Predicting Delamination in Concrete Bridge Decks from Ground Penetrating Radar Signals using Machine Learning

Authors

- **Ishfaq Aziz**
  0000-0001-5723-0331 · ishfaqa2

- **Jesus Castro**
  0000-0002-7684-7422 · jcastr54

- **Chirayu Kothari**
  0000-0003-4468-1138 · ck0103

- **Omar Abdelrahman**
  0000-0003-0618-8481 · oaa4

Introduction

The data set represents raw signals of Non-Destructive Evaluation (NDE) tests conducted by Ground Penetrating Radar (GPR) which were collected from in-service reinforced concrete bridge decks. The dataset was annotated using three classes in accordance with bridge deck repair: Class 1-No Delamination; Class 2-Delamination (delamination above top bar mat), and Class 3-Delamination (delamination below top bar mat).

The data will be obtained from https://commons.und.edu/data/19/ [1] which contains annotated dataset for the structural defects. Data is present in the. xlsx format. Each column in the dataset has an approximate length of 512 where the first column represents the time (ns), and the rest of the columns are all amplitudes of GPR signals. Hence, each raw signal (columns) is directly annotated to one of the three possible classes. In addition, some other information like the scan length and the coordinates of each signal scan is also provided.

Ground Penetrating Radar (GPR) is an imaging technique that uses wide-band sinusoidal electromagnetic waves, with frequencies ranging from 0.5 to 2 GHz, to produce high-resolution images of the subsurface materials. GPR is an effective tool for subsurface inspection and quality control on engineering construction projects. The survey method is rapid, nondestructive, and noninvasive. Interpretation of GPR data commonly helps to evaluate and measure different properties of a concrete structure.

The data collected from the GPR are usually presented in two different formats, A-scan and B-scan. Most modern devices have the capability to present both formats simultaneously. The A-scan is the raw signal of energy received by the antenna shown as a function of time and signal strength (amplitude). The received signal in Figure 1a is an example of an A-scan. The B-scan, also known as radargram, is constructed from the sequence of multiple A-scans related to the position of the antenna, as shown in Figure 1b, where the depth is represented on the y-axis and the survey distance is shown on the x-axis orthogonal to the y-axis. The amplitude of the received signal is often shown as a color-coded intensity plot, often in grey, as shown in Figure 1c. B-scans are usually visually inspected to identify and locate any delamination within the concrete or to locate the reinforcement rebars.
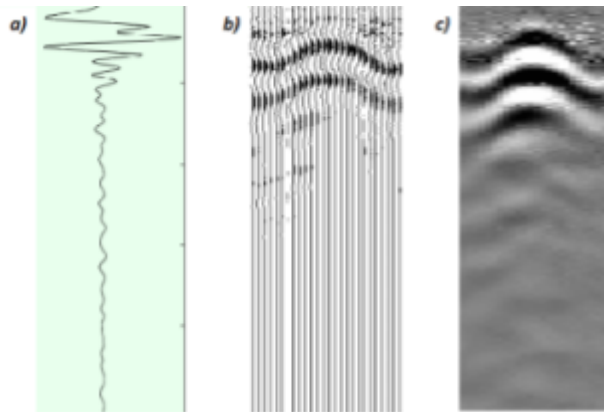

Figure 1: a) Single wave signal (A-scan). b) Collection of signals across a distance along the surface. c) Color intensity plot of (b) (B-scan).

Plan

The A-scan data is categorized into three classes based on the delamination conditions. Fast Fourier transform of the A-scans will be performed to obtain the peak frequency and peak width. These values will act as independent variables in the machine learning model to predict the delamination condition. The bridge health at each location where an A-scan is collected is categorized into three classes: Class 1 No delamination, Class 2 Delamination (above the top bar mat), and Class 3 Delamination (below the top bar mat). Around 10,000 labeled A-scans will be used to train the model

and 3000 labeled A-scans will be utilized for validation. Later, the trained model will be tested on 3000 A-scans to predict the delamination condition. Finally, B-scans will be generated and deck conditions at different locations of the bridge will be predicted.

