

Damage detection in concrete structures with impedance data and machine learning

Asraar Anjum¹, Meftah Hrairi^{1*}, Abdul Aabid², Norfazrina Yatim¹, and Maisarah Ali³

¹ Department of Mechanical and Aerospace Engineering, Faculty of Engineering, International Islamic University Malaysia, P.O. Box 10, 50728, Kuala Lumpur, Malaysia

² Department of Engineering Management, College of Engineering, Prince Sultan University, PO BOX 66833, Riyadh 11586, Saudi Arabia

³ Department of Civil Engineering, Faculty of Engineering, International Islamic University Malaysia, P.O. Box 10, 50728, Kuala Lumpur, Malaysia

Abstract. This study aims to evaluate the effectiveness of machine learning (ML) models in predicting concrete damage using electromechanical impedance (EMI) data. From numerous experimental evidence, the damaged mortar sample with surface-mounted piezoelectric (PZT) material connected to the EMI response was assessed. This work involved the different ML models to identify the accurate model for concrete damage detection using EMI data. Each model has been evaluated with evaluation metrics with the prediction/true class and each class is classified into three levels for testing and trained data. Experimental findings indicate that as damage to the structure increases, the responsiveness of PZT decreases. Therefore, examined the ability of ML models trained on existing experimental data to predict concrete damage using the EMI data. The current work successfully identified the approximately close ML models for predicting damage detection in mortar samples. The proposed ML models not only streamline the identification of key input parameters with models but also offer cost-saving benefits by reducing the need for multiple trials in experiments. Lastly, the results demonstrate the capability of the model to produce precise predictions.

Keywords: Concrete structures; damage; piezoelectric material; Electromechanical Impedance; Machine Learning.

1. INTRODUCTION

Engineers enhance structural safety and reduce failure costs, focusing on structural health monitoring (SHM) and damage detection. SHM involves monitoring techniques like spaced dynamic response measurements and analysing sensitive deterioration indicators via statistics to estimate system health. As concrete structures face constant degradation, early damage detection becomes vital to maintain integrity and extend their lifespan. Concrete cracks are viewed as an early indicator of structural deterioration and long-term durability. Crack examination is a typical aspect of normal maintenance in most industrialised countries. Detecting fractures or cracks serves as an indicator of durability challenges, including issues like water seepage due to rebar corrosion. Typically, the existing traditional repair methods that involve filling the gaps with adhesive or cement-based materials are widely used. Consequently, it is crucial to conduct a thorough quality assessment to ensure the effectiveness of crack repairs and to validate the materials used for the repairs. The EMI technique offers a cost-effective solution for evaluating the performance of crack repair materials, due to its ability to be embedded and its dual sensing and actuation capabilities.

The EMI technique potential in assessing repair performance has not been explored despite its use for civil engineering damage identification. A method involving PZT

sensors and impedance analysis was devised to monitor crack-healing material performance. By computing damage indicators from frequency responses, the material effectiveness can be tracked over time, necessitating appropriate frequency intervals. Frequency ranges were explored for correlations between damage indicators and degree [1]. Throughout the repair process conducted by the authors, a notable distinction emerged between the results obtained after damage and those after repair. The root-mean-square deviation (RMSD) decreased during repair when compared to the pre-repair RMSD, indicating a partial restoration of material properties characterized by reduced damping [2]. An impedance analyzer was employed to assess impedance and admittance characteristics, considering various debonding scenarios between the reinforcement and concrete. Statistical metrics, namely RMSD and mean absolute percentage deviation, were applied to quantify alterations in impedance patterns recorded at the PZT patches caused by debonding conditions [3]. The findings indicate that the EMI-based method for SHM effectively identifies debonding damage in both fiber-reinforced polymer and rebar-reinforced concrete structures.

ML addresses engineering demands, particularly in civil engineering structures. ML, a subset of artificial intelligence (AI), boosts accuracy by interpreting data and fitting models. It has been found that blending field and laboratory data for

training ML models has the potential to address the common issue of over-predicting when models trained solely on laboratory data are used to predict the compressive strength of field-placed concrete [4]. The seismic performance of reinforced concrete frame structures with masonry infill panels is significantly influenced by their failure modes. The validation results show that both adaptive boosting and support vector ML models achieved maximum accuracy making them the top-performing models in terms of accuracy [5]. The ML methods were applied in research related to shear design. The suggested approach proves to be both effective and robust for designing shear in reinforced concrete beams, whether they have stirrups or not, thus paving the path for intelligent construction practices [6]. Performed a sensitivity analysis to evaluate the significance of 16 input variables and highlight their correlation with the shear capacity of steel fiber-reinforced concrete beams. The investigation indicated that factors such as web width, effective depth, and the clear depth ratio had the most pronounced impact on determining shear capacity using ML techniques [7].

The articles explore advanced machine learning applications in concrete structural analysis [8]. They delve into concrete strength development, damage identification, and bond strength prediction in reinforced concrete using methods like EMI, CNNs, and ensemble learning models [9]. These studies validate their approaches through experiments on structures like concrete cubes and tunnels, highlighting the potential of machine learning in enhancing concrete assessment, monitoring, and rehabilitation strategies [10]. Key focuses include interpreting raw EMI signatures, comparing machine learning model performances, and improving non-destructive evaluation methods for concrete structures [11].

A comparison study was conducted to see how well the suggested fully convolutional network-based methodology performs when compared to a SegNet-based method and they showed the proposed method surpasses the competition and is capable of detecting various tangible faults at the pixel level in real-world circumstances [12]. To evaluate the effectiveness of the suggested deep convolutional neural network approach, we examined a five-level benchmark building featuring adaptive smart isolators subjected to seismic forces. The outcomes indicate that the proposed technique exceeds conventional ML methods in both its ability to generalize and accurately detect structural damage. Hence, it is acknowledged as a viable and effective means for detecting damage in intelligent structures [13]. An extensive review study has been conducted on the repair and control of damaged structures using a PZT actuator which illustrates the impact of using PZT's on civil structural applications to sense the damage area [14] and that can also control through the actuation and many more recent technologies though the PZT can be seen in studies conducted recently [15].

According to existing literature, it is found that the using ML models in combination with EMI technology for damage detection, even when material properties change as damage

improvements, as this can impact the sensitivity analysis. A typical practical method for discovering concrete fractures with durability difficulties is to use an alternative to fill the gaps. Since it is difficult to determine the detecting quality, an effective approach is needed to assess the quality of detection monitoring. Combining in-depth training with EMI to assess the quality of detection may also make it simpler to provide an on-site detection monitoring system method using the ML model in cement structures. The literature highlights the widespread use of EMI for purposes such as damage detection, characterization, and concrete monitoring. However, based on the authors' knowledge, there is a noticeable gap in the literature concerning the integration of EMI data with ML models. This combination represents an unexplored opportunity for optimizing data-sensing capabilities. Consequently, in this study the primary objective is to detect concrete structure flaws by leveraging EMI data collected through PZT actuators and applying various ML models for predictive analysis. The current investigation is extended towards the identification of concrete cracks using EMI and the development of an SHM system that employs ML models for fracture detection in concrete, thus highlighting the importance of this work in addressing the literature gap.

