Time series prediction of personalized insulin dosage for type 2 diabetics

Jisha G¹, Nikhila T. Bhuvan², Ritta Jerrard¹
¹Department of Computer Science and Engineering, Rajagiri School of Engineering and Technology, Kakanad, India
²Department of Information Technology, Rajagiri School of Engineering and Technology, Kakanad, India

ABSTRACT
Careful blood glucose monitoring and consistent insulin administration are necessary for managing diabetes. People with demanding schedules or little access to medical personnel may find this difficult. Fortunately, without having to visit a doctor every day, daily insulin dosage may now be customized to a person’s unique needs using technology and customised algorithms based on their food intake, exercise routines, and blood glucose levels. This information can be entered into a diabetes management app or device, where an algorithm will determine the proper insulin dosage and offer real-time feedback to assist maintain ideal blood glucose levels. A patient’s dietary preferences, degree of physical activity, and blood sugar are taken into account for determining the proper bolus and basal insulin dosages in this study. According to the tracked body data, a patient’s appropriate insulin dosage is predicted using artificial neural network (ANN)-based models. Based on patient activity, food intake, exercise, and past insulin administration, insulin projections are created. To forecast an individual’s basal and bolus insulin requirements, long short-term memory (LSTM) and random forest regression models are employed. Accuracy of both models are tested and random forest regression shows better accuracy which is used in the prediction system.

Keywords:
Bolus and basal
Insulin prediction
Long short-term memory
Neural network
Random forest

This is an open access article under the CC BY-SA license.

Corresponding Author:
Jisha G
Department of Computer Science and Engineering, Rajagiri School of Engineering and Technology
Rajagiri Valley, Ernakulam-682039, Kerala, India
Email: jishag@rajagiritech.edu.in

1. INTRODUCTION
The international diabetes federation reported that there were 30 million type 2 diabetics worldwide in 1985. According to the surveys conducted in 2021, international diabetes federation approximated that 537 million people between the ages of 20-79 are living with diabetes. The number of persons at risk for developing diabetes is expected to reach 643 million by 2030 and 783 million by 2045. According to the World Health Organization (WHO) and International Diabetes Foundation, the global prevalence of diabetes will rise in the next few years, pushing up health spending to at least $966 billion. Blood glucose levels that are excessively high can lead to the condition ‘diabetes’. Many with diabetics require insulin injections to maintain a normal blood glucose level. A diabetic patient’s insulin dosage is extremely important. There won’t be enough glucose in the blood if too much insulin is administered. When predicting insulin dosage or blood glucose levels (BGL), non-linear regression models like artificial neural networks must be used because there is a non-linear and dynamic link between input variables like a patient’s diet, physical activity, stress, etc. and the glucose level. A steady insulin injection can occasionally be lethal. Insulin-dependent individuals must therefore administer the correct dosage of insulin to their bodies. Basal and bolus insulin supplements are typically given by
injection to diabetic patients. Long-lasting insulin, or basal insulin, keeps blood sugar levels stable throughout the day. When bolus insulin is administered prior to each meal, the blood glucose level is immediately brought under control. It is not safe for diabetic people to routinely visit a hospital to change their insulin dosage during this covid epidemic. Based on a patient’s daily food, physical activity, and past dosage/blood glucose level, this research suggests a method to forecast the intake of insulin.

For training and forecasting the insulin dosage based on nutrition, exercise, and previous dosage, data from the OhioT1DM [1] dataset is used. The OhioT1DM dataset includes information on 12 persons who have type 1 diabetes over a period of 8 weeks. All of these patients were receiving insulin pump treatment and ongoing glucose monitoring (CGM). They submitted information on their blood sugar levels, insulin levels, self-reported life events, and information from physiological fitness bands. The dataset consists of a CGM blood glucose reading every 5 minutes, readings from periodic self-monitoring of blood glucose (finger sticks), bolus and basal insulin doses, self-reported meal times with carbohydrate estimates, self-reported times of exercise, sleep, work, stress, and illness, and physiological information from fitness bands.

