Question 4, all the AI/ML methods listed in the question description are already used or currently being used in the nuclear power industry. Two survey participants added that explainable AI and autocorrelation methods, which are not listed, are also important topics to be considered. Clustering algorithms, ANNs, and NLP are the most mentioned methods in the responses. Some of the methods, such as CNNs and clustering algorithms, have already been proven to be valuable through the existing applications.
6. Both the top-down approach and the case-by-case approach for developing and implementing AI/ML are deemed having their own pros and cons; no strong preference is demonstrated by the survey participants. The advantages and disadvantages of both approaches were extensively discussed by the participants in the responses to Survey Question 5.
    a. Commonly mentioned advantages of top-down approach include:
        i. Enabling a holistic and standardized framework.
        ii. Easier to generalize and save repetitive work.
        iii. Easier to share knowledge and experience.
        iv. Increasing business efficiency.
    b. Commonly mentioned disadvantages of top-down approach include:
        i. Difficulty in adapting the framework to a constantly changing technology landscape.
        ii. Challenge in developing a catchall strategy accommodating diverse applications.
        iii. Potential loss of innovative human inputs.

One survey participant also mentioned that the level of top-down guidance (i.e., from the NRC or within the company) could make a difference.

7. It is commonly believed that the nuclear power industry is in the early adopter phase of AI/ML technology adoption. According to the responses to Survey Question 6, a majority of survey participants identified either themselves or the nuclear power industry as an early adopter. A second most common belief is that the nuclear power industry is in the innovator phase. One survey participant considered the nuclear industry in either the late majority or laggard phase, while another survey participant considered ranging from early majority to late majority.
8. Most survey participants agreed that the high cost of developing and implementing AI/ML can be a challenge but that the net value of costs and benefits is also a significant decision driver. According to the responses to Survey Question 7, two survey participants that have experiences in completed or ongoing AI/ML applications mentioned that they were able to balance the development costs against expected plant improvements or have observed the benefits far outweighing the costs. But it is commonly believed that the costs are truly a concern when deciding future AI/ML development and implementation, since such costs are usually high and upfront while the benefits are neither timely achieved nor guaranteed.
9. The level of AI/ML expertise is overall well-versed/skilled in the nuclear power industry, and the sources of expertise (i.e., in-house, external, or a combination of both) is varied across the industry. According to the responses to Survey Question 8, the most common situation is developing in-house AI/ML talents and, in the meanwhile, obtaining expertise support from external entities, such as vendors, national laboratories, and universities. Several survey participants also mentioned that the level of expertise is varied with AI/ML capabilities, methods and tools, and application fields.
10. Most survey participants expected that AI/ML applications could improve nuclear power performance and cost efficiency and could boost its competitiveness in comparison to other power generation sources. According to the responses to Survey Question 9, the survey participants had consistent perspectives; they believed that applying AI/ML could improve nuclear power competitiveness and that the paths leading to these improvements seem to be clear. Some survey participants also mentioned that the nuclear power might benefit more from AI/ML applications when compared to other power generation sources, since the impacts of AI/ML on the nuclear power industry are usually concentrated on high-impact events, such as those related to nuclear fuel.
11. Most survey participants believed that AI/ML applications could improve the regulatory efficiency and effectiveness for nuclear power in direct or indirect ways. According to the responses to Survey Question 10, the direct ways of benefiting regulatory efficiency and effectiveness are on the NRC staff side, including using AI/ML to automate staff labors, such as reviewing plant documentation, using NLP to make the NRC ADAMS data more searchable, using surrogate modeling to reduce the computational cost of running simulation models submitted by the licensees, and adopting advanced oversight methods to streamline the regulatory process, such as coordinating diagnostics data with risk-informed categorization. The indirect ways of benefiting regulatory efficiency and effectiveness are on the utility side; one example is that AI/ML applications have potential in leading to safer plants with fewer events and thus reduce the number of regulatory activities. One survey participant also mentioned that integrating AI/ML into regulatory activities can be a learning process and that decreased efficiency and increased cost might be observed at the outset; but this is expected to be a temporary situation, and the costs will eventually stabilize, if not decline.

12. The survey participants believed that data security is critically important and must be well maintained; relevant protective policies are either in place or being actively developed. According to the responses to Survey Question 11, the most mentioned concerns related to data security include cyber-intrusion, proprietary information leakage, and loss of export control. Several survey participants mentioned that their organizations have mature sets of controlling policies in place and will continue to strengthen as new tools and capabilities become available. One participant mentioned that the data might have an inherent security—it would be difficulty to draw significant insights from the stolen data, unless the intruders had access to the original model and software. Another participant mentioned their organization is exploring AI platforms with collocated hardware and inspection systems to process data locally and minimize the need for data transfer.

7 EXPLORING POTENTIAL APPLICATIONS OF ADVANCED COMPUTATIONAL TOOLS AND TECHNIQUES TO OPERATING NUCLEAR PLANTS AND ADVANCED REACTORS

This section investigates potential applications of the advanced computational tools and techniques, including AI, ML, big data, and content analytics, to operating NPPs, advanced LWRs, and advanced NLWRs to improve plant safety and efficiency. Based on the literature reviews on the existing AI/ML applications in the nuclear industry covered in Section 4 and the insights derived from the federal register survey in Section 6, three main technological AFs are considered in this section:

AF 1: Plant safety and security assessments
- AF 1.1: Plant safety assessment - SSC reliability
- AF 1.2: Plant safety assessment - human reliability
- AF 1.3: Plant safety assessment - external events
- AF 1.4: Plant safety assessment - accidental radiological release and monitoring
- AF 1.5: Plant security assessment - cybersecurity and physical security

AF 2: Plant degradation modeling, fault and accident diagnosis and prognosis
- AF 2.1: Degradation modeling
- AF 2.2: Fault detection, diagnosis and prognosis (FDDP)
- AF 2.3: Accident detection, diagnosis and mitigation (ADDM)

AF 3: Plant operation and maintenance efficiency improvement
- AF 3.1: SSC operation and control optimization
- AF 3.2: Operator and SSC performance evaluation
- AF 3.3: SSC maintenance planning

Figure 7 illustrates how AI/ML applications in these three main technological AFs can make benefits to both NPP operators and the regulator for plant safety and efficiency. By introducing AI/ML techniques into these AFs, potential benefits for plant safety and efficiency include but are not limited to:

- Achievement of a better level of safety by
  - Removing/reducing failure sources
  - Developing better failure-preventing strategies
  - Developing better accident-mitigation strategies
- Enhancement of safety evaluation techniques by
  - Expanding safety evaluation scope
  - Improving safety evaluation accuracy
- Reduction of human and computational labor cost.

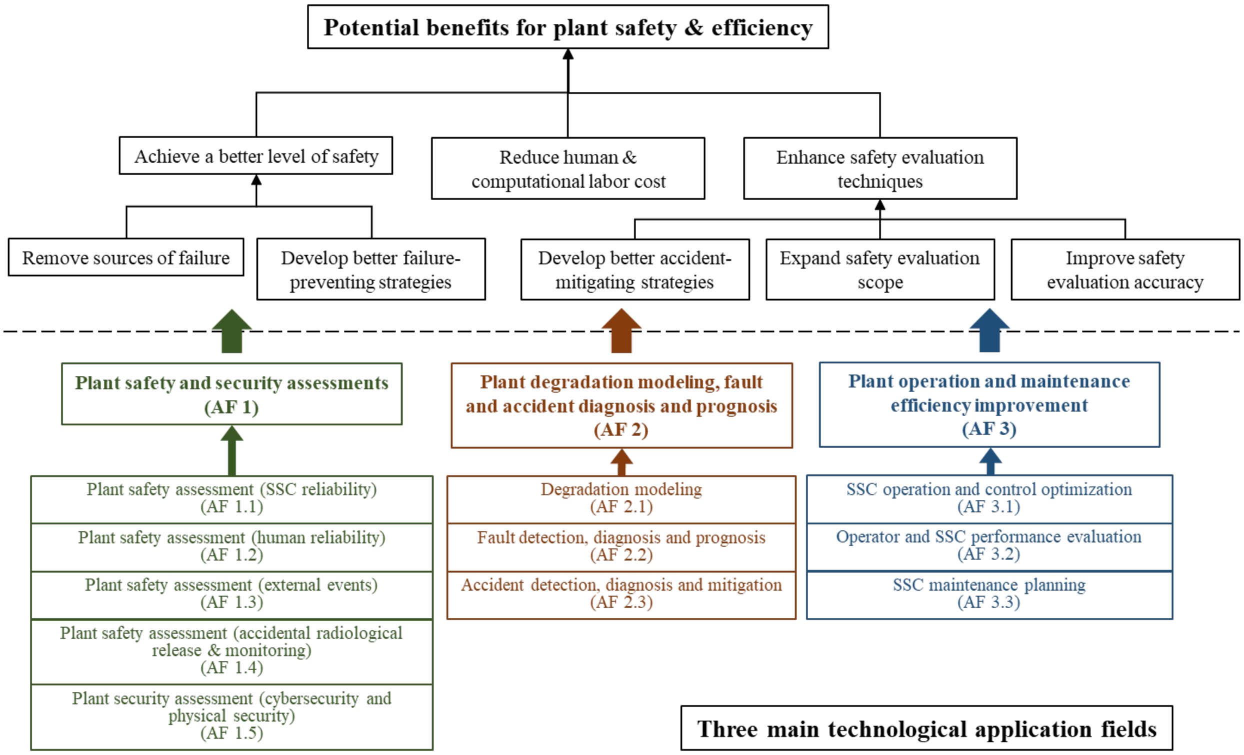

Figure 7 Potential Benefits for Plant Safety and Efficiency via AI/ML Applications in Three Main Technological Application Fields.[1]

This section evaluates the potential applicability of the new computational tools and techniques to inform and simplify the regulatory process on the operating NPPs and advanced reactors while simultaneously improving plant safety and efficiency and enhancing regulatory oversight. Details of these three main technological application fields are provided in Sections 7.1, 7.2, and 7.3, respectively.

