Using Electronic Health Records Data to Identify Incarcerated Persons at Increased Risk for HIV Acquisition

Session: 1213

Presenter: Alex Treacher, PhD

Continuum 2024 • June 9-11, 2024 • Puerto Rico
Team and collaborators

Development and successful deployment of a high-performance machine learning models requires an interdisciplinary team to maximize both performance, and clinical impact of a predictive model.

Ank Nijhawan, M.D.
Timothy Hogan, PhD
Robin Higashi, PhD
Song Zhang, PhD
Brynn Torres, MPH

Douglas Krakower, M.D.
Julia Marcus, PhD, MPH

Jackie Naeem, M.D.
Alex Treacher, PhD
Reshma Suresh M.A/M.S

Kenneth Dobbs

NIH Funding: R01 MH 129185
Incarcerated persons experience disproportionately high rates of HIV acquisition.

Development of tools to help identify incarcerated individuals with elevated HIV risk can optimize allocation of resources for testing, counseling, and prevention.

We developed a predictive risk model to estimate individuals' future HIV risk using electronic health records from the 8th largest US jail.
Final dataset:

- **Demographic**
- **Historical**
- **Medical data**

Includes data up until 3 days after start of latest incarceration, or incarceration prior to HIV+ diagnosis (when applicable).
**Modeling Approach**

**Goal:** Create an optimized classification model to predict patients with a future HIV+ diagnosis, and use the prediction probability to estimate risk.

- **End to end hyperparameter optimization**
  - Optimized parameters for processing and model development using Bayesian optimization (TPE)

- **Data split:** Stratified Multi-holdout 10-fold cross validation

- **ML Processing:**
  - Normalization
  - KNN imputation
  - Augmentation

- **Model development:**
  - Model selection: Logistic regression, Random forest, Support vector machine/classifier
  - Model parameters e.g. regularization

- **Model analysis:**
  - Final model performance
  - Model calibration
  - Feature importance (PFI)
  - Risk stratification
Results: Model Performance

<table>
<thead>
<tr>
<th>Test Performance</th>
<th>AUROC</th>
<th>.702</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bal Acc</td>
<td>.655</td>
</tr>
<tr>
<td></td>
<td>Sens</td>
<td>.687</td>
</tr>
<tr>
<td></td>
<td>Spec</td>
<td>.624</td>
</tr>
</tbody>
</table>

Test Performance:

- AUROC: 0.702
- Bal Acc: 0.655
- Sens: 0.687
- Spec: 0.624

<table>
<thead>
<tr>
<th>% Pop. High Risk</th>
<th>Bal Acc</th>
<th>Sens</th>
<th>Spec</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>.580</td>
<td>.208</td>
<td>.952</td>
</tr>
<tr>
<td>20</td>
<td>.672</td>
<td>.538</td>
<td>.805</td>
</tr>
<tr>
<td>50</td>
<td>.668</td>
<td>.831</td>
<td>.505</td>
</tr>
</tbody>
</table>

Graph showing density over predprob for True Label and HIV Pos vs HIV Neg.
Results: HPO and Feature Importance

Hyperparameter Optimization (HPO)

Feature Importance (PFI)

<table>
<thead>
<tr>
<th>Feature</th>
<th>PFI (Reduction in Bal. Acc.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.06</td>
</tr>
<tr>
<td>Hepatitis C Tested</td>
<td>0.04</td>
</tr>
<tr>
<td>STI Testing</td>
<td>0.02</td>
</tr>
<tr>
<td>Sex</td>
<td>0.01</td>
</tr>
<tr>
<td>Depression history</td>
<td>0.00</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>0.00</td>
</tr>
<tr>
<td>ETHNICITY</td>
<td>0.00</td>
</tr>
<tr>
<td>Substance Use</td>
<td>0.00</td>
</tr>
<tr>
<td>12M Days in Jail</td>
<td>0.00</td>
</tr>
<tr>
<td>Marital Status</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Balanced Acc. Lower 95% CI

Model count (ordered by performance)
Future Work

- Evaluation of performance on newly available data
- Develop/build deployment pipelines
- Prospective implementation to evaluate performance in identifying people entering the jail who need HIV and STI testing and may be potential PrEP candidates
- Integrate features available from Parkland Health EHR
Summary

• Created HIV dataset/cohort from the 8th largest jail in the US.
• To our knowledge, this is the first HIV prediction model developed for an incarcerated population.
• Predictive performance with good calibration is likely in a range that can improve efficiency for HIV prevention resources in jails.
• Limitations of the performance are likely due to limited information from jail EHR.
• Given the large population of individuals at risk for HIV who pass through US jails, the potential population-level impact of a jail HIV prediction model is substantial and warrants prospective evaluation.