Progress in Nuclear Energy 151 (2022) 104339

Contents lists available at ScienceDirect

Progress in Nuclear Energy

journal homepage: www.elsevier.com/locate/pnucene

Review

Deep learning for safety assessment of nuclear power reactors: Reliability, explainability, and research opportunities

Abiodun Ayodeji [a,*], Muritala Alade Amidu [b,c,**], Samuel Abiodun Olatubosun [d,***], Yacine Addad [b,c,****], Hafiz Ahmed [a,*****]

[a] Nuclear Futures Institute, Bangor University, Bangor, Gwynedd, LL57 1UT, United Kingdom
[b] Department of Nuclear Engineering, Khalifa University of Science and Technology, United Arab Emirates
[c] Emirates Nuclear Technology Center (ENTC), Khalifa University of Science and Technology, United Arab Emirates
[d] Department of Mechanical and Aerospace Engineering, The Ohio State University, Columbus, OH, 43210, USA

ARTICLE INFO

Keywords:
Deep learning
Uncertainty quantification
Reliability
Modeling and simulation
Nuclear reactor safety
Sensitivity analysis
Machine learning

ABSTRACT

Deep learning algorithms provide plausible benefits for efficient prediction and analysis of nuclear reactor safety phenomena. However, research works that discuss the critical challenges with deep learning models from the reactor safety perspective are limited. This article presents the state-of-the-art in deep learning application in nuclear reactor safety analysis, and the inherent limitations in deep learning models. In addition, critical issues such as deep learning model explainability, sensitivity and uncertainty constraints, model reliability, and trustworthiness are discussed from the nuclear safety perspective, and robust solutions to the identified issues are also presented. As a major contribution, a deep feedforward neural network is developed as a surrogate model to predict turbulent eddy viscosity in Reynolds-averaged Navier–Stokes (RANS) simulation. Further, the deep feedforward neural network performance is compared with the conventional Spalart Allmaras closure model in the RANS turbulence closure simulation. In addition, the Shapely Additive Explanation (SHAP) and the local interpretable model-agnostic explanations (LIME) APIs are introduced to explain the deep feedforward neural network predictions. Finally, exciting research opportunities to optimize deep learning-based reactor safety analysis are presented.

1. Introduction

Increasingly, energy policy decisions are shifting in favor of low-carbon sources. Amid global economic contraction from the effect of COVID-19, and the debate on global warming, nations are tilting energy policies toward net-zero emission technology that maximizes return on investment. Nuclear power reactors have proven to be a reliable, consistent, and low-carbon energy source with high energy density. Moreover, the long-operating time, low maintenance cost, and load-following capability present nuclear energy as a potential substitute for fossil-based sources. However, the construction cost, construction delays, safety concerns, and public acceptance remain major bottlenecks that restrict the utilization of nuclear energy. These bottlenecks also pose an existential threat to the operating and planned nuclear power plants.

To address some of these bottlenecks and improve the safety and economic competitiveness of nuclear power plants (NPP), next-generation reactors are being developed. These reactors are characterized by modularity, improved safety margin, and significantly less construction and operation costs, with support for localized microgrids and multiple energy applications. However, timely development of the next-generation reactors to meet global net-zero targets require significant investment and the adoption of advanced research tools and techniques. This is evident given the expensive experimental test facilities

* Corresponding author.
** Corresponding author. Department of Nuclear Engineering, Khalifa University of Science and Technology, United Arab Emirates.
*** Corresponding author.
**** Corresponding author. Department of Nuclear Engineering, Khalifa University of Science and Technology, United Arab Emirates.
***** Corresponding author.
E-mail addresses: ayod_abe@yahoo.com (A. Ayodeji), amidu.alade@ku.ac.ae (M.A. Amidu), abitubosun@gmail.com (S.A. Olatubosun), yacine.addad@ku.ac.ae (Y. Addad), hafiz.ahmed@bangor.ac.uk (H. Ahmed).

https://doi.org/10.1016/j.pnucene.2022.104339
Received 23 April 2022; Received in revised form 4 July 2022; Accepted 13 July 2022
Available online 2 August 2022

necessary for safety verification. In addition, researchers increasingly rely on computational tools that require a significantly long time to converge.

Recently, machine learning (ML) algorithms are being leveraged to analyze different engineering systems. In heavy industries, ML models have been used for reactor probabilistic safety modeling (Worrell et al., 2019), risk assessment techniques modernization (Moradi and Groth, 2020), and systems identification and control (Patan and Patan, 2020; Naimi et al., 2022). In engineering systems, ML models have also been used for early accident diagnosis (Tolo et al., 2019), to study the reliability of systems (Xu and Saleh, 2021), and for risk-based inspection (Rachman and Ratnayake, 2019). From optimizing fuel depletion to intelligent load forecasting, ML and data-driven algorithms are also being used as advanced predictive tools to enhance the development and optimization of next-generation energy systems (Duan and Luo, 2021). However, the conventional ML algorithms are sub-optimal, as significant data preprocessing and feature engineering are required. Moreover, latent and spatial generalization is not guaranteed.

Deep learning (DL), a subset of ML techniques that utilizes multiple layers of computational neurons (neural network) to learn the latent representation in a given input, is the state-of-the-art technology used to model complex, nonlinear industrial systems (Zhou et al., 2021). A DL model has the capability to approximate arbitrary functions, maximize accuracy, and handle high-dimensional and strongly nonlinear problems. In the nuclear industry, the DL model capability is being leveraged to accelerate research in thermal-hydraulic analysis, safety margin quantification, uncertainty analysis, neutron transport prediction, and autonomous control of next-generation reactors. DL techniques have also been applied in tasks such as the optimization of steam supply systems (Dong et al., 2020) and the prediction of leakage in reactor coolant pumps (Nguyen et al., 2021).

Further, DL models have been used as embedded surrogate models to compress complex and expensive computations. For instance, physics-constrained DL models are used for surrogate modeling of fluid flows (Sun et al., 2020), for parameter prediction of two-phase flows (Gao et al., 2020), for uncertainty quantification (Tripathy and Bilionis, 2018), and for modeling nuclear data uncertainties (Radaideh et al., 2021a). Deep learning models have also been used to identify flow regimes (Yang et al., 2017), predict heat transfer coefficients (Ma et al., 2017), calculate critical heat flux (Park et al., 2020), predict void fraction (Chu et al., 2021), and characterize two-phase flow (Gao et al., 2021), with satisfactory accuracy. Impressive results have also been reported for DL models utilized to forecast the loss of coolant accident (Radaideh et al., 2020; She et al., 2021), to predict reactor water level during a severe accident (Do Koo et al., 2019), for reactor parameters monitoring and data augmentation (Ayodeji and Liu, 2019; Ayodeji et al., 2019), component fault diagnosis (Kim et al., 2020; Zhao et al., 2021a), nuclear plant valve prognosis (Wang et al., 2020, 2021a) and nuclear component predictive maintenance (Liu et al., 2021; Wang et al., 2021b).

The current and potential application of DL algorithms in nuclear safety analysis is attractive. However, critical issues need to be addressed to sustain and improve its application, especially in a highly regulated, nuclear industry. First, the high-risk, safety-critical nuclear environment demands proven techniques, as the consequences of a wrong prediction are significant. In addition, implementing black box models would constrain DL applications in practical reactor analysis, this issue should be properly addressed. DL applications in nuclear safety are becoming prevalent, however, to the best of the authors' knowledge, research works that discuss the critical challenges with DL models from the reactor safety perspective are limited.

Towards an enhanced safety margin for the next-generation reactors, this paper presents the state-of-the-art applications of DL algorithms for nuclear power reactor safety analysis. First, this paper presents the state-of-the-art DL algorithms commonly applied for complex system modeling. Secondly, the paper critically reviews recent publications in DL algorithm application for nuclear safety analysis and the merits and demerits of the proposed approaches. Thirdly, the data availability, reliability, and explainability constraints of DL model application for nuclear safety analysis are enumerated. The available literature shows that the application of DL to two-fluid closure modeling is still nascent. Hence, as a major contribution, the application of the DL algorithm as a surrogate closure model in Reynolds-averaged Navier–Stokes (RANS) simulation is demonstrated. Then, the DL model convergence speed and accuracy is compared with the conventional Spalart Allmaras closure model. Further, recent explainability tools that could aid DL models' trustworthiness in reactor safety analysis are presented and demonstrated. This paper contributes to knowledge by addressing the following issues:

1. This work presents the state-of-the-art in deep learning application for nuclear power reactor safety analysis.
2. From a reactor safety analysis perspective, this study enumerates sources of uncertainty in DL-based models. This paper also discusses data availability, model reliability, and interpretability issues with the existing DL approach.
3. To aid the reproducibility of DL-based reactor safety analysis, this paper also demonstrates the application of the deep feedforward neural network (DFNN) model for turbulent eddy viscosity prediction using the Eulerian-Eulerian two-fluid model. Moreover, the convergence speed and accuracy of the DFNN-based turbulence simulation are compared with the conventional Spalart Allmaras model in RANS simulation.
4. Towards improved trust for DL-based reactor safety analysis results, two explainability tools – the Shapely Additive Explanation (SHAP) and the local interpretable model-agnostic explanations (LIME) – are used to explain the DFNN local and global predictions.
5. Finally, future research focus that could improve reactor safety margins, optimize reactor performance, and accelerate the development of next-generation reactors are discussed.

This work is arranged as follows: the next section discusses the state-of-the-art in DL applications for reactor safety assessment. Section 3 demonstrates the application of a DL model for RANS simulation optimization. Section 4 explains the model predictions obtained from section 3, and section 5 discusses the exciting research opportunities and potential directions towards optimized reactor safety analyses.

2. Applications of deep learning in reactor safety assessment: state-of-the-art

The safety assessment of a nuclear reactor is a systematic process that includes the safety analysis of the reactor. This safety assessment is performed to ensure that the reactor design meets the relevant safety requirements set by the operating organization and the regulators. The safety analysis involves the use of appropriate numerical tools (system analysis code, computational fluid dynamics (CFD) codes, engineering-level code, etc.) to establish and confirm the safety of the components of the reactor as well as the overall plant design. The numerical tools are made up of empirical or semi-mechanistic models that capture the behaviors of phenomena that are deemed important to safety in the plant. These models are traditionally developed from experimental observations. However, the traditional method of developing these empirical models is expensive and more so, the models may be limited by the conditions and configurations of the experiments from which they are derived. There is, therefore, the availability of big experimental data for different range of applications. These deficiencies hamper the application of these models in simulating new systems conditions and configurations. As a feasible substitute, a statistical data-driven modeling approach, especially modern DL techniques, could be used to reveal the functional relations behind the big experimental data and the resulting data-driven model can be employed instead of the traditional empirical

or semi-mechanistic models. Thus, current application of DL techniques to analyze key reactor components (reactor thermal-hydraulic analysis, reactor core neutronic analysis, fuel loading optimization, severe accident management, etc) are summarized in Table 1 and further presented in detail in the following sections.

2.1. Reactor thermal-hydraulics analysis

A nuclear reactor involves a large generation of heat, and the heat generation must be carefully analyzed to determine the temperature and density distribution of various core materials under steady-state and transient conditions. The temperature limitations of the core materials determine the maximum power the reactor core can produce. Thus, accurate prediction of the reactor core temperature through thermal-hydraulic analysis is paramount for the safe operation of the reactor. The thermal-hydraulic analysis determines the significant parameters (plant efficiency and system coolability) for reactor design and consists of three interdependent key parts: thermodynamic, fluid mechanics, and heat transfer. This analysis can be achieved with the use of computational fluid dynamics (CFD) code or system analysis codes as discussed in the following sub-sections.

Table 1
Summary of previous application of deep learning in nuclear safety assessment.

