AI-assisted design of low-carbon cost-effective ultra-high-performance concrete (UHPC)

Weina Meng, Ph.D.
Director, Advanced Concrete Technology (ACT) Laboratory
Department of Civil, Environmental and Ocean Engineering
Stevens Institute of Technology, Hoboken, New Jersey 07030

Yi Bao, Ph.D.
Director, Smart Infrastructure Laboratory
Department of Civil, Environmental and Ocean Engineering
Stevens Institute of Technology, Hoboken, New Jersey 07030
Design of UHPC: Current challenges

• Particle packing model-based method
  ✓ Design based on the packing density
  ✓ Other properties are not guaranteed

• Performance-based method
  ✓ Achieve the optimal performance based on step-by-step testing
  ✓ Extensive experimental testing (costly, time consuming, and labor intensive)

It is important to develop more efficient and effective methods to design UHPC

\[ P(D) = \left( \frac{D^q_{\text{max}} - D^q_{\text{min}}}{D^q_{\text{max}} - D^q_{\text{min}}} \right) \times 100\% \]

Select raw materials
Optimize binder combination
Determine water-to-binder ratio
Determine sand degradation
Determine sand-to-binder ratio
Determine fiber volume

Al-assisted design of UHPC

- Through a prediction-optimization framework, which was designed for auto-discovery of low-carbon cost-effective UHPC

How do machine learning models predict UHPC properties?

- Machine learning models are trained by using existing data
  - The prediction of UHPC properties is a typical regression task
  - High-fidelity machine learning models are required for the regression task

**Material design variables**
(e.g., water-to-cement ratio, sand-to-cement ratio, type of fibers, fiber content, etc.)

**Material properties**
(e.g., compressive strength, tensile strength, flowability, ductility, porosity, etc.)

Machine learning

The real problem is a high-dimensional (nD) problem.
What data are used to train the models?

- Design variables and key properties of UHPC

**Table 5** Proportioning of the designed UHPC mixtures (unit: kg/m³)

<table>
<thead>
<tr>
<th>Code</th>
<th>Cemen</th>
<th>SF</th>
<th>FAC</th>
<th>GGBS</th>
<th>Quartz sand</th>
<th>Fine sand</th>
<th>Sand A</th>
<th>Sand B</th>
<th>HRWR</th>
<th>Total water</th>
<th>Steel fibers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref.</td>
<td>712</td>
<td>231</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1020</td>
<td>211</td>
<td>–</td>
<td>164</td>
<td>156</td>
</tr>
<tr>
<td>G50SF5</td>
<td>548</td>
<td>42</td>
<td>–</td>
<td>535</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>6.5</td>
<td>164</td>
<td>156</td>
</tr>
<tr>
<td>G50</td>
<td>593</td>
<td>–</td>
<td>–</td>
<td>546</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>164</td>
<td>156</td>
</tr>
<tr>
<td>FAC40SF5</td>
<td>663</td>
<td>42</td>
<td>–</td>
<td>367</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>164</td>
<td>156</td>
</tr>
<tr>
<td>FAC60</td>
<td>486</td>
<td>–</td>
<td>556</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>164</td>
<td>156</td>
</tr>
</tbody>
</table>

