scientific reports

OPEN

Mealtime prediction using wearable insulin pump data to support diabetes management

Baiying Lu[1], Yanjun Cui[1], Prajakta Belsare[2], Catherine Stanger[3], Xia Zhou[4] & Temiloluwa Prioleau[1]✉

Many patients with diabetes struggle with post-meal high blood glucose due to missed or untimely meal-related insulin doses. To address this challenge, our research aims to: (1) study mealtime patterns in patients with type 1 diabetes using wearable insulin pump data, and (2) develop personalized models for predicting future mealtimes to support timely insulin dose administration. Using two independent datasets with over 45,000 meal logs from 82 patients with diabetes, we find that the majority of people (∼ 60%) have irregular and inconsistent mealtime patterns that change notably through the course of each day and across months in their own historical data. We also show the feasibility of predicting future mealtimes with personalized LSTM-based models that achieve an average F1 score of > 95% with less than 0.25 false positives per day. Our research lays the groundwork for developing a meal prediction system that can nudge patients with diabetes to administer bolus insulin doses *before* meal consumption to reduce the occurrence of post-meal high blood glucose.

Keywords Dietary monitoring, Diabetes, Insulin pump, Personalized modeling, Wearable medical device

Many patients with diabetes struggle with post-meal high blood glucose because of poor adherence to the daily task of administering meal-related insulin doses[1–7]. Sometimes patients skip or forget meal insulin doses (almost 40% of people with type 1 diabetes[8]), while other times they fail to administer this critical dose *at the right time* (i.e., ∼ 20 mins before meal consumption)[9,10]. The timing of meal insulin doses is critical to maintain post-meal glucose control because insulin taken “too early” or “too late” can lead to adverse blood glucose events (i.e., hypoglycemia or hyperglycemia)[9,10]. With less than 22% of patients with diabetes achieving the recommended glycemic target[8], there is immense potential for a meal prediction system that can *nudge* patients to administer the needed insulin dose *at the right time* to minimize the occurrence of post-meal high blood glucose. In fact, prior research has shown the potential to improve glycemic control (i.e., reduce hemoglobin A1C and increase time with blood glucose in the target range) through use of a smartwatch-based meal detection system that provides mealtime insulin reminders in an effort to reduce missed or late meal insulin doses[11].

Given the rise of advanced diabetes technology[8,12–14], our research investigates the potential of modeling dietary behavior using routinely collected data from wearable insulin pumps for the goal of predicting future mealtimes. As shown in Fig. 1, an insulin pump is a clinical-grade wearable medical device (FDA-approved) for administering insulin (i.e., the hormone needed to control blood glucose levels). The prevalence of insulin pump use is growing amongst people with diabetes and research shows that 65% of patients with type 1 diabetes already employ insulin pumps for daily management of their condition[8]. Despite the benefit of this innovative technology, a *key limitation* with current insulin delivery systems is that they all require users to manually log their meals in order to administer bolus doses of insulin needed to metabolize glucose from meals consumed[13,15]. This inherent requirement for insulin pump users to manually log their meals generates large amounts of food intake data (over months and years) that can be useful for modeling and understanding dietary behavior. However, this routinely collected data is significantly underutilized and rarely revisited, similar to other wearable device data from everyday diabetes technology[16–18].

Building on the state-of-the-art research on dietary monitoring and unmet needs in the diabetes domain[4,19–25], our research objectives are to: (1) investigate and model mealtime patterns using large volumes of meal logs from insulin pumps used by persons with type 1 diabetes, (2) develop and evaluate personalized models for

[1]Department of Computer Science, Dartmouth College, Hanover 03755, USA. [2]Integrated Science and Technology, James Madison University, Harrisonburg 22807, USA. [3]Center for Technology and Behavioral Health, Dartmouth College, Lebanon 03766, USA. [4]Department of Computer Science, Columbia University, New York 10027, USA. ✉email: tprioleau@dartmouth.edu

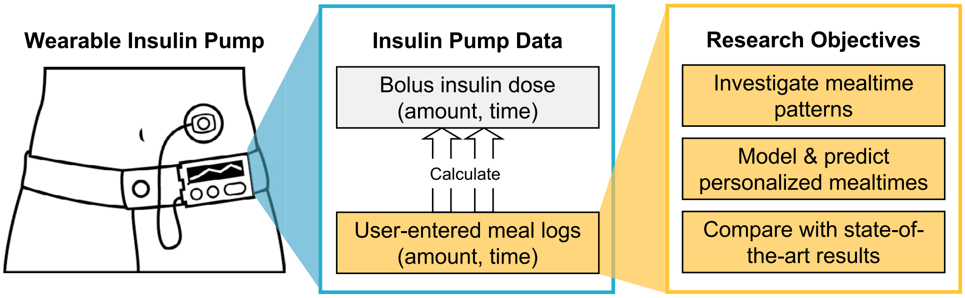


Fig. 1. Research overview. Insulin pumps used for daily management of diabetes are a rich source of data for understanding dietary patterns. This study leverages meal log data from wearable insulin pumps to study dietary patterns for the goal of modeling and predicting personalized mealtimes.

predicting future mealtimes based on population- and individual-level patterns learned from rich longitudinal meal log data from insulin pumps, and (3) compare the performance of personalized mealtime prediction from insulin pump data with results from related studies in the literature. Toward these objectives, we leverage two independent datasets with over 45,000 meal logs from more than 80 patients with diabetes to make the following innovative contributions:

1. We show the feasibility of learning dietary patterns and predicting future mealtimes using rich longitudinal meal log data from wearable insulin pumps. This contribution lays the groundwork for meal prediction systems that can nudge patients in real time to administer critical insulin doses *before* meals, and in so doing mitigate post-meal high blood glucose.
2. We characterize dietary patterns in a cohort of patients with diabetes and find that the majority ($\sim 60\%$) have irregular and inconsistent mealtime patterns. Consequently, a smaller percentage ($\sim 40\%$) of patients have regular mealtime routines with most of their meals occurring around the same time of day.
3. We implement, evaluate, and compare five data-driven methods for predicting mealtimes using historical meal logs from wearable insulin pumps. Our results show that future mealtimes can be predicted with an average F1 score of 95.33% and false positives per day (FP/day) of 0.25 using personalized 1-layer LSTM models. This result is significantly better (>70%) than what was achieved with three baseline methods and one ARIMA-based method.
4. We compare our research and results to related work in the literature[26–30] and show superior performance for the task of mealtime prediction with the added benefit of using *real patient* data as opposed to simulated data or data from a healthy population.

Results

Study and data description

Our research leverages retrospective wearable data from two independent cohorts with total of 88 patients with diabetes - see Table 3 in the Methods section. Dataset 1, also known as DiaTrend[31], is open-source and includes 54 subjects with type 1 diabetes (age = 33.6 ± 15.6 yrs.). This dataset includes 8220 days of insulin pump data with 23,678 meal logs (associated with > 0 g of carbs) across all 54 subjects (mean: 438 meal logs per subject). The DiaTrend dataset was collected by our research team through the process of recruiting patients with type 1 diabetes from a local hospital and online platforms, and downloading their retrospective diabetes device data for research. Our study was approved by the Committee for Protection of Human Subjects at Dartmouth College and all subjects provided informed consent. Conversely, dataset 2 comprises 34 subjects with type 1 diabetes (age = 42.3 ± 8.6 yrs.). Dataset 2 includes 5392 days of insulin pump data with 21,519 meal logs (associated with > 0g of carbs) across all 34 subjects (mean: 632 meal logs per subject). This dataset was collected by and licensed from Tidepool[32]—a non-profit diabetes organization. Across both datasets, there is an average of 3.5–4 meal logs per day and a range of 1–19 meal logs per day. In addition, the average quantity of each meal log is 40 ± 24.7 grams and 37 ± 24.9 grams in datasets 1 & 2, respectively.

Informed by prior research on the prevalence of missed bolus insulin doses and missed meal logs amongst patients with diabetes[4,7], we excluded subjects who had less than 15 days of data with at least 2 meal logs per day and days of data with less than 2 meal logs per day—see Supplementary Fig. S1. These criteria led to the exclusion of 6 subjects, namely, subjects 2, 26, 27, 29, 32, and 35, all from dataset 1. Additionally, the above criteria led to the exclusion of 2623 out of 8220 days (31.9%) from dataset 1 and 449 out of 5392 days (8.3%) from dataset 2. Following these exclusions, dataset 1 comprises 5597 days of insulin pump data with > 2 meal logs per day (total: 22,322 meal logs from 48 subjects in dataset 1), meanwhile, dataset 2 comprises 4943 days of insulin pump data with > 2 meal logs per day (total: 21,274 meal logs from 34 subjects in dataset 2). In total, our final dataset includes 10,497 days of insulin pump data with 43,596 meal logs from 82 subjects with type 1 diabetes.

