Mapana – Journal of Sciences
2025, Vol. 24, No. 4, *105-116*
ISSN 0975-3303 | https://doi.org/10.12723/mjs.75.7

Predictive Modeling for Blood Sugar Levels and Personalized Dietary Recommendations using Machine Learning Approach

S. Velmurugan*, K. Subramanian†

Abstract

In recent years, the prevalence of diabetes and other metabolic disorders has been increasing worldwide, highlighting the need for effective blood sugar management. One crucial factor in controlling these conditions is maintaining stable blood glucose through appropriate dietary choices. In this study, we propose a two-stage machine learning approach. The first model, developed using a nutritional dataset of 5,740 food items with 55 features from data.world, applies multiple linear regression to predict the sugar content of foods based on nutritional variables such as carbohydrate, starch, protein, fat, calcium, iodine, iron, and potassium. The second model recommends foods tailored to a patient's blood sugar range (e.g., fasting and post-meal thresholds) to support personalized dietary management. Data preprocessing steps addressed missing and duplicate values, and model evaluation was performed using train-test split and statistical measures including R^2, MAE, and MSE. The results show that the proposed framework predicts food sugar content with high reliability ($R^2 = 0.98$) and provides dietary recommendations with an overall accuracy of 97%. This demonstrates its potential for integration into mobile health applications and clinical decision support systems to aid diabetic patients in making informed food choices.

Keywords: Machine learning, diabetes management, personalized nutrition, predictive modeling, dietary recommendation.

* Department of Computer Science with Data Analytics, Kongunadu Arts and Science College, Coimbatore, Tamilnadu, India; svelmurugan_cs@kongunaducollege.ac.in

† School of Business Management, Christ University, Bangalore, India

I. Introduction

Particularly in those between the ages of 25 and 70, diabetes is one of the chronic diseases with the largest global growth rates [1] [2]. Around the world, 537 million persons aged 20 to 79 have diabetes, and that number is expected to rise to 643 million by 2030 and 783 million by 2045, according to the ninth edition of the IDF Diabetes Atlas [3]. An estimated 6.7 million adults will pass away from diabetes. With almost 77 million diabetics, India is the second-largest diabetes patient country in the world.

Serious complications from diabetes can include renal disease, vision loss, heart problems, nerve damage, etc. Diabetes is a chronic condition that needs to be managed with medicine, healthy food, consistent exercise, keeping a healthy BMI, and routine blood glucose level checks [4]. Healthy diet should be used to treat diabetes and pre-diabetes patients in order to maintain their blood sugar levels and lessen their need for medication.

The most fundamental need is for nourishing food, and keeping a healthy eating regimen can generally support excellent health. Food requirements vary by age, gender, weight, height, lifestyle, amount of activity, physical activity, and other factors. India is home to a large number of religious people, each of whom has their own eating preferences, making it difficult to choose foods that match individual needs and preferences [5].

Type 1 diabetes is a chronic disease that impairs insulin secretion because the immune system kills the beta cells in the pancreas [6]. The presence of type 1 diabetes was rapidly identified after the onset of symptoms. This can be managed with regular medical exams, using insulin as directed, eating well, and practicing meditation [7]. Type 2 diabetes manifests as a result of the body's inability to produce insulin. This is usually accurate for people who are between the ages of 25 and 70 [8]. This can be controlled with regular medical checks, using insulin as directed, eating appropriately, and practicing meditation [9]. Gestational diabetes, also known as type-3 diabetes, typically affects pregnant women but can be prevented by taking the right medications and eating a nutritious diet [10]. Having good diet is highly advised for the improvement of patient health because it is typically present in all three forms of diabetes [11].

