

INTELLIGENT TECHNIQUES FOR ANOMALY DETECTION IN NUCLEAR REACTORS

George Ioannou¹, Thanos Tagaris¹, Georgios Alexandridis¹,
and Andreas Stafylopatis¹

¹Institute of Communication and Computer Systems
National Technical University of Athens, Greece

geoioannou@islab.ntua.gr, thanos@islab.ntua.gr, gealexandri@islab.ntua.gr,
andreas@cs.ntua.gr

ABSTRACT

The safe operation of nuclear power plants is highly dependent on the ability of quickly and accurately identifying possible anomalies and perturbations in the reactor. Operational defects are primarily diagnosed by detectors that capture changes in the neutron flux, placed at various points inside and outside of the core. Neutron flux signals are subsequently analyzed with signal processing techniques in an effort to be better described (have their higher-order characteristics uncovered, locate transient events, etc). To this end, the application of intelligent techniques may be extremely beneficial, as it may assist and extend the current level of analysis. Besides, the combination of signal processing methodologies and machine learning techniques in the framework of nuclear power plant data is an emerging topic that has yet to show its full potential. In this context, the current contribution attempts at introducing intelligent approaches and more specifically, deep learning techniques, in neutron flux signal analysis for the identification of perturbations and other anomalies in the reactor core that may affect its operational capabilities. The obtained results of an initial stage of analysis on neutron flux signals captured at pressurized water reactors are encouraging, underlying the robustness and the potential of the proposed approach.

KEYWORDS: Anomaly Detection; Nuclear Power Plants; Deep Neural Networks; Multi-class Classification

1. INTRODUCTION

Anomaly detection is a well-established research field whose primary objective is to identify events or samples deviating from what it could be considered “ordinary” or “normal” behavior within an application domain. It has already found usage in such diverse areas as event detection in sensor networks and factory systems, intrusion detection in networks, health monitoring and even in ecosystem disturbances [1]. An increase in anomaly detection methodologies based upon intelligent techniques has also been witnessed lately [2].

The main focus of this work is the application of anomaly detection methodologies on data originating from *Nuclear Power Plants* (NPPs). It is obvious that operational safety in NPPs, as well as

performance improvement, is highly dependant upon condition monitoring and the timely detection of incipient faults. For example, a drift in steam generator feedwater flow sensors can cause a 3% reduction in a reactor's power output [3]. These faults are usually due to single or multiple component defects and their detection can be quite difficult, since the proper handling of the anomaly requires the timely identification and location of the aforementioned faults [4]. However, changes in the reactor core usually take place at a slow pace and by the time an anomaly becomes evident, it may be too late to face the underlying problem. For these reasons, it is necessary to develop techniques that are able to automatically monitor the state and conditions within nuclear reactors.

A nuclear reactor's state may be assessed through the inspection of neutron noise levels, which actually are the neutron flux fluctuations around a mean value [5]. Processes occurring within the reactor core (mechanical and fuel assembly vibrations, thermohydraulic oscillations, etc) result in changes of the neutron noise levels, which are captured by two types of sensors; *in-core* (located inside the reactor's core) and *ex-core* (located outside of the core).

More specifically, this work examines the *KWU pre-Konvoi Pressurized Water Reactors* (PWRs) [6], designed by Siemens AG. The aforementioned reactors have demonstrated high values of neutron noise since their inception [7]. Nevertheless, this activity has not been considered to be alarming, as it has not resulted in problems related to the high flux fluctuations. However, for safety reasons, it is important to frequently monitor these values.

In this direction, the introduction of intelligent techniques in reactor state monitoring is expected to be both beneficial for the safety of the NPP and also important for the whole system's performance [8]. In Section 2 below, related work on anomaly detection in NPP data along with intelligent techniques are outlined. The proposed methodology is discussed in Section 3 and the experimental framework is described in Section 4, along with the obtained results. Finally, the paper concludes in Section 5.

2. RELATED WORK

Most of the research on anomaly detection in NPP data is related to *fault detection and diagnosis* (FDD) methodologies [9]. The aforementioned methods may either be *model-based* or *model-free* and are further classified according to their adaptation to different NPP problems and intricacies. In [10], a real-time data-driven application, based upon *symbolic dynamic filtering* (SDF), is introduced and compared to *principal component analysis* (PCA) [11]. SDF is capable of extracting meaningful features from *time series* (like NPP data) through the construction of a probabilistic finite state automaton. The output of the system is a lower-dimensional vector that is subsequently used to train a classifier. In the said work, it has been demonstrated that SDF-based classifiers outperformed PCA-based ones in terms of the achieved test error rate.

