Nutrition 142 (2026) 112986

Contents lists available at ScienceDirect

Nutrition

journal homepage: www.nutritionjrnl.com

Applied nutritional investigation

Carbohydrate counting in traditional Turkish fast foods for individuals with type 1 diabetes: Can artificial intelligence models replace dietitians?

Volkan Özkaya [a,*], Erdal Eren [b], Şebnem Özgen Özkaya [a], Güven Özkaya [c]

[a] *Department of Nutrition and Dietetics, Kutahya Health Sciences University School of Health Sciences, Kütahya, Türkiye*
[b] *Department of Child Health and Diseases, Bursa Uludağ University Faculty of Medicine, Bursa, Türkiye*
[c] *Department of Biostatistics, Bursa Uludag University Faculty of Medicine, Bursa, Türkiye*

A R T I C L E I N F O

Article History:
Received 28 May 2025
Received in revised form 1 September 2025
Accepted 30 September 2025

Keywords:
Type 1 diabetes
Dietitian
Carbohydrate counting
Artificial intelligence
Turkish food

A B S T R A C T

Objectives: Carbohydrate counting is a recommended approach for achieving glycemic control in individuals with type 1 diabetes (T1D). This study aimed to compare the accuracy of carbohydrate content estimations for traditional Turkish fast foods made by artificial intelligence (AI) models and dietitian.
Methods: Children and adolescents with T1D were pretested to identify the 12 most preferred Turkish fast-food items. Standardized recipes were developed for these meals, and the meals were photographed under standardized angular and lighting conditions. The photos were then uploaded to AI applications (ChatGPT-4.0, DeepSeek, Gemini, and CarbManager) and each model was prompted to estimate the carbohydrate content of the respective food items. Dietitians were asked to estimate the carbohydrate content based on these photographs.
Results: Of the dietitians in the study (n = 40), 50% had postgraduate education, and 17.5% of those providing carbohydrate counting education (n = 20, 50.0%) had been doing so for more than 7 y. No significant difference was found between the carbohydrate estimates of dietitians who provided and those who did not provide carbohydrate counting training ($P > 0.05$). The intraclass correlation coefficient (ICC) between the AI models was 0.3554 (95% confidence interval [CI]: 0.0974–0.6801), indicating low reliability. The highest agreement with the estimates of dietitians who provided carbohydrate counting training (ICC = 0.417, 95% CI: 0.247–0.685) and those who did not (ICC = 0.307, 95% CI: 0.163–0.578) was observed with ChatGPT.
Conclusions: AI models can assist individuals with diabetes and healthcare professionals in estimating the carbohydrate content of foods, and consequently, can make a significant contribution to diabetes self-management.

Introduction

Type 1 diabetes mellitus (T1D) is a chronic autoimmune disease characterized by hyperglycemia, resulting from the autoimmune destruction of β-cells in the pancreas and subsequent insulin deficiency. Worldwide, 8.4 million people are living with T1D, and its prevalence is expected to increase to 13 to 17 million by 2040. T1D is frequently associated with a reduced quality of life, microvascular complications, and increased cardiovascular mortality [1–3]. To prevent these complications, postprandial glycemia should be maintained within appropriate target ranges. Glycemic control in T1D has been optimized through the use of insulin analogs, multiple daily injections (MDI), continuous subcutaneous insulin infusion (CSII), continuous glucose monitoring (CGM), and hybrid closed-loop (HCL) systems. Despite these technological advances, achieving postprandial glycemic control remains challenging for individuals with T1D, increasing the risk of diabetes-related complications. Many factors, including insulin dose, the timing of the insulin bolus, physical activity, meal timing and content, and carbohydrate quality can affect glycemic control. In particular, the amount of carbohydrate consumed remains the most important factor influencing postprandial glycemia [1,4,5].

Carbohydrate counting (CHOC) is a method of glycemic control based on balancing the amount of carbohydrates in a meal with the bolus insulin dose. It has long been recognized as a fundamental approach for achieving optimal glycemia in individuals with T1D undergoing intensive insulin therapy [4,6,7]. Carbohydrate

*Corresponding author. Tel.: +90 (274) 260 00 43.
E-mail address: volkan.ozkaya@ksbu.edu.tr (V. Özkaya).

https://doi.org/10.1016/j.nut.2025.112986
0899-9007/

counting is recommended as an ideal method for calculating meal-related insulin doses, as it provides individuals with T1D greater dietary flexibility and, in some cases, can reduce the burden of the disease [7,8]. Studies involving both children and adults have reported glycemic and lifestyle benefits of using carbohydrate counting as an intervention for people with diabetes. These benefits include lower glycosylated hemoglobin (HbA1c) levels, improved glycemic control, enhanced quality of life, better anthropometric outcomes, and more flexible eating [6–8].

