Annals of Nuclear Energy 180 (2023) 109452

Contents lists available at ScienceDirect

Annals of Nuclear Energy

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

Review

Data-driven machine learning for disposal of high-level nuclear waste: A review

Guang Hu *, Wilfried Pfingsten

Laboratory for Waste Management, Paul Scherrer Institut, 5232 Villigen, Switzerland

ARTICLE INFO

Keywords:
Data-driven
Machine learning
High-level nuclear waste
Deep geological repository

ABSTRACT

The application of the data-driven machine learning (DDML) for the disposal of the high-level nuclear waste (HLW) is of emerging interest in the recent years. This review aims to systematically elaborate, analyze, and summarize recent advances related to DDML in the area of disposal of the HLW. Firstly, a comprehensive work on the DDML for the disposal of the HLW is examined. Five DDML algorithms including the linear regression (LR), principle component analysis (PCA) and artificial neural network (ANN) are illustrated. Then, it summarizes the typical DDML algorithms and the main inputs/outputs for the deep geological repository (DGR). Furthermore, it is concluded that the hybrid DDML algorithms are efficient choices. Also, the DDML shows a great applicability for the simulation of the multiscale and multiphysics field. Lastly, the physical-informed DDML may enhance the performance of all algorithms.

1. Introduction

1.1. Disposal of High-level nuclear waste

Nuclear energy plays an important role in addressing the global warming and energy supply issues (Azam et al., 2021). An increasing interest on the nuclear energy comes with the disposal of the nuclear waste (Wisnubroto et al., 2021). Commonly, classification of the nuclear waste commonly depends on the radioactivity level and radionuclide half-life (Corkhill and Hyatt, 2018). It can be categorized into the high-level nuclear waste (HLW), intermediate-level nuclear waste (ILW) and low-level nuclear waste (LLW) (Othman et al., 2019). Among them, the HLW has the highest radioactivity level. It requires the radioactive shielding as well as consideration of the reliable containment of the HLW (El-Samrah et al., 2021). In particular, safety and security of the HLW would affect the development of the nuclear industry and public acceptance of the nuclear energy (Kurniawan et al., 2022). Hence, extra attention needs to be paid to the disposal of the HLW.

For the HLW, its disposal is quite problematic while the spent nuclear fuel (SNF) or reprocessing wastes is included. For instance, the radionuclide ^{99}Tc within the HLW has a half-life of around 211,100 years (Yamano et al., 2021). It may need to maintain the long-lived radionuclides isolated for 300,000 years. For the nuclear waste storage technologies, each technology including the vitrification, partitioning and transmutation, pyro-processing, and deep geological repository (DGR) has its own advantages and limitations. Moreover, the DGR provides an ultimate destination in a deep underground that permanently isolates the waste from inhabitants and the environment (Kurniawan et al., 2022). Consequently, the DGR for the disposal of the HLW is a relatively favor choice worldwide (Hall et al., 2021). However, the DGR which consists of substantial engineered barriers and nature barriers is quite complex (Konevnik et al., 2020). Furthermore, the thermal–hydraulic-mechanical-chemical multi-physics processes usually occur in the DGR (De Lucia et al., 2017; Abootalebi and Siemens, 2018). In addition, the pore size of the porous media differs and spans orders in magnitude (Wang and Sun, 2018). Hence, the DGR belongs to the typical complex system with different time scales and space scales. Traditional numerical tools may lack some physical processes and are quite computational expensive (Laloy and Jacques, 2019). It can hardly capture all the details which is also not easy to be implemented (Prasianakis et al., 2020).

1.2. Data-driven machine learning in HLW

In the recent years, data-driven models have drawn much attention in many fields including the disposal of the HLW. In Fig. 1, publication records and citations on both the machine learning and radioactive waste disposal in the nuclear industry searched on the Web of Science are presented. As can be seen in Fig. 1, the number of publications as well as the citations has increased rapidly in the last 10 years, and

* Corresponding author.
E-mail address: guang.hu@psi.ch (G. Hu).

https://doi.org/10.1016/j.anucene.2022.109452
Received 1 July 2022; Received in revised form 24 August 2022; Accepted 9 September 2022
Available online 15 September 2022

Nomenclature

ANN	Artificial neural network
BPNN	Back propagation neural network
BN	Bayesian network
CFD	Computational fluid dynamics
CM	Clustering method
CNN	Convolutional neural network
DBN	Deep belief network
DDML	Data-drive machine learning
DGR	Deep geological repository
DNN	Deep neural network
DQL	Deep Q learning network
DRL	Dimension reduction learning
DT	Decision tree
EL	Ensemble learning
FC	Fine-grained content
GC	Gravel content
GMDH	Group method of data handling
GPR	Gaussian process regression
Gs	Specific gravity
HLW	High-level nuclear waste
ILK	Intermediate-level nuclear waste
kNN	k-nearest neighbor
kRR	Kernel ridge regression
LLW	Low-level nuclear waste
LSTM	Long-short term memory
LR	Linear regression
WL	Liquid limit
WP	Plastic limit
MC	Monte Carlo
ML	Machine learning
NB	Naive Bayes
NN	Neural network
W_{opt}	Optimum moisture content
PCA	Principal component analysis
PCs	Principal components
PP	Pyro-processing
PT	Partitioning and transmutation
RBF	Radial basis function
RT	Random tree
SC	Sand content
SNF	Spent nuclear fuel
SVD	Singular value decomposition
Y	Output of ANN
$\rho_{measured}$	Measured glass density, g/cm^3
$\rho_{predicted}$	Predicted glass density, g/cm^3

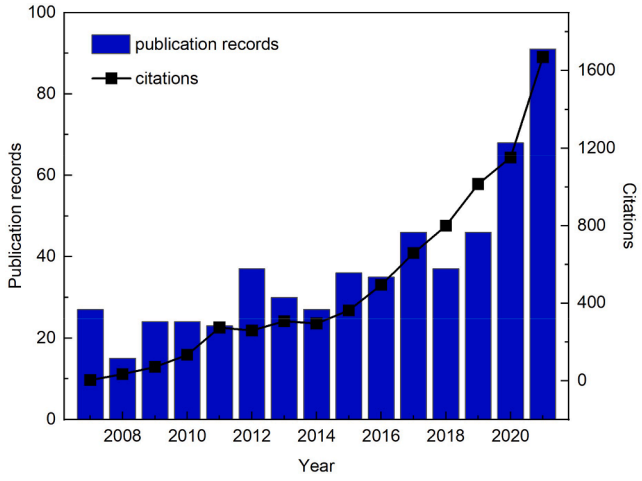

Fig. 1. Publication records and citations on the machine learning and radioactive waste disposal in the nuclear industry searched on the Web of Science (The search item is ((data driven or machine learning or ML or artificial intelligence or AI) and (nuclear waste or nuclear waste disposal or radioactive waste or radioactive waste disposal or high-level radioactive waste or HLW)) in the topic in Web of Science).

especially in the last 5 years. Combined with the machine learning (ML) methods, data-driven machine learning (DDML) appears to be quite powerful and can handle complex issues without quite much priori knowledge of the system (Hu et al., 2021a).

