

Fracture mode identification in cementitious materials using supervised pattern recognition of acoustic emission features



Alireza Farhidzadeh^a, Anastasios C. Mpalaskas^b, Theodore E. Matikas^b, Hamidreza Farhidzadeh^c, Dimitrios G. Aggelis^{d,*}

^a Smart Structures Research Laboratory (SSRL), Department of Civil, Structural, and Environmental Engineering, University at Buffalo, USA

^b Department of Materials Science and Engineering, University of Ioannina, 45110 Ioannina, Greece

^c Department of Computer Science and Engineering, University of South Florida, FL 33620, USA

^d Department of Mechanics of Materials and Constructions, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

HIGHLIGHTS

- Classification of AE features according to the original fracture mode based on pattern recognition algorithm.
- Quantification of the effect of propagation distance to the classification error.
- Evaluation of different parameters as to their classification power.

ARTICLE INFO

Article history:

Received 18 November 2013

Received in revised form 12 February 2014

Accepted 9 May 2014

Available online 6 June 2014

Keywords:

Acoustic emission

Mortar

Crack mode classification

Frequency

RA value

Support vector machines

ABSTRACT

Cracking in concrete as a ubiquitous cementitious material in civil structures has been a worldwide critical issue in the field of engineering. Acoustic emission (AE) has demonstrated promising outcomes in research and laboratory experiments for monitoring these structures that led to plethora of reports, articles and recommendations for concrete structures. Many of these studies focus on cracking mode detection to estimate the significance of damage because in general, shear-like phenomena indicate severe damage and occur after tensile (flexural) cracking. The distinctive signs of the cracking modes are embedded in some AE parameters like the RA-value and average frequency (AF). Signals emitted from shear fracture exhibit higher RA-values with smaller AF than tensile ones. However, there are no universally fixed boundaries for classification of these features due to the parameters like member geometry, material properties sensor location and response. In addition, although AE consists of a random set of data, the role of uncertainty is not fully taken into account in data processing. To overcome these deficiencies, this article proposes a pattern classifier technique titled support vector machines. Small-scale fracture experiments were carried out to impose controlled cracking modes, record AE data for each cracking mode, and evaluate the performance of classifiers. The results show that the classification boundaries for AE features and their associate uncertainties could be successfully estimated. The effect of sensor distance as an imperative parameter in variation of classification boundaries could be quantified. Furthermore, the adequacy of other feature sets (i.e., other than RA and AF) for classification was also examined.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

In the past three decades, significant efforts have been made towards the development of structural health monitoring (SHM) systems for concrete structures. A technique that demonstrates promising online monitoring is acoustic emission (AE) [1–6]. The

term AE describes the stress waves caused by sudden strain releases due to fracture of the material. Many studies have demonstrated that the modes of cracking (tensile or shear) in concrete structures emit different AE signatures [1,2,7–9]. While loading concrete structures until failure, tensile cracks (mode I) generally develop at moderate loading level (elastic behavior), while shear cracks (mode II) dominate at large loading levels (plastic behavior) [10]. Therefore, it could be beneficial to monitor the mode of cracks as a lead to estimate the structural damage state. Nevertheless, the

* Corresponding author. Tel.: +32 2 629 3541; fax: +32 2 6292928.

E-mail address: daggelis@vub.ac.be (D.G. Aggelis).

conventional AE analyses are mostly inadequate for real time monitoring and warning systems because they do not allow exploration of the uncertainties prevalent in the structural characterization problem.

Uncertainty is an important issue in damage classification for real world health monitoring systems. Variability in structure geometry, material properties, sensor characteristics, noise, temperature, humidity, measurement inaccuracy, and insufficient knowledge about the process of damage nucleation and evolution [11] are among imperative sources of uncertainty. The above factors increase the inherent variability of AE measurements which inevitably results from the random nature of fracture process. Indeed, each crack propagation event is unique and different from the previous or the next in terms of fractured area or released energy. This implies that all external sources of error or variability should be limited in order not to mask the anyway demanding classification attempted by the AE features. With the aid of ongoing computer science progress, statistical techniques and pattern classification are recently playing a significant role in the field of SHM; in particular, acoustic emission monitoring. It has been used to either understand the underlying source of mechanism or to define criteria for probabilistic decision making [2,12–18].

The issue of cracking mode characterization based on AE signals, has been treated quite adequately in laboratory by the “moment tensor analysis” (MTA) [19,20]. Despite the successful classification achieved in usually small scale experiments, application of MTA to larger scale is not straightforward mainly because each cracking event must be recorded by at least six transducers. This needs expensive instrumentation in an actual case of monitoring since the transducers are dispersed to cover a large volume and supply information for several different zones of the structure. The result is sensor separation distances of the order of meters which do not allow acquisition of AE waves from the same fracture event by as many as six sensors. Therefore, a new procedure is necessary to yield crucial and reliable information using less number of sensors.

A full definition of AE features can be found in literature, e.g. [21], however the definition of the most common features is also done herein. Fig. 1 depicts a typical waveform with its main features. A threshold is always defined by the user in order to avoid noise signals while the first time the waveform crosses the threshold (threshold crossing, count) is considered the onset of the waveform. One of the most important waveform parameters is Amplitude which is the voltage of the highest peak and is commonly measured in Volts or dB. Duration is the time window between the onset and the last threshold crossing, while rise time (RT) is the time between the onset and the maximum peak. “Energy” is another important parameter that measures the area under the rectified signal envelope (MARSE). A parameter taking

into account the initial rising angle of the signal is RA value which is RT over amplitude (Eq. (1)) and is measured in $\mu\text{s}/\text{V}$. Frequency indicators can be found in the form of “average frequency, AF” which is defined simply as the number of counts over duration (Eq. (2)) while “central frequency” and “peak frequency” correspond to the centroid and the frequency with the maximum magnitude of the spectrum after FFT of the waveforms.

$$\text{RA} = (\text{Rise time})/(\text{Peak amplitude}) \quad (1)$$

$$\text{AF} = (\text{Counts})/\text{Duration} \quad (2)$$

Energy-related features, like energy and amplitude are connected to the intensity of the cracking source. Additionally, waveform features like duration, RT, RA and AF have been shown to correlate well to the fracture mode and are proposed for crack characterization in concrete [22]. Specifically, it has been consistently observed that tensile mode of cracking results in AE of higher frequency content and shorter duration [23,24]. The actual reason is related to the elastic wave modes excited by the different motion of the tips of the cracks. At the tensile mode, due to the opposing displacement of the sides vertical to the crack plane, a volumetric change occurs in the vicinity of the crack tip emitting most of the energy in the longitudinal wave mode. On the other hand, under shear cracking, the sides of the crack move in opposite directions but in parallel to the crack plane, which introduces a change of shape instead of volume [1].

Due to the large proportion of shear wave energy and the lower velocity of shear waves, the meaningful content of the waveforms is delayed compared to the “tensile” waveforms. Simulation studies concerning through the thickness and surface cracks in concrete have confirmed these trends [24–26]. Additionally, in several dedicated experimental studies, AE parameters like AF and RA have shown a definite change when the fracture mechanism shifts from tensile (micro-) cracking to debonding, fiber pull-out or actual shear [1,2,8,18,22,23,27,28]. AF registers an average decrease of 50% and RA an even stronger increase, which allow identifying the different stages at laboratory scale [1].

