أكاديمية كاوست
KAUST ACADEMY

MedTwin: Biomedical Knowledge Graphs for Early Prognosis of Diabetes Using Digital Twin Technology

Leena Alshehri, Mozn Alshehri, Khadija Drira, Saja Alhirbish

Supervised by: Sakhaa Alsaedi

أكاديمية كاوست
KAUST ACADEMY

Contents

Abstarct....3
List of Figures and Tables....3
Introduction....5
  Background....5
  Problem Statement....5
  Objectives....5
  Scope....5
  Significance....6
Literature Review....6
Methodology....6
  Approach....6
  Data Collection....6
  Tools and Technologies....6
  Procedure....7
Results....7
Discussion....7
Conclusion....7
Recommendations....8
References....8

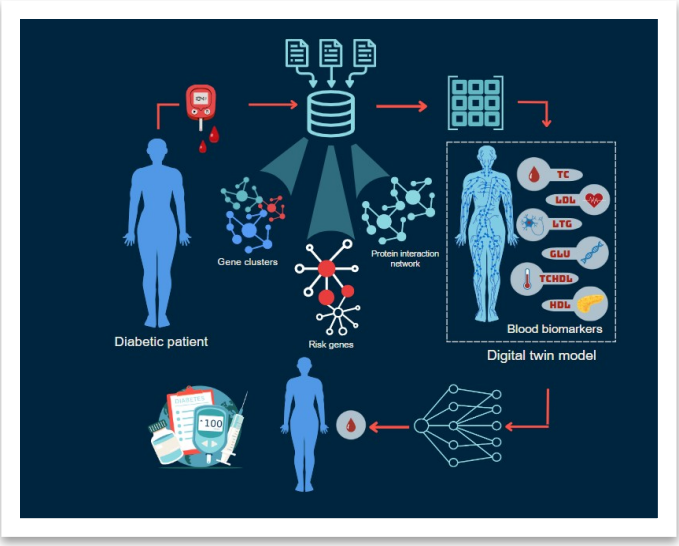
أكاديمية كاوست
KAUST ACADEMY

Abstarct

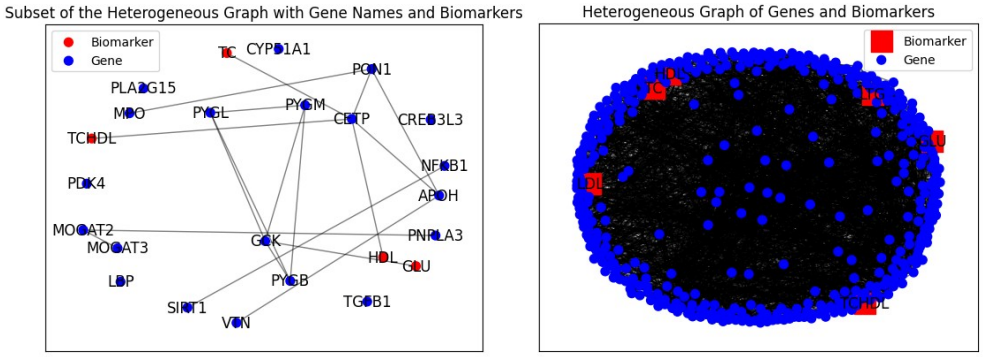
MedTwin is an AI-driven digital twin model designed to enhance diabetes disease severity prediction by integrating genomic data with biomarkers. The primary aim is to create a graph-based model that contributes to more effective personalized healthcare, improving patient outcomes and optimizing healthcare resources. The methodology involves constructing a comprehensive biomedical knowledge graph using curated data from diabetic patient blood biomarkers, gene clusters, protein interaction networks, and risk genes. The model successfully validated the digital twin concept, demonstrating potential in accurately predicting diabetes severity and enhancing patient experiences through personalized healthcare. Future work includes integrating drug data, real-time monitoring, and expanding the model to other chronic diseases.

