Project Summary

Overview

In this document the PIs describe an approach to identify trends in disease treatment by applying existing and novel machine learning techniques. Data sources will include both direct clinical sources and also pertinent text and metadata acquired from users of social media. High level information, analytical tools, and data resources produced through these analyses will be disseminated back into both clinical and social media venues to improve patient's direct medical outcomes and address social concerns. One specific application will focus on the use of antiretroviral therapy to preventatively treat individuals who are at-risk for HIV infection. This use of antiretroviral therapy is referd to as Pre-Exposure Prophylaxis (PrEP) and uses the small molecule nucleotide reverse transcriptase inhibitor trade named Truvada.

Intellectual Merit

The information produced through data mining and machine learning will be useful to provide clinicians and public health professionals feedback concerning successes and challenges associated with disease treatments. Pertinent tools and summary datasets will also be produced and made available to researchers. In addition novel applications of machine learning and data science methods will be developed as necessary, and shared with the data science community to assist and inspire related applications in other domain areas.

Broader Impacts

The proposed research will contribute to an increased scientific literacy through collaborations and dissemination of our findings through social media. This research will also lead to improvements of the health of all Americans, with in some cases such as PrEP, special emphasis on underrepresented groups such as the LGBT community. This work also has the potential to engage collaborations and partnerships between the social media industry, medical providers and pharmaceutical companies, and academia as we come together to improve public health.

<div align="center">

Project Description

</div>

1 Mining Pre-Exposure Prophylaxis Trends in Social Media

Propose research on PrEP and what additional steps can be taken and how you will know if you succeed and how this benifits society etc..

2 Arvind's secion (to be decided)

3 Broad application of biosurveillance and social media patient feedback

<div align="center">

Broader Impacts

</div>

References Cited

[1] Roy G. Biv, *Spectral Analysis of Colorized Arcs*, Journal at the End of the Rainbow, 2005.

Biographical Sketch: Patrick Breen

(a) Professional Preparation

Bowdoin College BA Biochemistry and Mathematics, Brunswick Maine, 2013
University of Georgia PhD Bioinformatics, Athens Georgia, (attending)

(b) Appointments

Oak Ridge National Laboratory Fellowship (2016)
President of Bioinformatics Graduate Student Association (2015)
University of Georgia Presidential Graduate Fellowship (2013)

(c) Products

Mining Pre-Exposure Prophylaxis Trends in Social Media. Patrick Breen, Jane Kelly, Timothy Heckman, Shannon Quinn. DSAA2016.
P2Y6 receptor antagonist, MRS2578, inhibits neutrophil activation and aggregated NET formation induced by gout-associated monosodium urate crystals. Payel Sil ... Patrick Breen ...

(d) Synergistic Activities

Interacted with the local scientific community by judging at high school science fair (2014)

Data Management Plan

Multiple filtered streams of twitter data relevant to diseases studied will be acquired using Twitter's public streaming API. This data includes tweets related to Pre Exposure Prophylaxis, Truvada, HIV, and AIDS as well as other infectious diseases, and commonly used prescription medications. In addition to twitter data we will also acquire other social media data such as Reddit or Facebook, either from an open data repository, or in the case of Reddit through the public API. Direct medical data will be acquired from Oak Ridge National Laboratory and other sources and will be used in accordance with institutional and national laws and regulations (including HIPAA) governing appropriate use of medical data.

Data will be analyzed on local researcher's desktop machines and on research server clusters including the Quinn research group cluster, the Georgia Advanced Computing Resource Center, and Oak Ridge's institutional computing resources. Code developed for analyses will be made openly available on github.com under open source licenses.

Most of the primary information acquired through Twitter's API or direct medical information cannot be shared directly due to Twitter's terms of service and federal regulations such as HIPAA. However anonymized summary information can and will be disseminated in the form of research article publications, and perhaps also through blog articles and other non-technical mediums.

Collaborators and Other Affiliations Information

Collaborators and Co-Editors

Shannon Quinn (University of Georgia)
Jane Kelly (Georgia Department of Public Health)
Timothy Heckman (University of Georgia)
    Arvind Ramanathan (Oak Ridge National Laboratory)

Graduate Advisors and Postdoctoral Sponsors

Shannon Quinn (University of Georgia)

Committee Members

Shannon Quinn (University of Georgia)
Jan Mrazek (University of Georgia)
Timothy Heckman (University of Georgia)
Keith Campbell (University of Georgia)