Several approaches have been put forth to predict blood glucose, ranging from traditional statistical techniques to models based on machine learning. The models based on machine learning (ML) and artificial intelligent (AI) began to outperform the conventional models. A feed-forward neural network model was developed for the 75-minute prediction horizon of real-time blood glucose (BG) prediction. The inputs included CGM values, insulin dosages, metered glucose values, nutritional value, lifestyle, and emotional aspects. An implementation of a deep neural network that predicts blood glucose using long short-term memory (LSTM) and bidirectional LSTM (Bi-LSTM) in [2] enables diabetic patients to respond before impending hyperglycemia and hypoglycemia. To forecast blood glucose levels for various predictive horizons, a sequential model containing a LSTM layer, a Bi-LSTM layer, and many fully connected layers were used. A Bluetooth low energy (BLE)-based paradigm for the internet of things was implemented in [3]. Real-time analysis of blood pressure, heart rate, weight, and blood glucose from sensor nodes was done using a machine learning method (BLE-based sensors). User sensor data is the input used in real-time data processing, which offers an early diagnosis of diabetes. An IoT-enabled diabetes monitoring system was used in [4] to determine the glucose level and provide alert messages to the patients about medication, food, and activity restrictions. The glucose prediction for both type 1 and type 2 diabetics was also done. Self-monitoring of blood glucose should be performed three or more times a day for the majority of type 1 diabetic patients using insulin pump therapy or multiple insulin injections, whereas it may be helpful for patients using non-insulin therapies or less frequent insulin injections to meet their glycemic targets.

Li and Fernando [5] developed a brand-new deep learning framework is put out for predicting blood glucose levels with edge inference in a microcontroller unit integrated in a system utilizing a recurrent neural network that is based on short-term memory. Personal models achieve cutting-edge performance on a clinical data set from 12 T1DM suffers by using a CGM sensor to detect glucose and a recurrent neural network that builds on long-term memory. Use of technology in predicting the bolus insulin dosages helps patients to plan more accurately their carbohydrate intake is improving [6]. It describes how artificial neural networks like recurrent, deep reinforcement models [7], [8] are used to forecast BGLs using data from previous BGLs, meal intake, and insulin injections. During the nocturnal phase of the daily cycle, artificial neural networks (ANNs) built on the Elman recurrent network and trained with the Levenberg-Marquardt algorithm were capable of making precise short- and long-term blood glucose predictions. Deep learning and machine learning models can be used for predicting diabetes and to determine the possible type of the diseases that can occur in the future due to diabetics [9]. Pernma et al. [10] proposed a learning vector quantization which uses genetic algorithms to detect diabetics.

Although there has been much research on blood glucose prediction, it is important to apply this prediction when estimating insulin dosage consumption to successfully drop BGL to the desired range. By effectively tracking the crucial information that depends on blood glucose level, physical activity, and diet, insulin dosage prediction for a diabetic patient can be carried out. Karim et al. [11] proposed a neural network model which uses the nutrient level of each meal along with previous doses of insulin and BGL. To predict insulin dosage, Zahran [12] proposed an ANN model that took height, weight, gender, and fast blood sugar level into account. Saha et al. [13] developed a way for improving control with more complex insulin administration for these individuals by interpreting data from two additional parameters, such as average fasting blood glucose and physical activity of the same patient population. The results were more accurate because the fuzzy system accumulated more patient-related factors. A shallow neural network model was developed for patient-specific prediction of glucose every 60, 90, 120, 180, and 240 min ahead, using features extracted from past CGM data and insulin logs [14]. An algorithm that is able to predict blood glucose evolution based on the personal nutrient intake and subcutaneous insulin injections data of diabetic out-patients using machine learning model is implemented [15]–[17] using a mobile-based sophisticated neural network that is accessible to patients with diabetes. By employing a mobile device to regulate daily insulin levels for patients who are less able to control their diabetes, the application aims to improve the knowledge and abilities of

**Time series prediction of personalized insulin dosage for type 2 diabetics (Jisha G)**
The aim of the proposed work is to reduce the risk of hyperglycemia and hypoglycemia in patients with comorbidities during the pandemic period. To prevent casualties and maintain a balanced insulin regimen in patients with comorbidities, basal and bolus insulin dosages must be adjusted daily with extreme caution. A diabetic patient receives 50% of their insulin from bolus and 50% from basal dose. The proposed prediction model’s system design is shown in section 2. Under this section, it is detailed how the LSTM model and random forest model are utilized to forecast a patient’s blood glucose level and insulin dosage depending on previous blood glucose levels and food habits. Based on the mean squared error (MSE) and coefficient of determination ($R^2$) values, the result section contrasts the efficiency of two models.