With the development of computational technologies, DDML has the potential to achieve high accuracy with relatively low computational cost (Shastri et al., 2021). Various nuclear filed stakeholders and research communities adapt it to overcome difficulties including the fault diagnosis, condition monitoring in the nuclear industry (Laikari and Backman, 2021). In addition, it has been considered as a promising future modelling direction from the encouraging results made by the recent studies for the disposal of the HLW (Asher et al., 2015; Birkholzer et al., 2019; Chinesta et al., 2020; Menke et al., 2021). However, to the best of our knowledge, there still lacks research on comprehensive reviewing the state of the art progress on the DDML for the disposal of the HLW in the nuclear industry (Kurniawan et al., 2022). As a result, it is necessary to consolidate this issue from scattered literature as well as many interdisciplinary areas.

1.3. Scope of this review

This review paper focuses on filling the gap in the body of literature and reflecting the current research status of DDML for the disposal of the HLW. It aims to systematically elaborate, analyze, and summarize recent advances concerned with DDML for the disposal of the HLW. In Section 2, algorithms of the DDML for the disposal of the HLW would be discussed, including the linear regression, principal component analysis and artificial neural network. Section 3 shows the applications, developments and future directions. Section 4 comes to the conclusions and remarks.

2. Machine learning algorithms for disposal of high-level nuclear waste

Currently, there are several ML algorithms for the disposal of HLW as presented in Fig. 2. On the one hand, it can be categorized into the supervised learning, unsupervised learning and reinforcement learning algorithms by the learning type. On the other hand, it can also be classified as the regression, instance-based learning, neural network, deep learning, dimension reduction and kernel-based leaning algorithms by the algorithm type (Mahesh, 2020). As for the DDML for the disposal of HLW, it is an emerging direction. It needs to be described in light of the current state of research.

In Fig. 3, the keyword extraction on the ML method and nuclear waste disposal in the nuclear industry searched on the Web of Science is shown. From Fig. 3, the linear regression (LR), principle component analysis (PCA), artificial neural network (ANN), genetic algorithm (GA) and clustering are the most commonly adopted algorithms for this topic. Consistent with Fig. 3, Table 1 presents the main algorithms with the DDML for the disposal of HLW. Therefore, the five algorithms mentioned

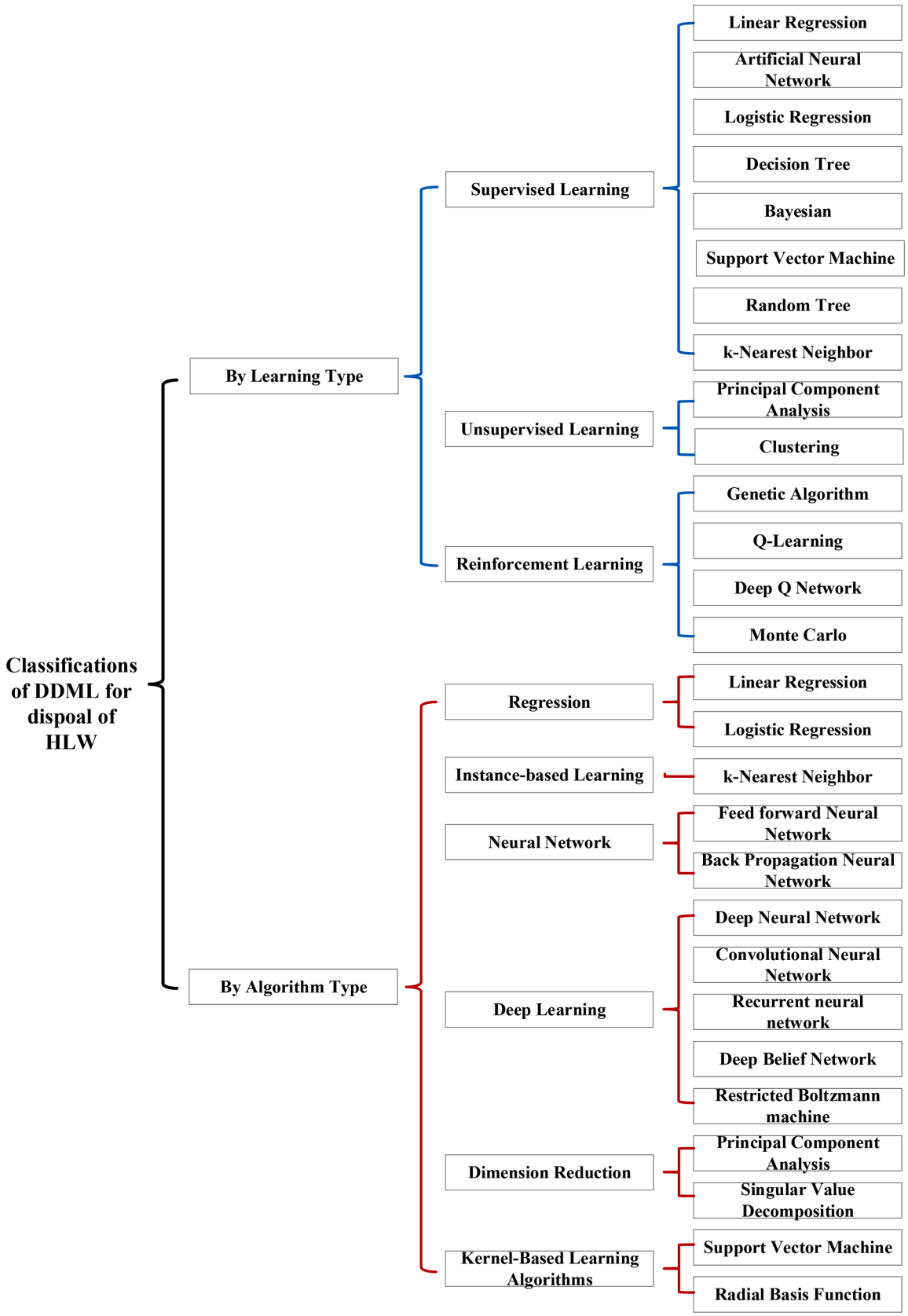

Fig. 2. Detailed classifications of DDML for disposal of HLW (partially from reference (Mahesh, 2020).

