Monitoring Crack Propagation using Data-Driven Causal Discovery and Supervised Learning

Rui Liu^ and Dr. Siddharth Misra^*

^*Harold Vance Department of Petroleum Engineering, College of Engineering, Texas A&M University*

**Department of Geology and Geophysics, College of Geosciences, Texas A&M University*

Abstract

Mechanical wave transmission through a material is influenced by the mechanical discontinuity in the material. The propagation of embedded discontinuities can be monitored by analyzing the wave-transmission measurements recorded by a multipoint sensor system placed on the surface of the material. In our study, robust monitoring of the propagation of a mechanical discontinuity is achieved by using supervised learning followed by data-driven causal discovery to process the multipoint waveform measurements resulting from a single impulse source. The new data-driven causal-discovery workflow jointly processes the nine 25-μs waveforms measured by the multipoint sensor system comprising 9 sensors. The proposed workflow can monitor the propagation of mechanical discontinuity through three stages, namely initial, intermediate, and final stages. The workflow considers the wave attenuation, dispersion and multiple wave-propagation modes. Among various feature reduction techniques ranging from decomposition methods to manifold approximation methods, the features derived based on statistical parameterizations of the measured waveforms lead to reliable monitoring that is robust to changes in precision, resolution, and signal-to-noise ratio of the multipoint sensor measurements. Causal signatures have been successfully identified in the multipoint waveform measurements. The numbers of zero-crossing, negative-turning, and positive in the waveforms are the strongest causal signatures of crack propagation. Higher order moments of the waveforms, such as variance, skewness and kurtosis, are also strong causal signatures of crack propagation. The newly discovered causal signatures confirm that the statistical correlations and conventional feature rankings are not always statistically significant indicators of causality.

***Keywords**: Signal Processing, Mechanical Discontinuity, Monitoring, Causal Inference, Feature Engineering*

1. Introduction

The word “discontinuity” is a collective term for any distinct break or interruption in the integrity of a material with zero or low tensile strength (ISRM, 1978). Mechanical discontinuity also signifies a sudden change in mechanical properties exhibiting high contrast along an interface. From a geophysical point of view, discontinuities refer to mechanical weakness, such as joints, fractures, and bedding planes, that occur at multiple scales (Osogba et al., 2020). Mechanical discontinuities are common geological phenomena with varying length, frequency and orientation. It plays an essential role as potential transport pathway and has dramatic effect on the bulk mechanical, physical, and chemical behavior of the material.

Many measurement techniques are available for describing mechanical discontinuities at different scales. During reservoir development, borehole image logs, surface geological analysis, well tests, well logging, seismic surveys, and cross-well tomography are used to identify large-scale fracture zones in formation. In laboratory experiments, non-destructive testing (NDT) is the identification and characterization of damage or defects on the surface and within a material without cutting or altering the material (Dwivedi et al., 2018). To date, numerous NDT techniques can be broadly categorized into five groups: 1) visual inspection; 2) acoustic wave-based techniques (e.g. Acoustic Emission) and Ultrasonic testing; 3) optical techniques, 4) imaging-based techniques (e.g. X-ray radiography); and 5) electromagnetic field-based techniques (Wang et al., 2020). In particular, direct NDT techniques such as acoustic emission and ultrasonic-wave testing are widely used for the detection of microstructural damage mechanisms in material. Even though acoustic emission testing allows for real-time damage detection and localization, this NDT technique is still qualitative and does not provide quantitative characteristics such as the dimensions of discontinuities. (Duchene et al., 2018). On the other hand, ultrasonic-wave testing operates in reflection, transmission and backscattering of elastic waves in the material systems (Ibrahim, 2014). It keeps the transducer and receiver off the surface, which is particularly beneficial when complex geometries do not allow contact (Gholizadeh, 2016).

With the development of machine learning, many studies have incorporated data-driven methods to locate mechanical discontinuities (e.g. Muratov et al., 2020; Chakravarty et al., 2021). In an earlier study, we developed a classifier-based workflow to categorically characterize certain bulk properties of the embedded crack clusters, such as orientation, dispersion, and spatial distribution, by processing multipoint compressional- and shear-wave arrival/travel times (Misra & Li, 2019; Liu and Misra, 2020; Li et al., 2021). However, in the previous works, we did not account for the mode conversion, reflection, attenuation, and dispersion of the wave propagation and ignored a large amount of the signal behind the first arrival during the analysis.

In the present study, multiple waveforms are simultaneously processed using a data-driven workflow to non-invasively visualize the propagation of mechanical discontinuity through three stages, namely initial, intermediate, and final. Supervised learning followed by causal discovery is used for a first-of-its-kind monitoring of the propagation of a mechanical discontinuity by jointly processing the multiple waveforms measured by the multipoint sensor system. Causal signatures have been successfully extracted from the measured waveforms that explain the causal relationship between the propagation of mechanical discontinuity and the measured multiple waveforms (i.e. multi-point signals).

2. Workflow Description

Our study has two interconnected purposes: 1) achieve robust monitoring of the propagation of mechanical discontinuity in a material by processing multiple waveforms resulting from wave transmission through the material due to a single, impulse source; and 2) discover causal signatures in the recorded multipoint waveforms due to the propagation of the mechanical discontinuity. **Figure 1** elaborates the data-driven workflow used in this study. The three main components of the workflow include: 1) wave-transmission data generation using k-Wave; 2) supervised learning to estimate the location, orientation, and length of the discontinuity as it propagates in the material through the three stages; and 3) data-driven causal discovery.

Section 3 describes the simulation of elastic wave propagation through a material with a propagating mechanical discontinuity. Physics-based open-source k-Wave toolbox is used to simulate wave propagation in a 2D material that undergoes three stages of propagation/growth of a mechanical discontinuity. The design of the transmitter-receiver configuration is presented in section 3.1. The transmitter-receiver (source-sensor) configuration is inspired by real-world experiments (Misra et al., 2019; Chakravarty et al., 2021). In this study, we will refer to discontinuities as cracks. More information about the properties of designed material and crack stages are present in section 3.2. To better simulate the experimental data, the effects of sensor sampling rate and precision of the sensors will also be discussed later. Multipoint waveforms corresponding to each stage of crack growth are recorded by 9 receivers placed around the 2D rectangular material. Each waveform is recorded for 25 microseconds, discretized into 2500 time steps. Overall, we generated a large dataset for 20,000 materials, each undergoing three stages of crack growth of varying length, orientation, and location of cracks. To overcome the curse of dimensionality, the regression analysis will start with the selection of feature reduction techniques.

In Section 4, we evaluate the performance of various feature reduction techniques and determine the best feature set to accurately monitor the location, orientation, and length of the crack as it evolves through the three stages of crack propagation. The selected feature sets are divided into training and test sets. The training dataset is used to build the data-driven models, while the test dataset is used to evaluate the

generalization of the trained data-driven models. It is very important to ensure that there are no common samples between the training and test sets. Moreover, during the data preprocessing, involving feature extraction and dimensionality reduction, it is of utmost importance to ensure that there is no information leakage between the training and testing datasets. The performance of the regressors are evaluated by the root mean square error (RMSE) between the actual and predicted properties of the crack embedded in the material. This metric helps compare the capabilities of the traditional regression models against neural-network based regressors.

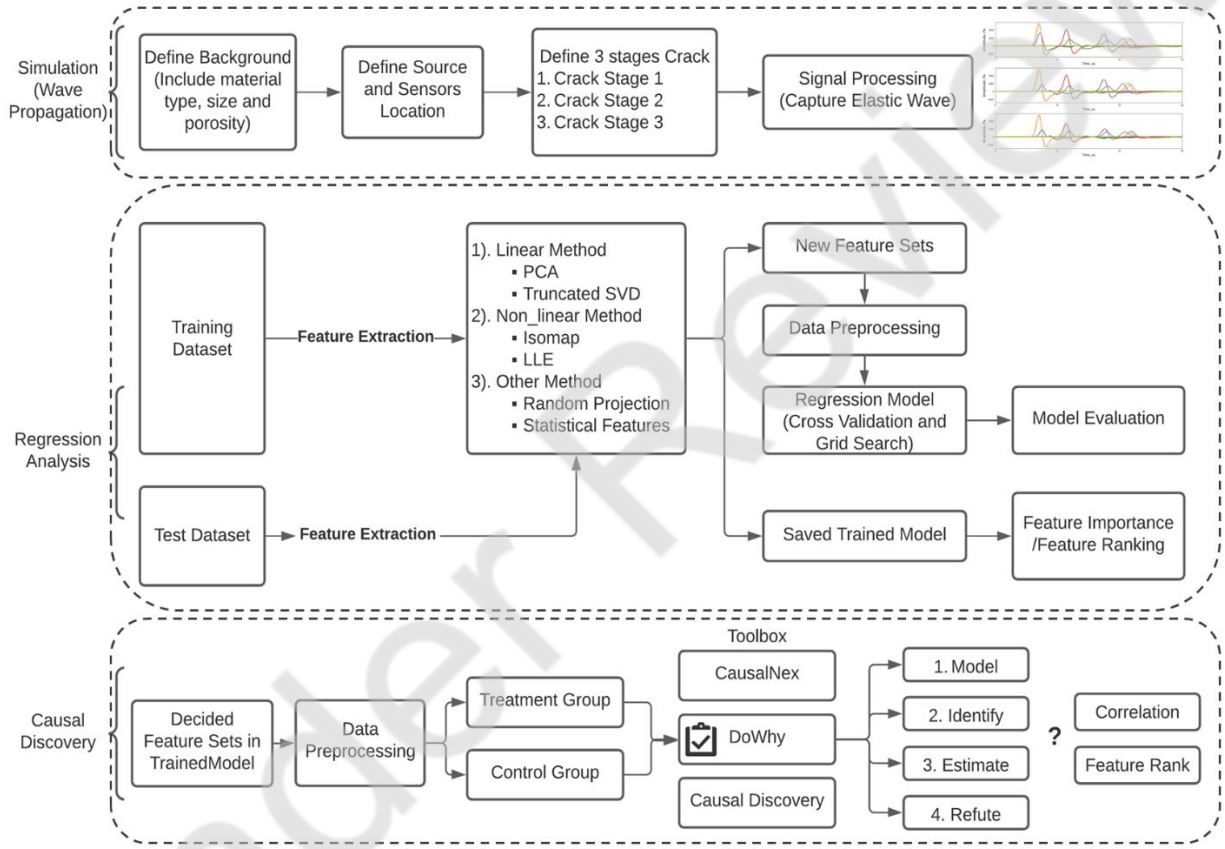


Figure 1. Brief overview of the data-driven workflow implemented in this study, which includes simulation of wave propagation, supervised learning, and data-driven causal discovery. Data is simulated to train the regressors to monitor the crack propagation. Feature extraction and data-preprocessing performed on the simulated data prior to training the regressors to reliably monitor the crack propagation. Finally, data-driven causal discovery technique is applied on the dataset to discover the causal signatures in the waveforms that are caused by the propagation of crack.