Exploratory analysis on daily meal and mealtime patterns

We examined daily meal and mealtime patterns across our dataset from patients with type 1 diabetes to enable comparison between our findings and those from prior work[23,33–37]. First, we investigated the population-level and individual-level distribution of mealtimes (based on the timestamp of each meal log which represents both meals and snacks in insulin pump data) versus time of day. It is important to note that meal logs from the insulin pump and those used in this study do not include carbohydrate supplements (e.g., carbohydrate intake to prevent hypoglycemia during exercise[38,39]) as these are typically not reported in the insulin pump because they do not require an accompanying insulin dose. Fig. 2a shows the population-level distribution of meal times summarized in 30-min blocks across the 24-h period of each day. Meanwhile, Supplementary Fig. S1b shows the individual-level distribution for each subject where the size of each bubble represents the normalized frequency of meal logs for each subject across various times of the day. From these figures, we observe notably irregular and inconsistent mealtime patterns that differ significantly from the standard breakfast–lunch–dinner pattern that might be expected. However, the population-level distribution of mealtimes shown in Fig. 2a highlights two prominent peaks, one at lunchtime around 12:00 PM and the other at dinner time around 6:30 PM. Based on our data in this study, the observed distribution was similar across weekdays and weekends, and also across U.S. holidays and non-holidays—See Supplementary Figs. S2 and S3. This finding is similar to what was found amongst adults without diabetes[34] and also in those with type 1 diabetes albeit with meals at different times of the day based on country norms[37]. However, in our dataset, meals around what might be considered the standard lunchtime and dinnertime represent around 4% per 30 mins block (or 8% per hour) of the total meal events. No subject *only* had meal logs in the standard breakfast–lunch–dinner hours of 7–9 AM for breakfast, 11 AM–1 PM for lunch and 6 PM–8 PM for dinner[33], and in total *only* 46.63% of meal logs were within these time periods.

From the individual-distribution of mealtimes shown in Supplementary Fig. S1, we observe that fewer subjects (39%) showed regular mealtime routines (i.e., majority of their meals happen around the same time of day), while more subjects (61%) showed irregular mealtime routines which varied significantly across hours of the day. For example, subjects 57, 58, 59, & 60 show more regular mealtime patterns on an individual-level with prominent but distinct mealtimes indicated by larger-sized bubbles. Meanwhile, many other subjects (e.g., subjects 1, 30, 45, & 56) show irregular and highly variable mealtime patterns with no prominent mealtimes. Nonetheless, as might be expected, we observe more meals during the wakeful hours (i.e., 93% of the total meal logs between 6 AM–12 Midnight) and fewer meals during the nighttime (i.e., 7% of the total meal logs between 12 AM–6 AM). Overall, the individual mealtime patterns observed in our cohort of patients with diabetes shows more variability and less regularity than what has been reported in prior work with subjects without diabetes[33].

To assess the number of meals/day, we combined meal logs that were close together (i.e., less than 30 mins apart) into one "meal" recorded at the first meal's timestamp. Fig. 2b shows that only 25% of days had exactly 3 meals, and the number of meals/day varied widely in our population of patients with type 1 diabetes, ranging from 1.81 meals/day for the 10th percentile to 6.12 meals/day for the 90th percentile. Overall, 48% of days in

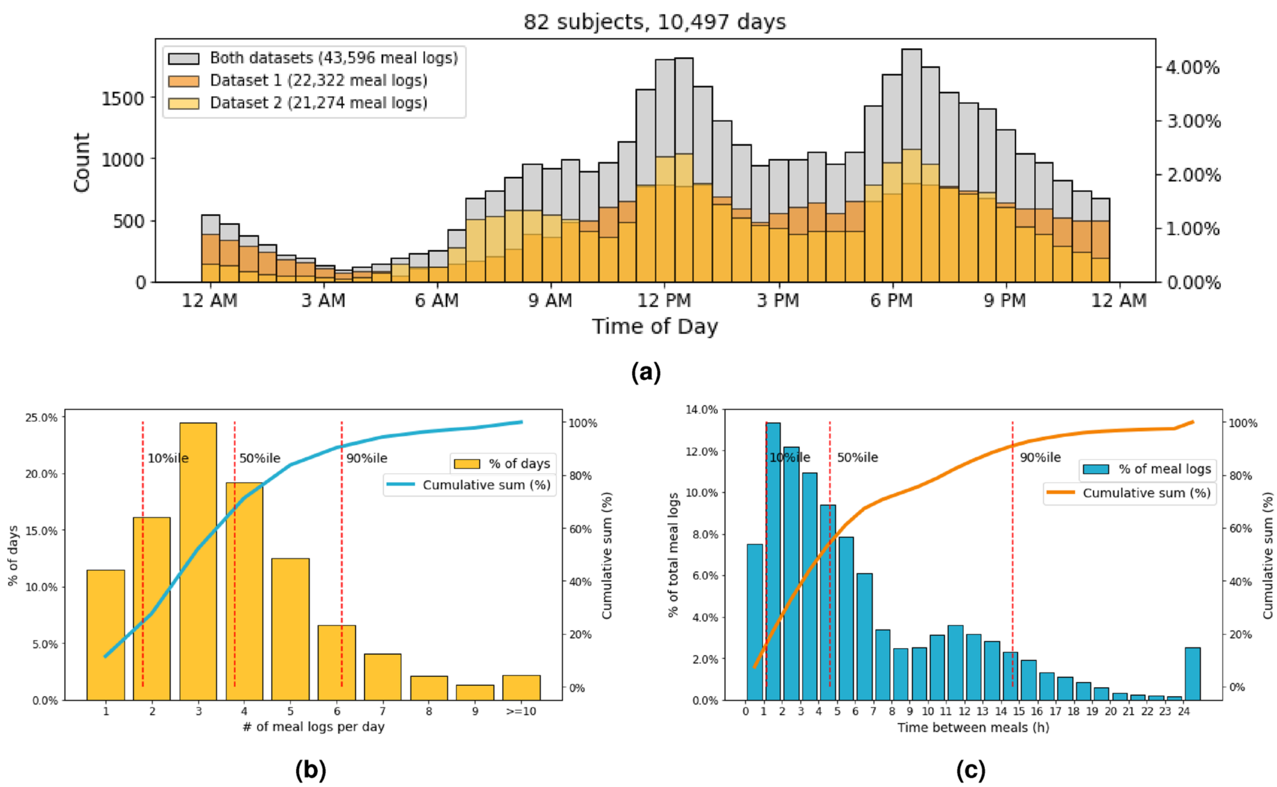


Fig. 2. Overview of mealtime patterns observed in 10,497 days of insulin pump data from 82 subjects with type 1 diabetes. (**a**) The mealtime distributions across all subjects show two prominent peak mealtimes, one at 12:00 PM and another at 6:30 PM. (**b**) A distribution plot showing the number of meal logs per day. (**c**) A distribution plot showing the time between meals.

our dataset correspond to days with 4 or more meals per day. We also assessed the time between meals at the population-level as shown in Fig. 2c. We found that 21% of all meals were within 2 h of another meal, and the median inter-meal interval was 4 hrs and 36 mins. These exploratory findings suggest that mealtime patterns are equally or sometimes even more variable in a population of patients with diabetes compared to what has been found in prior research on a population without diabetes[33,34]. Given that people with diabetes are prone to low blood glucose events which requires food intake to bring the blood glucose back to the normal/target range, it is reasonable that this population may have more eating events and/or more variable eating patterns than a population without diabetes[37].

Modeling and predicting mealtimes using wearable insulin pump data

Toward the goal of mealtime prediction, we implemented and evaluated five data-driven methods, including three baselines (i.e., standard-mealtime, prior-day, and peak-mealtime), one statistical method (i.e., SARIMA), and one deep learning based method (i.e., LSTM). The objective of all methods was to predict mealtimes ($M_{i,j}$) for subject i on day j using any length of historical data. The standard-mealtime baseline evaluates the assumption that most people eat 3 meals per day and informed by prior work on healthy subjects[33], it uses fixed times at 8:30 AM, 12:30 AM, and 6:30 PM for prediction. Meanwhile, the prior-day baseline builds on the premise that different people have different routines and yesterday's data from a specific individual could be useful for predicting today's behavior. Thus, a reasonable mealtime prediction for subject i on day j is their own mealtimes on the prior day $j-1$. Finally, the peak-mealtime baseline uses the k most frequent mealtimes from n prior days of subject's i data to predict their mealtimes on the next day (i.e., day $n+1$). Contrary to the aforementioned baseline methods and toward the goal of a meal prediction system that can nudge patients with diabetes to administer the needed insulin doses at the right time (i.e., ~ 20 mins before each meal[9]), we also formulate the mealtime problem as a task of predicting the probability of a meal 30 mins in advance of the current time t. This formulation enables the use of well-known statistical and deep learning based methods for time-series forecasting, such as autoregressive integrated moving average (ARIMA)[40] and long short-term memory model (LSTM)[41]. In addition, this formulation supports the use of insights from the earlier times in each day (e.g., time of last meal) to achieve a better prediction of when the next meal might occur.