The novelty of the research presented in this work can be articulated by emphasizing the innovative integration of EMI data with ML models to enhance damage detection in concrete structures. This unique combination has not been extensively explored in existing literature, offering a fresh perspective on SHM. The research further distinguishes itself by developing and evaluating machine learning models of varying complexity to establish the most accurate approach for this purpose. The practical implication of these models demonstrates cost-effectiveness and efficiency, marking a significant advancement in the field.

2. IMPEDANCE-BASED STRUCTURAL HEALTH MONITORING

Impedance-based techniques for SHM have emerged by employing the EMI and the properties of PZT materials [16]. This innovative approach constitutes a novel non-destructive testing evaluation method. Essentially, it revolves around the concept of tracking alterations in a structure mechanical impedance attributable to the presence of damage. Given the complexity of calculating a structure impedance directly, this method relies on the EMI of PZT materials to derive the system impedance. As damage influences the structure impedance, its occurrence induces changes in the EMI of PZT materials, whether they are affixed to or embedded within the structure [17].

When any physical changes affect the stiffness of a structure, it becomes possible to detect damage and loads by comparing them to the undamaged EMI signature of the structure. To enhance sensitivity in this approach, EMI measurements are conducted at high frequencies, where the excitation wavelength is small. This makes it feasible to detect even incipient damage that has a minimal impact on

the structure stiffness. Another advantage of using high frequencies is their low power requirement, typically in the micro-Watt range, for excitation. Eqn. (1) provides the solution to the wave equation when a PZT patch is bonded to the surface, and the electrical admittance is expressed as [18]:

$$Y(\omega) = i\omega a \left(\varepsilon_{33}^T (1 - i\delta) - \frac{Z_s(\omega)}{Z_s(\omega) + Z_a(\omega)} d_{3x}^2 Y_{xx}^E \right). \quad (1)$$

3. PROBLEM DEFINITION

In the current work, the prediction of damage detection was done using ML models from the experimental setup shown in Figure 1 that was utilised to capture the EMI response with THP51 PZT material type. Based on the experimental work conducted by Taha et al. [2], the sample was made with a single mortar beam of 100 mm in height and width and 500 mm in length. A 0.8-gram aluminium plate with the dimensions 15 mm, 10 mm, and 2 mm and a PZT patch on it were used. The aluminium plate was then fastened to the mortar beam surface at a 10 mm distance from its centre line. Utilized adhesive for bonding the PZT patch to the aluminium plate and attaching the aluminium plate to the beam. This approach aimed to mitigate the influence of potential local variations at the PZT-mortar interface, which might compromise the mechanical coupling efficiency.

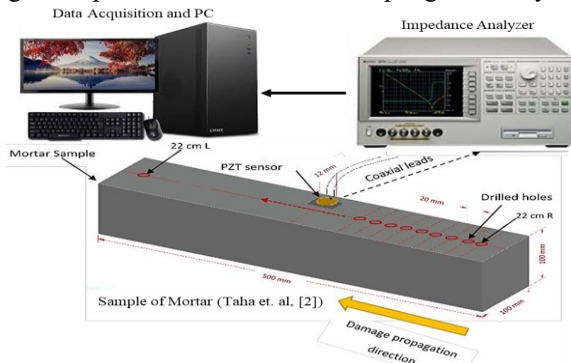


Fig.1. Setup for the electromechanical impedance technique (EMI) experiment.

The present problem has been defined based on the existing experimental work for prestige, stage 2 and stage 5 cases that were extracted with data and used ML models to enhance the energy-saving and cost-efficient approach. Figure 2 refers to

the current model which has three major stages experiment, models, and workflow. The aim is to achieve the optimal finding for concrete structure repair/control with PZT actuators and EMI technology. Figure 2 is named as the ML discovery model, and it defines the overall work of this current problem.

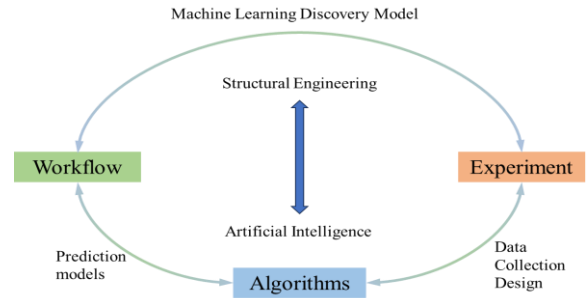


Fig.2. Diagram illustrating the overall configuration for Impedance data and ML work.

4. MACHINE LEARNING APPROACH

The soft computing approach has recently grown in favour of handling all kinds of engineering problems. In the past decade, ML has been used solve for the complex problems in engineering fields and has emerged as a crucial solution for both linear and non-linear issues. To solve structural difficulties in civil engineering problems, the present work is focused on ML approach (Figure 3).

A. Data collection.

The data collection for current work is from Taha et al. [2] experimental work for three stages, including the pristine stage, damage stage 2, and stage 5. As reported Table 1, there are three stages to damage detection that take frequency range vs. admittance into account. The trial-and-error method is used to find the frequency. The trial-and-error method is normally necessary for the EMI damage detection method when finding a sensitive frequency range to estimate the damage metrics after conducting several experiments. When the frequency is applied to the concrete specimen the impedance analyzer generates a signal or admittance.

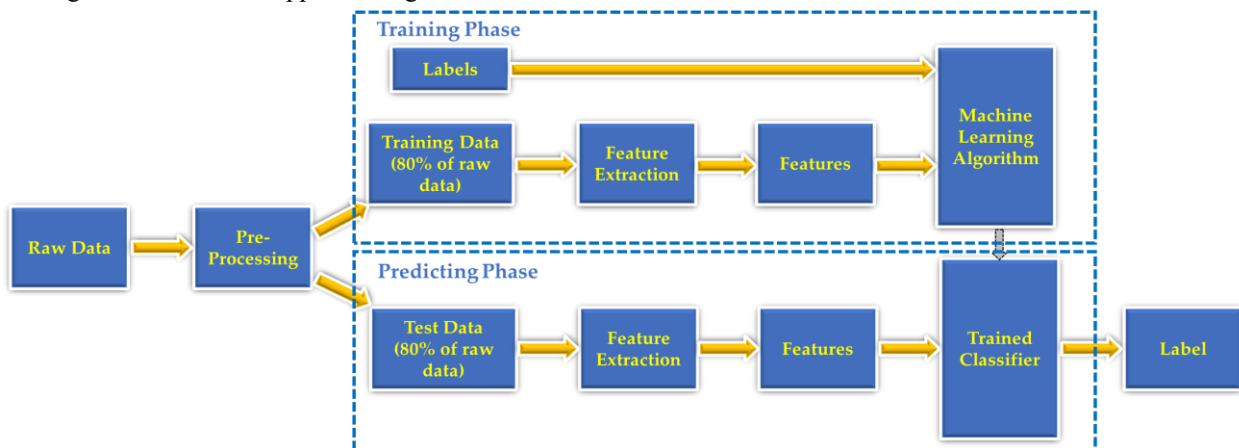


Fig.3. Machine learning process for the current application.

