Modeling of Safety Parameters for Low Level Radioactive Waste Repository Using Machine Learning Approaches

SARDAR ALI SHAH[*1], ABDUL MAJID[2], SAFDAR ALI[3]

[1] Pakistan Institute of Nuclear Science and Technology, Islamabad, Pakistan
[2] Pakistan Institute of Engineering and Applied Sciences, Islamabad, Pakistan
[3] Pakistan Atomic Energy Commission, Islamabad, Pakistan
**Corresponding author's email: sardarshah@gmail.com*

(Received on 20th April, 2016 Accepted on 6th September, 2016)

Abstract

The paper is aimed at the safety assessment of intermediate and low level radioactive waste (LLRW) disposal facility. Now a days extensive research is going on to develop safety assessment methodologies for radioactive waste disposal facilities. For disposal of low level radioactive waste, near surface disposal facility is assumed the preferred option. Safety assessment is helpful to get public confidence. The main objective of disposal of radioactive waste is to protect the human health and the environment from its worse effects. Therefore, it is necessary to manage the radioactive waste safely. In this work, machine learning (ML) approaches of support vector regression (SVR), generalized regression neural network (GRNN), artificial neural network (ANN) and multiple linear regressions (MLR) have been applied for the modeling of different safety parameters of LLRW disposal facility. Simulations have been performed to model the distribution coefficients (K_d), leaching rates (λ_l), and retardation factors (R_f) of radionuclide present in the RW. Experimentations are conducted in Matlab environment. Percentage absolute difference is used to evaluate the performance of the proposed models. The best results have been achieved by SVR and GRNN models with correlation coefficients R=0.99812 and 0.94773 for K_d, respectively. The performance of ML models is compared with conventional linear regression (LR) methods. Experiments highlights that the proposed ML models provide better results compared to conventional LR methods. This study is useful for the development and safety assessments of our national future assessment of low level radioactive waste disposal facility.

Keywords: Radioactive waste, Near surface disposal facility, Distribution coefficient, Retardation factor

INTRODUCTION

The proper disposal of radioactive waste is a problem which needs to be addressed properly to avoid any serious consequences. One of the main reasons is to protect humans and environment from the hazardous radiations emitted from waste. We are dealing with low level radioactive waste (LLRW) generated from nuclear power plants, medicine and research[1]. For the disposal of LLRW near surface disposal facilities are used contrary to high level radioactive waste (HLRW) which is disposed deep within the ground.

Theoretical methods use mathematical predictive models and these models are easy to simulate on computers. In this work, MLR, GRNN, ANN and SVR were simulated and the results are compared.

Since we are concerned mainly to near surface disposal facilities in this paper, so there is a possibility that water can enter into the disposal area in accidental scenario[2]. Due to which radio nuclides will be released either into rocks or into cover soil. In case they are released into rocks, radio nuclides released from the facility will become part of the underground fresh water. Now this water can affect human beings either directly or indirectly. This contaminated water may be used for farming which will affect us directly. Indirectly this water used for farming will affect us by using eatables like milk, meat etc of animals. In case water is released to cover soil, local residents will obviously be affected directly[3].

Distribution coefficient or partition coefficient is the ratio of the concentration of an element on a solid and the concentration in the liquid phase (water etc.) [4].

$$Distribution\ coefficient = \frac{Adsorbed\ Concentration}{Dissolved\ Concentration} \quad (1)$$

Retardation of the contaminant can be estimated using distribution coefficient K_d. K_d model is the simplest and robust model. K_d values are empirical and represent a very simplest model of sorption or attenuation on soil. In general all isotopes of an element have the same K_d value, because sorption is a chemical property which is not affected by atomic mass or nuclear emissions. Also K_d values are highly dependent on environmental factors such as pH, particle size distribution and temperature etc[5].

The retardation factor (R_f) is commonly used in transport models. It describes the chemical interaction between the contaminant and geological materials (such as soil,

sediments, rocks, simply referred to as soils). Specifically, retardation factor indicates the rate at which contaminant is transported from one place to other [6].

Higher the bulk density, higher is the retardation factor. For high adsorption coefficient, retardation coefficient is high. Similarly if water content is low, retardation factor is also low.

$$R_f = 1 + \frac{\rho_b K_d}{\theta} \quad (2)$$

where

ρ_b= bulk density of medium (Kg/m^3),
K_d= distribution or partition coefficient (m^3/Kg),
θ= Total porosity of the medium.

