Supporting Information for

Physics-Informed Machine Learning Enabled Virtual Experimentation for 3D Thermoplastics Printing

Zhenru Chen[1], Yuchao Wu[1], Yunchao Xie[2], Kianoosh Sattari[1], and Jian Lin[1*]

[1]Department of Mechanical and Aerospace Engineering, University of Missouri, Columbia, Missouri 65211, United States

[2]Department of Mechanical and Manufacturing Engineering, Miami University, Oxford, Ohio 45056, United States

*Email: linjian@missouri.edu

Supplementary Notes

Supplementary Note S1: Detailed information of physics informed descriptors.

Molecular Weight (MW): This descriptor represents the molecular weight of the compounds, a fundamental property affecting various material characteristics.

Lipophilicity (xlogp3): A measure of lipophilicity, xlogp3 indicates the distribution coefficient of the compound between water and a non-aqueous phase, impacting solubility and material interactions.

Hbond_donor and Hbond_acceptor: The number of hydrogen bond donors and acceptors in a molecule, crucial for understanding molecular interactions and binding capabilities.

Rotatable Bonds (Rot_bond): This parameter denotes the flexibility of a molecule, which can influence its mechanical and physical properties.

Polar Surface Area: Relating to the molecule's ability to interact with other molecules, the polar surface area is key in determining solubility and reactivity.

Heavy Atoms (HA): The count of heavy atoms within a molecule, providing insight into the molecular size and complexity.

The above descriptors are sources from PubChem. [1]

Complexity: A descriptor of the molecule's structural complexity, which can affect its physical behavior and interactions.[2]

Total Energy (dft_sp_E_RB3LYP): Calculated using Density Functional Theory (DFT), this value represents the total energy of the molecule, indicative of its stability and reactivity.[3]

Solubility_dipole: It refers to the solubility influenced by the molecule's dipole moment, a measure of the separation of positive and negative charges. It affects the interaction of the molecule with polar solvents like water.

Solubility_p: This indicates the solubility parameter, representing the cohesive energy density of a material. Substances with similar solubility parameters are generally soluble in each other, following the 'like dissolves like' principle.

Solubility_h: This descriptor relates to the hydrogen-bonding component of solubility, reflecting the compound's capacity to form hydrogen bonds and its consequent solubility in hydrogen-bonding solvents.

Solubility_sqrt_MJperm3: This is a measure of solubility expressed in terms of energy density (MJ/m^3). The square root transformation is applied for normalization or to linearize relationships in the data. The total solubility was calculated from Eq. 1 and Eq. 2.

$$\delta^2 = \delta_d^2 + \delta_p^2 + \delta_h^2 \quad (1)$$

$$\delta_d = \frac{\sum F_{di}}{V}; \ \delta_p = \frac{\sqrt{\sum F_{pi}^2}}{V}; \ \delta_h = \sqrt{\frac{\sum E_{hi}}{V}} \quad (2)$$

F_{di}, F_{pi}, and E_{hi} for different functional groups were extracted from Table 7.10 in the book by Krevelen.[4] V is the molar volume of the monomers.

The above solubility parameters were predicted from monomers' group contributions.[4]

Supplementary Note S2: Discussion in High Explained Variance of the First Principal Component (PC1)

In the datasets, each sample is composed of two-dimensional data, consisting of X (Strain) and Y (Stress) values. During the PCA process, the data is initially reshaped into a single row before the PCA analysis is applied. Similarly, to reconstruct the S-S curves from PCs, the single row of data is reshaped back into two rows. Given the dimensionality of 100 data points, there is a significant degree of freedom involved.

During PCA, the model first identifies a "collinearity" structure in the data. In this context, "collinearity" refers to the linear dependency between variables commonly encountered in statistics and machine learning, where one variable can be well predicted by a linear combination of another. For example, the first 50 values of the 100 data points, which correspond to the strain component, are continuously increased. Thus, the fact that PC1 accounts for 99% of the variance can be intuitively understood because these 100 values adhere to a foundational structure akin to a S-S curve. This interpretation is supported by a related work, where Yang et al. conducted PCA on the stress component of the S-S curves and found that the first three PCs could explain > 85% of the variance.[5] This suggests a generalization of the data's underlying structure by PCA.

