

## **FAILURE PREDICTION WITH (PSEUDO) ACOUSTIC EMISSION AND SUPERVISED ALGORITHM RANDOM FOREST – CASE STUDY OF FOUR NUMERICAL SANDSTONES**

Piotr KLEJMENT<sup>1</sup>

Institute of Geophysics, Polish Academy of Sciences, Warsaw, Poland

### **Abstract**

In this paper, an automated methodology for predicting the stress state and time to failure of a material during a uniaxial compression test was proposed. It was shown that, based solely on pseudo-acoustic emission, the supervised machine learning algorithm Random Forest can perform predictions with good or very good accuracy. The Coefficient of Determination  $R^2$  on the test dataset reached 84% (for axial stress prediction) and 73% (for time to failure prediction). This work was limited to predictions only in numerical modeling using the Discrete Element Method. Cylindrical samples with macroscopic parameters corresponding to four real sandstones were generated. SHapley Additive exPlanations (SHAP) was applied to show what is the contribution of individual features of pseudo-acoustic emission to the algorithm's output and its predictions.

**Keywords:** uniaxial compression, numerical modelling, supervised machine learning, random forest, discrete element method

### **1. INTRODUCTION**

Acoustic emission (AE) is related to material deformation and fracture, as well as all kinds of groaning, creaking, and chattering of continuous grain motions [1]. It is the release of strain energy, as a result of which elastic waves propagate through the material. Acoustic emission provides useful information in many fields of geosciences, such as industrial applications, glacial calving, stick-slip cycles, sheared granular fault experiments, earthquakes, and others. For example, Rouet-Leduc et al. [2] showed that using supervised machine learning and acoustic emission, it is possible to predict the time to the next slip in laboratory stick-slip cycles, also called laboratory earthquakes. Similar conclusions were reached by Bolton et al. [3].

AE is monitored by a network of sensors placed on the surface of the analyzed material. AE provides information about events inside the material, including the occurrence of fractures [4]. Due to

---

<sup>1</sup>Corresponding author: Institute of Geophysics Polish Academy of Sciences (IG PAS), 64 Księcia Janusza Str. 01-452 Warszawa, Poland, pklejment@igf.edu.pl, +48 22 6915-950

its versatility and usefulness, the AE technique is widely used in material monitoring. AE allows for real-time, non-destructive, in-situ continuous monitoring of the occurrence of cracks in brittle or quasi-brittle materials. Detection of cracking at an early stage of loading is extremely important in civil engineering in maintaining the safety of structures and constructions [4].

Despite numerous works showing the effectiveness of AE application in materials testing, precise indication of the moment of crack occurrence in the material still remains a challenge. Numerical modeling can provide new results that complement laboratory studies. Numerical models have the advantage that experiments can be performed in any configuration in a repeatable manner, without destroying the samples. Modeling of brittle materials is often performed using the Discrete Element Method (DEM). In this work, the DEM method was used to generate models of four sandstones. The model parameters were calibrated so that their selected macroscopic parameters corresponded to the macroscopic parameters of real sandstones. Several series of uniaxial compression test simulations were performed on such models. During the simulation, information on selected parameters describing the particle velocity and the rate of bond fracturing was continuously recorded. Dorostkar and Carmeliet [1] showed that the parameters related to the particle velocity in the DEM model can be referred to as Pseudo Acoustic Emission (PAE). The simplified numerical model does not contain all the complexities present in nature, therefore, it is impossible to talk about classical acoustic emission in the DEM model.

Machine learning algorithms are useful in detecting hidden, nonlinear, and non-obvious relationships between data. The application of machine learning algorithms to the analysis of acoustic signals gives the opportunity to discover new relationships. This opens a new field for research in many fields. In addition, huge datasets are created as a result of numerical simulations, which is a natural field for the application of machine learning [5]. In recent years, more and more applications of machine learning have appeared in this field, including numerical simulations.

This work aims to combine machine learning, acoustic emission, and DEM numerical modeling techniques to perform automatic failure prediction in uniaxial compression. The collected PAE and DEM data were used to train a supervised machine learning algorithm Random Forest (RF). The aim of this work was to propose an automatic method to estimate the Axial Stress (AS) state in a sample and the estimated Time to Failure (TtF) using supervised machine learning, based on continuous PAE signal monitoring. This work was performed numerically on the example of four numerical sandstones. However, such a methodology has the potential to be a universal method and also to be extended to experimental studies.

## **2. DATA COLLECTION AND METHODOLOGY**

### **2.1. Numerical Simulations – Discrete Element Method**

Discrete Element Method (DEM) is a numerical method based on particles [6]. The material in this method is represented as a set of interacting particles. DEM allows the analysis of the translations and rotations of each particle in the model at each time step. DEM easily reproduces fracturing in brittle and quasi-brittle materials. In DEM, the motion of each particle is divided into two independent parts: the translational motion of the center of mass and the rotations about the center of mass. The translational motion is described by Newton's equations and integrated by the conventional Molecular Dynamics scheme. The rotations are described by Euler's equations. DEM description can be found in [6,7,8,9,10,11]. The details of the DEM implementation in ESyS-Particle are in [7,8,10,11]. The number of particles in the model is the result of the particle packing algorithm and depends on the cylinder geometry and particle sizes. The software algorithm uses pre-prepared grids, the largest particles are

placed first, and the smaller ones in the free spaces [12]. At the beginning of the simulation, all particles are connected by bonds, so there are no microcracks in the sample.

The basic scheme of DEM simulation consists of two parts. At the beginning of each simulation, the model is generated, particles are packed, and the network of bonds between them is created. In the second part, the actual simulation starts, in which at each time step, Newton's second law of dynamics is applied to each of the discrete elements to calculate the changes in position and velocity due to the acting forces. The numerical stability of the simulation is ensured by the Courant condition for the time step. The Courant condition is used to ensure that the particles do not move too far in one time step and interact only with their immediate surroundings. The DEM method is computationally expensive, and modeling different phenomena requires many simplifications to ensure a reasonable computation time. In this work, DEM modeling was performed using the open-source software ESyS-Particle [7]. Two main advantages of ESyS-Particle are the implemented MPI parallel engine and a wide range of types of interactions between particles [10,11].

In this work, the *BrittleBeamPrms* particle bond was used [10,11]. This bond was designed specifically for brittle fracture simulation. This bond takes into account translational and rotational degrees of freedom, and two connected particle bonds interact through normal, shear, bending, and twisting forces. The bond breaks according to the Mohr-Coulomb failure criterion. Four parameters describe the bond properties: microscopic Young's modulus  $E_b$ , microscopic Poisson's ratio  $\nu_b$ , microscopic cohesion  $c_b$ , and microscopic internal friction  $\phi_b$ .

## 2.2. DEM model of the Uniaxial Compression Test

The uniaxial Compression test is a common laboratory test in which a cylindrical sample is compressed from the top and bottom [13]. By measuring the change in sample size along one of the axes and the applied load, it is possible to calculate a stress-strain profile. Stress is defined as  $\sigma=F/A$ , where  $F$  is the applied load and  $A$  is the initial cross-sectional area of the sample normal to the loading direction. Strain can be calculated as  $\varepsilon_e=\Delta L/L_0$  where  $\Delta L$  is the measured displacement and  $L_0$  is the initial sample length along a single axis. The slope of the strain-stress curve within the linear elastic regime is equal to Young's modulus, which is the ratio of the stress to axial strain  $E= \sigma/\varepsilon_z$ . Poisson's ratio is the ratio of the relative transverse strain normal to the applied load  $\varepsilon_{xy}$  and the relative axial strain  $\varepsilon_z$  in the direction of the applied load. Poisson's ratio can be expressed as  $\nu= \varepsilon_{xy}/\varepsilon_z$  [14].

DEM model calibration is an iterative process of adjusting macroscopic parameters of the numerical and experimental samples to be similar. Calibration is done by a trial-and error method – an iterative process of adjusting input parameters such that the macroscopic results are equivalent to experiments. Such a trial-and-error method is time-consuming and impractical, but it is the only effective one.

Calibration was performed on four types of sandstones from Polish quarries. Sandstones were selected because they are a commonly used material. Sandstones are used in industrial applications, such as crushed stone for road construction, cladding material, block stone, paving stones and so on. Laboratory data of sandstones were taken from the work [15], where four types of sandstones with different lithology from different regions of Poland (Brenna, Mucharz, Radków, Tumlin) were analyzed in a laboratory and in a numerical way. Four types of sandstones that were the object of study in [15] were:

- Sandstone 1: monoliths of Godula sandstone; origin - from the active quarry in Brenna in the Silesian Beskid (in the Carpathians); age - Cretaceous (Cenoman); structure - coarse and medium-grained flysch sandstones with carbonate-chlorite cement and grey colour.

- Sandstone 2: monoliths of Krosno sandstone; origin - from the active quarry in Mucharz in the Little Beskid Mountains (in the Carpathians); age - Paleogen age (Eocen/Paleocen); structure - medium-grained flysch sandstones with carbonate-silty cement and grey colour.

- Sandstone 3: monoliths of joint sandstone; origin - from the active quarry in Radków in the Table Mountains (Sudetes); age - upper Cretaceous; structure - medium and coarse sandstones with silica-clay cement and yellow colour.

- Sandstone 4: monoliths of an elliptical, laminated sandstone (Bunter sandstone); origin - from the active quarry in Tumlin in the Świętokrzyskie Mountains; age - lower Triassic; structure - medium-grained sandstones with a silica-clay-iron cement and red colour.