Fig. 2 shows a simple representation of classification using the two aforementioned parameters, AF and RA. Though this classification produces quite successful results in discriminating the different modes, it can only be safely applied in laboratory scale and only if the boundaries have been acquired by experiment at similar specimens. Applying similar classification in large scale is more complicated because of the influence of long propagation on the elastic waveforms. Concrete is heterogeneous and effectively scatters and damps the waves. Therefore, depending on the distance between the cracking source and the AE receiver, the wave will undergo changes in vital parameters like its frequency content and amplitude (downshifted) and RT and duration (increased for

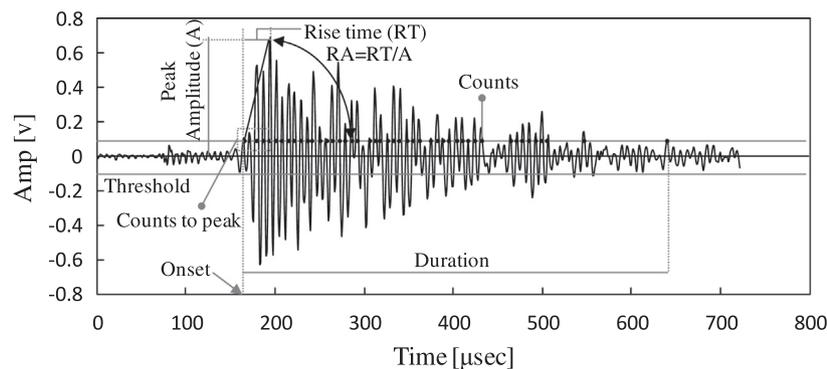


Fig. 1. A typical AE signal with some features $x = (\text{Amp}, \text{Dur})$.

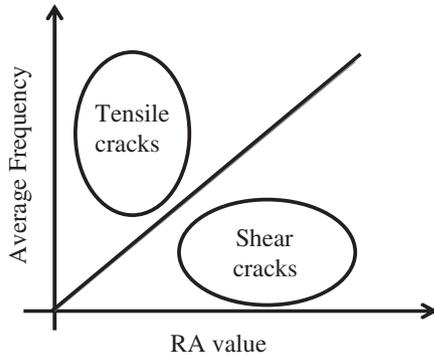


Fig. 2. Conventional crack classification in JCMS-IIIB5706 code [29].

moderate distances due to scattering). This leads to increase of the RA value as well. The changes will be stronger as the propagation distance increases. Therefore, any classification rule developed by experiments will only apply to the specific experimental conditions and should not be lightheartedly applied for AE signals recorded by sensors at different separation distances. The same applies for treating data from different sensors in a single population since waveforms collected at different locations will have different accumulated effect of scattering.

In this study, experimental results of a recent investigation on fracture of cementitious materials are analyzed. Experiments targeted the bending and shear mode of fracture showing that frequency decreases by 35–50% in average between tensile and shear mode and RA increases by several times. However, classification success is restricted if all data are considered in a single population as the tensile cracking signals collected at the furthest sensor from the crack tend to resemble the shear ones due to downshifting of their frequency and “stretching” in duration.

The objective of this article is twofold. First, the proportionality problem in conventional classification (i.e., identifying the straight line in Fig. 2) is addressed. Second, other AE features in which the characterization of cracking modes is highly likely and have not been extensively used are introduced. To approach this goal, the authors propose a pattern recognition technique for crack classification using AE signals, i.e., Support Vector Machine (SVM). Studying the effect of sensor distance from cracks, performance of the classifier, and possibility of using other AE features rather than conventional ones, are among the main objectives of this research.

The rest of this paper is organized as follows. A synopsis of pattern recognition along with mathematical background of SVM is presented in the next section. Then the experimental setup is described. This section is followed by the results of analyses for each classifier, cross-validation, and effect of sensor distance. Finally, the concluding remarks are given.

2. Pattern recognition

In machine learning, pattern recognition is: “the act of taking in raw data and taking an action based on the “category” of the pattern” [30]. The output is a set of labeled data associated to a set of classes. Pattern recognition problems are generally categorized based on the type of learning that generates the output, i.e., supervised and unsupervised. Supervised learning exploits a set of training data that consists of a set of labels clarifying to which class each training data belongs to. This training set is used to train (learn) a classifier as well as possible. In unsupervised learning, on the other hand, the training data does not have labels. Therefore, it attempts to discover the inherent patterns in the training data that can then be used to determine the hidden class boundaries. In either case, performance of the classifier is examined by a set of

validation (test) data to assess the generalization on new data sets. In case of unavailability of test data, *cross-validation* is applied. It consists of splitting initial training data set into two parts, the training part and the validation part. The classifier is trained on the training part and then to estimate the generalization accuracy, the classifier is tested on the validation part. It is imperative that the validation set not have points also in the training set.

With advent of advanced statistical techniques and high performance computing, pattern recognition of acoustic emission has recently attracted researcher’s attention [31–35]. In particular, to classify crack modes in large-scale reinforced concrete structures, unsupervised pattern classification techniques have been used when the source of AE events were unknown [2,17,31,32]. Gaussian Mixture Modeling and likelihood ratio test are among successful methods that could probabilistically discover the underlying hidden nature of AE sources [2]. In another investigation, k-means and Sifted *b*-value (*Sb*) analysis [17,33,34] were proposed to estimate the micro-to-macro crack transition stages for each principal crack mode when no information is available for the data labels. In this study, small-scale experiments with controlled cracking mode were conducted in order to obtain the “label” (i.e., crack mode) for each AE datum so as to clearly define the associated boundaries between the classes of shear or tensile. Support Vector Machine (SVM) was adopted to classify the labeled AE data as a robust linear classifier. SVM description and its mathematical background is briefly discussed as follows.

2.1. Support vector machine

Support vector machine is a linear discriminant classifier that is widely used in pattern recognition problem due to its decent generalization performance. SVM has been limitedly investigated for classification of AE in composite laminates [36], pipeline monitoring to discriminate leakage signals from other type signals [37], and elimination of noisy signals [38] but its application in AE classification for cementitious materials has not been investigated. The theory of SVM is well documented in textbooks [30,39]. In this section, the concept of this technique with a short review of the mathematical formulation is presented.

SVMs preprocess the data to map them to a sufficiently high dimension, usually higher than the original feature space because data from two classes can always be separated by a hyperplane if a suitable mapping (kernel) function $\phi()$ is applied [30,39]. Here we assume each pattern \mathbf{x}_k has been transformed to $\mathbf{y}_k = \phi(\mathbf{x}_k)$. The linear discriminant $g()$ in an augmented \mathbf{y} space is [30]:

$$g(\mathbf{y}) = \mathbf{a}^T \mathbf{y} \quad (3)$$

where \mathbf{a} is the weight vector and \mathbf{y} is the transformed pattern. Next, for each pattern \mathbf{x}_k , $k = 1, 2, \dots, n$, we assign $z_k = \pm 1$ to distinguish. Whether pattern k is in class ω_1 or ω_2 (i.e., $z = +1$ or $z = -1$). A discriminant hyperplane insures [30]

$$z_k g(\mathbf{y}_k) \geq 0 \quad k = 1, \dots, n \quad (4)$$

Support vector machine finds the separating hyperplane with the largest margin between hyperplane and the data from each class; see Fig. 3. The larger margin leads to better generalization of the classifier. The distance from a (transformed) pattern \mathbf{y} to the hyperplane is $r = |g(\mathbf{y})| / \|\mathbf{a}\|$ [30], and assuming that a positive margin b exists, Eq. (4) implies:

$$\frac{z_k g(\mathbf{y}_k)}{\|\mathbf{a}\|} \geq b \quad k = 1, \dots, n \quad (5)$$