In [12], outlier identification on neutron flow signals originating from a nuclear reactor channel is achieved via *support vector data description* (SVDD) [13]. Additionally, *deep learning* (DL) techniques [14] localizing the specific points in the reactor core where perturbations originate from, have been employed in [15]. Specifically, time signals transformed to the frequency domain are provided to a *convolutional neural network* (CNN) [14], with the overall architecture achieving satisfactory results. In the same work, a *denoising autoencoder* along with a clustering technique

are also employed for perturbation identification and localization. Finally, in [16], anomalies in the reactor core are identified through the combined use of wavelet-based scaleograms and DL methodologies.

3. PROPOSED METHODOLOGY

The objective of the proposed methodology is to be able to identify up to N co-occurring perturbations through the application of *multi-class* classification [17]. In particular, an *one-vs-all* (also known as *one-vs-rest*) architecture is used (Figure 1), with each binary classifier being independently trained to recognize only one of the N perturbations (classes).

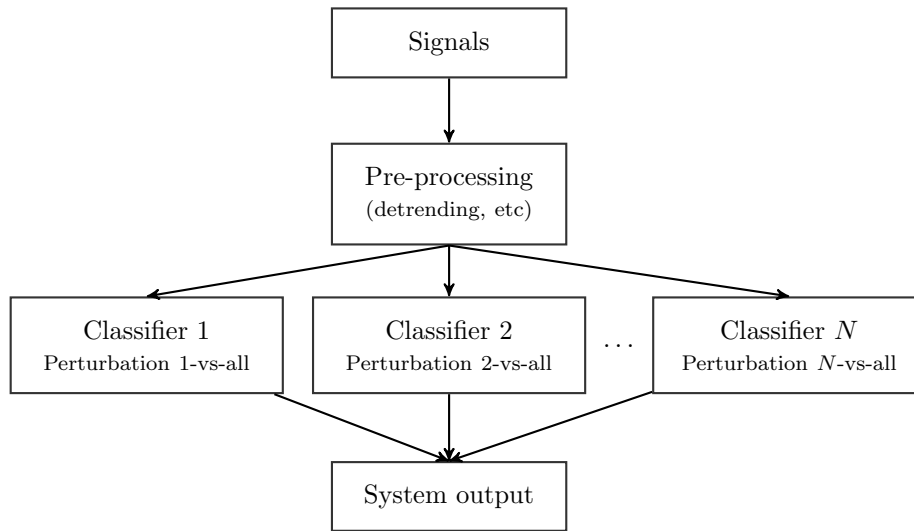


Figure 1: One-vs-All System Architecture

During inference, the preprocessed signal is provided as input to all N classifiers and their predictions are merged to generate the final output, which is an N -dimensional binary vector that can have an arbitrary number of ones and zeros, depending on the number of perturbations present in the signal.

Each classifier is itself an independent *deep neural network* (DNN) [14] that consists of *convolutional* and *recurrent* components. In principle, CNNs are types of DNN that are comprised of *convolutional*, *pooling* and *fully connected* (FC) layers and are commonly employed in image analysis. They use *weight sharing* and *sparse connectivity* to extract spatial information from images, significantly reducing the number of parameters per layer, which allows for the creation of even deeper networks. CNNs are considered to be the state-of-the-art for most image-related tasks and have been applied to many domains. In this particular case, a special type of CNN has been used, that includes *1-dimensional* (1-D) convolutional filters instead of the commonly-used 2-D ones. These were chosen in order to adhere to the single dimensional shape of the input signals.

A *recurrent neural network* (RNN) [14], on the other hand, is a DNN suitable for sequential data (e.g. time series). It is designed to recognize patterns and make associations through time. This is

achieved via the concurrent input of multiple sequence instances (present and past) to the network, thereby forming a short of memory, that can be used to analyze temporal relationships. Network training is attained via *backpropagation through time* [18], a variant of the standard neural network training algorithm. To overcome certain issues occurring during training (namely vanishing and exploding gradients) and to be able to handle long-term dependencies, a special type of component has been chosen, the *long short-term memory* (LSTM) [19] unit. In this study, RNNs with LSTM units have been employed to extract time-dependent information out of the input signals.

