Probabilistic Reasoning - Term Project

DiaCausal: Expert vs. Data-Driven Causal Discovery, Sensitivity Analysis, and Interventions for Diabetes Risk Factors

Abdul Haseeb (17132), Annayah Usman (18673)

June 11, 2025

Contents

1 Introduction

Diabetes Mellitus, commonly known as diabetes, is one of the fastest-growing public health concerns worldwide, affecting people of all ages, genders, and regions [10]. The 2016 study by the Noncommunicable Disease Risk Factor Collaboration found that less than 1% women-and even fewer men-have chances to prevent the rise in the incidence of diabetes by 2025 [1]. Although according to the Global Burden of Diseases 2019 report, ischemic heart disease and stroke were the first and second leading contributors to the global burden of disease in that year [12]. Diabetes is regarded a significant precursor for both these disorders, making it one of the most prevalent global causes of mortality and morbidity. Therefore, understanding the complex factors behind the onset and progression of diabetes by experts in the domain is crucial to enable timely and accurate diagnoses and guide the solution to this global challenge.

While traditional statistical and machine learning approaches have advanced in predicting disease risk, they often fall short in uncovering the underlying causal mechanisms. This limitation hinders the ability to design targeted interventions that address the root causes rather than merely correlated symptoms. The true potential for impactful healthcare advancements lies not just in prediction, but in understanding causality to guide effective action.

To facilitate this, we drew inspiration from recent research [16] to reimplement their approach and develop a decision support system grounded in causal reasoning. The aim is to offer an interactive platform for doctors, researchers, and medical students to explore and understand the complex relationships between various risk factors and diabetes using the Behavioral Risk Factor Surveillance System (BRFSS) 2015 dataset [5].

Through this system, users can define and evaluate causal graphical models informed by domain expertise, compare expert-defined structures with those learned from data using different algorithms, and simulate the effects of potential interventions on key risk factors. Additionally, the platform allows users to analyze the sensitivity of diabetes risk to different variables within a causal framework, ultimately fostering a dialogue between expert knowledge and data-driven insights to refine causal understanding and support more informed decision-making.

For a comprehensive overview of our Streamlit application, **DiaCausal: A Diabetes Decision Support**, please refer to the Appendix: Streamlit Application Interface. It provides additional details on the interface design, layout components, and interaction flow. The application can also be accessed via this link and the Github repository can be found here. Currently, caching button and model fitness page crashes the app due to limited hosting resources on Streamlit Cloud. So, it is recommended to run the app locally following the GitHub repo.

2 Dataset Overview

2.1 Data Source

The data were acquired from Centers for Disease Control and Prevention's (CDC), BRFSS, a leading health-related telephone survey system that gathers information on health behaviors, chronic conditions, and use of preventive services among US residents [5]. From this extensive repository, to maintain consistency with the study that inspired our work, we used the BRFSS-2015 dataset, which comprises 253,680 health records across 22 specifically selected variables.

2.2 Data Pre-processing

The selected variables included for this study can be found in Table 1 and are grouped as follows:

- **Non-modifiable Risk Factors (RFs):** Sex, Age
- **Modifiable RFs:** Body mass index (BMI), Blood Pressure (HighBP), Cholesterol (HighChol), General Health (GenHealth), Education, Heart Disease (HeartDiseaseorAttack), Physical Activity (PhysActivity)
- **Medical Condition:** Diabetes (Diabetes binary).

The pre-processing of the data involved encoding non-categorical variables and ensuring that existing categorical variables are restricted to a reasonable number of states without loss of information. The following variables were subjected to these steps, and their corresponding encodings and labels are also provided in Table 1:

- **Age grouping:** Age, in years, was categorised into six groups
- **General health:** With initially several states, was simplified into three categories
- **Education:** Education levels were condensed into three states: individuals who never attended school or attended from grade 1 through grade 8 (1), those who completed grade 9 through GED (2), and individuals with college education (3)
- **Income:** Income levels, originally in thousands of dollars, were condensed into four states
- **BMI categorisation:** Initially continuous, was categorised into three states
- **Physical and mental health duration:** Originally had values in days, ranging from 0 to 30, were transformed into three categories (2).

Variable	States	Marginal
Diabetes_binary	{0: No, 1: Yes}	{86%, 14%}
HighBP	{0: No, 1: Yes}	{57%, 43%}
HighChol	{0: No, 1: Yes}	{58%, 42%}
BMI	{0: 0-24, 1: 25-39, 2: $\geq$40}	{28%, 66%, 6%}
HeartDiseaseOrAttack	{0: No, 1: Yes}	{91%, 9%}
CholCheck	{0: No, 1: Yes}	{4%, 96%}
Stroke	{0: No, 1: Yes}	{96%, 4%}
Smoker	{0: No, 1: Yes}	{56%, 44%}
Fruits	{0: No, 1: Yes}	{37%, 63%}
Veggies	{0: No, 1: Yes}	{19%, 81%}
HvyAlcoholConsump	{0: No, 1: Yes}	{94%, 6%}
AnyHealthcare	{0: No, 1: Yes}	{5%, 95%}
NoDocbcCost	{0: No, 1: Yes}	{92%, 8%}
MentHlth	{0: 0-9, 1: 10-19, 2: $\geq$20}	{87%, 5%, 7%}
PhysHlth	{0: 0-9, 1: 10-19, 2: $\geq$20}	{84%, 5%, 10%}
DiffWalk	{0: No, 1: Yes}	{83%, 16%}
Sex	{0: Female, 1: Male}	{56%, 44%}
Age	{1: 18-29, 2: 30-44, 3: 45-54, 4: 55-64, 5: 65-74, 6: $\geq$75}	{5%, 10%, 14%, 23%, 26%, 22%}
Income	{1: <15, 2: 15-25, 3:25-50, 4: $\geq$50}	{8%, 14%, 25%, 53%}
Education	{1: 1-8, 2: 9-GED, 3: College}	{2%, 29%, 70%}
GenHlth	{1: Excellent, 2: Good, 3: Poor}	{53%, 42%, 5%}
PhysActivity	{0: No, 1: Yes}	{25%, 75%}

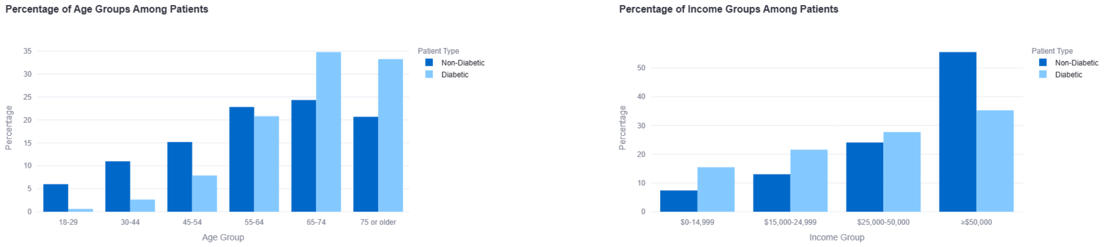
Table 1: Overview of all variables used in the analysis. Each variable is presented with its coded states and corresponding descriptive labels, along with their marginal percentages in the dataset.

3 Exploratory Data Analysis

An initial dive into the pre-processed data uncovers relationships between key variables and the target variable, diabetes. A few observations are as follows:

- The **distribution of age** in Figure 1 showed a higher prevalence of diabetes in individuals aged 55+
- Regarding socioeconomic factors, the **income levels** displayed a difference between diabetic and non-diabetic groups, with a higher proportion of non-diabetics falling into higher income categories as seen in Figure 1

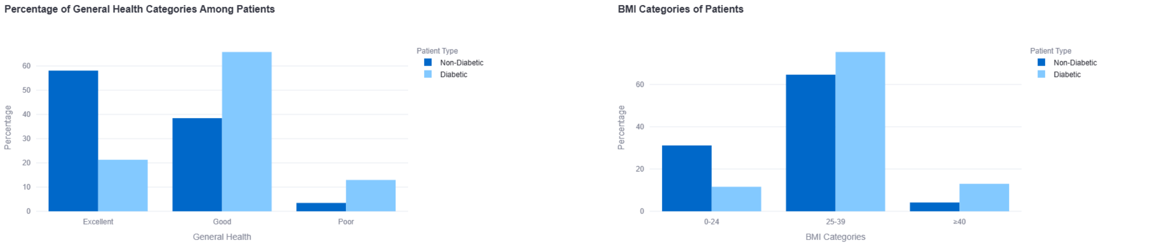
- In Figure 2, the **distribution of individuals across general health categories** shifted, with a decrease in the "Excellent" category and an increase in the "Poor" category amongst those with diabetes
- The **distribution of individuals across BMI categories** shifted towards higher levels in the diabetic group, highlighting the importance of monitoring individuals in the 25-39 BMI range as observed in Figure 2
- As found in Figure 3, the percentage of individuals with **high cholesterol** significantly increased in the diabetic group compared to non-diabetics
- The prevalence of **high blood pressure** was substantially higher in the diabetic group compared to the non-diabetic group as seen in Figure 3.



(a) Distribution of Age of diabetic and non-diabetic individuals.

(b) Distribution of Income of diabetic and non-diabetic individuals.

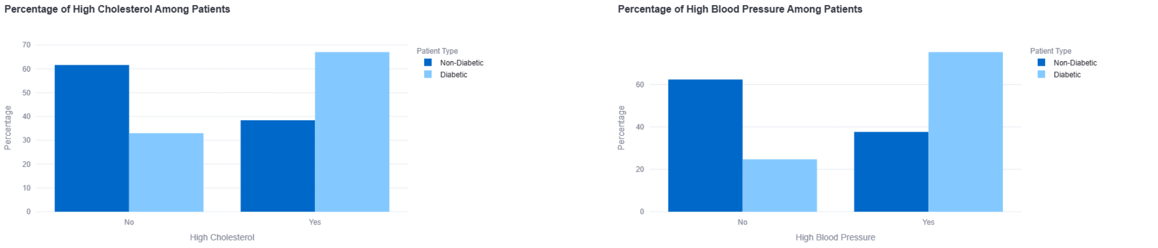
Figure 1: Age and Income distributions among diabetic and non-diabetic individuals.



(a) General Health distribution

(b) BMI distribution

Figure 2: Distributions of General Health and BMI for diabetic and non-diabetic individuals.



(a) High Cholesterol distribution

(b) High Blood Pressure distribution

Figure 3: Distributions of High Cholesterol and High Blood Pressure among diabetic and non-diabetic individuals.