The ratio of the amount of radio nuclides released from a given facility to the amount remaining in the waste is called leaching rate. We assume that leaching of radio nuclides (partitioned into the pore water) is due to the steady-state infiltration or drainage through the waste. Equation (3) can be used for leaching rate.

$$\lambda_l = \frac{q}{z(\theta_w + \rho_b K_d)} \quad (3)$$

where

q = the rate of drainage of water through the waste forms (m/y),
θ_w = the water filled porosity of the waste form ,
ρ_b = the bulk density of the waste form (kg/m^3),
K_d = the waste form distribution coefficient (m^3/ kg),
z = the height of the waste form (m).

Materials and Methods

Reduction and assessment of migration of radio nuclides present in waste is done by safety assessment of radioactive waste disposal facility. As safety assessment of radioactive waste is a very vast field, some safety parameters like distribution coefficient (K_d), retardation factor (R_f) and leaching rate (λ_l) are modelled in this paper. Different machine learning (ML) approaches like multiple linear regression (MLR), support vector regression (SVR), generalized regression neural networks (GRNN) & artificial neural networks (ANN) are used for future prediction of RW disposal facility.

Modeling of different safety parameters of low level or intermediate level waste is a problem because in nuclear power plants, medical and research centers, huge amount of radioactive waste (RW) is produced every year. It is essential to safely dispose this RW so that workers and other people can be protected. Disposal of RW is a very vast and multi-dimensional field. There are huge amount of radio nuclides present in the RW dumped in near surface disposal facilities. However in this paper radionuclide of our interest will be cesium (Cs) due to its concentration in the waste. Here in this paper a few safety parameters are modeled using different techniques. For modeling safety parameters like distribution coefficient K_d, retardation factor R_f and leaching rate λ_l, techniques used in this paper are MLR, SVR, GRNN and ANN. These models will help in safety assessment and to assess sorption of radio nuclides from near surface disposal facilities.

Our proposed model is of the following form:

$$Model = f(pH, CEC, SA, Aqueous\ cesium) \quad (4)$$

A set of known K_d values for cesium taken from the literature are analyzed considering it to be output parameter while other four parameters are considered to be input parameters. This cesium data set included 176 cesium K_d values. Two separate data sets were compiled. The first one included both soils and pure mineral phases. It contains 176 K_d values. The lowest K_d value is 0.6 ml/g while the largest K_d value is 52,000 ml/g. The average cesium K_d value is 2635 ± 530 ml/g. The second data set shown contain data which include only soil studies, that is, data from pure mineral phases and rocks were eliminated from the data set. This second data set contains 57 K_d values [7]. Statistics were also applied on this data set. These four methods were applied to predict output parameter K_d from four input parameters pH, Cation exchange capacity (CEC) in meq/100g, surface area (SA) in m^2/g aqueous cesium in µM. Matlab codes were written in Matlab 2013 [8] to model these parameters.

In this paper machine learning approaches like MLR, GRNN, ANN and SVR are used to model safety parameters [9]. The safety parameters are distribution coefficient K_d, retardation factor R_f, leaching rate λ_l. Researchers use ML approaches to make decisions for their problems. In the literature, there are many techniques but machine learning techniques are significant in a way that they can be applied to solve a large variety of problems. ML approaches learn from experience [10].

Matlab and matlab toolboxes were used for modeling K_d values [8]. A set of known K_d values for cesium taken from the literature are analyzed considering it to be output parameter, while other four parameters (i.e. pH, CEC, SA, aqueous cesium) are

considered to be input parameters. First the four techniques were used to model K_d values taken from the literature. Then these modeled values of K_ds were used to find values of retardation factor R_f and leaching rate λ_l using *Equations 2 and 3*. Other values of parameters like bulk density of medium (Kg/m^3) ρ_b, total porosity of the medium θ, the rate of drainage of water through the waste forms (m/y) q, the water filled porosity of the waste form θ_w, the height of the waste form (m) z, were taken from literature [2].

In this paper, our ML based methodology comprises three steps. First step is data set partitioning, in the second step models are constructed and in the last step models are tested. Performance is measured by finding percentage absolute difference (PAD) as:

$$PAD\% = \frac{K_{d_{\exp}} - K_{d_{pred}}}{K_{d_{\exp}}} \times 100 \quad (5)$$

In our case, out of 55 samples, for training 39 samples were selected, while for validation and testing 8 samples were selected respectively. Any machine learning training process consists of learning domain, training set, learning system and finally testing. Keeping these factors in mind, machine learning models can be developed [11].The four techniques used are discussed as below:

Multiple Linear Regression (MLR)

In this paper, MLR is used to predict the value of distribution coefficient K_d from a set of four predictors i.e. pH, Cation exchange capacity (CEC) in meq/100g, surface area (SA) in m^2/g, aqueous cesium in µM. In other words the four input parameters are pH, CEC, SA and aqueous cesium and the output parameter is distribution coefficient K_d. A matlab code was written for the prediction of K_d out of these four parameters.