The goal is to find the weight vector \mathbf{a} that maximizes b . The solution vector can be scaled arbitrarily while it still preserves the hyperplane. Therefore, the constraint $b \|\mathbf{a}\| = 1$ is imposed to

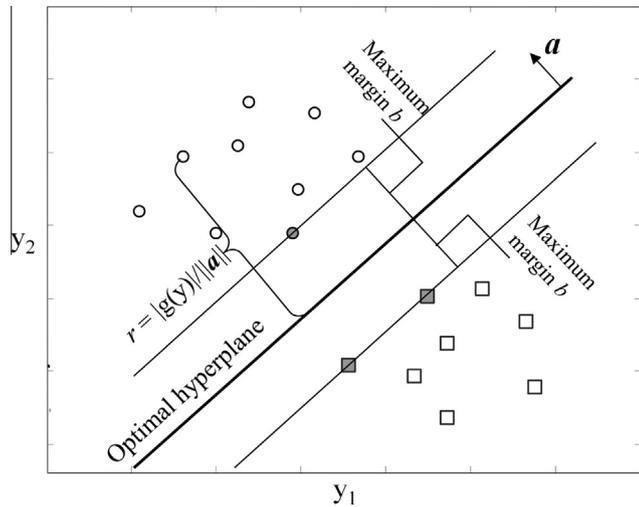


Fig. 3. SVM discriminant hyperplane.

insure unique solution [30]. Given that b should be maximized, this constrain demand the solution to minimize $\|a\|^2$. To solve Eq. (5) with the aforementioned constrain, the method of Lagrange undetermined multipliers is used in the following functional:

$$L(a, \alpha) = \frac{1}{2} \|a\|^2 - \sum_{k=1}^n \alpha_k [z_k a^T y_k - 1] \quad (6)$$

The optimized solution seek to minimize $L()$ with respect to a and maximize it with respect to the Lagrange multipliers α_k [30]. The last term in Eq. (6) represents the task of classifying the points correctly (see Eqs. (4) and (5)). The solution of Eq. (6) incorporates a reformulation process [40] to eventually find the support vectors. The support vectors are those transformed training patterns that represent equality in Eq. (4) (filled markers in Fig. 3). They are the closest data to the margin. The error rate is finally calculated via the following equation:

$$\varepsilon = \frac{1}{m} \sum_{k=1}^m H(-1 \times z_k g(y_k)); \varepsilon \leq 0.5 \text{ otherwise : } \varepsilon = 1 - \varepsilon \quad (7)$$

where m is the size of testing vector and $H(u)$ is the Heaviside step function. If the error from Eq. (7) was above 0.5, it means the clustering regions are reverse and we need to flip the sign of z_k and update the error as shown in the right hand side of Eq. (7). The error rate depends on the extent of overlapping within the clusters. It is not a parametric function of the feature space, but it can be quantified using Eq. (7). The advantage of SVM is its ability to present the classification boundary as a parametric function (not to present error as a parametric function). To quantify uncertainty (or to find the margin of error), cross-validation algorithm should be implemented by random selection of training data in a recursive fashion.

Similar to every clustering technique, SVM has some limitations. The most important governing parameter of SVM is selection of a proper kernel function. Based on the feature space with high dimensionality, we may not be able to choose a kernel function at the first glance. A prior information about the clusters (here Japan Construction and Material Standard for crack mode classification was used) is highly helpful for choosing an appropriate kernel function. In this study a linear kernel function was used and resulted in a robust classification. Another restriction of SVM is training large datasets which can result in slow performance. In this study, we did not encounter such a problem, as the number of data points was limited.

In addition, if a higher order (nonlinear) kernel function is used for SVM, special care should be taken about over-fitting when the boundary becomes biased. In such conditions, over-fitting can be resolved by greedy tuning of regularization parameters and using an appropriate kernel function. Since SVM is independent of the feature space dimension, we can have “resistant” marginal error by adjusting the SVM parameters of kernel function. Resistant means that SVM can handle outliers better or that it is insensitive to the outliers.

3. Experimental setup

This paper focuses more at the pattern recognition analysis. However, basic information about the experimental procedure is given here, while the interested reader can be directed for more details to [23]. The material was mortar with water to cement ratio 0.55 and sand to cement 3.5 by mass. The maximum sand grain size was 4.75 mm while the specimens were prisms of dimensions $40 \times 40 \times 160$ mm.

Six specimens were subjected to monotonic loading through three-point bending according to EN 13892-2: 2002 [41], see Fig. 4. The loading rate was 50 N/s. A slight modification was done for the other six specimens in order to study of the shear fracture mode. Specifically, the center load was transferred by a metal tab in order to be distributed while the supports were also distributed leaving only a small shearing zone. FEM analysis confirmed that with the specific setup the dominant stresses were shear as opposed to the tensile of bending [24].

For AE monitoring purposes, two broadband AE sensors with maximum sensitivity at 450 kHz (Pico, PAC) were attached to the same side of the specimen at specific distances from the cracking zone, as seen in Fig. 4a and b. The first sensor, S1, was placed 15 mm from the cracking zone while the second, S2, at 55 mm from the expected location of the crack which was secured by small notches, as seen in Fig. 4b for the shear mode. The waveforms were recorded in a two-channel monitoring board PCI-2, of PAC with a sampling rate of 5 MHz. The threshold and pre-amplifier gain was set to 40 dB. In the present analysis signal belonging to “AE events” were only considered to increase reliability. This means that for each signal captured by the nearby sensor (S1), another signal was captured by the second transducer (S2) within a limited time window (defined at 18 μ s) belonging to the same cracking incident. For each specimen a number of approximately 8–10 AE events were recorded from the location of the crack resulting in 116 hits for all tensile tests and 86 for the shear.

4. Results and discussion

To evaluate the performance of classifiers, the following AE features were initially extracted from AE signals: Counts, Energy, Duration, Amplitude, AF, Central-Frequency (CF), RA and Peak Frequency. In the first part of this section, classification performance is evaluated for the conventionally recommended features (i.e.,

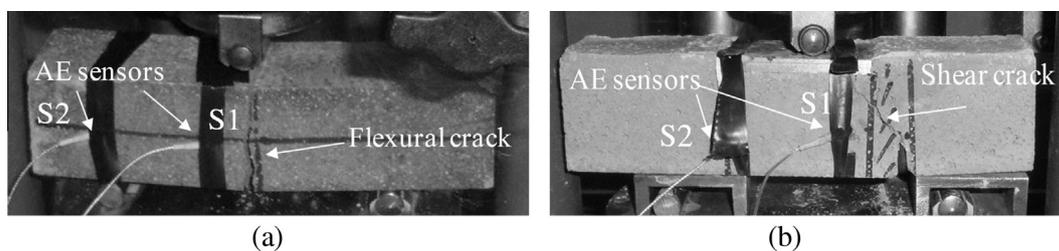


Fig. 4. Loading set up for (a) bending and (b) shear.