To calibrate the DEM models to real sandstones, tuning of the Young's modulus  $E_b$ , and cohesion  $c_b$  microparameters was performed, as they are responsible for the macroscopic Young's modulus  $E$  and Unconfined Compressive Strength UCS, which allows for fitting the laboratory and numerical stress-strain curves. The procedure for  $E$  calibration was as follows. There is a linear relationship between the microscopic Young's modulus  $E_b$  and the macroscopic Young's modulus  $E$ . This means that if the other microparameters are kept constant, a linear increase in  $E_b$  will cause a linear increase in  $E$ . An arbitrary value of  $E_{b1}$  was selected, for which one uniaxial compression simulation was performed and the value of  $E_1$  was calculated. Depending on whether the value of  $E$  was greater or less than the expected  $E$ , value of  $E_{b2}$  was selected, and for this new value of  $E_b$ , another uniaxial compression simulation was performed and the value of  $E_2$  was calculated.  $E_{b1}$  and  $E_{b2}$  should be selected so that the expected value of  $E$  was between  $E_1$  and  $E_2$ . Using the linear relationship between  $E_b$  and  $E$ , according to the formula  $y=ax+b$ , the system of equations was obtained  $E_1=a*E_{b1}+b$  and  $E_2=a*E_{b2}+b$ , from which were obtained the values of parameters  $a$  and  $b$ , so the general equation of the line was known  $E=a*E_b+b$ . From this equation it was possible to calculate for which  $E_b$  the expected  $E$  would be obtained. This method gives results with very good accuracy. Then the obtained  $E_b$  value was used in further simulations of the given material and it was possible to proceed to calibrate the microscopic cohesion  $c_b$  and macroscopic UCS. Cohesion does not affect (or affects negligibly)  $E$ , so the same procedure was repeated as above for  $c_b$  and UCS.

For small external loading in uniaxial compression, the material is an ideal elastic body. When loads are larger, bonds start to break and particles can move from their initial positions. Material behaves with plasticity. Higher stress eventually leads to the appearance and development of cracks. As a result of fragmentation, particles without bonds appear. Friction interactions occur between particles without bonds, so the broken bond is identical to the fracture surface. Static and dynamic friction on the microscopic scale in the DEM model results from the Coulomb-type friction law [9].

### 2.3. Supervised Machine Learning – Random Forest

Random Forest (RF) is a supervised machine learning algorithm [16]. The basis of this algorithm is a forest of independent and uncorrelated decision trees, which together create the final prediction of the algorithm. RF works for both regression and classification problems. Grouping individual trees into a single forest allows for better prediction quality by overcoming the disadvantages of individual random trees, such as bias or overfitting.

RF is an example of ensemble learning - combining the capabilities of multiple models. In ensemble learning, a set of models is used to make predictions, instead of a single model. In the case of RF, the capabilities of multiple Decision Tree models are combined. Each Decision Tree works independently of the other trees, improving the quality of the algorithm. The ensemble technique in RF is called bagging (or Bootstrap Aggregation). In this technique, multiple bootstrap samples are created from the original set. Bagging begins with randomly selecting subsets from the original set. Such subsets are called Bootstrap Samples. Sampling with replacement is allowed. In this way, multiple trees from

the forest are trained on different datasets. This introduces variability into the training process and makes the model more robust. The sampling process is called row sampling. The process of sampling with replacement itself is called Bootstrapping. Each tree is trained independently on its Bootstrap Sample. The final predictions are made by averaging the individual predictions of all independent trees in the forest (for regression) or by majority vote – the most popular result is chosen (for classification). The process of collecting votes of independent trees and calculating the final result is called Aggregation. Additionally, RF uses feature randomness (also called feature bagging or the random space method) to generate random subsets of features, ensuring low correlation between decision trees [16].

The algorithm requires several conditions to be met in order to work effectively. The first one is Independence of Trees, which means that decision trees in the forest must be independent of each other, which is achieved by bootstrap sampling and feature randomness. Additionally, sufficient data is needed, because RF needs a large amount of data to build diverse trees and achieve optimal performance. In addition, Balanced Trees are required, which means that individual trees reach deep enough to capture patterns in the data. When it comes to Noisy Data Handling, RF is able to handle noisy data, as long as the noise is randomly distributed and not systematic [16].

RF has a slightly slower computational time compared to other algorithms due to the computation of data for each decision tree in the forest, and is slightly more difficult to interpret than a single decision tree. The advantage of RF is the reduced risk of overfitting. Single decision trees are prone to overfitting, but for a whole forest of uncorrelated trees, the averaged final result has lower variance and prediction error [16]. In this work RF was applied to the regression problem.

#### 2.4. Model Training and Testing

Some algorithms are sensitive to large differences between data and require scaling, e.g., standardization or normalization. Graphical models and tree-based models (like RF) are indifferent to feature scaling, so in this case, scaling is omitted. 80% of the dataset is used for training and 20% for testing. On the training dataset, the algorithm learns the dependencies between the data. On the testing dataset, the learning result of the algorithm is checked [5]. The results on the test set are quantitatively evaluated using evaluation metrics. Training and testing, together with K-Fold Cross Validation, help to avoid underfitting or overfitting. In K-Fold Cross Validation, the dataset is divided into K parts (folds), and K-1 parts are used for training and 1 for testing. In each subsequent iteration, a different fold is used for testing, and the remaining folds are used for training, so each fold is used exactly once for testing. Three regression-specific metrics were used for the evaluation: Coefficient of Determination  $R^2$  (Eq.2.1), Mean Square Error MSE (Eq.2.2), and Mean Absolute Error MAE (Eq.2.3) [5]:

$$R^2 = 1 - \frac{\sum(y_i - y_{pred})^2}{\sum(y_i - y_{mean})^2} \quad (2.1)$$

$$MAE = \frac{1}{N} \sum |y_i - y_{pred}| \quad (2.2)$$

$$MSE = \frac{1}{N} \sum (y_i - y_{pred})^2 \quad (2.3)$$

where  $N$  is the number of points in the dataset,  $y_i$  are the simulation results, and  $y_{pres}$  are the prediction results of the algorithm.

The result of the algorithm depends on its characteristic parameters, called hyperparameters. As a result of hyperparameter tuning, it is possible to achieve better algorithm performance by searching for the architecture that provides the best result [5]. A hyperparameter tuning technique called *GridSearchCV* was used. Three important parameters of the RF algorithm were selected for tuning: *n\_estimators* (the number of trees in the forest), *max\_depth* (maximum number of levels in a decision tree, which determines how deep the tree can grow during the learning process), *min\_samples\_leaf* (the minimum number of samples required to be at a leaf node).

The evaluation of the influence of input features (dependent parameters) on the output feature (dependent parameter), i.e., the result, was performed using SHapley Additive exPlanations (SHAP):

$$\Phi_i = \sum_{S \subseteq \{1, \dots, p\} \setminus \{i\}} \frac{|S|!(p-|S|-1)!}{p!} [val(S \cup \{i\}) - val(S)] \quad (2.4)$$

where  $S$  is a subset of the features used in the model,  $p$  is the number of features,  $val(S)$  is the prediction for feature values in set  $S$  that are marginalized over features that are not included in set  $S$ , and the SHAP value for feature value  $j$  is the value of the  $j$ -th feature contributed  $\phi_j$  to the prediction of this particular instance compared to the average prediction for the dataset [17].

### 3. RESULTS

#### 3.1. Numerical sandstones

To apply machine learning, a sufficiently large dataset is needed. To generate such a dataset, several series of simulations were performed, 10 simulations for each numerical type of sandstone (40 in total). Each simulation had slightly different parameters, but they were as far as possible within the range corresponding to the macroscopic parameters of real sandstones tested in the laboratory. As mentioned above, the laboratory parameters of sandstones were obtained from the work [15]. In accordance with the standards, a 4:1 ratio was maintained between the height of the cylinder and its radius. The cylinder height and radius were smaller than the values usually used in the laboratory, but in the work [18] it was shown that macroscopic values do not change significantly with the radius in DEM models.

This may seem to contradict common knowledge that the physical properties of rocks have a strong size effect. The size or scale effects in intact rocks refer to the change in the mechanical behaviour of an intact rock from small (laboratory) to large (engineering) scales. Different theories have been developed to explain size effect in intact rocks including statistical [19], fracture energy [20] and fractality [21]. One of the most complex research studies related to this matter was published by Hoek and Brown [22, 23]. Research on this topic was continued by other authors. Xiao et al. [24] checked the size effect on mechanical behaviors of intact granite on an enlarged digital model. Thuro et al. [25] evaluated the impact of shape and size on common rock properties. Durmeková et al. [26] performed standard tests on specimens of different sizes to check how the size and shape of specimens influence the measured strength of rocks. Zhao and Hu [27] investigated the size effect of the elastic modulus of rock with parallel joints using numerical simulation and regression analysis. Manuel et al. [28] extensively investigated rock strength and deformability under various loading conditions and different sizes.

However, the presented DEM models were able to simulate the mechanical behavior of the sample as a whole, but their internal structure was simplified in comparison with the reality. It did not reflect all the nuances of the structure of real rocks. Such simplifications were necessary to perform simulations

in real time. In addition, such simplifications meant that with a good approximation some macroscopic parameters of the DEM model practically did not change with the sample size. In a series of numerical experiments in [18], it was shown that cylindrical numerical DEM models of 5 mm x 20 mm behave approximately the same as DEM models of 18.75 mm x 75 mm. This means that the results of calculations on the DEM numerical model of 5 mm x 20 mm and the DEM numerical model of 18.75 mm x 75 mm would be almost identical. Therefore, the DEM numerical model of 5 mm x 20 mm could be compared with the real sample of 18.75 mm x 75 mm, which was used in this work. Such an assumption was necessary, because a generally known problem among DEM practitioners is the computational burden of this method (even for modern computers), as well as the huge amounts of data created. Therefore, scientific work using the DEM method even currently is limited to models with a small number of particles (a few or several dozen thousand), often 2D models are applied.