Until now, in several engineering applications of data-driven methods, there has been a large emphasis on quantifying the feature ranking based on the correlation/association of the feature with target. There has been extremely limited attention on causality. Causal discovery in section 5 sets out to investigate the

usefulness of causality and explains the critical influences that lead to certain distinct causal signatures in the measured multipoint waveforms due to the crack propagation. We will also provide a detailed overview of the concepts and tools used for the data-driven causal discovery. Finally, we demonstrate new connections between the concepts of causality, correlation, and feature importance/ranking. The conclusions and discussions for future research are drawn in Section 6.

3. Simulation of Elastic Wave Propagation through a Material Undergoing Crack Growth

3.1. Experiment design

K-Wave is an open-source MATLAB toolbox designed for time-domain acoustic and ultrasound simulations in complex materials (Treeby & Cox, 2010). This toolbox can handle the propagation of elastic waves based on two coupled first-order equations describing the stresses and particle velocities within the material. Moreover, the effects of wave propagation, such as attenuation, dispersion, and multiple modes of propagation have been considered in k-wave simulation. In this work, we use K-wave to create a computationally efficient model for elastic wave propagation in porous sandstone-like materials. The elastic simulation functions (pstdElastic2D and pstdElastic3D) are invoked to perform the desired simulation with four input structures: kgrid, medium, source and sensor location. First, all simulations assume a 2D homogeneous material of size 60mm by 60mm discretized using 300 by 300 grids. The material discretization is called with the utility function makeGrid, which takes the number and spacing of the grid points in each Cartesian direction and returns an object of the kWaveGrid class (Treeby et al., 2014).

Following that, the material properties in the medium can be characterized by the elastic wave velocities and the mass density. Elastic waves can be divided into body waves and surface waves according to the way they propagate through a material. Compressional (P-wave) and shear (S-wave) waves as body waves are most often used for inspecting defects (He et al., 2019). Here, without considering spatial variations of water saturation and stress, the compressional wave velocity of the crack-bearing material is set to 3760 m/s and the shear wave velocity is set to 2300 m/s representing a water-filled porous sandstone (Hamada & Joseph, 2020).

Next, the locations of source and sensors are defined on the surface of the material. The numerical models of material containing discontinuities are inspired by the laboratory experiments conducted at the Integrated Core Characterization Center (Bhoumick et al., 2018; Misra & Li, 2019). In those studies, they placed multiple sources and sensors/receivers around porous cylindrical rock to observe the distribution of fractures inside the samples. In this manuscript, we modeled a similar simulation in K-Wave, with nine sensors and a pressure impulse source surrounding the 2D homogeneous material shown in **Figure 2**. The black dot marks the sensor and the red triangle is the impulse pressure source. The source and sensor locations in K-Wave are defined as a series of Cartesian coordinates within the computational grids. If the

Cartesian coordinates do not exactly match the coordinates of a grid point, the output values are calculated from the interpolation.

Finally, simulation time step is defined by user with kgrid.t_array (0:dt:t_end) or within simulation function (makeTime). The time array must be evenly spaced and monotonically increasing (Treeby et al., 2014). A good discussion is required on Courant-Friedrichs-Lewy (CFL) condition to balance the computational time and stability. It depends on both the maximum speed of sound in the medium and the speed of sound used inside the k-space operator (Mast et al., 2001). After the time loop is complete, model returns the field variables recorded at the sensor points. By default, the visualization of the propagating wave field and a status bar are displayed for both compressional and shear components during the simulation (Treeby et al., 2014). In total, we run simulations for 20,000 samples of 2D homogenous material undergoing three stages of crack propagation and store the 25-μs waveforms recorded by the 9 sensors for as each sample. The three stages of crack growth are described in detail in the following sections.

3.2. Three stages of crack propagation

There are no restrictions on the values for the material and crack properties. However, a large contrast between matrix and crack material properties may lead to unstable, non-convergent simulation. Air-filled cracks cause unstable simulations due to large contrast between the air velocity and matrix velocity. Cracks are assumed to be water filled and P-wave velocity of 1418m/s. Shear waves cannot propagate in liquids, the S-wave velocity through the water-filled crack is set to a very small value close to zero (1m/s). In our study, the crack propagation involves linear extension of a crack in the material. **Figure 2** shows two random samples with three stages of crack propagation. The first, second, and third column of the figure show the initial, intermediate, and final stage of the crack, respectively, and corresponding waveforms recorded by sensors are shown at the bottom row.

The crack locations, orientations and lengths are generated randomly using the Latin hypercube sampling (LHS) method having a width of 0.6mm (3 grids). For the first stage, the crack length ranges from 10mm to 20mm (50 to 100 grids). The orientation and location of the first stage can be anywhere inside the material. For the second stage of crack propagation, the crack can extend by a length of 6 to 12 mm. In addition, the direction of the propagation during the second stage will be a constant value limited to an angle smaller than 60-degree relative to the direction of the crack in the first stage. Similarly, the crack extension in the third stage is between 2mm to 8mm at an angle smaller than 60-degree relative to the direction of propagation during the second stage. The primary goal of our study is to monitor these propagations of these discontinuities over the three stages and also to identify causal signatures in the measured waveforms corresponding the crack propagation.

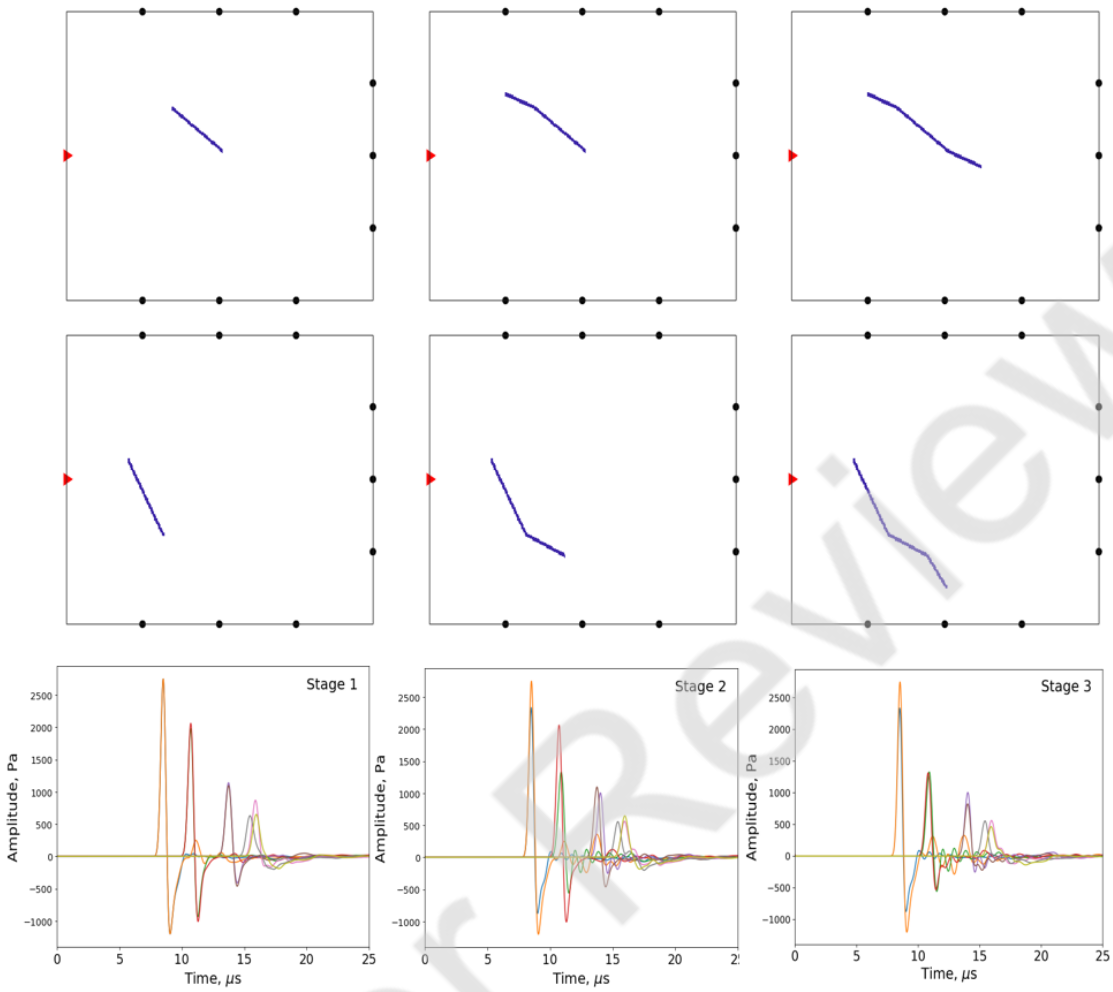


Figure 2. The first two rows of figures show two randomly chosen material samples undergoing the three stages of random crack growth. The left plots represent initial crack stage, the middle plots represent the intermediate crack stage, and the right plots represents the final crack stage. Elastic wave transmission through the material shown in the top row results in the multipoint waveforms shown at the bottom. The waveforms change with crack propagation. The changes are minute and hard to detect using our eyes. The red triangle in the left boundary is the impulse source, while the black dots around the material are the 9 sensors. Each sensor records a 25-μs waveform due to the impulse generated by the source. Different colors of the waveform indicate waveforms recorded by different sensors.

The whole dataset has 20,000 samples. Data generated for each sample comprises 25-μs waveforms recorded at 9 sensor locations during three stages of crack growth. Each 25-μs waveform is discretized into 2500-time steps. In total, eighteen 25-μs waveforms, comprising 2500 amplitudes corresponding to the 2500 discrete time steps, are recorded for each sample. In the 25μs window of recorded waveforms shown in the last row of **Figure 2**, the useful, relevant, and independent information are present only in certain regions of the waveform. Therefore, several feature extraction methods were applied to the dataset before training the regressors to dimensionally reduce the large dataset.

3.3. Sensor sampling rate and precision

The sample rate (or sampling rate) of a sensor is the number of samples measured per second. The units for sample rate are samples per second (SPS) or Hertz. The precision of the sensor represents the resolution of the sensor. Sensor sampling rate and precision are configurable to match the experimental objectives. Moreover, the physics of the sensor and the process under investigation limit the sampling rate and precision of the sensor. In our simulation, 2500 points are recorded in 25μs representing a receiver sampling rate of 10MHz to satisfy the simulation stability. We resample the simulated data to get closer to the real experimental data. In **Figure 3**, we compare the waveforms recorded for a material at the final stage of crack propagation for a sensor sampling rate of 10MHz, 2MHz and 1MHz. A sampling rate of 2MHz can be achieved by most sensors and that sampling rate guarantees preservation of the wave shape, as shown in Figure 3. It means that a 25-μs waveform comprises of 500 time steps instead of 2500 time steps. Precision represents the degree to which an instrument repeats the same value of measurement. Our pressure sensors are set up with the same precision as laboratory experiments that can detect pressure differences as small as 0.01 Pascal.