The fact that PC1 explains 99% of the variance does not imply that other PC values are unimportant. Since explained variance is a relative measure, it merely highlights that the variances of PCs following PC1 are comparatively smaller. In **Figure 4**, we conduct an interpretability analysis of the PCA results to further elucidate these observations.

Supplementary Note S3: Root Mean Squared Error

Root mean squared error (RMSE) is a standard metric used in statistical modeling to evaluate the differences between values predicted by a model and the observed values. RMSE represents the square root of the average of the squared differences between the predicted values and the actual values. This metric is particularly sensitive to large errors, as it disproportionately weighs these errors more heavily than smaller ones, making it a useful tool for highlighting significant prediction errors. Additionally, compared to Mean Squared Error (MSE), RMSE has a scale that is closer to the original data, making it easier to be interpreted in the context of the problem domain.

The mathematical formula for RMSE is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

where:

n is the number of observations,

y_i represents the actual observed values,

$\hat{y}_i$ represents the predicted values.

Supplementary Note S4: Details of the MLP Model

In this study, we utilized a Multilayer Perceptron (MLP) model to process a wide array of 85 inputs, encompassing both independent and cross-features. The model is structured with four hidden layers, each featuring a descending number of neurons (200, 100, 50, 25 respectively), ultimately leading to an output of 8 principal component (PC) values. These PC values are then reconstructed from these PCs to generate the corresponding stress-strain (S-S) curves. The model's training was facilitated using the Adam optimizer, characterized by a learning rate of 0.001 and a L1 regularization factor of 0.1. This configuration ensures effective learning and regularization to achieve accurate and reliable predictions of material properties. The architecture of the model, along with its dropout rate, learning rate, and L1 regularization, was fine-tuned through a process of grid search optimization. The performance metrics presented in Table 1 and Supplementary Table S2 represent the averages obtained over 10 experimental runs.

All computational tasks in this study were performed on a desktop computer configured with an Intel Core i7-12700K processor, an NVIDIA GeForce 2080 GPU, and 64GB of RAM. The operating system used was Ubuntu 22.04.2. Programming and implementation were carried out in Python 3.7.9. For handling data processing, we employed NumPy (version 1.19.2), Scikit-learn (version 1.0.2), and Pandas (version 1.2.1). The MLP model was developed using PyTorch version 1.13.1+cu117.

Given the relatively small size of the dataset and the simplicity of the model, training could be conducted using either CPU or GPU, with each training session taking less than 15 seconds to complete. The demonstration of 100,000 virtual experiments conducted in this study was performed using CPU inference, with the entire process taking less than 1 minute to complete.