DEM modeling (uniaxial compression test in this case) of samples with the actual laboratory size of 18.75 mm x 75 mm is practically and technically possible only with the use of a supercomputer (as shown in [15]), and even then to a limited extent. In the simulations presented in this work, dozens of uniaxial compression tests were performed, and during each simulation, huge amounts of data were collected for the training of the machine learning algorithm. If 18.75 mm x 75 mm models were used, then in addition to the computational burden, it would also be necessary to collect not gigabytes, but terabytes of data for a single simulation. For technical and practical reasons this was simply impossible in the presented work.

### 3.1.1 Numerical Sandstone1

Sandstone1 was a cylinder with a radius of 5 mm and a height of 20 mm. There were 1643 particles in the cylinder with radii ranging from 0.25 mm to 2.0 mm. At the beginning of the simulation, all particles were connected by bonds. The simulation time step was 1.17068e-05 s, and the sample was compressed at a speed of 0.2 mm/s. The compression speed is higher than the speed usually used in the laboratory, but slow enough to maintain quasi-static conditions. The compression speed was increased gradually at the beginning to provide quasi-static loading without generating acoustic waves. The approximate computation time on a desktop computer for one simulation was 225 s. The visualization of the sample at the beginning of the simulation (Fig.1A), and just after failure (Fig.1C), is shown in the image below, as well as the view of the bonds between particles at the beginning of the simulation (Fig.1B), and just after failure (Fig.1D).

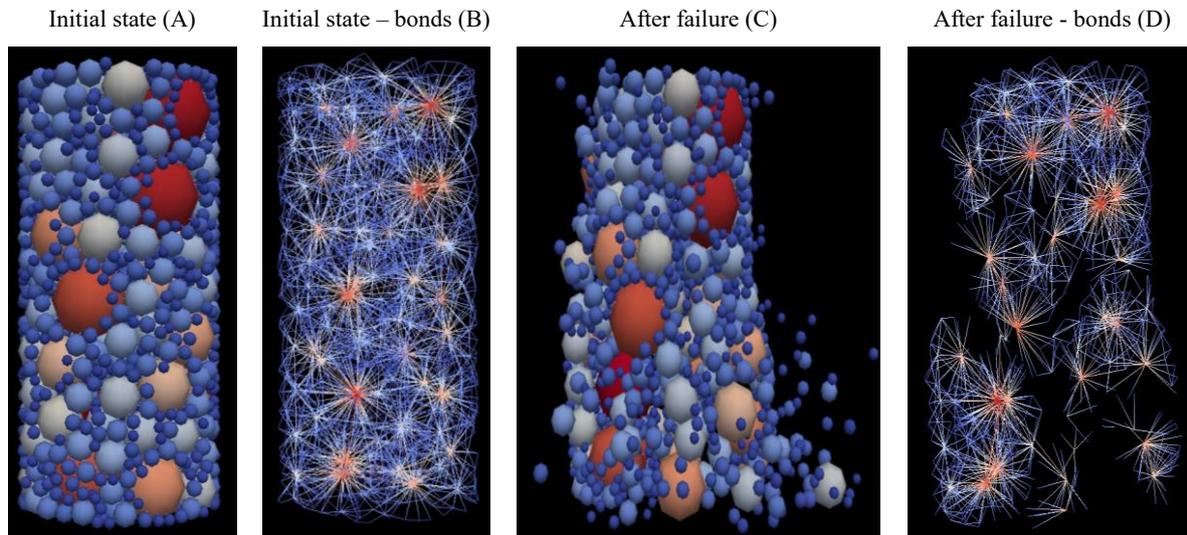


Fig. 1. Numerical Sandstone1 before and after failure – view of sample (A,C) and bonds (B,D)

As a result of the calibration process, microscopic bond parameters were identified such that the macroscopic sample parameters matched those observed in laboratory tests (Tab.1). The value of the microscopic parameter Poisson's ratio  $\nu_b$  was assumed: 0.25 [dimensionless] and for the tangent of internal friction angle  $\phi_b$ : 1.0 [dimensionless]. In order to generate an appropriate dataset for machine learning, 10 samples were generated whose macroscopic parameters were slightly different from each other, but still remained within the range of laboratory parameters for this type of sandstone. For this purpose, the microscopic parameters were changed: Young's modulus  $E_b$  in the range: 5790 - 5810 [MPa] and cohesion  $c_b$  in the range: 14.9 - 15.1 [MPa]. As a result, the dataset of generated 10 sandstones was characterized by the following macroscopic parameters: Young's modulus  $E$ : 20.94-21.14 [GPa] (in laboratory 20.2-21.8 [GPa]), Unconfined Compressive Strength UCS 97.44-103.13 [MPa] (in laboratory 99-107 [MPa]), Poisson's ratio  $\nu$  0.21-0.22 [dimensionless] (in laboratory 0.21-0.25 [dimensionless]), and critical strain  $\epsilon$  0.48-0.53 [%] (in laboratory  $\sim$  0.52 [%]). The particle density was constant and amounted to 2429 kg/m<sup>3</sup> (in laboratory 2391-2444 [kg/m<sup>3</sup>]).

Table 1. Macroscopic and microscopic parameters of numerical Sandstone1

Cylinder Geometry	Microscopic Bonds Parameters	Laboratory Macroscopic Parameters	Simulation Macroscopic Parameters
Cylinder height: 20 mm Cylinder radius: 5 mm Particles radii: 0.25 - 2.0 mm Particles density: 2429 kg/m <sup>3</sup> Initial number of particles: 1643 Time step: 1.17068e-05 s stroke rate: 0.2 mm/s computational time: $\sim$ 225 s	- Young's modulus $E_b$ : 5790 - 5810 [MPa] - Poisson's ratio $\nu_b$ : 0.25 [dimensionless] - cohesion $c_b$ : 14.9 – 15.1 [MPa] - tangent of internal friction angle $\phi_b$ : 1.0 [dimensionless]	E: 20.2-21.8 [GPa] UCS: 99-107 [MPa] $\nu$ : 0.21-0.25 [dimensionless] critical strain $\epsilon$ : $\sim$ 0.52 [%] d: 2391-2444 [kg/m <sup>3</sup> ]	E: 20.94-21.14 [GPa] UCS: 97.44-103.13 [MPa] $\nu$ : 0.21-0.22 critical strain $\epsilon$ : 0.48-0.53 [%]

The image below (Fig.2) shows an example stress-strain curve for one of the numerical Sandstones1.

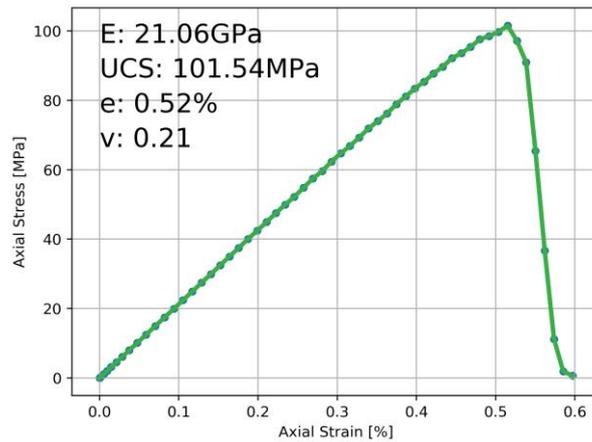


Fig. 2. Exemplary stress-strain curve for one of the Sandstones1

### 3.1.2 Numerical Sandstone2

Sandstone2 was a cylinder with a radius of 5 mm and a height of 20 mm. There were 5037 particles in the cylinder with radii from 0.25 mm to 0.5 mm. At the beginning of the simulation, all particles were connected by bonds. The simulation time step was  $7.33645e-06$  s, and the sample was compressed at a speed of 0.2 mm/s. The approximate computation time on a desktop computer for one simulation was 619 s. The visualization of the sample at the beginning of the simulation (Fig.3A), and just after failure (Fig.3C), is shown in the image below, as well as the view of the bonds between particles at the beginning of the simulation (Fig.3B), and just after failure (Fig.3D).

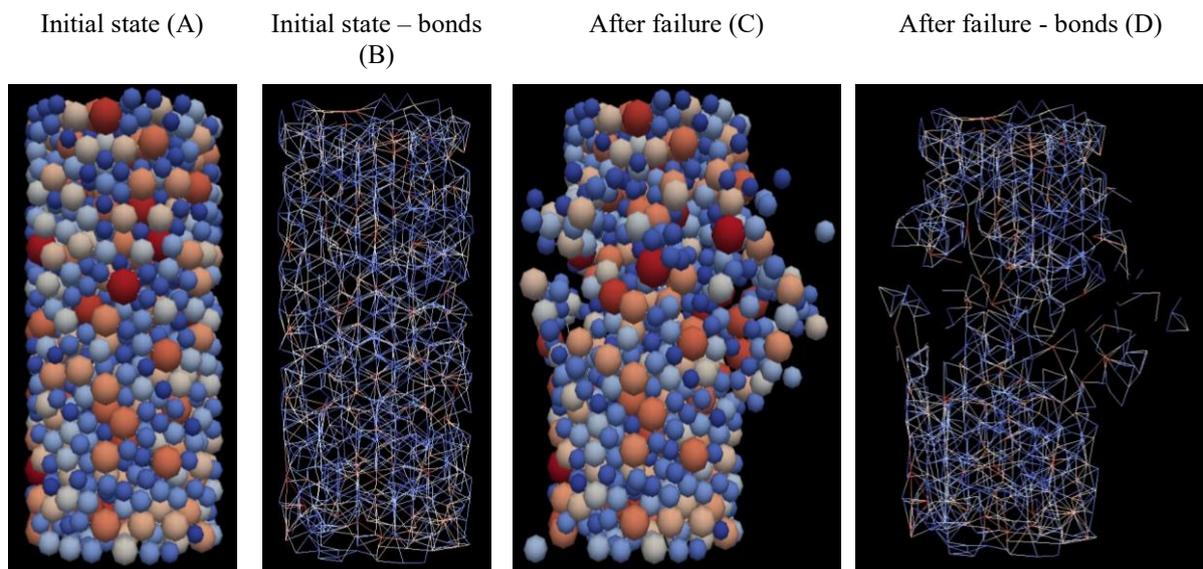


Fig. 3. Numerical Sandstone2 before and after failure – view of sample (A,C) and bonds (B,D)