TABLE 1. Numerical analysis data set and their properties [2]

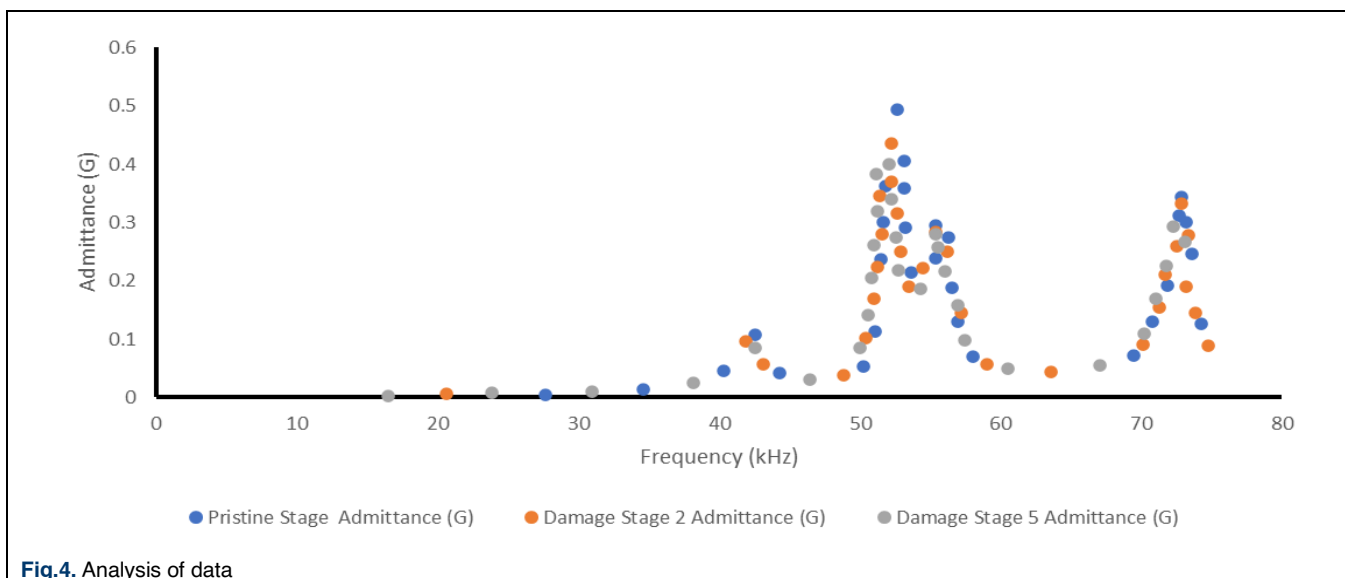
Pristine stage		Damage stage 2		Damage stage 5	
Frequency (kHz)	Admittance (G)	Frequency (kHz)	Admittance (G)	Frequency (kHz)	Admittance (G)
27.661	0.004264316	20.613	0.005411192	16.490	0.00280996
34.579	0.014507651	41.914	0.095644114	23.806	0.007996714
40.305	0.046249871	43.100	0.056423631	30.987	0.010649568
42.582	0.10831362	48.814	0.03879854	38.172	0.024694877
44.294	0.042519717	50.426	0.102116896	42.576	0.085528708
50.281	0.05401317	50.976	0.169214782	46.418	0.029897342

The signatures of frequency vs. admittance are extracted for the three different stages — the pristine stage, damage stage 2, and stage 5; To generate more data samples, seven different levels of gaussian noise- from 5dB to 35 dB with an increment of 5dB, is added to these admittance signatures using MATLAB. In every damage phase, 22 holes were employed on the beam surface. When using ML for prediction, only hole diameters of 5 mm and 10 mm were considered for damage stages 2 and 5, respectively, while referencing the pristine stag. Then, from this data extraction, the standard deviation (STD), skillfulness (SKW), lowest value, maximum value, mean value, and energy signal are determined. Finally, there are three sets listed for each damage stage: case 0 for the immaculate stage, case 1 for damage stage 2, and case 2 for damage stage 5. Each stage has eight admittance signatures, and from these signatures various parameters are obtained for training and testing the ML model.

B. Selection of models.

The primary objective of ML is the design and development of models that can recognize intricate patterns in experimental data [19]. ML models can be either descriptive or predictive or both [20]. When creating an ML model, the key design considerations revolve around the training experience, the specific function to be learned, the representation of this function, and the approach employed for learning from training data. These elements constitute the primary factors to be addressed in the model development. Depending on the training resources, ML is split into supervised, unsupervised, semi-supervised, and reinforcement learning categories.

When using supervised learning, a function (or model) is built using a training set of input instances and desired results to predict an unknown target output of future instances with accuracy. When want to predict continuously changing target values, the most suitable method is regression. When trying to predict discrete target values, this task is known as classification. Unsupervised learning involves starting with a training dataset containing input samples and then dividing these examples into clusters, where each cluster data shows a high degree of similarity. In this work, the beam which has been used for health monitoring has several holes and each hole has its effect on reducing the strength of the beam. Therefore, below illustrated models from ML have been chosen to predict the health monitoring conditions.



a. Support Vector Machine (SVM) classification.

Characterizing the typical feature space values or distribution of each class is the objective of parametric classification. SVM, in contrast, only considers the training samples that are located most closely to the ideal class border in the feature space. Finding the best border that maximizes the margin between the support vectors is the aim of SVM. The kernel trick is the term used to describe this projection to a higher dimension. There could be numerous different kernels, and each kernel might require a unique set of user-specified

parameters. According to the tests, the proposed damage detection approach has a 96.8% accuracy, which is much higher than that of an SVM. The findings indicate that the damage detection strategy outperforms the others in terms of damage detection. The suggested framework comprises three primary components: an initial data preprocessing module featuring data whitening, a sparse dimensionality reduction module aimed at reducing the dimensionality of the original input vector while retaining essential information, and a relationship learning module responsible for establishing a

connection between the compressed dimensional features and the structural stiffness reduction parameters [21].

b. Decision Tree (DT) Classification.

