M3LW-24OR0403025
Revision 0

Light Water Reactor Sustainability Program

Performance Comparison of Machine Learning Models for Ultrasonic Nondestructive Evaluation of Alkali-Silica Reaction in Concrete

Hongbin Sun
Oak Ridge National Laboratory
Samantha Sabatino
Oak Ridge National Laboratory

August 2024

U.S. Department of Energy
Office of Nuclear Energy

ORNL/SPR-2024/3463

Light Water Reactor Sustainability Program

Nuclear Energy and Fuel Cycle Division

PERFORMANCE COMPARISON OF MACHINE LEARNING MODELS FOR ULTRASONIC NONDESTRUCTIVE EVALUATION OF ALKALI-SILICA REACTION IN CONCRETE

Hongbin Sun
Samantha Sabatino

August 2024

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, TN 37831
managed by
UT-BATTELLE LLC
for the
US DEPARTMENT OF ENERGY
under contract DE-AC05-00OR22725

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AI	artificial intelligence
ASR	alkali-silica reaction
ASR-2D	alkali-silica reaction sample with 2D confinement
DNN	deep neural network
DWT	discrete wavelet transform
FFT	fast Fourier transform
LR	linear regression
ML	machine learning
NDE	nondestructive evaluation
NN	neural network
ReLU	rectified linear unit
RMSE	root mean square error
RNN	recurrent neural networks
SVM	support vector machine
SVR	support vector regression

EXECUTIVE SUMMARY

Alkali-silica reaction (ASR) causes concrete degradation, leading to cracking, rebar corrosion, and reduced structural integrity, which raises safety concerns. Ultrasonic nondestructive evaluation (NDE) effectively assesses concrete properties and monitors ASR progression. However, its deployment and analysis require specialized expertise and subjective interpretation. As computational power increases, artificial intelligence (AI) and machine learning (ML) algorithms are increasingly being used to automate NDE data analysis across various industries for AI-assisted automation. Regulatory agencies are adapting to this technological shift, prompting a need to evaluate current ML technologies' capabilities and limitations in assessing concrete material properties and damage.

This report presents a comparative analysis of four ML regression models for predicting concrete material damage induced by ASR expansion using long-term ultrasonic data monitoring. The models investigated include linear regression (LR), support vector regression (SVR), shallow neural networks (NN), and deep neural networks (DNN). LR, SVR, and shallow NN models use features extracted from ultrasonic signals, whereas the DNN model processes time-domain ultrasonic signals and frequency spectra directly.

The study systematically compared the models' performance from various perspectives, including model input, prediction performance, and generalization ability. The findings indicate significant variability in model performance, with some ML algorithms achieving very high or very low prediction accuracy depending on the preprocessing and feature engineering (extraction and selection) applied. Key insights include the observation that shallow ML models (LR, SVR, and shallow NNs) require meticulous preprocessing and feature extraction to achieve high accuracy. In contrast, the DNN model, although it bypasses the need for feature engineering, necessitates extensive preprocessing to mitigate noise and computational demands. The SVR model emerged as the top performer among the shallow models, and the DNN model exhibited superior performance on specific datasets but struggled with generalization across specimens from different batches. Additionally, the SVR model is sensitive to temperature variations, whereas the DNN model is robust in this regard. Using recurrent neural networks is recommended for future ASR expansion prediction studies. Recurrent neural networks' inherent ability to capture temporal dependencies and long-term patterns makes them well suited for analyzing sequential ultrasonic monitoring data.

Overall, the results and conclusions of this study could provide insights into the capabilities and effectiveness of ML when applied to ultrasonic NDE data and help identify best practices for using ML for ultrasonic NDE of concrete material properties.

1. INTRODUCTION

1.1 MOTIVATION

Alkali-silica reaction (ASR) is a type of material degradation found in concrete infrastructure. It occurs when certain types of silica, such as chert or microcrystalline quartz, present in the aggregate interact with alkali metals in the cement. This interaction forms hydrated sodic or K-bearing Ca silicate gels, known as ASR gels [1]. These gels absorb water and expand within the material's pores and preexisting cracks, causing microcracking and larger cracks in the concrete. ASR contributes to the deterioration of concrete, resulting in a loss of mechanical properties and increased penetration of harmful substances. This process can ultimately lead to the corrosion of steel reinforcement and negatively affect the structure's service life. In 2009, ASR was discovered in certain below-grade structures at the Seabrook Nuclear Power Plant. Subsequently, ASR was confirmed in multiple concrete structures, including the electrical tunnel, containment enclosure building, residual heat removal vault, emergency diesel generator building, and emergency feedwater building. The safety issues and concerns that arise from ASR-induced concrete cracks impede the license renewal of nuclear power plants. To ensure the continued safe operation of nuclear power plant concrete structures over extended periods, it is necessary to examine the long-term effects of ASR on the durability, serviceability, and safety of these structures [2].

Nondestructive evaluation (NDE) methods, such as ultrasonic testing, provide an effective approach for assessing ASR damage in concrete structures. Various ultrasonic wave parameters, including wave velocity, attenuation, and amplitude, are used to measure ASR damage [3], [4], [5]. As ASR progresses, ultrasonic wave amplitude and velocity decrease, and wave attenuation increases. Numerous studies have successfully employed ultrasonic NDE for long-term monitoring of ASR damage development in concrete, demonstrating the efficacy of these techniques for evaluating ASR-affected structures [2], [6], [7], [8], [9]. However, interpreting ultrasonic NDE data requires specialized knowledge in ultrasonic testing and material properties and relies heavily on engineering expertise. Recent advances in artificial intelligence (AI) and machine learning (ML) have transformed this process by offering automated solutions for analyzing and interpreting NDE data, including ultrasonic signals. The industrial sector is increasingly adopting ML algorithms to simplify NDE data analysis, with a growing focus on AI assistance and AI-assisted automation. Regulatory agencies are preparing for these changes, anticipating updates to relevant standards. Recent research has extensively investigated the application of ML in NDE, particularly for ultrasonic NDE.