As a result of the calibration process, it was possible to find microscopic bond parameters for which the macroscopic parameters of the sample correspond to the laboratory parameters (Tab.2). The value of the microscopic parameter Poisson’s ratio  $\nu_b$  was assumed: 0.25 [dimensionless] and for the tangent of internal friction angle  $\phi_b$ : 1.0 [dimensionless]. In order to generate an appropriate dataset for machine learning, 10 samples were generated whose macroscopic parameters were slightly different from each other, but still remained within the range of laboratory parameters for this type of sandstone. For this purpose, the microscopic parameters were changed: Young’s modulus  $E_b$  in the range: 64560-64580 [MPa] and cohesion  $c_b$  in the range: 299.9-300.1 [MPa]. As a result, the dataset of generated 10 sandstones was characterized by the following macroscopic parameters: Young’s modulus  $E$ : 28.90-29.09 [GPa] (in laboratory 26.6-31.8 [GPa]), Unconfined Compressive Strength UCS 143.83-149.93 [MPa] (in laboratory 138-153 [MPa] ), Poisson’s ratio  $\nu$  0.26 [dimensionless] (in laboratory 0.18-0.23 [dimensionless]), and critical strain  $\epsilon$  0.52-0.54 [%] (in laboratory  $\sim$  0.49 [%]). The particle density was constant and amounted to 2655 kg/m<sup>3</sup> (in laboratory 2627-2668 [kg/m<sup>3</sup>]).

Table 2. Macroscopic and microscopic parameters of numerical Sandstone2

Cylinder Geometry	Microscopic Bonds Parameters	Laboratory Macroscopic Parameters	Simulation Macroscopic Parameters
Cylinder height: 20 mm Cylinder radius: 5 mm Particles radii: 0.25 – 0.5 mm Particles density: 2655 kg/m <sup>3</sup> Initial number of particles: 5037 Time step: 7.33645e-06 s stroke rate: 0.2 mm/s computational time: 619 s	Young’s modulus $E_b$ : 64560-64580 [MPa] Poisson’s ratio $\nu_b$ : 0.25 [dimensionless] cohesion $c_b$ : 299.9-300.1 [MPa] tangent of internal friction angle $\phi_b$ : 1.0 [dimensionless]	E: 26.6-31.8 [GPa] UCS: 138-153 [MPa] $\nu$ : 0.18-0.23 $\epsilon$ : $\sim$ 0.49 [%] d: 2627-2668 [kg/m <sup>3</sup> ]	E: 28.90-29.09 [GPa] UCS: 143.83-149.93 [MPa] $\nu$ : 0.26 $\epsilon$ : 0.52-0.54 [%]

The image below (Fig.4) shows an example stress-strain curve for one of the numerical Sandstones2.

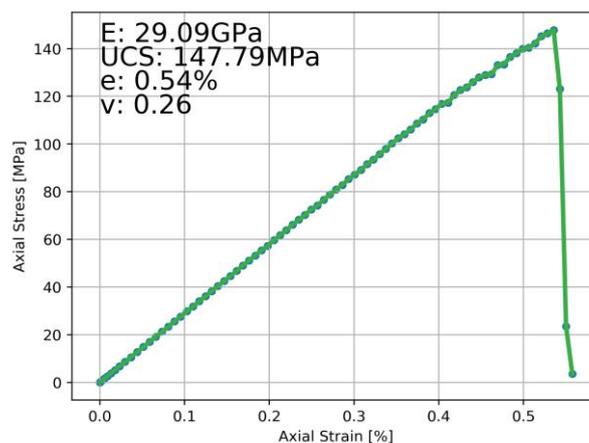


Fig. 4. Exemplary stress-strain curve for one of the Sandstones2

### 3.1.3 Numerical Sandstone3

Sandstone3 was a cylinder with a radius of 5 mm and a height of 20 mm. There were 2636 particles in the cylinder with radii from 0.25 mm to 1.0 mm. At the beginning of the simulation, all particles were connected by bonds. The simulation time step was  $1.05604e-05$  s, and the sample was compressed at a speed of 0.2 mm/s. The approximate computation time on a desktop computer for one simulation was 180 s. The visualization of the sample at the beginning of the simulation (Fig.5A), and just after failure (Fig.5C), is shown in the image below, as well as the view of the bonds between particles at the beginning of the simulation (Fig.5B), and just after failure (Fig.5D).

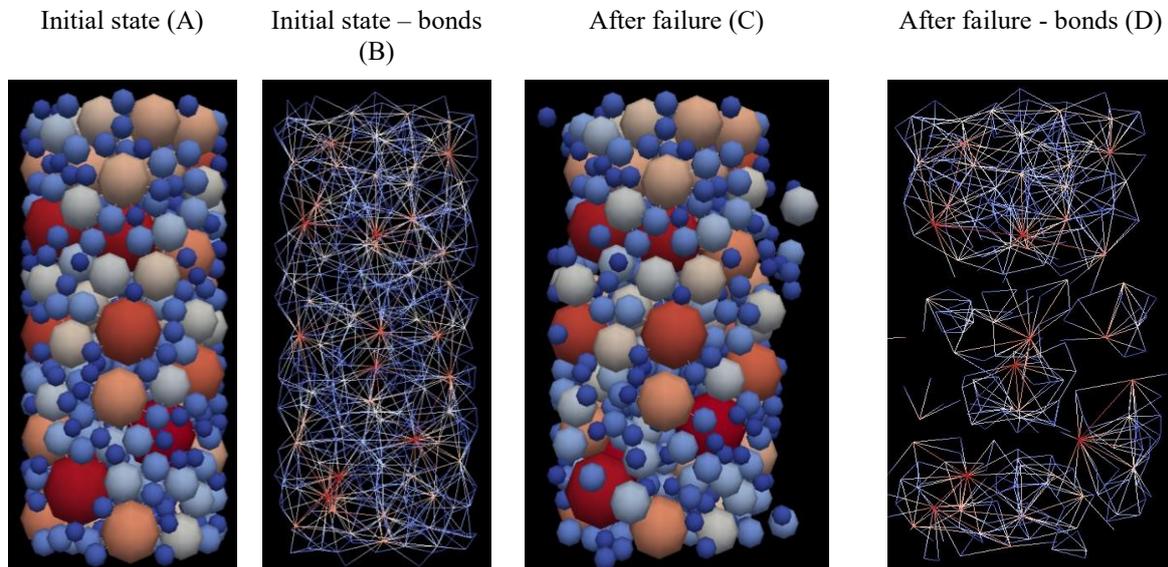


Fig. 5. Numerical Sandstone3 before and after failure – view of sample (A,C) and bonds (B,D)

As a result of the calibration process, microscopic bond parameters were found for which macroscopic sample parameters correspond to laboratory parameters (Tab.3). The value of the microscopic parameter Poisson's ratio  $\nu_b$  was assumed: 0.25 [dimensionless] and for the tangent of internal friction angle  $\phi_b$ : 1.0 [dimensionless]. In order to generate an appropriate data set for machine learning, 10 samples were generated whose macroscopic parameters were slightly different from each other, but still remained within the range of laboratory parameters for this type of sandstone. For this purpose, the microscopic parameters were changed: Young's modulus  $E_b$  in the range: 12561-12581 [MPa] and cohesion  $c_b$  in the range: 19.9-20.01 [MPa]. As a result, the dataset of generated 10 sandstones was characterized by the following macroscopic parameters: Young's modulus  $E$ : 23.46-23.58 [GPa] (in laboratory 22.5-24.9 [GPa]), Unconfined Compressive Strength UCS 56.93-59.10 [MPa] (in laboratory 55-60 [MPa]), Poisson's ratio  $\nu$  0.16 [dimensionless] (in laboratory 0.20-0.29 [dimensionless]), and critical strain  $\epsilon$  0.25-0.26 [%] (in laboratory  $\sim$  0.29 [%]). The particle density was constant and amounted to 2142 kg/m<sup>3</sup> (in laboratory 2119-2172 [kg/m<sup>3</sup>]).

Table 3. Macroscopic and microscopic parameters of numerical Sandstone 3

Cylinder Geometry	Microscopic Bonds Parameters	Laboratory Macroscopic Parameters	Simulation Macroscopic Parameters
Cylinder height: 20 mm Cylinder radius: 5 mm Particles radii: 0.25 - 1.0 mm Particles density: 2142 kg/m <sup>3</sup> Initial number of particles: 2636 Time step: 1.05604e-05 s stroke rate: 0.2 mm/s computational time: 180 s	Young's modulus $E_b$ : 12561-12581 [MPa] Poisson's ratio $\nu_b$ : 0.25 [dimensionless] cohesion $c_b$ : 19.9-20.01 [MPa] tangent of internal friction angle $\phi_b$ : 1.0 [dimensionless]	E: 22.5-24.9 [Gpa] UCS: 55-60 [MPa] $\nu$ : 0.20-0.29 $e$ : ~ 0.29 [%] $d$ : 2119-2172 [kg/m <sup>3</sup> ]	E: 23.46-23.58 [GPa] UCS: 56.93-59.10 [MPa] $\nu$ : 0.16 $e$ : 0.25-0.26 [%]

The image below (Fig.6) shows an example stress-strain curve for one of the numerical Sandstones3.