7.1 Application Field 1: Plant Safety and Security Assessments

7.1.1 Plant Safety Assessment - System, Structure, Component Reliability

Some efforts prove that AI/ML techniques can be introduced in the analysis, evaluation, and enhancement of SSC reliability by providing an efficient and accurate prediction of SSC failure probability or reliability. Traditionally, this task is performed using PRA tools or reliability modeling methods with conventional statistical methodologies, which may have the limitations of inapplicability in some extrapolated conditions and be expensive computationally. By developing surrogate models that may have a better scalability and predictive capability when ML training data is sufficient, AI/ML techniques have the potential to improve SSC reliability analysis and evaluation in plant safety assessments. There are some demonstrations in this field. For example, Santhosh et al. (Santhosh et al. 2018) presents an integrated approach to predict the lifetime and reliability of I&C cables by ANNs from the accelerated aging data. Fink, Zio, and

[1] Note that the relationship between the application fields and the benefits in the figure could be cross-connecting (i.e., one AF might bring multiple benefits to the plant).

Weidmann (Fink et al. 2014) proposed multilayer ANNs based on multi-valued neurons, a specific type of complex valued neural networks, for reliability and degradation prediction problems. Wang et al. (Wang et al. 2019) constructed a Kriging regression model to avoid a large number of thermal hydraulics simulations for the reliability assessment of passive residual heat removal systems.

Various ML models have been applied for the SSC reliability analysis, including ANN, some kernel regression models, NLP, unsupervised ML methods such as classification and clustering, and others. These ML models can be used for either developing a purely data-driven model or building up a physics-guided surrogate model to support existing physical models or tools.

The major technical issue to be solved in the application of AI/ML in SSC reliability analysis is the inconsistency of ML training data and the data in full-size prototypic conditions. While ML training data mainly consists of numerical simulation data, some available experimental data, and very limited operating data, scale distortion may exist in the simulated conditions for training data generation and the real full-size prototypic conditions. However, as this field has been studied extensively, the knowledge and empirical correlations gained in past efforts can be utilized to guide the development and assessment of AI/ML models.

7.1.2 Plant Safety Assessment - Human Reliability

Similar to the SSC reliability analysis, human reliability analysis (HRA) has been applied in NPP PRA to identify potential human failure events; to systematically estimate the probability of these events using data, models, or expert judgment; and to evaluate the impacts of these events to key plant performance. Considering that human operators are adversely affected by excessive physical and mental workloads, AI/ML has the potential to be recommended for supporting human operators in tasks that may place them in unsafe conditions, as well as for better understanding and investigating how and why human errors have occurred to improve human performance in future operations. For example, Ham and Park (Ham and Park 2020) used a big data analysis technique called CART for extracting HRA data from event investigation reports. Park, Kim, and Jung (Park et al. 2017) applied CART to analyze the relative importance of performance shaping factors from event investigation reports for estimating the human error probability of a given task environment extracted from event investigation reports of NPPs. Zou et al. (Zou et al. 2018) introduced data mining for identifying intrinsic correlations among human factors.

These efforts mainly apply to unsupervised ML methods like classification, clustering, or regression trees to analyze the factors relevant to human error or performance in NPPs. Unsupervised learning is a type of ML algorithm that learns and captures patterns from untagged data, then builds a compact internal representation to generate imaginative content. Therefore, suggestions for improving human performance and preventing human errors in NPPs can be provided by these efforts.

The major technical issue to be solved in the application of AI/ML in HRA is still the insufficient data and knowledge in complex human errors from NPP operating experience. Accordingly, AI/ML-guided automation has been recommended for replacing the human operator in areas where the speed and accuracy of plant control and management cannot be satisfied by human operator performance. This application field will be discussed in later sections.

7.1.3 Plant Safety Assessment - External Events

External events include both natural external events (e.g., earthquakes, high winds, and external flooding) and human-made external events (e.g., airplane crashes, explosions at nearby industrial facilities, and impacts from nearby transportation activities). External events PRA is separated from internal events PRA because it has unique and specialized analysis methods for various kinds of external events. These external events normally have wide-area effects that may cause common-cause malfunctions of SSCs or combined failures to the entire plant. For different NPP designs, respective strategies are needed to prevent and mitigate the related failures and accidents led by these external events. PRA tools have been widely applied to external event analyses and can provide sufficient information and knowledge for constructing control and management strategies. However, an external event PRA model, such as a fire PRA model, could be large and needs expensive computation power to quantify. Researchers have suggested the introduction of AI/ML to provide for a more efficient external event analysis in plant safety assessment. Worrell et al. (Worrell et al. 2019) applied ML to generate metamodel approximations of a physics-based fire hazard model to generate accurate and efficient metamodels to improve modeling realism in PRAs without significant computational burdens. Sainct et al. (Sainct et al. 2020) developed an efficient methodology for seismic fragility curves estimation using SVMs. Wang, Zentner, and Zio (Wang et al. 2018b) estimated fragility curves based on seismic damage data and numerical simulations by ANNs.

Existing efforts of applying AI/ML in external event analyses have applied various ML and advanced statistical methods, including k-nearest neighbor modeling, mean-iterative neural networks, simple ANNs or deep neural networks (DNNs), SVMs, and others for scenario analyses and classification, clustering and regression trees for identification of external events. The major technical issues in this application field include the lack of data or knowledge for some rare external events, particularly some combinations of external events.

7.1.4 Plant Safety Assessment - Accidental Radiological Release and Monitoring

The rapid and accurate estimation of accidental radiological release is very important for nuclear safety and accident control and management decision-making. The source term information is typically unknown and uncontrollable once radioactive materials are released into the atmosphere. Relevant monitoring of the spreading of accidental radiological release is necessary. The severe nuclear accident at Chernobyl in 1986, for example, resulted in extraordinary contamination of the surrounding territory, as the monitoring of accidental radiological release is still ongoing.

In past decades, researchers have started to apply ML methods to better estimate the release rate, amount, and area of source terms or radioactive materials from NPPs operations and accidents. Briechle et al. (Briechle et al. 2020) developed a method to detect radioactive waste sites based on high-resolution remote sensing data using the random forest method. The results showed a good estimation of area-wide unknown radioactive biomass burials in the Chernobyl Exclusion Zone. Cho et al. (Cho et al. 2021) proposed a reproduction strategy using CNNs for radiation maps to compensate for the loss of radiation detection data. Sasaki et al. (Sasaki et al. 2021) applied ANNs to develop a new method of visualizing the ambient dose-rate distribution around the Fukushima Daiichi NPPs. Sun et al. (Sun et al. 2020) developed a methodology for optimizing the monitoring locations of long-term radiation air dose-rate monitoring near the Fukushima Daiichi NPPs. Zhang and Hu (Zhang and Hu 2020) proposed a real-time method for radionuclide estimation in NPP wastewater using ANNs.

Different ML methods or data-driven frameworks have been applied in this field, including typical ANNs, GP, random forest, generic algorithm (GA) and some advanced algorithms such as CNNs and RNNs. The usage of ML methods depends on the complexity of the database and latent physics inside. The main challenge of applying ML in this field is the lack of data for the validation of ML models and frameworks; however, based on high-fidelity simulation data generated using high-power computation, uncertainty quantification and reduction of these ML predictions can be performed.

7.1.5 Plant Security Assessment - Cybersecurity and Physical Security

Digital I&C systems offer significant advantages over existing analog systems in monitoring, processing, testing, and maintenance. In the last few years, the U.S. nuclear power industry initiated the replacement of existing aging analog systems with digital I&C technology and developed new designs for advanced plants using digital I&C systems in integrated control rooms to provide modern control and protection. However, cyber-vulnerabilities in digital systems and networks are also introduced and should be well addressed, prevented, and mitigated, particularly considering that some cyberattacks can result in software/digital common-cause failures of multiple SSCs with similar designs. Meanwhile, physical security is also important for NPPs to prevent external physical intrusion and terroristic sabotage.

Currently, some efforts have been made to enhance cybersecurity or physical security using AI/ML techniques. Zhang, Hines, and Coble (Zhang et al. 2020) proposed an ML-aided cybersecurity solution platform to improve cybersecurity by integrating process data together with traditional host system and network data in a unified platform. Kim, Lim, and Kim (Kim et al. 2018) developed an image-based intelligent intrusion detection system with a virtual fence, active intruder detection, classification, and tracking, and motion recognition to detect physical intrusion to NPPs. Some efforts are focused on building up a coupled cyber-physical system for NPPs. Gawand, Bhattacharjee, and Roy (Gawand et al. 2017) introduced least square approximation and GA to secure a cyber-physical system in NPPs.

These efforts applied to various ML methods like CNN, RNN, random forest regression, least square approximation, GA and others and demonstrated that ML methods can provide necessary technical support for the cybersecurity and physical security analysis. However, the identification of unfamiliar features of cyber-failures or physical intrusions constitutes a technical challenge to ML-based approaches for cybersecurity and physical security analysis. Additional efforts are needed to fill this gap in the future.

7.2 Application Field 2: Plant Degradation Modeling, Fault, and Accident Diagnosis and Prognosis

7.2.1 Degradation Modeling

As complex engineering systems, NPPs present a very harsh environment to their internal interacting and interdependent mechanical components, which must tolerate high-temperature water, stress, vibration, and an intense neutron field. Degradation of materials in this environment might lead to degraded plant performance or an unplanned shutdown with a loss of power generation and negative economic impact (U.S. Department of Energy 2008). Therefore, degradation modeling and online monitoring is necessary to address component aging problems and provide an accurate prediction of their failure points or remaining useful life (RUL) for on-time maintenance or replacement. Although various models have been developed for estimating material or component degradation, these models generally have fixed model forms or parameters, which leads to their limited applicability for some extrapolated conditions. Besides, traditional methods which account for degradation rely heavily on prior physics knowledge and expertise and may have limited capacity to learn from massive measured or simulated data. That is why there is the potential to apply ML methods to construct data-driven surrogate models with flexible and self-improvable model forms which can benefit when new data is available for model improvement.

There are some recent efforts on this topic. Alamaniotis, Ikonomopoulos, and Tsoukalas (Alamaniotis et al. 2012) proposed a probabilistic kernel approach for the intelligent online monitoring of mechanical components, GP is applied to the distribution prediction of a component's degradation trend. Baraldi, Mangili, and Zio (Baraldi et al. 2015) also used GPs to develop a stochastic model of the equipment degradation evolution, then applied it for estimating the distribution of the RUL by comparing it to a failure criterion. Wang et al. (Wang et al. 2021) proposed a RUL prediction method for electric valves using a convolutional autoencoder to extract features and RNNs to deal with time-series data. Sirola and Julsund (Sirola and Hulsund 2021) introduced ML methods to classify aging features and create prognostic models. Zhao and Wang (Zhao and Wang 2018) used DNNs to automatically extract centrifugal pump bearing degradation features from massive amounts of vibration data.