Generalized Regression Neural Network (GRNN)

This is basically a neural network based function approximation or function estimation algorithm. It predicts the output of a given input data. In principle, neural network needs a training data to train itself. Training data should contain input-output mapping. Now if the network is trained with the training data set and a new testing data set is fed, it will accordingly give the output or predict the result.

The main advantage of GRNN models is that it estimates the appropriate regression model from the given data. The network is over-trained for low value of Gaussian function parameter σ, and the prediction error would be higher for evaluation

data. In a similar way, for large value of σ, the network remains under-trained and the prediction error would increase for training data.

Training procedure involves determination of the optimum value of spread constant σ. Best practice is that find the position where the Mean Squared Error (MSE) is minimum. First divide the whole training sample into two parts, training sample and test sample. Apply GRNN on the test data based on training data and find out the MSE for different [11, 12].

Artificial Neural Network (ANN)

ANN uses generalized mathematical models of human or neural biology. A general mathematical model of simple neuron is shown in figure 1. A simple neuron has many inputs but a single output. ANN model also needs a training sample set with desired output values. Back-propagation technique has been employed. Neural Network Toolbox of Matlab is employed for the training[8]. This toolbox is a built in set of MATLAB functions that provide an environment to develop not only feed forward neural networks but also recurrent neural networks. For optimal network performance, ten neurons are used in the hidden layer. Levenberg–Marquardt algorithm is used for weights adjustment. Initial weights and bias values are randomly selected. Tansig and pure linear activation functions are used for hidden and output layers. The values of four parameters and the desired K_d value of each sample point are fed as the network input [13].

Support Vector Regression (SVR)

Support vector machine is a practical learning method based on statistical learning theory. Like ANN, SVR is also first trained on training data samples and then it is tested for those data samples not present in the training dataset[12]. In using SVR methodology first data set is generated. After data set generation, parameters are optimized on training dataset. Then optimized model is validated using validation dataset and finally optimized model is validated using verification dataset. For developing SVR machine the parameters initialized are type and width of the kernel, Trade–off parameter, C and ε–insensitivity zone

Results

Correlation coefficients for the parameters are determined for both types of data sets. The parameter which have largest correlation coefficient with cesium K_d was CEC (R^2=0.3364). Similarly the correlation coefficients of other parameters i.e Ph and

aqueous Cs with cesium K_d values are 0.0356 and 0.0989 respectively. The correlation coefficient between parameter aqueous Cs and Cs K_d values is poor. This is due to the fact that this parameter contains concentration of the solution before and after contact with the soils. The correlation coefficients between the chosen input parameter CEC and K_d is shown in Figure 1. The correlation coefficient R values and corresponding equations of all four input parameters and K_d is shown in Table 1.

