Can we predict snacking behaviour just from previous instances?

Shaima Dammas, Tillman Weyde, Katy Tapper, Gerasimos Spanakis, Anne Roefs, Emmanuel M. Pothos

Submitted to: Journal of Medical Internet Research
on: February 20, 2024

Table of Contents

Can we predict snacking behaviour just from previous instances?

Shaima Dammas[1]; Tillman Weyde[2]; Katy Tapper[3]; Gerasimos Spanakis[4]; Anne Roefs[5]; Emmanuel M. Pothos[3]

[1]Department of Health Services and Hospitals Administration King AbdulAziz University Jeddah SA
[2]Department of Computer Science City, University of London London GB
[3]Department of Psychology City, University of London London GB
[4]Department of Advanced Computing Sciences Maastricht University Maastricht NL
[5]Faculty of Psychology & Neuroscience Maastricht University Maastricht NL

Corresponding Author:
Shaima Dammas
Department of Health Services and Hospitals Administration
King AbdulAziz University
Department of Health Services and Hospitals Administration
King AbdulAziz University
Jeddah
SA

Abstract

Background: Consuming too much food or drink with high levels of saturated fats, salt or sugar can be harmful for health. Many snack foods fall into this category (HFSS snacks). However, the palatability of these snacks means that people can sometimes struggle to reduce their intake. Machine learning algorithms could help by predicting the likely occurrence of HFSS snacking, so that just-in-time adaptive (JITAI) interventions can be deployed. However, HFSS snacking data has characteristics (such as sparseness and incompleteness), which make snacking prediction a challenging machine learning problem. Previous attempts have employed several potential predictor variables, achieving considerable success. Nevertheless, collecting information along several dimensions requires several potentially burdensome user questionnaires per day, so that this approach may be less acceptable among the general public.

Objective: Our aim is to consider the capacity of machine learning algorithms to predict HFSS snacking based on minimal data that can be collected in a mostly automated way: day of week, time of day, and location (coarsened as work, home, other).

Methods: A sample of 111 participants in the UK were asked to record HFSS snack occurrences and location category, over a period of 28 days, leading to a new dataset on HFSS snacks. Data collection was facilitated by a purpose-specific app. Additionally, we use a similar dataset from the Netherlands. For both datasets, we employ machine learning methods (random forests and neural networks).

Results: We report results concerning the ability of machine learning methods to predict the time of the next HFSS snack. The quality of the prediction depended on both the dataset and temporal resolution employed. In some cases, predictions were accurate to as few as 17 minutes on average.

Conclusions: We demonstrated that prediction of HFSS snacks on sparse data is possible to reasonable accuracy. We consider the type of prediction problem which may be most suitable for putative interventions in relation to HFSS snacking. While we think it was important to employ standard machine learning algorithms in this work, we also discuss ways to tailor both the machine learning algorithms and the prediction problem to better align with the unique characteristics of the problem.

(JMIR Preprints 20/02/2024:57530)
DOI: https://doi.org/10.2196/preprints.57530

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Can we predict snacking behaviour just from previous instances?

Shaima Dammas[1], Tillman Weyde[2], Katy Tapper[3], Gerasimos Spanakis[4], Anne Roefs[5], & Emmanuel M. Pothos[3]

1. Department of Health Services and Hospitals Administration, King Abdulaziz University, Jeddah, Saudi Arabia. smalghamdi2@kau.edu.sa
2. Department of Computer Science, City, University of London, t.e.weyde@city.ac.uk
3. Department of Psychology, City, University of London, katy.tapper.1@city.ac.uk
4. Department of Advanced Computing Sciences, Maastricht University, jerry.spanakis@maastrichtuniversity.nl
5. Faculty of Psychology & Neuroscience, Maastricht University, a.roefs@maastrichtuniversity.nl

Abstract

Background

Consuming too much food or drink with high levels of saturated fats, salt or sugar can be harmful for health. Many snack foods fall into this category (HFSS snacks). However, the palatability of these snacks means that people can sometimes struggle to reduce their intake. Machine learning algorithms could help by predicting the likely occurrence of HFSS snacking, so that just-in-time adaptive (JITAI) interventions can be deployed. However, HFSS snacking data has characteristics (such as sparseness and incompleteness), which make snacking prediction a challenging machine learning problem. Previous attempts have employed several potential predictor variables, achieving considerable success. Nevertheless, collecting information along several dimensions requires several potentially burdensome user questionnaires per day, so that this approach may be less acceptable among the general public.

Objectives

Our aim is to consider the capacity of machine learning algorithms to predict HFSS snacking based on minimal data that can be collected in a mostly automated way: day of week, time of day, and location (coarsened as work, home, other).

