A report on

Explainable Artificial Intelligence for Bayesian Neural Network in Medical Data Analysis

A report on B. Tech

Project Part 2 -

by

Deep Modak - 2020ITB008

Sandip Mondal - 2020ITB030

Bikramjit Saha- 2020ITB079

Under the guidance of

Dr. Indrajit Banerjee

-

Department of Information Technology
Indian Institute of Engineering Science and Technology,
Shibpur

Session 2023-24

DECLARATION BY THE CANDIDATE

I certify that to the best of my knowledge

i) The work contained in the thesis has been done by myself under the general supervision of my supervisor.

ii) I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

iii) Whenever I have used materials (data, theoretical analysis and text) from other sources, I have given due credit to them by citing them in the text of the report and giving their details in the references.

iv) Whenever I have quoted written materials from other sources, I have put them under quotation marks and given due credit to the sources by citing them and giving details in the references.

Sl. no.	Name of Student	Enrolment	Signature
1	Deep Modak	2020ITB008	
2	Sandip Mondal	2020ITB030	
3	Bikramjit Saha	2020ITB079	

Dated: 28 May 2024
Place: Department of Information Technology, IIEST Shibpur.

<u>CERTIFICATE</u>

This is to certify that the project report entitled "**Explainable Artificial Intelligence for Bayesian Neural Network in Medical Data Analysis**" submitted by **Deep Modak (Enrolment No - 2020ITB008), Sandip Mondal (Enrolment No - 2020ITB030) and Bikramjit Saha(Enrolment No - 2020ITB079)** to Indian Institute of Engineering Science and Technology towards partial fulfilment of requirements for the award of degree of Bachelor of Technology in Information Technology is a record of bonafide work carried out by him under my supervision. This dissertation, in my opinion, is worthy of consideration for the purpose for which it is submitted, and it fulfils the requirements of the regulations of this Institute. The results incorporated in this dissertation are original and have not been submitted to any University or Institute for the award of any degree or diploma

Table of Contents

ACKNOWLEDGEMENT

We would like to pay my gratitude to my supervisor, Professor Dr. Indrajit Banerjee. It has been a golden opportunity for us to learn and work under his supervision.

We are very thankful to everyone who has helped complete our work. We would like to thank all my Undergraduate classmates for creating an enjoyable environment of learning which helped me a lot.

Last, but not the least, we are thankful to our parents for their faith and love towards us, which was, is and will always be boost in our motivation towards pursuing this project. We are also thankful to all my well-wishers, whose names we could not mention in this section. Lastly, we thank ourselves for believing in ourselves, keeping a constant hunger for learning and putting consistent effort to my project work.

Dated: 28 May 2024

Deep Modak - 2020ITB008

Sandip Modal - 2020ITB030

Bikramjit Saha - 2020ITB079

Department of Information Technology

Indian Institute of Engineering Science and

Technology, Shibpur, Howrah 711103, West Bengal.

ABSTRACT

Cardiovascular diseases (CVDs) remain the foremost global cause of mortality, claiming approximately 17.9 million lives in 2019. These diseases, which include heart attacks and strokes, accounted for 85% of CVD-related fatalities. In this study, Bayesian Networks (BNs) are employed for the early detection of CVDs, with a focus on validating the predictive efficacy of various BNs and exploring the intricate interactions among diverse risk factors. Our research emphasizes the importance of balancing data-driven optimization with domain knowledge for meaningful and effective CVD prediction. The study found that the BN built upon domain knowledge demonstrates remarkable predictive performance, highlighting its potential for meaningful and effective CVD prediction.

