Estimating Calories using CGM and Image Data

*Note: Sub-titles are not captured in Xplore and should not be used

1st Shyam Sankar
Department of Computer Science and Engineering
Texas A&M University
College Station, United States of America
shyam.sankar@tamu.edu

***Abstract*—Monitoring dietary intake is essential for managing diabetes, a chronic disease that affects more than ten percent of the world's adult population. For individuals with Type 1 and Type 2 diabetes, managing dietary habits and thereby controlling the blood sugar level is vital to preventing cardiovascular diseases. While the traditional methods for tracking dietary intake are often inaccurate and inconvenient, recent advancements have shown that data collected from sensors along with photographs of meal can be analyzed to streamline the process. This project aims to develop a multi-modal machine learning model, that integrates modalities developed from sensor data, demographic information and meal images to improve and automate the monitoring of dietary intake.**

***Index Terms*—Diabetes, Calorie Estimation, Continuous glucose monitors, Machine Learning**

I. Introduction

Diabetes occurs when the human body becomes incapable of regulating the glucose level in blood, which can result in either Type 1 or Type 2 diabetes. Insulin is a hormone produced by the pancreas that is used by the human body to regulate glucose levels. An individual is said to have Type 1 diabetes if the pancreas does not produce enough insulin. An individual is said to have type 2 diabetes when the cells become less responsive to insulin. Diabetes leads to high glucose levels, this can result in cardio vascular diseases, neuropathy, retinopathy and kidney diseases. [1]

The blood glucose levels are heavily impacted by the food consumed. Carbohydrates in the form of sugars and grains are broken down into glucose during digestion and absorbed into blood streams, resulting in momentary increse in glucose levels. Hence, for diabetes patients, it is important that they monitor type, timing and quantity of food consumed, to prevent causing any drastic fluctuations in glucose levels.

Clinical glucose meters (CGM) is a device that measures the concentration of glucose, from a small blood sample. A CGM uses the electrical current produced by a reaction between glucose in the blood and glucose oxidase to measure the amount of glucose present in the sample. CGMs allow diabetes patients to get real time information on blood glucose levels.

A. Related Work

Huo, Zepeng et al. (2019) introduces a multitask neural network capable of estimating the macro nutrient composition of meals from CGM signal data. The area under the curve at five different points in the CGM data was extracted to be used as features. The multi-tasking neural network provides better performance in predicting macro nutrient composition when compared against baseline least squares regression model. The neural network was designed to have a fully connected layer to learn shared features, along with a task specific layer for macro nutrient estimation. [2]

A. Das et al. (2021) uses the post-prandial glucose response(PPGR), the response seen in glucose levels after a meal to predict the macro nutrient content in the meal. The experiment uses Lasso regularization to represent the PPGR of consumed meal as a sparse combination of PPGRs in a dictionary. [3]. M. Yang et al. (2021) introduces a model using Siamese Neural Network (SNN) architecture, to learn PPGR embeddings using a contrastive loss function, designed to maximized distance between meals of similar macro nutrient composition and minimized the distance between meals with different macro nutrient composition. [4]

L. Zhang et al. introduces a multi-modal apporach for predicting calories from a model, improving upon the models using data gathered from CGM devices, by combining it with photographs of meals. The model used Gaussian based kernels to extract gAUC features from the CGM signal data, an attention based transformer is used for processing the data. The multi-modal approach used a transformer for CGM data and a vision transformer for the image data, showcasing a Normalized Root Mean Squared Error(NRMSE) of 0.34 and correlation of 0.52. [6]

B. Dataset

The training dataset is from an experiment conducted where 36 individuals where provided with meals of known composition, CGM sensor data measuring their glucose levels were collected every five minutes. Additionally data related to their health and demographics is available. Images were taken prior to breakfast and lunch as well.