**Table 6** Characteristics of the UHPC mixtures

<table>
<thead>
<tr>
<th>Code</th>
<th>Ref.</th>
<th>G50SF5</th>
<th>G50</th>
<th>FAC40SF5</th>
<th>FAC60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow time (s)</td>
<td>12</td>
<td>30</td>
<td>37</td>
<td>39</td>
<td>46</td>
</tr>
<tr>
<td>HRWR demand (%)</td>
<td>0.69</td>
<td>1.38</td>
<td>1.06</td>
<td>1.01</td>
<td>0.51</td>
</tr>
<tr>
<td>Mini slump flow (mm)</td>
<td>275</td>
<td>280</td>
<td>285</td>
<td>285</td>
<td>285</td>
</tr>
<tr>
<td>Yield stress (Pa)</td>
<td>39</td>
<td>35</td>
<td>37</td>
<td>34</td>
<td>30</td>
</tr>
<tr>
<td>Plastic viscosity (Pa s)</td>
<td>23</td>
<td>39</td>
<td>50</td>
<td>44</td>
<td>29</td>
</tr>
<tr>
<td>Air content (%)</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>3.5</td>
</tr>
<tr>
<td>Specific gravity</td>
<td>2.47</td>
<td>2.45</td>
<td>2.43</td>
<td>2.44</td>
<td>2.41</td>
</tr>
<tr>
<td>Initial setting (h)</td>
<td>5</td>
<td>2</td>
<td>6</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Final setting (h)</td>
<td>10</td>
<td>6</td>
<td>12</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>1 days—standard curing (MPa)</td>
<td>53</td>
<td>52</td>
<td>64</td>
<td>65</td>
<td>69</td>
</tr>
<tr>
<td>28 days—standard curing (MPa)</td>
<td>135</td>
<td>125</td>
<td>124</td>
<td>124</td>
<td>120</td>
</tr>
<tr>
<td>28 days—heat curing (MPa)</td>
<td>202</td>
<td>178</td>
<td>170</td>
<td>168</td>
<td>136</td>
</tr>
<tr>
<td>Splitting tensile strength (MPa)</td>
<td>12</td>
<td>14</td>
<td>12</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>Unit costs normalized by compressive strength ($/m²/MPa)</td>
<td>14.8</td>
<td>4.7</td>
<td>4.2</td>
<td>4.3</td>
<td>3.5</td>
</tr>
<tr>
<td>Modus of elasticity (GPa)</td>
<td>53</td>
<td>50</td>
<td>50</td>
<td>52</td>
<td>46</td>
</tr>
<tr>
<td>Flexural performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First cracking load (kN)</td>
<td>22</td>
<td>21</td>
<td>24</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>Peak load (kN)</td>
<td>21</td>
<td>29</td>
<td>33</td>
<td>31</td>
<td>28</td>
</tr>
<tr>
<td>δ₁ (mm)</td>
<td>0.092</td>
<td>0.085</td>
<td>0.080</td>
<td>0.093</td>
<td>0.089</td>
</tr>
<tr>
<td>δ₂ (mm)</td>
<td>0.701</td>
<td>0.690</td>
<td>0.653</td>
<td>0.820</td>
<td>0.635</td>
</tr>
<tr>
<td>Peak strength (MPa)</td>
<td>19.7</td>
<td>20.2</td>
<td>22.8</td>
<td>21.3</td>
<td>20.1</td>
</tr>
<tr>
<td>T50 (J)</td>
<td>40.4</td>
<td>48.8</td>
<td>51.5</td>
<td>51.1</td>
<td>49.4</td>
</tr>
<tr>
<td>Surface conductivity (kΩ cm)</td>
<td>45</td>
<td>30</td>
<td>28</td>
<td>38</td>
<td>34</td>
</tr>
<tr>
<td>Durability factor (%)</td>
<td>99.8</td>
<td>99.8</td>
<td>99.8</td>
<td>99.7</td>
<td>99.7</td>
</tr>
<tr>
<td>Autogenous shrinkage at 28 days (µm/m)</td>
<td>731</td>
<td>602</td>
<td>253</td>
<td>545</td>
<td>593</td>
</tr>
<tr>
<td>Drying shrinkage at 98 days (µm/m)</td>
<td>600</td>
<td>430</td>
<td>56</td>
<td>466</td>
<td>500</td>
</tr>
</tbody>
</table>

Challenges of AI-assisted design of UHPC

- Challenges of data (“the source of knowledge”)
  ✓ How can we efficiently collect data and update the dataset?
  ✓ How can we identify and remove anomalous data?
  ✓ How can we select relevant variables from many variables?

- Challenges of machine learning models
  ✓ How can we select or develop the most appropriate machine learning model?

- Challenges of design optimization
  ✓ How can we optimize concrete design by considering multiple design objectives?

- Challenges of various wastes
  ✓ How can we deal with the large variations in the physical properties and chemical compositions of wastes?
Our research (AI designer)

- Challenges of data
  - Self-updatable data collection (**AI data collector**)
  - Artificial data generation (**AI data generator**)
  - Data cleaning and variable selection (**AI data processor**)
- Challenges of machine learning models
  - Automatic generation of machine learning model (**AI auto-learner**)
- Challenges of design optimization
  - Multi-objective optimization (**AI optimizer**)
- Challenges of various wastes
  - Artificial language for data presentation (**AI data presenter**
An approach was developed to automatically collect available data from published documents (e.g., journal papers, conference proceedings, reports, etc.).

The collected database can be automatically updated through tracing and extracting data from new publications.
Why do we use the AI data collector?

