MACHINE LEARNING IN NUCLEAR SCIENCE AND ENGINEERING APPLICATIONS[1]

ÓSCAR CABELLOS
Full Professor in Nuclear Engineering
UNIVERSIDAD POLITÉCNICA DE MADRID
oscar.cabellos@upm.es

DENISE NEUDECKER
Staff Scientist
LOS ALAMOS NATIONAL LABORATORY
dneudecker@lanl.gov

BACKGROUND

Many generations of Spanish nuclear engineers were educated in quantum mechanics by Professor Guillermo Velarde[2] at the Polytechnical University of Madrid. Professor Velarde taught the fundamental principles of modern quantum mechanics jointly with the history of those pioneers of quantum physics. He highlighted the importance of the following skills for nuclear engineers in a technical EU report saying: "*future engineers, besides English and a programming language (Fortran), should learn, obviously, Quantum Mechanics*". Nowadays, we could rephrase this sentence adding "*Python and Machine Learning & Artificial Intelligence*".

Strictly speaking, Machine Learning (ML) is an application of Artificial Intelligence (AI). AI is a broader concept which deals with the use of computers to imitate the cognitive functions of humans. ML provides the tools to build a "mathematical" model using "training" data and use these data to learn and improve applications being as varied as large on-line stores analyzing their purchase data for patterns to nuclear engineers assessing the safety of reactors. There are multiple definitions of ML, the simplest one is given by Stanford University in his on-line ML course: "*ML is the science of getting computers to act without being explicitly programmed*".

There are two basic approaches within ML, the supervised learning and unsupervised learning **[J. Delua, 2021]**. The supervised learning is an approach defined by the use of labelled data which are designed to train or "supervise" algorithms. This supervised learning is suitable for classification (e.g., linear classifiers, support vector machines, decision trees and random forest) and regression problems (e.g., linear regression, logistic regression and polynomial regression). The unsupervised learning uses algorithms to analyse and cluster unlabelled data sets and unsuitable for regression type problems. Unsupervised learning is suitable for clustering phenomena (e.g. k-means, k-medois, DBSCAN,…), dimensionality reduction and association rule learning.

ML has been around since the 1950's. However, only recently there has been a resurgence of interest in ML with countless ML and AI applications transforming technology; an example is the pedestrian-detection vision systems of self-driving vehicles. Currently, ML/AI are tackling issues across different science disciplines such as biology, medicine, health care and climate science. Among the main factor factors explaining the ML/AI resurgence are:

- The emergence of "Big-Data" has revolutionized the traditional data processing capabilities. An enormous amount of data is being generated, with an unparalleled velocity and variety, that can now be used to feed ML algorithms to address engineering problems at a scale never tackled before.
- The sudden availability of ``Open-and-Public Data'' underscores transparency in the way that data-driven approaches are used, and therefore their accountability and trustworthiness.
- The Open-Science and Open-Source tools (e.g. scikit-learn **[Scikit-Learn, 2011]**,…, KNIME Analytics Platform **[KNIME, 2009]** (Figure 1), etc…) which constitute a temptingly low barrier to entry for building ones own first machine learning models.
- There are different considerations for the selection of the right ML technique: linearity, training time, accuracy, interpretability, etc. However, the type of data at hand is one major factor that will determine which ML technique is suitable for a successful application.

[1] LA-UR-21-28449

[2] Guillermo Velarde passed away on January 14, 2018

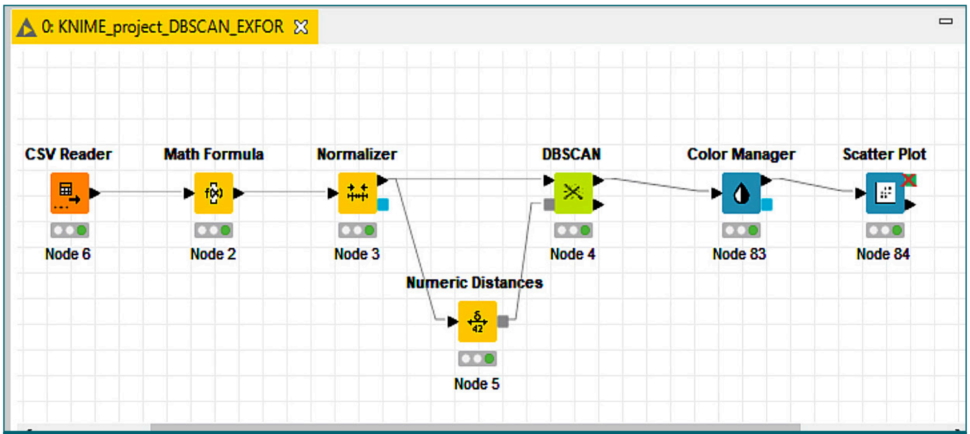

Figure 1. The user interface of KNIME[3]. Example of KNIME flowchart using ML DBSCAN solver to identify potential outliers in EXFOR [N. Otsuka, 2014] database.

