# Micro-mechanics and machine learning based approach for the prediction of tensile strength of cementitious materials

**Jinane MURR, Syed Yasir ALAM and Frederic GRONDIN**

*Nantes Université, Ecole Centrale Nantes, CNRS, GeM, UMR 6183, F-44000 Nantes, France*
*Corresponding author: Jinane MURR (e-mail: jinane.murr@ec-nantes.fr)*

**ABSTRACT** Optimization of the carbon footprint of cementitious materials remains insufficient if the effects on their mechanical properties are not considered. Multi-criteria optimization is required. This demands knowledge of the correlations between these mechanical properties and the composition of the material. In this context, it is possible to use artificial intelligence (AI) methods to predict the mechanical properties from the chemical components of the material. In this work, the mechanical property considered is the tensile strength at the scale of ordinary cement paste. The proposed methodology is a a coupling of a hydration model with a micro-mechanical computational model to feed data into an AI regression algorithm. This coupling will serve to fill the lack of data in the literature on the relationship between the mechanical behaviour of hardened cement paste and its chemical composition. This methodology will then be extended to characterize cement pastes with mineral substitutions and to account for other elastic and inelastic properties.

**Keywords** Cement paste, neural networks, tensile strength, micro-mechanical modeling

## I. Introduction

The prediction of the degradation of concrete structures is of great interest for safety as it concerns buildings, dams, bridges, etc. Tensile strength is an important property to consider for the performance of a cementitious material. However, compressive strength and Young modulus are considered the main performance indicators for concrete in most standards such as Eurocode2 and ACI 318. For applications where knowledge of the tensile strength is required, the standards propose formulas to derive it from the compressive strength of concrete, as in ACI 363R-92 (Russell & al. (1998)), ACI 318-99 (Wallace & Orakcal (2002)), and Eurocode EN 1992-1-1. Therefore, most of the models found in the literature are aimed at calculating the compressive strength of cementitious materials. There are emperical formulas with statisticly fitted coefficients, such as Powers (1958) and Nielsen (1993). These models take into account the dependence of compressive strength on the microstructural composition of the hardened cement paste, more specifically, on solid to pores ratio. In his model, Powers demonstrated the dependence of the compressive strength of a cement paste on the gel-space ratio of the cement paste. However, the Power's model does not take into account the actual proportions of the phases in the gel fraction of the cement paste, so it is limited in terms of blended cement pastes (Gartner & Hirao (2015)). Moreover, the relationship between the compressive strength and the tensile strength of concrete is not as simple as described in the empirical formulas. For a more reliable estimate, other parameters must be considered, such as the chemical composition of the cement paste, curing, and water/binder ratio. The importance of a better estimate of the tensile streng of cement paste is related to the role it plays in predicting the cracking behaviour of the cement paste. is well known that cracking affects the durability of the cement paste, and high compressive strength not an indicator of good fracture behaviour. Finding the most important parameters related to the fracture properties of cement pastes is a complex problem due to synergy between these parameters. Therefore, an

approach for the computation of the damage properties of cement pastes that is both simple and dependent on the microstructural composition is needed.

Later, 3D simulations of hydrated cement pastes enabled micro-mechanical studies to calculate mechanical properties. Micro-mechanical models take into account the heterogeneity of the cement paste microstructure. However, such models usually require some knowledge of one or more chemical and mechanical models and are sometimes difficult to master and may require a high computational cost.

Recently, machine learning methods have been introduced for predicting properties of cementitous materials. Reasons why they are suitable for this type of applications are their flexibility and their ability to recognize complex relationships between different inputs. These types of methods can compensate for several drawbacks of linear and non-linear regression methods. First, they can consider a large number of parameters and the number and type of parameters can be more easily modified than for traditional regression models. In addition, their applicability extends to a larger and more diverse range of paste compositions.

Knowledge of the homogenized properties of cement pastes is an important input for deriving the mechanical properties of mortar and concrete. Studies on the fracture properties of cement pastes are only available to a limited extent in the literature. In addition, limited studies consider the effect of the chemical composition of the hydrated cement paste on the final mechanical properties at a given curing age.