Depending on the complexity of involved physics, sufficiency of data, and internal dependency of degradation features, different ML methods have been introduced and demonstrated for degradation modeling, including GPs, simple ANNs, DNNs, RNNs, unsupervised learning, and support vector regression. Users should be careful regarding the selection of ML methods, and it is always advisable to have adequate training on the ML models and be guided by the physics at work.

7.2.2 Fault Detection, Diagnosis, and Prognosis

FDDP has been widely performed in existing NPPs to improve and ensure the reliability and availability of SSCs of nuclear reactors for plant safety and efficiency. Many FDDP techniques have been developed and applied to NPPs, which can be classified as physics-based and data-driven approaches. The physics-based approaches for FDDP are developed based on available measured or simulated data with limited applicability, they mainly rely on prior physics knowledge and expertise and do not require large amounts of data. But these methods may be not able to accurately predict the faults or NPP states under some unfamiliar and abnormal conditions.

In contrast, the data-driven approaches using ML methods can explore deep, complex, and highly nonlinear patterns from large amounts of data. They also have wide applicability and self-improving capability enabled by flexible ML models when new data becomes available. However, the explainability, interpretability, and trustworthiness of these ML-based data-driven models need more studies. The integration of data-driven and physics-based approaches (or hybrid physics-guided data-driven approaches) are promising to fill the knowledge and technical gaps by leveraging the advantages of each approach.

Many ML methods have been demonstrated, from supervised to unsupervised learning, from simple GPs, SVMs, ANNs to complicated DNNs, CNNs and RNNs. The main technical issues in the AI/ML applications in FDDP remain in how to improve their explainability, interpretability, and trustworthiness, especially considering that their deployment may affect the performance of highly safety-related safety-significant I&C systems.

7.2.3 Accident Detection, Diagnosis, and Mitigation

Similar to FDDP, ADDM plays an important role in NPP safety control and management. Fast and accurate ADDM can detect tiny and/or rare abnormal events in reactors that would be difficult for operators to identify, and support operators search for accident-mitigation strategies in the early phase of accident progression. Unlike FDDP, ADDM should be able to provide a long-term and rapid prediction of plant and system behaviors and states during various accidents. This long-term and rapid prediction of complex systems for unfamiliar accident scenarios suggests a significant requirement on the development and deployment of ADDM techniques.

Data-driven ADDM techniques have been developed and demonstrated to identify different accidental events, or predict their potential sequences, or provide suggestions on accident mitigation based on risk and cost estimation. A few ML-based data-driven techniques break the requirements regarding the need for data to training or develop the model. For example, Farber and Cole (Farber and Cole 2020) developed an automated fault detection tool to detect very small loss-of-coolant accidents in PWRs using only nominal operating data without the need for loss-of-coolant accidents data. Some tools provide integrated support for accident diagnosis, prognosis, and accident mitigation. Lin et al. (Lin et al. 2021a) developed and demonstrated a nearly autonomous management and control system for advanced reactors using ML-based digital twin technology. Lee, Seong, and Kim (Lee et al. 2018a) developed an autonomous algorithm including superior functions to monitor, control and diagnose automated subsystems.

Recently, developing ML-based digital twins for achieving autonomous control and operation became a trend in the nuclear industry and academic research. As automation levels increasingly rely on AI/ML techniques, the explainability, transparency, reliability, and trustworthiness of these digital-twin-enabled autonomous operation and control systems needs continual enhancement. An uncertainty quantification and software risk analysis is needed to evaluate the uncertainty of digital twins and their impacts on the reactor safety and efficiency. Reference (Lin et al. 2021b) provides a comprehensive review on relevant uncertainty quantification techniques and software risk analysis methods that may be suitable for ML-based digital twins for the development and assessment of (semi-)autonomous operation and control systems. To ensure consistency and transparency for the development of digital twins, a development and assessment process was suggested in (Lin et al. 2021b) to guide digital twin development and assessment according to target expectations as set out in the planning stage. It also indicated that "crucial software common-cause failures may occur in different ML-based digital twins for different intended uses in an autonomous system or in redundant digital twins for the same intended use, during the operation of ML-based digital twins and respective autonomous control systems that are designed for safety purposes."

7.3 Application Field 3: Plant Operation and Maintenance Efficiency Improvement

7.3.1 System, Structure, Component Operation and Control Optimization

To reduce operator workload from normal SSC operations in existing NPPs, such as reactor startup and shutdown, core optimization, load following, and pressurizer control, some AI/ML applications have been developed to optimize SSC operation and control processes and to provide advisory support to operators. These benefits to plant safety and efficiency cannot be achieved by traditional manual control. For example, Hosseini et al. (Hosseini et al. 2020) designed and applied a supervisory control using ANN-based controllers for the pressurizer system. Koo et al. (Koo et al. 2019a) developed an AI framework based on RNNs for startup and shutdown operation of NPPs. Norouzi et al. (Norouzi et al. 2013) introduced a Parallel Integer Coded Genetic Algorithm to obtain the best configuration for core optimization. Hui et al. (Hui et al. 2021) developed an adaptive backstepping control strategy with an extended state observer for the load following of NPPs.

Efforts in this field normally addresses one or two component-level operations. ML methods can help with the prediction of key parameters based on the training of relevant operating data. Different ML methods have been applied, such as GA, ANNs, RNNs, and NLP. A significant technical issue is that they rarely take system-level factors into consideration and usually only focus on component-level factors affecting performance of a specific operation. In some conditions, the impacts of system-level factors, such as the interactions between this operation function and other components or systems may affect the prediction accuracy of these ML models. Find a way to extend the applicability of these ML applications in different conditions requires more study, especially when system-level factors have significant impacts to the target operation function.

7.3.2 Operator and System, Structure, Component Performance Evaluation

In the main control room of nuclear reactors, human operators' attention level determines their performance on the task. Insufficient operator attention is one of the main causes of human error. To improve the efficiency of human operators during normal NPP operations, AI/ML methods have the potential to address the performance of human operators and SSCs. Progress on this topic has been made. Kim et al. (Kim et al. 2020b) investigated the development of quantitative indicators that can identify an operator's attention, and diagnose or detect a lack of operator attention thus preventing potential human errors in advanced control rooms. Experiments were designed to collect the electroencephalography and eye movement of the subjects who were monitoring and diagnosing nuclear operator safety-relevant tasks. Choi and Seong (Choi and Seong 2020) introduced an unsupervised learning technique, hierarchical clustering analysis, to find meaningful characteristics in the measured data. Wu et al. (Wu et al. 2020) used ANNs to predict the mental workload of an operator in nuclear reactors. The validity, sensitivity, and the relationship between the indices of eye tracking of both experts and nonexperts when they were operating the state-oriented procedure system in NPPs were analyzed. Yan, Yao, and Tran (Yan et al. 2021) also applied ANNs for predicting and evaluating the situation awareness of the operators. Kusumoputro, Sutarya, and Lina (Kusumoputro et al. 2013) developed an intelligent technique to classify the fuel pellet quality using ensemble back propagation neural networks.

Compared with the AI/ML applications in other application fields, the training of ML models for operator performance evaluation should have more data available. The main technical issue is the difficulty in collaborating with other sciences (e.g., biology, psychology, or sociobiology). Studies and experience in HRA can be leveraged here.

7.3.3 System, Structure, Component Maintenance Planning

The existing nuclear fleet relies on labor-intensive and time-consuming preventive maintenance programs to operate and maintain plant systems, resulting in high operation and maintenance costs. The implementation of an efficient predictive maintenance strategy is critical for the long-term safe and economical operation of plant systems. An application of ML-based solutions can lead to major cost savings, improved predictability, and the increased availability of plant systems.

Section 7.2.1 of this report has described and discussed how ML methods can be applied for degradation modeling, particularly the estimation of the RUL of plant SSCs, which provides insights into when the SSCs fail, and, accordingly respective maintenance planning can be made. Section 7.2.2 introduced how FDDP can provide benefits from the applications of ML methods, which give insights on the planning of maintenance actions after failures. ML methods can also be applied for normal plant SSC maintenance planning. For example, Dupin and Talbi (Dupin and Talbi 2020) developed a ML-guided dual heuristics and new lower bounds for the refueling and maintenance-planning. Gohel et al. (Gohel et al. 2020) used SVMs to explore and compare rare events that could occur in nuclear infrastructure to support the development of a predictive maintenance architecture. Musabayli, Osman, and Dirix (Musabayli et al. 2020) proposed a predictive maintenance mechanism for small steam sterilizers using classification models that categorized the health condition of two critical components in small steam sterilizers.

One potential issue is the application of ML-guided maintenance actions that may impact many separate plant SSCs. A comprehensive ML-guided maintenance strategy that covers the maintenance of all relevant plant SSCs should be beneficial to the integrated plant operation and efficiency.

8 CONCLUSIONS

This report presents the project INL conducted for the NRC to explore the advanced computational tools and techniques, such as AI, ML, and other analytics for operating NPPs and developing advanced computational predictive capabilities in nuclear OpE. The report first looks at the nuclear data that may be available and could be used in advanced computational tools and techniques. A categorization of nuclear data sources is presented, which focuses on different types of OpE data that may be applied through advanced computational tools and techniques. Section 3 presents an overview of advanced computational tools and techniques, including the relationships between statistics and AI/ML, and the most widely used AI/ML algorithms in both supervised and unsupervised learning. Section 4 reviews the existing applications of advanced computational tools and techniques, including AI/ML, in various fields of the nuclear industry, such as reactor system design and analysis, plant operation and maintenance, and nuclear safety and risk analysis. Section 5 provides the insights on the three questions under the purpose of Task 1 (i.e., the types of advanced computational tools and techniques that may be employed in nuclear industry, the aspects of advanced tools and techniques that could contribute to the increased understanding of safety and risk, and the types and quantities of information that would be needed for the new tools and techniques to generate safety and risk implications).