- Powerful computer hardware in a well-developed (robust and explainable) ML/AI computer science which allows rapid, accurate and trusted predictions.
- Stimulating and trans-disciplinary collaboration between researchers and experts in ML/AI and applications that accelerates advancements both in developing new algorithms and the desired application.
- The increased visibility of outstanding ML/AI research and achievements is attracting young talent to this scientific area furthering ML/AI.

A REVIEW OF ML ACTIVITIES TO THE FIELD OF NUCLEAR SCIENCE AND ENGINEERING

A first comprehensive catalogue of applications that involve the use of "expert systems" (neural networks, fuzzy logic, etc.) applied to the design, management and operation of research, test and commercial power reactors was compiled in a monograph published in 1989 **[J. A. Bernard, T. Washio, 1989]**. Later, in 1990s and 2000s there have been published review articles, **[R. E. Uhrig, L. H. Tsoukalas, 1999]** among them, and special issues in scientific journals **[D. Ruan et al, 2005]** which give a good overview of the potential use of ML and AI on nuclear engineering applications. More recently, a thorough and comprehensive review on the status of research and development of learning-based approaches in nuclear science and engineering applications can be found in **[M. Gómez-Fernández, 2020]**. This paper summarizes the recent interest of the nuclear industries using ML techniques to improve safety, reliability, and availability of assets.

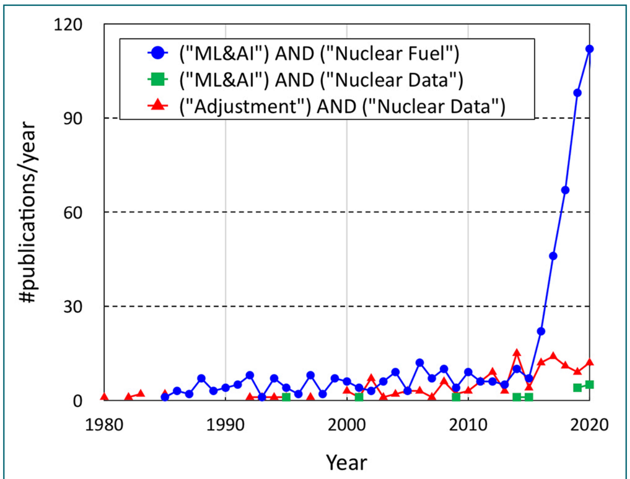

Figure 2. The number of papers containing terms "ML/AI" and "Nuclear Fuel for in-core management" are plotted in blue. For the topic "nuclear data evaluation", the number of papers shown are for conventional adjustment and ML/AI techniques.

For the nuclear engineering applications, first genetic algorithms for optimization of in-core fuel management design were reported in 1991 **[Y. Tahara, 1991]**. In view of the limitations of using a trial-and-error conventional way of generating an appropriate fuel loading pattern, Tahara's paper remarked on the usability of the AI techniques: "*The fuel shuffling problem is considered to be suitable for artificial intelligence techniques because of its heuristic nature*". Figure 2 shows the number of publications during the last four decades containing "ML/AI" and "Nuclear Fuel for in-core management" from Web of Science. It can be seen that more than 100 publications were published in 2020; that is more than ten times the average publications per year in the last decades! This distinct upsurge in publications showcases that ML/AI approaches are readily adopted to aid in fuel loading pattern design—both in conventional and new reactor and fuel cycles designs.

Contrary to that, Figure 2 shows also that the field of nuclear data and adjustment is slow to adopt "ML/AI" as indicated by the low number of publications containing the terms "ML/AI" or "adjustment" and "nuclear data evaluation". However, some recent ML/AI initiatives and movements have been initiated that may change the scenario for this community.

In this context, to foster ML/AI activities in Nuclear Science and Engineering (NS&E) applications the NEA/OECD Nuclear Science Committee has added in his new 2021-2026 mandate the new technical activity topic on "*Big Data and Machine Learning's knowledge in nuclear science*". This international collaborative framework will contribute to the identification of promising new algorithms to support the safe, reliable and economic operation of current and next-generation nuclear systems.