Carbohydrates ingested with food are broken down into glucose and other monosaccharides, and approximately 90% of them are converted into glucose and enter the bloodstream within 1 to 2 h after consumption [9]. This transition time is influenced by the macronutrient distribution and quantity of the meal. For this purpose, various methods have been developed for carbohydrate counting, such as portion or exchange systems (10 g or 15 g carbohydrate units), gram-based carbohydrate estimation, and experiential learning [9,10]. However, the optimal method remains controversial. Therefore, the carbohydrate counting approach may vary depending on the cognitive ability of the individual with T1D, as well as the type and frequency of insulin used. Moreover, the skill and accuracy of carbohydrate counting affect postprandial glycemia and HbA1c levels, as well as the quality of life and the daily diabetes management burden of both the individual with T1D and their family [7,11]. Studies have reported that a carbohydrate calculation error of 10 to 15 g per meal does not significantly affect blood glucose levels, but an error of ±20 g per meal may lead to postprandial hypo- or hyperglycemia [7,9,12]. Additionally, studies have also shown that carbohydrate counting is challenging for both health professionals and individuals with diabetes, even when they are trained and experienced, and many people with diabetes make inaccurate counts. Accurately estimating portion sizes, reading food labels, weighing or measuring foods, and determining carbohydrate content require a high level of nutritional literacy, numeracy skills, and extensive nutritional knowledge. Therefore, the ability of individuals with T1D and their families to accurately measure carbohydrates determines the effectiveness of carbohydrate counting. This is why methods are needed to estimate the carbohydrate content and amount of food or meals as accurately as possible [13–15]. The continuous growth and development of smartphone technologies, software, mobile health apps, and artificial intelligence (AI) offer the opportunity for better diabetes control. Some apps designed for this purpose have been shown to reduce errors in carbohydrate counting estimation [5,16,17]. Computer-based tools outperform traditional methods in estimating the carbohydrate content of foods [18]. On the other hand, it has been reported that the biggest obstacle in predicting carbohydrate content is whether the source database used by AI or augmented reality applications includes local foods and beverages specific to the individual's country [5,15].

This study aims to compare the estimations of carbohydrate content in traditional Turkish fast foods made by AI tools and experienced dietitians. It is hypothesized that AI tools will show limited reliability in estimating the carbohydrate content of traditional Turkish fast foods compared to dietitians; however, the estimates of dietitians who provide carbohydrate counting education are expected to demonstrate higher agreement with AI predictions.

Methods

Study design and participants

Before the study, Ethics Committee Approval No. 2024-13/24, dated August 21, 2024, was obtained from the Bursa Uludağ University Faculty of Medicine Clinical Research Ethics Committee. The study was conducted between September and December 2024. A pretest was conducted to determine the most preferred traditional Turkish fast food items among individuals with T1D. Children and adolescents aged 7 to 18 y, diagnosed with T1D at least 1 y prior and followed up at the Pediatric Endocrinology Department of the Bursa Uludağ University Health Application Research Center, were included in the pretest. Before starting the study, an information form was provided to the parents of children with T1D for the pretest, and written informed consent was obtained from both parents and children. Participants with parental permission and individual consent were included in the study. The study was conducted in accordance with the provisions of the Declaration of Helsinki.

In the pretest, participants were asked to identify their five most preferred traditional Turkish fast foods. Based on their responses, 12 Turkish fast foods were selected and included in the study. The portion size (1 serving) of the foods identified in the pretest was determined using Standard Food Recipes [19]. Each food was weighed using a household scale (Medisana, model 37400). After deciding the standard portion sizes, the food items were analyzed using the Nutrition Information System (BeBiS 9.0), and the actual carbohydrate content was calculated. BeBiS 9.0 is a comprehensive and integrated food composition database that provides access to over 20 000 foods and more than 130 nutrients. Researchers, dietitians, and health professionals use it as an information resource.

After the standardized recipes were created, one portion of each food was photographed using a smartphone. A plain white standard serving plate was used during the photography (Fig. 1). A portable photo studio with fixed lighting (PDX Plus Studio Tent 60 × 60 cm, Color Temperature: 3200K / 5800K) was used to ensure standardization in the food photographs. During photography, as determined in previous studies, 90° and 75° angles were used to achieve optimal visualization of the food [20,21]. A standard resolution was set for each food, and six images were taken for each item.

Estimation of dietitians

The study was conducted with dietitians working in Türkiye who have at least a bachelor's degree in Nutrition and Dietetics, including those who actively teach carbohydrate counting (CHOC) education (Educational Dietitians – ED) and those who do not (Non-Educational Dietitians – nED). A questionnaire was sent to dietitians randomly selected from those who volunteered to participate in the study, and they were asked to estimate the carbohydrate amounts by evaluating food images. Dietitians were instructed to report a single carbohydrate value for each food image. The study was conducted in accordance with the Declaration of Helsinki, and informed consent was obtained from all dietitians.

Estimation of AI tools

ChatGPT-4.0, DeepSeek (DeepThinking R1), Carb Manager, and Gemini 2.0 Flash were used as AI tools to estimate the carbohydrate content of meals. All prompts were provided by a single researcher through one account. For this purpose, a new chat was initiated on ChatGPT-4.0, DeepSeek (DeepThinking R1), and Gemini 2.0 Flash, with all AI tools receiving the same prompt: "I have some photos of Turkish food that I want you to estimate the nutritional content of. Specifically, I would like you to estimate the total carbohydrates in the food shown in the attached photo. Please provide your best point estimate and do not give a range." After the prompt, photos taken from the best angle were uploaded for each food. This process was repeated for the photos of other foods. Only one estimation was requested for each photo, and the queries were completed in a single conversation. The process was repeated separately for all foods and AI tools, and the results

Fig. 1. Sample images of the prepared foods.

were recorded. However, it should be noted that the training data for the AI tools used in our study has not been made public by the developers, and it is therefore not possible to confirm whether traditional Turkish dishes were specifically included.

The Carb Manager application (Wombat Apps LLC) was downloaded free of charge from the Apple Store and Google Play. The researcher created an account, and the carbohydrate content of the food was determined by uploading the images prepared for the study and following the application instructions. Unlike the generative AI models used in this study, Carb Manager operates primarily by matching user-uploaded images to entries within its pre-existing food composition database, rather than relying on real-time computer vision analysis for autonomous nutrient estimation. Wombat Apps LLC did not support the project, and the research team covered the program membership fee.