However, limited literature exists on using AI/ML for ultrasonic NDE of concrete damage evaluation. The growing variety of ML techniques highlights the importance of evaluating the current capabilities in ML-supported analysis of NDE data, especially in the ultrasonic NDE of concrete. Previous work has included a literature review of research on the use of AI/ML for processing ultrasonic data in evaluating concrete damage, particularly for characterizing ASR damage [10], [11]. Key takeaways from the literature review are summarized in a former report [11]. Other work has examined various ML models for processing long-term ultrasonic monitoring data to predict ASR damage. The studies identified the capabilities and limitations of these ML models and investigated the effects of various factors on their performance. These studies' findings provide insights into the effectiveness of ML applications in ultrasonic NDE data and help identify best practices for using ML in the ultrasonic NDE of concrete. However, each ML model has its own advantages and disadvantages when applied to ultrasonic NDE of ASR damage in concrete. This report summarizes the procedures for using each previously studied ML model and compares their performances in predicting ASR damage.

1.2 OBJECTIVES

This work compared the performance and effectiveness of four ML models—linear regression (LR), support vector regression (SVR), shallow neural networks (NN), and deep neural networks (DNN)—in predicting ASR expansion based on extensive, long-term monitoring of ultrasonic data across multiple specimens. The assessment of these four ML models was partially based on results from previously published reports [10], [11].

The models were evaluated from several key perspectives relevant to the application of ultrasonic NDE for ASR damage, including the following:

- **Raw data:** raw data used for model input,
- **Preprocessing:** the preprocessing procedure used to process the raw data for feature extraction,
- **Model input:** the types and features of input data required by each model,
- **Data preprocessing:** the steps needed to prepare the ultrasonic data for analysis,
- **Model complexity:** the structural complexity and computational demands of each model,
- **Model-training difficulties:** the challenges associated with training each model, including convergence and overfitting issues,
- **Model output:** the accuracy and reliability of the predictions generated by each model.

The reference ultrasonic data were obtained from four distinct concrete specimens cast with artificially induced ASR, which are discussed in detail in previous reports [10], [11]. This report involves the use of ML terminology, such as data training, data testing, model optimization, model tuning, etc. This background information can be found in multiple works of literature [12], [13], [14] and we also provided an introduction to these commonly used ML terminologies in a previous report [15]. This work focused on assessing and comparing the performance of different ML models for regression based on ultrasonic signals. This report summarizes the capabilities and limitations of different ML models for predicting ASR expansion using ultrasonic long-term monitoring data. The results of these assessments are expected to be useful to industries and the US Department of Energy for future activities involving NDE of concrete with ASR.

2. MACHINE LEARNING MODELS

2.1 LINEAR REGRESSION

LR is a linear statistical approach to modeling the relationship between the response and one or more variables. By fitting a linear equation to the observed data, this model helps predict the value of the dependent variable based on the values of the independent variables. The simplicity and interpretability of LR make it a widely used technique in various fields. It provides insights into how changes in the independent variables influence the dependent variable, enabling researchers and analysts to make informed decisions.

The core idea of LR is finding the best-fit line through the data points, minimizing the discrepancies between the observed and predicted values. This line, known as the regression line, is characterized by its slope and intercept, which are estimated from the data. The slope indicates the direction and strength of the relationship between the variables, and the intercept represents the expected value of the dependent variable when all independent variables are zero. Through this model, one can quantify the effect of each predictor on the outcome, assess the goodness of fit, and make predictions for new observations. This model is represented by the following equation:

$$y = X\beta + \varepsilon,\#(1)$$

where y is the response (i.e., the volumetric expansion of the concrete specimens), X is the vector of the variables (i.e., the extracted features), β is the parameter vector, and ε is the error term.

One key advantage of LR is its interpretability. The coefficients of the model directly represent the relationship between the dependent and independent variables, allowing for straightforward interpretation. Additionally, LR models are relatively easy to implement and computationally efficient, making them suitable for large datasets. However, it is essential to check for the assumptions underlying LR, such as linearity, independence, homoscedasticity, and normality of residuals. Violations of these assumptions can lead to biased or inefficient estimates, thereby affecting the model's reliability.

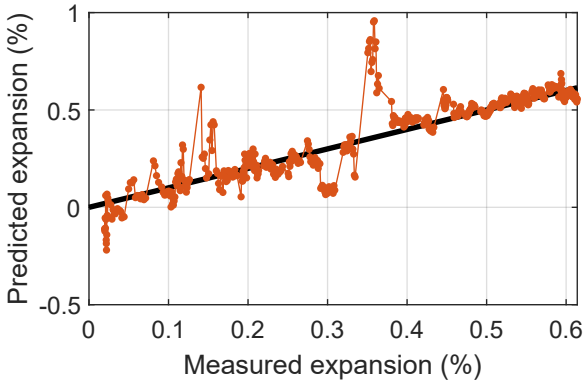
In previous work [10], the parameter vector, β, was fitted using data from ASR specimen experiments, and the fitted LR model was tested using the specimen data from the alkali-silica reaction sample with 2D confinement (ASR-2D). The input features included 13 variables: mean amplitude, maximum amplitude, energy, and attenuation in three different frequency ranges, and absolute wave velocity. The LR model was optimized in MATLAB and included an intercept term, linear and squared terms for each variable, and all products of pairs of distinct variables.

The testing results are shown in

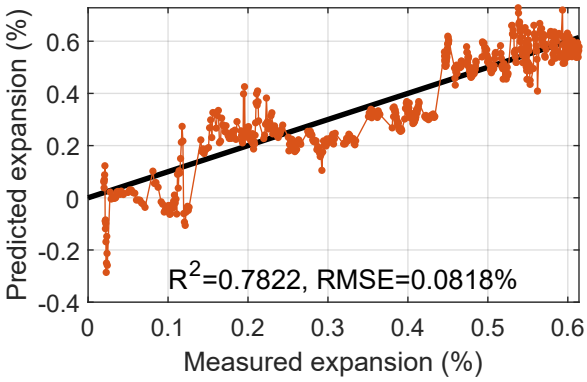
Figure 1(a). Although the overall prediction accuracy was acceptable, large errors were observed at expansion levels of approximately 0.14% and 0.35%. Examination of the expansion data and original time-domain signals revealed significant variations during the conditioning periods caused by sensor reinstallation and chamber malfunction. The model also indicated overfitting during the training process. After feature selection and removal of less correlated features, the accuracy and prediction errors improved slightly, as seen in

Figure 1(a). The large prediction errors at approximately 0.14% and 0.35% expansion were also reduced. However, large prediction errors remained at the initial points when the expansion was very small, both with and without feature selection.