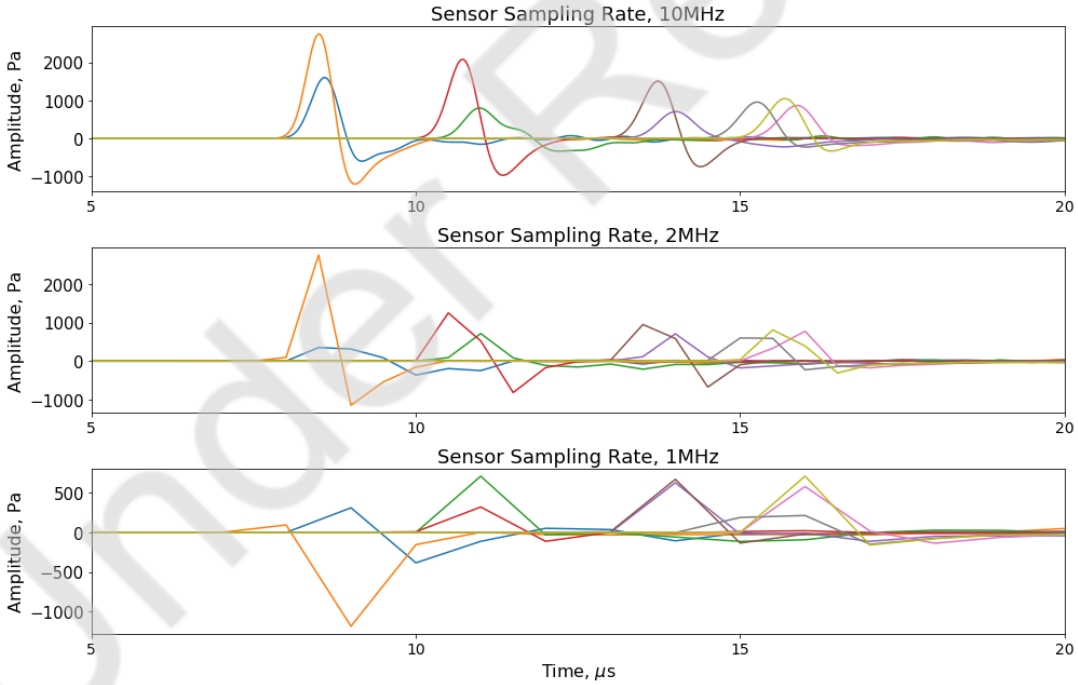


Figure 3. Comparison of the waveforms recorded by the 9 sensors at the final stage of crack growth for various sensor sampling rates. The top, middle and bottom figures identify the sampling rate of 10MHz, 2MHz and 1MHz, respectively. A sampling rate of 2 MHz and precision of 0.01 Pa are chosen for all waveform measurements used in this study. Sensors have limitations in the sampling rate and precision; consequently, the data used for training the regressors honor the physical limitations of sampling rate and precision.

4. Regression Analysis

4.1. Feature reduction techniques

Feature selection/extraction is a critical step in data analysis and machine learning for reducing computational costs, reducing uncertainty due to noise, and improving the model generalization (Bolón-Canedo et al., 2015; Misra and Yu, 2019). Designing the optimal dimensionality reduction workflow for specific predictive modeling requires extensive numerical testing. Principal component analysis (PCA) is a popular multivariate statistical technique, which maps data onto a linear subspace and reduces the dimensionality of the variable space by representing it with few orthogonal vectors (Abdi & Williams, 2010). Similar to PCA, Truncated support vector decomposition (SVD) is a matrix factorization technique suited for linear dimensionality reduction. However, it operates on sample vectors directly instead of on a covariance matrix. Moreover, Isomap and locally linear embedding (LLE) are well-known manifold approxmation algorithms for nonlinear dimensionality reduction (Chakravarty and Misra, 2021). Isomap attempts to preserve local topology on all scales, mapping nearby points close and distant points far away from each other (De Silva & Tenenbaum, 2002). LLE does the same basic thing as Isomap, except that small neighborhoods are stitched together in a different way (Ventura, 2008). Random projection includes gaussian random projection (GRP) and sparse random projection (SRP) can also reduce the dimensionality by projecting the original input space on a randomly generated matrix. Finally, we also extract twenty optimal statistical features from each waveform per sensor per stage of crack growth. This led to 20×9×3 equal to 540 features. The 20 statistical parameters extracted from each waveform includes measurements of shape (skewness, kurtosis), central tendency (mean, medium), position (percentile, zero crossing), impurity (entropy) and other statistical parameters. The detailed list of all 20 statistical parameterization-based features and their definitions are elaborated in the Appendix A. Similar statistical parameterizations have been used by Foster et al. (2021) to assist machine learning methods.

4.2. Comparison of the generalization performance of regression methods for various feature extraction techniques

Regression analysis can discover relationships between features and continuous-valued multiple targets. Features need to be extracted from raw data and the dimensionally reduced prior to training the regression model. After feature extraction, the dataset corresponding to each stage of crack growth can be visualized as a table/matrix with 20,000 sample instances as rows, N features as columns, and T continuous-valued targets as columns. The multiple continuous-valued targets are the x and y coordinates that define each stage of the crack, i.e., eight target values (four x coordinates and four y coordinates that define the crack propagation from the initial to intermediate to final stage). In this study, we use random forest (RF), k-nearest neighbor (KNN), gradient boosting (GB) and neural network (NN) to perform the desired regression to relate the features extracted from the waveforms measured during the three stages of crack

propagation to the location, orientation, and length of crack during each stage of propagation. The 20000 samples are split into 14000 training samples and 6000 testing samples. It is of utmost importance to ensure there is no leakage of information from the testing data into the training data during the feature extraction and other data-preprocessing steps.

RF is an ensemble learning method for both classification and regression. It is a bagging technique that constructs several decision trees and aggregates the predictions of all the trees to generate the final prediction. KNN is a non-parametric machine learning method that approximates the association between features and targets by averaging the targets of the *K* nearest neighbors in the feature space. It is a lazy algorithm that needs all the training data during the testing and deployment phases (Song et al., 2017). KNN delivers simple, flexible and adequate model for nonlinear problems, but is computationally expensive and requires feature scaling. GB is another ensemble learning method, wherein each subsequent learner fixes the shortcomings of the previous weak learners by using gradient descent optimization of a loss function (Misra et al., 2019). Random forest trains and deploys the trees in parallel, whereas the gradient boosting trains and deploys the trees in series. Gradient boosting is susceptible to overfitting. Neural networks (NN) are complex node systems with strong adaptive learning capabilities. Single layer NN has three components: input layer, hidden layer, and output layer. Data is fed to the input layer, and predictions are made on the output layer, also called the visible layer. There can be one or more non-linear layers, referred to hidden layers, between the input and the output layer. Each layer of the neural network has multiple nodes as computational units (Misra and He, 2019). A node on one layer can be fully or partially connected to the nodes on the next layer. A single-layer network can be extended to a multiple-layer network by adding hidden layers. A multilayer perceptron (MLP) is a specific neural network with multiple hidden layers. It is often applied to supervised learning problems that train feature-target pairs and learn to map the statistical correlation between those features and targets. Neural networks are susceptible to overfitting and hard to optimally design.

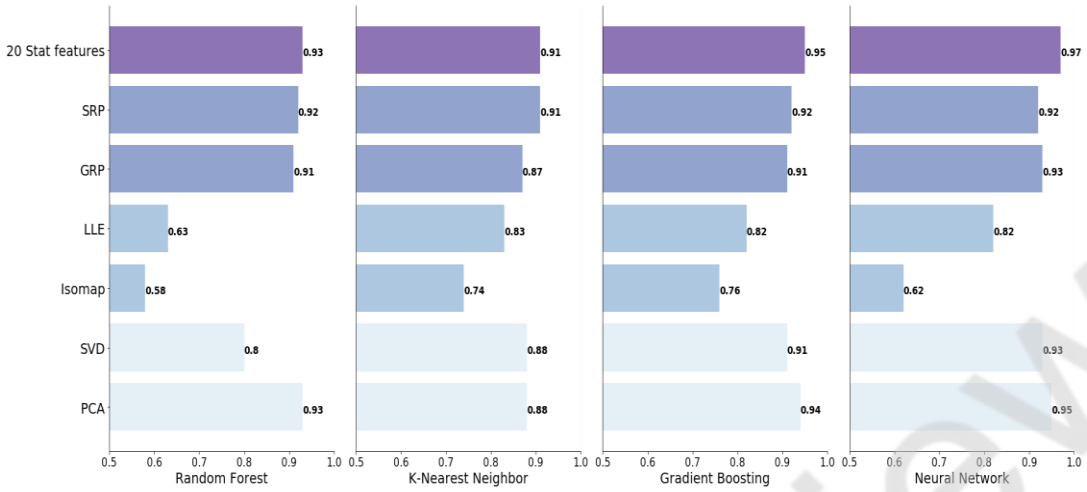


Figure 4. Generalization performances of the four regression methods for the seven different feature extraction techniques. Neural network and gradient boosting enable the best monitoring of crack propagation. Statistical parameterization and principal component analysis-based features enable the best monitoring of crack propagation.

Eventually, 14000 samples undergoing three stages of crack growth are used as training data to build the regressors. Hyperparameter tuning is applied with grid search to optimize the trained regressors for the best performance on the test data. Additional 6000 samples undergoing three stages of crack growth were kept separate to test the trained regressors. **Figure 4** compares the testing performances of RF, KNN, GB and NN for the seven different feature extraction methods discussed in the previous subsection. The numbers labeled in the figure are provided to assess the performance of the regressors, in terms of R-squared, which provides a measure of how well the predictions replicate the crack properties in the testing samples as a proportion of the total variations in crack properties.

The performances presented in **Figure 4** indicate that neural networks and gradient boosting are more accurate than KNN and random forest for the prediction problem under investigation. Regarding the feature extraction techniques, features obtained through the statistical parameterization and those obtained using principal component analysis lead to more accurate regression performance as compared to the remaining five techniques. For most of the feature extraction techniques, KNN has the lowest performance. Topology based feature extraction lead to lower regression performance, where Isomap has the lowest performance and LLE has slightly higher performance than Isomap.

