ACI Convention (Spring)
Monday, March 25, 2024

Deep learning from physicochemical information for AI-assisted design of low-carbon cost-effective concrete

Yi Bao, Ph.D.
Director, Smart Infrastructure Laboratory

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

NAE Grand Challenges in the 21st Century

- Aging infrastructure
  - ✓ America's Infrastructure Report Card 2021: C-
  - ✓ Adverse effects on safety, mobility, and economy
- Climate change
  - ✓ Magnify the challenge of aging infrastructure
  - ✓ Aggravated by infrastructure construction/repair
  - ✓ Cement is responsible for 8% the total emissions
- Water pollution
  - ✓ Landfill of wastes (e.g., construction wastes)

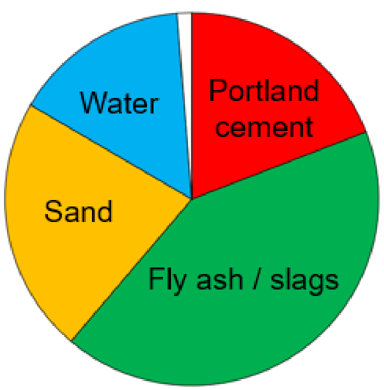
These grand challenges converge to the idea of <u>utilizing wastes to produce **green concrete**</u> with high performance and high sustainability

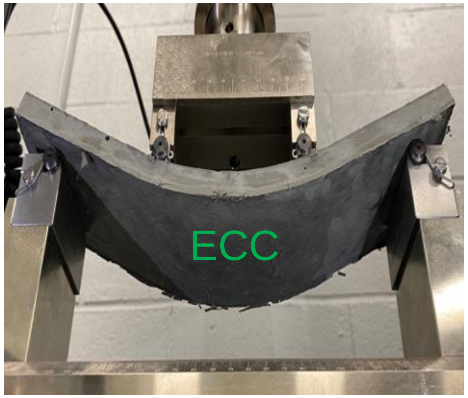
Utilize wastes to produce concrete

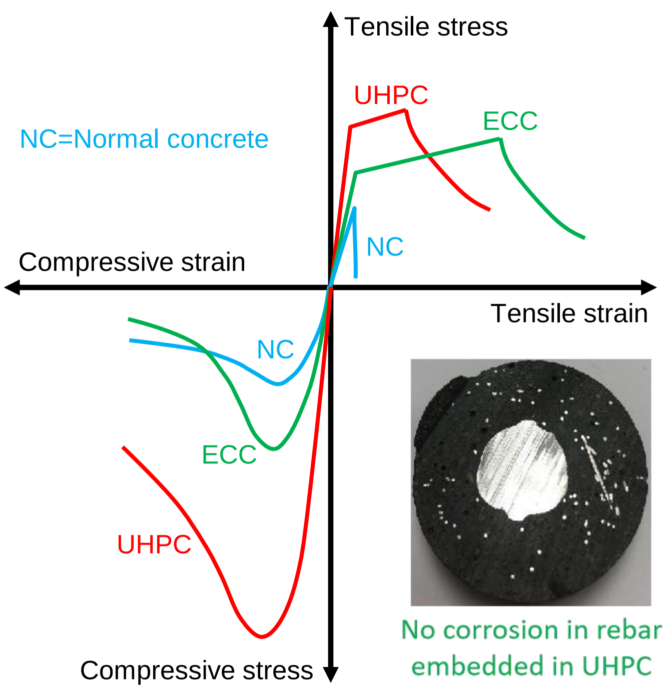
- Representative types of wastes
  - ✓ Industrial by-products (e.g., fly ash, slag, etc.)
  - ✓ Municipal wastes (e.g., waste glass, waste plastics, etc.)
  - ✓ Construction and demolition wastes (e.g., waste concrete)

Replace virgin ingredients

- Advanced concrete
  - ✓ Ultra-high-performance concrete (UHPC)
  - ✓ Engineered cementitious composite (ECC)





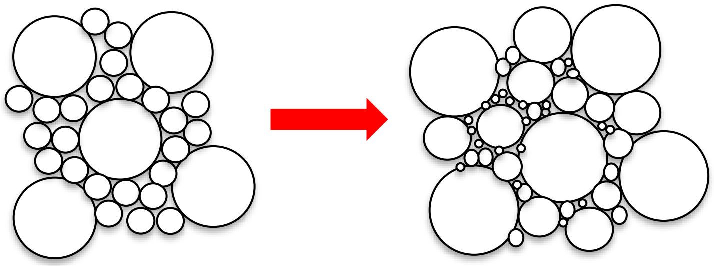


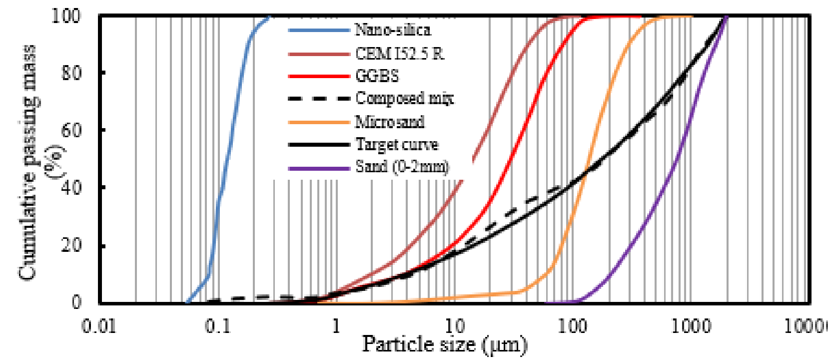
No corrosion in rebar embedded in UHPC

Design of UHPC: Current challenges

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

$$P(D) = \left(\frac{D^q - D^q_{min}}{D^q_{max} - D^q_{min}}\right) \times 100\%$$





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

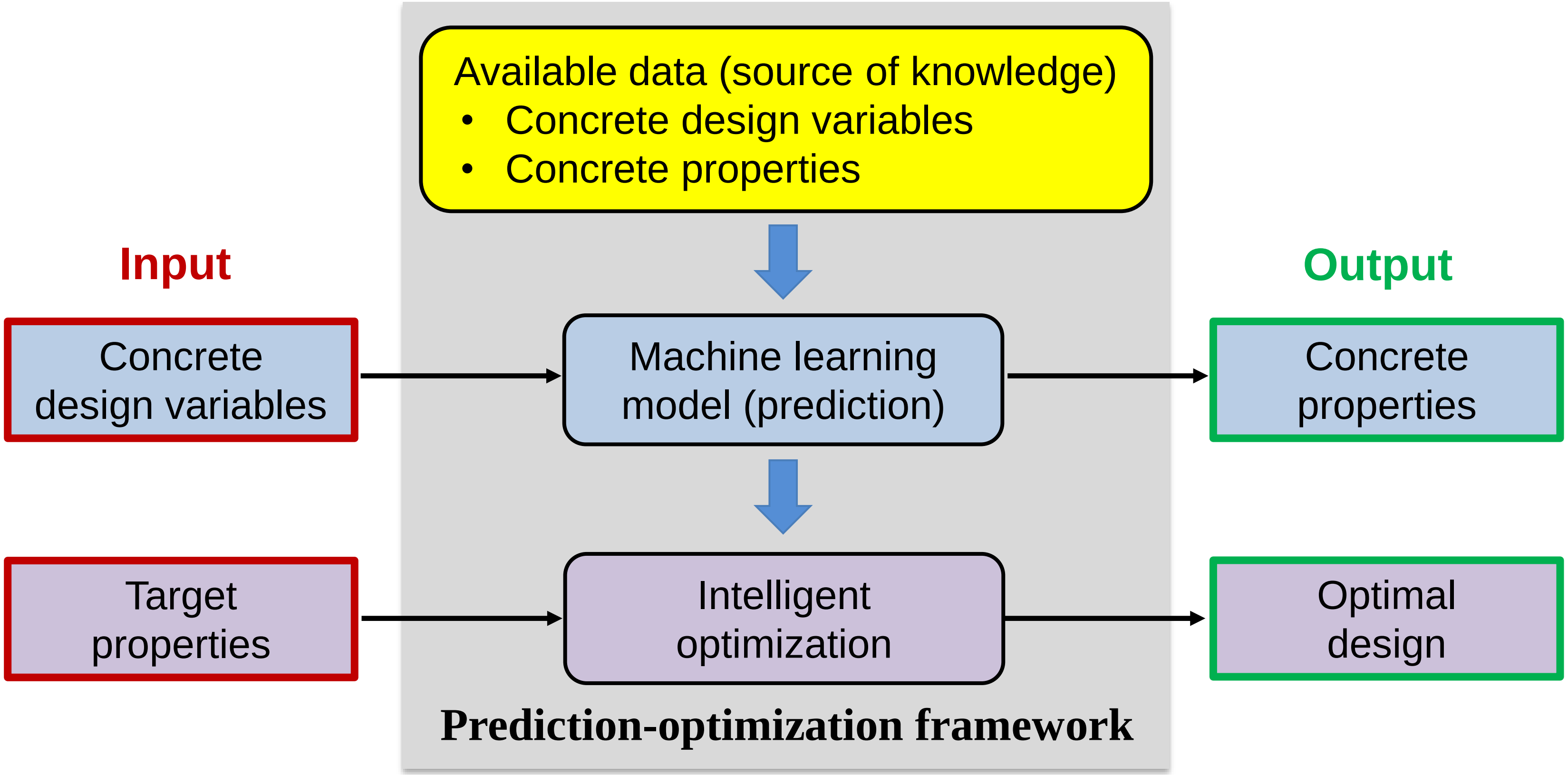
Select raw materials
↓
Optimize binder combination
↓
Determine water-to-binder ratio
↓
Determine sand degradation
↓
Determine sand-to-binder ratio
↓
Determine fiber volume

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

Du, J., Meng, W., Khayat, K.H., et al., 2021. New development of ultra-high-performance concrete (UHPC). *Composites Part B: Engineering*, *224*, p.109220.

AI-assisted design of concrete

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



Mahjoubi, S., Barhemat, R., Meng, W. and Bao, Y., 2023. AI-guided auto-discovery of low-carbon cost-effective ultra-high performance concrete (UHPC). ***Resources, Conservation and Recycling***, *189*, p.106741.

How do machine learning models predict concrete properties? This is a regression problem.

