

# 기계학습법을 통한 압축 벤토나이트의 열전도도 추정 모델 평가

윤 석\* · 방헌태\*\* · 김건영\*\*\* · 전해민\*\*\*\*

Yoon, Seok\*, Bang, Hyun-Tae\*\*, Kim, Geon-Young\*\*\*, Jeon, Haemin\*\*\*\*

## Evaluation of a Thermal Conductivity Prediction Model for Compacted Clay Based on a Machine Learning Method

### ABSTRACT

The buffer is a key component of an engineered barrier system that safeguards the disposal of high-level radioactive waste. Buffers are located between disposal canisters and host rock, and they can restrain the release of radionuclides and protect canisters from the inflow of ground water. Since considerable heat is released from a disposal canister to the surrounding buffer, the thermal conductivity of the buffer is a very important parameter in the entire disposal safety. For this reason, a lot of research has been conducted on thermal conductivity prediction models that consider various factors. In this study, the thermal conductivity of a buffer is estimated using the machine learning methods of: linear regression, decision tree, support vector machine (SVM), ensemble, Gaussian process regression (GPR), neural network, deep belief network, and genetic programming. In the results, the machine learning methods such as ensemble, genetic programming, SVM with cubic parameter, and GPR showed better performance compared with the regression model, with the ensemble with XGBoost and Gaussian process regression models showing best performance.

**Key words :** Engineered barrier system, Compacted bentonite, Thermal conductivity, Machine learning method

### 초 록

완충재는 고준위 방사성 폐기물을 처분하기 위한 공학적 방벽 시스템에서 중요한 구성요소 중 하나이며 사용 후 핵연료가 담긴 처분용기와 암반 사이에 채워지는 물질이기 때문에 지하수 유입으로부터 처분용기를 보호하고, 방사성 핵종 유출을 방지하는 중요한 역할을 수행한다. 따라서 공학적 방벽 시스템의 처분용기로부터 발생하는 고온의 열량은 완충재를 통하여 전파되기에 완충재의 열전도도는 처분시스템의 안전성 평가에 매우 중요하다. 본 연구에서는 국내에서 생산되는 압축 벤토나이트 완충재의 열전도도 예측을 위한 경험적 회귀 모델의 정확성을 검증하고 정확도를 높이기 위해 예측모델의 구축에 기계학습법을 적용해 보았다. 벤토나이트의 건조밀도, 함수비 및 온도 값을 바탕으로 열전도도를 예측하고자 하였으며, 이때 다항 회귀, 결정 트리, 서포트 벡터 머신, 앙상블, 가우시안 프로세스 회귀, 인공신경망, 심층 신뢰 신경망, 유전 프로그래밍과 같은 기계학습 기법을 적용하였다. 기계학습 기법을 이용하여 예측한 결과, 부스팅 기반의 앙상블 기법, 유전 프로그래밍, 3차 함수 기반의 SVM, 가우시안 프로세스 회귀의 기계학습기법을 활용한 모델이 선형 회귀 분석 기법에 비해 좋은 성능을 보였으며, 특히 앙상블의 부스팅 기법과 가우시안 프로세스 회귀 기법을 사용한 모델들이 가장 좋은 성능을 보였다.

**검색어 :** 공학적 방벽, 벤토나이트 완충재, 열전도도, 기계학습법

\* 정희원 · 한국원자력연구원 방사성폐기물처분연구부 책임연구원 (Korea Atomic Energy Research Institute · syoon@kaeri.re.kr)

\*\* 정희원 · 한밭대학교 토목공학과 석사과정 (Hanbat National University · htbang@hanbat.ac.kr)

\*\*\* 한국원자력연구원 방사성폐기물처분연구부 책임연구원 (Korea Atomic Energy Research Institute · kimgy@kaeri.re.kr)

\*\*\*\* 정희원 · 교신저자 · 한밭대학교 건설환경공학과 부교수 (Corresponding Author · Hanbat National University · hjeon@hanbat.ac.kr)

Received May 25, 2020/ revised June 26, 2020/ accepted November 11, 2020

## 1. 서론

원자력 발전소의 연료로 사용된 핵연료는 방사선 준위가 높은 고준위 폐기물로써 지하 500~1,000 m 깊이의 암반층에 공학적 방벽의 개념에 의해 안전하게 처분되어야 한다. 공학적 방벽은 Fig. 1과 같이 사용 후 핵연료가 담긴 처분용기(Canister), 뒷채움재(Backfill), 완충재(Buffer), 근계영역 암반(Near-Field rock) 등으로 구성되어 있다(Choi et al., 2014). 이 중 완충재는 처분용기와 암반 사이에 채워지는 물질로써 물리적 충격으로부터 처분용기를 보호하고 지하수 유입을 최소화하여 처분용기 내의 핵종이 암반으로 유출되는 것을 저지하는 역할을 한다(Villar et al., 2006; Choi et al., 2014). 또한 완충재는 처분용기로부터 발생하는 붕괴열을 주위 암반으로 발산시키는 역할을 하는데 만약 붕괴열을 효과적으로 발산시키지 못하면 처분용기 내에 열이 축적되어 폐기물 및 처분용기의 장기건전성에 악영향을 끼친다(JNC 1999; Cho et al., 2011).

이에 따라 완충재에 적합한 후보물질을 선정하기 위해 다양한 연구가 이루어졌으며, 그 결과 주로 몬모릴로나이트의 벤토나이트가 가장 적합한 물질인 것으로 알려져 있다(Karland, 2010; Lee et al., 2011). 몬모릴로나이트는 알루미늄 8면체의 내부치환에 의해 충전하가 발생하는 광물이며, 층간 사이에 존재하는 교환성 양이온의 종류에 따라 Na형 벤토나이트와 Ca형 벤토나이트로 구분된다(Madsen, 1988). 현재 우리나라의 경우 경주에서 생산되는 Ca형 벤토나이트를 기반으로 연구를 수행하고 있다(Lee et al., 2014; Yoo et al., 2016). 압축 벤토나이트 완충재가 가져야 할 성능은 다양하며 이 중 완충재의 열전도도는 공학적 방벽의 전체 성능평가에 있어 매우 중요하기에 국내외적으로 벤토나이트 완충재의 열전도도에 관한 다양한 연구가 이루어지고 있다(Tang et al., 2008; Yoon et al., 2018; Cho, 2019; Xu et al., 2019).