In this paper, a model is presented combining a micro-mechanical model and a thermochemical model to create a database for training and validating a machine learning prediction tool. The input to the model is the composition of the hydrated cement paste and other features related to the hydration kinetics and porosity of the hardened cement paste. As output, the model provides the value of the tensile strength $f_t'$ (MPa) of the cement paste.

## II. Overview of the used models

### *A. Artificial Neural Networks*

#### 2.1.1 Overview of neural networks

Artificial Neural networks (ANN) are powerful machine learning algorithms that can be used to detect non-linear relations between multiple inputs and one or more outputs. In such cases, the problem to be solved is called a regression. A good presentation of the use of ANN for regression applications can be found in Géron (2017). In general, an ANN consists of an input layer, one or more hidden layers, and an output layer. The output of each hidden layer is the input of the next layer until the output layer is reached and the predicted final value is computed. This process is called the forward propagation step. Once the forward step has been applied to all input examples, a predetermined loss function such as the mean square error is calculated. The training is done in the backpropagation step, which consists in finding the gradient of the loss function with respect to each weight and bias for each example, and then updating the old weights of the network. An epoch is reached each time all training examples take a pass through the network. The number of epochs, the activation functions, and the learning rate are all hyperparameters to be determined during training.

#### 2.1.2 Training, validation and testing

Tuning a neural network is about finding the best set of hyperparameters for the regression problem. Each time one of the hyperparameters changes, a new regression model is created. $K$-fold cross validation is used

to evaluate the performance of the different models. The data are shuffled and split into $K$ subsets, then the model is trained $K$ times. At each training iteration, $K-1$ subsets are used for training and the omitted subset is used for validation. The average of the metric over the $K$ validation subsets is then calculated and is considered the final performance metric of the model. Here $K$ is taken to be equal to $5$. The metrics used for performance evaluation are mean squared error (mse), mean absolute error (mae), and R-squared, which are presented below.

Validation allows for the detection of bias problems. A network with high bias is one that has overfitted the training data. As a result, the network performs very well on the training data, but performs poorly on the validation data. To prevent overfitting, a regularization method should be used during the training process. Dropout is a well-known regularization method Hinton et al. (2012). $p$ is the dropout rate, another hyperparameter of the network to be determined.

Comparison and evaluation of the performances of neural networks requires the definition of some metrics. Examples of metrics used:
The mean square error

$$mse = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^{pred})^2. \tag{1}$$

The mean absolute error

$$mae = \frac{1}{n}\sum_{i=1}^{n}|y_i - y_i^{pred}|, \tag{2}$$

where $y_i$ and $y_i^{pred}$ are the expected and predicted outputs respectively. Another evaluation factor is the correlation factor that assess the fit of the final model:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i^{pred})^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}, \tag{3}$$

where $n$ is the number of examples, and $\bar{y}_i$ is the mean of the expected outputs.

### *B. Micro-mechanical model for the estimation of the mechanical properties*

Rhardane et al. (2020) has developed a micro-mechanical approach to determine the damage properties of a cement paste. The approach is based on the isotropic damage model of Fichant et al. (1999). The 3D microstructure of the hydrated cement paste is generated using VCCTL Bentz (1997) based on the hydration simulation model of CEMHYD3D. CEMHYD3D works on the principle of cycles. One of the methods proposed by the model to convert the cycle into a physical curing time is to provide the model with a curve expressing the accumulated heat released during the reaction as a function of time. Experimentally, this curve is measured using an isothermal calorimetric test on a cement paste with the same water-cement ratio as the simulated microstructure. However, since the cement used to simulate the microstructure is not always available for experimental testing, a numerical approximation of the curve is required. For this purpose, the hydration kinetics model presented in section 2.3 is used to calculate the heat release during hydration of the paste.

