Proceedings of the INMM & ESARDA Joint Annual Meeting
August 23-26 & August 30-September 1, 2021

DATA SCIENCE MEETS NUCLEAR – WHAT DATA ANALYTICS, COMPUTATIONAL INTELLIGENCE AND MACHINE LEARNING CAN CONTRIBUTE TO NUCLEAR WASTE MANAGEMENT AND NUCLEAR VERIFICATION

Lisa Beumer, Irmgard Niemeyer
Forschungszentrum Jülich GmbH, Germany

ABSTRACT

Data science is multidisciplinary field that deals with the study of all aspects of data right from its generation to processing to converting it into valuable knowledge source. While data science has a wide range of applications, to what extent have new data science methods made their way into research related to nuclear waste management and nuclear verification? And which further research questions in these fields would particularly benefit from the use of new data science methods? In this line, this paper has two objectives: First, to highlight the state-of-the-art of data science in nuclear waste management and nuclear verification. Second, to discuss the potential use of data science. Ideas for data science in nuclear waste management include, e.g., i) facilitating integration, analytics and visualization of data in the comparative selection process for a geological repository site, ii) creating a virtual geological repository system, iii) geological repository monitoring over its life cycle phases. In nuclear verification, data science can make a significant contribution to i) unattended monitoring by using, e.g., seals/tags, surveillance (optical, 2D/3D laser, gamma, etc.), radiation measurements, etc.; ii) perimeter monitoring through surveillance (optical, gamma, thermal, etc., radiation measurements, etc.), and iii) wide area monitoring using satellite imagery, geophysical monitoring, environmental sampling, etc.

INTRODUCTION

Data Science is an interdisciplinary field of science that enables the extraction of insights, patterns, and conclusions from both structured and unstructured data through the application of scientifically sound methods, processes, algorithms, and systems. Data science is concerned with how very large amounts of data ('big data') can be collected, processed, prepared, and analyzed. The focus of Data science is not on the data itself, but on how it is processed, prepared, and analyzed. Data science is concerned with purpose-oriented data analysis and the systematic generation of decision-making aids and bases in order to be able to achieve competitive advantages. [1-3]

Big data refers to data volumes that are too large, too complex, too fast-moving, or too weakly structured to be analyzed using manual and conventional methods of data processing. Big data is often explained by the 4 to 6 V's: Large *volume* of data, the speed with which the data is generated and transferred (*velocity*), the range of data types and sources (*variety*), and the authenticity of data (*veracity*) are the distinguishing features of this data. This definition is often expanded to include added business *value* and the assurance of data quality (*validity*). [1-3]

Data science can be divided into four core areas: 1) *Data engineering* comprises all methods and processes required for the storage, access, and traceability of data. 2) *Data analytics* deals with data analysis. 3) *Data prediction* deals with the prediction of topics and situations based on empirical knowledge. 4) *Machine learning* (ML) is a cross-cutting area to the other three areas and stands for the development of algorithms that learn from data (experiential knowledge), thereby recognizing patterns, generating models and, based on that predict topics and situations based on these patterns and models. [4]

Other terms to be mentioned in this context are *computational intelligence* (CI) [5] and *artificial intelligence* (AI). CI is the theory, design, application, and development of biologically and linguistically motivated computational paradigms. Traditionally, the three main pillars of CI are neural networks, fuzzy systems, and evolutionary algorithms. However, over time, many nature-inspired computational paradigms have evolved. AI plays an important role in the development of successful intelligent systems, including games and cognitive development systems. In recent years, there has been an explosion of research on *deep learning*, especially deep convolutional neural networks. Today, deep learning has become the core method for artificial intelligence. In fact, some of the most successful AI systems are based on CI.

While data science has a wide range of applications, to what extent have new data science methods made their way into research related to nuclear waste management and nuclear verification? And which further research questions in these fields would particularly benefit from the use of new data science methods? In this line, this paper has two objectives: First, to highlight the state-of-the-art of data science in nuclear waste management and nuclear verification. Second, to discuss the potential use of data science.

DATA SCIENCE IN NUCLEAR WASTE MANAGEMENT

In radioactive waste management, waste separation into, for example, combustible and non-combustible takes place, which increases operational workflow and causes manual errors. Deep learning methods can be used to automate waste categorization to maximize categorization efficiency. Studies such as [6] demonstrate that a waste detection system based on a Residual Network (ResNet), trained on video data recorded from a sorting facility is able to detect four typical radioactive wastes (vinyl, rubber, cotton, and paper) with no object with hands (no object) and without hands (empty) with an accuracy of 99.67% on a test set.

