PNNL-31699

Report on Next-Gen AI for Proliferation Detection Workshop: Domain-Aware Methods

February 2021

Tammie L Borders[1,2]
Maria F Glenski[3]
Thomas F Grimes[3]
Jennifer M Mendez[3]
Brienne N Seiner[3]
Angela M Sheffield[2]
Ashley J B Shields[1]
Jesse Ward[3]
Marc A Wonders[2,3]

1 Idaho National Laboratory
2 National Nuclear Security Administration
3 Pacific Northwest National Laboratory

U.S. DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062
www.osti.gov
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: http://www.ntis.gov

Report on Next-Gen AI for Proliferation Detection Workshop: Domain-Aware Methods

February 2021

Tammie L Borders[1,2]
Maria F Glenski[3]
Thomas F Grimes[3]
Jennifer M Mendez[3]
Brienne N Seiner[3]
Angela M Sheffield[2]
Ashley J B Shields[1]
Jesse Ward[3]
Marc A Wonders[2,3]

1 Idaho National Laboratory
2 National Nuclear Security Administration
3 Pacific Northwest National Laboratory

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Executive Summary

Introduction

The emergence of artificial intelligence (AI) and machine learning (ML) in the modern world has impacted nearly every application imaginable. This includes nuclear proliferation detection, which offers the potential to improve existing capabilities as well as create new ones. Proliferation detection seeks to detect and characterize attempts by state and non-state actors to acquire nuclear weapons or associated technology, materials, or knowledge. Such a mission is vitally important for global stability and security but is notoriously difficult. By leveraging advances in AI, exciting opportunities exist to enhance the proliferation detection regime.

The Data Science and AI portfolio within the National Nuclear Security Administration's Office of Defense Nuclear Nonproliferation Research and Development (DNN R&D) seeks to leverage the capabilities of the Department of Energy's (DOE's) national laboratories and other partners to develop AI systems that can accomplish otherwise impossible tasks in support of proliferation detection. As part of its efforts, the portfolio has created a series of workshops on *Next-Gen AI for Proliferation Detection* to help define the requirements for suitable AI systems, share successful research and best practices, and foster connection and understanding between the relevant parties including researchers and end-users. Each workshop in the series focuses on a specific and critical aspect of AI to enable it to accomplish proliferation detection objectives. The first workshop focused on explainability techniques; the second workshop and the topic of this report, covers methods for incorporating domain awareness into AI. The *Next-Gen AI for Proliferation Detection Workshop: Domain-Aware Methods* took place virtually over two days in February 2021 and included four keynote presentations, 22 technical presentations, and a concluding panel. The presentations, discussions, and workshop findings are summarized in this report.

Requirements and Opportunities for Next-Gen AI for Nuclear Proliferation Detection

While existing AI has demonstrated impressive performance in many industry and academic scenarios, such systems do not necessarily translate to successful systems in proliferation detection. The environment in which proliferation detection algorithms operate is substantially different than those in which many AI systems are created, and many challenges exist. Challenges include:

- Complex and noisy environments: signals of interest are typically faint and hidden among challenging backgrounds.
- Sparse data and rare events: limited training data exist, and events of interest occur infrequently.
- Robust deployment and decision support: accounting for all possible scenarios is impossible, and the environments in which algorithms are deployed differ from those in which they are trained.
- Early proliferation detection and signature discovery: novel signatures are desired to advance proliferation detection to early stages.

These themes define the sessions covered in this workshop, the structure of which was chosen to emphasize the importance of considering these problems and the development of AI technologies in the context of mission applications. Domain-aware methods offer one

opportunity to address these challenges. Such methods can combine the well-established expertise in proliferation detection that has been developed over decades with the power of AI in a synergistic manner that avoids ignoring hard-earned knowledge and understanding.

Keynote Presentations on Actors in Proliferation Detection

Two keynote presentations described in more detail the proliferation detection mission space and the roles of various actors in it. Within the United States, the DOE, the Department of Defense (DoD), the Department of Homeland Security (DHS), the intelligence community, and other Executive Branch entities work together to contribute to the vital mission. Each of these organizations apply the tools of proliferation detection and demonstrate the scope of end-users for which new AI technologies can benefit. The second keynote presentation dove further into the intelligence community and its requirements. It was shared that the intelligence community is actively seeking to incorporate AI into its work and is ready to do so. Nonetheless, a key message was that intelligence analysis is inherently a human activity, and all approaches must remain human-centric, even as the purview of AI expands. All other methods are doomed to failure.

Methods and Applications for Implementing Domain-Aware Techniques

Despite the importance of domain-aware techniques and the extended length of time they have been applied, there remains no accepted taxonomy. Two introductory and keynote presentations delved into this absence in more detail and attempted to provide some structure to domain-aware techniques. Five categories were highlighted:

- Expert knowledge
- Synthetic data generation
- Inclusion of non-traditional AI/ML methods into traditional AI/ML models
- Semantic or constraint-based methods
- Soft labels

Examples of each were provided, particularly in a nuclear applications context, and it was also found that many of these techniques are typically used together rather than in isolation from other methods of domain awareness. Challenges associated with a lack of data and opportunities to leverage domain awareness were also introduced and motivated the forthcoming “Sparse Data and Rare Events” session.

Complex and Noisy Environments

This technical session focused on how to extract valid and useful information from complex and noisy data sets. Three themes emerged. First, it was clearly shown that the incorporation of domain knowledge greatly improves model performance. Second, combined approaches led to better results than those of single analysis methods. Third, even if the volume of available data is large, events of interest are rare and consequently the data is still “sparse.” These discoveries were found in various applications of nuclear proliferation detection from the analysis of radiation detection data to the analysis of nuclear material transfers.

Early Proliferation Detection and Signature Discovery

Ultimately, the goal of proliferation detection and a major opportunity for AI is to detect proliferation at the earliest stage possible. With the increase in volume and variety of publicly available data, discovery of early-stage proliferation activities is being realized. It is likely this will require the development of novel signatures enabled by AI and these additional data. Presentations in this session focused on non-traditional data sources, open-source data sources, and combining these with traditional detection approaches to lead to new insights about proliferation activities. A key finding was that incorporating domain-aware methods with data-driven approaches yields contextual information for ongoing analysis. Incorporating social sciences and psychology was identified as a future opportunity.

Sparse Data and Rare Events

This session focused heavily on the problem of having insufficient data and the associated shortcomings of purely data-driven methods. This frequently manifests as the “small n, large p” problem, whereas there are few events of interest usable for training but a multitude of data and variables to process. Domain-aware methods have been critical to addressing this problem, for example by reducing the scale of measurements and enabling the creation of new data. Nonetheless, these activities require careful consideration in their successful use. Techniques are thoughtfully selected to mathematically represent all information in the data sets and preventing overfitting is a key focus. In this session, nearly all presenters performed some type of feature engineering. Regardless of the method chosen, the application of domain awareness turned an otherwise intractable problem into a solvable one. Ultimately, it was stated that reliance on data-driven models without domain knowledge will lead to spectacular failures in the wild.

Robust Deployment and Decision Support

A key theme that emerged in this session was the necessity of integrating guidance, feedback, or knowledge representations from end-users as a requirement for building robust models. This echoes similar emphases throughout the workshop to incorporate domain experts early and often in the lifecycle to develop and deploy AI-based solutions. Human-in-the-loop feedback was a key component of the systems presented. Additionally, several presentations illustrated how domain-aware methods can help validate AI models. Finally, other presentations highlighted how domain awareness can improve the quality of data used, for example by reducing noise and removing inconsequential features.

Panel Discussion on Domain-Aware Methods

Following the many technical presentations to share research and specific approaches to domain-aware AI, a panel concluded the workshop by focusing on overarching considerations and to refocus on the mission itself. In the panel, the importance of creating collaborations between data science and domain experts was highlighted. This follows from the realization that purely data-driven techniques fail in proliferation detection, even despite the success of such techniques in other areas. Fortunately, domain-aware methods are rapidly being developed and incorporated into operational AI. Specific opportunities for AI were discussed, such as in enabling decision-making superiority, and again, the intelligence community is actively preparing to increase the presence of AI in its operations.

Conclusion

While there remain challenges in developing deployable AI for national security applications, there are tremendous possibilities moving forward. Analytics methods that are solely data-driven are insufficient in national security because data is sparse, incomplete, and noisy. Data-driven approaches forego inclusion of key mission-relevant information found in subject matter expertise, computational simulations, mission requirements, and other traditional domain-aware methods and data sources. This workshop demonstrated a variety of ways in which domain-aware methods can be used to overcome these shortcomings. Further, domain-aware approaches are key to improving generalizability and transferability and to ensure the creation of useful and robust models suitable for high-consequence missions. The time is ripe to expand the role of AI in nuclear proliferation detection. Not only is progress being made on the technology, but end-users recognize the potential of AI to create new capabilities. AI systems do not necessarily need to be perfect to be helpful, and through interactions with end-users, researchers can identify opportunities to make substantial impacts, including in the near-term.

Acknowledgments

The workshop organizers would like to thank all workshop participants and attendees for their engagement that created a successful workshop. Special thanks to Boian Alexandrov, Zoe Gastelum, Becky Olinger, and Stefan Hau-Riege for chairing the technical sessions; Emma Hague and Kary Myers for inspiring keynote presentations; and Patria Smith for her role in managing the meeting logistics. They also thank Mark Greaves for his role in creating a successful workshop and to Kelly Machart for graphic design and support. Finally, they would like to thank Sarah Logue, Susan Tackett, and Jan Haigh for their valuable technical editing and reference support of the report. The organizers would also like to thank Brienne Seiner for managing the document through the production.

Acronyms and Abbreviations

ABC	attribution-based confidence
ADAPD	Advanced Data Analytics for Proliferation Detection
AGAT	Attribute-Guided Adversarial Training
AI	artificial intelligence
ANN	artificial neural network
BEADS	baseline estimation and denoising
CBRN	chemical, biological, radiological, and nuclear
CDF	cumulative density function
COTS	commercial off-the-shelf
CWMD	countering weapons of mass destruction
DAG	Dry Alluvium Geology
DANN	domain adversarial neural network
DHS	Department of Homeland Security
DNN	Defense Nuclear Nonproliferation
DNN R&D	Office of Defense Nuclear Nonproliferation Research and Development
DoD	U.S. Department of Defense
DOE	U.S. Department of Energy
DOE-IN	Office of Intelligence and Counterintelligence
FFS	Forward Feature Selection
GAN	generative adversarial network
HFIR	High Flux Isotope Reactor
HMM	hidden Markov models
IAEA	International Atomic Energy Agency
IC	intelligence community
ICA	intelligent cognitive assistants
KMS	knowledge management system
LINAC	linear accelerator
LOCO	Leave One Covariant Out
LWIR	longwave infrared
MINOS	Multi-Informatics for Nuclear Operations Scenarios
ML	machine learning
NILM	non-intrusive load monitoring
NNSA	National Nuclear Security Administration
ORNL	Oak Ridge National Laboratory
PGD	projected gradient descent
REDC	Radiochemical Engineering Development Center

ROC	receiver operating characteristic
SME	subject matter expert
SNM	special nuclear material
UAV	unmanned aerial vehicle
WMD	weapons of mass destruction

Contents

Figures

1.0 Introduction

Nuclear proliferation detection faces a constantly changing landscape including both geopolitical developments and technological advances. Perhaps the most significant technological development, in this context and beyond, is the emergence of artificial intelligence (AI) and machine learning (ML), which is already pervasive throughout everyday life. AI, enabled by continuously increasing computing power, offers the ability to analyze data in new ways and to revolutionize approaches to tackling a myriad of problems. For proliferation detection, examples of unique opportunities arising from the use of AI include the discovery of new signatures, leveraging of underused data modalities, and the ability to analyze far more information than possible by a human. Of particular interest is the potential to detect nuclear proliferation at early stages. For example, the acquisition of certain types of expertise or technology discovered via text-based data sources using AI, may indicate a would-be proliferator's intent to pursue nuclear weapons and be detected earlier than is possible via traditional methods.

While commercially and academically developed AI is achieving transformative performance for many applications, these commercial-off-the-shelf (COTS) solutions are typically inadequate for the proliferation detection mission. While a more in-depth description of proliferation detection can be found in Alexander et al. 2020, the mission centers on the ability to detect and monitor emerging and ongoing nonproliferation challenges around the world. Therefore, nonproliferation is an high-consequence domain, where the impact of a single incorrect conclusion can have a devastating impact, a key difference from many other applications that use AI.

Many of the most successful AI applications have benefited from the availability of enormous volumes of training data. As limited nuclear proliferation exists worldwide and events of interest are rare, this makes AI use much more challenging. Similarly, efforts to characterize proliferation activity within controlled situations may uncover patterns that do not translate easily across the variety of scenarios that could lead to a nuclear weapon. Finally, the nature of nuclear activities creates information in physical modalities such as radiation and acoustics, where the signatures of interest may be complex, faint, and/or hidden amidst noisy data streams that have time dependencies, which are best correlated with similarly hidden signatures found in different modalities. These features of proliferation detection create the challenging environment in which AI systems must operate and will be discussed more in Section 2.0.

The Data Science and AI portfolio in the National Nuclear Security Administration's (NNSA) Office of Defense Nuclear Nonproliferation Research and Development (DNN R&D) is driving the development of next-generation AI systems to detect and characterize foreign nuclear proliferation activities. By leveraging the expertise of the national laboratories and academic partners, new technologies are being developed to enhance the U.S. government's nonproliferation and nuclear security capabilities. To meet proliferation detection requirements, the Data Science and AI portfolio has identified the following research focus areas:

- Interpretability and explainability techniques
- Domain-aware methods
- Robust AI models
- Highly specialized AI systems
- Curating and generating relevant data sets.

As part of its efforts to develop next-generation AI for proliferation detection, the portfolio has created a series of workshops devoted to these specific research topics with the following goals:

- share successful research in these challenging focus areas
- promote best practices for AI and analytics in national security
- create connections and collaborations
- define what next-generation AI entails
- bring together researchers, end-users, and stakeholders to foster understanding of how best to operationalize AI systems for proliferation detection.

The first workshop in this series was attended virtually via WebEx by over 170 participants in September 2020 and focused on explainability techniques. It included an introduction to explainability, three keynote presentations, eight technical presentations, and two context-focused panels (Alexander et al. 2020). The second workshop, Next-Gen AI for Proliferation Detection: Domain-Aware Methods, was held virtually via Webex on February 23 and 24, 2021 and expanded on the first workshop's structure to include two concurrent tracks of 22 technical presentations, which were competitively selected following a call for papers. Four keynote presentations discussed next-generation AI for proliferation detection, perspectives on AI from the intelligence community, a survey of domain-aware approaches, and a discussion of the growth of domain-aware AI as well as current work. A panel concluded the workshop and generated insights into the potential of domain-aware AI to enhance proliferation detection capabilities. Over 250 participants attended the workshop, with attendees coming from the national laboratories, academia, and partner U.S. government agencies. As with the first workshop in the series, inclusivity and diversity were ensured in recruiting participants that resulted in improved breadth of perspectives and more comprehensive coverage of technical expertise. Student presentations were also included in the second workshop, which were absent in the first.

The workshop was structured by AI challenge rather than technique, in part, to promote best practices and share effective approaches for the specific need but also to reinforce that next-generation AI for nuclear proliferation detection requires alignment to a mission problem throughout the entire research lifecycle. Each session encompassed a variety of domain-aware approaches/techniques to address a particular challenge theme. The track themes were as follows:

- Noisy and Complex Data
- Early Proliferation Detection and Signature Discovery
- Sparse Data and Rare Events
- Robust Deployment and Decision Support.