4. EXPERIMENTAL FRAMEWORK

4.1. Dataset Description

The proposed methodology has been trained and evaluated upon data generated by the SIMULATE-3K tool [20,21], modelling some basic types of perturbations occurring in nuclear reactors. The aforementioned tool is used to model time-dependent situations which, in the present case, are induced by adding different perturbations, such as *fuel assembly vibrations* (of both single assemblies and of clusters), *coolant flow* and *temperature oscillations*, as compared to the steady-state operation of the reactor. Specifically, the 3 main perturbation types of Table 1 have been considered in the experiments that follow, for a 4-loop Westinghouse PWR mixed core, utilized in the OECD/NEA transient benchmark [22] (Boron concentration: 408 ppm, mass flow rate: 100%, system pressure: 2250 psi, inlet coolant temperature: $286.7^{\circ}C$).

Table 1: Perturbation types & scenarios considered in the experiments

Scenario	Perturbation Type	Description	Frequency	Amplitude
1	I	5×5 Central Cluster of Fuel Assemblies synchronously vibrating in the x -direction	white noise	$1mm$
2				$0.5mm$
3			$1.5Hz$	$1mm$
4				$0.5mm$
5	II	Coolant Flow Oscillations (Mean value: 100%)	Random	$\pm 1\%$
6	III	Coolant Temperature Oscillations (Mean value: $286.7^{\circ}C$)	Random	$\pm 1^{\circ}C$

Every simulated scenario included measurements from 48 in-core and 8 ex-core detectors. An initial experimentation revealed that the contribution of the latter category of detectors in the prediction task was marginal, due to the spatial attenuation of the induced neutron noise between the core and outside of the core. As a result, the analysis that follows has been solely based on in-core detector signals.

More specifically, the in-core sensors were located at 8 different axial locations, each taking measurements at 6 different heights [23]. Each signal had a duration of 100secs and a sampling rate of 100Hz, resulting in a length of 10,000 data points. In all signals, linear trend has been detected (pre-processing step) and subsequently removed.

4.2. Experimental Procedure

As outlined in Section 3, the one-vs-all classification strategy has been used. Thus, 3 different DNN binary classifiers have been trained independently, one for each perturbation type. After thorough experimentation with different network types and sizes, the optimum architecture for identification of perturbations of Type II & III is illustrated in Figure 2. The first two convolutional layers have 64 and 32 filters respectively, while the LSTM layer is comprised of 20 units. Obviously, the FC layers have a single sigmoid-activated output, denoting whether the signal belongs or not to the respective class (perturbation).



Figure 2: Deep Neural Network Architecture

Type I perturbations proved to be harder to learn (compared to the rest), requiring a smaller architecture, as larger ones were prone to overfitting. The optimal network included two 1D convolutions (with 48 and 24 filters, respectively), an average pooling layer and two FC ones (the first with 10 units and the final with 1).

All models have been trained for 20 epochs, using the *Adam optimizer* [24] and *binary cross-entropy* as the loss function. The batch size was set to 24 and due to the imbalance in the number of samples per class (signals per perturbation type), class weights have been used. Particularly, for the first network, a weighting scheme of 2 : 1 was used (i.e. perturbations of Type I were twice as important as perturbations of Types II & III), while the second and third classifiers have been trained on a 1 : 5 weight ratio.

4.3. Results

Initially, each classifier has been evaluated on its own task, i.e. classifier 1 for classifying perturbations of Type I and so forth. The results of this single-label classification, on a held-out test set are presented in Table 2. The accuracy for the latter two classifiers is exceedingly high, but it drops off a bit for the first. Nevertheless, the overall results are considered to be quite satisfactory.

The mean validation accuracy per perturbation type of the overall architecture (multi-label classification) is outlined in Table 3. The detection accuracy for the coolant oscillation perturbations (Types II & III) is competent. The performance on perturbations of a cluster of fuel assembly synchronously vibrating (Type I), on the other hand, is sub-par. This is attributed, in part, to the fact that the task is more difficult in nature, causing the models to be more prone to overfitting.