References	Nuclear safety assessment application	Method	Training Data Generation	Predicted parameter
Chang and Dinh (2019)	Vertical boiling channel	4-layer FNN	Two-fluid CFD simulation	Slip ratio of a mixture CFD model
Hanna et al. (2020)	Turbulent flow in a cavity	1-layer FNN	DNS CFD simulation	Coarse Grid -CFD error
Bao et al. (2019)	Turbulent mixing	3-layer FNN	RANS CFD simulation	Mesh optimization
Kang et al. (2004)	Rod bundle	10-layer DNN with POD	RANS CFD simulation	Rod bundle flow field
Ling et al. (2016b)	Turbulent flow	10-layer	DNS simulation	Reynold stress anisotropy
Tian et al. (2018)	Loss of coolant accident (LOCA)	2- layer FNN	RELAP simulation	Break size in LOCA
Kim et al. (2019)	Reactor protection system (alarm)	CNN	RELAP simulation	Reactor states with alarm
Saeed et al. (2020)	Fault diagnosis	LSTM & CNN	RELAP simulation	Fault types
Lu et al. (2021a)	Analysis of nuclear reactor core and steam generator	DNN	RELAP simulation	Thermal hydraulic parameters
Guillen et al. (2020)	Analysis of drywell cooling fan failure	LSTM	RELAP simulation	Fan coil unit outlet temperature
Do Koo et al. (2019)	Severe accident monitoring	GA-DNN	MAAP simulation	Reactor vessel water level
Lee et al. (2020)	Severe accident management	DNN & CNN	DPSA/DPRA and MELCOR simulation	Online operator support tool
Zhao et al. (2021b)	MDNBR margin	DNN	Experimental dataset	Departure from nucleate boiling
Shriver et al. (2021)	Core neutronic analysis	CNN	VERA simulation	Pin power and k_{eff}
Saleem et al. (2020)	Core neutronic analysis	DNN	PARCS core simulation	Power peaking factor, control rod level, and cycle length
Bae et al. (2020)	Spent nuclear fuel analysis	DNN	UNF-ST&DARDS Unified Database	Spent nuclear fuel composition and decay heat

2.1.1. Application of deep learning in computational fluid dynamics (CFD)

The application of computational fluid dynamics methods to the study of nuclear reactor thermal hydraulics has become prominent over the years. For transient flows, the application of the basic equations of flow, also known as the Navier-Stokes equations, to the simulation of turbulent flows is called direct numerical simulation (DNS). This technique is ideal for the investigation of the basic turbulence mechanisms, but its use is limited to weak turbulence flows because the computational costs become prohibitive at a high Reynolds number (representing the intensity of the turbulence) as reported by Stefan (Heinz, 2003). Regrettably, it is apparent today that DNS in its current form cannot be deployed for flow predictions of technological or environmental relevance. Thus, the DNS method is confined to simple problems of relatively low Reynolds number (≤ 600) as compared to the Reynolds number of $\sim 1 \times 10^8$ inside the primary piping of a nuclear reactor. To resolve this problem of inapplicability of the basic equations of fluid flows to most industrial problems, the basic equations are averaged through ensemble means to obtain mean equations which are called the Reynolds-Averaged Navier-Stokes (RANS) equations. A typical example of this RANS formulation is the classical two-fluid model.

In RANS formulation, some transient details are lost due to the ensemble averaging process from which they are derived. These lost details are compensated for through closure relations which are physics-based models used to capture the microscopic phenomena (Ishii and Hibiki, 2010; Amidu et al., 2020), among others. Over the years, empirical or semi-mechanistic correlations derived from experimental observations and data have been used as closure relations in the RANS model. However, the traditional method of developing these empirical correlations is expensive and more so, the resultant closure term may be limited by the conditions and configurations of the experiments from which they are derived. There is, therefore, the availability of big experimental data for a different range of applications. These deficiencies hamper the application of the computational fluid dynamics methods in simulating new systems conditions and configurations. As a feasible substitute, a statistical data-driven modeling approach especially DL technique could be used to reveal the functional relations behind the big experimental data and the resulting data-driven closure model can be employed instead of the traditional empirical or semi-mechanistic closure relationship.

For the large database to be relevant in data-driven modeling, numerous preprocessing activities need to be performed on the data (Ackoff, 1989). The first step is to collect and categorize the data such that its archive is readily accessible. Thereafter, the contents of the data should be evaluated vis-à-vis their importance to the closure model conditions under consideration. Subsequently, statistical learning methods (including machine learning) are applied to the refined and organized data to detect and identify functional relations behind the data. The developed data-driven closure model can then be adopted in the simulation of thermal fluid problems to enhance the accuracy of the simulation. The framework of the data-driven modeling approach is depicted in Fig. 1 which involves Data as a Service (DaaS) concept, and the Method as a Service (MaaS) concept as can be found in the article written by Mell and Grance (2011). These two concepts (DaaS and MaaS) provide quality assurance and closure relation formulation, respectively. In addition, a Platform as a Service (PaaS) concept (shown in Fig. 1) provides a simulation platform that stores thermal-fluid-model with distinct DL-base closures that may require a specific numeric scheme. Thus, the PaaS concept provide the platform for model selection.

As far as closure surrogates' models are concerned, Chang and Dinh (2019) categorized the ML frameworks into five viz: physics-separated, physics-integrated, physics-evaluated, physics-recovered, and

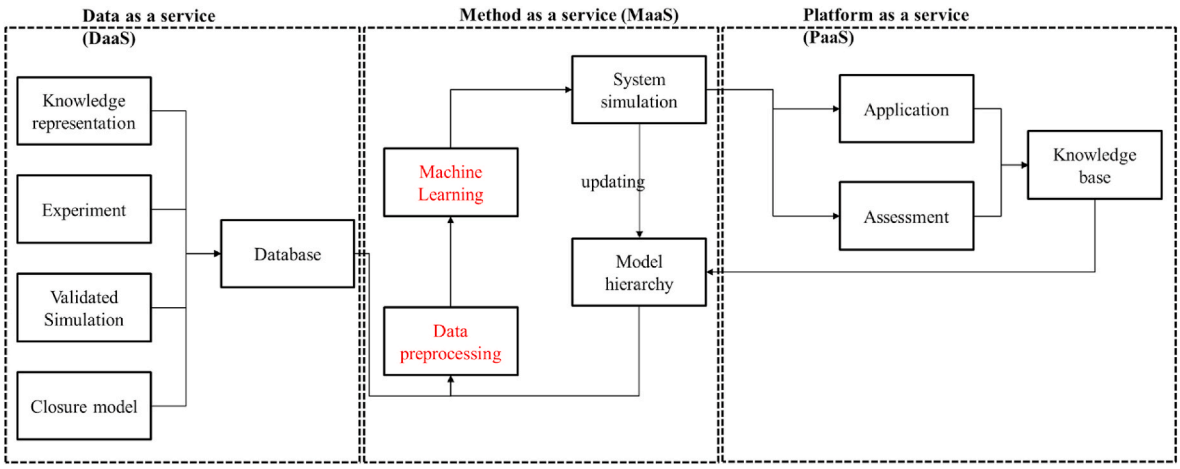

Fig. 1. Deep learning-based safety analysis modeling framework.

physics-discovered ML frameworks. The physics-separated framework requires separation of scales, and the closure relations are separately built from separate effect test (SET) data which is then implemented in the RANS conservation equations. The separation of the closure model from the conservation equations could lead to model biases in case the assumption of separation of scale becomes unachievable. The architecture of the physics-separated ML framework includes forward data-driven modeling without feedback. This indicates that the performance of the closure relation built under physics-separated architecture is largely influenced by the quality of the training data. Whereas the physics-integrated ML approach does not assume scale separation. Rather, data-driven closure relations are incorporated and trained within the system dynamics (RANS conservation equations). The data used in this approach could be obtained from separate effect experiments or integral effect experiments. Therefore, the physics-integrated approach shifts the paradigm from separate effect experimental data to the integral effect experimental data. In this approach, the closure relations are closely coupled with the conservation equation which means that heavy computational cost must be paid when the approach is used even though it promises more accurate prediction of complex thermal fluid problems where separation of scales could result in significant errors.

In addition, physics-evaluated ML is targeted at minimizing uncertainty in the conservation equations, and prior knowledge is required in the selection of closure models to predict the thermal fluid behavior. With this technique, high-fidelity data is used to inform a low-fidelity simulation to minimize the uncertainty in the low-fidelity simulation. This way, the non-linear thermal fluid physics behind the high-dimensional data can be captured. Moreover, another algorithm is the physics-recovered ML which is aimed at recovering the exact form of the governing equations. This framework constructs a candidate library that includes components of governing equations such as time derivative, advection, diffusion, and higher-order terms (Kutz et al., 2016). And lastly, physics-discovered ML is the extreme case that requires neither governing equations nor prior knowledge of the closure models. It completely relies on data for the discovery of efficient predictive models, and this could be instrumental for a new modeling paradigm for complex thermal fluid systems.

DL-based methods can be used to address the numerical errors resulting from discretization during the application of CFD to the simulation of nuclear thermal-hydraulic phenomena. Towards this end, Hanna et al. (2020) conducted a feasibility study on the application of ML techniques to produce a proxy model that can compute the coarse-grid CFD (CG-CFD) local errors to correct a specific variable under consideration. The principal focus of their work was the correction of discretization error in coarse-grid CFD computation while disregarding the possible model errors that may happen in the application of CFD to thermal-hydraulic analysis. This approach was then applied to identify parameters that influence nuclear reactor containment thermal-hydraulic (CTH) phenomena (Hanna et al., 2020). By using the CG-CFD to minimize the computation cost, a surrogate model was developed to predict the CG-CFD local errors that can be used to correct the fluid flow variables. This CG-CFD surrogate model was able to correct coarse grid results and provide reasonable predictions of new cases of 3-dimensional turbulent flow in a lid-driven cavity.

Later on, a data-driven framework called Optimal Mesh/Model Information System (OMIS) was developed by Bao et al. (2019) to provide error predictions and suggest the optimal mesh size and models for system-level thermal-hydraulic simulation. The development of the OMIS framework is targeted at thermal-hydraulic codes with some specific features. For instance, it is appropriate for CFD-like codes that use coarse mesh sizes and simplified boundary-layer correlations whose range of applicability depends on the respective characteristic lengths scale. It is also deemed appropriate for coarse-mesh Reynolds-averaged Navier–Stokes (RANS) methods with wall functions. The OMIS framework has been successfully applied in both adiabatic fluid and thermal heating fluid dynamics. Moreover, a previous demonstration of the OMIS predictive capability has been showcased for the turbulent mixing case by Bao et al. (2018). In addition, the computational cost of CFD simulation of some time-consuming scenarios (such as sensitivity analysis and uncertainty quantification) is a problem that can be addressed using DL techniques. For instance, Kang et al. (2004) proposed a reduced-order model (ROM) which combines proper orthogonal decomposition (POD) and DL to solve this kind of computational cost problem. Using a nuclear reactor fuel rod bundles configuration, the evaluation of the ROM showed an accurate description of the flow fields with high resolution in only a few milliseconds.

Another key closure parameter that is being targeted with the deep-learning-based method in the RANS model application to the simulation of nuclear thermal hydraulics is the turbulence model. DL-based methods have been used to improve turbulence modeling in RANS simulations using high-fidelity simulations and experimental data (Kutz, 2017). The main approach behind the DL-based method is to establish a relationship between the unclosed terms with other known features (extracted from data) and use the neural network to fit the unknown parameters for this relationship. For instance, Ling and Templeton (2015) applied different DL techniques to evaluate RANS accuracy compared to DNS data and to identify regions where RANS simulations fail while deep neural networks (Ling et al., 2016a) were used to model the Reynolds stress anisotropy eigenvalues. These approaches are limited because they cannot be applied to new flow configurations since they use shallow neural networks with one or two layers. Later on, Ling

et al. (2016b) developed a deep neural network (8–10 layers) able to predict the full anisotropy tensor. The same approach was used in other research works (Zhao et al., 2020). Alternatively, new terms (time and space-dependent) were added to the turbulence transport equations (mainly production) and are learned directly from data (Zhang and Duraisamy, 2015; Wang et al., 2017; Wu et al., 2018).

2.1.2. Application of deep learning in the thermal-hydraulic system codes

Several system thermal-hydraulics codes have been developed over the years to simulate the transient characteristics of nuclear power plants. These codes provide the best-estimate analysis of nuclear reactor transient conditions, and they are based on the conservation equations for two-phase flow which are resolved in Eulerian coordinates. Examples of such best-estimate thermal-hydraulic codes are RELAP (**R**eactor **E**xcursion and **L**eak **A**nalysis **P**rogram), TRAC (**T**ransient **R**eactor **A**nalysis **C**ode), ATHLET (**A**nalysis of **TH**ermal-hydraulics of **LE**aks and **T**ransients), etc. Traditionally, these codes operate in an iterative mode which requires the repetition of calculation, evaluation, and corrections steps. The entire process could be very tedious and takes a lot of computational time. The successful application of DL techniques in several fields has awoken serious interest in new data-driven innovations for the thermal-hydraulic design of nuclear reactor systems. For instance, the thermal-hydraulic parameters of the two key components (reactor core and tube-in-tube-through steam generator) of a nuclear reactor (KLT-40 S) were predicted using a DL algorithm (Lu et al., 2021a). The RELAP code was used to generate data for the training of the DL model and the prediction of the thermal-hydraulic parameters using the trained DL model was able to reproduce RELAP results in a very rapid manner. Many studies (Tian et al., 2018; Kim et al., 2019; Saeed et al., 2020) have been performed in recent times using thermal-hydraulic system code to generate data for the training of deep learning models that can be used for the diagnostics, fault-monitoring, and prediction of thermal-hydraulic parameters of nuclear reactors in a more efficient and timely manner.

In addition, the DL model can also be used to complement thermal-hydraulic system codes to improve their predictive accuracies. This has been demonstrated in previous work (Guillen et al., 2020) where a long-short term memory (LSTM) deep neural network model was used to complement RELAP code for the analysis of the failures of two drywell cooling fans in a nuclear power plant. The RELAP code was used to simulate four fan coil units (FCUs) each of which comprises a water-cooled heat exchanger and a centrifugal fan. The results (outlet temperature of the FCUs) of the RELAP simulation of the FCUs using the measured FCU inlet temperature as input was compared with that of LSTM-generated inlet temperature as input. The predicted outlet temperatures using measured FCU inlet temperature and LSTM-generated inlet temperature fall within 7.35% and 5.16% of the reported values by the plant process information system, respectively. Thus, the use of RELAP5-3D with the LSTM model provides a better physics-based anomaly detection model for the drywell FCUs.