Decision trees (DTs) are among the most fundamentally simple classifiers. The leaf values of a regression tree represent a continuous variable, whereas the leaf values of a classification tree represent classes. The model rationale can be encapsulated through a collection of if-then rules. DTs can handle categorical data effectively, and once the model is constructed, classification tasks can be executed swiftly, as they do not necessitate complex mathematical computations. To mitigate the latter issue, tree pruning is commonly applied, involving the elimination of one or more split levels (branches). Pruning frequently enhances accuracy in dealing with unknowns while diminishing accuracy in detecting training data [22]. Therefore, this is one of the reasons on selecting DT for health monitoring purpose.

c. kNN classification.

The k-NN classifier is different because it does not make a model during training. Instead, it compares each new sample directly to the original training data [23]. The k-training samples that share the most similarities in feature space with the unknown sample are assembled into the dominant class. As a result, a low k will provide a very convoluted decision boundary, whereas a high k will produce better generality. Because a trained model has not yet been established, it is believed that more training samples are added. Mask R-structure kNNs have three stages that make them great for finding different types of damage in concrete (like cracks, efflorescence, rebar exposure, and spalling) [24].

C. Classification of model evaluation.

To evaluate the categorization model, many metrics are available [25]. In this study, the classification model was evaluated using the confusion matrix, sensitivity (recall), precision, F-score, and Matthew's correlation coefficient (MCC). Classification is the process of classifying a given collection of data into distinct categories in ML. The confusion matrix is used in ML to assess the effectiveness of a classification model. The confusion matrix is one way of summarising a classifier performance in binary classification situations (or error matrix). As indicated in Table 2, the projected values are positive (1) and negative (0), while the actual values are true (1) and false (0).

Accuracy is a measure of how many correct predictions there are compared to the total number of data points. The mathematical expression of accuracy is represented by:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN}. \quad (2)$$

Sensitivity (recall) gives the true positive rate and is the ratio of the number of accurate positive predictions to the total positive datasets. The mathematical expression of sensitivity is given by:

$$Sensitivity = \frac{TP}{TP+FN}. \quad (3)$$

Precision is the ratio of the number of accurate positive predictions to the total number of positive predictions. The mathematical expression of precision is given by:

$$Precision = \frac{TP}{TP+FP}. \quad (4)$$

F-Score is the measure of the accuracy of the model. It is determined based on the precision and reminders and is given by:

$$F - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \quad (5)$$

MCC is the correlation between predicted class and basic truth. MCC is often seen as a balanced metric that may be applied even when the classes have varied sizes. It is determined based on the values of the confusion matrix and is given by:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}. \quad (6)$$

TABLE 2. Classification of model evaluation

Class Designation		Actual Class	
		True (1)	False (0)
Predicted Class	Positive (1)	True Positive (TP)	False Positive (FP)
	Negative (0)	False Negative (FN)	True Negative (TN)

D. Analysis of Variance (ANOVA) Method.

The analysis of variance (ANOVA) approach is also used for investigation in the current work. The most frequent way is to use ANOVA to compare the average results of a single run or multiple runs. The primary goal is solely optimization. As a result, the orthogonal array inquiry was finished using ANOVA analysis. An ANOVA test was used to compare the mean square against treatment and errors at pre-set confidence levels to examine the effect of factors and their relationships. The ANOVA assists in determining the impact of each element on the overall variance of the results. The ANOVA table shows the effect of each factor, their error, and all potential factors which can be seen in the result section.

5. RESULTS AND DISCUSSION

In this work, several measures were implemented to prevent overfitting in the developed models. Firstly, used a robust cross-validation technique, dividing the data into training and validation sets to ensure that the models generalize well to unseen data. Additionally, regularization methods were employed to penalize model complexity. Also monitored the performance metrics on both training and validation datasets throughout the training process to detect any signs of overfitting. This cautious approach ensured that our models maintained high predictive accuracy without overfitting to the training data.

A. Detailing of data extraction.

The MATLAB program is used to implement the ML model for damage detection with classification approaches. The information was derived from a plot of frequency vs. admittance for the pristine stage, damage stage 2, and stage 5. The admittance for damage stage 5 is greater than for the pristine stage and damage stage. Whereas the admittance for damage stage 2 is higher than that of the pristine stage.

Presently, the admittance for the pristine stage is low as compared to the admittance for damage stage 2 and damage stage 5. From this, it is determined that the admittance is low for the pristine or healthy stage, however, for the damage stage, it is increased as per the applied frequency. Figure 4 shows the specifics of this comparison.

The pristine stage data can be changed for ML prediction for damage detection. This data, however, can be varied by adding 5 dB for each stage. This approach enhances the model ability to detect damage across a range of severities, making it more robust and applicable in practical scenarios. So, for a perfect stage, a total of eight data sets are used. The total number of steps for reading frequency and admittance for this task is 140. The frequency and admittance data were manually retrieved by selecting places in the origin for the pristine stage. Table 3 shows the details for pristine stage data.

TABLE 3. Pristine stage [2]

0dB	Frequency (kHz)	27.66	34.58	40.31	42.58	44.29
	Admittance (G)	0.00	0.01	0.05	0.11	0.04
5dB	Frequency (kHz)	27.96	35.61	39.03	43.07	44.47
	Admittance (G)	0.64	0.21	0.12	0.12	0.10
10dB	Frequency (kHz)	27.74	34.28	39.89	42.87	44.29
	Admittance (G)	0.04	0.19	0.42	0.16	0.30
15dB	Frequency (kHz)	27.48	34.53	40.23	42.51	44.47
	Admittance (G)	0.07	0.01	0.12	0.06	0.10
20dB	Frequency (kHz)	27.67	34.39	40.27	42.53	44.20
	Admittance (G)	0.02	0.02	0.02	0.19	0.03
30dB	Frequency (kHz)	27.66	34.57	40.33	42.60	44.31
	Admittance (G)	0.02	0.01	0.07	0.16	0.03
35dB	Frequency (kHz)	27.67	34.56	40.29	42.57	44.29
	Admittance (G)	0.02	0.01	0.02	0.11	0.01

The data for damage stage 2 is a type of plot between frequency and admittance for the pristine stage. Damage stage 2 data can be varied to anticipate ML damage detection. The damaging stage 2 data was gathered after specimen damage. In contrast, it was a healthy specimen when it was in its prime. Drilling equipment was utilized to induce stage 2 damage due to a 5 mm diameter hole on the surface of the concrete specimen at a specific location. This data, however, can be changed by adding 5 dB for each stage. As a result, for damage stage 2, a total of eight data sets are used. This assignment requires a total of 140 steps for frequency and admittance readings. Data for the frequency and admittance for damage stage 2 were manually extracted by selecting locations in the origin. Table 4 contains information on damage stage 2 data.