2. METHOD

On the basis of past data, personalized insulin prediction includes predicting a person’s insulin intake. To forecast patients’ insulin consumption, a time series analysis of CGM value, food intake, and physical activity is carried out here. A time series is made up of a sequence of measurements taken over a period of time. A diabetic patient’s accurate insulin dosages are very important, and it should be ensured that they receive the appropriate number of doses based on their individual bodies and daily routines. Two dosage types basal and bolus are used in dosage computation. The information on daily exercise or physical activities, expected carbohydrate intake, and current blood glucose levels can be used to forecast insulin dosages more accurately.

To prevent catastrophes, patients with comorbidities must maintain tight control over their blood glucose levels. Insulin prediction can be useful for different categories of people. For example, during an fluorodeoxyglucose (FDG) PET scan, a tailored insulin prediction can be helpful in identifying metabolically active malignant tumors [20]. The appropriate amount of insulin that should be injected into the body prior to treatment is therefore predicted based on age, gender, BMI, insulin intake, diabetic history of the patient, and other factors. The F-FDG intake can result in hyperglycemia that can negatively affect the diabetic patient. For the purpose of reducing hyperglycemia before an FDG PET/computed tomography (CT) scan, a customized insulin calculator can be used to calculate the appropriate dose of IV insulin. For optimum blood glucose management during glucocorticoid [21] therapies during an arthritic attack or following an organ transplant, individualized insulin prediction is another use case. Since these medications encourage the production of glucose in the liver and decrease cell sensitivity to insulin, one of the side effects of glucocorticoid treatment is an increase in blood glucose. As the side effects vary from person to person depending on the dosage of glucocorticoids prescribed, it is necessary to check blood glucose levels more frequently than usual to monitor the drug’s effect on the management of diabetes.

The study is conducted to predict insulin dosage to enhance the health and welfare of diabetic’s people. In this study, two distinct prediction models, LSTM and random forest regression models are used for predicting basal and bolus insulin dosages for a day using the OhioT1DM dataset [1]. Figure 1 represents the overall architecture of the insulin prediction model. As per Figure 1, the training process of two models, stacked LSTM and random forest, is conducted using data collected from OhioT1DM dataset. The OhioT1DM dataset used in the study, consists of eight weeks of continuous glucose monitoring, insulin dosage levels, and self-reported life event data for each of the 12 patients with type 1 diabetes [2]. The dataset includes blood glucose readings from a CGM every 5 minutes; blood glucose readings from periodic self-monitoring of blood glucose (finger sticks); insulin doses, both bolus and basal; self-reported meal times with carbohydrate estimates; self-reported times of exercise, sleep, work, stress, and illness; and 1-minute or 5-minute aggregations of heart rate, galvanic skin response (GSR), skin temperature, air temperature, and step count.

The data used for prediction includes current blood glucose level, current insulin doses, meal time, carbohydrate intake and physical activity rate that is taken from the OhioT1DM dataset. These data are preprocessed to standardize the values. The data preprocessing involves label encoding which refers to converting the labels into numeric form so as to convert them into machine-readable form. Machine learning algorithms can then decide in a better way how those labels must be operated. Sklearn provides a very efficient tool for encoding the levels of categorical features into numeric values. LabelEncoder encodes labels with a value between 0 and nclasses-1 where n is the number of distinct labels. If a label repeats it assigns the same value as assigned earlier. During preprocessing, meal-type is converted from text to numeric form. Then the dataset is normalized using minmax scalar which transforms features by scaling each feature to a given range. This estimator scales and translates each feature individually such that it is in the given range on the training set, e.g. between 0 and 1 the real-time data like current blood glucose level, meal type (breakfast, lunch, snack and dinner), carbohydrate intake and physical activity details are collected from the user. The carbohydrate intake can be either collected directly from the user or calculated based on the meal type. Physical activity includes the self-reported time of exercise, level of stress, and illness. Self-reported times of exercise is based
on exercise intensity rate perceived exertion (RPE) which ranges between 1-10 and the duration of exercise.