Data analysis

Comparison of sociodemographic and professional characteristics

Independent samples t-tests were used to compare age and occupational duration between dietitians who provided Carbohydrate Counting (CHOC) Training and those who did not. For nonnormally distributed data, the Mann−Whitney *U* test was applied to assess differences in dietitians' estimates of carbohydrate content across various foods. Categorical variables, including gender and educational status, were analyzed using the Fisher−Freeman−Halton test and Fisher's exact chi-square test, as appropriate. To control for potential confounding effects, multivariate analyses using generalized linear modeling (GLM) was conducted to compare the agreement between dietitians with and without CHO counting education and AI model predictions, with adjustment for participant education level and years of professional experience.

Descriptive statistics are reported as mean ± standard deviation (SD) or median with interquartile range (IQR) for continuous variables and as frequency (*n*) and percentage (%) for categorical variables.

Evaluation of AI predictions versus reference measurements

A paired t-test assessed the difference between AI-predicted carbohydrate content and actual CHO measurements. Additionally, Bland-Altman plots were generated to evaluate agreement visually and systematic bias between AI predictions and reference measurements [22,23]. The 95% limits of agreement (LoA) for mean differences are presented in the plots.

Inter-rater reliability analysis

Intraclass correlation coefficients (ICCs) were computed using a two-way random-effects model (absolute agreement, single rater) to determine the agreement between:

1. AI predictions and dietitians who provided CHO Count Training ($n = 20$), and
2. AI predictions and dietitians who did not provide training ($n = 20$).

Separate analyses were performed for each AI method (total raters: 21, including the AI model).

Statistical Software

All analyses were performed using IBM SPSS Statistics (Version 29.0.2.0; IBM Corp., Armonk, NY, USA), except for Bland−Altman analyses, which were conducted in MedCalc Statistical Software (Version 23.2.1; MedCalc Software Ltd, Ostend, Belgium). The statistical significance level was set at $\alpha = 0.05$.

Results

The demographic characteristics of the dietitians are shown in Table 1. Ninety-five percent of the participants were female; mean age was 35.1 ± 7.3 y. Fifty percent of the dietitians held a bachelor's degree, 40.0% had a master's degree, and 10% had a PhD degree. There was no significant difference between ED and nED based on educational status. The professional experience of the dietitians was 12.0 ± 7.7 y. There was no statistical difference between ED and nED in terms of professional experience ($P > 0.05$). Regarding carbohydrate counting education, 10% of the dietitians provided education for 1 y or less, 10% for 3 to 5 y, and 17.5% for more than 7 y.

Table 2 compares of the carbohydrate content estimates of foods by dietitians who teach and do not teach CHOC education. No statistically significant difference was observed between the ED ($n = 20$) and nED ($n = 20$) groups in terms of the carbohydrate content estimates of foods ($P > 0.05$). However, in the Food7 estimation, the median value for dietitians in the ED group was 30 (27.5−40), while in the nED group, it was 35 (30−45), approaching the significance level cut-off ($P = 0.063$). After adjusting for education level and years of professional experience, no statistically significant differences were observed between dietitians with and without CHO counting education in their carbohydrate content estimates across all food items (all adjusted $P > 0.05$).

Figure 2 shows the actual carbohydrate values of the fast foods used in the study and the carbohydrate estimates made by the ED, nED, and AI tools.

The relationship between the actual carbohydrate content of the foods and the estimates made by AI models and Carb Manager is presented in Table 3. According to the findings, there was no statistically significant difference between the actual carbohydrate values and the AI estimations. The mean difference between ChatGPT estimates and actual measured values was −5.88 ($P = 0.322$), between DeepSeek estimates and actual values was 10.58 ($P = 0.070$), between Carb Manager estimates and actual values was 4.08 ($P = 0.425$), and between Gemini estimates and actual values was 10.99 ($P = 0.051$). These results indicate that the estimates made by the AI models are very close to the actual values and show no statistically significant difference. The ICC was calculated to measure the absolute agreement among the evaluations of four different AI models (ChatGPT, DeepSeek, CarbManager, and

Table 1
Demographic characteristics of the dietitians

		Provision of CHO counting education		
	Total	Yes ($n = 20$)	No ($n = 20$)	*P*-value
Age	35.1 ± 7.3	34.2 ± 7.0	36.1 ± 7.7	0.465
Gender				1.000
Female	38 (95.0)	19 (95.0)	19 (95.0)	
Male	2 (5.0)	1 (5.0)	1 (5.0)	
Education				1.000
Bachelor's degree	20 (50.0)	10 (50.0)	10 (50.0)	
Master's degree	16 (40.0)	8 (40.0)	8 (40.0)	
PhD degree	4 (10.0)	2 (10.0)	2 (10.0)	
Professional experience	12.0 ± 7.7	11.1 ± 7.5	13.0 ± 8.0	0.489
Duration of the provision of CHO counting education				
1 y or less	4 (10.0)	4 (10.0)	-	-
1−3 y	3 (7.5)	3 (7.5)	-	-
3−5 y	4 (10.0)	4 (10.0)	-	-
5−7 y	2 (5.0)	2 (5.0)	-	-
>7 y	7 (17.5)	7 (17.5)	-	-

Descriptive statistics were mean ± standard deviation, frequency (*n*), and percentage (%).