Supplementary Figures

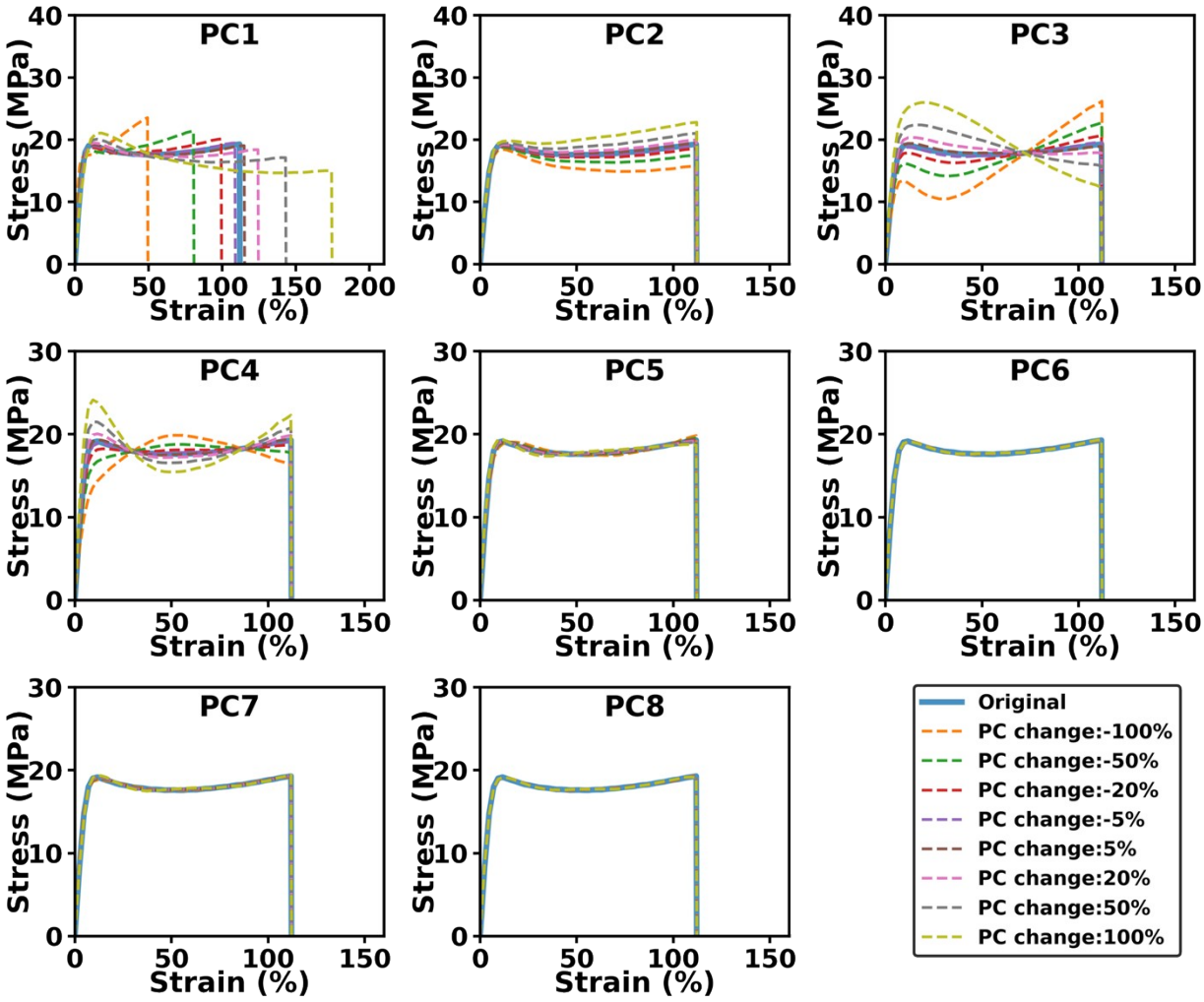

Figure S1. The impact of the changes in each principal component (PC1 to PC8) on the reconstructed stress-strain curves of a strong/tough sample.