When faced with these complexities, domain awareness becomes a critical aspect for AI systems. Many COTS AI systems are data-driven, meaning that they seek to draw conclusions based purely on leveraging large amounts of data to define the relationships used to arrive at the systems' outputs. Such approaches are severely limited in proliferation detection and shifting toward more model-driven AI can help address the difficulties associated with realizing successful AI systems. More importantly, enormous amounts of expertise have been developed in proliferation detection, and domain-aware AI offers an ideal opportunity to combine the power of AI with the substantial domain knowledge and tools developed for proliferation detection over

decades. Advancing domain-aware methods for AI is therefore critical to enabling AI to fulfill its potential and find success in proliferation detection. This report is intended to summarize the *Next-Gen AI for Proliferation Detection Workshop: Domain-Aware Methods*, describing the content of each keynote or session in turn.

2.0 Requirements and Opportunities for Next-Gen AI in Nuclear Proliferation Detection

Angie Sheffield, Senior Program Manager, Data Science, DNN R&D

The opening keynote address described challenges and requirements for modeling and analytics technologies for nuclear proliferation detection. Nuclear proliferation detection, which focuses on the use of technologies and scientific capabilities to detect and characterize observable activities and resources related to nuclear weapons development, is a notoriously difficult task. While AI presents new opportunities to transform nuclear proliferation detection and reduce the threat of nuclear weapons, commercial and strictly data-driven AI are insufficient for its highly specialized and high-consequence missions. Informed by requirements for nuclear proliferation detection and the science of AI, DNN R&D seeks to advance the state of the art in AI to develop AI systems suitable for the unique challenges and requirements of national security missions.

In particular, there are several challenges and opportunities for domain-aware AI in nuclear proliferation detection. For example, adversaries may intentionally disguise their activities such that indicators that reveal proliferation pursuits are hard to detect, and their signatures are faint against a complex and noisy background. Further, many of the signals of interest in proliferation detection include complex physical and time dependencies and are produced by systems that do not behave like typical AI features such as pixels in vast, static datasets of generic images. Domain-aware AI techniques offer the potential to combine heterogenous data sources, modeled predictions, and ML to increase sensitivity to these faint and obscured signals.

Another characteristic difficulty for nuclear proliferation detection and national security applications more broadly, is data sparsity. Targets of interest are rare, and extensive examples of proliferation attempts to provide large volumes of training data do not exist. In contrast, the available data that may include information relevant to proliferation detection is enormous. However, much of this information may not be machine readable and may span a variety of modalities across space and time with sparse signals of interest among the vast information to process.

To generate data, system developers can apply domain knowledge to create training data using simulations or fabricated experimental surrogates for activities of interest. However, such sources of training data cannot include all possible proliferation routes. Further, training data will invariably have different characteristics based on its source than the environments in which the system will be applied, and so the assumption of independent and identically distributed random variables used as the basis for strictly data-driven techniques does not hold. AI for proliferation detection must therefore detect new indicators that are not present in training data and do so in different environments than are used to create training data “in the wild.”

Data available for proliferation detection are imperfect. Nonetheless, models must be made to work with what is available. Domain-aware methods are key to developing innovative techniques and custom pipelines that combine expert knowledge, models of nuclear weapons development, and ML to augment sparse samples or provide underlying structure that cannot be obtained by data alone. Similarly, domain-aware methods can ensure robust deployments of the systems by characterizing model stability, validating and explaining signatures used, and to predict performance in new and uncharacterized settings.

Deployment of AI technologies for nuclear proliferation detection will support decision making including diplomacy, military operations, or economic measures of national and strategic importance. Such high-consequence decisions have no room for error and little tolerance for false positives. Assessments of the performance of AI systems must be informed by the domain to understand opportunity and limitations to support decision making. Still, the benefits of AI, including the discovery and application of novel signatures, can provide new capabilities in proliferation detection that can support decision making, particularly in detecting early weapons development activities. Domain awareness maximizes the performance of AI for proliferation detection and provides a basis for its application to decision making.

Domain-aware methods are key to the development of next-generation AI technologies suitable for the unique challenges of nuclear proliferation detection and national security. These techniques combine the knowledge and longstanding capability resident within the nuclear security enterprise and the transformative power of AI to build AI systems that leverage domain information, overcome issues of data sparsity, discover new features indicative of nuclear proliferation, and predict performance in new and uncharacterized environments.

Summarizing the discussion above, the challenges focused on in this workshop for which domain-aware methods offer solutions include:

- Complex and Noisy Environments: Domain-aware AI techniques to combine heterogeneous data sources, modeled predictions, and ML may increase sensitivity to faint signals of interest. Domain-informed techniques that train ML to model source and propagation characteristics may produce an entirely new class of detection methods that no longer rely on minimizing signal-to-noise.
- Sparse Data and Rare Events: Events of interest are rare, and assumptions of independent and uniformly distributed samples do not hold. Methods to combine expert knowledge, models of the nuclear weapons development process, and ML may be used to augment sparse samples or provide underlying structure that cannot be learned directly from the data.
- Robust Deployment and Decision Support: Domain-aware methods may be used to validate that an AI model accurately predicts system behavior, not just fits the data. Model performance must be domain-informed to assess the limits of an AI system to support decision making and build AI systems that perform predictably in new and uncharacterized settings.
- Early Proliferation Detection and Signature Discovery: Advances in ML and the availability of new data sources present new opportunities to detect early indicators of nuclear proliferation from large and unstructured data. Domain-aware techniques may help to direct exploitation of these new data sources or interpret the signatures identified by ML models.

The first *Next-Gen AI for Proliferation Detection* workshop was held in September 2020 and focused on the use of explainability methods. Explainability techniques are another critical enabling feature to develop AI systems for proliferation detection, including domain-aware models; more information can be found in Alexander et al. 2020. A particular takeaway from the workshop was that approaches used should be driven by the end use and researchers must work closely with mission partners to understand requirements for AI systems that can be used. This is relevant in developing domain-aware systems and for this domain-aware-focused workshop.

The development of domain-aware AI systems is possible. Indeed, ongoing research in the national laboratories and academia applies domain knowledge to improve AI systems, and some successes are already being achieved. One example is in fusing heterogenous data sets together. Research funded by DNN R&D has found that it is most promising to combine different data sources within a unifying model structure, such as a Bayesian net or an activity-based model, instead of simply combining the data sources without consideration for their context and hoping to discover a novel purely data-driven signature among the numerous unprocessed input features. This is one generalizable domain-aware approach that can be used across research pursuits and mission applications. This workshop aims to identify further repeatable approaches that are well-matched to nuclear proliferation detection.

3.0 Keynote Presentations on Actors in Proliferation Detection

3.1 U.S. Government Missions and Agencies that Use Nuclear Proliferation Detection Capabilities

Angie Sheffield, Senior Program Manager, Data Science, DNN R&D

Nuclear proliferation detection capabilities support and enable an incredibly diverse set of missions executed by a broad set of partners across the U.S. government. Leveraging the capabilities of the DOE national laboratories and working with partners in the nuclear security enterprise, DNN R&D leads U.S. government efforts in AI research for national security (Subcommittee on Nuclear Defense Research and Development 2019). This keynote described the interconnected roles of various agencies to carry out the essential mission of proliferation detection.

NNSA's Office of Defense Nuclear Nonproliferation (DNN) applies proliferation detection capabilities and technologies to reduce nuclear threats, ensure peaceful nuclear uses, and to enable verifiable nuclear reductions. More concretely, DNN actively applies proliferation detection technologies to support the International Atomic Energy Agency (IAEA) to apply safeguards that detect and deter illicit diversion of nuclear material, to build domestic and global capacity to combat illicit trafficking of nuclear material and technology, to monitor and verify compliance with and within export control programs, and to develop technologies that can facilitate verifiable arms control treaties.

The intelligence community (IC) also employs proliferation detection capabilities to detect, characterize, and disrupt activities of state and non-state actors engaged in the proliferation of weapons of mass destruction (National Intelligence Strategy, 2019). Intelligence tradecraft uses a characteristic six-step cycle of defining requirements, planning and direction, collection, processing and exploitation, analysis and production, and dissemination. Dr. Emma Hague's presentation during the workshop described key requirements for AI-enabled technologies used in all-source analysis, and she provided the lingering message that in this stage, "high-impact models are human-centered… and all the rest die." However, next-gen AI can contribute to other stages in the cycle including processing, planning, and requirement definition. In the latter two stages, use of next-generation AI can reduce downstream demands on exploitation and analysis. A key motivator for domain-aware AI is the nature of the data available. The IC is faced with an abundance of data produced during collection, but these data are sparse with respect to examples of activities and targets of interest, and the National Intelligence Strategy has identified this particular challenge as an opportunity for AI (NIS, 2019).

The DoD also has responsibility to disrupt weapons of mass destruction (WMD) threats and to inform operations to address aggression from adversary states (Summary of the 2018 National Defense Strategy of the United States of America, 2018). These tasks leverage proliferation detection capabilities to achieve success. Most prominently for the DoD, it relies on NNSA and the national laboratories for deterrence through the U.S. nuclear weapons stockpile. U.S. military strategy faces a new paradigm and recognizes the role for AI and analytics to support adaptation to changing environments. Next-generation AI can improve situational awareness and decision-making for counter-WMD missions and improve predictive modeling and analytics

of countering WMDs (CWMD) threats while reducing the time that operators and analysts spend evaluating data.

DNN R&D works with DHS, who leverages proliferation detection capabilities in their mission to prevent nuclear and radiological threats against the U.S. This includes the mission of nuclear forensics, which provides attribution in the case of a wide variety of misuses of nuclear material, incentivizing responsible behavior and deterring nuclear-based destruction. Additionally, DHS supports the emergency response mission to prevent and respond to nuclear accidents and disasters.

DNN R&D also collaborates with the White House Office of Science and Technology Policy and the National Security Council to support and accelerate the use of AI technologies across the national security enterprise. Finally, within the DOE, DNN R&D works with the Office of Science and Defense Programs to meet requirements and opportunities for high-performance computing and hybrid infrastructure for data intensive computing across the national laboratory complex and develop next-generation AI foundational capabilities.

While the variety of missions and tasks may seem intimidating, DOE's 17 national laboratories make them possible. The national laboratories have the responsibility to address large scale, complex research and development challenges to support national security, and their multidisciplinary and innovative capabilities make them well-matched for the demanding requirements of the various proliferation detection missions. DNN R&D draws on the national laboratories to address the diverse proliferation detection challenges and has the unique perspective within the U.S. government science and technology community to identify challenges, gaps, and opportunities where advances in the math and science of AI can transform our capability to detect proliferation and to set the research direction for AI and analytics in the U.S. government. By working together, the complex will be able to transform nuclear proliferation detection through advancements the field of AI.

3.2 A Perspective from the Analytic Intelligence Community

Emma Hague, Chief Data Scientist, Foreign Nuclear Programs Division, DOE Office of Intelligence

This keynote presentation shared perspectives from the multiple vantages that Dr. Hague has been privy to throughout her career, particularly from her current position in the intelligence community, and aimed to provide insights on how to ensure success in developing new AI technologies. Characteristics of intelligence work and the environment were described to help researchers better understand the context in which their systems may be applied. The importance of connecting human and machine domains, as well as researchers and analysts, was emphasized.

Before joining the intelligence community, Dr. Hague worked in nuclear emergency response where she focused primarily on medium- and high-threat radiological searches, including both tactical and maritime missions. The radiological search mission is well-suited to ML as in-situ radiation detection in the highly variable backgrounds that are commonly encountered is a challenging task. To address difficult applications such as with radiological search, an integrated team of diverse specialists is required, which is a recurring theme in the presentation. This work led her toward the application of adaptive techniques and then to her current role in the DOE's Office of Intelligence and Counterintelligence (DOE-IN).

DOE-IN has two primary missions:

1. Provide policy makers with the best possible information

2. Foster collaboration between laboratories and the intelligence community.

Consequently, in partnership with national laboratory partners (the Field Intelligence Elements), analysts produce all-source intelligence. Such assessments are created by deeply skeptical and devoted truth-seekers using a rigorous tradecraft. In turn, absolute conclusions are avoided and communication takes place in the language of likelihoods. Ultimately, while the description of prior events can be nice, predictive capability is the required objective, and intelligence uses data, training, and knowledge to predict the likelihood of future events. This naturally calls to mind ML and provides fertile ground for the productive application of AI technologies. Nonetheless, intelligence analysis is the product of a sequence of events including collection, processing and exploitation, and analysis and production that are inherently carried about by humans. Indeed, intelligence is by definition what people think and so the creation of intelligence must always be human-centric.

The central theme of the keynote then is that high-impact models are human-centered, and all other models die. To turn this insight into a guiding principle, analytic-centric development is proposed, and the advantages of creating a shared mental model between all necessary parties are discussed. A shared mental model facilitates understanding of how a model might be used, and familiarity with the competing tradeoffs of different methods is important in design. For example, there are non-trivial compromises between precision and recall, and the relative weighting of false negatives and false positives in evaluating systems requires consideration that should be carried out using the shared mental model. Analysts will typically prioritize high recall in a model, but this may not always be the case and mutual discussion should take place to find the right balance of model characteristics. The systems will evolve over time as priorities or the data available change, and this further motivates the creation of a shared mental model so that models can be dynamically adapted together for the greatest impact. Models must work with people to be useful and to generate the desired insight. To support this researcher-analyst and human-machine connection, the user interface and experience must be designed into the research at an early stage and be an inherent consideration.

Fortunately, the IC has prioritized the incorporation of AI into its capabilities and to have a baseline understanding of AI by 2025 (NSCAI 2021). Doing so is urgent as decision makers are already “data smart,” with an understanding of confidence intervals and the difference between statistics and real-world factors. Further, the U.S. is now operating in a near-peer global environment with AI. Guidance on the leveraging of AI can be found in the National Security Commission on Artificial Intelligence’s Final Report, particularly Chapter 5, which urges analysts to change risk management and jump start technology adaptations (NSCAI 2021). A noted challenge in the report was that of recruiting and retaining technology experts, with the main deterrent being challenges and waiting times associated with clearance awards.

Overall, the time is ripe to grow the role of AI in national security and intelligence, but it is critical to remember that intelligence will always be a human endeavor. As such continual consideration of human-machine teaming and how models will be used is imperative, and domain-aware methods can help bridge the human-machine gap. Worth considering is that, in the IC, humans are the domain to be aware of. Data scientists and analysts should strive to create a shared mental model and reach out to each other early and often to understand capabilities and needs. Finally, systems do not need to achieve perfection to be helpful, and unrealistic objectives

should not be allowed to impede research. By working with diverse teams of experts, researchers can make a difference now in improving intelligence and national security capabilities.

4.0 Methods and Applications for Implementing Domain-Aware Techniques

4.1 Survey of Domain-Aware Methods

Ashley Shields, Cloud Software Engineer and Digital Twin Data Scientist, Idaho National Laboratory

Put simply, domain-aware methods find ways to incorporate world knowledge into models to enhance performance; however, no concrete definition or structured methodology exists to do so. An objective of the *Next-Gen AI for Proliferation Detection Workshop: Domain-Aware Methods* was to provide more structure to this topic specifically in the proliferation detection mission space. One source of guidance comes from a recent report on basic research needs for scientific ML (Baker et al. 2019). Broadly considered, domain-aware methods are methods that enhance the accuracy or interpretability of models by applying physical principles, constraints, conservation laws, or other knowledge representations.