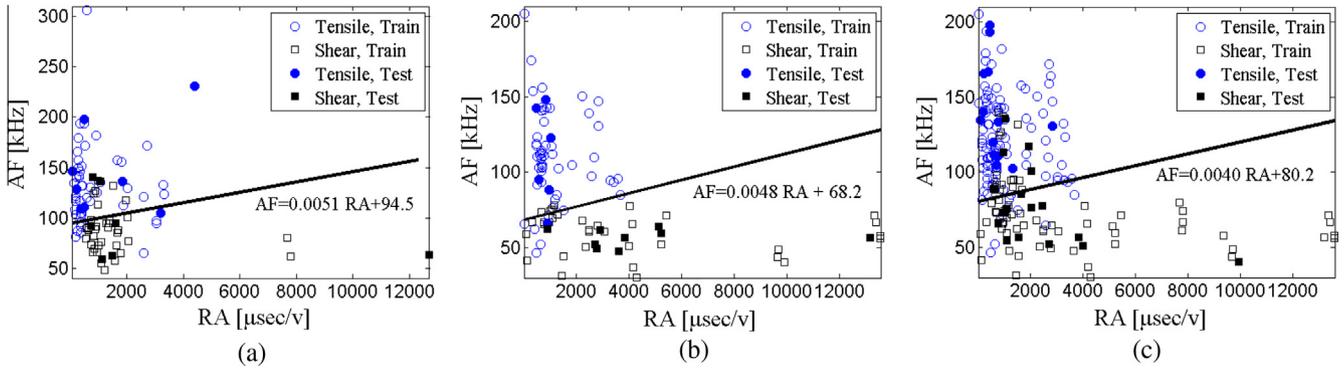


Fig. 5. Training SVM for $x = (RA, AF)$ and $\lambda = 15\%$ in (a).

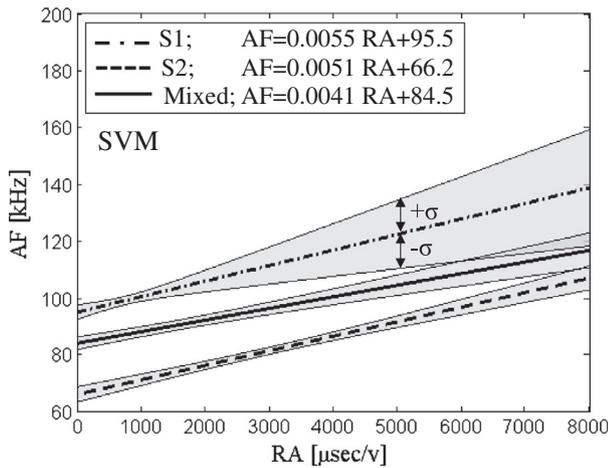


Fig. 6. Decision boundaries and margin of uncertainty for crack mode classification using support vector machine and conventional AE features.

RA and AF) for crack classification. To discover other possible feature pairs, all the pair-wise selection of features for all three cases (data from first sensor S1, data from S2, and mixed data from both sensors) were studied in the next part of this section.

4.1. Conventional features for crack classification

RA and AF populated the features vector $x = (RA, AF)$ for classification. A moving average with window span of 5 hits was then applied on the data to reduce scatter. To train each classifier, the data set D was divided in two subsets of training and testing. To study the robustness of SVM to the size of training and testing data, 5–50% (with 2.5% increment) of the total number of data in D were sampled as testing set, λ . For each validation test, random re-sampling carried out $\beta = 30$ times. The average of β computed error rates was then considered as the error rate associated to each testing size λ .

Fig. 5 illustrates results of SVM for $\lambda = 15\%$ as an example for one of the 30 resamplings. The hollow and solid marks represent the training and testing data, respectively. The linear discriminant function also shows that the boundary varies for sensor distance.

To study the uncertainty region for classification decision boundaries and sensitivity to the sensor distance, Fig. 6 compares the boundaries for each sensor and mixed data. The thick lines show the average of boundaries and the gray area represents spread parameter (or standard deviation σ). The equation of linear discriminants for the averaged boundaries are given in the legend. The results are highlighting the effect of propagation distance on

the AE classification. The boundary that best separates the two populations (tensile and shear signals) is actually moving according to the propagation distance of the AE waves to the sensors. When classification is based on the furthest sensor (55 mm from the crack as opposed to 15 mm) the line is translated to lower AF and higher RA values. When the data of both sensors are treated as a single population the line boundary is basically within the two. Furthermore, the spread parameter is maximum for the case where data are from S1. This means that the classification boundary is less reliable using a very close sensor to location of crack. On the other hand, the data from S2 result in a more robust classification boundary with less spread. When the data are mixed, the spread parameter is between the two. The promising point is that, for each boundary, the increasing spread moves toward high RA and AF values where the classification is less critical and easier due to smaller population of data (see Fig. 5). This is the first effort to quantify the influence of propagation distance on the classification behavior and not only on the waveform parameters.

To investigate the effect of λ on the error rate, Fig. 7 depicts variation of error rate versus λ for all the cases. When data from S2 is deployed into analysis, the minimum error is obtained. On the other hand, the maximum error belongs to S1 (closest sensor). The error almost oscillates around a certain value and thus there is no significant correlation between λ and error rate. To approximate quantitative values for errors in each case, the results were averaged and reported in Table 1 along with standard deviation of error. The least error rate of 8.4% is obtained with the least variation (standard deviation = 0.3) using data from S2 (furthest sensor). The fact that generally S2 exhibits less error rates than S1 cannot be securely evaluated at this time. It is possible that

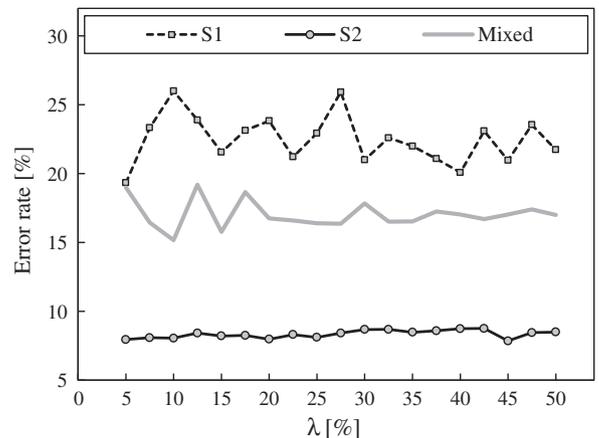


Fig. 7. Variation of error rate versus λ by cross-validation.

Table 1
Average of error rates for testing data.

Sensor distance	Error (%)	σ (%)
15 mm (S1)	22.5	1.8
55 mm (S2)	8.4	0.3
Mixed	17.0	1.0

there is a variability related to near field effects since the emitted wavelengths are typically longer than 10 mm and comparable to the distance to the first sensor.

4.2. Other feasible features for crack classification

This study goes one step further than recent literature. The objective in classification is to find the feature spaces that minimize the classification error. Although the classification error of the order of 10% (using SVM) is considered quite successful, other AE features provided much better performance, even for the closer sensor.

Table 2 reports the classification error rates for various couples of AE parameters extracted from S1. Tables 3 and 4 contain the error rates for S2 and mixed data, respectively. The error rates less than 10% are bolded and those less than 15% are underlined. In each case there are some feature pairs that result in less error than AF–RA, the highlighted cell.