After its creation, a 2D section of the virtual microstructures is extracted and meshed in the finite element code CAST3M (Dureisseix (1997)). It is assumed that all cement paste phases follow the same damage model. The elastic properties of the cement phases are considered intrinsic for each component and can be found in Rhardane et al. (2020) along with the hypothesis of the method. The 2D section is subjected to uniaxial incremental displacement. The stress is then plotted as a function of the width of the fracture opening and the tensile strength $f't$ is calculated from the curve.

### C. *Thermochemical model for the cement paste*

When in contact with water, cement constituents start to dissolve and hydration products start to form. The rate at which each of the anhydrous cement main components reacts is different and evolves with time. In their approach, Parrot & Killoh (1984) established a set of empirical expressions to describe the evolution with time of rate of hydration $\dot{\alpha}$ of each clinker phase. At each time step, the hydration rate of a phase $x$ of clinker is controlled by the minimum of the three following equations:

*Nucleation*

$$\dot{\alpha_{1x}} = \frac{K_{1x}}{N_{1x}}(1-\alpha_x)(-ln(1-\alpha_x))^{(1-N_{1x})}. \tag{4}$$

*Diffusion*

$$\dot{\alpha_{2x}} = \frac{K_{2x}(1-\alpha_x)^{2/3}}{1-(1-\alpha_x)^{1/3}}. \tag{5}$$

*Shell formation*

$$\dot{\alpha_{3x}} = (1-\alpha_x)^{K_{3x}}. \tag{6}$$

The $K_{ix}$ and $N_{ix}$ are determined by Parrot and Killoh Parrot & Killoh (1984). $x$ refers to the cement anhydrous phase, $x \in \{C_3S, C_2S, C_3A, C_4AF\}$.

For each of the four main reactants, the rate of hydration is calculated using the following formula

$$\dot{\alpha_x} = min(\dot{\alpha_{1x}}, \dot{\alpha_{2x}}, \dot{\alpha_{3x}}) f_{w/c} \beta_{RH} \frac{A}{A_0} \exp\left(\frac{E_a}{R}\left(\frac{1}{T_0} - \frac{1}{T}\right)\right), \tag{7}$$

where $\beta_{RH}$ takes into account the influence of the relative humidity and is given by

$$\beta_{RH} = \left(\frac{RH - 0.55}{0.45}\right)^4. \tag{8}$$

The term $A/A_0$ is introduced to take into account the effect of the blaine fineness. $A_0 = 385m^2/Kg$ and A is the actual value of the blaine fineness of the cement. $f_{w/c}$ is the term that takes into consideration the effect of the water-cement ratio and is given by

$$\begin{cases} f_{w/c} = (1 + 3.333(H_x w/c - \alpha_x))^4 & if\ \alpha_{tot} \geq H_x w/c \\ f_{w/c} = 1 & if\ \alpha_{tot} < H_x \end{cases}. \tag{9}$$

which means that w/c ratio starts affecting the hydration degree after $\alpha_{tot}$ reaches $H_x w/c$ and the free space available limits the hydration. In this equation, $H_x$ is a parameter given in table **??**. The last term is introduced to take into account the effect of temperature. T and $T_0$ are the actual and reference temperatures respectively and $T_0 = 293.5K$. $R = 8.314J/mol - K$ refers to the universal gas constant.

The next step consists on computing the degree of hydration of each phase at a time $t_i$ using the following equation

$$\alpha_{x,t_i} = \alpha_{x,t_{i-1}} + \dot{\alpha_{x,t_i}}(t_i - t_{i-1}). \tag{10}$$

Finally, the total degree of hydration of the hydrating cement at time $t_i$ is calculated by

$$\alpha_{tot} = f_{C_3S} \times \alpha_{C_3S} + f_{C_2S} \times \alpha_{C_2S} + f_{C_3A} \times \alpha_{C_3A} + f_{C_4AF} \times \alpha_{C_4AF}, \tag{11}$$

where $f_x$ represents the volume fraction of the component $x$.