4.3. Generalization performance of the Neural-Network based regression model

Neural network trained on the 540 statistical-parametrization-based features exhibited the best generalization performance. 20 statistical parameters were extracted from each of the 9 waveforms recorded by the 9 sensors for the 3 stages of crack growth, i.e. $20 \times 9 \times 3 = 540$ features. The neural network learns to

relate the 540 features with the 8 targets, i.e. 4 *x*-coordinates and 4 *y*-coordinates. The optimal neural network model with best generalization performance has two hidden layers with 200 and 100 nodes, respectively. The training of neural network was on 14000 samples for 800 epochs with early stopping to reduce computational cost and overfitting. The batch size is set as 100 samples. Dropout layer is added to avoid overfitting by turning off a certain user-defined percentage of neuron units in each layer for every training batch. Twenty percent of training dataset, 2800 samples, is used as validation set to allow an unbiased estimate of the performance of the model during the training to identify early stopping that prevents overfitting. Unlike the validation dataset, the test dataset helps evaluate the generalization performance of the fully trained and optimized model prior to the deployment on new dataset.

The performance of the neural network model on testing data is summarized in **Figure 5**. Notably, the proposed model can precisely estimate the location, orientation, and length of the three cracks as the crack grows. In **Figure 5(a)**, the red line segments represent the known crack path as the crack propagated from stage 1 to stage 3, while the blue line segments are the neural network model estimations based on the data-driven processing of the multipoint waveform data. The known and the predicted crack paths show a striking similarity. The performance of the neural network model is evaluated using root mean squared error (RMSE), which is the standard deviation of the residuals (prediction errors). Compared to mean absolute error, RMSE is preferred when large errors are particularly undesirable. **Figure 5(b)** is boxplot displaying the distribution of RMSE of the predictions obtained using the neural-network regressor. The boxplot includes five important vertical lines: minimum, first quartile (Q1), median, third quartile (Q3), and maximum. In the boxplot, the outliers are shown as individual points. It is apparent from boxplots that 75% of errors in the neural-network predictions are lower than 2.5mm. The other three prediction models also perform well with the maximum error of less than 10mm. In the end, the RMSE distribution of the neural-network predictions is shown in **Figure 5(c)**. All in all, the neural-network regressor performs the best both in terms of accuracy and computational cost.

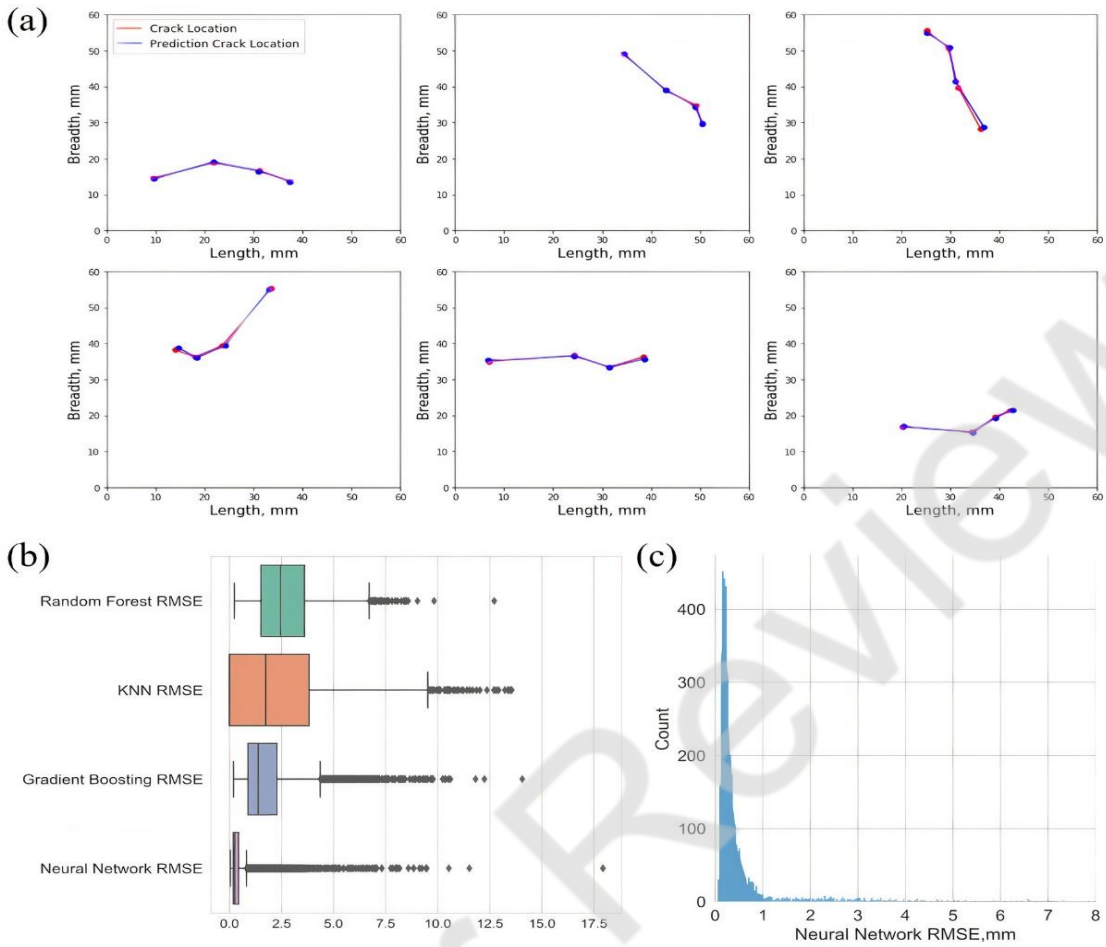


Figure 5. (a) Data-driven monitoring of crack propagation from initial to final stages using the neural-network based regressor to process the multipoint waveforms recorded by the 9 sensors. (b) Box plot of root mean squared errors for various regression methods for the task of monitoring the crack propagation. (c) Distribution of RMSE of the neural-network based predictions of crack path when monitoring the crack propagation for the 4000 testing samples.

4.4. Sensitivity to Noise

In signal processing, white noise is random noise with the same spectral density in all frequency bands. Gaussian white noise means the noise signal follows a normal distribution with mean μ and variance σ, N (μ, σ). In this work, the elastic waveform recorded by each sensor is corrupted with Gaussian white noise of different variances and zero mean. The noise is not directly added to the feature set, but to the raw waveforms (include train, validation, and test sets). The resulting waveform (e.g. stage 3) with Gaussian noise N(0,10), N(0,50) and N(0,100) are depicted in **Figure 6(a)**. Furthermore, to evaluate the effects of noise on regression model performance, the accuracy of RF, KNN, GB and NN (trained on noise-bearing training and validation data) are quantified using R2 metric when applied to the noise-bearing testing data,

as shown in **Figure 6(b)**. Neural Network and KNN regressors are relatively robust to Gaussian noise, whereas Gradient boosting method is the most susceptible to noise. An increase in the variance of Gaussian noise beyond 50 adversely affects the performances of all regressors. Overall, the four regressors exhibit accuracies higher than 0.82 under high noise conditions.

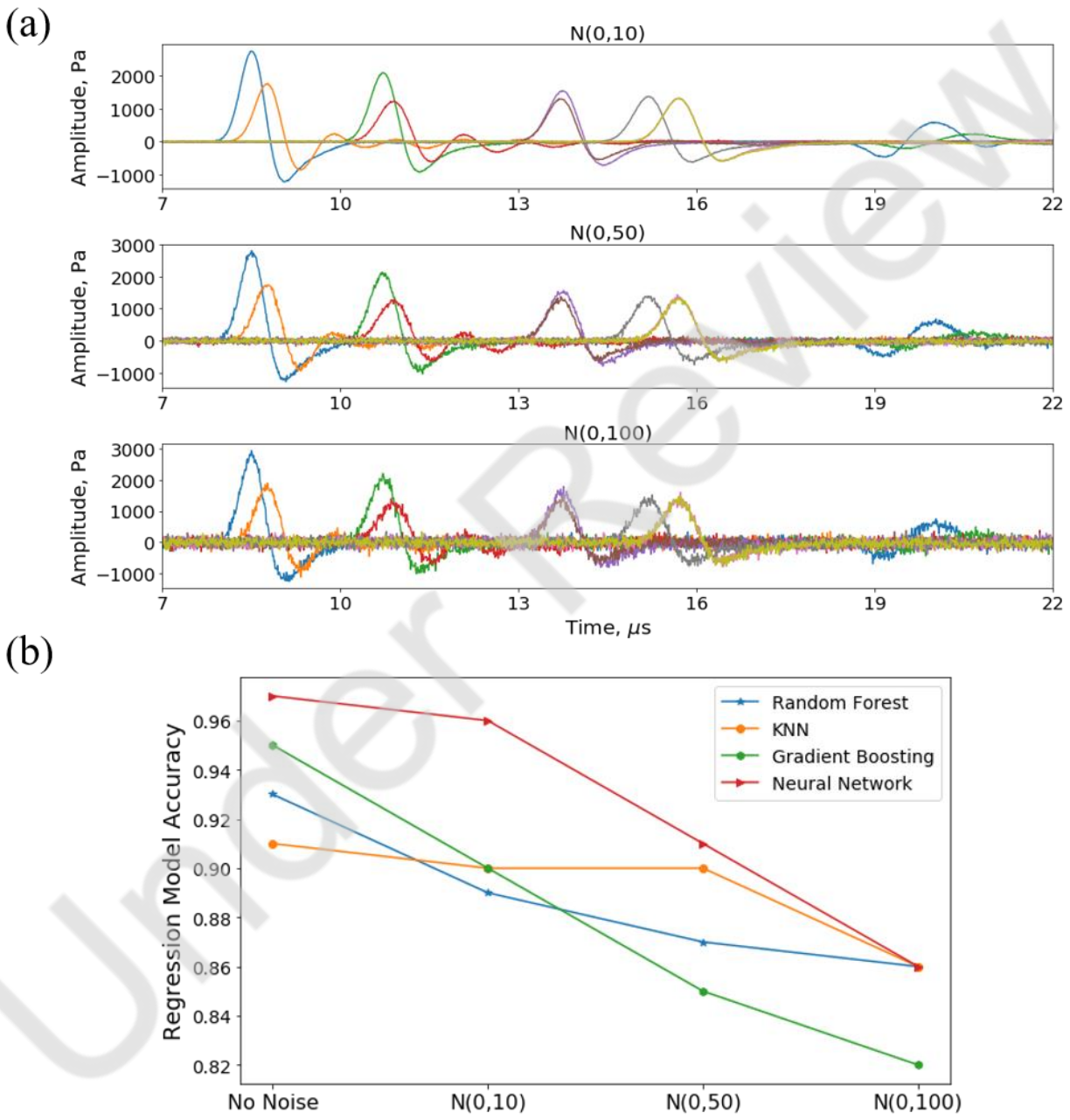


Figure 6. Sensitivity of the data-driven workflow to the noise when monitoring the crack growth. (a) Waveforms recorded by the 9 sensors at the final stage of crack growth containing Gaussian white noise of various variances: N(0,10), N(0,50), and N(0,100), respectively. (b) Change in the accuracy of the regression model for various variances of Gaussian white noise in the recorded waveforms.