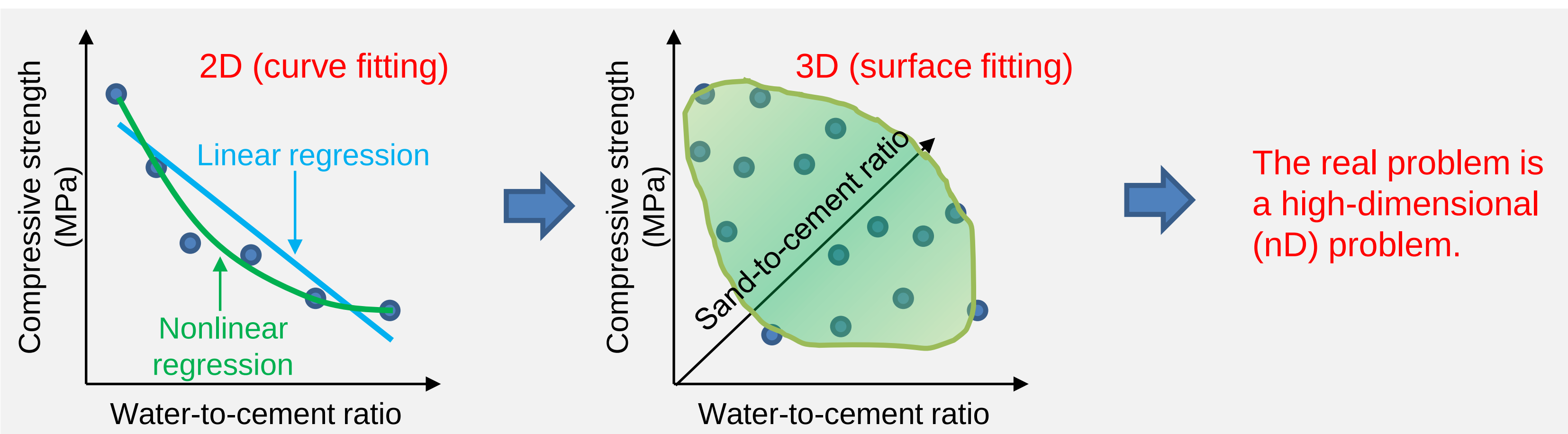
- Machine learning models are trained by using existing data
  - ✓ The prediction of concrete properties is a typical regression task
  - ✓ High-fidelity machine learning models are required for the regression task

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

Link

Machine learning

<u>Material properties</u>
(e.g., compressive strength, tensile strength, flowability, ductility, porosity, etc.)



What data are used to train the models?

- Design variables and key properties of concrete

Table 5 Proportioning of the designed UHPC mixtures (unit: kg/m^3)

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

Mixture design variables

- Types of ingredients
- Mixture variables
- Processing schemes

UHPC properties

- Fresh properties
- Hardened properties

Table 6 Characteristics of the UHPC mixtures

Code	Ref.	G50SF5	G50	FAC40SF5	FAC60
Flow time (s)	12	30	37	39	46
HRWR demand (%)	0.69	1.38	1.06	1.01	0.51
Mini slump flow (mm)	275	280	285	285	285
Yield stress (Pa)	39	35	37	34	30
Plastic viscosity (Pa s)	23	39	50	44	29
Air content (%)	4	5	5	4	3.5
Specific gravity	2.47	2.45	2.43	2.44	2.41
Initial setting (h)	5	2	6	10	6
Final setting (h)	10	6	12	15	12
1 days—standard curing (MPa)	53	52	64	65	69
28 days—standard curing (MPa)	135	125	124	124	120
28 days—heat curing (MPa)	202	178	170	168	136
Splitting tensile strength (MPa)	12	14	12	12	10
Unit costs normalize by compressive strength ($/m^3/MPa)	14.8	4.7	4.2	4.3	3.5
Modus of elasticity (GPa)	53	50	50	52	46
Flexural performance					
First cracking load (kN)	22	21	24	21	20
Peak load (kN)	21	29	33	31	28
δ_1 (mm)	0.092	0.085	0.080	0.093	0.089
δ_p (mm)	0.701	0.690	0.653	0.820	0.635
Peak strength (MPa)	19.7	20.2	22.8	21.3	20.1
T150 (J)	40.4	48.8	51.5	51.1	49.4
Surface conductivity (kΩ cm)	45	30	28	38	34
Durability factor (%)	99.8	99.8	99.8	99.7	99.7
Autogenous shrinkage at 28 days (μm/m)	731	602	253	545	593
Drying shrinkage at 98 days (μm/m)	600	430	56	466	500

Meng, W., Valipour, M. and Khayat, K.H., 2017. Optimization and performance of cost-effective ultra-high performance concrete. *Materials and Structures*, *50*(1), pp.1-16.

Challenges of AI-assisted design of concrete

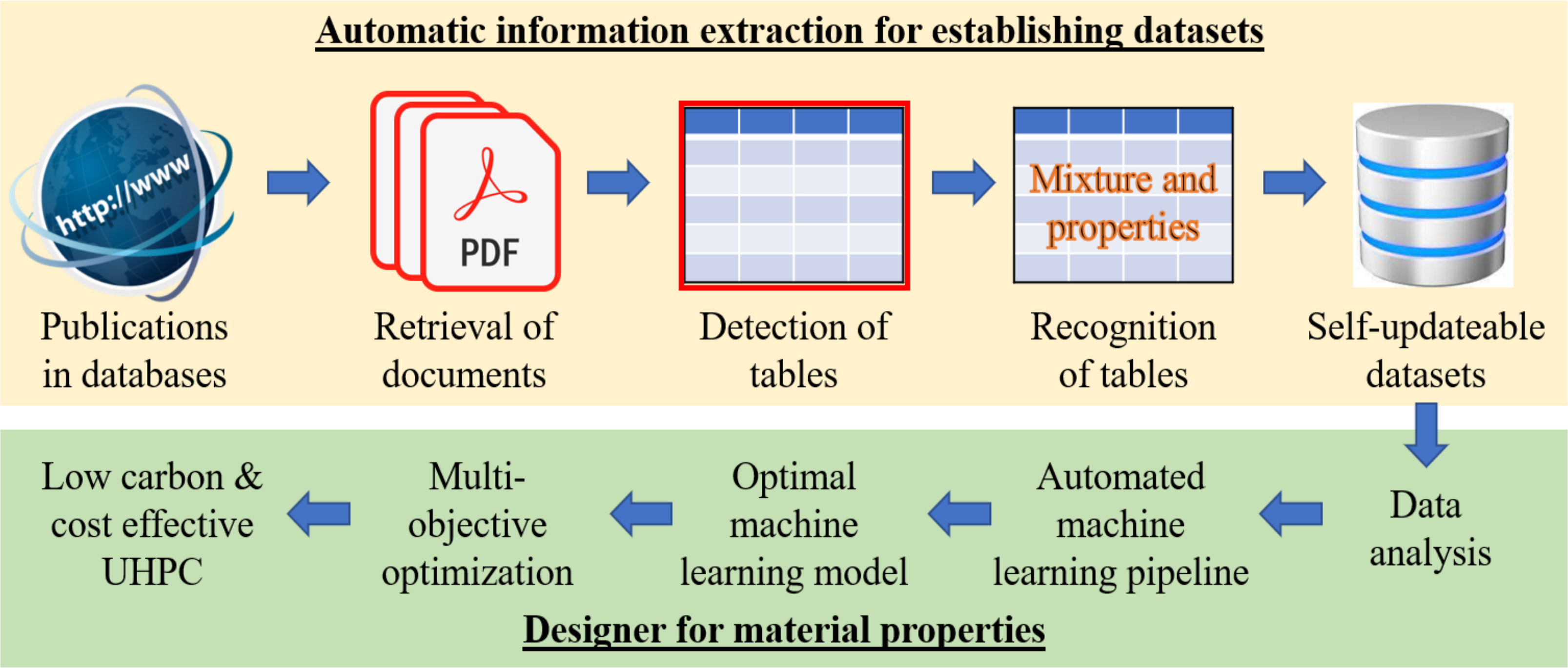
- Challenges of data ("the source of knowledge")
  - ✓ How can we efficiently collect data and update the dataset?
  - ✓ How can we identify and remove anomalous data?
  - ✓ How can we select relevant variables from many variables?
- Challenges of machine learning models
  - ✓ How can we select or develop 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 translator**)

Self-updatable AI data collector

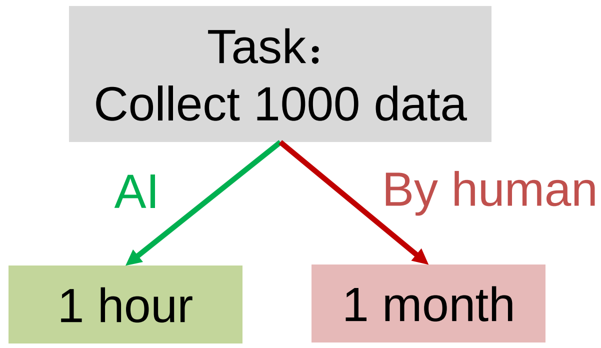
- An approach was developed to automatically collect available data from published documents (e.g., journal papers, conference proceedings, reports, etc.)



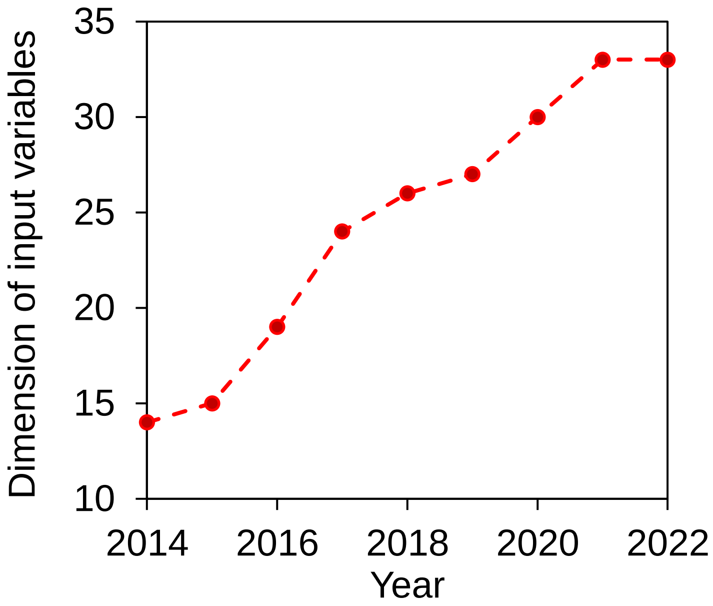
- The collected database can be automatically updated through tracing and extracting data from new publications.

