Supplementary materials

S.1 Architecture of VAE model

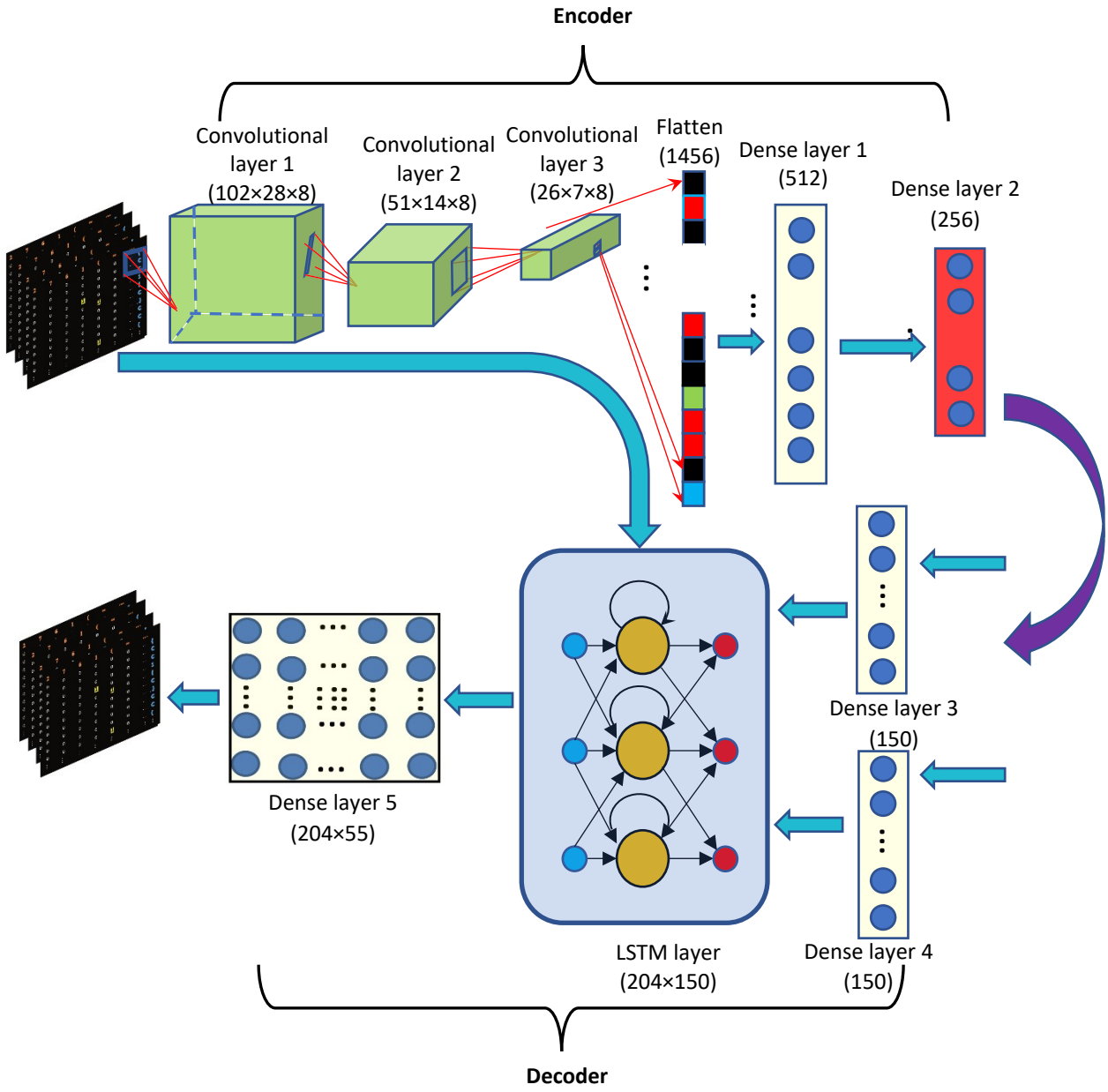

Fig. S1 Basic pipeline structures for the network of VAE model (the number in the bracket represent the number of neurons; the red color of the dense layer represents the output of latent space). All the activation functions for all layers are "Relu" except the last dense layer ("Softmax").

S.2 Hyperparameters selection

1. TL-VAE Model: Two crucial parameters, namely, the dimension of the latent space and the loss function, were optimized for the TL-VAE model. (a) The dimension of the latent space is a pivotal

hyperparameter in the VAE model. Conventionally set as 2, implying a one-dimensional Gaussian distribution with one parameter each for mean and standard deviation, it may fall short in capturing all features of complex inputs. For a multivariate Gaussian distribution, an adjusted dimension of space might be more appropriate in such cases. To validate this, we introduced a metrics that we applied before, i.e., average cosine similarity, which can be defined as

$$sim_{avg} = \frac{\sum_{i=1}^{n} \frac{\mathbf{v}_{inp} \cdot \mathbf{v}_{out}}{\left\|\mathbf{v}_{inp}\right\| \left\|\mathbf{v}_{out}\right\|}}{n} \qquad \backslash* \text{ MERGEFORMAT (1)}$$