When using LR for ASR prediction based on ultrasonic signals, it is crucial to first extract and select features correlated with the expansion. These features serve as the model input. The quality of the original ultrasonic signals significantly affects model performance. Factors such as transducer reinstallation, changes in coupling conditions, and significant environmental changes must be considered when preparing the training and testing datasets. Additionally, the LR model should be reoptimized during each training session to improve performance.



(a)



(b)

Figure 1. Measured expansion vs. expansion predicted using LR on the ASR-2D specimen (testing) (a) without feature selection and (b) with feature selection [10].

2.2 SUPPORT VECTOR REGRESSION

SVR is an ML algorithm used for regression tasks. It is based on the principles of support vector machines (SVMs). An SVM is a supervised ML algorithm primarily used for classification tasks, but it can also be applied to regression tasks. The basic principle of the SVM is to find the optimal hyperplane (decision boundary) that best separates data points belonging to different classes in the feature space. For a linear model with dataset (x, y), the hyperplane, $f(x)$, is expressed as

$$f(x) = wx + b = 0, \#(2)$$

where w is the weight vector, and b is the bias term. The hyperplane should be subject to the following constraints:

$$y_i f(x_i) = y_i(wx_i + b) = 0, \#(3)$$

whereas the cost function J is

$$J = \frac{1}{2}\|w\|^2 .\#(4)$$

The constraint and the cost function are for separable data. If the data are not separable, then a soft margin is used with a cost function:

$$J = \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{m}\xi_i ,\#(5)$$

with constraint

$$y_i f(x_i) = y_i(wx_i + b) \geq 1 - \xi_i, \xi_i \geq 0, \#(6)$$

where ξ_i is a positive slack variable that indicates that the sample training allows a small number of erroneous samples, and c is the penalty parameter. The SVM model has several hyperparameters: the kernel function, kernel scale, and penalty parameter. Linear kernels [17], Gaussian radial basis functions [18], [19], [20], polynomials [21], [22], and sigmoid functions are among the most common kernel functions. The kernel scale defines how far the influence of a single training example reaches. The penalty parameter controls the trade-off between minimizing and maximizing the classification margin and training error. Instead of predicting discrete class labels as in classification using SVM, SVR predicts continuous output values. Similar to SVM for classification, SVR aims to find a hyperplane in the feature space that best fits the training data while minimizing the error, known as the epsilon-insensitive loss function. This hyperplane is determined by support vectors, which are the data points closest to the hyperplane and influence its position. SVR allows for flexibility in modeling nonlinear relationships between input features and output values by using kernel functions to map the input data into higher-dimensional feature spaces where linear separation can be achieved. SVR aims to minimize error and ensure that the deviations of the actual output values from the predicted values within the epsilon tube do not exceed a predefined threshold. This approach makes SVR robust to outliers and noisy data, making it suitable for regression tasks in various domains.

In previous work [10], [11], extracted features from time-domain and frequency-domain signals were used as inputs for the SVR model. The model was trained with data from one of the specimens (ASR, ASR-2D, small ASR, or small ASR-2D) and then tested with data from other specimens. Several key findings emerged from the results [11]. First, like the LR model, the SVR model achieved good prediction performance using the extracted features with appropriate procedures, as shown in Figure 2. These procedures included data cleaning, feature extraction and selection, and model optimization during training.

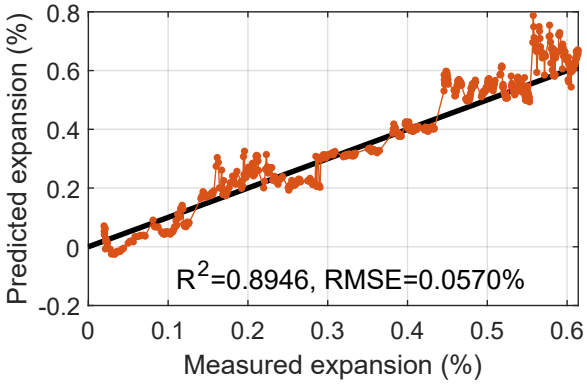


Figure 2. Measured expansion vs. expansion predicted using SVR on the ASR-2D specimen (testing) with feature selection [10].

Additionally, the data range in the training and testing datasets significantly affected the SVR model's performance. When the data in the testing dataset fell outside the range of the training dataset, the SVR model performed poorly. Therefore, the training dataset should be as large and diverse as possible to ensure the trained model generalizes the testing data well. Signal preprocessing, such as amplitude normalization, also influences SVR performance. Preprocessing should be consistent for both training and testing datasets. Normalizing signal amplitudes to a constant value is suggested if they show large variations during long-term monitoring. The SVR model uses wave features and parameters as inputs, and these parameters are easily affected by environmental temperatures when ultrasonic signals are collected. Thus, the effect of temperature should be considered, especially if training and testing data are collected at very different temperatures. The SVR model trained with data from one batch of specimens often performed well in the early stage but performed poorly in the late stage when tested on data from a different batch. Figure 3(a) and (b) illustrate the results of testing on the small ASR and small ASR-2D data using the SVR model trained with ASR data. Both figures show poor performance in the later stages of expansion. This issue appears to be common across the ML models studied in this and previous works. Improving the generalization ability of ML models across different batches of specimens remains a challenge for future research.