A survey on the current state of commercial nuclear power operations relative to the use of AI/ML tools as well as the role of AI/ML tools in nuclear power operations was published by the NRC in FRN NRC-2021-0048 in April 2021. Section 6 provides a summary of the survey, including the survey questions, survey participants, survey responses, and the conclusions and insights derived from the survey. The survey results could be used to enhance the understanding of the short- and long-term applications of AI and ML in nuclear power industry operations and management, as well as potential pitfalls and challenges associated with their applications.

Based on the literature reviews on existing AI/ML applications in nuclear industry covered in Section 4 and the insights derived from the federal register survey in Section 6, Section 7 provides an overview and insights into potential AI/ML applications aimed at improving NPP safety and efficiency. Three main application fields are considered: plant safety and security assessments; plant degradation modeling, fault, and accident diagnosis and prognosis; and plant operation and maintenance efficiency improvement. By introducing AI/ML techniques into these application fields, potential benefits for plant safety and efficiency include but are not limited to the achievement of a better level of safety; the enhancement of safety evaluation; and the reduction of human and computational labor costs. For each application field, the justification for using AI/ML methods, efforts and technical challenges is discussed.

9 REFERENCES

Agrež, M., et al. (2020). "Entropy and exergy analysis of steam passing through an inlet steam turbine control valve assembly using artificial neural networks." International Journal of Heat and Mass Transfer **156**: 119897. www.https://www.doi.org/10.1016/j.ijheatmasstransfer.2020.119897

Akkoyun, S., et al. (2013). "An artificial neural network application on nuclear charge radii." Journal of Physics G: Nuclear and Particle Physics **40**(5): 106-112. https://www.doi.org/10.1088/0954-3899/40/5/055106

Al Rashdan, A., et al. (2019). "Data integration aggregated model and ontology for nuclear deployment (DIAMOND): preliminary model and ontology, INL/EXT-19-55610," Idaho National Laboratory. https://www.lwrs.inl.gov/Advanced%20IIC%20System%20Technologies/DIAMOND_Prelimina ry_Model_and_Ontology.pdf

Al Rashdan, A. and S. St. Germain (2019). "Methods of data collection in nuclear power plants." Nuclear Technology **205**(8): 1062-1074. www.https://www.doi.org/10.1080/00295450.2019.1610637

Alamaniotis, M., et al. (2012). "Probabilistic kernel approach to online monitoring of nuclear power plants." Nuclear Technology **177**(1): 132-145. https://www.doi.org/10.13182/NT12-A13333

An, Y., et al. (2020). "Critical flow prediction using simplified cascade fuzzy neural networks." Annals of Nuclear Energy **136**: 107047. https://www.doi.org/10.1016/j.anucene.2019.107047

Ankerst, M., et al. (1999). "OPTICS: ordering points to identify the clustering structure." ACM Sigmod record **28**(2): 49-60. https://www.doi.org/10.1145/304181.304187

Athanassopoulos, A., et al. (2004). "Nuclear mass systematics using neural networks." Nuclear Physics A **743**(4): 222-235. https://www.doi.org/10.1016/j.nuclphysa.2004.08.006

Atwood, C. L., et al. (2003). "Handbook of parameter estimation for probabilistic risk assessment, NUREG/CR-6823." U.S. Nuclear Regulatory Commission. https://www.nrc.gov/reading-rm/doc-collections/nuregs/contract/cr6823/index.html

Ayo-Imoru, R. and A. Cilliers (2018). "Continuous machine learning for abnormality identification to aid condition-based maintenance in nuclear power plant." Annals of Nuclear Energy: 61-70. https://www.doi.org/10.1016/j.anucene.2018.04.002

Baldi, P. and K. Hornik (1989). "Neural networks and principal component analysis: Learning from examples without local minima." Neural Networks **2**(1): 53-58. https://www.doi.org/10.1016/0893-6080(89)90014-2

Bao, H., et al. (2019). "A data-driven framework for error estimation and mesh-model optimization in system-level thermal-hydraulic simulation." Nuclear Engineering and Design **349**: 27-45. https://www.doi.org/10.1016/j.nucengdes.2019.04.023

Bao, H., et al. (2020a). "Using deep learning to explore local physical similarity for global-scale bridging in thermal-hydraulic simulation." Annals of Nuclear Energy. https://www.doi.org/10.1016/j.anucene.2020.107684

Bao, H., et al. (2020b). "Computationally efficient CFD prediction of bubbly flow using physics-guided deep learning." International Journal of Multiphase Flow **131**: 103378. https://www.doi.org/10.1016/j.ijmultiphaseflow.2020.103378

Bao, H., et al. (2021). "Deep learning interfacial momentum closures in coarse-mesh CFD two-phase flow simulation using validation data." International Journal of Multiphase Flow **135**: 103489. https://www.doi.org/10.1016/j.ijmultiphaseflow.2020.103489

Baraldi, P., et al. (2015). "A prognostics approach to nuclear component degradation modeling based on Gaussian Process Regression." Progress in Nuclear Energy **78**: 141-154. https://www.doi.org/10.1016/j.pnucene.2014.08.006

Bensi, M. and K. Groth (2020). "On the value of data fusion and model integration for generating real-time risk insights for nuclear power reactors." Progress in Nuclear Energy **129**(0149-1970): 103497. https://www.doi.org/10.1016/j.pnucene.2020.103497

Berkan, R., et al. (1991). "Advanced automation concepts for large-scale systems." IEEE Control Systems Magazine **11**(6): 4-12. https://www.doi.org/10.1109/37.92985

Boroushaki, M., et al. (2003). "An intelligent nuclear reactor core controller for load following operations using recurrent neural networks and fuzzy systems." Annals of Nuclear Energy **30**(1): 63-80. https://www.doi.org/10.1016/S0306-4549(02)00047-6

Boser, B. E., et al. (1992). "A training algorithm for optimal margin classifiers", in Proceedings of Proceedings of the Fifth Annual Workshop on Computational Learning Theory, ACM Press. 10.1145/130385.130401https://www.dx.doi.org/10.1145/130385.130401

Breiman, L. (2001). "Random forests." Machine Learning **45**(1): 5-32. https://www.doi.org/10.1023/a:1010933404324

Breiman, L., et al. (1984). Classification and regression trees, CRC press.

Briechle, S., et al. (2020). "Detection of radioactive waste sites in the Chernobyl exclusion zone using UAV-based lidar data and multispectral imagery." ISPRS Journal of Photogrammetry and Remote Sensing **167**: 345-362. https://www.doi.org/10.1016/j.isprsjprs.2020.06.015

Bzdok, D., et al. (2018). "Statistics versus machine learning." Nature Methods **15**(4): 233-234. https://www.doi.org/10.1038/nmeth.4642

Cetiner, M. and P. Ramuhalli (2019). "Transformational challenge reactor – Autonomous control system framework and key enabling technologies, ORNL/SPR-2019/1178." Oak Ridge National Laboratory. https://www.doi.org/10.2172/1530084

Chang, C.-W. and N. Dinh (2019). "Classification of machine learning frameworks for data-driven thermal fluid models." International Journal of Thermal Sciences **135**: 559-579. https://www.doi.org/10.1016/j.ijthermalsci.2018.09.002

Chang, D., et al. (2019). "Accident diagnosis of a PWR fuel pin during unprotected loss of flow accident with support vector machine." Nuclear Engineering and Design **352**. https://www.doi.org/10.1016/j.nucengdes.2019.110184

Cheng, J. Z., et al. (2016). "Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in US images and pulmonary nodules in CT scans." Sci Rep **6**: 24454. https://www.doi.org/10.1038/srep24454

Cho, W., et al. (2021). "Reproduction strategy of radiation data with compensation of data loss using a deep learning technique." Nuclear Engineering and Technology. https://www.doi.org/10.1016/j.net.2021.01.012

Choi, G., et al. (2016). "Prediction of hydrogen concentration in nuclear power plant containment under severe accidents using cascaded fuzzy neural networks." Nuclear Engineering and Design **300**: 393-402. https://www.doi.org/10.1016/j.nucengdes.2016.02.015

Choi, M. K. and P. H. Seong (2020). "A methodology for evaluating human operator's fitness for duty in nuclear power plants." Nuclear Engineering and Technology **52**(5): 984-994. https://www.doi.org/10.1016/j.net.2019.10.024

Christian, R., et al. (2020). "Dynamic PRA-based estimation of PWR coping time using a surrogate model for accident tolerant fuel." Nuclear Technology. https://www.doi.org/10.1080/00295450.2020.1777035

Ciregan, D., et al. (2012). "Multi-column deep neural networks for image classification". Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE. https://www.doi.org/10.1109/CVPR.2012.6248110

Cireşan, D., et al. (2012). "Multi-column deep neural network for traffic sign classification." Neural Networks **32**: 333-338. https://www.doi.org/10.1016/j.neunet.2012.02.023

Clark, J. and H. Li (2006). "Application of support vector machines to global prediction of nuclear properties." International Journal of Modern Physics B **20**: 5015-5029. https://www.doi.org/10.1142/S0217979206036053

Colorado, D., et al. (2011). "Heat transfer using a correlation by neural network for natural convection from vertical helical coil in oil and glycerol/water solution." Energy **36**(2): 854-863. https://www.doi.org/10.1016/j.energy.2010.12.029

Cortes, C. and V. Vapnik (1995). "Support-vector networks." Machine Learning **20**(3): 273-297. https://www.doi.org/10.1007/bf00994018

Costrirs, N., et al. (2020). "A global model of β−-decay half-lives using neural networks." HNPS Advances in Nuclear Physics **15**(2654-0088): 210-217. http://www.dx.doi.org/10.12681/hnps.2640

Deng, L., et al. (2013). "New types of deep neural network learning for speech recognition and related applications: An overview". Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE. https://www.doi.org/10.1109/ICASSP.2013.6639344

Di Maio, D., et al. (2016a). "Transient identification by clustering based on Integrated Deterministic and Probabilistic Safety Analysis outcomes." Annals of Nuclear energy **87**: 217-227. https://www.doi.org/10.1016/j.anucene.2015.09.007

Di Maio, F., et al. (2016b). "A semi-supervised self-organizing map for post-processing the scenarios of an integrated deterministic and probabilistic safety analysis". Proceedings of the 13th International Conference on Probabilistic Safety Assessment and Management Conference, Seoul, South Korea. https://www.iapsam.org/PSAM13/program/Abstract/Oral/A-007.pdf

Dinh, N., et al. (2013). "Perspectives on nuclear reactor thermal hydraulics". Proceedings of the 15th International Topical Meeting on Nuclear Reactor Thermal Hydraulics (NURETH-15).