열전도도는 재질의 열전달능력을 나타내는 물성으로써, 완충재는 붕괴열을 신속하게 주변 암반으로 전달시켜야하기 때문에 높은

열전도도를 가져야 한다(Cho, 2019). 벤토나이트와 같은 흙은 흙 입자, 물, 그리고 공기로 구성된 삼상 구조이며 흙의 열전도도는 물의 함량을 나타내는 함수비와 흙 입자의 무게와 관련된 건조 밀도에 큰 영향을 받는다. 이렇듯 벤토나이트 완충재의 열전도도는 크게 함수비와 건조밀도에 영향을 받기에 각국의 벤토나이트 특성을 반영하여 건조밀도와 함수비에 따른 다양한 열전도도 회귀 모델들이 제시되었다(JNC, 1999; Tang et al., 2008; Yoon et al. 2018). 이러한 열전도도 회귀 모델들은 공학적 방벽의 성능 평가를 위한 중요한 입력 변수로 적용되며 처분시스템에서 매우 중요한 역할을 하므로 이에 대한 정확성 검증이 요구된다.

따라서 본 논문에서는 열전도도에 미치는 다양한 변수와 제한된 개수를 가지는 데이터 세트에 활용되는 기계학습 기법들을 적용하여 그 성능을 선형적 회귀모델과 비교 및 검증하고자 하였다. 기계학습은 학습대상이 되는 데이터의 입력변수와 출력변수 사이에 존재하는 경향을 학습하여 모델을 구축하는 방법이다. 기존 회귀모델을 사용한 예측모델의 경우 데이터의 분포나 이상치에 의해 쉽게 영향을 받는 경향이 있었다. 반면에 기계학습기법은 예측 모델을 구축하기 위한 학습 과정에서 학습의 대상이 되는 데이터에 과적합되어 새로 입력되는 데이터의 예측 성능이 저하되는 현상을 방지하기 위해 다양한 방식의 규제들을 적용한다. 이러한 규제의 사용을 통하여 데이터 세트의 불확실성에도 불구하고 이상치에 강인한 모델의 구축이 가능하다.

## 2. 연구방법

### 2.1 벤토나이트 완충재

국내에서는 경주, 울산, 포항 일대에서 Ca형 벤토나이트가 생산되고 있기에, Ca형 벤토나이트는 국내 고준위폐기물 처분장의 공학적 방벽의 구성요소 중 하나인 완충재로 매우 유력한 후보물질이라 할 수 있다. 한국원자력연구원에서는 국내 Ca형 벤토나이트에 대해 2015년도 이전까지 적용된 것은 KJ- I 으로 2015년도 이후에 적용된 것을 KJ-II로 명명하고 있으며, 본 연구에서는 KJ-II 벤토나이트를 사용하였다. 국내 Ca형 벤토나이트 분말을 통일분류법에 의거하여 분류한 결과 고소성 점토인 CH로 분류되었으며, 액성한계와 소성한계는 각각 146.7 %, 28.4 %로 나타났다(Yoon et al., 2018). 국내 벤토나이트의 광물 조성 분석 결과 몬모릴로나이트가 약 60 %를 차지하였으며, 장석은 약 20 %, 석영은 약 5 %를 나타냈다(Yoo et al., 2016). 또한 화학적 조성 분석 결과 SiO<sub>2</sub>가 약 60 %로 가장 많은 성분을 차지하였으며, CaO는 NaO 성분보다 약 5~6배 높은 비율을 나타냈다(Yoo et al., 2016). 벤토나이트가 완충재로 적용 시에는 분말 상태가 아닌 분말을 압축한 블록의 형태로 적용된다.

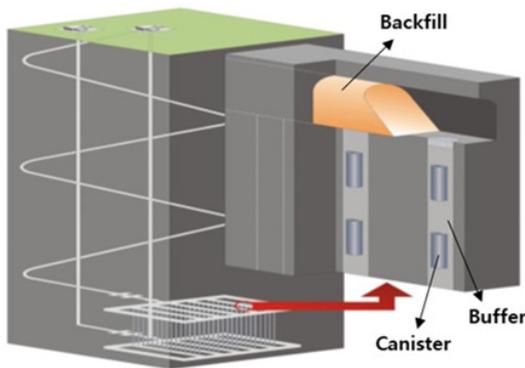


Fig. 1. Engineered Barrier System for HLW Disposal (Yoon and Kim, 2021)

## 2.2 열전도도 회귀 모델

벤토나이트 완충재의 열전도도는 처분시스템의 안정해석을 위한 가장 중요한 인자 중 하나이다. 벤토나이트 완충재의 건조밀도 및 함수비에 따른 열전도도 특성 규명에 관한 연구들이 많이 진행되어 있지만, 실제 처분 환경에서의 고온의 온도변화에 따른 열전도도 변화에 대한 연구는 거의 진행된 적이 없기에 한국원자력연구원에서는 상온에서 80-90도 까지 온도 변화에 따른 벤토나이트 완충재의 열전도도 변화를 측정하여 총 147개의 데이터 세트를 구축하였으며, 이를 사용하여 Eq. (1)과 같이 건조밀도, 함수비 및 온도에 따른 열전도도 회귀식을 제안하였다.

$$\lambda = -1.198 + 2.628\omega + 1.055\gamma_d + 0.030\ln(T) \quad (1)$$

여기서  $\omega$ 는 함수비,  $\gamma_d(\text{g/cm}^3)$ 는 건조밀도를 뜻하며  $T(^{\circ}\text{C})$ 는 온도를 나타낸다. 본 회귀식은 모든 회귀분석의 가정 조건을 충족하였고 R<sup>2</sup>값은 0.925였으며, RMSE 값은 0.054 값을 나타냈다. Eq. (1)에 사용된 147개 데이터의 통계량은 Table 1과 같으며, 벤토나이트 완충재의 열전도도는 Fig. 2와 같이 건조밀도, 함수비 및 온도변화에 비례한 것으로 알려져있다(Xu et al., 2019).