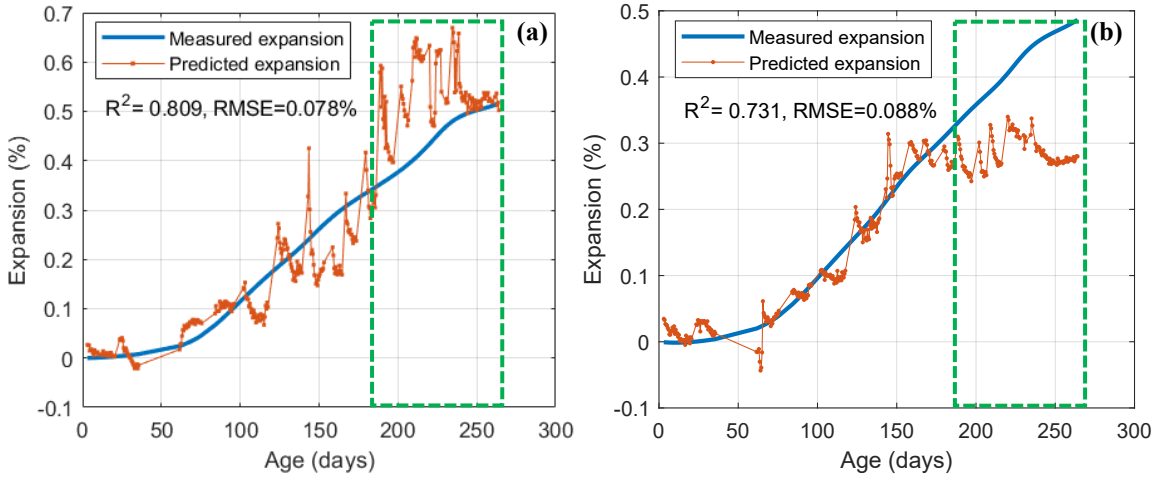


Figure 3. Results of testing using the SVR model trained with ASR data and tested on (a) small ASR data and (b) small ASR-2D data [11]. The areas enclosed by green dashed lines are poor predictions with large errors.

2.3 SHALLOW NEURAL NETWORK

A shallow NN is a type of artificial NN characterized by having only one or two hidden layers between the input and output layers. This simplicity makes it a suitable choice for regression tasks in which the goal is to predict continuous values. Shallow NNs are advantageous because of their relatively straightforward architecture, which requires less computational power and is easier to train than deeper networks. They can efficiently model complex relationships in the data, capturing nonlinear patterns that traditional LR might miss. This makes them particularly useful in fields like finance, engineering, and natural sciences, in which accurate predictions are essential.

In regression tasks, the objective of a shallow NN is to learn a mapping from input features to continuous target values. This is achieved through the training process, during which the network adjusts its weights based on the error between its predictions and the actual values. Various activation functions can be used in the hidden layer to introduce nonlinearity, enhancing the network's ability to model intricate data patterns. Despite their simplicity, shallow NNs can perform robustly and reliably for many regression problems, especially when the dataset is not excessively large or complex. A shallow feed-forward NN typically has a straightforward structure comprising one input layer, one or two intermediate (hidden) layers, and one output layer. Lippmann [23] suggests that a multilayer perceptron with two hidden layers is sufficient for creating classification regions of any desired shape. Another critical hyperparameter is the activation function, which defines the output of a node given the input or a set of inputs. Common activation functions include Gaussian, sigmoid, hyperbolic tangent, and radial basis functions. Optimizing the parameters in the hidden layers is crucial for achieving the best model performance.

Figure 4 illustrates the shallow NN structure used in this work. The optimized NN has two hidden layers in addition to the input and output layers. The first hidden layer contains 16 neurons, the second hidden layer has 8 neurons, and both layers use rectified linear unit (ReLU*)* activation functions. A randomly selected 25% of the training data was used as validation data to assess the model's performance during training. The input to the shallow NN comprised extracted features from the original ultrasonic data, which were also used as input for the LR and SVR models. The model was initially trained with the ASR data and then was tested on the ASR-2D data. Like the LR and SVR models, the shallow NN initially exhibited poor prediction performance on the ASR-2D data without feature selection, with large prediction errors observed near expansions of 0.35% and 0.02%. However, after feature selection (see Figure 5), the prediction performance significantly improved, showing that the shallow NN model requires essential preprocessing procedures, such as data cleaning and feature selection, to achieve acceptable performance. With feature selection, the shallow NN model performed similarly to the SVR model, and both models performed better than the LR model.

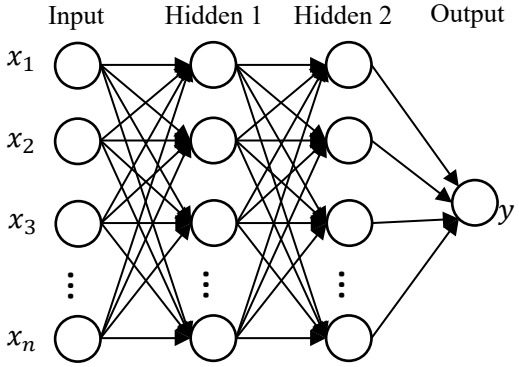
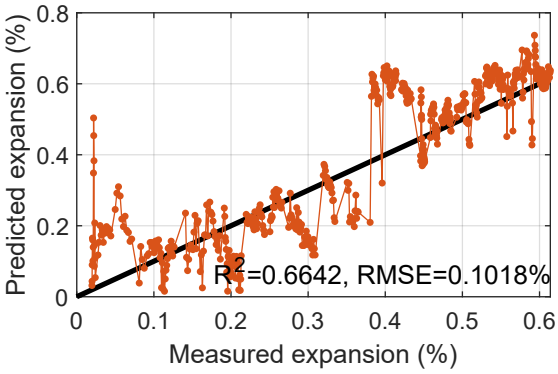
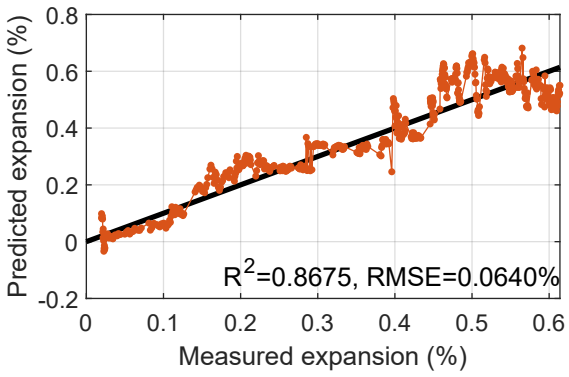


Figure 4. The network structure of the shallow NN used.