**TABLE 1. Statistical analysis of the cement composition and hydration properties used as input for the neural network database. The statistical properties are determined on the 149 examples of the database. The first 29 properties are the input features. The tensile strength of index 29 is the considered output feature.**

| Index | Property | Mean | Standard deviation | Minimum | Maximum |
|---|---|---|---|---|---|
| 0 | Curing time $(h)$ | 569.39 | 202.66 | 333 | 981.24 |
| 1 | Density of cement | 3.14 | 0.01 | 3.13 | 3.16 |
| 2 | w/c | 0.44 | 0.05 | 0.4 | 0.5 |
| 3 | DOH | 0.64 | 0.08 | 0.51 | 0.82 |
| 4 | Gel/space ratio | 0.63 | 0.04 | 0.55 | 0.68 |
| 5 | Non-evaporable water on an unignited cement mass | 0.18 | 0.02 | 0.14 | 0.22 |
| 6 | Non-evaporable water on an ignited cement mass | 0.18 | 0.02 | 0.14 | 0.22 |
| 7 | Chemical shrinkage $(\mu L/g)$ | 54.04 | 4.69 | 46.26 | 66.19 |
| 8 | pH | 13.24 | 0.13 | 13.16 | 13.59 |
| 9 | Conductivity $(S/m)$ | 8.16 | 1.28 | 7.28 | 11.9 |
| 10 | Fraction of total filled porosity | 0.25 | 0.03 | 0.19 | 0.32 |
| 11 | Fraction of connected porosity | 0.65 | 0.19 | 0.15 | 0.86 |
| 12 | Fraction of connected solids | 0.98 | 0.0051 | 0.97 | 0.98 |
| 13 | Fraction of C3S | $5.8 \times 10^{-2}$ | $2.41 \times 10^{-2}$ | $7.3 \times 10^{-3}$ | $1.3 \times 10^{-1}$ |
| 14 | Fraction of C2S | 0.04 | 0.02 | 0.02 | 0.08 |
| 15 | Fraction of C3A | $7.78 \times 10^{-3}$ | $4.76 \times 10^{-3}$ | 0 | $2.5 \times 10^{-2}$ |
| 16 | Fraction of C4AF | $2.25 \times 10^{-2}$ | $6.21 \times 10^{-3}$ | $1.17 \times 10^{-2}$ | $3.79 \times 10^{-2}$ |
| 17 | Fraction of $K_2SO_4$ | $4.34 \times 10^{-3}$ | $3.18 \times 10^{-3}$ | 0.0 | $1.14 \times 10^{-2}$ |
| 18 | Fraction of $Na_2SO_4$ | $1.09 \times 10^{-3}$ | $9.57 \times 10^{-4}$ | 0.0 | $4.15 \times 10^{-3}$ |
| 19 | Fraction of gypsum | $4.89 \times 10^{-3}$ | $3.86 \times 10^{-3}$ | $6 \times 10^{-4}$ | $2.0 \times 10^{-2}$ |
| 20 | Fraction of calcite | $1.3 \times 10^{-2}$ | $1.83 \times 10^{-2}$ | $8.5 \times 10^{-4}$ | $7.06 \times 10^{-2}$ |
| 21 | Fraction of inert filler | $2.85 \times 10^{-3}$ | $3.91 \times 10^{-3}$ | 0.0 | $1.58 \times 10^{-2}$ |
| 22 | Fraction of portlandite | 0.12 | 0.01 | 0.09 | 0.15 |
| 23 | Fraction of CSH | 0.3 | 0.04 | 0.22 | 0.38 |
| 24 | Fraction of hydrogarnet | 0.03 | 0.02 | 0.00 | 0.05 |
| 25 | Fraction of ettringite | 0.03 | 0.02 | 0.00 | 0.06 |
| 26 | Fraction of iron hydroxide | $6.45 \times 10^{-3}$ | $2.22 \times 10^{-3}$ | $2.82 \times 10^{-3}$ | $1.25 \times 10^{-2}$ |
| 27 | Fraction of absorbed gypsum in CSH | $2.58 \times 10^{-3}$ | $1.01 \times 10^{-3}$ | $2.50 \times 10^{-4}$ | $4.02 \times 10^{-3}$ |
| 28 | Fraction of empty porosity | 0.07 | 0.01 | 0.04 | 0.10 |
| 29 | Tensile strength (MPa) | 5.88 | 1.98 | 1.88 | 10.87 |

## III. Application of the ANN for the prediction of the tensile strength of cement paste

### *A. Database creation*

A numerical database was created. The 29 inputs are summarized in table 1. The label or output is the tensile strength.