Table 1 shows the average recall, precision, FP/day, FN/day, and F1 score for each method across 82 subjects in this study. From this table, we observe that the F1 score was between 10.84 to 23.66% for all three baseline methods and seasonal ARIMA (SARIMA), with the peak-mealtime method having the highest baseline performance due to a higher recall of 44.26% but relatively low precision of 17.33%. Comparatively, the personalized 1-layer LSTM performed significantly higher by over 70% and showed the best performance results for the task of mealtime prediction with an average F1 score of 95.33% with 0.25 FP/day. Furthermore, we show a box-plot comparison in Fig. 3 to assess the within-method variability in performance across our subject cohort. From this figure, we observe similar variability with an average interquartile range of 8.23% for all three baseline methods and SARIMA. Conversely, we observe notably less variability with an interquartile range of 2.03% with the 1-layer LSTM for the task of mealtime prediction. These findings show that LSTM achieved more stable results across subjects which is more desirable. However, it is worth noting that even with LSTM lower performance was observed amongst some subjects as shown by the outliers in Fig. 3a. More specifically, the minimum F1-score was 57.14% for subject 16. To further understand our results, we examined the insulin pump data from subjects with lower performance (i.e., < 70%) on the task of mealtime prediction and found that most of these subjects had large periods of missing data (i.e., no meal log for many consecutive days) within their insulin pump data (e.g., subjects 4, 16, 21, and 83). For example, the F1-score for subject 83 was 64.7%, however this subject had 186 days of insulin pump data with only 37 valid days (i.e., days with more than 2 meal logs per day).

Understanding how much training data is needed for mealtime prediction

Building on the above results which show that the 1-layer LSTM as described outperforms comparative methods for the task of mealtime prediction (i.e., predicting meals before they occur), we sought to investigate the effect of varying amounts of training data (i.e., number of days with past meal logs) on this performance. Toward this task, we trained personalized 1-layer LSTM models with 32 units and 5 epochs with varying amounts of training data ranging from 3 to 90 days of past meal log data and evaluated the performance on the same test dataset (i.e., the last 20% of each subject's data). As outlined in Supplementary Table S1, the sequence length of each input window used for training the LSTM models was 96 (i.e., using the past 48-h of data). Therefore, when training with only 3

	Method	Recall (%)	Precision (%)	FP/day	FN/day	F1 score (%)
Baseline	Standard-mealtime	10.21	11.99	2.64	3.14	10.84
	Prior-day	14.62	14.65	2.94	2.95	14.63
	Peak-mealtime	44.26	17.33	8.46	1.92	23.66
Rolling Forecast	SARIMA	35.95	10.38	11.23	2.18	16.00
	1-layer LSTM	**97.72**	**93.51**	**0.25**	**0.07**	**95.33**

Table 1. Performance comparison across methods for mealtime prediction. These results show the average of each performance metric on the held-out test dataset from all 82 subjects with type 1 diabetes. The best performance for each metric is in bold. The best possible value for recall, precision, and F1 is 100%. The best possible value for FP/day and FN/day is 0.

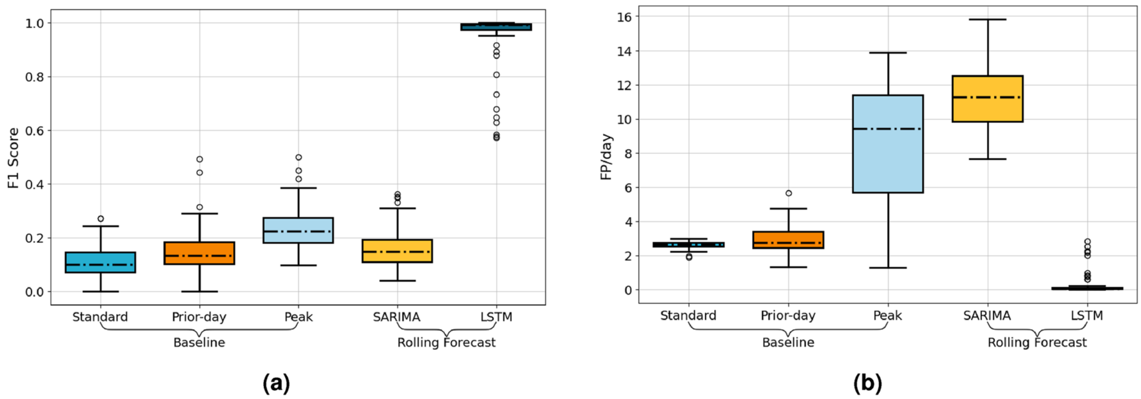


Fig. 3. Overview of the within-method variability in performance across subjects based on the held-out test dataset. (**a**) The F1 score obtained with each method for the task of predicting future mealtimes (the optimum F1 score is 1). (**b**) False positives per day obtained with each method for the task of predicting future mealtimes (the optimum FP/day is 0).

days of data (i.e., 144 observations) and sliding the window one step for each iteration as shown in Fig. 6b, there will be a total of 47 iterations (i.e., 144 observations - 96 input features - 1 output), thus yielding a training matrix of 96 input features x 47 iterations to effectively train the proposed LSTM model. Fig. 4 shows the result of our analysis with varying amounts of training data. As expected, we observe better performance with more training data as evident from the increasing F1 scores and decreasing variability in performance across subjects. More specifically, our result shows that 45 days of training data is sufficient to achieve an average F1 score of 91.15% with an interquartile range of 7.72%, while 60 days of training data is sufficient to achieve an average F1 score of 93.75% with an interquartile range of 3.60%. It is also important to note that the F1 score plateaus after 60 days thereby suggesting that 60 days of training data is optimal for the task of mealtime prediction.

Performance comparison across subject subgroups

The final part of our analysis focused on comparing the performance of all methods described in this work across subject subgroups for the task of mealtime prediction. First, we compared the performance of each method across datasets, i.e., dataset 1 (DiaTrend[31]) and dataset 2 (licensed from Tidepool[32]). Fig. 5a presents the result from this analysis. From this figure, we observe that all five methods demonstrated better performance with a higher F1 score that was on average 4.5% higher on dataset 2 compared to dataset 1. Upon further examination, we observed that dataset 2 has more meal logs/day, more meal logs/subject, and the total meal logs across 34 subjects in dataset 2 is only about 1000 less than the total meal logs across 54 subjects in dataset 1 in spite of there being a similar number of days/subject—Table 3. These observed differences show that dataset 2 has less missing data and consists of more subjects who are adherent to logging meals into their insulin pumps to administer the needed doses of bolus insulin compared to subjects in dataset 1 (DiaTrend[42]). This finding suggests that the ability to model dietary patterns and predict future mealtimes using meal log data from wearable insulin pumps is influenced by the quality and density of the initial dataset. It is expected and intuitive that the performance of mealtime prediction will be worse when a dataset includes more missing data.

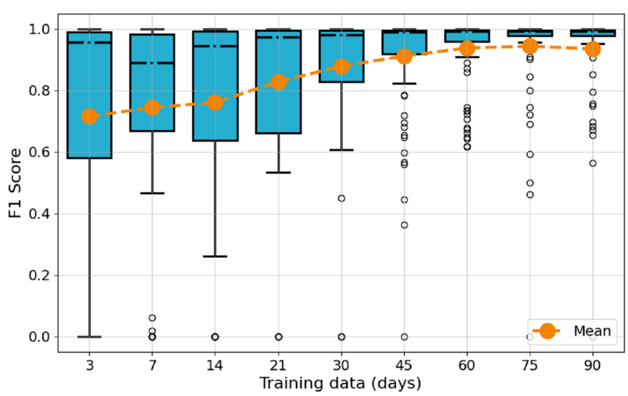
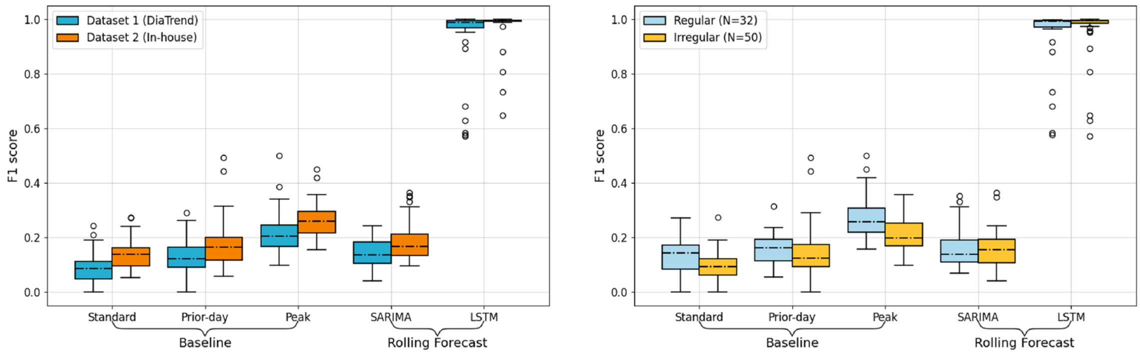


Fig. 4. The effect of varying amounts of training data on a 1-layer LSTM for mealtime prediction and evaluated on the held-out test dataset. We observe that the performance increases with increasing amounts of training data. However, at 60 days, the performance plateaus thereby showing that 60 days of meal log data from wearable insulin pumps is sufficient for learning and predicting future mealtimes.



