Research proposal:
AI-Driven Causal Inference for Tailored Treatment Strategies in Chronic Disease Management

Fatima Zahra Ouardirhi
Final-Year Data Engineering student
Ecole Nationale Supérieure des Mines de Rabat
Expected graduation: September 2025
Email: fatimazahra.ouardirhi@enim.ac.ma or
FatimaZahra.OUARDIRHI@student.umons.ac.be

***Abstract*—Subject 2: Chronic diseases such as diabetes, cardiovascular disorders, and hypertension continue to be leading causes of morbidity and mortality globally. This phenomenon is particularly alarming given the increasing prevalence of these conditions in diverse populations, driven by factors such as urbanization, lifestyle changes, and aging demographics. Personalized medicine offers the promise of tailoring treatments to individual patient profiles, but traditional predictive models often fail to distinguish between correlation and true causation. This research project aims to develop and validate AI-enhanced causal inference models — including causal forests, Bayesian networks, and meta-learners — capable of estimating individualized treatment effects (ITE) using multimodal electronic health data. By integrating clinical, demographic, and laboratory data while applying machine learning, causal discovery algorithms, and explainable AI (XAI) techniques, the objective is to develop models trained using the UK BioBank that not only predict patient outcomes but also identify the optimal treatment strategies for each individual. Previous research using MIMIC-IV and eICU public datasets has shown great potential for benchmarking and validation. However, a key contribution would be improving the Moroccan electronic health records (EHRs) to help ensure causal models are population-representative and ethically robust. Hence, this work introduces a next-generation AI framework that moves beyond predictive analytics to deliver trustworthy, personalized, and equitable treatment recommendations, fundamentally reshaping clinical decision-making and improving healthcare precision.**

Keywords: Causal Inference; Causality; Explainable AI (XAI); Personalized Medicine; Multimodal Data; Healthcare Precision; Clinical Decision-Making

I. Introduction

Chronic diseases such as diabetes, cardiovascular disorders, and hypertension are among the top causes of mortality and morbidity globally. The burden of these diseases continues to rise, particularly in developing regions, due to urbanization, sedentary lifestyles, and aging populations. Conventional predictive analytics in medicine, while useful, often fail to uncover true causal relationships and cannot reliably guide personalized treatment strategies. In this context, causal inference, the process of identifying cause-effect relationships, offers a promising avenue for improving healthcare outcomes.

This proposal outlines a research project aimed at developing and validating artificial intelligence (AI)-driven causal inference models to estimate individualized treatment effects (ITE) using multimodal electronic health records (EHRs). Our goal is to enable precision medicine through trustworthy, explainable, and ethically robust AI tools.

II. Literature Review

In recent years, causal inference has emerged as a key discipline in data-driven healthcare. One notable development is the parametric causal forest for estimating heterogeneous treatment effects, which extends Breiman's widely used random forest algorithm. These models are pointwise consistent for the true treatment effect and exhibit an asymptotically Gaussian and centered sampling distribution. Experiments demonstrate that causal forests outperform classical methods such as nearest-neighbor matching, especially in the presence of irrelevant covariates [3]. Additionally, Bayesian networks have been employed to leverage prior knowledge and probabilistic reasoning. By capturing conditional dependencies, they reveal underlying causal structures [2].

The availability of large-scale datasets containing fine-grained information on individuals has further intensified interest in understanding how treatment effects vary across populations and contexts. To address this, Künzel et al. introduced the X-learner, a meta-algorithm that is particularly efficient when treatment groups are imbalanced. The X-learner, along with T-learners and S-learners, frames individualized treatment effect (ITE) estimation as a supervised learning problem, making it adaptable to various data modalities [1]. On the other hand, explainable AI (XAI) methods such as SHAP and LIME are increasingly employed to interpret model predictions, a critical requirement for clinical transparency.

Despite the validation of these approaches on public datasets such as MIMIC-IV and eICU, their application to diverse and underrepresented populations—such as those in Morocco—remains scarce. Integrating Moroccan electronic health records (EHRs) into these methodologies will enable a deeper

evaluation of their generalizability and the ethical considerations involved in real-world clinical settings.