Six different OPC cements are considered. These cements were chosen to cover a wide range of possible cement compositions. The mineralogical compositions of the simulated cements can be found in table 2.

The creation of an input example of the dataset is done using a combination of the micro-mechanical model and the hydration model presented in section 2.2 and section 2.3 respectively. A w/c of 0.4, 0.45 and 0.5 is considered. Curing ages of 14, 21, 28 and 31 days are selected for each cement at each w/c ratio.

The final database includes 149 examples. The statistical distribution of the various features is shown in table 1. Dataset sizes of similar applications typically vary between 100 and 1000 examples, as noted

**TABLE 2. Mineralogical composition of the six different OPCs used for the creation of the database**

| | Primary constituents | | | | | | | Secondary constituents | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Label | $C_3S$ | $C_2S$ | $C_3A$ | $C_4AF$ | $KS$ | $NS$ | Sulfates | Calcium carbonate | Free lime | Inerts |
| CEMI-A | 67.5 | 15.53 | 9.26 | 7.77 | 0 | 0 | 6 | 0 | 0 | 0 |
| CEMI-B | 62.07 | 15.87 | 9.34 | 10.8 | 1.53 | 0.39 | 3.41 | 0.85 | 1.41 | 0.94 |
| CEMI-C | 54.49 | 25.05 | 7.67 | 11.28 | 1.19 | 0.32 | 6.35 | 1.34 | 2.67 | 1.34 |
| CEMI-D | 62.91 | 16.16 | 8.35 | 10.63 | 1.58 | 0.37 | 4.8 | 1.8 | 0.45 | 0.9 |
| CEMI-E | 60.5 | 20.09 | 8.93 | 10.48 | 0 | 0 | 7.92 | 0.21 | 3.16 | 0.84 |
| CEMI-F | 67.33 | 17.39 | 0 | 15.28 | 0 | 0 | 5.2 | 0 | 0 | 10.23 |

byBen Chaabene et al. (2020), therefore the numerically determined dataset considered falls within this range, but could be increased.

### *B. Dimensionality reduction of the input variables*

When a dataset has a large number of input parameters or features, it is refered to as high dimensional. High dimensionality often makes the modelling predictions difficult because the required amount of data grows exponentially with the number of dimensions. In addition, high dimensionality increases the risk of overfitting. In machine learning applications for predicting the properties of cementitious materials, it has been shown that reducing the number of input variables or dimensions has a positive effect on the performance factors of the model Wan et al. (2021).

One method to determine if unnecessary features are present is to filter them based on a statistical measure. Features are ranked based on the Pearson correlation coefficient (when linear relationships are considered) and the mutual information coefficient (when nonlinear relationships are considered). The initial 29 features are ranked using three different coefficients, presented below

1. Pearson coefficient, which represents the linear correlation between each feauture with respect to the output variable,
2. Mutual Information Coefficient (MI) using the Joint Mutual Information (JMI) method proposed by Yang & Moody (1999) and Meyer et al. (2008).
3. Mutual information coefficient (MI) using the Joint Mutual Information Maximization Method (JMIM) proposed by Bennasar et al. (2015).

The JMI and JMIM methods allow to select the features that, together, are relevant to the prediction of the output. Indeed, these methods take into account not only the MI of the features with respect to the output variable, but also the MI of the features with respect to the features already selected in the previous steps. The Python implementation of the JMI and JMIM methods is provided by Homola (2017).