Five different categories of domain-aware methods were identified as examples:

1. Expert knowledge
2. Synthetic data generation
3. Inclusion of non-traditional AI/ML methods into traditional AI/ML models
4. Semantic or constraint-based methods
5. Soft labels.

Several interesting and recent examples of each of these categories exist in the literature, and it was found that these categories are frequently used in parallel, as shown in Figure 4.1. Examples of each category are described in more detail here, particularly with an emphasis on applications related to proliferation detection.

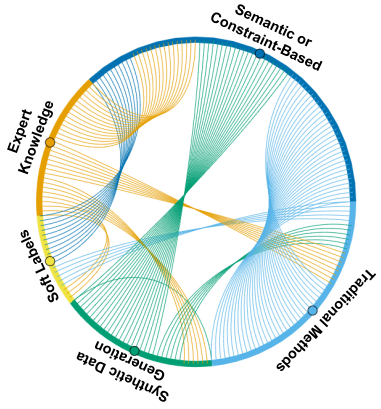


Figure 4.1. Mapping of the connections in the literature between different categories of domain-aware methods. The number of chords between different categories is proportional to the number of publications returned on Google Scholar with keywords associated with the domain-aware approaches connected by those chords. Taken with permission from the presentation being summarized (4.1. “Survey of Domain-Aware Methods”).

One example of incorporating expert knowledge was found in nuclear power plant monitoring, in which an interactive, human-in-the-loop method was developed to reduce costs while addressing trust, security, and explainability (Rashdan et al. 2019, Versino and Lombardi 2011, Gastelum et al. 2019). A camera was set up in a nuclear power plant, and a decision tree model was developed in which training examples were provided and the model identified relevant features in the imagery. The visual feedback was presented to an expert reviewer who confirmed or rejected the results, sending the information back into the algorithm as additional training data. This led to accuracy improvements by generating additional training data and led to improved explainability and accessibility of the AI system.

Synthetic data generation is often required for situations where data scarcity is a problem. Synthetic data generation improves the data quality by increasing the size and diversity of the dataset. Modern methods can generate data that is indistinguishable from real-world data. Typically, this is done using a simulator to generate a synthetic image, and real-world data is then used to refine the quality of the data synthesizer's fabricated images (Shrivastava et al. 2017, Yu et al. 2019). Approaches typically use generative adversarial network (GAN) structures, which consist of a generator to create synthesized data, and a discriminator to distinguish synthesized data from real-world data. If the discriminator concludes that the synthetic data is real data, then that data can be incorporated into the data set. Otherwise, the image is returned to the refiner and placed back into the cycle. When the discriminator can no longer tell the difference between the synthetic and real data, the synthetic data is of high enough quality to train other ML models. The discriminator and generator are typically trained together resulting in a push and pull, or adversarial, approach to modify input data generation in such a way to become highly realistic.

An example of augmenting traditional AI/ML models with domain knowledge is demonstrated in a recent study using geospatial analysis to determine which buildings belong to the same facility. This approach used a convolutional neural network to flag the buildings in a petroleum facility (Brost et al. 2014). The model was constrained using geospatial information on building size and shape parameters and triangulation to map out the facility footprint. The resulting model could perform well on multiple size scales and correctly detect both large and small facilities. In addition to the multi-scale analysis capability, advantages include semantic consolidation in which discontinuous features can be interpreted as being part of a continuous object based on conditional requirements.

Constraint-based approaches can be used to help control the movements of unmanned aerial vehicle (UAV) swarms (Wang et al. 2018). In UAV swarms, it is critical to coordinate the movements between individual units so that there are no accidents. Typically, a set of rules are implemented to automate search and tracking behavior. A minimum distance constraint ensures that individual UAVs do not collide, while a fuel constraint ensures that UAVs are able to return to a fueling station before they run out of power. Further constraints ensure that units work together for a common goal. For example, if a UAV detects a nearby UAV that is in tracking mode, it will join in and coordinate on a search. This approach resulted in the automation of dynamic search and tracking based on situational information.

Finally, soft labels assign not only a label to a datum, but a probability or confidence interval to the assignment. A recent example of the use of soft labels is a recurrent neural network system that takes in a sequence of video frames as its input and assigns probabilities of a given action taking place as an output (Hu et al. 2019). The soft labels allow for detection of actions mid-execution, which facilitates an early response. Predictions can take place in nearly real-time.

In summary, domain-aware methods are diverse, and there is no concrete definition or methodology for domain-aware approaches. However, domain-aware methods have proven to be useful in a variety of contexts. These methods should be employed to develop new nuclear proliferation indicators and to enable earlier detection of emerging threats.

4.2 Domain-Aware AI: There and Back Again

Kary Myers, Statistical Sciences Group, Los Alamos National Laboratory

This keynote presented by Dr. Kary Myers, provided an overview of current opportunities and challenges for domain-aware AI in nonproliferation, highlighting domain-aware AI use and limitations when used by the Multi-Informatics for Nuclear Operations Scenarios (MINOS) project led by Oak Ridge National Laboratory (ORNL).

It was proposed that methodologies often encounter resurgences in popularity that far out-scale their initial popularity, frequently due to improvements in underlying technologies supporting the given method. For example, neural networks resurged in popularity recently due to the increase in data availability and computation technologies needed to train and deploy increasingly effective models that would not have been possible when neural networks were first introduced. However, there is a danger for domain-aware techniques wherein if ML models make a sufficiently bad mistake or exhibit biased or low performance, domain experts can lose trust in ML completely, which would be difficult to regain even if the method gains popularity or improves efficacy later.

Unfortunately, when relying on data-driven AI methods that use large-scale collections of available data, unintended biases are often found for several reasons. One such bias is due to convenience sampling, which describes the case when one trains on data from or about a subset of the population and applies the model to the entire population. In this case, there are significant biases and errors that are likely to occur. A possible solution is to incorporate domain knowledge to identify biases that are present and mitigate those biases. Dr. Myers posits that “if we rely on data-driven AI methods without domain knowledge, we risk creating tools that will fail spectacularly in the wild.”

To highlight the effectiveness of incorporating domain knowledge to bolster AI in the context of nuclear nonproliferation, the MINOS project was used as a successful example that dealt with key challenges and opportunities. MINOS leverages domain-aware AI, focusing on learning as much as possible with both calibrated measurements and rich ground-truth knowledge to incorporate domain knowledge to better understand and improve methodology, and evaluate both what is and is not suitable to be transferred to other facilities. Key research questions for MINOS focused on (1) identifying signals of interest, (2) combining data sets, and (3) generalizing approaches to proliferations scenarios. However, as with many applications, a significant challenge was that there are often few data points for events of interest and large amounts of data features from measurements to consider – described as “small n, large p.”

Domain-aware methodology can be used to alleviate this challenge of “small n, large p” through a variety of ways: (1) using domain-aware AI to reduce the scale of measurements under consideration (e.g., feature selection) and using domain knowledge to constrain AI predictions to those which are physically plausible, (2) increase the number of events of interest observed using simulations or by defining defensible subproblems that have more events of interest. An important caveat to simulations, however, is that simulations cannot be treated as equivalent to

raw data but can be leveraged alongside domain knowledge to increase the sample size of events.

One challenge that was specifically noted was the application of domain-aware techniques when ground truth is not available, a challenging but realistic scenario. One possible solution is to examine a particular scenario that does have ground truth, as was done with MINOS. Another recommendation is to build a community of people who are conversant in the necessary areas – understanding the data science (the ML or AI methodology under consideration) and the application space (e.g., reactor operations) in order to leverage domain knowledge and experts in way to best mitigate the negative impacts of not having access to ground truth.

5.0 Complex and Noisy Environments

The first technical session focused on how domain-aware ML methods can extract valid and useful information from complex and noisy datasets, whereas often in this context a data-driven approach would fall short.

Three themes emerged among the six presentations during this session. First, incorporating domain knowledge into the model can greatly improve results. This was well-illustrated in the second talk on harmonic structure, in which incorporating a specific harmonic layer greatly improved network classification accuracy, depicted in Figure 5.1. The third talk also showed that a new network architecture can be developed to specifically cancel the effect of the background. The fourth talk showed how realistic source trajectories could only be calculated from radiation detector data when using a model that included information on the roadways and speed limits.

A second theme was that a combined approach worked better than a single analysis method. The third presentation showed how an ensemble of eight distinct classifiers performed better in terms of error minimization than using any single analysis method. Another presentation demonstrated how a complex network architecture with six distinct subcomponents performed more reliably than a single neural network.

Finally, often the data are “sparse” in the sense that events of interest are rare, even if the volume of data is large. The final presentation touched upon how the nuclear material transfers of interest happen infrequently, and the fourth talk showed a situation where even with an array of detectors, often zero or one of those detectors are collecting significant signal at any given moment. The sparsity of the data drives the need to generate synthetic datasets. Data synthesis is also useful for situations where it is difficult to establish clean ground-truth data—the synthetic pileup pulses from the fifth talk illustrate this. How representative these synthetic datasets are of real-world data and how well methods trained on synthetic datasets will transfer to a real-world scenario remain open questions.

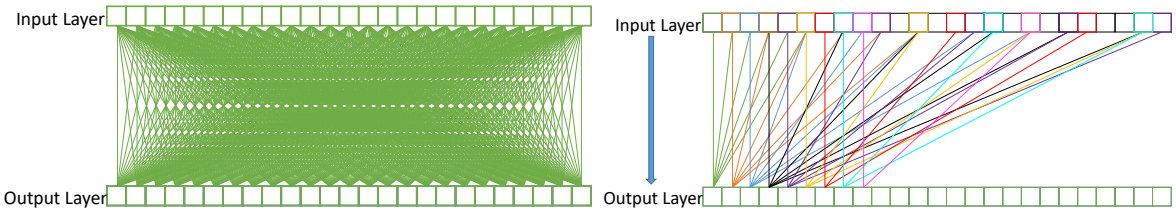


Figure 5.1. Example of incorporating domain knowledge directly into network structure whereby fully connected layers (left) are replaced by those in which connections are made based on connecting base frequencies to associated harmonics (right). Taken with permission from 5.1, “Imposing Harmonic Structure in Neural Networks.”

5.1 Imposing Harmonic Structure in Neural Networks

Mark Adams, Oak Ridge National Laboratory

Harmonics are alternating current (AC) voltages or currents present in an electrical system at a frequency other than the fundamental. Harmonics have a variety of causes and usually indicate a power quality problem. It is important to be able to detect these harmonics because they can

cause heating and damage to electrical components. This presentation covered the development of neural networks that can identify harmonics in non-intrusive load monitoring (NILM) data of a building's electrical system.

Standard fully connected neural networks connect every input data point to every neuron in the first hidden layer, and the output of every hidden layer neuron to the input of every neuron in the next hidden layer. This structure provides a lot of flexibility that can be helpful in discovering potentially important features that the researcher might not have thought of. However, it also requires many parameters, since each connection has an associated weight that needs to be trained. This greatly increases the training time.

Instead, Adams et al. take an approach inspired by acoustic signal processing and introduce harmonic connections between layers (Zhang et al. 2020). In this approach, the input is the Fourier transform of the collected raw time series data. Not every point in the input data is connected to the input of every neuron in the first hidden layer; rather, each neuron in the first hidden layer is connected to a base frequency and multiple harmonics. In this way, the harmonic structure of the data is preserved, which makes it easier to detect the harmonic signals of interest against a noisy background. Furthermore, this significantly reduces the number of trainable parameters—in this case, a reduction of 1 billion parameters to 3 million.

To test this approach, different waveforms with well-defined harmonic structure (e.g., square wave, triangle wave, etc.) were injected into the building power system at a facility at ORNL. NILM was used to measure the AC signal at various points in the building. Three different neural network architectures (dense, ResNet, and U-Net), either with or without a harmonic layer, were trained to identify the type of injected signal.

Without the harmonic layer, both the dense network and the ResNet network show poor classification accuracy—17% and 20% accuracy, respectively. The deeper U-Net network shows better performance at 60% classification accuracy. However, all three network architectures showed dramatic improvements when a harmonic layer was used. The dense network classification accuracy improved to 70%, the ResNet network improved to 81% accuracy, and the U-Net network achieved 82% classification accuracy with a harmonic layer. Furthermore, when these networks did misclassify waveforms, the assigned labels at least belonged to the right type (e.g., a square waveform at a given base frequency would tend to be assigned to a square waveform with a different base frequency). This demonstrates how including a harmonic layer can improve classification accuracy against a noisy environment.

5.2 Domain Adversarial Networks and Explainability Assessment

Thomas Grimes, Pacific Northwest National Laboratory

Neural network classifiers can solve problems in ways that the data scientist might not expect. Networks take shortcuts in making their assignments. If there is any component of the background that correlates with the data, the network will try to take advantage of it, even if that background is irrelevant or not easily interpretable from a human perspective. Examples include making assignments based on correlated noise-like features, learning the background of an image rather than the image subject, or making assignments based on fiducial markers or metadata.

A taxonomy of these shortcut learning methods has been identified in the literature. A "ρ-useful" feature is noisy and present in every dataset, but fragile with respect to making assignments

(Ilyas et al. 2019). On the other hand, “γ-useful” features are robust and human-interpretable, but not what a human would deem relevant when it comes to making assignments (Xiao et al. 2020). For example, the background of an image can serve as a γ-useful feature if the image subject is often photographed against a similar background.

During training, the effect of ρ-useful features can be minimized by adding random noise to each image at the start of each epoch (Madry et al. 2017). This effectively converts each data point into a “ball” of data, which makes for much more robust decision boundaries.

Eliminating γ-useful features is more involved but can be done with a domain adversarial neural network (DANN) (Grimes et al. 2020). Regular neural networks can be considered as having two sections: a feature extractor and a class predictor. A DANN adds another fork to the end of the feature extractor to try to predict a background label. During training, the gradient of this branch is reversed to effectively cancel the effect of the background on the extracted features. This produces features that are background-orthogonal that can be used to predict the image subject alone. The relative magnitudes of the updates to both branches of this fork are controlled by a new hyperparameter, λ, which must be tuned to optimize performance.

In addition to producing background-orthogonal features, if λ is set to 0 the DANN branch can be used to assess the impact that background information has on network class assignments. In one trial, with the gradient reversal turned off, the background labels could predict the proper class with ~80% accuracy. With gradient reversal turned on, the background labels alone performed no better than random guessing.

This approach was tested in a scenario where building electrical system data was used to train a neural network to determine the on/off state of a turbo pump powered by that system. These measurements are complicated by the fact that different buildings host different electronics, which make different demands on the electrical system. However, these demands on the electrical system form a background that can be taken advantage of using a domain adversarial approach. Twenty networks, either with or without a DANN branch canceling the effects of background introduced by different buildings, were trained on this data. The networks that included the DANN branch were found to have a 25% lower misclassification rate.

5.3 Dissolution Event Classification Using Isotope Decay Chains and Half-Life Estimates

Nageswara Rao, Oak Ridge National Laboratory

The goal of this work was to develop a machine learning model that would detect whether a particular dissolution event took place at a nuclear processing facility (Nageswara et al. 2020). Specifically, this work attempted to identify when the production of plutonium-238 (Pu-238) from a neptunium-237 (Np-237) target is taking place. This is a complex problem that requires knowledge of the physics and chemistry of dissolution.