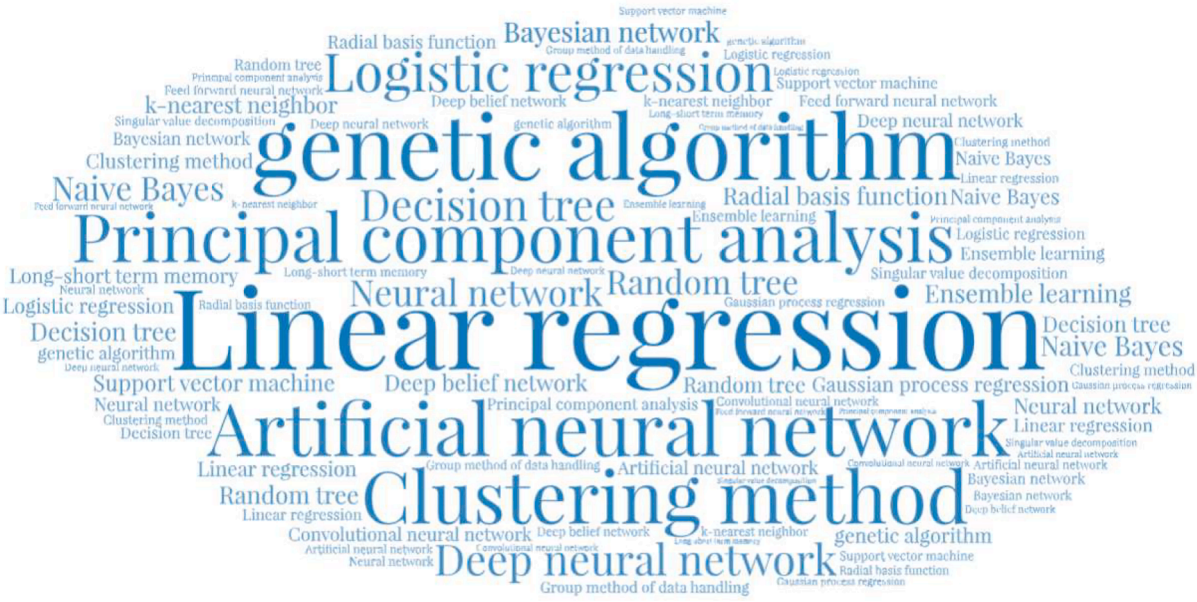

Fig. 3. Keyword extraction on the machine learning algorithm and radioactive waste disposal in the nuclear industry searched on the Web of Science.

Table 1
Main research on the status of the DDML for the disposal of HLW.

Author/year	Algorithms	Descriptions
(Gong et al., 2021); (Bang et al., 2020); (Trivelpiece et al., 2020); (Neeway et al., 2018)	Linear regression (LR)	Supervised learning type
(Kim et al., 2020); (Wang et al., 2020); (Tudorache et al., 2018)	Principal component analysis (PCA)	Unsupervised learning type
(Lu et al., 2021); (Sirdesai et al., 2019); (Yoon et al., 2019)	Artificial neural network (ANN)	Supervised learning type
(Solans et al., 2021); (Elodie et al., 2020); (Tsai et al., 2019)	Genetic algorithm (GA)	Supervised or Unsupervised learning type
(Suh et al., 2020); (Xu et al., 2020)	Clustering method (CM)	Unsupervised learning type
(Stanfill et al., 2020); (Suh et al., 2018)	Logistic regression	Supervised learning type
(Kim et al., 2020)	Deep neural network (DNN)	Deep learning type
(Bang et al., 2020)	Decision tree (DT)	Supervised learning type
(Tosoni et al., 2019)	Bayesian network (BN)	Deep learning type
(Sun et al., 2020)	Support vector machine (SVM)	Supervised learning type
(Lu et al., 2021)	Gaussian process regression (GPR)	Supervised learning type

above will be described in detail.

2.1. Linear regression

Firstly, the LR provides a means to model a straight line relationship between two variables which belongs to the supervised learning type. In supervised learning algorithms, the output for the given input is known. For example, the measured glass density $\rho_{measured}$ versus predicted glass density $\rho_{predicted}$ based on the LR algorithm is shown in Fig. 4 (Trivelpiece et al., 2020).

In this case, the input parameter is the additivity of partial molar volumes of individual oxide components of the glass while the output parameter is the glass density. In Table 1, Gong et al. (2021) adopted it for the prediction of the leaching behaviors of A2B2O7-structures in the glass. Additionally, both the prediction of the thermal conductivity of

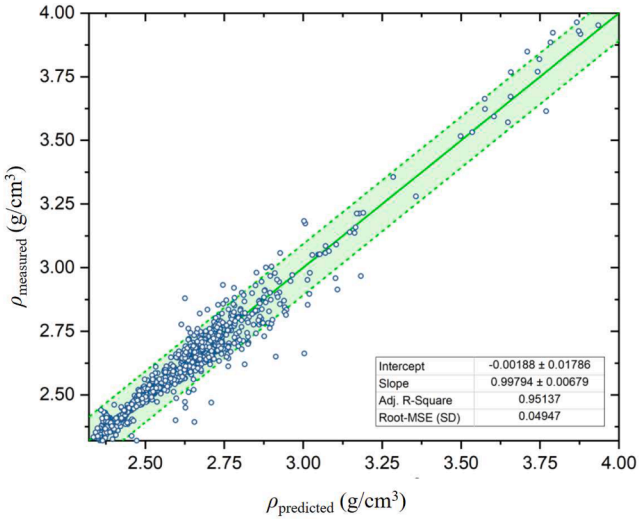

Fig. 4. Linear regression (LR): measured and predicted glass density (Trivelpiece et al., 2020).

bentonite (Bang et al., 2020) and dissolution behavior of borosilicate glasses in far-from equilibrium conditions (Neeway et al., 2018) are based on the LR algorithm.

2.2. Principal component analysis

Then, the PCA converts a large set of variables into a smaller one while retaining the main features of the large set (Kim et al., 2020; Wang et al., 2020; Tudorache et al., 2018). It is also a kind of the unsupervised learning type. In unsupervised learning algorithms, the output for the given input is unknown.

Fig. 5 shows an illustration of the loading of the 16 hydro-chemical variables and scores of the 3D spaces of principle components (PCs) (Kim et al., 2020). In Fig. 5(a), four in-site variables (pH, temperature, electrical conductivity, alkalinity), four calculated variables (dissolved solids, CO_2, HCO_3^-, CO_3^-) and eight major ions are considered as the input parameters. In Fig. 5(b), the scores of the deep groundwater samples in the DQR are presented with the three PCs.