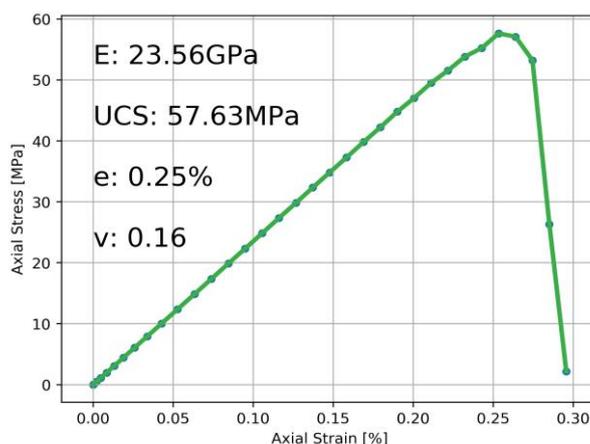


Fig. 6. Exemplary stress-strain curve for one of the Sandstones3

### 3.1.4 Numerical Sandstone4

Sandstone 4 was a cylinder with a radius of 3 mm and a height of 12 mm. In the case of sandstone 4, a smaller cylinder size was used because this sandstone consisted of the smallest particles. This resulted in an irrational increase in computational time, so it was decided to use a smaller cylinder. There were 4394 particles in the cylinder with radii from 0.125 mm to 0.5 mm. At the beginning of the simulation, all particles were connected by bonds. The simulation time step was 5.13344e-06 s, and the sample was compressed at a speed of 0.2 mm/s. The approximate computational time on a desktop computer for one simulation was 698 s. The visualization of the sample at the beginning of the simulation (Fig.7A), and just after failure (Fig.7C), is shown in the image below, as well as the view of the bonds between particles at the beginning of the simulation (Fig.7B), and just after failure (Fig.7D).

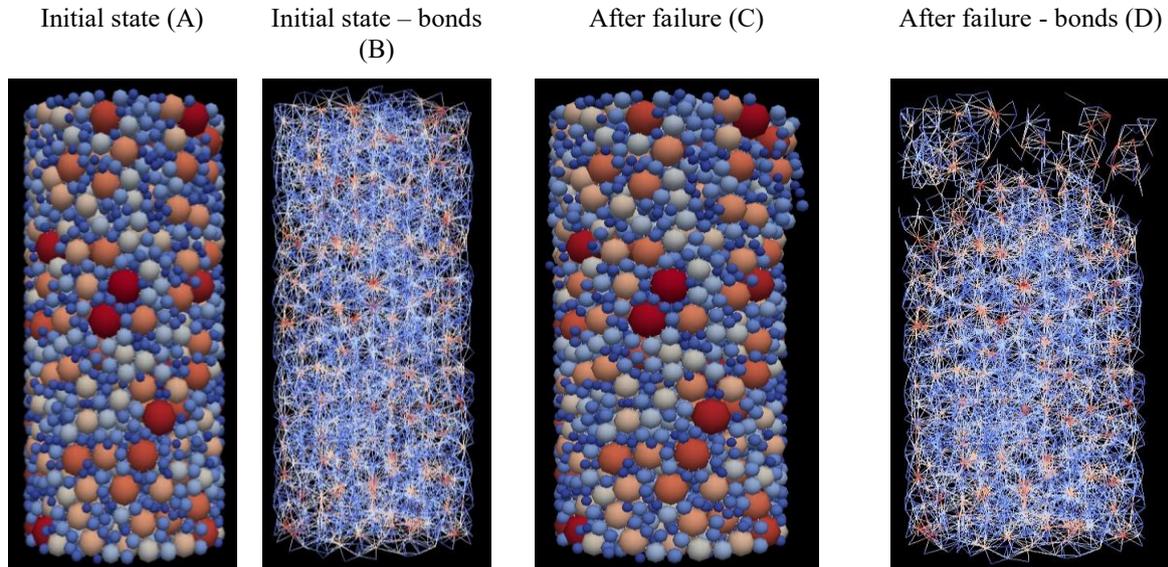


Fig. 7. Numerical Sandstone4 before and after failure – view of sample (A,C) and bonds (B,D)

As a result of the calibration process, it was possible to find microscopic bond parameters for which the macroscopic parameters of the sample correspond to the laboratory parameters (Tab.4). The value of the microscopic parameter Poisson's ratio  $\nu_b$  was assumed: 0.25 [dimensionless] and for the tangent of internal friction angle  $\phi_b$ : 1.0 [dimensionless]. In order to generate an appropriate dataset for machine learning, 10 samples were generated whose macroscopic parameters were slightly different from each other, but still remained within the range of laboratory parameters for this type of sandstone. For this purpose, the microscopic parameters were changed: Young's modulus  $E_b$  in the range: 15190-15210 [MPa] and cohesion  $c_b$  in the range: 57.9-58.1 [MPa]. As a result, the dataset of generated 10 sandstones was characterized by the following macroscopic parameters: Young's modulus  $E$ : 27.05-27.28 [GPa] (in laboratory 26.2-29.4 [GPa]), Unconfined Compressive Strength UCS 118.78-123.86 [MPa] (in laboratory 115-128 [MPa]), Poisson's ratio  $\nu$  0.17- 0.19 [dimensionless] (in laboratory 0.22-0.28 [dimensionless]), and critical strain  $\epsilon$  0.44-0.50 [%] (in laboratory  $\sim$  0.51 [%]). The particle density was constant and amounted to 2448 kg/m<sup>3</sup> (in laboratory 2355-2480 [kg/m<sup>3</sup>]).

Table 4. Macroscopic and microscopic parameters of numerical Sandstone4

Cylinder Geometry	Microscopic Bonds Parameters	Laboratory Macroscopic Parameters	Simulation Macroscopic Parameters
Cylinder height: 12 mm Cylinder radius: 3 mm Particles radii: 0.125 – 0.5 mm Particles density: 2448 kg/m <sup>3</sup> Initial number of particles: 4394 Time step: 5.13344e-06 s stroke rate: 0.2 mm/s computational time: 698 s	Young's modulus $E_b$ : 15190-15210 [MPa] Poisson's ratio $\nu_b$ : 0.25 [dimensionless] cohesion $c_b$ : 57.9-58.1 [MPa] tangent of internal friction angle $\phi_b$ : 1.0 [dimensionless]	E: 26.2-29.4 [Gpa] UCS: 115-128 [MPa] $\nu$ : 0.22-0.28 $\epsilon$ : $\sim$ 0.51 [%] d: 2355-2480 [kg/m <sup>3</sup> ]	E: 27.05-27.28 [Gpa] UCS: 118.78-123.86 [MPa] $\nu$ : 0.17-0.19 $\epsilon$ : 0.44-0.50 [%]

The image below (Fig.8) shows an example stress-strain curve for one of the numerical Sandstones4.

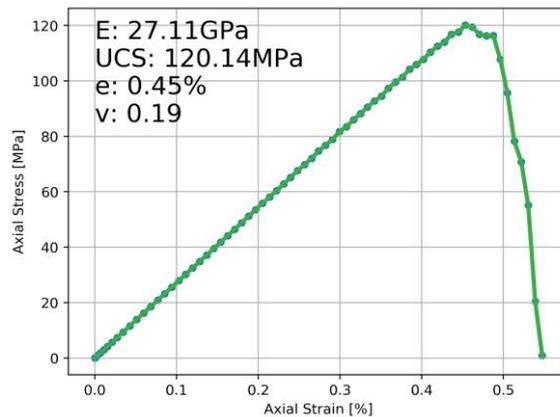


Fig. 8. Exemplary stress-strain curve for one of the Sandstones1

### 3.2 Moving time window

During each of the 40 simulations, data useful for the machine learning algorithm were collected. For this purpose, the concept of a moving time window (MTW) was used (Fig.9). The moving time window is a technique known in the literature. The moving time window consists of recording selected parameters describing the state of the system at selected moments of the analyzed process (checkpoints). For example, it was used by [2] in a laboratory experiment during stick-slip cycles to continuously collect acoustic emission data for the machine learning algorithm. The trained algorithm then predicted the time to failure with good accuracy based on the acoustic signal. The moving time window was also used by [29] in a numerical DEM experiment of stick-slip cycles, also to generate a dataset for the machine learning algorithm. Here, (Pseudo) Acoustic Emission based on continuous monitoring of the particle velocity signal was used. This definition of (Pseudo) Acoustic Emission was used for example, by [1] in DEM modeling.

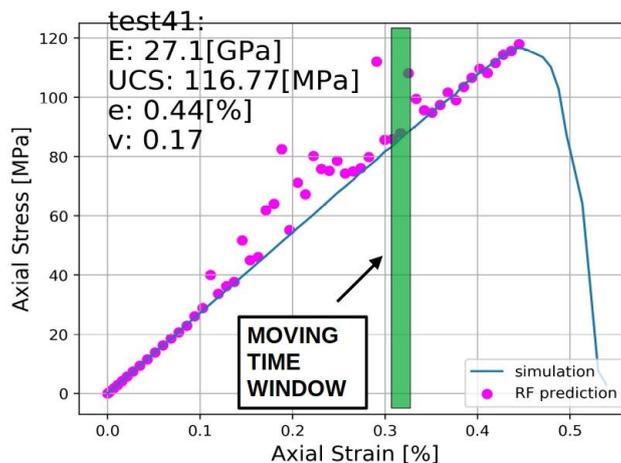


Fig. 9. Schematic representation of the Moving Time Window

Referring to the definition of PAE used in [1] for DEM simulations, in this work, MTW was used to collect 7 features approximately describing PAE at selected time steps (so-called checkpoints – every 1000 time steps). The following features describing the mean particle state were used:

- mean velocity in x direction:  $vx\_mean$
- standard deviation of the mean velocity in the x direction:  $vx\_std$
- mean velocity in y direction:  $vy\_mean$
- standard deviation of the mean velocity in the y direction:  $vy\_std$
- mean velocity in z direction:  $vz\_mean$
- standard deviation of the mean velocity in the z direction:  $vz\_std$
- ratio of broken bonds:  $bonds$

The above features were treated as independent variables in the language of machine learning. The dependent variables were AS and TtF (where failure is understood as the time of the main crack). The data obtained in this way during each of the 40 simulations were used to generate one dataset used to train and test the RF algorithm. Training and testing were performed for each of the two dependent variables separately.

