2024

A Machine Learning Health System to Integrate Care for Substance Misuse and HIV Treatment and Prevention among Hospitalized Patients

FINAL PROGRESS REPORT

Multiple Principal Investigators: Dr. Philip Held*, Dr. Hale Thompson

Team Members: Brihat Sharma, Mennefer Blue, Uddyalok Banerjee, Philip Held, Hale Thompson, Majid Afshar, Connor McCluskey, Ketzel Feasley, Aniruddha Hazra, Dale Smith

Organization: Rush University Medical Center

Inclusive Dates of the Project: 09/01/2021 – 08/31/2024

Federal Project Officer: Sheena Patel

Acknowledgement of Agency Support: Supported by Agency for Healthcare Research and Quality (AHRQ)

Grant Number: R21 HS28511

Abstract

Purpose:

Develop a machine learning (ML) classifier leveraging natural language processing (NLP) techniques to identify high-risk HIV acquisition or transmission amongst hospitalized patients.

Scope:

This study focused on patients hospitalized at Rush University Medical Center (RUSH) as a reachable moment for HIV prevention, particularly those with substance misuse and/or high-risk sexual behaviors.

Methods:

We trained multiple ML classifiers on electronic health records (EHR). Our best model was a binary convolutional neural network (CNN)-based classifier trained on clinical notes obtained from 1507 unique patient's hospital encounters. The dataset was manually annotated at the MRN level for HIV risk factors and then stratified by four levels of risk. To develop our models, we utilized a platform that processed unstructured data of all encounters into concept unique identifiers (CUIs). CUIs served as inputs to ML algorithms for identifying patients, within the first 24 hours of admission, at high risk for acquiring or transmitting HIV. Model predictions were validated against retrospective encounter data validated by a team of experts.

Results:

For previously unseen labeled data (n=383) our best performing model achieved an ROC-AUC of 0.6846 and PR-AUC of 0.5158. Class 0 (low-risk HIV) demonstrated significantly better precision, recall, and F1 scores compared to Class 1 (high-risk HIV) patients.

Conclusion:

Current charting practices of unplanned admissions are insufficient for modeling highly accurate ML classifiers of HIV acquisition or transmission risk. More research is needed to develop and implement HIV treatment interventions for acute care settings.

Key Words:

Natural language processing, convolutional neural networks, substance misuse

Purpose

Objective 1: Develop, train, and test an ML classifier to identify risk for HIV acquisition or transmission among hospitalized patients with substance misuse.

Machine learning provides a data-driven solution to these challenges by leveraging advanced models, including convolutional neural networks (CNNs), to analyze complex patient data. CNNs, specifically, excel at detecting patterns and relationships within data, enabling them to uncover indirect evidence of HIV risk from clinical documentation. Natural language processing (NLP) allows the analysis of unstructured clinical notes housed within Electronic Health Records (EHRs)- an information rich but underutilized source of behavioral and social health data. A term-frequency, inverse document-frequency (tf-idf) transformation can place weights on text features of the training dataset and normalize these values for the ML classifiers to then perform numerical assessments.

Our objective was to standardize medical terms into UMLS concept unique identifiers (CUIs) and furthermore analyze their relationships to capture nuanced patterns that structured data models may overlook.

Objective 2: Integrate the ML classifier in the EHR infrastructure to test predictive validity and test classifier efficacy.

We utilized our best performing classifier to make predictions on previously unseen data labeled via manual chart reviews by a team of expert panelists. Traditional ML metrics such ROC-AUC, Precision, Recall, PR-AUC were used to confirm that the developed classifier performed acceptably.

Objective 3: Improve HIV Prevention Framework

The primary objective of this study was to address limitations in existing HIV prevention strategies, particularly in acute care settings. While pre-exposure prophylaxis (PrEP) has a proven track record of reducing HIV incidence, its uptake amongst injection drug-using populations has been low. Patients hospitalized with substance use disorders within one of Chicago's largest health systems (Rush University Medical Center) present a unique opportunity for intervention. Current screening methods overlook high-risk groups such as patients residing within high-risk zip codes, neglect gender stratifications, underemphasize history of injection drug use and can often omit individuals with high-risk heterosexual behaviors. The integration of ML classifiers and utilizing data to make decisions addresses this gap by providing a systemic and unbiased approach to identifying at-risk individuals. By identifying patients who might otherwise be overlooked, this study fills a critical void in the current HIV prevention framework.