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)

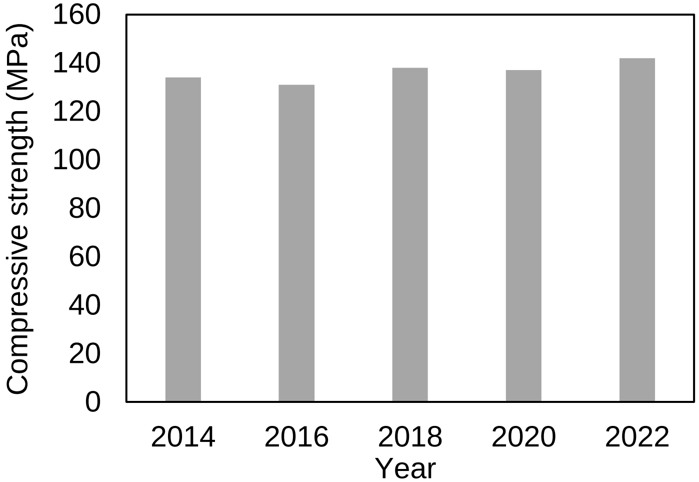


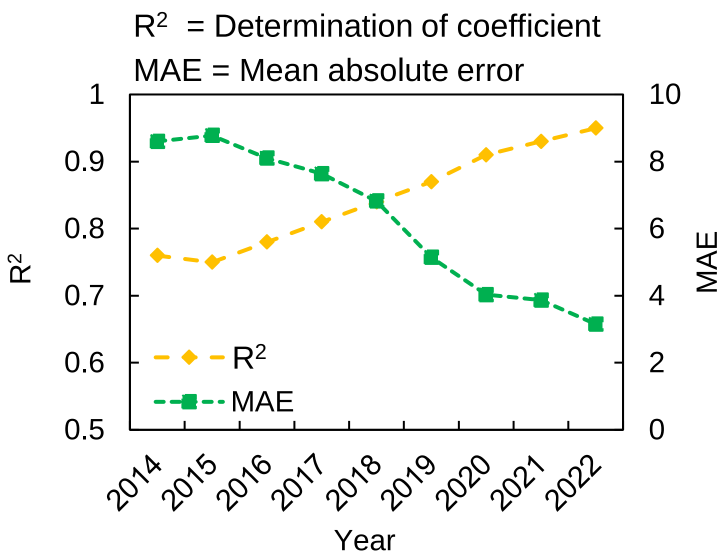
- Self-updatability
  - ✓ Improve accuracy by increasing the database size
  - ✓ Enable the consideration of new materials (e.g., new solid wastes)
  - ✓ Enlarge the design space for lower carbon footprint and lower cost



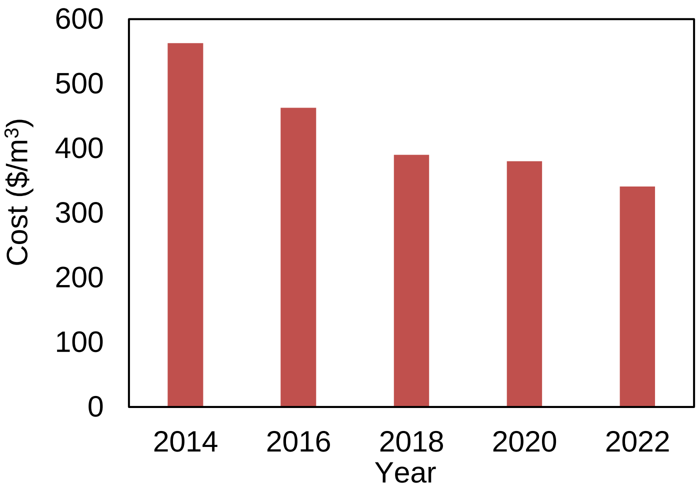
Self-updatability enhances the design capability

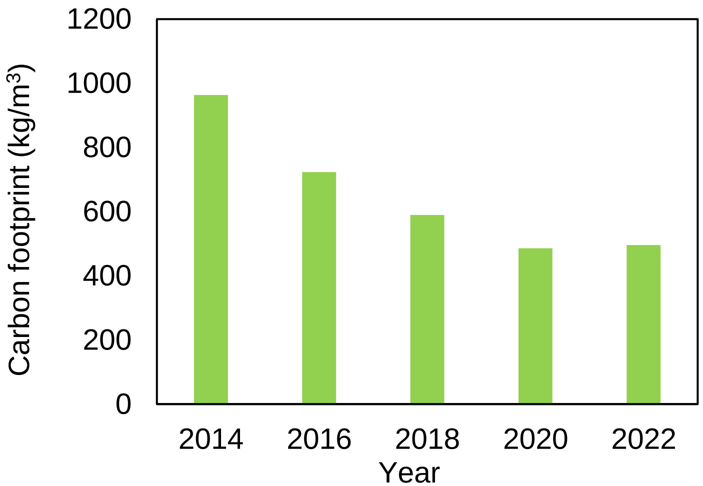
- The accuracy increases with time





- 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 ε_{cu} is tensile strain capacity; L_f is the fiber length; d_f is the fiber diameter; and V_f is the fiber content.

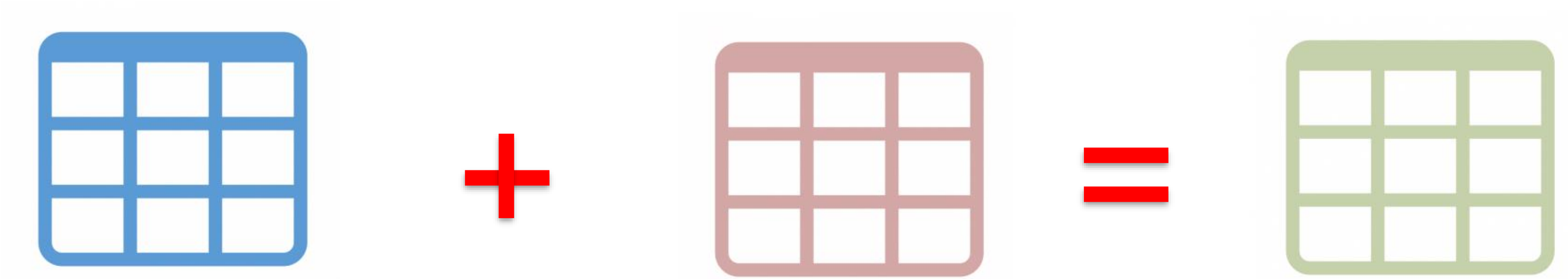
- Method 2: Use advanced machine learning techniques such as generative adversarial networks (GANs)

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

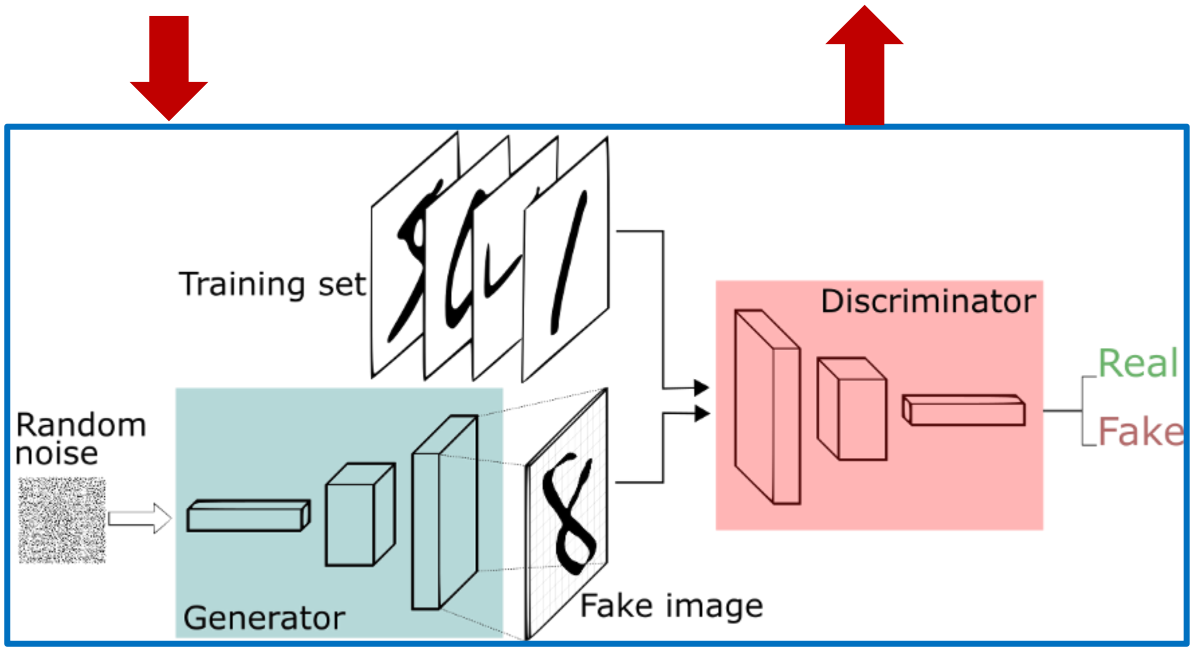
Guo, P., Meng, W., Xu, M., Li, V.C., Bao, Y., 2021. Predicting mechanical properties of high-performance fiber-reinforced cementitious composites by integrating micromechanics and machine learning. *Materials*, 14(12), p.3143.