### 3.3 Algorithm training and testing

The graph below presents the results of the RF algorithm test set. This is a comparison of the RF algorithm predictions with the actual values (numerical simulation results). For the dependent variable AS (Fig.10A), the metrics R2 were obtained: 84%+/-6.0%; MEA: 9.87+/-2.5; MSE: 246.95+/-103.24. Whereas for the dependent variable TtF (Fig.10B), the metrics R2 were obtained: 73%+/-13.0%; MEA: 0.05+/-0.02; MSE: 0.01+/-0.0.

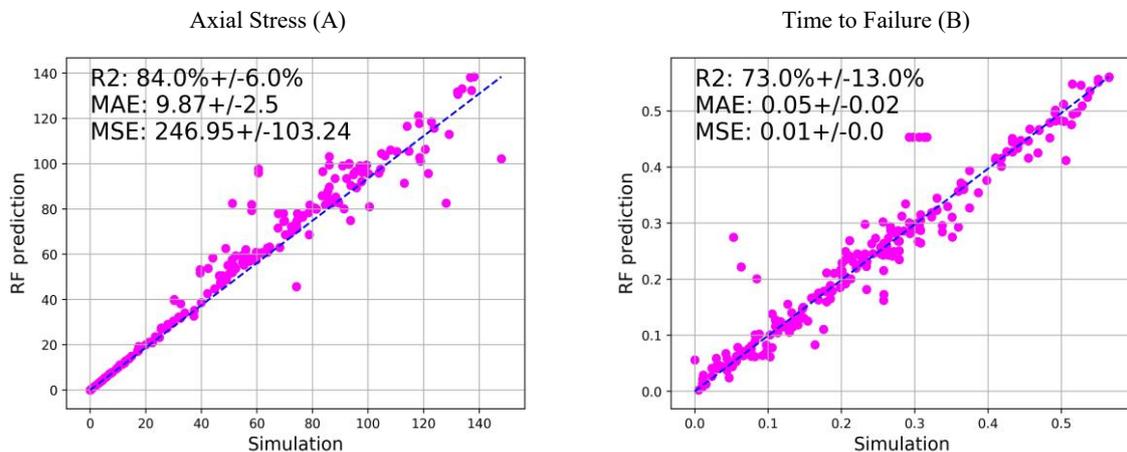


Fig. 10. Algorithm performance for predictions of the Axial Stress (A) and Time to Failure (B) on the test dataset

### 3.4 Evaluation

The RF algorithm took 7 different features as input to its predictions, but it is not known what the exact contribution of each feature is to the algorithm's results. To determine this, SHapley Additive exPlanations (SHAP) can be used. The first row of the figure shows the average SHAP value for each of the input features. For AS (Fig.11A), the highest contribution to prediction is made by features related to the standard deviation of the mean velocity value in the y, x, and z directions, respectively. The smallest, zero contribution comes from the ratio of broken bonds. In the case of TtF (Fig.11B), the highest contribution was given by the feature of the standard deviation of the mean value in the x direction, the subsequent features have a much lower contribution, and the ratio of broken bonds has a zero contribution, similarly to the previous dependent variable.

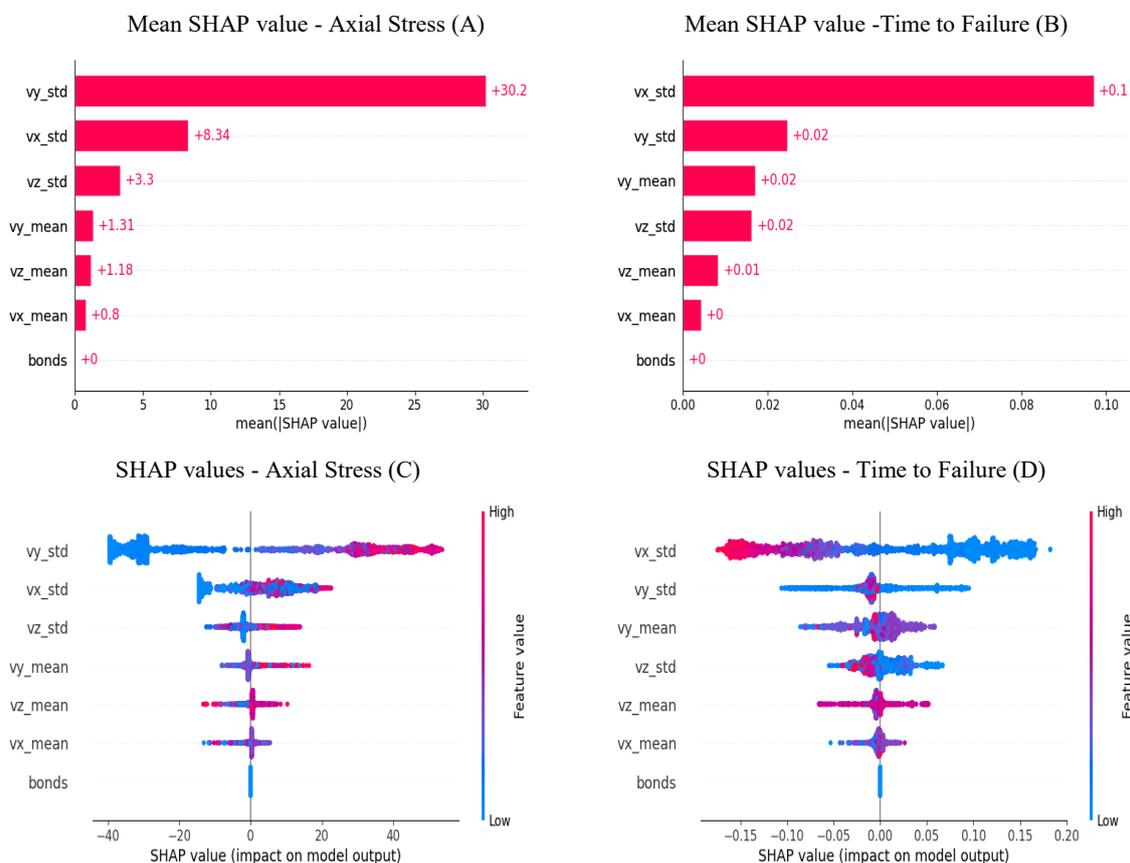


Fig. 11. Mean SHAP values of the algorithm for Axial Stress (A) and Time to Failure (B) and individual SHAP values of the algorithm for Axial Stress (C) and Time to Failure (D)

The second row of the figure shows the SHAP calculated for individual data points for each feature. If (within a given feature) data points have high values, they are marked in a warm color. Points with low values are marked in a cool color. The individual SHAP value of each point results from the position of that point on the x-axis. The order of the features presented in the second row is not random, it is the same as in the first row, because the features are arranged in accordance with their influence on the result of the algorithm. The greatest influence on the RF algorithm had *vy\_std* for the AS (Fig.11C) prediction and *vx\_std* for TtF (Fig.11D), the average SHAP value for these features was several times higher than for the other features. As for the distribution of individual points along the x-axis for *vy\_std*, points with low values had a negative effect on the prediction, and points with high values had a positive effect on the prediction. In the case of *vx\_std*, it was exactly the opposite. In the case of the remaining features, it is difficult to discern clear relationships.

### 3.5 Model Verification on New Data

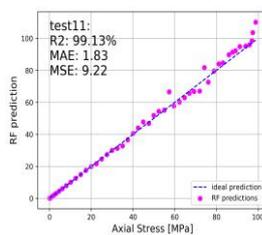
In the previous sections, it was shown that the RF algorithm was able to predict AS and TtF on the training and testing datasets based only on the PAE in the numerical model. The metrics of these predictions were good or very good. However, it is known that the machine learning algorithm can be overtrained and show good efficiency on the training and testing datasets, but fail on new data. In order

to verify the performance of the previously prepared model, several series of new simulations were prepared. These were simulations with parameters different from those previously seen by the algorithm. Based on the sandstone models prepared earlier, 8 new uniaxial compression tests were performed. For sandstone 1, tests test11 and test12 were performed, for sandstone 2, tests test21 and test22, for sandstone 3, tests test31 and test32, and for sandstone 4, tests test41 and test42. Tests test11, test21, test31, test41 were characterized by a large similarity of parameters to the previously performed simulations, but they were not the same. However, for tests test12, test22, test32, test42, one of the microparameters was modified so that one of the macroparameters significantly differed from the simulations performed earlier. For these 8 new simulations, training and testing of the algorithm were not performed, but the RF model trained and tested on 40 previous simulations was used. In this way, it was checked how the ready model would cope with AS and TtF predictions on completely new data.

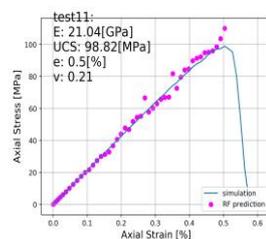
### 3.5.1 Axial Stress predictions

The scheme of the figure below looks as follows for AS predictions. In each row, two tests performed for each of the numerical sandstones are placed. The first test was performed in a slightly modified version. The slightly modified tests were test11 (Fig. 12A, B), test21 (Fig. 12E, F), test 31 (Fig. 12I, J), and test41 (Fig. 12M, N). The second test was performed in a significantly modified version. The significantly modified tests were test12 (Fig. 12C, D), test22 (Fig. 12G, H), test 32 (Fig. 12K, L), and test42 (Fig. 12O, P). As for the columns, the first and third columns present a comparison between the actual values (on the x-axis), i.e., those derived from the numerical simulations, and the predictions of the RF model (y-axis). The dashed line indicates the line of ideal prediction. In turn, the second and fourth columns show the AS predictions performed against the stress-strain curve for a given numerical experiment.

