UNIVERSITY OF HELSINKI
FACULTY OF MEDICINE

Network pharmacology for
precision medicine
Faculty of Medicine,
University of Helsinki

Techniques for Deep Fusing Multimodal Data

Based on *Deep Multimodal Data Fusion*[1]

Qinhan Hou, Aug 28th, Helsinki

[1] Fei Zhao, Chengcui Zhang, and Baocheng Geng. 2024. Deep Multimodal Data Fusion. ACM Comput. Surv. 56, 9, Article 216 (September 2024), 36 pages. https://doi.org/10.1145/3649447

Introduction

What is multimodal data, and why we need the fusion of them

- Multimodal data:
  - The world is represented by information in different mediums.
  - They share the same semantic information - Information Redundancy.
  - But also complementary.
  - Multimodality interpretation deliver the “fuller picture” of observed activity, making the model more robust and reliable.

Background

The evolution of multimodality fusion

Network pharmacology for precision medicine
Faculty of Medicine,
University of Helsinki

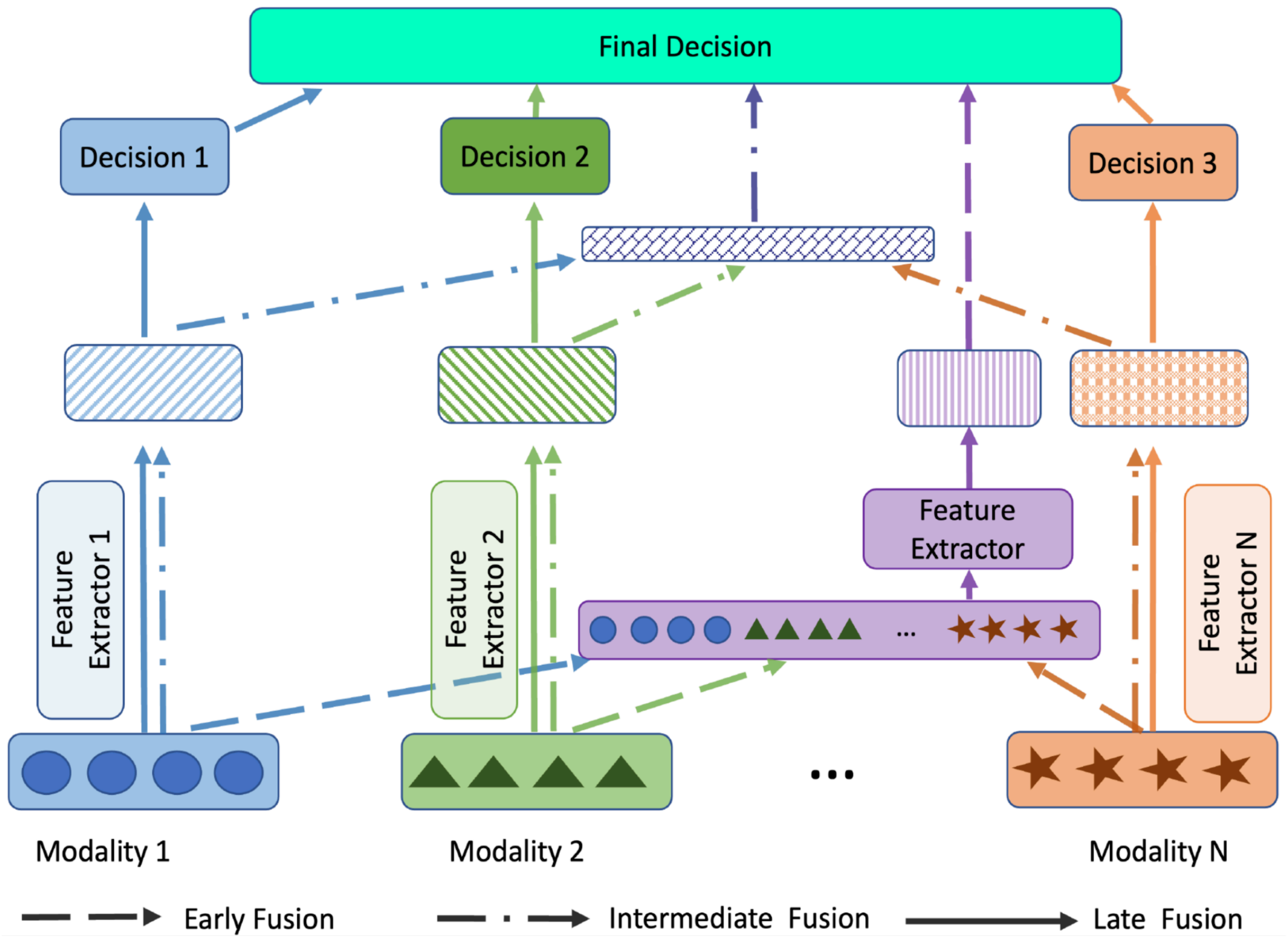

Fig 1. The conventional taxonomy categorizes fusion methods into three classes.

- Classical Machine Learning:
  - Hand-made feature engineering
  - Hard to capture the redundancy and complementation
- Deep Learning:
  - Data-driven feature representation
  - Implicit, mostly interlaced with feature representation
  - Rely on well-designed network topology and loss function

Deep Multimodal Fusion

Five classes of deep multimodal fusion techniques

UNIVERSITY OF HELSINKI
FACULTY OF MEDICINE

Network pharmacology for precision medicine
Faculty of Medicine,
University of Helsinki

- Deep multimodal fusion schema:
  - Encoder-decoder-based
  - Attention-based
  - Generative neural network-based
  - Graph neural network-based
  - Constraint-based

Deep Multimodal Fusion
Encoder-Decoder-based

UNIVERSITY OF HELSINKI
FACULTY OF MEDICINE
Network pharmacology for precision medicine
Faculty of Medicine,
University of Helsinki