TABLE 4. Damage stage 2 [2]

0dB	Frequency (kHz)	20.61	41.91	43.10	48.81	50.43
	Admittance (G)	0.01	0.10	0.06	0.04	0.10
5dB	Frequency (kHz)	20.20	42.36	42.00	47.71	51.34
	Admittance (G)	0.52	1.35	0.68	0.47	0.06
10dB	Frequency (kHz)	20.91	41.98	43.22	48.92	51.28
	Admittance (G)	0.27	0.40	0.57	0.09	0.11
15dB	Frequency (kHz)	20.81	42.03	43.01	48.73	50.06
	Admittance (G)	0.35	0.35	0.14	0.05	0.17
20dB	Frequency (kHz)	20.79	41.66	43.05	49.06	50.38
	Admittance (G)	0.13	0.10	0.07	0.07	0.18
25dB	Frequency (kHz)	20.71	41.90	43.15	48.83	50.53
	Admittance (G)	0.09	0.15	0.08	0.01	0.10

30dB	Frequency (kHz)	20.61	41.88	43.10	48.81	50.50
	Admittance (G)	0.01	0.11	0.04	0.01	0.13
35dB	Frequency (kHz)	20.62	41.89	43.11	48.78	50.41
	Admittance (G)	0.02	0.10	0.07	0.07	0.11

Next, the data from damage stage 5 can be mosaicked for ML damage detection prediction. The damage stage 5 data was collected following specimen damage. It is comparable to damage stage 2 since both damage stages 2 and 5 include damage to the concrete specimen at several locations. Damage stage 5 was made by drilling a 10 mm diameter hole on the surface of the concrete specimen. This data, however, can be changed by adding 5 dB for each stage. As a result, for damage stage 5, a total of eight data sets are employed. The frequency and admittance readings for this work total 140 steps taken. Data for the frequency and admittance for damage stage 5 were manually acquired by picking locations in the origin. Table 5 contains information about damage to stage 5.

TABLE 5. Damage stage 5 [2]

0dB	Frequency (kHz)	16.49	23.81	30.99	38.17	42.58
	Admittance (G)	0.00	0.01	0.01	0.02	0.09
5dB	Frequency (kHz)	16.60	24.51	30.33	39.09	42.96
	Admittance (G)	0.96	0.27	0.02	0.40	0.49
10dB	Frequency (kHz)	16.39	23.88	30.89	37.64	42.39
	Admittance (G)	0.22	0.06	0.05	0.54	0.18
15dB	Frequency (kHz)	16.40	23.90	30.78	38.31	42.58
	Admittance (G)	0.07	0.16	0.01	0.03	0.27
20dB	Frequency (kHz)	16.42	23.68	30.93	38.12	42.53
	Admittance (G)	0.06	0.04	0.06	0.09	0.35
25dB	Frequency (kHz)	16.42	23.73	31.00	38.28	42.59
	Admittance (G)	0.03	0.01	0.02	0.01	0.12
30dB	Frequency (kHz)	16.47	23.77	30.93	38.16	42.59
	Admittance (G)	0.01	0.04	0.02	0.03	0.07
35dB	Frequency (kHz)	16.49	23.81	30.96	38.19	42.59
	Admittance (G)	0.00	0.02	0.01	0.04	0.06

Tables 4 and 5 show the detailing of the pristine stage, damage stage 2, and stage 5 concerning the signals, where count data is used for comparing means of three or more binary data.

B. Parameters of Conductance Results for ML Techniques

The parameters listed below are critical for determining the ML model forecast: skewness, standard deviation, minimum, maximum, and mean values, as well as energy signals for specific cases (Table 6). Whereas skewness measures the extent to which a distribution of a random variable differs from a perfectly symmetrical normal distribution on both sides. This skewness can be either to the left or to the right. Skewness risk arises when we apply a symmetric distribution to skewed data. On the other hand, the mean-squared value (MSV) or MS of a signal is its average squared value over a specific interval. Because a signal squared value is assumed to reflect its instantaneous power, the MSV is also known as the average power.

The term energy signal refers to a signal whose total energy is finite, and it is noted as 0 E. An energy signal average power P is also 0. Non-periodic signals are examples of energy signals. Following the acquisition of the data for

the pristine stage, damage stage 2, and stage 5 admittance and frequency, the data is then analyzed to determine the standard deviation (STD), skewness (SKW), minimum, maximum, mean, and energy signal. Finally, cases 0 for the pristine stage, 1 for damage stage 2, and 2 for damage stage 5 are indicated for each stage 8 set that is mentioned. The ML system may conclude the damage detection prediction utilizing these 8 groups of different cases.

TABLE 6. Extracted EMI data [2]

STD	SKW	Minimum	Maximum	Mean	Energy	Case
1.536	1.883	0.004	7.047	1.245	533.411	0
1.512	1.949	0.005	7.908	1.324	551.119	0
1.559	1.883	0.012	7.237	1.276	553.260	0
1.555	1.875	0.014	6.946	1.234	537.295	0
1.541	1.870	0.008	7.061	1.264	541.910	0
1.526	1.902	0.014	7.142	1.258	533.444	0
1.534	1.885	0.011	7.017	1.246	532.432	0
1.539	1.880	0.011	7.062	1.245	534.317	0
1.585	1.618	0.005	6.872	1.398	617.680	1
1.548	1.743	0.016	7.560	1.500	644.364	1
1.587	1.628	0.036	6.835	1.446	639.716	1
1.593	1.616	0.006	6.899	1.418	629.840	1
1.588	1.625	0.007	6.939	1.396	617.533	1
1.579	1.617	0.006	6.862	1.398	615.624	1
1.585	1.613	0.006	6.827	1.396	616.840	1
1.583	1.620	0.017	6.859	1.402	618.081	1
1.568	1.690	0.003	6.547	1.291	616.124	2
1.633	1.552	0.010	6.613	1.441	708.711	2
1.551	1.652	0.003	6.824	1.347	630.618	2
1.571	1.669	0.008	6.767	1.308	624.503	2
1.560	1.667	0.000	6.545	1.296	614.402	2
1.558	1.686	0.004	6.498	1.287	610.002	2
1.572	1.696	0.004	6.585	1.288	616.980	2
1.567	1.687	0.005	6.536	1.291	615.655	2

C. Prediction of results based on selected models

In practical applications, precisely locating damage in a deteriorated beam can be challenging. Therefore, the researchers employed a SHM system, integrating PZT transducers and EMI methods to assess structural health. The experimental findings led to the development of frequency charts that consider impedance and admittance, enabling the identification of a structure condition, as well as predicting damage locations through peak values. However, it remains uncertain which parameters are most effective in achieving optimal results in such scenarios. To address this, the study leveraged ML models to determine the most effective predictive approach for damaged civil concrete structures, choosing the models based on relevant prior work addressing similar challenges.