The DL techniques can also be used for severe accident analysis in nuclear power plants. Severe accidents are nuclear accidents that are beyond design basis accidents (DBA) that involve substantial core damage due to melting. An example of this accident is the Fukushima nuclear power accident of 2011. Although, this kind of accident has a very low probability of occurrence recent event of Fukushima has shown that it can nonetheless happen. Thus, safety analysis of nuclear reactors is also performed under this kind of rare accident scenario. The progression of a severe accident encapsulates several complex interdependent phenomena (Kang et al., 2004; Henry and Fauske, 1993; Theofanous and Syri, 1997; Zhang et al., 2010; Amidu et al., 2021) and DL application to severe accident analysis could be very beneficial. For instance, Koo et al. (Do Koo et al., 2019) have used a deep neural network to predict the reactor vessel water level during a severe accident scenario in a nuclear power plant. The water level in the reactor vessel is one of the parameters that are monitored during a severe accident since it is directly connected to the reactor coolability and mitigation of core exposure. To train the deep neural network, modular accident analysis program (MAAP) code was used to simulate postulated loss of coolant accident (LOCA) at hot legs and cold legs, and steam generator tube rupture. This deep neural network was found to have small RMSE (root mean square error) in the predicted reactor vessel water level and superior performance over the cascaded fuzzy neural network was also observed. With this DL model, supporting information on reactor vessel water levels can be provided to plant operators to aid their management of the severe accident progression especially when the instrumentation signals from the nuclear power plant cannot be accurately acquired due to instrumentation damage under severe accident condition.

Furthermore, DL methods can also be used to aid nuclear power plant operators in the implementation of severe accident management guidelines (SAMG). This way, the decision-making by the operators as the severe accident progress can be expedited. For example, an operator support tool (OST) based on the DL technique has been proposed by Lee et al. (2020) to aid operators in decision-making. The data-driven OST model was trained using the data generated from dynamic probabilistic safety/risk assessment (DPSA/DPRA) of station blackout (SBO) of a pressurized water reactor. The data generation process combines three simulation codes: dynamic event tree (DET) generator, MELCOR (accident simulator), and radiological assessment system for consequence analysis (RASCAL). Subsequently, the data-driven OST was able to predict possible offsite doses at 2-mile and 10-mile radius for emergency response planning.

Notably acknowledging the significance of DL in the nuclear field, several ML/DL libraries have been integrated into a Risk Analysis Virtual Environment (RAVEN) framework, a tool developed by Idaho National Laboratory (Rabiti et al., 2021). The RAVEN code provides a modular and pluggable environment for different programming languages (python, c++, etc) and several thermal-hydraulic and severe accident codes such as MELCOR, RELAP7, MAAP5, TRACE, RATTLESNAKE, MAMMOTH, SCALE, etc (Alfonsi et al., 2022). In addition, RAVEN has interfaces for several ML/DL libraries but the one that is very relevant to the subject of this article is the TensorFlow-Keras Deep Neural Networks. Thus, with RAVEN, data-driven closure models can easily be deployed in many thermal-hydraulic and severe accident codes.

2.1.3. Application of deep learning in multiphase heat transfer phenomena

Prediction of multiphase heat transfer phenomena is another key component of the thermal-hydraulic analysis of a nuclear power system. Such phenomena are nucleate boiling (Hari and Hassan, 2002; Dhir, 2006; Podowski, 2012; Yoo et al., 2014; Amidu et al., 2018; Amidu, 2021), and critical heat flux (CHF) conditions (Katto, 1994; Liang and Mudawar, 2018; Devahdhanush and Mudawar, 2021; Lu et al., 2021b). These phenomena are germane to the determination of thermal design limits that need to be imposed to maintain the integrity of the cladding in the fuel rod. As an illustration, the CHF phenomenon results from a relatively abrupt deterioration of the heat transfer capability of the two-phase coolant, and the resulting thermal design limit is expressed in terms of the departure from nucleate boiling (DNB) for pressurized water reactors (PWRs) and critical power (CP) condition for boiling water reactors (BWRs). The significance of CHF has elicited extensive experimental and theoretical studies over several years. Many predictive tools for CHF, ranging from empirical correlations and look-up tables (LUTs) to physics-driven mechanistic models have been suggested in the literature. However, most of these empirical correlations are specific to experimental datasets from which they were derived and most of the time could not be used for different systems outside their range of validities. Thus, DL techniques can be leveraged to achieve better prediction of these phenomena instead of empirical correlations. The use of DFNN for the prediction of CHF has been presented in several relevant previous studies (Nafey, 2009; Cong et al., 2013; Moon et al., 1996; Su et al., 2002). Unfortunately, these studies neither differentiated DNB

from dryout when selecting data sources nor assessed their deep learning network architecture through cross-validation techniques.

To address the deficiency of previous application of DFNN to DNB, a mechanistic CHF model (based on the concept of liquid sublayer dryout and bubble crowding) was incorporated as the physics-informed component of the physics-informed machine learning-aided framework (PIMLAF) to obtain a superior prediction of DNB for a rod bundle (Zhao et al., 2021b). This way, DL leverages the domain knowledge in the field to capture the undiscovered information from the mismatch between the actual and domain knowledge predicted output. This hybrid framework (PIMLAF) was able to predict DNB in varieties of heater geometries covering a wide range of flow conditions without recalibration of the model. With this framework, experimental data or high-fidelity numerical data can be leveraged to achieve reduced thermal margin in the minimum DBN ratio (MDNBR) for reactor designs.

2.2. Reactor core neutronic, fuel loading optimization, and spent fuel storage

The reactor core neutronic analysis is customarily carried out using either high-fidelity approaches (such as method of characteristics, Monte Carlo method, and finite element method) or low-fidelity methods such as the nodal diffusion approach. The high-fidelity simulation of light water reactor neutronic requires high computing power and may take several days to complete. Whereas the low-fidelity methods do not need much computational capability as they are intended to run on personal computers but they provide low-level detail of the neutronic characteristics. For design optimization, this low-fidelity method (diffusion method) is conservative enough to estimate reactor core parameters at a low computational cost. Since the low-fidelity method is desirable for design optimization, DL techniques can be used to leverage a high-level detail provided by the high-fidelity method for the enhancement of the low-fidelity approach to achieve the best estimate of the core neutronic parameters. In light of this, a DL architecture based on a convolutional neural network (CNN) was proposed by Shrivel et al. (Shriver et al., 2021) to predict the normalized pin powers and k_{eff} of a two-dimensional reflective pressurized water reactor assembly model. In their study, the data generated and used for training the DL model was based on a 2-D hot-zero-power infinite fuel lattice with several pin geometries and materials compositions. A high-fidelity simulation was performed with VERA (virtual environment for reactor applications) code suits to obtain the training dataset. The DL model predicts the neutronic parameters with high accuracy (high-fidelity) but at a low computational cost. Similarly, Jinyoung and Yongduk (Jinyoung and Younduk, 2019) have used very deep CNN to predict assembly powers and individual pin powers by interpolating on pre-computed form factor tables which makes it different from Shrivel et al. (Shriver et al., 2021) approach where these high-fidelity features (individual pin powers) are predicted without any pre-computed libraries.

The deep learning techniques have also found application in the estimation of neutronic parameters of high-dimensional large nuclear reactors. Due to the heterogeneity of large nuclear reactor cores (e.g. boiling water reactors) in terms of the composition of the fuel, insertion of the control rod, and flow regimes, high order symmetry are not possible which estimates the neutronic parameters for large spaces of possible loading patterns difficult. This challenge can be resolved using the DL method. To this end, Saleem et al. (2020) have used a deep neural network trained and optimized using a combination of manual and Gaussian processes, for the prediction of neutronic parameters (power peaking factor, control rod bank level, and cycle length) of Ringhal-1 BWR unit. The training data were generated by Purdue Advanced Reactor Core Simulator (PARCS) code applied to the half-symmetry core by shuffling 196 fuel assemblies. The deep neural network performed creditably well with absolute errors of ~0.2, ~0.2, and ~0.5 for maximum, radially averaged, and axially averaged power peaking factors, respectively.

In addition, the application of the DL method has also been extended to the nuclear fuel transmutation in a fuel cycle as a flexible, quick, and medium-fidelity method to predict the PWR spent fuel composition with time-varying burnup and enrichment. The deep neural network was trained using the Used Nuclear Fuel Storage, Transportation & Disposal Analysis Resource and Data System (UNF-ST&DARDS) Unified Database (UDB) as the ground truth (Bae et al., 2020). The trained deep neural network was very quick taking about 0.27 s for 100 predictions with very good accuracy (1% error for used nuclear fuel inventory decay heat and 2% error for major isotopic inventory). A balance between high fidelity and speed of calculation can be reached with a well-trained deep neural network. This is crucial because most nuclear fuel cycle simulators using high-fidelity models face prohibitive computation expenses on one hand, while simpler and quicker methods do not provide high-level detailed calculations.

3. Demonstration of a deep learning model for RANS simulation optimization

To demonstrate the applicability of physics-separated DL algorithms for robust RANS simulation, this section presents a deep feedforward neural network (DFNN) model for the prediction of turbulent eddy-viscosity and its integration for optimal RANS simulation. The python framework used is Tensorflow 1.14 with C-backend, integrated into OpenFOAM 5.0. The physics-separated development framework for the DL-based closure model is used in this demonstration, as illustrated in Fig. 2 (Chang and Dinh, 2019). This demonstration also follows the 2D backward-facing workflow presented by Maulik et al. (2021), as shown in Fig. 3. The purpose of this demonstration is to show the improvement of DL-based RANS simulation, as opposed to solving an additional partial differential equation to close the Spalart Allmaras model in RANS simulation. A widely tested flow configuration passing over a backward-facing step is used to generate steady-state turbulent eddy viscosities data in RANS simulation using a one-equation Spalart Allmaras model as a closure. These data are subsequently used to train the DFNN. The equation of the Spalart Allmaras utilized can be found in Spalart and Allmaras (1992). To perform this RANS simulation, a solver (simpleFoam in OpenFOAM 5.0) is used in this study.

Using the configuration shown in Fig. 2, airflow over the backward step is simulated for Reynolds number (Re) determined by the step height (h) which is fixed at 1.27 cm, and the free stream velocity. The definition of the Reynolds number is given in Eq. (1).

$$Re = \frac{Uh}{v} \tag{1}$$

where U is the freestream velocity and v is the kinematic viscosity (with the value for air given as $1.5 \times 10^{-5} \mathrm{m}^2/\mathrm{s}$).

The DFNN has five input units u_x, u_y, x_c, y_c, and h representing horizontal velocity, vertical velocity, horizontal component, and vertical component of the cell-centered coordinates of the grid in the domain and the step height respectively. The DFNN also has five hidden layers and 60 neurons in each layer. A fully connected layer and a single output neuron representing the turbulent eddy viscosity form the model output. The network is trained with simulated data provided by Maulik et al. (2021). The data is generated from simulation of ten scenario with different free stream velocities of 40 m/s, 41 m/s, 42 m/s, 43 m/s, 44 m/s, 45 m/s, 46 m/s 47 m/s, 48 m/s, and 49 m/s corresponding to Reynolds number ranging between 34,000 and 41,500.

The DFNN is built using the ReLU activation function for both the input and hidden layers. The network is trained using Adam optimizer with a learning rate of 0.01 while the model convergence criterion is based on the Log-Cosh function which is the logarithm of the hyperbolic cosine of the prediction given by Eqn. (2).

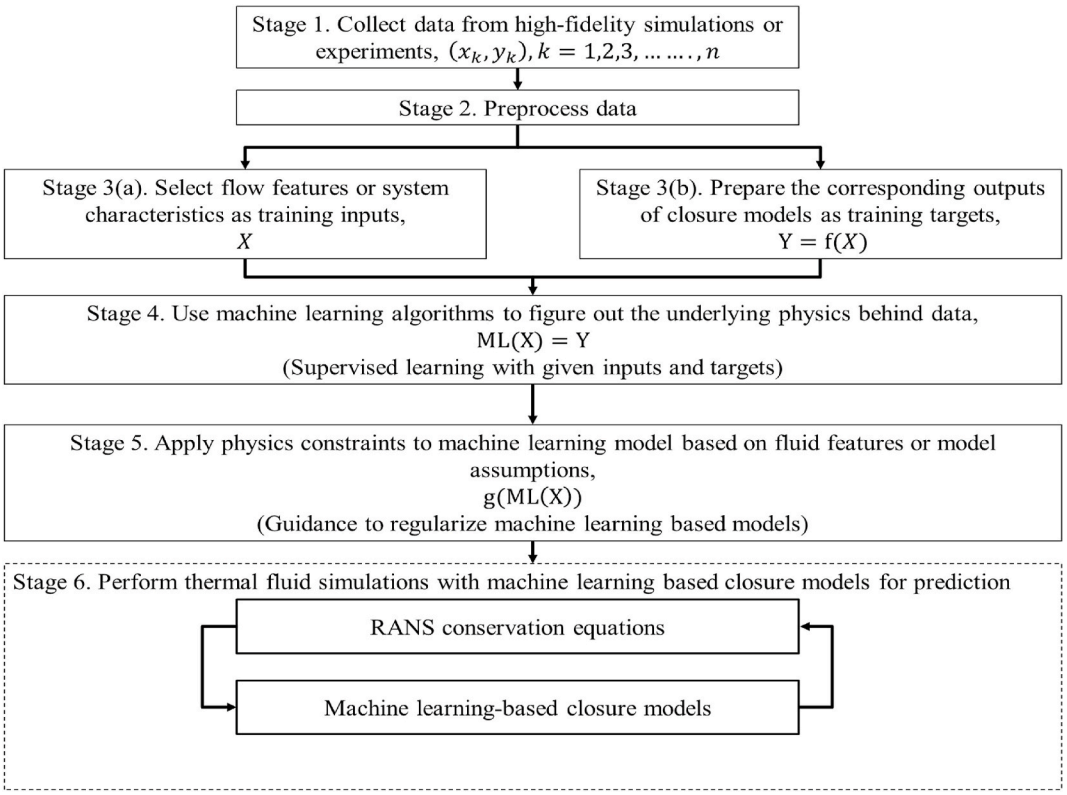

Fig. 2. Physics-separated DL framework (Chang and Dinh, 2019).