GANs learn from existing real data

- To generate artificial but reasonable and useful data



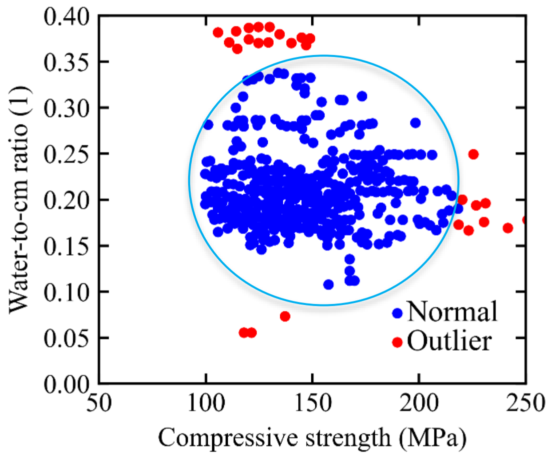
Test data + AI data = Synthetic data



Generative adversarial network (GAN)

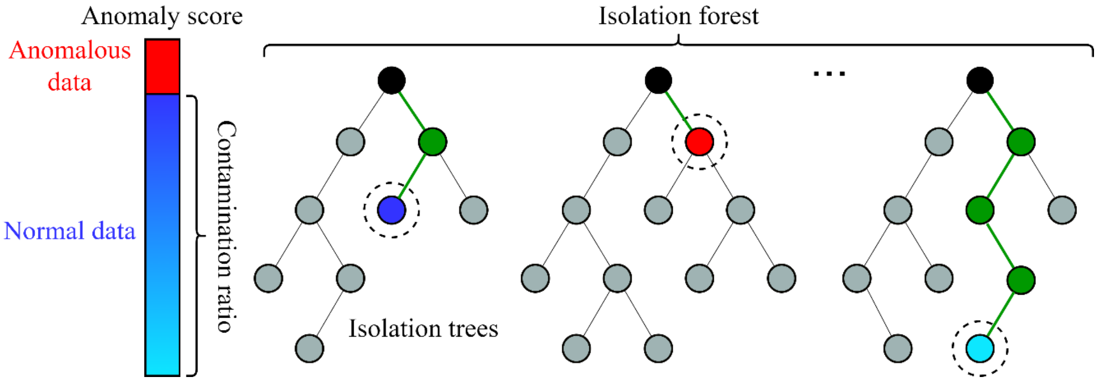
Anomalous data

- 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
- The data-driven identification of anomalous data may treat normal data as anomalous data
- Data are **ranked** by their normalness through supervised or unsupervised learning



Supervised anomaly detection based on bivariate analysis

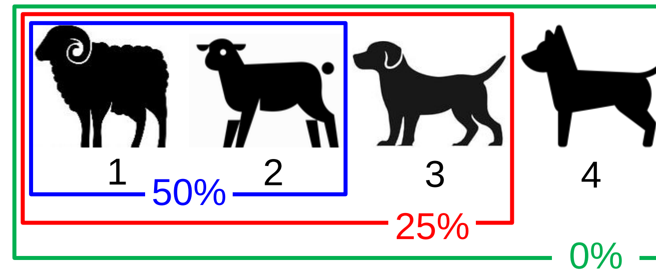
Dog? Sheep?



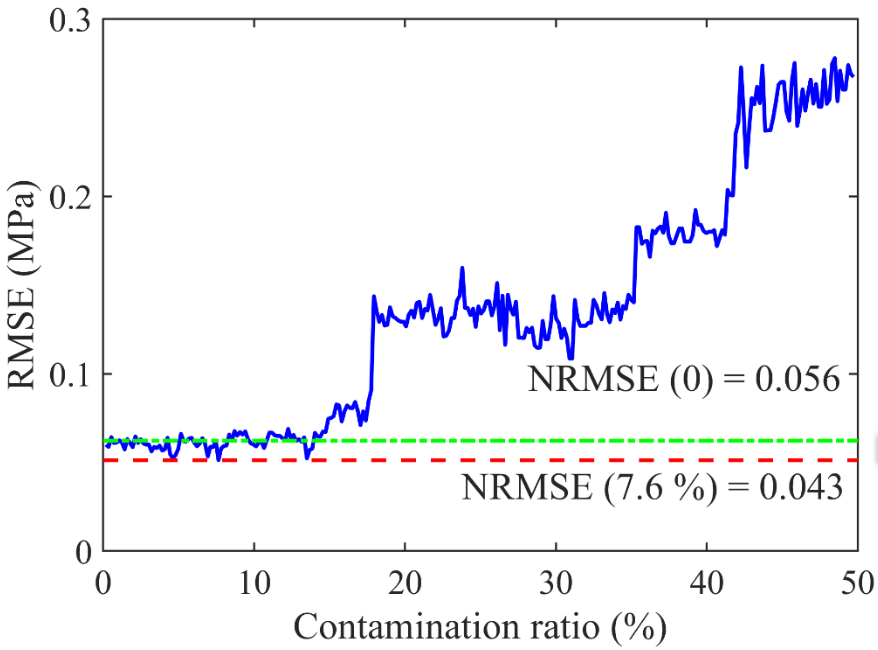
Unsupervised anomaly detection using isolation forest

Removing anomalous data improves accuracy

- **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)



If we have four data, we rank each data, and vary the CR from 0% to 50%



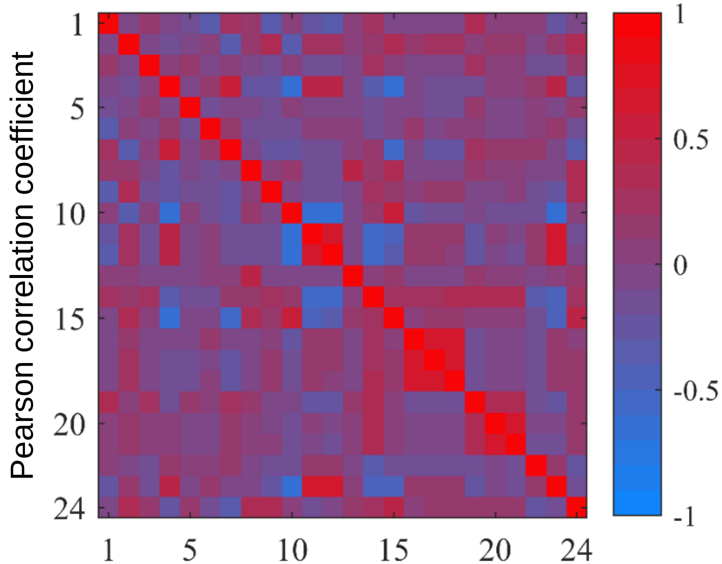
If CR=0, RMSE=0.056

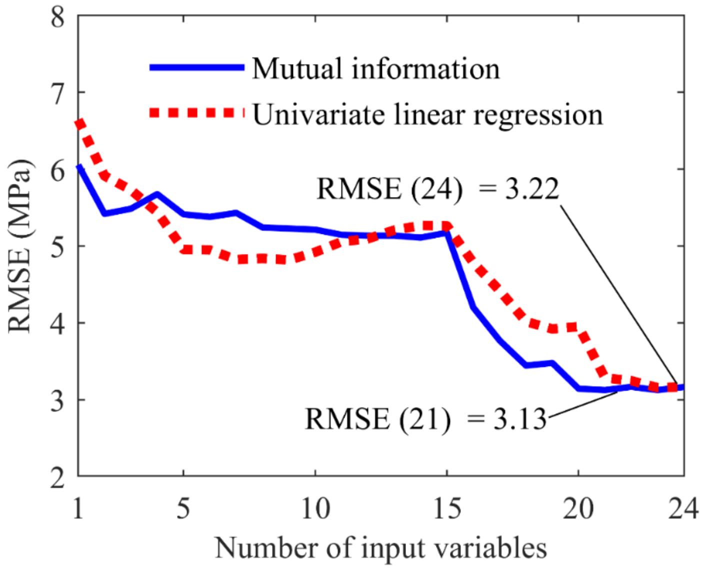
If CR=7.6%, RMSE=0.043

The minimum error

Variable selection

- How to select appropriate design variables?
  - ✓ <u>Problem</u>: 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.
  - ✓ <u>Criteria:</u>
    1. The design variables are independent of each other (low correlation)
    2. The design variables are highly correlated to the concerned concrete properties

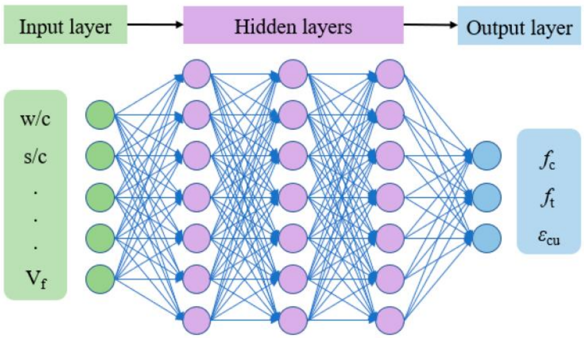


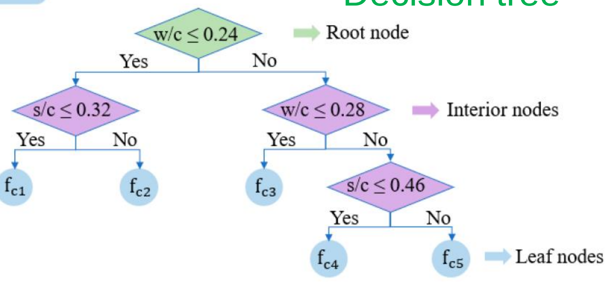


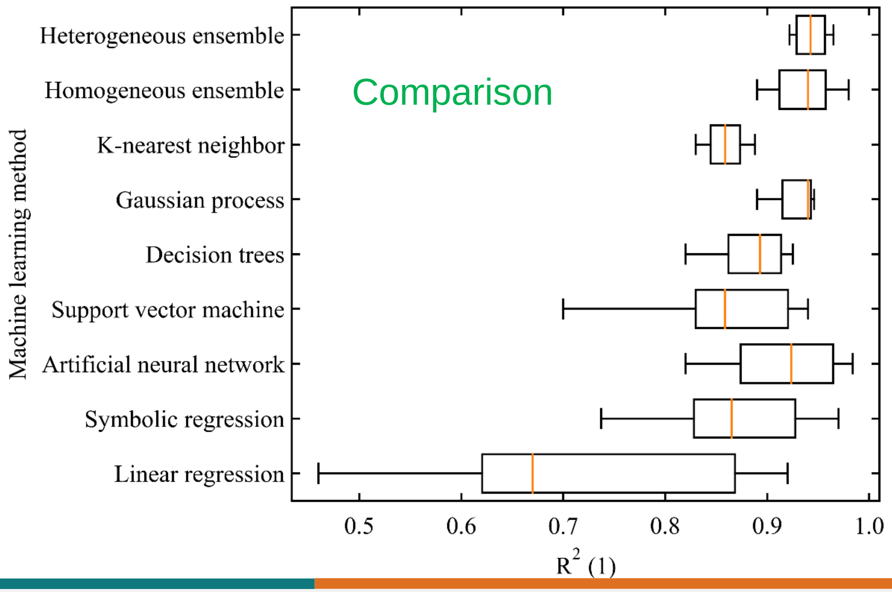
Variable selection based on correlation (mutual information and univariate linear regression)