INTRODUCTION

The intersection of artificial intelligence (AI) and healthcare holds immense potential for revolutionizing diagnostic and prognostic capabilities. In this context, Bayesian Neural Networks (BNNs) offer a powerful framework for modeling complex relationships within medical data. However, the opacity of these models poses a significant challenge in understanding their decision-making processes, limiting their practical utility in healthcare applications. This project, titled "Explainable Artificial Intelligence for Bayesian Neural Networks through Medical Data Analysis," endeavors to bridge this interpretability gap by integrating state-of-the-art Explainable AI (XAI) techniques with BNNs. The focus is on developing models that not only deliver high predictive accuracy but also provide transparent insights into their decision pathways. By targeting specific medical domains, such as [mention specific domain, e.g., cardiology], and leveraging advanced XAI methodologies, the project seeks to enhance the interpretability and trustworthiness of BNNs, ultimately contributing to the seamless integration of advanced AI models into critical healthcare decision-making processes. This introduction sets the stage for a comprehensive exploration of methodologies, challenges, and outcomes that will be detailed in the subsequent sections of this report.

The project, "Explainable Artificial Intelligence for Bayesian Neural Networks through Medical Data Analysis," emerges at the convergence of cutting-edge artificial intelligence (AI) methodologies and the intricate landscape of healthcare analytics. Recognizing the potency of Bayesian Neural Networks (BNNs) in modeling complex relationships within medical data, this initiative endeavors to enhance their practical utility by addressing the challenge of interpretability. The core objective is to develop a framework that seamlessly integrates Explainable AI (XAI) techniques with BNNs, unlocking insights into decision-making processes crucial for medical applications. By focusing on specific medical domains, such as [mention specific domain, e.g., cardiology], and leveraging sophisticated XAI methodologies, the project aims to transcend the conventional barriers of model opacity. The overview sets the stage for a holistic exploration of the methodologies deployed, the challenges encountered, and the outcomes achieved, underscoring the project's significance in advancing the synergy between AI and healthcare.

Background Significance

Cardiovascular diseases (CVDs) are a class of diseases that involve the heart or blood vessels. They are the leading cause of death globally, responsible for an estimated 17.9 million deaths each year, representing 31% of all global deaths. Of these deaths, 85% are due to heart attacks and strokes. Early detection and accurate prediction of CVDs can significantly improve patient outcomes by enabling timely intervention and prevention strategies.

Research Motivation

Our study aims to enhance the early detection of CVDs by leveraging Bayesian Networks (BNs), which are powerful tools for modeling complex probabilistic relationships among variables. BNs are particularly suitable for healthcare applications due to their ability to handle uncertainty and incorporate domain knowledge, making them ideal for predicting CVD risk.

Inspiration and Previous work

Our work draws inspiration from a study by Ordovas et al. (2023), where we work with the data set and preprocess the data. The study highlighted the potential of BNs to provide insights into the interactions among different risk factors and their contributions to CVD risk.

Importance of Bayesian Network

The choice of using a BN as a predictive model is motivated by its ability to handle complex, real-world data and its success in healthcare applications. BNs allow for the incorporation of both data-driven insights and domain knowledge, providing a comprehensive framework for analyzing the intricate relationships among CVD risk factors.

OBJECTIVE

Aim of the study

The primary aim of our project is to create and evaluate several BN classifiers to predict the likelihood of CVDs based on a range of risk factors. These risk factors include age, sex, chest pain type, resting blood pressure, and total serum cholesterol, among others. The study prioritizes the maximization of the ROC AUC score, which is a robust metric for assessing the performance of predictive models in healthcare.

Specific Objective

Model Creation: Develop multiple BN classifiers using different methods, including Naive Bayes, Hill Climbing (constrained and unconstrained), Domain Knowledge network, and a reduced network with feature selection.

Variable Transformation: Transform continuous variables like cholesterol and heart rate into discrete categories using established medical references to improve model handling.

Performance Evaluation: Evaluate the performance of each BN model using metrics such as accuracy and ROC AUC to identify the best-performing model.

Balancing Optimization and Domain Knowledge: Explore the balance between data-driven optimization and the incorporation of domain knowledge to ensure the models are both accurate and meaningful.