4 Methodology

This section presents the main blueprint of the methodological framework adopted in this project, including an overview of the interactive Streamlit application. We begin by introducing the three-tiered domain expert knowledge graphs, which serve as ground truths. This is followed by a detailed discussion of the structure learning algorithms, parameter learning configuration, and the model-averaging strategy, along with their integration into the application. The section further introduces the interventional analysis employed for causal interpretation and concludes by defining the metrics that will be used for the evaluation of results.

4.1 Knowledge Graphs

The original paper began the methodology by jointly constructing a causal graph in collaboration with a medical expert. Together, they drew cause-and-effect maps based on existing medical knowledge and classified possible causal relationships at three levels of confidence: high, moderate, and low.

- The **high confidence graph** in Figure 4 provides a structured view of the most trusted connections within the knowledge domain
- The **moderate confidence graph** in Figure 5 includes both high and moderate confidence level relationships, offering insights in connections that are at least moderate if not high confidence
- The **low confidence graph** in Figure 6 provides a comprehensive overview of potential connections, even when confidence is low.

The high-, moderate-, and low-confidence graphs already provided as CSVs are collected from actual domain experts by the authors of the reference paper and so we have used them as well. The graphs were preloaded and integrated directly into the Streamlit application interface to view interactively.

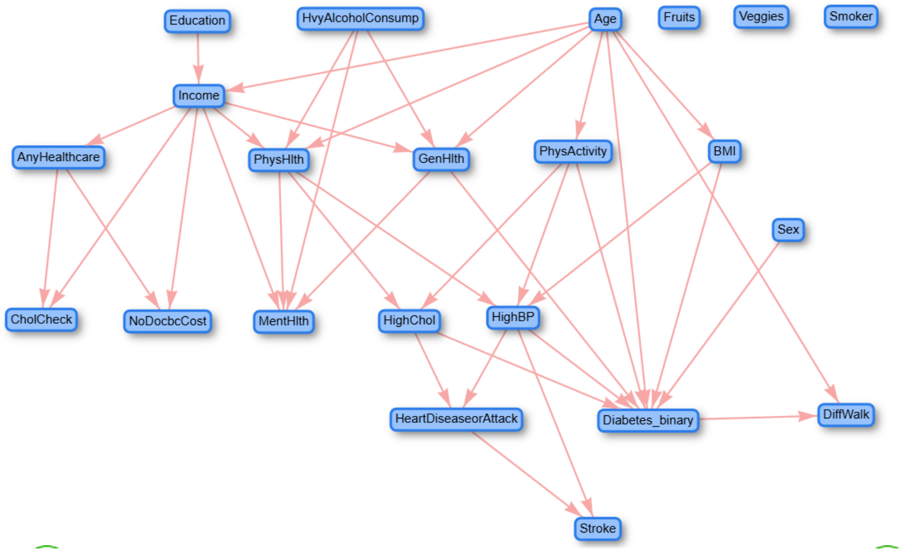


Figure 4: High confidence causal graph. It contains the 22 variables and 37 directed edges as identified in the original paper.

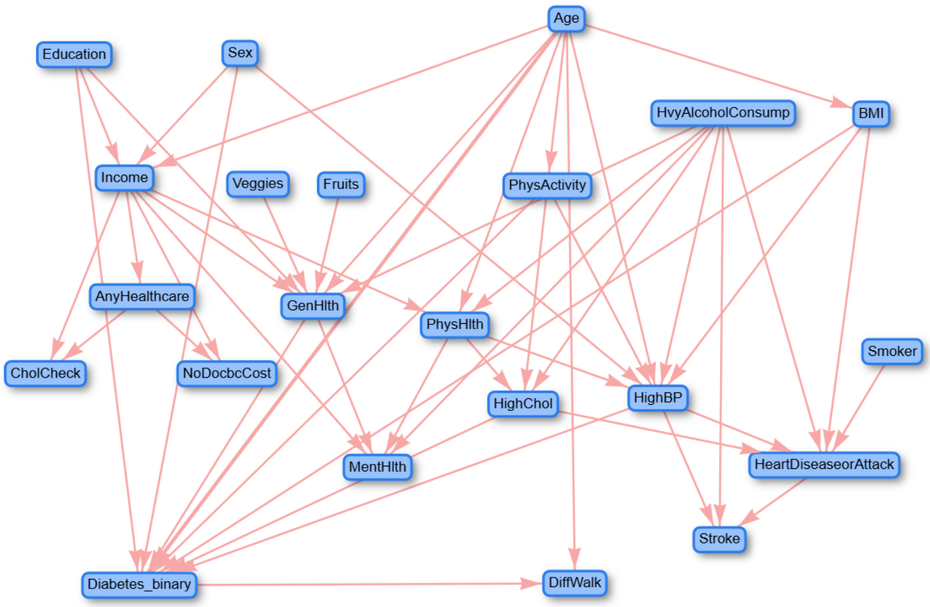


Figure 5: Moderate confidence causal graph. It contains the 22 variables and 50 directed edges as identified in the original paper.

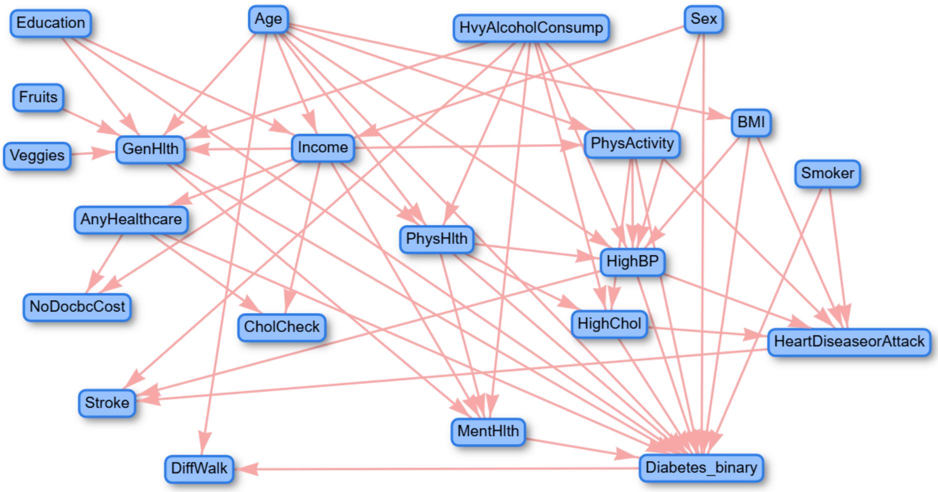


Figure 6: Low confidence causal graph. It contains the 22 variables and 55 directed edges as identified in the original paper.

Our application offers several ways to generate a clinical knowledge graph. One option allows users to upload an expert-defined network in CSV format. Upon upload, the corresponding confidence-level graph—along with its nodes and edges—is rendered in the interface. By default, the Graph Editor remains in a view-only mode in this case, and any newly uploaded network will overwrite the existing one. Alternatively, users can choose the "Define a New Network" option, which anables the Graph Editor's Manual Controls. These controls allow users to construct or modify the graph by selecting source and target nodes, specifying the confidence level (high, moderate, or low), and choosing whether to add or remove an edge.

4.2 Structure and Parameter Learning

In our project, we focus majorly on score-based structure learning algorithms, as they allow for flexible exploration of the structure space without requiring any tests for assumptions. By limiting our approach to mainly this class of algorithms, we maintain methodological consistency and enhance the interpretability and comparability of the resulting network structures. Ultimately, this gives way to a more controlled and robust examination of the underlying relationships within our data. In addition to these, we explored one constraint-based algorithm to compare its output against score-based methods and assess the consistency of discovered causal structures under differing learning paradigms.

The algorithms implemented include: Peter-Clark algorithm, Greedy Equivalence Search, Simulated Annealing, Hill Climb Search, Evolutionary Algorithm, Multi-Agent Genetic Algorithm and Particle Swarm Optimization and their details are as follows:

- **Peter-Clark algorithm (PC):** PC [14], the only constraint-based algorithm explored in this study, starts by forming a fully connected undirected graph and eliminates edges using conditional independence tests. It identifies the causal relationships between variables based on the absence of subsets of variables that render them independent when conditioned upon
- **Greedy Equivalence Search (GES):** GES [6] operates in two phases: it first adds edges to an initially empty equivalence class to incorporate dependencies, then removes edges to optimize the structure. Both phases rely on local changes, but the algorithm's scalability is limited by the exponential cost of evaluating head-to-head node subsets—particularly in networks with large cliques or high-degree nodes. The initial phase is usually the most computationally demanding
- **Hill Climb Search (HCS):** HC algorithm [11] typically starts with an empty graph and greedily searches the space of graphs by adding, removing and reorienting edges iteratively. It uses a scoring function to score each graph visited, and, at each iteration, moves to the graph that maximizes the scoring function. It terminates when no neighboring graph further increases the score
- **Simulated Annealing (SA):** SA algorithm [9] refines an initial DAG by making local changes—adding, deleting, or reversing edges—and evaluating each candidate using a scoring function. Unlike greedy methods, SA occasionally accepts worse-scoring DAGs at higher "temperatures," allowing it to escape local optima in the large search space. As the temperature decreases, the algorithm increasingly favors better-scoring structures, guiding the search toward a globally optimal DAG until a stopping criteria, such as minimum temperature is met
- **Evolutionary Algorithm (EA):** EA [15] begins with a random population of DAGs evaluated using a fitness function. Through crossover between parent graphs to combine structural features, then iterative mutation (adding, removing, or reversing edges) and selection, it evolves the population while ensuring acyclicity, aiming to converge on the most optimal DAG over multiple generations
- **Multi-Agent Genetic Algorithm (MAGA):** MAGA [4] unlike standard genetic algorithms, employs a lattice-based multi-agent system where agents interact with neighbors through crossover and mutation, maintaining diversity and avoiding premature convergence. A self-learning operator further refines the best solution in each generation. This design enables efficient exploration of the DAG space to identify a structure that best fits the data
- **Particle Swarm Optimization (PSO):** In PSO [8], each particle represents a candidate graph. To accommodate the discrete nature of the problem, customized rules for velocity and position updates are introduced, often incorporating stochastic mutation and crossover operations. A score-based fitness function evaluates each structure, guiding the swarm toward high-scoring network configurations through iterative updates based on individual and global best solutions.