(a)



(b)

Figure 5. Results of testing on the ASR-2D data using the shallow NN model trained with ASR data (a) without feature selection and (b) with feature selection [10].

2.4 DEEP NEURAL NETWORK

Another model used in this work was a DNN. A DNN is a type of artificial NN with multiple layers between the input and output layers. DNNs are characterized by their depth, meaning they have many hidden layers compared with traditional shallow NNs, which may have only one or two hidden layers. Each layer in a DNN typically comprises multiple neurons, also known as nodes or units, which perform mathematical operations on the input data. The output of each layer serves as the input to the next layer, and the final output layer produces the prediction or classification. DNNs can learn intricate patterns and representations from complex data, making them particularly effective in tasks such as image and speech recognition, natural language processing, and other areas in which the input data have a high degree of complexity. Training a DNN involves feeding it with labeled training data and adjusting the weights and biases of the connections between neurons through a process called backpropagation to minimize the difference between the predicted output and the true output. This training process often requires a large amount of data and computational resources but can result in highly accurate models capable of handling complex tasks.

In this work, time-domain signals and frequency spectra were used as inputs for the DNN models. Figure 6 shows the network structure of the DNN used; it has three hidden layers with ReLU activation functions. The input layer has 512 neurons when time-domain signals are used as input, and the second hidden layer has 128 neurons. The third and fourth hidden layers have 32 and 8 neurons, respectively. Dropout layers were added to the network to prevent overfitting, with dropout rates of 0.4 for the first and second dropout layers and 0.3 for the last dropout layer. If frequency spectra are used as the model input,

the input layer has 256 neurons, and then the following hidden layer has a different number of neurons accordingly.

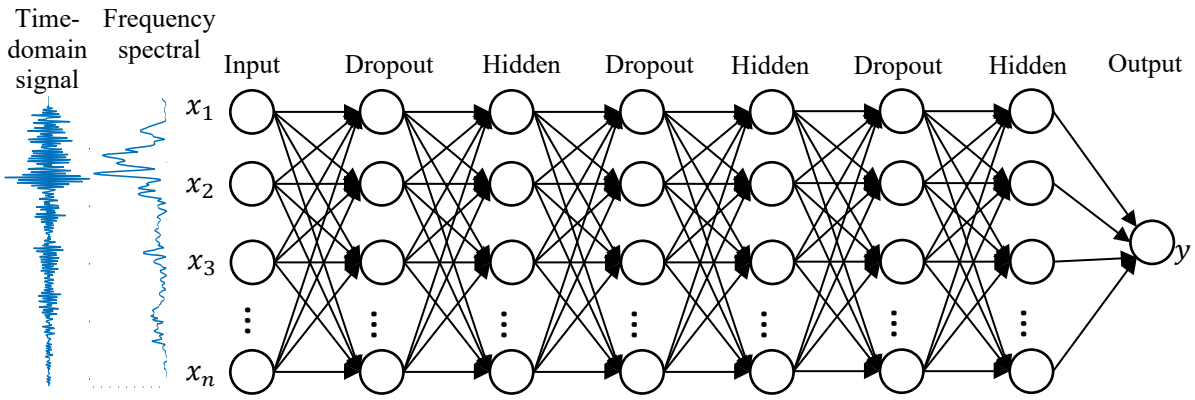
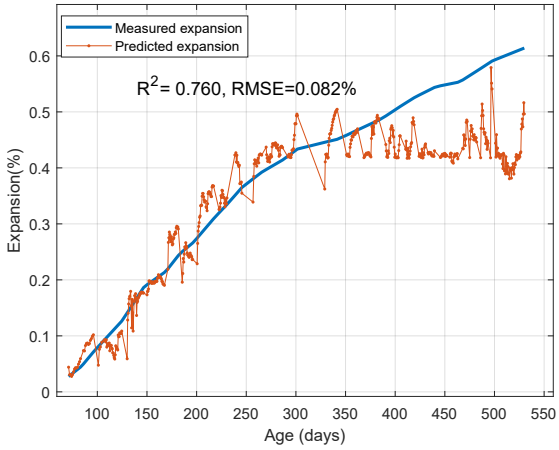
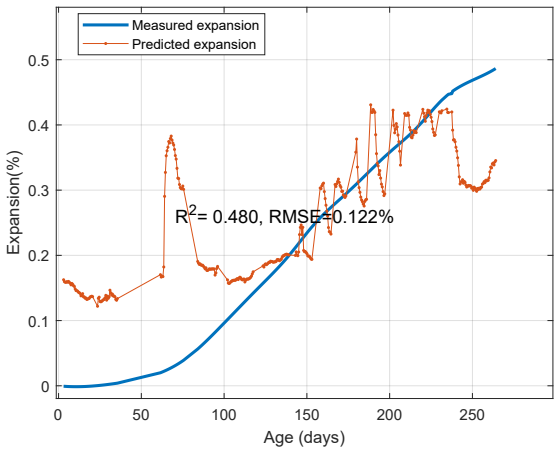


Figure 6. The network structure of the DNN model.

When the DNN model was used for ASR expansion prediction, both time-domain signals and frequency spectra were downsampled before being used as model inputs to reduce computational costs. Figure 7(a) shows an example of the results from testing on the ASR-2D data using the DNN model trained with ASR time-domain signals. Although the overall performance was acceptable, large prediction errors were observed after 300 days. Additionally, the model exhibited poor generalization performance on the small ASR-2D data from a different batch. In contrast, the LR, SVR, and shallow NN models demonstrated better generalization performance on data from different batches of specimens. A possible reason for this discrepancy is that the DNN model is more prone to overfitting during the training process than the shallower models (LR, SVR, and shallow NNs).



(a)



(b)

Figure 7. Results of testing on (a) the ASR-2D data and (b) small ASR-2D data using the DNN model trained with ASR time-domain signals [11].