- High efficiency and high accuracy
  - Automate the data collection process (without human intervention, free of human errors)

- Self-updatability
  - Improve accuracy by increasing the database size
  - Enable the consideration of new materials (e.g., new solid wastes)
  - Enlarge the design space for lower carbon footprint and lower cost

Task:
Collect 1000 data

AI: 1 hour
By human: 1 month

- Dimension of input variables

- Dimension of input variables

- 35
- 30
- 25
- 20
- 15
- 10
Self-updatability enhances the design capability

- The accuracy increases with time

- The life-cycle cost and carbon footprint are reduced (large design space)
Artificial data generation

Two methods were developed to generate new data (Generative AI)

- Method 1: Use established theories or equations
  \[ \varepsilon_{cu} = 6.6 \ln \left( \frac{L_f}{d_f} V_f \right) - 10.7 \]
  where \( \varepsilon_{cu} \) is tensile strain capacity; \( L_f \) is the fiber length; \( d_f \) is the fiber diameter; and \( V_f \) is the fiber content.

- Method 2: Use advanced machine learning techniques such as generative adversarial networks (GANs)

Transform an image to the style of Van Gogh’s starry night paint

GANs learn from existing real data

- To generate artificial but reasonable and useful data


*Stevens Institute of Technology*
Anomalous data

- Anomalous data can be generated by many reasons (e.g., error in experiments, data entry, and post-processing)
- Anomalous data have different features from normal data
- Data are ranked by their normalness through supervised or unsupervised learning

Supervised anomaly detection based on bivariate analysis

Unsupervised anomaly detection using isolation forest

Removing anomalous data improves accuracy

- The data-driven identification of anomalous data may treat normal data as anomalous data.

- **Contamination ratio** (CR) is defined as the percentage of anomalous data in a dataset.

- The optimal contamination ratio is obtained through a parametric analysis, to minimize the errors (i.e., maximize the accuracy).

\[
\begin{align*}
\text{If CR}=0, \text{RMSE}&=0.056 \\
\text{If CR}=7.6\%, \text{RMSE}&=0.043
\end{align*}
\]

The minimum error is achieved at a contamination ratio of 7.6%, with a RMSE of 0.043.

---

Variable selection

• How to select appropriate design variables?
  ✓ Problem: When extra variables are included, the machine learning model will be complex and inaccurate. When the necessary variables are not included, the machine learning model will be inaccurate too.
  ✓ Criteria:
    1. The design variables are independent of each other (low correlation)
    2. The design variables are highly correlated to the concerned concrete properties

Variable selection based on correlation (mutual information and univariate linear regression)
Representative types of machine learning models

- **Individual models**
  - ✓ Linear regression
  - ✓ Symbolic regression
  - ✓ K-nearest neighbor
  - ✓ Artificial neural network
  - ✓ Support vector machine
  - ✓ Decision tree

- **Ensemble models**
  - ✓ XGBoost
  - ✓ LightGBM
  - ✓ Gradient boosting
  - ✓ Random forest

Different models have different performance, depending on the specific problem.
Automated machine learning

- Automates the development of high-fidelity machine learning models
- The machine learning model development tasks:
  - Model selection and combination
  - Hyperparameter optimization
  - Model complexity minimization

Different types of models are combined to achieve high accuracy.

Current practice

Raw data → Normalization → Train an ML model

Auto-tune machine learning

Raw data → Data processing → ML selection and control → High-fidelity model → Train and validation

Auto-tune machine learning shows high accuracy

- The Taylor diagram compares the accuracy of different machine learning methods

<table>
<thead>
<tr>
<th>No.</th>
<th>Machine learning method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ridge</td>
</tr>
<tr>
<td>2</td>
<td>Passive aggressive</td>
</tr>
<tr>
<td>3</td>
<td>Multi-layer perceptron</td>
</tr>
<tr>
<td>4</td>
<td>Support vector machine</td>
</tr>
<tr>
<td>5</td>
<td>Partial least squares</td>
</tr>
<tr>
<td>6</td>
<td>Random forest</td>
</tr>
<tr>
<td>7</td>
<td>LightGBM</td>
</tr>
<tr>
<td>8</td>
<td>Azure Microsoft</td>
</tr>
<tr>
<td>9</td>
<td>Proposed method</td>
</tr>
</tbody>
</table>

The proposed method had the lowest errors.

Al-assisted design of UHPC

- Through a prediction-optimization framework, which was designed for auto-discovery of low-carbon cost-effective concrete.
Multi-objective optimization

- How to simultaneously optimize environmental, economical, mechanical, and durability properties?