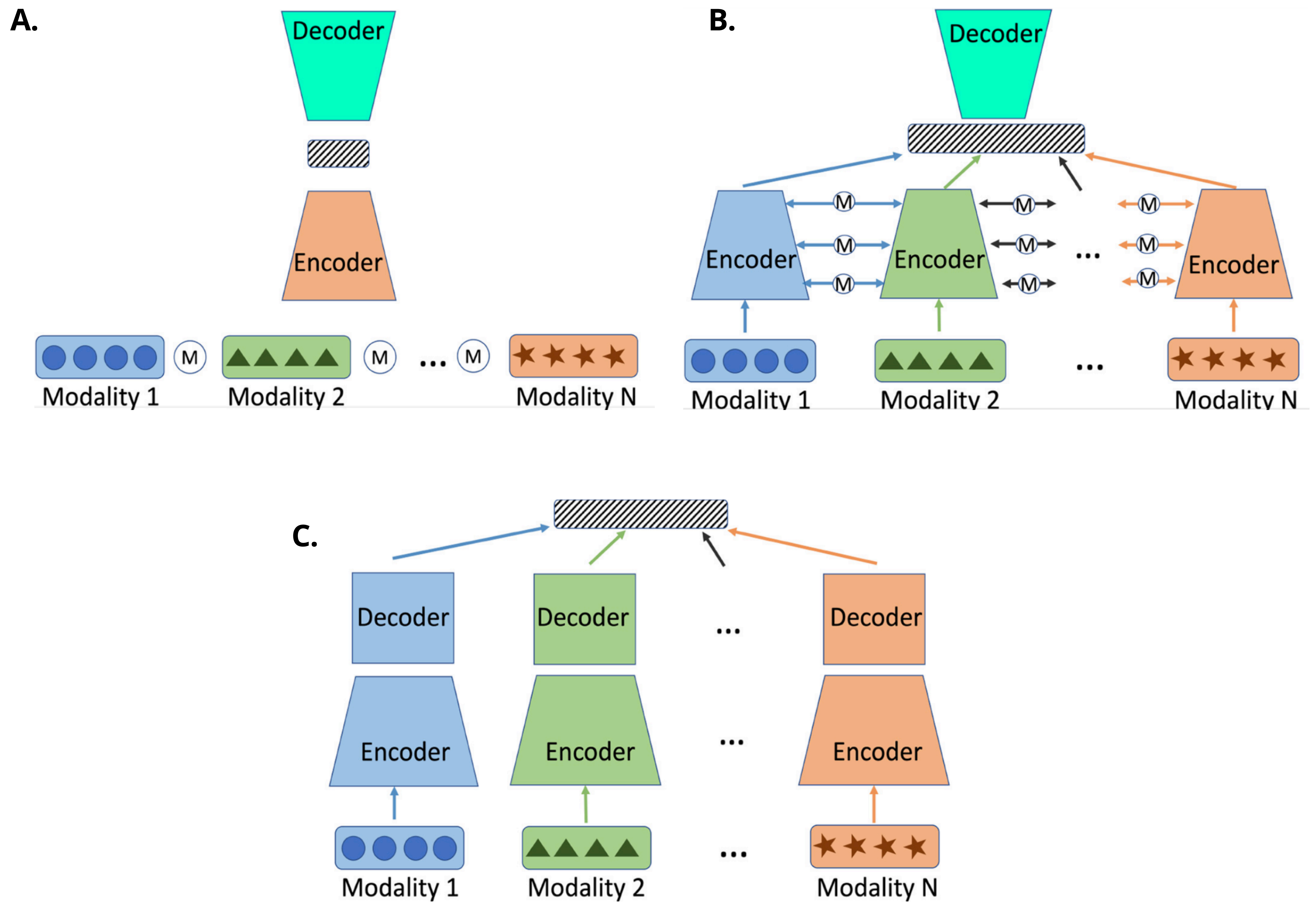

Fig 2. Visualizations of different fusion strategy in encoder-decoder-based scheme.
A) The raw-data-level fusion. B) The hierarchical feature fusion. C) The decision-level fusion.

- Encoder-Decoder:
  - Encoder: high-level feature extractor
  - Decoder: generate “prediction” from laten representations.
- Categories:
  - Raw-data-level fusion.
  - Hierarchical feature fusion.
  - Decision-level fusion.
- Key operation - merge:
  - Addition / Multiplication
  - Concatenation
  - Cross product

Deep Multimodal Fusion
Attention-based

Network pharmacology for precision medicine
Faculty of Medicine,
University of Helsinki