Using frequency spectra as the model input, the DNN also achieved good prediction performance. As shown in Figure 8, the DNN model trained with small ASR frequency spectra demonstrated good prediction accuracy on the small ASR-2D data. Similarly, the model trained with small ASR-2D data performed well when tested on the small ASR data. The small ASR and small ASR-2D data were very consistent in both expansion and ultrasonic features, which likely contributed to the strong performance of the DNN models. However, the DNN model did not achieve acceptable performance on other specimen data when frequency spectra were used as the input, as shown in Figure 9. Although the DNN can achieve good prediction performance using either time-domain signals or frequency spectra, it requires more model-tuning effort, and its generalization ability may not be as robust as those of well-trained shallow ML models (e.g., LR, SVR, and shallow NNs).

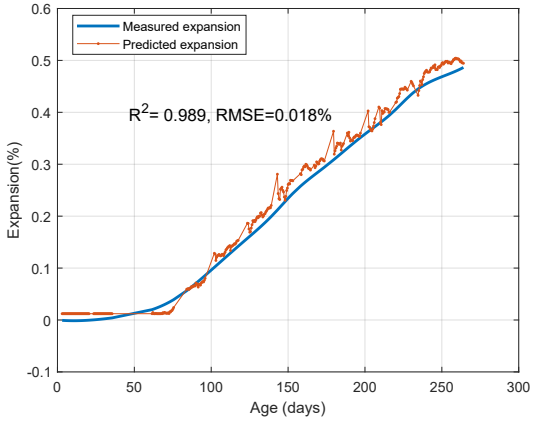


Figure 8. Results of testing on small ASR-2D data using the DNN model trained with small ASR frequency-domain spectra [11].

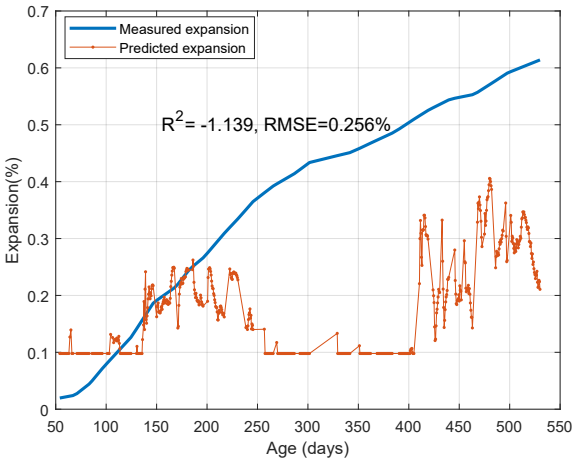


Figure 9. Results of testing on the ASR-2D data using the DNN model trained with ASR frequency-domain spectra [11].

3. COMPARISON OF DIFFERENT MACHINE LEARNING MODELS

The performances of various ML models used for predicting ASR-induced expansion were compared across multiple aspects, including model inputs, complexity, training, and prediction accuracy. The detailed comparisons are summarized in Table 1. Conclusions were drawn from comparing the four ML models studied in both the current and previous works. Some of these conclusions may not be generalizable since they are based on the specific results of this study.

The raw data for all models used in this work are the long-term monitoring ultrasonic signals and the measured expansion data for each specimen. The ultrasonic signal is a 1D time-domain waveform.

Preprocessing: For the LR, SVR, and shallow NN models, the raw signals need to be normalized and then transformed using a fast Fourier transform (FFT) or discrete wavelet transform (DWT) for feature extraction. Wave features, such as velocity, attenuation, and amplitude, can be calculated after these preprocessing procedures. For the DNN model, minimum preprocessing procedures are needed, and the time-domain signals or frequency spectra may be truncated and downsampled to reduce computational cost.

Feature extraction: For LR, SVR, and shallow NN models, extracting relevant wave features (such as velocity, amplitude, attenuation, etc.) and statistical parameters is essential for effective model inputs. This process demands expertise in ultrasonic NDE, as the selection of features is often guided by domain-specific knowledge and practical experience. In contrast, the DNN model can use the time-domain waveforms or frequency spectra directly as inputs without explicit feature extraction. However, like LR and SVR, the DNN model is able to accept the extracted features as the input, although this is less common in ultrasonic NDE applications.

Feature selection: To avoid too many irrelevant feature inputs, performing feature selection for the LR, SVR, and shallow NN models is recommended. This procedure can improve model performance and prevent overfitting during training. The DNN typically does not require feature selection.

The DNN model uses the time-domain signal or frequency spectra as inputs, whereas the other models use extracted features.

Model complexity:

- LR is relatively straightforward, involving a simple linear equation and minimal computational requirements. The model complexity is low.
- SVR has a moderate model complexity, involving finding the optimal hyperplane in a higher-dimensional space and tuning various hyperparameters.
- The shallow NN has one or two hidden layers, making it a less complex model than a DNN.
- The DNN is the most complex, comprising many hidden layers and nodes, requiring extensive data and computational resources, and often involving sophisticated techniques for training and optimization.

Model training: The training LR, SVR, and shallow NN models is relatively simple, involving optimizing a limited number of parameters. DNN model training is more complex because of the large number of model parameters.

Model optimization: LR optimization is simple and involves finding the best-fit line's coefficients. SVR and the shallow NN introduce moderate complexity in optimization. SVR requires tuning three

hyperparameters (kernel function, penalty parameter C, and epsilon parameter ε), and the shallow NN needs to optimize the model structure and corresponding hyperparameters. The DNN requires careful tuning of hyperparameters, extensive computational resources, and sophisticated optimization algorithms like Adam or RMSprop. Overfitting is a major concern, necessitating the use of regularization techniques such as dropout, L2 regularization, and early stopping.

