Using Analytics to Help in the Early Identification of At-Risk Diabetes Patients and the Provision of Evidence-based Intervention, Treatment and Remediation Strategies

Tyrone Grandison
Proficiency Laboratory International, Ashland, OR, USA, tgrandison@proficicenylabs.com

Varun Bhagwan
University of California – Santa Cruz, Santa Cruz, CA, USA, vbhagwan@gmail.com

ABSTRACT

Diabetes is one of the top causes of death for adults in Jamaica. It is manageable condition. We propose to take a technology-based approach to help the people of Jamaica and the world detect the disease early and implement evidence-based strategies in its intervention, treatment and remediation. We propose to use the disease predictors, outlined in a diabetes risk score (DRS) model to build as complete, comprehensive and current profile of the patient. Our system ingests predictor data points through a diverse number of techniquesand once a patient profile is completed, analyses and calculates the patient's DRS. For patients "at risk" or close to the threshold, the system flags an intervention event and a cohort of similar peer patients is analyzed for possible positive remediation and treatment pathways.

Keywords: Diabetes, Intervention, Natural Language Processing, Risk Prediction

1. INTRODUCTION

Jamaica is the largest English-speaking nation in the Caribbean. The 2009 Basic Indicators of Health report from the World Health Organization (WHO) states that Jamaica's health service is very fragmented and there is no healthcare information system (HIS) policy and strategic framework. Additionally, general shortage of health care providers on key areas of health service delivery due in large part to a high attrition rate of skilled personnel, a low adoption rate for Electronic Medical Records (EMRs) and Electronic Health Records (EHRs), a propensity for the government to address issues with social and policy mechanisms (Davidson, 2013) and a willingness of the media to not report the severity of health concerns (Chambers, 2010).

In 1999, diabetes, hypertension, cancer and cardiovascular diseases were the causes of the majority of deaths in Jamaicans over the age of 20 (PAHO, 2009). The Pan American Health Organization (PAHO) stated that diabetes accounted for one of every nine deaths. The rate of diabetes among women increased from 51.8 per 100,000 population in 1990 to 59.9 per 100,000 population in 1999; with the expectation of this trend continuing. It was also noted that that was high among women age 20-59.

Given that a vast of majority of the data in the Jamaican healthcare system is still paper-based, i.e. it is unstructured free text written in English, and that the top two mortality agents (i.e. diabetes, hypertension) are detectable and manageable, we propose a technological system that helps in the early detection and treatment of these conditions. Our approach is to create a system that uses open-source Optical Character Recognition (OCR) technology to convert paper records into their digitized form, then develop customized Natural Language Processing (NLP) modules to make sense of the content, store the elements in a persistent, secure and private data store, use this information, along with other available information to develop a patient model, then generate a risk score that determines the likelihood that the patient is at-risk of getting the disease, and finally utilizing a cohort of the patient's peers (with positive treatment outcomes) to help determine a course of action.

In this paper, we start by presenting the system approach (section 2), grounding our discussion in the context of diabetes. In section 3, we review the architecture and workflow of the system. We conclude in section 4.

2. SYSTEM APPROACH

Our system integrates data from structured, unstructured, sensory and mobile sources, whether patient-supplied, provider-supplied or Open Data, in order to create a complete and comprehensive patient profile. We create a longitudinal patient record, and then execute a suite of targeted analytics and information extractors on this data to compute quantifiable and tangible measurements, values and patient-datum. Our approach employs NLP techniques developed specifically for healthcare data to perform these extractions.

We choose the Framingham study as the baseline for the model for our risk score because a significant amount has been done on it that enables it to be used to determine risk for diabetes, hypetension and cardiovascular diseases.

The goal with our analytics is twofold. One, for pre-diabetics or population at large, we wish to use Diabetes Risk Calculators to compute a risk score, including warning patients *on-the-cusp* of being at-risk, to enable them to take adequate precautions and care to help delay the onset and or severity of disease. Two, for diabetics, we would like to ensure detailed and timely monitoring of patient condition, disease progression and trending, and continued care.

3. SYSTEM PROPOSAL

Figure 1 illustrates our system architecture and the flow of data across the various modules of our system. The ingest module supports data fetching using a variety of push-pull based techniques, including patient supplied data, sensory data from medical devices, data ingested via EMR/EHR and legacy systems, as well as data from unstructured sources, progress notes, discharge summaries etc. The unstructured data also incudes free-text data that can be found embedded within EMR systems. Data ingestion is performed on a per-patient basis, such that each data artifact is associated with a patient and used to compile the longitudinal patient record.

Upon extraction of relevant predictors, we apply one or more Diabetes Mellitus risk models via the Diabetes Risk Score (DRS) Calculator module. This module performs two key functions. One, it computes a per-patient risk score, such as the Framingham 8-year Diabetes risk score. Second, the DRS module also determines if a patient is on-the-cusp of being at-risk, and identifies the causal factors that are likely to cause a significant increase in the patients' risk score. This output is fed to the Intervention and Strategy Analysis module, which also interacts with the Cohort Selection & Analysis module to perform aggregate analysis of both at-risk and asymptomatic patients. External modules for Wellness Strategy as well as Clinical guidelines are incorporated to provide appropriate guidance and care management for patients. Finally, this data is exposed via an API and a visualization layer to a multitude of end-users, including Payers, Providers and medical professionals, as well as Patients themselves.