Production of specific isotopes has two parts: First, a target is inserted into a fuel rod and neutron-irradiated at a nuclear reactor. Next, targets are moved to another facility for extraction and shipment. In principle, the off-gas from the extraction process can be analyzed to decide whether a dissolution event of interest took place.

Production was monitored by collecting the gamma spectrum near an effluent conduit. Ground truth was established from the facility production logs. Collections took place at 1-hour intervals

over the duration of the study. Spectra showed distinct peaks but were quite noisy. Peak fitting was used to get counts for 15 different isotopes, including isotopes of iodine, xenon, krypton, cesium, and barium. Knowledge of the proper isotopes to monitor was informed by knowledge of the fission decay chain. Furthermore, the proper sampling time intervals were informed by the knowledge of the isotope half-lives.

The classification strategy used eight different classifiers of various types (e.g., support vector machines, decision trees, naïve Bayes, etc.) to predict when a dissolution event took place. The errors of each classifier were determined, then the best three were fused together to produce a final prediction. No universal best classifier exists, even in theory. Therefore, the fused classifier achieved a lower error rate than any single classifier.

Different time windows between 1 hour and 1 month were tested and receiver operating characteristic (ROC) curves (a plot of the false positive rate vs. true positive rate) were generated. Increasing the time window produced a better ROC curve up until 24 to 48 hours. Beyond around 24 hours, signatures became more unstable because many different processes aside from the process of interest occur at these facilities and the signatures of these irrelevant background events mix with the signatures of the event of interest.

Performance was tested by including only certain subsets of the monitored isotopes (e.g., using only the iodine isotopes, or only the krypton isotopes). Some isotope groups performed better than others; however, using all the isotopes together gave the lowest classification error rate.

5.4 Inferring the Dynamic Location of an Environmentally-Constrained Radiative Source with a Network of Detectors

Dave Osthus, Los Alamos National Laboratory

This presentation described an effort that took place under the MINOS project, where instrumentation and detectors were set up throughout a nuclear processing facility at ORNL to determine how to combine multiple data streams. The facility itself was not part of the project, in the sense that it carried on operations as normal throughout the duration of the study. This allowed for researchers to study the feasibility of various methods for detecting events of interest against a noisy background of irrelevant events.

This work attempted to identify and localize a source moving in a constrained environment. Specifically, this work focused on using six different gamma radiation detectors located throughout the ORNL nuclear processing facility. This data is sparse and noisy, and it has many unknowns. However, certain constraints can be applied. For example, radiation sources are often transported by vehicle, so the constraint that the moving source must be located on a road can be applied.

While, the radiation source was moving, the gamma detectors were fixed. Each detector generated a time series of counts that was integrated over all energy channels, and the raw data can be broken into signal and noise components. Because the signal follows an inverse square distance model, in principle, three detectors can triangulate the location of the source. However, due to the noise levels the detectors have an effective limited range, so in most cases fewer than three detectors are picking up a reliable signal at any given time. In the real-world sample transfer data, there was no point where more than one gamma detector (out of six) was able to collect a significant signal at any one time, and there are many time points when none of the detectors were able to pick up a usable signal at all.

As an illustration, a moving radiation source taking a left turn on a road was simulated by computer so that the ground truth would be well known. During this simulation, there are time periods where none of the simulated detectors are receiving a signal above noise level.

The talk demonstrated that it is relatively easy to generate a source trajectory that "agrees" with the detector data, in the sense that this trajectory generates a best fit. However, it is more difficult to generate a trajectory that both fits the data and is physically plausible. Adding constraints to the model increases plausibility of the generated trajectories. When a simple Bayesian model was used to fit the detector data, the generated source trajectory skipped around the simulated field erratically, even though the model fit the detector data well. Localization of the source was good at the beginning and ends of the simulated trajectory, where the signal was strong. However, at intermediate time points the model showed high uncertainty. Therefore, it was necessary to add a constraint that correlates source positions in time.

When the simple model was replaced with a Markov model, where each time point in the trajectory needed to be located close to previous time point (essentially encoding a speed limit), the result showed both good agreement with the gamma detector data as well as an improved trajectory prediction. However, this trajectory still meandered and was not realistic for how vehicles actually move. Adding another constraint to the model that encodes the road locations produced a much better path prediction.

All three models (simple, simple + speed limit, simple + speed limit + road) fit the gamma detector data well, but only the highly constrained model both fits the ground truth and was easy to interpret. The simulations were followed up with a real-world measurement of a radiation source traveling around a building on the ORNL campus. The constrained model was able to produce an accurate reconstruction of the real-world source path.

5.5 An Artificial Neural Network System for Special Nuclear Material Detection in Photon Based Active Interrogation Scenarios

Abbas Jinia, University of Michigan

In nuclear nonproliferation, analysts need to be able to detect radioactive sources. Passive interrogation techniques are frequently not robust enough to detect small quantities. Active interrogation is better for detecting small-quantity or shielded materials. The work presented in this talk used a linear accelerator (LINAC) along with a spontaneous fission source to test a stilbene-based neutron detector.

Stilbene is known and was shown to possess excellent pulse shape discrimination capability to discriminate between photon and neutron detection events. However, during active interrogation, there is an intense photon environment that makes it more difficult to measure the neutron rate due to photon pileup and noise (Fu et al. 2018). The presenter's work constructed an artificial neural network (ANN) to recover these neutron detection events from the photon noise.

An ANN system was developed that used six neural networks that work in conjunction to improve accuracy. Each network had a specific purpose—one performs sorting of pulses into single pulses or pileup pulses, one classifies single pulses as photon or neutron, one classifies pileup pulses according to how close the pulses are, etc. All classifiers had a classification accuracy that was greater than or equal to 99%.

In addition, the ANN included two cleansers to flag misclassified data. The cleansers were based on an autoencoder and decoder pair. The cleansers effectively denoised the data and then detected whether a pulse was a true single pulse or a misclassified pileup pulse (or vice versa).

A californium-252 (Cf-252) source was used to get clean single pulse training data. Pileup pulses were synthesized from the single pulse data to establish reliable ground-truth training data and represents one example of synthetic data generation.

Performance was evaluated in the presence of intense photon flux by taking measurements on a Cf-252 source while the LINAC was turned on or off. The Cf-252 source has a known, established particle emission rate. When the LINAC is turned on, the photon pulses from the LINAC account for most of the detection events in the detector. The researchers found that their ANN was able to recover some information that would have normally been lost, since typically all pileup pulses would be eliminated from the dataset. The ANN could recover 19% of the neutron pulses that would normally be lost within the photon pileup noise. Follow-up studies on a depleted uranium target also found that 19% of the pileup pulses could also be recovered and correctly classified.

5.6 Spectral Signatures for Shielded Sources

Jaston Hite, Oak Ridge National Laboratory

This presentation covered work that was carried out under the MINOS project, a multi-lab venture to conduct multi-modal, multisensor measurements of a nuclear facility at Oak Ridge National Laboratory. Specifically, this work tracked the movement of special nuclear material (SNM) at this site. Irradiation targets are made at the Radiochemical Engineering Development Center (REDC), loaded into a container (called the “Q-Ball”), transported to the High Flux Isotope Reactor (HFIR) for irradiation, loaded back into the Q-Ball, and transported back to the REDC. The goal was to detect when these transfers happen and by what route.

This work was complicated by the fact that there is a lot of activity going on at this facility, with many different vehicles. Other vehicles (e.g., the laundry truck) showed detectable radiation levels, but were not of interest. A metric more targeted to the specific Q-Ball movements needed to be developed. The work was also complicated by the fact that the frequency of Q-Ball movements was rare enough where the data could be considered “sparse”, even though a large stream of data is generated by all of the detectors operating continuously.

The researchers were primarily interested in Np/Pu and Cm/Cf transfers. The Q-Ball is composed of water, concrete, and steel layers and is effective at shielding, so that the gamma spectrum typically does not show distinct peaks. However, there is a shift in the gamma spectrum toward high energy due to neutrons interacting with the shielding and producing secondary gamma rays. The researchers proposed using this shift in spectral density as a signal of interest.

A signature was developed to detect this shift in the gamma spectrum. A baseline estimation and denoising (BEADS) algorithm was used to smooth out noise in spectrum (Ning et al. 2014). The smoothed spectra were integrated to get a cumulative density function (CDF), which accentuates the shift in the spectrum toward higher photon energy. The head-to-tail ratio of the CDF served as the signature of interest and appears as a sharp spike in the presence of the Q-Ball. Other factors such as weather conditions (e.g., rain) have a detectable effect on this

metric; however, these effects were easily distinguished from the event of interest. The time series of spikes of this metric could be used to determine vehicle routes, or at least to identify routes that were not taken.

The presenter finished by describing an effort that is underway to generate synthetic data using a long-short term memory neural network architecture that the researchers expect will generate realistic-enough data to train a method to identify material transport paths in the real world.

6.0 Early Proliferation Detection and Signature Discovery

This session focused on detecting proliferation attempts as early as possible and the signatures that can help facilitate that detection. It remains critical to move detection capabilities earlier in the timeline. Specifically, in this session, it was discussed how advanced scientific methods in concert with the specific domain expertise can be used to maximize the limited data that are available.

A key focus of these approaches was to exploit non-traditional data sources. One approach searched public records for activities that might impact local sensor networks. Another studied not just the content of individually published papers but included the metadata and relationships of authors of published literature to glean insight, as illustrated in Figure 6.1. Another approach attempted to detect user intent from a sequence of internet search queries.

A talk on Plutonium Attribution Methodology focused on improving an approach which had previously been successful but was vulnerable to spoofed samples. Domain knowledge was required to identify the existence of this problem and was further required to identify the solution. This domain-aware-motivated improvement to the methodology serves as a reminder to question traditional models in the context of the applied domain and to continually improve them.

In “Applying Domain-Aware Artificial Intelligence on the CBRN [Chemical, Biological, Radiological, and Nuclear] Battlefield,” a reminder was provided that applying these techniques across the entire proliferation monitoring process is also useful. Opportunities for which AI solutions would be welcome were presented and included network strength monitoring, image characterization, and route planning. Finally, it was noted by the session chair that incorporating search behavior and author network groups begins a welcome integration of the social sciences.

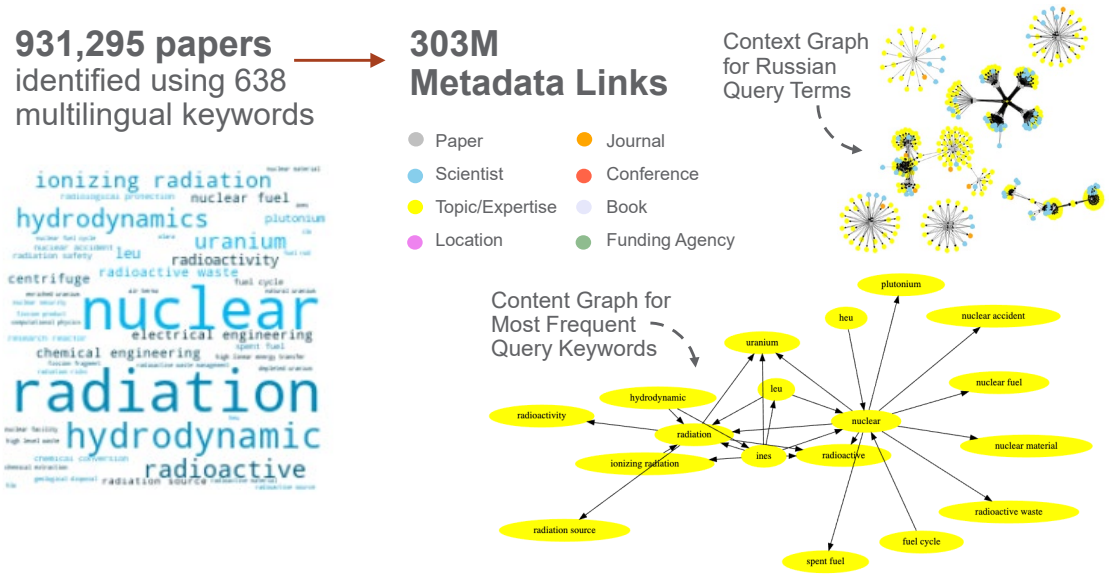


Figure 6.1. Illustration of the use of multilingual keywords along with metadata and other relationships to characterize nuclear expertise as a potential proliferation indicator. Taken with permission from 6.3, “Extracting Dynamic Proliferation Expertise and

Capability Representations from Heterogenous Multilingual Open-Source Data Streams."

6.1 Open-Source Data Analytics Value Quantification to Inform and Explain Radiological Source Detection Localization and Tracking

Sannisth Soni, Pacific Northwest National Laboratory; Svitlana Volkova, Pacific Northwest National Laboratory; and Ellyn Ayton, Pacific Northwest National Laboratory

This work focused on open-source data analytics to augment radiation detection sensors. The combination of both descriptive and predictive types of analytics gives the approach its leverage and is exercised on both dynamic sensors and static sensors. Logistic regression and random forest models that used characteristics of construction permits as dynamic sensors found that those which contained excavation and indoor complaints were the most predictive. Though the results that were presented show a high F1 score for some alerts, it is unclear what those alerts actually predict, and chance correlation cannot be disregarded at this point as the causal link between excavation and alerts remains unclear.

The described natural language processing work showed the ability to predict detections of cesium-137 (Cs-137) based on the type of construction work being performed and based on the work permits submitted and their associated location. In this case, the application of natural language processing aims for predictive analytics. Specifically, given a particular alert, identifying the source which could have generated this alert is of interest.

Pattern of life analysis was also applied to three medical isotopes: positron emitters, iodine-131 (I-131) and metastable technetium-99 (Tc-99m), for which signals from each are distinct. 27 sensor specific models were compared with nine ensemble models. The ensemble approach gave the best performance for source prediction but did not work well for sensor prediction. To establish normal background noise or baseline patterns, a pattern of life analysis was performed first to discover specific sensor and isotope signatures across locations. For example, it was found, as expected, that there are less alerts on the weekends when construction activities are limited.

6.2 Applying Domain-Aware Artificial Intelligence on the CBRN Battlefield

Adam Seybert, U.S. Army Nuclear and CWMD Agency

In many ways, the pursuit of early proliferation detection mirrors the hunt for rapid post-detonation characterization methods. While early proliferation detection results in a reduced threat of strategic surprise, early fallout characterization reduces risk to first responders during consequence management operations and allows greater operational freedom of movement for forces on a nuclear battlefield. However, extracting the useful information continues to challenge even the most knowledgeable person when presented with the sheer volume of available data, sources, and formats.

This presentation highlighted how solutions in the nonproliferation domain can be applied to CBRN response. Both domains encounter the same data sparsity and similar source and sensor variability. In turn, domain awareness is critical, although these domains face different domain challenges such as time scales, data volume, data types, and quality.

During the WMD defeat stage of CBRN response, there is more time for data quality control (Joint Publications 2019). In CBRN response, information about the data quality or where it came from does not always exist. Another key difference is the type of decisions being made. Data volume dramatically increases in a response situation. Though nonproliferation work traditionally focuses on the WMD defeat stage, there are many related AI opportunities in the CBRN response stage. Key opportunities focus on awareness and execution.