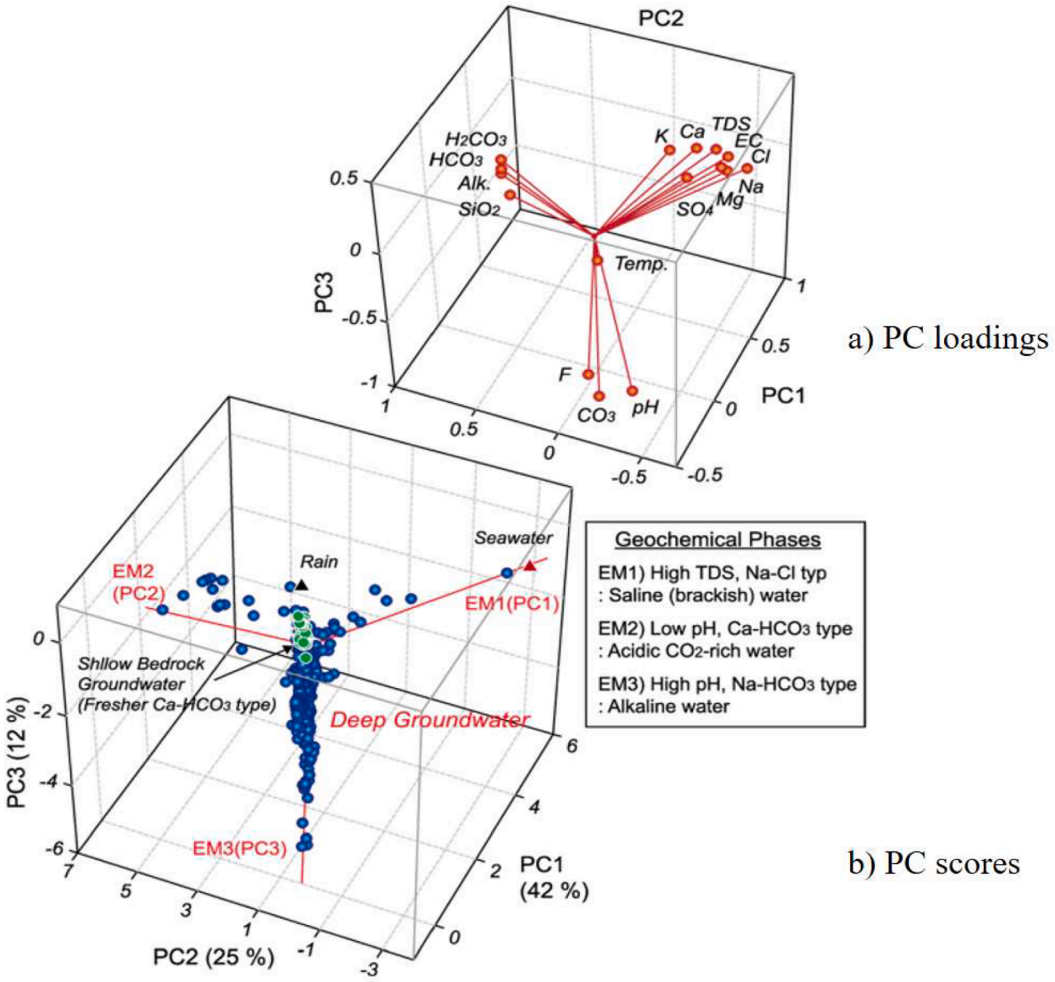

Fig. 5. Principal component analysis (PCA): loading of 16 hydro-chemical variables and scores of 3D spaces of principle components (PCs) (Kim et al., 2020).

2.3. Artificial neural network

In addition, the neuron of the ANN algorithm processes the input data of different layers with suitable weights and biases. It is a typical supervised learning type algorithm. A basic mathematical formula of the principle of the ANN is shown in Eq. (1) (Jain et al., 1996).

Sirdesai et al. (2019) adopted the ANN algorithm to predict the strength properties of the thermally treated sandstone of the DGR. Five input parameters including the bulk density and temperature are considered as the input parameters as presented in Fig. 6. It is also adopted to predict the nepheline precipitation from compositions (Lu et al., 2021) and the specific heat capacity of bentonite buffer materials (Yoon et al., 2019).

$$Y = \sum_{j=1}^{m} W_j f\left(\sum_{i=1}^{n} W_{ij} X_i + b_i\right) \quad (1)$$

where Y is the output; W_j ($j = 1, 2,\ldots, m$), W_{ij} are the model parameters; m is the number of the hidden layers; f is the transfer function; X_i ($i = 1, 2,\ldots, n$) is the input variable; n is the number of the input variables; b_i is the model biasness.

2.4. Genetic algorithm

Furthermore, for the GA algorithm, it reflects the process of the natural selection where the fittest individuals are selected for reproduction in order to produce the offspring of the next generation. Especially, it can be unsupervised/supervised learning type algorithm.

A GA algorithm is then developed to optimize simultaneously the effective neutron multiplication factor k_{eff} and decay heat of the SNF canister as shown in Fig. 7 (Solans et al., 2021). In Fig. 7, the GA algorithm ends with 10,000 iterations. Initially, the emplacement of each SNF is randomly determined. Afterwards, the mean value and standard deviation of k_{eff} change with the iterations. This algorithm is also adopted for the optimization of both the sensor arrangements (Elodie et al., 2020) and cutting plan of the SNF (Tsai et al., 2019).

2.5. Clustering method

For the clustering method (CM), it is an unsupervised learning algorithm. It is a quite popular method which identifies the similar groups in a dataset. Xu et al. (Xu et al., 2020) estimated the geometric characteristics of fracture traces with CM. By the way, the clustering analysis was also used to identify topics in nuclear waste treatment patents (Suh et al., 2020).

In Fig. 8, the clustering results for the total discontinuity set using the CM is presented. To investigate the applicability, the CM was applied with totally 286 discontinuity sets. It automatically identify the most unpopular discrete values (marked black in Fig. 8) with the determined threshold value. At the same time, this paper recommends that the discrete values should be better removed to accelerate the simulation accuracy in most engineering cases. As the dataset is enough for common engineering cases, it is no need to explore the influence of the discrete values.

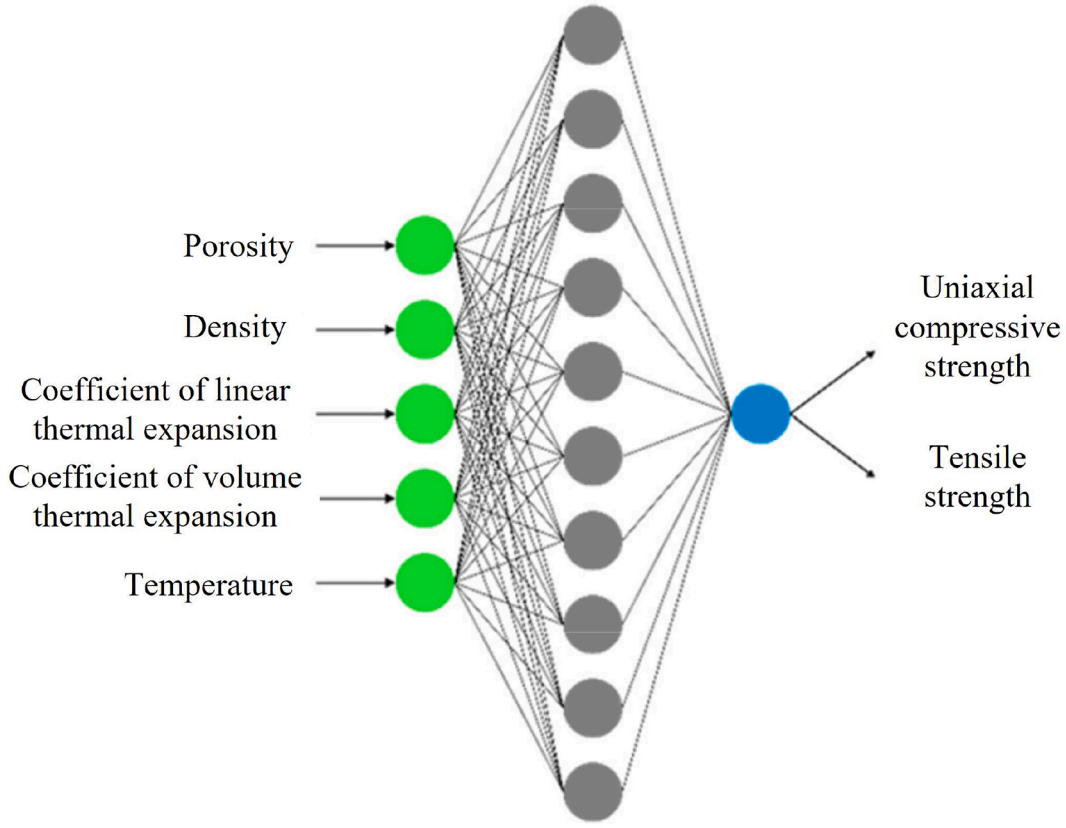

Fig. 6. Artificial neural network (ANN): using five input parameters to predict strength properties of thermally treated sandstone (Sirdesai et al., 2019).