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





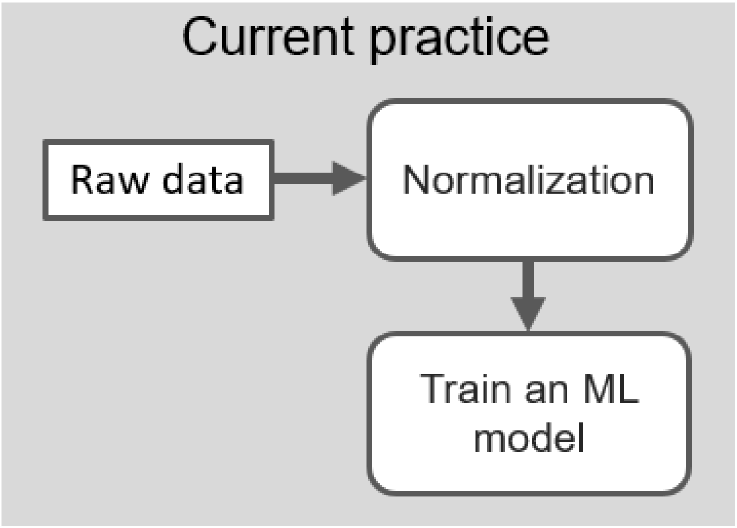


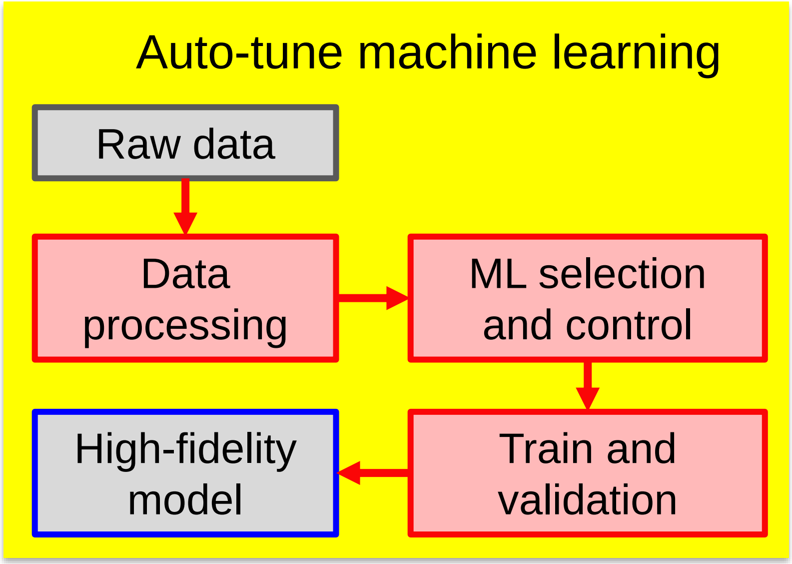
Different models have different performance, depending on the specific problem.

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

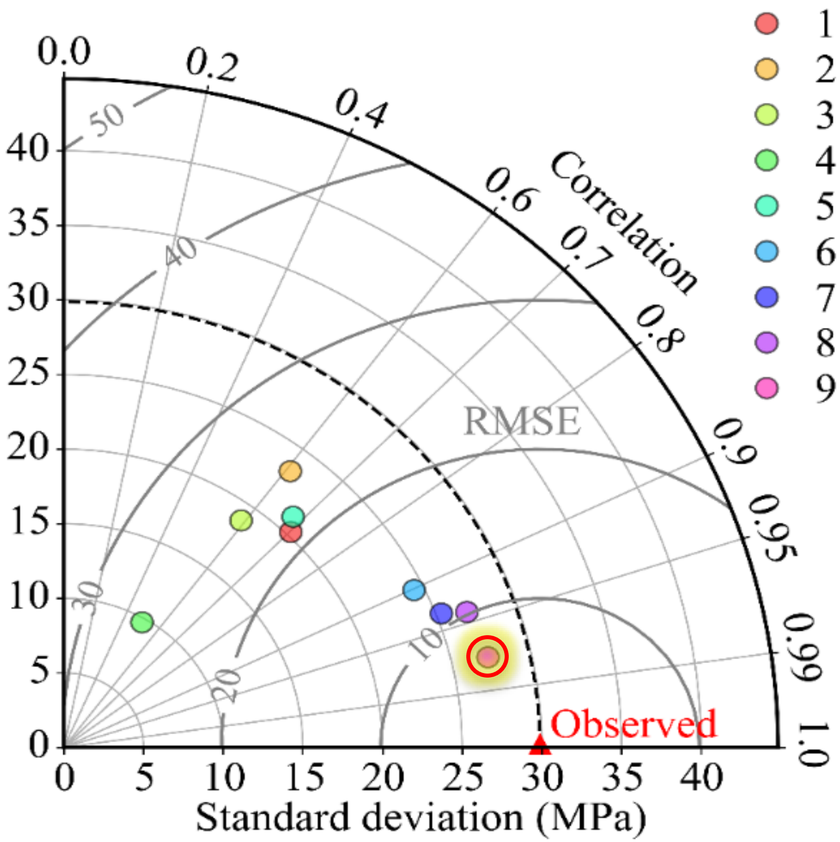
Different types of models are combined to achieve high accuracy.





Auto-tune machine learning shows high accuracy

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

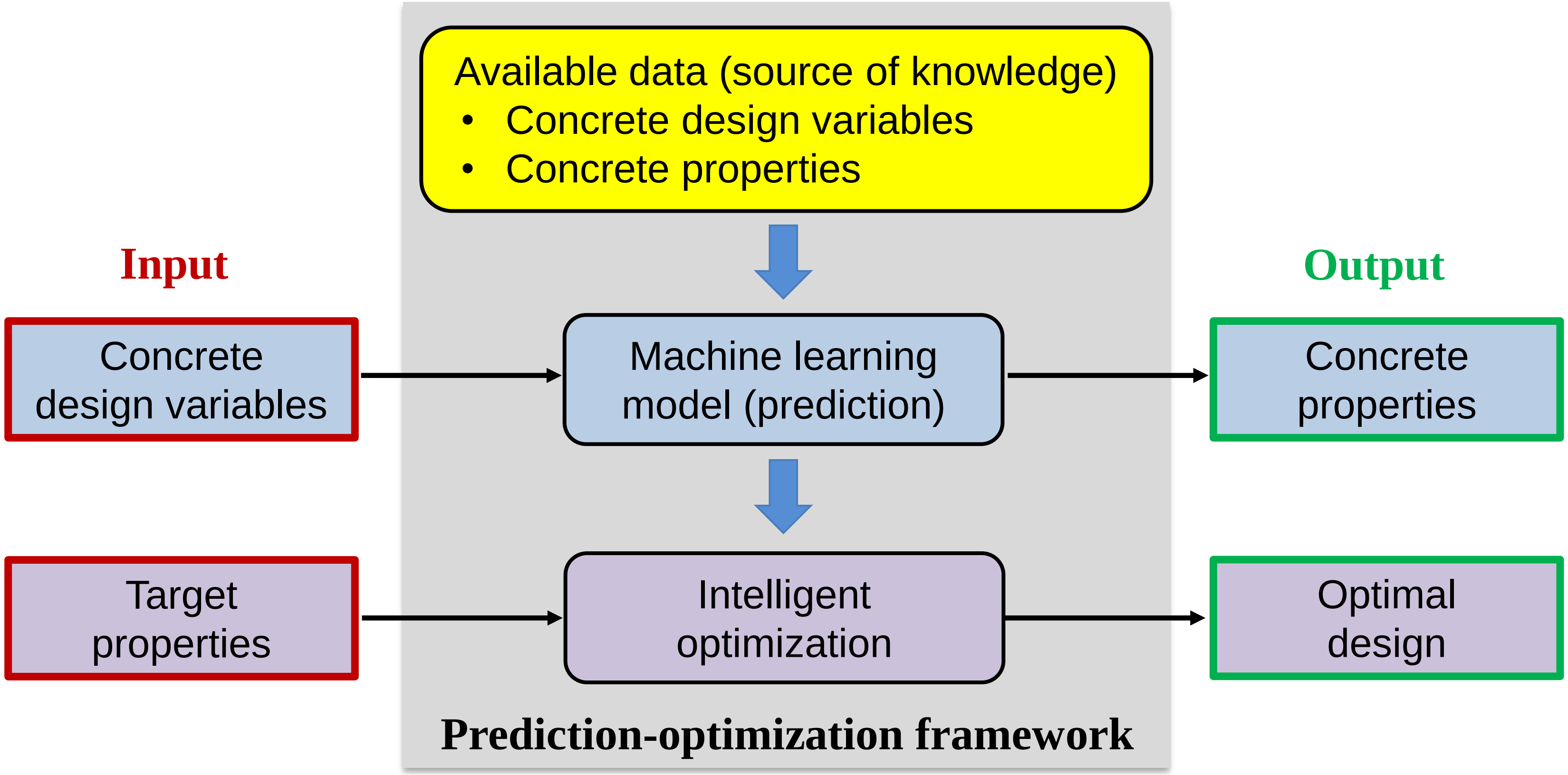


No.	Machine learning method
1	Ridge
2	Passive aggressive
3	Multi-layer perceptron
4	Support vector machine
5	Partial least squares
6	Random forest
7	LightGBM
8	Azure Microsoft
9	**Proposed method**

The proposed method had the lowest errors (highest accuracy)