Exploratory Data Analysis

Preliminary Analysis

The dataset consists of 16,384 GPR scans that were taken at different locations along the bridge. Each scan has a total of 512 points representing a scan time of 12 ns. The scans are categorized into three main classes: Class 1 – no delamination, class 2 – delamination above the top rebar, and class 3 – delamination below the top rebar, representing 8427, 6812, and 1144 scans, respectively, as shown in Table 1. Other properties of the GPR scans are summarized in Table 2.

Table 1. Summary of classes/labels of the dataset

Classes	Class 1	Class 2	Class 3
No of scans	8427	6812	1144
Meaning of each class	Sound Concrete	Concrete with delamination above top rebar	Concrete with delamination below top rebar
Peak amplitude of mean signal	10121	9813	9649

Table 2. Properties of the GPR scans

Parameter	Value
Total number of scans	16,383
Total Length of scan	363.3 ft.
Interval between subsequent scans	0.0222 ft.
Time duration of each scan	12 nano seconds
Data points per scan	512
Sample rate (calculated)	42.67 GHz

The amplitudes of the signals in the dataset were found to have high numerical values. The average amplitude of the whole dataset is around 33,000. So, in order to make the data more symmetric around the x-axis, the average of the first few nanoseconds of readings was subtracted from the whole dataset, resulting in scans that start with amplitudes close to zero. Figure 2 shows one randomly selected scan from each of the three classes.

Figure 2: Randomly selected signals from each of the three classes

Based on the reference paper that was used to obtain the data [1] the B-scan, which is a visual representation of a combination of individual scans stitched together, showed that the bottom rebar reflection was detected at around 7 ns. Thus, it was decided that the readings after around 8 ns would not be useful for the purpose of our model and the remaining 4 ns were removed. Also, it was observed that the initial reflection from the top surface of the concrete is detected around 2 ns, so the first 2 ns were also removed from our data. Sample scans from the resulting signal are shown in Figure 3.

Figure 3: Randomly selected signals from each class, showing only the data points between 2 and 8 ns.

Mean Plots

To visually distinguish between the three different classes of data, the mean all scans from each class are plotted in Figure 4. The figure shows that the three means look almost identical to each other, so there is no distinctive feature in the time domain signal that can help assign new signals to any of the three classes. Thus, deeper levels of data analysis are required to identify any distinctive features that can help classify new data.

Figure 4: The mean of all signals for each class.

Fast Fourier Transform (FFT)

Another approach to visualizing sinusoidal time-series datasets is by identifying dominant cyclic patterns. This can be conducted with the Fast-Fourier Transform, a computationally efficient way to calculate the discrete Fourier Transform from a dataset. This algorithm transforms information on the time domain into the frequency domain.

The time-series datasets from the three classes (no delamination, delamination above rebar, and delamination below rebar) were all transformed into the frequency domain to identify patterns that could easily differentiate one class from another. As observed in Figure 5, most of the signal amplitude peaks are observed at a frequency of approximately 0.8 GHz, which is in the standard operating frequency range of GPR antennas for concrete testing.

Figure 5: Amplitude vs Frequency Class 1 with no delamination

In Figure 6 it is observed that in the sections where delamination was identified above the top rebar layer, the amplitude peaks were higher in magnitude than the observed peaks in the sections with non-delaminated sections at similar frequencies. On the other hand, the plot corresponding to the amplitudes of sections identified with delamination below the top rebar later (Figure 7.) displays amplitude peaks with magnitudes lower than in Figure 5 and Figure 6.