Table 1. Comparison of different ML models for ultrasonic NDE of ASR in concrete

ML model	**LR**	**SVR**	**Shallow NN**	**DNN**
Preprocessing	Extensive preprocess (e.g., normalization, noise filtering, FFT, DWT, windowing)	Extensive preprocess (e.g., normalization, noise filtering, FFT, DWT, windowing)	Extensive preprocess (e.g., normalization, noise filtering, FFT, DWT, windowing)	Minimum preprocessing (e.g., normalization, FFT, windowing)
Feature extraction	Required, based on domain expertise	Required, based on domain expertise	Required, based on domain expertise	Using time-domain signals or frequency spectra directly
Feature selection	Recommended for improved performance	Recommended for improved performance	Recommended for improved performance	Not required
Model development complexity	Low	Moderate	Moderate	High
Model training	Finding the best-fit line	Finding the hyperplane to fit the data	Adjusting the weights and biases	Training multiple layers with many parameters
Model optimization	Simple, best-fit line's coefficients	Simple, kernel function, C, and ε	Simple, model structure and hyperparameters	Complex, requires careful tuning
Model performance	Acceptable	Good	Good	Good
Generalization ability on different batch specimens	N/A	Depends on data ranges	N/A	Poor
Temperature influence	Not studied	Sensitive	Not studied	Less sensitive

Model performance: The LR models in this work achieved an acceptable prediction performance with an R^2 of 0.78 and an RMSE of 0.082% after feature selection. The SVR model performed better, with an R^2 of 0.89 and an RMSE of 0.057%, but only when feature selection was applied; without feature selection, the SVR model's performance was poor. The shallow NN model achieved prediction performance comparable to the SVR model when feature selection was used. However, without feature selection, the LR model outperformed the others. The DNN model also performed comparably to the SVR and shallow NN models, using both time-domain signals and frequency spectra as inputs. Nevertheless, the DNN model's performance varied across different training and testing combinations.

Generalization ability: The SVR model performed well on data from different batch specimens, although its effectiveness depended on the data ranges in the training and testing datasets. In contrast, the DNN models demonstrated poor generalization ability on data from different batch specimens, performing well only on data from the same batch. The generalization abilities of the LR and shallow NN models across different batch specimens were not studied in this work.

Temperature influence: Ultrasonic signals are sensitive to temperature changes in the data-collection environment, which affect features such as amplitude and velocity. A previous study [11] suggests that training and testing datasets should be collected at similar temperatures when using temperature-sensitive features as inputs. Because the SVR model relies on these extracted features, its prediction performance is sensitive to temperature variations between the training and testing data. Conversely, the DNN model uses time-domain signals or frequency spectra as inputs, and because these inputs are amplitude normalized, the DNN model's performance is less affected by temperature variations.

Temporal dependencies: The four ML models studied in previous work and this study treat each data point independently during the training process and do not inherently account for temporal relationships or long-term dependencies. However, ASR damage is progressive and accumulative, suggesting that there should be temporal dependencies among the ultrasonic signals collected at different ages. Therefore, using recurrent neural networks (RNNs) for ASR expansion prediction based on ultrasonic monitoring signals is recommended in future studies.

ASR expansion prediction involves analyzing ultrasonic data collected over time. RNNs are specifically designed to handle sequential data, making them ideal for capturing the progression of ASR damage. RNNs can effectively capture the temporal dependencies and relationships in the data, which are crucial for predicting how ASR expansion will evolve over time. Additionally, RNNs can learn long-term dependencies and patterns in the data, allowing for more accurate predictions of future expansion based on historical data. RNNs maintain a hidden state that evolves as new data points are processed. This state retention allows RNNs to incorporate past information into current predictions, which is essential for understanding the cumulative effect of ASR over time.

RNNs are well suited for ASR expansion prediction using long-term monitoring ultrasonic data because of their ability to model temporal dependencies, learn long-term patterns, retain dynamic states, and effectively handle time-series data. These advantages make RNNs a powerful tool for predicting the progression of ASR damage and supporting proactive maintenance strategies.

4. SUMMARY

This report compares the performance of four ML regression models in predicting concrete material damage induced by ASR expansion based on long-term ultrasonic data monitoring. The models examined include LR, SVR, a shallow NN, and a DNN. Whereas the LR, SVR, and shallow NN models use features extracted from ultrasonic signals as inputs, the DNN model uses processed time-domain ultrasonic signals and frequency spectra directly as inputs.

Various aspects of these models' performance, including model inputs, prediction accuracy, and generalization ability, were systematically assessed and compared. The results indicate significant variation in performance, with ML algorithms capable of achieving either very high or very low prediction accuracy.

Collectively, the results suggest the following:

- **Shallow ML models (LR, SVR, and shallow NN):** These models achieved high prediction accuracy only with appropriate preprocessing procedures and feature engineering.
- **DNN model:** The DNN model, which uses time-domain signals and frequency spectra as inputs, did not require feature engineering. However, preprocessing of these inputs was necessary to remove noise and reduce computational cost.
- **Modeling complexity:** The shallow ML models had lower modeling complexity, whereas the DNN model required additional knowledge and more effort in training and optimization to achieve satisfactory prediction performance.
- **Prediction performance:** Among the shallow ML models, the SVR model demonstrated better prediction performance than the LR and shallow NN models. The DNN model achieved superior prediction performance on specific datasets.
- **Generalization ability:** The shallow ML models demonstrated stable generalization abilities across datasets from different batches of specimens (e.g., ASR and small ASR). In contrast, the DNN model demonstrated poor prediction performance on data from different batch specimens.
- **Temperature sensitivity:** The SVR model's performance was sensitive to the temperature of the training and testing data. The DNN model was less sensitive to temperature variations because it used the whole waveform as the model input.
- **Future work—RNNs:** RNNs are recommended for ASR expansion prediction based on ultrasonic monitoring signals in future studies because of their ability to model temporal dependencies, learn long-term patterns, retain dynamic states, and effectively handle time-series data.

These findings underscore the importance of model selection, preprocessing, and feature engineering in achieving high prediction accuracy and generalization ability in ML models for ASR-induced concrete damage prediction.

ACKNOWLEDGMENT

This research was sponsored by the US Department of Energy Office of Nuclear Energy's Light Water Reactor Sustainability Program (LWRS) under contract DE-AC05-00OR22725 with UT Battelle LLC/Oak Ridge National Laboratory. Work was done under the Materials Research Pathway as part of the Concrete NDE of Civil Structures Task. The experimental ultrasonic data were shared by the University of Nebraska–Lincoln (supported by the US Nuclear Energy University Programs). Xiang (Frank) Chen, Elena Tajuelo Rodriguez, and Pradeep Ramuhalli are acknowledged for their invaluable efforts in project management and their insightful feedback on earlier drafts of this report.