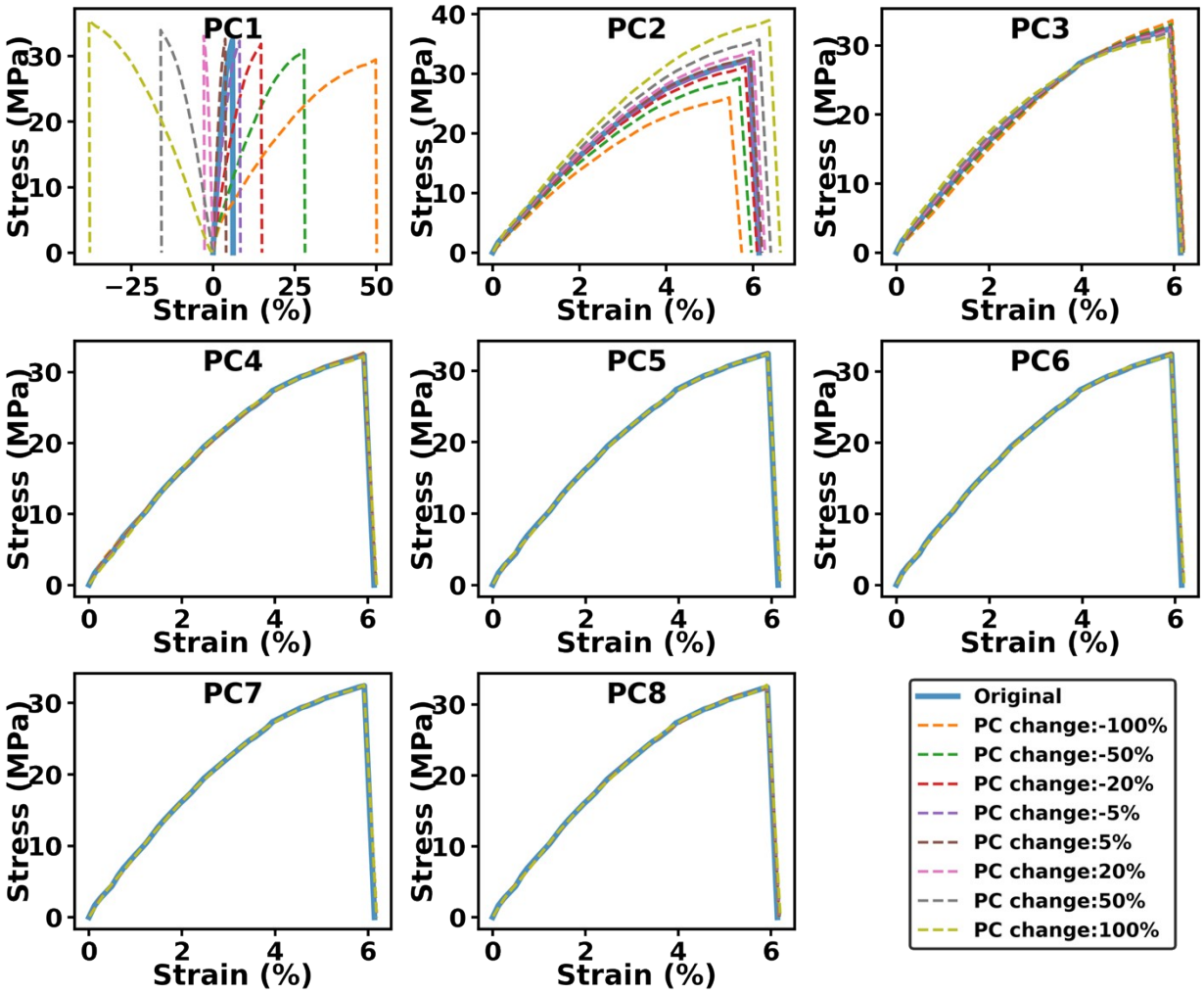

Figure S2. The impact of the changes in each principal component (PC1 to PC8) on the reconstructed stress-strain curves of a hard/brittle sample.

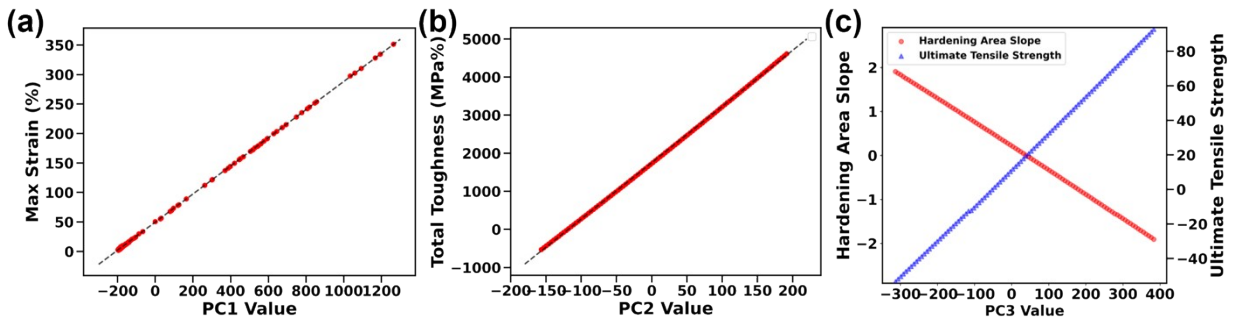

Figure S3. (a) Relationship between various PC1 values and the corresponding maximum strain across the entire dataset. **(b)** Relationship between different PC2 values and the corresponding toughness. Based on S-S curves in **Figure 4** and **Figure S1**, we control the other PC values and modify PC2 values to calculate the corresponding toughness. **(c)** Relationship between different PC3 values vs. the corresponding slope of the strain Hardening Area (red) and Yield Strength (blue). Based on S-S curve data in **Figure 4** and **Figure S1**, control the other PC values and modify PC3 values to calculate the corresponding features.