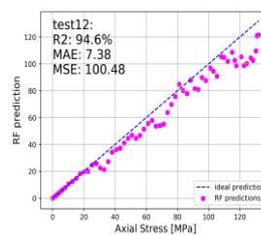
Numerical vs Predicted Axial Stress - Test11 (A)



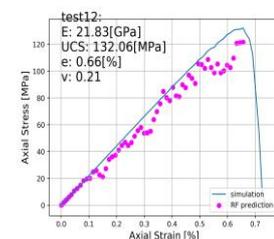
Predictions vs Stress-Strain Curve - Test11 (B)



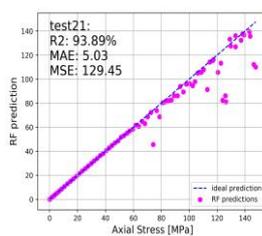
Numerical vs Predicted Axial Stress - Test12 (C)



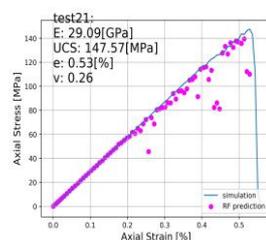
Predictions vs Stress-Strain Curve - Test12 (D)



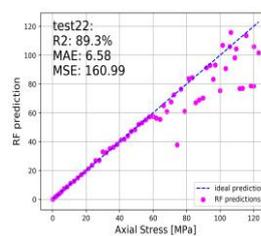
Numerical vs Predicted Axial Stress - Test21 (E)



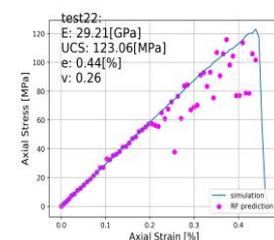
Predictions vs Stress-Strain Curve - Test21 (F)



Numerical vs Predicted Axial Stress - Test22 (G)



Predictions vs Stress-Strain Curve - Test22 (H)



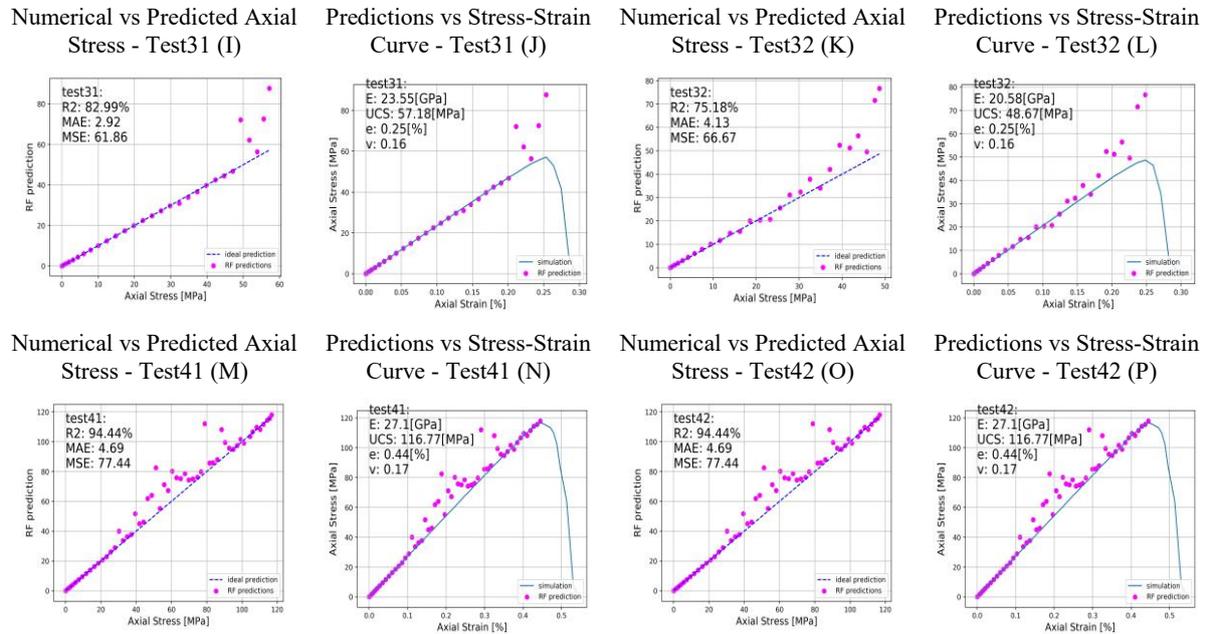


Fig. 12. Mean SHAP values of the algorithm for Axial Stress (A) and Time to Failure (B) and individual SHAP values of the algorithm for Axial Stress (C) and Time to Failure (D)

Using the R2 metric as an example, the results of AS predictions can be traced. For tests test11, test21, test31, and test41 (i.e., for slightly modified numerical sandstones), very good R2 statistics were obtained, 99.13%, 93.89%, 82.99%, and 94.44%, respectively. The second column of graphs shows that AS predictions with good accuracy correspond to the stress-strain curve until failure. For tests test12, test22, test32, and test42 (i.e., for significantly modified numerical sandstones), slightly worse, but still satisfactory, R2 statistics results were obtained - 94.6%, 89.3%, 75.18%, and 94.44%, respectively. The visualization in the fourth column shows that AS predictions in this case deviate slightly more from the shape of the curve, but still reproduce it with a good approximation.

### 3.5.2 Time to Failure predictions

The scheme of the figure below looks as follows for the TtF prediction. In each row, two tests performed for each of the numerical sandstones are placed. The first test was performed in a slightly modified version. The slightly modified tests were test11 (Fig. 13A, B), test21 (Fig. 13E, F), test31 (Fig. 13I, J), test41 (Fig. 13M, N). The second test was performed in a significantly modified version. The significantly modified tests were test12 (Fig. 13C, D), test22 (Fig. 13G, H), test32 (Fig. 13K, L), test42 (Fig. 13O, P). As for the columns, the first and third columns present a comparison between the actual values (on the x-axis), i.e., those derived from the numerical simulations, and the predictions of the RF model (y-axis). The dashed line indicates the line of ideal prediction. In turn, the second and fourth columns show the performed TtF predictions against the background of the time to failure in the axial strain function for a given numerical experiment.

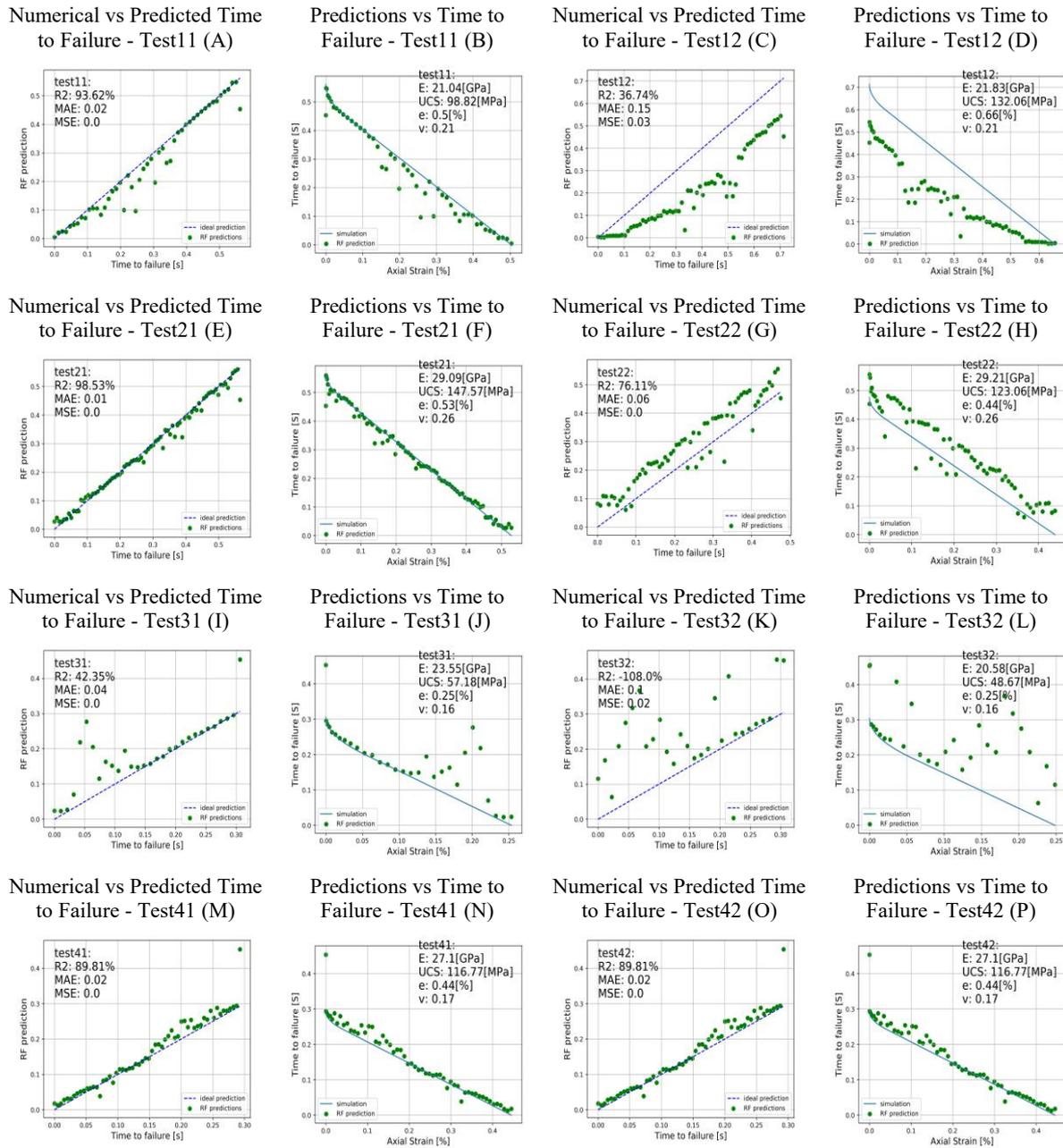


Fig. 13. Mean SHAP values of the algorithm for Axial Stress (A) and Time to Failure (B) and individual SHAP values of the algorithm for Axial Stress (C) and Time to Failure (D)

Using the R2 metric as an example, the results of TtF predictions can be followed. For tests test11, test21, test31, and test41 (i.e. for slightly modified numerical sandstones), the R2 statistics were obtained at 93.62%, 98.52%, 42.35%, and 89.81%, respectively. Apart from test31, these are very good results. The second column of graphs shows that TtF predictions are made with good accuracy. For tests test12, test22, test32, test42 (i.e., for significantly modified numerical sandstones), the results of the R2

statistics were obtained at 36.74%, 76.11%, -108%, and 89.81%, respectively. The results are very diverse. In some cases the model performed well, in others it did not perform at all (test32) or very poorly (test12). The visualization in column four shows that the TtF predictions for poor R2 results are chaotic and unrelated to the shape of the curve.