List of Figures and Tables

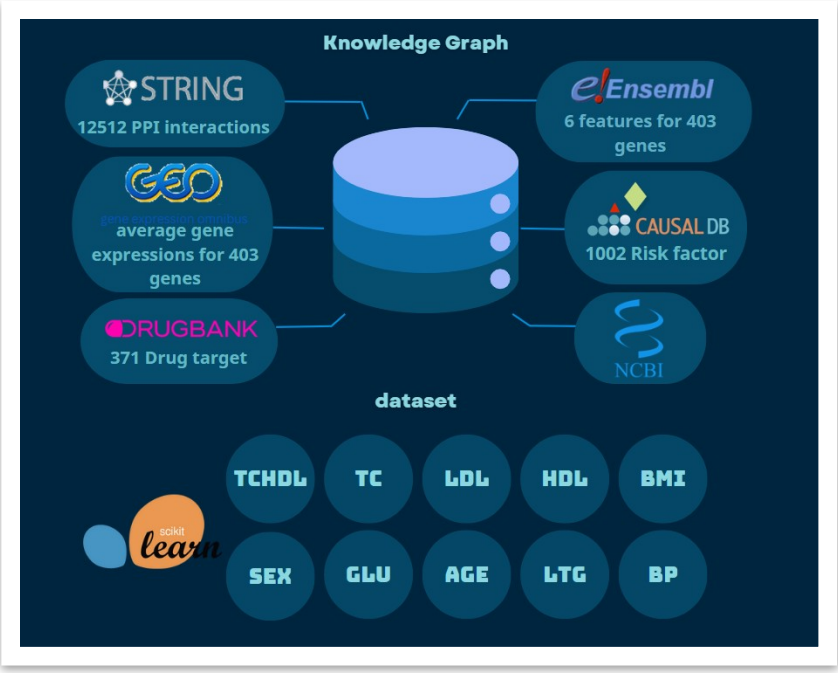
- Figure 1: Digital Twin Model Workflow



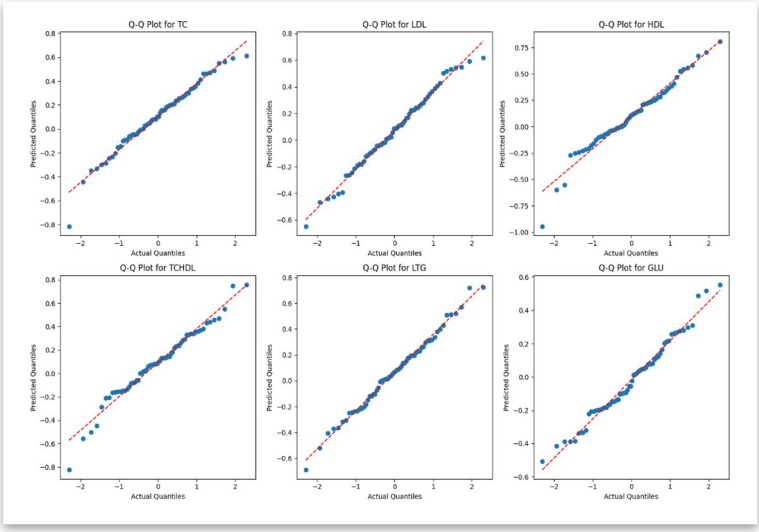
- Figure 2: Knowledge Graph Analysis



- Table 1: Summary of Curated Data



- Table 2: Model Performance Metr



Q-Q Plot for each biomarker

Introduction

Background

Diabetes is a global health crisis affecting over 537 million adults, placing immense pressure on healthcare systems. Accurate prediction of disease progression is critical for effective management and treatment of diabetes. The advent of digital twin technology offers a novel approach to personalized healthcare by simulating a patient's physiological state using AI-driven models.

Problem Statement

The current methods for predicting diabetes severity fail to integrate comprehensive genomic data, limiting their effectiveness in personalized healthcare.