a. Decision Tree Model

Tables 7 and 8 depict the confusion matrix for the eight data points representing the pristine stage, damage stage 2, and stage 5 for the tree model for testing and training model respectively. These eight data points are based on experimental work for predicting damage detection in concrete specimens. Decision trees are also known as classification and regression trees that expect data responses. To forecast a reaction, follow the decisions in the tree from the root (beginning) node down to a leaf node. The answer can be found in the leaf node. Classification trees provide

nominal answers such as true or false. A decision tree model is a form of decision tree model. In the decision tree model, all data points are true for the pristine stage since there is no indication of damage in a beam (healthy structures), however for damage stage 2, seven data points are true and one is false because the beam is unhealthy and damaged between stages 2 and 3. The structures were discovered to be damaged as a result of these results, but the accuracy in terms of PZT still has to be enhanced by modifying any of the selected parameters. Finally, damage stage 5, where one is false and seven are true, indicates that the damage stage has damage that does not disclose the entirely true worth of the class. This reflects the frequency results of structures that have been damaged in some areas once again, and it can be termed unhealthy in stage 5 with one false result. Where "0" represents the pristine stage, "1" represents the damage stage 2 case, and "2" represents the damage stage 5 case.

TABLE 7. Confusion matrix of training data for Decision Tree Model

		Predicted Class		
		0	1	2
True Class	0	4	2	
	1	2	4	
	2	1	1	5

TABLE 8. Confusion matrix of test data for Decision Tree Model

		Predicted Class		
		0	1	2
True Class	0	2		
	1	1	1	
	2			1

b. Fine kNN Model

The confusion matrix for the eight data points is shown in Tables 9 and 10 for the fine kNN model for training and testing model. Based on the existing experimental data, these eight data points were examined. Where kNN is a form of supervised ML a kNN classification model is a nearest-neighbor classification model that allows you to change both the distance measure and the number of neighbors.

Because it keeps training data, a classification kNN classifier can be used to make resubstituting predictions. In the fine kNN model, all data points are true for the pristine stage, all data points are true for damage stage 2, one is false and seven are true for damage stage 5. In comparison to the fine tree model, the kNN model predicts both the pristine stage and damage stage 2 to be true, whereas in the fine tree model, the pristine stage is true for all data points, damage stage 2 is true for seven data points, and one is false, and for damage stage 5, both models are similar. Table 9 reveals that six data points are true, and one data point is false of the training data. Predict whether the location of specified stages has been damaged using the kNN model once more. Because the structures are thought to be damaged in a hidden area, and PZT can detect such locations using EMI, stage 5 displays true and false variations, whereas stage 2 displays just true values. It does not recognize the erroneous value in this

prediction. Meanwhile, it demonstrates that stage 2 has some variety in healthy structures, with one untruth because no indication of damaged structures was detected in stage 2, which may arise from PZT at this stage having a low impact.

TABLE 9. Confusion matrix of training data for Fine kNN Model

True Class	Predicted Class		
	0	1	2
0	6		
1		6	
2		1	6

TABLE 10. Confusion matrix of test data for Fine kNN Model

True Class	Predicted Class		
	0	1	2
0	2		
1		2	
2			1

c. Quadratic Support Vector Machine (SVM)

A quadratic support vector machine made the most recent prediction for damage in a concrete specimen and is suitable when data has precisely two categories. They work by identifying the ideal hyperplane that effectively separates all data points from one category and those from the other. The best hyperplane for an SVM is the one that maximizes the gap between the two categories. Tables 11 and 12 depict the true class vs. anticipated class for the eight data points for training and testing data. In a quadratic SVM model, all data points for the pristine stage are true, whereas all data points for the damage stage 2 are true, one is false and seven are true for the damage stage 5. In terms of true class vs. expected class, the SVM model prediction is extremely comparable to the Fine kNN for the pristine stage, damage stage 2, and stage 5. This SVM theoretical explanation for the damaged beam is comparable to the kNN model. These variations demonstrate that the ML models use a novel way to identify damage in damaged structures and can forecast where the damage will occur. Indeed, previous research demonstrated that the PZT detected the area of damage using the EMI technique, however, the time necessary to estimate the accuracy was longer. Doing the trials with the fewest number of tests, on the other hand, can be used in an ML strategy to produce the best findings, which can be a cost-effective and energy-saving approach.

TABLE 11. Confusion matrix of training data for Quadratic SVM

True Class	Predicted Class		
	0	1	2
0	6		
1		6	
2		1	6

Table 13 illustrates the various evaluation metrics for different classes using different methodologies. The table clearly shows that the kNN and SVM approaches outperform the DT model. This means that, depending on the models

chosen to anticipate the location of damaged structures in a beam, some have lower ideal results than others, whereas the current work is shown with three distinct models. It is also necessary to determine if the problem is linear or nonlinear; the techniques will be employed accordingly. Because of the large quantity of admittance and impedance value of the damaged structure, we considered the problem to be nonlinear in this study. Furthermore, the number of steps, frequency ranges, PZT dimensions, bonded materials, and many other parameters can influence the results; thus, a nonlinear model is assumed to be used to predict the results.

TABLE 12. Confusion matrix of test data for Quadratic SVM

True Class	Predicted Class		
	0	1	2
0	2		
1		2	
2			1

In comparing the applied methods, we assessed their performance based on several key parameters: sensitivity, precision, F-Score, and MCC. Each method was evaluated against these criteria, offering a comprehensive understanding of their strengths and weaknesses. To facilitate a direct comparison, we have included an evaluation table (Table 13 and 14) for training and testing data, which succinctly outlines these aspects. This section also discusses the practical implications of each method, considering factors like cost-effectiveness and scalability. By contrasting these methods with similar approaches in existing literature, we underscore the unique contributions and limitations of our approach, providing a balanced and nuanced perspective.