(a) (b)

Fig. 5. Comparison of mealtime prediction performance across subject subgroups. (**a**) Performance comparison across two datasets (i.e., the open-source Dataset 1 (DiaTrend)[31] versus Dataset 2 (licensed). (**b**) Performance comparison across subjects with regular versus irregular mealtime routines.

In addition to the above analysis, we also compared the performance of each method across subjects with varying mealtime routines. From our exploratory analysis of daily meals and mealtime patterns, we observed notably varying mealtime patterns across patients with type 1 diabetes as shown in Supplementary Fig. S1. However, we also observed that some subjects have more regular mealtime routines (i.e., the majority of their meals happen around the same time of day), such as subjects 59 and 60, while others have more irregular mealtime routines such as subjects 1 and 30 as shown in Supplementary Fig. S1. Building on this insight and empirical observations for our dataset, we classified subjects into two subgroups based on their mealtime patterns (i.e., regular vs. irregular mealtime routine). To achieve this, we defined four intuitive features, namely, *f*1 - the number of prominent peaks (prominence > 0.3), *f*2 - the number of minor peaks (prominence > 0.05), *f*3 - area under the distribution curve, and *f*4 - the average time between peaks, all of which were calculated from the mealtime distribution plots for each subject. We manually selected five example subjects whose mealtime distributions clearly showed regular and irregular mealtime routines, respectively. The five subjects selected as having regular mealtime routines are subjects 57, 58, 59, 60, 65, while the five subjects selected as having irregular mealtime routines are subjects 56, 64, 66, 74, 80—see Supplementary Fig. S1 for reference. Then, data from these subjects was used to train a logistic regression classifier for the binary task of classifying the remaining subjects into one of these two subgroups (i.e., regular or irregular mealtime routine). From this analysis, we found that only 32 (or 39%) had regular mealtime routines, while 50 subjects (or 61%) had irregular mealtime routines.

Building on this stratification, we evaluated the aforementioned five data-driven methods for the task of mealtime prediction. Fig. 5b shows the performance of each method for subjects in each subgroup. From this figure, we observe an average of $\sim$ 3% difference in mealtime prediction performance between subjects in the regular routine subgroup versus in the irregular routine subgroup for all baseline methods. For example, with the prior-day baseline, the average F1 score for mealtime prediction is 16% for subjects in the regular routine subgroup and 14% for subjects in the irregular routine subgroup. This difference in performance across both subgroups was not as prominent as in the rolling forecast methods. For example, with the 1-layer LSTM, the average F1 score was 94% and 96% for subjects in the regular and irregular routine subgroups, respectively. Findings from this analysis show that mealtime prediction is generally more accurate for subjects with regular routines compared to subjects with irregular routines, however, this difference is not observed with the LSTM model.

Discussion

Our research work in this paper shows the feasibility of learning dietary patterns and predicting future mealtimes using rich longitudinal meal log data from wearable insulin pumps. This study builds on prior work in literature centered around increasing the utility of large-scale digital data from advanced diabetes technology[16,18,43–47]. In relation to prior research on sensor-based methods for dietary monitoring[11,19–21,48,49], there are several unique strengths to the work presented here. First, this study leverages routinely collected data from wearable insulin pumps already used by patients for daily management of diabetes. Since our datasets comprise of retrospective and observational device data, our research has the advantage of understanding dietary behaviors in the natural environment without the potential bias that comes from enrolling participants in a prospective research study. Second, this study combines two independent datasets with a total of over 45,000 meal logs for developing and evaluating data-driven models for mealtime prediction. Our dataset is orders of magnitude larger than what can be found in related work in this field[33,34,36,37]. Third, this study uses real data from patients with diabetes for modeling and prediction as opposed to data from healthy subjects or simulated data from computer software[26–29,33,34]. Additionally, we show that for the task of mealtime prediction, personalized 1-layer LSTM models can achieve an average F1 score of 95.33% and FP/day of 0.25 when evaluated on two independent datasets, including a public dataset and a distinct licensed dataset (total: 82 subjects, 10,540 days, 43,596 meal logs). This research is critical toward the development of a meal prediction system that can nudge patients with diabetes to administer critical insulin doses *before* meals, and in so doing mitigate post-meal high blood glucose.

Study	Task	Data source	# Subjects	# Meals	Recall	Precision	FP/day	FN/day	F1 score
2009[26]	Detection	Simulated	200	800	82%	92%	0.2	-	90%
2016[27]	Detection	Simulated	30	180	76%	84%	-	-	-
2018[28]	Detection	Simulated	10	15,000	82%	92%	-	-	90%
2020[29]	Detection	Simulated	100	15,900	81.3%	-	0.15	0.02	-
2022[30]	Detection	Real Patients	50	13257	92.3%	96.17%	-	-	93.97%
This work	Prediction	Real Patients	82	43,596	97.72%	93.51%	0.25	0.07	95.33%

Table 2. Comparison with related work on mealtime detection or prediction. Dashes ('-') are used when a certain metric was not reported in prior work.

		Dataset 1 (DiaTrend[31])		Dataset 2 (In-house)	
	Attributes	**Mean ± SD**	**Range**	**Mean ± SD**	**Range**
Demographics	# Subjects (# excluded)	54 (6)	–	34 (0)	–
	Age (years)	33.6 ± 15.6	19–70	39.8 ± 8.7	24–52
	Time since diagnosis (years)	18.5 ± 12.4	2–56	18.4 ± 10.6	2–48
Insulin pump data	Total number of days	8220	–	5392	–
	Number of days / subject	152.2 ± 151.2	31–780	158.6 ± 40.2	98-268
	Total meal logs (# excluded)	23,678 (1356)	–	21,519 (245)	–
	Meal logs / subject	438.5 ± 479.7	1–2310	632.9 ± 223.0	144–1039
	Meal logs / day	3.5 ± 2.2	1–19	4.1 ± 2.1	1–19
	Quantity logged / meal (grams)	40.0 ± 24.7	0.5–666	37.0 ± 24.9	1–240
	Excluded days w/ < 2 meal logs	2623	–	449	/
	Remaining days w/ >= 2 meal logs	5597	–	4,943	–
	Total meal logs used in this study	22,322	–	21,274	–

Table 3. Data description. Overview of datasets 1 & 2 with over 12,000 days of insulin pump data containing over 45,000 meal logs from 88 subjects with diabetes. After the exclusion of 6 subjects who had less than 15 days of insulin pump data with at least 2 meal logs and all days in the remaining subjects with less than 2 meals logs per day, our final dataset includes 10,497 days of insulin pump data with 43,596 meal logs from 82 subjects with diabetes.

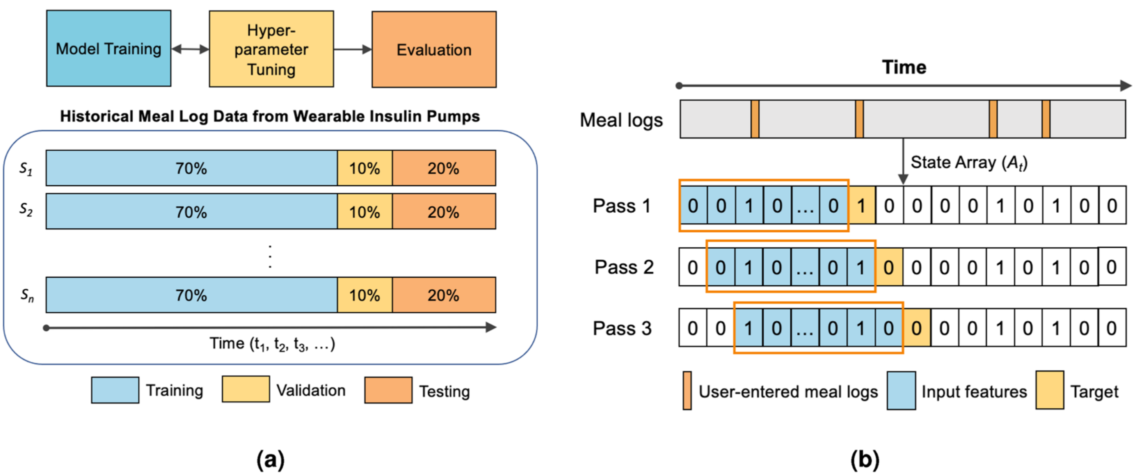


Fig. 6. Rolling forecast methods for modeling and predicting mealtimes. (**a**) The partitioning of each subject's data used for training the models, hyperparameter tuning (validation set), and evaluation (testing set). (**b**) An illustration of the mapping from user-entered meal log data to the state arrays used as inputs to the rolling forecast models.