Table 2
Comparison of dietitians' estimates of carbohydrate content of foods

	Provision of CHO counting education		Measured carbohydrate values (g)	*P*-value	Adjusted *P*-value
	Yes (*n* = 20)	No (*n* = 20)			
Food1	40 (35–45)	30 (30–45)	67	0.231	0.885
Food2	55 (45–70)	82.5 (37.5–90)	61	0.231	0.948
Food3	50 (45–65)	47.5 (45–72.5)	42	0.738	0.170
Food4	60 (48.75–80)	50 (45–60)	60	0.102	0.954
Food5	60 (52.5–60)	60 (45–60)	35	0.461	0.738
Food6	40 (35–50.5)	50 (35–65)	34	0.341	0.469
Food7	30 (27.5–40)	35 (30–45)	55	0.063	0.082
Food8	30 (23.75–30)	30 (30–42.5)	77	0.174	0.684
Food9	30 (30–30)	30 (30–30)	30	0.799	0.571
Food10	60 (52.5–75)	60 (60–90)	40	0.678	0.979
Food11	60 (57.5–70)	60 (60–92.5)	80	0.289	0.586
Food12	60 (50–72.5)	60 (45–65)	45	0.495	0.617

Descriptive statistics were given as median (interquartile range). Adjusted p-values were calculated using GLM analysis, controlling for education level and years of professional experience as covariates.

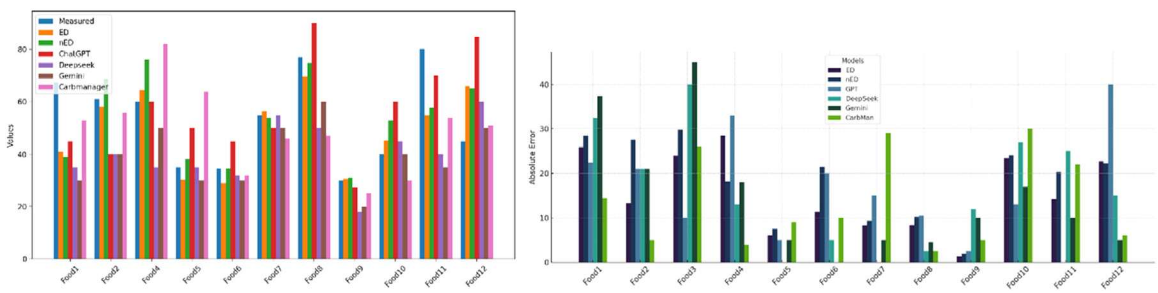

Fig. 2. Grouped bar chart of carbohydrate estimate values by dietitians and AI tools.

Table 3
The relationship between the actual carbohydrate content of the foods and the estimates made by AI models

Paired samples	Mean	Std. deviation	95% confidence interval of the difference		*t*	Effect size	*P*
			Lower	Upper			
Measured–ChatGPT	−5.88	19.66	−18.38	6.61	−1.037	0.278	0.322
Measured–DeepSeek	10.58	18.27	−1.03	22.18	2.005	0.538	0.070
Measured–CarbManager	4.08	17.05	−6.76	14.91	0.828	0.222	0.425
Measured–Gemini	10.99	17.38	−0.05	22.04	2.191	0.588	0.051

Gemini) on 12 different fast foods. The ICC value among the four methods was 0.3554, with a 95% confidence interval ranging from 0.09741 to 0.6801. Based on established interpretation guidelines, this ICC value corresponds to "fair" agreement (0.21–0.40), indicating that the reliability of the evaluators considered individually is low to moderate.

Figure 3 illustrates the agreement between the four different AI models and the actual CHO values using Bland–Altman plots. In terms of limits of agreement, the widest range of dispersion was observed in the ChatGPT model, with a 95% confidence interval ranging from −66.40 to 32.65. The CarbManager (−48.41 to −10.28), DeepSeek (−45.66 to −4.81), and Gemini (−42.51 to −3.65) models had narrower limits of agreement, indicating relatively less variability in their estimates.

Table 4 presents the results of the Bland–Altman analysis performed to compare the CarbManager, DeepSeek, Gemini, and Claude methods to the reference method, the actual CHO count. The intercept values in the regression analysis were examined, and the lowest deviation was observed in the ChatGPT model, with a value of −34.79. Considering the 95% confidence intervals, the CarbManager, DeepSeek, and Gemini models also showed substantial but comparable deviation levels. Slope values ranged between 0.55 and 0.83. It is important to note that the slope values were significantly different from 1 in all methods. In particular, there are significant differences in the p-values, the slope values are statistically significant in the DeepSeek (P = 0.003) and Gemini (P = 0.008) models. This suggests that these two methods exhibit a systematic difference and should be evaluated carefully in terms of concordance. On the other hand, there is no significant difference in the CarbManager (P = 0.055) and ChatGPT (P = 0.116) models, suggesting that these methods demonstrate a better fit in terms of averages. As a result, it can be concluded that there is a systematic difference between the DeepSeek and Gemini methods, leading to a lower agreement. In contrast, better agreement can be achieved with the CarbManager and ChatGPT methods, as the systematic difference is insignificant. The wide limits of agreement observed in the ChatGPT model indicate a higher variability in its estimations, which should be considered in clinical settings due to the potential for significant deviations from actual values.