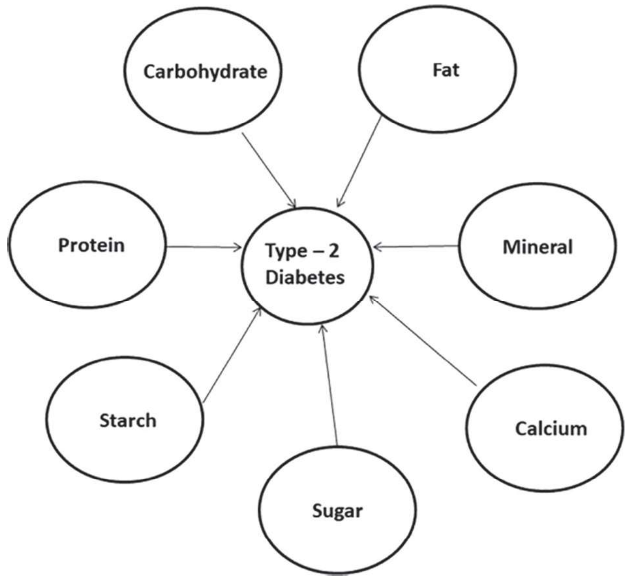

Figure1: Food Causes Type-2 Diabetes

We have proposed a machine learning model based on expert knowledge to detect the sugar level in food based on the nutrition in the food, and a second model is used to recommend the food to diabetes patients from nutrition dataset based on the patient's sugar level. This model assists the patient in controlling his diabetes by advising him to consume fewer or more sugary foods.

This paper proposed two models: the first to determine the sugar level in food based on the nutritional value, and the second to recommend healthy foods based on the patient's sugar level. The model focuses on maintaining the same sugar level in a diabetes patient for the whole day on the recommended foods from the nutrition dataset.

In this work, the first model predicts the sugar level of foods based on their nutritional composition, while the second model provides personalized dietary recommendations based on the patient's reported blood sugar range. This dual approach makes it possible to connect food composition with patient-specific glucose management. The rest of the paper is organized as follows: Section II is a related work. Section III presents the feature selection. Section IV discusses the machine learning algorithm. Section V is result and Section VI conclusion and future work.

II. Related Work

Sallian Shafi Bhat et al. (2021) proposed various machine learning algorithms for the detection of diabetes and a pre-processing approach to improve the accuracy of the model. The proper food recommended model is random forest algorithm, which gives accuracy upto 93% [12].

Kushan De Silva et al. (2021) proposed a model for the diagnosis of type-2 diabetes in adults and recommended clinical nutrition food. The model is working with accuracy up to70.88%[13].

Hyperkim et al. (2020) proposed a deep neural network model to find the nutrition intake effect on type-2 diabetes patients and the model accuracy of 62.48 % respectively [14].

Yajie Guo et al. (2020) discuss the role of nutrition in the prevention of type-2 diabetes and also discus show the main five nutrition's are protein, carbohydrate, fate, vitamin and mineral are responsible for developing type- 2 diabetes [15].

Phuong Ngo et al. (2019) proposed a machine learning model for food recommendations fortype-1 diabetes patients. In this paper he discussed two methods based on feed forward neural network and the reinforcement learning method and this method applied to short physical activity and long physical activity person. Found that reinforcement learning performs better for long physical activity and food intake spread out during exercise [16].

B.RajKumaretal.(2015) implemented the food recommendation system based on ontology and clustering analysis for food recommendations. The aim of the model recommendation to users of whether the suggested system of the food menu that it has declared based on the user profile condition is satisfied [17].

Irshad Faiz et al. (2014) proposed a prototype model for recommendation of balanced food and proper exercise for diabetes patients. This prototype uses semanticontology – based rules and healthy food and exercise recommendations for diabetes patients [18].

III. Methodology

This study is based on two models. The first model is used to find the sugar level in the food based on the nutrition value of the food and the second model is used for effective food recommendation based on the sugar level of foods provided in the nutrition dataset. The first model we have developed is considering various nutrients like carbohydrate, starch, protein, fat, calcium, iodine, iron, potassium and sugar levels of food which are the cause of diabetes. This model used a nutrition dataset which is collected from https://data.world/datasets has 55 feature value and 5740 number of foods, we select only highly effective nutrition which cause for the diabetes, like carbohydrate, starch, protein, fat, calcium, iodine iron, potassium and sugar value. This nutrient issues to develop a model by using a multi-linear regression machine learning model which gives the unknown foods sugar level. The second model is developed for recommendation of food based on the sugar level of diabetes patients from a nutrition database.