The application offers a flexible interaction of both structure and parameter learning for resulting graphs. In the "Evidence Based Learning" page, users can start by selecting the learning parameters that guide the structure learning process across all aforementioned algorithms, after the "Learn Structures" button is invoked. These options for these configurations include:

- **Initializing DAG Options:** Random, Sparse Random, Empty, Tree (TAN), Tree (Chow-liu)
- **Scoring Functions:** BIC, AIC, K2, BDeu, BDs.

For parameter learning, that is to generate the Bayesian Networks from the structures learnt above, the "Learn Parameter" entails two estimator options:

- **Maximum Likelihood Estimation (MLE)** which is default
- **Bayes with equivalent sample size** parameter to initialize uniform priors.

This learns CPTs for both structures learned from above algorithms and expert-defined knowledge graphs.

4.3 Model-Averaging Approach

We also incorporate a model-averaging strategy [7] to synthesize outputs from multiple algorithms. In this context, model-averaging provides a more holistic representation of the potential causal structure inferred from the data, rather than depending on a single model selected by one method. This aims to counteract biases and improve the stability and trustworthiness of the derived causal insights. It works by the following steps:

Step 1: Add High-Frequency Directed Edges

- Iterate through directed edges in order of decreasing frequency.
- For each edge $X \to Y$:
  - Skip if the reverse edge $Y \to X$ is already in the graph.
  - If adding $X \to Y$ creates a cycle:
    - Reverse the edge to $Y \to X$ and store it in set **C**.
  - Otherwise, add $X \to Y$ to the graph.

Step 2: Add High-Frequency Undirected Edges

- Iterate through undirected edges in order of decreasing frequency.
- For each edge $X - Y$:
  - Skip if either $X \to Y$ or $Y \to X$ is already present in the graph.
  - Otherwise, add $X - Y$ to the graph.

Step 3: Add Reversed Edges from Set C

- Iterate through edges in set **C** in order of decreasing frequency.
- For each edge $Y \to X$:
  - Add it only if it does not form a cycle.
  - Skip if it introduces a cycle.

In the app, hyperparameter of minimum edge frequency is incorporated when selecting the configurations for structure learning. It can be set to any integer value from 1 to 7, inclusive. It uses all of the learned structures from the above algorithms to construct an averaged structure from them. For example, if minimum edge frequency is 4 then for an edge to be considered, it must appear in at least 4 out of 7 graphs.

4.4 Interventional Analysis

To understand how changing one variable directly affects others in a system, researchers use a method, termed interventional analysis, that simulates hypothetical interventions. The aim is to understand causal effect rather than correlated impact through some confounder as is in evidence-based belief propagation. Conceptually this is done by do-calculus [13], which mathematically models such interventions by disconnecting a variable from its usual causes and forcing it into a new state.

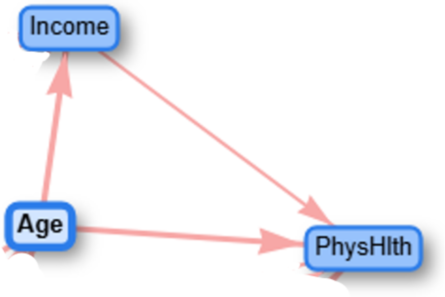


Figure 7: Sample 3 node network for explanation of interventional analysis.

For example given the following network in Figure 7. To understand the causal impact of income on physical health, we remove the edge from age to income, isolating income from confounding influence from its parent. Then by setting evidence on income and observing changes in the marginals of physical health, we study the causal impact only rather than inclusion of correlated impact through age node.

In the original paper, this is implemented using the GeNIe software [3] which employs do-calculus while our approach utilizes the Python package, pgmpy [2]—both tools built upon the same theoretical foundations. Furthermore, our app, on the "Interventional Analysis" pages, provides an option to view the cumulative impact of intervention on a variable to another variable. The cumulative effect is derived by intervening on each state of the target node, observing the resulting changes in marginal probabilities (as percentage changes), and then summing these effects across all states of the intervened node. In addition, users can simulate manual interventions on individual nodes at specific states to observe localized effects. Cumulative effect results display in descending order.

4.5 Evaluation Metrics

We employ three metrics to assess how closely the DAG structures generated by the algorithms, the model-averaging approach, and the domain expert knowledge graphs align with each other. The metrics include: Structural Hamming Distance, F1 Score and the Balanced Scoring Function with their details as follows:

- **Structural Hamming Distance (SHD):** SHD assesses how different two graphs are. It quantifies the number of structural changes—such as adding, removing, or reversing edges—required to convert a learned graph into the reference or ground truth graph. A lower SHD score (better) indicates a closer match between the two graph structures, whereas a higher score reflects greater structural differences
- **F1 Score:** The F1 balances both precision and recall to evaluate the accuracy of predicted edges in a graphical structure. It is calculated as:
$$F1 = \frac{2 \cdot R \cdot P}{R + P}$$
where R (recall) is the proportion of correctly predicted edges out of all true/actual edges, and P (precision) is the proportion of correctly predicted edges out of all predicted edges. A higher F1 Score indicates better predictive performance.
- **Balanced Scoring Function (BSF):** BSF metric takes into account the difficulty of discovering both the presence and absence of edges, to generate a score that is balanced, relative to the difficulty of discovering the presence of an edge. The BSF score is calculated as:
$$\text{BSF} = 0.5\left(\frac{TP}{a} + \frac{TN}{i} - \frac{FP}{i} - \frac{FN}{a}\right)$$
where TP and TN represent true positives (correctly identified edges) and true negatives (correctly identified absent edges), FP and FN are false positives (incorrectly identified edges that are not present in the ground truth) and false negatives (edges that are present in the ground truth but not predicted by the model), a is the number of edges in the assumed ground truth, and i is the number of independencies in the assumed ground truth, calculated using the formula:
$$i = \frac{|V|(|V|-1)}{2}$$
where $|V|$ is the total number of variables in the graph. A higher BSF score indicates better overall performance.

5 Results and Discussion

5.1 Inference-based Evaluation: Model Fitness

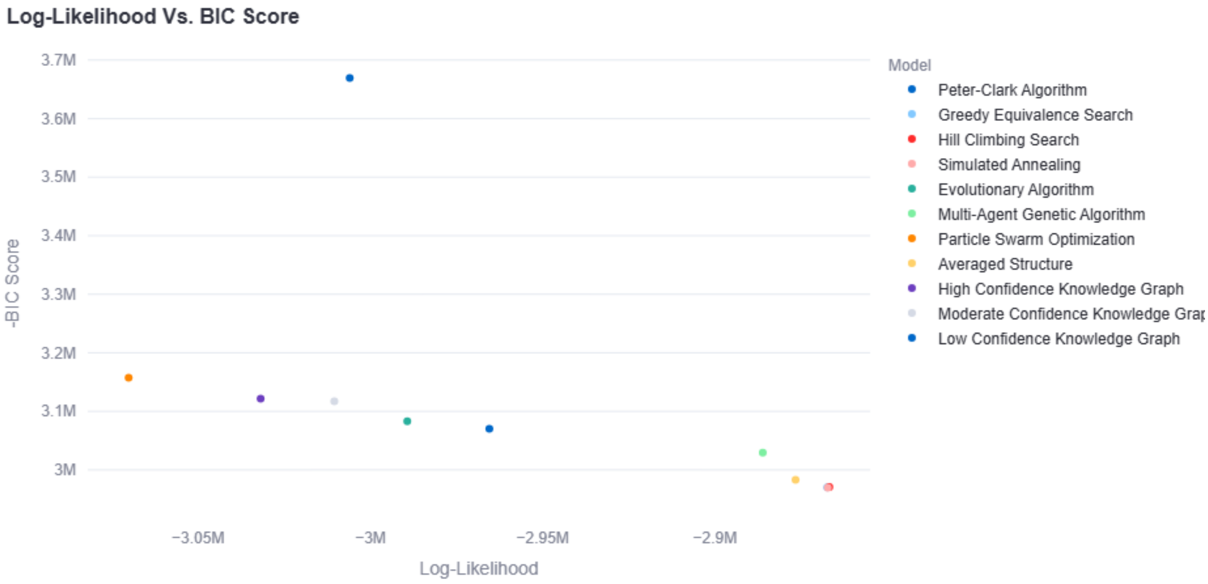


Figure 8: Comparison of BIC and Log-Likelihood Scores Across Causal Structure Learning Algorithms and Knowledge Graphs through a scatterplot. Y-axis values need to read as negative to correctly interpret the BIC.

The performance of the learned networks is assessed using two primary metrics: the Bayesian Information Criterion (BIC) and Log-Likelihood (LL). BIC aids in model selection by balancing model accuracy with complexity; it imposes a penalty on models with more parameters to reduce the risk of overfitting. In contrast, LL evaluates how well the model explains the observed data by measuring the likelihood of the data under the model's structure. A higher LL value indicates that the model provides a better fit to the data, and, in this implementation, a higher BIC is preferred but it should be noted that in Figure 8 BIC scale is negated. Together, these metrics offer a comprehensive evaluation of model quality, considering both goodness of fit and parsimony.

Figure 8 illustrates the performance of various learning algorithms based on their LL and BIC scores. The results indicate that the HCS, SA, and GES algorithms perform best, achieving the highest BIC scores and highest Log-Likelihood (LL) values, suggesting an optimal balance between model complexity and data fit. The Averaged Structure closely follows this top-performing group, which is consistent with findings from the original paper. Its strong performance may be attributed to the ensemble effect, where the aggregation of multiple structures captures stable and high-confidence edges. MAGA also performs well, trailing just behind the Averaged Structure.

In contrast, the low-confidence graph stands out as an outlier, displaying a relatively low LL but an extremely negative BIC, indicating poor model efficiency and complexity. PSO yields the lowest LL, suggesting a poor fit, although its BIC is moderately acceptable. Lastly, both the moderate- and high-confidence graphs show progressively decreasing LL values, respectively, but maintain moderate BIC scores, indicating a trade-off between likelihood and simplicity. Seeing low to moderate BIC scores with the knowledge graphs is not a major disappointment, as stipulated in the original paper, since their primary objective is to represent prior knowledge rather than to optimize model selection criteria.