5. Data-Driven Causal Discovery

The following section focusses on the use of various data-driven causal discovery techniques on the multipoint waveform data to identify and quantify the causal relationships between the propagation of crack (the cause) and the resulting changes in the multipoint waveforms recorded by the 9 sensors (the effect). The data contains the information about the crack propagation from stage 1 to stage 3 and the nine 25-μs waveforms captured by the 9 sensors due to an impulse source for the 20000 2D material samples. The primary objective is to identify causal signatures in the multipoint waveform measurements that arise due to the crack propagation. Intrinsically, causality facilitates robust predictions and data-driven modeling. Causal relationships are crucial for building explainable machine learning models. Causality reveals the underlying mechanisms of a process or a system. However, the current evaluation of data-driven models and machine learning algorithms are primarily focused on the statistical correlation between the features and targets rather than potential causality and causal signatures.

5.1. Cause and effect

In the eighteenth century, philosopher David Hume pointed out the conceptions of causality in terms of repeated "conjunctions" of events (Hume, 1751). Hume claimed that the labeling of causality relies on the empirical regularities involving previous phenomena (Eagleman & Holcombe, 2002). In other words, we cannot confirm the inevitability between cause and effect, instead we can only understand the repeated connection between cause and effect through observations. Judea Pearl (Pearl, 1998; Pearl, 2009) provided a comprehensive causation study with significant applications in the fields of statistics, artificial intelligence, economics, cognitive science and health science. In fact, there are several well-established and operational causal models in many fields that accurately reflect our intuitive understanding of cause and effect and can be described in precise mathematical terms.

In the simplest terms, a cause and effect relationship (or causality) is between two events, where an occurrence of an event (e.g. change in a property) causes an occurrence of another event, such that the cause precedes the effect. It is important to note that the effect may have multiple causes in the past, the cause can create multiple effects in future, and an effect may be the cause for multiple effects in the future. The essence of cause-effect relationship is the generation or determination of one event by another event. Donald Rubin proposed a potential outcome framework, also called Rubin causal model to understand causal mechanisms (Robin, 2005). The inability to simultaneously know the potential outcome/effect and observed outcome/effect is the fundamental problem of causal discovery. To overcome this problem, we need to design a treatment group and a control group to compare the differences in outcomes with and without the cause. In general, treatments are applied to the samples in the treatment group, while no

treatments are applied to the samples in the control group. The treatment serves as the cause. The presence and absence of treatment causes varying effects/outcomes.

In our study, there are 20000 2D material samples. Each sample undergoes crack propagation from stage 1 to stage 3. Samples undergoing a total change in length greater than 8 mm during the crack propagation from stage 1 to 3 are used in our study on data-driven causal discovery. The state of these samples at stage 1 along with the multipoint waveform measurements constitute the control group. The state of the sample at stage 3 along with the multipoint waveform measurements constitute the treatment group. The treatment is the extension/propagation of crack to a length greater than 8mm. The treatment causes changes in the multipoint waveform measurements. We quantify the causality between crack propagation and the changes in the multipoint waveforms using the average treatment effect (ATE), which measures the average difference between the outcomes for treatment group and the outcomes for the control group. The outcomes of the control group and the treatment group are the multipoint waveforms measured for each sample at stage 1 or stage 3 of the crack propagation, respectively. ATE enables the quantification of causal relationships when performing the data-driven causal discovery. It is expressed as:

$$\mathrm{ATE} = \frac{\sum_{i=1}^{n}(\mathrm{Y}_1(i) - Y_0(i))}{n}$$

where n is the number of samples in the treatment or control group, i is an individual sample, Y_0 is the outcome for a sample from the control group (before the treatment), and Y_1 is the outcome for a sample from the treatment group (after the treatment). The outcomes of the control group and the treatment group are the multipoint waveforms measured for each sample at stage 1 or stage 3 of the crack propagation, respectively. In our study, we use ATE to quantify the causality. Another metric, average treatment effect on the treated, ATT, quantifies the causal effect of the treatment for individuals in the treatment group. For example, ATT tells us what is the effect with different drug treatment, while ATE tells us what the effects of people taking or not taking the drugs are. Medical studies typically use ATT because they often only care about the causal effect of administering a drug for various patients. In reality, we do not always have both the control and treatment groups. For example, when a person takes the medication, then the result of that person not taking the medication does not exist. Therefore, in non-experimental studies, causal discovery is fraught with uncertainties.

5.2. Methods for Data-Driven Causal Discovery

In recent years, there has been several studies to learn the connection between machine learning and causality (Scholkopf, 2019; Guo et al., 2020; Moraffah et al., 2020). Although many toolboxes have provided diverse frameworks or methods in the field of causal inference, most of them lack stable validation and implementations. **Table 1** presents an overview of well-documented computational tools for causal

inference in Python. Each library has its advantages and disadvantages as well as the scope of application. Causal discovery, CausalNex and DoWhy are able to provide graphs through analysis to describe and visualize causality of the dataset.

Table 1: Python toolbox for causal analysis with pros and cons

Library	Description	Advantages	Disadvantages	License
Causal Discovery	Causal Discovery Toolbox (CDT) is an end-to-end package aimed at learning causal graphs from observations.	• Unify pairwise and score-based multi-variate approaches within a single package.	• Hard to validate causal graph. • Does not take advantages of machine learning.	MIT License
CausalNex	CausalNex uses Bayesian Networks to uncover structural relationships in data by combining machine learning and domain expertise.	• Able to visualize the causality. • Able to adjust structural model easily.	• Hard to validate the structure graph.	QuantumBlack
EconML	EconML applies the beauty of machine learning algorithms to measure the causal effects in observational or experimental data.	• Algorithms improve the computational cost of causal analysis.	• No causal graph.	Microsoft
DoWhy	DoWhy is a causal inference library that focuses on modeling causal assumptions and validating them.	• Effective even with unobservable confounders. • Combine features from EconML. • Focus causal inference on identification and frees up estimation using any available statistical estimator. • Automated robust check on the obtained estimate.	• Reliable assumption needed.	Microsoft
Causal ML	Causal ML provides tree-based and meta-based algorithms to estimate causal impact of intervention on outcome.	• Stable and allow to estimate CATE and ITE.	• Fewer functions. • No causal graph.	Uber Technologies

Our study aims to identify specific signatures in the recorded multipoint waveforms that have strong causal relationships with crack propagation. Here, we use DoWhy toolkit to evaluate the causal relationship between the crack propagation and the signatures in the multipoint waveforms. For our dataset, all samples undergoing a crack propagation greater than 8mm in length are considered for the data-driven causal discovery. The multipoint waveforms measured for samples in the first stage and final stage of crack propagation constitute the control group and the treatment group, respectively. We extract several statistical-parameterization based features from the multipoint waveforms (listed in Appendix A) and scale them to constitute the effects for a sample in the control or treatment group. The differences in these scaled statistical-parameterization based features between the control group and the treatment group are a result of the crack propagation. Pre-processing of the data is required to ensure that all the features in the control and treatment groups have the same range. This ensures that the ATE values, which quantify the causality corresponding to various effects, are comparable between the 20 highest-causality features (elaborated in the Appendix A).

In this study, the treatment effect quantifies the effect of crack propagation from Stage 1 to Stage 3 on the multipoint waveforms measured by the 9 samples. Treatment effect can be measured in four steps: **(1) model, (2) identify, (3) estimate, and (4) refute.** In step one, DoWhy creates an underlying causal graphical model for a given problem including confounders and instrumental variables with assumptions. Confounding variables (or confounders) are often defined as extraneous/irrelevant variables whose presence affect the variables under study, leading to an erroneous obscuring or emphasizing of their relationship (MacKinnon, et al., 2000). For instance, temperature confounds (blurs or confuses) the positive correlation between ice cream consumption and the number of sunburns. You may find a cause-and-effect relationship that does not actually exist between ice cream sales and sunburns because the effect is caused by the confounding variable: high temperature. As a result, any observed correlation between treatment (cause) and outcome (effect) may be only due to the existence of confounding variables and not due to a cause-and-effect relationship. Since confounders obscure the real effect of a treatment on the outcome, the confounders need to be eliminated as much as possible (Jager et al., 2008). There are various ways to exclude or limit the influence of confounders, which include randomization, restriction and matching (Pourhoseinghoil et al., 2012). Besides confounder, instrumental is another group of special variables that may mislead the discovery of cause and effect. They are factors that influence the cause but do not directly affect the outcome. Instrumental variables affect the cause, which subsequently leads to variations in outcome.

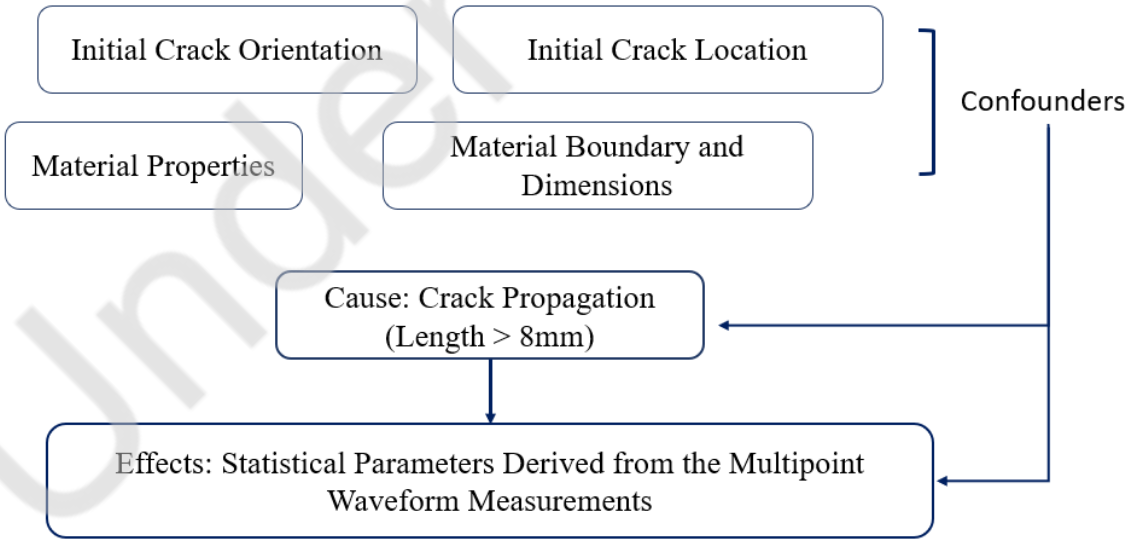


Figure 7. Simplified causal graph analysis of our dataset reveals certain confounders that influence the effects of crack propagation on the statistical-parameterization-based features extracted from the multipoint waveforms.