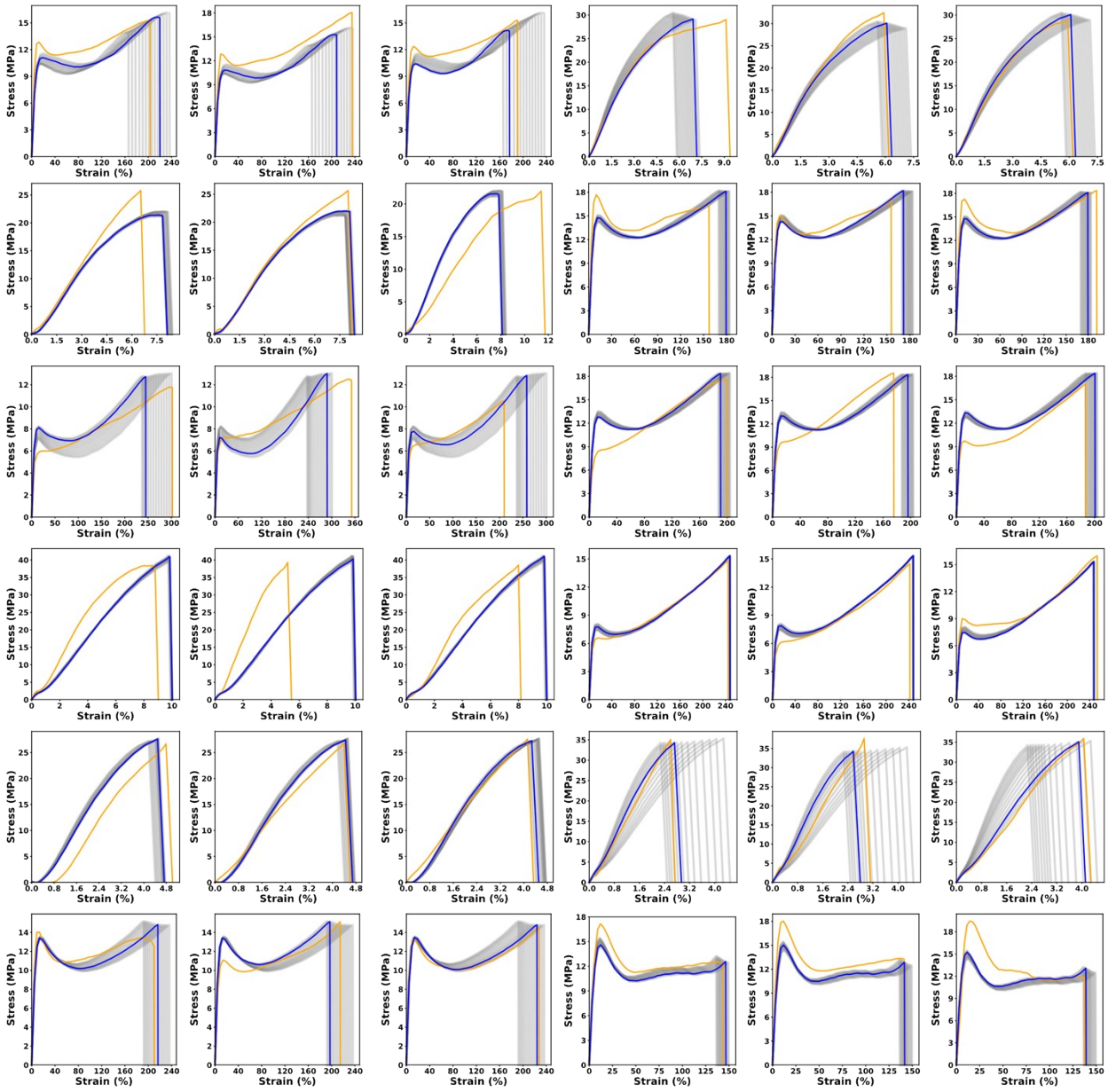

Figure S4. Comparison between 36 original (yellow) and the predicted (blue) S-S curves of the stress-strain curves (12 ink formulations). The predicted curves were reconstructed from the predicted PCs by the MLP model with the physics-informed descriptors. To effectively adapt to the variations originated from the experimental and testing conditions, the e_values were varied from -2 to 2 to reconstruct the S-S curves (grey lines).