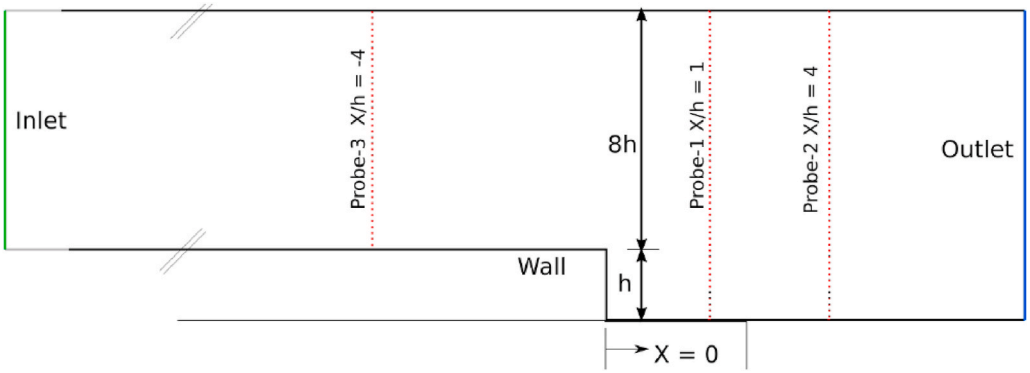

Fig. 3. Configuration of the 2D backward-facing step (Maulik et al., 2021), used for CFD simulation showing the boundary conditions and probe locations for data to assess the data-driven model.

$$Log - Cosh\ Loss = log\left[\frac{exp\ (y^{pred} - y^{truth}) + exp\ (y^{truth} - y^{pred})}{2}\right] \quad (2)$$

The generated data contained 205,390 samples and 90% of these samples were used as the training set while the remaining 10% of the samples were used for model validation. Additional out-of-sample test data is generated, with an inlet velocity condition of 44.2 m/s. Three probe locations shown in the simulation geometry (Fig. 3) are used to assess the prediction accuracy of DFNN on the test set.

The model converges after about forty-five epochs, which seems to be adequate for parameterizing the input-output relationship, and the R^2 value tends to 1. Further, a statistical assessment of the model is performed to underscore its effectiveness. The scatter plot shown in Fig. 4 (a) depicts that the predicted turbulent eddy viscosity by the DFNN reasonably agrees with the true magnitudes of the turbulent eddy viscosity, particularly for the higher magnitudes. Moreover, Fig. 4(b) shows a good agreement in the probability distribution plots for the predicted values and ground truth.

Analyzing the result of the DFNN model, it is observed that converged solutions of the steady-state solver were obtained after 17,800 iterations and 1990 iterations for the actual Spalart Allmaras model and DFNN model, respectively. That is, the iteration convergence is sped up by a factor of ~9 when DFNN is used. This shows that the DL model is significantly faster than the conventional Spalart Allmaras model. As expected, the DFNN model speeds up the simulation because the additional partial differential equation for the calculation of the turbulent eddy viscosity has been eliminated. Moreover, the gain in speed does not compromise the model's predictive accuracy as shown in Fig. 5 where the DFNN model predicted velocities profiles (left) and eddy viscosity fields (right) closely agree with the prediction using the actual turbulent model of Spalart Allmaras. Furthermore, a qualitative assessment also lends credence to the predictive accuracy of the DFNN model as shown in Fig. 6 and Fig. 7 for the respective contours of the velocity magnitude and turbulent eddy viscosity predicted by the DL

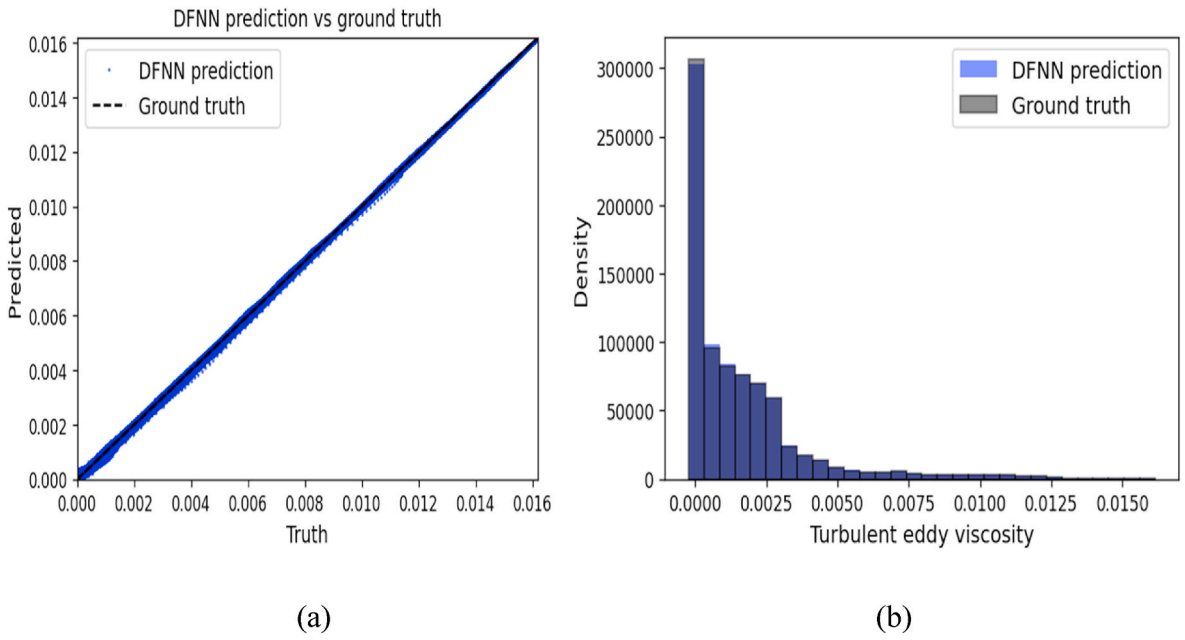

Fig. 4. Assessment of the trained DFNN model: (a) a plot of the DFNN prediction and the ground truth, and (b) probability density distribution.

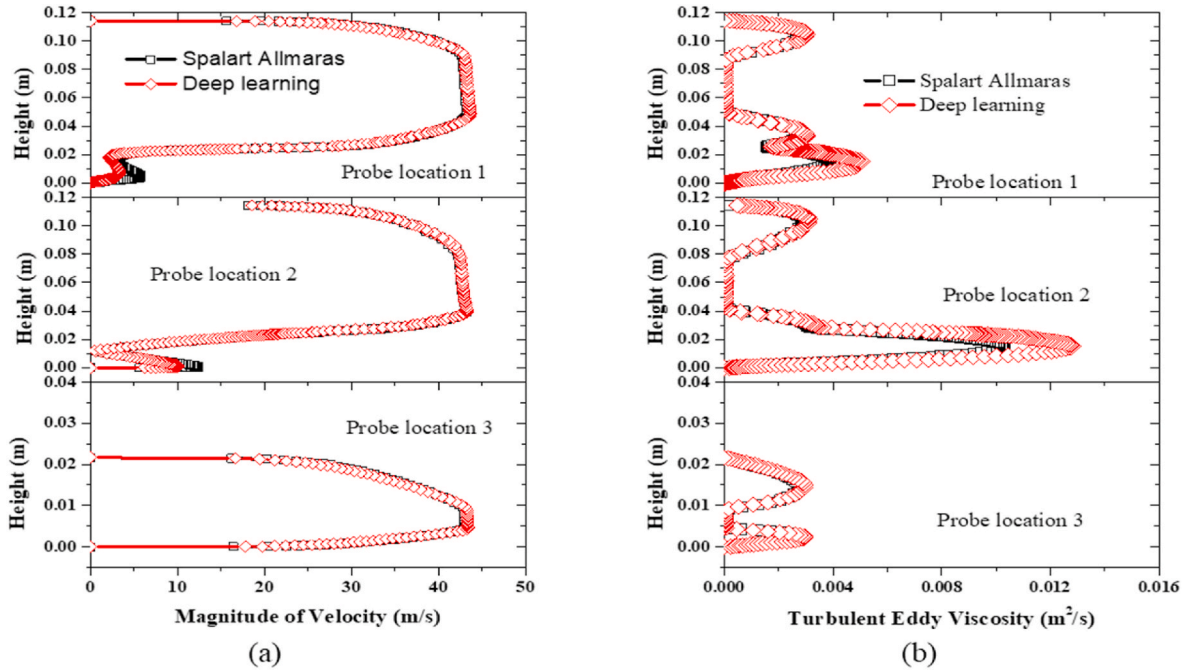

Fig. 5. Comparison of the DFNN model prediction and the actual model (Spalart-Allmaras model) prediction for the velocity profiles (a) and turbulent eddy viscosity fields (b) at three different probe locations (1, 2, and 3) shown in the simulation geometry.

model closely match those predicted by the actual turbulent (Spalart Allmaras) model. However, in this case, the DFNN inherits the shortcomings of the RANS model since it was trained using RANS simulation data.

4. Issues, challenges, and proposals for robust DL-based reactor safety analysis

DL applications have several advantages which prompt their rapid adoption in various fields. However, there are issues associated with the DL models which require careful attention and adequate resolution to obtain the ever-required, dependable, realistic, and reliable results from the application of DL in complex and safety-critical systems such as nuclear reactors.

Generally, modern DL algorithms still require a large curated benchmarked dataset to build a useable model. For many tasks, DL models are unscalable and rotation invariant. State-of-the-art DL algorithms also have limited ability to understand contextual information, thereby restricting their broader application. Moreover, save for advanced application of message-passing graph neural networks, many DL architectures are weak in tasks involving extreme generalization, making ontological inference, recognizing causal relationships, and learning abstract ideas and symbols (Tsimenidis, 2020).

Current DL algorithms are limited in what they can represent, and a

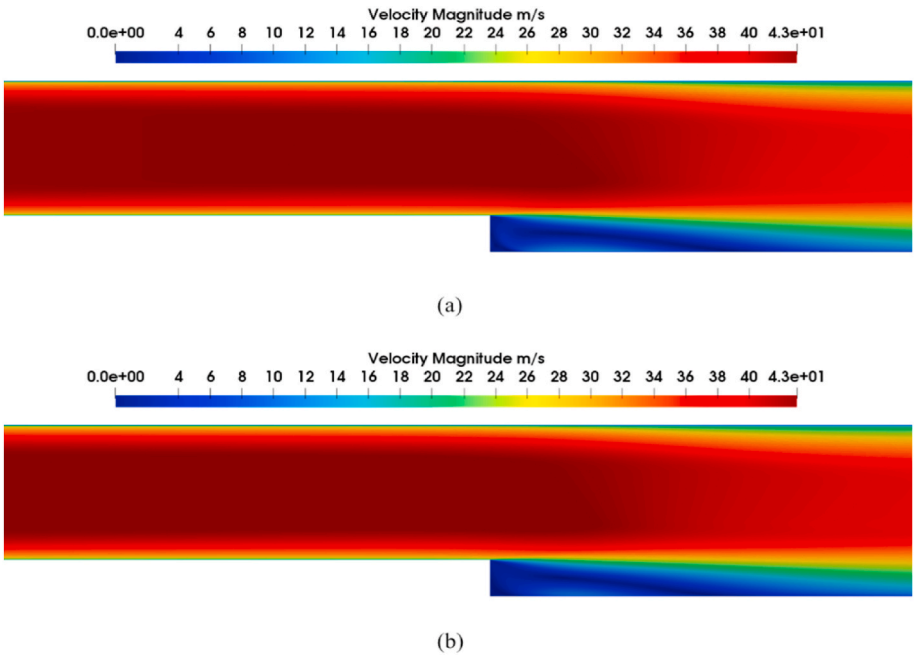

Fig. 6. Comparison of the predicted velocity contours: (a) by the DFNN model, and (b) by the actual Spalart Allmaras model.