The same domain-aware and traceable AI technologies that enable early detection of nuclear weapons development should be applied to post-detonation nuclear modeling to bridge the gap in data discrimination. Three awareness-focused opportunities were presented. First, network strength monitoring was discussed. Assessing types of network disruption in multi-domain operations could help determine whether the disruption is normal or an attack. Second, image characterization and recognition though traceability was described as key. It is important for analysts to understand why an image processing algorithm classifies a result as a particular solution compared to another. The third awareness opportunity concerned CBRN route planning, which considers specific domain constraints that optimize for safety or mission success.

Execution opportunities were presented as well. Redefining combat power in CBRN response is really about the effectiveness of solving problems. Combat power in this sense is increased when the experts can increase the speed of decision-making. Domain-aware AI allows us to place the tools that these experts need in their hands so that they can be effective and frees up other experts to do other things that AI cannot. These methods may also provide refined analysis enabling more informed and rapid operations in a contaminated battlespace. Applying these methods will reduce decision support timelines for critical consequence management and contamination avoidance missions.

6.3 Extracting Dynamic Proliferation Expertise and Capability Representations from Heterogenous Multilingual Open-Source Data Streams

Maria Glenski, Pacific Northwest National Laboratory; Svitlana Volkova, Pacific Northwest National Laboratory; and Emily Saldanha, Pacific Northwest National Laboratory

Detecting and anticipating proliferation signatures such as expertise and capabilities from unstructured and dynamically evolving real-world data is a challenging but highly desired task that supports the nuclear nonproliferation mission. Existing efforts primarily focus on the detection of proliferation expertise in English bibliometric data via co-citation network analysis, which completely ignores content. In this presentation, a novel AI-driven mixed-method approach was presented that carries out two tasks. First, it fuses a variety of multilingual, heterogenous open-source data streams and converts unstructured data into knowledge. Second, it uses these dynamically evolving proliferation expertise and capability representations to enable predictive modeling and counterfactual reasoning.

Understanding and reasoning in real-time was the objective and was accomplished by summarizing gigabytes of publicly available data with ML, natural language processing, and insights to make the data useful to end-users. These experimental results were demonstrated on in-domain and out-of-domain evaluation, respectively the nuclear and AI domains. Paper content and metadata were both evaluated for various keywords and topics. The relationships

between the content and context were fused to generate concept vectors to quantify the similarity between the various concepts and showed relationships between the different ideas.

This research supplements traditional nonproliferation efforts by detecting, forecasting, and reasoning about illicit proliferation though adding strong multilingual, knowledge representation and summarization, and inference components. These results show representation from many languages and country participation. The topic of papers, specifically nuclear or non-nuclear, can be identified. Taking authorship and other metadata into account, this approach can show insight about the expertise of teams.

6.4 Plutonium Attribution Methodology Development Using Machine Learning Techniques

Patrick O'Neal, Texas A&M University

A nuclear forensics methodology capable of identifying the source of an undeclared plutonium sample would act as a deterrent to potential nuclear proliferation. Previous work at Texas A&M University, developed a methodology able to determine a plutonium sample's reactor of origin, burnup, and the time since irradiation by comparing a set of intra-element isotopic ratios against a database of isotopic ratios using a straightforward maximum-likelihood calculation.

The maximum-likelihood surface generated is different for each reactor. The most likely burnup scenario can be modeled, but a spoofed sample would not fit these typical model profiles. Spoofed samples could come from two different scenarios, but a maximum-likelihood approach will assign a much higher probability for one than the other, and this work therefore focused on a different approach.

To improve the robustness of the methodology, the attribution step used models trained using ML techniques in lieu of the maximum-likelihood calculation. The presented ML approach consisted of a support vector classifier to resolve the reactor of origin and a set of gaussian process regression models to quantify the sample's burnup while the time since irradiation was quantified analytically. This change allowed the methodology to better leverage knowledge about how each isotopic ratio is related to the three parameters of interest as well as to scale the methodology to handle plutonium samples with more complex characteristics.

The previous library relied on the varied isotopic ratios caused by different separation efficiencies in various scenarios. Specifics about the separation physics were required to extrapolate. This work removed the need for that insider knowledge. Using ML instead of maximum likelihood allowed only the ratios that contribute to the determination to be used, resulting in a much leaner approach to solve the attribution. The performance of the ML method was similar to the previous maximum-likelihood method which demonstrated that the method maintains the success of the overall approach while removing a vulnerability.

6.5 Toward Early Intent Detection of Search Queries with Transformers and Experts in the Loop

Adithya V. Ganesan, Stony Brook University

Most technical report search engines retrieve information for a single given search query, but one might be able to infer the end goal or general intent by considering the sequence of multiple

queries. In fact, such intent may often be inferred early in a sequence, before a search is complete, enabling, for example, detection of intent to perform illicit activities (Chen et al. 2019, Hashemi et al. 2016). The goal of this work was to produce a toolkit that can detect user intent based on sequential searches where various AI techniques can be swapped in and out. The approach used both supervised and self-supervised approaches, making it more generalized.

An early event detection algorithm and pipeline to classify sequences were proposed, using only an early subsequence as well as deriving the sequence length necessary to confidently make such classifications. A self-supervised approach was described to classify these sequences of queries by attempting to classify the eventual cluster a sequence will belong to before the sequence is complete. This aids an “expert-in-the-loop” process whereby topical expertise can inform the cluster objectives of the model. In turn, the predictive models seek to be able to learn the patterns in the sequences associated with such expert information to induce the final intent at an early stage.

Specific approaches and their benefits were presented. Transformers that are pre-trained on scientific documents can incorporate domain awareness by using annotations from the experts and include more domain-specific knowledge. Short-text clustering could be useful because it is not a supervised task and can learn attributes. Clustering of a sequence of queries can identify a higher-level representation of queries to be graphed. Clustering can be replaced with other things like topic modeling approaches or linguistic modeling. An expert can be added where class models are assigned to a cluster label, which can then get fed into the next step.

6.6 Proliferation Monitoring with Hidden Markov Models

Andrew Hollis, North Carolina State University

A longstanding goal in nonproliferation research has been the monitoring of development, manufacturing, or testing processes that might present a proliferation risk. For a particular process, it is desired to determine what activity is underway by using a combination of observed data and subject matter expertise about the process. In many cases, the data gathered from standard monitoring and surveillance systems do not yield direct knowledge of the activities underway.

This presentation described a model that allows for the inference of the process activity based on what is observed. Using hidden Markov models (HMM) (Rabiner and Juang 1986), a probabilistic model was developed that encodes subject matter knowledge about the process and can be used to infer and characterize processes.

This model described the unobserved process of interest, the observed process data, and the relationship between the process and the data. Given specific observations, the model could infer the most likely activity at a given time. Like any statistical model, the parameters must be fully specified for the initial state, observation, and transition.

The case study presented looked at the Dry Alluvium Geology (DAG) experiment. Observation data included equipment in use at certain time intervals. Domain awareness was incorporated using a discrete event simulator which incorporated a process model built by experts specifically for the DAG experiment. This allowed estimation probabilities to be determined for activity completion times. This resulted in a determination of the most likely scenario and characterization of other possible scenarios with uncertainty. Additionally, this model can produce a likely sequence of activities, including predictions of the process start and end.

7.0 Sparse Data and Rare Events

This session focused heavily on the problem of having insufficient data to constrain the solution via purely data-driven methods. This problem often does not stem from a lack of data. In fact, many of the projects described large datasets consisting of extended periods of monitoring using large numbers of sensors and multiple modalities of sensors. Instead, this problem stems from a severe class imbalance in the dataset. While a large amount of data is collected, only an small percentage actually describes the phenomenon of interest that researchers are attempting to identify or characterize.

This problem imposes restrictions on the methods that can be used to model the data. First, the method chosen must be able to deal with having a smaller than ideal quantity of data related to the phenomenon of interest. This means choosing techniques that preserve information and/or bring in information known from outside the data. Second, the method chosen must deal with the severe balance issues. This means choosing techniques that prevent overoptimizing on the null data without also introducing an unacceptable level of false positives.

The problem of extreme amounts of data, but scant amounts of data related to phenomena of interest is an common one in the DNN space and is often a driver of incorporating domain-aware methods into neural analysis. The problem was elegantly described in Dr. Myers' keynote as the "small n, large p" problem. While this problem is felt across many if not most presentations in the workshop as a whole, perhaps it is felt most acutely by some of the presentations in this session.

To combat this problem, presenters turned to a wide variety of methods to make their problem tractable. Some presentations hugged very close to data-driven methods with only small tweaks to provide domain awareness such as domain-aware data segmentation. Some of the tweaks were larger such as domain-aware data augmentation strategies, an example of which is shown in Figure 7.1. Almost all presenters turned to (at least minimally) domain-aware feature engineering in order to reduce the dimensionality of their data stream – although the methods by which this was done varied. Finally, some of the work begins to hint at moving toward neural symbolic equation discovery and equation driven state estimation – a rapidly evolving domain-aware literature. Regardless of the method chosen, all of these domain-aware methods were invaluable in making otherwise unsolvable problems tractable.

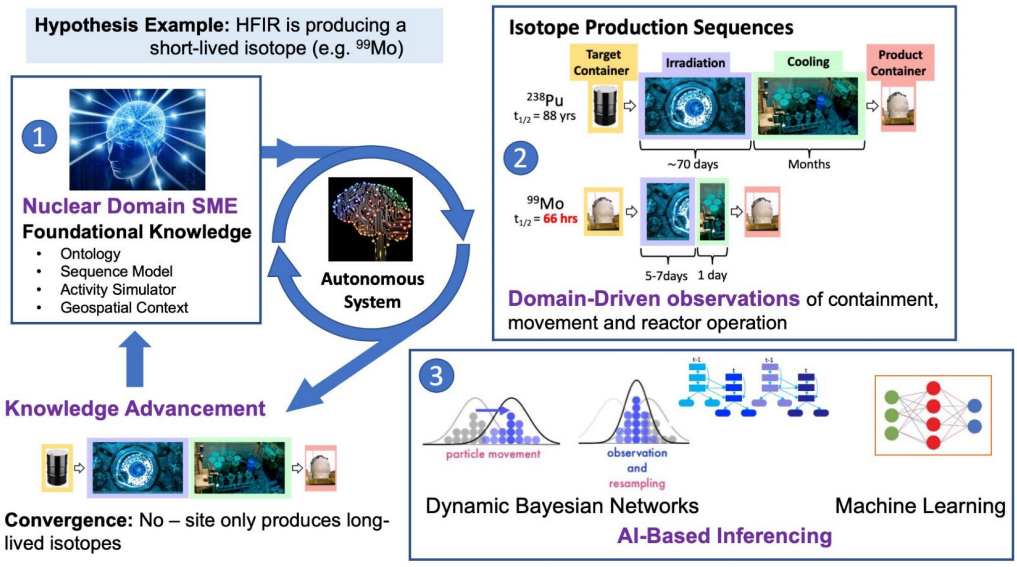


Figure 7.1. Example of cyclically combining multiple types of domain knowledge to inform the collection of new information, followed by the use of AI for analysis. Taken with permission from 7.3, "Persistent DyNAMICS: Remote Sensing Based on Domain-Informed Analytics."

7.1 Constraining Data-Driven Models for Detection of Sparse Temporally Correlated Events

Garrison Flynn, Los Alamos National Laboratory

The MINOS dataset includes data from a very large number of sensors spanning several modalities including radiation, effluent, electromagnetic, thermal imagery, biota, and seismo-acoustic. The sensors considered include one current monitor for EM data, one high-purity germanium detector for radiation data (placed near the stack), three infrasound sensors for acoustic data (placed near the cooling towers), one tri-axial geophone for seismic data (placed near the cooling towers), and one FLIR camera for infrared imaging (placed near the cooling towers). From these raw data streams, features were engineered to provide on the order of tens of features per modality across time. In her work, Dr. Flynn seeks to use a subset of these sensors to estimate the power level of HFIR around which the MINOS sensors are placed.

The method chosen to model the data was a Naïve Bayes classifier (Rish 2005). This model gives the probability of the reactor being in a given power state given the observations from the sensing modalities by directly applying Bayes theorem. The classifier was broken into two stages where the first stage assessed if the reactor was at 0%, 100%, or in between and the second stage predicted the specific power level between 0 and 100%. The data available was limited with only five events where power levels were between 0 and 100% being available. Because there are strong correlations in intra-event data, in a domain-aware fashion four events were used for training and the fifth was held out for testing. It was also important to include the correct combination of modalities. It was found that some modalities provide very useful information and adding other modalities with different information was generally beneficial while adding highly correlated modalities was generally detrimental. The most successful model used thermal, electromagnetic, and effluent data.

Beyond making single predictions about the narrow 30 second windows, the group had the insight that events sequential in time are highly correlated. Therefore, a model that made use of nearby information would likely outperform the single timestep classifier. To this end, the group employed Hidden Markov Models and Naïve Bayes Sequential models. These model forms outperformed Naïve Bayes for the vast majority of combinations of held-out cycles and held-out modalities. Future work looks to change the structure of the hierarchical model to make even better use of known characteristics of reactor behavior and find ways to place additional weight on events that are the most similar to the event in the test set in order to eliminate errors due to event mismatch.

7.2 Domain-Informed Assessment of Nuclear Reactor Operations

Tom Reichardt, Sandia National Laboratories

Dr. Reichardt and his team are also part of the MINOS venture and worked on a very similar problem to that of Dr. Flynn's team, but the methods chosen were significantly different. His team relied on the same infrasound, but rather than staying with a primarily data-driven model, the team-built physics models attempting to connect the signatures collected to what was known about the system that they were investigating.

The infrasound data was converted into a spectrogram via Fourier transforms and then decomposed into spectral and temporal factors via non-negative matrix factorization. Already in these spectral components it is possible to see the effects of some of the changes in fan and pump activity. To attain more specific information about the intensity of the activity, it was necessary to use physics interpretations of the changes in observed frequency. Because the acoustic emanations are highly nonlinear with speed, it was necessary to train a decision tree that related blade passing frequency and intensity to the fan speed and number of fans at that speed. This model was >96% successful. Future models look to model fan speed via equation discovery rather than decision trees with the hopes of being even more accurate and being able to characterize states in addition to 0%, 50%, and 100% (Udrescu and Tegmark 2020, Brunton et al. 2016).

Despite being able to diagnose changes in pump and fan behavior, these activities did not directly correlate with reactor activity. To bridge from cooling behavior to reactor power, it was necessary to model the flows of heat in the system. Heat generated by the reactor needs to be rejected into the environment by the cooling tower. However, the efficiency at which it does so is highly impacted by surrounding conditions. Successful modeling of power behavior required having a model of the heat exchanges of the system as well as knowing the local meteorological conditions to set the magnitude of the loss terms and efficiency of heat exchange in the cooling tower. After applying this heat exchange model, the predictions of the system on reactor power become well-calibrated. Future work looks to continue to refine this model in order to improve the accuracy with which the heat flows are modeled.

7.3 Persistent DyNAMICS: Remote Sensing Based on Domain-Informed Analytics

Thomas Kulp, Sandia National Laboratories; and Sidharth Manay, Lawrence Livermore National Laboratory

The Persistent DyNAMICS Venture is developing a multi-modal activity detection system also using HFIR as a testbed. Sensed quantities include the state, motor state, controller state and movement of key hardware, the pipe state, pump state, water state, flow rate, valve movement, valve state of the water flow, tower state, fan state, plume presence, plume size, and temperature of the heat disposal, occupancy, door movement, and door state of the entrance, presence, movement, and state of targets, and the presence and movement of vehicles and containers. The sensors are cued to make observations 'at the right place and time' based on a 'dynamic persistence' model and processing occurs at the edge such that the transmitted information is reduced to compact textual sensed information. These compact transmissions avoid site-specific info and instead focus on transmitting science-constrained activities in a domain agnostic Lexicon. All of these observations are orchestrated and synthesized by PD-LIVE, a system that infers site operational state and processes from the data.