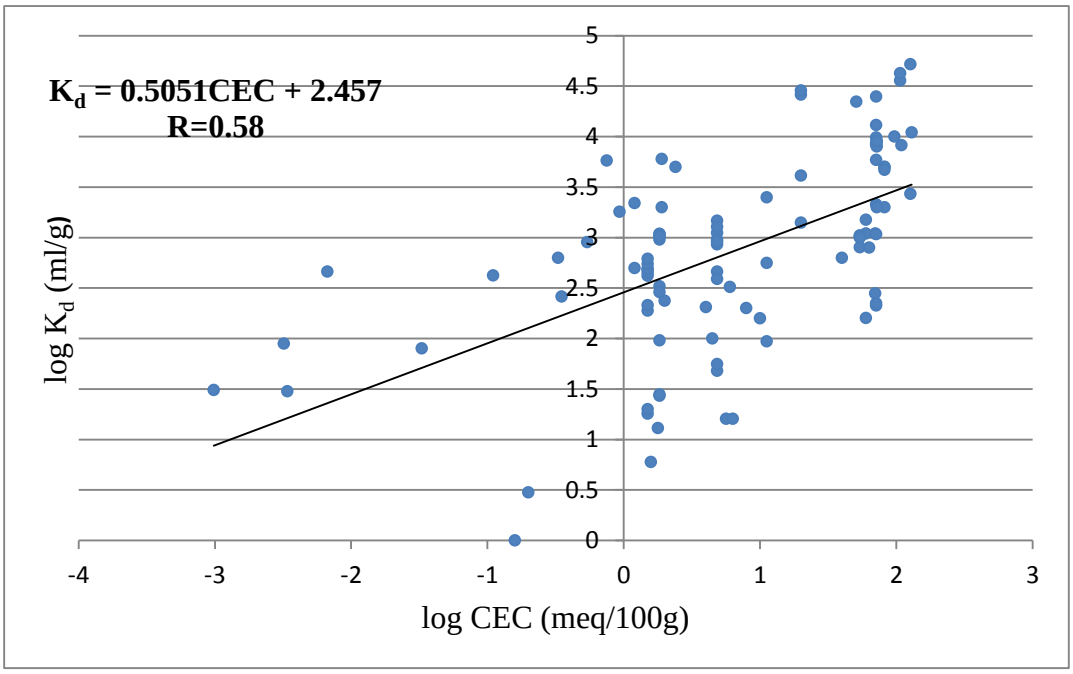

Figure 1: Relationship between Cs K_d values and CEC.

Table 1: Input parameter correlation coefficient values and corresponding K_d equations.

Sr. No.	Input Parameter	R^2 value	Equation
1	CEC (meq/100g)	0.336	K_d = 0.5051 CEC + 2.457
2	SA (m^2/g)	0.176	K_d = 0.5679 SA + 1.841
3	Aqueous cesium (µm)	0.102	K_d= -0.1354 AqCs + 2.0512
4	pH	0.036	K_d= 1.6236 pH + 1.0251

K_d Prediction Using MLR

Four independent variables (pH, CEC, SA, Aq Cs) are the input parameters and K_d is the corresponding output variable. The coefficients of independent variables are computed under ordinary least squares criterion, using training data, as follows:

$$K_d^{MLR} = 8107.9 - 910.7\,X_1 + 129.8\,X_2 + 3.1\,X_3 - 3.2\,X_4 \quad (6)$$

where

X_1=pH, X_2= Cation exchange capacity (CEC), X_3= surface area (SA) and X_4= Aqueous cesium (Aq Cs). i.e. intercept b_0= 8107.9

Regression coefficients are b_1= -910.7, b_2= 129.8, b_3= 3.1, b_4= -3.2

Linear correlation between theoretical and predicted K_d is shown in Figure 2.

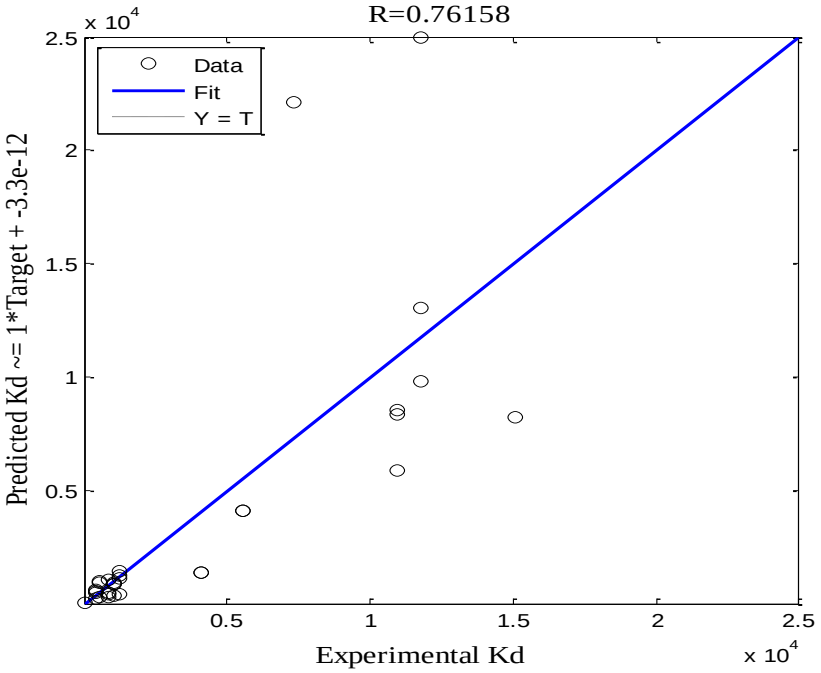

Figure 2: Linear correlation graph between experimental and predicted K_d for MLR

K_d Prediction Using GRNN

For GRNN, Gaussian function parameter σ is tuned and the best results of this network are obtained by finding the optimal spread values of σ. In our case σ =0.1025, 0.0275 for four parameters, respectively. Figure 3 shows the linear correlation graph between the two variables.