Objectives

The primary objective of this project is to develop a graph-based digital twin model that enhances the prediction of diabetes disease severity by integrating gene-biomarker relationships.

Scope

The study focuses on using biomedical knowledge graphs to predict the severity of diabetes. The model is limited to the integration of genomic and biomarker data, and future work may extend to other chronic diseases.

Significance

This project has significant implications for personalized healthcare, offering a more precise method for predicting diabetes severity, thus enabling tailored

treatment plans that can improve patient outcomes and reduce healthcare costs.

Literature Review

Previous studies have explored the use of biomarkers and genomic data in predicting diabetes, but most have not fully integrated these data types into a comprehensive model.

The use of knowledge graphs in biomedical research has shown promise in identifying complex relationships between genes and biomarkers.

However, there is a gap in applying these techniques specifically to diabetes prognosis, which this project aims to address.

Methodology

Approach

The project employs a graph-based approach to integrate various data sources, constructing a biomedical knowledge graph that serves as the foundation for the digital twin model.

Data Collection

The dataset includes diabetic patient blood biomarkers, gene clusters, protein interaction networks, and risk genes.

Specific metrics include 371 drug targets, 1002 risk factors, average gene expressions and other features for 403 genes, 12,512 PPI interactions, and 6 biomarkers.

Tools and Technologies

- Programming Languages: Python

- Libraries: NetworkX, Pandas, PyTorch Geometric, Scikit-learn, Node2Vec.

- Machine Learning Algorithms: Heterogeneous Graph Neural Networks (HetGNNs), Deep Neural Network.

Procedure

1. Data Preprocessing: Cleaning the collected data for integration into the knowledge graph.
2. Knowledge Graph Construction: Building the graph by mapping gene-biomarker relationships and integrating gene interaction data.
3. Model Training: Training the model using the knowledge graph and validating its accuracy in predicting diabetes severity.
4. Evaluation: Assessing the model's performance by using MSE, and predicting each biomarkers individually by using q-q plot.

Results

The MedTwin model successfully predicted the severity of diabetes, with performance metrics indicating a high level of accuracy and reliability.

The knowledge graph revealed critical insights into gene-biomarker relationships, and the model's predictions were validated against real-world patient data.

The following are key results:

- Validation of the Digital Twin Concept: The model demonstrated its ability to simulate patient conditions.
- Enhanced Patient Experience: The model provided insights into the patient’s health.

Discussion

The results indicate that integrating genomic data with biomarkers in a knowledge graph improves the accuracy of diabetes severity predictions.

The findings align with previous research but offer a more comprehensive approach by utilizing digital twin technology.

However, the model's reliance on curated data may limit its applicability in diverse populations, and further validation is needed to ensure generalizability.

Conclusion

The MedTwin project successfully developed a graph-based digital twin model that enhances diabetes disease stage prediction.

The model's integration of gene-biomarker relationships provides a more accurate and personalized approach to healthcare,

with the potential to improve patient outcomes and optimize healthcare resources.

Recommendations

- Future Work: Integrate drug data into the model to predict medication efficacy and dosage for individual patients.

- Real-Time Monitoring: Expand the model to support real-time monitoring using wearable devices.

- Telemedicine Integration: Incorporate the model into telemedicine platforms to provide remote healthcare support.

- Extension to Other Diseases: Apply the digital twin technology to other chronic conditions beyond diabetes.

References

[1] STRING db, Available: https://string-db.org/.
[2] Ensembl db, Available: https://www.ensembl.org/index.html.
[3] Causal db, Available: https://maayanlab.cloud/Harmonizome/dataset/Causal+Network.
[4] NCBI, Available: https://www.ncbi.nlm.nih.gov/.
[5] GEO db, Available: https://www.ncbi.nlm.nih.gov/geo/.
[6] Drug Bank db, Available: https://www.drugbank.ca/.