III. Problem Statement

The increasing use of machine learning in healthcare has enabled significant advancements in disease prediction, diagnosis, and treatment recommendations. However, many existing predictive models remain limited in their ability to support individualized clinical decision-making due to their lack of causal interpretability. These models often rely on correlations within electronic health records (EHRs) and fail to account for the underlying causal relationships between patient characteristics, treatments, and outcomes. As a result, they tend to generate generalized treatment suggestions that may not be effective, or even safe, for all individuals—particularly for underrepresented or diverse populations.

This issue is further compounded in regions such as Morocco, where local healthcare data is sparse, and most models are trained on Western populations, leading to poor generalization and potential disparities in care. Without models that can estimate Individual Treatment Effects (ITE) reliably and transparently, clinicians are left without the evidence-based tools needed to tailor interventions to specific patients, which may undermine trust and clinical adoption.

There is thus an urgent need to develop trustworthy, interpretable, and population-sensitive causal inference models capable of uncovering true treatment effects across heterogeneous groups. These models must also integrate Explainable AI (XAI) techniques to ensure that their predictions are not only accurate but also understandable and actionable by medical professionals. Addressing this gap is essential for advancing equitable and personalized healthcare delivery, both globally and in Morocco.

IV. Research Objectives

This research aims to develop causal inference models capable of estimating individualized treatment effects (ITE) from multimodal electronic health records (EHRs). By integrating explainable AI (XAI) techniques, the study ensures that the resulting models are interpretable and clinically usable. The models will be validated using both UK BioBank and anonymized Moroccan EHR datasets to assess performance and generalizability across populations. Ultimately, the project will provide policy recommendations to support ethical, fair, and representative implementation of AI in Moroccan healthcare systems.

V. Research Design and Methods

a) Data Sources: This study will utilize two primary data sources. The first is the UK BioBank, a large-scale biomedical database providing rich, high-quality longitudinal data, which will serve for model training, development, and benchmarking. The second data source comprises electronic health records (EHRs) collected from Moroccan hospitals, in collaboration with the Faculty of Medical Sciences at UM6P Hospitals. These records will be anonymized and rigorously preprocessed to ensure data privacy and quality, enabling evaluation of the model's generalizability to a North African population, which remains underrepresented in current causal inference research. Additional data sources may also be incorporated as needed throughout the research to enhance model robustness and applicability.

b) Data Types: Multimodal data will be incorporated to capture comprehensive patient profiles. This includes demographic data such as age, gender, and geographic location, clinical information encompassing diagnoses, prescribed medications, treatment regimens, and observed outcomes, as well as laboratory test results reflecting patients' physiological status. The integration of these heterogeneous data types aims to enrich the causal models and improve individualized treatment effect estimation.

c) Causal Inference Methods: We will employ state-of-the-art causal inference methodologies tailored to heterogeneous treatment effect estimation. Causal Forests will be applied to leverage ensemble tree-based methods for robust and interpretable effect estimation across subpopulations. Bayesian Networks will be constructed to model the underlying causal structures explicitly, allowing the incorporation of expert domain knowledge and conditional dependencies. Furthermore, meta-learners including T-learners, S-learners, and X-learners will be implemented to provide flexible frameworks capable of handling diverse data distributions and facilitating precise individualized treatment effect (ITE) prediction.

d) Machine Learning Tools: Implementation will leverage widely used, open-source machine learning libraries and frameworks to ensure reproducibility and community support. Scikit-learn will provide foundational algorithms and utilities, while EconML and DoWhy will be used for specialized causal inference modeling. Pyro, a probabilistic programming library built on PyTorch, will enable advanced Bayesian modeling and uncertainty quantification within the causal framework.

e) Explainability Techniques: To enhance clinical trust and usability, explainability methods will be integrated into the modeling pipeline. SHAP (Shapley Additive Explanations) will be used to quantify the contribution of each feature to individual predictions, providing global and local interpretability. LIME (Local Interpretable Model-Agnostic Explanations) will complement SHAP by generating locally faithful explanations of model decisions, helping clinicians understand the rationale behind treatment effect estimates in specific cases.