Apart from the added side-effect of seeding EHR and EMT systems with data (and thus boosting electronic healthcare adoption) in Jamaica, our concept provides patients and their providers with 1) a predictive analysis of their diabetes risk, 2) early warning for those being close to at-risk but currently asymptomatic, and 3) intervention, remediation and treatment pathways that can lead to positive outcomes. We enable better decision-making by building a more complete, comprehensive and up-to-date profile of the patient than readily available and by using similar patients with positive trajectories to influence the plan for successfully combating diabetes. We recognize that DRS models will evolve, and thus our design is model-agnostic and extensible. As future work, we desire to translate the DRS into meaningful concepts for patients, e.g. chance of diabetes diagnosis at 50 = 90% (similar to the heart age calculator developed by Unilever).

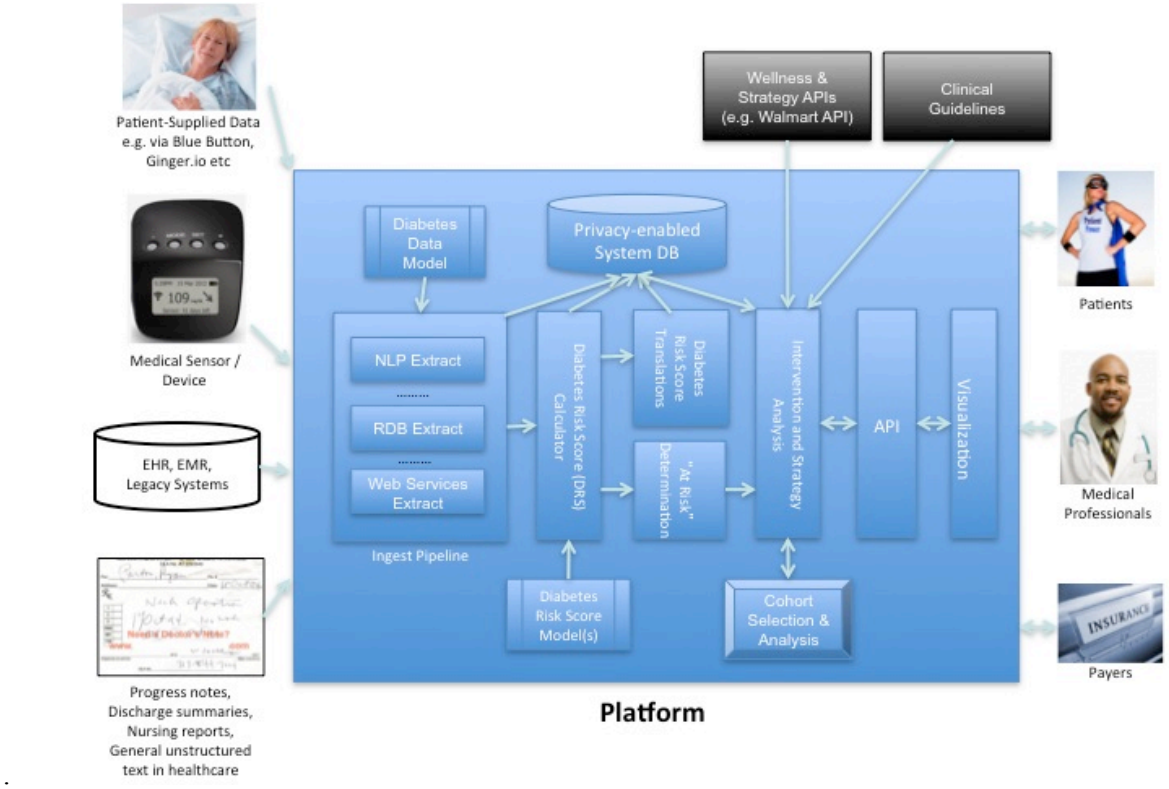

Figure 1: System Architecture

4. CONCLUSION

Jamaica has 1) a problem with its healthcare information systems, 2) a need for new ways to tackle persistent ailments, such as diabetes, that is ravaging its adult population, and 3) a need to find different solutions from those taken in the past. This is further complicated by the fact that a lot of the healthcare data is untapped and locked away in unstructured text. Our concept focuses on early warning by accurate tracking of diabetes risk scores ensures that appropriate preventive measures can be undertaken for patients who are "on the cusp" of being at-risk.

REFERENCES

Chambers, M. (2010). "National Media Coverage of Health Issue in Jamaica: Whose Agenda?" World Journalism Education Congress, South Africa, July 5-7, 2010. Retrieved on May 8, 2013 from http://wjec.ru.ac.za/index.php?option=com_rubberdoc&view=doc&id=117&format=raw.

Davidson, W. (2013). "Review Of Health Care In Jamaica". The Jamaica Gleaner. January 8th, 2013. Retrieved on May 8th, 2013 from http://jamaica-gleaner.com/gleaner/20130108/lead/lead91.html.

Meigs, W., Fox, S., and D'Agostino, N. (2007). "Prediction of Incident Diabetes Mellitus in Middle-aged Adults: The Framingham Offspring Study". Archives of Internal Medicine.

PAHO (Pan American Health Organization). (2009). "Basic Health Indicator". Retrieved on May 8th, 2013 from http://www1.paho.org/english/dd/ais/cp_388.htm.

WHO (World Health Organization). (2009). "Basic Indicators of Health". Retrieved on May 8th, 2013 from http://www.who.int/countryfocus/cooperation_strategy/ccsbrief_jam_en.pdf.

Authorization and Disclaimer

Authors authorize LACCEI to publish the paper in the conference proceedings. Neither LACCEI nor the editors are responsible either for the content or for the implications of what is expressed in the paper.