Methods

A sample of 111 participants in the UK were asked to record HFSS snack occurrences and location category, over a period of 28 days, leading to a new dataset on HFSS snacks. Data collection was facilitated by a purpose-specific app. Additionally, we use a similar dataset from the Netherlands. For both datasets, we employ machine learning methods (random forests and neural networks).

Results

We report results concerning the ability of machine learning methods to predict the time of the next HFSS snack. The quality of the prediction depended on both the dataset and temporal resolution employed. In some cases, predictions were accurate to as few as 17 minutes on average.

Discussion

We demonstrated that prediction of HFSS snacks on sparse data is possible to reasonable accuracy. We consider the type of prediction problem which may be most suitable for putative interventions in relation to HFSS snacking. While we think it was important to employ standard machine learning algorithms in this work, we also discuss ways to tailor both the machine learning algorithms and the prediction problem to better align with the unique characteristics of the problem.

Keywords: high fat, salt, and sugar snacks; machine learning algorithms; internet data collection; just in time interventions

Introduction

The biggest threats to health today are from non-communicable diseases such as cardiovascular disease, cancer, and chronic respiratory disease (Institute for Health Metrics and Evaluation, 2020). These are heavily influenced by behaviors such as poor diet, physical inactivity, and smoking (e.g., Bolnick et al., 2020; OECD, 2019). Changing the obesogenic environment would be most effective at bringing about wide scale change (Marteau, Fletcher, Munafo & Hollands, 2021). However, such environmental changes are unlikely to occur anytime soon unfortunately. Therefore, most interventions seek to change people's response to their environment. Approaches to achieve these changes have ranged from mass media campaigns at the population level, to group-based and individual healthy lifestyle coaching. Over the last decade, digital technologies for supporting a healthy lifestyle have been on the rise, but there is still much room for improvement (Lau et al., 2020; Myers-Ingram et al., 2023; O'Boyle & Davidson, 2022).

One potential avenue for improving the effectiveness of technology-assisted interventions is providing just-in-time adaptive interventions (JITAIs). JITAIs are designed to predict the points at which a person is likely to most need, and be most receptive to, reminders or assistance to change a target behavior (Nahum-Shani et al., 2018). For example, someone engaged in a smoking quit attempt could be reminded of their goal at the point at which they are most likely to lapse (Naughton et al., 2016; 2017). This may be effective because motivation for behavior fluctuates varies over time, as does the risk for lapses (Pimpini et al., 2023). Additionally, health-related behaviors may be elicited by cues in the environment (e.g., a certain time of day or walking past a bakery store) and be largely habitual (Cleobury & Tapper, 2014; Neal et al., 2011; Verhoeven et al., 2012; Gardner, 2015). This can make them hard to change unless behaviors elicited by these cues are disrupted. Helping a person identify those cues may help them to either avoid or adjust them or to consciously monitor their behavior to ensure they respond to the cues in a different way (Gardner, 2015; Jenkins & Tapper, 2014). JITAIs could help people achieve this.

There is increasing research into the use of JITAIs (e.g., Hardeman et al., 2019; Perski et al., 2022). Their development has been greatly boosted by the widespread adoption of powerful, personal smartphones. For example, in 2021, 88% of all adults (aged 16+) possessed a smartphone in the UK (https://www.uswitch.com/mobiles/studies/mobile-statistics/), with similar statistics throughout Western Europe and North America. Modern smartphones allow increasingly complex data collection from their owners including date, time, ambient temperature, and location. Furthermore, the increased computational capabilities of modern smartphones mean that many of the required computations (e.g., for prediction) can be carried out locally, rather than rely on distant servers and active internet access.

Our focus for the current work is snacking on foods and drinks that are high in saturated fats, salt and/or sugar (HFSS snacks). Snacking can be defined as "food and beverage intake between meals, including products such as potato chips, chocolate, and soft beverages" (p.82, Almoraie et al., 2021). HFFS foods contribute to poorer health (Afshin et al., 2019) and many snack foods fall into this category. Indeed, one study found that people with overweight or obesity ate an average of 1.3 snacks per day, with 79% of these being high in either fat or sugar (Cleobury & Tapper, 2014). However, reducing HFSS snacking poses many challenges, not least because it can be triggered by emotional or environmental factors (McArthur et al, 2012). It can also occur in an automatized (reflexive) way making it less amenable to conscious efforts to control it (Cleobury & Tapper, 2014, Neal et al., 2011; Verhoeven et al., 2012). Additionally, snacks that are high in sugar may make a person crave further sugary foods because they can lead to a spike and subsequent dip in blood sugar levels (Ehrampoush et al, 2020; Wyatt et al., 2021). Indeed, feelings of hunger and food preoccupation are key reasons cited for snacking (Cleobury & Tapper, 2014).