5.2 Graphical Structure Evaluation

Learned Network	High Confidence	Moderate Confidence	Low Confidence
PC	45, 0.129, 0.084	58, 0.107, 0.055	63, 0.1, 0.043
GES	91, 0.228, 0.102	97, 0.235, 0.096	102, 0.227, 0.082
HCS	91, 0.24, 0.112	97, 0.246, 0.105	102, 0.238, 0.092
SA	91, 0.198, 0.077	97, 0.209, 0.071	102, 0.201, 0.058
EA	62, 0.132, 0.053	72, 0.135, 0.05	76, 0.128, 0.038
MAGA	93, 0.195, 0.073	99, 0.206, 0.067	104, 0.199, 0.054
PSO	59, 0.118, 0.052	68, 0.148, 0.091	72, 0.14, 0.08
Averaged Structure	84, 0.2, 0.084	90, 0.211, 0.083	95, 0.203, 0.07

Table 2: Evaluation of learned network structures against varying confidence levels in expert graphs. Each cell reports the Structural Hamming Distance (SHD), F1 score, and Balanced Scoring Function (BSF) score, respectively.

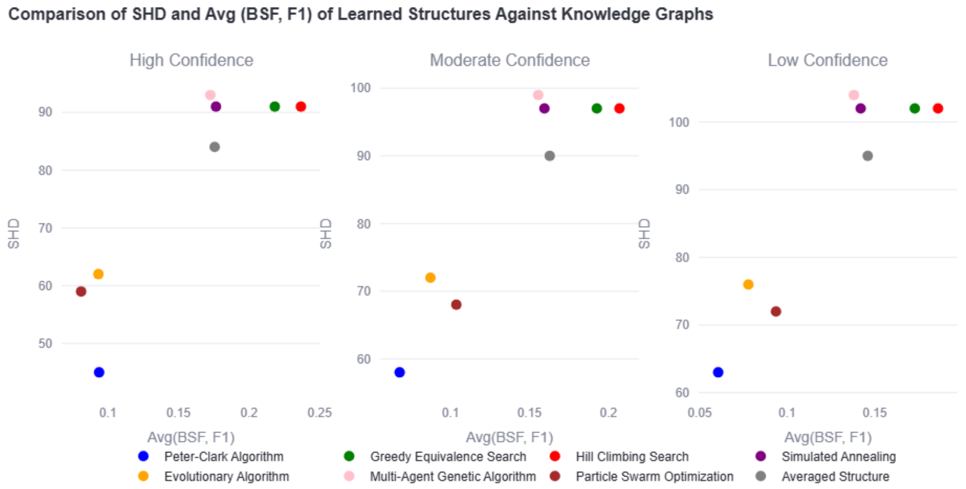


Figure 9: Scatter Plot of SHD vs. Average(F1, BSF) Across Learning Algorithms, AVeraged Structure by the Knowledge Graphs

Table 2 presents the triplet tuples of evaluation metrics: SHD, F1 score, and BSF for each model evaluated against each of the expert knowledge graphs. Recall that higher F1 and BSF values and lower SHD indicating better performance. The accompanying Figure 9 visualizes these results. It must be noted that taking the average of F1 and BSF helps smooth out metric-specific fluctuations and offers a more balanced view of both precision-recall tradeoffs and score stability—beyond just adhering to the original paper, it serves to combine complementary strengths of these metrics for more holistic assessment. The visual shows a consistent pattern across all knowledge graphs:

- PC, EA, and PSO cluster in the bottom-left region of the scatter plot, reflecting lower SHD but weaker performance on the Avg F1-BSF metric
- In contrast, GES, HCS, SA, MAGA, and the Averaged Structure form a cluster in the top-right quadrant, indicating stronger Avg F1-BSF values at the cost of a higher SHD
- This suggests that algorithms like PC, EA, and PSO may generate sparser graphs, minimizing structural errors (low SHD), but at the expense of missing meaningful edges, thus lowering F1 and BSF. This is also indicated in the original paper with the mention that SHD is biased towards sparser graphs.

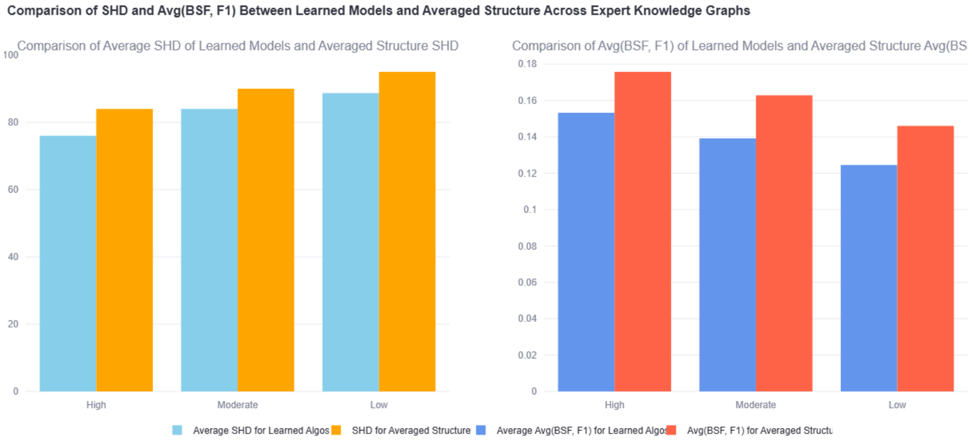


Figure 10: Comparative Barchart of Averaged SHD and Average(F1, BSF) Between Learned Algorithms and the Averaged Structure

Figure 10 highlights that the averaged SHD and the average of the combined F1-BSF scores across all learned algorithms are consistently lower than those of the Averaged Structure alone. While this aligns with findings from the original research, it also points to an interesting trade-off. The Averaged Structure, by combining multiple models, tends to include a broader set of edges, capturing more true positives and a higher F1 and BSF. However, this also leads to a denser graph, which may diverge more from the true structure, thereby increasing SHD. On the other hand, individual learned algorithms produce sparser graphs resulting in lower SHD but also potentially missing some true edges, which brings down their F1 and BSF scores.

5.3 Intervention Analysis

Network	**Intervention Variables Impact on Diabetes-Most to Least Impactful**
PC	GenHlth (334%), PhysHlth (164%), MentHlth (65%), Rest (0%)
GES	BMI (249%), HighBP (112%), HighChol (105%), Veggies (29%), AnyHealthcare (13%), Rest (0%)
HCS	BMI (249%), HighBP (112%), HighChol (105%), Veggies (29%), AnyHealthcare (13%), Rest (0%)
SA	BMI (246%), Veggies (42%), Rest (0%)
EA	HighBP (153%), Stroke (48%), HighChol (44%), HeartDiseaseorAttack (12%), CholCheck (8%), DiffWalk (7%), BMI (6%), PhysHlth (6%), AnyHealthcare (1%), NoDocbcCost (0.05%), Sex (0.001%), Rest (0%)
MAGA	BMI (231%), HighBP (113%), HighChol (107%), Veggies (30%), AnyHealthcare (13%), Rest (0%)
PSO	HeartDiseaseorAttack (176%), HighBP (17%), HighChol (13%), Rest (0%)
Averaged Structure	BMI (247%), HighBP (134%), HighChol (37%), Veggies (32%), Rest (0%)
High Confidence	Age (274%), GenHlth (234%), BMI (198%), Income (95%), HighBP (75%), HighChol (53%), Education (42%), PhysHlth (38%), PhysActivity (28%), Sex (17%), HvyAlcoholConsump (6%), Rest (0%)
Moderate Confidence	Age (319%), GenHlth (224%), BMI (201%), Education (139%), Income (79%), HighBP (74%), HighChol (52%), PhysHlth (32%), PhysActivity (23%), Sex (17%), Veggies (7%), Fruits (6%), HvyAlcoholConsump (5%), Rest (0%)
Low Confidence	GenHlth (264%), Age (250%), Education (210%), PhysHlth (162%), BMI (161%), Income (127%), MentHlth (72%), HighBP (63%), AnyHealthcare (57%), HighChol (47%), PhysActivity (29%), Sex (16%), Veggies (6%), Fruits (5%), HvyAlcoholConsump (4%), Smoker (3%), Rest (0%)

Table 3: Ranked impact of intervention variables on the effect variable, Diabetes across learned and expert graph structures. The impact is cumulative and is reported in percentage in descending order.

We follow the original paper and stick to the intervention variables as provided: High Blood Pressure (HighBP), High Cholesterol (HighChol), Heart Disease or Heart Attack (HeartDiseaseorAttack), Body Mass Index (BMI), Education level (Education), and General Health (GenHlth). To recall, for each we assess the effect of the intervention variables by their cumulative impact on the effect variable. The cumulative impact is calculated as the sum of the net absolute percentage change in the marginal probabilities of the effect variable for all states of each intervention variable. The higher the cumulative impact, the more the intervention variable affects the effect variable.

Table 3 shows the intervention variables and their impacts on Diabetes_binary effect variable for each graph structure, both learned, averaged and domain-expert provided. Overall, we find that BMI consistently has a high impact on Diabetes across all graphs, except for EA and PSO. In the learned algorithms, the interventions of Education and Heart Disease are also negligible or have no impact. It can also be observed, that though we list all the interventions in Table 3, the main intervention variables can be found across majority graph structures.

Effect Variable \Intervention	High BP	High Chol	BMI	Heart Disease	Education	Gen Hlth
HighBP	-	113%	167%	0	0	0
HighChol	0	-	81%	0	0	0
BMI	0	0	-	0	0	0
HeartDiseaseorAttack	114%	111%	114%	-	5%	267%
Education	98%	47%	114%	0	-	511%
GenHlth	170%	117%	320%	0	0	-

Table 4: Intervention for specific variables, using Averaged Structure. Reported percentages.

Effect Variable \Intervention	High BP	High Chol	BMI	Heart Disease	Education	Gen Hlth
HighBP	-	0	148%	0	14%	0
HighChol	0	-	0	0	11%	0
BMI	0	0	-	0	0	0
HeartDiseaseorAttack	121%	91%	44%	-	7%	0
Education	0	0	0	0	-	0
GenHlth	0	0	0	0	300%	-

Table 5: Intervention for specific variables using High Confidence. Reported percentages.