Subject matter expertise regarding site industrial process is encoded into the KMS [knowledge management system]. This process allows the system to take that information and make decisions in a maximally domain-aware fashion. This formatted knowledge is used to support testing of hypotheses about the facility and the activities that are occurring using AI-based inferencing tools. The KMS interfaces between the hypothesis and the autonomous system to make observations, employ inferential tools, and update site knowledge based on the inferences in a process termed the 'knowledge update cycle.'

A sample use case for this system might include using the sensed data to tell if a target facility is consistent with production of a short-lived medical isotope or with the production of a plutonium isotope for a radioisotope thermoelectric generator. Current focus on characterization of industrial activities focuses on determining what activity they are doing and where they are in the process.

Functional tools used to make these decisions include a sequence model which informs dynamic Bayesian networks and case-based reasoning and a high-level process model which is a generative simulator of state vectors for ML. Finally, the data provided is used to train data-driven random forest models which implicitly embed the foundational knowledge extracted by the components earlier in the system.

7.4 Node and Region Importance for Classifying Nuclear Operations using Multisensor Arrays

Jake Tibbetts, University of California Berkeley

Like the presentations for Dr. Flynn and Dr. Reichardt, this presentation focused on data taken in the vicinity of HFIR with the goal of classifying the reactor power level (this time looking at on/off instead of transient power level). However, the data that was used was different than that collected by the MINOS team. Instead, this data was collected by the SNITCHES team via 12 Merlyn multisensory platforms. These platforms collect data describing magnetic field,

acceleration, pressure, temperature, and ambient light at 16 Hz and report the mean and variance at 10-minute intervals. These sensors were deployed in April 2019.

The nodes were deployed in a spatially distributed manner and the project sought to identify the nodes most useful to prediction accuracy. In order to perform this assessment, the project looked to wrapper methods including Leave One Covariant Out (LOCO) (Lei et al. 2016) and Forward Feature Selection (FFS) (Guyon and Elisseeff 2003). LOCO involves iteratively training a network and removing one input stream at a time to determine the effect on network accuracy. FFS involves iteratively training a network and greedily selecting one input at a time to maximize network accuracy. These wrapper methods were performed both for single nodes and for regional groups of nodes. This regional grouping allowed the wrapper methods to identify the importance of a geographic area in a domain-aware fashion.

Results suggest that nodes near the cooling tower were the most influential to increasing performance followed by nodes near the processing facility. Nodes between HFIR and REDC, near the target processing facility, and near the main complex entrance were detrimental to model performance. Because the wrapper method provided explanations about node importance to an otherwise uninterpretable model, it is now possible for subject matter experts (SMEs) to view the wrapper methods in a domain-aware fashion to make hypotheses about why the various regions had the predictive power indicated by the results.

7.5 One-shot Target Detection via Physics-Informed Training

Natalie Klein, Los Alamos National Laboratory

Longwave infrared (LWIR) hyperspectral data detected by an airborne sensor yields signals that are starkly different than the emissivity signals that leave the ground. Blackbody radiance, downwelling radiance, atmospheric transmission, upwelling radiance, etc. all work to modify the signal such that the detected spectrum differs from the spectrum at the origination. It is desired to be able to determine the source of the signal corresponding to any detected spectrum and to be able to match that source spectrum to a library of materials.

Using purely data-driven methods would require a prohibitive number of labeled emissivity/detection pairs. However, the physics of the relationship is known such that for a given emissivity it is possible to use a physics model to construct many physically realizable detection spectra that span the space of expected detections for that emission source. Thus, it was possible to generate physically informed, domain-aware input-output pairs for a network to learn.

The actual architecture instantiated to do the learning was a Paired (see also Matching, Siamese) Network. Paired networks present the network with two inputs identified as either belonging to the same class or as belonging to a different class. Those inputs are put through identical encoders to form latent vectors. The loss function incentivizes like inputs to be put close together in latent space and unlike inputs to be put far apart in latent space. These networks are popular elsewhere for allowing semi-supervised training – with the challenge that finding optimal negative examples is an open science question (LeCun and Misra 2021). For this problem, Paired networks were chosen because they excel at one-shot learning. This has been shown previously on image, language, and even hyperspectral data (Anderson et al. 2019, Koch et al. 2015, Vinyals et al. 2016).

Once the network was trained on the synthetic signal pairs, it was possible to apply the network to new materials under novel conditions. The test data obtained ROC scores very similar to that of the training data – showing excellent separation between pairs of the same material and pairs of different material. By applying principal component analysis to the latent vectors, Dr. Klein's team was able to show tight clustering in the latent space for groups of the same material – including groups formed from materials not seen during testing. Future work looks to apply explainability tools (e.g., LIME), extend to applying the method to gases, and extending the physics model generating the input-output pairs to accommodate the new inputs.

8.0 Robust Deployment and Decision Support

A key theme that emerged in this session is the necessity of integrating guidance, feedback, or knowledge representations from SMEs as a requirement for building robust and effective models to deal with applications in the nuclear, and related, domains. This echoes similar emphases from keynote presenters and panel discussions on the benefit and need for incorporating domain experts early and often in the lifecycle of development and deployment of AI-based solutions.

Presentations in this session introduced approaches for the development or evaluation of domain-aware AI methodology that incorporated human-in-the-loop feedback from domain experts or included domain knowledge in the foundation of the ML models and algorithms used as a means to bolster model performance in the face of limited training data, biased datasets, and new environments encountered in testing or deployment. It is well known that AI models are often brittle and fail when expected to perform on out-of-domain inputs or in new environments, examples of which are shown in Figure 8.1, and current approaches to increase the robustness of these models rely on increasing the diversity or generalizability of training data. One method to do so is to increase the scale of data, and attempt to sample from all environments that may be encountered. However, this is difficult or infeasible for many domains and in particular the nuclear domain where representative or large-scale ground-truth datasets simply do not exist.

Several presentations illustrated how domain-aware methods can be used to validate AI models or AI components of systems that predict behavior and to determine whether the AI predictions are robust or whether the accuracy in test beds is more reflective of overfitting to the training or testing data. For example, the third presentation highlighted the benefit of incorporating domain knowledge to evaluate model performance across the wide range of variations that would be encountered in the wild. The final talk presented a corruption recovery approach to transform new, unexpected inputs to variations of inputs that a model was trained on, using a transformation that can be adapted on demand to increase robustness in unseen environments.

Others highlighted the ways in which domain expertise can be leveraged to enhance the data provided to models, by reducing noise and removing inconsequential data features. For example, the first talk presented an approach that uses tensor decomposition and subject matter expertise guidance to reduce large-scale measurement signals from seismic and power sensors to meaningful representations needed to identify when industrial activities such as firesets, emplacement, and stemming occur. This approach leverages the inclusion of SMEs during development (a workflow style that Dr. Hague advocated for during her keynote “A Perspective from the Analytic Intelligence Community”) to reduce or remove the noise in measurement datasets, which is an approach to mitigate the “small n, large p” issue discussed by Dr. Myers in her keynote “Domain-Aware AI: There and Back Again.”

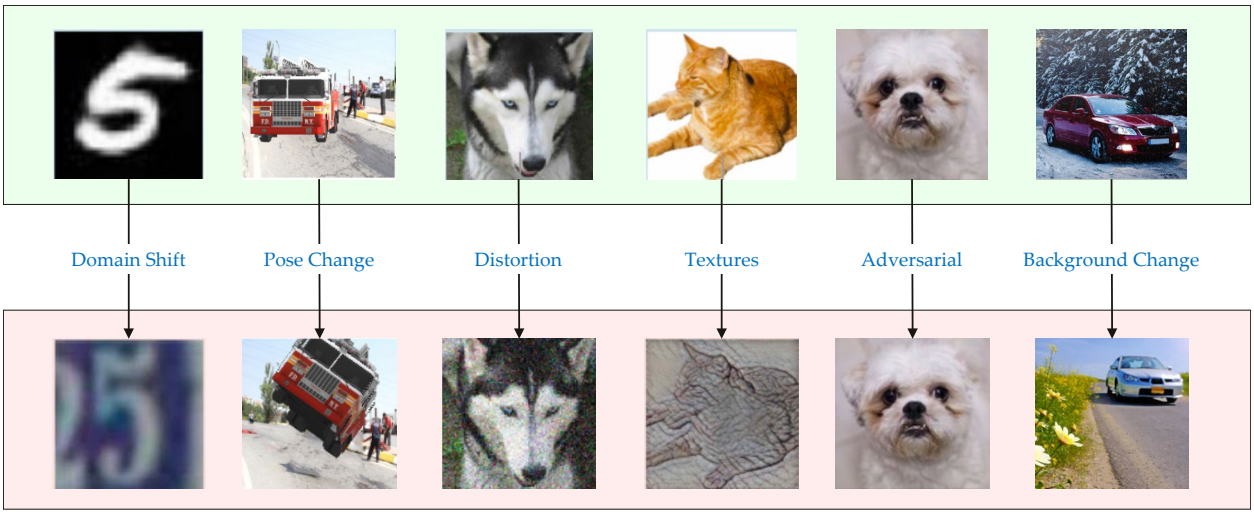


Figure 8.1. Examples of variations in an image context that are found in the wild that may not be present in the training data, leading to unpredictable behavior of AI systems. Taken with permission from 8.3, "Robustness in the Wild using Domain-Aware Surrogate Functions." Originally adapted from Geirhos et al. 2020.

8.1 Annotation Transfer for Prediction of Industrial Operations

Erik Skau, Los Alamos National Laboratory

This presentation highlighted the physical explainability of latent features and their application in guiding feature selection in downstream predictive tasks, aided by SME feedback, or as a preprocessing methodology for large-scale datasets to support transferability of annotations to new target datasets. Using matrix tensor decompositions to approximate data inputs (structured as tensors) as summations of tensor products of triplets, enables a reduction in dimensionality and can be used to denoise the data. However, the number of fundamental components to reduce to is a key parameter to be tuned – overestimating leads to fitting noise in the data while underestimating the number results in loss of fundamental information. Communication with SMEs or the use of supervised techniques to understand the decomposition of features is key to avoid either extreme and can be used to improve predictive models or interpretability of models that are fed the decomposition of features as input. This technique of incorporating tensor decomposition and SME feedback on what the decomposition represents was applied to data from the DAG test bed, using data made available by the Advanced Data Analytics for Proliferation Detection (ADAPD) project.

In this application, there is seismic data from geophones and data collected from electric power meters, and the downstream predictive models seek to predict industrial activities of interest (firesets, emplacement, and stemming) for which there is an annotated calendar of known industrial events. Examining the latent tensor features for the seismic data, SMEs were able to identify the decomposition was a combination of the spectral pattern, temporal pattern, and angular pattern; effectively the decomposition represented what was happening, where, and when in a very manageable and interpretable representation for SMEs. Further, after interpreting this decomposition, SMEs were able to provide further guidance to clean the decomposition and remove irrelevant signals. Support Vector Machine models for stemming and emplacement using the seismic data decompositions were found to outperform similar models that did not leverage the technique, as illustrated by higher performance according to

ROC curves. After identifying that seismic information was able to transfer from one experiment (DAG 2) to another (DAG 3), a combination of latent feature fusion and graph interval techniques was used to transfer information between seismic and power. Using resulting models that use both seismic and power signals was found to outperform the models that rely on seismic inputs alone (Prasad et al. 2020) illustrating how combining unsupervised tensor decompositions with supervised approaches can reduce complexity, improve performance, and increase the interpretability of the supervised algorithm. Physically interpretable decompositions such as those described in this talk can be related to SMEs to provide meaningful guidance or feedback on what to discover or investigate to understand large datasets and guide the development of interpretable predictive models.

8.2 Automated Synthesis of Soft Labels using Neural Stochastic Differential Equations and Attribution-Based Confidence

Sumit Kumar Jha, University of Texas at San Antonio

Widely used benchmark image datasets often have hard labels—i.e., singular labels that should clearly annotate each example—for each image but the reality is often that these datasets have examples where hard labels that are difficult to justify or where the image does not have a clear, single label. In contrast, they may have soft labels where there are multiple labels or a level of uncertainty in labels where it may be hard to distinguish between more than one label, which can cause confusion. When dealing with this issue, there are two key questions that were covered in this presentation:

1. How can we algorithmically detect images in large data sets that clearly require soft labels? (e.g., in ImageNet where the scale of the dataset makes it hard to identify which need multiple soft labels, creating challenges for a manual approach)
2. How can we algorithmically identify candidate soft labels for images that require them, in a way that SMEs can tweak?

Dr. Jha addressed how to detect images that require soft labels using ensembles of neural stochastic differential equations and a novel attribution-based confidence (ABC) metric (Jha et al. 2019) that is used to compute the probabilities of soft labels. The ABC metric calculates the probabilities for each of the soft labels by focusing not just on the input image but also on the explanations for each class, i.e., why an input should belong to a certain class. These explanations can include succinct explanations of class compatibility from existing explainability tools for machine learning predictions, such as LIME (Ribeiro et al. 2016). Finally, several avenues of future work in the proposed approach were introduced including improving the soft labels that are identified using neural networks trained on the initial soft labels, controlling the training process to remove errors such as those caused by adversarial examples, and incorporating domain knowledge into the process as a means to improve performance.

8.3 Robustness in the Wild using Domain-Aware Surrogate Functions

Jay Thiagarajan, Lawrence Livermore National Laboratory

When developing ML models, it is difficult to impossible to consider every variation in order to handle data not only in the real-world usage but also new environments. It was noted that “machine learning models behave unpredictably when they are not exposed to the variations

expected in the wild" during their training process. This presentation focused around how to use domain knowledge to identify the problems that would not be expected in development, with the emphasis that achieving robustness in the wild requires systematic integration of domain knowledge in the training process. He illustrated that the issue of model robustness goes beyond answering the question of whether the model behaves similarly to how humans have labeled the known data (e.g., the annotations on which the model was trained, and would be expected to mimic) but also whether the model behaves as you would it expect it to in new applications. For some conditions, such as scrambling image pixels, where models generalize more effectively than humans. However, for other conditions such as domain shift, distortion, adversarial attacks, or background changes in image inputs, models fail whereas humans are naturally able to adapt.

Although many existing techniques on model robustness focus on methods such as adversarial training which is popular in the general computer vision community, used to increase model generalizability by introducing more and more variations of the data inputs. However, these approaches are often limited to simple pixel variations, which are insufficient to model the wide range of variations encountered in the wild. Attribute-Guided Adversarial Training (AGAT) (Gokhale et al. 2020) is the proposed solution for robust model development and design.