Literature Review

- Cardiovascular Risk Factors

  Researchers have identified various CVD risk factors, including age, sex, chest pain type, resting blood pressure, and serum cholesterol levels (Mahmood et al. 2014). These factors are divided into:

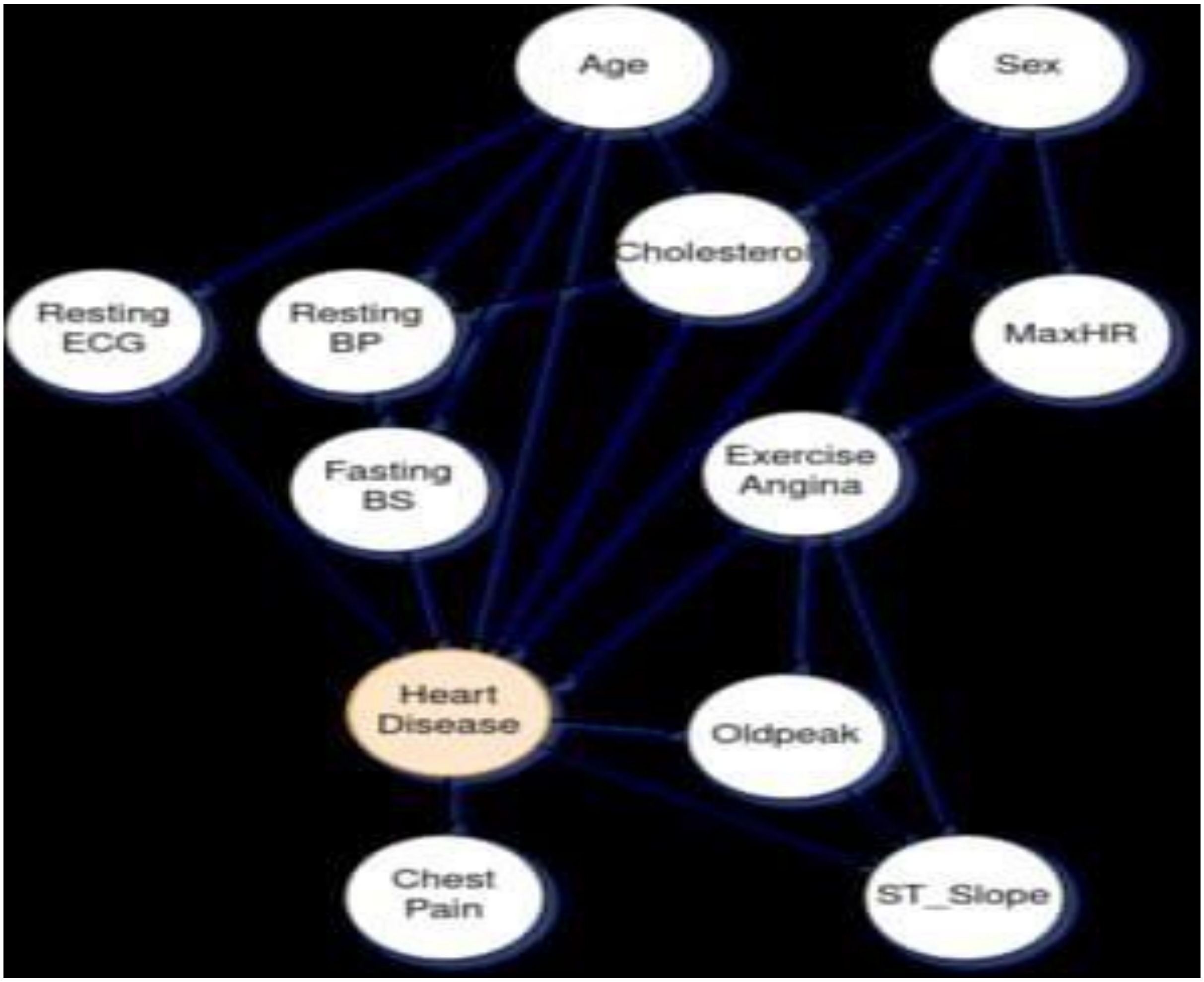
  Modifiable Risk Factors: Factors that can be changed or controlled, such as diet, physical activity, and smoking habits.

