Supplementary Information

Large language model-enabled machine learning for high-performance Nd-Fe-B permanent magnet design

Zhi Yang [a, b, *], Xiaoxiao Zhang [a, b], Weiqiang Liu [a, b], Jie Zhang [a, b], Mengya Tian [a, b], Yingying Ma [a, b], Haibo Xu [c], Xiao Chi [c], Ming Yue [a, b, *]

[a] *State Key Laboratory of Materials Low-Carbon Recycling, Beijing University of Technology, Beijing 100124, China.*

[b] *Key Laboratory of Advanced Functional Materials, Ministry of Education of China, College of Materials Science and Engineering, Beijing University of Technology, Beijing 100124, China.*

[c] *Ganjiang Innovation Academy, Chinese Academy of Sciences, Ganzhou 341119, China.*

*Corresponding author: yangzhi@bjut.edu.cn (Zhi Yang), yueming@bjut.edu.cn (Ming Yue).

Section S1. Schematic representation of dual-tower neural network architectures

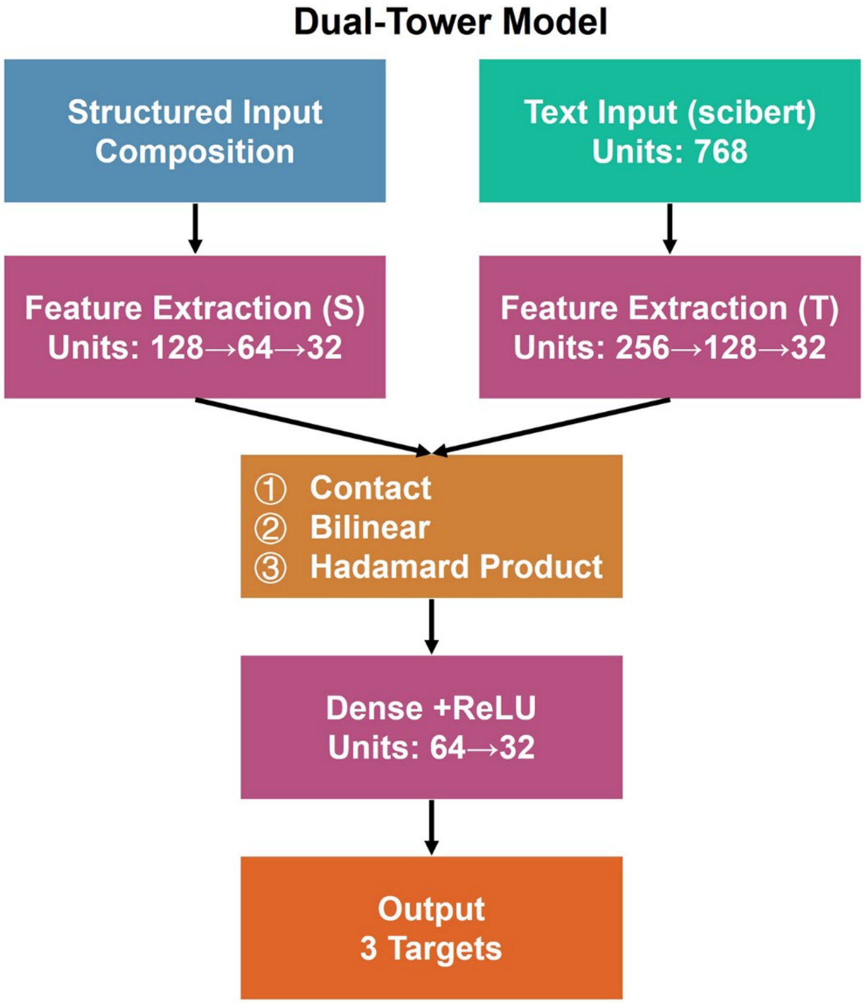

Figure S1. Schematic representation of dual-tower neural network architectures employing concatenation, bilinear, and Hadamard-product fusion mechanisms.