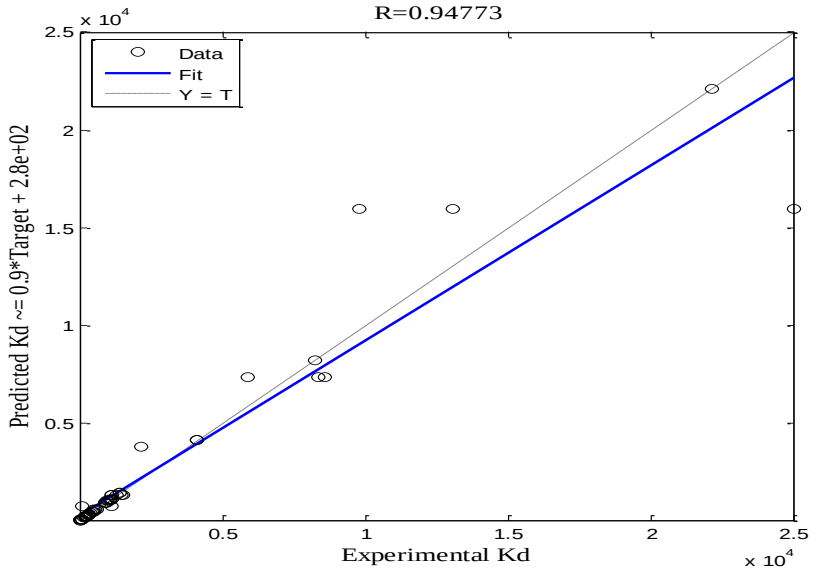

Figure 3: Linear correlation graph between experimental and predicted K_d for GRNN.

K_d Prediction Using ANN

Neural Network Toolbox of Matlab [8] is employed for the training. For optimal network performance, ten neurons are used in the hidden layer. Predicted K_d is shown in Figure 4.

K_d Prediction Using SVR

Gaussian Kernel function is employed for this problem. To find the optimal values of these parameters, selected ranges for these three parameters are as: C=[80,120], σ=[0.000001,0.1] and ε = [6, 1.8]. The combination for which the average PAD value is minimum for both the training and validation set is chosen to be optimal. The optimal values for the trade-off parameter C, the kernel width σ and ε–insensitivity zone are found to be 100, 1.7255 and 1.0 x 10^{-6} respectively in our case. The correlation coefficient in this case is 0.99812.

A comparison of PAD values is given in Table 2. The performance comparison in terms of correlation coefficient is shown in Table 3.

Now using these predicted Kd values, two other parameters i.e. retardation factor Rf and leaching rate of radio nuclides are modeled. Using Equation (2) retardation factor is calculated. The values of bulk density ρ_b and total porosity θ of the medium are taken from the literature. The graph between K_d and R_f is shown in Figure 5.