Figure 6: Amplitude vs Frequency Class 2 with delamination above the rebar

Figure 7: Amplitude vs Frequency Class 3 with delamination below the top rebar

From this data processing approach (FFT), a sound differentiation from the three classes (beyond the observed magnitudes in amplitude at approximately 0.8 GHz) cannot be concluded. Furthermore, the maximum amplitude of the FFT data and the frequency corresponding to the maximum amplitude was plotted (Figure 8). It was observed that the maximum amplitude of most of Class 1 and Class 2 was similar. However, some of the FFT spectra in Class 2 had higher amplitude than that of Class 1. Also, some of the FFT spectra in Class 2 had amplitude close to the mean of Class 3. It can potentially be attributed to the labeling of data.

Figure 8: Violin plots representing the variation in maximum FFT amplitude of the three classes.

Similarly, the frequency at maximum amplitude was observed for the FFT spectra in the three classes (Figure 9). The mean frequency at maximum amplitude for class 3 was lower than that of class 2 and class 1. The Violin plot for class 1 and class 2 look similar. However, some of the FFT spectra in class 2 have a frequency at a maximum amplitude lower than that of class 1's data.

Figure 9: Violin plot representing the variation in frequency at maximum FFT amplitude of the three classes.

It can be easier to separate out Class3 from Class 1 and Class 2 based on the FFT analysis.

Principal Component Analysis

Principal component analysis (PCA) was performed on the dataset to change the basis of the data and improve its interpretability. The number of modes was selected to be 2. The results of the first and second modes of the PCA data of the whole dataset are plotted in Figure 10. Since the number of data points for each class is different, normalizing the number of points might provide a better visual representation of the scatter of data of different classes. So, 1000 random points were selected from each class and then PCA was performed on this sub-dataset. The results of the two modes of this PCA are shown in Figure 11.

Figure 10: The PCA data for all the datasets.

Figure 11: PCA data of two different 1000 randomly selected points from each class.

Predictive Modeling

Based on the exploratory data analysis it can be concluded that the differentiation of the classes is challenging. The analytical approach to determining key features for distinguishing different classes is not conclusive and overall, the data delineation is not observed. Hence, four different types of models including the Convolutional Neural Network (CNN) were used to classify the delamination into three categories. The dataset was divided into training and testing data.

The four models used in the preliminary predictive modeling are 'Decision Tree', 'Random Forest', 'Regression', and '1-D Convolutional Neural Network'. The explanations of each of these models and the corresponding results are discussed below.

1. Decision Tree

The first predictive model applied was the 'Decision tree' model, which is a non-parametric supervised learning method that can be used for classification or regression. Decision trees learn from data to approximate a sine curve with a set of if-then-else decision rules. The deeper the tree, the more complex the decision rules, and the fitter the model.

Decision tree in machine learning has the advantage of being simple to understand, interpret and visualize. The model can be visually represented by a flowchart of decisions. It can also implicitly perform variable screening or feature selection to determine the most dominant variables that can govern the classification more than the others. However, it also comes with some limitations. For instance, decision tree learners can create over-complex trees that do not generalize the data well, causing data overfitting. In addition, the predictions of decision trees are neither smooth nor continuous, so they do not perform well with extrapolation.

The decision tree classifier was created using 3 labels -one for each class- and 12 variables. These variables are the 10 first PCA modes of the data, the dominant FFT frequency, and its corresponding amplitude. The first 10 PCA modes were selected as they cover more than 90 % of the variance. Then, leaves having more than 90% combined purity were merged before applying the model to avoid overfitting. About 60 % of the data from each class was used for training the model. Then, the model was tested on the complete dataset.