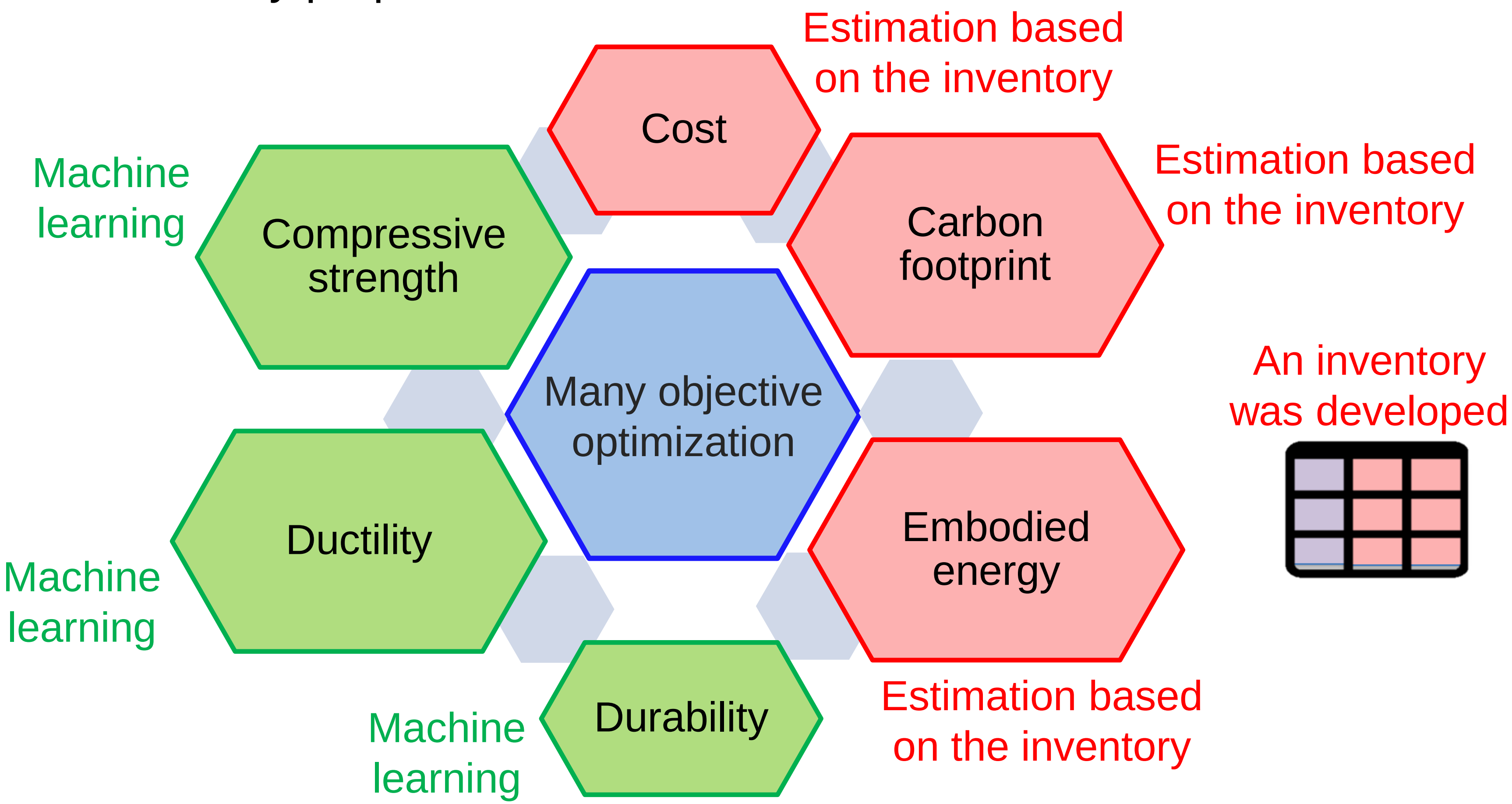
AI-assisted design of concrete

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

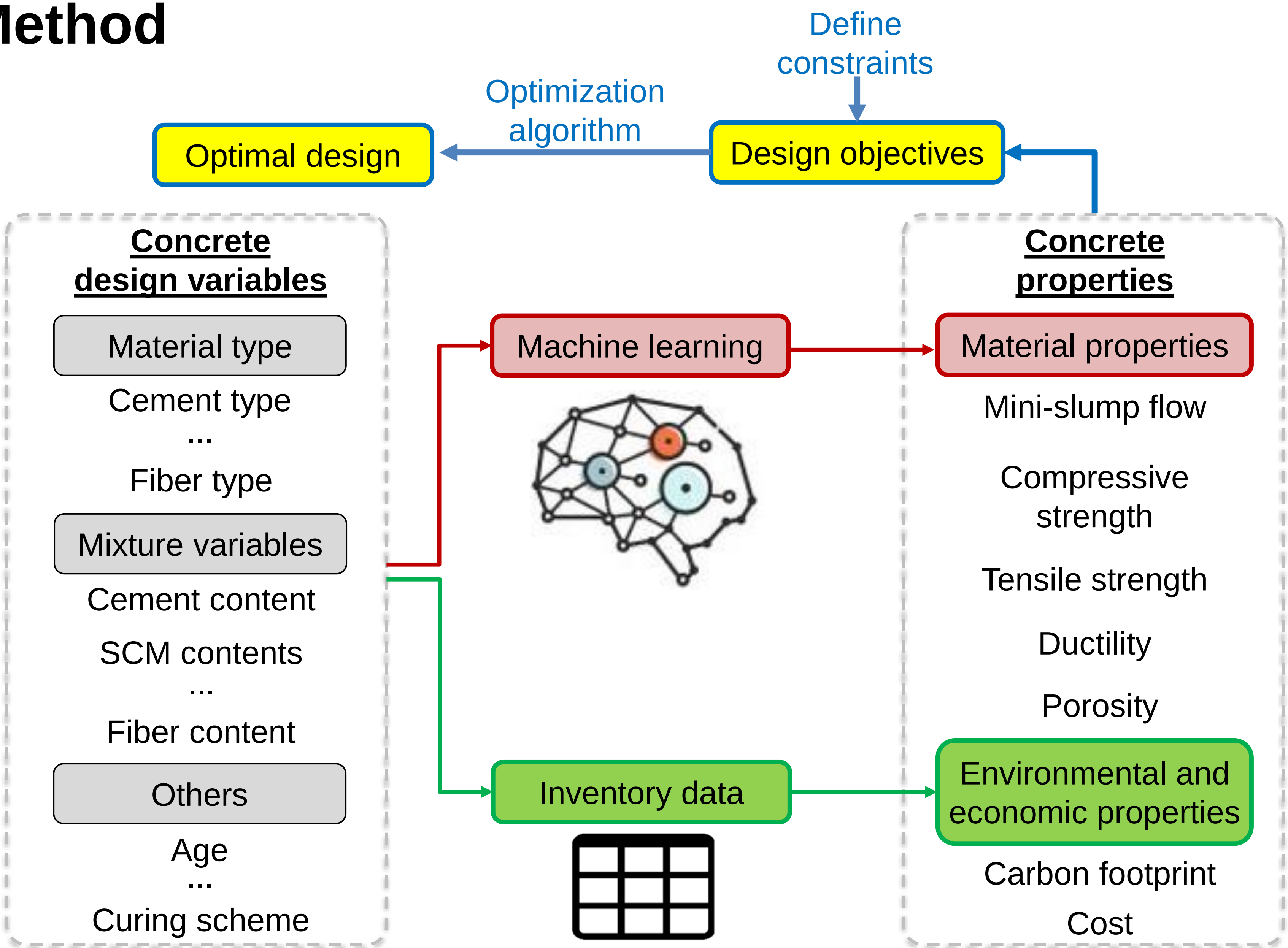


Multi-objective optimization

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



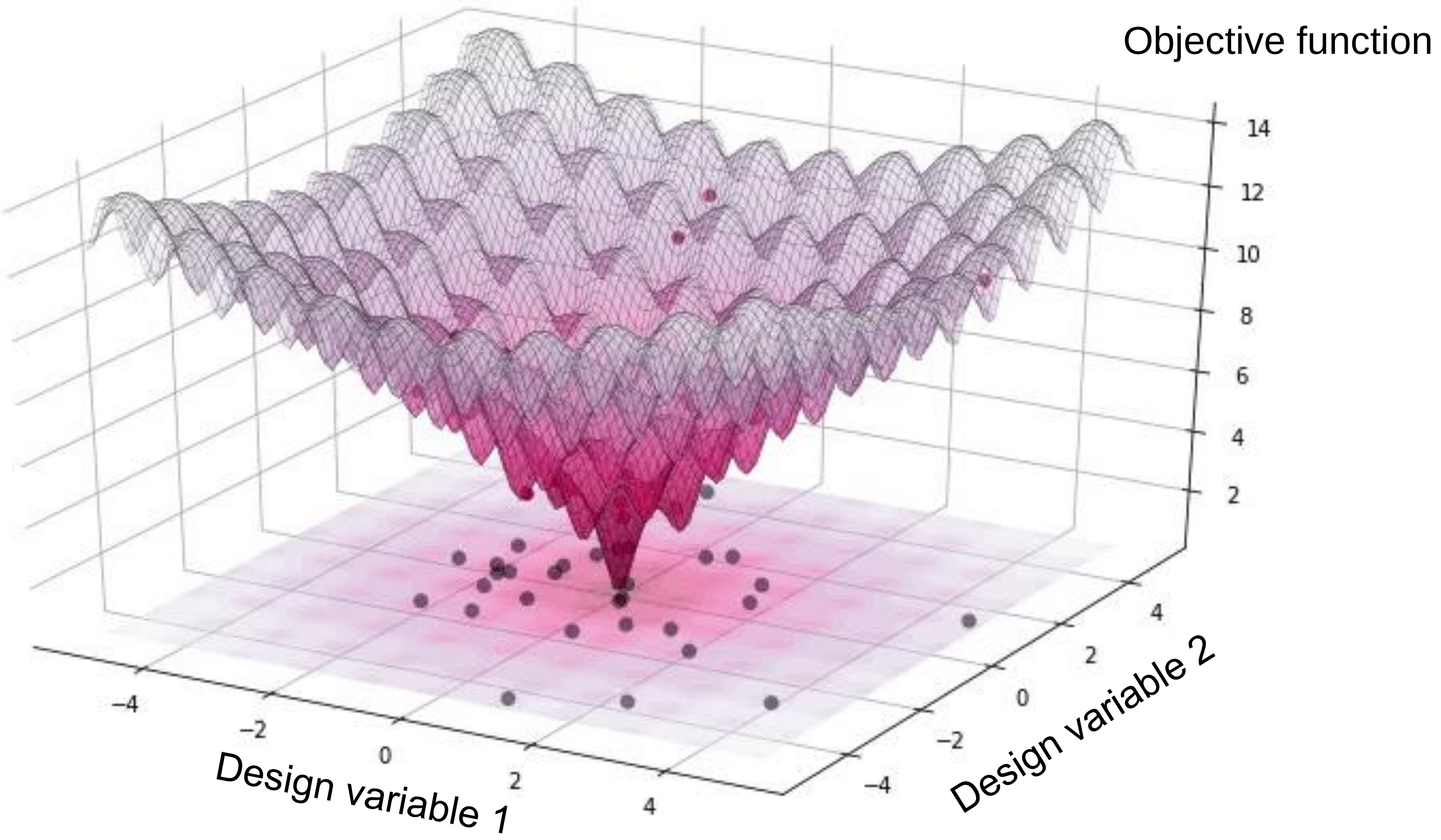
Method



Define constraints
Optimization algorithm
Optimal design
Design objectives
Concrete design variables
Material type
Cement type
...
Fiber type
Mixture variables
Cement content
SCM contents
...
Fiber content
Others
Age
...
Curing scheme
Machine learning
Inventory data
Concrete properties
Material properties
Mini-slump flow
Compressive strength
Tensile strength
Ductility
Porosity
Environmental and economic properties
Carbon footprint
Cost