The proposed model predicts the patient’s insulin dosage using inputs such as blood glucose level, carbohydrate intake, meal-type, and exercise intensity. The investigation revealed that, when compared to the LSTM model, random forest regression produced the best results; this technique employed ensemble learning for regression by building numerous trees and merging them to generate the best result.

Figure 1. Overall architecture of the insulin prediction model

2.1. LSTM based neural network model for insulin dosage prediction

Memory cells that can store state information over time make up the LSTM of a neural network, while a gate structure manages and regulates cell state data. Recurrent neural networks of the LSTM can learn order dependence in sequence prediction issues. The LSTM model is created with the goal of learning from prior patient data, extracting patterns, and forecasting future value. Current blood sugar levels, meal type, carbohydrate intake, and physical activity, including length and intensity, were all provided. LSTM is a type of recurrent neural network (RNN) that has shown promising results in time series prediction tasks, including bolus insulin prediction. LSTM network was used to deliver an accurate glucose prediction output in the patients with diabetes management devices that rely on CGM sensors which monitor glucose levels from the interstitial fluid (ISF) [22].

Preprocessing involves removing all invalid values and normalizing the dataset using MinMax scaler and label encoder. The process of data framing prepares the dataset for training the model. To make the model predict based on previous time steps lookback was set to 3 and delay was set to indicate number of time steps in future target. The hyperparameters such as optimization algorithm, batch size, activation functions were all set in the best way that model predicts the most accurate value. Here a stacked LSTM [23] is used with an input layer comprising 30 neurons, a hidden layer with 20 neurons and an output layer to predict the insulin dosage. It was found that models with activation layer rectified linear units (ReLU) [24] and optimizer Adam provided the best results. The model was trained with an epoch of 50 to monitor the loss. A h5 file with output was generated in each run for later evaluation. The model performance is evaluated using the loss function. 20% data is used for validation split up. Validation loss obtained for the model is 0.0265 and the training loss obtained is 0.0211.

2.2. Random forest regression model for insulin dosage prediction

Random forest regression is a machine learning algorithm that can be used for predicting continuous variables, such as the amount of bolus and basal insulin needed. It is a regression model that takes into account all of the inputs to forecast the bolus and basal insulin doses. The association between scalar dependent and independent variables may be discovered using the regression model [25]. Random forest consists of a collection of different regression trees. It predicts using a set of trees rather than a single tree. The logic behind a set of trees for prediction is to mitigate the instability issue of each tree by combining the prediction of multiple trees. In an experiment for predicting diabetes mellitus using decision tree, random forest and neural
network, the prediction with random forest could reach the highest prediction accuracy [26]. That is the driving force for comparing decision tree with LSTM. The method uses ensemble learning methods for regression by constructing multiple trees and combining them to produce best result. Here, 20 decision trees are used for predicting bolus doses and 10 decision trees are used for predicting basal dosages. The results from the various decision trees are aggregated. The one with the most voted prediction result is the final prediction result. ‘d’ sets of trees for prediction is used to mitigate the instability issue of each tree by combining the prediction of multiple trees.

3. RESULTS AND DISCUSSION

This section provides a detailed analysis of two models used for predicting insulin dosage for patients who regularly use insulin to maintain the blood glucose level. The two models used for analysis are LSTM and random forest regression. Analysis was done for both basal and bolus insulin dosage prediction.

3.1. Bolus insulin prediction using LSTM and random forest regression—a comparison

Bolus is an insulin used to reduce the high blood glucose spike that might occur when people include a lot of carbs in their food. The bolus doses are typically administered three times a day, at least, before each meal. Blood sugar levels, meal type, carbohydrate intake, and physical activity, including duration and intensity, are taken into account from the OhioT1DM dataset as input features to predict the bolus insulin. The comparison of prediction results using both models is shown in Figure 2. Figure 2(a) depicts the bolus prediction using stacked LSTM and Figure 2(b) shows bolus prediction using random forest regression. From the Figures 2(a) and 2(b), it is evident that random forest regression offers better predictions than LSTM, hence this model was chosen and integrated into the system.