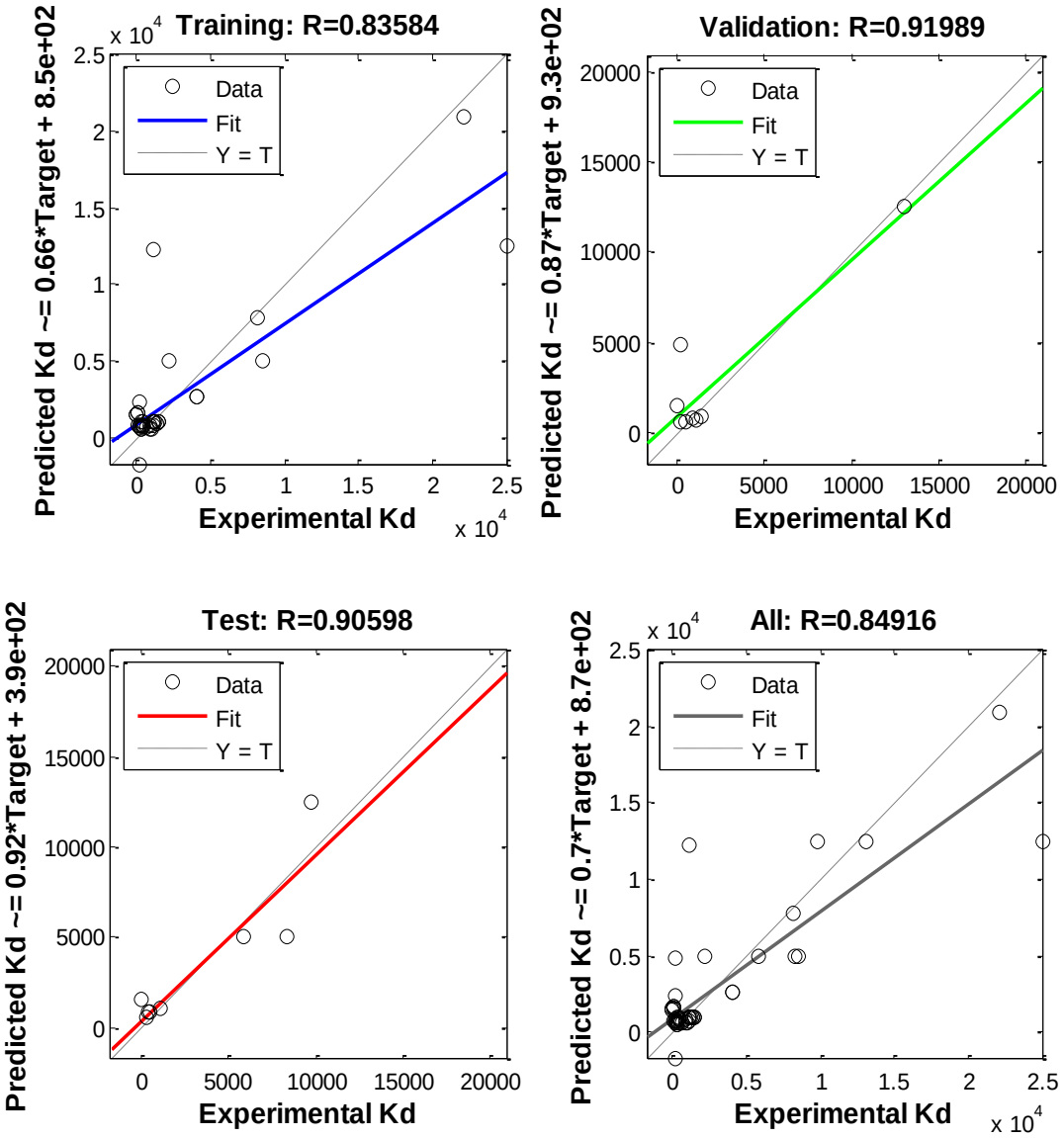

Figure 4: Linear correlation graph between experimental and predicted K_d for ANN

Similarly Equation (3) is used to calculate leaching rate λ_l. The values of rate of drainage of water q, water filled porosity of waste θ_w, bulk density of waste ρ_b and height of waste form z is taken from the literature. The graph between K_d and leaching rate λ_l is shown in Figure 6.

Table 2: Comparison of ML approaches

PAD for SVR (%)	PAD for GRNN (%)	PAD for ANN (%)	PAD for MLR (%)
1.333	9.891	12.916	18.374
12.500	9.903	15.743	16.851
1.818	9.892	14.146	15.607
0.487	9.890	9.090	13.143
1.428	9.891	15.405	15.805
1.818	9.892	9.146	9.324
7.142	9.897	15.609	9.148
8.438	9.899	14.052	15.515
0.243	9.890	13.792	14.590
6.153	9.896	12.209	19.062
0.090	9.890	13.128	11.959
6.079	9.896	11.916	12.871
1.838	9.892	14.542	15.656
1.845	9.892	14.675	12.205
2.103	9.892	7.913	17.346
1.956	9.892	4.913	19.839

Table 3: Performance comparison of ML approaches in terms of correlation coefficient

Input Dataset	SVR Model	GRNN Model	ANN Model	SVR Model
Train. Data	0.998	0.947	0.835	0.761
Valid. Data	0.983	0.950	0.919	0.741
Test. Data	0.975	0.942	0.905	0.752
Mean R-value	0.985	0.946	0.887	0.751