f) Evaluation Metrics: The performance of the causal inference models will be rigorously evaluated using multiple complementary metrics. Precision in individualized treatment effect estimation will be assessed using the Precision in Estimation of Heterogeneous Effect (PEHE), as well as average treatment effect (ATE) and average treatment effect on the treated (ATT). Additionally, fairness metrics will be computed to ensure that model performance is consistent across demographic subgroups such as gender and geographic regions, addressing potential biases. Finally, clinical validation will be conducted through expert review to assess the plausibility and practical relevance of the estimated treatment effects.

VI. Contribution to Knowledge

This project contributes to the field in multiple ways. First, it proposes a novel framework that combines causal inference and explainable AI (XAI), specifically tailored for chronic disease treatment. It also establishes ethical AI guidelines for healthcare deployment in low- and middle-income countries (LMICs), with a focus on population-specific model validation using Moroccan EHRs. Additionally, the project will deliver a reusable pipeline for integrating causal modeling with multimodal healthcare data, promoting broader applicability and reproducibility in real-world clinical settings.

VII. Research Schedule

The following table outlines the proposed timeline for the research activities over a period of 24 months:

TABLE I
Research schedule over 24 months

Activity	M1-M3	M4-M6	M7-M9	M10-M12	M13-M18	M19-M24
Literature Review and Problem Definition	✓	✓				
Data Collection and Preprocessing		✓	✓			
Model Development			✓	✓		
Model Validation and Testing				✓	✓	
Policy Recommendation and Impact Analysis					✓	✓
Thesis Writing and Submission						✓

VIII. Budget

To successfully carry out this research, I propose a well-considered budget that reflects the core needs of the project while maintaining efficiency and responsible use of resources. First, a portion of the budget will be allocated to obtaining high-quality, anonymized EHR data from Moroccan healthcare institutions, particularly in collaboration with the Faculty of Medical Sciences and UM6P Hospitals. This includes potential expenses related to data cleaning, standardization, and integration to ensure compatibility with existing models and tools.

Given the computational demands of training causal inference models, I will need access to robust infrastructure—such as GPU-enabled servers or cloud computing platforms. I plan to allocate resources for this to ensure efficient experimentation and model development.

I also believe it is essential to support the human capital behind the project. The budget will therefore include stipends for graduate researchers involved in the study, as well as partial funding for a research assistant with expertise in data engineering, whose contributions will be crucial for managing and processing the multimodal data.

To ensure that the research has international visibility and benefits from external insights, I am proposing funds for travel to relevant conferences. Presenting the work and engaging with global experts will strengthen the project and enhance its impact. Additionally, I will reserve funds for open-access publication fees, so that the findings are freely accessible to both the global research community and local Moroccan stakeholders.

IX. Limitations

While this research aims to develop robust and interpretable causal inference models, several limitations should be acknowledged. Firstly, the availability and quality of electronic health record (EHR) data from Moroccan healthcare institutions may be variable, potentially affecting model performance and generalizability. Secondly, causal inference methods rely on assumptions such as no unmeasured confounding, which might not fully hold in real-world datasets. Additionally, computational resource constraints may limit the complexity or scale of models that can be practically trained and validated within the project timeline. Finally, although efforts will be made to ensure ethical use of data and model fairness, some biases inherent in historical healthcare data may persist, impacting the applicability of results to underrepresented patient groups. Addressing these limitations will be an ongoing part of the research process.

X. Conclusion

This project aims to bridge the gap between advanced AI techniques and practical clinical needs in managing chronic diseases. By focusing on causal inference and ethical AI, we aspire to produce not only accurate but also trustworthy and population-representative treatment recommendations that can inform real-world clinical decisions, particularly in underrepresented healthcare systems like that of Morocco.

References

[1] Sören R Künzel, Jasjeet S Sekhon, Peter J Bickel, and Bin Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*, 116(10):4156–4165, 2019.

[2] Judea Pearl. *Causality*. Cambridge university press, 2009.

[3] Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018.