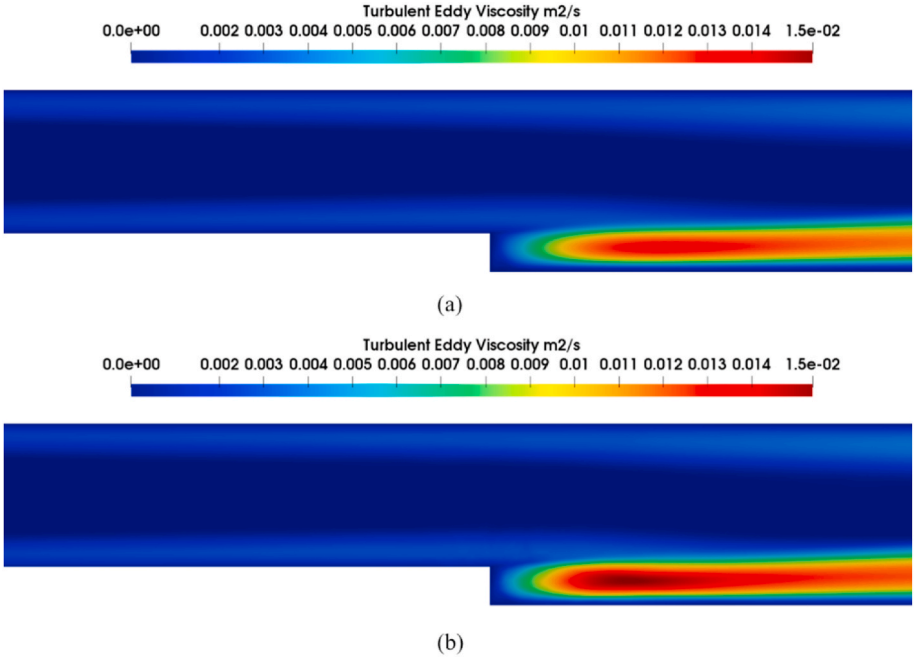

Fig. 7. Comparison of the predicted turbulent eddy viscosity contours: (a) by the DFNN model, and (b) by the actual Spalart Allmaras model.

bias/variance tradeoff is necessary to obtain a useful model. However, extreme tradeoffs are unacceptable in high-risk and highly regulated fields such as the nuclear industry. Moreover, there are specific issues with DL usage for nuclear safety analysis. This section discusses the issues and proposes solutions to the problems identified.

4.1. Dataset availability, sampling, and reconstruction

The key element that determines the optimum performance of any data-driven model is the quality of the data used in its development. Effective application of the DL approach to reactor safety requires significant training data. Also, proper implementation and reliability of the DL-based safety assessment approach depend on the sources and credibility of the dataset. Commonly used databases for DL-based reactor safety analysis are mostly derived from simulation and experimental testbeds, and a few other repositories provided by the International Energy Agency (IAEA).

To the best of the authors' knowledge, there is no robust open-source experimental dataset that could aid rapid DL-based reactor safety analysis. This is partly because nuclear data is subjected to export control bottlenecks, for security reasons. To obtain a useful nuclear dataset for data-driven modeling, researchers are actively utilizing integral and best estimate simulation codes (Ayodeji et al., 2018). Best-estimate codes are useful for the acquisition of high-fidelity multivariate time

series datasets used for modeling complex industrial systems, as demonstrated in (Ayodeji et al., 2021). However, simulated datasets face certain reliability challenges, which invariably constrain the generalization and utility of such datasets. First, the value and ultimate reliability of a particular simulated dataset depend on the subjective experience, acquisition framework, and application expertise of the modeler. Secondly, the static nature of simulated data may not reflect the composition and dynamics in real-world systems.

Several open-source simulation data for thermal fluid flow are available. However, the data are not standardized for DL model training, and there are no suitable benchmarked results to evaluate the performance of various DL architectures on the dataset. Although costly, experimental data from expert evaluators would still perform a significant role in providing curated input to DL models. This is supported by the fact that some repositories with simulated data are unlabeled and require significant preprocessing to be useful for model development. Considering the size of data needed for developing a robust DL algorithm, some repositories contain inadequate data size to train a robust DL model. Although this work does not attempt to compare the reliability of the available data, a casual study shows that the open-source simulated dataset is not sufficiently representative, and heterogeneity requirements are not defined. This needs to be addressed for the application of DL models to achieve its full potential, especially in tasks involving data-driven closure models in RANS simulation and other reactor safety analysis problems.

Another major performance controller in DL training is the data sampling and reconstruction method utilized, as the input to DL models are prone to uncertainties. Data sampling refers to the method used to select a (sub)set of a big dataset to ensure speedy model convergence and reduce computational cost. Data reconstruction refers to the method used in obtaining desired properties in a particular dataset. Data sampling and reconstruction are also necessary to correct imbalance distribution in dataset, for sensitivity analysis, and to reduce uncertainty. A number of data sampling methods such as the random sampling (i.e. grid sampling and Monte Carlo approach), cluster sampling, stochastic collocation approach and adaptive methods such as stratified sampling, importance sampling, Latin hypercube approach and their variants have been proposed to reduce sampling error and to ensure that the data subset used for training adequately represent the whole data. Detailed discussion of the sampling methods and data reconstructions approaches are discussed in (Katharopoulos and Fleuret, 2018). Also, the importance of data reconstruction for DL models is discussed in (Liu et al., 2021). Further, section 4.2 below discusses model reliability and uncertainty quantification approaches useful for improved DL model performance.

4.2. Model reliability and uncertainty quantification

The reliability concerns that emanate from insufficient data often prompts the use of simulated datasets and their associated static nature. There are also reliability issues caused by internal structural defects of DL with significant effects (Zhang and Xiao, 2020). Reliability issues could also cause a small transformation (rotations or translations) in the input to have a significant influence on the classification result (Azulay and Weiss, 2018). This section discusses the reliability and uncertainty concerns in DL models and the strategies for improving the reliability of the DL-based reactor safety analysis.

4.2.1. Reliability concerns

Reliable predictions through DL could greatly facilitate their utilization in safety-critical applications. An extensive study by Zhang and Xiao (2020) reveals that the reliability of DNNs has attracted serious attention due to potential structural defects in DL. Reliability translates to effective performance, and safe operation of complex systems (such as nuclear reactors and related installations) is crucial. Hence, the data-driven approach is increasingly being introduced to enhance the safety, reliability, and availability of nuclear systems. DL-based methods are now being applied to key nuclear safety fields such as systems health monitoring and control, radiation detection, and optimization (Gomez-Fernandez et al., 2020). Thus, the reliability of DL applications requires careful consideration.

Generally, there are many factors that affect complex systems' performances that are necessary for consideration in reliability analysis. Some of the factors include human reliability factors, equipment, materials, the environment, and the management procedures (Zhao et al., 2020). The issues associated with those factors are continuously being resolved through different strategies and approaches to which DL is also a significant contributor, as DL is currently the most advanced method in artificial intelligence (Szegedy et al., 2013).

Three primary factors that often affect the reliability of the DL models are (Krizhevsky et al., 2012):

- the input-output mapping discontinuity caused by the deep neural network model's nonlinearity.
- the over-fitting caused by the inefficient model training; and
- insufficient regularization.

Different external contributors to reduce the reliability of DL models are extensively discussed in Zhang and Xiao (2020) which include adversarial attacks that should be carefully studied. Details about the different types of attacks and the possible mitigating approaches were also discussed in Zhang and Xiao (2020). Several novel contributions are being made to improve the reliability of DL models and frameworks over time. Among them is the study by Zhang et al. (2019) which proposed a new class of reliability analysis methods that needs no simulation data. The method is based on recent advances in DL and adopts tools such as automatic differentiation. With the method, the unknown response is represented by a DL model and a physics-informed loss function can be used to obtain the unknown parameters. With the loss function, there is no need for training data and the neural network parameters are directly computed through the ordinary differential equation and partial differential equation (ODE/PDE) that describe the system. Because simulation data is not required, the computational cost associated with reliability analysis is greatly reduced. In addition, since the network parameters are trained by using a physics-informed loss function, the neural network solutions observe the physical principles (e.g., the conservation laws). Furthermore, the method provides prediction at every spatial and temporal location which makes it very appropriate for handling time-dependent reliability analysis.

4.2.2. Uncertainty quantification

4.2.2.1. Uncertainty quantification: sources and types of uncertainty. A mismatch of the test and training data is the main source of uncertainty in DL, while data uncertainty occurs because of the class overlap or the presence of noise in the data. However, estimating knowledge uncertainty is much more difficult than estimating data uncertainty. In general, two broad types of uncertainty exist which are aleatory and epistemic uncertainties (Hüllermeier and Waegeman, 2021). The irreducible uncertainty in data that gives rise to uncertainty in predictions is the aleatory uncertainty which is also referred to as data uncertainty. Aleatory uncertainty is not associated with the characteristics of the model, but rather it is an inherent property of the data distribution, and thus, irreducible. In contrast, epistemic uncertainty (also known as knowledge uncertainty) occurs due to inadequate knowledge. More detailed background, discussion, and insights into these two broad groups of uncertainty were given in a generalized form to engineering systems in Zhang and Olatubosun (2019) and Abdar et al. (2021). The predictive uncertainty, as applicable to DL, is therefore a combination of the two broad categories of uncertainty, viz:

Predictive uncertainty (*PU*)

$$= Epistemic\ uncertainty(EU) + Aleatory\ uncertainty\ (AU) \tag{3}$$

Any prediction made without uncertainty quantification usually lacks integrity and thus, is not dependable (Abdar et al., 2021). The uncertainty quantification in DL-based applications is thus essential as the various uncertainty quantification methods adopted in conjunction with different DL techniques can significantly raise the integrity of their results (Abdar et al., 2021). A more robust framework for classifying the sources of uncertainty based on a detailed review of uncertainty analysis in different fields was given by Radaideh and Kozlowski (2019). The classification framework is generic, concise, and captures most of the uncertainties applicable in nuclear engineering and reactor safety specifically. Based on the classification, five major categories were identified and discussed in detail (Radaideh and Kozlowski, 2019), which are now summarized using Table 2. Furthermore, based on the different sources of uncertainty (as highlighted in Table 2), equation (3) can be explicitly defined as:

$$PU(as\ applicable\ to\ DL) = EU\begin{pmatrix} surrogate\ uncertainty, \\ model-form\ uncertainty, \\ model\ deficiency \end{pmatrix} + AU\begin{pmatrix} Parametric\ and\ input\ uncertainty, \\ output\ uncertainty \end{pmatrix} \tag{4}$$

The essential literature gaps and open issues in the application of uncertainty quantification methods are extensively discussed in Abdar et al. (2021) and further related gaps are presented in (Mashlakov et al., 2021).

Table 2
Generalized uncertainty sources.

	Source of uncertainty	Category of uncertainty	Brief description	Examples
1	Parametric and input uncertainty	Aleatory uncertainty	The most common and most analyzed uncertainty. The sub-types are explanatory parameters and model parameters.	Material properties, calibration parameters, code or user inputs, physical constants, etc.
2	Output/ observation uncertainty	Aleatory uncertainty	Experimental or measured uncertainty of the output	Measured or experimental uncertainty
3	Interpolation/ surrogate uncertainty	Epistemic uncertainty	This uncertainty arises when the real model is substituted by a reduced-order model basically to reduce the high computational costs of the original model.	Reduced-order models, Gaussian process (GP) surrogates, etc.
4	Uncertainty due to model deficiency or discrepancy	Epistemic uncertainty	Referred to as model discrepancy or model deficiency. It arises due to approximations and assumptions.	Model assumptions, inadequate physics knowledge, etc.
5	Model-form uncertainty	Epistemic uncertainty	The least common source of uncertainty in both engineering modeling and specifically, nuclear reactor modeling. It is a form of a multi-dimensional version of the predictive uncertainty.	Model averaging, model selection, etc.

4.2.2.2. Uncertainty quantification of deep learning models and methods. The results (predictions) obtained from the application of DL often fail to provide the reliability or confidence level of such predictions despite the several advantages associated with the DL-based approaches, especially when applied to proffer solutions to practical problems (Radaideh and Kozlowski, 2019). To resolve this issue, interpretation of the model parameters is often carried out using Bayesian deep learning (BDL) and Bayesian neural networks (BNNs) (Foong et al., 2019). Both BNNs and BDL are resourceful in handling overfitting problems and can be trained on both small and large datasets (Kucukelbir et al., 2017).

Since wrong predictions can be disastrous in critical applications such as nuclear power reactors, it is pertinent to properly handle uncertainty quantification, especially for practical applications (Lakshminarayanan et al., 2016). The integrity of predictive uncertainty evaluation is tasking as there is no empirical evidence of uncertainty estimates in general in most cases (Abdar et al., 2021). Calibration and domain shift are the two common uncertainty evaluation measures often applied and are usually inspired by the practical applications of neural networks. Calibration measures the discrepancy between long-run frequencies and subjective forecasts. Another thought generalizes predictive uncertainty to a domain shift that estimates whether the network under study knows what it knows (Abdar et al., 2021).

Uncertainty quantification methods, therefore, help in drastically reducing the effect of uncertainties, especially in optimization and decision-making in most fields. Bayesian approximation and ensemble learning techniques are two widely adopted types of uncertainty quantification methods. Abdar et al. (2021) presented in detail the recent advances in the uncertainty quantification methods used in DL, investigates the application of these methods in reinforcement learning and highlights fundamental research challenges and future research directions in uncertainty quantification.