Dow, E. and Q. Wang (2011). "Quantification of Structural Uncertainties in the k-w Turbulence Model". Proceedings of the 52nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Denver, Colorado, AIAA. https://www.doi.org/10.2514/6.2011-1762

Dupin, N. and E. G. Talbi (2020). "Machine learning-guided dual heuristics and new lower bounds for the refueling and maintenance planning problem of nuclear power plants." Algorithms **13**(8). https://www.doi.org/10.3390/A13080185

Eide, S. A., et al. (2007). "Industry-average performance for components and initiating events at U.S. commercial nuclear power plants, NUREG/CR-6928" U.S. Nuclear Regulatory Commission. https://www.nrc.gov/reading-rm/doc-collections/nuregs/contract/cr6928/index.html

Electric Power Research Institute (2020a). "Automated analysis of remote visual inspection of containment buildings", https://www.epri.com/research/products/000000003002018419

Electric Power Research Institute (2020b). "Quick insight — Power industry dictionary for text-mining and natural language processing application: a proof of concept", https://www.epri.com/research/products/000000003002019609

Electric Power Research Institute (2021a). "Quick insight brief: Leveraging artificial intelligence for nondestructive evaluation", https://www.epri.com/research/products/000000003002021074

Electric Power Research Institute (2021b). "Quick insight brief: Leveraging artificial intelligence for the nuclear energy sector", https://www.epri.com/research/products/000000003002021067

Ester, M., et al. (1996). "A density-based algorithm for discovering clusters in large spatial databases with noise", in Proceedings of Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf?source=post_page

Estivill-Castro, V. (2002). "Why so many clustering algorithms: a position paper." ACM SIGKDD explorations newsletter **4**(1): 65-75. https://www.doi.org/10.1145/568574.568575

Farber, J., et al. (2018). "Using kernel density estimation to detect loss-of-coolant accidents in a pressurized water reactor." Nuclear Technology **205**(8): 1043-1052. https://www.doi.org/10.1080/00295450.2018.1534484

Farber, J. A. and D. G. Cole (2020). "Detecting loss-of-coolant accidents without accident-specific data." Progress in Nuclear Energy **128**. https://www.doi.org/10.1016/j.pnucene.2020.103469

Fernandez, M., et al. (2017). "Nuclear energy system's behavior and decision making using machine learning." Nuclear Engineering and Design **324**: 27-34. https://www.doi.org/10.1016/j.nucengdes.2017.08.020

Fink, O., et al. (2014). "Predicting component reliability and level of degradation with complex-valued neural networks." Reliability Engineering and System Safety **121**: 198-206. https://www.doi.org/10.1016/j.ress.2013.08.004

Frey, B. J. and D. Dueck (2007). "Clustering by passing messages between data points." Science **315**(5814): 972-976. https://www.doi.org/10.1126/science.1136800

Friedman, J. H. (1991). "Multivariate adaptive regression splines." The Annals of Statistics: 1-67. https://www.jstor.org/stable/2241837

Gao, W., et al. (2020). "Component detection in piping and instrumentation diagrams of nuclear power plants based on neural networks." Progress in Nuclear Energy **128**. https://www.doi.org/10.1016/j.pnucene.2020.103491

Gawand, H., et al. (2017). "Securing a cyber physical system in nuclear power plants using least square approximation and computational geometric approach." Nuclear Engineering and Technology **49**(3): 484-494. https://www.doi.org/10.1016/j.net.2016.10.009

Glauner, P. O. (2015). "Deep convolutional neural networks for smile recognition." arXiv preprint. https://www.arxiv.org/abs/1508.06535

Gohel, H. A., et al. (2020). "Predictive maintenance architecture development for nuclear infrastructure using machine learning." Nuclear Engineering and Technology **52**(7): 1436-1442. https://www.doi.org/10.1016/j.net.2019.12.029

Goodfellow, I., et al. (2016). Deep learning, MIT press.

Goodfellow, I., et al. (2014). "Generative adversarial nets". Proceedings of the 2014 Advances in Neural Information Processing Systems. https://www.proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3- Paper.pdf

Grishchenko, D., et al. (2016). "Development of a surrogate model for analysis of ex-vessel steam explosion in Nordic type BWRs." Nuclear Engineering and Design **310**: 311-327. https://www.doi.org/10.1016/j.nucengdes.2016.10.014

Ham, D. and J. Park (2020). "Use of a big data analysis technique for extracting HRA data from event investigation reports based on the Safety-II concept." Reliability Engineering & System Safety **194**: 106232. https://www.doi.org/10.1016/j.ress.2018.07.033

Hanna, B., et al. (2020a). "An application of ASP in nuclear engineering: Explaining the Three Mile Island nuclear accident scenario." Theory and Practice of Logic Programming **20**(6): 926-941. https://www.doi.org/10.1017/S1471068420000241

Hanna, B. N., et al. (2020b). "Machine-learning based error prediction approach for coarse-grid Computational Fluid Dynamics (CG-CFD)." Progress in Nuclear Energy **118**: 103140. https://www.doi.org/10.1016/j.pnucene.2019.103140

He, K., et al. (2016). "Deep residual learning for image recognition", in Proceedings of Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

He, M. and Y. Lee (2018). "Application of machine learning for prediction of critical heat flux: Support vector machine for data-driven CHF look-up table construction based on sparingly distributed training data points." Nuclear Engineering and Design **338**: 189-198.

Hinton, G. E., et al. (2006). "A fast learning algorithm for deep belief nets." Neural Computation **18**(7): 1527-1554. https://www.doi.org/10.1162/neco.2006.18.7.1527

Hobold, G. M. and A. K. da Silva (2018). "Machine learning classification of boiling regimes with low speed, direct and indirect visualization." International Journal of Heat and Mass Transfer **125**: 1296-1309. https://www.doi.org/10.1016/j.ijheatmasstransfer.2018.04.156

Hochreiter, S. and J. Schmidhuber (1996). "LSTM can solve hard long time lag problems." Advances in Neural Information Processing Systems **9**: 473-479. https://www.proceedings.neurips.cc/paper/1996/file/a4d2f0d23dcc84ce983ff9157f8b7f88- Paper.pdf

Hosseini, S. A., et al. (2020). "Design and application of supervisory control based on neural network PID controllers for pressurizer system." Progress in Nuclear Energy **130**. https://www.doi.org/10.1016/j.pnucene.2020.103570

Huang, G., et al. (2017). "Densely connected convolutional networks", in Proceedings of Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Hubel, D. H. and T. N. Wiesel (1968). "Receptive fields and functional architecture of monkey striate cortex." The Journal of Physiology **195**(1): 215-243. https://www.doi.org/10.1113/jphysiol.1968.sp008455

Hui, J., et al. (2021). "Adaptive backstepping controller with extended state observer for load following of nuclear power plant." Progress in Nuclear Energy **137**. https://www.doi.org/10.1016/j.pnucene.2021.103745

Jae, M. and J. Moon (2002). "Use of fuzzy decision-making method in evaluating severe accident management strategies." Annals of Nuclear Energy **29**(13): 1597-1606. https://www.doi.org/10.1016/S0306-4549(01)00125-6

James, G., et al. (2013). An introduction to statistical learning, Springer.

Kass, G. V. (1980). "An exploratory technique for investigating large quantities of categorical data." Applied Statistics **29**(2): 119. https://www.doi.org/10.2307/2986296

Kelly, D. and C. Smith (2011). Bayesian inference for probabilistic risk assessment: A practitioner's guidebook, Springer Science & Business Media.

Khajavi, M., et al. (2002). "A neural network controller for load following operation of nuclear reactors." Annals of Nuclear Energy **29**(6): 751-760. https://www.doi.org/10.1016/S0306- 4549(01)00075-5

Kim, D., et al. (2015). "Estimation of minimum DNBR using cascaded fuzzy neural networks." IEEE Transactions on Nuclear Science **62**(4): 1849-1856. https://www.doi.org/10.1109/TNS.2015.2457446

Kim, J., et al. (2020a). "Dynamic risk assessment with bayesian network and clustering analysis." Reliability Engineering & System Safety **201**(0951-8320): 106959. https://www.doi.org/10.1016/j.ress.2020.106959

Kim, J. H., et al. (2020b). "Biosignal-based attention monitoring to support nuclear operator safety-relevant tasks." Frontiers in Computational Neuroscience **14**. https://www.doi.org/10.3389/fncom.2020.596531

Kim, K. and E. Bartlett (1996). "Nuclear power plant fault diagnosis using neural networks with error estimation by series association." IEEE Transactions on Nuclear Science **43**(4): 2373-2388. https://www.doi.org/10.1109/23.531786

Kim, S. H., et al. (2018). "Intelligent intrusion detection system featuring a virtual fence, active intruder detection, classification, tracking, and action recognition." Annals of Nuclear Energy **112**: 845-855. https://www.doi.org/10.1016/j.anucene.2017.11.026

Kohonen, T. (1982). "Self-organized formation of topologically correct feature maps." Biological Cybernetics **43**(1): 59-69. https://www.doi.org/10.1007/BF00337288

Kohonen, T. and T. Honkela (2007). "Kohonen network." Scholarpedia **2**(1): 1568. http://www.scholarpedia.org/article/Kohonen_network

Koo, S. R., et al. (2019a). "Development of ai framework based on RNN for startup and shutdown operation of nuclear power plant." Journal of Institute of Control, Robotics and Systems **25**(9): 789-794. https://www.doi.org/10.5302/J.ICROS.2019.19.0104

Koo, Y., et al. (2019b). "Nuclear reactor vessel water level prediction during severe accidents using deep neural networks." Nuclear Engineering and Technology **51**(3): 723-730. https://www.doi.org/10.1016/j.net.2018.12.019

Kortelainen, J., et al. (2020). "Artificial intelligence for the support of regulator decision making" VTT Technical Research Centre of Finland.