In addition to examining the effects of interventions on diabetes, we extend our analysis to include the other variables introduced at the start of this section, focusing solely on the Averaged Structure and High Confidence graphs. Consistent with the original paper, we use the same set of intervention variables throughout. The purpose is to compare the aggregated output of the learned models with expert knowledge to evaluate how closely the inferred causal structure aligns with the expert-defined understanding. The findings from the Average Structure and the High Confidence in Table 4 and Table 5, respectively are summarized below:

- **Intervening on HighBP:** HighBP has a strong influence on HeartDisease which is common in both the Average Structure and High Confidence graph. However, unlike the High Confidence, it also impacts Education and General Health very strongly. The link between HighBP and HeartDisease is also stipulated in the original paper

- **Intervening on HighChol:** In the Averaged Structure, HighChol has pronounced effect on HighBP, Heart Disease and General Health with moderate impact on Education. However, in High Confidence it only effects Heart Disease

- **Intervening on BMI:** For the Averaged Structure, BMI is the only variable that has substantial influence on all the effect variables. While in High Confidence, it exhibits a pronounced effect for HighBP and moderate impact for Heart Disease

- **Intervening on HeartDisease:** Consistent in both structures, and also in line with results from the paper, Heart Disease has negligible impact on all variables upon intervening

- **Intervening on Education:** Though Education only shows impact in the High Confidence structure, its most profound effect is on General Health
- **Intervening on GenHlth:** Showing no effects in the High Confidence graph on these variables, it just exhibits a high impact on Education and Heart Disease in the Averaged Structure.

Overall, BMI serves as the intervening variable with the most pronounced impact on the effect variables in question across the Averaged and High Confidence structures.

5.4 Sensitivity Analysis

Sensitivity analysis involves evaluating how changes in input variables affect the output of a model. In other words, it analyzes how sensitive the probability of Diabetes is to perturbations in its ancestors' CPDs (Conditional Probability Distributions) within a selected Bayesian Network. In our case, we focus on the parent and ancestor nodes of the diabetes node within each structure. We perturb the marginal distributions of these influencing nodes using a defined perturbation ratio (set as 0.5), and then observe the resulting changes in the marginal distribution of the diabetes node. The overall impact is quantified by calculating the cumulative sum of absolute percentage changes across both states of the diabetes node.

Network	**Sensitivity Impact on Diabetes Given Perturbed Ancestors**
PC	GenHlth (3.257%), MentHlth (0.3689%), PhysHlth (0.3143%)
GES	BMI (2.2943%), AnyHealthcare (0.0744%), HighBP (0.0223%), Veggies (0.0206%), HighChol (0.0124%)
HCS	BMI (2.2943%), AnyHealthcare (0.0744%), HighBP (0.0223%), Veggies (0.0206%), HighChol (0.0124%)
SA	BMI (1.5499%), Veggies (0.9446%)
EA	HighBP (2.1734%), Stroke (0.4632%), Sex (0.1979%), BMI (0.1796%), CholCheck (0.0533%), PhysHlth (0.0404%), HeartDiseaseorAttack (0.0222%), DiffWalk (0.0208%), HighChol (0.0058%), AnyHealthcare (0.005%), NoDocbcCost (0.0006%)
MAGA	HighChol (4.5206%), BMI (2.0218%), AnyHealthcare (0.0751%), HighBP (0.0192%), Veggies (0.0181%)
PSO	HighBP (1.1816%), HighChol (0.8983%), HeartDiseaseorAttack (0.0091%)
Averaged Structure	BMI (2.8594%), HighChol (1.7103%), Veggies (0.7247%), HighBP (0.6654%)
High Confidence	Sex (1.0215%), HighBP (0.2565%), BMI (0.1555%), Age (0.1366%), HighChol (0.1075%), HvyAlcoholConsump (0.0735%), Income (0.0385%), Education (0.0318%), PhysActivity (0.0265%), GenHlth (0.0215%), PhysHlth (0.0087%)
Moderate Confidence	Sex (1.0858%), Fruits (0.2367%), BMI (0.174%), Education (0.1561%), Veggies (0.1484%), HighChol (0.1203%), Age (0.0963%), HvyAlcoholConsump (0.0638%), PhysActivity (0.041%), PhysHlth (0.0095%), HighBP (0.0004%), GenHlth (<0.0001%), Income (<0.0001%)
Low Confidence	Sex (1.1358%), Education (0.2887%), Fruits (0.2242%), Smoker (0.2116%), HighChol (0.1795%), BMI (0.1557%), Veggies (0.1495%), AnyHealthcare (0.0608%), HvyAlcoholConsump (0.0519%), Age (0.043%), MentHlth (0.0215%), PhysHlth (0.0095%), PhysActivity (0.0027%), Income (0.0017%), HighBP (0.001%), GenHlth (0.0003%)

Table 6: Sensitivity analysis of the probability of Diabetes to perturbations in the CPDs of its ancestral nodes within the selected Bayesian Network. A perturbation ratio of 0.5 was used, and the cumulative percentage change in the marginals of Diabetes is reported in descending order.

The findings of Table 6 show that in all the learned models, Diabetes is highly sensitive to BMI, which is corroborated by the findings of the original article. An increase in BMI is associated with a higher likelihood of developing diabetes. However, the learned structures do not exhibit Age or General Health as a variable Diabetes is consistently sensitive to, the way the expert knowledge graphs show. Furthermore, the highest senstivity with Diabetes is exhibited by HighChol in the MAGA model, followed by General Health in PC.

5.5 Causal Relation Analysis

We further carry out a qualitative evaluation to assess how well the learned causal structures align with expert knowledge. Focusing on the high-confidence relationships identified by the domain expert, we compare the extent to which each structure learning algorithm is able to recover these expert-verified links. This evaluation provides a measurable indication of the alignment between data-driven discovery and established domain understanding, with a summary count highlighting how many expert-defined connections each algorithm successfully captures.

Connection from High Confidence	**PC**	**GES**	**HCS**	**SA**	**EA**	**MAGA**	**PSO**	**Averaged**
$HvyAlcoholConsump \rightarrow PhysHlth$	No	No	No	No	No	No	No	No
$Sex \rightarrow Diabetes\_binary$	No	No	No	No	No	No	No	No
$Age \rightarrow Income$	No	Yes	Yes	Yes	No	No	No	No
$Age \rightarrow GenHlth$	No	No	No	No	No	No	Yes	No
$Age \rightarrow Diabetes\_binary$	No	No	No	No	No	No	No	No
$Age \rightarrow BMI$	No	No	No	No	No	No	No	No
$Age \rightarrow PhysActivity$	No	No	No	No	No	No	No	No
$Age \rightarrow DiffWalk$	No	Yes	Yes	Yes	No	No	No	No
$Age \rightarrow PhysHlth$	No	No	No	No	No	No	No	No
$Income \rightarrow GenHlth$	No	No	No	No	No	No	No	No
$Income \rightarrow MentHlth$	No	No	No	No	Yes	No	Yes	No
$Income \rightarrow PhysHlth$	No	No	No	No	No	No	No	No
$Income \rightarrow AnyHealthcare$	No	No	No	No	No	No	No	No
$Income \rightarrow NoDocbcCost$	No	No	No	No	No	No	No	No
$Income \rightarrow CholCheck$	No	No	No	No	No	No	No	No
$HvyAlcoholConsump \rightarrow MentHlth$	No	No	No	No	No	No	No	No
$HvyAlcoholConsump \rightarrow GenHlth$	No	No	No	No	No	No	No	No
$AnyHealthcare \rightarrow NoDocbcCost$	Yes	Yes	Yes	Yes	No	Yes	No	Yes
$AnyHealthcare \rightarrow CholCheck$	No	Yes	Yes	Yes	Yes	Yes	No	Yes
$BMI \rightarrow Diabetes\_binary$	No	Yes	Yes	Yes	No	Yes	No	Yes
$BMI \rightarrow HighBP$	No	Yes	Yes	Yes	No	Yes	No	Yes
$HighBP \rightarrow HeartDiseaseorAttack$	No	Yes	Yes	Yes	No	Yes	Yes	Yes
$HighBP \rightarrow Stroke$	No	No	Yes	No	No	Yes	No	No
$HighBP \rightarrow Diabetes\_binary$	No	Yes	Yes	No	Yes	Yes	No	Yes
$HighChol \rightarrow HeartDiseaseorAttack$	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
$HighChol \rightarrow Diabetes\_binary$	No	Yes	Yes	No	No	Yes	No	No
$GenHlth \rightarrow Diabetes\_binary$	Yes	No	No	No	No	No	No	No
$GenHlth \rightarrow MentHlth$	No	Yes	Yes	Yes	Yes	Yes	No	Yes
$Diabetes\_binary \rightarrow DiffWalk$	No	No	No	No	No	No	No	No
$Education \rightarrow Income$	Yes	Yes	Yes	Yes	Yes	No	No	Yes
$PhysActivity \rightarrow Diabetes\_binary$	No	No	No	No	No	No	No	No
$PhysHlth \rightarrow MentHlth$	No	Yes	Yes	Yes	No	Yes	No	Yes
$PhysHlth \rightarrow HighBP$	No	No	No	No	No	No	No	No
$PhysHlth \rightarrow HighChol$	No	No	No	No	No	No	No	No
$PhysActivity \rightarrow HighChol$	No	No	No	No	No	No	No	No
$PhysActivity \rightarrow HighBP$	No	No	No	No	No	No	No	No
$HeartDiseaseorAttack \rightarrow Stroke$	No	Yes	Yes	Yes	No	Yes	No	Yes
Count of Same Links:	4	14	15	12	5	12	4	11
Total Number of Links:	33	23	22	25	32	25	33	26

Table 7: Comparison of inferred causal links across algorithms, highlighting matches with the high-confidence graph.

Table 7 shows that most of the structure learning algorithms, with the poor exception of PC, EA and PSO, have generally performed well in identifying many of the relationships provided in the high-confidence knowledge graph. Though the Averaged Structure graph provided a decent match with the High Confidence graph, the result could have been improved had a weighted average been taken rather than an equally-weighted average.

Lastly, beyond empirical evaluation, it is important to incorporate qualitative reasoning to assess whether the matched connections are logically sound—supported by domain expertise and existing literature.

6 Conclusion

This project reimplements and extends a prior study to develop a comprehensive decision support system aimed at uncovering and analyzing causal relationships in the context of diabetes. The resulting Streamlit-based application provides a user-friendly interface for visualizing, comparing, and interpreting causal structures derived from both data-driven algorithms and expert knowledge. Key accomplishments of this work include:

- Development of a practical decision support system that facilitates causal exploration and aids researchers in understanding complex relationships among diabetes-related variables
- Comparative evaluation of seven structure learning algorithms, covering both deterministic and stochastic methods, to assess their performance in recovering meaningful causal structures
- Integration of model-averaged graphs, combining outputs from multiple algorithms to derive a more robust and stable representation of underlying causal relationships
- Transformation of causal graphs into Bayesian networks by parameter learning, thereby enabling simulation of hypothetical interventions and sensitivity analyses to better understand risk factors and outcomes
- Qualitative analysis of alignment between learned and expert knowledge graphs, highlighting consistent patterns and discrepancies to validate model outputs

References

[1] Ehtasham Ahmad et al. “Type 2 diabetes”. In: *The Lancet* 400.10365 (2022), pp. 1803–1820.