The following are examples of applications to show the potential use of ML/AI in NS&E, more detailed information can be found in **[M. Gómez-Fernández, 2020]**:

- Reactor-oriented applications
  - ML/AI can aid in the field of reactor safety and controls, that ensure plants health, by: (a) detecting anomalies in large amount of data that would otherwise be hard to

[3] https://www.knime.com/

Technical	Economical
• Reduce radiation exposure to personnel • Enhance equipment reliability • Avoid actuation of safety systems • Assist with correct and timely decision making • Enhance safety margins	• Optimize the maintenance schedule • Improve plant availability • Avoid escalation of minor problems into major events • Support power uprate and life extension

Table 1. List of potential objectives using ML/AI in reactor-oriented applications

search for conclusive patterns, (b) indicating those input variables most predictive of the state of the plant, (c) diagnostics by transient identification, (d) monitoring of performance and efficiency, (e) on-line estimation of safety margins and (f) flow regime identification.

- Monitoring and diagnostics with ML-augmented capabilities that recognize rapidly patterns in time-records as a function of several variables, or undertake input-to-output modelling of various phenomena pertaining to the safe operation of a reactor, among them e.g., vibration monitoring and analysis, sensor/instrumentation surveillance and calibration verification, validations of inferential sensing and virtual instrumentation as well as any phenomena recorded by equipment monitoring operability.
- ML/ AI can be applied to optimize in-core nuclear fuel management in reloading, core shuffling, emulating fuel configuration calculations, core parameter prediction and criticality optimization. These techniques can also be connected to improved use of robotics and controls.

• Detection-oriented application

- Gamma and neutron spectroscopy can leverage ML/AI to improve identification of particles via classification algorithms used for ensuring optimal operation; but ML/AI for is also used for calibrations, and automation of activity prediction which requires reduced measurement time. ML also allows solving rather challenging inverse problems such as spectrum or isotope unfolding.
- ML/AI methods can be applied to radioactive environmental monitoring, and for nuclear security applications.

• Nuclear data application

- ML/ AI can support compilation as well as analysis of differential and integral experimental data used for evaluation and validation. For instance, natural-language processing could be used in searching through many journal articles for relevant information on experiments. On the other hand, ML methods can be put to task to identify outliers and reasons for those. Both tasks are labor-intensive and ML/AI would significantly cut down on human time needed.
- ML/AI has been successfully applied to identify nuclear data needs and critical issues across the nuclear data pipeline. Their applications range from finding the best parameter sets for nuclear reaction codes to describe a reaction to highlighting the need of new evaluations of nuclear data.
- ML/AI can be used to enhancing the processing and encoding of nuclear data for real applications such as, e.g., selecting adequate group structures of nuclear data.

Table 1 shows examples of promising ML/AI applications with both technical and economical benefits in reactor-oriented applications **[M. Gómez-Fernández, 2020]**.

RECENT ACTIVITIES AND EVENTS OF ML/AI IN NS&E

ML/AI innovation in NS&E applications is involving a broad range of actors of the nuclear community, such as R&D performers (public research organisations, industrial research organisations, universities), utilities and industries (fuel companies, nuclear vendors) and safety authority bodies. A review of recent domestic events will help to highlight the importance of ML/AI activities for the Spanish "actors":

1) In 2021, the Polytechnical University of Madrid organized a series of Seminars on Machine Learning within the Master of Nuclear Science and Technology program to promote ML/AI activities in our graduate and master students. These Seminars were scheduled within the INGENIA[4] course devoted to the topic "Nuclear Reactor Design and Simulations". The speakers in these Seminars gave a good overview of ML applications in different areas of NS&E[5,6,7].
2) The Industry and Academia (CEIDEN[8]-UPM) jointly organized the first Workshop on "Machine Learning in Nuclear Science and Technology Applications" held virtually on May 27, 2021. This Workshop served the users and developers in the field of nuclear data and reactor physics to discuss and exchange their expertise using different Machine Learning techniques. A total of 14 technical presentations were scheduled. There were more than 150 virtual attendants from 19 different countries and 55 institutions.
3) The Nuclear Safety Authorities are giving attention to the current state and future trends of ML/AI tools in the various phases of nuclear power reactors, from operational experience to plant management in predictive reliability and safety assessment. This is the first time that the Spanish Nuclear Safety Council (CSN) has funded a project explicitly in matters related to "Machine Learning, Big Data and Artificial Intelligence Applications in Nuclear Safety" (R&D Subventions, BOE 132, Sec. III, pp68169-68176, June 3, 2021).

