An Intelligent Mobile System to Predict Blood Sugar Level for Gestational Diabetes Patients using Machine Learning

Shiyu Sara Huang[1], Chou in Luk[2], Liping Zhou[3], Yu Sun[2]

[1]Shanghai High School International Division, Shanghai, China 201204
[2]California State Polytechnic University, Pomona, CA 91768
[3]Amazon.com, Irvine, CA 92618

***Abstract*— Gestational diabetes patients have to closely monitor their blood sugar levels four times a day using the traditional finger pricks, which often causes extra pains and inconvenience during the pregnancy. The monitoring approach without using finger pricks has not been widely used due to the accuracy and cost. In this project, we address this problem by using mobile computing and machine learning. A mobile app has been developed to collect the patient's diet and the tested blood sugar level. Once the sufficient amount of data has been collected, the system is able to train the machine learning model and predict the patient's blood sugar level based on the diet. Experiments show that our prediction without finger prick monitoring can reach to 91% accuracy when the patient is under a regular and routine diet and daily activity.**

***Index Terms*— Gestational diabetes, Machine Learning, Mobile Computing.**

I. INTRODUCTION

Gestational diabetes, or gestational diabetes mellitus (GSM), is a symptom where women who do not have diabetes develops high blood sugar levels when having a baby [1]. Gestational diabetes often starts when hormonal changes in the women's body (particularly the placenta) block insulin release, which results in the lack of chemicals in the body that regulates its metabolism and turns sugar into usable energy. Therefore, glucose levels in the blood may increase drastically, leading to the increased rate of gestational diabetes between week 24 and week 28 [4].

Some women may have higher risks of developing gestational diabetes if they are previously overweight, older, have a family history of glucose intolerance or gestational diabetes. As gestational diabetes is the inability for the body to maintain regular homeostasis, women who are overweight have higher probabilities of developing this illness because the extra weight affect insulin's ability to maintain stable blood sugar levels [2]. Gestational diabetes affects 1% of women between the age of 15 and 19, and 13% of women aged between 44 and 49. The overall rate of gestational diabetes is 0.05% of people or 4-7% of women with pregnancy.

Under most circumstances, the presence of gestational diabetes would be first noticed by routine pregnancy screening tests, but in some cases of severe gestational diabetes, the patient may have unusual thirst or hunger, frequent urination, or abnormal levels of fatigue. During pregnancy, the baby might grow overweight due to the abundance of nutrients, or sugar that the mother provides, which could result in the necessity of a C- section (Cesarean Section), which results in longer time for the women to recover from childbirth. High blood pressure (preeclampsia) and low blood sugar (hypoglycemia) in the baby after delivery may result from gestational diabetes, which could be both harmful for the baby and the mother.

Gestational diabetes increases the risk of the mother developing type 2 diabetes, which is a disorder where the body cannot make enough insulin to keep normal metabolism. Other potential side effects of gestational diabetes are babies suffer from yellowish pigmentation of skin, or jaundice due to abnormal breakdown of blood waste products like bilirubin. The baby's liver is unable to remove older red blood cell, leading to the buildup of bilirubin, which may require a blood transfusion due to the poisonous nature of bilirubin.

The most important practice for Gestational diabetes patients is to closely monitor their blood sugar levels during the pregnancy [3]. The traditional approach to test the blood sugar level is to use finger prick to collect a small amount of blood followed by quantifying the sugar level in a test meter. Doctors normally suggested four tests per day - fasting, after breakfast, after lunch, and after dinner. In addition, the patients have to follow a very strict diet plan to keep their blood sugar level low. Gestational diabetes patients should maintain moderate-intensity daily exercise that is at least 30 minutes, including swimming, brisk walking, and other types of exercises. A healthy diet is also encouraged for patients, preferably the prescribed diet for diabetes patients.

Patients with diabetes should avoid consuming too much carbohydrates because they raise the blood sugar level quickly, leading to higher stress on the pancreas, where insulin is produced [5]. Especially, sweets, fruit, milk, yogurt, bread, cereal, rice, pasta, potatoes should be monitored. Sweets and deserts contain large amounts of carbohydrate, which should be substituted or avoided. Some effective methods to control a patient's diet include: eating several small portions rather than large meals, limit milk and fruit consumption, and having breakfast with whole grain cereals. Keeping a food record is highly recommended as people can have a clearer sense of their calorie and carbohydrate intake. Foods with fibers, like vegetables, whole grain bread, brown rice, help control blood sugar, hence increased intake can be beneficial if taken slowly [6,7].