The proposed work can be sub divided into four classes

1. Data pre-processing
2. Feature selection
3. Normalization of Data
4. Model Algorithm
5. Statistical tool for performance measurement

3.1. Data pre-processing

In the data, pre-processing the mechanism is responsible for all operations related to data cleaning and data balancing [19]. In this paper we used a nutrition dataset is having 5740 number of foods and 55 feature column, some of the values in the feature column are missed or duplicate value. To make an efficient and unbiased machine learning model, the dataset should not have missing or duplicate value in the row, will give a biased result. Missing value in the dataset was replaced with a mean value and deleted the duplicate row from the dataset.

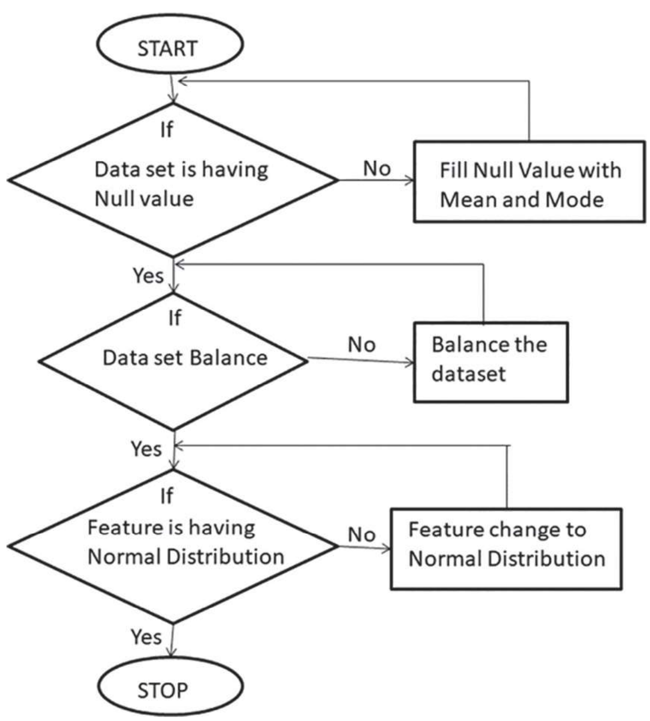

Figure 2: Flow chart for Data pre-processing mechanism

3.2. Feature Selection

To increase the efficiency of machine learning algorithm then collection of effective features and discard meaningless features from the dataset [20]. The given nutrition dataset has 5740 rows of foods and 55 feature columns. In this data set we identified some nutrition like carbohydrate, starch, protein, fat, calcium, iodine, iron, potassium and sugar which, cause diabetes in the human body as shown in fig 3. Several algorithms, including decision trees and random forest, were tested, but multiple linear regression achieved the most consistent performance for sugar content prediction. Data imbalance and missing values were handled through mean imputation and duplicate removal. For evaluation, the dataset was split into 80% training and 20% testing. Performance was measured using R^2, Mean Absolute Error (MAE), Mean Squared Error (MSE), Median Absolute Error, and Variance Score.

```
Data columns (total 10 columns):
 #   Column                                     Non-Null Count  Dtype
---  ------                                     --------------  -----
 0   FoodName                                   5740 non-null   object
 1   Protein(g)                                 5740 non-null   float64
 2   Totalfat(g)                                5740 non-null   float64
 3   Availablecarbohydrateswithsugaralcohols(g) 5740 non-null   float64
 4   Starch(g)                                  5740 non-null   float64
 5   Totalsugars(g)                             5740 non-null   float64
 6   Calcium(Ca)(mg)                            5740 non-null   int64
 7   Iodine(I)(g)                               5740 non-null   float64
 8   Iron(Fe)(mg)                               5740 non-null   float64
 9   Potassium(K)(mg)                           5740 non-null   int64
```

Figure 3: Dominated nutrition for diabetes mellitus

Table 1: which provide and summarized of mean, std. deviation, min, max etc., which is helpful for the development of machine learning model.

3.3. Normalization of data

In this method, all selected feature column values should be scaled in the same way to improve model accuracy [21]. The value of the feature column transformed into a standard normal distribution scaled. The following fig 4 shows the distribution of the feature column value.