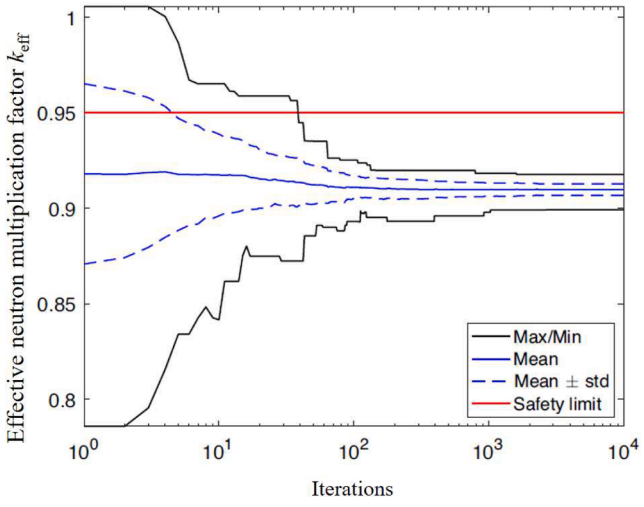

Fig. 7. Genetic algorithm (GA): mean value and standard deviation of k_{eff} change with iterations (Solans et al., 2021).

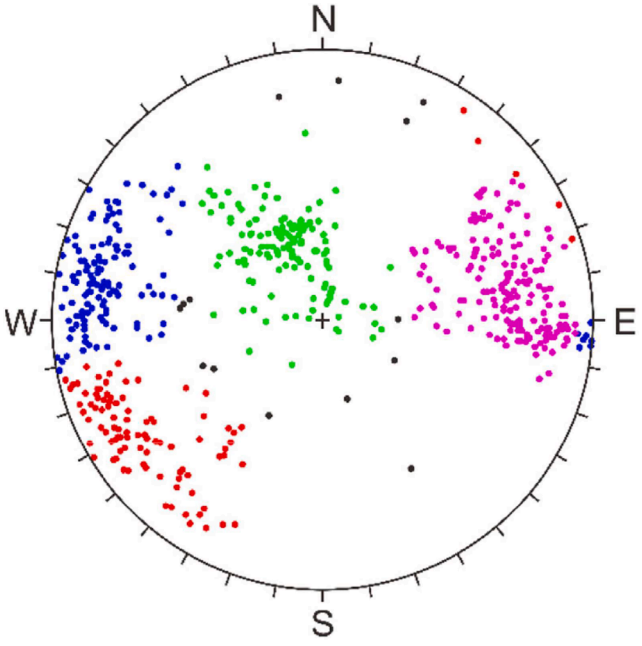

Fig. 8. Clustering method (CM): results for total discontinuity set (Xu et al., 2020).

2.6. Other algorithms

Lastly, there are also several other algorithms adopted for the disposal of the HLW including the logistic regression, deep neural network (DNN), neural network (NN) and support vector machine (SVM). Among them, the logistic regression and SVM are the supervised learning type algorithms while the DNN is a deep learning type.

For instance, the logistic regression is applied for the modeling binary response data from a mixture experiment (Stanfill et al., 2020) and identification of major factors affecting nuclear decommissioning strategy decision (Suh et al., 2018). Kim et al. (2020) carried out the waste recognition system based on the DNN algorithm. Tosoni et al. (2019) conducted the safety assessment of nuclear waste repositories with the NN method. Sun et al. (2020) determined the hydraulic aperture of rough rock fractures using the SVM.

3. Application of DDML for disposal of high-level nuclear waste

At present, huge achievements have been obtained for the disposal of the HLW with the above DDML algorithms. Hence, there is necessary to summarize the latest applications of the DDML in this field. Table 2 and Table 3 elaborate the latest applications of the DDML for the DGR and vitrification respectively. Table 4 summarizes the future development of the DDML for the disposal of the HLW.

Table 2
Latest application of DDML for DGR.

Author/ year	Algorithms	Descriptions	Inputs and outputs
(Jeong et al., 2018)	ANN, SVM	Uncertainty prediction of subsurface models	• Input parameters: matrix of data vectors of all prior models for observation times; Output parameters: matrix of data vectors of all prior models for forecast times;
(Prasianakis et al., 2020)	NN + Modelling	Geochemical reactive transport simulations	• Input parameters: master species concentrations of Sr^{2+}, SO_4^{2}; Output parameters: saturation index of celestine
(Benbouras and Lefilef, 2021)	DNN, RT	Prediction of the soil compaction parameters	• Input parameters: FC, SC, GC, Gs, WL, WP; Output parameters: W_{opt}, ρ_{dmax}
(Tian et al., 2021)	GA, ANN + Modelling	Permeability prediction of porous media	• Input parameters: pore structure parameters; Output parameters: permeability
(Menke et al., 2021)	DT	Upscaling multimodal porosity–permeability relationships	• Input parameters: porosity of sub-volumes; Output parameters: permeability
(Frankel et al., 2021)	ANN	Prediction of the crystallographic cell parameters	• Input parameters: average ionic radius, elements electronegativity; Output parameters: crystal channel size
(Hu et al., 2021b)	CNN	Mesh generation from rock fracture images	• Input parameters: rock matrix, fractures; Output parameters: generated mesh
(Schmeide et al., 2021)	ANN	Prediction of the carbon steel corrosion product	• Input parameters: spectral similarity; Output parameters: fractions, pH values
(Solans et al., 2021)	GA	Optimization of the canister effective neutron multiplication factor k_{eff} value and fuel decay heat	• Input parameters: 38 isotopic concentrations; Output parameters: k_{eff}, Decay heat

where FC is the fine-grained content; SC is the sand content; GC is the gravel content; Gs is the specific gravity; WL is the liquid limit; WP is the plastic limit; W_{opt} is the optimum moisture content; ρ_{dmax} is the maximum dry density.

Table 3
Main research of the DDML for vitrification.