Evolutionary optimization algorithms

- Search for the optimal solution through minimizing the objective function defined based on the design objective

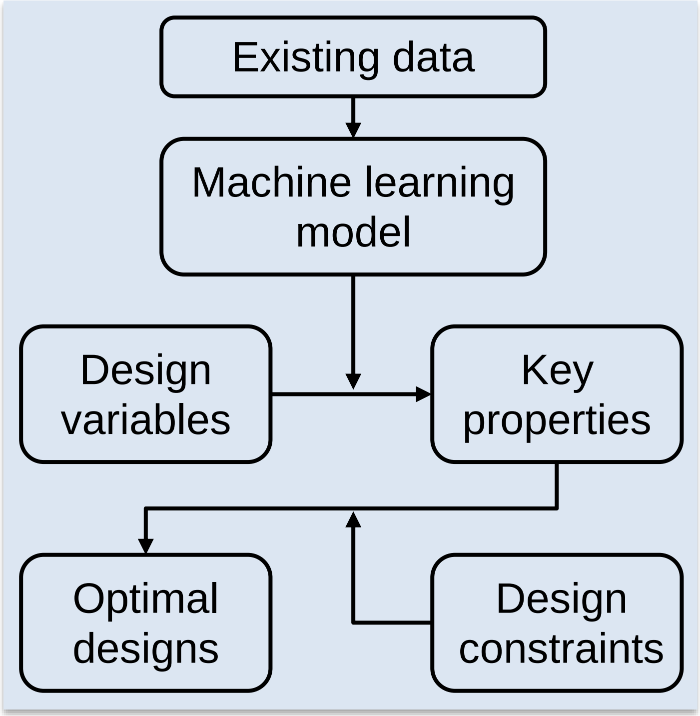


Example: AI-assisted design of green UHPC

- 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

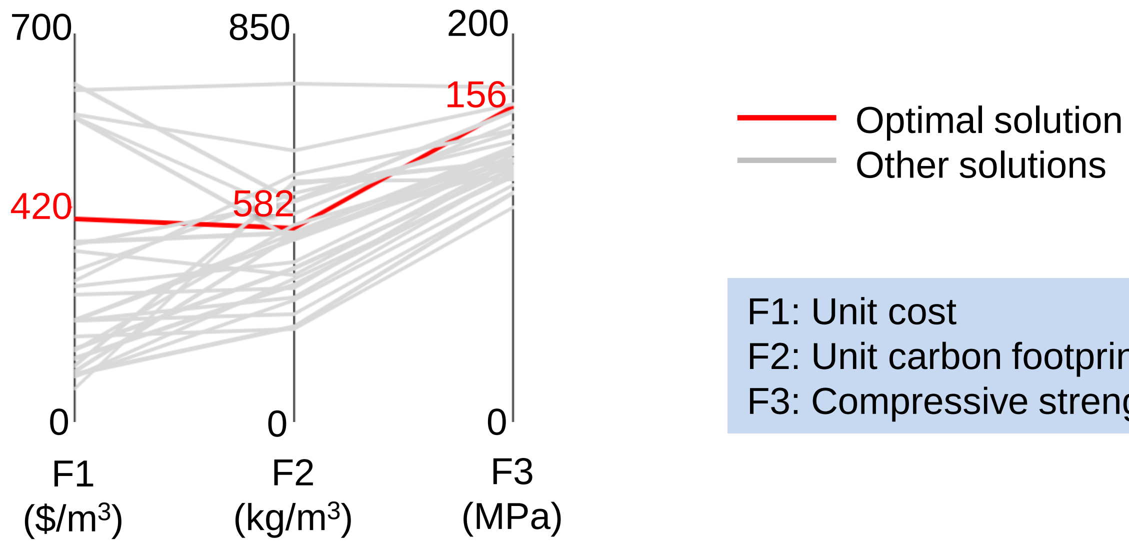
- Design constrains
  - ✓ Mini-slump flow ≥ 260 mm
  - ✓ 28-day compressive strength ≥ 120 MPa

- 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%



F1: Unit cost
F2: Unit carbon footprint
F3: Compressive strength

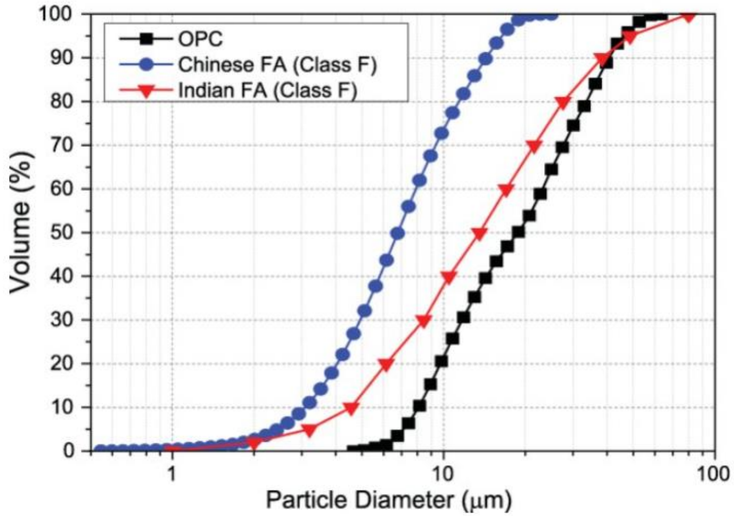
	AI	UHPC-1	UHPC-2	UHPC-3
Cost (USD/m^3)	420	1,204	1,134	942
Carbon footprint (kg/m^3)	582	1,312	1,128	773
Compressive strength (MPa)	156	154	154	157

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.)

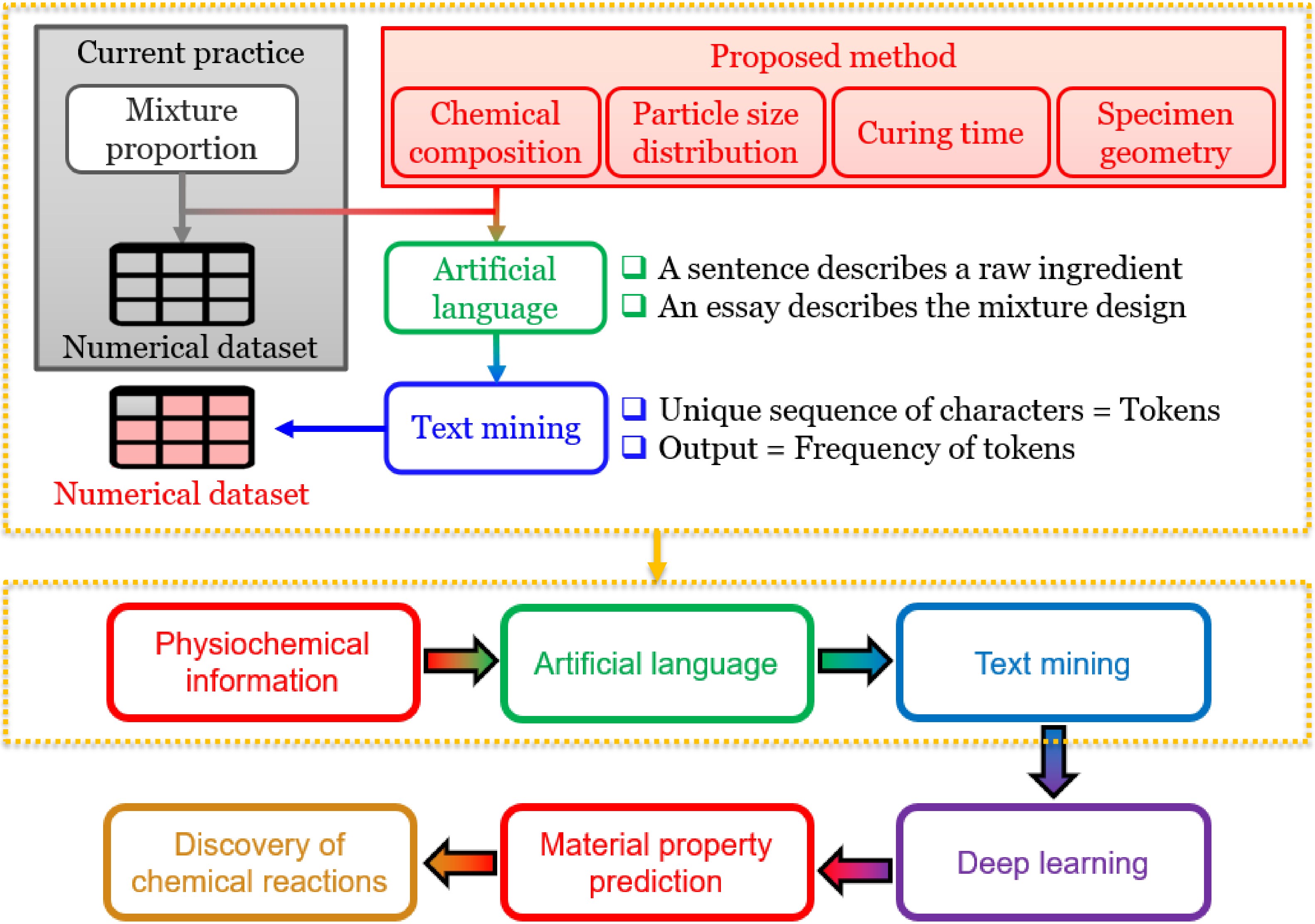
Chemical differences	Class F	Class C
$SiO_2 + Al_2O_3 + Fe_2O_3$, minimum %	70.00	50.00
SO_3, maximum %	5.00	5.00
Moisture content, maximum %	3.00	3.00
LOI, maximum %	6.00	6.00
Available alkalis (as Na_2O), maximum %	1.50	1.50

Source: ASTM standard C 618 – 95; composition requirement for fly ash classes



Mahjoubi, S, Barhemat, R., Meng, W., and Bao, Y., 2023. ***Resources, Conservation and Recycling***, 90, p.106870.