The 29 features are ranked from the most relevant to the least relevant using the three methods mentioned above. Afterwards, the number of selected features is varied from 10 to 28 features for the JMI and the Pearson methods and 10 to 26 features for the JMIM methods. Knowing that the ranking of the features obtained from different ranking methods is different, two datasets having the same number of features as inputs determined from two different methods are not identical. Therefore, a total of 53 different sizes and combinations of possible input features are evaluated. Each of the 53 different datasets is used to train and validate a neural network according to the method presented in section 2.1.2. The neural network consists

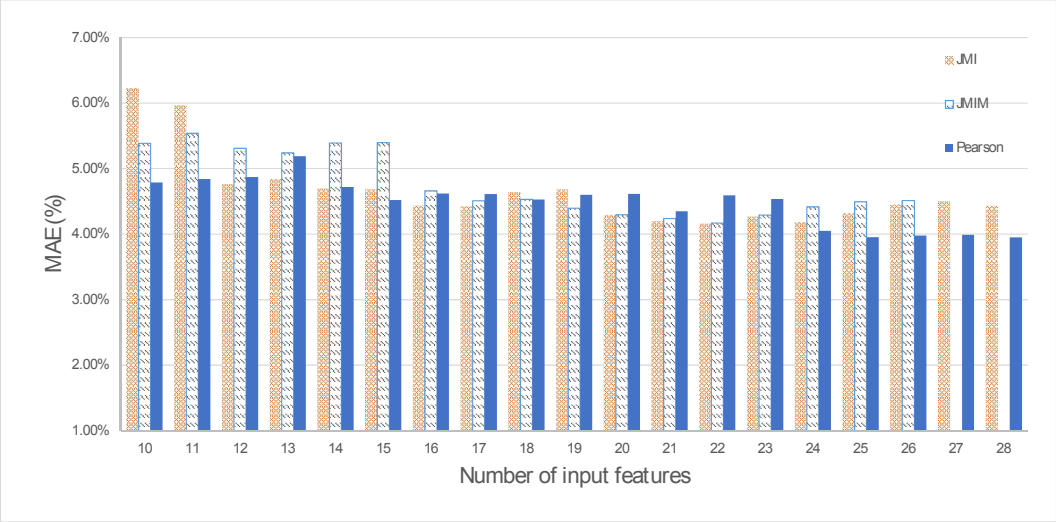


**FIGURE 1. The average of the $mae$ on the $K$ validation sets as a function of the size of the input layer.**

of an input layer of varying number of units (equal to the number of features taken as inputs), 2 hidden layers of 80 units each, and an output layer of one unit. The learning rate is taken to be equal to $1 \times 10^{-4}$, and the activation function applied to all the connections between the hidden layers is the "Leaky-ReLU". To avoid overfitting, a dropout rate of 0.3 is applied on the two hidden layers. The number of epochs is fixed to 5000 epochs and the Nadam optimization is applied. The average of the Mae and the R-squared of the validation sets is used to compare the performances of the different models.

## IV. Results and discussion

### *A. Choice of input features*

To compare the performance of the different models having different input features, the average values of the $mae$ on the $K$ validation subsets as a function of the input layer size are shown in the figure 1. Note that the $mae$ is calculated on the normalized data and multiplied by 100 to be taken as a percentage. A lower value of the$mae$ indicates better accuracy of the model.

The results of the dimensionality reduction, following the Pearson coefficient method, suggest that including the 29 input features gives the most accurate model for the prediction of the tensile strength. The average of the $mae$ on the $K$ validation subsets is equal to $3.89\%$ and the $R - squared$ is equal to 0.94. As for the JMI method, the best performance is obtained for 22 input features where the average of the $mae$ on the $K$ validation subsets reaches a minimum of $4.16\%$ and the $R - squared$ is equal to 0.92. A similar result is found with the JMIM method, where the $mae$ reaches a minimum of $4.17\%$ for 22 input features and the $R - squared$ is equal to 0.92.

For the JMI and the JMIM methods, the six common input features that could be discarded for an optimal model are those having an index of $20, 7, 0, 1, 11$ and 28 shown in table 1.

Those results should be interpreted with caution since the hyperparameters of the model are determined for a model with 29 input features and are kept constant for all the other input layer sizes. Therefore, two final models could be considered, model A with 29 input features and model B with 22 input features determined according to the JMI method.