In case of S1, the minimum error rate belongs to Amp (amplitude)–CF (center frequency) with error of 9.2%. In this case, neither RA nor AF perform well in conjunction with other parameters because their combination with any other feature shows relatively high error (e.g., see column of RA in Table 2). On the other hand, features like amplitude and center frequency perform better in conjunction with other features. The pairs of Amp–Eng (energy) and Amp–Dur (duration) seem to be other appropriate candidates for classification purpose.

In case of S2, Amplitude–energy possesses the minimum error for classification. Furthermore, it is interesting to note that features of AF and PF (peak frequency) generally outweigh other features because they result in minimum error in combination with all other features (see column and row of AF or column of PF in Table 3). The feature vector of (AF, PF) shows error in order of 6%. It is remarkable to note that the feature candidates for closer sensor, i.e., Amp–Eng, Amp–Dur, and Amp–CF, have acceptable performance on S2 as well (error \leq 15%).

When the data from the sensors are mixed, it is hard to find small errors (i.e., less than 15%). PF–Dur or PF–AF could be potential candidates to be used in case of mixed data. There are some features that irrespectively of data set origin (S1, S2 or mixed), result in almost similar and low error rates around 15% or below. They are printed in italic format in these tables. Amp–Dur, Amp–CF, AF–CF, and engagement of PF with many other features are among these informative feature selection.

Table 2
Error rates for S1 (%).

FEATURE	Eng	Dur	Amp	AF	CF	RA	PF
Count	21.9	23.1	24.5	20.8	14.4	37.8	19.3
Eng		32.0	<u>10.8</u>	22.3	18.6	33.1	21.1
Dur			<u>10.7</u>	23.5	20.9	33.0	16.5
Amp				15.3	9.2	22.2	<u>14.8</u>
AF					17.9	22.5	<u>14.6</u>
CF						18.6	17.9
RA							15.1

Table 3
Error rates for S2 (%).

FEATURE	Eng	Dur	Amp	AF	CF	RA	PF
Count	<u>13.3</u>	27.8	18.2	8.0	24.1	19.5	9.7
Eng		33.3	5.7	9.1	34.0	20.3	9.4
Dur			15.0	9.1	32.7	26.2	<u>12.1</u>
Amp				8.5	<u>13.6</u>	<u>10.1</u>	<u>10.3</u>
AF					9.1	8.4	6.6
CF						20.2	9.5
RA							<u>11.2</u>

Table 4
Error rates for mixed data (%).

FEATURE	Eng	Dur	Amp	AF	CF	RA	PF
Count	22.0	22.2	27.7	15.9	22.2	33.1	16.3
Eng		36.5	19.1	16.3	26.5	27.9	15.6
Dur			16.8	16.1	28.4	23.0	<u>14.8</u>
Amp				13.7	15.6	27.4	15.8
AF					15.3	17.0	<u>14.5</u>
CF						23.5	16.0
RA							16.8

The fact that lowest error rates are achieved by separate sensors (mostly the 55 mm sensor), vividly demonstrates the paramount importance of data handling. Even the most effective algorithm cannot produce accurate results if the data are treated without respect to the propagation distance between the crack and the sensor. It should be kept in mind that these changes are noticed in small scale specimens. Despite the small scale and the anyway limited accumulated effect of distortion, classification results are substantially improved when treating separate AE data populations according to the distance from the crack. This is because data collected at the same distance (sensor) from the crack have undergone approximately the same amount of distortion (reduction of amplitude, frequency, duration elongation due to scattering) canceling out its effect. When data from several sensors are included in the same population, the effect of distortion is differential, meaning that some of the signals have undergone strong distortion and some weaker which leads to higher amount of classification error. Apart from demonstrating how heterogeneity influences the propagating waves, it is also shown how this influence is transferred to the final classification results of AE signals.

In the rest of this section, we study the variation of discriminant function for four pairs of Amp–Energy, Amp–CF, Amp–Dur, and PF–AF.

4.2.1. Amplitude–energy

Fig. 8 shows the boundaries between Amplitude and Energy for the different populations. At certain energy zones (e.g. below 10) the amplitude of tensile signals is higher; because in shear type signals, the energy is distributed more along the time of signal while in tensile signals it is mostly concentrated at the beginning part, as discussed in the introduction. Thus, given two signals with the same amplitude, the higher energy signal is likely shear. The translation of the line is shown in Fig. 9. Regarding the error, it is minimized for S2 reaching 5.5%, while for S1 and mixed data it is 10.8% and 19.1%, respectively (see Tables 3 and 4). It is worth to note that these error percentages are much lower than the AF–RA classification and the uncertainty region (with respect to the data domain) is much narrower particularly for S1 (compare Fig. 9 with Fig. 6). It means that if energy and amplitude construct

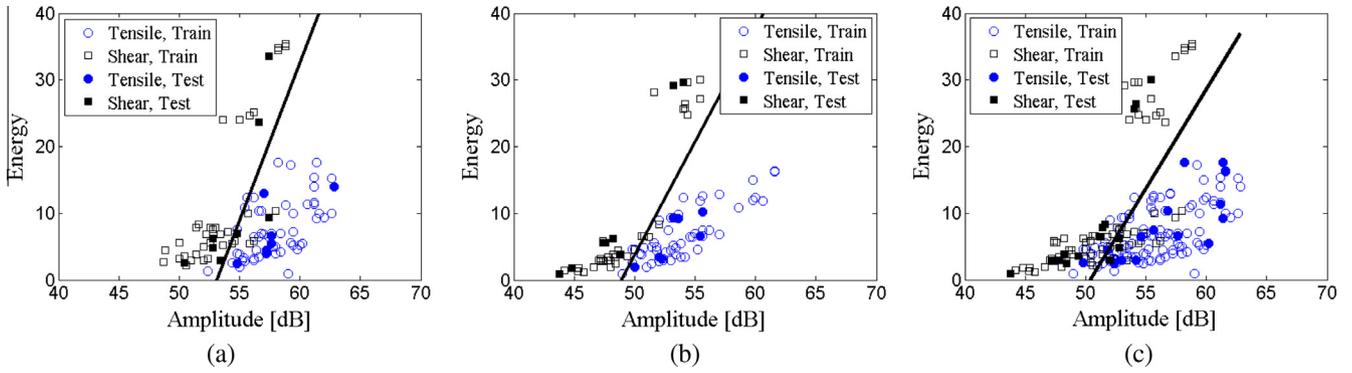


Fig. 8. Training SVM for $x = (\text{Amp-Eng})$ and $\lambda = 15\%$ in (a) S1, (b) S2 and (c) mixed.

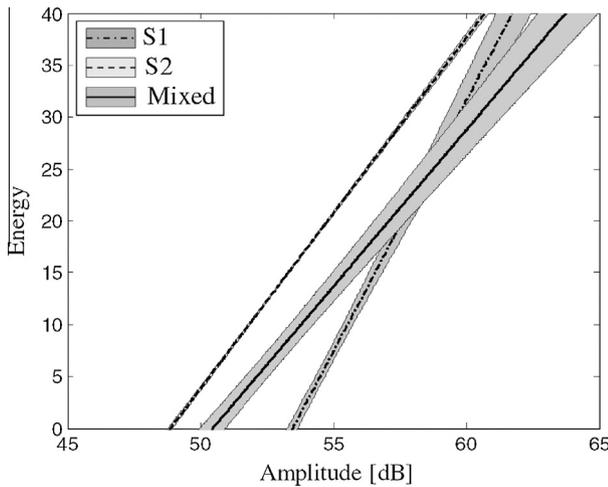


Fig. 9. Comparison of boundaries for Amp–Eng and $\lambda = 15\%$.

