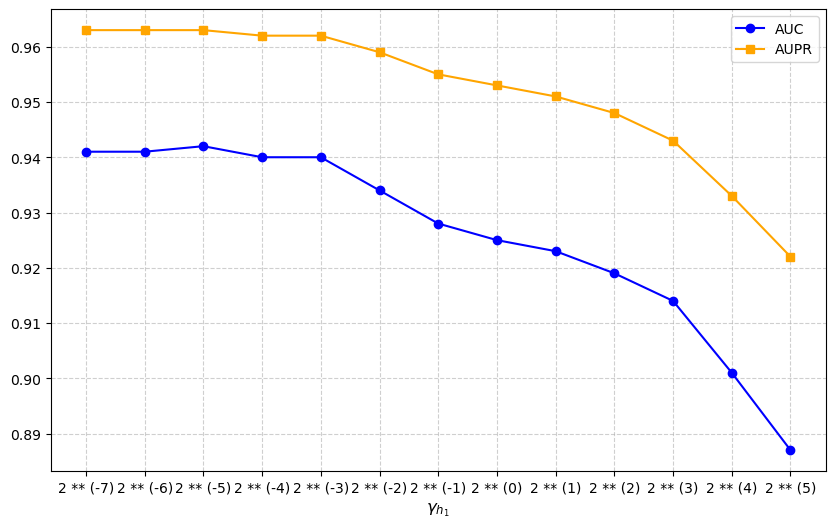
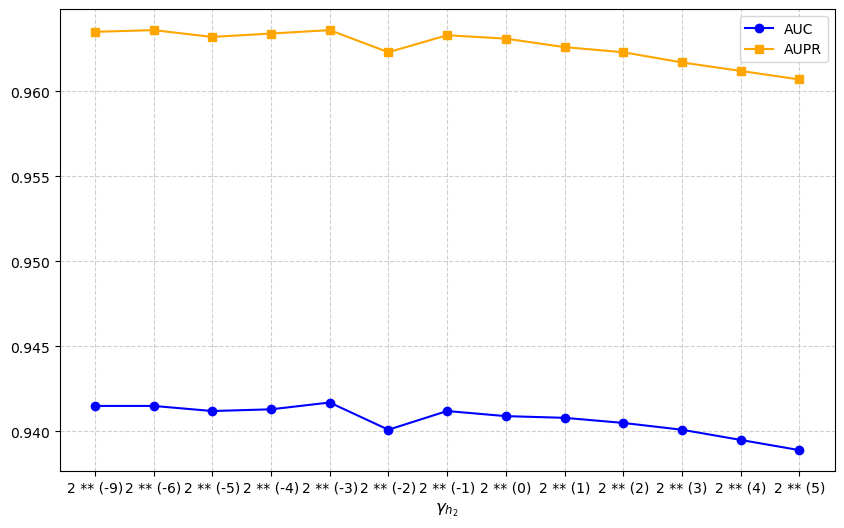
Below are the supplementary information of the MKDTI: Predicting drug-target interactions via multiple kernel fusion on graph attention network:

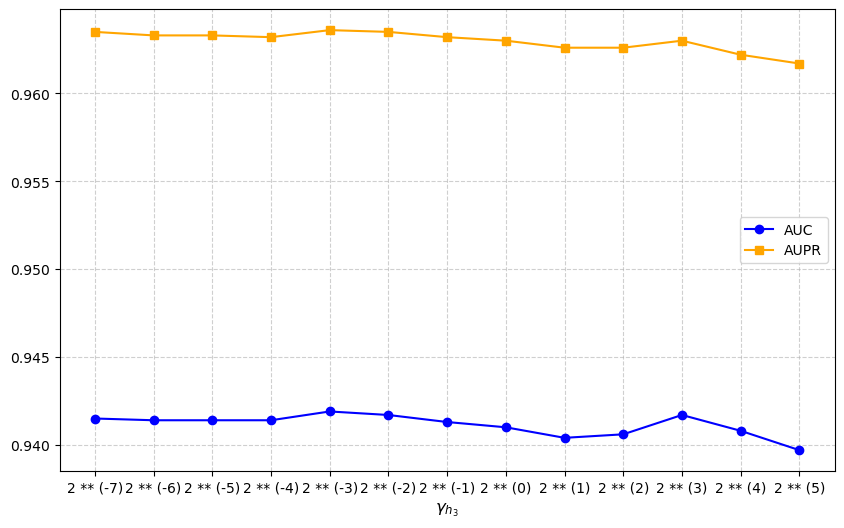
## Supplementary Fig



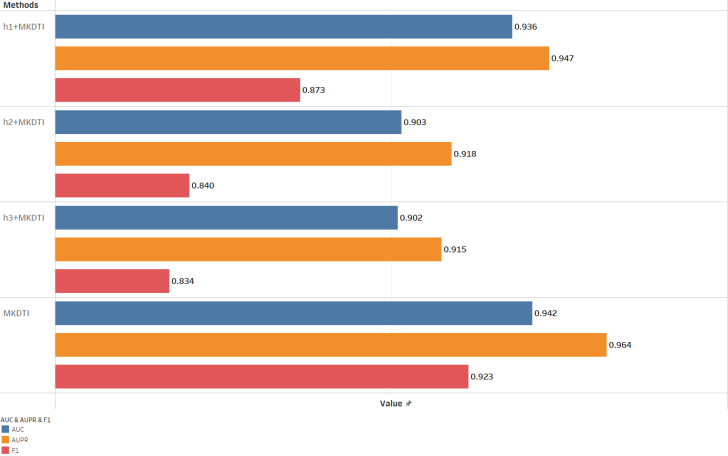
**Supplementary Fig. 1.** Effects of parameter on MKDTI performance measured by AUPR and AUC.



**Supplementary Fig. 2.** Effects of parameter on MKDTI performance measured by AUPR and AUC.



**Supplementary Fig. 3.** Effects of parameter on MKDTI performance measured by AUPR and AUC.



**Supplementary Fig. 4.** Performance of the model under different kernel

## Supplementary Tables

**Supplementary Table 1.** Performance of the model with different embedding dimensions

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Methods** |  |  |  | **AUPR** | **AUC** |
| MKDTI-1 | 1024 | 512 | 256 | 0.961 | 0.939 |
| MKDTI-2 | 512 | 256 | 128 | 0.963 | 0.942 |
| **MKDTI3** | **384** | **192** | **96** | **0.964** | **0.942** |
| MKDTI-4 | 256 | 128 | 64 | 0.963 | 0.941 |
| MKDTI-5 | 128 | 64 | 32 | 0.963 | 0.940 |

**Supplementary Table 2.** Performance of the model under different kernel

|  |  |  |  |
| --- | --- | --- | --- |
| Methods | AUC | AUPR | F1 |
| h1+MKDTI | 0.936 | 0.947 | 0.873 |
| h2+MKDTI | 0.903 | 0.918 | 0.840 |
| h3+MKDTI | 0.902 | 0.915 | 0.834 |
| MKDTI | 0.942 | 0.964 | 0.923 |

## Supplementary paragraph

Link prediction module and loss function

To further enhance the prediction performance of DTIs, Ding et al. [29] suggest a Dual Laplacian Regularized Least Squares (DLapRLS) framework, which is inspired by LapRLS [34]. We use DLapRLS framework to predict drug-target interactions.

Firstly, we compute the normalised Laplace matrices  *a*nd by:

Here, and are the fusion kernels. and are diagonal matrices and are computed by:

Let be the F-paradigm, be the association matrix in the training set; and be the trainable matrices. The loss function is shown below:

Using the trainable matrices and combined with the information from the fusion kernel, the drug-target association prediction results are as follows:

During the forward propagation stage, we first establish the model with all trainable parameters and then compute the model's loss function. Using the Adam optimizer [35] to optimize the graphical attention network's parameters and the partial derivatives of the loss function as the DLapRLS framework to iteratively optimize the function are the two types of optimization techniques we employ during back propagation to optimize the loss function.

The following is the formula for the DLapRLS loss function's partial derivative with regard to parameter :

Let be equal to 0:

Similarly, the following is the formula for the DLapRLS loss function's partial derivative with regard to parameter :

Let be equal to 0:

Multi-kernel fusion based on graph attention network

We take rows of the embedding vectors of graph attention network l-layer as drug embedding vectors and rows as target embedding vectors , and use Gaussian interaction profile to obtain the kernel matrix of the corresponding embedding vectors. Set the corresponding bandwidth parameter .

By collecting the extracted kernel matrices from different layers, we have a collection of kernel matrices for drugs and targets.

As we mentioned before, these kernel matrices contain semantic information about various hop-neighbor aggregations, thereby enhancing the model's predictive capability. We use weighting coefficient to combine different kernels, with the following equation:

Similarly, the target fusion kernel can be computed by the following equation:

The and are matrices of the ith kernels in the collection of kernel matrices for drugs and targets, and and are the weights corresponding to each kernel matrices, which in this study is set .

Self-augmented multi-headed graph attention networks

Using the self-enhancing attention mechanism, the current node is treated as its own neighbor to obtain a single-head attention node embedding for each layer:

Next, node features are computed using k independent attention mechanisms and then aggregated by concatenation or averaging so that the information contained in the features can be extended:

Or

In our proposed MKGCN model, ​​multiple kernel fusion​​ is performed via ​​average weighting​​. Here, means the normalized attention coefficient of the k th attention head, means the normalization function sigmoid, || means the concatenation operation.

Here, means the normalized attention coefficient of the k th attention head, means the normalization function sigmoid, || means the concatenation operation.