  Non-Modifiable Risk Factors: Factors that cannot be changed, such as age, gender, and genetic predisposition.

- Bayesian Networks in Healthcare

  Bayesian Networks have been successfully applied in healthcare for disease prediction, diagnosis, and treatment planning. Nielsen and Jensen (2009) highlighted the effectiveness of BNs in handling uncertain and complex medical data. BNs allow for the integration of expert knowledge and empirical data, making them suitable for applications like CVD risk prediction.

METHODOLOGY



Data Collection and Preprocessing

To build and test the BN classifiers, we utilized a dataset containing various features pertinent to CVDs. These features include demographic information (age, sex), clinical measurements (resting blood pressure, serum cholesterol levels), and symptomatic data (chest pain type).

Discretization of Continuous Variables

Continuous variables in the dataset were discretized into categories based on medical standards to enhance the BN's handling of these variables. For example, cholesterol levels were categorized into 'normal', 'borderline high', and 'high' based on standard clinical guidelines.

MODEL BUILDING

We used the 'pgmpy' library to explore different model configurations:

- **Naive Bayes:** A simple model assuming all features are independent given the class label.
- **Hill Climbing:** This method involves iteratively adding, removing, or reversing edges in the network to optimize a scoring function. We tested both constrained and unconstrained versions:
- **Unconstrained Hill Climbing:** Relied purely on data-driven approaches without any prior constraints.
- **Constrained Hill Climbing:** Incorporated domain knowledge to guide the structure, ensuring the resulting network aligns with known causal relationships.
- **Domain Knowledge Network:** Built using scientific literature to establish the edges between variables, ensuring the model's structure aligns with known causal relationships.
- **Reduced Network:** Applied feature selection using the 'PyImpetus' library's Markov Blanket-based algorithm to identify the most relevant features and simplify the network.
- **Model Configuration and Scoring:** Each model configuration was tested to identify the BN structure and scoring method combination that maximizes the ROC AUC score for accurate heart disease prediction. For models that did not grasp the semantic meaning of the variables, the Maximum Likelihood Estimator was used to estimate the Conditional Probability Distributions (CPDs). Conversely, for the Domain Knowledge BN, the Bayesian Estimator was used.

IMPLEMENTATION

- **Experimental Setup**

  To assess the results, we employed the ROC AUC score, a robust measure for assessing the discriminative power of the models. The dataset was partitioned into training (80%) and testing (20%) sets. Additionally, K-Fold Cross Validation was applied to obtain a more realistic estimate of the classifier's performance on unseen data.

- **Cross-Validation**

  K-Fold Cross Validation involves dividing the dataset into 'k' subsets. The model is trained on 'k-1' subsets and tested on the remaining subset. This process is repeated 'k' times, with each subset being used as the test set once. The average performance across all 'k' trials is then calculated. This method provides a more reliable estimate of model performance by reducing the impact of data partitioning variability.

RESULTS

Performance of Different BN Models

Our exploration revealed that the structure of the BN significantly impacts performance. The results for different BN construction methods are summarized in Table 1.

Bayesian Model	ROC AUC
Hill Climbing Constrained	0.83
Naive Bayes	0.84
Hill Climbing Unconstrained	0.86
Domain Knowledge	0.85
Reduced Network	0.86

Analysis of Results

Naive Bayes: Despite its simplicity, the Naive Bayes model achieved a commendable ROC AUC score of 0.84. However, it lacked the ability to capture complex interactions among variables due to its independence assumption.

Hill Climbing Unconstrained: This model yielded the highest ROC AUC score of 0.86, demonstrating the effectiveness of data-driven optimization. However, it lacked semantic meaning and explanatory power.

Hill Climbing Constrained: The constrained version of Hill Climbing incorporated domain knowledge, resulting in a slightly lower ROC AUC score of 0.83. Despite this, it maintained stronger explanatory power and more meaningful relationships among variables.