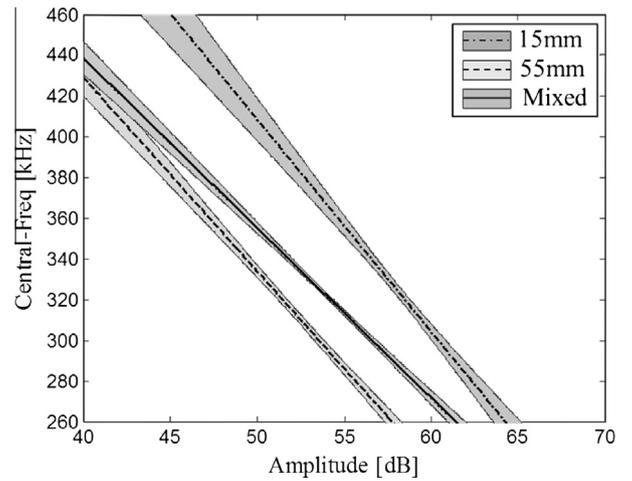


Fig. 11. Comparison of boundaries for Amp–CF and $\lambda = 15\%$.

the feature vector for classification, decision boundaries will be more consistent and robust, no matter how the testing data are sampled for cross-validation.

4.2.2. Amplitude–central frequency

Fig. 10 shows the central frequency versus amplitude. Shear signals demonstrate lower central frequency with smaller amplitude. The advantage of this pair is that both central frequency and amplitude are threshold-independent variables (in contrast to AF that both counts and duration are threshold-dependent). Therefore, this engagement could be more robust for classification purposes in noisy environment or floating threshold. To study the variation of decision boundary with sensor distance, the SVM

decision boundaries are exhibited in Fig. 11. They follow the same trend as before, i.e., the data from S2 push the boundary towards lower values and when the data are a mixture of S1 and S2, the decision boundary lies between S1 and S2.

4.2.3. Amplitude–duration

Another beneficial feature vector is $x = (\text{Amp}, \text{Dur})$. It is evident from Fig. 12 that there is a correlation between amplitude and duration meaning that higher amplitude signals usually have longer duration. However, duration of tensile signals is typically shorter than that of shear given a similar amplitude. Therefore, SVM could be well-trained so that no matter how the feature vector is populated (from S1, S2, or mixed), the estimated error is

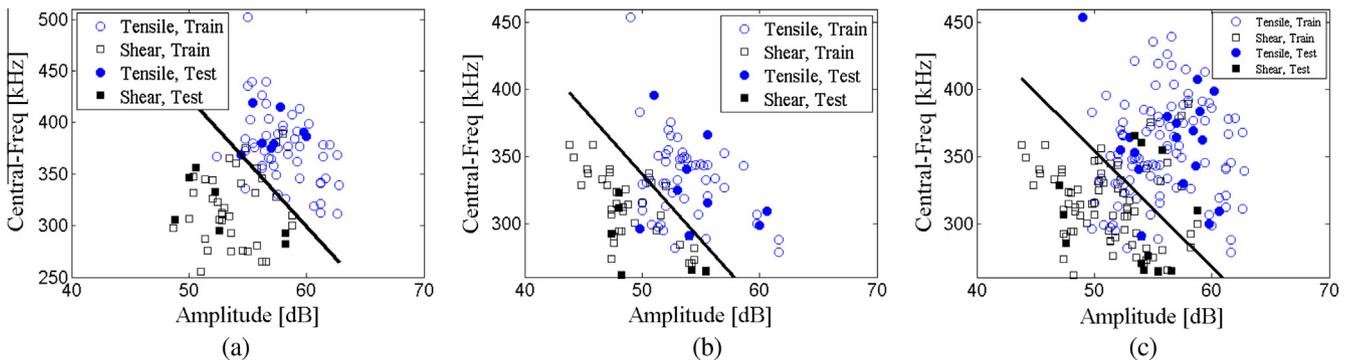


Fig. 10. Training SVM for $x = (\text{Amp}, \text{Eng})$ and $\lambda = 15\%$ in (a) S1, (b) S2 and (c) mixed.

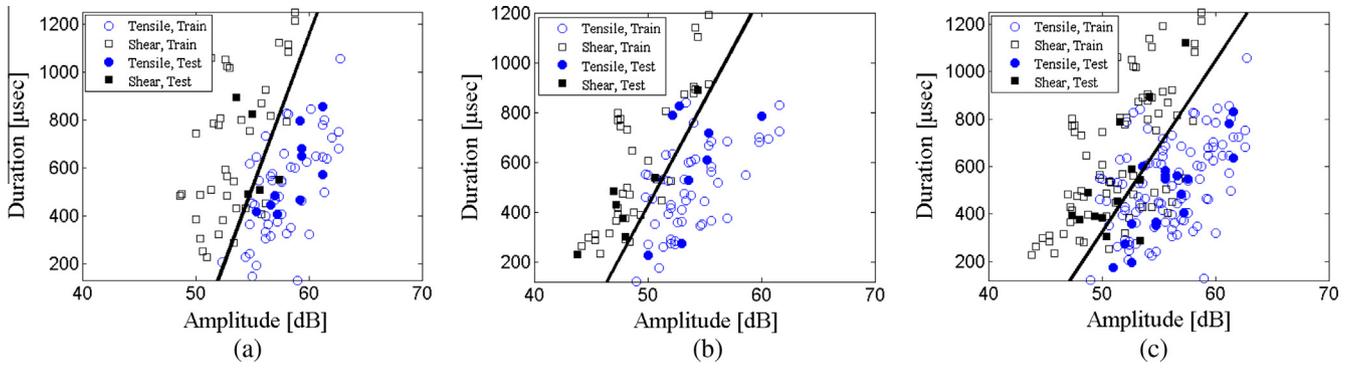


Fig. 12. Training SVM for $\mathbf{x} = (\text{Amp and Dur})$ and $\lambda = 15\%$ in (a) S1, (b) S2 and (c) mixed.

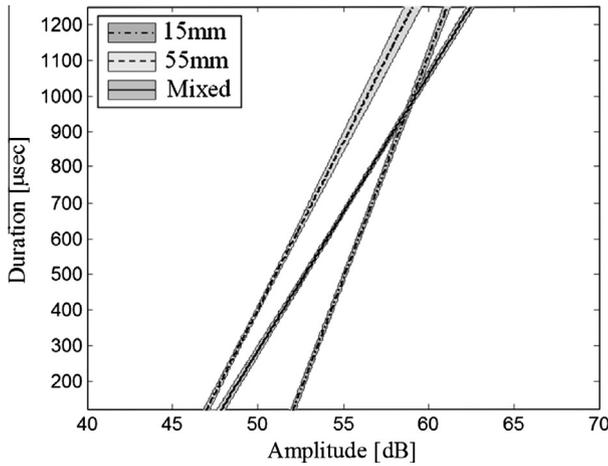


Fig. 13. Comparison of boundaries for Amp–Dur and $\lambda = 15\%$.

also investigated resulting in the lowest error for the 55 mm sensor in our laboratory study shown in Table 3. Combination of PF with other features led to a very small error as well. The distribution pattern is illustrated in Fig. 14. The figures present very well-separated clusters. The tensile signals possess higher peak frequency besides higher average frequency. It is interesting to note that for the closer sensor, there is more overlapping between clusters (see the group of shear signals with PF around 300 kHz in Fig. 14(a)) which clears out for the furthest sensor. As a result, Fig. 14b shows that the cloud of shear signals are pushed to lower values, leading to less cluster overlapping, and consequently minimum error rate at 55 mm. Fig. 15 depicts the variation of SVM decision boundaries by data set and sensor distance. Similar to

around 10–16%. Re-sampling repetition for cross-validation resulted in the same error range. The reason is depicted in Fig. 13; the spread parameter for decision boundaries is negligible. Consequently, this feature pair can be potentially adopted for crack classification with two fold advantages, (1) consistent error with different sensor distances, (2) less uncertainty about the decision regions. The first advantage could be due to the effect of attenuation on both amplitude and duration. Less uncertainty, on the other hand, is due to the aforementioned association of amplitude to duration that almost scatter around a linear correlation.