Author/ year	Algorithms	Descriptions	Inputs and outputs
(Krishnan et al., 2018)	ANN	Prediction of the dissolution rates of silicate glasses	• Input parameters: glass composition, initial solution pH value, pH value at the time of measurement; Output parameters: SiO_2 leaching rate
(Bishnoi et al., 2019)	GPR	Prediction of Young's modulus for silicate glasses	• Input parameters: molar percentage composition of the oxide components, glass density; Output parameters: Young's Modulus of the glass compositions
(Hartnett et al., 2019)	RT	Prediction of formable and thermodynamically stable iodine-containing apatites	• Input parameters: 13 crystal chemistry descriptors; Output parameters: categorical binary variable (Yes or No)
(Lillington et al., 2020)	NN, SVM, GPR	Prediction of the glass leaching behavior	• Input parameters: 17 + 13 variable combinations; Output parameters: normalized boron (B) release
(Lu et al., 2021)	kNN, GPR, ANN, SVM, etc.	Prediction of the nepheline precipitation from compositions	• Input parameters: compositions of 332 glasses; Output parameters: nepheline precipitation from glass compositions
(Gong et al., 2021)	LR, kernel ridge regression (KRR)	Prediction of leaching behaviors of A2B2O7-structures	• Input parameters: 30 glass compositions Output parameters: leaching behaviors of A2B2O7-structures

3.1. Latest application of DDML for DGR

Currently, the DGR method is recognized as one of the most reliable methods for the long-term disposal of the HLW (Hall et al., 2021). As shown in Table 2, the DDML has been adopted for the DGR for the disposal of the HLW.

Firstly, Jeong et al. (2018) proposed the ANN and SVM for the predictive uncertainty of subsurface models. In addition, Prasianakis et al. (2020) combined the NN with the geochemical reactive transport simulations. Then, Benbouras and Lefilef (Benbouras and Lefilef, 2021) predicted the soil compaction parameters with the DNN and RT. Afterwards, Tian et al. (2021) adopt the GA, ANN algorithm for the Permeability prediction of the porous media with the computational fluid dynamics (CFD). Menke et al. (2021) upscaling multimodal porosity–permeability relationships with the DT algorithm and Darcy–-Brinkman–Stokes model. Moreover, the ANN algorithm is adopted for the prediction of both the crystallographic cell parameters (Frankel et al., 2021) and carbon steel corrosion product (Schmeide et al., 2021). Hu et al. (2021b) conducted the CFD mesh generation with the CNN. At

Table 4
Development of the DDML for the disposal of HLW.

Author/year	Characteristics	Descriptions
(Bang et al., 2020)	LR, DT, SVM, EL, GPR, ANN, DBN	Prediction of thermal conductivity model for compacted bentonite
(Lillington et al., 2020)	NN, SVM, GPR	Prediction of the glass leaching behavior
(Lu et al., 2021)	kNN, GPR, ANN, SVM, etc.	Prediction of the nepheline precipitation from compositions
(Gong et al., 2021)	LR, kRR	Prediction of the chemical durability of A2B2O7 pyrochlore and fluorite
(Benbouras and Lefilef, 2021)	DNN, RT	Prediction of the soil compaction parameters
(Ebiwonjumi et al., 2021)	GPR, SVM, NN	Prediction of the fuel assembly decay heat
(Birkholzer et al., 2019) (Viswanathan et al., 2022)	ML + Multiscale and multiphysics field, fracture	Modeling coupled THMC processes
(Prasianakis et al., 2020)	NN based digital twin + Multiscale and multiphysics field	Neural network based on the results of microscopic geochemical reactive transport simulations
(Menke et al., 2021)	DT + Multiscale and multiphysics field	Combination with DT from the pore scale to the Darcy-scale
(Himanen et al., 2019)	ML + Multiscale and multiphysics field, fracture	Acceleration of physics-based models by orders of magnitude
(Krishnan et al., 2018)	ANN + Physical-informed modelling	Prediction of the dissolution kinetics of silicate glasses
(Hu et al., 2021)	CNN + Physical-informed modelling	Mesh generation from rock fracture images with CNN
(Tian et al., 2021)	GA, ANN + Physical-informed modelling	Permeability prediction of porous media
(Solans et al., 2021)	GA + Physical-informed modelling	GA is developed to optimize simultaneously the canister effective neutron multiplication factor k_{eff} value and fuel decay heat

last, Solans et al. (2021) utilized the GA approach for the optimization of the canister effective neutron multiplication factor k_{eff} value and fuel decay heat. For the input and output parameters, those are described in detail in Table 2.

3.2. Latest applications of DDML for vitrification

Especially, the vitrification technique is the most preferable solution for immobilizing HWL of various origins (Kurniawan et al., 2022). In Table 3, the parameter predictions for the vitrification with various algorithms are presented.

Initially, Krishnan et al. (2018) applied the physical-informed ANN algorithm for the prediction of the dissolution rates of silicate glasses. Then, Bishnoi et al. (2019) carried out the prediction of Young's modulus for silicate glasses with the GPR algorithm. Also, Hartnett et al. (2019) conducted the prediction of formable and thermodynamically stable iodine-containing apatites. Lillington et al. (2020) adopted the NN, SVM and GPR for the Prediction of the glass leaching behavior. It should be noted that there are 17 and 13 input variable combinations for the static and dynamic glass leaching behaviors correspondingly. Gong et al. (2021) conducted the leaching behaviors of A2B2O7-structures with the LR and kernel ridge regression (KRR) method. Finally, Lu et al. (2021) predicted the nepheline precipitation from compositions with the kNN, GPR, ANN, SVM, etc.

3.3. Further development of DDML for disposal of HLW

The DDML is of emerging direction for the disposal of the HLW. As described above, many efforts have been taken into the predictions of parameters of the DGR or vitrification. In particular, the future development of the DDML for the disposal of the HLW can be concluded based on Table 4.

(1) hybrid DDML algorithms.

Firstly, the hybrid DDML algorithms are nice choices as presented (Lu et al., 2021; Benbouras and Lefilef, 2021; Gong et al., 2021; Bang et al., 2020; Lillington et al., 2020; Ebiwonjumi et al., 2021). The accuracy of the DDML is related to many factors including the learning algorithm and model parameters. Hence, the hybrid DDML algorithms can yields notably improved results even with no changes in the input dataset.

(2) multiscale and multiphysics simulation.

Furthermore, the DDML shows a great superiority for the simulation of the multiscale and multiphysics field (Prasianakis et al., 2020; Menke et al., 2021; Viswanathan et al., 2022; Himanen et al., 2019). For the disposal of the HLW in the DGR, it belongs to a typical multiscale and multiphysics process. It spans from microscopic, mesoscopic, to macroscopic scales. Also, the thermal, hydraulic, mechanical and chemical processes are frequently involved. Hence, it is difficult for the traditional numerical calculations to overcome these issues.

(3) physical-informed DDML.

Finally, the DDML coupled with physical constraints, theoretical equations and relations can enhance the DDML performance (Solans et al., 2021; Tian et al., 2021; Hu et al., 2021a; Krishnan et al., 2018). Standard DDML algorithms lack important physics and mechanisms. The so-called physical-informed DDML can fill the gaps of the traditional DDML algorithms. Taking the physical-informed modelling into account, it is observed that the calculation results of all algorithms are improved. Combined with the physical-informed modelling, the computational accuracy of even simple models can be comparable to that of previous complex models.

Compared with the traditional methods, the DDML is adoptable and presents new possibilities to handle the complex systems. Nevertheless, as a DDML, its performance largely rely on the quality and quantity of the dataset. For the issues of the small dataset, one option could be the public scientific literature or database (Karniadakis et al., 2021). Therefore, literatures may contain many kinds of the dataset. Another option may be the physical-informed DDML. Especially, the models of the physical-informed DDML can be effectively constrained on a lower-dimensional manifold and hence trained with a small data regime (Krems, 2019). In addition, physical-informed DDML can not only the interpolation, but also the extrapolation [59]. The third option could be the combination of the physical modelling and DDML (Prasianakis et al., 2020). For complex computational models, traditional calculations are time-consuming. In case the accuracy of the model is verified, we can perform a few necessary calculations to obtain the data needed for the DDML. In this way, it is possible to combine the advantages of the physical computation as well as the DDML.