In general, this shows that the prepared RF model performs very well on samples with characteristics similar to those used in the training and testing process. This was also the case for heavily modified samples for AS prediction. In the case of TtF, a larger range of data used for training and testing the model is needed.

#### **4. CONCLUSIONS**

In this work, (pseudo) acoustic emission was used to continuously monitor the uniaxial compression process of four numerical sandstones. The numerical sandstones were obtained using DEM modeling. By tuning selected microscopic parameters in the numerical models, a calibration process was carried out, i.e., the macroscopic parameters of the numerical models were matched to the macroscopic parameters of the real sandstones.

Several series of tests were carried out for each of the numerical sandstones. During each simulation, data on selected model parameters were collected, which were a numerical approximation of the real acoustic emission. The collected dataset was used to train the Random Forest machine learning algorithm. The algorithm was applied to predict the Time to Failure and Axial Stress at any moment of the duration of the numerical uniaxial compression test. The algorithm was verified by calculating evaluation metrics during training the algorithm and during its testing. Very good accuracy and efficiency of the algorithm were achieved. The ready and trained algorithm was applied to several series of new tests, for four samples similar to those used during training, and for four significantly modified samples. The evaluation metrics during the verification showed that the algorithm performed well or very well on new but similar samples, while for new, heavily modified samples, the results were very diverse. In general, better results were achieved for the prediction of Axial Stress than Time to Failure.

To sum up, the most important results of this work were as follows:

- it was shown that it is possible to predict (in numerical modeling) Axial Stress and Time to Failure at any time during a uniaxial compression test, based only on pseudo-acoustic emission, with good or very good accuracy,
- the use of a wider set of samples during training of the algorithm would allow predictions for a wider range of materials,
- as part of this work, a methodology was prepared, which in the next step can be used for automatic prediction of failure based on the acoustic signal in real laboratory tests,
- this work is part of the research for new ways of determining the moment of failure in a material, in an automated way, based on acoustic emission.

#### **ADDITIONAL INFORMATION**

##### **Acknowledgements**

The author would like to thank for the support from the statutory funds of the Institute of Geophysics, Polish Academy of Sciences. The author would also like to thank all the authors of ESyS-Particle Tutorial [7]. The laboratory work from [15] was performed in the Laboratory of Applied Geology,

Faculty of Geology, University of Warsaw, with dr hab. Paweł Łukaszewski and mgr Robert Dziejczak. The author would like to thank the Anonymous Reviewers. The author would like to thank the Scientific Information and Publishers Department of IG PAS for the language correction.

### Computer Code

Numerical simulations were performed using ESyS-Particle, GenGeo and Paraview. Analysis and visualization of results and application of the machine learning algorithm were performed using Python3 and its libraries (Matplotlib, Seaborn, Pandas, Scikit-learn and others).

### REFERENCES

1. Dorostkar, O and Carmeliet, J 2019. Grain friction controls characteristics of seismic cycle in faults with granular gouge. *Journal of Geophysical Research: Solid Earth* **124**, 6475–6489.
2. Rouet-Leduc, B, Hulbert, C, Lubbers, N, Barros, K, Humphreys, CJ and Johnson, PA 2017. Machine learning predicts laboratory earthquakes. *Geophysical Research Letters* **44**, 9276–9282.
3. Bolton, DC, Shreedharan, S, Rivière, J and Marone, C 2020. Acoustic energy release during the laboratory seismic cycle: Insights on laboratory earthquake precursors and prediction. *Journal of Geophysical Research: Solid Earth* **125**, e2019JB018975.
4. Rucka, M, Knak, M and Nitka, M 2023. A study on microcrack monitoring in concrete: discrete element method simulations of acoustic emission for non-destructive diagnostics. *Engineering Fracture Mechanics* **293**, 109718.
5. Géron, A 2019. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd Edition, O'Reilly Media, Inc.
6. Cundall, PA and Strack, ODL 1979. A discrete numerical model for granular assemblies. *Geotechnique* **29**, 1, 47–65.
7. Abe, S, Boros, V, Hancock, W and Weatherley, D 2014. *ESyS-particle tutorial and user's guide. Version 2.3.1*, <https://launchpad.net/esys-particle>.
8. Weatherley, DK, Boros, VE, Hancock, WR and Abe, S 2010. *Scaling benchmark of ESyS-Particle for elastic wave propagation simulations*, 2010 IEEE Sixth Int. Conf. E-Science, Brisbane, Australia, 277–283.
9. O'Sullivan, C 2011. *Particulate Discrete Element Modelling: A Geomechanics Perspective*, Spon Press/Taylor and Francis, London.
10. Wang, Y, Abe, S, Latham, S and Mora, P 2006. Implementation of Particle-scale Rotation in the 3-D Lattice Solid Model, *Pure and Applied Geophysics* **163**, 1769-1785.
11. Wang, YC 2009. A new algorithm to model the dynamics of 3-D bonded rigid bodies with rotations. *Acta Geotechnica* **4**, 2, 117-127.
12. Mora, P and Place, D 2002. Stress correlation function evolution in lattice solid elasto-dynamic models of shear and fracture zones and earthquake prediction. *Pure and Applied Geophysics* **159(10)**, 2413-2427.
13. Sochor, M 1998. *Strength of Materials I*, CTU Publishing House, Prague.
14. Łukaszewski, P 2003. Development of fracture processes in Silesian Carboniferous sandstones. *Geological Quarterly* **47**, (1), 29-38.
15. Klejment, P, Dziejczak, R, Łukaszewski, P 2021. Strength of industrial sandstones modelled with the Discrete Element Method. *Studia Geotechnica et Mechanica* **43(4)**, 346–365.
16. Breiman, L 2001. Random Forests. *Machine Learning* **45**, 5–32.

17. Strumbelj, E and Kononenko, I 2014. Explaining prediction models and individual predictions with feature contributions. *Knowledge and information systems* **41.3**, 647-665.
18. Klejment, P 2021. Application of supervised machine learning as a method for identifying DEM contact law parameters, *Mathematical Biosciences and Engineering* **18(6)**, 7490–7505.
19. Zhao, T and Hu, G 2023. Study on the size effect of elastic modulus of rock considering the joint spacing. *Frontiers in Materials* **10**, 1142683.
20. González-Fernández, MA, Estévez-Ventosa, X, Pérez-Rey, I, Alejano, LR, Masoumi, H 2024. Size effects on strength and deformability of artificially jointed hard rock. *International Journal of Rock Mechanics and Mining Sciences* **176**, 105696, ISSN 1365-1609.
21. Weibull, W 1939. *A Statistical Theory of the Strength of Materials*, 1–45, Stockholm, Sweden: Royal Swedish Academy of Engineering Science, 1939.
22. Bazant, Z 1997. Scaling of quasibrittle fracture: hypotheses of invasive and lacunar fractality, their critique and Weibull connection. *International Journal of Fracture* **83**, 41–65.
23. Carpinteri, A, Chiaia, B, Ferro, G 1995. Size effects on nominal tensile strength of concrete structures: multifractality of material ligaments and dimensional transition from order to disorder. *Materials and Structures* **28(6)**, 311–317.
24. Xiao, H, He, L, Shi, R, Song, J, Tang, Y, Zhang, Q and Shi, P 2021. Size Effect on Mechanical Behaviors of Intact Rocks Based on Enlarged Digital Models. *Frontiers in Earth Science* **9**, 803406.
25. Thuro, K, Plinninger, RJ, Zah, S, and Schütz, S 2001. Scale Effect in Rock Strength Properties. Part 1: Unconfined Compressive Test and Brazilian Test. *Rock Mechanics—A Challenge for Society*, Sarkka et Eloranta, Eds. Swets et Zeitinger Lisse.
26. Tatiana, D, Bednarik, M, Dikejová, P, Adamcová, R 2022. Influence of specimen size and shape on the uniaxial compressive strength values of selected Western Carpathians rocks, *Environmental Earth Sciences* **81**, 247.
27. Hoek, E, Brown, ET 1980. *Underground excavations in rock*. The Institution of Mining and Metallurgy, Stephen Austin and Sons Ltd., Hertford, London, p. 527.
28. Hoek E, Brown, ET 1997. Practical estimates of rock mass strength. *International Journal of Rock Mechanics and Mining Sciences* **34(8)**, 1165–1186.
29. Ren, CX, Dorostkar, O, Rouet-Leduc, B, Hulbert, C, Strebel, D, Guyer, RA et al. 2019. Machine learning reveals the state of intermittent frictional dynamics in a sheared granular fault. *Geophysical Research Letters* **46**, 7395–7403.