Details regarding subjects from the training data are as follows:

TABLE I
NUTRITION DETAILS PER DAY

Day	Calories	Carbs (g)	Fat (g)	Protein (g)
1	830	92	42.0	17
2	435	16	14.0	66
3	555	94	13.0	12
4	355	19	15.0	32
5	1180	81	54.5	88
6	830	92	42.0	17
7	435	16	14.0	66
8	555	94	13.0	12
9	355	19	15.0	32

- The age of subjects ranged from 22 to 69 years with an average of 50 years.
- The subjects comprised of 24 males and 12 females.
- The weight of the subjects ranged from 117lbs to 285lbs with an average of 179lbs.
- The height of the subjects ranged from 59 to 72 inches with an average height of 64 inches.
- There are 10 subjects with diabetes status 1, 15 with diabetes status 2 and 11 with diabetes status as 3.
- The subjects had an average baseline fasting glucose level of 122.
- The insulin level of the subjects ranged from 2.5 to 46.4 with an average of 14.5.

II. METHODS

A. Data Preparation

We are dealing with three different sets of data.

- The first dataset is a time-series data collected from a subject's continuous glucose meter through the day. For the purpose of this project, we have extracted the glucose values, ignoring the timestamps. In order to make the data more suitable for numerical operation, we convert the data into NumPy array.
- The second dataset we have is the health and demographic information for each subject involved in the study. Numerical features available were standardized in this dataset while ignoring data on race and biome data, transforming them to have mean zero and standard deviation one. Normalization was done to prevent machine learning models from being sensitive to unfair magnitude between different features.
- The third dataset contains images of lunch and breakfast meals. Ignoring the images available on breakfasts, we checked each image for having the expected 3D shape, resized them to 224x224 pixel size. If the image had incorrect dimensions, or was not available, they were replaced with a default image.

B. Model

a) Image Model: Here we use the pre-trained ResNet-50 model. The final layer of the ResNet model, typically a fully connected layer is replaced with a linear layer, which gives us a feature vector of size 512 as output. The model serves as the feature extractor for the image data.

b) CGM Model: With the CGM data being a sequential data, a recurrent neural network was designed which uses Long Short Term Memory (LSTM). The LSTMs allow us to capture temporal dependencies in the CGM data.

c) Biometric Model: The biometric model is defined as a fully connected neural network. The model consists of three fully connected layers with a Rectified Linear Unit (ReLU) activation function after each layer to introduce non-linearity. The final output has 64 dimensions.
The Rectified Linear Unit is defined as follows,

$$\mathrm{ReLU}(x) = \max(0, x)$$

d) Multimodal setup: The muti-modal calorie predictor integrates the data from the three modalities to predict the calorie intake. Unique features are extracted from sub-models as defined above. The predictor is designed to have a series of fully connected layers to learn complex features from the concatenated features.

C. Model Training

a) Loss Function: Root Mean Squared Relative Error is a loss function that measures the loss by taking the average of the relative difference between the predicted and true values.

$$\mathrm{RMSRE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_{\text{true},i} - y_{\text{pred},i}}{y_{\text{true},i} + \epsilon}\right)^2}$$

Where:

- $y_{\text{true},i}$ represents the actual value for the i-th sample,
- $y_{\text{pred},i}$ is the predicted value for the i-th sample,
- ϵ is a small constant (e.g., 10^{-6}) added to avoid division by zero when $y_{\text{true},i}$ is very small or zero,
- n is the number of data points in the dataset.

b) Optimization: Adaptive Moment Estimation abbreviated as Adam is an optimization algorithm that combines AdaGrad and RMSProp, adapting the learning rate based on past gradients and using a moving average of squared gradients to scale the learning rate. Because of the large nature of the dataset and the complex nature of the multi-modal machine learning model, Adam serves as a great choice for optimization as the algorithm adjusts the learning rate for each parameter individually.

The key update rule for the parameters is given by the following formula:

$$\hat{m}_t = \beta_1 \hat{m}_{t-1} + (1 - \beta_1) g_t$$

$$\hat{v}_t = \beta_2 \hat{v}_{t-1} + (1 - \beta_2) g_t^2$$

$$\theta_t = \theta_{t-1} - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}$$

Where:

- g_t is the gradient of the loss function with respect to the parameter at time step t,
- $\hat{m}_t$ and $\hat{v}_t$ are the bias-corrected estimates of the first and second moments at time step t,
- β_1 and β_2 are hyperparameters that control the exponential decay rates of the moving averages,
- α is the learning rate,
- ϵ is a small constant (e.g., 10^{-6}) added to prevent division by zero.

