The detailed structure of the iEnhancer-DCLA model is presented in Figure S1. The sequence length we input is 200bp. Sequence is represented as 200*100-D matrix by embedding layer, and transformed into 50*128-D matrix by sequential (CNN) and bidirectional LSTM modules, then transformed into 128-D feature vectors by attention mechanism module, and finally into the fully connected layers for prediction.

| Layer (type) | Output Shape | Param # |
|---|---|---|
| input_5 (InputLayer) | [(None, 200)] | 0 |
| embedding_4 (Embedding) | (None, 200, 100) | 1638500 |
| sequential_4 (Sequential) | (None, 50, 128) | 468864 |
| bidirectional_4 (Bidirection | (None, 50, 128) | 98816 |
| att_layer_4 (AttLayer) | (None, 128) | 8320 |
| batch_normalization_18 (Batc | (None, 128) | 512 |
| dropout_18 (Dropout) | (None, 128) | 0 |
| dense_8 (Dense) | (None, 64) | 8256 |
| batch_normalization_19 (Batc | (None, 64) | 256 |
| activation_14 (Activation) | (None, 64) | 0 |
| dropout_19 (Dropout) | (None, 64) | 0 |
| dense_9 (Dense) | (None, 1) | 65 |

Figure S1. Dimension changes of iEnhaner-DCLA under each module.

UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction) is a dimension reduction technique. As shown in Figure S2, the left figure shows the two-dimensional feature representation of the sequence at the embedding layer. The right figure shows the feature representation after the action of the attention mechanism.

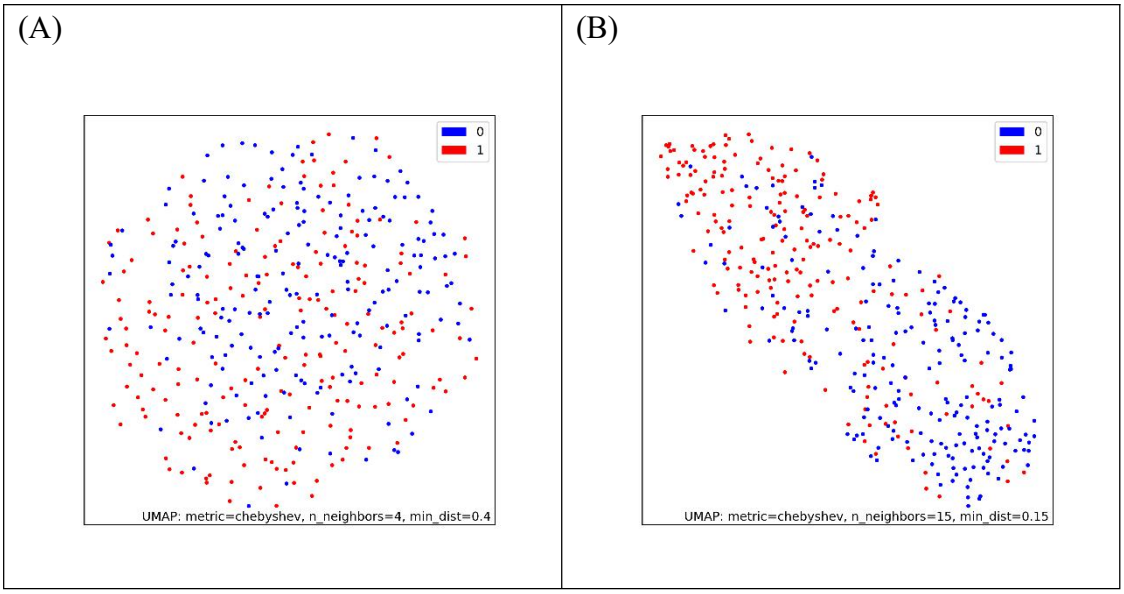


Figure S2. Two-dimensional feature representation of enhancers and non-enhancers' data before and after model training.