High-level nuclear waste is vitrified in borosilicate glass and then stored in an underground repository. To ensure proper storage, tests are performed with different conditions and glass compositions. Based on these data, machine learning methods can be applied to determine the static and dynamic class leaching behavior of radioactive waste glasses, and to predict missing data and time forecasts. The bagged random forest method was used, which can predict very accurate predictions for static leaching even when the glass composition or initial condition variables such as glass density are not specified. [7]

The multidisciplinary collaborative repository research project iCross: Integrity of Radioactive Waste Repository Systems - Cross-Scale Systems Understanding and Systems Analysis addresses the issue of safe and long-lived nuclear waste disposal [8]. The cross-scale research concept of the project includes novel approaches using high-end research infrastructures. For example, neural networks are used to address the challenges of multiscale-multiphysics-modeling of geochemical processes, which is challenging due to the complexity of the chemistry, the heterogeneous microstructure, and the different spatial and temporal scales of the processes [9]. The neural network gave excellent results compared to exact solution methods for the calculation of the chemical speciation, with better computational efficiency and memory demands. NN are thus promising possibilities for large-scale 3D reactive transport simulations of complex systems.

Deep Learning also expands study of nuclear waste remediation. The monitoring of nuclear waste at a decommissioned nuclear power plant or waste repository requires a limited number of drilling wells in order to obtain data on geological properties and groundwater using sensors [5]. However, the limited number of boreholes causes a lack of sensor data, making it difficult to determine flow equations of especially heterogeneous subsurface environments. However, when physics, expert knowledge and machine learning methods are combined, it is possible to create a virtual representation of such locations based on a small amount of data and to estimate parameters and quantify uncertainty in the subsurface flow. As [10] revealed, physics informed GANs, which allow a combination of stochastic computational models and observational data, are suitable for this purpose. GANs have the advantage of being an unsupervised learning method, learning density distributions of data, and even generating data that is similar to real data. This allows them to have many different uses, such as simulation problems, in the real world. For synthetic data, it has been shown that a physics-constrained GAN architecture can generate spatial fields consistent with our knowledge of physics [10]. However, adding real sensor data and real conditions is promising.

DATA SCIENCE IN NUCLEAR VERIFICATION

The IAEA's basic verification measure is nuclear material accountancy (MA), in which inspectors perform independent measurements such as counting fuel assemblies and measuring as well as verifying radiological signatures using non-destructive analysis (NDA) techniques. The monitoring of radiological signatures is of great importance not only for the control of potential contamination and the associated health risks but also for the detection of illegal nuclear material and the discovery of unwarranted changes in storage and production.

Today a broad range of NDA instruments are available, the IAEA has already authorized over 100 different types for inspections. The most widely used NDA instruments rely on the detection of nuclear radiation. Several studies have already shown how neural networks can favor NDA techniques. Single-photon emission tomography (SPECT) techniques are used for safety monitoring and quantitative verification of the amount of nuclear material contained in a spent fuel assembly. Due to the high Z materials contained in a spent fuel assembly, the quality of the projection image is quite low. However, deep learning methods, such as the convolutional autoencoder (CAE) model-based de-noised image reconstruction technique, enable de-noised image reconstruction and faster image acquisition, resulting in an increased overall fuel assembly inspection time [11]. In addition to the application of Compton cameras for SPECT, advanced radiation imaging with rotation scatter masks (RSM), an inexpensive method with a large field-of-view for identifying the direction of a gamma-ray emitting sources, have proven advantages in inspecting containers or transport casks, imaging large contaminated areas and nuclear waste characterization. To make the single detector imaging system RSM a real imaging device, image reconstruction algorithms such as convolutional neural networks can be applied. With these methods the problem of source images that cannot be reconstructed accurately can be solved, since application-specific expected source distributions can be added for training the model [12]. The potential of using CNNs was demonstrated by [12] using a simple, fast network capable of reconstructing simplified RSM data very accurately. The ongoing developments in deep learning promise the extension of the simple network to a more complex model used for real RSM data and therefore a promising use of RSM with respect to radiation material monitoring and localization. Radiation detectors can also be used to characterize gamma ray emissions from a sample of interest. Studies indicate that the machine learning algorithms k-nearest neighbor (kNN) and support vector machines (SVMs) are able to detect and localize removed materials from a given radioactive sample of interest „even when gamma ray emissions are different than modeled or expected“ [13].