As we discussed, model inference problem using assumptions is the first step with toolbox DoWhy. **Figure 7** lays out the causal graph in our case, where the cause is the crack propagation and effects are statistical parameters change derived from the multipoint waveform measurements. Other variables such as initial crack location, initial crack orientation, sample dimension, material boundaries, and material properties are known confounders that influence both the cause and the effects. Once a user defines possible causes and effects, DoWhy tests the rest of variables in the dataset as potential confounders. Next, in the second step, supported identification criteria (back-door criterion, front-door criterion, instrumental variables, and mediation) is used to recognize the cause and effects based on the given model. Identification can be achieved with graph-based criteria and do-calculus without access to the data. Then, in the third step, the target estimand (ATE/ATT) identified in previous step is computed using statistical estimators. To model non-linear data, the DoWhy tool box also provides machine learning-based methods like gradient boosting tree to learn the relationships between the outcome and confounders, and the treatment and confounders, and then finally compares the residual variation between the outcome and treatment (Sharma & Kiciman, 2019). Finally, in the fourth step, DoWhy offers multiple refutation methods to check the robustness of the causal estimate. The common refutation methods checks the sensitivity of the causal estimate when the true treatment variables are replaced with a simulated random dataset similar to the treatment variables. The effect of the treatment should go to zero when a true treatment is replaced by an independent random variable.

5.3. Causal Signatures due to Crack Propagation

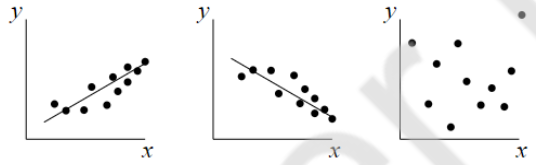
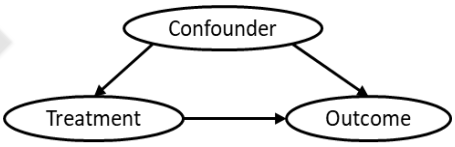
The most important part of this work is illustrating whether the changes in our statistical features derived from the multipoint waveform measurements stems from the crack propagation. Each statistical feature has 9 corresponding ATE calculated for the 9 waveforms recorded by the 9 sensors placed on the surface of the material. The detail ATEs are displayed as bar plots in Appendix B. Any feature that fails the refutation test is assigned a value of zero, which means no causation found. The causal effect can be positive or negative. Positive causation means the change in the effect variable is in the same direction as the change in the cause. By calculating the average of the absolute ATEs for statistical features, it is possible to compare their causal intensity and to determine the causal signatures of crack propagation.

The data-driven causal discovery reveals that the number of zero-crossing, negative-turning, and positive turning are the top three causal signatures of crack propagation. Relevance of these causal signatures are quantified in the Appendix B. A zero-crossing occurs when the sign of the signal changes as the signal oscillates in time. Positive/negative turning is the number of positive/negative turning points for the entire 25-μs waveform. Here, we report the impact of crack propagation on the elastic waveform, which

does not significantly affect the wave shape or crest, but triggers unstable fluctuations near the zero point. That abnormal fluctuations in the waveform sign the presence of the crack growth in materials.

In addition, significant changes in higher order moments of the waveforms, such as variance, skewness and kurtosis, are also caused by crack propagation in the material. This is evident from the bar plots in Appendix B. It was originally thought that the times of arrival of the highest peak and maximum dip in the waveform, as well as their amplitude, would change dramatically due to the effects of crack growth. The amplitude versus time (0 to 25 μs) information in the waveform has been expressed as amplitude versus index (0 to 2500). So, the time of arrival has been quantified in terms of the index. The causality of the index of maximum amplitude (peak) and the index of the minimum amplitude (dip) were negligible. In terms of sensors, the signal from sensor 8, located on the opposite side of the source, has the highest causality with the crack propagation. One unanticipated finding was that the sensor 2 and sensor 1, located very close to the source, have higher causality as compared to sensor 7 and sensor 9 placed next to sensor 8. The sensors 7 and 9 exhibited higher correlation but lower causality as compared to the sensors 1 and 2. This emphasizes the need for data-driven causal discovery.

Table 2: Overview of differences between correlation and causation

Correlation	**Causation**
Statistical and mathematical formulation	Mechanistic relationship
	
x and y can be interchanged	Cause and effect cannot be interchanged
• Linear Correlation • Non-linear correlation	• A causes B (direct causation) • B causes A (reverse causation) • A and B are both caused by C • A causes B and B causes A • No connection between A and B
• Correlation does not imply causation • Causation requires proof that no confounder variable is contributing to false causality	

5.4. Causality versus correlation

Correlation and causality can be deceptively similar and can simultaneously exist. However, correlation is not the same as causation (Duesberg, 1989; Conn, 2017). It is easier to find correlations compared to proving causality. Correlations can be easily quantified in terms of statistical or mathematical formulations that is applied on the data. As compared to correlation, a quantification of causality allows us to make better predictions about the future, explain the past, and intervene to change the outcomes. The differences

between causation and correlation are summarized in **Table 2**. A correlation between x and y measures association and can be interchanged between the two. When x is correlated to y, y is automatically correlated to x. On the contrary, when "A causes B", we cannot say that "B causes A".

Most commonly used correlation coefficients measure the linear (Pearson) or monotonic (Spearman) relationships. Mutual Information (MI) between random variables is considered more general than correlation coefficients and handles nonlinear dependencies. The concept of mutual information is directly linked to the entropy of random variables. Specifically, it measures the average amount of information in a random variable about another random variable (Paninski, 2003). In practice, MI is zero when and only when the two random variables are strictly independent (Kraskov et al., 2004). A higher MI represents a high dependency. Average MI between each feature and eight targets can be computed with sklearn in python. The eight targets represent the crack location, orientation, and length as the crack propagates from stage 1 to stage 3. The average MI values for each feature with the 8 targets are summarized in **Figure 8**, where we zoom into the top and bottom ten features based on the mutual information ranking. A feature having higher MI value has stronger association with the 8 targets to be predicted for the robust monitoring. The features have been labeled to represent the statistical parameter and the sensor number. Bar plot ranking reveals that variance, standard deviation and energy from sensor 7, 8, 9 are highly correlated with the targets. Median of waveforms are least correlated with the crack propagation.

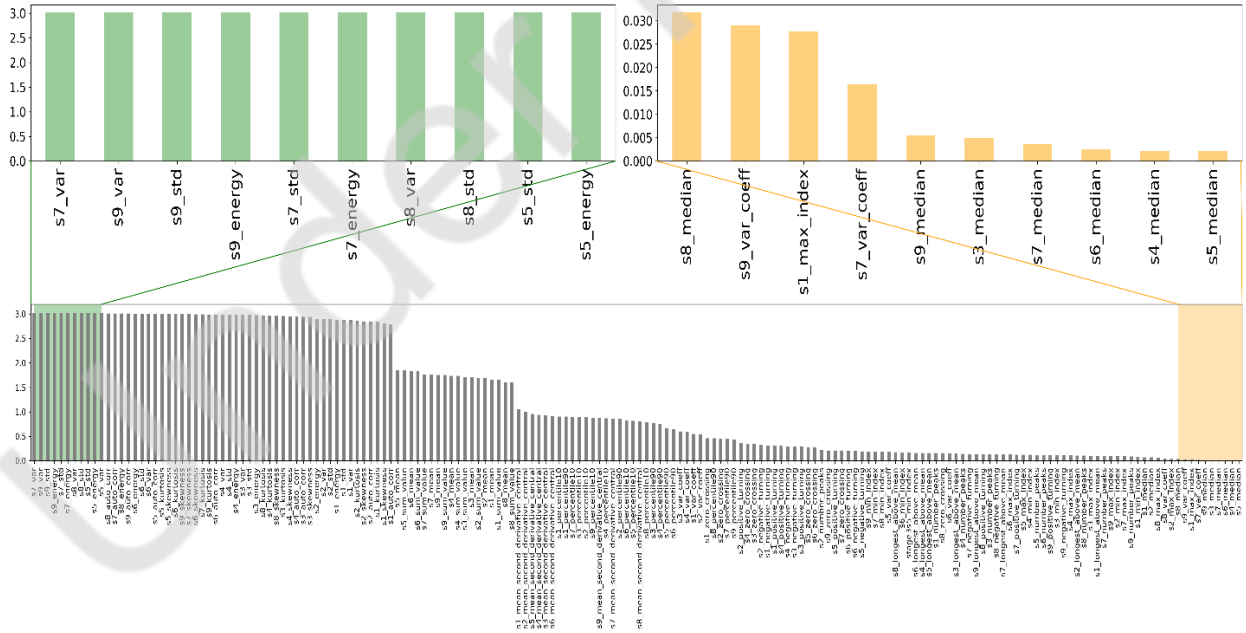


Figure 8. Mutual information ranking for all the statistical-parameterization based features derived from the multipoint waveforms measurements. Feature label indicates the sensor number and statistical parameter (e.g. s7_var refers to the variance of sensor 7 signal).

Following that, we compare the correlation quantified using mutual information with the causality. **Table 3** subdivides the features with high/low mutual information (correlation with crack propagation) into high/low causality. Notably, the feature S5_var is highly correlated with the crack propagation but does not have high causality due to crack propagation. 9 of the 10 features that were highly nonlinearly correlated with the crack propagation have high causality due to crack propagation. Further, all features that are poorly correlated with the crack propagation exhibit low or negligible causality. In this study, correlation may indicate an underlying potential causation, but further research is needed to establish it.

