

DLHS Research Snapshot

V.G. Vinod Vydiswaran, PhD, works on leveraging Big Data in health

“Small crumbs of information about our public health is everywhere – in electronic health records, online health forums, social media and in biomedical literature. Connecting the dots between these sources and filling the gaps will help us better understand our health and take steps to improve it.”

The Problem: Text data is abundant in health care but approximately 85% of health data is unstructured and free text. Traditionally, relevant medical knowledge resides in books, journals, and electronic health records authored and consumed by medical professionals. Consumer-facing health information is on the rise where patients interact with doctors through health portals and with each other through a variety of online media. Health-related text data generated by diverse online sources present opportunities to better understand health factors.

The Promise: To be able to assimilate health related insights from multiple sources – electronic health records, social media, and peer-reviewed literature – and make derived inference to help improve public health.

The Project: Developing novel text mining and natural language processing algorithms and tools, then applying them on diverse health data sources to produce actionable health insights. Three themes define the work:

(i) Deriving medical information from Electronic Health Records, (ii) Analyzing health-related social media for actionable insights, and (iii) Text Mining and Big Data approaches for health services research.

Learning Health System in Action: This work resides on the Data to Knowledge arc of [The Learning Cycle](#) and also has an implication in the Performance to Data arc by facilitating change in documentation to make health data more complete and accurate.

High Tech Health: Newer free-text processing approaches can increase understanding of social, environmental, and geolocation-based factors that affect individual health. These can be more effective to overcome factors like the sometimes poor response rates and inaccurate data that can skew traditional patient survey instruments. Algorithms are being developed to detect risky substance use, smoking status, diet, sleep, and mental health status from clinical notes. Further, analyzing food-related tweets and those that indicate sedentary and physical activities may reveal if health is influenced by the neighborhood characteristics of where they are tweeted from.

Vydiswaran is also developing algorithms to identify components of a prescription directive (drug name, strength, dosage, frequency, route) and comparing prescriber with pharmacist-written prescriptions to assist in automated transcription and potentially reducing pharmacy-related errors.

The Bottom Line: Health can be improved by developing novel text mining and natural language processing algorithms and tools then applying them on diverse health data sources to derive actionable health insights.

KEY WORDS

Big Data
Algorithms
Medical Text
Social Media
Trustworthiness
Clinical Notes
NLP

My research interests