Table 2: Validation Accuracy for Single-label Classification

Classifier	1	2	3
Accuracy	84.02%	97.61%	98.35%

Table 3: Mean Validation Accuracy for Multi-class Classification

Perturbation Type	I	II	III
Overall architecture	62.4%	85.5%	79.2%

Finally, the performance of the overall architecture is further evaluated upon a set of *information retrieval* metrics (Table 4) [25]. *Precision* is the probability of correctly identifying a perturbation type. *Recall* is the probability of identifying all instances of a perturbation type. The *F1-score* is their harmonic mean, while *support* designates the number of occurrences of each perturbation in the validation set. Their average (5th row) is computed on each sample individually and then averaged over all samples (*sample average*). All metrics, apart from the last one, are given on the percentage scale. In general, the desirable classifier performance is to exhibit the same levels of precision & recall for each class (perturbation), with both metrics being as high as possible. From the first 3 rows of Table 4 it is evident that the proposed architecture achieves more than satisfactory results, with the minor exception of the recall metric on Type I perturbations.

Table 4: Detailed Report for Multi-Label Classification

	Precision	Recall	F1-Score	Support
Type I	91%	64%	75%	6,144
Type II	84%	92%	88%	3,840
Type III	87%	75%	80%	3,840
Average	89%	74%	79%	13,824

5. CONCLUSION

In this paper, a novel methodology for anomaly detection in nuclear reactors, based on measuring neutron noise levels, has been presented. The outlined technique employs one-vs-all multi-class classification, combining some state-of-the-art DNN architectures, such as CNNs and RNNs, to achieve perturbation identification. The aforementioned models have the capability of uncovering complex relationships in signals and for this reason, they achieve significant performance.

Experiments have been performed on simulated data for PWRs, modelling 3 different perturbation types. The obtained results on a variety of relevant metrics, both in single and multi-class classification cases, have been encouraging so far and call for the test of the overall approach on real plant measurements.

Even though Type I perturbations may be considered a very special case, both in nature and in location, the applicability of the method is not affected, as it is actually a modular architecture for anomaly detection that can easily be extended to account for even more perturbation types than those considered in the experiments.

ACKNOWLEDGEMENTS

The research conducted has been made possible through funding from the Euratom research and training programme 2014-2018 under grant agreement No 754316 for the “CORE Monitoring Techniques And EXperimental Validation And Demonstration (CORTEX)” Horizon 2020 project, 2017-2021.

We express our gratitude to the Paul Scherrer Institute (PSI), Switzerland, for supplying the data used in our experiments, within the framework of the aforementioned CORTEX project. The data have been generated by PSI using their SIMULATE-3K tool [21,20].

REFERENCES

- [1] V. Chandola, A. Banerjee, and V. Kumar. “Anomaly Detection: A Survey.” *ACM Comput Surv*, **volume 41**(3), pp. 15:1–15:58 (2009). URL <http://doi.acm.org/10.1145/1541880.1541882>.
- [2] A. L. Buczak and E. Guven. “A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection.” *IEEE Communications Surveys Tutorials*, **volume 18**(2), pp. 1153–1176 (2016).
- [3] A. Chan and A. Ahluwalia. “Feedwater flow measurement in US nuclear power generation stations.” Technical report, Electric Power Research Inst. (1992).
- [4] C. Demaziere and G. Andhill. “Identification and localization of absorbers of variable strength in nuclear reactors.” *Annals of Nuclear Energy*, **volume 32**(8), pp. 812 – 842 (2005). URL <http://www.sciencedirect.com/science/article/pii/S0306454905000319>.
- [5] I. Pázsit and C. Demazière. *Noise Techniques in Nuclear Systems*, pp. 1629–1737. Springer US, Boston, MA (2010). URL https://doi.org/10.1007/978-0-387-98149-9_14.
- [6] L. Torres, D. Chionis, C. Montalvo, A. Dokhane, and A. García-Berrocal. “Neutron noise analysis of simulated mechanical and thermal-hydraulic perturbations in a PWR core.” *Annals of Nuclear Energy*, **volume 126**, pp. 242–252 (2019).
- [7] E. Laggiard and J. Runkel. “Detection of subcooled boiling in a PWR using noise analysis and calculation of the steam void fraction.” *Annals of Nuclear Energy*, **volume 24**(1), pp. 49–54 (1997).
- [8] F. D. S. Ribeiro, F. Caliva, D. Chionis, A. Dokhane, A. Mylonakis, C. Demaziere, G. Leontidis, and S. Kollias. “Towards a Deep Unified Framework for Nuclear Reactor Perturbation Analysis.” In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 120–127 (2018).