Table 3: Comparison of correlation against causality of top 10 and the bottom 10 features in terms of the correlation of the feature with crack propagation quantified using mutual information

	High Causation (>0.2)	**Low Causation (<0.2)**
High Mutual Information (Non-linear)	S7_var, S9_var, S7_std, S7_energy, S9_std, S8_std, S8_var, S8_energy, S9_energy	S5_var
Low Mutual Information (Non-linear)	None	S3_median, S4_median, S5_median, S6_median, S7_median, S8_median, S9_median, S1_max_index, S7_variation_co, S9_variation_co

5.5. Causality versus feature importance

Feature ranking is the process of sorting the importance of features for a supervised learning task. In our case, the feature importance reflects the significance of a feature for the task of robust monitoring of crack propagation. We could use the permutation-importance testing or the SHAP (Shapley Additive explanations) impact for computing the feature importance/ranking. This ranking informs which feature has the highest importance for the desired monitoring of the crack propagation pathway. Permutation-based feature ranking is defined as the reduction in model's prediction performance on the testing dataset when a specific feature is randomly shuffled (Breiman, 2001) while preserving the originality of all other features. When the permutation is repeated, the permutation importances tend to exhibit slight variations. SHAP or Shapley value is computed using a method from coalitional game theory (Shapely, 1971; Shapely, 1988). While the Shapley value was originally used to quantify the contribution of each player to a game, and later developed to quantify the contribution of each feature to the prediction of the model. SHAP values determine the importance of feature by comparing the results of different combinations of the feature. For example, for three features (A, B, C) in a dataset, 2^3 $(\emptyset; A; B; C; AB; AC; BC; ABC)$ possible combinations exist for predicting the target. Each different combination trains a unique prediction model. Thus, we will

have 2^3 models and corresponding predictions. Shapely value is the average marginal contribution of an instance of feature among all possible coalitions. To evaluate the contribution of feature A, the marginal contribution for subsets with/without feature A are calculated as weighted average:

$$Shap\ (A) = w1 * MC_{A,\{A\}} + w_2 * MC_{B,\{A,B\}} + w_3 * MC_{C,\{A,C\}} + w_4 * MC_{\{B,C\},\{A,B,C\}}$$

where MC is the marginal contribution measured by the difference of two outcomes with two subsets. w_1, w_2, w_3, w_4 are weights ($w_1 + w_2 + w_3 + w_4 = 1$).

SHAP Explainer can provide an explanation for many different ML algorithms such as tree-based models with TreeExplainer, linear models with LinearExplainer, and neural network models with KernalExplainer. The superiority of SHAP value over permutation feature importance is that tree-based SHAP is fast implementation with positive and negative impact. The idea behind SHAP feature ranking is simple that features with large absolute SHAP value are important. **Figure 9** shows the results obtained from the global or local importance of statistical features in testing dataset.

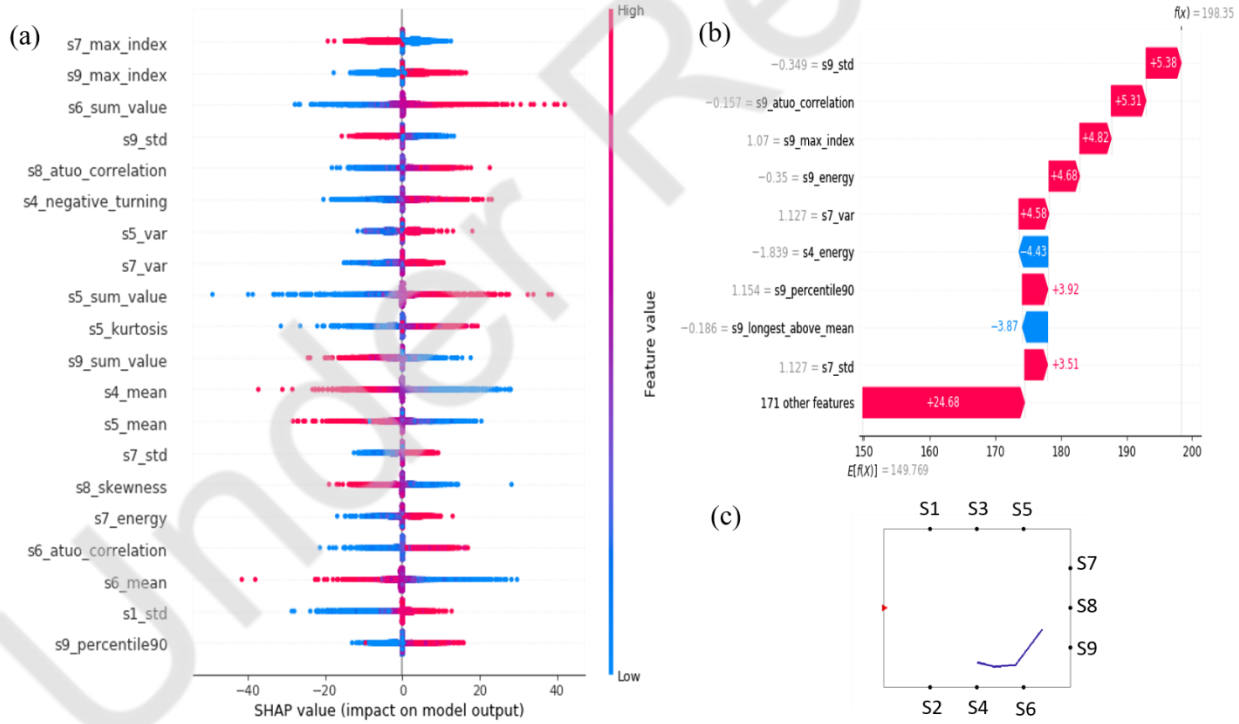


Figure 9. Feature importance with SHAP: (a) The summary plot combines feature importance with feature effects to show the global feature importance. (b) Waterfall plot with SHAP values for a random sample in testing dataset shows the local feature importance. (c) Diagram of locations for sensors 1 to 9.

Figure 9(a) sorts the global importance by the sum of SHAP value per feature across all samples in the testing set with the neural network-based explainer. The summary plot in Figure 9a combines feature

importance from the most important at the top to the least important at the bottom with feature effects. The top features in feature ranking contain significant useful information for prediction. It consists of a number of sample points with vertical positions showing the feature name, horizontal positions showing the impact on the model output, and color indicating the feature values from low (blue) to high (pink). Many overlapping points are jittered in y-axis direction hints at the distribution of SHAP value for each feature. In contrast to the mutual information based quantification of non-linear correlation, the top three features importance to the target became index of wave peak, sum of signal amplitudes, and standard deviation. From the sensor perspective, it is notable that S7, S8, S9 contain more information for monitoring the crack propagation in terms of both mutual information correlation and feature ranking. As the sensor locations are labeled in Figure 9(c), sensors 7 to 9 are the sensors located at the opposite boundary from the pressure source. Waterfall plot presented in Figure 9(b) gives the importance of feature based on a randomly selected sample. This ranking varies sample to sample as a local feature importance. Combination of local importances, e.g. shown in Figure 9b, generates the Figure 9a, which represents the global feature importance. **Table 4** shows that the features with high feature importance do not always exhibit high causality. In other words, several features critical for the regression-based monitoring of crack pathway do not exhibit high causality due to the crack propagation.

Table 4: Comparison of feature importance against causality of top 20 features based on global feature ranking on testing dataset

	High Causality (>0.2)	**Low Causality (<0.2)**
High Feature Ranking (ranking top 20)	S9_std, S8_auto_correlation, S4_negative_turning, S7_var, S5_kurtosis, S7_std, S8_skewness, S7_energy, S6_auto_correlation, S1_std	S9_max_index, S7_max_index, S6_sum, S5_var, S5_sum, S9_sum, S4_mean, S5_mean, S6_mean, S9_percentile90

Recently, there have been several studies about developing feature selection algorithms based on causality, since causal relationship suggests the underlying intrinsic nature of a problem. There is a growing body of literature from 2018 to 2021 that recognizes the importance of causality and discuss the need for causality-based feature selection (Ling et al., 2019; Wu et al., 2020; Wang et al., 2020). In our work, we observe that the features needed to accurately predict the crack propagation are not always causally related with the crack propagation. Not all highly correlated features exhibit high causality. Also, several high importance features for a prediction task exhibit low causality. Consequently, feature selection is a complex and time-consuming task that should involve consideration of different relationships between features and

objectives, rather than a single consideration. This study reminds that we cannot ignore the causal signatures and over-rely on correlation or importance based features when investigating the underlying mechanisms of a process or a system.

6. Conclusions

Supervised learning successfully monitored the propagation of mechanical discontinuity by processing multipoint sensor measurements of waveforms resulting from the interaction of a wave transmission with the mechanical discontinuity. The data-driven workflow can successfully monitor the location, orientation, and length of a mechanical discontinuity as it propagates through three stages. The monitoring of three stages of evolution of the mechanical discontinuity has a median RMSE lower than 2.5mm in material with dimension 60mm by 60mm. Neural network and gradient boosting based regression model outperforms all other regression techniques in the absence of noise. The presence of white noise with a standard deviation of 100Pa in the measured waveforms degrades the monitoring performance that still stays higher than 0.82, in terms of R-squared. Neural network model is relatively robust to the presence of Gaussian white noise, such that the monitoring performance is higher than 0.90 at noise variance of 10Pa and 50Pa. Regarding the feature extraction techniques, features obtained through the statistical parameterization and those obtained using principal component analysis lead to more robust monitoring. Topology based feature extraction, such as Isomap and Locally Linear Embedding, lead to low monitoring performance.

Causal discovery successfully identified signatures that exhibit strong causal relationship with the propagation of mechanical discontinuity. The newly developed causal discovery workflow provides a promising approach to evaluate true causal impacts with robustness check. The numbers of zero-crossing, negative-turning, and positive-turning in the waveforms are the strongest causal signatures of crack propagation. Higher order moments of the waveforms, such as variance, skewness and kurtosis, are also strong causal signatures of crack propagation. We observe that the features needed to accurately monitor the crack propagation are not always causally related with the crack propagation. Not all highly correlated features exhibit high causality. Also, several high importance features for monitoring task exhibit low causality. This study reminds that we cannot ignore the causal signatures and over-rely on correlation or importance based features when investigating the underlying mechanisms of a process or a system.

Acknowledgement

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Chemical Sciences, Geosciences, and Biosciences Division under the Award Number DE-SC0020675.

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley interdisciplinary reviews: computational statistics, 2(4), 433-459.

Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. Heliyon, 4(11), e00938.

Bhoumick, P., Sondergeld, C., & Rai, C. S. (2018, June). Mapping hydraulic fracture in pyrophyllite using shear wave. In 52nd US Rock Mechanics/Geomechanics Symposium. OnePetro.

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2015). Recent advances and emerging challenges of feature selection in the context of big data. Knowledge-based systems, 86, 33-45.

Bradford, A. (1965). Association or causation. Proc R Soc Med, 58, 295-300.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Chakravarty, A., Misra, S., & Rai, C. S. (2021). Visualization of hydraulic fracture using physics informed clustering to process ultrasonic shear waves. International Journal of Rock Mechanics and Mining Sciences, 137, 104568.

Chakravarty, A., & Misra, S. (2021). Unsupervised learning tracks spatiotemporal evolution of hydraulic fractures. https://www.essoar.org/doi/abs/10.1002/essoar.10506964.1

Conn, V. S. (2017). Don't rock the analytical boat: Correlation is not causation.