The evaluation of dietitians' agreement with AI models is presented in Table 5. The fit of the ED and nED groups (n = 20

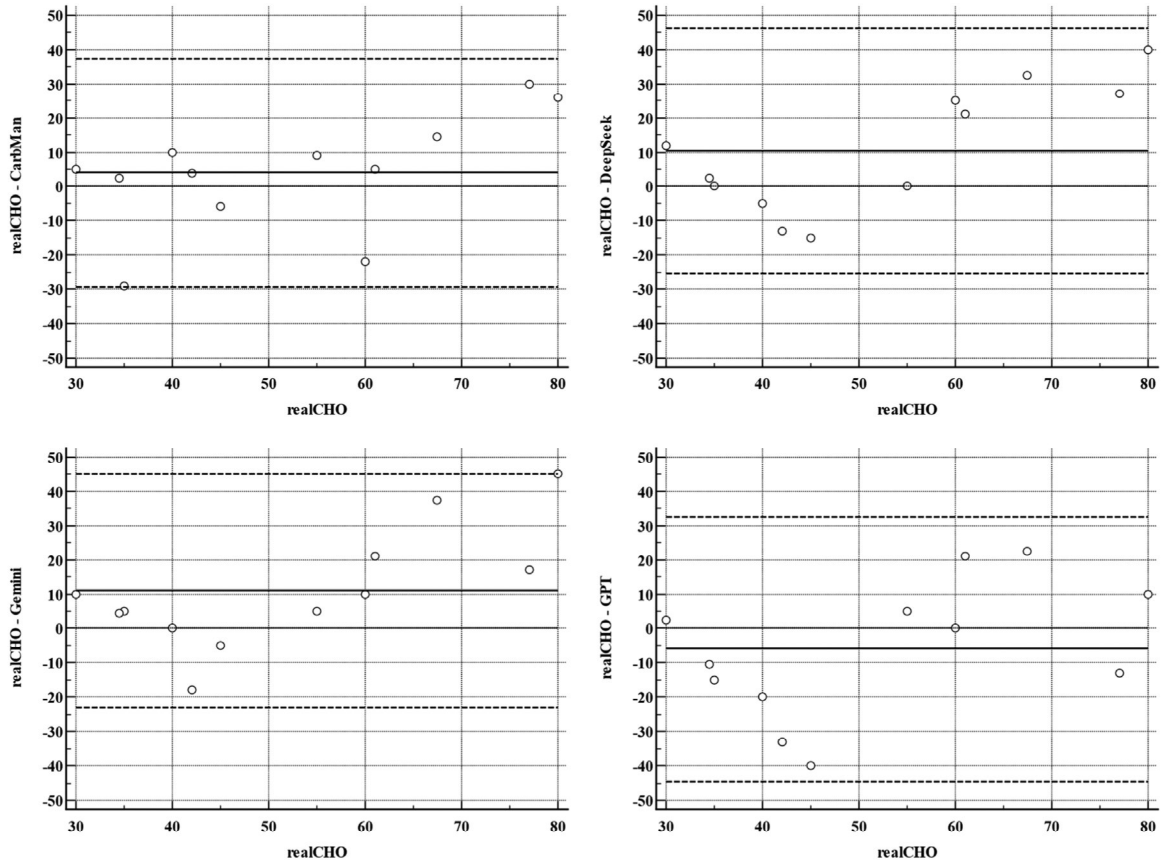

Fig. 3. Bland−Altman plot for interrater agreement analysis AI methods with the reference real CHO (n = 12).

Table 4
Bland−Altman analysis result

Variable		CarbMan	DeepSeek	Gemini	GPT
Limits of agreement	Lower limit	−29.35	−25.24	−23.08	−44.42
	95% CI	−48.41 to −10.28	−45.66 to −4.81	−42.51 to −3.65	−66.40 to −22.44
	Upper limit	37.50	46.39	45.06	32.65
	95% CI	18.43 to 56.56	25.96 to 66.81	25.63 to 64.49	10.67 to 54.63
Regression	Intercept	−25.66	−32.9013	−27.62	−34.79
	95% CI	−57.55 to 6.24	−59.19 to −6.61	−54.92 to −0.32	−74.00 to 4.42
	Slope	0.57	0.83	0.74	0.55
	95% CI	−0.01 to 1.15	0.35 to 1.31	0.24 to 1.24	−0.16 to 1.27
	P	0.055	0.003	0.008	0.116

participants each) with the four different AI models was analyzed using the ICC. Accordingly, ICC values were calculated separately for the two groups based on whether dietitians provided carbohydrate counting training. The ICC values for the ED group ranged between 0.400 and 0.417, while those for the nED group ranged between 0.297 and 0.307. Overall, the ED group demonstrated higher agreement with the AI models.

Table 5
The evaluation of dietitians' agreement with AI models

	Provision of CHO counting education	
	Yes (n = 20) ICC (95% CI)	No (n = 20) ICC (95% CI)
ChatGPT	0.417 (0.247−0.685)	0.307 (0.163−0.578)
DeepSeek	0.409 (0.240−0.677)	0.297 (0.156−0.567)
Gemini	0.415 (0.244−0.682)	0.303 (0.160−0.574)
CarbManager	0.400 (0.233−0.670)	0.300 (0.158−0.570)

Among dietitians who received CHO counting education (n = 20), the agreement with AI models showed fair to moderate reliability, with ICC values ranging from 0.400 to 0.417, indicating relatively consistent performance across all AI platforms. The highest agreement was observed with the ChatGPT model (ICC = 0.417, 95% CI: 0.247−0.685). For dietitians without CHO counting education (n = 20), the agreement with AI models demonstrated fair reliability, with ICC values ranging from 0.297 to 0.307, suggesting lower but still acceptable consistency compared to the trained group. The highest agreement was again found with the ChatGPT model (ICC = 0.307, 95% CI: 0.163−0.578). The lowest agreement was observed between the nED group and the DeepSeek model (ICC = 0.297, 95% CI: 0.156−0.567).