Open Problem: Although the frequent tests give the patients and doctors update-to-date results for the blood sugar level, it causes extra pains and inconvenience for the patients. Many patients start to test their blood sugar level since week 26 during their pregnancy, and they have to do the finger pricking four times a day for the rest of 14 weeks. The test has to be conducted at specific time (exactly one hour after the meal), so it is challenging for the patients to consistently follow the plan and schedule the tests for such a long period of time. In many cases, patients either forget about the test and meal time, or perform the tests with the wrong time or procedure.

Solution => Automated prediction of the blood sugar level using mobile computing and machine learning. In this paper, we present a new approach for patients to know their blood sugar level without finger pricking. Since most of the patients follow a standard diet and routine activities, their blood sugar tent to follow a pattern that is based on their own personalized lifestyle. Summarizing the pattern will allow us to predict their blood sugar levels. Thus, we have developed a mobile application that enables the patients to input their diet, activities, and the normal blood sugar test results. Once the application collects enough data, it trains a model using machine learning algorithms which could be used to predict the blood sugar level based on the given input. The mobile application is supported by a backend running in the cloud that collects data, trains the model, and makes predictions. As the patients input more data with real test results, the accuracy of the tests will improve accordingly.

The rest of the paper is organized as follows: Section 2 details the challenges in this research project; Section 3 presents the solution, followed by showing the experimental results in Section 4; we compare the related works in Section 5, and Section 6 offers the conclusion remarks and future work directions.

II. CHALLENGES

A few challenges exist in training the patient's blood sugar level model and making the accurate predictions.

Challenge 1. Deciding the exact set of factors to include in the training model is difficult. A patient's blood sugar level is affected by a number of factors such as the diet, the specific amount of food, the types of the meal, the time of the day, and the time spend on exercises etc. Getting the right factors may require some statistical analysis and model selection techniques, and a lot of experiments. While too many factors may lead to overly complicated models that have less precise predictions, only using a few factors would lead to overlooking potential lurking variables and confounding variables. Therefore, first listing all potential factors may be helpful, as we can eliminate the lesser factors in the process of training the machine.

Challenge 2. The blood sugar level is very specific to individuals, which means that the prediction has be to personalized based on each patient's situation. The model and the training approach might be adapted to the different individual patients. Through collecting data, we can categorize patients according to their family history or medical history so as to have higher accuracy when predicting the outcome. We plan to provide more options for users to select as they input their data, subsequently they can see their predictions using people in similar circumstances as reference. As the system continues to take in more data, each level of severity and background can be categorized to provide higher confidence in the predicted outcome and lower the range of the predictions.

Challenge 3. The model training needs to be updated frequently with the new data. During the pregnancy, a patient's physical situation changes from time to time. The system needs to provide a way to easily collect new data and keeps improving the training model by itself using a technique known as reinforcement learning [8]. There may be significant outliers as each individual has vastly different conditions; therefore, the system's ability to react immediately to new data and decipher the significance is crucial to its development. We intend to give users an easy and efficient way of utilizing the app in order for more people to be willing to spend time in putting their data.

III. SOLUTION

A. Overview of the Solution

An overview of the system is presented in Figure 1. Patients use the mobile client to input their daily diet, exercise time, and activity information, followed by manually testing the blood sugar value. These data build the training dataset, so each time a new record is inputted by a patient, it triggers the training process and build the updated prediction model. Once the dataset grows over a certain threshold, patients can start to input the information and get the predicted blood sugar value.

B. Machine Learning Model and Feature Selection

Table 1 shows the training features selected in this machine learning process, which are described as follows:

Age: Age of the patients

Pregnancy week: the of pregnancy the patient is at

Carbs: grams of carbohydrates patients consumed

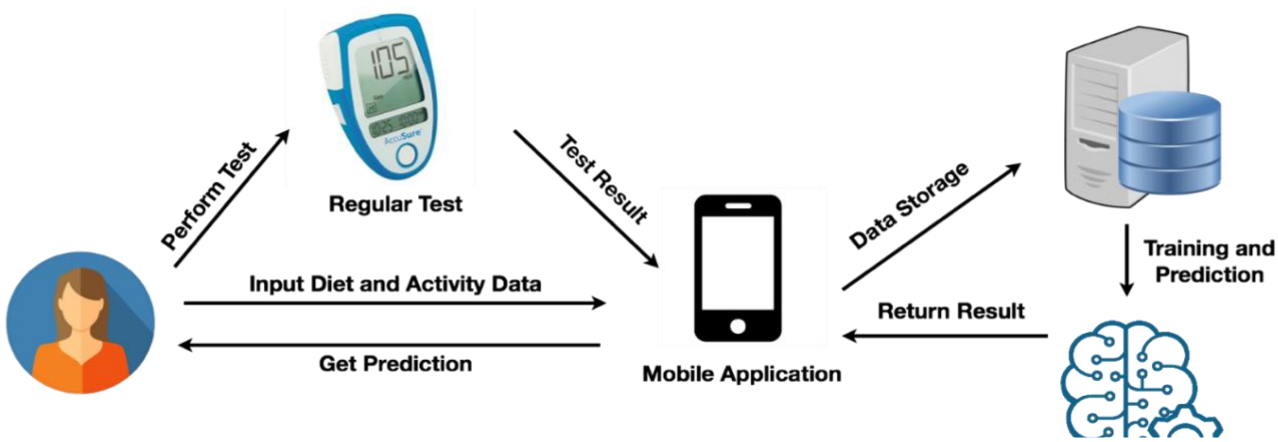

Figure 1. The Overview of the Solution

Protein: grams of protein patients consumed
Fiber: grams of fiber patients consumed
Time: Categorical variable on meal time
Exercise: Minutes spend on exercise by patients
Blood Sugar Value: milligrams of sugar per deciliter of blood

C. Training and Prediction

We used the standard machine learning library scikit-learn [9] to training and prediction the model. Scikit-learn provides different machine learning procedures such as classification, regression, cluster dimensionality reduction. It is built on strong mathematical packages such as NumPy and SciPy. In order to get the most accurate approach, we have used 3 different training algorithms to test the results, followed by some comparisons. The approaches we have chosen are: *Linear regression, Polynomial regression, and Gaussian Process Regression*. Linear regression attempts to minimize the sum square of residuals, and is a simple approach for unimodal problems [10]. Polynomial regression is a form of linear regression with polynomial terms of features, offering more complexity [11]. The Gaussian Process Regression interpolates values between data points with Gaussian process with prior covariance information [12].

D. Mobile Application

The frontend mobile application is developed using MIT App Inventor [13]. MIT App Inventor is a web-based integrated development environment for developing Android mobile applications. MIT App Inventor provides an interactive and real time debugging and testing environment that accelerates building simple applications.

As shown in Figure 2, the app has one dedicated screen that enables users to type and input their diet and blood sugar values. The data will be saved in the cloud database, which automatically triggers the training process.

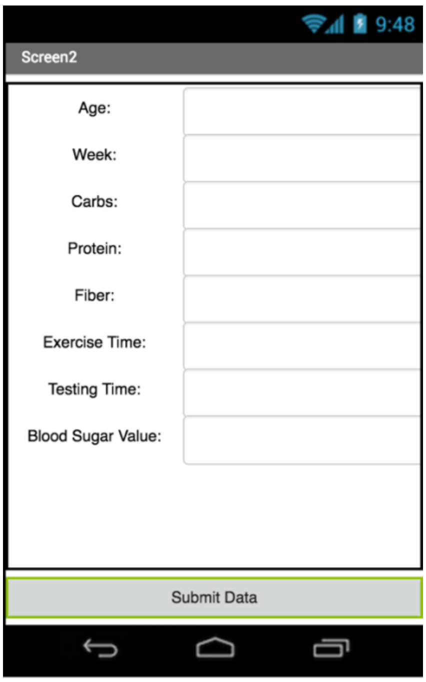

Figure 2. The Screen that Allows the User to input the Data

On the other hand, the patient can also start to make predictions of their blood sugar value by triggering the other screen of the app as shown in Figure 3. In this screen, the patient needs to input everything except the blood sugar level, which will be predicted using the latest trained model in the cloud.