An effective mealtime prediction system has the potential to significantly improve state-of-the-art technology in the diabetes domain, i.e., the hybrid closed loop (HCL) system for insulin delivery and blood glucose control[11,13,50]. Today's technology still fall short compared to what is desired in the field because current HCL systems depend on patients to manually log their meals to administer bolus doses of insulin needed to maintain postprandial (or post-meal) glucose control[51]. Consequently, ample research shows that blood glucose control around mealtimes remains a primary challenge for patients with type 1 diabetes[4,7]. Thus, technology-based solutions that accurately predict mealtimes (and quantities) have the potential to radically transform the diabetes domain and advance the state-of-the-art technology from a hybrid closed loop to a fully closed loop system[5,52–55]. In addition, such a solution can contribute to novel decision support solutions for advanced diabetes management, even for patients who prefer open-loop therapy[56].

Comparison with related work on mealtime detection or prediction

Given the importance of mealtime detection and/or prediction in the diabetes domain, there are several studies in the literature that have worked toward these tasks. To position our work and provide an anchor for our results, we compare this research with others in the field. Table 2 presents a comparison of key parameters from related work for mealtime detection or prediction using diabetes device data. From this table, we can observe that the majority of prior work in the field focus on the task of detecting unannounced meals *after* the meal is consumed and majority of these studies use simulated data sources[25–30,54,55,57] . Meanwhile, this study focuses on the task of predicting future mealtimes (i.e., *before* the meal is consumed) and we use real patient data for modeling and evaluation. Only a few studies in literature have worked toward the predicting or anticipating future meals in the context of diabetes management[58–61], but the methods/metrics used for evaluation do not allow for comparison with this study. However, prior work has shown improvements in hemoglobin A1C and time with blood glucose in the target range (70–180 mg/dL) when mealtime reminders are combined with HCL systems to reduce missed or late meal insulin doses[11], and improvements in time with blood glucose below the target range (i.e., < 70 mg/dL) when mealtime anticipatory systems are combined with fully-closed loop systems (FCL) systems[58]. Nonetheless, the datasets used in this study have significantly more meal events (i.e., user-entered meal logs) than those in prior work. Additionally, the performance scores obtained on both datasets in this study (i.e. 82 subjects) is similar or greater than what has been reported in all prior studies. From this analysis, we also observed that several studies did not report comprehensive metrics which may give a false sense of their full results if not assessed carefully.

Comparison of SARIMA and LSTM for time series forecasting

Time series forecasting is an important task in many domains such as health, economics, business, and finance[40,41,62–65]. Given the prevalence and promise of well-known stochastic models (e.g., ARIMA and SARIMA) and deep learning based models (e.g., LSTM and BiLSTM), open questions remain around understanding which forecasting methods offer the best predictions for various domain-specific tasks. Previous research studies have found that LSTM can outperform ARIMA on various forecasting tasks with a lower error rate of ∼ 85–100%[40,41,63]. Similarly, our research shows that LSTM also outperformed SARIMA (a variant of the standard ARIMA model) for predicting mealtimes. In particular, SARIMA performed very poorly with an average F1 score of *only* 16% compared to 95.33%, as shown in Table 1. One key contributor to the poor performance of SARIMA for the task of predicting mealtimes relates to the nature of meal data from wearable insulin pumps. In particular, user-entered meal logs create a discrete and binary dataset with only two states (meal or no meal) as shown in Supplementary Fig. S4. Given this, ARIMA-based models are not fitting, meanwhile, LSTM models are more effective for the task of modeling and predicting mealtimes from wearable insulin pump data.

Limitations and future work

Although there are several strengths to the work presented in this paper, there are also some important limitations to note. First, given that meal logs from wearable insulin pumps are a form of self-reported data, this will inherently include human error such as missed meal logs or inconsistency in the timing of entries. Although, the clinical recommendation is for patients with diabetes to log each meal in their insulin pump around 20 mins before meal initiation[9], there is often variability in when exactly people enter their meals into their insulin pump (e.g., 10-mins before meal initiation or 5-mins after meal initiation)[6]. Given this knowledge, we used a 30-mins buffer window around the exact meal log timestamps to assess the performance of our mealtime prediction models. A second limitation of this work is that mealtime prediction alone may not be sufficient to initiate bolus insulin delivery based on today's hybrid closed loop technology, another important factor is the meal quantity which is necessary to determine the appropriate amount of insulin needed for that meal. However, as seen in prior work mealtime prediction and/or detection systems can be used to provide reminders that reduce the occurrence of missed or late meal insulin doses[11]. Additionally, meal anticipation systems have been evaluated with fully closed loop systems by providing a fixed insulin bolus at mealtime or increased insulin in the hours leading up to a potential meal[58]. With this knowledge, our future work will investigate the ability to model and predict meal quantities as well as other key factors that influence insulin delivery and dosage such as exercise[54,55,66,67]. A third limitation of this work is that we leveraged retrospective data for model development and evaluation, thus more research is needed to evaluate such a system in a real-time/online fashion.

Method

Study and data description

As described above, this study leverages retrospective wearable data from two independent cohorts with total of 88 patients with diabetes as shown in Table 3. Dataset 1 (DiaTrend[31]) is an open-source dataset that includes 54 subjects with type 1 diabetes (age = 33.6 ± 15.6 yrs.). This dataset was collected by our research team through the process of recruiting patients with type 1 diabetes from a local hospital and online platforms, and downloading their retrospective diabetes device data for research. Our study was approved by the Committee for Protection of Human Subjects at Dartmouth College, all subjects provided informed consent, and all experiments were performed in accordance with the relevant guidelines. Conversely, dataset 2 comprises 34 subjects with type 1 diabetes (age = 42.3 ± 8.6 yrs.). This dataset was collected by and licensed from Tidepool[32]—a non-profit diabetes organization. Both datasets represent retrospective and observational device data (i.e., data that is *not* influenced by recruiting subjects into a prospective study).

Building on prior research[4,7], we excluded subjects who had less than 15 days of data and days of data with less than 2 meal logs per day—see Supplementary Fig. S4. These criteria led to the exclusion of 6 subjects, namely, subjects 2, 26, 27, 29, 32, and 35, all from dataset 1. Additionally, this criteria led to the exclusion of 2623 out of 8220 days (31.9%) from dataset 1 and 449 out of 5392 days (8.3%) from dataset 2. Following these exclusions, dataset 1 comprises 5597 days of insulin pump data with > 2 meal logs per day (total: 22,322 meal logs from 48 subjects in dataset 1), meanwhile, dataset 2 comprises 4943 days of insulin pump data with > 2 meal logs per day (total: 21,274 meal logs from 34 subjects in dataset 2). In total, our final dataset includes 10,497 days of insulin pump data with 43,596 meal logs from 82 subjects with type 1 diabetes.

Modeling and predicting mealtimes

Toward the goal of modeling and predicting mealtimes from routinely collected insulin pump data, we implemented, evaluated, and compared five data-driven methods (three baselines and two machine learning methods) for the task. Our baseline methods are similar to those in prior work on continuous trajectory prediction tasks[68–70], while our machine learning methods include both statistical and deep learning-based methods for time-series forecasting[40,41].

Baseline methods for predicting mealtimes

As a baseline, we evaluated the effectiveness of predicting mealtimes ($M_{i,j}$) for subject i on day j using: (1) a standard-mealtime baseline, (2) a prior-day baseline, and (3) a peak-mealtime baseline. The aforementioned methods represent reasonable baselines that leverage both population-level and individual-level insights from this study and from prior work[33,34]. For performance assessment, we use an evaluation window of 30 mins to quantify the prediction performance across all methods. Therefore, 15 mins before and 15 mins after any prediction with the baseline methods is compared with the ground truth timestamp of meal logs to assess correctness—see Supplementary Fig S5.

The standard-mealtime baseline is informed by prior work which found a clear breakfast-lunch-dinner pattern in a population of subjects without diabetes, with the first meal at 8:29 AM, second meal at 12:42 PM, and third meal at 6:30 PM[33]. Building on these findings, we selected standard mealtimes at the closest half-hour point, namely 8:30 AM, 12:30 PM, and 6:30 PM. This choice is also (partly) supported by the peak mealtimes observed in our population-level analysis shown in Fig. 2a. Hence, the standard-mealtime baseline evaluates the assumption that most people with diabetes eat 3 meals per day and it uses fixed times at 8:30 AM, 12:30 PM, and 6:30 PM for prediction. This baseline is similar to the "most frequent visit" baseline from prior work on location/mobility prediction[68–70].