Discussion

In this study, the carbohydrate content estimates of 12 different traditional Turkish fast-food items were examined and compared based on the evaluations made by dietitians who actively provide or do not provide CHOC training, AI models, and the CarbManager application. Dietitians who provided CHOC training had higher ICC values for all AI models than those who did not, suggesting that carbohydrate counting training may enhance assessment accuracy when using AI tools. As a result, the agreement between AI models and dietitians was found to be low to moderate. However. dietitians who provide carbohydrate counting training showed higher agreement with AI models. This finding suggests that the education level of dietitians may be an important factor in AI-assisted nutritional assessments.

Carbohydrate counting is widely used by individuals with T1D to adjust insulin doses based on estimated carbohydrate content. CHOC contributes to better glycemic control and an improved quality of life. However, accurately estimating the carbohydrate content of meals is challenging due to individual variability and the complexity of the food matrix. Several factors, including experience, knowledge, health literacy, lifestyle, food type, and cooking methods influence the accuracy of individuals' carbohydrate estimates. Gurnani et al. found that among 140 adolescents with T1D who had experience with carbohydrate counting, 42% correctly estimated the carbohydrate content of sample meals, 44% estimated inaccurately, and 14% made significantly incorrect estimations [24]. These inaccurate estimates can be made not only by people with diabetes but also by health professionals [4,7,13]. Kawamura et al. evaluated the accuracy of carbohydrate content predictions in foods made by experienced and inexperienced health professionals (physicians and dietitians) in carbohydrate counting and reported that, although the experienced group made more accurate estimations than the inexperienced group, the difference was not statistically significant. The same study found that even participants with CHOC experience tended to underestimate the carbohydrate content in high-carbohydrate foods [13]. Similarly, Chotwanvirat et al. reported that the error rates in the closest, lowest, and highest carbohydrate estimates made by experienced registered dietitians through images of Thai dishes were 0.7%, 25.5%, and 7.6%, respectively [25]. We found no statistically significant difference between the carbohydrate content estimates of dietitians who actively provided CHOC training and those who did not. Emphasizing CHOC as an important component in diabetes management within dietitian education curricula in Türkiye and considering it as a standard component in diabetes management may have reduced individual differences in estimates. In addition, the emphasis on CHOC education in diabetes management guidelines and continuing professional development programs in Türkiye may have contributed to this consistency.

Studies to date have focused on evaluating the ability of AI models to identify food, estimate portion sizes, and determine nutrient content (energy and macro- and micronutrients) from food photos [15,17,26–28]. In particular, image recognition models developed for identifying food from photographs have been reported to be effective [29]. O'Hara et al. found that ChatGPT demonstrated good accuracy (93.0%) in correctly identifying foods in photographs compared to real meals. However, they reported that the accuracy decreased as the portion size of the food increased [27]. Similarly, Lo et al. found that GPT-4V was more successful in identifying food when information about the origin of the food was provided. They also found that the error in portion size estimation ranged from 1.1% to 85%. Energy, carbohydrate, protein, and fat were also shown to be less accurately estimated in meals where the portion size of the food was underestimated [26]. Lu et al. compared the accuracy of goFOOD and dietitians in predicting portion size, energy, and macronutrients using food visualizations. They found that the median absolute error in carbohydrate predictions was 27 g for dietitians and 7.2 g for goFOOD. Additionally, goFOOD demonstrated significantly better prediction performance compared to dietitians. They also reported that the correlations between model-generated predictions and actual values for energy and macronutrients ranged from 0.50 to 0.89 [21]. Chotwanvirat et al. compared the performance of a deep learning-based system with that of experienced dietitians in estimating the amount of carbohydrates from food images. They found that the prediction error of the system was 4%, while the prediction errors of dietitians ranged between 0.7% and 25.5%. Additionally, the system exhibited an estimation error below 10 grams in 13 of the 20 images analyzed. Based on these findings, they concluded that the developed system demonstrated satisfactory accuracy in estimating carbohydrate content [25]. In our study, on the other hand, we calculated the mean differences between the actual carbohydrate values and the estimates made by ChatGPT, DeepSeek, Gemini, and CarbManager as −5.88 (P = 0.322), 10.58 (P = 0.070), 10.99 (P = 0.051), and 4.08 (P = 0.425), respectively, and observed no statistically significant difference. However, the evaluation of the ICC among the four methods revealed low reliability. The broad limits of agreement observed in the ChatGPT model increase variability and suggest that caution should be exercised in clinical use. Examination of the agreement between dietitians and AI methods showed that the highest agreement for both ED and nED groups was with the ChatGPT model. In contrast, the lowest agreement was observed between the nED group and the DeepSeek model. The limited access of the AI models used in the study to direct food composition databases, combined with the diversity of nutrients in traditional Turkish foods (such as rice, vegetables, grains, bulgur, meat, phyllo, flour, yogurt, sauce, etc.) due to historical and cultural influences, may have impacted the accuracy of their estimations.