AGAT focuses on emulating the data variations in a natural manner, parameterizing the input space with relevant attributes, and going beyond the training examples to maximize exposure to combinations of attributes without having access to the test domain. As a result, AGAT can support a broad range of specifications for domain-specific applications and use domain expertise via human-in-the-loop approaches where experts can specify the surrogate functions used to transform images to reflect required changes to the example – integrating domain knowledge in the training process. It was illustrated that AGAT can effectively outperform other approaches when applied for object-specific attribute changes, geometric transformations, and natural image distortions (noise, blur, weather, pixelation). When queried, he identified that the most challenging aspect of this approach is creating the connections between domain knowledge and that this solution is domain independent – it will be effective for any domain for which one can create a domain knowledge representation of the space, although data or problems with better understandings of what variations can be expected that can be converted to clear mathematical formalisms have an advantage over others.

8.4 A Computational Framework for Deterrence Assessment Analyses

Michelle Quirk, NNSA Office of Advanced Simulation and Computing (NA-114)

Deterrence operations "convince adversaries not to take actions that threaten U.S. vital interests by means of decision influence over their decision-making," essentially acting to prevent bad actors from doing bad things. This presentation introduced a computational framework that leverages intelligent cognitive assistants (ICA), domain knowledge, and cognitive sciences for the application of deterrence assessment and analysis. The presentation highlighted that successful policies are considered in regard to both the military domain and political, or socio-economic impacts. The most common deterrent persuades an adversary not to carry out intended actions because of costly consequences. Deterrence analyses consider adversary calculations consisting of (1) benefits of the given course of actions, (2) the costs of the course of action, and (3) the consequences of restraint, i.e., the costs and benefits of not taking the

course of action sought to deter. These key features of adversary calculations are the core of the framework presented.

With the aid of protoforms, the framework is able to formalize deterrence analyses that transform generalized queries in the form of “Does adversary A consider the use of weapon W against U.S. interest/target T in the context C?,” where the context can be geopolitical, economic, or social and reflect the complexity of the problem, to simplified structures “If (A and W and T and C) then D?” (where D represents deterrence options) for which an automated ranking of deterrence options can be computed. Often these analyses are static and do not support significant reuse of results, with one method being the manual creation of colored tables of scores and risks using the typical high (red), medium (yellow), low (green) color mapping. In comparison, the framework presented uses protoforms and computational “perception of the value of the threat” functions, represented as theta functions where theta is bounded by 0 and 1 to provide a more continuous representation of the high/medium/low categorization of threats to automate adversary calculations in a manner that is generalizable and reusable. This framework can incorporate not only analyst perceptions of adversaries, but also adversary’s perceptions of analyst perceptions and analyst’s perceptions of adversary’s perceptions of analyst perceptions – what adversaries think analysts know about them and what analysts think adversaries think they know about analysts.

The iterative approach for perceptions and adversary calculations enables this framework to fully embed strategic and cultural aspects. However, there is a tradeoff between losing information (e.g., narratives or details that are found in longer reports of 500 pages and more) and the speed and structural gains when automated. The need of continuous evaluation with input from deterrence experts, decision makers, and the knowledge engineers throughout the deterrence analysis lifecycle was emphasized. Doing so allows the framework to leverage automation gains as well as domain knowledge of deterrence experts and cognitive sciences. When queried about the use of probability distributions for ranking deterrence, it was noted that the proposed framework does not just use probabilities but supports the inclusion of analyst’s domain knowledge or intuition; One cannot just assign a number to a deterrent easily because the context matters. Another benefit to the framework is that it is dynamic, where knowledge of adversaries can be updated as dynamics change or more is learned about the adversary evolving over time.

8.5 On-the-Fly Robustness in the Wild via Data-Driven Generative Priors

Rushil Anirudh, Lawrence Livermore National Laboratory

AI models are typically designed under controlled training settings before being deployed in the wild, where, in contrast, there is little to no control over how inputs may change or become distorted. It is well known that ML models fail when they are deployed outside the training data, as they were not tested against the kinds of variations of inputs that they will encounter in the wild when tested in such a controlled setting. Some of the uncontrollable distortions that sensor-based systems can encounter include environmental distortions caused by weather or changes in lights or sensor distortions caused by missing or broken sensors. Pre-trained models are often trained using cleaned datasets that were collected under ideal controlled conditions. It is not always possible to make changes to the classifier to make it more robust, e.g., because data is changing over time, which would require constant retraining and redeployment or because the classifier is proprietary, or the result of several workflows combined. In this presentation, a

model independent approach to robustness, MimicGAN, is described that leverages signal recovery on inputs in deployment to enable classifiers to remain robust even when the inputs shift (Anirudh et al. 2020; Anirudh et al. 2021).

In a motivating example, it was illustrated how a corruption agnostic data recovery process was able to significantly improve model performance for a deep neural network image classifier: 7.2% accuracy increasing to 46.4% for negative blur distortions and a 25.5% accuracy increasing to 45.5% for dropped pixel distortions. For these corruption models, every sample in the wild is assumed to be a corrupted version of the clean dataset. MimicGAN performs data cleaning using a manifold projection, which is able to clean noisy data encountered in deployment by solving a noise model formula and projecting it onto an approximated space of the clean data. One approach to solving the noise model formula is to use projected gradient descent (PGD), however this optimization is not robust when the corruption function is unknown. This causes PGD to fail when the test distribution differs from the training distribution.

In comparison, the MimicGAN approach is able to adapt “on the fly” to new distortions and leverages an iterative approach to estimate the clean solution and corruption until the estimations converge. More robust performance was shown using input images with varying degrees of rotation, where MimicGAN consistently reconstructed the image without rotation and other existing approaches (PGD, ResNet+PGD, iGAN) were not as successful. In discussion after the presentation, Dr. Anirudh highlighted that there is an assumption that every layer in the MimicGAN network is some corruption of interest, and that there is a guarantee that the approach is robust to all combinations of those functions. However, there has not been a clear study comparing leveraging domain knowledge versus noise distributions, so the inclusion of domain knowledge is still dependent on the application.

9.0 Panel Discussion: Requirements and Opportunities for Domain-Aware Methods in Proliferation Detection

A guiding principle of the DNN R&D portfolio and a key takeaway from the first *Next-Generation AI for Proliferation Detection Workshop* is the importance of tying the research as directly to the mission as possible. To reinforce this idea, a panel was held at the end of the workshop to refocus the discussion from specific technical approaches to mission considerations. Panelists were chosen from the keynote speakers and session chairs to provide representation from end-user, researcher, and research sponsor communities.

Panelists:

- Emma Hague, Chief Data Scientist, DOE Office of Intelligence Foreign Nuclear Programs Division
- Kary Myers, Statistical Sciences Group, Los Alamos National Laboratory
- Becky Olinger, Portfolio Manager for Nuclear Threats Detection, Nuclear Detection Division, Defense Threat Reduction Agency
- Angie Sheffield, Senior Program Manager for AI and Data Science, Nonproliferation Research and Development, National Nuclear Security Administration

Moderated by Tammie Borders, Technical Advisor for AI and Data Science, Nonproliferation Research and Development, National Nuclear Security Administration

Each panelist was given time for a brief opening statement and then a series of questions to prompt discussion were proposed. Questions are shown below.

Moderator questions:

1. If you ask five people about the definition of domain-aware techniques for AI, you are likely to get five definitions. What do you think of when you hear domain-aware techniques?
2. For our mission partners, what are the most pressing needs that you believe can be addressed by incorporating domain-aware techniques into AI systems or analytics methodologies?
3. Working to bring the mission and research closer together, what are barriers to using AI models to support the intelligence community?

Audience questions:

1. What are your thoughts on role of AI/ML to achieve levels of shared autonomy between machines and human experts in national security? For example, from level 0 (no automation) to level 5 (full automation).
2. The fast pace of innovation and development in AI can be partially attributed to the combination of open-source software, collaborative software tools, and a strong emphasis on open and reproducible work. This environment enables comparison and development even in niche fields. For our niche field, security and classification boundaries complicate things. Do you have thoughts about how the national labs could think about adapting to take

advantage of these concepts (i.e., better share, review, and develop as a community) rather than in a siloed fashion?

3. Have you found any resistance to presenting results to stakeholders with uncertainty? Do they understand how to interpret it or does it diminish the results if they are presented as "uncertain"?

In the opening remarks, it was discussed that domain-aware AI presents opportunities to create collaboration between data science and domain experts. This is important as while it was initially thought that data-driven methods could provide special insights, particularly given the success of data-driven AI in the business world, this has been found to not translate to nonproliferation. Fortunately, the workflow to operationalize domain-aware approaches are close to being "solved" and teams can focus on more time-consuming or difficult tasks, such as new observation or data acquisition opportunities for a challenge such as treaty verification. Next, opportunities to apply these techniques must be further identified. On the DoD side, AI warfare won't center on one technology, but rather integration of multiple technologies. An open question is how to understand what is "steady state" and what rises above a threshold to need an expert opinion. Common myths and misconceptions about AI/ML were also shared as reflected in the previous workshop report (Alexander et al. 2020).

It was suggested that mere "pattern-matching" does not work for proliferation detection and that domain-aware AI requires collaboration and time investment to ensure that the data meets the assumptions of the method being used and the mathematics are applied in an informed way. It is important that the training data span the space of interest. Pressing problems in proliferation detection as directed by DoD that were noted include the need to acquire decision-making superiority or information superiority. In other words, there is a desire to make faster, more reliable decisions and to share information more effectively between agencies. These are challenges that domain-aware AI can help address, and while a new defense strategy may be developed, the priorities will remain the same.

Turning to the intelligence community, "augmentation" and "automation" are stressed. To help facilitate the adoption of AI, tools need to be demystified, and relationships between researchers and analysists need to be built and strengthened. Keeping an open mind and ensuring a two-way communication flow will dramatically help. Automation is on the way, and all levels of automation will eventually be implemented depending on the application, ranging from minimal to full. There is reason for optimism about capabilities that stop short of full automation, for example looking at a computational feedback loop. It was noted that AI can accomplish a lot, but that it will not solve everything.

Another key aspect of operationalizing AI is to intentionally develop it to be shareable. This means planning open-source projects with best practices and working with technology transition offices and partners. Good documentation is important, and code should be shared as much as reasonably allowed. Standards for open-source software are also worthwhile to consider. Ending on an especially promising note, the panel reported that very little resistance is encountered when presenting results from algorithms with uncertainty. Decision makers understand that the rigor required is never going to provide a completely certain answer. Decisions will never be made solely on an algorithms output, and so answers with uncertainties are helpful to contribute to an improved decision-making process and outcome.

10.0 Conclusions

Over the course of the two-day workshop, many presentations and panel discussions demonstrated applications of domain-aware AI technologies to proliferation detection challenges and identified potential opportunities for future research. While there remain challenges in developing deployable AI for national security applications, there are tremendous possibilities moving forward.

Further, domain-aware approaches are key to improving generalizability and transferability and to ensure the creation of useful and robust models suitable for high-consequence missions. The time is ripe to expand the role of AI in nuclear proliferation detection. Not only is progress being made on the technology, but end-users recognize the potential of AI to create new capabilities. AI systems do not necessarily need to be perfect to be helpful, and through interactions with end-users, researchers can identify opportunities to make substantial impacts, including in the near-term.

Key Finding: AI Techniques for National Security Must Adhere to High-Consequence Mission Requirements

National security missions must employ AI methods across multiple data types including sensor, technical, scientific, signals and others. Standard machine learning models largely focus on image and text data sets. An emerging challenge is the advancement in data fusion algorithms for all-domain operations and AI-enabled, accelerated discovery and decision intelligence. Many operational environments will have degraded or uncharacterized conditions and developing robust models that perform predictably will require new mathematical approaches, including domain-aware methods. AI methods for proliferation detection must support high-consequence mission requirements and perform well in complex and noisy environments, excel at rare event discovery particularly in sparse data, be robust and predictable when deployed for decision support, and discover signatures for early proliferation detection beyond what is possible today.

Key Finding: Domain-Aware Methods Combined with Machine Learning Have Great Promise

Analytics methods that are solely data-driven are insufficient in national security because data is sparse, incomplete, and noisy. Data-driven approaches forego inclusion of key mission-relevant information found in subject matter expertise, computational simulations, mission requirements, and other traditional domain-aware methods and data sources. This workshop demonstrated a variety of ways in which domain-aware methods can be used to overcome these shortcomings.

Despite the importance of domain-aware techniques and their history of use, there remains no accepted taxonomy. Five categories were discussed: expert knowledge, synthetic data generation, inclusion of non-traditional AI/ML methods into traditional AI/ML models, semantic or constraint-based methods, and soft labels. Further, it was observed these techniques are more often used in combination rather than in isolation.

Key Finding: Operational AI for Decision Intelligence Requires a System of Systems Solution

> The workshop reinforced a few considerations that are essential in developing AI systems. In particular, understanding the context for their application is key, and the formation of shared mental models between system developers and end-users ensures the usefulness of models developed. A key input to the mental model is the understanding that AI systems will continue to be used by humans and so must be human-centric or else inevitably have a high probability of failing. Components of the system of systems solution include the human in or on the loop, AI-based, and traditional approaches integrated into a decision intelligence framework.

As researchers and system developers create game-changing AI technologies, this workshop makes it evident that one of the first considerations in developing AI systems should be how to leverage and incorporate domain knowledge. Such domain knowledge can come in many forms such as outputs from computational models and simulations; scientific, engineering, and operational requirements and constraints; subject matter expertise; and other data sources. While the sources of domain awareness may vary, a constant fact is that ignoring these additional sources of information sells short the impact that AI can have on proliferation detection.

11.0 References

Alexander FJ, T Borders, A Sheffield, and M Wonders. 2020. “Workshop Report for Next-Gen AI for Proliferation Detection: Accelerating the Development and Use of Explainability Methods to Design AI Systems Suitable for Nonproliferation Mission Applications.” doi:10.2172/1768761.

Anirudh R, JJ Thiagarajan, B Kailkhura, et al. 2020. “MimicGAN: Robust Projection onto Image Manifolds with Corruption Mimicking,” *International Journal of Computer Vision*, Vol. 128, pp. 2459–2477.

Anirudh R, S Lohit, and P Turaga. 2021. “Generative Patch Priors for Practical Compressive Image Recovery,” *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2535-2545.

Anderson DZ, JD Zollweg, and BJ Smith. 2019. “Paired neural networks for hyperspectral target detection.” In *Proceedings Volume 11139*, *Applications of Machine Learning*, 111390J, *International Society for Optics and Photonics*, San Diego, California. https://doi.org/10.1117/12.2531310.

Baker N, F Alexander, T Bremer, et al. 2019. “Workshop Report on Basic Research Needs for Scientific Machine Learning: Core Technologies for Artificial Intelligence.”

Brost RC, WC McLendon, O Parekh, MD Rintoul, D Strip, and D Myung-kyung Suh. 2014. *Facility Search in Remote Sensing Data Using Geospatial Semantic Graphs*. SAND2014-1362C, Sandia National Laboratories, Albuquerque, New Mexico. https://www.osti.gov/servlets/purl/1141268.

Brunton SL, JL Proctor, and JN Kutz. 2016. “Discovering governing equations from data by sparse identification of nonlinear dynamical systems.” In *Proceedings of the national academy of sciences* 113(15):3932-3937. https://doi.org/10.1073/pnas.1517384113.

Chen Q, Z Zhuo, and W Wang. 2019. “BERT for Joint Intent Classification and Slot Filling.” *arXiv:1902.10909*.

Fu C, A Di Fulvio, SD Clarke, D Wentzloff, SA Pozzi, and HS Kim. 2018. “Artificial neural network algorithms for pulse shape discrimination and recovery of piled-up pulses in organic scintillators,” *Annals of Nuclear Energy*, Vol. 120, pp. 410–421.