Several other novel contributions have been made to address the issues of uncertainty quantification in DL. An extensive comparison of the common methods for uncertainty quantification was made in Caldeira and Nord (2020) in which BNNs, concrete dropout (CD), and deep ensembles (DEs) are compared to the standard analytic error propagation. The issues with the adoption of these methods were highlighted which include when the variation of noise in the training set is small, all methods predicted the same relative uncertainty irrespective of the inputs. This issue is particularly hard to avoid in BNN. On the other hand, when the test set contains samples far from the training distribution, no methods sufficiently increased the uncertainties associated with their predictions. This issue was particularly glaring with CD. The results obtained by Caldeira and Nord (2020) informed the following recommendations for handling uncertainty issues in DL-based reactor safety analysis:

- To obtain an accurate estimate of aleatoric statistical uncertainty, there must be a sufficiently wide variation of the noise present in the training set. This will prevent the model from being stuck by avoiding the prediction of the same relative uncertainty for all points.
- For systematic uncertainties, the model can only infer the typical uncertainty in each region of inputs from the training set, since the uncertainty cannot be statistically derived from the inputs.
- For epistemic uncertainties, all methods failed to detect the extents to which the inputs had moved from the training distribution. Specifically, CD converged to a very low dropout probability in training, which makes it capable of predicting very low epistemic uncertainties. While DE and BNN could be used to detect out-of-distribution data, their quantitative estimates of epistemic uncertainty are not reliable for that case.
- DE was recommended as the best approach as its results are comparable to the best in all the experiments, which also corroborates the existing findings (Ovadia et al., 2019).

4.3. Model output explainability

DL model explainability refers to the ability of the model to relate the feature values to the predicted output for a particular instance in an interpretable manner. Explainability defines the model characteristics that enable interpretable output, and improved trust in the model prediction. Moreover, the unexplained model output results in an incomplete problem formalization, creating a fundamental barrier to optimization and evaluation (Doshi-Velez and Kim, 2017). Current explainability approaches are divided into two: gradient-based and perturbation-based. The working principles, computational complexity, and the essential properties of interpretability in the DL model are discussed in Robnik-Šikonja and Bohanec (2018).

In the reactor safety analysis, the importance of explainability is evident. First, the high-risk, safety-critical nuclear environment demands proven techniques, as the consequences of wrong predictions are significant. Moreover, implementing black box models would inherently constrain the DL applications in practical reactor analysis. Also, technologies based on black-box models may delay reactor licensing approval. Besides, explainability is critical in quantifying the uncertainties and failure modes of DL models. Considering the complexity involved in reactor development, and the strong coupling of multiple codes for safety analysis, interpretable DL modeling is critical to understanding the coupling effect on other components and for effective safety verification.

To improve trust in model outputs, shapely additive explanations (SHAP) (Lundberg and Lee, 2017), and local interpretable model-agnostic explanations (LIME) (Ribeiro et al., 2016), are two commonly-used tools. SHAP's capacity to explain the predictors' relationship with the target (global interpretability) and LIME's value-based explanation of prediction from each predictor (local interpretability) is important to understand local phenomena in the reactor safety analysis. The global interpretations provided by SHAP explain the input features and the entire model relationship with the prediction, which is an approximation. The LIME's local interpretation explains the model predictions for a single observation or groups of observations from any ML model. The two tools incorporate advanced visualizations that aid model interpretability, a critical characteristic in reactor safety applications. To demonstrate the importance of advanced explainability tools in DL-based reactor safety analysis, sections 4.3.1 and 4.3.2 below discuss the global and local interpretation of the output from the DFNN surrogate model presented in section 3.

4.3.1. Feature importance estimation with shapely additive explanation (SHAP) metric

Shapely Additive Explanation (SHAP) is a python-based library that uses shapely values in cooperative game theory. The API is used to define the optimum contribution of each feature to the marginal deviation in model output. The library is used to interpret black-box models for which predictions are known. The explainers in the SHAP library are one of the most common tools for a global and local explanation of each feature's contribution to the model prediction, although SHAP computes all permutations globally to get local accuracy. The SHAP library has been used to interpret DL model output in both classification and regression tasks, and the explanation is specified as:

$$g(z^{'}) = \varnothing_o \sum_{j=1}^{M} \varnothing_j \tag{5}$$

Where g is the explanation model, $z^{'} \in \{0,1\}^M$ is the simplified features, M is the maximum coalition size and $\varnothing_j \in R$ is the feature attribute for input j. A detailed description of the SHAP library can be found in Lundberg and Lee (2020) and Molnar (2020).

This section utilizes the SHAP tool to rank the features that contribute to the DFNN predictions in the RANS closure surrogate model presented in section 3. The demonstration utilizes a similar framework used to develop the model obtained in section 3. Then SHAP functions are used to estimate the contributions of the horizontal velocity, vertical velocity, horizontal component, vertical component, and the step height to the model prediction. Because of the training set size, the dataset was summarized with a set of weighted average values. For global interpretation, the SHAP *KernelExplainer* takes the trained model and the summarized training set as input. The output is the explainer that serves as the input to estimate expected SHAP values. Thereafter, the *shap.summary_plot* function is used to visualize all the DFNN model predictions in the test set. The SHAP summary plot describing the feature importance is shown in Fig. 8.

Fig. 8 above ranks the importance of each feature to the model predictions. The Figure shows that the most important feature for predicting turbulent eddy viscosity is the vertical components (y_c) and the horizontal components (x_c) respectively. This is closely followed by the horizontal velocity (u_x), the step height (h) and vertical velocity (u_y). The demonstration shows the effect of the vertical and horizontal components in turbulent eddy viscosity prediction, useful for improved RANS closure modeling, and to further the understanding of the thermal-hydraulic phenomena in the reactor core.

4.3.2. Local prediction explanations with LIME API

Feature importance estimated with SHAP in section 4.3.1 explains the global contribution of each feature to the DFNN predictions. However, it does not give insight into the average direction that a feature affects a prediction for a particular observation. To further explain the DFNN model prediction, the local interpretable model-agnostic explanations (LIME) API is used to interpret the local prediction of the DFNN model. LIME explains model prediction through local approximation with an interpretable surrogate model, by perturbing data around an individual prediction to build the model.

Similar experimental settings and frameworks used for the SHAP demonstration in section 4.3.1 are retained. First, the DFNN model is used to predict the first one hundred inputs in the test set. This is used to specify the observation for which explanation is required. Then, the LIME distinguishes an interpretable representation from the original feature space used by the DFNN. This is achieved with LIME's *LimeTabularExplainer* function, an explainer that takes the training set and feature names as key input parameters. Then the function *explainer.explain_instances* is used to explain a particular observation in the test dataset to get their probability values for each prediction. Then LIME assigns probability, provides an explanation as to the reason for assigning the probability, and compares the probability values to the actual value of the target variable for that prediction.

The first prediction in the test set is used as the observation of interest to demonstrate the tool and for a legible explanation. LIME perturbs the data by sampling from $N \in (0,1)$ and scaling the data according to the means and standard deviations in the training set. Fig. 9 and Fig. 10 show how LIME ranks the features according to their contribution to the observed local prediction at indices 200 and 5000 respectively. For the 200th local variable (index-200), the LIME output explains each feature contribution to or against the observed local

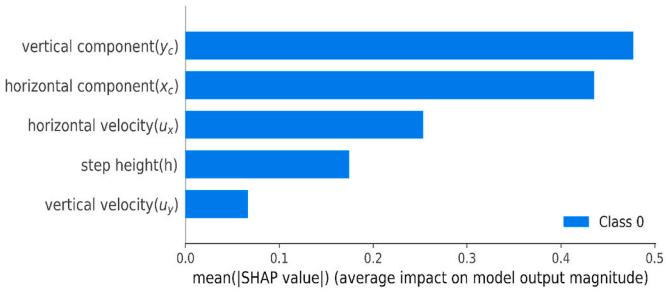

Fig. 8. The DFNN input feature importance, estimated with SHAP.

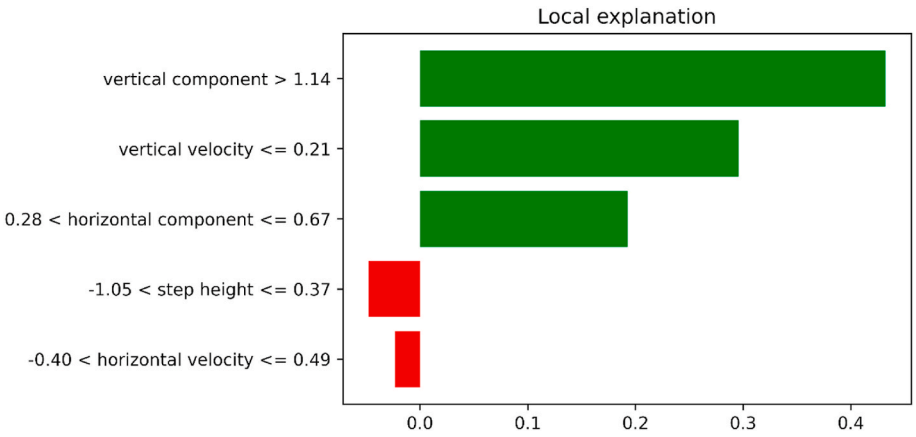

Fig. 9. Lime model prediction bar chart interpretation for a local variable at index 200.

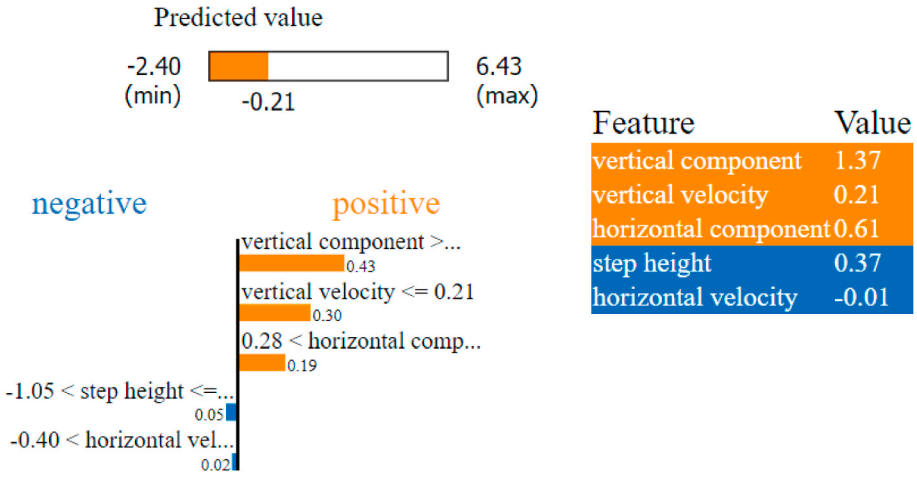

Fig. 10. Lime model prediction interpretation for a local variable at index 200.

prediction. The explanation is visualized as a horizontal bar with feature importance values. For a particular local variable, the feature that increases the prediction probability is indicated by a green bar to the right, with a positive feature value, while the feature that decreases the prediction probability is indicated by a red bar to the left, with a negative feature value. Fig. 9 shows a discretized feature with the corresponding range and the feature importance value. The Figure shows that for the predicted local variable at index 200, the vertical component has the highest influence on the prediction, followed by the vertical velocity, etc.

Fig. 9 also shows that the vertical component, vertical velocity, and the horizontal component positively correlate with the target (turbulent eddy viscosity), while the step height and horizontal velocity have a negative correlation. Fig. 10 shows with colored bars the discretized feature and their weights. Fig. 10 also explains the influence of each feature on the model's local prediction. For instance, the step height < -1.05 shows that its value negatively correlates with turbulent eddy viscosity, and the corresponding weight for the feature is 0.37. This means

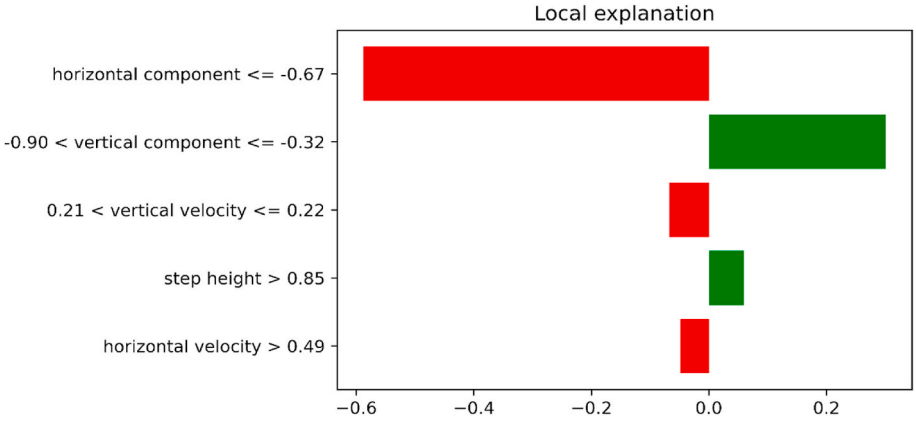

Fig. 11. Lime model prediction bar chart interpretation for a local variable at index 5 k.

if the horizontal component value is $\geqslant 1.05$, on average, this prediction would be 0.37 less negative. The table on the bottom-right contains each input feature and its corresponding original values. Similarly, Figs. 11 and 12 explain the local prediction at index 500. In Fig. 11, the horizontal component ranks higher than all other features. It is also observed that the horizontal component, vertical velocity, and horizontal velocity all have a positive influence on the value of the local variable at index 5000. Apart from interpreting results from surrogate closure models, LIME and SHAP are critical tools useful for explaining different outputs in DL-based reactor safety analysis. The tools are also useful for the assessment of relationships between experimental features and nuclear metadata, to capture systematic bias and outliers in nuclear data, among others.