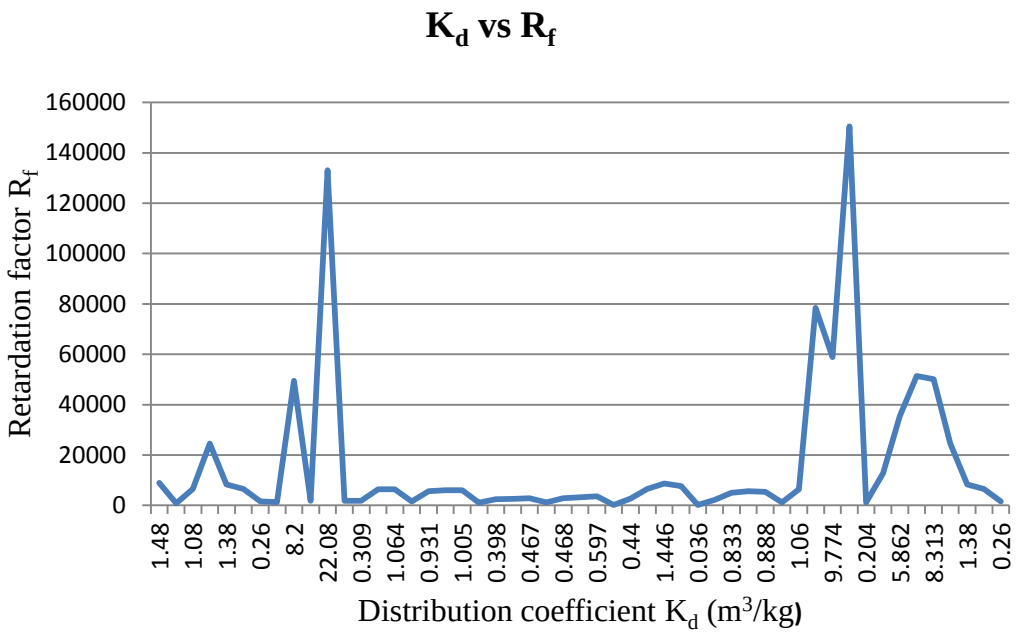

Figure 5: Graph between Distribution coefficient and Retardation factor

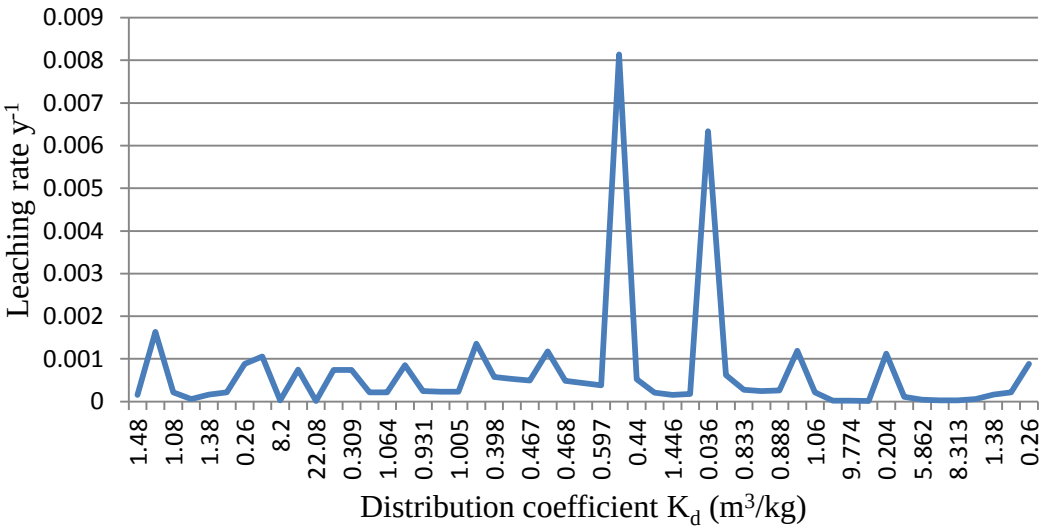

Figure 6: Graph between distribution coefficient (K_d) and leaching rate (λ_l)

Discussion

The correlation coefficient value of each parameter with K_d is calculated to determine which parameter has the largest correlation coefficient. The correlation coefficients of parameters CEC, SA, pH and aqueous cesium with K_d was 0.58, 0.42, 0.19 and 0.32 respectively. Hence the parameter with the largest correlation coefficient with cesium K_d was CEC (R = 0.58).