Machine learning

- Compressive strength
- Ductility
- Durability

Cost

- Carbon footprint
- Embodied energy

Material cost of a unit mass

Many objective optimization

Estimation based on the inventory

An inventory was developed

Method

Concrete design variables
- Material type
- Cement type
- Fiber type
- Cement content
- SCM contents
- Fiber content
- Others
  - Age
  - Curing scheme

Concrete properties
- Mini-slump flow
- Compressive strength
- Tensile strength
- Ductility
- Porosity
- Environmental and economic properties
  - Carbon footprint
  - Cost

Material properties
- Material properties

Define constraints
- Optimization algorithm
- Design objectives
- Optimal design

Machine learning

Inventory data
Evolutionary optimization algorithms

• Search for the optimal solution through minimizing the objective function defined based on the design objective
Example: AI-assisted design of green UHPC

- The following types of materials are available
  ✓ Portland cement, Class C fly ash, silica fume, slag, rice husk, oil tailing powder, limestone, waste glass, waste concrete, quartz powder, quartz sand, river sand, masonry sand, oil tailing aggregate, straight steel fibers, superplasticizer, water

- Design constrains
  ✓ Mini-slump flow $\geq 260$ mm
  ✓ 28-day compressive strength $\geq 120$ MPa

- Design objectives
  ✓ F1: Unit cost (minimization)
  ✓ F2: Unit carbon footprint (minimization)
  ✓ F3: Compressive strength (maximization)
Multiple-criterion decision making

- Reduce the cost by 65%, and reduce the carbon footprint by 56%

<table>
<thead>
<tr>
<th></th>
<th>AI</th>
<th>UHPC-1</th>
<th>UHPC-2</th>
<th>UHPC-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost ($/m³)</td>
<td>420</td>
<td>1,204</td>
<td>1,134</td>
<td>942</td>
</tr>
<tr>
<td>Carbon footprint (kg/m³)</td>
<td>582</td>
<td>1,312</td>
<td>1,128</td>
<td>773</td>
</tr>
<tr>
<td>Compressive strength (MPa)</td>
<td>156</td>
<td>154</td>
<td>154</td>
<td>157</td>
</tr>
</tbody>
</table>
Problem: The applicability of ML models is limited

- There are different solid wastes in different locations
  ✓ Various machine learning predictive models
- Different wastes have different properties
  ✓ Particle size distribution, chemical composition
- The different physicochemical properties are not considered
  ✓ Materials are designated with their engineering names (e.g., fly ash, slag, etc.)

<table>
<thead>
<tr>
<th>Chemical differences</th>
<th>Class F</th>
<th>Class C</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiO₂ + Al₂O₃ + Fe₂O₃, minimum %</td>
<td>70.00</td>
<td>50.00</td>
</tr>
<tr>
<td>SO₃, maximum %</td>
<td>5.00</td>
<td>5.00</td>
</tr>
<tr>
<td>Moisture content, maximum %</td>
<td>3.00</td>
<td>3.00</td>
</tr>
<tr>
<td>LOI, maximum %</td>
<td>6.00</td>
<td>6.00</td>
</tr>
<tr>
<td>Available alkalis (as Na₂O), maximum %</td>
<td>1.50</td>
<td>1.50</td>
</tr>
</tbody>
</table>

Source: ASTM standard C 618 – 95; composition requirement for fly ash classes
Method: Create a language to describe wastes

Current practice
- Mixture proportion
- Numerical dataset

Proposed method
- Chemical composition
- Particle size distribution
- Curing time
- Specimen geometry

Artificial language
- A sentence describes a raw ingredient
- An essay describes the mixture design

Text mining
- Unique sequence of characters = Tokens
- Output = Frequency of tokens

Physiochemical information
Artificial language
Text mining

Discovery of chemical reactions
Material property prediction
Deep learning

Example of the artificial language

- Various symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>H2O, SiO2, ...</td>
<td>Water, Silicon dioxide, ...</td>
</tr>
<tr>
<td>SP</td>
<td>Superplasticizer</td>
</tr>
<tr>
<td>SF</td>
<td>Steel fiber</td>
</tr>
<tr>
<td>d</td>
<td>Days</td>
</tr>
</tbody>
</table>