TABLE 13. Evaluation metrics of various models for training data

Model		Evaluation Metrics				
		Accuracy	Sensitivity	Precision	F-Score	MCC
DT	Class 0	0.7368	0.6667	0.5714	0.6154	0.4942
	Class 1	0.7368	0.6667	0.5714	0.6154	0.4942
	Class 2	0.8947	0.7143	1.0000	0.8333	0.7825
kNN	Class 0	1.0000	1.0000	1.0000	1.0000	1.0000
	Class 1	0.9474	1.0000	0.8571	0.9231	0.8895
	Class 2	0.9474	0.8571	1.0000	0.9231	0.8895
SVM	Class 0	1.0000	1.0000	1.0000	1.0000	1.0000
	Class 1	0.9474	1.0000	0.8571	0.9231	0.8895
	Class 2	0.9474	0.8571	1.0000	0.9231	0.8895

TABLE 14. Evaluation metrics of various models for test data

Model		Evaluation Metrics				
		Accuracy	Sensitivity	Precision	F-Score	MCC
DT	Class 0	0.8000	1.0000	0.6667	0.8000	0.6667
	Class 1	0.8000	0.5000	1.0000	0.6667	0.6124
	Class 2	1.0000	1.0000	1.0000	1.0000	1.0000
kNN	Class 0	1.0000	1.0000	1.0000	1.0000	1.0000
	Class 1	0.7143	1.0000	0.5000	0.6667	0.5477
	Class 2	1.0000	1.0000	1.0000	1.0000	1.0000
SVM	Class 0	1.0000	1.0000	1.0000	1.0000	1.0000
	Class 1	0.7143	1.0000	0.5000	0.6667	0.5477
	Class 2	1.0000	1.0000	1.0000	1.0000	1.0000

D. ANOVA Optimization

The ANOVA table for pristine stages at 0 dB, 5 dB, and 10 dB about frequency and admittance is provided below. Table 15 for damage stage 2 with frequency and admittance for 0 dB, 5 dB, and 10 dB was also extracted, and the last reading mentioned frequency and admittance regarding damage stage 5 for 0 dB, 5 dB, and 10 dB. We may determine the average value, sum of all group values, and variance which can be determined from the ANOVA analysis. Calculated the sum of squares treatment (SS) for splitting into two components using the above data. The degrees of freedom (DF) were then computed, and they represent the number of independent bits of information (DF).

TABLE 15 Frequency and Admittance for pristine stage, damage stage 2 and 5

Signals	Count	Sum	Average	Variance
<i>Pristine stage</i>				
Frequency (0 dB)	137	20252.22	147.8264	5639.118
Admittance (0 dB)	137	170.6066	1.245304	2.359955
Frequency (5 dB)	137	20257.98	147.8685	5624.312
Admittance (5 dB)	137	181.4086	1.324151	2.286075
Frequency (10 dB)	137	20249.06	147.8034	5642.219
Admittance (10 dB)	137	174.7679	1.275678	2.428768
<i>Damage stage 2</i>				
Frequency (0 dB)	137	19962.93	145.7148	4730.743
Admittance (0 dB)	137	191.5222	1.397971	2.567799
Frequency (5 dB)	137	19966.03	145.7375	4741.063
Admittance (5 dB)	137	205.4733	1.499805	2.450447
Frequency (10 dB)	137	19964.57	145.7268	4728.662
Admittance (10 dB)	137	198.0998	1.445984	2.582753
<i>Damage stage 5</i>				
Frequency (0 dB)	137	18221.66	133.0048	4042.476
Admittance (0 dB)	137	186.0916	1.358333	2.63499
Frequency (5 dB)	137	18218.99	132.9853	4037.76
Admittance (5 dB)	137	207.9385	1.517799	2.840424
Frequency (10 dB)	137	18218.05	132.9785	4042.155
Admittance (10 dB)	137	194.561	1.420153	2.559818

In an ANOVA study, the variance of the mean squares is calculated by dividing the sum of squares by the corresponding DF. The F value in the ANOVA test determines the P value, which shows the possibility of receiving a result at least as extreme as the one observed. In some cases, p values are given in ANOVA tables. The null hypothesis that a single source, model, or parameter is not significant is rejected more frequently as the p-value for a given ratio decreases. The F-critical value is: The F statistic is obtained because of an ANOVA test. It demonstrates the significance of variable groupings. The F critical value is also known as the F-statistic. All this derived information is required for MATLAB parameters. The calculated results are shown in Table 16.

TABLE 16 F-critical value

Source of Variation	SS	DF	MS	F	P-value	F crit
Treatment	12274161	17	722009.5	300.481	0	1.626945
Error	5882166	2448	2402.846			
Total	18156327	2465				

The study evaluated various ML models of increasing complexity, including fine tree models, KNN, and SVM. Previous studies have also used ML techniques for SHM, but the specific combination and comparison of these models may offer new insights. For example, research on the use of fully convolutional networks for concrete structure damage detection and the use of deep learning for structural damage identification have been conducted, showing the growing interest in applying advanced ML techniques in this field. The effectiveness of ML models in detecting concrete damage in the study is significant, with models demonstrating strong predictive accuracy. This aligns with other research findings where ML models have been shown to be effective in various civil engineering applications, such as predicting the compressive strength of concrete, shear capacity of beams, and structural damage detection.

The current approach to blending field and laboratory data for training ML models is an important aspect. Previous research has also highlighted the issue of over-predicting when models trained solely on laboratory data are applied in real-world scenarios. This approach of combining data sources could lead to more robust and accurate models. One of the highlighted benefits of the study is the cost-saving aspect due to reduced need for multiple trials in experiments. This aligns with the general trend in SHM research where the focus is on developing cost-effective and efficient methods for monitoring and maintaining structural integrity.

6. CONCLUSION

To improve damage detection methods, systematically trained and tested ML models. The goal was to identify the most effective model for detecting concrete damage using EMI data. For this, the performance of DT, KNN, and SVM models were examined. Current findings revealed that the fine tree model exhibited strong predictive accuracy for detecting damage in pristine or healthy concrete specimens. However, its performance deteriorated for damage stages 2 and 5. Next, the KNN model demonstrated accurate predictions for both the pristine stage and damage stage 2 but exhibited inconsistencies across all data points in the concrete specimen. In contrast, quadratic SVM proved highly accurate for damage detection in both pristine and damage stages but experienced a decline in performance at the later damage stage. ML prediction models, coupled with EMI analyzers, can effectively detect damage in concrete specimens. Researchers can leverage these insights to enhance their work in damage concrete – surface mounted PZT material with EMI and ML model predictions. Moreover, the study underscores the significance of ML models in developing innovative concepts, particularly in the context of distinct ML methods. In addition to ML, explored the utility of ANOVA optimization to evaluate the impact of treatment values on prestige and damage stages. This analysis identified the optimal frequency range, streamlining the process and reducing the need for extensive manual intervention.

This work acknowledges the limitations of the proposed models. While the current approach demonstrates significant advancements in damage detection using EMI data and selected ML constraints. These include dependency on the quality and quantity of data, potential overfitting, and generalizability to different types of concrete structures. Future work will focus on addressing these limitations by exploring more diverse datasets, implementing regularization techniques to prevent overfitting, and testing the models across various structural conditions. This ongoing research aims to enhance the robustness and applicability of current ML models in real-world scenarios.