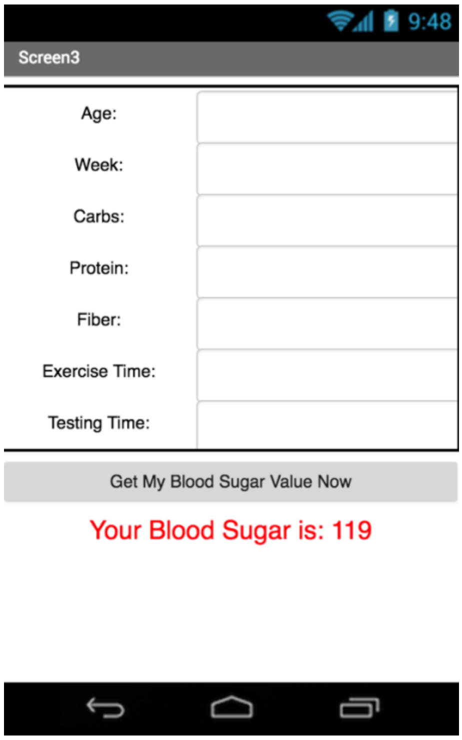

Figure 3. The Screen that Predicts the User's Blood Sugar Value

IV. EXPERIMENTS

To evaluate the accuracy of our approach, we have collected 500 real dataset from 4 gestational diabetes patients. In order to get the compare the approaches, we conducted experiments to verify two aspects: the accuracy of using different machine learning models, and the different selection of the training features.

A. Comparison of Different Machine Learning Algorithms

Figure 4 illustrates the accuracy of the prediction by using three different machine learning algorithms. As can be seen, the polynomial regression has the best result, due to its capability of handling the non-linear factors in the training process. By comparison, the linear repression has the lowest accuracy, which proves that the blood sugar level prediction in this case is not a linear situation.

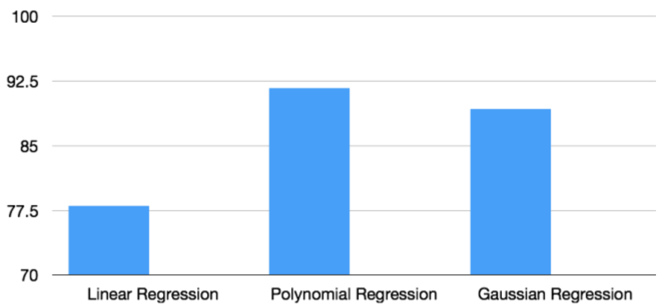

Figure 4. The Accuracy of Different Machine Learning Algorithms

B. Comparison of Different Sets of Training Features

We also looked into how the selection of the different training set will affect the accuracy. It turns out that the selection does have an impact on the accuracy. For instance, the Feature Set

Age	Pregnancy Week (1-40)	Carbs (g)	Protein (g)	Fiber (g)	Time (fasting, breakfast, lunch, dinner)	Exercise Time (mins)	Blood Sugar Value (mg/dL)
34	31	17	60	40	Breakfast	0	115
34	31	25	55	60	Lunch	10	129
34	31	23	100	80	Dinner	20	145
34	32	9	50	70	Breakfast	0	109
36	29	15	75	70	Fasting	0	128
36	29	14	70	65	Lunch	20	120

Table 1. A Sample of the training Dataset

1 and Feature Set 2 both have more features but the accuracy is lower, which indicts that both sets have the problem of over training. The Feature 3 is the set we showed in this project as shown in Table 1.

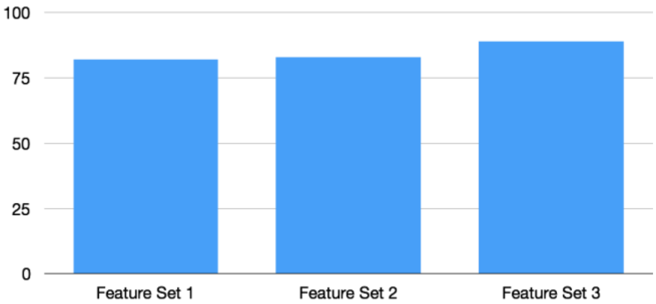

Figure 5. The Accuracy of Prediction using Different Feature Set

V. Related Work

Mani, S. et al [14] presented a machine learning approach to predict diabetes risk using data from electronic medical records. Syahputra, M. F., et al [15] calculated the number of caloric needs using Harris-Benedict equation and proposed genetic algorithm for scheduling diet for DM patient. We are using machine learning model to predict blood sugar value based on patients' specific data. Al Jarullah, A. A. [16] applied data mining to extract knowledge from information stored in database and generated clear and understandable description of patterns. In this work, we are able to train a model using machine learning algorithms which could be used to predict the blood sugar level based on patients' diet. Huang, Y., et al. [17] identified significant factors influencing diabetes control by applying feature selection to assist with classification. By comparison, we are using feature selection to do the model prediction. Plis, K. et al [18] described a solution that uses a generic physiological model to predicting blood glucose levels. This paper compares three approaches that are Linear regression, Polynomial regression and Gaussian Process Regression to train the machine learning model in order to get most accurate results.