The summary of the results is shown in Table 3. The accuracy achieved on training data for classes 1, 2, and 3 were around 86%, 83%, and 46%, respectively. This resulted in an overall accuracy of about 82% on the training data. Due to the lower number of Class 3 data points, the model prediction accuracy is lower for class 3 than for the other two classes. The confusion matrices achieved by using this model are shown in Figure 12 and Figure 13. When applied to the complete dataset, the achieved accuracy was around 67%, which means that the accuracy of the unseen 40% of data was about 45%.
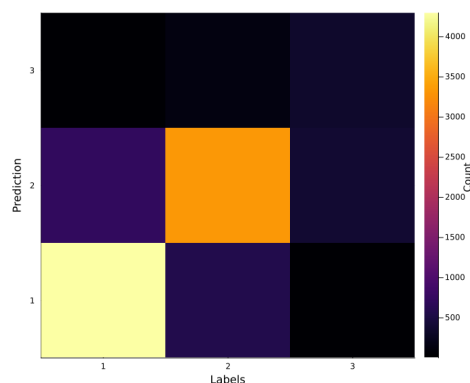

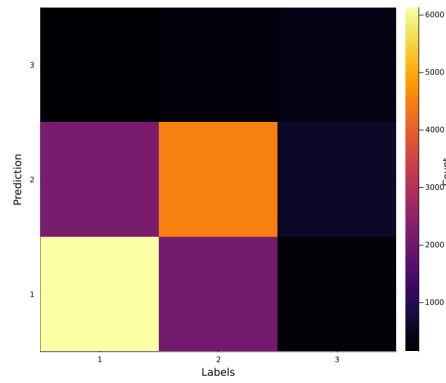**Figure 12: Confusion matrix of the training data for the 'Decision Tree' model**

Figure 13: Confusion matrix of full dataset for the 'Decision Tree' model

2. Random Forest

To avoid the problem of overfitting, we have also tried to apply the 'Random Forest' classifier. A random forest is a meta-estimator that fits several decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. The parameters used for building our model are listed below:

• Number of random features = 6

• Trees = 10

• Portions of samples per tree = 0.6

• Maximum tree depth = 17

The accuracy obtained by the model using the numbers mentioned above was around 64 %. Values of these parameters were changed to default values (default of the package: DecisionTree.jl), and then the new model was generated and tested. With the default values, the achieved accuracy was around 66 %. The confusion matrix obtained from this model is shown in Figure 14.
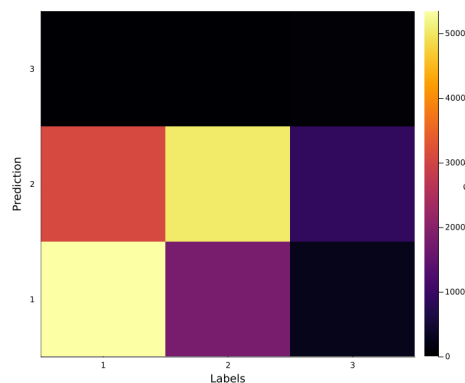

Figure 14: Confusion matrix of full dataset for the 'Random Forest' model

3. Regression

Regression is a method for creating a model of the relationship between one or more independent variables and one or more dependent variables, where variation in the dependent variables is used to explain variation in the dependent variables. The regression method prepared for this project considered a total of 260 independent variables and 1 dependent variables (classification). The regression model was optimized by minimizing a maximum square error function, used a learning

rate of 0.001 and 10,000 steps. 60% of the entire data set was used as training for the regression model.

Ideally, regression model results should match actual observations. As observed in Figure 15, similar values were predicted for the three observed classes, which unables this system alone to accurately classify the predictions. This finding confirms that regresion model is not the most adequate tool for a classification model.
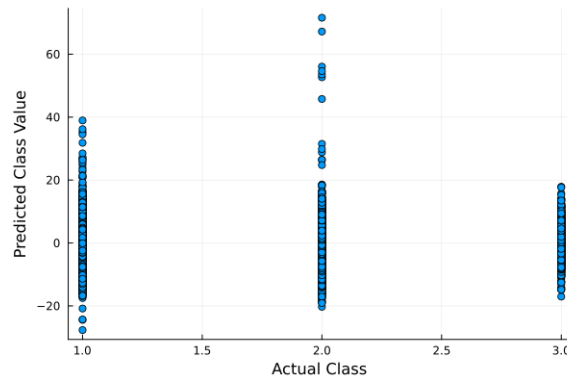