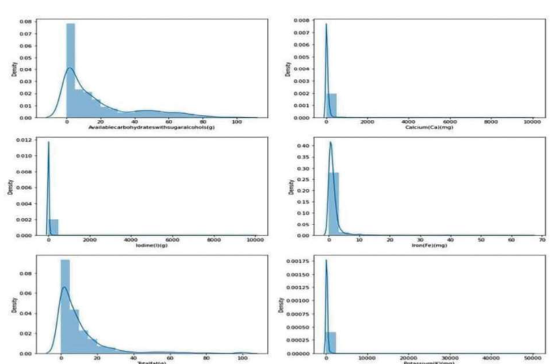

Figure 4: Distribution of feature data value

3.4. Algorithm for model

A. First Model–Sugar level of the Food

The model's feature selection is critical in determining the sugar level of the foods. The given dataset contains 50 features, of which eight are selected: carbohydrate, starch, protein, fat, calcium, iodine, iron, potassium, and one output, which is the sugar level of the food. The dataset is split into two parts: train and test, 80% of the data is used to train the model and 20% of the data is used to test the model. The model was developed using multiple linear regression machine learning algorithms based on the selected feature value. In this patient or user enter the numeric feature value and to get the sugar level of the food.

Mathematically represent the model as shown equation (1).

Sugar=α0+α1Carbohydrate+α2Starch+α3Protein+α4Fat+α5Calcium+ α6Iodine+α7Iro n+ α8potassium (1)

To get the maximum accuracy of the model coefficient value of the equation should be shown below table 1.

Coefficient	Value
α0	0.353346174337803
α1	8.86665296e-01
α2	-8.69471253e-01
α3	-1.37865278e-02
α4	1.19835962e-02
α5	6.56220117e-04
α6	-2.66105677e-06
α7	-4.73739754e-02
α8	3.40728006e-05

Table 1: Multiple linear regression coefficient value

In this model having multiple independent variable and variance influence factor (VIF) is used to measure the multi-collinearity of the independent variable. Multi- collinearity exists when there is a correlation between multiple independent variables in the above equation. Thus, the VIF can estimate how much the variance of a regression coefficient is influenced due to multi-collinearity, as shown below in Table 2.

Feature	VIF Score
Carbohydrate	2.6185
Starch	2.4501
Protein	1.1807
Fat	1.0093
Calcium	1.1092
Iodine	1.0015
Iron	1.2308
Potassium	1.0275

Table 2: VIF score table

B. Second model–To recommend the food to diabetes patient

The second model is used to recommend foods based on the sugar level of diabetes patients. The normal person's blood sugar level after fasting is 70 to 100 mg/dl, and the level after a meal should not go above 180 mg/dl. The best time to check the sugar level of any person is after 8 to 10hours of fasting or after waking up in the morning. The following flowchart shows the model working procedure.

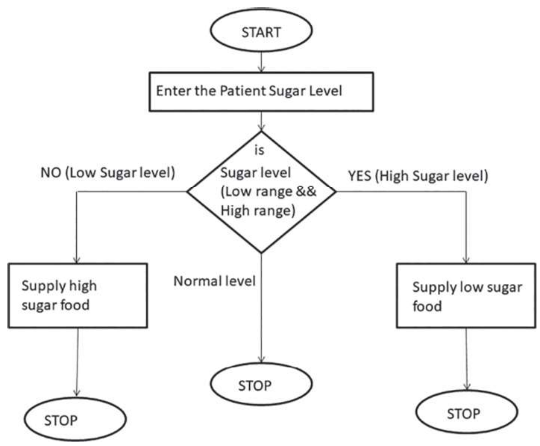

Figure 5 : Flow chart for food recommendation model

The model takes the diabetes patient's input sugar level, compares it to the normal sugar value, and recommends foods for high/low diabetes patients. The recommended foods tried to maintain the patient's normal sugar level for the whole day.

1.1. Statistical tools for performance measurement of algorithm

A. R^2 (R Square)

R square is a measure that provide information about the goodness of fit of a model. Mathematically represented in equation (1).

Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (1)$$

R-squared measure the strength of the relationship between your model and the dependent variable and its range from 0 to 100%.