Meanwhile, the prior-day baseline builds on the premise that different people have different routines and yesterday's data from a specific individual could be useful for predicting today's behavior. Thus, a reasonable mealtime prediction for subject i on the day j is their own mealtimes on the prior day $j-1$. For example, if the meal log data for subject i on the day j shows mealtimes at 10:56 AM and 5:10 PM, then the mealtime predictions for subject i on the day $j+1$ will be 11 AM and 5 PM. This prior-day baseline is similar to the "last week trajectory" or "same place" baseline methods from prior work on location/mobility prediction[68,70].

Thirdly, the peak-mealtime baseline also builds on the premise that different people have different routines. However, it uses the most frequent pattern in recent days for each individual to predict their future behavior. More specifically, this baseline uses the k most frequent mealtimes from n prior days of subject's i data to predict their mealtimes on the next day (i.e., day $n+1$). Supplementary Fig. S5 shows an example of this method with $k=3$ based on 15 days of prior data for subject 68. As can be observed from this figure, the most frequent mealtimes for this unique subject based on their own historical data are 5:30 AM, 11:30 AM, and 6 PM; hence these times directly inform the mealtime prediction for this subject on day 16. While this might be a reasonable approach for mealtime prediction for some subjects, we observed that this method has the potential to perform poorly for subjects who show irregular mealtime routines (i.e., the majority of subjects in this study as shown in Supplementary Fig. S1).

Rolling forecast methods for predicting mealtimes

Toward the goal of a meal prediction system that can nudge patients with diabetes to administer the needed insulin doses *at the right time* (i.e., $\sim$ 20-mins before each meal[9,10]), we formulate the problem here as a task of predicting the probability of a meal 30-mins in advance of the current time t. This formulation enables the use of well-known statistical and deep learning-based methods for time-series forecasting, such as autoregressive integrated moving average (ARIMA) and long short-term memory model (LSTM)[40,41]. In addition, this formulation supports the use of insights from earlier times in each day (e.g., the time of the last meal) to achieve better prediction of when the next meal will occur. It is important to note that the proposed approach is different from the above-mentioned baseline methods that aim to predict mealtimes $M_{i,j}$ for subject i on day j using meal logs from prior days.

For the rolling forecast methods implemented in this work, we start by apportioning each subject's data into a training, validation, and testing set using a 70/10/20 split as shown in Fig. 6a (i.e., we develop personalized models). Then, we transform each subject's past meal log data into a 1-D state array (A_t), where A represents the state: meal (1) or no meal (0), and t is the time index of each consecutive (i.e. non-overlapping) 30-mins window— see Fig. 6b. In the case where two adjacent meal logs were within 30 mins of each other but fell in different window segments, these meal logs were combined and only the first meal log was used as the ground truth timestamp. Following this, we apply LSTM - a state-of-the-art neural network algorithm that is well-established for long sequence modeling, and seasonal ARIMA (SARIMA) - a well-known variation of the traditional ARIMA model fitting for seasonal data[71,72].

Since LSTM models learn a function that maps a sequence of past observations as input to predict a target[71], we define a sliding window of length N to allocate a portion of past observations $[A_{t-N+1}, A_{t-N+2}, \dots, A_t]$ as input features to predict a target A_{t+1}. We evaluate a 1-layer LSTM and a 2-layer LSTM with various unit numbers (16, 32, 64, and 128) and various number of epochs (2 to 7), while maintaining a fixed learning rate of 0.0003. Fig. 7a and b show the results of this analysis on the validation set. Additionally, we evaluate different window sizes N = [24, 48, 96, 144], which represent the prior 12-h, 24-h, 48-h, and 72-h of past observations as input features. Fig. 7c shows the result of this analysis on the validation set. Based on this analysis, we select a 1-layer LSTM architecture with 32 units, 5 training epochs, and a window size of N = 96 (i.e., the past 48-h) as input features. Supplementary Table S1 provides an outline of the parameter settings for the proposed 1-layer LSTM. Upon implementation, this model outputs the probability of a meal occurring in a target time index within the state array based on prior observations, hence, we leverage a simple peak detection method to identify the highest probabilities and annotate the mealtime predictions—see Supplementary Fig. S6 for an example.

To support a comparison of two well-known time-series forecasting methods, we also applied SARIMA for the task of predicting mealtimes. Similar to the implementation with LSTM, we defined a sliding window of length N to allocate a portion of past observations $[A_{t-N+1}, A_{t-N+2}, \dots, A_t]$ as input features to predict a target A_{t+1}. In this work, we used a fixed value of N = 14 days to provide two full weeks of past observations on mealtimes in order to create a balance between the computational efficiency of the model and performance. Our choice of setting N = 14 days was also informed by our exclusion/inclusion criteria described earlier which shows that our dataset retained 82 subjects with a minimum of 15-days of insulin pump data. As described in prior work[62,72], SARIMA is formed by adding seasonal terms to the traditional ARIMA model and thus can be described as:

$$SARIMA(p, d, q)(P, D, Q)_m \tag{1}$$

where (p, d, q) represents the non-seasonal part of the model, particularly, autoregression (p), the difference (d), and moving average (q), while $(P, D, Q)_m$ represents the seasonal part of the model, particularly, the seasonal autoregression (P), the seasonal difference (D), and seasonal moving average (Q). Lastly, the seasonality parameter m represents the number of time periods in a single seasonal period. In this work, we set the seasonality parameter (m) to 24-h (i.e., window size N = 48 for 30-min windows) and we used grid search on the validation dataset to determine the order of the non-seasonal and seasonal parts of the model. Based on our analysis, we select the parameters for SARIMA(3,1,0)(2,1,0). Finally, we defined and used a threshold of 0.158 (the best performing threshold evaluated on the validation set of 10 randomly selected subjects) to annotate mealtime predictions from the probability outputs of SARIMA. Our choice of using a threshold-based approach (as opposed to the peak detection method) to assess the probability output from SARIMA was informed by the fact that we did not observe any obvious structure in the probability outputs from SARIMA during development and testing.

Performance metrics and evaluation

To evaluate the performance of the above mealtime prediction methods, we leverage standard metrics such as recall (or sensitivity), precision, and F1 score[73]. Recall represents the true positive rate, that is the proportion of real mealtimes that were correctly predicted. Conversely, precision represents the true positive accuracy (or

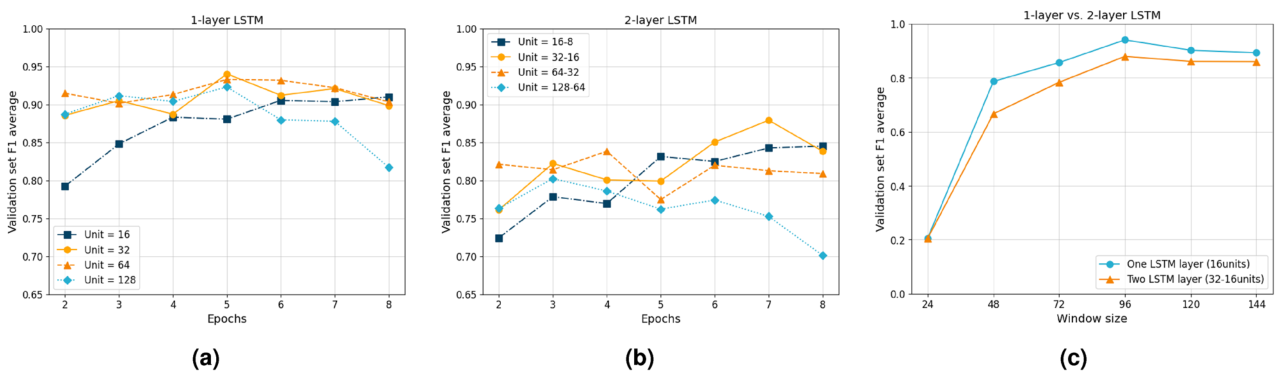


Fig. 7. Comparing the performance of a 1-layer LSTM vs. a 2-layer LSTM with varying number of units, training epochs, and window sizes. (**a**) The average F1 score on the validation set with a 1-layer LSTM. (**b**) The average F1 score on the validation set with a 2-layer LSTM. (**c**) The average F1 score on the validation set with different window sizes.

confidence), that is the proportion of correct mealtimes (per the ground truth) amongst the predicted mealtimes, as shown in Supplementary Fig. S4. Meanwhile, the F_1 score is a cumulative metric that represents the harmonic mean between recall and precision. Since the above metrics focus on positive events, we complement these with other metrics that quantify the error rate. In particular, the false positive rate, which describes the proportion of real negatives (i.e., no meal on a given day and time) that are predicted as a mealtime, and the false negative rate (or miss rate), which describes the proportion of real positives (i.e., mealtimes on a given day and time) that are *not* predicted. We calculate the false positive and false negative rates per day (i.e. FP/day and FN/day) to enable easy comparison of our results with related work[28,29,54,55,74].