The strength of the study is that it represents the first comparison of the image-based estimations of dietitians who are actively teaching CHOC with those who are not, alongside AI models such as ChatGPT, DeepSeek, Gemini, and CarbManager. Additionally, the standardization of portion sizes and visuals of the sample meals also enhances the study. However, this study has several limitations. While the estimations from the AI models did not differ significantly from the actual carbohydrate values, their internal consistency was low. The limited number of sample meals and the complex nutritional content of Turkish fast foods make it challenging to generalize our findings. Furthermore, it should be noted that the training data for the AI tools used in our study has not been made public by the developers, and it is not possible to confirm whether traditional Turkish dishes were specifically included. This limitation may have impacted the accuracy of the estimations, particularly given the diversity of nutrients in traditional Turkish foods. The cultural specificity of our foods and the single-center design may also limit the generalizability of our findings. To validate these results and assess their external validity, multicenter studies involving foods from various cultures are required. A notable limitation is that a baseline assessment was not conducted regarding participants' frequency of technology use and digital literacy levels. Given that simulation-based training was employed in our intervention, learning outcomes and performance measures may have been influenced by individual differences in comfort and familiarity with digital technologies. This information could have provided valuable insights into participant characteristics that moderate training effectiveness and should be considered in future research designs.

Conclusions

This study compared the estimations of carbohydrate amounts in traditional Turkish fast food made by dietitians with and without active CHOC training and AI models. Although the estimations made by AI models are not significantly different from the actual carbohydrate values, their internal consistency is low. In terms of dietitians' agreement with AI methods, the highest agreement was observed with ChatGPT for both ED and nED, while the lowest agreement was found between nED and the DeepSeek model. AI tools may assist individuals with T1D practicing the CHOC method in reducing the burden of counting carbohydrates. The moderate agreement observed between AI tools and dietitians suggests a potential role for these applications as supplementary educational tools. They could be integrated into structured carbohydrate counting training programs to provide immediate feedback. However, they should not replace expert dietetic guidance due to their limitations with culturally complex foods. The underrepresentation of traditional foods in international databases remains a major challenge. To address this, developers should collaborate with local institutions to expand databases and incorporate user-driven content generation features for region-specific foods. The complexity of Turkish meals, in particular, makes it challenging to estimate carbohydrate content accurately. Therefore, caution should be when relying solely on AI for CHOC. The potential for AI systems to adapt to user inputs and instructions should also be considered. While this feature enhances human-computer interaction, it may also facilitate the spread of manipulative and persuasive disinformation.

Funding

The authors declared that this study received no financial support.

Data availability statement

The datasets presented in this article are not readily available due to restrictions (e.g., their containing information that could compromise the privacy of research participants). Requests to access the datasets should be directed to V¨

O, volkan.ozkaya@ksbu.edu.tr

Ethics approval and consent to participate

The studies were approved by Bursa Uludag University Faculty of Medicine Clinical Research Ethics Committee approved the study protocol with decision number 2024-13/24, dated 21.08.2024. The studies were conducted in accordance with the local legislation and institutional requirements. All participants provided informed consent, aligning with the Helsinki Declaration's principles.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Volkan Özkaya: Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Erdal Eren:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization. **Şebnem Özgen Özkaya:** Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Güven Özkaya:** Writing – review & editing, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Acknowledgments

We appreciate the participation of all the study's participants.