[2] Ankur Ankan and Abinash Panda. *pgmpy: Python Library for Probabilistic Graphical Models*. Accessed: 2025-06-05. 2015. URL: https://pgmpy.org/.

[3] BayesFusion. *GeNIe Modeler USER MANUAL*. https://support.bayesfusion.com/docs/GeNIe.pdf. [Online; accessed 21-April-2023]. 2023.

[4] João PAF Campos, Itallo G Machado, and Michel Bessani. “Multi-Agent Genetic Algorithm for Bayesian networks structural learning”. In: *Knowledge-Based Systems* 310 (2025), p. 113025.

[5] Centers for Disease Control and Prevention. *Behavioral Risk Factor Surveillance System (BRFSS) 2015 Annual Data*. https://www.cdc.gov/brfss/annual_data/annual_2015.html. Accessed: 2025-06-03. 2016.

[6] David Maxwell Chickering. “Optimal structure identification with greedy search”. In: *Journal of machine learning research* 3.Nov (2002), pp. 507–554.

[7] A Constantinou. “The Bayesys user manual”. In: *Queen Mary University of London, London, UK.[Online]. Software available: http://bayesian-ai. eecs. qmul. ac. uk/bayesys* (2019).

[8] Soheila Gheisari and Mohammad Reza Meybodi. “Bnc-pso: structure learning of bayesian networks by particle swarm optimization”. In: *Information Sciences* 348 (2016), pp. 272–289.

[9] Alireza Sadeghi Hesar. “Structure learning of bayesian belief networks using simulated annealing algorithm”. In: *Middle-East Journal of Scientific Research* 18.9 (2013), pp. 1343–1348.

[10] Md Jamal Hossain, Md Al-Mamun, and Md Rabiul Islam. “Diabetes mellitus, the fastest growing global public health concern: Early detection should be focused”. In: *Health Science Reports* 7.3 (2024), e2004.

[11] Dimitris Margaritis and Sebastian Thrun. “Bayesian network induction via local neighborhoods”. In: *Advances in neural information processing systems* 12 (1999).

[12] Kanyin Liane Ong et al. “Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021”. In: *The Lancet* 402.10397 (2023), pp. 203–234.

[13] Judea Pearl. “Causal diagrams for empirical research (with Discussions)”. In: *Probabilistic and causal inference: The works of Judea Pearl*. 2022, pp. 255–316.

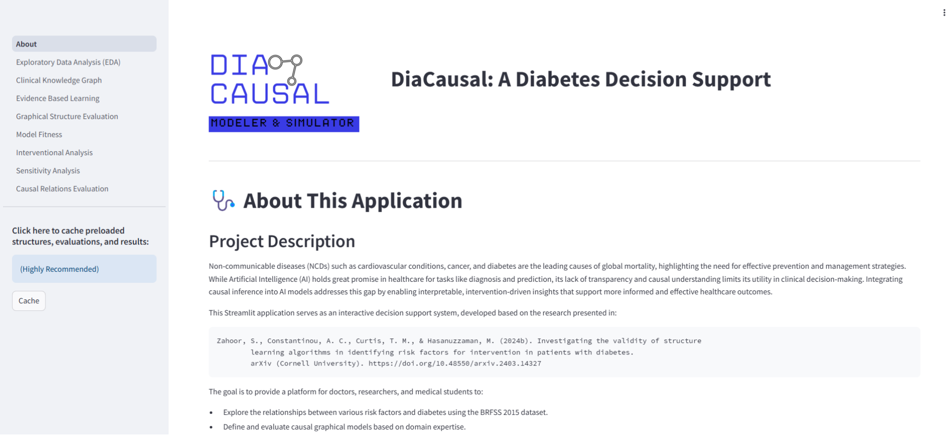
[14] Peter Spirtes and Clark Glymour. “An algorithm for fast recovery of sparse causal graphs”. In: *Social science computer review* 9.1 (1991), pp. 62–72.

[15] Man Leung Wong and Kwong Sak Leung. “An efficient data mining method for learning Bayesian networks using an evolutionary algorithm-based hybrid approach”. In: *IEEE transactions on evolutionary computation* 8.4 (2004), pp. 378–404.

[16] Sheresh Zahoor et al. “Investigating the validity of structure learning algorithms in identifying risk factors for intervention in patients with diabetes”. In: *arXiv preprint arXiv:2403.14327* (2024).

A Appendix: Streamlit Application Interface

A.1 About Page



DiaCausal: A Diabetes Decision Support

About This Application

Project Description

Non-communicable diseases (NCDs) such as cardiovascular conditions, cancer, and diabetes are the leading causes of global mortality, highlighting the need for effective prevention and management strategies. While Artificial Intelligence (AI) holds great promise in healthcare for tasks like diagnosis and prediction, its lack of transparency and causal understanding limits its utility in clinical decision-making. Integrating causal inference into AI models addresses this gap by enabling interpretable, intervention-driven insights that support more informed and effective healthcare outcomes.

This Streamlit application serves as an interactive decision support system, developed based on the research presented in:

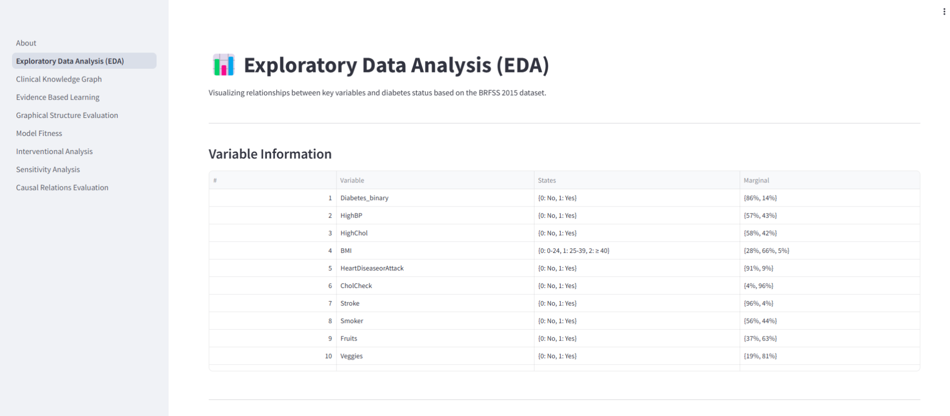
```
Zahoor, S., Constantinou, A. C., Curtis, T. M., & Hasanuzzaman, M. (2024b). Investigating the validity of structure
        learning algorithms in identifying risk factors for intervention in patients with diabetes.
        arXiv (Cornell University). https://doi.org/10.48550/arxiv.2403.14327
```

The goal is to provide a platform for doctors, researchers, and medical students to:

- Explore the relationships between various risk factors and diabetes using the BRFSS 2015 dataset.
- Define and evaluate causal graphical models based on domain expertise.

Figure 11: Opening interface of the Streamlit application. Advised to cache before running the subsequent pages.

A.2 EDA



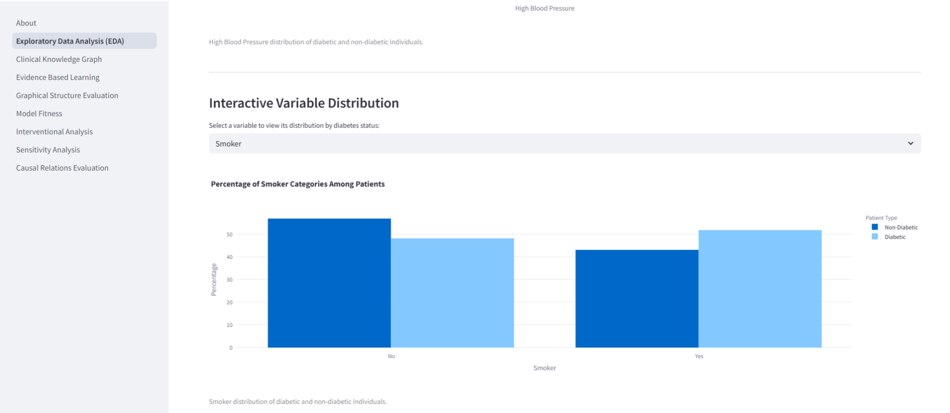
Exploratory Data Analysis (EDA)

Visualizing relationships between key variables and diabetes status based on the BRFSS 2015 dataset.

Variable Information

#	Variable	States	Marginal
1	Diabetes_binary	{0: No, 1: Yes}	{86%, 14%}
2	HighBP	{0: No, 1: Yes}	{57%, 43%}
3	HighChol	{0: No, 1: Yes}	{58%, 42%}
4	BMI	{0: 0-24, 1: 25-39, 2: ≥ 40}	{28%, 66%, 5%}
5	HeartDiseaseorAttack	{0: No, 1: Yes}	{91%, 9%}
6	CholCheck	{0: No, 1: Yes}	{4%, 96%}
7	Stroke	{0: No, 1: Yes}	{96%, 4%}
8	Smoker	{0: No, 1: Yes}	{56%, 44%}
9	Fruits	{0: No, 1: Yes}	{37%, 63%}
10	Veggies	{0: No, 1: Yes}	{19%, 81%}

Figure 12: First half of the EDA page; shows dataset's variable information.



High Blood Pressure distribution of diabetic and non-diabetic individuals.

Interactive Variable Distribution

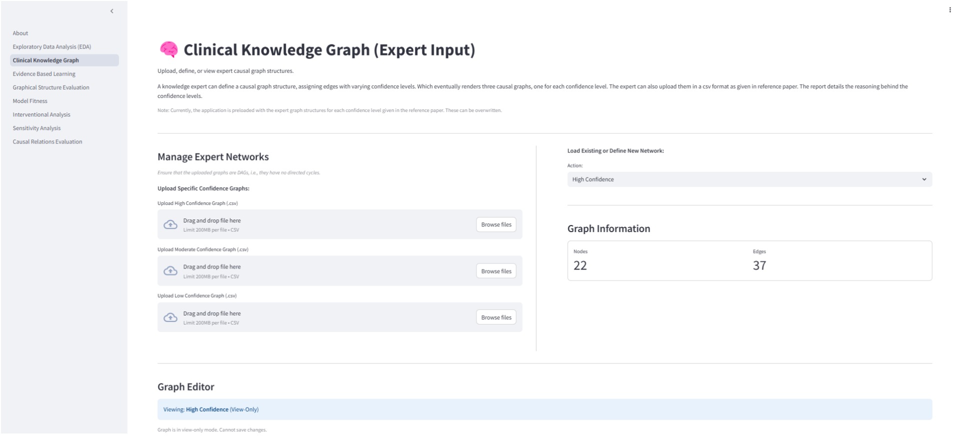
Select a variable to view its distribution by diabetes status: Smoker

Percentage of Smoker Categories Among Patients

Smoker distribution of diabetic and non-diabetic individuals.