## 2.3 기계학습법 이론

기계 학습은 특정 목적을 위해 입력된 데이터 세트를 바탕으로 학습을 진행하여 시스템을 형성하고 개선해 나가는 프로세스로 정의할 수 있고(Mitchell, 1997), 학습 데이터 세트의 특성에 따라 입력되는 데이터 세트에 목표로 하는 값을 포함하여 학습하는 지도 학습과 목표로 하는 값을 학습하지 않고 데이터 세트의 분포에 기반을 두어 군집을 만들거나 특징을 추출하는 비지도 학습으로 분류된다. 지도 학습에서는 입력 데이터를 목표 값에 매핑(mapping)하여 대상 데이터의 경향과 분포에 대한 학습을 수행하고 예측 시스템을 구축하며, 사용하는 목표 값의 유형에 따라 다시 분류와 회귀로 나뉜다. 비지도 학습은 목표 데이터가 학습 데이터에 포함되지 않는 방법으로, 입력 데이터를 이용하여 군집을 만들거나 데이터의 분포 또는 특성에 따라 관계를 분석하는 방식으로 학습을 한다(Géron, 2017). 본 논문에서는 벤토나이트의 열전도도 예측모델 구축을 위하여 지도 학습에 기반을 둔 모델 학습 알고리즘을 제안하고 각각의 기법들의 성능을 검증한다. 기존의 선형 회귀 모델에 대비하여 성능을 검증하고 개선하기 위하여 다항 회귀, 결정 트리, 서포트 벡터 머신(Support Vector Machine, SVM), 앙상블, 가우시안 프로세스 회귀(Gaussian Process Regression, GPR), 인공신경망, 심층 신뢰 신경망(Deep Belief Network, DBN) 및 유전 프로그래밍과 같은 다양한 기계학습 기법을 적용하였다. 함수비, 건조밀도 및 온도의 입력 데이터에 따른 열전도도의

데이터는 한국원자력연구원에서 실험을 통해 구축된 데이터 세트를 사용하였으며, 과적합(overfitting)을 방지하기 위해 5겹 교차검증을 사용하여 검증을 실시하였다.

### 2.3.1 다항 회귀

다항 회귀 모델은 입력 변수들과 목표 값의 관계에 대한 평균 제곱 오차(Mean Square Error, MSE)를 최소화하는 회귀 방정식을 추정하여 모델을 구축한다. 모델을 구축하고 최적화하기 위해 임의의 값들을 사용하여 초기 모델을 구축한 후 평균 제곱 오차를 사용한 비용함수를 정의하고, 학습을 반복하면서 비용함수를 최소

Table 1. Statistical Quantities for the Compacted Bentonite Dataset

	Minimum	Maximum	Mean	Standard deviation
Dry density (g/cm <sup>3</sup> )	1.314	1.836	1.636	0.116
Water content	0.000	0.234	0.072	0.067
Temperature (°C)	25	87	44.847	22.203
Thermal conductivity (W/(m·K))	0.388	1.536	0.834	0.227

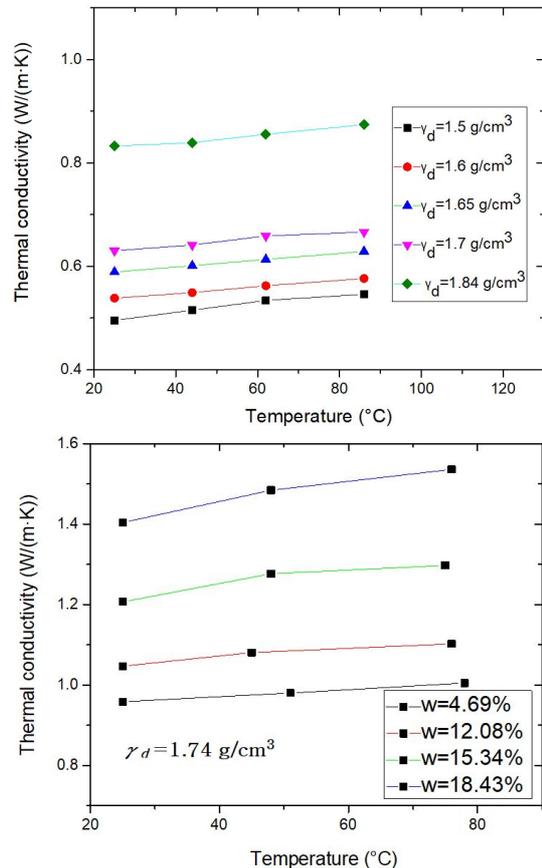


Fig. 2. Thermal Conductivity Variation of the Compacted Bentonite

화하는 방향으로 회귀 방정식을 수정하여 최적의 계수들을 추정한다. 회귀 방정식을 추정하는 과정에서 과적합을 방지하고 입력변수와 목표값 간의 관계성을 고려하기 위해 다양한 방법을 사용하게 된다(Bang et al., 2020).

### 2.3.2 결정 트리

결정 트리 모델은 나무가 단일 줄기에 여러 개의 가지를 만드는 것과 같은 방식으로 데이터를 분류하는 방법으로, 데이터의 특정 숫자 값 또는 조건을 기반으로 분기를 만들어 데이터를 나누는 프로세스를 반복하여 회귀 분석을 수행한다. 결정 트리 방법의 경우 회귀분석을 위해 입력데이터를 일정한 기준으로 나누고 나누어진 그룹을 다시 일정한 기준을 사용하여 세분화하는 과정을 반복하여 목표값과 가장 잘 대응되는 기준을 가진 모델을 찾는 방식으로 모델의 최적화가 이루어진다. 이때 데이터를 나누기 위한 수치 또는 조건은 지니 불순도(Gini Impurity) 또는 교차엔트로피(Cross Entropy)를 사용하여 정보이득을 계산하여 결정한다. 회귀모델의 경우 MSE를 사용하여 기준을 정하기도 한다(Géron, 2017).

### 2.3.3 서포트 벡터 머신(Support Vector Machine, SVM)

SVM 모델은 학습하는 데이터 세트에서 특정 범위의 분산에 가지는 선형 함수를 추정하고 추정된 함수가 가지는 특정 범위의 분산에 최대한 많은 데이터가 포함되는 방향으로 최적화를 수행한다. 이를 위해 데이터 세트에서 경계가 되는 위치에 있는 점을 지지 벡터로 정의하고 이를 바탕으로 함수의 추정과 최적화를 수행한다(Smola and Schölkopf, 2004). 따라서 SVM 모델은 지지 벡터로 설정된 점들 사이에 가장 많은 데이터가 포함될 수 있도록 최적의 지지 벡터를 선정하고 지지 벡터의 중심을 통과하는 선형의 함수를 추정하는 방식으로 모델을 생성한다. 선형의 함수를 사용하여 모델을 제작하기 때문에 비선형의 데이터에 적용하기 위해 커널을 설정하여 차원을 변경하거나 가우시안 방사 기저함수를 통하여 비선형의 데이터에도 사용가능한 모델을 구축할 수 있다.