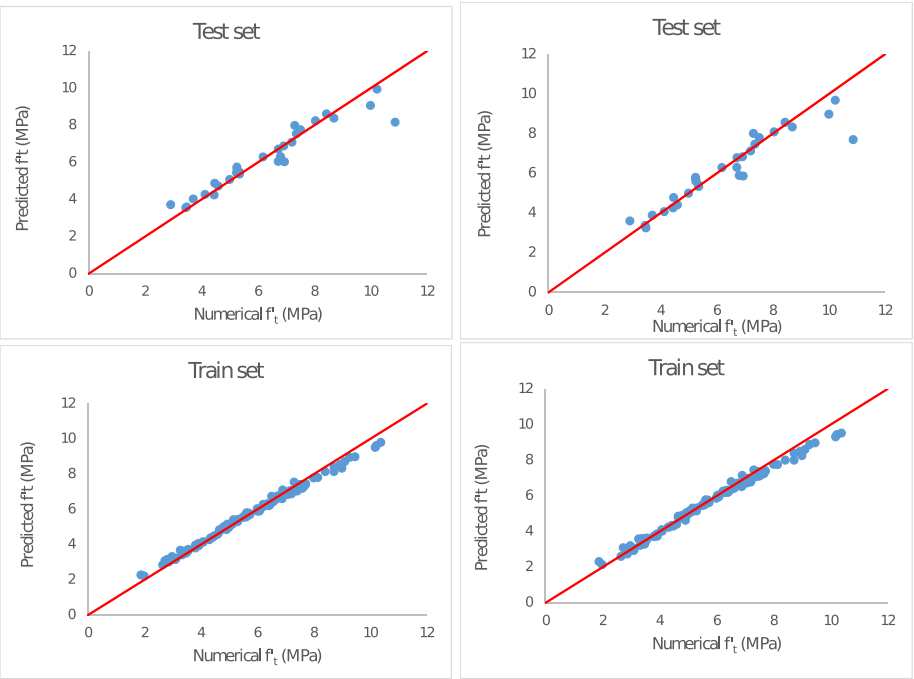


**FIGURE 2. The performance of the ANN during training and testing for the prediction of the tensile strength. On the right, 29 input features are considered (model A). On the left 22 input features are considered (model B).**

**TABLE 3. Evaluation metrics of the two final models for the prediction of the tensile strength.**

| Model | | RMSE | MAE | R-squared |
|---|---|---|---|---|
| A | Train | 0.001881 | 0.03128 | 0.99 |
| | Test | 0.00555 | 0.0456 | 0.9 |
| B | Train | 0.002007 | 0.03403 | 0.98 |
| | Test | 0.007192 | 0.0488 | 0.88 |

### *B. Predicting the tensile strength of a cement paste using an ANN*

For the prediction of the tensile strength, the dataset is split into a train and test set with a ratio of 20 % for the testing set. The same hyperparameter determined with the 5-fold cross validation in section 3.2 are taken. The figure 2 shows the performance of the two models A and B for training and testing.

The evaluation metrics of the two models are presented in table 3. As expected, the model A has a better performance than model B. This result suggests that predicting the tensile strength from the composition of the cement paste is possible. A strong correlation exists between the cement paste property and its composition. Moreover, the prediction is more accurate when the specific phases of the solid are considered separately. Information about porosity and the connection of the porosity are also required.

## V. Conclusion

In this article, the creation of a numerical database is presented. It was used in to train an artificial neural network for the prediction of the tensile strength of the cement paste based on properties such as its com-

position and the porosity volume fraction. The input features are obtained from a hydration model. As for the tensile strength, it is computed using a micro-mechanical model. This model will help fill the lake of knowledge about the relation between the cement paste chemical composition and its tensile strength. The results have shown that:

- In the case of lake of available data in the literature, a micro-mechanical model is an effective tool for the creation of a database. This approach is interesting in the case of use of machine learning algorithms since the size of the database could be increased if needed.
- Linear and non-linear correlations between the chemical composition of a cement paste and its tensile strength makes it possible to get a good prediction of the $f'_t$. This prediction method using ANN is less costly than the use of micro-mechanical modelling once the ANN is trained. In this paper, only the prediction of the tensile strength has been presented. However, it is possible to use the same methodology for the prediction of other elastic and inelastic properties such as the compressive strength and the fracture energy.