4. Conclusions and future remarks

Nuclear energy plays an important role in energy supply and stability. Moreover, safety and security of the high-level nuclear waste (HLW) would affect the development of the nuclear industry and public acceptance of the nuclear energy. In particular, the data-drive machine learning (DDML) for the disposal of the HLW is of emerging interest in the recent years. This work reviews the state-of-art DDML for the disposal of the HLW. The main conclusions are obtained:

(1) Firstly, a comprehensive work on the DDML for the disposal of the HLW is elaborated. Five DDML algorithms including the linear regression (LR), principle component analysis (PCA), artificial neural network (ANN), genetic algorithm (GA) and clustering method (CM) are the most commonly adopted algorithms in this filed.

(2) Then, applications of the DDML for the disposal of the HLW are described. It summarizes the typical DDML algorithms and the main

inputs/outputs for the deep geological repository (DGR). Among them, many efforts have been taken into the predictions of parameters of the DGR or vitrification.

(3) Afterwards, it is concluded that the hybrid DDML algorithms are nice choices. Also, the DDML presents a great superiority for the simulation of the multiscale and multiphysics field due to the improved efficiency and accuracy. In addition, the physical-informed DDML can be effectively constrained on a lower-dimensional manifold and hence dealt with a small dataset. Lastly, the DDML can also cooperate with the physical modelling.

Compared with the traditional numerical calculations, the DDML shows the superior advantage in the calculation efforts and accuracy. In addition, it has been considered as a promising future modelling direction from the encouraging results made by the recent studies. Therefore, this review which focuses on elaborating the DDML, introducing the applications and illustrating the further development of the disposal of the HLW can provide the guidance to related research as comprehensive as possible.

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.

Data availability

Data will be made available on request.

Acknowledgments

This work was partially funded by the European Union's Horizon 2020 research and innovation programme **Eu**ropean Joint Programme on **Rad**ioactive Waste Management in the EURAD MODATS (**Mo**nitoring equipment and **Da**ta **T**reatment for **S**afe repository operation and staged closure) work package [Grant No. 847593].

References

Abootalebi, P., Siemens, G., 2018. Thermal properties of engineered barriers for a Canadian deep geological repository. Can. Geotech. J. 55 (6), 759–776.
Asher, M.J., Croke, B.F., Jakeman, A.J., et al., 2015. A review of surrogate models and their application to groundwater modeling. Water Resour. Res. 51 (8), 5957–5973.
Azam, A., Rafiq, M., Shafique, M., et al., 2021. Analyzing the effect of natural gas, nuclear energy and renewable energy on GDP and carbon emissions: A multi-variate panel data analysis. Energy 219, 119592.
Bang, H.T., Yoon, S., Jeon, H., 2020. Application of machine learning methods to predict a thermal conductivity model for compacted bentonite. Ann. Nucl. Energy 142, 107395.
Benbouras, M.A., Lefilef, L., 2021. Progressive machine learning approaches for predicting the soil compaction parameters. Transport. Infrastr. Geotechnol. 25, 1–28.
Birkholzer, J.T., Tsang, C.F., Bond, A.E., et al., 2019. 25 years of DECOVALEX-Scientific advances and lessons learned from an international research collaboration in coupled subsurface processes. Int. J. Rock Mech. Min. Sci. 122, 103995.
Bishnoi, S., Singh, S., Ravinder, R., et al., 2019. Predicting Young's modulus of oxide glasses with sparse datasets using machine learning. J. Non-Cryst. Solids 524, 119643.
Chinesta, F., Cueto, E., Abisset-Chavanne, E., et al., 2020. Virtual, digital and hybrid twins: a new paradigm in data-based engineering and engineered data. Arch. Comput. Methods Eng. 27 (1), 105–134.
Corkhill, C., Hyatt, N., 2018. Nuclear Waste Management. IOP Publishing.
De Lucia, M., Kempka, T., Jatnieks, J., et al., 2017. Integrating surrogate models into subsurface simulation framework allows computation of complex reactive transport scenarios. Energy Procedia 125, 580–587.
Ebiwonjumi, B., Cherezov, A., Dzianisau, S., et al., 2021. Machine learning of LWR spent nuclear fuel assembly decay heat measurements. Nucl. Eng. Technol. 53 (11), 3563–3579.
Elodie, C., Aurélie, T., Alaa, C., et al., 2020. Sensors position optimization for monitoring the convergence of radioactive waste storage tunnel. Nucl. Eng. Des. 367, 110778.
El-Samrah, M.G., Tawfic, A.F., Chidiac, S.E., 2021. Spent nuclear fuel interim dry storage; Design requirements, most common methods, and evolution: A review. Ann. Nucl. Energy 160, 108408.
Frankel, G.S., Vienna, J.D., Lian, J., et al., 2021. Recent advances in corrosion science applicable to disposal of high-level nuclear waste. Chem. Rev. 121 (20), 12327–12383.
Gong, B., Yang, K., Lian, J.A., et al., 2021. Machine learning-enabled prediction of chemical durability of A2B2O7 pyrochlore and fluorite. Comput. Mater. Sci. 200, 110820.
Hall, D.S., Behazin, M., Binns, W.J., et al., 2021. An evaluation of corrosion processes affecting copper-coated nuclear waste containers in a deep geological repository. Prog. Mater Sci. 118, 100766.
Hartnett, T.Q., Ayyasamy, M.V., Balachandran, P.V., 2019. Prediction of new iodine-containing apatites using machine learning and density functional theory. MRS Commun. 3, 882–890.
Himanen, L., Geurts, A., Foster, A.S., et al., 2019. Data-driven materials science: status, challenges, and perspectives. Adv. Sci. 6 (21), 1900808.
Hu, M., Rutqvist, J., Steefel, C.I., 2021b. Mesh generation and optimization from digital rock fractures based on neural style transfer. J. Rock Mech. Geotech. Eng. 13 (4), 912–919.
Hu, G., Zhou, T., Liu, Q.F., 2021a. Data-driven machine learning for fault detection and diagnosis in nuclear power plants: A review. Front. Energy Res., 663296
Jain, A.K., Mao, J., Mohiuddin, K.M., 1996. Artificial neural networks: A tutorial. Computer 29 (3), 31–44.
Jeong, H., Sun, A.Y., Lee, J., et al., 2018. A learning-based data-driven forecast approach for predicting future reservoir performance. Adv. Water Resour. 118, 95–109.
Karniadakis, G.E., Kevrekidis, I.G., Lu, L., et al., 2021. Physics-informed machine learning. Nat. Rev. Phys. 3 (6), 422–440.
Kim, J.G., Jang, S.C., Kang, I.S., et al., 2020. A study on object recognition using deep learning for optimizing categorization of radioactive waste. Prog. Nucl. Energy 130, 103528.
Kim, K.H., Yun, S.T., Yu, S., et al., 2020. Geochemical pattern recognitions of deep thermal groundwater in South Korea using self-organizing map: Identified pathways of geochemical reaction and mixing. J. Hydrol. 589, 125202.
Konevnik, Y.V., Zakharova, E.V., Shiryaev, A.A., et al., 2020. Mineral-specific heterogeneous neptunium sorption onto geological repository rocks in oxic and anoxic conditions and various temperatures. Chem. Geol. 545, 119654.
Krems, R.V., 2019. Bayesian machine learning for quantum molecular dynamics. PCCP 21 (25), 13392–13410.
Krishnan, N.A., Mangalathu, S., Smedskjaer, M.M., et al., 2018. Predicting the dissolution kinetics of silicate glasses using machine learning. J. Non-Cryst. Solids 487, 37–45.
Kurniawan, T.A., Othman, M.H., Singh, D., et al., 2022. Technological solutions for long-term storage of partially used nuclear waste: A critical review. Ann. Nucl. Energy 166, 108736.
Laikari, A., Backman J. Industrial Internet of Things in Nuclear: Feasibility study. Energiforskrapport, 2021: 726.
Laloy, E., Jacques, D., 2019. Emulation of CPU-demanding reactive transport models: a comparison of Gaussian processes, polynomial chaos expansion, and deep neural networks. Comput. Geosci. 23 (5), 1193–1215.
Lillington, J.N., Goût, T.L., Harrison, M.T., et al., 2020. Predicting radioactive waste glass dissolution with machine learning. J. Non-Cryst. Solids 533, 119852.
Lu, X., Sargin, I., Vienna, J.D., 2021. Predicting nepheline precipitation in waste glasses using ternary submixture model and machine learning. J. Am. Ceram. Soc. 104 (11), 5636–5647.
Mahesh, B., 2020. Machine learning algorithms-a review. Int. J. Sci. Res. (IJSR) 9, 381–386.
Menke, H.P., Maes, J., Geiger, S., 2021. Upscaling the porosity-permeability relationship of a microporous carbonate for Darcy-scale flow with machine learning. Sci. Rep. 11 (1), 1–10.
Neeway, J.J., Rieke, P.C., Parruzot, B.P., et al., 2018. The dissolution behavior of borosilicate glasses in far-from equilibrium conditions. Geochim. Cosmochim. Acta 226, 132–148.
Othman, H., Sabrah, T., Marzouk, H., 2019. Conceptual design of ultra-high performance fiber reinforced concrete nuclear waste container. Nucl. Eng. Technol. 51 (2), 588–599.
Prasianakis, N.I., Haller, R., Mahrous, M., et al., 2020. Neural network based process coupling and parameter upscaling in reactive transport simulations. Geochim. Cosmochim. Acta 291, 126–143.
Schmeide, K., Rossberg, A., Bok, F., et al., 2021. Technetium immobilization by chukanovite and its oxidative transformation products: Neural network analysis of EXAFS spectra. Sci. Total Environ. 770, 145334.
Shastri, B.J., Tait, A.N., Ferreira de Lima, T., et al., 2021. Photonics for artificial intelligence and neuromorphic computing. Nat. Photonics 15 (2), 102–114.
Sirdesai, N.N., Singh, A., Sharma, L.K., et al., 2019. Development of novel methods to predict the strength properties of thermally treated sandstone using statistical and soft-computing approach. Neural Comput. Appl. 31 (7), 2841–2867.
Solans, V., Rochman, D., Brazell, C., et al., 2021. Optimisation of used nuclear fuel canister loading using a neural network and genetic algorithm. Neural Comput. Appl. 33 (23), 16627–16639.
Stanfill, B.A., Piepel, G.F., Vienna, J.D., et al., 2020. Nonlinear logistic regression mixture experiment modeling for binary data using dimensionally reduced components. Qual. Reliab. Eng. Int. 36 (1), 33–49.
Suh, Y.A., Hornibrook, C., Yim, M.S., 2018. Decisions on nuclear decommissioning strategies: Historical review. Prog. Nucl. Energy 106, 34–43.
Suh, J.W., Sohn, S.Y., Lee, B.K., 2020. Patent clustering and network analyses to explore nuclear waste management technologies. Energy Policy 146, 111794.
Sun Z., Wang L., Zhou J. Q., et al. A new method for determining the hydraulic aperture of rough rock fractures using the support vector regression. Eng. Geol., 2-020(271): 105618.