There are also international events organized/sponsored by the safety and regulatory bodies that will allow accelerate the implementation of ML/AI in NS&E in a more safe and effective way:

[4] http://www.etsii.upm.es/estudios/masteres/ingenia/14/index.es.htm

Seminars organized within the Master of Nuclear Science and Technology, Universidad Politécnica de Madrid, course 2020-2021:

[5] Mario E. Gómez-Fernández, "Understanding the Application and Benefits of Learning-based Methods in Nuclear Science and Engineering Seminar ", October 19,2020.

[6] Denise Neudecker, "Using Machine Learning Algorithms for Large-scale Nuclear-data Validation", October 26, 2020.

[7] Tadahiro Kin, "Machine Learning in Radiation Metrology: Applications of gamma Ray and Cosmic Ray Measurement", February 22, 2021.

[8] CEIDEN - The Spanish Nuclear Fission Energy Technology Platform (https://ceiden.com/en/)

- The U.S Office of Nuclear Regulatory sponsored the technical session on “Analytics, Machine Learning, and Artificial Intelligence for Nuclear Power Plant Activities” on March 11, 2021. Shortly after, the U.S. NRC launched an action to know the potential “Role of Artificial Intelligence Tools in U.S. Commercial Nuclear Power Operations”.
- The IAEA is promoting meetings on ML/AI to foster international cooperation in this specialised area: “The Future of Atoms: Artificial Intelligence for Nuclear Applications”, 64th IAEA General Conference, 23 Sep 2020. “Technical Meeting on Artificial Intelligence for Nuclear Technology and Applications”, IAEA, 25 – 29 Oct 2021.

ML ACTIVITIES IN THE NUCLEAR DATA COMMUNITY

In the last years, many references are found to foster ML/AI methods in the pipeline of nuclear data. This effort will increase soon the number of scientific publications of ML/AI in the nuclear data science (Figure 2).

- The first explicit reference is given by the *“Collaborative International Evaluated Library Organisation (CIELO)”* project working under the auspices of the NEA/WPEC. In the summary report, they concluded: *“With the rapid advancement of Big Data and Machine Learning techniques, the opportunity to apply them to a fully encapsulated nuclear data system would be one of the most natural directions to pursue.”* **[CIELO, 2019]**.
- Recently, the Coordination Group of the Joint Evaluated Fission and Fusion (JEFF) Data Evaluation Project pointed out how *“Machine Learning applications could be useful in automatically flagging outliers, if implemented systematically (by R. Capote (IAEA/NDS)”*. **[JEF-DOC-2026, 2020]**.
- There is a consensus in this community that “AI/ML approaches have tremendous potential to address critical short-term and long-term needs across the nuclear data pipeline.” **[V. Sobes, 2020]**, and more effort is needed, “*to explore to which extent methods of machine learning and statistics can help us to get a better understanding of nuclear data and identify potential directions of development.*” **[G. Schnabel, 2020]**.

A variety of ML techniques have been applied recently to solve open issues in nuclear physics and nuclear data activities. A detailed summary of applications can be found in **[D. Neudecker, 2020]** and **[D. Neudecker, 2021]**. Here, we briefly describe two applications: the detection of outliers in experimental datasets being based on measurement information, and the identification of groups of correlated nuclear data that are likely describing a nuclear-physics observable imperfectly (e.g. keff).

1) The presence of outliers in several datasets for a nuclear physics observable may culminate in poor nuclear data and, hence, impaired predictive modelling performance for application. Identifying and removing outliers is thus a key but challenging task for nuclear data evaluators. Past efforts were conducted to perform a comprehensive review of cross-section data in the EXFOR database using statistical verification and validation techniques **[A. Koning, 2014]**.

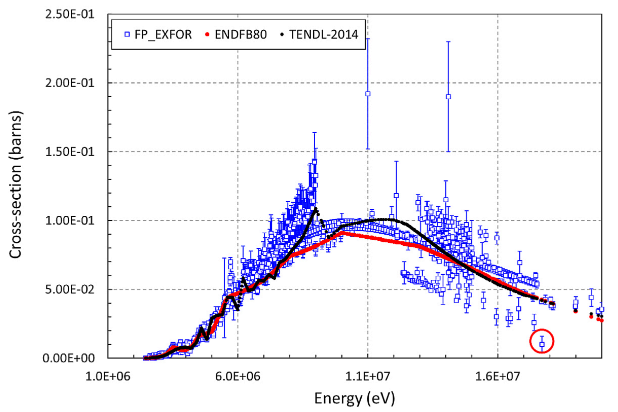

Figure 3. Potential outlier identification using DBSCAN solver for the experimental data of ^{27}Al(n,p) reaction in EXFOR database. The suspicious value is highlighted inside red-circle corresponding to the entry number “X11494.004” at neutron energy of 1.77E+07 eV. [O. Cabellos, 2021].