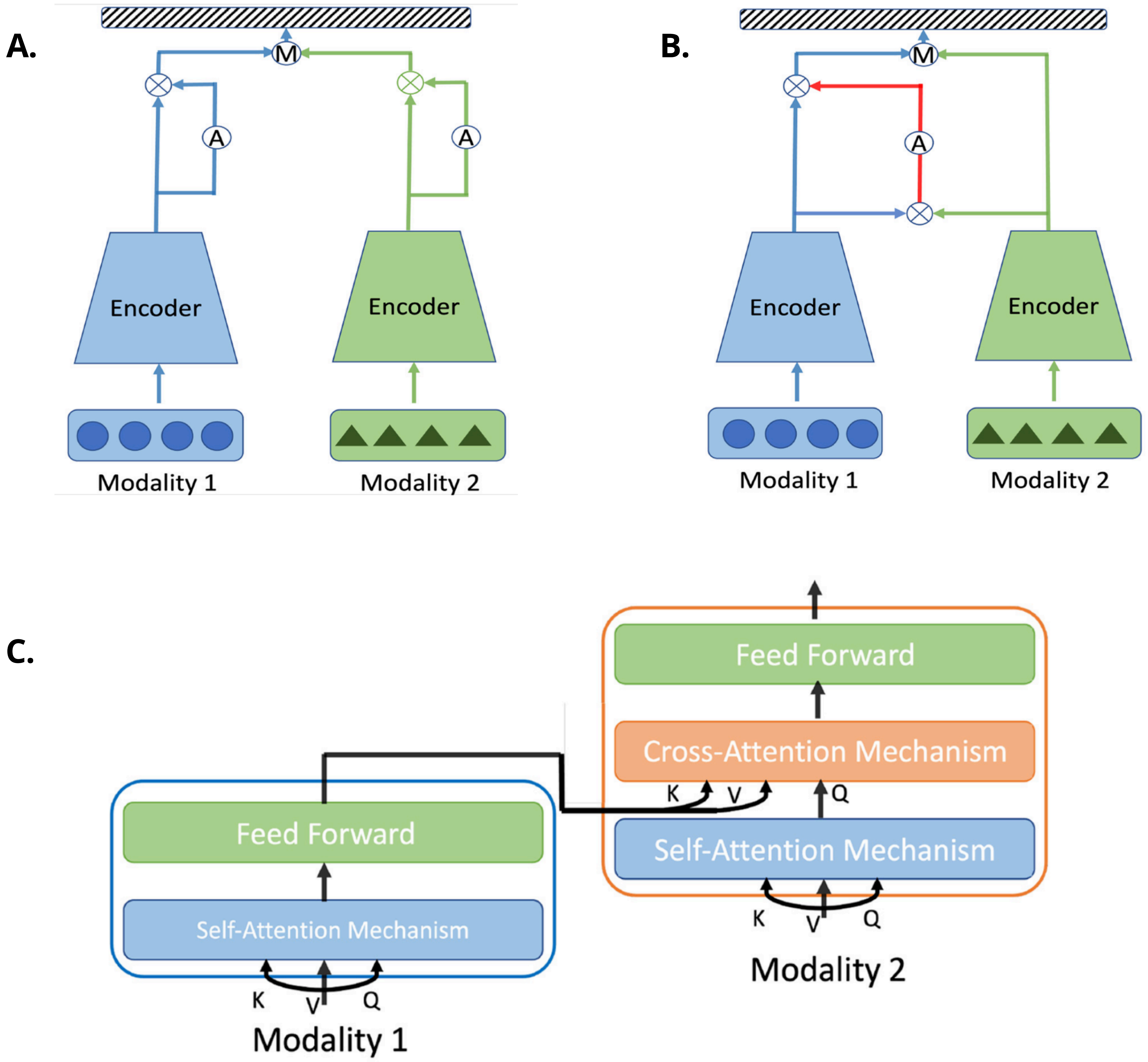

Fig 3. Visualizations of different fusion strategy in attention-based scheme.
A) The intra-modality attention. B) The inter-modality attention. C) The transformer-style attention.

- Attention mechanism:
  - Enable models to assign different weights on different parts in input data
- Categories:
  - Intra-attention
  - Inter-attention
  - Transformer-based
- Limitations:
  - Capacity - the number of modalities
  - Computation complexity

Deep Multimodal Fusion

Graph-based

Network pharmacology for precision medicine
Faculty of Medicine,
University of Helsinki