5. Research opportunities/future direction

Although modern DL architectures have the potential to improve reactor safety analysis, their proper application requires not only a well-curated dataset, but also a deep understanding of the domain, and a knowledge-driven representation of the problem. Moreover, the common drawbacks of DL – such as its inability to produce dynamic representation and flexible inference mechanism – also provide future direction and research opportunities. For instance, the explainability tools presented above are interesting solutions to the black box problem of DL. Also, the legacy reactor analysis codes embody generations of work from experts and renewing the codes and incorporating DL-based solutions is an exciting research opportunity. Other research opportunities and potential future directions are discussed in this section.

Over the years, empirical or semi-mechanistic correlations derived from experimental observations have been used as closure relations in the RANS model. However, the traditional method of developing these empirical correlations is expensive. Also, the resultant closure term may be limited by the conditions and configurations of the experiments from which they are derived. These deficiencies hamper the application of the computational fluid dynamics methods in simulating new systems conditions and configurations. For researchers with access to big experimental/simulation data, DL models are feasible substitutes. DL algorithms could reveal the functional relations of features in big experimental data. The resulting DL closure model can be employed instead of the traditional empirical or semi-mechanistic closure approach. For instance, turbulence remains one of the oldest physics problems not yet solved or well understood, due to the nonlinear behavior of the governing (Navier-Stoke) equations. A wide range of scales (spatial and temporal) are generated during a turbulent flow, and they present the major barrier to solving the fluid flow problem either experimentally or numerically. DL models can reconstruct fully resolved turbulent flows, optimize the simulation and replace conventional

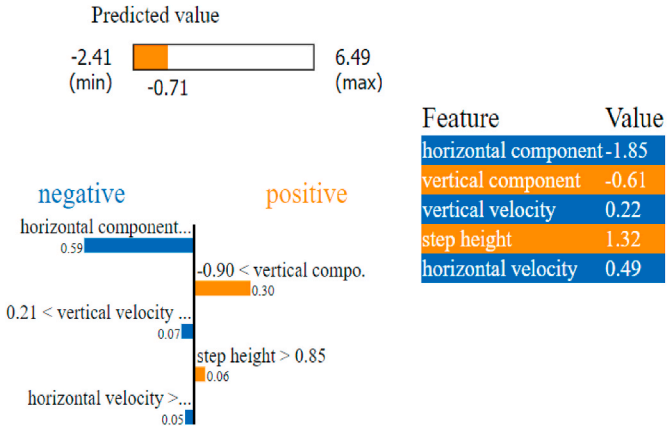

Fig. 12. Lime model prediction interpretation for a local variable at index 5 k.

model tuning and calibrations. Additionally, DL models can be used to determine optimum selections of mesh size and models in coarse mesh CFD codes to reduce the errors due to simplifying assumptions and mathematical approximation.

Further, research that utilizes DL algorithms to predict the properties of appropriate sacrificial materials for in-vessel core catchers would provide an exciting approach to severe accident management. DL models can also be utilized for uncertainty quantification and Bayesian system updating by developing a learning-based parametric model to learn functional relations using the aleatory and epistemic parameters as inputs. DL algorithms can also be utilized to predict parameters of a Bayesian probability distribution, detect bias, and identify outliers in both experimental and simulated nuclear data. The data validation capability of the DL algorithm provides recursive advantages. That is, DL algorithm output could be applied to detect imperfections in the input deck, replace manual data comparison, provide cost-effective nuclear data evaluation, detect discrepancies in differential experimental data, and produce a robust dataset for advanced DL modeling. DL algorithms are also valuable for criticality benchmarking and to maximize sensitivity with the least code execution.

DL algorithms are also useful in estimating the effective neutron multiplication factor (K_{eff}) and fission mass yield of nuclear materials with full posterior distributions without making assumptions about the shape of the posterior distribution. Many K_{eff} catalogs are known to have biased inputs and DL algorithms can predict the K_{eff} bias in the reactor of interest. In addition, DL models can be used to study the spatio-temporal behavior of neutrons in the reactor core. The attributes of message-passing graph neural networks can be leveraged to model fixed source and criticality neutron transport, and to study effective neutron distribution in complex geometries. DL models can also be used to optimize both deterministic and probabilistic neutron transport codes, using the framework provided in section 3.

Perturbation localization in nuclear reactors is another exciting application of DL algorithms. DL models can also be developed to detect fluctuations in neutron flux, and to identify, localize and size fluctuation in the reactor core. DL algorithms can be used to predict neutron energy, detect effective fission cross-section for new fuel materials, measure gamma spectrum, and predict in-core radioactive material activities for channel and energy ranges where data is not available.

DL models are also useful in detecting resonance parameters and covariance in experimental nuclear data, fitting and analyzing inconsistent data, and correcting experimental uncertainties. Moreover, DL models can also be used for verifying and benchmarking nuclear data. As shown in other works with shallow models (Neudecker et al., 2020), DL models can be used to deepen the data verification and benchmarking frontier. Given sufficient experimental data, DL algorithms can also estimate resonance parameters for heavy elements, and predict the level densities (Dwivedi, 2019). DL algorithms can also be applied to optimize reactor fuel assembly. As demonstrated by Radaideh et al. (2021b), DL-based optimization tools such as the NeuroEvolution Optimization with Reinforcement Learning (NEORL) can be used to optimize fuel cell design and reactor control. The diverse algorithms in the NEORL tool are also important for broader performance evaluation of different algorithms on a specific optimization task. Further, as demonstrated by Whyte and Parks (2021), DL models could be used to optimize the power peaking factor (PPF) and to determine the position of the hottest pin at the beginning of the cycle (BOC) in a PWR core. DL algorithms can also be used for loading pattern optimization and other fuel management issues.

To enhance scalability, robustness and explainability in DL models for nuclear reactor safety, causality also needs to be considered. The conventional DL approach learns the correlation in training data to make predictions. Moreover, traditional DL assumes independent and identically distributed input. However, the learned correlation may be spurious or unstable, and the training data may be out of distribution. Hence, abstract representation (causal) learning is useful to properly

capture the casual information in the measured or observed reactor parameters used as training data. Moreover, leveraging causality would aid the integration of reactor domain knowledge, which is critical for better prediction. Besides, sampling bias problems can be reduced and prediction reliability can be significantly improved, to obtain a robust predictor that can generalize well in real-world analysis. Lastly, integrating unique features of causal models, such as latent structural learning and improved knowledge representation, into safety analysis of nuclear reactors would present greater utility beyond the conventional DL approach.

6. Conclusion

DL applications in complex systems have many advantages, which prompt their rapid adoption in various fields. A DL algorithm eliminates over-reliance on domain knowledge, does not require manual feature extraction, offers an end-to-end learning process from raw data, and automatically learns hierarchical representations in large-scale data. It is thus an innovative and robust tool for optimized safety analysis from high volume and multi-dimensional nuclear reactor data. This has prompted its adoption in both numerical and analytical reactor safety research and development. However, obtaining dependable, realistic, and reliable safety analysis results from DL models requires careful attention to some issues.

This paper presents the state-of-the-art deep-learning application in the nuclear reactor safety analysis and the critical issues and challenges with the model. This work also proposes robust solutions to the identified challenges, toward a safe and efficient reactor development and safety analysis. The reactor safety optimization with a DL-based surrogate model demonstrated in this work, and the explainability tools used to interpret the model output are important contributions that support an open, robust, and explainable DL application. The discussion in this work would also support the development of effective DL-based modeling tools and methods with a significant reduction in associated uncertainties. The literature reviewed in this work, and the DL implementation framework provided would also guide researchers and engineers in applying DL models for independent verification and interpretation of safety analysis results. Moreover, the open research challenges, opportunities, and future direction itemized would guide advanced research in DL applications for reactor safety. Considering the complexity and safety requirements of nuclear reactors, regulatory approval for safety analysis performed with innovative tools may be problematic. Implementing the recommendations in this work could improve the trustworthiness and explainability of DL-based safety analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The work of AA and HA are funded through the Sêr Cymru II 80761-BU-103 project by Welsh European Funding Office (WEFO) under the European Development Fund (ERDF). MAA, SAO and YA received no funding for this work.

References

Abdar, M., et al., 2021. A Review of Uncertainty Quantification in Deep Learning: Techniques, Applications and Challenges. Information Fusion.
Ackoff, R.L., 1989. From data to wisdom. J. Appl. Syst. Anal. 16 (1), 3–9.
Alfonsi, A., et al., 2022. Risk analysis virtual ENvironment for dynamic event tree-based analyses. Ann. Nucl. Energy 165, 108754.
Amidu, M.A., 2021. Toward mechanistic wall heat flux partitioning model for fully developed nucleate boiling. J. Heat Tran. 143 (11), 114503.
Amidu, M.A., Jung, S., Kim, H., 2018. Direct experimental measurement for partitioning of wall heat flux during subcooled flow boiling: effect of bubble areas of influence factor. Int. J. Heat Mass Tran. 127, 515–533.
Amidu, M.A., Addad, Y., Riahi, M., 2020. A hybrid multiphase flow model for the prediction of both low and high void fraction nucleate boiling regimes. Appl. Therm. Eng. 178, 115625.
Amidu, M.A., et al., 2021. Investigation of the pressure vessel lower head potential failure under IVR-ERVC condition during a severe accident scenario in APR1400 reactors. Nucl. Eng. Des. 376, 111107.
Ayodeji, A., Liu, Y.-k., 2019. PWR heat exchanger tube defects: trends, signatures and diagnostic techniques. Prog. Nucl. Energy 112, 171–184.
Ayodeji, A., Liu, Y.-k., Xia, H., 2018. Knowledge base operator support system for nuclear power plant fault diagnosis. Prog. Nucl. Energy 105, 42–50.
Ayodeji, A., et al., 2019. Acoustic signal-based leak size estimation for electric valves using deep belief network. In: 2019 IEEE 5th International Conference on Computer and Communications (ICCC). IEEE.
Ayodeji, A., et al., 2021. Causal augmented ConvNet: a temporal memory dilated convolution model for long-sequence time series prediction. ISA Trans. 123, 200–217. https://doi.org/10.1016/j.isatra.2021.05.026.
Azulay, A., Weiss, Y., 2018. Why Do Deep Convolutional Networks Generalize So Poorly to Small Image Transformations? arXiv preprint arXiv:1805.12177.
Bae, J.W., et al., 2020. Deep learning approach to nuclear fuel transmutation in a fuel cycle simulator. Ann. Nucl. Energy 139, 107230.
Bao, H., et al., 2018. Study of data-driven mesh-model optimization in system thermal-hydraulic simulation. In: ANS Winter Meeting.
Bao, H., et al., 2019. A data-driven framework for error estimation and mesh-model optimization in system-level thermal-hydraulic simulation. Nucl. Eng. Des. 349, 27–45.
Caldeira, J., Nord, B., 2020. Deeply uncertain: comparing methods of uncertainty quantification in deep learning algorithms. Mach. Learn.: Sci. Technol. 2 (1), 015002.
Chang, C.-W., Dinh, N.T., 2019. Classification of machine learning frameworks for data-driven thermal fluid models. Int. J. Therm. Sci. 135, 559–579.
Chu, W., et al., 2021. Study on measure approach of void fraction in narrow channel based on fully convolutional neural network. Nuc. Power Plant Equip. Prognostic Health Manag. Based Data-Driven Methods 1, 300.
Cong, T., et al., 2013. Applications of ANNs in flow and heat transfer problems in nuclear engineering: a review work. Prog. Nucl. Energy 62, 54–71.
Devahdhanush, V., Mudawar, I., 2021. Critical heat flux of confined round single jet and jet array impingement boiling. Int. J. Heat Mass Tran. 169, 120857.
Dhir, V.K., 2006. Mechanistic Prediction of Nucleate Boiling Heat Transfer–Achievable or a Hopeless Task?.
Do Koo, Y., et al., 2019. Nuclear reactor vessel water level prediction during severe accidents using deep neural networks. Nucl. Eng. Technol. 51 (3), 723–730.
Dong, Z., et al., 2020. Multilayer perception based reinforcement learning supervisory control of energy systems with application to a nuclear steam supply system. Appl. Energy 259, 114193.
Doshi-Velez, F., Kim, B., 2017. Towards a Rigorous Science of Interpretable Machine Learning arXiv preprint arXiv:1702.08608.
Duan, H., Luo, X., 2021. A novel multivariable grey prediction model and its application in forecasting coal consumption. ISA Trans. 120, 110–127. https://doi.org/10.1016/j.isatra.2021.03.024.
Dwivedi, N.R., 2019. Trees and Islands–Machine Learning Approach to Nuclear Physics arXiv preprint arXiv:1907.09764.
Foong, A.Y., et al., 2019. On the Expressiveness of Approximate Inference in Bayesian Neural Networks arXiv preprint arXiv:1909.00719.
Gao, Z., et al., 2020. Multitask-based temporal-channelwise CNN for parameter prediction of two-phase flows. IEEE Trans. Ind. Inf. 17 (9), 6329–6336. https://doi.org/10.1109/TII.2020.2978944.
Gao, Z.-K., et al., 2021. A novel complex network-based deep learning method for characterizing gas–liquid two-phase flow. Petrol. Sci. 18 (1), 259–268.
Gomez-Fernandez, M., et al., 2020. Status of research and development of learning-based approaches in nuclear science and engineering: a review. Nucl. Eng. Des. 359, 110479.
Guillen, D.P., et al., 2020. A RELAP5-3D/LSTM model for the analysis of drywell cooling fan failure. Prog. Nucl. Energy 130, 103540.
Hanna, B.N., et al., 2020. Machine-learning based error prediction approach for coarse-grid Computational Fluid Dynamics (CG-CFD). Prog. Nucl. Energy 118, 103140.
Hari, S., Hassan, Y.A., 2002. Improvement of the subcooled boiling model for low-pressure conditions in thermal-hydraulic codes. Nucl. Eng. Des. 216 (1–3), 139–152.
Heinz, S., 2003. Statistical Mechanics of Turbulent Flows. Springer Science & Business Media.
Henry, R.E., Fauske, H.K., 1993. External cooling of a reactor vessel under severe accident conditions. Nucl. Eng. Des. 139 (1), 31–43.
Hüllermeier, E., Waegeman, W., 2021. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. Mach. Learn. 110 (3), 457–506.
Ishii, M., Hibiki, T., 2010. Thermo-fluid Dynamics of Two-phase Flow. Springer Science & Business Media.
Jinyoung, L., Younduk, N., 2019. Convolutional neural network for 2-D assembly-wise pin power peaking factor prediction in PWRs. Transactions 121 (1), 1569–1571.
Kang, K.-H., et al., 2004. Thermal behavior of the reactor vessel penetration under external vessel cooling during a severe accident. Nucl. Technol. 145 (1), 57–66.