Nowadays, the objective is to apply modern ML techniques to detect outliers in data. As it is shown in Figure 3, the unsupervised ML algorithm DBSCAN[9] is used to identify this anomalous data in EXFOR database. In addition, this methodology may be extended to other standard databases used in the validation of nuclear data (e.g., the Handbook of criticality safety benchmark experiments **[ICSBEP, 2018]**, etc …).

Databases such as EXFOR or ICSBEP not only provide the raw data for evaluations; they also provide a wealth of information on the measurement stored as metadata. By feeding these meta-data to ML algorithms, one can identify possible correlations between specific meta-data (e.g., like a detector type) and data being outliers **[Whewell, 2020]**. Such results can provide essential insight into the physics reason why data might be outlying that is otherwise obscured by large amount of heterogeneous data.

2) In **[D. Neudecker, 2021]**, the random forest and SHAP (SHapley Additive exPlanations) metric are used with differential and integral data to partially disentangle the individual contribution of any nuclear data to predict the bias between simulated and experimental values of integral experiments representative of various NS&E applications.

This work demonstrates that one cannot unambiguously disentangle which of the several nuclear observables is biased just with one type of integral input (e.g., keff-criticality). To this end, one needs a combination of different integral observables (e.g., pulsed spheres responses, neutron transmission experiments, kinetic experiments, post irradiation experiments, etc.) with the key feature that they have different sensitivities to those nuclear-data observables in the same energy range. In addition, information from differential experiments is highlighted as being able to aid in solving this confounding if accurate and precise data are available. The resulting information provides valuable input on what future measurements and theory developments are needed.

[9] DBSCAN is Density-Based Spatial Clustering of Applications with Noise.

CONCLUSION

ML/AI methods have been introduced in nuclear reactor applications during the last decades; they contribute today to enhance the safety of the current fleet of reactors. However, these statistic techniques are becoming increasingly more practical and powerful in recent years. Hence, these current developments foretell an increase of applying ML/AI-based methods to solving issues pertaining to nuclear science and technology.

The adoption of the very same ML techniques across the nuclear data pipeline has been slower compared to the field of reactor applications. First tentative steps of applying ML/AI to nuclear data have indicated the algorithm's potential to generate better nuclear data files for the community. For the nuclear&science community, ML/AI algorithms have proven to be extremely helpful in these first initial studies to sieve through large amounts of heterogeneous data that are too large to comprehend for a human mind on its own. However, it also became clear that it takes experts to feed the algorithms carefully curated data and correctly interpret the result.

Fostering collaborations between nuclear data researchers and experts in the ML/AI community and attracting young talent to this scientific area will be critical for a successful outcome. In this sense, Universities will play an important role updating nuclear engineering academic programs with ML/AI disciplines which ensure competent graduate and undergraduate nuclear engineers familiarized with these techniques.

ACKNOWLEDGMENTS

Óscar Cabellos thanks CEIDEN for his continuous support of the UPM/INGENIA course organizing jointly the annual Workshop of Spanish Users on Nuclear Data and Codes.

Nuclear data activities at UPM are supported by the SANDA project (grant agreement 847552) of EU Horizon 2020 framework program.

Work at LANL was carrried out under the auspices of the National Nuclear Security Administration (NNSA) of the U.S. Department of Energy (DOE) under contract 89233218CNA000001. Research reported in this publication was partially supported by the U.S. DOE LDRD program at LANL. We gratefully acknowledge partial support of the Advanced Simulation and Computing program at LANL and the DOE Nuclear Criticality Safety Program, funded and managed by NNSA for the DOE.