### 2.3.4 앙상블

앙상블 모델은 다양한 모델을 결합하여 단일 모델의 성능을 향상시킨다. 여러 개의 모델을 구축하고 그 결과를 평균 등의 방식으로 종합하거나 여러 번 반복적으로 모델을 구축하고 학습시키면서 이전 모델의 결과를 반영하여 가중치를 계산하고 학습한다. 결정 트리를 사용하여 앙상블 모델을 구축할 경우 배깅 및 부스팅과 같은 기법이 주로 사용된다. 배깅(Bagging)은 데이터 세트에서 반복을 허용하여 무작위로 데이터를 추출하고 여러 개의 결정 트리를 만들어 그 결과를 결합하는 기법이다. 부스팅(Boosting)은 데이

터를 반복적으로 추출하여 약한 학습기를 만들고 여러 단계에 걸쳐 가중치를 추가하여 학습에 적합하게 데이터를 변환하는 방식이다(Polikar, 2006).

### 2.3.5 가우시안 프로세스 회귀(Gaussian Process Regression, GPR)

가우시안 프로세스 회귀(GPR) 모델은 비모수 기반의 확률론적 모델로, 불확실한 분포를 고려하여 데이터 세트의 분포에서 예측되는 평균과 공분산의 함수를 이용하여 모델을 구축한다(Rasmussen and Williams, 2006). GPR 모델은 하나의 모델을 설정하여 예측을 하는 방식이 아닌 한 지점에서 유사한 분포를 지닌 데이터의 확률을 계산하여 예측한다. 그로인해 데이터가 가지는 분포를 표현하는 공분산의 함수가 중요하며, 공분산 함수 가정에 따라 모델의 성능이 결정된다.

### 2.3.6 인공신경망 및 심층 신뢰 신경망(Deep Belief Network, DBN)

인공신경망은 인간의 뇌를 구성하는 신경망 구조를 모방한 모델이다. 신경망을 구성하는 뉴런은 상호 간에 전기 신호를 교환하고, 인간은 이를 통해 물체나 사건을 인식하고 판단한다. 인공 신경망은 이러한 신경망의 구조를 가중치와 편향을 이용하여 인위적으로 구성하고 이를 바탕으로 예측을 수행한다. 데이터가 신경망을 통과하면, 인공 신경망을 구성하는 뉴런들을 거치며 가중치가 부여되고 이를 갱신하는 형태로 학습을 반복하며 가중치가 업데이트된다(Géron, 2017). 인공신경망을 구성하는 뉴런들을 노드라 부르는데 데이터가 인공신경망에 입력되면 노드들을 거치면서 가중치와 편향을 부여받게 된다. 이렇게 부여된 가중치와 편향이 반영된 값들은 연결된 노드에 추가적인 가중치와 편향을 부여받으며 전달되거나 최종 결과로서 출력되게 된다. 이때 노드를 거친 값의 최종적인 출력을 결정하기 위해 활성화 함수가 사용된다. 심층 신뢰 신경망은 인공신경망을 다층으로 확대하면서 가중치의 학습이 원활하게 이루어지지 않는 문제를 해결하기 위해 제안된 학습방법이다. 학습을 위해 입력된 데이터 세트의 변수들을 재구성 하면서 데이터 세트가 가지는 분포를 파악하고, 이를 인공신경망에 통과시켜 추가적인 미세조정을 수행하여 예측 모델을 구축한다(Fischer and Igel, 2014).

### 2.3.7 유전 프로그래밍

유전 프로그래밍은 생물학적 진화 과정에서 착안된 기법으로 생물학적 유전 알고리즘의 방식과 유사하게 선택, 교차, 변이, 대치의 과정을 거치면서 최적의 해 집단을 선정한다(Eberhart and Shi, 2011). 유전 프로그래밍이 최적의 해 집단을 결정하는 과정은 초기에 임의의 해 집단을 선택한 후 적합도를 기준으로 집단의 구성원 중 일부를 선택하고 이들을 사용하여 새로운 구성원을

만들어 기존의 구성원 일부를 대체하는 방식을 반복하면서 이루어진다. 이 과정에서 잘못된 값으로 수렴하는 것을 방지하기 위해 새로운 구성원을 만드는 과정에서 일정한 확률로 변이가 발생하게 하여 다양성을 높이는 방법을 사용한다.

### 3. 적용 결과

기계학습법을 사용하여 구축된 벤토나이트 완충재의 열전도도 예측모델의 성능측정을 위해 데이터 분석에서 많이 사용되는 성능 평가 지표 중 하나인 평균 제곱근 오차(Root Mean Square Error, RMSE)를 계산하여 예측 성능을 평가하였다. 예측 모델의 구축을 위해 선형 회귀모델 구축에 사용되었던 Table 1과 같은 통계량을 가지는 건조밀도, 함수비, 온도의 입력변수에 따른 열전도도의 값을 저장한 147개의 데이터 세트를 사용하였다. 입력변수들 각각이 가지는 값의 차이가 상대적으로 큰 경우에는 하나의 입력변수에 상대적으로 큰 가중치가 부여될 위험이 있으므로, 예측 모델의 정확도 향상을 위해 온도에 자연로그를 취해 다른 입력변수들과 비슷한 크기를 가지도록 조정하였다. 학습을 위해 다항 회귀, 결정 트리, SVM, 앙상블, 가우시안 프로세스 회귀, 인공신경망, 심층 신뢰 신경망, 유전 프로그래밍의 학습 알고리즘이 사용되었고 각각의 알고리즘에 따라 매개변수가 다른 모델을 구축하여 성능을 평가하였다. 성능 평가의 정확도를 향상하고 검증과정에서 모델에 과적합이 발생하는 것을 방지하기 위해 데이터를 K개의 그룹으로 나누어 하나의 그룹을 테스트를 위한 그룹으로 사용하고 나머지를 학습을 위한 그룹으로 사용하여 K번 반복하여 검증하는 K겹 교차 검증을 사용하여 결과의 유효성과 과적합 여부를 검증하였다. K를

5로 설정하고 147개의 데이터 세트를 5개로 나누어 테스트를 위해 추출하는 데이터를 바꾸어 가면서 성능평가를 수행하였다. 최종 성능은 5개의 테스트그룹 각각에서 RMSE를 산출하여 그 평균을 구하여 평가하였다. 앙상블의 랜덤포레스트와 XGBoost를 사용한 모델과 인공신경망 및 심층 신뢰 신경망 모델의 구축을 위해서는 Python의 라이브러리를 사용하였고 나머지 기계학습 모델은 MATLAB으로 구축하였다.