De Silva, V., & Tenenbaum, J. B. (2002, December). Global versus local methods in nonlinear dimensionality reduction. In NIPS (Vol. 15, pp. 705-712).

Duchene, P., Chaki, S., Ayadi, A., & Krawczak, P. (2018). A review of non-destructive techniques used for mechanical damage assessment in polymer composites. Journal of materials science, 53(11), 7915-7938.

Duesberg, P. H. (1989). Human immunodeficiency virus and acquired immunodeficiency syndrome: Correlation but not causation. Proceedings of the National Academy of Sciences, 86(3), 755-764.

Dwivedi, S. K., Vishwakarma, M., & Soni, A. (2018). Advances and researches on non destructive testing: A review. Materials Today: Proceedings, 5(2), 3690-3698.

Eagleman, D. M., & Holcombe, A. O. (2002). Causality and the perception of time. Trends in cognitive sciences, 6(8), 323-325.

Foster, J., Misra, S., FALOLA, Y., & Bhatia, M. (2021). Preemptive Detection of High Water-Cut Wells in Delaware Basin using a Joint Unsupervised and Supervised Learning Approach. https://www.essoar.org/doi/abs/10.1002/essoar.10507756.2

Gholizadeh, S. (2016). A review of non-destructive testing methods of composite materials. Procedia Structural Integrity, 1, 50-57.53, 1 – 37.

Guo, R., Cheng, L., Li, J., Hahn, P.R., & Liu, H. (2020). A Survey of Learning Causality with Data. ACM Computing Surveys (CSUR),

Hamada, G., & Joseph, V. (2020). Developed correlations between sound wave velocity and porosity, permeability and mechanical properties of sandstone core samples. Petroleum Research, 5(4), 326-338.

He, J., Li, H., & Misra, S. (2019). Data-driven in-situ sonic-log synthesis in shale reservoirs for geomechanical characterization. SPE Reservoir Evaluation & Engineering.

Holland, P. W. (1986). Statistics and causal inference. Journal of the American statistical Association, 81(396), 945-960.

Hume, D. (1751). Philosophical essays concerning human understanding (No. 35). M. Cooper.

Ibrahim, M. E. (2014). Nondestructive evaluation of thick-section composites and sandwich structures: A review. Composites Part A: Applied Science and Manufacturing, 64, 36-48.

ISRM, I. (1978). Suggested methods for the quantitative description of discontinuities in rock masses. Commission on the standardization of Laboratory and Field Tests in Rock Mechanics, ISRM.

Jager, K. J., Zoccali, C., Macleod, A., & Dekker, F. W. (2008). Confounding: what it is and how to deal with it. Kidney international, 73(3), 256-260.

Kraskov, A., Stögbauer, H., & Grassberger, P. (2004). Estimating mutual information. Physical review E, 69(6), 066138.

Li, H., Misra, S., & Liu, R. (2021). Characterization of mechanical discontinuities based on data-driven classification of compressional-wave travel times. International Journal of Rock Mechanics and Mining Sciences, 143, 104793.

Ling, Z., Yu, K., Wang, H., Liu, L., Ding, W., & Wu, X. (2019). BAMB: A Balanced Markov Blanket Discovery Approach to Feature Selection. *ACM Trans. Intell. Syst. Technol., 10*, 52:1-52:25.

Liu, R., & Misra, S. (2020). Data-Driven Classification of Materials with Open or Closed Mechanical Discontinuities Based on Multipoint, Multimodal Travel-Time Measurements.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. Prevention science, 1(4), 173-181.

Mast, T. D., Souriau, L. P., Liu, D. L., Tabei, M., Nachman, A. I., & Waag, R. C. (2001). A k-space method for large-scale models of wave propagation in tissue. IEEE transactions on ultrasonics, ferroelectrics, and frequency control, 48(2), 341-354.

Misra, S., & Li, H. (2019). Noninvasive fracture characterization based on the classification of sonic wave travel times. Machine Learning for Subsurface Characterization, 243.

Misra, S., Chakravarty, A., Bhoumick, P., & Rai, C. S. (2019). Unsupervised clustering methods for noninvasive characterization of fracture-induced geomechanical alterations. Machine Learning for Subsurface Characterization, 39.

Misra, S., & Wu, Y. (2019). Machine learning assisted segmentation of scanning electron microscopy images of organic-rich shales with feature extraction and feature ranking. Machine Learning for Subsurface Characterization, 289.

Misra, S., Li, H., & He, J. (2019). Machine learning for subsurface characterization. Gulf Professional Publishing.

Misra, S., & He, J. (2019). Stacked neural network architecture to model the multifrequency conductivity/permittivity responses of subsurface shale formations. Machine Learning for Subsurface Characterization, 103.

Misra, S., Li, H., & He, J. (2020). Noninvasive fracture characterization based on the classification of sonic wave travel times. In Machine Learning for Subsurface Characterization (pp. 243-287). Gulf Professional Publishing.

Moraffah, R., Karami, M., Guo, R., Raglin, A., & Liu, H. (2020). Causal Interpretability for Machine Learning - Problems, Methods and Evaluation. ACM SIGKDD Explorations Newsletter, 22, 18 - 33.

Muratov, M.V., Biryukov, V., & Petrov, I. (2020). Solution of the Fracture Detection Problem by Machine Learning Methods. Doklady Mathematics, 101, 169-171.

Osogba, O., Misra, S., & Xu, C. (2020). Machine learning workflow to predict multi-target subsurface signals for the exploration of hydrocarbon and water. Fuel, 278, 118357.

Paninski, L. (2003). Estimation of entropy and mutual information. Neural computation, 15(6), 1191-1253.

Pearl, J. (1998). Graphs, causality, and structural equation models. Sociological Methods & Research, 27(2), 226-284.

Pearl, J. (2009). Causality. Cambridge university press.

Pourhoseingholi, M. A., Baghestani, A. R., & Vahedi, M. (2012). How to control confounding effects by statistical analysis. Gastroenterology and hepatology from bed to bench, 5(2), 79.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. Journal of the American Statistical Association, 100(469), 322-331.

Scholkopf, B. (2019). Causality for Machine Learning.

Shapley, L. (1971). Cores of convex games. International Journal of Game Theory, 1, 11-26.

Shapley, L. (1988). A Value for n-person Games.

Sharma, A., & Kiciman, E. (2019). DoWhy: A Python package for causal inference. In KDD 2019 workshop.

Song, Y., Liang, J., Lu, J., & Zhao, X. (2017). An efficient instance selection algorithm for k nearest neighbor regression. Neurocomputing, 251, 26-34.

Treeby, B. E., & Cox, B. T. (2010). k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields. Journal of biomedical optics, 15(2), 021314.

Treeby, B. E., Jaros, J., Rohrbach, D., & Cox, B. T. (2014, September). Modelling elastic wave propagation using the k-wave matlab toolbox. In 2014 IEEE international ultrasonics symposium (pp. 146-149). IEEE.

Ventura, D. (2008). Manifold Learning Examples–PCA, LLE and ISOMAP. Department of Computer Science, Brigham Young University, Provo, UT, USA.

Wang, B., Zhong, S., Lee, T. L., Fancey, K. S., & Mi, J. (2020). Non-destructive testing and evaluation of composite materials/structures: A state-of-the-art review. Advances in mechanical engineering, 12(4), 1687814020913761.

Wu, X., Jiang, B., Yu, K., Miao, C., & Chen, H. (2020). Accurate Markov Boundary Discovery for Causal Feature Selection. *IEEE Transactions on Cybernetics, 50*, 4983-4996.

Wu, X., Jiang, B., Yu, K., Miao, C., & Chen, H. (2020). Accurate Markov Boundary Discovery for Causal Feature Selection. *IEEE Transactions on Cybernetics, 50*, 4983-4996.

Yu, K., Guo, X., Liu, L., Li, J., Wang, H., Ling, Z., & Wu, X. (2020). Causality-based feature selection: Methods and evaluations. ACM Computing Surveys (CSUR), 53(5), 1-36.

Appendix A: Feature Set Based on Statistical Parametrizations

1) Mean: the average pressure amplitude of the waveform.
2) Median: the median pressure amplitude of the waveform.
3) Variance: measures variability of pressure amplitude from the average or mean.
4) Standard deviation: measures the dispersion of dataset relative to its mean, the square root of variance.
5) Skewness parameter: calculate with the adjusted Fisher-Pearson standardized moment coefficient G1.
6) Kurtosis: calculated with the adjusted Fisher-Pearson standardized moment coefficient G2.
7) Energy: absolute energy of the time series which is the sum over the squared values.
8) Maximum-amplitude index of the signal: index of signal maximum point.
9) Minimum-amplitude index of the signal: index of signal minimum point.
10) 10th Percentiles: measures to indicate the pressure amplitude value below which lies 10% of observations.
11) 90th Percentiles: measures to indicate the pressure amplitude value below which lies 90% of observations.
12) Zero crossing: number of crossings of x on zero.
13) Number of peaks: number of peaks of at least support n in the time series x.
14) Longest period above mean: returns the length of the longest consecutive subsequence in x that is bigger than the mean of x.
15) Variance coefficient: returns the variation coefficient (standard error / mean, give relative value of variation around mean) of x.
16) Sum Value: the sum value over the time series values.
17) Autocorrelation: the autocorrelation of the specified lag.
18) Mean value of second derivation: returns the mean value of a central approximation of the second derivative.
19) Positive turning: number of positive turning points of the signal.
20) Negative turning: number of negative turning points of the signal.

Appendix B: Average Causal Estimates using the Average Treatment Effect

Figure B1: ATE bar plot for the 20 statistical-parameterization based features computed from the 9 waveforms captured by 9 sensors placed around the surface of the material. All zeros mean that no causal relationship was found between the feature and the target and the feature failed the refutation test. A positive ATE means the crack propagation has a positive impact on the feature. The group of high/low causality in Table 3 and Table 4 in the main text are drawn based on the ATE values presented in this plot. The numbers of zero-crossing, negative-turning, and positive in the waveforms are the strongest causal signatures of crack propagation. Higher order moments of the waveforms, such as variance, skewness and kurtosis, are also strong causal signatures of crack propagation. The causality of the time of arrival or the index of maximum amplitude (peak) and that of the minimum amplitude (dip) is negligible. In terms of sensors, the signal from sensor 8, located on the opposite side of the source, has the highest causality with the crack propagation. Sensors 1 and 2, located very close to the source, have higher causality as compared to sensors 7 and 9 placed next to sensor 8. The sensors 7 and 9 exhibited higher correlation but lower causality as compared to the sensors 1 and 2. This emphasizes the need for data-driven causal discovery.