Containment and surveillance (C/S) techniques and unattended monitoring supplement the nuclear material accounting verification measure by providing additional information to detect undeclared access to nuclear material or its movement. A variety of C/S techniques are applied, primarily optical surveillance. In addition to images, video recordings allow a deeper understanding of the situation, as the image sequences provide further information about actions. Real-time video processing is an essential technology in surveillance systems. To support the visual inspection of surveillance videos in terms of locating and identifying objects and activities of interest, deep machine learning and transfer learning approaches can be used.

Besides original slow CNN approaches, such as sliding window or two-step methods (prediction followed by classification), YOLO, a single shot object detector for multiple object detection, has been established. Both localization and classification are performed by a single CNN and an image only needs to be viewed once. Because YOLO is an efficient CNN, it allows real-time execution on an inspector's laptop. Other constraints, such as differences between nuclear facilities and objects of interest and the limited amount of training data, can be overcome by using YOLO and transfer learning. For example, deep learning applied to image sequences of surveillance videos can help to properly detect objects of interest such as objects of a waste repackaging facility and objects at simulated spend fuel pools. Furthermore, it can mark them with bounding boxes which could significantly increase the efficiency of the review process of an inspector. But deep learning methods are not only very promising in the area of object detection. Autoencoders, for example, can be used to detect abnormal scenes in a video sequence in real time and to trigger an alarm. [14]

Environmental sampling is also part of the IAEA's verification measures, since the absence of even minute traces of a specific nuclear material can guarantee that no activities involving the material occurred in the area where environmental samples were taken. For example, automatic detection and characterization of anomalous radiological signatures by means of persistent radiation detection and mobile detection systems is of great importance. Due to numerous background fluctuations, it is important to discriminate between illicit and innocent sources of radioactive material transfer. To extract robust features from noisy, high-dimensional datasets, denoising autoencoders have proven to be very useful. Autoencoders are neural networks that compress the input information and use this information for a correct reconstruction. This principle is also exploited by the Autoencoder Radiation Detection Anomaly (ARAD) model, which learns statistical regularities and key components of gamma-ray spectra obtained in dynamic background radiation environments [15]. The reconstruction accuracy can then be used to detect radiation anomalies. The advantage of learning techniques in this case is that the model is able to accurately reconstruct the input spectrum without adding detailed detector information as it learns it from the training data.

Like environmental sampling, satellite imagery is also part of remote monitoring. There are many applications of satellite imagery in the field of nuclear verification because they provide analysts with clear insights into nuclear facilities and nuclear activities worldwide. For the IAEA, commercial satellite imagery has become " a very important information source, especially regarding places where the IAEA does not have access." [16]. This allows the accuracy and completeness of information supplied by States to be verified, for change detection and activity monitoring at nuclear fuel related sited and the identification of undeclared sites. Various studies have shown that the use of satellite imagery for safeguards purposes can provide valuable information. [17-21]. Since there are an increasing number of satellite operators, the data volume is rising rapidly. However, the deluge of data, as well as the variety of associated metadata, entails further automation of pre-processing and analysis. Deep learning has also proven its strength in the context of earth observation. Thanks to the capabilities of CNNs, many change detection, image segmentation and object detection approaches have already been developed with great success. For example Smartt et al. [22] developed a machine learning algorithm based on CNNs to automatically detect, count and label vehicles to assist human analysts.

Similar to satellite imagery, ground-based photographs provide safety-related information such as the operational status of a nuclear facility, traffic patterns of heavy vehicles, or construction activities within a nuclear facility. Photo contingents, such as those taken during an on-site inspection by an IAEA safety inspector, can be usefully supplemented by information from open sources, such as news images or images in social media feeds [23]. The increasing availability of open-source images on the Internet makes manual verification difficult. As [23] have shown, machine learning based automated methods for image search and classification or prioritization can be applied to identify security relevant features in image collections, making open source data a useful resource for security monitoring. Transfer learning with CNNs demonstrated great potential for image classification of data from a cooling tower dataset created on search terms related to the nuclear fuel cycle from the website Flickr. Regarding the two-class classification problem (image contains hyperbolic cooling tower or not), the approach achieved 90.4% accuracy on a test set. Open-source data and social media data thus provide a source of information for international verification of nuclear safeguards when used in conjunction with inspection data, government declarations, or other sources of information available to the IAEA.

The complexity and diversity of facilities containing safeguarded nuclear material require a correspondingly diverse set of verification techniques and equipment. If the different data types, image, text or video, are not considered separately, a powerful system of information is created. The amount of data is huge, and you do not know in which entry you will find which information and you cannot prioritize the data within an acceptable time frame. To support the analyst, there are large-scale multimodal retrieval systems, which include open-source information, technology, and news data in the context of the nuclear fuel cycle. The difficulty is to design them in such a way that they are easy to search and filter, and that data can be prioritized based on appropriate queries. Since data science has already enabled effective analyses of unstructured, heterogeneous, and complex data, it is also a promising solution in this case, as [24] showed. The development of a large-scale system of deep neural networks (DNNs) enabled to map and retrieve multimodal data proximal in a multimodal feature space. The DNN algorithm applied to open-source science, technology, and intelligence-based multimodal datasets enabled the identification of indicators of nuclear proliferation capabilities and activities.