![Figure 2](image_url)

Figure 2. Target Vs predicted graph of bolus prediction using two models (a) LSTM and (b) random forest regression
3.2. Basal insulin prediction using random forest regression

Basal is a long-acting insulin that patients should ideally take just once daily. Since there are fewer patient historical records, there are also fewer basal values overall. As random forest regression model provides better prediction of bolus insulin, basal insulin prediction also was done using random forest method. Although the bulk of a patient’s basal value falls between 0 and 5, it may vary from patient to patient. In contrast to bolus, the basal value is constant. The basic actual and projected graph is shown in Figure 3, and the model is assessed using RMSE, MSE, and $R^2$. Figure 3 provides a comparison between actual and predicted insulin doses generated during different intervals. Using a random forest regression model both predicted and actual basal values are almost similar which is evident from the Figure 3 and the RMSE, MSE, and $R^2$ values shown in Table 1.

![Figure 3. Target Vs predicted graph of random forest regression in case of basal insulin prediction](image)

Table 1 shows the performance matrix of LSTM and random forest models for predicting bolus and basal dosages. The evaluation is done using performance evaluation factors RMSE, MSE, and $R^2$. As random forest outperformed the LSTM model the basal insulin prediction was done only using the random forest method which provided a $R^2$ of 0.98.

<table>
<thead>
<tr>
<th>Performance evaluation factors</th>
<th>Bolus insulin dosage prediction using LSTM</th>
<th>Bolus insulin dosage prediction using random forest</th>
<th>Basal insulin dosage prediction using random forest</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>2.13</td>
<td>0.58</td>
<td>0.13</td>
</tr>
<tr>
<td>MSE</td>
<td>5.21</td>
<td>0.2302</td>
<td>0.02</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.02</td>
<td>0.93</td>
<td>0.98</td>
</tr>
</tbody>
</table>

4. CONCLUSION

The paper proposes a system using ANN based model to predict the proper amount of insulin dosage for the diabetic patients. The system will be useful for an individual to analyze the pattern of food intake and exercise to find proper insulin dosage which is taken on a daily basis. In most of the related works, the dataset was collected from Ohio university from which prediction was done based on food, exercise, activity of patients and previous insulin dosage taken using an insulin pump. Two-time series forecasting model’s LSTM and random forest regression are used in this work. The results can be improved by incorporating more historical data of the patients and considering other factors including their stress, sleep, heart rate, medicine intake etc. LSTM and random forest were used to predict both basal and bolus insulin dosage of a person. The random forest appeared to be the best in predictions using time series data. Accumulating longer term insulin data and medication details of patients with comorbidities from clinical servers can also be used to train and predict more accurate results.
REFERENCES


Time series prediction of personalized insulin dosage for type 2 diabetics (Jisha G)

BIOGRAPHIES OF AUTHORS

Dr. Jisha G is Assistant Professor in the Department of Computer Science and Engineering, Rajagiri School of Engineering and Technology, Kochi, India. She holds a Ph.D. Degree in information and communication engineering with specialization in Heterogeneous Networks. She is research guide under APJ Abdul Kalam Technological University, Kerala, India. She has got 20 years of teaching experience. She has supervised more than 20 masters and 4 Ph.D. students. She has a patent published under the application No.: 2021410420 for the title “IoT enabled step to ramp for mobility impaired”. She had publications in reputed Journals and presented papers in prestigious conferences. Her research areas are wireless networks, image processing, machine learning, and its applications. She can be contacted at email: jishag@rajagiritech.edu.in.

Dr. Nikhila T. Bhuvan is Assistant professor in the Department of Information Technology at Rajagiri School of Engineering and Technology, Kakanad, India. She was awarded Ph.D. in faculty of Engineering (CS) from Cochin University of Science and Technology, Kerala, Kochi, India in 2021. She completed her B. Tech from the Kannur University and M. Tech from IGNOU. She has got 14 years of teaching experience. She is the chapter author for the book, “data mining and warehousing”. She has a patent published under the application No.: 202141042022 for the title “IoT enabled step to ramp for mobility impaired”. Her areas of interest are cloud computing, information retrieval, AI, and machine learning. She can be contacted at email: nikhilabhuvan@rajagiritech.edu.in.

Ms. Ritta Jerrard completed B. Tech in Computer Science and Engineering from Rajagiri School of Engineering and Technology in the academic years 2017-2021 under APJ Abdul Kalam Technological University. Her area of interest includes artificial intelligence and its applications, currently focusing on advancing a career in the field of AI and is working towards that. She can be contacted at email: rittajerrad@gmail.com.