BIBLIOGRAPHY

– **[A. Koning, 2014]** A. Koning, "*Statistical Verification and Validation of the EXFOR database: (n,n'), (n,2n), (n,p), (n,α) and other neutron-induced threshold reaction cross-sections*", NEA-DB-DOC-2014-3.
– **[CIELO, 2019]** M. Chadwick et al., "*International Co-operation in Nuclear Data Evaluation: An Extended Summary of the Collaborative International Evaluated Library Organisation (CIELO) Pilot Project*", NEA No. 7498, 2019.
– **[D. Neudecker, 2020]** D. Neudecker, M. Grosskopf, M. Herman, W. Haeck, P. Grechanuk, S. Vander Wiel, M.E. Rising, A.C. Kahler, N. Sly, P. Talou, "*Enhancing nuclear data validation analysis by using machine learning*", Nuclear Data Sheets, Volume 167, pp. 36-60 (2020).
– **[D. Neudecker, 2021]** D. Neudecker, O. Cabellos, A. R. Clark, M. J. Grosskopf, W. Haeck, M. W. Herman, J. Hutchinson, T. Kawano, A. E. Lovell, I. Stetcu, P. Talou, S. Vander Wiel, "*Informing Nuclear Physics via Machine Learning Methods with Differential and Integral Experiments*", Phys. Rev. C 104, 034611 (2021).
– **[D. Ruan et al, 2005]** D. Ruan, J. Wesley Hines, I. Pázsit, *"Computational Intelligence in Nuclear Applications: Lessons Learned and Recent Developments"*, Special Issue in Progress in Nuclear Energy, Vol. 46, Issues 3-4, (2005).
– **[G. Schanbel, 2020]** G. Schnabel, *"Consultants' Meeting on Machine Learning for Nuclear Data"*, 8-11 December 2020, IAEA, Vienna.
– **[ICSBEP, 2018]** *"International Handbook of Evaluated Criticality Safety Benchmark Experiments"*, NEA/NSC/DOC(95)/03, OECD-NEA, Paris, France (2018).
– **[J. A. Bernard, T. Washio, 1989]** J. A. Bernard, T. Washio, *"Expert Systems Applications Within the Nuclear Industry"*, Ed. American Nuclear Society, LaGrange, Park Il, (1989).
– **[J. Delua, 2021]** J. Delua, *"Supervised vs. Unsupervised Learning: What's the Difference?"*, https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning (2021).
– **[JEFDOC-2026, 2020]** F. Michel-Sendis, *"Summary Record of the JEFF Co-ordination Group meeting"*, JEFF Meeting, November 2020.
– **[KNIME, 2009]** M.R. Berthold et al., *"KNIME - the Konstanz information miner: version 2.0 and beyond"*, ACM SIGKDD Explorations Newsletter, Vol. 11 Issue 1June 2009 pp 26–31 (2009).
– **[M. Gómez-Fernández, 2020]** M. Gómez-Fernández, K. Higley, A. Tokuhiro, K. Welter, W-K. Wong, H. Yang, *"Status of research and development of learning-based approaches in nuclear science and engineering: A review"*, Nuclear Engineering and Design, Volume 359, 1, (2020).
– **[N. Otsuka, 2014]** N. Otuka et al., *"Towards a More Complete and Accurate Experimental Nuclear Reaction Data Library (EXFOR): International Collaboration Between Nuclear Reaction Data Centres (NRDC)",* Nucl. Data Sheets 120, 272 (2014).
– **[O. Cabellos, 2021]** O. Cabellos, *"Overview of Processing, Verification and Benchmarking activities at UPM"*, JEF-DOC-2041, JEFF Meeting, April 2021.
– **[R. E. Uhrig, L. H. Tsoukalas, 1999]** Robert E. Uhrig, Lefteri H. Tsoukalas, *"Soft Computing Technologies in Nuclear Engineering Applications"*, Progress in Nuclear Energy, Vol.34, no.1, pp-13-75 (1999).
– **[Scikit-Learn, 2011]** Pedregosa F et al., *"Scikit-learn: machine learning in python"*, J. Mach. Learn. Res. 12, 2825–2830, (2011).
– **[V. Sobes, 2020]** V. Sobes, M. Grosskopf, K. A. Wendt, D. Brown, M. Smith, P. Talou, *"WANDA: AI/ML for Nuclear Data"*, Workshop on Applied Nuclear Data Activities 2020, May 22, 2020.
– **[Y. Tahara, 1991]** Y. Tahara, K. Hamamoto, M. Takase, and K. Suzuki, *"Computer Aided System for Generating Fuel shuffling Configurations Based on Knowledge Engineering"*, J. of Nuclear Science and Technology, 28 (5), pp.399-408 (1991)
– **[B. Whewell, 2020]** B Whewell, M Grosskopf, Denise Neudecker, *"Evaluating 239Pu (n, f) cross sections via machine learning using experimental data, covariances, and measurement features"*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 978, pp. 164305 (2020).■