다양한 기계학습의 방법론을 사용하여 벤토나이트 완충재의 열전도도 예측 모델을 구축하고 각각의 성능을 RMSE로 평가하여 비교한 결과 Table 2와 같이 선형 회귀모델의 RMSE인 0.054에 비하여 부스팅 기법 기반의 앙상블, 유전프로그래밍, 3차 함수 기반의 SVM, 가우시안 프로세스 회귀를 포함한 기계학습 모델에서 높은 성능을 보여주었다. 또한, 각각의 기계학습 방법론에서 가장 좋은 성능을 보여준 모델의 예측 값과 실험을 통해 측정된 값은 Fig. 3과 같은 관계를 보여주었다. 특히 부스팅 기법에 바탕을 둔 방법론을 사용한 앙상블과 5/2를 곱한 Matern함수, 유리2차 함수 등의 공분산 함수를 사용한 가우시안 프로세스 회귀 및 유전 프로그래밍 모델에서 전반적으로 높은 성능을 보여주었다.

#### 3.1 다항 회귀 모델

다항 회귀 방법을 이용하여 최적의 예측 모델을 구축하기 위해 최소제곱법을 사용한 상호작용항 추가, 로버스트 가중치 추가, 단계적 변수적용 방법을 사용하여 세 개의 모델을 구축한 결과 Eqs. (2)-(4)와 같은 회귀식이 도출되었다. 식에서  $\omega$ 는 함수비,  $\gamma_d$ 는 건조밀도( $g/cm^3$ )이고  $T(^\circ C)$ 는 온도이며  $\lambda$ 는 열전도도 ( $W/mK$ )이다.

Table 2. Root Mean Square Error of Thermal Conductivity Estimation Using the Machine Learning Methods

Method	Parameter	RMSE	Method	Parameter	RMSE
Empirical Regression	—	0.054	Support Vector Machine	Linear	0.073
Polynomial Regression	interactions	0.060		Quadratic	0.051
	robust	0.072		Cubic	0.048
	stepwise	0.060		Gaussian RBF similarity : 0.35	0.175
Decision Tree	Min. leaf size : 4	0.103		Gaussian RBF similarity : 1.4	0.066
	Min. leaf size : 12	0.111		Gaussian RBF similarity : 5.7	0.079
	Min. leaf size : 36	0.174	squared exponential	0.048	
Ensemble	Bagging	0.221	Gaussian Processes Regression	5/2 Matern	0.044
	Random Forest	0.083		Exponential	0.049
	LSBoost	0.045		Rational Quadratic	0.044
	XGBoost	0.043	Neural Network	(20, 25)	0.064
Genetic Programming	tournament	0.045	Deep Belief Network	(100)	0.062

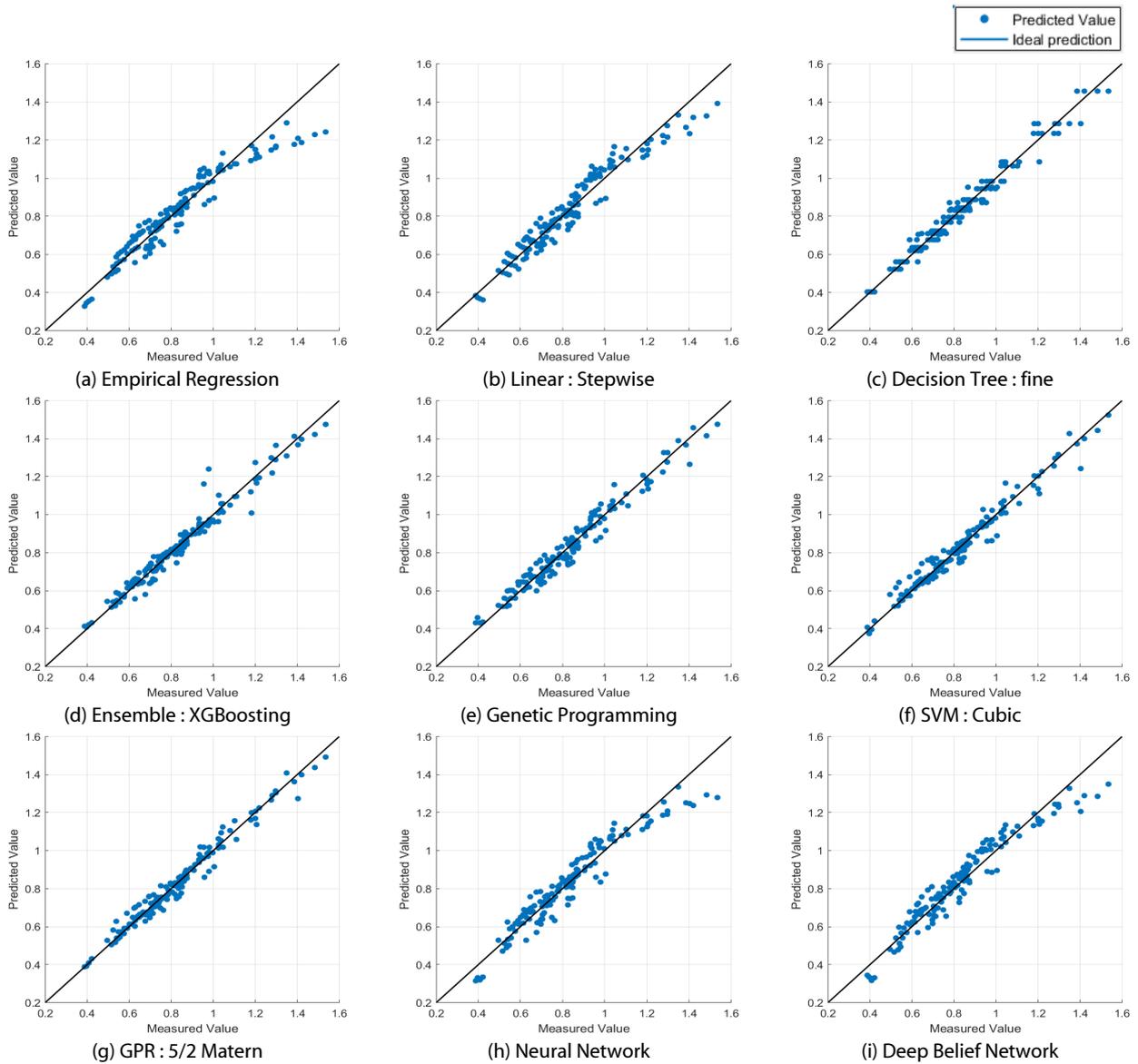


Fig. 3. Estimation Results of the Best Performance Models On Each Machine Learning Method

$$\lambda = -0.054 - 5.618\omega + 0.464\gamma_d - 0.210\ln(T) + 3.170\omega\gamma_d + 0.896\omega\ln(T) + 0.116\gamma_d\ln(T) \quad (2)$$

$$\lambda = -1.180 + 2.645\omega + 1.046\gamma_d + 0.028\ln(T) \quad (3)$$

$$\lambda = -0.785 - 5.452\omega + 0.906\gamma_d - 0.018\ln(T) + 3.122\omega\gamma_d + 0.873\omega\ln(T) \quad (4)$$