Autonomous systems, AI and ML could significantly impact IAEA safeguards verification activities in the next decade. [25] Based on an inventory consisting of 14 AI-based methods that could potentially be applied to improve a variety of safeguards challenges, two methods were selected to explore their application to the safeguards operating environment: 1) unsupervised machine learning (ML) One-Class Support Vector Machine (OCSVM) for analysis of large amounts of unattended monitoring data; and 2) Convolutional Neural Network. The authors concluded that autonomous systems, AI and ML provide important opportunities to improve the effectiveness and efficiency of IAEA safeguards. They have the potential to reduce the time and resources needed to implement safeguards measures and can help provide important insights, such detecting patterns or anomalies that would not otherwise have been observed by human inspectors or analysts. However, AI and ML present non-trivial challenges and risks for practical implementation. In particular, the way how systems learn, decide, and act must be carefully understood to overcome any barriers to deployment. This requires robust testing and evaluation to increase trust, transparency, and acceptance of AI in nuclear waste management and nuclear verification.

CONCLUSION

Based on the current state-of-the-art of data science in nuclear waste management and nuclear verification, further complex data analysis problems in these areas that may potentially be mitigated or solved by data science are being identified. At the same time, data science methods and techniques that were established in non-nuclear sectors are being studied about their potential suitability for nuclear waste management and nuclear verification. Following the analysis and prioritization of needs and objectives of promoting data science in the nuclear domain, specific data science methods and techniques will need to be further developed and evaluated.

Ideas for data science in nuclear waste management include, e.g., i) facilitating integration, analytics and visualization of data in the comparative selection process for a geological repository site, ii) creating a virtual geological repository system, iii) geological repository monitoring over its life cycle phases. In nuclear verification, data science can make a significant contribution to i) unattended monitoring by using, e.g., seals/tags, surveillance (optical, 2D/3D laser, gamma, etc.), radiation measurements, etc.; ii) perimeter monitoring through surveillance (optical, gamma, thermal, etc., radiation measurements, etc.), and iii) wide area monitoring using satellite imagery, geophysical monitoring, environmental sampling, etc.

REFERENCES

[1] U. Qamar, U., Raza, M.S. (2020). *Data Science Concepts and Techniques with Applications*. Springer, Singapore, 2020, https://doi.org/10.1007/978-981-15-6133-7

[2] Ley, C. and Bordas, S.P.A. (2018). What makes Data Science different? A discussion involving Statistics 2.0 and Computational Sciences. *International Journal of Data Science and Analytics* Volume 6, pp. 167–175, 2018. https://doi.org/10.1007/s41060-017-0090-x.

[3] Cao, L. (2018). *Data Science Thinking, The Next Scientific, Technological and Economic Revolution*. Springer, Cham, 2018, https://doi.org/10.1007/978-3-319-95092-1

[4] acatech (Hrsg.) (2018). Pilotphase Nationales Kompetenz-Monitoring (NKM): Bericht: Data Science. Auswahl, Beschreibung, Bewertung und Messung der Schlüsselkompetenzen für das Technologiefeld Data Science (acatech DISKUSSION)

[5] Bezdek, J. C. (2014). (Computational) Intelligence: What's in a Name? *IEEE Systems, Man, and Cybernetics Magazine*, Volume 2. pp. 4-14, 2016, 10.1109/MSMC.2016.2558778.

[6] Jeong-Guk Kim, J.-G., Jang, S.-C., Kang, I.-S., Lee, D.-J., Lee, J.-W., Park, H.-S. (2020). A study on object recognition using deep learning for optimizing categorization of radioactive waste. *Progress in Nuclear Energy*, Volume 130. doi: 10.1016/j.pnucene.2020.103528.

[7] Lillington, J., Goût, T., Harrison, M., Farnan, I. (2020). Predicting radioactive waste glass dissolution with machine learning. *Journal of Non-Crystalline Solids*, Volume 533. doi:10.1016/j.jnoncrysol.2019.119852.