VI. Conclusion and Future Work

In this project, we proposed an intelligent approach to address the problem of frequent blood sugar level tests for gestational patents using mobile computing and machine learning. A mobile app has been developed to collect the patient's diet and the tested blood sugar level. Once the sufficient amount of data has been collected, the system is able to train the machine learning model and predict the patient's blood sugar level based on the diet. Experiments show that our prediction without finger prick monitoring can reach to 91% accuracy when the patient is under a regular and routine diet and daily activity.

As for the future work, we will investigate other machine learning algorithms to keep improving the accuracy. We also would like to explore the possibility of applying deep learning in this problem domain.

In addition, one limitation related with the app is that it does not suggest the sufficient threshold of the training dataset. One feature we plan to add in the next version of the app is to automatically evaluate the accuracy of the training process and prompts the sufficiency of the dataset to users automatically.

References

[1] Stöppler, Melissa Conrad. "7 Gestational Diabetes Causes, Signs, Symptoms, Diet & Tests."MedicineNet, 5 Dec. 2016, www.medicinenet.com/gestational_diabetes/article.htm#what_are_risk_factors_for_gestational_diabetes.

[2] Donovan, Peter J, and H David McIntyre. "Drugs for Gestational Diabetes | Australian Prescriber." NPS MedicineWise, 1 Oct. 2010, www.nps.org.au/australian-prescriber/articles/drugs-for-gestational-diabetes.

[3] Berggren, Erica K., Boggess Kim A., Stuebe, Alison M., Jonsson Funk, Michele. "National Diabetes Data Group versus Carpenter-Coustan Criteria to Diagnose Gestational Diabetes." American Journal of Obstetrics and Gynecology, U.S. National Library of Medicine, 15 June 2011, www.ncbi.nlm.nih.gov/pmc/articles/PMC3670957/.

[4] "Pregnancy." Centers for Disease Control and Prevention, 16 Sept. 2015, www.cdc.gov/pregnancy/diabetes-gestational.html.

[5] Schmidler, Cindy. "Gestational Diabetes-Will My Baby Be Alright?" HealthPages.org, 22 Nov. 2017, www.healthpages.org/health-a-z/gestational-diabeteswill-baby-be-alright/.

[6] "The Facts About Carbs, Fiber, and Diabetes." Edited by Laura J Martin, WebMD, 14 May 2016, www.webmd.com/diabetes/guide/understanding-carbohydrates-fiber.

[7] " Dietary Recommendations for Gestational Diabetes." UCSF Medical Center, www.ucsfhealth.org/education/dietary_recommendations_for_gestational_diabetes/.

[8] Kaelbling, Leslie Pack, Michael L. Littman, and Andrew W. Moore. "Reinforcement learning: A survey." *Journal of artificial intelligence research* 4 (1996): 237-285.

[9] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." *Journal of Machine Learning Research* 12.Oct (2011): 2825-2830.

[10] Kutner, Michael H., Chris Nachtsheim, and John Neter. *Applied linear regression models*. McGraw-Hill/Irwin, 2004.

[11] Theil, Henri. "A rank-invariant method of linear and polynomial regression analysis." *Henri Theil's contributions to economics and econometrics*. Springer, Dordrecht, 1992. 345-381.

[12] Quiñonero-Candela, Joaquin, and Carl Edward Rasmussen. "A unifying view of sparse approximate Gaussian process regression." *Journal of Machine Learning Research* 6.Dec (2005): 1939-1959.

[13] Pokress, Shaileen Crawford, and José Juan Dominguez Veiga. "MIT App Inventor: Enabling personal mobile computing." *arXiv preprint arXiv:1310.2830* (2013).

[14] Mani, S., et al. "Type 2 diabetes risk forecasting from EMR data using machine learning." *AMIA annual symposium proceedings*. Vol. 2012. American Medical Informatics Association, 2012.

[15] Syahputra, M. F., et al. "Scheduling Diet for Diabetes Mellitus Patients using Genetic Algorithm." *Journal of Physics: Conference Series*. Vol. 801. No. 1. IOP Publishing, 2017.

[16] Al Jarullah, Asma A. "Decision tree discovery for the diagnosis of type II diabetes." *Innovations in Information Technology (IIT), 2011 International Conference on*. IEEE, 2011.

[17] Huang, Y., et al. "Feature selection and classification model construction on type 2 diabetic patients' data." *Artificial intelligence in medicine* 41.3 (2007): 251-262.

[18] Plis, Kevin, et al. "A Machine Learning Approach to Predicting Blood Glucose Levels for Diabetes Management." *AAAI Workshop: Modern Artificial Intelligence for Health Analytics*. 2014.