Figure 15: Class Predictions vs Actual Classes, from regresion-based machine learning model

4. Convolutional Neural Network

The Convolutional Neural Network (CNN) was created for the GPR data at different locations. Out of the approximately 16000 signals, around 9800 random signals were used to train the model. Each signal has 260 data points. The 1D CNN is composed of several layers. The first layer was a convolution layer with a filter size = 5 with 1 input channel to 8 output channels of training data. The padding was of size= 2 and Relu was the activation function. The next layer was a Maxpool layer of size= 4. The third layer is another convolutional layer with a filter of a size = 5 and padding of size= 2 with 8 channels to 3 channels. The fourth layer with a Maxpool layer of size =5 and the fifth layer is a dense layer of size (39,3) based on the input size and the number of channels. Finally, a Softmax layer is applied to discriminate between different classes. The batch size was 10 and learning rate was 0.001.

Figure 16 show some diagnostics of our model training, include loss and accuracy curves as a function of gradient descent step and a confusion matrix of the final results. The accuracy of 52% is achieved based on the convolutional neural network. As observed in the figure , the accuracy becomes constant after certain epochs and does not improve. The model was further executed for different channels as well as different activation functions like Relu, Tanh, and its combination. However, the output was close to 52% accuracy with no improvement. The rest 40% of the total data was tested based on the CNN model. The accuracy was again 52% and the correlation matrix for the testing data was similar to that of the correlation matrix for the training data(scales on the color bar are different as the number of datasets for training and testing data is different) Further modification and improvements in the CNN model are required.
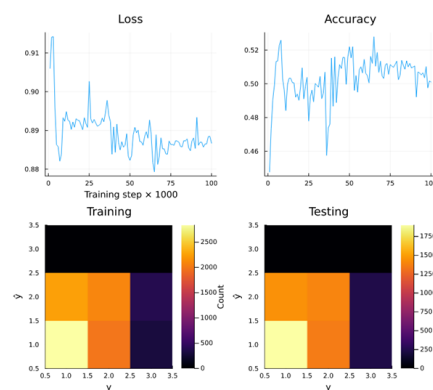
Figure 16: Performance of the CNN model. It shows the loss and accuracy curves as a function of gradient descent step and a confusion matrix for both the training data and the testing data

Discussions

As mentioned, the data is classified into three (3) different labels/classes. However, the amount of data samples for the three classes is not the same. For example, class 3 represents only about 7.5% of the entire dataset, while class 1 is approximately 50%. Table 3 provides a accuracy/performance based summary of the effectiveness of each of the studied predicitve models.The accuracy of predictive models is linked to the sample size of each class. An imbalanced dataset may cause predictive models to be biased towards learning more about the dominant classes causing higher accuracy of prediction of these classes. Moreover further improvement in CNN model by optimizing the model parameters and layers is required. While Regresion model performance was measured in terms of MSE, which may not be comparable to other accuracy metrics. Decision Tree and Random Forest models provided relatively high accuracy levels on training data (above 75%), 1-D CNN model provided the highest accuracy on unseen data with 52%.

Table 3. Summary of results of predictive modeling

Models	Training data	Accuracy On Training data	Accuracy On Unseen data
Decision Tree	60 %	82 %	45 %
Random Forest	60 %	76 %	49 %
Regression	60 %	6.1 (MSE)	5.5 (MSE)
1-D CNN	60%	52 %	52 %

References

[1] Ichi, E., & Dorafshan, S. (2021). SDNET2021: Annotated NDE Dataset for Structural Defects, 10.31356/data019

[2] Scheers, B., 2001. Chapter 2. Ground penetrating radar. In: Ultra-Wideband Ground Penetrating Radar, with Application to the Detection of Anti-Personnel Landmines. Brussels: s.n., pp. 2-1 -2-43.