Tian, J., Qi, C., Sun, Y., et al., 2021. Permeability prediction of porous media using a combination of computational fluid dynamics and hybrid machine learning methods. Eng. Comp. 37 (4), 3455–3471.
Tosoni, E., Salo, A., Govaerts, J., et al., 2019. Comprehensiveness of scenarios in the safety assessment of nuclear waste repositories. Reliab. Eng. Syst. Saf. 188, 561–573.
Trivelpiece, C.L., Edwards, T.B., Johnson, F.C., et al., 2020. Method for estimating the density of high-level nuclear waste glass. Int. J. Appl. Glass Sci. 11 (4), 641–648.
Tsai, P.C., Huang, X., Hung, Y.H., et al., 2019. The computer aided cutting planning of components using genetic algorithms for decommissioning of a nuclear reactor. Ann. Nucl. Energy 130, 200–207.
Tudorache, A., Marin, C., Ionita, D.E., et al., 2018. Assessing barium and strontium concentrations and speciation in groundwater from the area of the future weak and medium radioactive waste repository Saligny-Romania, Carpath. Carpathian J. Earth Environ. Sci. 13, 57–66.
Viswanathan, H.S., Ajo-Franklin, J., Birkholzer, J.T., et al., 2022. From fluid flow to coupled processes in fractured rock: recent advances and new frontiers. Rev. Geophys. 60 (1) e2021RG000744.
Wang, K., Sun, W., 2018. A multiscale multi-permeability poroplasticity model linked by recursive homogenizations and deep learning. Comput. Methods Appl. Mech. Eng. 334, 337–380.
Wang, H., Wang, X., Tuo, X., et al., 2020. Experimental research on a Raman-based distributed temperature sensor assisted by PCA for locating the temperature abnormal event of nuclear waste drums. Appl. Opt. 59 (2), 579–588.
Wisnubroto, D.S., Zamroni, H., Sumarbagiono, R., et al., 2021. Challenges of implementing the policy and strategy for management of radioactive waste and nuclear spent fuel in Indonesia. Nucl. Eng. Technol. 53 (2), 549–561.
Xu, W., Zhang, Y., Li, X., et al., 2020. Extraction and statistics of discontinuity orientation and trace length from typical fractured rock mass: A case study of the Xinchang underground research laboratory site, China. Eng. Geol. 269, 105553.
Yamano, N., Inakura, T., Ishizuka, C., et al., 2021. Estimation of uncertainty in transmutation rates of LLFPs in a fast reactor transmutation system via an estimation of the cross-section covariances. J. Nucl. Sci. Technol. 58 (5), 567–578.
Yoon, S., Jeon, J.S., Kim, G.Y., et al., 2019. Specific heat capacity model for compacted bentonite buffer materials. Ann. Nucl. Energy 125, 18–25.