Multiple linear regression (MLR), generalized regression neural network (GRNN), artificial neural network (ANN) and support vector regression (SVR) are used to predict K_d and results obtained from these methods are compared. Correlation coefficient for these machine learning approaches MLR, GRNN, ANN and SVR are 0.76158, 0.94773, 0.84916, and 0.99812, respectively. Thus the results obtained from SVR and GRNN are better than other two approaches. The correlation coefficients for SVR and GRNN are 0.99812 and 0.94773 respectively.
Similarly Performance is measured by finding percentage absolute difference (PAD) as follows:

$$PAD_{SVR} < PAD_{GRNN} < PAD_{ANN} < PAD_{MLR} \quad (7)$$

After modeling K_d values using different machine learning approaches, the results of best approach is taken which in our case is that of SVR for which R=0.99812. These K_d values are plotted against retardation factor (R_f). The results are shown in Figure 5, which shows a smooth variation between K_d and R_f. The values of K_d are constant for a given range of R_f values but at some points some peaks are observed which shows that at these points one must be careful in selecting these K_d values. For example in Figure 5, the value of retardation factor is constant in the range of 0.308 up to 13.022 but at points 0.307 and 0.204 there is a peak showing abnormal variation of R_f with K_d.

In a similar way, K_d values are also plotted against leaching rate and the results are shown in Figure 6. These graphs also shows smooth variation between K_d and R_f but at some particular points there are peaks showing abnormal behavior.

Conclusion

Four different models are used for modeling K_d values in this work and their performance is tested using % PAD and coefficient of performance R. The best results have been achieved by SVR and GRNN models with correlation coefficients R=0.99812 and 0.94773 for K_d, respectively. Predicted K_d values are then plotted against retardation factor and leaching rate, which shows smooth variation with K_d except some particular points. So we should avoid taking those K_d values at those particular points.

Recommendations and Future Work

In future distribution coefficient K_d, retardation factor R_f and leaching rate λ_l prediction can be modelled for other radionuclides against different disposal concepts. This study could be used for development of radionuclides transport models in different media. This study will also integrate with entire safety assessment activities of near surface disposal facilities.

REFERENCES

[1] R. H. Little, *et al.*, "Assessment of radioactive and non-radioactive contaminants found in low level radioactive waste streams," in *Proc. Waste Manag. Conf.*, Tucson, AZ, United States, 2003.

[2] IAEA, "Safety assessment methodologies for near surface disposal facilities: Results of a co-ordinated research project," IAEA, Vienna2004.

[3] E. Tolentino and C. C. O. d. Tello, "An overview of technical requirements on durable concrete production for near surface disposal facilities for radioactive wastes," in *Proc. INAC 2013 Int. Nucl. Atl. Conf.*, Recife, PE, Brazil, 2013.

[4] S. C. Sheppard, *et al.*, *Solid/liquid partition coefficients (Kd) and plant/soil concentration ratios (CR) for selected soils, tills and sediments at Forsmark*, 1st ed.: Swedish Nuclear Fuel and Waste Management Company, 2011.

[5] M. Bragea, *et al.*, "Influence of distribution coefficients on the transfer of radionuclides from water to geological formation," *Chem. Bull. Politehnica Univ. Timisoara,* vol. 53, pp. 1-2, 2008.

[6] S. A. Adeleye, *et al.*, "Sorption of caesium, strontium and europium ions on clay minerals," *J. Mater. Sci.*, vol. 29, pp. 954-958, 1994.

[7] K. M. Krupka, *et al.*, "Review of geochemistry and available Kd values, for cadmium, cesium, chromium, lead, plutonium, radon, strontium, thorium, tritium (3H), and uranium," in *Understanding Variation in Partition Coefficient, Kd, Values* vol. II, 1999.

[8] MatLab. Natick, Massachusetts, United States: The MathWorks Inc., 2013.

[9] S. Kwon and W. J. Cho, "A sensitivity analysis of design parameters of an underground radioactive waste repository using a backpropagation neural network," *J. Korean Soc. Rock Mech.*, vol. 19, pp. 203-212, 2009.

[10] F. L. d. Lemos, *et al.*, "Uncertainty / sensitivity methodologies for safety assessments of low-level waste disposal facilities," in *Waste Manag. Conf.*, 1999.

[11] A. Majid, *et al.*, "Predicting lattice constant of complex cubic perovskites using computational intelligence," *Comput. Mater. Sci.*, vol. 50, pp. 1879-1888, 2011.

[12] A. Majid, *et al.*, "Combination of support vector machines using genetic programming," *Int. J. Hybrid Intell. Syst.*, vol. 3, pp. 109-125, 2006.

[13] A. Ceguerra and I. Koprinska, "Automatic fingerprint verification using neural networks," in *Proc. ICANN 2002: Int. Conf. Artif. Neural Netw.*, Berlin, Heidelberg, 2002, pp. 1281-1286.