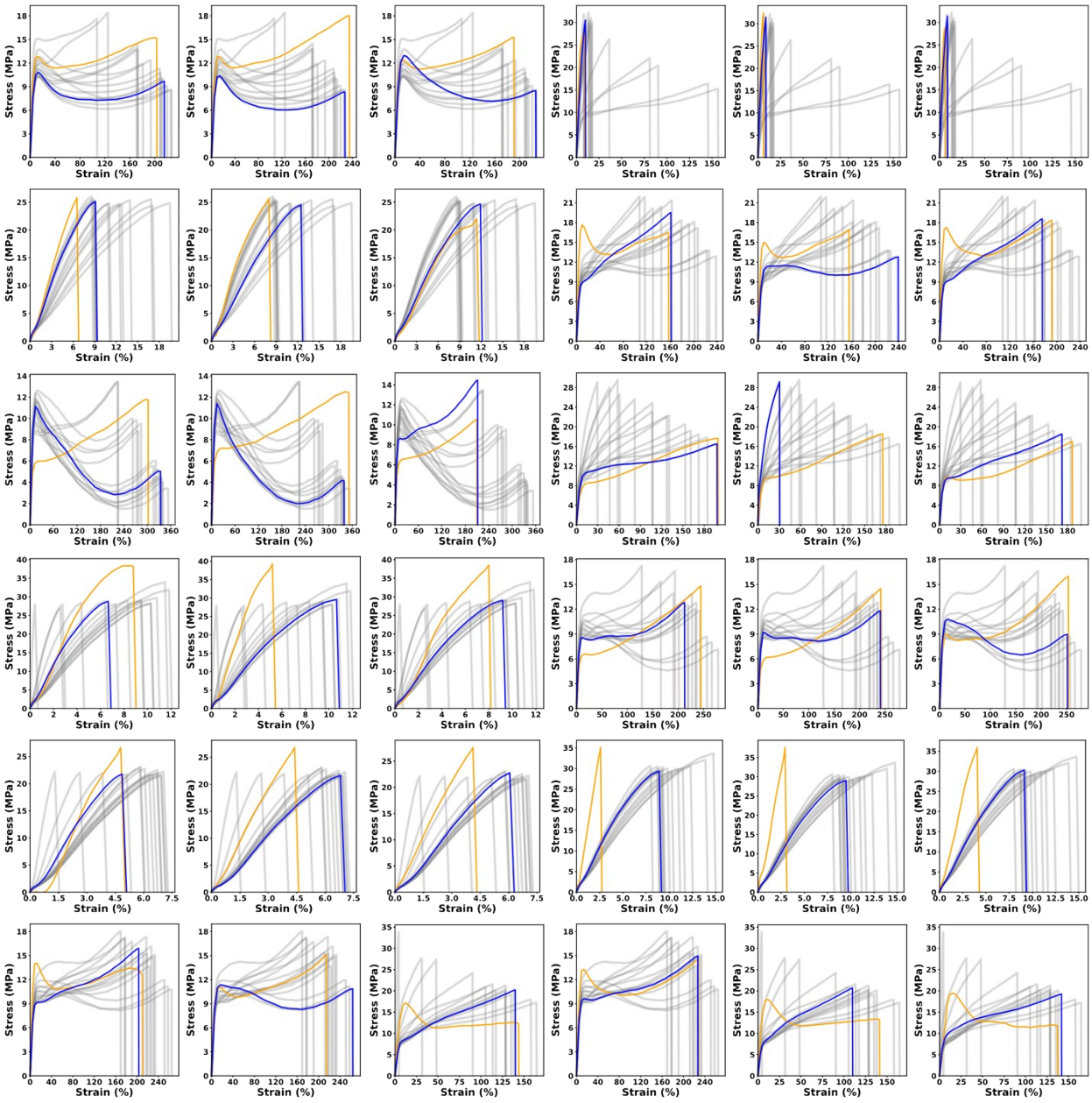

Figure S5. The comparison between original curve (yellow) and reconstructed (blue) stress-strain curves of the test set (12 ink formulation consisting of 36 stress-strain curves) using MLP model without the physics-informed descriptors. To effectively adapt to the variations originated from the experimental and testing conditions, the e_value varying from -2 to 2 were further incorporated to reconstruct the stress-strain curves (grey lines).