B. Mean Absolute Error

This error basically is the absolute difference between true value and the value that are predicated. Mean absolute error takes the average of this error from individual sample in the dataset and gives the output. Mathematically represent in the equation (2).

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad (2)$$

C. Mean Squared Error–

The error is calculated by taking the average of the difference between the actual value and predicated value. It is mathematically represented in equation (3).

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (3)$$

It is also known as mean squared deviation and also represented mean squared residual.

D. Median Absolute error – The median absolute error is calculated by taking the median of all absolute difference between the true value and predicted value. It is showing in following equation (4).

$$\text{Median Absolute Error} = \text{median}\left(|y_i - \hat{y}_i|\right) \qquad (4)$$

E. Variance Score–

It is the difference between actual sample of the dataset and the predicated value of the model. It is used to find the variance by the machine learning model. It is represented in equation(5).

$$e_i = y_i - \hat{y}_i \qquad (5)$$

IV. Result and Discussion

In order to demonstrate how the sugar level of the food model works, we have built multiple linear regression machine learning models which give the sugar level of the food based on nutrition value. Table 3 shows the performance of a nutrition dataset with multiple linear regression machine learning models with respect to multiple research methods. The pre-processing methods improve the overall performance of the model. The reported accuracy of 97% refers specifically to the sugar content prediction in foods using the first model. In addition, the food recommendation model was validated by comparing its outputs against standard dietary guidelines for diabetic patients, showing strong alignment with medically accepted thresholds (fasting 70–100 mg/dl; postprandial <180 mg/dl).

Performance measure parameter	Value
R2 score	98 %
Mean absolute error	0.79
Mean squared error	2.77
Median absolute error	0.33
Variance score	0.98

Table 3: Performance measurement of model

The second model is used to recommend the food to the diabetes patient based on the sugar level of the patient. In this model, the patient will enter the sugar level into the model and the model uses a conditional loop based on sugar level and searches for the food from the nutrition dataset and recommended to the patient. The recommended food is helpful to the diabetes patient for maintaining the sugar level of the body for the whole day.

V. Conclusion and Future work

This paper proposed two models, the first model is used to find the sugar level of the unknown food. The model used multiple linear regression algorithms by considering various nutrients and calculating the sugar level of the food. This model is very helpful to the patient to predict the sugar level of the unknown food for the patient can easily choose the healthy food based on predicted sugar value. The second model is used for the recommendation of the food based on the sugar level of the patient. Furthermore, the proposed model will be deployed into the mobile application for easily predication of sugar level of the food and recommended the food as per diabetes patient sugar level value.

Limitations and Future Work:

Although the current model effectively predicts sugar content and provides dietary recommendations, personalization is currently limited to blood sugar values. Other important factors such as age, weight, BMI, activity level, and medication use are not yet incorporated. In the future, integration with continuous glucose monitoring (CGM) devices and mobile health applications will enable real-time feedback and adaptive recommendations. Additionally, incorporating larger clinical datasets will enhance the robustness and allow for broader clinical adoption.

Funding

This research received no external funding.

Acknowledgement

The authors would like to thank the Department of Computer Science with

Data Analytics, Kongunadu Arts and Science College, Coimbatore, for providing technical support and necessary facilities to carry out this research work.

Author Contributions

Conceptualization & Methodology: Dr. S. Velmurugan

Data Curation, Model Development & Experimentation: Dr. S. Velmurugan Supervision & Technical Support: Dr. R. Saravana Moorthy

Writing – Original Draft Preparation: Dr. S. Velmurugan Writing – Review & Editing: Dr. R. Saravana Moorthy

VI. Reference

[1] Roshan Birjais and Ashish Kumar Mourya, "Prediction and diagnosis of future diabetes risk: a machine learning approach, "*Springer Nature Switzerland*, vol.1, no.1112,August2019.

[2] Tabish SA, "Is Diabetes Becoming the Biggest Epidemic of the Twenty-first Century?," *Int JHealthSci (Qassim)*, July 2007.

[3] Saeedi P and Petersohn I, "Global and regional diabetes prevalence estimates for 2019 andprojections for 2030 and 2045: Results from the International Diabetes Federation DiabetesAtlas,9th edition,"*Elsevier B.V.*, vol.157, September2019.