REFERENCES

[1] M. Kawamura and K. Iwahori, "ASR gel composition and expansive pressure in mortars under restraint," *Cement and Concrete Composites*, vol. 26, no. 1, pp. 47–56, Jan. 2004.

[2] J. Zhu *et al.*, "Online Monitoring System for Concrete Structures Affected by Alkali-Silica Reaction," Univ. of Nebraska, Lincoln, NE (United States), DOE-UNL-NE8544, Dec. 2021. doi: 10.2172/1838356.

[3] F. Saint-Pierre, P. Rivard, and G. Ballivy, "Measurement of alkali–silica reaction progression by ultrasonic waves attenuation," *Cement and Concrete Research*, vol. 37, no. 6, pp. 948–956, Jun. 2007.

[4] E. R. Giannini, "Evaluation of concrete structures affected by alkali-silica reaction and delayed ettringite formation," thesis, 2012.

[5] T. Ju, J. D. Achenbach, L. J. Jacobs, M. Guimaraes, and J. Qu, "Ultrasonic nondestructive evaluation of alkali-silica reaction damage in concrete prism samples," *Materials and Structures*, vol. 50, no. 1, p. 60, Aug. 2016.

[6] H. Sun, Y. Tang, C. Malone, and J. Zhu, "Long-term ultrasonic monitoring of concrete affected by alkali-silica reaction," *Structural Health Monitoring*, p. 14759217231169000, Apr. 2023, doi: 10.1177/14759217231169000.

[7] H. Song, S. B. Feldman, and J. S. Popovics, "In situ detection and characterization of alkali-silica reaction damage in concrete using contactless ultrasonic wavefield imaging," *Cement and Concrete Composites*, vol. 133, p. 104661, Oct. 2022, doi: 10.1016/j.cemconcomp.2022.104661.

[8] C. Malone, J. Zhu, J. Hu, A. Snyder, and E. Giannini, "Evaluation of alkali–silica reaction damage in concrete using linear and nonlinear resonance techniques," *Construction and Building Materials*, vol. 303, p. 124538, 2021.

[9] C. Malone, H. Sun, and J. Zhu, "Nonlinear Impact-Echo Test for Quantitative Evaluation of ASR Damage in Concrete," *Journal of Nondestructive Evaluation*, vol. 42, no. 4, p. 93, Oct. 2023, doi: 10.1007/s10921-023-01003-2.

[10] H. Sun and S. Sabatino, "Machine Learning for Processing Ultrasonic Data from Long-Term Monitoring of Concrete with Alkali-Silica Reaction (ASR)," Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), ORNL/SPR--2023/2948, M3LW-23OR0403025, 2076195, Jun. 2023. doi: 10.2172/2076195.

[11] H. Sun and S. Sabatino, “Assessment of Machine Learning for Ultrasonic Nondestructive Evaluation of Alkali–Silica Reaction in Concrete,” Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), ORNL/TM-2024/3295, Mar. 2024.

[12] S. Marsland, *Machine Learning: An Algorithmic Perspective*. CRC Press, 2011.

[13] T. Hastie, J. Friedman, and R. Tibshirani, *The Elements of Statistical Learning*. in Springer Series in Statistics. New York, NY: Springer, 2001. doi: 10.1007/978-0-387-21606-5.

[14] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O’Reilly Media, Inc., 2022.

[15] H. Sun, P. Ramuhalli, and R. Meyer, “An Assessment of Machine Learning Applied to Ultrasonic Nondestructive Evaluation,” ORNL/SPR--2023/3245, 2320378, Dec. 2023. doi: 10.2172/2320378.

[16] M. Muhlheim, P. Ramuhalli, A. Huning, A. Guler, and A. Saxena, “Status Report on Regulatory Criteria Applicable to the Use of Artificial Intelligence (AI) and Machine Learning (ML),” Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), ORNL/SPR-2023/3072, Sep. 2023.

[17] V. Matz, M. Kreidl, and R. Smid, “Classification of ultrasonic signals,” *International Journal of Materials and Product Technology*, vol. 27, no. 3–4, pp. 145–155, Jan. 2006, doi: 10.1504/IJMPT.2006.011267.

[18] G. Guarneri, “Weld discontinuities classification using principal component analysis and support vector machine,” presented at the Brazilian Symposium on Intelligent Automation, Fortaleza, Brazil, Oct. 2013.

[19] K. Virupakshappa and E. Oruklu, “Ultrasonic flaw detection using Support Vector Machine classification,” in *2015 IEEE International Ultrasonics Symposium (IUS)*, Taipei, Taiwan: IEEE, Oct. 2015, pp. 1–4. doi: 10.1109/ULTSYM.2015.0128.

[20] S. A. Shevchik *et al.*, “Laser Welding Quality Monitoring via Graph Support Vector Machine with Data Adaptive Kernel,” *IEEE Access*, vol. 7, pp. 93108–93122, 2019, doi: 10.1109/ACCESS.2019.2927661.

[21] A. Al-Ataby, W. Al-Nuaimy, C. R. Brett, and O. Zahran, “Automatic detection and classification of weld flaws in TOFD data using wavelet transform and support vector machines,” *Insight - Non-Destructive Testing and Condition Monitoring*, vol. 52, no. 11, pp. 597–602, Nov. 2010, doi: 10.1784/insi.2010.52.11.597.

[22] N. A. Akram, D. Isa, R. Rajkumar, and L. H. Lee, “Active incremental Support Vector Machine for oil and gas pipeline defects prediction system using long range ultrasonic transducers,” *Ultrasonics*, vol. 54, no. 6, pp. 1534–1544, Aug. 2014, doi: 10.1016/j.ultras.2014.03.017.

[23] R. Lippmann, “An introduction to computing with neural nets,” *IEEE ASSP Magazine*, vol. 4, no. 2, pp. 4–22, Apr. 1987, doi: 10.1109/MASSP.1987.1165576.