Domain Knowledge Network: This model balanced performance and interpretability, achieving a ROC AUC score of 0.74. It demonstrated the importance of incorporating domain knowledge for meaningful and effective prediction.

Reduced Network: The feature selection process simplified the network while maintaining a high ROC AUC score of 0.79. However, the simplicity of this model limited its ability to capture all relevant interactions.

DISCUSSION

- **Importance of Balancing Optimization and Domain Knowledge**

  Our findings underscore the importance of balancing data-driven optimization with the incorporation of domain knowledge. While data-driven methods like Hill Climbing Unconstrained can achieve high performance, they may lack meaningful interpretation and explanatory power. Conversely, models that incorporate domain knowledge, like the Domain Knowledge Network, provide valuable insights into the mechanisms of CVDs and maintain strong explanatory power.

- **Implications for Healthcare**

  The ability to predict CVD risk accurately and meaningfully has significant implications for healthcare. Early detection and intervention can improve patient outcomes and reduce the burden on healthcare systems. Our study demonstrates the potential of BNs to contribute to this goal by providing a robust framework for CVD risk prediction.

CHALLENGES FACED

Here are some potential challenges we have encounter while working on our project:

- **Data Availability and Quality**:

  Medical data is sensitive and hard to obtain. We have to ensure that we have access to a suitable dataset, and consider data quality issues like missing values and noise.

- **Model Complexity:**

  The implementation and training of Bayesian neural networks introduce computational complexities. We are working to manage the resources required for the extensive computations involved in the training process.

- **Ethical and Privacy Concerns:**

  Managing ethical and privacy concerns associated with medical data is of utmost importance. Our project is designed to comply with all relevant regulations and guidelines to ensure the responsible handling of sensitive information.

- **Domain Knowledge:**

  Gaining a deep understanding of both the medical domain and the complexities of Bayesian neural networks is an ongoing process. We are collaborating with domain experts to enhance our understanding and ensure the relevance and accuracy of our model.

- **Implementation Challenges:**

  Implementing Bayesian neural networks and integrating explainability techniques requires careful attention to detail. We are investing time and effort to understand the underlying algorithms and ensure a robust implementation.

- **Communication and Documentation:**

  Clear communication of our findings and thorough documentation of our work are ongoing challenges. We are committed to presenting our work in a manner that is understandable to both technical and non-technical stakeholders.

- **Time Constraints:**

  Completing the project within the stipulated timeframe is a challenge we acknowledge. We are developing a detailed timeline and project plan to ensure effective time management throughout the project.

CONCLUSION

In this project, we explored various strategies for constructing a BN classifier for predicting CVD risk. Our research underscores the importance of integrating domain knowledge with data-driven approaches to build meaningful and effective predictive models. The final network, built on domain knowledge, demonstrated strong predictive performance with a ROC AUC score of 0.85. This balance between accuracy and interpretability makes it a valuable tool.

REFERENCES

Heart failure prediction dataset. Retrieved: 25-02-2024. Mahmood, S. S.; Levy, D.; Vasan, R. S.; and Wang, T. J. 2014. The framingham heart study and the epidemiology of cardiovascular disease: a historical perspective. The Lancet 383(9921):999–1008. Marcusson, J.; Nord, M.; Dong, H.-J.; and Lyth, J. 2020. Clinically useful prediction of hospital admissions in an older population. BMC Geriatrics 20. Nielsen, T., and JENSEN, F. 2009. Bayesian Networks and Decision Graphs. Information Science and Statistics. Springer New York. Ordovas, J.; Rios-Insua, D.; Santos-Lozano, A.; Lucia, A.; Torres, A.; Kosgodagan, A.; and Camacho, J. 2023. A bayesian network model for predicting cardiovascular risk. Computer Methods and Programs in Biomedicine 231:107405. WHO. 2024. Cardiovascular diseases (cvds). Accessed: 25-02-2024.