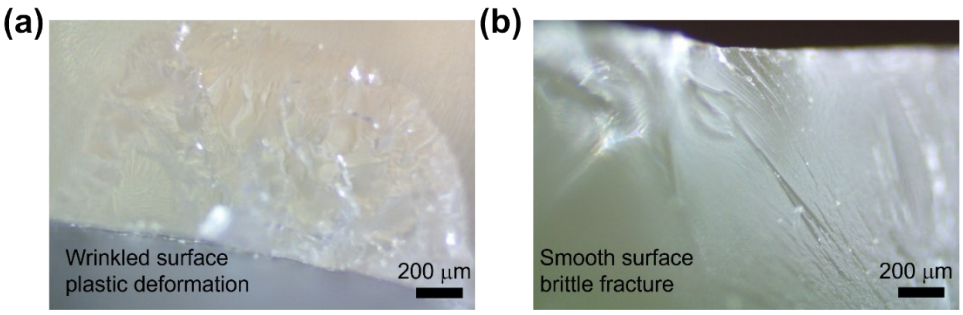

Figure S6. Digital microscope images of fracture surfaces for samples corresponding to the formulations shown in Figures 6b and Figure 6d, respectively. (a) An image taken on surface of a ductile sample shown in Figure 6b, revealing a ductile fracture surface. (b) An image taken on surface of a brittle sample shown in Figure 6d, exhibiting a smoother surface indicative of a brittle failure mechanism.

Supplementary Tables

Table S1: Extracted and calculated descriptors for the six monomers.

Physics-informed descriptors	HA	IA	NVP	AA	HEAA	IBOA
Molecular Weight	222.24	184.27	111.14	72.06	115.3	208.3
Lipophilicity	1.8	4.2	0.4	0.3	-0.6	3.9
Hbond_donor	1	1	0	1	2	0
Hbond_acceptor	4	2	1	2	2	2
Rot_bond	7	1	1	1	3	3
Polar Surface Area	55.8	37.3	20.3	37.3	49.3	26.3
Heavy Atoms	16	5	8	5	8	15
Complexity	221	55.9	120	55.9	90.4	306
Total Energy	-766.6	-581.7	-364	-267.2	-401.2	-657.9
Solubility_dipole	17.49	15.57	17.87	16.04	17.34	17.99
Solubility_p	5.66	4.14	10.39	13.39	9.07	4.10
Solubility_h	12.09	4.89	8.09	17.91	15.55	4.86
Solubility_sqrt_MJperm3	22.00	16.83	22.20	27.52	25.00	19.08

Table S2: MLP model performance evaluation based on PCs without physics-informed descriptors.

PCs	R^2	RMSE	MAE	Max	Min	Range	RMSE/Range
1	0.91	132.66	63.36	1263.16	-197.45	1460.61	9.08%
2	-0.11	21.03	15.94	65.84	-23.65	89.49	23.50%
3	0.37	10.77	8.79	31.38	-26.13	57.51	18.72%
4	0.13	6.66	5.16	16.21	-12.54	28.76	23.15%
5	-0.37	3	2.11	3.4	-6.33	9.72	30.86%
6	-0.37	1.29	0.99	2.84	-2.96	5.79	22.27%
7	-0.1	1.71	1.17	6.96	-2.71	9.67	17.68%
8	-0.4	0.67	0.46	1.13	-1.61	2.74	24.45%

Table S3: MLP model performance evaluation based on the tensile strength and toughness without physics-informed descriptors.

Metric	R^2	RMSE	MAE	Max	Min	Range	RMSE/Range
Fracture strength	0.52	6.27	4.91	39.29	11.76	27.53	28.70%
Toughness	0.38	1.50	1.21	10.48	4.03	6.45	21.95%

References

(1) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; et al. PubChem 2023 update. *Nucleic Acids Research* **2022**, *51* (D1), D1373-D1380. DOI: 10.1093/nar/gkac956 (acccessed 7/6/2023).

(2) Bertz, S. H. The first general index of molecular complexity. *Journal of the American Chemical Society* **1981**, *103* (12), 3599-3601. DOI: 10.1021/ja00402a071.

(3) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *Journal of Cheminformatics* **2011**, *3* (1), 33. DOI: 10.1186/1758-2946-3-33.

(4) Van Krevelen, D. W.; Te Nijenhuis, K. Chapter 7 - Cohesive Properties and Solubility. In *Properties of Polymers (Fourth Edition)*, Van Krevelen, D. W., Te Nijenhuis, K. Eds.; Elsevier, 2009; pp 189-227.

(5) Yang, C.; Kim, Y.; Ryu, S.; Gu, G. X. Prediction of composite microstructure stress-strain curves using convolutional neural networks. *Materials & Design* **2020**, *189*, 108509.