One additional criteria leveraged for performance evaluation in this study relates to using a buffer window to quantify mealtime prediction results. This criteria was set in place because there is extensive debate in the literature on *when* exactly is mealtime. For example, is the mealtime at the start, middle, or end of an eating episode (e.g., having breakfast)? Additionally, in the diabetes domain, the clinical recommendation is for patients to log their meals ~ 20-mins before meal consumption to initiate the needed insulin dose, and thus mitigate post-meal high blood glucose events[9,10]. However, in practice, there is often variability in when patients log their meals into the insulin pump. Hence, in this study, we use a 30-mins buffer window around the exact timestamp of each user-entered meal log (i.e., ground truth) to quantify true positives, false positives, and false negative labels for mealtime predictions. More specifically, only when a prediction is within the window of 15-mins before or 15-mins after the exact timestamp of a ground truth meal log is it labeled as a true positive event.

Data availability

A subset of the data analyzed in this work (i.e., Dataset 1 also known as DiaTrend[31]) is available for request and direct access through Synapse at https://doi.org/10.7303/syn38187184. Meanwhile, data sharing agreements prohibit the authors from making the Dataset 2 publicly available. However, access to Dataset 2 can be requested from and provided by Tidepool[32] pending scientific review and a completed material transfer agreement.

Code availability

All the code generated in this work was done with Python 3.10. In this study, we leveraged packages Scipy and Scikit-learn for statistical analysis and Tensorflow 2.15 for LSTM model building. The code generated for this study is available upon request to the Augmented Health Lab at Dartmouth College (www.ah-lab.cs.dartmouth.edu/).

Received: 29 November 2023; Accepted: 29 August 2024
Published online: 09 September 2024

References

1. O'Connell, M., Donath, S. & Cameron, F. Poor adherence to integral daily tasks limits the efficacy of CSII in youth. *Pediatr. Diabetes* **12**, 556–559 (2011).
2. Patton, S. R. *et al.* Frequency of mealtime insulin bolus as a proxy measure of adherence for children and youths with type 1 diabetes mellitus. *Diabetes Technol. Ther.* **15**, 124–128 (2013).
3. Burdick, J. *et al.* Missed insulin meal boluses and elevated hemoglobin a1c levels in children receiving insulin pump therapy. *Pediatrics* **113**, e221–e224 (2004).
4. Gingras, V., Taleb, N., Roy-Fleming, A., Legault, L. & Rabasa-Lhoret, R. The challenges of achieving postprandial glucose control using closed-loop systems in patients with type 1 diabetes. *Diabetes Obes. Metab.* **20**, 245–256 (2018).
5. Mosquera-Lopez, C. *et al.* Enabling fully automated insulin delivery through meal detection and size estimation using artificial intelligence. *npj Digit. Med.* **6**, 39 (2023).
6. Robinson, S., Newson, R. S., Liao, B., Kennedy-Martin, T. & Battelino, T. Missed and mistimed insulin doses in people with diabetes: A systematic literature review. *Diabetes Technol. Ther.* **23**, 844–856 (2021).
7. Olinder, A. L., Kernell, A. & Smide, B. Missed bolus doses: Devastating for metabolic control in CSII-treated adolescents with type 1 diabetes. *Pediatr. Diabetes* **10**, 142–148 (2009).
8. Foster, N. C. *et al.* State of type 1 diabetes management and outcomes from the t1d exchange in 2016–2018. *Diabetes Technol. Ther.* **21**, 66–72 (2019).
9. Cobry, E. *et al.* Timing of meal insulin boluses to achieve optimal postprandial glycemic control in patients with type 1 diabetes. *Diabetes Technol. Ther.* **12**, 173–177 (2010).
10. Slattery, D., Amiel, S. & Choudhary, P. Optimal prandial timing of bolus insulin in diabetes management: A review. *Diabet. Med.* **35**, 306–316 (2018).
11. Corbett, J. P. *et al.* Smartwatch gesture-based meal reminders improve glycaemic control. *Diabetes Obes. Metab.* **24**, 1667–1670 (2022).
12. McAdams, B. H. & Rizvi, A. A. An overview of insulin pumps and glucose sensors for the generalist. *J. Clin. Med.* **5**, 5 (2016).
13. Leelarathna, L. *et al.* Hybrid closed-loop therapy: Where are we in 2021?. *Diabetes Obes. Metab.* **23**, 655–660 (2021).
14. Taylor, P. J., Thompson, C. H. & Brinkworth, G. D. Effectiveness and acceptability of continuous glucose monitoring for type 2 diabetes management: A narrative review. *J. Diabetes Investig.* **9**, 713–725 (2018).
15. Sora, N. D., Shashpal, F., Bond, E. A. & Jenkins, A. J. Insulin pumps: Review of technological advancement in diabetes management. *Am. J. Med. Sci.* **358**, 326–331 (2019).
16. Bartolome, A., Shah, S. & Prioleau, T. Glucomine: A case for improving the use of wearable device data in diabetes management. *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.* **5**, 1. https://doi.org/10.1145/3478109 (2021).
17. Wong, J. C., Neinstein, A. B., Spindler, M. & Adi, S. A minority of patients with type 1 diabetes routinely downloads and retrospectively reviews device data. *Diabetes Technol. Ther.* **17**, 555–562 (2015).
18. Belsare, P., Lu, B., Bartolome, A. & Prioleau, T. Investigating temporal patterns of glycemic control around holidays. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 1074–1077 (IEEE, 2022).
19. Prioleau, T., Moore, E. & Ghovanloo, M. Unobtrusive and wearable systems for automatic dietary monitoring. *IEEE Trans. Biomed. Eng.* **64**, 2075–2089 (2017).
20. Bell, B. M. *et al.* Automatic, wearable-based, in-field eating detection approaches for public health research: A scoping review. *npj Digit. Med.* **3**, 1–14 (2020).