Objective 4: Address Critical Gaps in Risk Assessment

Substance misuse is a major factor in HIV acquisition and transmission; however, existing risk assessment tools often undermine the threat it poses. Individuals with SUDs are at increased risk due to behaviors such as needle-sharing, condomless sex, which are exacerbated by the challenges such as addiction, limited access to care and social stigma. This project acknowledges that hospitalized patients with SUDs represent a vulnerable population that traditional assessments fail to adequately address. By developing an interoperable ML classifier, the study aims to integrate HIV risk assessment seamlessly into electronic health record (EHR) systems. This innovation not only improves the accuracy of risk stratification but also ensures that care providers are equipped with actionable insights, enabling them to deliver timely, targeted interventions.

Scope

Background and Context:

At the crossroads of substance misuse and HIV exists a critical public health challenge, particularly in light of the opioid epidemic that has been ongoing since the 1990s. This epidemic has not only contributed to a dramatic rise in overdose deaths but has also exacerbated HIV transmission risks among vulnerable populations, including people who inject drugs (PWID). The 2015 HIV outbreak in Indiana and subsequent clusters across the United States underscore the urgent need for integrated prevention and care strategies. Furthermore, the COVID-19 pandemic has compounded these challenges by increasing risk-taking behaviors and disrupting follow-up care for individuals undergoing HIV treatment.

Substance misuse significantly increases the risk of HIV acquisition and transmission. This occurs through behaviors such as needle-sharing and unprotected sex, as well as through broader systemic issues like stigma and limited access to care. While pre-exposure prophylaxis (PrEP) has been shown to effectively reduce HIV incidence, its uptake remains insufficient among key populations, including those with substance use disorders (SUDs). Hospitalized patients with SUDs experience higher rates of hospitalization and longer stays compared to the general population. These hospitalizations present a crucial "reachable moment" for initiating HIV prevention and treatment, particularly for non-treatment seeking individuals. However, existing HIV risk assessments often overlook substance misuse as a significant factor, additionally manual screening methods are often impractical in busy acute care settings.

Participants and Settings:

The source cohort was derived from Rush University Medical Center's (RUMC) EHR data repository of unplanned hospital encounters, specifically from January 1, 2017, through December 31, 2019 (N=120,628), with inclusion criteria for patients aged 15 to 89 at the time of admission. To enrich our sample for annotation from the source cohort, we employed topic modeling, a form of unsupervised machine learning. This technique synthesizes voluminous textual data, such as electronic health record (EHR) clinical notes, into more manageable concepts and organizes them into topics based on clustered patterns of concepts. Our experiment utilized topic modeling on the corpus of EHR clinical notes from admissions. We searched for Concept Unique Identifiers (CUIs) related to HIV transmission or acquisition, identified by our team, which included infectious disease and addiction medicine physicians. Additionally, we identified encounters devoid of any CUIs related to HIV transmission or acquisition. This process yielded a cohort of 2,932 unique patients, half of whom had encounters with documented HIV risk-related information, while the other half comprised encounters lacking such documentation. To ensure objectivity during the annotation process, the two batches were randomly mixed. The final training dataset was created by annotating 1,507 unique patients who had at least one admission during the experience period.

Incidence and Prevalence:

Approximately 6% of the patients with SUDs at Rush exhibited documented HIV risk factors, such as sexually transmitted infections (STIs) associated with HIV. However, this figure likely underrepresents the true prevalence due to limitations associated with manual chart review methods. Nationally, the

HIV prevalence among PWID is estimated to be 17%, compared to 0.34% within the general population, highlighting the disproportionate burden faced by this cohort.

Methods

Study Design:

For Phase 1 of this study, a multidisciplinary panel of physician researchers conducted an iterative review to develop a ground truth annotation tool, based on federal guidelines and clinical expertise, for HIV screening and Pre-Exposure Prophylaxis (PrEP) eligibility. This tool stratified patients with unplanned admissions by their potential risk for HIV transmission or acquisition.

