**1. 數據來源與特徵提取（上半部分）**

這部分展示了模型如何從不同數據庫獲取 **miRNA 和疾病的相關信息**，並計算它們之間的相似性。

**數據來源**

**miRNA sequence（miRNA 序列）**

來自基因數據庫，通過 **k-mers 方法** 提取特徵。

**MISIM database**

包含 miRNA 功能相似性（FS）。

**HMDD v3.0 dataset**

包含已驗證的 **miRNA-疾病關聯（A）**。

**MeSH database**

包含疾病的語義相似性（Sim1 & Sim2），通過 **有向無環圖（DAG）** 來建立疾病之間的關聯。

**特徵計算**

**miRNA functional similarity（FS）**

計算 miRNA 之間的功能相似性。

**Gaussian interaction profile kernel similarity（KM, KD）**

KM：計算 miRNA 之間的相互作用相似性。

KD：計算疾病之間的相互作用相似性。

**Integrate similarity for miRNA（SM）和疾病（SD）**

綜合以上不同的相似性數據，形成完整的特徵矩陣。

這些特徵共同組成 **miRNA 和疾病的綜合特徵表示**，用於後續的學習和預測。

**2. 多源特徵融合（Fusion of Multi-Source Features，中間部分）**

在獲取 **miRNA 和疾病的相似性特徵後**，這些特徵會被融合，形成模型的輸入數據。

這些特徵涵蓋 **序列相似性、功能相似性、疾病語義相似性**，讓模型能夠更準確地學習 miRNA-疾病的潛在關聯。

這一步類似於 **將不同來源的數據進行標準化與整合**，以便後續的機器學習模型能夠處理這些信息。

**3. Broad Learning System（BLS）建模與增量學習（下半部分）**

這部分展示了 **Broad Learning System（BLS）** 的結構，它是一種能夠支持增量學習的機器學習框架。

**BLS 結構**

**Instance（輸入層）**：來自多源特徵融合的輸入樣本。

**Z1 ~ Zn（特徵映射層）**：

這些是通過 BLS 進行學習的 **基礎特徵**。

這些特徵會被傳遞到增強層進行進一步的學習。

**Enhancement Nodes（增強節點，H1 ~ Hn）**：

這些節點可以根據新數據進行 **動態擴展**，支援增量學習。

與傳統神經網絡不同，BLS 允許添加新的特徵，而不會影響原有模型的穩定性。

**增量學習機制**

**Add Additional Enhancement Nodes**

當新數據進入時，BLS 允許動態擴展 **增強節點**，而不需要重新訓練整個模型。

這解決了 **災難性遺忘（Catastrophic Forgetting）** 的問題，使模型可以隨著新數據的增加而持續學習。

**4. 預測結果（Final-class）**

最終的預測結果 **Y** 代表 **miRNA 是否與某種疾病相關**。

BLS 模型會根據新的數據不斷學習，提高 **miRNA-疾病關聯預測的準確性**。

**持續學習新知識，提升預測準確率**（Continuously learning new knowledge from new samples to predict potential miRNA-disease associations with high accuracy）。

**總結**

這張圖展示了 **MISSIM 的整體架構與學習流程**：

**從多個數據庫獲取 miRNA 和疾病的相似性特徵**（功能、序列、語義、相互作用）。

**融合不同來源的特徵**，形成完整的 miRNA-疾病關聯數據。

**使用 BLS 進行學習**，透過 **增強節點與增量學習** 來不斷適應新的生物數據。

**最終輸出 miRNA-疾病預測結果**，提供潛在的疾病標誌物候選清單。

這種方法相比傳統機器學習模型，**更能適應生物數據的增長，並保持較高的準確率**。

**1. 構建訓練數據集**

**正樣本（Positive Samples）**：從 HMDD v3.0 數據集中提取 **32226 個已驗證的 miRNA-疾病關聯對**。

**負樣本（Negative Samples）**：隨機選擇 **850 種疾病** 和 **1057 個 miRNA**，確保這些 miRNA-疾病組合 **不在正樣本集中**，然後組成負樣本。

**2. 特徵表示（Feature Representation）**

每個 **疾病** 和 **miRNA** 都被轉換為 **數值向量**：

**疾病特徵向量（Disease Feature Vector, SD）**： SD(d(a))=(v1,v2,v3,…,v850)SD(d(a)) = (v\_1, v\_2, v\_3, …, v\_{850})SD(d(a))=(v1​,v2​,v3​,…,v850​)

**miRNA 特徵向量（miRNA Feature Vector, SM）**： SM(m(a))=(w1,w2,w3,…,w1057)SM(m(a)) = (w\_1, w\_2, w\_3, …, w\_{1057})SM(m(a))=(w1​,w2​,w3​,…,w1057​)

每個 **miRNA-疾病對（miRNA-Disease Pair）** 都被表徵為一個 **1907 維的向量**：

Fsim=(SD(d(a)),SM(m(a)))F\_{sim} = (SD(d(a)), SM(m(a)))Fsim​=(SD(d(a)),SM(m(a)))

**3. 特徵降維（Feature Reduction）**

為了提高計算效率並減少冗餘信息：

使用 **自動編碼器（Autoencoder）**，將 **F\_sim 從 1907 維降至 32 維**。

另一個特徵矩陣 **F\_seq 也從 64 維降至 32 維**。

最終，**每個 miRNA-疾病對被壓縮為 64 維的向量**： F=(Fsim′,Fseq′)F = (F\_{sim}', F\_{seq}')F=(Fsim′​,Fseq′​)

**4. 訓練模型並預測 miRNA-疾病關聯**

**Broad Learning System（BLS）** 被用來訓練模型。

**標籤定義（Label Definition）**：

**正樣本（已知 miRNA-疾病關聯）= 1**

**負樣本（隨機 miRNA-疾病組合）= 0**

訓練完成後，系統會為 **新的 miRNA-疾病對打分**：

**得分越高，代表該 miRNA 與該疾病的關聯可能性越大**。