- Sentence-like elements

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: {curing time}† d</td>
<td>Curing time</td>
</tr>
<tr>
<td>B: SP = {superplasticizer content}</td>
<td>Admixtures content</td>
</tr>
<tr>
<td>C: SF = {steel fiber content}</td>
<td>Fiber content</td>
</tr>
<tr>
<td>S: {mix proportion}, D10: {d10}, D50: {d50}, D90: {d90}, CC1: {1st chemical composition}, ...</td>
<td>Mix proportion, particle size distribution, and chemical composition of an ingredient</td>
</tr>
</tbody>
</table>

- Essay: A sequence of sentences: [A][B][C][S1][S2]…[Sn]

Input: 60d [5.30:SP][2.53:SF] 416.00 d10: 2.77, d50: 11.66, d90: 44.37, SiO2: 21.16, Al2O3: 6.04, Fe2O3: 3.15, SO3: 2.88, CaO: 63.96, MgO: 0.87, Na2O: 0.05, K2O: 0.54 [S2][S3][S4]…[Sn] = [A][B][S1][S2]…[Sn]

Output: 91.2 MPa (compressive strength)

Perform deep learning to predict concrete properties

<table>
<thead>
<tr>
<th></th>
<th>MAE (MPa)</th>
<th>RMSE (MPa)</th>
<th>$R^2(1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>3.06</td>
<td>5.08</td>
<td>0.99</td>
</tr>
<tr>
<td>Validation</td>
<td>5.79</td>
<td>7.46</td>
<td>0.97</td>
</tr>
<tr>
<td>Test</td>
<td>6.31</td>
<td>9.15</td>
<td>0.96</td>
</tr>
</tbody>
</table>

$$MAE(P, A) = \sum_{i=1}^{n} |p_i - a_i|^2$$

$$R^2(P, A) = 1 - \frac{\sum_{i=1}^{n}(p_i - a_i)^2}{\sum_{i=1}^{n}[a_i - mean(a_i)]^2}$$

Investigate chemical reactions

- Evaluate the interactions between different physicochemical properties

\[
F(x, y) = \frac{|f(x, y) - f(x) - f(y)|}{\max(F)}
\]

Most significant interactions

(1) \( \text{CaO} - \text{Al}_2\text{O}_3 \) interaction:

\[
\begin{align*}
\text{C}_3\text{A} + 3\text{C}\text{SH}_2 + 26\text{H} &\Rightarrow \text{C}_6\text{AS}_3\text{H}_{32} \\
\text{C}_3\text{A} + \text{CH} + 21\text{H} &\Rightarrow \text{C}_4\text{AH}_{22} \\
\text{C}_3\text{AF} + 4\text{CH} + 22\text{H} &\Rightarrow \text{C}_4\text{AH}_{13} + \text{C}_4\text{FH}_{13} \\
\text{CA} + 3\text{C}\text{SH}_2 + 26\text{H} &\Rightarrow \text{C}_6\text{AS}_3\text{H}_{32}
\end{align*}
\]

(2) \( \text{CaO} - \text{SiO}_2 \) interaction:

\[
\begin{align*}
\text{C}_3\text{S} + (3 - x + y) \text{H} &\Rightarrow \text{C}_x\text{SH}_y + (3 - x) \text{Ca(OH)}_2 \\
\text{C}_2\text{S} + (2 - x + y) \text{H} &\Rightarrow \text{C}_x\text{SH}_y + (2 - x) \text{Ca(OH)}_2
\end{align*}
\]

Conclusions

- The machine learning-based **prediction-optimization framework** can predict the key properties and optimize the design of green UHPC
- The **AI data collector** and **generator** are effective in producing and updating datasets
- The **AI data processor** facilitates data cleaning and variable selection
- The **AI auto-learner** enables automatic generation of machine learning models with high accuracy and high generalization performance
- The **AI optimizer** is effective in multi-objective optimization of design
- The **artificial language** enables the consideration of physicochemical properties of various solid wastes, facilitate the design of concrete with various solid wastes and high performance
- It is possible to use the AI methods to investigate the physicochemical reactions of new concrete systems
References


Acknowledgement

The presented research was performed by a team

Prof. Weina Meng
Director, ACT Lab

Prof. Yi Bao
Director, SI Lab

S. Mahjoubi
(Ph.D. student, 19-23)

P. Guo
(Ph.D. student, 20-24)

The research was funded by external sponsors

National Science Foundation
U.S. Department of Transportation
U.S. Department of Energy
U.S. Army
Thank you!

weina.meng@stevens.edu