Given the lack of a validated HIV-risk assessment tool for identifying substance misusers at risk for HIV, we used current CDC guidelines to build a reference dataset of non-cases and cases of HIV acquisition or transmission risk. This dataset was validated via manual chart review. Cases included adults who were HIV-positive or met one of several risk criteria, such as injection drug use, shared drug equipment, enrollment in methadone treatment, sexual partner HIV-positive, unprotected anal or vaginal sex, multiple concurrent sexual partners, pregnancy, or bacterial/viral STIs related to HIV acquisition.

Data Sources/Collection:

The training dataset was sourced from unplanned admissions to Rush University Medical Center (RUSH) and Rush Oak Park Hospital between 2017 and 2019, encompassing 89,720 unique patient records. Topic modeling on the clinical notes of this cohort revealed that approximately 10% contained concepts related to HIV transmission or acquisition, such as injection drug use or sexual risk behaviors and related comorbidities. From the source cohort, we extracted 2,932 unique patient records, including 1,466 records from patients potentially at risk for HIV acquisition or transmission (identified through the presence of HIV-related concepts in their clinical notes) and 1,466 records with no such concepts. These records were randomly mixed and assigned in batches of 100 to a team of five annotators. The annotation team reviewed and labeled the EHRs of 1,507 unduplicated patients, holding weekly meetings with the PI and our consulting Infectious Disease/HIV/Addiction Medicine physician (Aniruddha Hazra, MD) to discuss and resolve any unique and difficult cases. Regular interrater agreement checks were conducted with a Kappa threshold set at 80%.

Natural Language Processing (NLP) pipeline:

Using a Fast Healthcare Interoperability Resources (FHIR)-based Apache NiFi pipeline, we performed linguistic processing of clinical notes and structured data. All clinical documents were processed through the open-source Apache clinical Text Analysis and Knowledge Extraction System (cTAKES), which recognizes medical terms and understands relations between identified concepts. This NLP method standardizes terms into Concept Unique Identifiers (CUIs) from the Unified Medical Language System (UMLS) Metathesaurus. The CUI features from all processed clinical notes served as inputs to machine learning algorithms for identifying patients at high risk for acquiring or transmitting HIV.

Machine Learning (ML) Model Development:

Multiple ML models such as Decision Trees, Random Forests, Support Vector Machines and CNN-based models were developed and tested for efficacy of identifying vulnerability to HIV acquisition or transmission. Our best performing natural language processing (NLP) model, a binary convolutional neural network (CNN) classifier included embedding, convolutional, and regularization layers to capture underlying patterns in the data and reduce the impact of confounding variables by representing them in a lower-dimensional space and encouraging the model to learn simpler patterns that are less likely to be influenced by noise or irrelevant variables.

For Phase 2, our best performing convolutional neural network (CNN)-based binary classifier from phase 1 was applied to encounter data from approximately 106,000 patients seen at RUSH between January 1, 2020, and December 31, 2022. The classifier made predictions based on clinical documentation created within the first twenty-four hours of admission to the acute care setting. Patients with a probability above 25% were flagged as high-risk for potential HIV acquisition or transmission - 25% threshold was chosen due to the low prevalence of clear EHR documentation of HIV vulnerability during acute hospital care, ensuring that providers err on the side of caution by screening those flagged as high-risk.

Measures:

Below are some of the metrics used to evaluate the developed classifier:

- ROC-AUC (Receiver Operating Characteristic – Area Under the Curve): This metric measures the ability of a classifier to distinguish between positive (high HIV risk) and negative classes (low to medium HIV risk). An ROC-AUC value closer to 1 indicates better performance.
- Precision: The proportion of true positive predictions among all positive predictions made by the model.
- Recall: The proportion of true positive predictions among all actual positive instances in the dataset.
- F1 score: harmonic mean of precision and recall, providing a single metric that balances both measures – useful when class distribution is imbalanced.
- Macro Average: The average of precision or recall across all classes, giving each class equal weight.
- Weighted Average: The average of precision or recall across all classes, with each class weighted by its support (the number of true instances).
- PR-AUC (Precision-Recall Area Under the Curve): Unlike ROC-AUC, PR-AUC focuses on the precision and recall of the classifier and measures the trade-off between precision and recall. Higher values signify better performance.

