Maximizing the Potential of EHR Data: Forging New Paths to Predict Postoperative Complications

Olivia Rennie[1], William M. K. Pitman[1], Alexander Fricke[1], Finn O'Hara[1], Raqib Hayder[1], Abdallah El-Falou[1], Nour Helwa[1]

1. FluidAI Medical (Formerly NERv Technology Inc.), Kitchener, Ontario, Canada

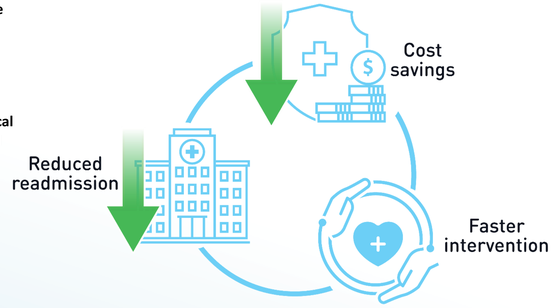
Background

Electronic health record (EHR) data presents an **unparalleled opportunity to enhance healthcare delivery**, including the optimization of postoperative patient outcomes.

While EHR systems have been widely adopted, they remain largely underutilized for **advanced clinical decision-making.**

This project aims to demonstrate how EHR data, when harnessed effectively, can **serve as a critical tool for predicting and mitigating the impact of postoperative complications in gastrointestinal surgery through machine learning (ML) approaches.**

Through a comprehensive literature review and initial development of ML models utilizing EHR data, we outline key steps and best practices for leveraging artificial intelligence (AI) and EHRs to optimize postoperative management.



Methods

A literature review was conducted by examining databases including **PubMed, Web of Science, OVID Embase, Google Scholar, and Cochrane library databases.** Searches were performed with the initial keywords: electronic health record/EHR, postoperative complications, gastrointestinal surgery, machine learning/ML, artificial intelligence/AI.

Publications that were retrieved underwent further assessment to ensure other relevant publications were identified and included. The best practices for creating ML models outlined in the papers reviewed were abstracted and defined to allow for the development of predictive models within healthcare data.

Conclusion

Despite the clear advantages of using EHR data for predictive modelling, the current healthcare landscape has yet to fully embrace this resource for machine learning applications.

By illustrating best practices for leveraging EHR data—from cohort identification to feature selection and model construction—this research emphasizes that EHRs are not just passive records but are an active, underexploited resource.

Unlocking the full potential of EHRs can lead to a revolution in precision medicine, particularly in the domain of postoperative care, enabling the prediction, prevention, and early detection of complications.

Scan our QR code for references

Results

While there has been growing research into integrating ML/AI with EHR data for improving gastrointestinal surgery outcomes**, significant gaps remain.** EHRs are still underutilized as a powerful resource.

Our findings highlight seven critical steps for leveraging EHR data in ML model development: (1) Defining the research question; (2) Data acquisition; (3) Data preparation; (4) Defining cohorts/labels; (5) Defining features (variables) of interest; (6) Model training, testing, and parameter tuning; and (7) External Validation and Deployment.

Defining the Research Question

Clearly articulate the clinical or research question to guide data use and model development.

Data Acquisition

EHR data varies in quality, scope, cost, and usability (with a need for prepocessing). Ideally, EHR data that contains both structured and unstructured de-identified health data and allows you to address the research question must be obtained and utilized.

Data Preparation

EHR data is often complex and may require significant transformation before use. When integrating data from multiple sources, harmonization is necessary. Addressing missing data and correcting errors are essential steps in preparing the dataset for analysis.

Defining Cohorts/Labels

This can be done using various methods, including using medical nomenclature only (ICD codes, CPT codes, etc.), employing a sequence of medical codes and account for clinical workflows, or utilizing text mining and natural language processing (NLP) methods to determine the patient's surgical procedure (cohort) and outcome (label).

Defining Features (variables) of Interest

Begin by identifying potential features from the literature. Use statistical tools like feature importance plots, partial dependence plots (PDP), and Friedman's H-statistic to evaluate the relationship between predictor variables and outcomes. Various feature selection methods can help prioritize the most relevant variables.

Model Training, Testing, and Parameter Tuning

The ML problem should be clearly defined to guide model selection. The model should be trained on a subset of data and then validated on an independent test set. Considerations such as model interpretability, fairness, overfitting, and real-world applicability are vital. Parameter tuning is necessary to optimize model performance.

External Validation and Deployment

Once the model is fine-tuned, it must be validated through techniques such as data splitting, testing on external datasets, or prospective deployment. Model deployment in clinical settings requires thorough evaluation to ensure it functions effectively in real-world conditions.