Method: Create a language to describe wastes



Mahjoubi, S, Barhemat, R., Meng, W., and Bao, Y., 2023. ***Resources, Conservation and Recycling***, 90, p.106870.

Example of the artificial language

- Various symbols

Symbol	Meaning
H2O, SiO2, …	Water, Silicon dioxide, …
SP	Superplasticizer
SF	Steel fiber
d	Days

- Sentence-like elements

Sentence	Meaning
A: {curing time}† days	Curing time
B: SP = {superplasticizer content}	Admixtures content
C: SF = {steel fiber content}	Fiber content
S: {mix proportion}, D10: {d10}, D50: {d50}, D90: {d90}, CC1: {1st chemical composition}, …	Mix proportion, particle size distribution, and chemical composition of an ingredient

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

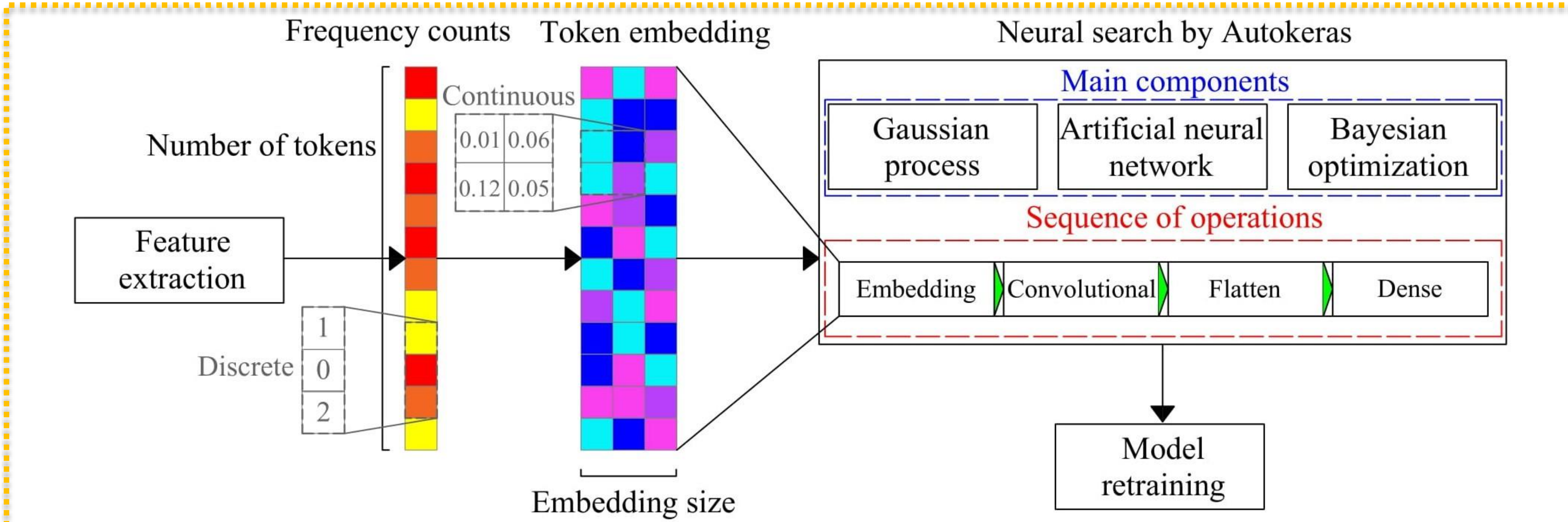
[A] [B] [C] [S1]

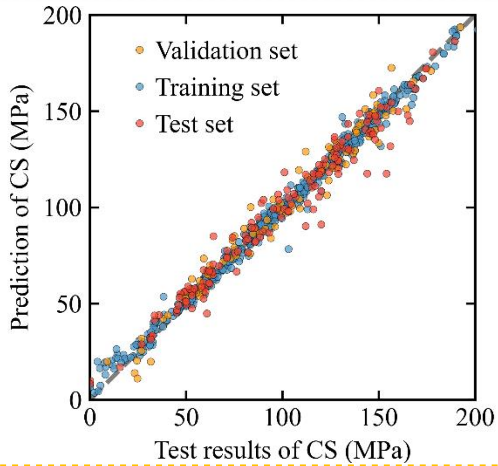
Input:[60d][5.30:SP][2.53:SF][416.00,d10:2.77,d50:11.66,d90:44.37,SiO2:21.16,Al2O3:6.04,Fe2O3:3.15,SO3:2.88,CaO:63.96,MgO:0.87,Na2O:0.05,K2O:0.54][S2][S3][S4]...[S*n*]=[A][B][S1][S2]…[S*n*]

Output: 91.2 MPa (compressive strength)

Mahjoubi, S, Barhemat, R., Meng, W., and Bao, Y., 2023. ***Resources, Conservation and Recycling***, 90, p.106870.

Perform deep learning to predict concrete properties





	MAE (MPa)	RMSE (MPa)	R^2(1)
Training	3.06	5.08	0.99
Validation	5.79	7.46	0.97
Test	6.31	9.15	0.96

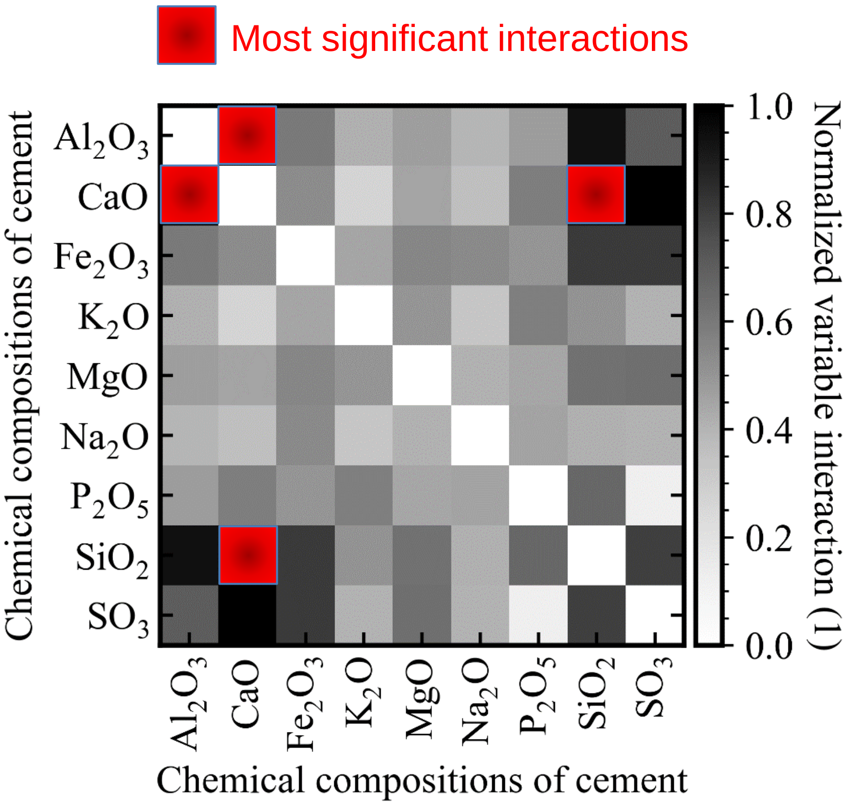
$$MAE(P,A) = \sum_{i=1}^{n} |p_i - a_i|^2$$

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

Investigate chemical reactions

- Evaluate the interactions between different physicochemical properties

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



(1) CaO – Al_2O_3 interaction:

$$C_3A + 3C\bar{S}H_2 + 26H \Rightarrow C_6A\bar{S}_3H_{32}$$

$$C_3A + CH + 21H \Rightarrow C_4AH_{22}$$

$$C_3AF + 4CH + 22H \Rightarrow C_4AH_{13} + C_4FH_{13}$$

$$CA + 3C\bar{S}H_2 + 26H \Rightarrow C_6A\bar{S}_3H_{32}$$

C = CaO
A = Al_2O_3
S = SiO_2
$\bar{S}$ = SO_3
F = Fe_2O_3
H = H_2O

(2) CaO – SiO_2 interaction:

$$C_3S + (3 - x + y)\,H \Rightarrow C_xSH_y + (3 - x)\,Ca(OH)_2$$

$$C_2S + (2 - x + y)\,H \Rightarrow C_xSH_y + (2 - x)\,Ca(OH)_2$$

Conclusions

- The machine learning-based **prediction-optimization framework** can predict the key properties and optimize the design of green concrete
- 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

[1] Mahjoubi, S., Barhemat, R., Meng, W. and Bao, Y.*, 2023. AI-guided auto-discovery of low-carbon cost-effective ultra-high performance concrete (UHPC). ***Resources, Conservation and Recycling***, 189, p.106741.

[2] Mahjoubi, S, Barhemat, R., Meng, W., and Bao, Y.*, 2023. Deep learning from physicochemical information of concrete with an artificial language for property prediction and reaction discovery. ***Resources, Conservation and Recycling***, 90, p.106870.

[3] Guo, P., Mahjoubi, S., Liu, K., Meng, W. and Bao, Y.*, 2023. Self-updatable AI-assisted design of low-carbon cost-effective ultra-high-performance concrete (UHPC). ***Case Studies in Construction Materials***, 19, p.e02625.

[4] Mahjoubi, S., Meng, W. and Bao, Y.*, 2022. Auto-tune learning framework for prediction of flowability, mechanical properties, and porosity of ultra-high-performance concrete (UHPC). ***Applied Soft Computing***, 115, p.108182.

[5] Mahjoubi, S., Meng, W. and Bao, Y.*, 2022. Logic-guided neural network for predicting steel-concrete interfacial behaviors. ***Expert Systems with Applications***, 198, p.116820.

[6] Mahjoubi, S., Barhemat, R., Guo, P., Meng, W. and Bao, Y.*, 2021. Prediction and multi-objective optimization of mechanical, economical, and environmental properties for strain-hardening cementitious composites (SHCC) based on automated machine learning and metaheuristic algorithms. ***Journal of Cleaner Production***, 329, p.129665.

[7] Guo, P., Meng, W., Xu, M., Li, V.C. and Bao, Y.*, 2021. Predicting mechanical properties of high-performance fiber-reinforced cementitious composites by integrating micromechanics and machine learning. ***Materials***, 14(12), p.3143.

Acknowledgement

The presented research was performed by a team

Prof. Weina Meng
Director, ACT Lab

Prof. Yi Bao
Director, SI Lab

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

Mr. 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

New Jersey Department of Transportation

Sto Corp.

Thank you!

yi.bao@stevens.edu