For future work, other machine learning (ML) algorithms will be explored and compared to the artificial neural network (ANN) used in this study. A feature importance study, such as a sensitivity analysis, will be performed to identify the most influential parameters on the tensile strength of the hardened cement paste. The database size will be increased and cement pastes with mineral additions will be included, following the same methodology but modified to be suitable for such materials. An experimental validation will be performed, comparing the numerical results with an experimental estimation of the tensile strength of hydrated cement paste using the three point bending test.

## Acknowledgements

The authors would like to acknowledge the financial support of ADEME (Agence de la transition écologique) and Région Pays de la Loire in France.

## References

Ben Chaabene, W., Flah, M. & Nehdi, M. L. (2020), 'Machine learning prediction of mechanical properties of concrete: Critical review', *Construction and Building Materials* **260**, 119889.

Bennasar, M., Hicks, Y. & Setchi, R. (2015), 'Feature selection using joint mutual information maximisation', *Expert Systems with Applications* **42**(22), 8520–8532.

Bentz, D. P. (1997), 'Three-dimensional computer simulation of portland cement hydration and microstructure development', *Journal of the American Ceramic Society* **80**(1), 3–21.

Dureisseix, D. (1997), Une Approche Multi-échelles pour des Calculs de Structures sur Ordinateurs à Architecture Parallèle, Theses, École normale supérieure de Cachan - ENS Cachan.

Fichant, S., La Borderie, C. & Pijaudier-Cabot, G. (1999), 'Isotropic and anisotropic descriptions of damage in concrete structures', *Mechanics of Cohesive-frictional Materials* **4**(4), 339–359.

Gartner, E. & Hirao, H. (2015), 'A review of alternative approaches to the reduction of co2 emissions associated with the manufacture of the binder phase in concrete', *Cement and Concrete Research* **78**, 126–142. Keynote papers from 14th International Congress on the Chemistry of Cement (ICCC 2015).

Géron, A. (2017), *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems*, O'Reilly Media.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. R. (2012), 'Improving neural networks by preventing co-adaptation of feature detectors'.

Homola, D. (2017), Integration and visualisation of clinical-omics datasets for medical knowledge discovery, PhD thesis, Imperial College London.

Meyer, P. E., Schretter, C. & Bontempi, G. (2008), 'Information-theoretic feature selection in microarray data using variable complementarity', *IEEE Journal of Selected Topics in Signal Processing* **2**(3), 261–274.

Nielsen, L. F. (1993), 'Strength development in hardened cement paste: examination of some empirical equations', *Materials and Structures* .

Parrot, L. & Killoh, D. (1984), Prediction of cement hydration., number 35, p. 41 – 53. Cited by: 171.

Powers, T. C. (1958), 'Structure and physical properties of hardened portland cement paste', *Journal of the American Ceramic Society* **41**(1), 1–6.

Rhardane, A., Grondin, F. & Alam, S. Y. (2020), 'Development of a micro-mechanical model for the determination of damage properties of cement pastes', *Construction and Building Materials* **261**, 120514.

Russell, H. G. & al. (1998), Aci 363 r-92 ( reapproved 1997 ) state-ofthe-art report on high-strength concrete reported by aci committee 363.

Wallace, J. & Orakcal, K. (2002), 'Aci 318-99 provisions for seismic design of structural walls', *ACI Structural Journal* **99**, 499–508.

Wan, Z., Xu, Y. & Šavija, B. (2021), 'On the use of machine learning models for prediction of compressive strength of concrete: Influence of dimensionality reduction on the model performance', *Materials* **14**(4).

Yang, H. H. & Moody, J. E. (1999), 'Feature selection based on joint mutual information'.