Krizhevsky, A., et al. (2017). "ImageNet classification with deep convolutional neural networks." Communications of the ACM **60**(6): 84-90. https://www.doi.org/10.1145/3065386

Ku, C., et al. (1992). "Improved nuclear reactor temperature control using diagonal recurrent neural networks." IEEE Transactions on Nuclear Science **39**(6): 2298-2308. https://www.doi.org/10.1109/23.211440

Kusumoputro, B., et al. (2013). "Nuclear power plant fuel's quality classification using ensemble back propagation neural networks." Advanced Materials Research **685**: 367-371. https://www.doi.org/10.4028/www.scientific.net/AMR.685.367

Langley, P. (2011). "The changing science of machine learning." Mach Learn **82**: 275-279. https://www.doi.org/10.1007/s10994-011-5242-y

LeCun, Y., et al. (1989). "Backpropagation applied to handwritten zip code recognition." Neural Computation **1**(4): 541-551. https://www.doi.org/10.1162/neco.1989.1.4.541

Lee, D., et al. (2018a). "Autonomous operation algorithm for safety systems of nuclear power plants by using long-short term memory and function-based hierarchical framework." Annals of Nuclear Energy **119**: 287-299. https://www.doi.org/10.1016/j.anucene.2018.05.020

Lee, J., et al. (2018b). "Use of dynamic event trees and deep learning for real-time emergency planning in power plant operation." Nuclear Technology **205**(8): 1035-1042. https://www.doi.org/10.1080/00295450.2018.1541394

Lee, S. and J. Huh (2019). "An effective security measures for nuclear power plant using big data analysis approach." The Journal of Supercomputing **75**: 4267-4294. https://www.doi.org/10.1007/s11227-018-2440-4

Li, J., et al. (2014). "Sensitivity analysis of CHF parameters under flow instability by using a neural network method." Annals of Nuclear Energy **71**: 211-216. https://www.doi.org/10.1016/j.anucene.2014.03.040

Lin, L., et al. (2021a). "Development and assessment of a nearly autonomous management and control system for advanced reactors." Annals of Nuclear Energy **150**. https://www.doi.org/10.1016/j.anucene.2020.107861

Lin, L., et al. (2021b). "Uncertainty quantification and software risk analysis for digital twins in the nearly autonomous management and control systems: A review." Annals of Nuclear Energy **160**: 108362. https://www.doi.org/10.1016/j.anucene.2021.108362

Ling, J. and J. Templeton (2015). "Evaluation of machine learning algorithms for prediction of regions of high Reynolds averaged Navier Stokes uncertainty." Physics of Fluids **27**: 085103. https://www.doi.org/10.1063/1.4927765

Liu, Y., et al. (2018). "Data-driven modeling for boiling heat transfer: using deep neural networks and high-fidelity simulation results." Applied Thermal Engineering **144**: 305-320. https://www.doi.org/10.1016/j.applthermaleng.2018.08.041

Lloyd, S. (1982). "Least squares quantization in PCM." IEEE Transactions on Information Theory **28**(2): 129-137. https://www.doi.org/10.1109/tit.1982.1056489

Long, J., et al. (2015). "Fully convolutional networks for semantic segmentation", in Proceedings of Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

https://www.openaccess.thecvf.com/content_cvpr_2015/html/Long_Fully_Convolutional_Netw orks_2015_CVPR_paper.html

Ma, D., et al. (2017). "Supercritical water heat transfer coefficient prediction analysis based on BP neural network." Nuclear Engineering and Design **320**: 400-408. https://www.doi.org/10.1016/j.nucengdes.2017.06.013

Maljovec, D., et al. (2015). "Analyzing simulation-based PRA data through traditional and topological clustering: A BWR station blackout case study." Reliability Engineering & System Safety **145**: 262-276. https://www.doi.org/10.1016/j.ress.2015.07.001

Mandelli, D., et al. (2013a). "Dynamic PRA: an overview of new algorithms to generate, analyze and visualize data". Transactions of the American Nuclear Society, Washington, DC, American Nuclear Society.

Mandelli, D., et al. (2018). "Cost risk analysis framework (CRAFT): An integrated risk analysis tool and its application in an industry use case, INL/EXT-18-51442." Idaho National Laboratory. https://www.doi.org/10.2172/1495190

Mandelli, D., et al. (2013b). "Scenario clustering and dynamic probabilistic risk assessment." Reliability Engineering and System Safety **115**: 146-160. https://www.doi.org/10.1016/j.ress.2013.02.013

Matsugu, M., et al. (2003). "Subject independent facial expression recognition with robust face detection using a convolutional neural network." Neural Networks **16**(5): 555-559. https://www.doi.org/10.1016/S0893-6080(03)00115-1

Merriam-Webster (2021). "Data",

Moura, R., et al. (2017). "Learning from accidents: Interactions between human factors, technology and organisations as a central element to validate risk studies." Safety Science **99**: 196-214. https://www.doi.org/10.1016/j.ssci.2017.05.001

Musabayli, M., et al. (2020). "Classification model for predictive maintenance of small steam sterilisers." IET Collaborative Intelligent Manufacturing **2**(1): 1-13. https://www.doi.org/10.1049/iet-cim.2019.0029

Na, M., et al. (2003). "Sensor monitoring using fuzzy neural network with an automatic structure constructor." IEEE Transactions on Nuclear Science **50**(2): 241-250. https://www.doi.org/10.1109/TNS.2003.809471

Na, M., et al. (2006). "Design of a model predictive power controller for an SPH-100 space reactor." Nuclear Science and Engineering **154**(3): 353-366. https://www.doi.org/10.13182/NSE06-A2638

Nabeshima, K., et al. (2012). "Real-time nuclear power plant monitoring with neural network." Journal of Nuclear Science and Technology **35**(2): 93-100. https://www.doi.org/10.1080/18811248.1998.9733829

Ng, A. Y., et al. (2002). "On spectral clustering: Analysis and an algorithm", in Proceedings of Proceedings of the 2002 Advances in neural information processing systems.

Norouzi, A., et al. (2013). "Nuclear reactor core optimization with parallel integer coded genetic algorithm." Annals of Nuclear Energy **60**: 308-315. https://www.doi.org/10.1016/j.anucene.2013.05.013

Nuclear Energy Institute (2011). "Industry guideline for monitoring the effectiveness of maintenance at nuclear power plants, NUMARC 93-01",

Nuclear Energy Institute (2018). "Monitoring the effectiveness of nuclear power plant maintenance, NEI 18-10",

Park, H., et al. (2020). "Wall temperature prediction at critical heat flux using a machine learning model." Annals of Nuclear Energy **141**: 107334. https://www.doi.org/10.1016/j.anucene.2020.107334

Park, J., et al. (2017). "Use of a big data mining technique to extract relative importance of performance shaping factors from event investigation reports". Proceedings of the AHFE 2017 International Conference on Human Error, Reliability, Resilience, and Performance, Los Angeles, California, USA, Springer, Cham.

Park, J. and P. Seong (2002). "An integrated knowledge base development tool for knowledge acquisition and verification for NPP dynamic alarm processing systems." Annals of Nuclear Energy **29**(4): 447-463. https://www.doi.org/10.1016/S0306-4549(01)00054-8

Pastore, A., et al. (2017). "A new statistical method for the structure of the inner crust of neutron stars." Journal of Physics G: Nuclear and Particle Physics **44**: 094003. https://www.doi.org/10.1088/1361-6471/aa8207

Pence, J., et al. (2020). "Data-theoretic approach for socio-technical risk analysis: Text mining licensee event reports of U.S. nuclear power plants." Safety Science **124**. https://www.doi.org/10.1016/j.ssci.2019.104574

Podofillini, L., et al. (2010). "Dynamic safety assessment: Scenario identification via a possibilistic clustering approach." Reliability Engineering & System Safety **95**(5): 534-549. https://www.doi.org/10.1016/j.ress.2010.01.004

Poolsappasit, N., et al. (2012). "Dynamic security risk management using Bayesian attack graphs." IEEE Transactions on Dependable and Secure Computing **9**(1): 61-74. https://www.doi.org/10.1109/TDSC.2011.34

Quinlan, J. R. (1986). "Induction of decision trees." Machine Learning **1**(1): 81-106. https://www.doi.org/10.1007/bf00116251

Quinlan, J. R. (1993). C4.5: Programs for machine learning, Elsevier Science.

Ramaswamy, P., et al. (1993). "An automatic tuning method of a fuzzy logic controller for nuclear reactors." IEEE Transactions on Nuclear Science **40**(4): 1253-1262. https://www.doi.org/10.1109/TNS.1993.8526778

Rokach, L. and O. Maimon (2005). Clustering methods. Data mining and knowledge discovery handbook, Springer**:** 321-352.

Rostamifard, D., et al. (2011). "Empirical correlation study of dryout heat transfer at high pressure using high order neural network and feed forward neural network." Heat Mass Transfer **47**: 439-448. https://www.doi.org/10.1007/s00231-010-0733-0

Roweis, S. T. and L. K. Saul (2000). "Nonlinear dimensionality reduction by locally linear embedding." Science **290**(5500): 2323-2326. https://www.doi.org/10.1126/science.290.5500.2323

Rumelhart, D. E., et al. (1986). "Learning representations by back-propagating errors." Nature **323**(6088): 533-536. https://www.doi.org/10.1038/323533a0

Sainct, R., et al. (2020). "Efficient methodology for seismic fragility curves estimation by active learning on Support Vector Machines." Structural Safety **86**. https://www.doi.org/10.1016/j.strusafe.2020.101972

Santhosh, T. V., et al. (2018). "An approach for reliability prediction of instrumentation & control cables by artificial neural networks and Weibull theory for probabilistic safety assessment of NPPs." Reliability Engineering and System Safety **170**: 31-44. https://www.doi.org/10.1016/j.ress.2017.10.010

Sasaki, M., et al. (2021). "New method for visualizing the dose rate distribution around the Fukushima Daiichi Nuclear Power Plant using artificial neural networks." Scientific Reports **11**(1). https://www.doi.org/10.1038/s41598-021-81546-4

Schölkopf, B., et al. (1998). "Nonlinear component analysis as a kernel eigenvalue problem." Neural Computation **10**(5): 1299-1319. https://www.doi.org/10.1162/089976698300017467

Sekimizu, K., et al. (1992). "Knowledge representation for automated boiling water reactor startup." Nuclear Technology **100**(3): 295-309. https://www.doi.org/10.13182/NT92-A34726

Shin, J., et al. (2017). "Cyber security risk evaluation of a nuclear I&C using BN and ET." Nuclear Engineering and Technology **49**(3): 517-524. https://www.doi.org/10.1016/j.net.2016.11.004

Simonyan, K. and A. Zisserman (2014). "Very deep convolutional networks for large-scale image recognition." arXiv preprint. https://www.arxiv.org/abs/1409.1556

Singh, A. P., et al. (2017). "Machine-learning-augmented predictive modeling of turbulent separated flows over airfoils." AIAA Journal **55**: 2215-2227. https://www.doi.org/10.2514/1.J055595

Sirola, M. and J. E. Hulsund (2021). "Machine-learning methods in prognosis of ageing phenomena in nuclear power plant components." International Journal of Computing **20**(1): 11-21. https://www.doi.org/10.47839/ijc.20.1.2086

Siu, N., et al. (2013). "Knowledge engineering tools–an opportunity for risk-Informed decision making?". Proceedings of the ANS PSA 2013 International Topical Meeting on Probabilistic Safety Assessment and Analysis, Columbia, SC,, Americal Nuclear Society.

