E-Health: Type 2 Diabetes Management Analytical Engine

Members:	Lama Yassin Kassab (100969142) Leen Yassin Kassab (100969141) Pooria Shafia (100759037) Joe Samuel (100998314) Mohamed Hozayen (100997960)	Supervisor:	Dr. Ibnkahla

Objective

To develop a smart machine learning algorithm which helps Type 2 Diabetes Mellitus patient to manage their disease by giving them real-time recommendation based on factors such as their current glucose level, their sleep data, exercise, stress and diet.

Background

Type 2 Diabetes Mellitus (T2DM) is a metabolic disease that results from impaired insulin secretion or insulin resistance and is characterized by high glucose levels in the blood (hyperglycemia) [1]. The risk factors for T2DM may include but are not limited to: family history, cardiovascular disease and a lack of exercise [2]. Untreated diabetes can result in damage to several organs in the body including the eyes, kidneys, as well as other organs. A recent report published in 2016 by the World Health Organization (WHO) has suggested that an estimate of 422 million adults (8.5% in the adult population) have suffered from diabetes in 2014 as compared to the 108 million adults (4.7% in the adult population) in 1980 [3]. This figure is expected to keep rising over the coming decades.

Due to T2DM's rising trend and its devastating effects on people and on the healthcare, it is imperative that steps be taken for the early diagnosis and management of the disease. The typical treatment options for the disease include either hospitalization to monitor the patient's glucose levels, or out of the hospital smart wearable for glucose and physiological monitoring [4]. Hospitalization is inconvenient for the patient, and the smart wearables can cost around $10,000 USD [4]. Additionally, past research suggests that self-management plays a crucial role in managing diabetes. A systematic review of the efficacy of self-management in T2DM by Norris, Engelgau and Narayan on 72 studies on the subject reveals that self-management of diabetes, individually and in groups, help patients with diabetes in controlling their glucose levels [5].

Therefore, the aim of this project is to provide a practical, inexpensive approach for T2DM monitoring for youth patients to help them self-manage their disease. This goal can be achieved by developing a data analytics engine that gives real-time suggestions to the patients to help them regulate their glucose levels based on their individual situations.

Preliminary survey of the current literature revealed that current research is mostly focused on the diagnosis of T2DM. Among the many articles, our team found two articles that are especially relevant to our project. An article by Chen et al. proposed a 5G-Smart Diabetes system which combines technologies such as wearables, machine learning, and big data to generate comprehensive sensing and analysis for patients suffering from diabetes. Their intelligent solution was able to help with diagnosis and treatment of patients with a cost-effective design [4]. Furthermore, a study reviewing the emerging technologies for the management of T2DM examined the latest accomplishments in sensors for glucose and lifestyle monitoring to facilitate self-management of the disease [6]. The results from the paper suggested that the use of data from the sensor devices is extremely important as the provide a non-invasive method for accurate glucose monitoring and enhancing the lives of people with diabetes.

Our Contribution

Based on the most recent published paper on 5G Smart Diabetes [4], the paper discussed a mechanism and a platform to assist diabetes patients, however, this assistance is based on prevention and treatment. We are considering our potential contribution toward examining algorithms to predict the potential for adverse events such as hypoglycemia or hyperglycemia. In addition, we are incorporating recommendations following predictions to help patients manage the disease.

5G Smart Diabetes' authors have collected and used the same data we are interested in, such as blood sugar, diet, exercise, sleep, and time. They used the following machine learning algorithms: decision tree, support vector

machine, and artificial neural network. Our team shall learn, implement, add, and compare to these algorithms, and investigate advancing these algorithms to output accurate, real-time predictions.

Through developing a quality algorithm that predicts events and helps patients manage their diabetes, we are aiming to potentially publish our work. We could also apply for NCUR, a conference in April in the USA to present our work, should our team reach the goals of the project.

System Architecture

This section provides the proposed high-level system architecture for our design. While the focus of the project will be on developing a machine learning algorithm that classifies the input data and provides recommendations to the user, our team would like to provide an Amazon Alexa interface for inputting the data into our algorithm.