최소제곱법은 목표값과 예측값 사이의 차이를 최소화하기 위해 평균 제곱 오차를 구하고 그 값을 최소화하는 방법을 사용하여 데이터에 적합한 회귀식을 추정하는 방법으로서 이를 기반으로 다항 회귀 모델을 구축하였다. Eq. (2)는 입력변수 상호간의 곱을 새로운

특성으로 추가하여 입력변수들 간의 상호작용을 반영한 회귀식을 도출하였다. Eq. (3)의 로버스트 가중치 방법은 최소제곱법 과정에서 산출되는 오차에 따른 가중치를 계산하고, 가중치를 고려한 최소제곱법에 기반하여 회귀식을 추정하는 방법으로 특이치에 대한 고려가 반영된다는 점에서 Eq. (1)과 차이가 있다. 각 오차의 제곱에 주는 가중치는 바이스퀘어(bisquare) 가중치 함수를 이용하여 반복적으로 계산되어 회귀식의 최적화를 수행하는데 사용된다. 이러한 방법을 통해 데이터가 가진 특이치의 영향을 줄임으로서 이상치에 강인한 회귀모델을 구축할 수 있다. Eq. (4)는 출력변수인 열전도도와 입력변수들 간의 상관도를 산출하고 상관도가 높은 입력변수부터 단계적으로 입력변수를 반영하여 도출된 회귀식이다.

각각의 방법을 사용하여 학습을 진행하고 성능을 평가한 결과 상호작용항 추가 방법과 단계적 변수적용 방법을 사용한 모델이 다른 다항 회귀 모델에 비하여 좋은 결과를 보여주었다.

### 3.2 결정 트리 모델

결정 트리는 특정한 기준에 맞추어 데이터를 다수의 집단으로 분류하는 방식으로 회귀 분석을 수행한다. 구축되는 결정 트리 예측 모델의 성능은 최종적으로 분류된 데이터의 집단을 구성하는 최소한의 데이터 수와 분류된 데이터 집단을 세부적으로 반복하여 분류하는 횟수에 의해 결정된다.

본 논문에서는 데이터 집단을 반복하여 분류하는 횟수에는 제한을 두지 않았고 최종적으로 분류된 데이터 집단을 구성하는 최소한의 데이터의 수를 나무의 잎사귀에 비유하여 최소 리프 크기라고 하는데, 이 최소 리프 크기를 3가지 조건으로 차이를 두어 예측 모델을 구축하였다. 최종 분류된 데이터 집단을 구성하는 최소의 데이터를 각각 4, 12, 36개로 설정하여 3개의 예측모델을 구축하였고 최종 구성 분류 그룹에서 최소 4개의 데이터를 가지는 조건의 모델이 가장 좋은 결과를 보여주었다.

### 3.3 서포트 벡터 머신(Support Vector Machine, SVM) 모델

서포트 벡터 머신은 대상이 되는 데이터 세트에 잡음과 이상치가 있을 경우를 고려하여 일정량의 오차를 허용할 수 있는 소프트 마진 SVM을 사용하였다. 일정량의 오차를 결정하고 제어하기 위해 슬랙변수를 도입하여 Eq. (5)과 같은 목표함수를 최소화하는 회귀함수를 설정하였으며, 분산의 범위는 Eq. (6)과 같이 설정하여 학습하였다(Vapnik, 1995). Eqs. (5) and (6)에서  $x_i$ 와  $y_i$ 는 각각 데이터 세트의 입력변수와 출력변수이고( $i = 1:n_d$ ,  $n_d$ 는 데이터의 개수),  $W$ 와  $b$ 는 각각 데이터 공간상의 최적 평면에서 입력 변수인  $x_i$ 가 가지는 가중치와 편향이 된다.  $\varepsilon$ 는 데이터의 경계를 나타내는 서포트 벡터들과 회귀식 간의 거리로 정의되고  $W$ ,  $b$ 로 구성되는 식이  $\varepsilon$ 보다 작은 오차를 가지도록 하는 편차의 크기이다.  $\xi_i, \xi_i^*$ 는 데이터 세트 내의 모든 데이터에서 0보다 큰 값을 가지며, 데이터 세트 내의 잡음을 고려하기 위해 정의된 슬랙변수로서 허용가능한 일정량의 오차를 고려한  $\varepsilon$ 값을 설정할 수 있게 한다. Eq. (5)의  $C$ 는 상수인 매개변수로 슬랙변수가 허용할 수 있는 잡음과 편차를 제어한다.

본 논문에서는 실험 데이터가 가우시안 분포를 따른다고 가정하고  $\varepsilon$ 와  $C$ 를 학습하는 데이터가 가지는 표준편차  $\sigma$ 를 사용하여 각각  $0.1\sigma$ 와  $\sigma$ 로 설정하였다. 또한 서포트 벡터머신의 회귀식이 비선형을 가질 수 있도록 확장하기 위해 입력변수를 변환하거나 가우시안 방사 기저 함수를 이용한 유사도에 관련한 입력변수를 추가하여 모델을 구축하였다.

$$\text{Minimize } \frac{1}{2} W^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (5)$$

$$\text{Subject to } y_i - \langle W, x_i \rangle + b \leq \varepsilon + \xi_i \\ \langle W, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \quad (6)$$

따라서 선형적 모델 이외에 2차, 3차로 비선형 회귀식을 가질 수 있도록 확장한 모델과 가우시안 방사 기저 함수(Gaussian Radial Basis Function, Gaussian RBF)를 이용한 입력변수를 추가하기 위해 방사 기저 함수가 가지는 유사도를 0.35, 1.4, 5.7로 제한한 모델을 포함하는 총 6개의 서로 다른 조건을 가진 모델을 구축하고 각각의 성능을 비교하였다. 그 결과 3차 함수로 확장 가능하도록 입력변수를 변환한 모델에서 가장 좋은 성능을 보여주었다.