4.2.4. Peak frequency–average frequency

The last feature pair to be studied is AF–PF. In the search for the best classification in two dimensional space, this combination was

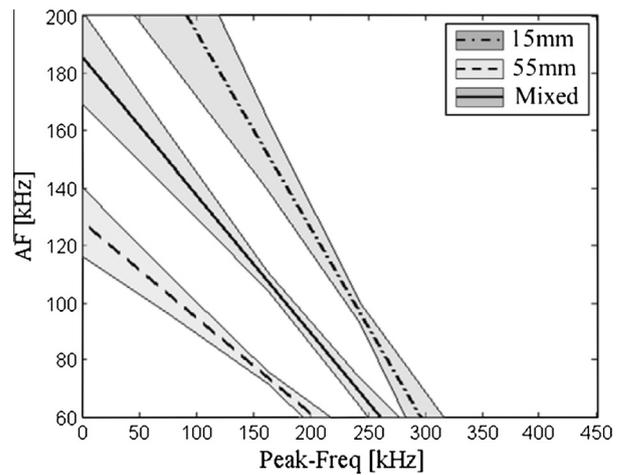


Fig. 15. Comparison of boundaries for AF–PF and $\lambda = 15\%$.

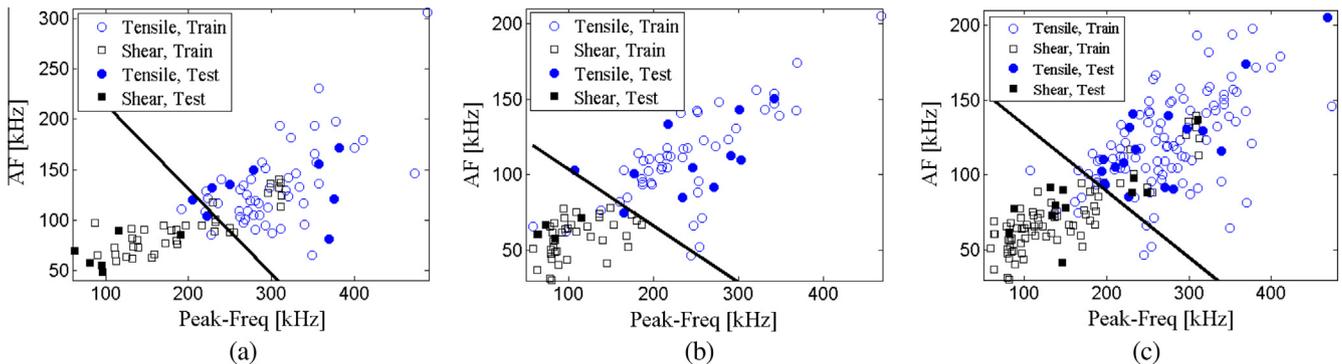


Fig. 14. Training SVM for $\mathbf{x} = (\text{PF, AF})$ and $\lambda = 15\%$ in (a) S1, (b) S2 and (c) mixed.

other feature pairs, when data set is a mixture of all sensors, the boundary lies between the boundaries for S1 and S2. The decision boundaries for different distances are quite highly apart; however, the spread parameter is quite larger than other feature pairs, meaning the boundary is more sensitive to re-sampling for cross-validation. In other words, despite the excellent classification performance for the data of S2 (only 6.6% error) the generalization performance of this pair is not straightforward.

5. Conclusion

The present work concerns classification of AE signals depending on the corresponding fracture mode with the aim of investigating a simple scheme for structural health monitoring purposes. The raw data come from bending and shear tests on cementitious mortar specimens using two sensors at specific distances from the crack. Support Vector Machine (SVM) is applied on pairs of AE features and their performance on classification is interrogated. Parameters like the population of the training and testing sets are also tested. Results quantitatively demonstrate that the boundaries between tensile and shear signals are translated according to the distance between the sensor and the cracking source. Other classifiers (i.e. Bayesian Decision Boundary) produce non-linear boundaries and have also been examined in the framework of this study. Their classification results (in terms of error percentage) are similar to SVM with linear kernel function. They are not reported here since the algorithm explanation and presentation of results would take much space and would make the manuscript tiresome.

Higher success rates are accomplished when the data of each sensor are treated separately rather than mixed in a single population. While the effect of propagation through concrete on the wave parameters is well known mainly in the framework of ultrasonic studies, this is one of the first approaches to quantify this effect in the field of AE. It shows that combining powerful algorithms and adequate treatment of the data may easily increase the characterization at least in laboratory conditions. Furthermore, new feature pairs are discussed that enhance the classification performance.

Based on this study, if a cementitious material is tested with AE monitoring, decision boundaries have to be modified for some different sensor distances. The source of AE has to be localized first and then classified as shear or tensile based on the closest pre-defined boundary. It is found that several pairs of descriptors perform quite well in terms of classification error regardless of the data set (population of close sensor, population of further sensor and mixed populations). The most indicative are AF–CF, Duration–Ampl, Amp–PF and Amp–CF. The last pairs are also threshold-independent. However, the best (lowest) error rate, is reached by the pair AF–RA for the data of the 2nd sensor alone. Still more experimental work needs to be conducted in order to upgrade the work. First in terms of scale to apply in larger and realistic geometries closer to the dimensions of actual concrete elements. Also the material will be concrete with large aggregates which is assumed to have even stronger scattering effect on the waves. Additionally, a larger number of sensors should be used in order to check the efficiency of the algorithms for several different distances.

Acknowledgments

This research project has been partially funded by the European Union (European Regional Development Fund-ERDF) and Greek National Funds through the Operational Program “Thessaly Mainland Greece and Epirus-2007-2013” of the National Strategic Reference Framework (NSRF 2007-2013).