- [9] J. Ma and J. Jiang. “Applications of fault detection and diagnosis methods in nuclear power plants: A review.” *Progress in nuclear energy*, **volume 53**(3), pp. 255–266 (2011).
- [10] X. Jin, Y. Guo, S. Sarkar, A. Ray, and R. M. Edwards. “Anomaly detection in nuclear power plants via symbolic dynamic filtering.” *IEEE Transactions on Nuclear Science*, **volume 58**(1), pp. 277–288 (2010).
- [11] H. Hotelling. “Relations Between Two Sets of Variates.” *Biometrika*, **volume 28**(3/4), pp. 321–377 (1936). URL <http://www.jstor.org/stable/2333955>.
- [12] C. K. Maurya and D. Toshniwal. “Anomaly detection in nuclear power plant data using support vector data description.” In *Proceedings of the 2014 IEEE Students’ Technology Symposium*, pp. 82–86. IEEE (2014).
- [13] D. M. Tax and R. P. Duin. “Support vector data description.” *Machine learning*, **volume 54**(1), pp. 45–66 (2004).
- [14] Y. LeCun, Y. Bengio, and G. Hinton. “Deep learning.” *nature*, **volume 521**(7553), p. 436 (2015).
- [15] F. Caliva, F. S. De Ribeiro, A. Mylonakis, C. Demaziere, P. Vinai, G. Leontidis, and S. Kollias. “A deep learning approach to anomaly detection in nuclear reactors.” In *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. IEEE (2018).
- [16] T. Tagaris, G. Ioannou, M. Sdraka, G. Alexandridis, and A. Stafylopatis. “Putting Together Wavelet-Based Scaleograms and Convolutional Neural Networks for Anomaly Detection in Nuclear Reactors.” In *Proceedings of the 2019 3rd International Conference on Advances in Artificial Intelligence, ICAAI 2019*, p. 237243. Association for Computing Machinery, New York, NY, USA (2019). URL <https://doi.org/10.1145/3369114.3369121>.
- [17] C. M. Bishop. *Pattern recognition and machine learning*. springer (2006).
- [18] P. J. Werbos et al. “Backpropagation through time: what it does and how to do it.” *Proceedings of the IEEE*, **volume 78**(10), pp. 1550–1560 (1990).
- [19] S. Hochreiter and J. Schmidhuber. “Long short-term memory.” *Neural computation*, **volume 9**(8), pp. 1735–1780 (1997).
- [20] D. Chionis, A. Dokhane, L. Belblidia, M. Pecchia, G. Girardin, H. Ferroukhi, and A. Pautz. “SIMULATE-3K analyses of neutron noise response to fuel assembly vibrations and thermal-hydraulics parameters fluctuations.” In *M&C 2017-International Conference on Mathematics & Computational Methods Applied to Nuclear Science & Engineering, at Jeju, Korea* (2017).
- [21] G. Grandi. “SIMULATE-3K Input specification (SSP-98/12, Rev. 17).” (2015).
- [22] T. Kozłowski and T. J. Downar. “OECD/NEA and US NRC PWR MOX/UO₂ core transient benchmark - Final Specifications.” techreport Revision 2, OECD Nuclear Energy Agency, Nuclear Science Committee (2003). URL https://engineering.purdue.edu/PARCS/MOX_Benchmark/Benchmark_Description/mox_bench_spec.pdf.
- [23] P. Bernitt. *In-Core Neutron Noise Analysis for Diagnosis of Fuel Assembly Vibrations*. Chalmers University of Technology (2008).
- [24] D. P. Kingma and J. Ba. “Adam: A method for stochastic optimization.” *arXiv preprint arXiv:1412.6980* (2014).
- [25] A. Kent, M. M. Berry, F. U. Luehrs Jr, and J. W. Perry. “Machine literature searching VIII. Operational criteria for designing information retrieval systems.” *American documentation*, **volume 6**(2), pp. 93–101 (1955).