Figure 13: Second half of the EDA page; features interactive dropdowns that allow users to visualize how each variable is distributed across diabetes and non-diabetes groups.

A.3 Clinical Knowledge Graphs



🧠 Clinical Knowledge Graph (Expert Input)

Upload, define, or view expert causal graph structures.

A knowledge expert can define a causal graph structure, assigning edges with varying confidence levels. Which eventually renders three causal graphs, one for each confidence level. The expert can also upload them in a csv format as given in reference paper. The report details the reasoning behind the confidence levels.

Note: Currently, the application is preloaded with the expert graph structures for each confidence level given in the reference paper. These can be overwritten.

Manage Expert Networks

Ensure that the uploaded graphs are DAGs, i.e., they have no directed cycles.

Upload Specific Confidence Graphs:

Upload High Confidence Graph (.csv)

Drag and drop file here
Limit 200MB per file • CSV

Upload Moderate Confidence Graph (.csv)

Drag and drop file here
Limit 200MB per file • CSV

Upload Low Confidence Graph (.csv)

Drag and drop file here
Limit 200MB per file • CSV

Load Existing or Define New Network:

Action:
High Confidence

Graph Information

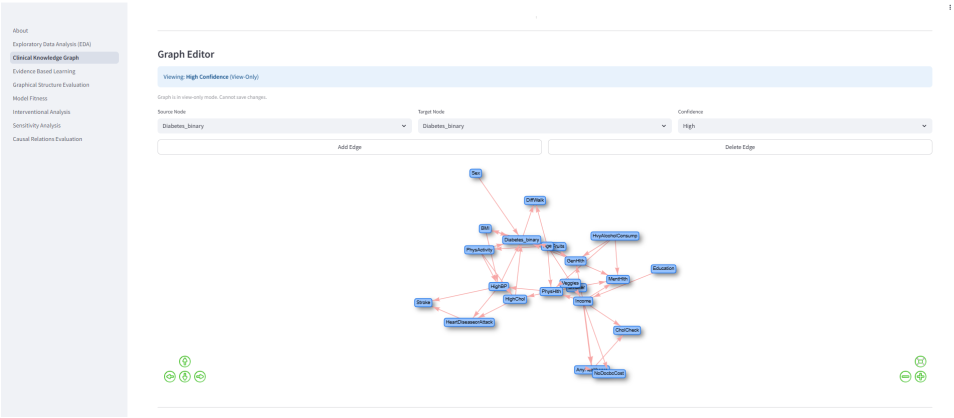
Nodes	Edges
22	37

Graph Editor

Viewing: **High Confidence** (View-Only)

Graph is in view-only mode. Cannot save changes.

Figure 14: Configuration panel for domain-expert knowledge graphs; users can upload CSVs of defined networks and view graph information about nodes and edges in the display or they can select the pre-defined/pre-loaded knoweldge graphs by selecting from the dropdown.



Graph Editor

Viewing: **High Confidence** (View-Only)

Graph is in view-only mode. Cannot save changes.

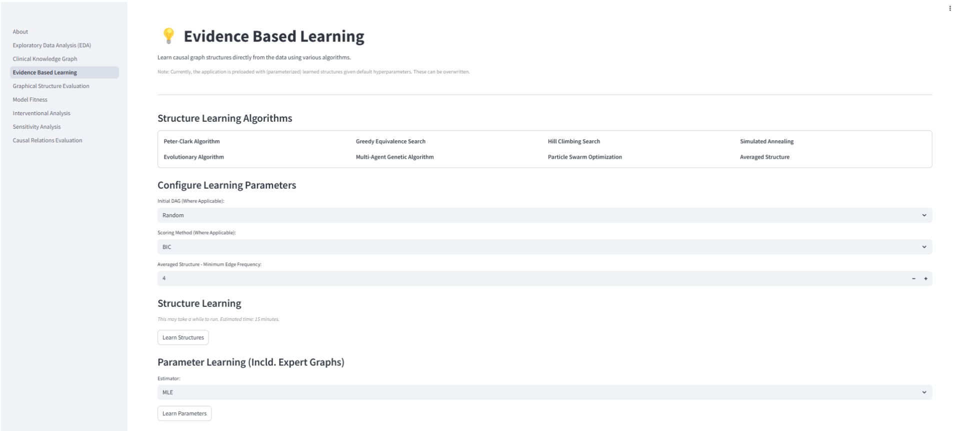
Source Node: Diabetes_binary

Target Node: Diabetes_binary

Confidence: High

Figure 15: Graph editor for viewing knowledge graphs in view-only mode.

A.4 Structure and Parameter Learning



💡 Evidence Based Learning

Learn causal graph structures directly from the data using various algorithms.

Note: Currently, the application is preloaded with (parameterized) learned structures given default hyperparameters. These can be overwritten.

Structure Learning Algorithms

Peter-Clark Algorithm	**Greedy Equivalence Search**	**Hill Climbing Search**	**Simulated Annealing**
Evolutionary Algorithm	**Multi-Agent Genetic Algorithm**	**Particle Swarm Optimization**	**Averaged Structure**

Configure Learning Parameters

Initial DAG (Where Applicable):
Random

Scoring Method (Where Applicable):
BIC

Averaged Structure - Minimum Edge Frequency:
4

Structure Learning

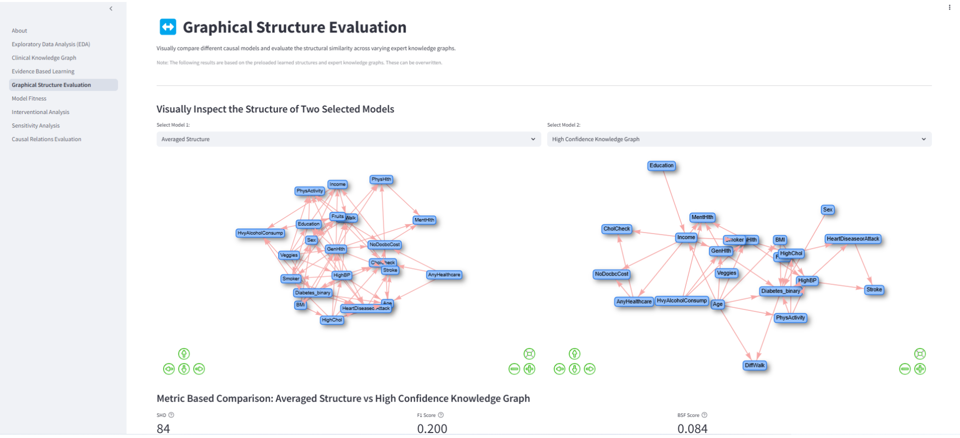
This may take a while to run. Estimated time: 15 minutes.

Parameter Learning (Incld. Expert Graphs)

Estimator:
MLE

Figure 16: Structure and Parameter Learning interactivity; models are pre-loaded however selecting the desired configurations where applicable and invoking the Learn Structures and Learn Parameters button, learns all the models together and overwrites the pre-loaded ones.

A.5 Graph Structure Evaluation



Graphical Structure Evaluation

Visually compare different causal models and evaluate the structural similarity across varying expert knowledge graphs.

Note: The following results are based on the preloaded learned structures and expert knowledge graphs. These can be overwritten.

Visually Inspect the Structure of Two Selected Models

Select Model 1: Averaged Structure

Select Model 2: High Confidence Knowledge Graph

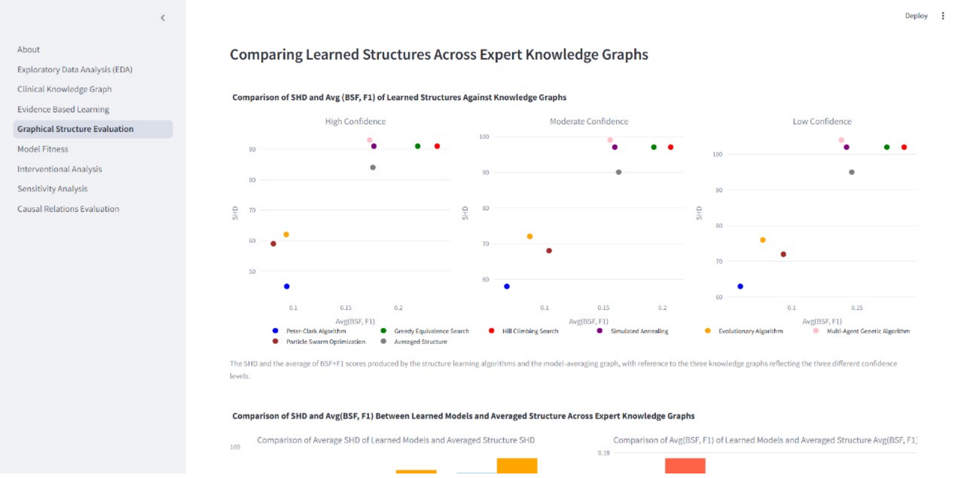
Metric Based Comparison: Averaged Structure vs High Confidence Knowledge Graph

SHD 84

F1 Score 0.200

BSF Score 0.084

Figure 17: Comparison of graphs and display of the evaluation metrics: SHD, F1-Score and BSF.



Comparing Learned Structures Across Expert Knowledge Graphs

Comparison of SHD and Avg (BSF, F1) of Learned Structures Against Knowledge Graphs

The SHD and the average of BSF+F1 scores produced by the structure learning algorithms and the model-averaging graph, with reference to the three knowledge graphs reflecting the three different confidence levels.