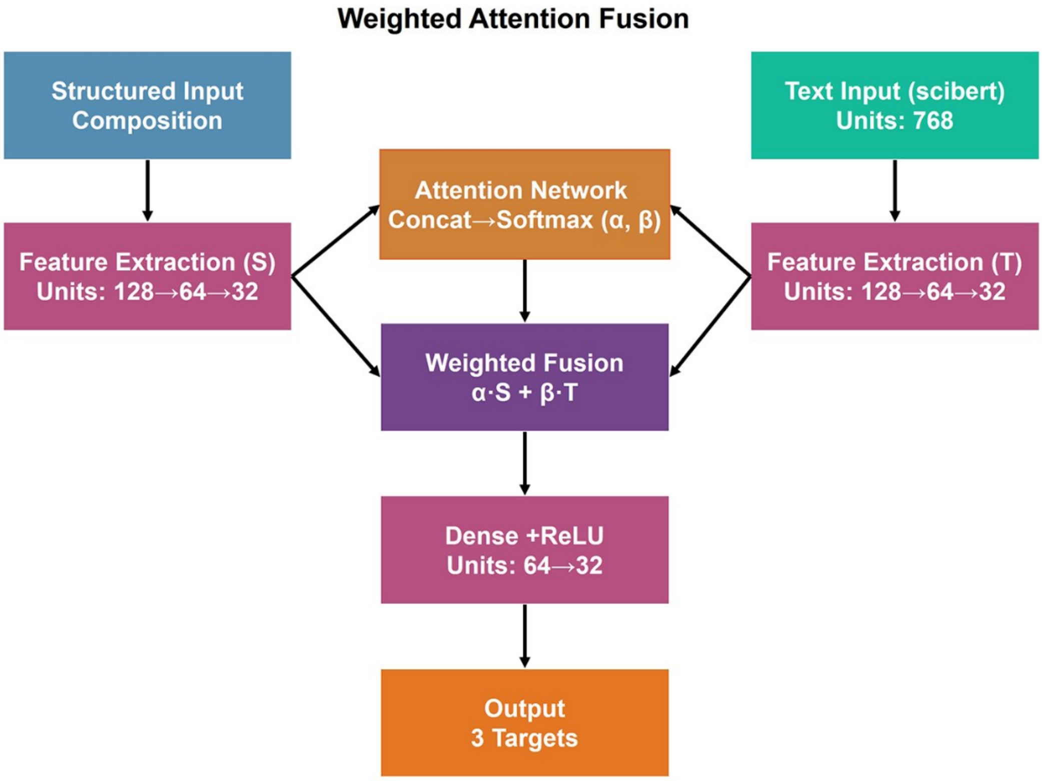

Figure S2. Dual-tower network architecture implementing weighted-attention fusion, where learnable attention coefficients modulate the relative contributions of each modality.