Siu, N., et al. (2016). "Advanced knowledge engineering tools to support risk-informed decision making: final report (public version)" U.S. Nuclear Regulatory Commission. https://www.nrc.gov/docs/ML1635/ML16355A373.pdf

Suman, S. (2020). "Artificial intelligence in nuclear industry: Chimera or solution?" Journal of Cleaner Production: 124022. https://www.doi.org/10.1016/j.jclepro.2020.124022

Sun, D., et al. (2020). "Optimizing long-term monitoring of radiation air-dose rates after the Fukushima Daiichi Nuclear Power Plant." Journal of Environmental Radioactivity **220-221**. https://www.doi.org/10.1016/j.jenvrad.2020.106281

Szegedy, C., et al. (2015). "Going deeper with convolutions", in Proceedings of Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Tenenbaum, J. B., et al. (2000). "A global geometric framework for nonlinear dimensionality reduction." Science **290**(5500): 2319-2323. https://www.doi.org/10.1126/science.290.5500.2319

Tian, D., et al. (2018). "A neural networks design methodology for detecting loss of coolant accidents in nuclear power plants." Applications of Big Data Analytics: 43-61. https://www.doi.org/10.1007/978-3-319-76472-6_3

Tin Kam, H. (1998). "The random subspace method for constructing decision forests." IEEE Transactions on Pattern Analysis and Machine Intelligence **20**(8): 832-844. https://www.doi.org/10.1109/34.709601

Tracey, B., et al. (2013). "Application of supervised learning to quantify uncertainties in turbulence and combustion modeling". Proceedings of the 51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition, Grapevine (Dallas/Ft. Worth Region), Texas.

Trontl, K, et al. (2008). "Machine learning of the reactor core loading pattern critical parameters." Science and Technology of Nuclear Installations **2008**: 695153. https://www.doi.org/10.1155/2008/695153

U.S. Department of Energy. (2008). "Materials degradation in light water reactors: Life after 60." https://www.energy.gov/ne/downloads/materials-degradation-light-water-reactors-life-after-60

U.S. Nuclear Regulatory Commission (2021). "Role of artificial intelligence tools in U.S. commercial nuclear power operations", https://www.federalregister.gov/documents/2021/04/21/2021-08177/role-of-artificial-intelligence-tools-in-us-commercial-nuclear-power-operations

Utama, R., et al. (2016). "Nuclear charge radii: density functional theory meets bayesian neural networks." Journal of Physics G: Nuclear and Particle Physics **43**(11): 114002. https://www.doi.org/10.1088/0954-3899/43/11/114002

Vaddi, P., et al. (2020). "Dynamic bayesian networks based abnormal event classifier for nuclear power plants in case of cyber security threats." Progress in Nuclear Energy **128**: 103479. https://www.doi.org/10.1016/j.pnucene.2020.103479

Van Der Maaten, L., et al. (2009). "Dimensionality reduction: a comparative." J Mach Learn Res **10**(66-71): 13.

Vinod, S., et al. (2003). "Symptom based diagnostic system for nuclear power plant operations using artificial neural networks." Reliability Engineering & System Safety **82**(1): 33-40. https://www.doi.org/10.1016/S0951-8320(03)00120-0

Wang, C., et al. (2019). "Reliability assessment of passive residual heat removal system of IPWR using Kriging regression model." Annals of Nuclear Energy **127**: 479-489. https://www.doi.org/10.1016/j.anucene.2018.12.040

Wang, H., et al. (2021). "Remaining useful life prediction techniques for electric valves based on convolution auto encoder and long short term memory." ISA Transactions **108**: 333-342. https://www.doi.org/10.1016/j.isatra.2020.08.031

Wang, J., et al. (2017). "Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data." Physical Review Fluids **2**: 034603. https://www.doi.org/10.1103/PhysRevFluids.2.034603

Wang, Z., et al. (2018a). "Seismic fragility analysis with artificial neural networks: Application to nuclear power plant equipment." Engineering Structures **162**: 213-225. https://www.doi.org/10.1016/j.engstruct.2018.02.024

Wang, Z., et al. (2018b). "A Bayesian framework for estimating fragility curves based on seismic damage data and numerical simulations by adaptive neural networks." Nuclear Engineering and Design **338**: 232-246. https://www.doi.org/10.1016/j.nucengdes.2018.08.016

Ward Jr, J. H. (1963). "Hierarchical grouping to optimize an objective function." Journal of the American Statistical Association **58**(301): 236-244. https://www.doi.org/10.1080/01621459.1963.10500845

Worrell, C., et al. (2019). "Machine learning of fire hazard model simulations for use in probabilistic safety assessments at nuclear power plants." Reliability Engineering and System Safety **183**: 128-142. https:/www./doi.org/10.1016/j.ress.2018.11.014

Wu, J., et al. (2017). "Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework." Physical Review Fluids **2**: 034603. https://www.doi.org/10.1103/PhysRevFluids.3.074602

Wu, Y., et al. (2020). "Using artificial neural networks for predicting mental workload in nuclear power plants based on eye tracking." Nuclear Technology **206**(1): 94-106. https://www.doi.org/10.1080/00295450.2019.1620055

Yan, S., et al. (2021). "Using artificial neural network for predicting and evaluating situation awareness of operator." IEEE Access. https://www.doi.org/10.1109/ACCESS.2021.3055345

Young, J., et al. (2004). "LER data mining pilot study final report, PNNL-14910" Pacific Northwest National Laboratory. https://www.doi.org/10.2172/15020763

Zhang, F., et al. (2020). "A robust cybersecurity solution platform architecture for digital instrumentation and control systems in nuclear power facilities." Nuclear Technology **260**(7): 939-950. https://www.doi.org/10.1080/00295450.2019.1666599

Zhang, Y. J. and L. S. Hu (2020). "Real time estimation of radionuclides in the receiving water of an inland nuclear power plant based on difference gated neural network." Radiation Physics and Chemistry **176**. https://www.doi.org/10.1016/j.radphyschem.2020.109019

Zhao, L. and X. Wang (2018). "A deep feature optimization fusion method for extracting bearing degradation features." IEEE Access **6**: 19640-19653. https://www.doi.org/10.1109/ACCESS.2018.2824352

Zhao, Y., et al. (2019). "Automated identification of causal relationships in nuclear power plant event reports." Nuclear Technology **205**(8): 1021-1034. https://www.doi.org/10.1080/00295450.2019.1580967

Zou, Y., et al. (2018). "A Data Mining Framework Within the Chinese NPPs Operating Experience Feedback System for Identifying Intrinsic Correlations Among Human Factors." Annals of Nuclear Energy **116**: 163-170. https://www.doi.org/10.1016/j.anucene.2018.02.038

APPENDIX A
RECENT APPLICATIONS OF ADVANCED COMPUTATIONAL TOOLS AND TECHNIQUES IN NUCLEAR INDUSTRY

Table A-1 Review of Applications of Advanced Computational Tools in Reactor System Design and Analysis after 2000

1. Applications in Reactor Thermal Hydraulics				
Reference	Adv. Comp. Tool	Data Category	Objective	
(Chang and Dinh 2019)	FNN, CNN	Simulated data	Classification of ML applications in nuclear thermal hydraulics	
(Hobold and da Silva 2018)	FNN, SVM	Experimental data	Flow regime identification	
(Bao et al. 2021)	FNN	Simulated data; experimental data	Model/code uncertainty analysis/error estimation	Two-phase flow
(Ling and Templeton 2015)	SVM, DT, RF	Simulated data		Turbulence flow
(Singh et al. 2017)	FNN	Simulated data		Turbulence flow
(Wu et al. 2017)	RF	Simulated data		Turbulence flow
(Wang et al. 2017)	RFs	Simulated data		Turbulence flow
(Dow and Wang 2011)	GP	Simulated data		Turbulence flow
(Tracey et al. 2013)	Kernel Regression	Simulated data		Turbulence flow
(Hanna et al. 2020b)	FNN	Simulated data		Turbulence flow
(Liu et al. 2018)	FNN	Simulated data		Boiling
(Bao et al. 2019)	FNN, RF	Simulated data		Mixed convection
(Bao et al. 2020a)	FNN	Simulated data		Mixed convection
(Bao et al. 2020b)	FNN, RF	Simulated data		Two-phase flow

Table A-1 Review of Applications of Advanced Computational Tools in Reactor System Design and Analysis After 2000 (continued)

(He and Lee 2018)	SVM	Simulated data	Closure model development	Critical heat flux prediction
(Li et al. 2014)	FNN	Experimental data		Sensitivity analysis of CHF parameters
(Rostamifard et al. 2011)	FNN	Experimental data		Dryout heat transfer coefficient prediction
(Colorado et al. 2011)	FNN	Experimental data		Heat transfer coefficient prediction
(Ma et al. 2017)	FNN	Experimental data		Supercritical water heat transfer coefficient prediction
2. Applications in Reactor Physics				
Reference	Adv. Comp. Tool	Data Category	Objective	
(Akkoyun et al. 2013)	FNN	Experimental data	Model development for prediction	Mass-dependent nuclear charge radii formula
(Athanassopoulos et al. 2004)	FNN	Experimental data		Statistical model of nuclidic masses
(Costrirs et al. 2020)	FNN, SVM	Experimental data		Beta-decay half-lives model
(Trontl et al. 2008)	SVM	Simulated data		Reactor core loading pattern critical parameters
(Clark and Li 2006)	SVM	Experimental data		Prediction of nuclear properties
(Utama et al. 2016)	BNN	Experimental data		Nuclear charge radii
(Pastore et al. 2017)	GP	Simulated data		Proton component of the inner crust of a neutron star
3. Applications in Reactor System Performance				
Reference	Adv. Comp. Tool	Data Category	Objective	
(Agrež et al. 2020)	FNN	Real process data from the thermal power plant	Entropy and exergy analysis for steam in valves	
(Park et al. 2020)	FNN	Simulated data	Wall temperature at critical heat flux	
(Lee et al. 2018b)	CNN	Simulated data	Expected offsite release & timing	
(Chang et al. 2019)	SVM	Simulated data	Peak cladding temperature and flow change during LOFA	
(An et al. 2020)	Simplified Cascade Fuzzy NNs	Simulated data	Critical mass flowrate	
(Kim et al. 2015)	Cascade Fuzzy NNs	Simulated data	Minimum departure from nucleate boiling ratio prediction	

