**1. Introduction**

Multimorbidity is the occurrence of two or more chronic conditions in a single individual. The complexity of multimorbidity presents significant healthcare challenges. A chronic condition causes substantial changes in the lives of patients. The presence of two or more conditions often results in a compounded negative effect on the health and lifestyle of the patient.

A promising solution is in machine learning(ML). Advances in ML promise new methods of analyzing health data from different sources such as Electronic Health Records(EHR), Social Determinants of Health (SDOH), wearable data, and multi-omics datasets (e.g., genomics, transcriptomics, proteomics). Integrating these different data sources into predictive and analytical methods can yield insights.

The solution explores disease-disease networks and predicts multimorbidity risk by leveraging matrix factorization, deep learning, and topological data analysis (TDA). Insights from large-scale datasets yield interpretable disease networks.

**2. Biological and Technical Framework**

**2.1 Biological Background**

Multimorbidity isa result of complex interactions between biological, environmental, and lifestyle factors:

* Biological complexity includes genetic predispositions, transcriptomic dysregulations, and proteomic variations.
* Environmental and social determinants (e.g., socioeconomic status) affect disease onset and progression.
* The temporal progression of diseases introduces situations in analyzing evolving patterns and causal pathways (Delanerolle et al., 2021) (Hassaine et al., 2020).

**2.2 Data Sources**

This solution uses:

* EHR Data: Chronic disease diagnosis codes and treatment records.
* SDOH Data: Demographics, income, education, and lifestyle factors.
* Wearable Data: Heart rate, step count, sleep metrics etc.
* Multi-omics Data:
* Genomics: DNA variations.
* Transcriptomics: Gene expression levels.
* Proteomics: Protein abundance measurements.
* These datasets allow for multi-scale analysis, capturing both macro-level (clinical and lifestyle) and micro-level (molecular) determinants of multimorbidity(Li et al., 2022) (Hassaine et al., 2020)

**3. Algorithms and Analytical Workflow**

**3.1 Matrix Factorization**

Non-negative Matrix Factorization (NMF) is applied to co-occurrence matrices to reduce dimensionality and uncover latent disease clusters. It identifies how diseases co-occur within a population, revealing underlying multimorbidity patterns (Hassaine et al., 2020).

**3.2 Deep Learning**

Neural Networks:

* Multi-layer Perceptrons (MLP) predict multimorbidity risk based on extracted features from all datasets.
* Autoencoders reconstruct omics data, identifying biomarkers associated with specific disease clusters.
* Integration with Precision Medicine:

Enables the personalization of treatments by predicting disease trajectories for individual patients(Delanerolle et al., 2021).

**3.3 Topological Data Analysis (TDA)**

TDA extracts global features of disease networks, leveraging persistent homology to detect disease clusters and patterns across time (Hassaine et al., 2020).

**4. Disease-Disease Interaction Network**

The interactive disease graph visualizes co-occurrence patterns among diseases:

Construction:

* Nodes represent diseases; edges represent co-occurrence strength.
* Edge thickness is proportional to the strength of the association.
* Visualization:
  + The network is built using Pyvis, enabling dynamic exploration of disease connections.
  + Users can add or remove diseases and inspect co-occurrence relationships.
  + The network also incorporates matrix factorization results, emphasizing latent relationships beyond direct co-occurrence.

**5. Workflow Pipeline**

* Data Preprocessing:
* Normalize features across datasets.
* Handle missing values using imputation methods.
* Feature Engineering:
* Extract multimodal features (e.g., social, molecular, clinical).
* Perform dimensionality reduction (e.g., PCA, NMF).
* Disease Network Construction:
* Use the co-occurrence matrix from EHR data.
* Apply TDA and machine learning for advanced insights.
* Prediction Model:
* Train ML models (e.g., Random Forest, Neural Networks) for risk prediction.
* Evaluate models using metrics like accuracy, AUROC, and RMSE.
* Visualization:
* Generate disease-disease networks with Pyvis.
* Present interactive outputs in a Streamlit application.

**6. Potential Impact**

**Clinical Insights.**

* Early prediction of multimorbidity risk enables preventive interventions, improving patient outcomes.
* Identifying hidden patterns in multimorbidity supports precision medicine, tailoring treatment based on individual disease trajectories(Delanerolle et al., 2021)(Hassaine et al., 2020).

**Scalability.**

* Applicable to diverse populations and scalable for large healthcare datasets.
* Integration of multi-omics data enables applications in biomarker discovery and drug target identification (Li et al., 2022).

**Challenges**

* Data Integration: Merging heterogeneous datasets from varied sources.
* Interpretability: Ensuring that models are explainable for clinical use.
* Longitudinal Analysis: Addressing temporal disease progression in complex networks​ (Hassaine et al., 2020).

**7. Conclusion**

This project integrates cutting-edge machine learning algorithms and omics technologies to tackle multimorbidity complexity. By leveraging the strengths of matrix factorization, deep learning, and TDA, the approach facilitates:

* Comprehensive disease network visualization.
* Accurate prediction of multimorbidity risk.
* Enhanced understanding of multimorbidity patterns.
* Future extensions include integrating real-world patient feedback, testing the system on longitudinal cohorts, and enhancing computational efficiency for broader adoption in clinical settings.