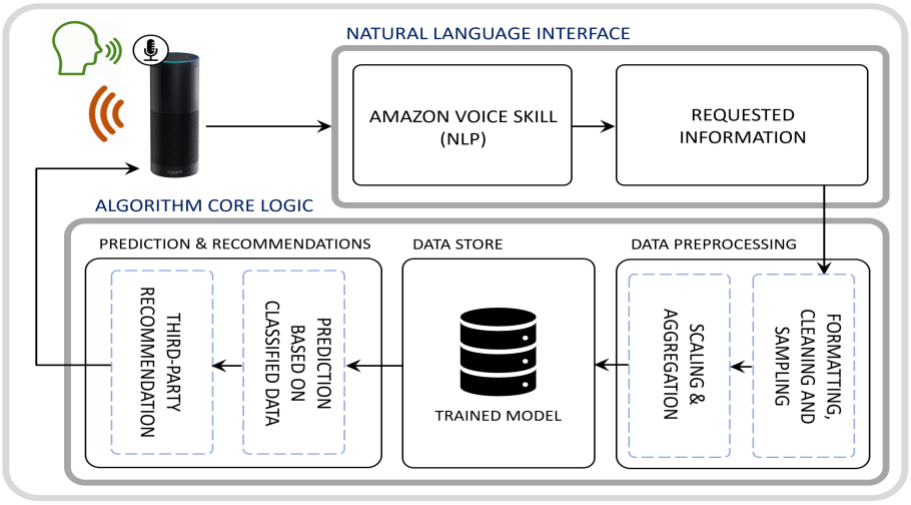

Figure 1: The proposed system architecture for the analytic engine

Input Data and the Expected Format

The following table summarizes the input data formats used to formulate the machine learning algorithm. A part of the preprocessing includes converting the input data into the expected formats.

Table 1: The desired input data with accordance to its format

INPUT DATA	UNIT
SLEEP	Hours
EXERCISE	Kcal burned
DIET	Kcal ingested
GLUCOSE	mmol/L or mg/dL
STRESS	Scale from 0 to 10

Outputs

Based on the classification of each data point, a prediction will be made and using the prediction, an output message will follow. The following output messages are samples and can be in the form of notifications, pop-up messages and Alexa voice messages.

Table 2: Sample outputs for the analytical engine

CATEGORY	SAMPLE SUGGESTIONS
EXERCISE	"How about you go for a walk." "What about a quick run instead!"
STRESS MANAGEMENT	"An hour of yoga could make you feel more relaxed." "How about deep breathing exercises to make you feel better!"
SLEEP	"Let's call it a night and go to sleep to get your target sleep hours!"
DIET	"Today's breakfast could be (information from linked diet apps)" "We have the following menu for lunch: (information from linked diet apps)"

Risk Management

According to the NIST Computer Security Handbook [NIST95], Computer Security is the protection afforded to an automated information system in order to attain applicable objectives of preserving the integrity, availability, and confidentiality of information system resources including hardware, software, firmware, information/data, and telecommunication.

This definition introduces three key objectives: Confidentiality, Integrity, Availability. Usually referred to these three concepts as the CIA triad. Likely, there are cloud services that already provide the CIA triad such as Amazon Web Services, Microsoft Azure, Google Cloud, and IBM Cloud. The team is to select a cloud platform to reside the data analytics engine on.

In this project, computer security plays an important aspect as we are working with patients' data. The system security shall be in accordance to the HIPAA compliance including personal health information (PHI) separation from non-PHI information. The team will take this into account developing the system design and incorporate it with a selected cloud security (AWS is highly to be used).

Milestones

The following table summarizes the major milestones set by the team. Furthermore, Appendix A contains a Gantt Chart showing a bird's eye view of the envisioned milestones for the project. A more detailed task breakdown can be provided upon request.

Milestone	Proposed Due Date	Who's Responsible	Details
Resource Acquisition*	October 5th	Leen, Lama	Involves gathering/generation of datasets required to build machine learning models, acquiring necessary equipment.
Initial Algorithm Development*	October 12th	Joe, Pooria, Mohamed	Formulating the core algorithm logic for the data analytics engine.
Data Analytics Engine Development	December 3rd	Leen, Lama, Joe, Pooria, Mohamed	Developing a prototype for the data analytics engine based on the initial algorithm draft.
Integration of Amazon Alexa with Data Analytics Engine	January 21st	Leen, Lama, Joe, Pooria, Mohamed	Integrating the Data Analytics Engine with Amazon's Alexa voice assistant for voice-based data inputs.
Testing – Verification and Validation	February 18th	Leen, Lama, Joe, Pooria, Mohamed	Performing verification techniques and validating results through a variety of testing techniques.