Table A-1 Review of Applications of Advanced Computational Tools in Reactor System Design and Analysis After 2000 (continued)

(Koo et al. 2019b)	FNN	Simulated data	Reactor vessel water level prediction during severe accidents
(Choi et al. 2016)	Cascade Fuzzy NNs	Simulated data	Hydrogen concentration
(Grishchenko et al. 2016)	FNN	Simulated data	Steam explosion

Table A-2 Review of Applications of Advanced Computational Tools in Plant Operation and Maintenance

1. Applications in Plant Operation and Maintenance			
Reference	Adv. Comp. Tool	Data Category	Objective
(Lin et al. 2021a)	FNN, RNN	Simulated data	Autonomous management and control
(Cetiner and Ramuhalli 2019)	Bayesian network	Simulated data	Supervisory control system
(Hanna et al. 2020a)	Answer Set Programming	Simulated data	Operator support system
(Lee et al. 2018a)	Long Short-Term Memory	Simulated data	Autonomous operation algorithm for core damage prevention
(Na et al. 2006)	Genetic Algorithm; PCA	Simulated data	Autonomous control system for a space reactor system
(Sekimizu et al. 1992)	Knowledge-based techniques	Simulated data	Automated startup of BWR
(Berkan et al. 1991)	Reconstructive inverse dynamics, fuzzy logic and FNN	Simulated data, plant sensor signal	Automated startup control system
(Ramaswamy et al. 1993)	Fuzzy logic	Simulated data	An automatically tuned fuzzy logic controller for optimal plant performance
(Ku et al. 1992)	RNN	Simulated data	Nuclear reactor temperature control
(Khajavi et al. 2002)	FNN	Simulated data	Load following operation
(Boroushaki et al. 2003)	RNN	Simulated data	Core controller
(Jae and Moon 2002)	Fuzzy logic	Simulated data	Emergency operating procedure
(Park and Seong 2002)	Knowledge-base tool	Tested using plant sensor data	Alarm processing system
(Kim and Bartlett 1996)	FNN	Simulated data	Fault detection and identification
(Na et al. 2003)	Fuzzy NN	Plant sensor data	Sensor validation and detection
(Fernandez et al. 2017)	FNN	Sensor data from test facility	Prediction of plant behaviors
(Vinod et al. 2003)	FNN	Simulated data	Symptom based diagnostic system
(Nabeshima et al. 2012)	FNN	Sensor data from test facility	Real-time nuclear power plant monitoring
(Ayo-Imoru and Cilliers 2018)	FNN	Simulated data	Abnormality identification
(Gao et al. 2020)	FNN	Simulated data	Component detection

Table A-2. Review of Applications of Advanced Computational Tools in Plant Operation and Maintenance (continued)

2. Applications in Plant Cyber Security			
Reference	Adv. Comp. Tool	Data Category	Objective
(Zhang et al. 2020)	k-nearest neighbor; DT; bootstrap aggregating; RT; auto-associative kernel regression; PCA	Simulated data	Cybersecurity solution platform
(Gawand et al. 2017)	least squares approximation	Simulated data	Cyber physical system
(Poolsappasit et al. 2012)	Bayesian methods	Simulated data	Dynamic security risk management
(Shin et al. 2017)	Bayesian network	Simulated data	Cyber security risk evaluation
(Lee and Huh 2019)	Unsupervised ML (classification); reinforcement learning	Simulated data	Security measures
(Vaddi et al. 2020)	Bayesian network	Simulated data	Abnormal event classifier

Table A-3 Review of Applications of Advanced Computational Tools in Nuclear Safety and Risk Analysis

Reference	Adv. Comp. Tool	Data Category	Objective
(Christian et al. 2020)	GP, SVM, k-nearest-neighbor classifier and regressor, Shepard's method, and spline interpolation method	Simulated data	Estimation of PWR coping time
(Mandelli et al. 2013b)	Mean-Shift methodology	Simulated data	Scenario clustering
(Kim et al. 2020a)	Dynamic Bayesian network and clustering	Simulated data	Risk assessment of a dynamic system
(Podofillini et al. 2010)	Unsupervised ML (clustering)	Simulated data	Scenario identification
(Zhao et al. 2019)	Natural language processing	Unstructured free text	Extracting the causal relationships from free-text reports
(Park et al. 2017)	CART	Structured data from reports	Extracting relative importance of performance shaping factors
(Zou et al. 2018)	A data mining framework combining three statistical approaches (i.e., correlation analysis, cluster analysis and association rule mining)	Structured data from reports	Identifying intrinsic correlations among human factors
(Farber et al. 2018)	Kernel Density Estimation	Simulated data	Loss-of-coolant accidents detection
(Bensi and Groth 2020)	Bayesian networks	Different data sources as sensor data, operational data, Simulated data	Generating real-time risk insights
(Moura et al. 2017)	Unsupervised ML (clustering)	Unstructured Free Text	Validating risk studies using information from past major accidents
(Di Maio et al. 2016a)	Unsupervised ML (clustering)	Simulated data	Transient identification
(Mandelli et al. 2013a)	Unsupervised ML (clustering)	Simulated data	Dynamic PRA
(Maljovec et al. 2015)	Unsupervised ML (clustering)	Simulated data	Analyzing simulation-based PRA data
(Tian et al. 2018)	FNN	Simulated data	Detecting loss of coolant accidents
(Worrell et al. 2019)	k-nearest neighbor, SVM, DT	Simulated data	Fire hazard model simulation
(Wang et al. 2018a)	FNN	Simulated data	Seismic fragility analysis

Table A-3 Review of Applications of Advanced Computational Tools in Nuclear Safety and Risk Analysis (continued)

Reference	Adv. Comp. Tool	Data Category	Objective
(Di Maio et al. 2016b)	Semi-Supervised ML (clustering)	Simulated data	Post-processing the multi-valued dynamic scenarios
(Ham and Park 2020)	CART	Simulated data	Extracting HRA data from event investigation reports
(Mandelli et al. 2018)	Supervised ML (Classification)	Unstructured free text	Cost risk analysis
(Young et al. 2004)	Unsupervised ML (clustering)	Unstructured free text	Data mining pilot study
(Siu et al. 2016)	Natural language processing	Unstructured free text	Supporting risk-informed decision making

NRC FORM 335 (12-2010) NRCMD 3.7 — U.S. NUCLEAR REGULATORY COMMISSION	1. REPORT NUMBER (Assigned by NRC, Add Vol., Supp., Rev., and Addendum Numbers, if any.)
BIBLIOGRAPHIC DATA SHEET *(See instructions on the reverse)*	NUREG/CR-7294

2. TITLE AND SUBTITLE	3. DATE REPORT PUBLISHED	
Exploring Advanced Computational Tools and Techniques with Artificial Intelligence and Machine Learning in Operating Nuclear Plants	MONTH: February	YEAR: 2022
	4. FIN OR GRANT NUMBER: IAA-31310019N0006	

5. AUTHOR(S)	6. TYPE OF REPORT
Z. Ma, H. Bao, S. Zhang, M. Xian, A. Mack	Technical
	7. PERIOD COVERED (Inclusive Dates)

8. PERFORMING ORGANIZATION - NAME AND ADDRESS (If NRC, provide Division, Office or Region, U. S. Nuclear Regulatory Commission, and mailing address; if contractor, provide name and mailing address.)

Idaho National Laboratory
1955 Fremont Ave
Idaho Falls ID 83402-1510

9. SPONSORING ORGANIZATION - NAME AND ADDRESS (If NRC, type "Same as above", if contractor, provide NRC Division, Office or Region, U. S. Nuclear Regulatory Commission, and mailing address.)

Division of Risk Analysis
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, D.C. 20555-0001

10. SUPPLEMENTARY NOTES

11. ABSTRACT (200 words or less)

This report presents the project Idaho National Laboratory conducted for the Nuclear Regulatory Commission (NRC) to explore the advanced computational tools and techniques, such as artificial intelligence (AI) and machine learning (ML), for operating nuclear plants. The report reviews the nuclear data sources, with the focus on operating experience data, that could be applied by advanced computational tools and techniques. Plant-specific and generic (national and international) data from different sources are described. The report describes the relationships between statistics and AI/ML and then introduces the most widely used AI/ML algorithms in both supervised and unsupervised learning. The report reviews the recent applications of advanced computational tools and techniques in various fields of nuclear industry, such as reactor system design and analysis, plant operation and maintenance, and nuclear safety and risk analysis. The report presents the insights from the project on the potential applicability of AI/ML techniques in improving advanced computational capabilities, how the advanced tools and techniques could contribute to the understanding of safety and risk, and what information would be needed to provide meaningful insights to decision makers.

12. KEY WORDS/DESCRIPTORS (List words or phrases that will assist researchers in locating the report.)	13. AVAILABILITY STATEMENT
PRA, artificial intelligence, machine learning, nuclear plant, probabilistic risk assessment, operating experience, data	unlimited
	14. SECURITY CLASSIFICATION
	(This Page) unclassified
	(This Report) unclassified
	15. NUMBER OF PAGES
	16. PRICE

UNITED STATES
NUCLEAR REGULATORY COMMISSION
WASHINGTON, DC 20555-0001

OFFICIAL BUSINESS

@NRCgov

NUREG/CR-7294

Exploring Advanced Computational Tools and Techniques with Artificial Intelligence and Machine Learning in Operating Nuclear Plants

February 2022