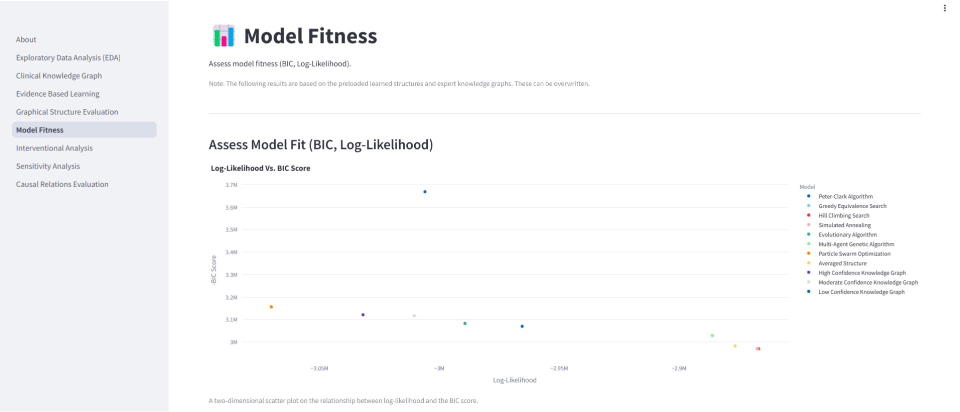
Comparison of SHD and Avg(BSF, F1) Between Learned Models and Averaged Structure Across Expert Knowledge Graphs

Comparison of Average SHD of Learned Models and Averaged Structure SHD

Comparison of Avg(BSF, F1) of Learned Models and Averaged Structure Avg(BSF, F1)

Figure 18: Comparing learned structures across expert knowledge graphs by stratified scatterplots and barcharts.

A.6 Model Fitness page



Model Fitness

Assess model fitness (BIC, Log-Likelihood).

Note: The following results are based on the preloaded learned structures and expert knowledge graphs. These can be overwritten.

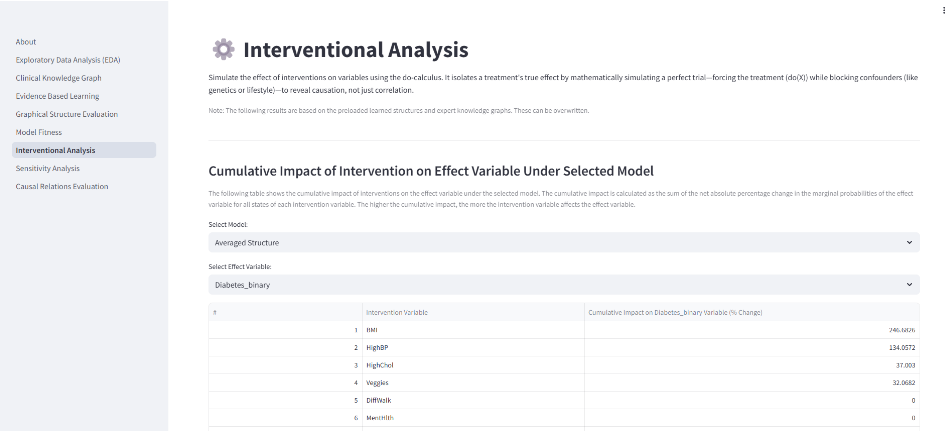
Assess Model Fit (BIC, Log-Likelihood)

Log-Likelihood Vs. BIC Score

A two-dimensional scatter plot on the relationship between log-likelihood and the BIC score.

Figure 19: Display of Model Fitness for all learned models; visualized by the scatterplot of BIC against LL.

A.7 Interventional Analysis



Interventional Analysis

Simulate the effect of interventions on variables using the do-calculus. It isolates a treatment's true effect by mathematically simulating a perfect trial—forcing the treatment (do(X)) while blocking confounders (like genetics or lifestyle)—to reveal causation, not just correlation.

Note: The following results are based on the preloaded learned structures and expert knowledge graphs. These can be overwritten.

Cumulative Impact of Intervention on Effect Variable Under Selected Model

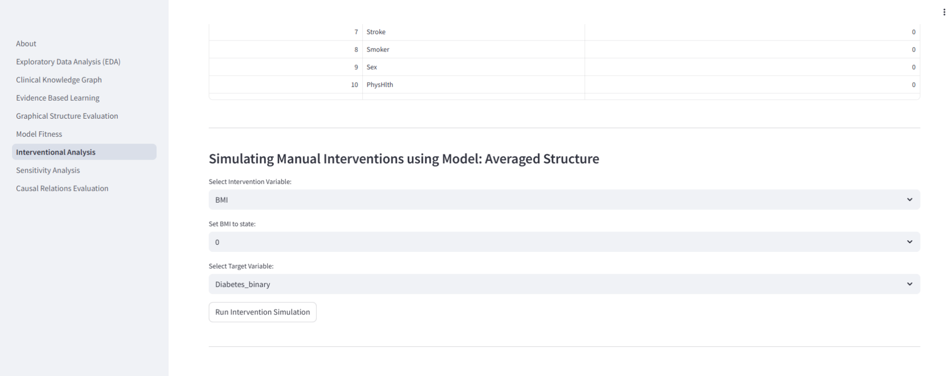
The following table shows the cumulative impact of interventions on the effect variable under the selected model. The cumulative impact is calculated as the sum of the net absolute percentage change in the marginal probabilities of the effect variable for all states of each intervention variable. The higher the cumulative impact, the more the intervention variable affects the effect variable.

Select Model: Averaged Structure

Select Effect Variable: Diabetes_binary

#	Intervention Variable	Cumulative Impact on Diabetes_binary Variable (% Change)
1	BMI	246.6826
2	HighBP	134.0572
3	HighChol	37.003
4	Veggies	32.0682
5	DiffWalk	0
6	MentHlth	0

Figure 20: Tabulated version of cumulative impacts of interventions for selected effect variable under chosen model.


About
Exploratory Data Analysis (EDA)
Clinical Knowledge Graph
Evidence Based Learning
Graphical Structure Evaluation
Model Fitness
Interventional Analysis
Sensitivity Analysis
Causal Relations Evaluation

7	Stroke	0
8	Smoker	0
9	Sex	0
10	PhysHlth	0

Simulating Manual Interventions using Model: Averaged Structure

Select Intervention Variable: BMI

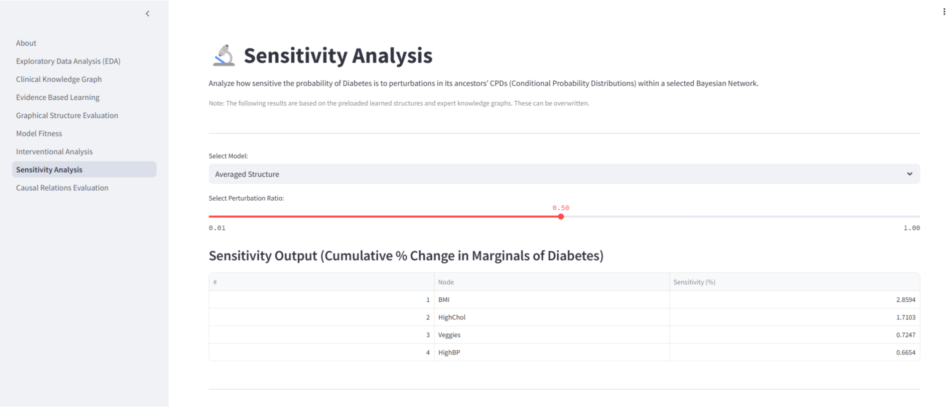
Set BMI to state: 0

Select Target Variable: Diabetes_binary

Run Intervention Simulation

Figure 21: Simulating manual interventions by changing intervention, its state and the effect variable under a selected algorithm

A.8 Sensitivity Analysis



Sensitivity Analysis

Analyze how sensitive the probability of Diabetes is to perturbations in its ancestors' CPDs (Conditional Probability Distributions) within a selected Bayesian Network.

Note: The following results are based on the preloaded learned structures and expert knowledge graphs. These can be overwritten.

Select Model: Averaged Structure

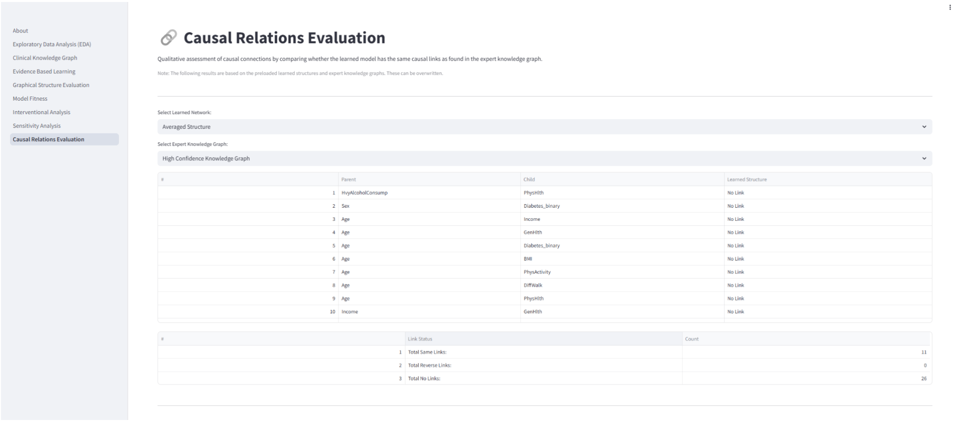
Select Perturbation Ratio: 0.50

Sensitivity Output (Cumulative % Change in Marginals of Diabetes)

#	Node	Sensitivity (%)
1	BMI	2.8594
2	HighChol	1.7103
3	Veggies	0.7247
4	HighBP	0.6654

Figure 22: Interactivity of Sensitivity Analysis page that defines the perturbation ratio by the slider and tabublates the cumulative percentage change in the marginals in Diabetes.

A.9 Causal Relations Evaluation



Causal Relations Evaluation

Qualitative assessment of causal connections by comparing whether the learned model has the same causal links as found in the expert knowledge graph.

Note: The following results are based on the preloaded learned structures and expert knowledge graphs. These can be overwritten.

Select Learned Network: Averaged Structure

Select Expert Knowledge Graph: High Confidence Knowledge Graph

#	Parent	Child	Learned Structure
1	HvyAlcoholConsump	PhysHlth	No Link
2	Sex	Diabetes_binary	No Link
3	Age	Income	No Link
4	Age	GenHlth	No Link
5	Age	Diabetes_binary	No Link
6	Age	BMI	No Link
7	Age	PhysActivity	No Link
8	Age	DiffWalk	No Link
9	Age	PhysHlth	No Link
10	Income	GenHlth	No Link

#	Link Status	Count
1	Total Same Links:	11
2	Total Reverse Links:	0
3	Total No Links:	26

Figure 23: Interactivity of Evaluating Causal Relations between a pair of Structures which shows the types of matches and their aggregate.