Limitations:

Sample imbalance and representation: The classifier may not adequately predict underrepresented high-risk populations due to sample imbalances. During manual chart reviews or bias assessments, if imbalances or misclassifications were identified, we used oversampling for the "minority class" and under-sampled for the "majority class" to enhance prediction accuracy. For the NLP model, we weighted CUIs linked to underrepresented patients or removed CUIs associated with biased classifications. Nevertheless, learning curves were generated to investigate the effect of sample size on model performance as an approach to assess adequacy of statistical power.

Manual annotation variability: Variability in manual annotations could introduce inconsistencies in the ground truth data. We conducted regular interrater reliability checks and provided thorough training for annotators to reduce variability. Additionally, we implemented a double-blind review process for contentious cases and utilized consensus meetings to enhance annotation consistency.

Ethical and Privacy Concerns: Implementing an ML-based HIV risk prediction tool raises ethical and privacy concerns. We ensured compliance with HIPAA and other relevant regulations. Additionally, RUSH uses robust data encryption and has adopted a secure data storage methodology. In the future, conducting ethical reviews and establishing guidelines for the ethical use of the tool will also be critical in addressing these concerns.

Data Quality and Completeness: The accuracy of the classifier relies on the quality and completeness of the EHR data. Implementing stringent data preprocessing steps to clean and standardize the data can mitigate this issue. Encouraging consistent, timely, and thorough documentation practices among healthcare providers is also crucial. Periodic audits of the data quality helped identify and rectify inconsistencies or gaps.

Temporal Changes in Clinical Practice: Changes in clinical practices and documentation standards over time may affect the classifier's performance. Establishing a feedback loop with clinicians to monitor and adjust the model based on real-world performance will ensure that the model remains relevant and accurate.

Threshold Selection: The chosen 25% threshold for identifying high-risk patients may not be optimal for all settings. Conducting sensitivity and specificity analyses to determine the optimal threshold for different populations and settings can improve the model's performance. Adjusting the threshold based on feedback and outcomes from clinical implementation can also help tailor it to specific needs.

NLP Limitations: NLP models, for example based on CUIs, may not capture all nuances in clinical language. Continuously improving the NLP model by incorporating advanced techniques such as context-aware embeddings (e.g., BERT) and other large language models can enhance its understanding of clinical language. Collaborating with domain experts to refine the model and validate its outputs can ensure it captures the relevant clinical nuances.

Results

Performance Metrics:

The machine learning classifier, trained to identify HIV acquisition and transmission risks, used data from the first 24 hours of patient admission. The classifier's performance was evaluated on a holdout dataset (n=262) and a separate, previously unseen labeled dataset (n=383).

Holdout Dataset Results – Holdout dataset is a subset of the available data that is withheld from the training process and used solely for evaluating the performance of the trained model. It helps assess how well the model generalizes to new, unseen data.

For the holdout dataset (n=262), our model achieved an ROC-AUC of 0.70 and PR-AUC of 0.55. Class 0 exhibited precision, recall, and F1 score of 0.7688, 0.76, and 0.7644 respectively, while Class 1 showed corresponding values of 0.5281, 0.5402, and 0.5341. The macro averages of precision and recall were 0.6485 and 0.6501 respectively, while the weighted averages were 0.6870 and 0.6853.

Results on previously unseen labeled data (n=383), revealed an ROC-AUC of 0.6846 and PR-AUC of 0.5158. Class 0 demonstrated precision, recall, and F1 score of 0.8924, 0.8185, and 0.8538, while Class 1 showed 0.4000, 0.5507, and 0.4634 respectively. The macro averages of precision and recall were 0.6462 and 0.6846 respectively, while the weighted averages were 0.8037 and 0.7703.

Validation process showed inter-rater reliability exceeding 90%, with 18% of encounters identified as high-risk. The HIV classifier successfully predicted 55% of these cases (sensitivity), and 82% of encounters labeled as low to medium risk (specificity).