### 3.4 앙상블 모델

개별적인 다수의 모델을 조합하여 회귀 모델을 구축하는 다양한 앙상블 기법 중, 본 논문에서는 배깅과 부스팅의 기법에 기반한 4개의 모델을 구축하여 그 성능을 비교하였다. 첫 번째 모델은 학습을 위한 데이터 세트에서 반복을 허용하여 무작위로 데이터를 추출하고 이를 이용해 여러 개의 결정트리를 만드는 기본적인 배깅 기법을 적용하였다. 두 번째 모델은 배깅 기법에 기반을 두고 트리에서 데이터를 분할하는 기준을 무작위로 선정하게 하여 무작위성을 고려한 다수 개의 결정트리를 만들고 결합하는 랜덤포레스트 기법을 사용하였다. 세 번째 모델은 학습 데이터를 이용해 결정 트리 모델을 만들고 예측 값과 목표 값 간의 오차를 계산한 후, 그에 따른 가중치를 입력변수에 추가하는 경사하강법을 적용하여 모델의 구축과 가중치의 갱신을 반복함으로써 최적화하는 방식으로 학습을 진행하는 LSBoost기법을 사용하였다. 네 번째 모델은 약한 결정 트리 모델을 만들고 예측을 수행한 후 오분류가 발생한 데이터에 가중치를 추가하여 재학습하는 과정을 반복하는 기법인 XGBoost기법을 사용하였다.

이러한 4개의 서로 다른 방식을 사용한 앙상블 모델을 동일한 조건 하에서 학습을 진행하고 그 성능을 비교하였다. 본 논문에서 사용된 앙상블 방식은 다수의 결정 트리 모델을 구축하고 결합하는 방식으로 예측 모델을 구축하므로 다양한 모델을 빠르게 구축하기 위해 최종적으로 분류된 집단의 데이터 개수를 제한하지 않고 분류된 데이터 집단의 재분류를 반복하는 횟수를 3회로 제한하여 모델을 학습하였다. 또한 의사결정나무를 만들거나 가중치를 갱신하여 학습하는 과정은 100회를 반복하였으며 부스팅 기법을 사용하여 학습을 진행한 모델의 경우 학습률을 0.1로 설정하여 학습을 진행하였다. 구축한 4개의 모델 중 XGBoost기법을 이용하여 학습을 진행한 앙상블 모델에서 가장 좋은 결과를 보여주었다.

### 3.5 가우시안 프로세스 회귀(Gaussian Process Regression, GPR) 모델

가우시안 프로세스 회귀(GPR)는 학습되는 데이터 세트의 분포에 기반을 두어 모델을 구축하기 때문에 평균과 공분산의 함수를 정의하는 것이 모델의 성능에 큰 영향을 주게 된다. 데이터 세트에 의해 평균의 함수는 결정되지만 공분산의 함수의 경우 데이터 세트 분포의 불확실성을 표현하는 함수로서 매개변수로 정의되게 된다. 따라서 공분산 함수의 결정은 모델의 성능을 결정하는 핵심적인 요소가 된다. 본 논문에서는 제곱지수, 5/2를 곱한 Matern, 지수, 유리 2차의 4가지 형태의 공분산 함수를 가정 한 후 예측 모델을 학습하고 성능을 비교하였다. 그 결과 5/2를 곱한 Matern 함수를 공분산의 함수로 가정했을 때 가장 좋은 성능을 보임을 확인하였다.

### 3.6 인공신경망 및 심층 신뢰 신경망(Deep Belief Network, DBN) 모델

인공신경망은 네트워크의 구조를 형성하는 노드나 층의 개수에 따라 구성되는 가중치와 편향의 개수와 모델의 복잡도가 달라진다. 모델의 복잡도가 과도하게 높은 경우 가중치의 갱신이 원활하게 이루어지지 못하거나 과적합이 일어날 가능성이 있으므로 학습하는 데이터의 숫자를 고려하여 한 개의 층에 30개 이하의 노드를 가지는 2-3층의 네트워크를 가정하고 휴리스틱하게 최적의 조건을 찾은 결과 2개의 층을 가지고 각 층마다 (20, 25)개의 노드를 가진 인공신경망 모델이 가장 좋은 성능을 보여 주었다.

심층 신뢰 신경망(DBN)은 제한 볼츠만 머신(Restricted Boltzmann Machine, RBM)을 사용하여 출력값에 따른 입력변수를 재구성하여 데이터 세트의 분포를 추정하고 인공신경망을 통해 미세조정된 거처 예측 모델을 구축하게 된다. 이 중의 과정을 거치기 때문에 복잡성을 줄이기 위해 1개의 층과 100개의 노드를 가진 네트워크를 휴리스틱하게 결정하여 학습을 진행하였다. 학습을 위한 DBN내부에서 동작하는 RBM의 매개변수로서 학습률과 학습 횟수는 각각 0.01과 20으로 설정하였고 RBM을 거친 모델의 미세조정을 위해 사용되는 인공신경망의 학습률과 학습 횟수는 각각 0.01과 200으로 설정하였다. 각각의 과정에서 사용되는 활성화 함수로는 Relu (Rectified Linear Unit)를 사용하였고 데이터 수가 제한적이므로 한 번에 학습하는 데이터의 수는 1개로 설정하여 학습을 수행하고 모델을 구축하였다.

### 3.7 유전 프로그래밍 모델

유전 프로그래밍은 선택, 교차, 변이, 대치의 과정을 거치면서 모델의 최적화를 수행하게 된다. 이때 주요한 매개변수로 작용하는 것이 하나의 세대를 이루는 구성원의 수와 다음세대를 구성하기 위한 교차를 위한 구성원의 선택이다.

본 논문에서는 하나의 세대를 이루는 구성원의 수를 150으로 설정하였다. 다음세대의 구성을 위한 교차의 경우 토너먼트 방식을 사용하고 구성원 중 4개의 개체를 선정하여 그 중 다음세대를 위한 개체를 선택하는 것을 반복하여 다음세대를 구성하였으며, 새로운 세대를 구성하며 250회를 반복하여 최적화를 수행하여 예측 모델을 구축하였다.