21. Vu, T., Lin, F., Alshurafa, N. & Xu, W. Wearable food intake monitoring technologies: A comprehensive review. *Computers* **6**, 4 (2017).
22. Katz, D. S., Price, B. A., Holland, S. & Dalton, N. S. Data, data everywhere, and still too hard to link: Insights from user interactions with diabetes apps. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12 (2018).
23. Popp, C. J. *et al.* Objective determination of eating occasion timing (oreo): Combining self-report, wrist motion, and continuous glucose monitoring to detect eating occasions in adults with pre-diabetes and obesity. *J. Diabetes Sci. Technol.*[SPACE]https://doi.org/10.1177/193229680900300511 (2023).
24. van den Brink, W. J., van den Broek, T. J., Palmisano, S., Wopereis, S. & de Hoogh, I. M. Digital biomarkers for personalized nutrition: Predicting meal moments and interstitial glucose with non-invasive, wearable technologies. *Nutrients* **14**, 4465 (2022).
25. Turksoy, K., Bayrak, E. S., Quinn, L., Littlejohn, E. & Cinar, A. Multivariable adaptive closed-loop control of an artificial pancreas without meal and activity announcement. *Diabetes Technol. Ther.* **15**, 386–400 (2013).
26. Lee, H., Buckingham, B. A., Wilson, D. M. & Bequette, B. W. A closed-loop artificial pancreas using model predictive control and a sliding meal size estimator (2009).
27. Xie, J. & Wang, Q. A variable state dimension approach to meal detection and meal size estimation: In silico evaluation through basal-bolus insulin therapy for type 1 diabetes. *IEEE Trans. Biomed. Eng.* **64**, 1249–1260 (2016).
28. Ramkissoon, C. M., Herrero, P., Bondia, J. & Vehi, J. Unannounced meals in the artificial pancreas: Detection using continuous glucose monitoring. *Sensors* **18**, 884 (2018).
29. Meneghetti, L., Facchinetti, A. & Del Favero, S. Model-based detection and classification of insulin pump faults and missed meal announcements in artificial pancreas systems for type 1 diabetes therapy. *IEEE Trans. Biomed. Eng.* **68**, 170–180 (2020).
30. Askari, M. R. *et al.* Meal and physical activity detection from free-living data for discovering disturbance patterns of glucose levels in people with diabetes. *BioMedInformatics* **2**, 297–317 (2022).
31. Prioleau, T., Bartolome, A., Comi, R. & Stanger, C. Diatrend: A dataset from advanced diabetes technology to enable development of novel analytic solutions. *Sci. Data* **10**, 556 (2023).
32. Tidepool. Tidepool big data donation project. https://www.tidepool.org/bigdata.
33. Biel, J.-I., Martin, N., Labbe, D. & Gatica-Perez, D. Bites 'n'bits: Inferring eating behavior from contextual mobile data. *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.* **1**, 1–33 (2018).
34. Gill, S. & Panda, S. A smartphone app reveals erratic diurnal eating patterns in humans that can be modulated for health benefits. *Cell Metab.* **22**, 789–798 (2015).
35. Aiello, E. M., Toffanin, C., Magni, L. & Nicolao, G. D. Model-based identification of eating behavioral patterns in populations with type 1 diabetes. *Control Eng. Pract.* **123**, 105128. https://doi.org/10.1016/J.CONENGPRAC.2022.105128 (2022).
36. Camerlingo, N., Vettoretti, M., Del Favero, S., Facchinetti, A. & Sparacino, G. Mathematical models of meal amount and timing variability with implementation in the type-1 diabetes patient decision simulator. *J. Diabetes Sci. Technol.* **15**, 346–359. https://doi.org/10.1177/1932296820952123 (2021).
37. Aiello, E. M., Toffanin, C., Magni, L. & De Nicolao, G. Model-based identification of eating behavioral patterns in populations with type 1 diabetes. *Control Eng. Pract.* **123**, 105128 (2022).
38. Riddell, M. C. & Milliken, J. Preventing exercise-induced hypoglycemia in type 1 diabetes using real-time continuous glucose monitoring and a new carbohydrate intake algorithm: an observational field study. *Diabetes Technol. Ther.* **13**, 819–825 (2011).
39. Riddell, M. C. *et al.* Exercise management in type 1 diabetes: A consensus statement. *Lancet Diabetes Endocrinol.* **5**, 377–390 (2017).
40. Kaushik, S. *et al.* AI in healthcare: Time-series forecasting using statistical, neural, and ensemble architectures. *Front. Big Data* **3**, 4 (2020).
41. Siami-Namini, S., Tavakoli, N. & Namin, A. S. A comparison of ARIMA and LSTM in forecasting time series. In *2018 17th IEEE international conference on machine learning and applications (ICMLA)*, 1394–1401 (IEEE, 2018).
42. Prioleau, T., Bartolome, A., Comi, R. & Stanger, C. Diatrend: A dataset from advanced diabetes technology to enable development of novel analytic solutions. https://doi.org/10.7303/syn38187184(2023).
43. Belsare, P., Bartolome, A., Stanger, C. & Prioleau, T. Understanding temporal changes and seasonal variations in glycemic trends using wearable data. *Sci. Adv.* **9**, eadg2132 (2023).
44. Bartolome, A. & Prioleau, T. A computational framework for discovering digital biomarkers of glycemic control. *npj Digit. Med.* **5**, 1–9 (2022).
45. Morton, S., Li, R., Dibbo, S. & Prioleau, T. Data-driven insights on behavioral factors that affect diabetes management. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 5557–5562 (IEEE, 2020).
46. Vhaduri, S. & Prioleau, T. Adherence to personal health devices: A case study in diabetes management. In *Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 62–72 (2020).
47. Zhu, T. *et al.* Enhancing self-management in type 1 diabetes with wearables and deep learning. *npj Digit. Med.* **5**, 78 (2022).
48. Bedri, A. *et al.* Earbit: Using wearable sensors to detect eating episodes in unconstrained environments. *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.* **1**, 1–20 (2017).
49. Thomaz, E., Bedri, A., Prioleau, T., Essa, I. & Abowd, G. D. Exploring symmetric and asymmetric bimanual eating detection with inertial sensors on the wrist. In *Proceedings of the 1st Workshop on Digital Biomarkers*, 21–26 (2017).
50. Tauschmann, M. & Hovorka, R. Technology in the management of type 1 diabetes mellitus-current status and future prospects. *Nat. Rev. Endocrinol.* **14**, 464–475 (2018).
51. Barnard, K. D. *et al.* Future artificial pancreas technology for type 1 diabetes: What do users want?. *Diabetes Technol. Ther.* **17**, 311–315 (2015).
52. Fox, I., Lee, J., Pop-Busui, R. & Wiens, J. Deep reinforcement learning for closed-loop blood glucose control. In *Machine Learning for Healthcare Conference*, 508–536 (PMLR, 2020).
53. Thabit, H. & Hovorka, R. Coming of age: The artificial pancreas for type 1 diabetes. *Diabetologia* **59**, 1795–1805 (2016).
54. Samadi, S. *et al.* Automatic detection and estimation of unannounced meals for multivariable artificial pancreas system. *Diabetes Technol. Ther.* **20**, 235–246 (2018).
55. Samadi, S. *et al.* Meal detection and carbohydrate estimation using continuous glucose sensor data. *IEEE J. Biomed. Health Inform.* **21**, 619–627. https://doi.org/10.1109/JBHI.2017.2677953 (2017).
56. Vettoretti, M., Cappon, G., Facchinetti, A. & Sparacino, G. Advanced diabetes management using artificial intelligence and continuous glucose monitoring sensors. *Sensors* **20**, 3870 (2020).
57. Askari, M. R. *et al.* Adaptive-learning model predictive control for complex physiological systems: Automated insulin delivery in diabetes. *Annu. Rev. Control.* **50**, 1–12 (2020).
58. Garcia-Tirado, J. *et al.* Assessment of meal anticipation for improving fully automated insulin delivery in adults with type 1 diabetes. *Diabetes Care* **46**, 1652–1658 (2023).
59. Cameron, F. *et al.* Anticipatory behavior in blood glucose control: using meal prior probabilities to prepare for future meal disturbances. In *IFAC Proceedings Volumes* Vol. 44 3771–3776 (Elsevier, 2011).
60. Hughes, C., Patek, S. D., Breton, M. & Kovatchev, B. P. Anticipating the next meal using meal behavioral profiles: A hybrid model-based stochastic predictive control algorithm for t1dm. *Comput. Methods Programs Biomed.* **102**, 138–148 (2011).
61. Corbett, J. P., Garcia-Tirado, J., Colmegna, P., Diaz Castaneda, J. L. & Breton, M. D. Using an online disturbance rejection and anticipation system to reduce hyperglycemia in a fully closed-loop artificial pancreas system. *J. Diabetes Sci. Technol.* **16**, 52–60 (2022).

62. Chen, P., Niu, A., Liu, D., Jiang, W. & Ma, B. Time series forecasting of temperatures using SARIMA: An example from Nanjing. In *IOP Conference Series: Materials Science and Engineering* Vol. 394 052024 (IOP Publishing, 2018).
63. Siami-Namini, S., Tavakoli, N. & Namin, A. S. A comparative analysis of forecasting financial time series using arima, lstm, and bilstm. arXiv preprint arXiv:1911.09512 (2019).
64. Siami-Namini, S., Tavakoli, N. & Namin, A. S. The performance of lstm and bilstm in forecasting time series. In *2019 IEEE International Conference on Big Data (Big Data)*, 3285–3292 (IEEE, 2019).
65. Jin, X. *et al.* Prediction for time series with cnn and lstm. In *Proceedings of the 11th International Conference on Modelling, Identification and Control (ICMIC2019)*, 631–641 (Springer, 2020).
66. Askari, M. R. *et al.* Detection of meals and physical activity events from free-living data of people with diabetes. *J. Diabetes Sci. Technol.* **17**, 1482–1492 (2023).
67. Turksoy, K. *et al.* Multivariable artificial pancreas for various exercise types and intensities. *Diabetes Technol. Ther.* **20**, 662–671 (2018).
68. Sadri, A. *et al.* What will you do for the rest of the day? An approach to continuous trajectory prediction. *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.* **2**, 1–26 (2018).
69. Do, T. M. T. & Gatica-Perez, D. Where and what: Using smartphones to predict next locations and applications in daily life. *Pervasive Mob. Comput.* **12**, 79–91 (2014).
70. Do, T. M. T., Dousse, O., Miettinen, M. & Gatica-Perez, D. A probabilistic kernel method for human mobility prediction with smartphones. *Pervasive Mob. Comput.* **20**, 13–28 (2015).
71. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735 (1997).
72. Box, G. E., Jenkins, G. M., Reinsel, G. C. & Ljung, G. M. *Time Series Analysis: Forecasting and Control* (Wiley, 2015).
73. Powers, D. M. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. arXiv preprint arXiv:2010.16061 (2020).
74. Dassau, E., Bequette, B. W., Buckingham, B. A. & Doyle, F. J. III. Detection of a meal using continuous glucose monitoring: Implications for an artificial β-cell. *Diabetes Care* **31**, 295–300 (2008).

Acknowledgements
The authors would like to thank Tidepool[32] for data contribution and support that make this and other diabetes-related research possible. In addition, this work is partly supported by the National Science Foundation under grant numbers 2322879 & 2127309, and the National Institute of Diabetes and Digestive and Kidney Diseases under grant number R01DK124428.

Author contributions
T.P. conceived the study, contributed to data collection, data analysis, data interpretation, manuscript preparation and funding acquisition. C.S. contributed to data collection and funding acquisition. B.L. led data analysis and interpretation. Y.C., and P.B. contributed data analysis and interpretation. T.P. and B.L. wrote the manuscript. X.Z. contributed to funding acquisition. All authors edited and reviewed the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-71630-w.

Correspondence and requests for materials should be addressed to T.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2024