* - Parallelizable task

Team Members Profiles

Lama Yassin Kassab (100969142) is with the department of Systems and Computer Engineering. Her high level of mathematical ability and knowledge of programming provide her with the necessary tools to improve the Diabetes Coach System by developing an advanced big data analytics tool. Her good knowledge of tools like regression, machine learning, Bayesian statistics and descriptive statistics give her a strong foundation to perform solid analyses. Her previous coop experience as a test engineer which involved conducting product testing prior to shipment and then debugging would help in ensuring that the system works after implementation.

Leen Yassin Kassab (100969141) is with the department of Systems and Computer Engineering, studying Biomedical and Electrical Engineering. Her ability to analyze, model and interpret data as well as her problem-solving skills and methodical and logical approaches make her eligible to improve on the machine learning part of the Diabetes Coach System. Additionally, her 20 months of research experience at both the National Research Council and Health Canada have enabled her to work with big data sets and apply filtering techniques as well as perform statistical analyses on them. This project aims to design a practical solution for patients with Type 2 diabetes through the development of a data analytical engine. Therefore, the project has both a biomedical and a systems component as per degree teachings.

Pooria Shafia (100759037) is with the Department of Systems and Computer Engineering. In the past, Pooria has worked on implementing a virtual reality platform allowing researchers to study human vision and perception, providing him with hands on experience with coding using object-oriented programing languages such as Python. Furthermore, Pooria worked as a part of a team in the Integrated Circuits and Systems Lab on developing a wireless EEG monitor capable of acquiring brain signals, removing motion artifact on-site, digitizing and transmitting the signals to any Bluetooth-connected device providing a platform for detection and classification of seizures using machine learning algorithms. Through his past experience, his passion for Biomedical Engineering and his skills in programming, Pooria is an ideal candidate to work in the E-Health group to develop a machine learning algorithm that helps patients with Type 2 Diabetes self-manage their disease through the My Diabetes Coach system.

Joe Samuel (100998314) is an undergraduate Software engineering student with the Department of Systems and Computer Engineering specializing in business, marketing, and design alongside technology. Joe has also worked on research projects involving building a temporal event annotation framework and has published a conference position paper for IBM CASCON 2018. His skills in user-centric UI/UX design, process management, mobile, and web application development provide him with the skill set required to implement an advanced machine learning based data analytics engine as part of the Diabetes Coach project.

Mohamed Hozayen (100997960) is a fourth-year student in the department of Systems and Computer Engineering. Mohamed have been conducting research in the department of Systems and Computer Engineering for over the past three years. He has presented in several conferences across Canada and the United States, and his outstanding contribution is the development of a real-time system that imports data off a patient monitor which is currently being used at CHEO for experimental purposes. Moreover, He has experience in problem solving and systems computers which provide him with the necessary skills to implement a big data analytics tool that would help diabetes patients with insightful predictions and recommendations. Currently, Mohamed is learning Machine Learning models in theory and in practice by coding. The theory part helps him choosing and manipulating convenient models to fit the data analytics engine. On the other hand, the coding part helps him with implementation.

References

[1] D. Of and D. Mellitus, "Diagnosis and classification of diabetes mellitus," *Diabetes Care*, vol. 37, no. SUPPL.1, pp. 81–90, 2014.

[2] H. Wu, S. Yang, Z. Huang, J. He, and X. Wang, "Type 2 diabetes mellitus prediction model based on data mining," *Informatics Med. Unlocked*, vol. 10, no. December 2017, pp. 100–107, 2018.

[3] World Health Organization, "Global Report on Diabetes," *Isbn*, vol. 978, p. 88, 2016.

[4] M. Chen, J. Yang, J. Zhou, Y. Hao, J. Zhang, and C. H. Youn, "5G-Smart Diabetes: Toward Personalized Diabetes Diagnosis with Healthcare Big Data Clouds," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 16–23, 2018.

[5] S. L. Norris, M. M. Engelgau, and K. M. Narayan, "Effectiveness of self-management training in type 2 diabetes: a systematic review of randomized controlled trials.," *Diabetes Care*, vol. 24, no. 3, pp. 561–587, 2001.

[6] K. Zarkogianni *et al.*, "A Review of Emerging Technologies for the Management of Diabetes Mellitus.," *IEEE Trans. Biomed. Eng.*, vol. PP, no. 99, p. 1, 2015.

Appendix A – Project Timeline Gantt Chart

Bird’s-Eye View