REFERENCES

- [1] H. Kim, X. Liu, E. Ahn, M. Shin, S.W. Shin, and S.H. Sim, "Performance assessment method for crack repair in concrete using PZT-based electromechanical impedance technique," *NDT E Int.*, vol. 104, no. April, pp. 90–97, 2019, doi: 10.1016/j.ndteint.2019.04.004.
- [2] H. Taha, R.J. Ball, and K. Paine, "Sensing of Damage and Repair of Cement Mortar Using Electromechanical Impedance," *Materials (Basel)*, vol. 12, no. 23, p. 3925, 2019, doi: 10.3390/ma12233925.
- [3] W. Li, S. Fan, S.C. M. Ho, J. Wu, and G. Song, "Interfacial debonding detection in fiber-reinforced polymer rebar-reinforced concrete using electro-mechanical impedance technique," *Struct. Heal. Monit.*, vol. 17, no. 3, pp. 461–471, 2018, doi: 10.1177/1475921717703053.
- [4] M.A. DeRousseau, E. Lafitchiev, J. R. Kasprzyk, B. Rajagopalan, and W.V. Srubar, "A comparison of machine learning methods for predicting the compressive strength of field-placed concrete," *Constr. Build. Mater.*, vol. 228, p. 116661, 2019, doi: 10.1016/j.conbuildmat.2019.08.042.
- [5] H. Huang and H. V Burton, "Classification of in-plane failure modes for reinforced concrete frames with infills using machine learning," *J. Build. Eng.*, vol. 25, p. 100767, 2019, doi: 10.1016/j.jobte.2019.100767.
- [6] J. Zhang, Y. Sun, G. Li, Y. Wang, J. Sun, and J. Li, "Machine-learning-assisted shear strength prediction of reinforced concrete beams with and without stirrups," *Eng. Comput.*, vol. 38, no. 2, pp. 1293–1307, 2022, doi: 10.1007/s00366-020-01076-x.
- [7] H.B. Ly, T.T. Le, H.L. Thi Vu, V.Q. Tran, L.M. Le, and B.T. Pham, "Erratum: Computational hybrid machine learning based prediction of shear capacity for steel fiber reinforced concrete beams [Sustainability 12 (2020) (2709)]," *Sustain.*, vol. 12, no. 17, 2020, doi: 10.3390/su12177029.
- [8] M. Khudhair and N. Gucunski, "Integrating Data from Multiple Nondestructive Evaluation Technologies Using Machine Learning Algorithms for the Enhanced Assessment of a Concrete Bridge Deck," *Signals*, vol. 4, no. 4, pp. 836–858, 2023, doi: 10.3390/signals4040046.
- [9] D. Ai and J. Cheng, "A deep learning approach for electromechanical impedance based concrete structural damage quantification using two-dimensional convolutional neural network," *Mech. Syst. Signal Process.*, vol. 183, no. May, p. 109634, 2023, doi: 10.1016/j.ymsp.2022.109634.
- [10] G. Li, M. Luo, J. Huang, and W. Li, "Early-age concrete strength monitoring using smart aggregate based on electromechanical impedance and machine learning," *Mech. Syst. Signal Process.*, vol. 186, no. May, p. 109865, 2023, doi: 10.1016/j.ymsp.2022.109865.
- [11] D. Ai, F. Mo, J. Cheng, and L. Du, "Deep learning of electromechanical impedance for concrete structural damage identification using 1-D convolutional neural networks," *Constr. Build. Mater.*, vol. 385, no. March, p. 131423, 2023, doi: 10.1016/j.conbuildmat.2023.131423.
- [12] S. Li, X. Zhao, and G. Zhou, "Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network," *Comput. Civ. Infrastruct. Eng.*, vol. 34, no. 7, pp. 616–634, 2019, doi: 10.1111/mice.12433.
- [13] Y. Yu, C. Wang, X. Gu, and J. Li, "A novel deep learning-based method for damage identification of smart building structures," *Struct. Heal. Monit.*, vol. 18, no. 1, pp. 143–163, 2019, doi: 10.1177/1475921718804132.
- [14] A. Aabid et al., "A review of piezoelectric materials based structural control and health monitoring techniques for engineering structures : challenges and opportunities," *Actuators*, vol. 10, no. 5, p. 101, 2021, doi: 10.3390/act10050101.
- [15] A. Aabid et al., "A Systematic Review of Piezoelectric Materials and Energy Harvesters for Industrial Applications," *Sensors*, vol. 21, pp. 1–28, 2021, doi: 10.3390/s21124145.
- [16] F.P. Sun, Z. Chaudhry, C. Liang, and C.A. Rogers, "Truss Structure Integrity Identification Using PZT Sensor-Actuator," *J. Intell. Mater. Syst. Struct.*, vol. 6, no. 1, pp. 134–139, 1995, doi: 10.1177/1045389X9500600117.
- [17] G. Park and D.J. Inman, "Impedance-based structural health monitoring," in *Damage Prognosis: For Aerospace, Civil and Mechanical Systems*, Wiley, pp. 275–292, 2005.
- [18] C. Liang, F.P. Sun, and C.A. Rogers, "Coupled Electro-Mechanical Analysis of Adaptive Material Systems -- Determination of the Actuator Power Consumption and System Energy Transfer," *J. Intell. Mater. Syst. Struct.*, vol. 5, no. 1, pp. 12–20, Jan. 1994, doi: 10.1177/1045389X9400500102.
- [19] V.M. Karbhari and L.S.W. Lee, *Vibration-based damage detection techniques for structural health monitoring of civil infrastructure systems*. Woodhead Publishing Limited, 2009. doi: 10.1533/9781845696825.1.177.
- [20] OxfordSparks, "What is Machine Learning? - YouTube," 2017 [Online]. Available: https://www.youtube.com/watch?v=f_uwKZIAeM0.
- [21] C.S.N. Pathirage, J. Li, L. Li, H. Hao, W. Liu, and R. Wang, "Development and application of a deep learning-based sparse autoencoder framework for structural damage identification," *Struct. Heal. Monit.*, vol. 18, no. 1, pp. 103–122, 2019, doi: 10.1177/1475921718800363.
- [22] A.E. Maxwell, T.A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review," *Int. J. Remote Sens.*, vol. 39, no. 9, pp. 2784–2817, 2018, doi: 10.1080/01431161.2018.1433343.
- [23] F. Maselli, G. Chirici, L. Bottai, P. Corona, and M. Marchetti, "Estimation of Mediterranean forest attributes by the application of k-NN procedures to multitemporal Landsat ETM+ images," *Int. J. Remote Sens.*, vol. 26, no. 17, pp. 3781–3796, 2005, doi: 10.1080/01431160500166433.
- [24] B. Kim and S. Cho, "Automated multiple concrete damage detection using instance segmentation deep learning model," *Appl. Sci.*, vol. 10, no. 22, pp. 1–17, 2020, doi: 10.3390/app10228008.
- [25] Ž. Vujović, "Classification Model Evaluation Metrics," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, pp. 599–606, 2021, doi: 10.14569/IJACSA.2021.0120670.