[4] Osama Gheith and Nashwa Farouk, "Diabetic kidney disease: world wide difference ofprevalence," *JournalofNephropharmacology*,vol.5,pp.49-56,October2015.

[5] VatikaSibal, "Food: identity of culture and religion," *ResearchGate*, vol. 6, pp. 10908-10915, September2018.

[6] Kimber M Simmons and Aaron W Michels, "Type 1 diabetes: A predictable disease," *WorldJournalofDiabetes*,vol.6(3),pp.380-390, April2015.

[7] Lindy Kahanovitz, Patrick M. Sluss, and Steven J. Russell, "Type 1 Diabetes - A ClinicalPerspective,"*Pointcare*, pp. 37-40,March2017.

[8] Benito-Vicente A, Jebari S, Larrea-Sebal A, Siddiqi H, Uribe KB, Ostolaza H, Martín C. Galicia-GarciaU,"PathophysiologyofType2Diabetes Mellitus," *Int JMolSci.*, August 2020.

[9] Asif Mohammad, "The prevention and control the type-2 diabetes by changing lifestyle and dietary pattern,"*J Educ HealthPromot*,Feb2014.

[10] Xiang AH, Page KA Buchanan TA, "Gestational diabetes mellitus: risks and management during and after pregnancy," *NatRevEndocrinol*, pp.639-649,November2012.

[11] Adelaide A. Bampoe-Addo, Stephen K. Armoo, Firibu K. Saalia, Francis Gatsi, Baffour Sarkodie-Mensah Robert A. Sowah, "Design and

Development of Diabetes Management System UsingMachine Learning," *International Journal of Telemedicine and Application*, vol. 2020, p. 17, July2020.

[12] SalliahShafi Bhat and Gufran Ahmad Ansari, "Predictions of Diabetes and Diet Recommendation System for Diabetic Patients using Machine Learning Techniques," *International Conference for Emerging Technology*, pp.1-5, May2021.

[13] Kushan De Silva, Siew Lim, and Aya Mousa, "Nutritional markers of undiagnosed type 2 diabetesin adults: Findings of a machine learning analysis with external validation and benchmarking," *PLOS ONE*,p. 21, May2021.

[14] Hyerim Kim, Dong Hoon Lim, and Yoona Kim, "Classification and Prediction on the Effects of Nutritional Intake on Overweight/Obesity, Dyslipidemia, Hypertension and Type 2 Diabetes Mellitus Using Deep Learning Model: 4–7th Korea National Health and Nutrition Examination Survey," *International Journal of Environmental Research and Public Health*, vol. 18, no. 11, p.18,May2021.

[15] Yajie Guo, Zihua Huang, and Dan Sang, "The Role of Nutrition in the Prevention and Intervention of Type 2 Diabetes," *Frontiers in Bioengineering and Biotechnology*, vol. 8, p. 15, September2020.

[16] "Food Recommendation Using Machine Learning for Physical Activities in Patients with Type 1Diabetes," *17th Scandinavian Conference on Health Informatics*, pp.45-49, Nov2019.

[17] Raj Kumar B. and Latha K., "DFRS: Diet food recommendation system for diabetic patients based on ontology," *ternational Journal of Applied Engineering Research*, vol. 10, pp. 2765-2770,January2015.

[18] FaizI., Mukhtar H., and Khan S., "An integrated approach of diet and exercise recommendations for diabetes patients," *2014 IEEE 16th International Conference on e-HealthNetworking,Applications and Services (Healthcom)*, pp. 537- 542, 2014.

[19] Esra'aAlshdaifat, Doa'aAlshdaifat, and Ayoub Alsarhan, "The Effect of PreprocessingTechniques, Applied to Numeric Features, on Classification Algorithms' Performance," *MDPI*, p.23,January2021.

[20] Jie Cai, Jiawei Luo, and Shulin Wang, "Feature selection in machine learning: A newperspective,"*Neurocomputing*, vol. 300,pp.70-79,2018.

[21] Sevilla Joaquin, "Importance of input data normalization for the application of neural networks to complex industrial problems," *Nuclear Science, IEEE Transactions on*, vol. 44, pp. 1464-1468,July1997.