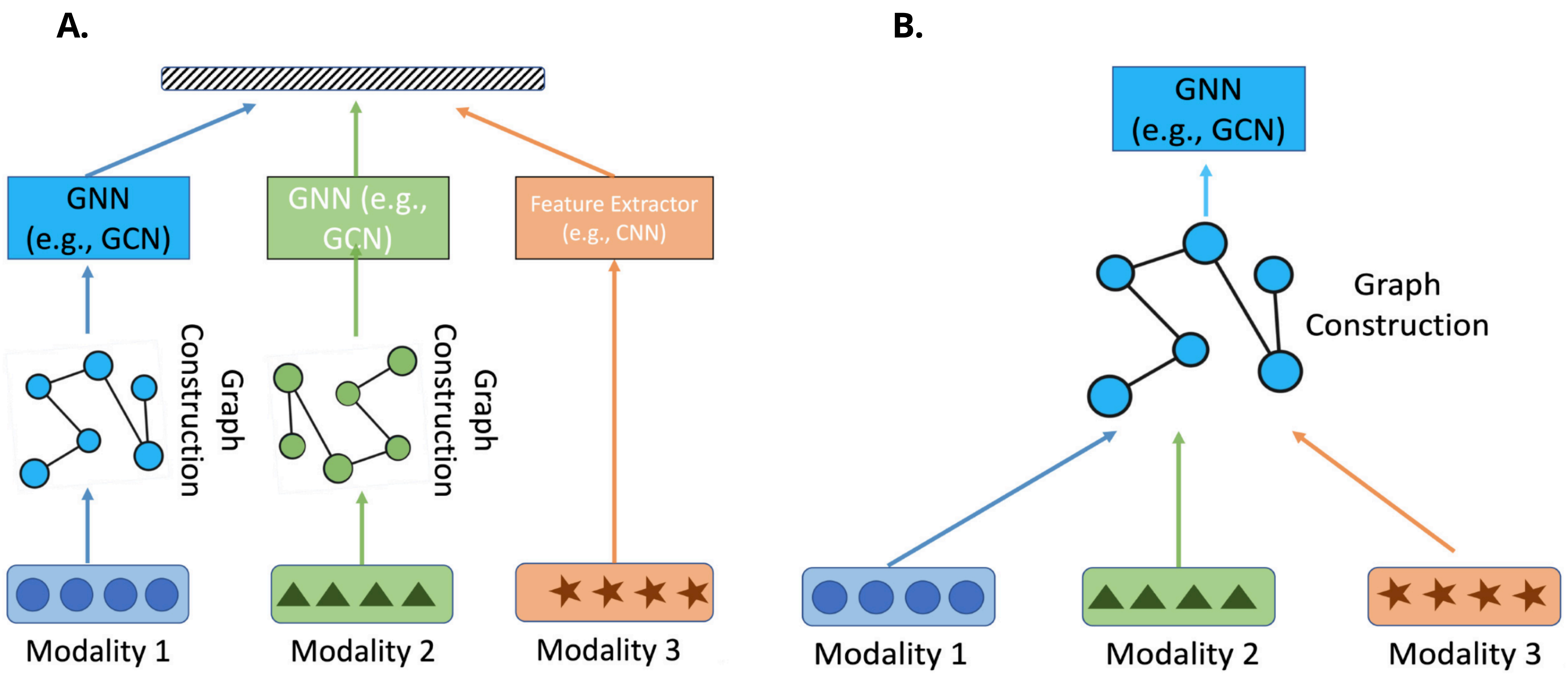

Fig 4. Visualizations of different fusion strategy in graph-based scheme. A) General schema. B) Fusion in graph construction.

- Graph-based method:
  - Handling relations between datapoint
- Categories:
  - General: graph built separately
  - Fused graph construction:
- Limitations:
  - The graph construction process depends on prior knowledge.
  - Time- and space-consuming.
  - Hard to be generalized.

Deep Multimodal Fusion

Generative neural network (GNN)-based

Network pharmacology for precision medicine
Faculty of Medicine,
University of Helsinki