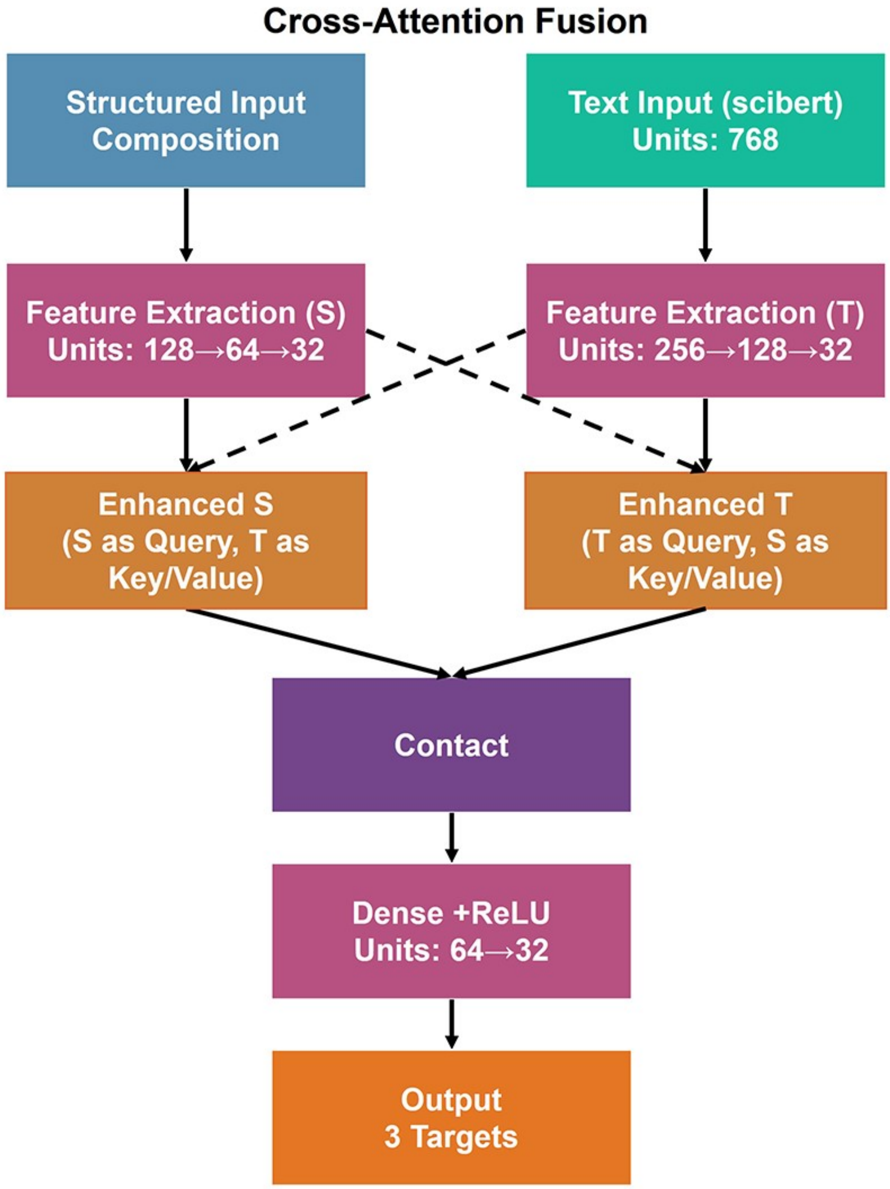

Figure S3. Dual-tower network architecture utilizing cross-attention fusion, enabling bidirectional information exchange between structured and textual representations.

Section S2. Hyperparameter sensitivity analysis

The analysis investigated the effects of batch size, dropout rate, learning rate, and network depth, with each parameter varied independently while keeping the others constant. Model performance was assessed using the average R^2 value from simultaneous predictions of three magnetic properties. Detailed structural parameters for each configuration are summarized in Table S1.