Key Findings:

1. Critical opportunity for intervention:
   Acute care settings were reaffirmed as critical spaces for prioritizing HIV prevention. High-risk populations such as people who inject drugs (PWID), heterosexual and sexually active Black women, formerly incarcerated persons, transgender individuals, sex workers, and individuals exposed to violence can particularly benefit from targeted screening, education, and treatment referral during hospital stays.
2. Model strengths and limitations:
   The classifier demonstrated strong performance in identifying low-risk HIV cases, as reflected in the high precision, recall, and F1 scores for Class 0 (low-risk HIV). However, its ability to identify high-risk cases (Class 1) remains limited. This limitation underscores the need to improve sensitivity for high-risk populations, as early identification is crucial for timely intervention.

Conclusions:

The study demonstrated the potential of leveraging natural language processing (NLP) to identify HIV risk factors using unstructured clinical data. While the classifier proved effective in identifying low-risk cases, it highlighted the challenges of detecting high-risk encounters in hospital settings. Future enhancements, such as acute care patient self-assessments of HIV risk, improvements in unstructured EHR data capture and clinical charting, developing ensemble models and expanding the dataset, will further optimize its utility for clinical workflows and ending the HIV epidemic.

Significance and Broader Impact:

This classifier represents a significant step toward proactive HIV risk identification and intervention. By integrating the model into hospital electronic health record (EHR) systems, it could facilitate opt-in HIV testing and linkage to PrEP treatment and HIV care during emergency department (ED) admissions and hospitalizations. Though EHR and ML enhancements are needed, this strategy aligns with public health goals to increase early case detection and reduce HIV transmission rates in high-prevalence communities.

In the long term, scaling the model across multiple healthcare settings, such as neighboring hospitals and community clinics, will maximize its reach. By addressing gaps in HIV risk assessment and care delivery, this approach has the potential to transform urban healthcare systems, improve patient outcomes, and contribute to broader public health efforts to curb HIV transmission.

List of Publications and Products

Posters:

1. Thompson, HM, Sharma B, Feasley K, Blue M, McCluskey C, Hazra A, Afshar M. A machine learning health system to integrate care for substance misuse and HIV risk." Annual American Public Health Association Meeting 2020 (virtual).
2. Blue M, Sharma B, Banerjee U, McCluskey C, Feasley K, Held P, Thompson, H. Development of a machine learning classifier to identify HIV acquisition and transmission risk among hospitalized patients with substance misuse. Presented at Rush University Medical Center, Chicago, IL, and Howard Brown, Department of Social and Behavioral Health, Chicago, IL. 2023.
3. Becker H, Banerjee U, Chaudry S, Guess A, Blue M, Held P, Thompson H. Validation of a machine learning classifier to identify vulnerability to HIV acquisition or transmission among hospitalized patients at an academic health center. Presented at: Rush Research Day; 2024; Rush University Medical Center, Chicago, IL.

Manuscripts in process:

1. Thompson HM, Sharma B, Banerjee U, Hazra A, Ridgway J, Stanford K, Lin A, Boley R, Feasley K, McCluskey C, Blue M, Held P, Karnik N, Afshar M. Can Supervised Machine Learning Models Screen Patients in Acute Care Settings for Vulnerability to HIV Acquisition and Transmission? Developing a high-quality training dataset without a gold standard.
2. Banerjee U, Becker, HN, Sharma B, Blue M, Thompson HM, Held, P. Machine Learning Classifiers to Enhance HIV Risk Identification in Hospitalized Patients at Urban Acute Care Settings

Computers used:

1. GPU Machines powered via Azure Databricks. ML cluster with specifications: Databricks runtime version 13.3 LTS ML with GPU support, Scala 2.12.15 and Spark 3.4.1. The cluster consisted of 2-14 workers with 448 GB memory and 4 GPUs, accompanied by a driver with similar configurations, facilitating efficient processing and training of the NLP classifier.
2. NVIDIA DGX multi-threaded parallelizable linux machine to develop the model and perform computations, 20 cores/80 processors (Red Hat Enterprise Linux Server release 7.6 – Maipo).
3. Source code of binary-CNN classifier and label training dataset available upon request.