where $\boldsymbol{v}_{inp}$ and $\boldsymbol{v}_{out}$ represent the binary matrices corresponding to the input SMILES and output SMILES, respectively. Next, we experimented with three latent space dimensions (8, 128, and 256) and evaluated them using 220 unique SMP or vitrimer monomers, considering the average cosine similarity. Table R1 illustrates that the respective average similarities are 58.92%, 70.17%, and 93.48%. We introduced a metric: percentage of valid monomer SMILES, signifying the percentage of mapped samples with a cosine similarity exceeding 80% compared to input samples. The VAE model with 256 latent dimensions showcased superior performance, prompting its selection for vector dimensions in the latent space. Furthermore, considering the complexity of SMPs relative to drug molecules, for which a latent dimension of 196 was used in previous VAE encoding [Gómez-Bombarelli et al., 2018], we added 60 dimensions to the latent space.

2. ANN Model: For the ANN model, we utilized an automatic parameter optimization method, specifically, the gridsearchcv method (from Scikit-learn). Employed parameters are itemized in Table R2, with "mean absolute error" employed as the scoring metric during the process.

Table R1 Performance comparison for 3 different training datasets. Percentage of valid monomer SMILES means the percentage of the mapped samples possessing the cosine similarity exceeds 80% with input samples.

Latent dimension	Average similarity (%)	Percentage of valid monomer SMILES (%)
8	58.92	0.45
128	70.17	32.72
256	93.48	93.18

Table R2 Employed parameters in gridsearch (random state=24).

Item	Values in gridsearchcv	Optimized parameter
Batch size	32, 64, 128, 256	32
Optimizer	Adam, Rmsprop, SGD	Adam
Split ratio	70/30,80/20	80/20

This part was then integrated into the supplementary materials.

S.3 DMA and DSC test results

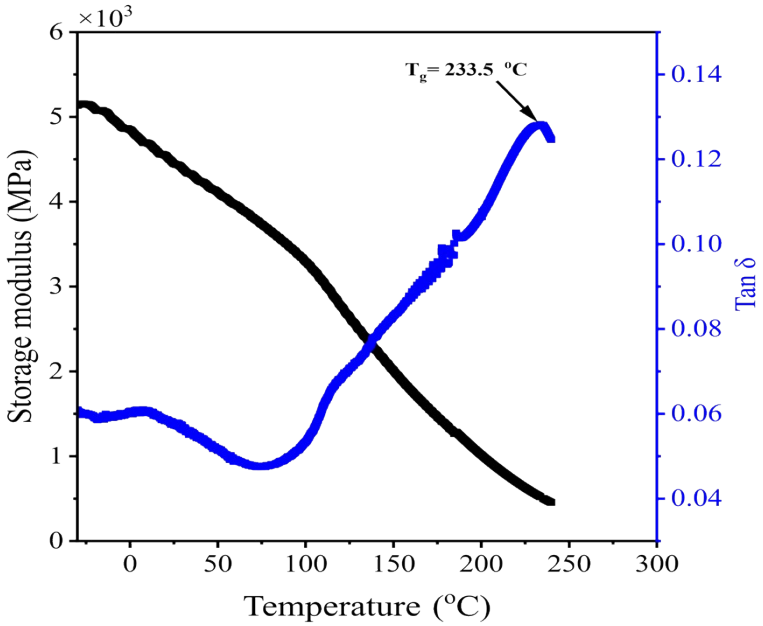

Fig. S2. DMA Test Results

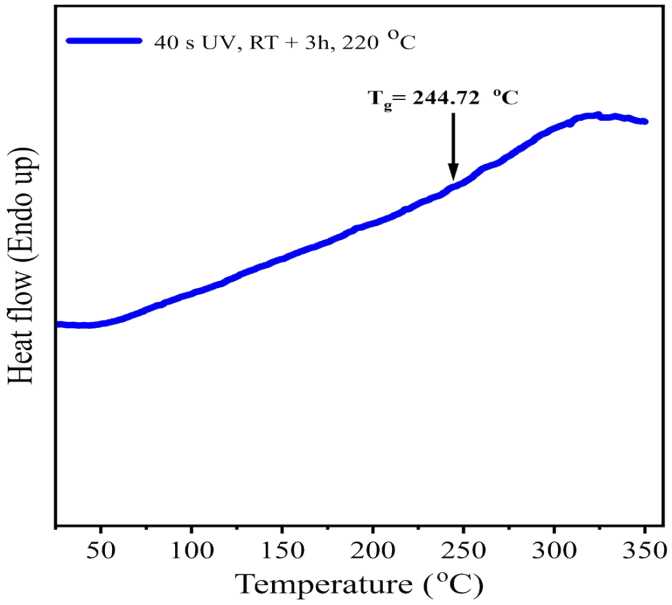

Figure S3. DSC Test Results

The details can be found in Excel files.