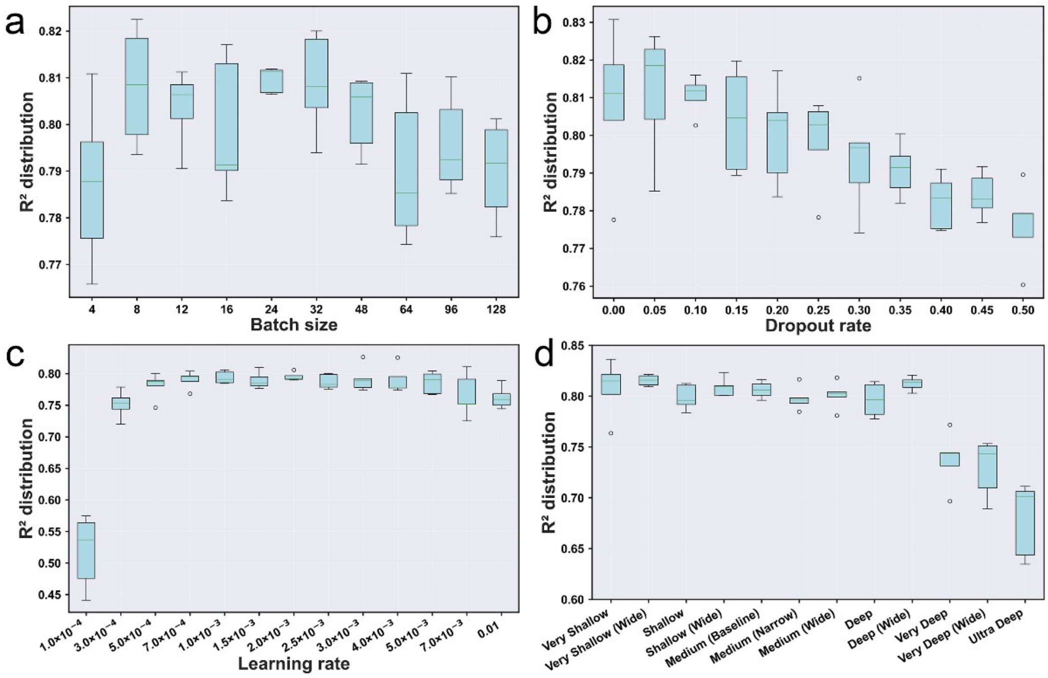

Figure S4. Hyperparameter sensitivity analysis for the gated fusion dual-tower neural network. Box plots showing the influence of (a) batch size, (b) dropout rate, (c) learning rate, and (d) network depth on model performance. Each configuration was evaluated across five independent runs with different random seeds.

Section S3. Cosine similarity analysis

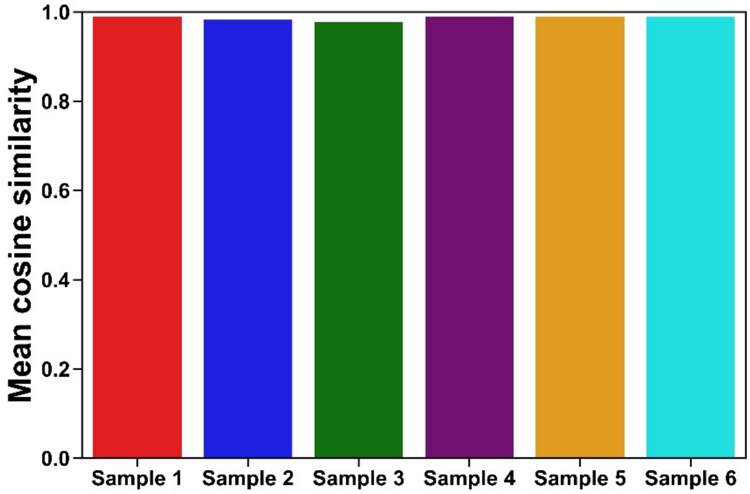

Figure S5. Mean cosine similarity between each of the six experimentally validated compositions and their five most similar counterparts in the training dataset.

Table S1. Architectural configurations of dual-tower neural networks with different depths. The table summarizes the number of layers and the corresponding number of neurons in each layer for both the composition and text input branches.

Configuration name	Composition input	Text input
Very Shallow	[32]	[64]
Very Shallow (Wide)	[64]	[128]
Shallow	[64, 32]	[128, 32]
Shallow (Wide)	[96, 32]	[192, 32]
Medium (Baseline)	[128, 64, 32]	[256, 128, 32]
Medium (Narrow)	[96, 64, 32]	[192, 128, 32]
Medium (Wide)	[160, 80, 32]	[320, 160, 32]
Deep	[128, 96, 64, 32]	[256, 196, 128, 32]
Deep (Wide)	[256, 128, 64, 32]	[512, 256, 128, 32]
Very Deep	[128, 96, 64, 48, 32]	[256, 192, 128, 96, 32]
Very Deep (Wide)	[256, 192, 128, 64, 32]	[512, 384, 256, 128, 32]
Ultra Deep	[128, 96, 80, 64, 48, 32]	[256, 192, 160, 128, 96, 32]