References

[1] Addissouky TA, Ali MMA, El Sayed IET, Wang Y. Type 1 diabetes mellitus: Retrospect and prospect. Bull Natl Res Centre 2024;48(1):42. https://doi.org/10.1186/s42269-024-01197-z.
[2] Gregory GA, Robinson TIG, Linklater SE, Wang F, Colagiuri S, de Beaufort C, et al. Global incidence, prevalence, and mortality of type 1 diabetes in 2021 with projection to 2040: a modelling study. Lancet Diabetes Endocrinol 2022;10(10):741–60. https://doi.org/10.1016/S2213-8587(22)00218-2.
[3] Lin X, Xu Y, Pan X, Xu J, Ding Y, Sun X, et al. Global, regional, and national burden and trend of diabetes in 195 countries and territories: an analysis from 1990 to 2025. Sci Rep 2020;10(1):14790. https://doi.org/10.1038/s41598-020-71908-9.
[4] Amorim D, Miranda F, Santos A, Graça L, Rodrigues J, Rocha M, et al. Assessing carbohydrate counting accuracy: current limitations and future directions. Nutrients 2024;16(14):2183. https://doi.org/10.3390/nu16142183.
[5] Tatulashvili S, Dreves B, Meyer L, Cosson E, Joubert M. Carbohydrate counting knowledge and ambulatory glucose profile in persons living with type 1 diabetes. Diabetes Res Clin Pract 2024;210:111592. https://doi.org/10.1016/j.diabres.2024.111592.
[6] AlBabtain SA, AlAfif NO, AlDisi D, AlZahrani SH. Manual and application-based carbohydrate counting and glycemic control in type 1 diabetes subjects: a narrative review. Healthcare (Basel) 2023;11(7):934. https://doi.org/10.3390/healthcare11070934.
[7] H Ibrahim SM, Shahat EA, Amer LA, Aljohani AK. The impact of using carbohydrate counting on managing diabetic patients: a review. Cureus 2023;15(11):e48998. https://doi.org/10.7759/cureus.48998.
[8] Wiyono L, Ghitha N, Clarisa D, Larasati A. Carbohydrate counting implementation on pediatric type 1 diabetes mellitus: systematic review and meta-analysis. Ann Pediatr Endocrinol Metab 2023;28(3):206–14. https://doi.org/10.6065/apem.2244242.121.
[9] Bell KJ, King BR, Shafat A, Smart CE. The relationship between carbohydrate and the mealtime insulin dose in type 1 diabetes. J Diabetes Complications 2015;29(8):1323–9. https://doi.org/10.1016/j.jdiacomp.2015.08.014.
[10] Brazeau AS, Mircescu H, Desjardins K, Leroux C, Strychar I, Ekoe JM, et al. Carbohydrate counting accuracy and blood glucose variability in adults with type 1 diabetes. Diabetes Res Clin Pract 2013;99(1):19–23. https://doi.org/10.1016/j.diabres.2012.10.024.
[11] Donzeau A, Bonnemaison E, Vautier V, Menut V, Houdon L, Bendelac N, et al. Effects of advanced carbohydrate counting on glucose control and quality of life in children with type 1 diabetes. Pediatr Diabetes 2020;21(7):1240–8. https://doi.org/10.1111/pedi.13076.
[12] Roversi C, Vettoretti M, Del Favero S, Facchinetti A, Sparacino G. Modeling carbohydrate counting error in type 1 diabetes management. Diabetes Technol Ther 2020;22(10):749–59. https://doi.org/10.1089/dia.2019.0502.
[13] Kawamura T, Takamura C, Hirose M, Hashimoto T, Higashide T, Kashihara Y, et al. The factors affecting on estimation of carbohydrate content of meals in carbohydrate counting. Clin Pediatr Endocrinol 2015;24(4):153–65. https://doi.org/10.1297/cpe.24.153.
[14] Stechova K, Hlubik J, Pithova P, Cikl P, Lhotska L. Comprehensive analysis of the real lifestyles of T1D patients for the purpose of designing a personalized counselor for prandial insulin dosing. Nutrients 2019;11(5):1148. https://doi.org/10.3390/nu11051148.
[15] Vasiloglou MF, Mougiakakou S, Aubry E, Bokelmann A, Fricker R, Gomes F, et al. A comparative study on carbohydrate estimation: GoCARB vs. dietitians. Nutrients 2018;10(6):741. https://doi.org/10.3390/nu10060741.
[16] Joubert M, Meyer L, Doriot A, Dreves B, Jeandidier N, Reznik Y. Prospective independent evaluation of the carbohydrate counting accuracy of two smartphone applications. Diabetes Ther 2021;12(7):1809–20. https://doi.org/10.1007/s13300-021-01082-2.
[17] Shehab M, Cohen RM, Brehm B, Bakas T. Accuracy and feasibility of using a smartphone application for carbohydrate counting versus traditional carbohydrate counting for adults with insulin-treated diabetes. J Diabetes Sci Technol 2024:19322968241248606. https://doi.org/10.1177/19322968241248606.

[18] Hegar K, Heiber S, Christ E. Carbohydrate counting of food. Swiss Med Wkly [Internet] 2011;141(2728):w13224. Available at: https://smw.ch/index.php/smw/article/view/1314.
[19] Rakıcıoğlu N, Tek AN, Ayaz A, Pekcan G. Photograph catalog of food and dishes: portion sizes and amounts. 3rd ed. Ankara: Ata Ofset Pub.; 2012.
[20] Dehais J, Anthimopoulos M, Shevchik S, Mougiakakou S. Two-view 3D reconstruction for food volume estimation. IEEE Trans. Multimed 2017;19:1090–9. https://doi.org/10.1109/TMM.2016.2642792.
[21] Lu Y, Stathopoulou T, Vasiloglou MF, Pinault LF, Kiley C, Spanakis EK, et al. goFOOD™: an artificial intelligence system for dietary assessment. Sensors (Basel) 2020;20(15):4283. https://doi.org/10.3390/s20154283.
[22] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476):307–10.
[23] Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. Lancet 1995;346(8982):1085–7. https://doi.org/10.1016/s0140-6736(95)91748-9.
[24] Gurnani M, Pais V, Cordeiro K, Steele S, Chen S, Hamilton JK. One potato, two potato,... assessing carbohydrate counting accuracy in adolescents with type 1 diabetes. Pediatr Diabetes 2018;19(7):1302–8. https://doi.org/10.1111/pedi.12717.
[25] Chotwanvirat P, Hnoohom N, Rojroongwasinkul N, Kriengsinyos W. Feasibility study of an automated carbohydrate estimation system using Thai food images in comparison with estimation by dietitians. Front Nutr 2021;8:732449. https://doi.org/10.3389/fnut.2021.732449.
[26] Lo FP, Qiu J, Wang Z, Chen J, Xiao B, Yuan W, et al. Dietary assessment with multimodal ChatGPT: a systematic analysis. IEEE J Biomed Health Inform 2024;28(12):7577–87. https://doi.org/10.1109/JBHI.2024.3417280.
[27] O'Hara C, Kent G, Flynn AC, Gibney ER, Timon CM. An evaluation of ChatGPT for nutrient content estimation from meal photographs. Nutrients 2025;17(4):607. https://doi.org/10.3390/nu17040607.
[28] Rhyner D, Loher H, Dehais J, Anthimopoulos M, Shevchik S, Botwey RH, et al. Carbohydrate estimation by a mobile phone-based system versus self-estimations of individuals with type 1 diabetes mellitus: a comparative study. J Med Internet Res 2016;18(5):e101. https://doi.org/10.2196/jmir.5567.
[29] Suddul G, Seguin JFL. A comparative study of deep learning methods for food classification with images. Food Humanit 2023;1:800–8. https://doi.org/10.1016/j.foohum.2023.07.018.