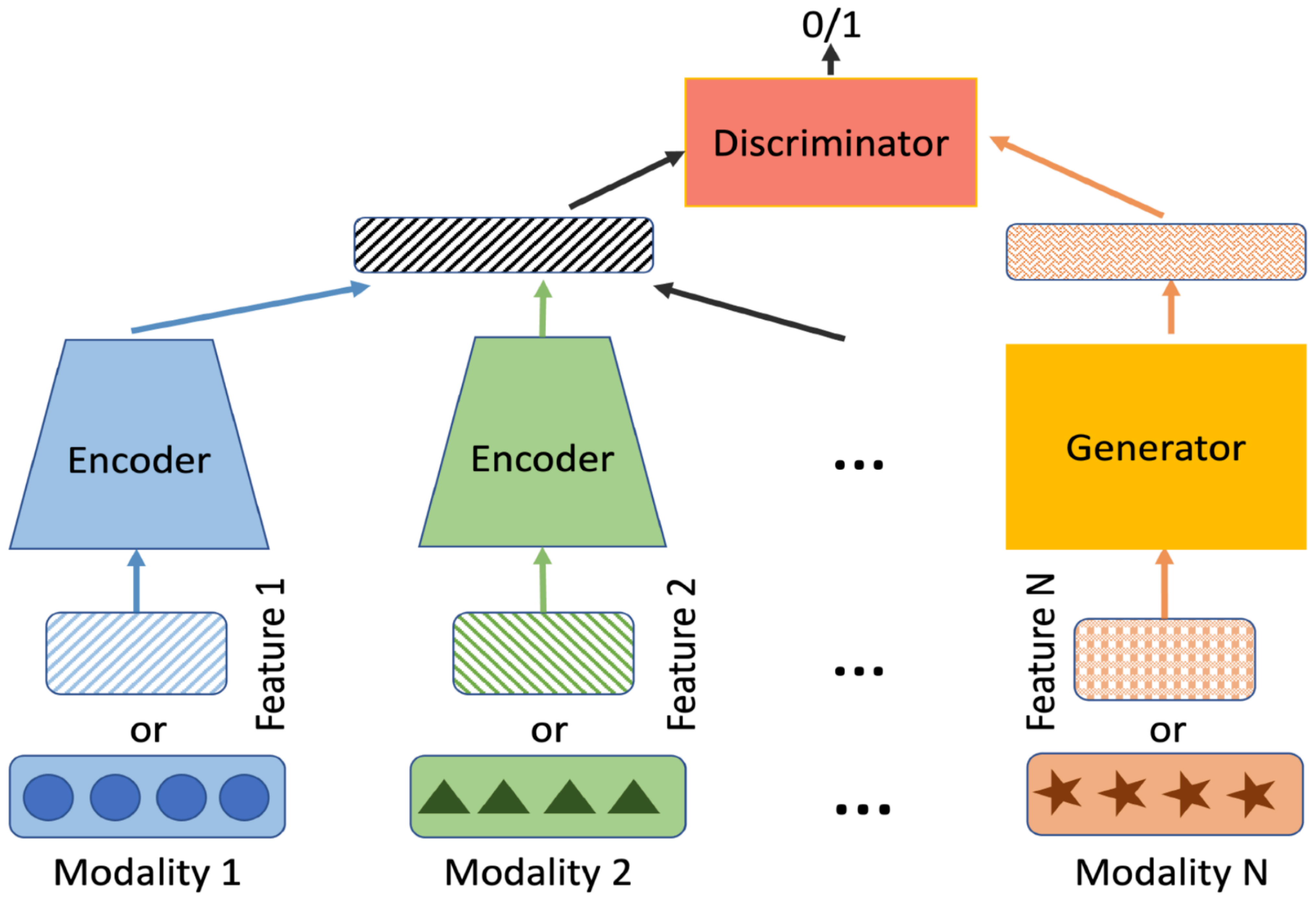

Fig 5. Visulization of general architecture of GNN-based methods.

- Generative neural network (GNN)
  - Learning data distributions for generation tasks / scenarios
  - Aim at handling missed, noisy or incomplete data
- GNN-based multimodal frameworks:
  - Synthesize the missing modality based on the other modalities
  - Compatible for merging the other generation methods, e.g. Diffusion
  - Tricky in training

Deep Multimodal Fusion

Constraint-based

UNIVERSITY OF HELSINKI
FACULTY OF MEDICINE
Network pharmacology for precision medicine
Faculty of Medicine,
University of Helsinki