Gastelum Z, MR Smith, and M Hamel. 2019. *Detecting Anomalies in Safeguards Surveillance Data*, Sandia National Laboratories. https://www.osti.gov/servlets/purl/1643860.

Geirhos R, J Jacobsen, C Michaelis, R Zemel, W Brendel, M Bethge, and FA Wichmann. 2020. “Shortcut learning in deep neural networks,” *Nature Machine Intelligence*, Vol. 2, pp. 665-673.

Gokhale T, R Anirudh, B Kailkhura, et al. 2020. “Attribute-Guided Adversarial Training for Robustness to Natural Perturbations.” *arXiv:2012.01806*.

Grimes T, E Church, W Pitts, et al. 2020. "Adversarial Training for EM Classification Networks." *arXiv:2011.10615*.

Guyon I and A Elisseeff. 2003. “An Introduction to Variable and Feature Selection,” *Journal of Machine Learning Research 3*, pp. 1157-1182.

Hashemi HB, A Asiaee, and R Kraft. 2016. “Query Intent Detection Using Convolutional Neural Networks,” *Proceedings of WSDM QRUMS 2016 Workshop*.

Hu J, W Zheng, L Ma, G Wang, J Lai, and J Zhang. 2019. “Early Action Prediction by Soft Regression.” In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 11, pp. 2568-2583, 1 Nov. 2019. doi: 10.1109/TPAMI.2018.2863279.

Ilyas A, S Santurkar, D Tsipras, et al. 2019. “Adversarial examples are not bugs, they are features.” *arXiv:1905.02175*.

Jha S, S Raj, S Fernandes, SK Jha, S Jha, B Jalaian, G Verma, and A Swami. 2019. “Attribution-based confidence metric for deep neural networks.” In *Advances in Neural Information Processing Systems* 32 (NeurIPS 2019). https://susmitjha.github.io/papers/neurips19.pdf.

Joint Publication 3-40, “Joint Countering Weapons of Mass Destruction” available from https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_40.pdf, 2019.

Koch G, R Zemel, and R Salakhutdinov. 2015. “Siamese neural networks for one-shot image recognition.” In *ICML Deeplearning Workshop*, Volume 2. Vauban Hall at Lille Grande Palais, France. http://www.cs.toronto.edu/~gkoch/files/msc-thesis.pdf.

LeCun Y and I Misra. 2021. “Self-supervised learning: The dark matter of intelligence.” Facebook AI. March 4, 2021. https://ai.facebook.com/blog/self-supervised-learning-the-dark-matter-of-intelligence

Lei J, M G’Sell, A Rinaldo, RJ Tibshirani, and L Wasserman. 2016. “Distribution-Free Predictive Inference for Regression.” *arXiv:1604.04173*.

Madry A, A Makelov, L Schmidt, D Tsipras, A Vladu. 2017. "Towards deep learning models resistant to adversarial attacks." *arXiv:1706.06083*.

Nageswara S, V Rao, C Greulich, Satyabrata Sen, et al. 2020. “Classification of Dissolution Events Using Fusion of Effluents Measurements and Classifiers.” *2020 Institute of Nuclear Materials Management Annual Meeting*.

National Intelligence Strategy of the United States of America, 2019. https://www.dni.gov/files/ODNI/documents/National_Intelligence_Strategy_2019.pdf

Ning X, IW Selesnick, and L Duval. 2014. “Chromatogram baseline estimation and denoising using sparsity (BEADS),” *Chemometrics and Intelligent Laboratory Systems,* Vol. 139, pp. 156-167.

NSCAI. 2021. *NSCAI Final Report*, National Security Commission on Artificial Intelligence https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf.

Prasad L, BS Alexandrov, and BT Nebgen. 2020. “Weak matching of temporal interval graphs of sensors for robust multimodal event detection in noise.” In *Proceedings Volume 11423*, *Signal*

Processing, Sensor/Information Fusion, and Target Recognition XXIX, vol. 11423, p. 114230N. International Society for Optics and Photonics. https://doi.org/10.1117/12.2558683.

Rabiner L and B Juang. 1986. “An introduction to hidden Markov models.” *IEEE ASSP Magazine*, Vol. 3, Iss. 1, pp. 4-16.

Rashdan A, M Griffel, R Boza, and D Guillen. 2019. *Subtle Process-Anomalies Detection Using Machine Learning Methods*. Light Water Reactor Sustainability Program, INL/EXT-19-55629, Idaho National Laboratory, Idaho Falls, Idaho. https://lwrs.inl.gov/Advanced%20IIC%20System%20Technologies/Subtle_Process-Anomalies_Detection_Using_Machine-Learning_Methods.pdf.

Ribeiro MT, S Singh, and C Guestrin. 2016. "'Why should i trust you?' Explaining the predictions of any classifier." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135-1144. https://doi.org/10.1145/2939672.2939778

Rish I. 2005. “An empirical study of the naïve Bayes classifier.” *IJCAI workshop on empirical methods in artificial intelligence*.

Shrivastava A, T Pfister, O Tuzel, J Susskind, W Wang, and R Webb. 2017. “Learning from Simulated and Unsupervised Images through Adversarial Training.” In *2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2107-2116. Hawai‘i Convention Center, Honolulu, HI., https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8099724.

Subcommittee on Nuclear Defense Research and Development. 2019. “Nuclear Defense Research and Development Strategic Plan for Fiscal Years 2020-2024.” Committee on Homeland and National Security, of the National Science and Technology Council. https://fas.org/nuke/guide/usa/r&d-plan.pdf

Summary of the 2018 National Defense Strategy of the United States of America, 2018. https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf

Udrescu S-M and M Tegmark. 2020. “AI Feynman: A physics-inspired method for symbolic regression.” *Science Advances* 6(16): eaay2631. doi: 10.1126/sciadv.aay2631.

Versino C and P Lombardi. 2011. “Filtering Surveillance Image Streams by Interactive Machine Learning.” W Lin, D Tao, J Kacprzyk, Z Li, E Izquierdo, H Wang (eds) *Multimedia Analysis, Processing and Communications. Studies in Computational Intelligence*, vol 346. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-19551-8_10.

Vinyals O, C Blundell, T Lillicrap, K Kavukcuoglu, and D Wierstra. 2016. “Matching networks for one shot learning.” In *Advances in Neural Information Processing Systems* 29:3630-3638.

Wang J, Y Tang, J Kavalen, AF Abdelzaher, and SP Pandit. 2018. “Autonomous UAV Swarm: Behavior generation and simulation.” In *2018 International Conference on Unmanned Aircraft Systems (ICUAS)*, pp. 1-8. doi: 10.1109/ICUAS.2018.8453464.

Xiao K, L Engstrom, A Ilyas, and A Madry. 2020. "Noise or signal: The role of image backgrounds in object recognitio." *arXiv:2006.09994*.

Yu S, H Dong, F Liang, Y Mo, C Wu, and Y Guo. 2019. “Simgan: Photo-Realistic Semantic Image Manipulation Using Generative Adversarial Networks.” In *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 734-738. doi: 10.1109/ICIP.2019.8804285.

Zhang Z, Y Wang, C Gan, et al. 2020. "Deep Audio Priors Emerge From Harmonic Convolutional Networks." *International Conference on Learning Representations*.

Appendix A – Workshop Agenda

The attached agenda lists presenters in order of presentation by session.

National Nuclear Security Administration

Next-Gen AI for Proliferation Detection:
DOMAIN-AWARE METHODS WORKSHOP
FEB 23-24

Complex & Noisy Environments
Early Proliferation Detection & Signature Discovery
Sparse Data & Rare Events
Robust Deployment & Decision Support

Sessions 1,3: Sessions 2,4: Link to Session:

Day 1: Tuesday February 23, 2021

Session 1		
1200	Welcome	**Marc Wonders** NGFP/PNNL
1205	Next Generation AI for Nuclear Proliferation Detection	**Angie Sheffield** | NNSA
1230	Introduction to Domain Aware Methods	**Ashley Shields** | INL
1250	A Perspective from the Analytic Intelligence Community	**Emma Hague** | DOE-IN
1320	**Break**	

Complex and Noisy Environments

Session Chair: **Boian Alexandrov** | LANL
Note Taker: **Jesse Ward** | PNNL

Session 1: Continuing (Same as Plenary Session)		
1330	Imposing Harmonic Structure in Neural Networks	**Mark Adams** ORNL
1350	Domain Adversarial Networks and Explainability Assessment	**Tom Grimes** PNNL
1410	Dissolution Event Classification Using Isotope Decay Chains and Half-Life Estimates	**Nageswara Rao** ORNL
1430	**Break**	

Early Proliferation Detection and Signature Discovery

Session Chair: **Zoe Gastelum** | SNL
Note Taker: **Jen Mendez** | PNNL

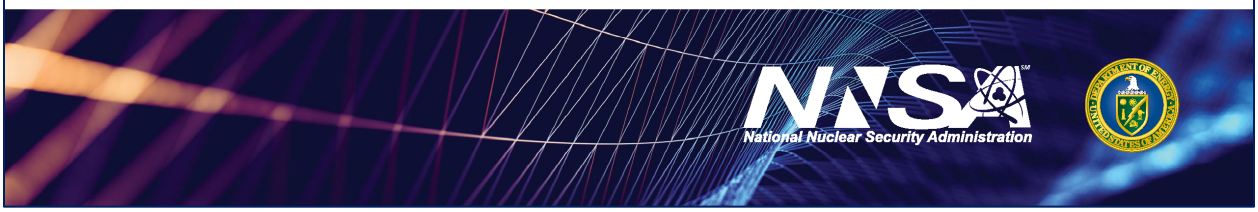
Session 2		
1330	Open-Source Data Analytics Value Quantification to Inform and Explain Radiological Source Detection Localization and Tracking	**Sannisth Soni** **Svitlana Volkova** **Ellyn Ayton** PNNL
1350	Applying Domain-Aware Artificial Intelligence on the CBRN Battlefield	**Adam Seybert** US Army Nuclear and CWMD Agency
1410	Extracting Dynamic Proliferation Expertise and Capability Representations from Heterogenous Multilingual Open-Source Data Streams	**Maria Glenski** **Svitlana Volkova** **Emily Saldanha** PNNL
1430	**Break**	

1440	Inferring the Dynamic Location of an Environmentally-Constrained Radiative Source with a Network of Detectors	**Dave Osthus** LANL
1500	An Artificial Neural Network System for Special Nuclear Material Detection in Photon Based Active Interrogation Scenarios	**Abbas Jinia** University of Michigan
1520	Spectral Signatures for Shielded Sources	**Jason Hite** ORNL
1540	**Closing Day 1** (in session)	**Boian Alexandrov** LANL

1440	Plutonium Attribution Methodology Development Using Machine Learning Techniques	**Patrick O'Neal** Texas A&M University
1500	Toward Early Intent Detection of Search Queries with Transformers and Experts in the Loop	**Adithya V Ganesan** Stony Brook University
1520	Proliferation Monitoring with Hidden Markov Models	**Andrew Hollis** NC State University
1540	**Closing Day 1** (in session)	**Zoe Gastelum** SNL

Day 2: Tuesday February 24, 2021

Session 3		
1200	Welcome	**Marc Wonders** NGFP/PNNL
1205	Challenges and Opportunities for AI in Nuclear Proliferation Detection	**Angie Sheffield** NNSA
1215	Domain Aware AI: There and Back Again	**Kary Myers** LANL
1300	**Break**	



Sparse Data and Rare Events

Session Chair: **Becky Olinger** | DTRA
Note Taker: **Tom Grimes** | PNNL

Session 3: Continuing
(Same as Day 2 Plenary Session)

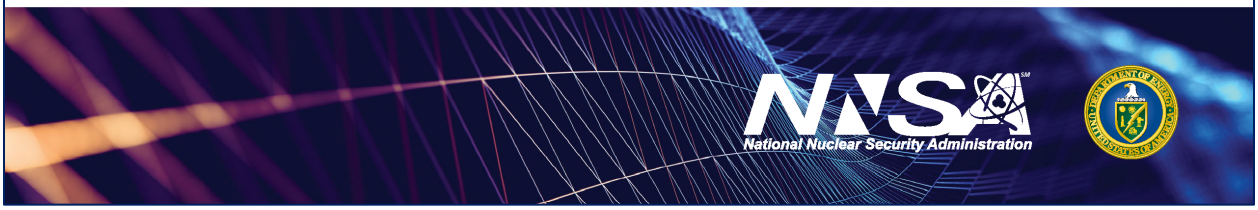
1310	Constraining Data Driven Models for Detection of Sparse Temporally Correlated Events	**Garrison Flynn** LANL
1330	Domain-Informed Assessment of Nuclear Reactor Operations	**Tom Reichardt** SNL
1350	Persistent DyNAMICS: Remote Sensing Based on Domain-Informed Analytics	**Thomas Kulp** SNL **Sidharth Manay** LLNL
1410	**Break**	
1420	Node and Region Importance for Classifying Nuclear Operations using Multisensor Arrays	**Jake Tibbetts** UC Berkeley
1440	One-shot Target Detection via Physics-Informed Training	**Natalie Klein** LANL
1500	**Closing Session 3** (followed by break)	**Becky Olinger** DTRA

Robust Deployment and Decision Support

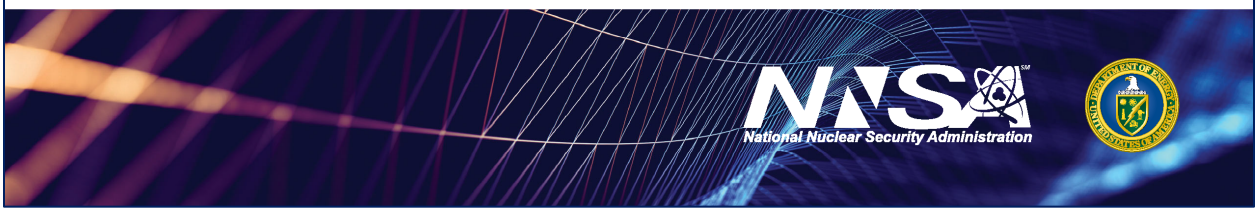
Session Chair: **Stefan Hau-Riege** | LLNL
Note Taker: **Maria Glenski** | PNNL

Session 4

1310	Annotation Transfer for Prediction of Industrial Operations	**Erik Skau** LANL
1330	Automated Synthesis of Soft Labels using Neural Stochastic Differential Equations and Attribution Based Confidence	**Sumit Kumar Jha** University of Texas at San Antonio
1350	Robustness in the Wild using Domain-Aware Surrogate Functions	**Jay Thiagarajan** LLNL
1410	**Break**	
1420	A Computational Framework for Deterrence Assessment Analyses	**Michelle Quirk** NA-114
1440	On-the-Fly Robustness in the Wild via Data-Driven Generative Priors	**Rushil Anirudh** LLNL
1500	**Closing Session 4** (followed by break)	**Stefan Hau-Riege** LLNL



1510	**Panel: Requirements and Opportunities for Domain Aware Methods in Proliferation Detection**	
Track 1: Continuing (Same as Day 2 Plenary Session)		
Moderator: **Tammie Borders** | NNSA/INL		
Panelists: **Emma Hague** | DOE-IN **Kary Myers** | LANL		**Angie Sheffield** | NNSA **Becky Olinger** | DTRA
1550	Final Thoughts	**Angie Sheffield** NNSA



Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354
1-888-375-PNNL (7665)

www.pnnl.gov