c) Training: The multi-modal calorie predictor was trained using the Adam optimizer and RMSRE loss function. After expeimenting with hyper-parameters, a learning rate of 0.001 and batch size of 32 was chosen. The model converges after seven epochs with a training RMSRE 0f 0.32 and validation RMSRE of 0.33. The model converges

TABLE II
RMSRE FOR TRAIN AND VALIDATION AT DIFFERENT EPOCHS

Epoch	Train RMSRE	Validation RMSRE
1/7	0.9828	0.8242
2/7	0.7473	0.3368
3/7	0.3795	0.8054
4/7	0.3433	0.3899
5/7	0.3418	0.4448
6/7	0.3391	0.3772
7/7	0.3254	0.3327

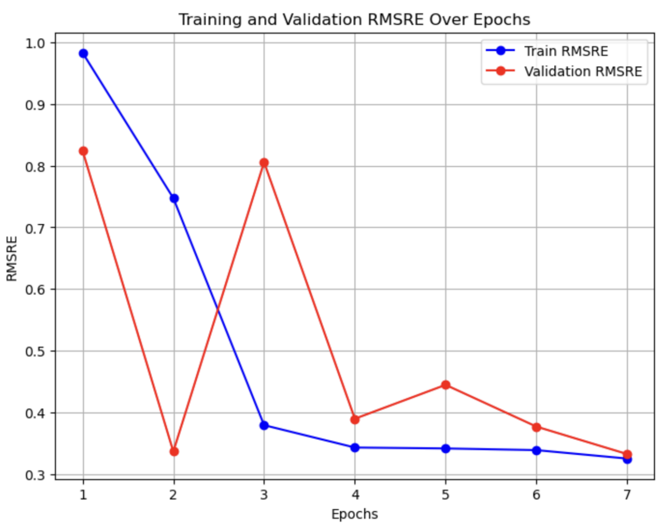

Fig. 1. RMSRE for train and validation for epochs

The above table and image on the training and validation performance of the model over seven epochs show a mix of improvement and fluctuations in the RMSRE (Root Mean Square Relative Error). Initially, in Epoch 1, the model shows a relatively high RMSRE for both training (0.98) and validation (0.82), indicating room for significant improvement. As training progresses through Epochs 2 to 4, there is a notable decrease in the training RMSRE, suggesting that the model is learning and fitting better to the training data. However, the validation RMSRE exhibits some instability, with a sharp drop in Epoch 2 (0.33) followed by an increase in Epoch 3 (0.80). From Epoch 4 onwards, both the training and validation RMSREs stabilize, with the model continuing to improve slightly in terms of training error. By the final epoch (Epoch 7), the training RMSRE reaches its lowest point at 0.32, and the validation RMSRE settles at 0.33, suggesting that the model has learned to good generalization.

d) Testing: Predictions where made using the test features provided. A testing RMSRE score of 0.3412 was obtained (Kaggle Score).

III. CONTRIBUTIONS

All the work on the implementation was carried out by Shyam Sankar. Dataset was provided by Dr. Bobak J. Mortazavi.

IV. CONCLUSION

Hence, the multi-modal machine learning model was developed using CGM, demographic and health data as well as Image data. The model has a testing RMSRE of 0.34, an improvement of 0.18 from the benchmark.

REFERENCES

[1] Wikipedia contributors. ”Diabetes.” Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 4 Dec. 2024.
[2] Huo, Zepeng et al. “Predicting the meal macronutrient composition from continuous glucose monitors.” 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI) (2019): 1-4.
[3] A. Das et al., ”A Sparse Coding Approach to Automatic Diet Monitoring with Continuous Glucose Monitors,” ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 2900-2904.
[4] M. Yang et al., ”A Metric Learning Approach for Personalized Meal Macronutrient Estimation from Postprandial Glucose Response Signals,” 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), Athens, Greece, 2021, pp. 1-5.
[5] L. Zhang et al., ”Joint Embedding of Food Photographs and Blood Glucose for Improved Calorie Estimation,” 2023 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), Pittsburgh, PA, USA, 2023, pp. 1-4
[6] Das, Anurag et al. ”Macronutrient Constraints and Priors Improve Carbohydrate Prediction From Continuous Glucose Monitors. Current Developments in Nutrition”, 2024