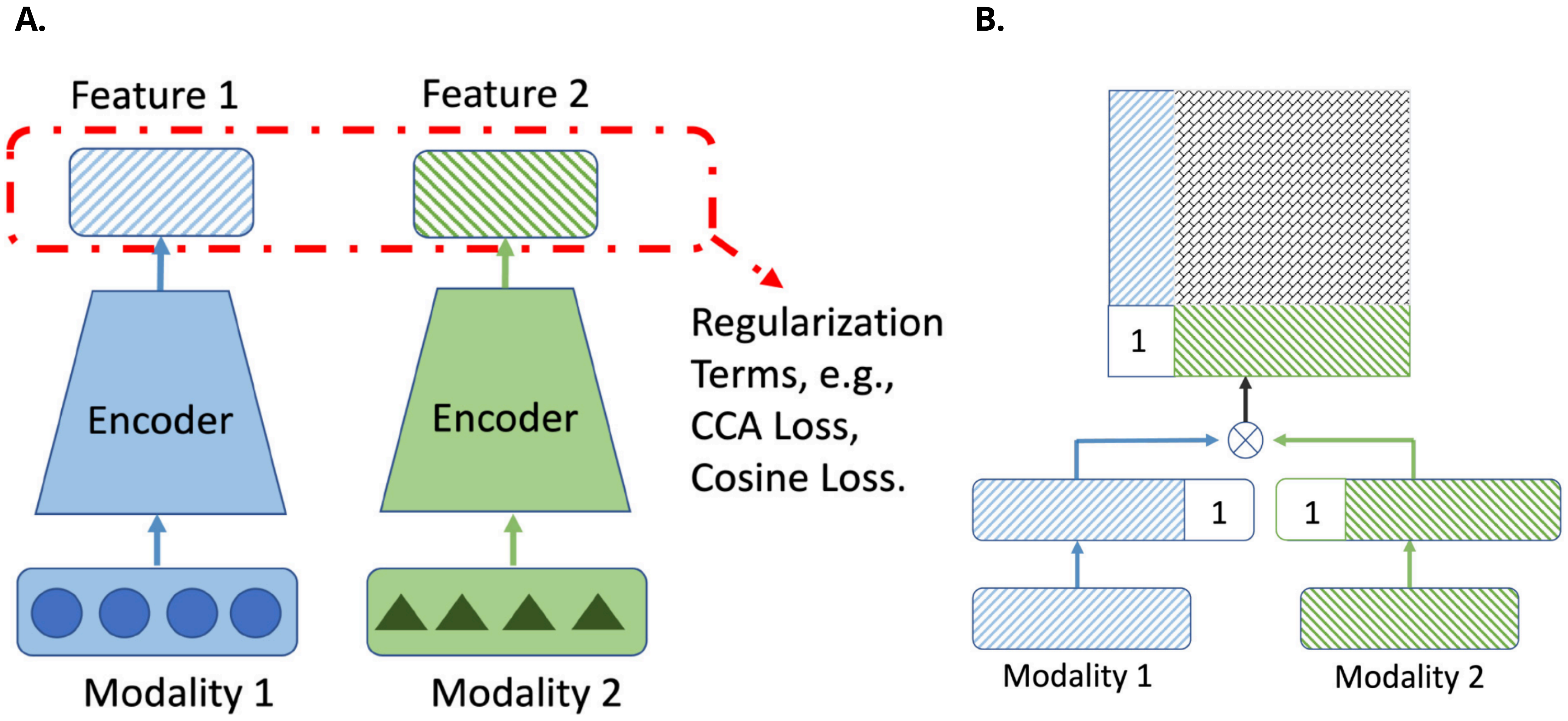

Fig 6. Visualizations of different fusion strategy in constraint-based scheme. A) Coordinated representations. B) Tensor fusion mechanism.

- Constraint-based methods:
  - learns separated but coordinated representations of each modality under certain constraints.
- Categories:
  - Coordinated representation
  - Tensor fusion
- Limitations:
  - Hard to extend the large amount of modalities
  - Highly relied on constraint design

Deep Multimodal Fusion

Applications & Challenges

UNIVERSITY OF HELSINKI
FACULTY OF MEDICINE

Network pharmacology for precision medicine
Faculty of Medicine,
University of Helsinki

- Applications:
  - Vision and languages, vision and sensors…
  - Others…
    - In biomedical field - multi-omics, different biomedical devices records
- Challenges:
  - Missing modality, or unbalanced modality contribution
  - Lack of data - high alignment requirements
  - Interpretability of the model