## 4. 결론

본 연구에서는 고준위의 방사성 폐기물을 심층 처분하는데 있어서 가장 중요한 구성요소 중 하나인 벤토나이트 완충재의 열전도도 추정 모델에 기계학습법을 적용하여 제시하고자 하였다. 이를 통해 선형적 회귀모델과 비교하여 기계학습을 적용한 모델의 성능을 검증하고 이를 이용하여 향후 열-수리-역학 복합거동 모델개발의 활용가능성을 확인하고자 하였다. 함수비, 건조밀도, 그리고 온도 변화를 독립변수로 이용하여 총 147 케이스에 대한 열전도도 회귀모델을 제시하였으며, 앙상블 모델의 XGBoost 기법이 가장 적은 RMSE 값(0.043)을 보임을 확인하였다. 이는 선형 회귀모델의 RMSE값(0.054)보다 20 %이상 감소된 값으로, 기계학습법 적용 결과 정확도가 보다 향상된 열전도도 모델을 도출할 수 있음을 알 수 있었다. 또한 선형 회귀모델이 열전도도가 높아질수록 성능이 저하되는 데 반해 기계학습을 사용한 모델은 대부분의 구간에서 비교적 동일한 성능을 보여주었다. 특히 부스팅 기법에 바탕을 둔 방법론을 사용한 앙상블과 가우시안 프로세스 회귀(GPR), 유전 프로그래밍 모델에서 상대적으로 높은 성능을 보여주었다. 부스팅 기법, 가우시안 프로세스 회귀(GPR), 유전 프로그래밍 모델은 모두 학습하는 데이터에 따른 세부적인 최적화가 가능한 모델이다. 부스팅 기법의 경우 결정 트리 모델을 구축하고 모델에서 잘못 예측된 값들에 가중치를 부여하거나 오치의 함수를 합산하여 학습을 진행하는 방식이고, 가우시안 프로세스 회귀(GPR)의 경우는 주변 데이터를 사용하여 데이터의 분포를 추정하는 방식으로 예측을 수행한다. 유전 프로그래밍의 경우도 세대를 반복하면서 최적화를 진행하면서 데이터세트 전체를 사용한 최적화가 진행된다. 본 논문에서는, 상대적으로 높은 열전도도를 가지는 실험 데이터가 부족하여 충분히 학습이 불가능 했던 부분을 기계학습의 학습 기법을 통해 보완할 수 있는 장점을 가진 모델들에서 상대적으로 높은 성능을 보인 것으로 확인된다. 이러한 결과를 바탕으로 기계학습의 방법을 사용할 경우 상대적으로 적은 수의 데이터를 사용하여 선형 회귀모델을 개선할 수 있음을 보여준다. 향후에는 기계학습법을 통하여 벤토나이트 완충재의 열전도도뿐만 아니라 공학적방법제의 열-수리-역학 복합거동 모델을 개발하는데도 충분히 활용될 수 있을 것으로 기대된다.

## 감사의 글

본 연구는 과학기술정보통신부 한국연구재단의 원자력연구개발사업(2021M2C9A1018633)과 개인기초연구사업(2020R1F1A072379) 지원에 의해 수행되었으며, 이에 깊은 감사를 드립니다.

## References

- Bang, H. T., Yoon, S. and Jeon, H. (2020). "Application of machine learning methods to predict a thermal conductivity model for compacted bentonite." *Annals of Nuclear Energy*, Vol. 142, 107395.
- Cho, W. J. (2019). *Bentonite barrier material for radioactive waste disposal*, Korea Atomic Energy Research Institute Report, KAERI/GP-535/2019 (in Korean).
- Cho, W. J., Lee, J. W. and Kwon, S. K. (2011). "An empirical model for the thermal conductivity of compacted bentonite and a bentonite-sand mixture." *Heat and Mass Transfer*, Vol. 47, pp. 1385-1393.
- Choi, H. J., Kim, K. S., Cho, W. J., Lee, J. O. and Choi, J. W. (2014). *HLW long-term management system development: Development of engineered barrier system performance*, Korea Atomic Energy Research Institute Report, KAERI/TR-3859/2014 (in Korean).
- Eberhart, R. C. and Shi, Y. (2011). *Computational intelligence: concepts to implementations*, Elsevier, Amsterdam, Nederland.
- Fischer, A. and Igel, C. (2014). "Training restricted Boltzmann machines: An introduction." *Pattern Recognition*, Vol. 47, No. 1, pp. 25-39.
- Géron, A. (2017). *Hands-on machine learning with Scikit-Learn and TensorFlow*, O'Reilly Media, Inc., California, USA.
- Japan Nuclear Cycle Development Institute (JNC) (1999). *H12 project to establish technical basis for HLW disposal in Japan*, Supporting Report 2. Japan Nuclear Cycle Development Institute.
- Karnland, O. (2010). *Chemical and mineralogical characterization of the bentonite buffer for the acceptance control procedure in a KBS-3 repository*, Svensk Kärnbränslehantering AB Report, SKB TR-10-60.
- Lee, J. O., Cho, W. J. and Kwon, S. K. (2011). "Thermal-hydro-mechanical properties of reference bentonite buffer for a Korean HLW repository." *Tunnel and Underground Space*, Vol. 21, No. 4, pp. 264-273 (in Korean).
- Lee, J. O., Lee, M. S., Choi, H. J., Lee, J. Y. and Kim, I. Y. (2014). *Establishment of the concept of buffer for an HLW repository: An approach*, Korea Atomic Energy Research Institute Report, KAERI/TR-5824 (in Korean).
- Madsen, F. T. (1988). "Clay mineralogical investigation related to nuclear waste disposal." *Clay Minerals*, Vol. 33, No. 1, pp. 109-129.
- Mitchell, T. (1997). *Machine learning*, McGraw Hill, New York, USA.
- Polikar, R. (2006). "Ensemble based systems in decision making." *IEEE Circuits and systems magazine*, Vol. 6, No. 3, pp.21-45.
- Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian processes for machine learning*, MIT Press, Cambridge, Massachusetts.
- Smola, A. J. and Schölkopf, B. (2004). "A tutorial on support vector regression." *Statistics and Computing*, Vol. 14, No. 3, pp. 199-222.
- Tang, A. M., Cui, Y. J. and Le, T. T. (2008). "A study on the thermal conductivity of compacted bentonite." *Applied Clay Science*, Vol. 41, pp. 181-189.
- Vapnik, V. (1995). *The nature of statistical learning theory*, Springer, New York, USA.
- Villar, M. V., Martin, P. L. and Barcala, J. M. (2006). "Modification of physical, mechanical and hydraulic properties of bentonite by thermo-hydraulic gradients." *Engineering Geology*, Vol. 81, No. 3, pp. 284-297.
- Xu, Y., Sun, D., Zeng, Z. and Lv, H. (2019). "Temperature dependence of apparent thermal conductivity of compacted bentonites as buffer material for high-level radioactive waste repository." *Applied Clay Science*, Vol. 174, pp. 10-14.
- Yoo, M., Choi, H. J., Lee, M. S. and Lee, S. Y. (2016). "Measurement of properties of domestic bentonite for a buffer of an HLW repository." *Journal of the Korean Radioactive Waste Society*, Vol. 14, No. 2, pp. 135-147 (in Korean).
- Yoon, S., Cho, W. H., Lee, C. S. and Kim, G. Y. (2018). "Thermal conductivity of Korean compacted bentonite buffer materials for a nuclear waste repository." *Energies*, Vol. 11, No. 9, 2269.
- Yoon, S. and Kim, K. Y. (2021). "Measuring thermal conductivity and water suction for variably saturated bentonite." *Nuclear Engineering and Technology*, Vol. 53, pp. 1041-1048.