References

- Aggelis DG. Classification of cracking mode in concrete by acoustic emission parameters. *Mech Res Commun* 2011;38:153–7.
- Farhidzadeh A, Salamone S, Singla P. A probabilistic approach for damage identification and crack classification in reinforced concrete structures. *J Intell Mater Syst Struct* 2013. <http://dx.doi.org/10.1177/1045389X13484101>.
- Kageyama K, Murayama H, Ohsawa I, Kanai M, Nagata K, Machijima Y, et al. Acoustic emission monitoring of a reinforced concrete structure by applying new fiber-optic sensors. *Smart Mater Struct* 2005;14:S52–9.
- Ohtsu M, Uchida M, Okamoto T, Yuyama S. Damage assessment of reinforced concrete beams qualified by acoustic emission. *ACI Struct J* 2002;411–7.
- Shiotani T, Aggelis DG, Makishima O. Global monitoring of large concrete structures using acoustic emission and ultrasonic techniques: case study. *J Bridge Eng* 2009;14(3):188–92.
- Carpinteri A, Lacidogna G, Niccolini G, Puzzi S. Critical defect size distributions in concrete structures detected by the acoustic emission technique. *Meccanica* 2008;43:349–63.
- Beck P, Lark RJ, Holford KM. Moment tensor analysis of acoustic emission in concrete specimens failed in four-point bending. *Key Eng Mater* 2003;245–246:443–50.
- Ohno K, Ohtsu M. Crack classification in concrete based on acoustic emission. *Constr Build Mater* 2010;24:2339–46.
- Shigeishi M. Acoustic emission moment tensor analysis: development for crack identification in concrete materials. *Constr Build Mater* 2001;15:311–9.
- Yuyama S, Li Z, Ito Y, Arazoe M. Quantitative analysis of fracture process in RC column foundation by moment tensor analysis of acoustic emission. *Constr Build Mater* 1999;13:87–97.
- Zhou W, Kovvali N, Reynolds W, Papandreou-Suppappola A, Chattopadhyay A, Cochran D. On the use of hidden Markov modeling and time-frequency features for damage classification in composite structures. *J Intell Mater Syst Struct* 2009;20:1271–88.
- Godin N. Clustering of acoustic emission signals collected during tensile tests on unidirectional glass/polyester composite using supervised and unsupervised classifiers. *NDT and E Int* 2004;37:253–64.
- Sause MGR, Gribov A, Unwin AR, Horn S. Pattern recognition approach to identify natural clusters of acoustic emission signals. *Pattern Recogn Lett* 2012;33:17–23.
- De Oliveira R, Marques AT. Health monitoring of FRP using acoustic emission and artificial neural networks. *Comput Struct* 2008;86:367–73.
- Sause MGR, Horn S. Quantification of the uncertainty of pattern recognition approaches applied to acoustic emission signals. *J Nondestruct Eval* 2013. <http://dx.doi.org/10.1007/s10921-013-0177-9>.
- Kahirdeh A, Khonsari MM. Criticality of degradation in composite materials subjected to cyclic loading. *Compos B: Eng* 2013; doi: <http://dx.doi.org/10.1016/j.compositesb.2013.06.048>.
- Farhidzadeh A, Dehghan-Niri E, Salamone S, Luna B, Whittaker A. Monitoring crack propagation in reinforced concrete shear walls by acoustic emission. *ASCE J Struct Eng* 2012;139(12):1–10.
- Calabrese L, Campanella G, Proverbio E, Engineering M. Use of cluster analysis of acoustic emission signals in evaluating damage severity in concrete structures. *J Acoust Emission* 2010;28:129–41.
- Ohtsu M, Kawasaki Y. AE-SiGMA analysis in Brazilian test and accelerated corrosion test of concrete. *J Acoust Emission* 2010;28:204–14.
- Ohtsu M, Okamoto T, Yuyama S. Moment tensor analysis of the acoustic emission source in the rock damage process. *ACI Struct J* 2005;15:609–13.
- Physical Acoustics Corporation. Chapter 1-Introduction. SAMOS AE SYSTEM USER'S MANUAL. Princeton Junction (NJ, USA): MISTRAS, GROUP; 2005, p. 28–32.
- Ohtsu M. Recommendation of RILEM TC 212-ACD: acoustic emission and related NDE techniques for crack detection and damage evaluation in concrete. *Mater Struct* 2010;43:1187–9.
- Aggelis DG, Verbruggen S, Tsangouri E, Tysmans T, Van Hemelrijck D. Characterization of mechanical performance of concrete beams with external reinforcement by acoustic emission and digital image correlation. *Constr Build Mater* 2013;47:1037–45.
- Aggelis DG, Mpalaskas AC, Matikas TE. Investigation of different fracture modes in cement-based materials by acoustic emission. *Cem Concr Res* 2013;48:1–8.
- Polyzos D, Papacharalampopoulos A, Shiotani T, Aggelis DG. Dependence of AE parameters on the propagation distance. *J Acoust Emission* 2011;29:57–67.
- Aggelis DG, Shiotani T, Papacharalampopoulos A, Polyzos D. The influence of propagation path on acoustic emission monitoring of concrete. *Struct Health Mon* 2012;11(3):359–66.
- Aggelis DG, Soulioti DV, Gatselou EA, Barkoula N-M, Matikas TE. Monitoring of the mechanical behavior of concrete with chemically treated steel fibers by acoustic emission. *Constr Build Mater* 2013;48:1255–60.
- Shahidan S, Pulin R, Muhamad Bunnori N, Holford KM. Damage classification in reinforced concrete beam by acoustic emission signal analysis. *Constr Build Mater* 2013;45:78–86.
- JCMS-IIIB5706. Japan Construction Material Standards. Monitoring Method for Active Cracks in Concrete by Acoustic Emission. Japan: The Federation of Construction Material Industries; 2003.
- Duda RO, Hart PE, Stork DG. Pattern Classification. 2nd ed. John Wiley & Sons, Inc.; 2001.

- [31] Farhidzadeh A, Salamone S, Luna B, Whittaker A. Acoustic emission monitoring of a reinforced concrete shear wall by *b*-value based outlier analysis. *Struct Health Monit – An Int J* 2013;12:3–13.
- [32] Farhidzadeh A, Dehghan-Niri E, Salamone S. Gaussian mixture modeling of acoustic emissions for structural health monitoring of reinforced concrete structures. *SPIE Smart Structures/NDE, San Diego: SPIE*; 2013.
- [33] Farhidzadeh A, Salamone S. Introducing sifted *b*-value analysis and a new crack classification for monitoring reinforced concrete shear walls by acoustic emission. 54th Acoustic emission working group meeting, Princeton, NJ, (Student paper award); 2012, p. 55–7.
- [34] Farhidzadeh A, Salamone S, Dehghan-Niri E, Luna B, Whittaker A. Damage assessment of reinforced concrete shear walls by acoustic emission. *NDE/NDT for highways and bridges: Structural Materials Technology (SMT)*; 2012. p. 74–81.
- [35] Rippengill S, Worden K, Holford KM, Pullin R. Automatic classification of acoustic emission patterns. *Strain* 2003;39:31–41.
- [36] Zhao J, Wang K. Acoustic emission signals classification based on support vector machine. In: 2010 2nd International conference on computer, engineering and technology; 2010:V6–300–V6–304.
- [37] Haifeng Z, Zhenlin L, Zhongli J, Hongxing L, Mingxiao L. Application of acoustic emission and support vector machine to detect the leakage of pipeline valve. In: 2013 Fifth international conference on measuring technology and mechatronics automation; 2013. p. 283–6.
- [38] Yu Y, Zhou L. Acoustic emission signal classification based on support vector machine. *Telkommnika* 2012;10:1027–32.
- [39] Bishop CM. *Pattern recognition and machine learning*. New York (USA): Springer; 2006.
- [40] Cherkasskyand V, Filip Mulier. *Learning from data: concepts, theory, and methods*. New York (NY): Wiley; 1998.
- [41] 13892-2:2002 E. *Methods of test for screed materials—Part 2: determination of flexural and compressive strength*; 2002.