[8] New project iCross - taking a holistic view on safe repository systems: Integrity of nuclear waste repository systems – Cross-scale system understanding and analysis. URL: https://www.fz-juelich.de/iek/iek-6/EN/News/More-News/Documents/2018-09-06%20iCross.html

[9] Prasianakis, N. I., Haller, R., Mahrous, M., Poonoosamy, J., Pfingsten, W., Churakov, S. V. (2020). Neural network based process coupling and parameter upscaling in reactive transport simulations. *Geochimica et Cosmochimica Acta*, Volume 291. pp. 126-143. doi: 10.1016/j.gca.2020.07.019

[10] Yang, L. et al. (2019). Highly-Scalable, Physics-Informed GANs for Learning Solutions of Stochastic PDEs. Proc. 2019 IEEE/ACM Third Workshop on Deep Learning on Supercomputers (DLS). doi: 10.1109/dls49591.2019.00006

[11] Choi, S.H., Choi, H.J., Min, C.H.,Chung, Y.H., Ahn, JJ. (2021). Development of de-noised image reconstruction technique using Convolutional AutoEncoder for fast monitoring of fuel assemblies. *Nuclear Engineering and Technology*, Volume 53, Issue 3. pp. 888-893. doi: 10.1016/j.net.2020.08.020.

[12] Olesen, R. J., Holland, D. E., Brubaker, E., Cole, J., and Bevins, J. E. (2019). Advanced Radiation Imaging Algorithms with Rotating Scatter Masks. Proc. INMM 2019 Annual Meeting.

[13] Durbin, M., Lintereur, A. (2020). Implementation of machine learning algorithms for detecting missing radioactive material. *J Radioanal Nucl Chem*, Volume 324. pp. 1455–1461. doi: 10.1007/s10967-020-07188-4

[14] Cui, Y. (2018). Using Deep Machine Learning to Conduct Object-Based Identification and Motion Detection on Safeguards Video Surveillance. Symposium on International Safeguards: Building Future Safeguards Capabilities, Vienna.

[15] Ghawaly Jr, J. M., Archer, A., Hall. H. (2019). An Unsupervised Radiation Anomaly Detection Algorithm Using a Deep Neural Autoencoder. Proc. INMM Annual Meeting.

[16] Quevenco, R. (2016). Completing the picture: using satellite imagery to enhance IAEA safeguards capabilities. *IAEA Bulletin* 2016 57-2. pp. 24-25.

[17] Jo S.Y., Shin, J.S. (2010). Status and Application of Spatial Information Technology as a Nuclear Verification Tool. Transactions of the Korean Nuclear Society Spring Meeting Pyeongchang, Korea.

[18] Leslie, R., Riggs, P., Bragin, V. Truong, Q.S.B., Borstad, G., Neville, R., Staenz, K. (2002). Satellite imagery for safeguards purposes: utility of panchromatic and multispectral imagery for verification of remote uranium mines. Proc. INMM Annual Meeting.

[19] Zhang, H. (2001). Strengthening IAEA safeguards using high-resolution commercial imagery. Proc. IAEA Symposium on International Safeguards: Verification and Nuclear Material Security, Vienna.

[20] Patton, T., Lewis, J., Hanham, M., Dill, C., Vaccaro, L. (2016). Emerging Satellites for Non-Proliferation and Disarmament Verification. Vienna Center for Disarmament and Non-Proliferation (VCDNP).

[21] Glaser, A., Niemeyer, I. (2021). Nuclear Monitoring and Verification Without Onsite Access. In: Toward Nuclear Disarmament: Building up Transparency and Verification. Ed. by M. Göttsche and A. Glaser. German Federal Foreign Office, 2021, pp. 86–115. doi: 10.18154/RWTH-2021-04949.

[22] Smartt, H.A., Lee, D.J., LaCasse, C.F., Bobeck, J., Bornhorst, H., Puccioni, A. (2019). Detection via Persistence: Leveraging Commercial Imagery from Small Satellites. Proc. INMM Annual Meeting.

[23] Shead, T.M., Gastelum, Z.N. (2017). Preliminary Application of Neural Networks to Support Assessment of Ground-Based Imagery for International Safeguards Analysis. Proc. INMM Annual Meeting.

[24] Feldman, Y., Arno, M., Carrano, C., Ng, B., and Chen, B (2018). Toward a Multimodal-Deep Learning Retrieval System for Monitoring Nuclear Proliferation Activities. J*ournal of Nuclear Materials Management. Journal*, Volume: 46; Journal Issue: 3; Journal ID: ISSN 0893-6188.

[25] Haddal, R, Hayden, N., Frazar, S. (2018). Autonomous systems, artificial intelligence and safeguards. Proc. IAEA Safeguards Symposium: Building Future Safeguards Capabilities, Vienna.