Katharopoulos, A., Fleuret, F., 2018. Not all samples are created equal: deep learning with importance sampling. In: International Conference on Machine Learning. PMLR.
Katto, Y., 1994. Critical heat flux. Int. J. Multiphas. Flow 20, 53–90.
Kim, T.K., et al., 2019. Deep-learning-based alarm system for accident diagnosis and reactor state classification with probability value. Ann. Nucl. Energy 133, 723–731.
Kim, J.M., et al., 2020. Abnormality diagnosis model for nuclear power plants using two-stage gated recurrent units. Nucl. Eng. Technol. 52 (9), 2009–2016.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 25, 1097–1105.
Kucukelbir, A., et al., 2017. Automatic differentiation variational inference. J. Mach. Learn. Res. 18 (1), 430–474.
Kutz, J.N., 2017. Deep learning in fluid dynamics. J. Fluid Mech. 814, 1–4.
Kutz, J.N., et al., 2016. Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems. SIAM.
Lakshminarayanan, B., Pritzel, A., Blundell, C., 2016. Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles arXiv preprint arXiv:1612.01474.
Lee, J.H., et al., 2020. An online operator support tool for severe accident management in nuclear power plants using dynamic event trees and deep learning. Ann. Nucl. Energy 146, 107626.
Liang, G., Mudawar, I., 2018. Pool boiling critical heat flux (CHF)–Part 2: assessment of models and correlations. Int. J. Heat Mass Tran. 117, 1368–1383.
Ling, J., Templeton, J., 2015. Evaluation of machine learning algorithms for prediction of regions of high Reynolds averaged Navier Stokes uncertainty. Phys. Fluids 27 (8), 085103.
Ling, J., Jones, R., Templeton, J., 2016a. Machine learning strategies for systems with invariance properties. J. Comput. Phys. 318, 22–35.
Ling, J., Kurzawski, A., Templeton, J., 2016b. Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. J. Fluid Mech. 807, 155–166.
Liu, Y.-k., et al., 2021. A multi-layer approach to DN 50 electric valve fault diagnosis using shallow-deep intelligent models. Nucl. Eng. Technol. 53 (1), 148–163.
Lu, Q., et al., 2021a. Prediction method for thermal-hydraulic parameters of nuclear reactor system based on deep learning algorithm. Appl. Therm. Eng. 196, 117272.
Lu, D., et al., 2021b. Overview on critical heat flux experiment for the reactor fuel assemblies. Ann. Nucl. Energy 163, 108585.
Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems.
Ma, D., et al., 2017. Supercritical water heat transfer coefficient prediction analysis based on BP neural network. Nucl. Eng. Des. 320, 400–408.
Mashlakov, A., et al., 2021. Assessing the performance of deep learning models for multivariate probabilistic energy forecasting. Appl. Energy 285, 116405.
Maulik, R., et al., 2021. A turbulent eddy-viscosity surrogate modeling framework for Reynolds-Averaged Navier-Stokes simulations. Comput. Fluid 227, 104777.
Mell, P., Grance, T., 2011. The NIST Definition of Cloud Computing.
Molnar, C., 2020. Interpretable Machine Learning. Lulu. com.
Moon, S.K., Baek, W.-P., Chang, S.H., 1996. Parametric trends analysis of the critical heat flux based on artificial neural networks. Nucl. Eng. Des. 163 (1–2), 29–49.
Moradi, R., Groth, K.M., 2020. Modernizing Risk Assessment: A Systematic Integration of PRA and PHM Techniques, vol. 204. Reliability Engineering & System Safety, 107194.
Nafey, A.S., 2009. Neural network based correlation for critical heat flux in steam-water flows in pipes. Int. J. Therm. Sci. 48 (12), 2264–2270.
Naimi, A., et al., 2022. Nonlinear model predictive control using feedback linearization for a pressurized water nuclear power plant. IEEE Access 10, 16544–16555.
Neudecker, D., et al., 2020. Enhancing nuclear data validation analysis by using machine learning. Nucl. Data Sheets 167, 36–60.
Nguyen, H.-P., Baraldi, P., Zio, E., 2021. Ensemble empirical mode decomposition and long short-term memory neural network for multi-step predictions of time series signals in nuclear power plants. Appl. Energy 283, 116346.
Ovadia, Y., et al., 2019. Can You Trust Your Model's Uncertainty? Evaluating Predictive Uncertainty under Dataset Shift arXiv preprint arXiv:1906.02530.
Park, H.M., Lee, J.H., Kim, K.D., 2020. Wall temperature prediction at critical heat flux using a machine learning model. Ann. Nucl. Energy 141, 107334.
Patan, K., Patan, M., 2020. Neural-network-based iterative learning control of nonlinear systems. ISA Trans. 98, 445–453.
Podowski, M.Z., 2012. Toward mechanistic modeling of boiling heat transfer. Nucl. Eng. Technol. 44 (8), 889–896.
Rabiti, C., et al., 2021. Raven User Manual. Idaho National Lab.(INL), Idaho Falls, ID (United States).
Rachman, A., Ratnayake, R.C., 2019. Machine learning approach for risk-based inspection screening assessment. Reliab. Eng. Syst. Saf. 185, 518–532.
Radaideh, M.I., Kozlowski, T., 2019. Combining simulations and data with deep learning and uncertainty quantification for advanced energy modeling. Int. J. Energy Res. 43 (14), 7866–7890.
Radaideh, M.I., et al., 2020. Neural-based time series forecasting of loss of coolant accidents in nuclear power plants. Expert Syst. Appl. 160, 113699.
Radaideh, M.I., Price, D., Kozlowski, T., 2021a. Modeling nuclear data uncertainties using deep neural networks. In: EPJ Web of Conferences. EDP Sciences.
Radaideh, M.I., et al., 2021b. Neorl: Neuroevolution Optimization with Reinforcement Learning arXiv preprint arXiv:2112.07057.
Ribeiro, M.T., Singh, S., Guestrin, C., 2016. Why should i trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
Robnik-Šikonja, M., Bohanec, M., 2018. Perturbation-based explanations of prediction models. In: Human and Machine Learning. Springer, pp. 159–175.
Saeed, H.A., et al., 2020. Online fault monitoring based on deep neural network & sliding window technique. Prog. Nucl. Energy 121, 103236.
Saleem, R.A., Radaideh, M.I., Kozlowski, T., 2020. Application of deep neural networks for high-dimensional large BWR core neutronics. Nucl. Eng. Technol. 52 (12), 2709–2716.
She, J., et al., 2021. Diagnosis and prediction for loss of coolant accidents in nuclear power plants using deep learning methods. Nuc. Power Plant Equip. Prognostic Health Manag. Based Data-Driven Methods 9. https://doi.org/10.3389/fenrg.2021.665262.
Shriver, F., Gentry, C., Watson, J., 2021. Prediction of neutronics parameters within a two-dimensional reflective PWR assembly using deep learning. Nucl. Sci. Eng. 195 (6), 626–647.
Spalart, P., Allmaras, S., 1992. A one-equation turbulence model for aerodynamic flows. In: 30th Aerospace Sciences Meeting and Exhibit.
Su, G., et al., 2002. Application of an artificial neural network in reactor thermohydraulic problem: prediction of critical heat flux. J. Nucl. Sci. Technol. 39 (5), 564–571.
Sun, L., et al., 2020. Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data. Comput. Methods Appl. Mech. Eng. 361, 112732.
Szegedy, C., et al., 2013. Intriguing Properties of Neural Networks arXiv preprint arXiv:1312.6199.
Theofanous, T., Syri, S., 1997. The coolability limits of a reactor pressure vessel lower head. Nucl. Eng. Des. 169 (1–3), 59–76.
Tian, X., et al., 2018. A study on the robustness of neural network models for predicting the break size in LOCA. Prog. Nucl. Energy 109, 12–28.
Tolo, S., et al., 2019. Robust on-line diagnosis tool for the early accident detection in nuclear power plants. Reliab. Eng. Syst. Saf. 186, 110–119.
Tripathy, R.K., Bilionis, I., 2018. Deep UQ: learning deep neural network surrogate models for high dimensional uncertainty quantification. J. Comput. Phys. 375, 565–588.
Tsimenidis, S., 2020. Limitations of Deep Neural Networks: a Discussion of G. Marcus' Critical Appraisal of Deep Learning arXiv preprint arXiv:2012.15754.
Wang, J.-X., Wu, J.-L., Xiao, H., 2017. Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data. Physical Review Fluids 2 (3), 034603.
Wang, H., et al., 2020. Remaining useful life prediction based on improved temporal convolutional network for nuclear power plant valves. Front. Energy Res. 8, 296.
Wang, H., et al., 2021a. Advanced fault diagnosis method for nuclear power plant based on convolutional gated recurrent network and enhanced particle swarm optimization. Ann. Nucl. Energy 151, 107934.
Wang, H., et al., 2021b. Remaining useful life prediction techniques for electric valves based on convolution auto encoder and long short term memory. ISA Trans. 108, 333–342.
Whyte, A., Parks, G., 2021. Surrogate model optimization of a 'micro core'pwr fuel assembly arrangement using deep learning models. In: EPJ Web of Conferences. EDP Sciences.
Worrell, C., et al., 2019. Machine learning of fire hazard model simulations for use in probabilistic safety assessments at nuclear power plants. Reliab. Eng. Syst. Saf. 183, 128–142.
Wu, J.-L., Xiao, H., Paterson, E., 2018. Physics-informed machine learning approach for augmenting turbulence models: a comprehensive framework. Physical Review Fluids 3 (7), 074602.
Xu, Z., Saleh, J.H., 2021. Machine Learning for Reliability Engineering and Safety Applications: Review of Current Status and Future Opportunities. Reliability Engineering & System Safety, 107530.
Yang, Z., et al., 2017. Application of convolution neural network to flow pattern identification of gas-liquid two-phase flow in small-size pipe. In: 2017 Chinese Automation Congress (CAC). IEEE.
Yoo, J., Estrada-Perez, C.E., Hassan, Y.A., 2014. A proper observation and characterization of wall nucleation phenomena in a forced convective boiling system. Int. J. Heat Mass Tran. 76, 568–584.
Zhang, Z.J., Duraisamy, K., 2015. Machine learning methods for data-driven turbulence modeling. In: 22nd AIAA Computational Fluid Dynamics Conference.
Zhang, Z., Olatubosun, S.A., 2019. Uncertainties associated with the reliability of thermal-hydraulic nuclear passive systems. J. Nucl. Sci. Technol. 56 (1), 17–31.
Zhang, Y., Xiao, C., 2020. Reliability on deep learning models: a comprehensive observation. In: 2020 6th International Symposium on System and Software Reliability (ISSSR). IEEE.
Zhang, Y., et al., 2010. Analysis of safety margin of in-vessel retention for AP1000. Nucl. Eng. Des. 240 (8), 2023–2033.
Zhang, D., et al., 2019. Quantifying total uncertainty in physics-informed neural networks for solving forward and inverse stochastic problems. J. Comput. Phys. 397, 108850.
Zhao, F., et al., 2020. A machine learning methodology for reliability evaluation of complex chemical production systems. RSC Adv. 10 (34), 20374–20384.
Zhao, X., et al., 2021a. Prognostics and health management in nuclear power plants: an updated method-centric review with special focus on data-driven methods. Front. Energy Res. 9, 294.
Zhao, X., Salko, R.K., Shirvan, K., 2021b. Improved departure from nucleate boiling prediction in rod bundles using a physics-informed machine learning-aided framework. Nucl. Eng. Des. 374, 111084.
Zhou, Y., Li, B., Lin, T.R., 2021. Maintenance Optimisation of Multicomponent Systems Using Hierarchical Coordinated Reinforcement Learning. Reliability Engineering & System Safety, 108078.