Several sophisticated approaches to predict various aspects of maladaptive eating behavior have already been proposed, using Ecological Momentary Assessment (EMA). For example, Arend et al. (2023) studied binge-eating episodes in clinical participants. The authors reported excellent predictive accuracy, based on an EMA protocol with 36 items, including emotional and environmental variables. Based on initial testing, they were subsequently able to identify a smaller subset (5-9) of highly valid individualized predictors (EMA items), thereby reducing the need for an extensive EMA protocol. Forman et al. (2019) similarly investigated the predictive adequacy of a large number of variables, concerning dietary lapses. Some questions were answered as many as four times a day (such as those relating to cravings or affect), and others only once. Finally, Spanakis et al. (2017) tracked several individual states, such as emotions, cravings etc., which might predict 'unhealthy eating events' (including unhealthy snacking, but also other events, such as eating of a high calorie food as part of a meal) in people who were overweight or obese. Participants were questioned as many as 10 times a day. Based on the collected data, a bottom-up clustering algorithm was used to arrive at six different subgroups of participants characterized by a specific pattern of eating behavior (e.g., the evening at home eaters), to enable tailoring interventions to a specific profile, which was implemented in a randomized controlled trial (Boh et al., 2016).

Research based on EMA is valuable because the prediction of a behavior as complex as eating can potentially only be accomplished by considering a multitude of variables (including environmental, psychological, and physiological). Moreover, the assessment of these variables takes place in daily life, contributing to ecological validity. However, health interventions based on EMA typically require long periods to train the machine learning algorithms for prediction, as well as considerable commitment and motivation from participants. There is therefore interest in exploring whether prediction of a particular behavior can proceed based on information that is both minimal and easily available, without much effort from participants.

In the domain of mental health, research on so-called digital phenotyping has recently started to develop. Digital phenotyping uses the smartphone as a tool for objective and ecologically valid measurements. This method includes passively obtained data, without needing input from the user. Digital biomarkers such as sensor technology, geolocation, characteristics of voice and speech, and human-computer interaction are obtained (Brietzke et al., 2019; Insel, 2018; Oudin et al., 2023; Prakash et al., 2021). Imagine for example that by this method you discover a pattern, over several weeks, that a person takes long to respond to messages, is browsing online until late at night, and is mostly at home. You may then suspect that it is not going particularly well with this person, and your suspicion may be increased by the tone, timing, and content of this person's social media posts. For example, research has shown that mood states in mood disorders could be predicted based on digital biomarkers based on the circadian rhythm (Cho et al., 2019).

The current research project takes a step in the direction of digital phenotyping for the prediction of unhealthy eating behaviour. Specifically, to what extent can HFSS snacking be predicted just on the basis of prior occurrences of HFSS, combined with information that can be automatically or easily collected from a smartphone (date, time, location)? It is possible this endeavor will fail due to temporal resolution requirements. If too much precision is needed, due to the intrinsic stochasticity of eating behavior, this is inevitable. Another problem is the degree of accuracy that can be achieved after modest training periods, because participants may not have the patience for long training (e.g., Tulu et al., 2017) -- even with mostly passively collected data, participants would need to, minimally, indicate instances of HFSS.

On the positive side, machine learning has progressed to such an extent that modern algorithms have many characteristics desirable for the present application, including the capacity to deal with sparse data and efficient learning of time series. For example, as an alternative to recurrent neural networks (which are well suited to time series data), ensemble methods have good ability in dealing with sparse data by reducing the impact of noise and outliers (Lee et al., 2020). Therefore, our aim was to compare a selection of machine learning algorithms, with a view to identify a good algorithm for predicting instances of HFSS snacking, just on the basis of prior instances, coded in terms of time, day, and location (the latter encoded in terms of broad categories). The algorithms were chosen to reflect complementary characteristics and be representative of the range of options available.

First, the Random Forest regressor (RFreg) is a tree-based ensemble method that trains many decision trees simultaneously with bootstrapping followed by aggregation, collectively referred to as bagging. Bootstrapping involves the training of several individual decision trees (here, between 100 and 300), on several subsets of the dataset, using various subsets of the available features (Lee et al., 2020). Aggregation means that the outputs from the distinct decision trees are combined into a single decision. RFreg is considered to generalize well and be resistant to overfitting, as well as produce high prediction accuracy, because of ensemble learning (Fawagreh et al., 2014).

Second, eXtreme Gradient Boosting regressor (XGBreg) is another tree-based ensemble method, which uses a form of gradient boosting, relying on the idea that correcting the model's earlier errors and learning from them helps performance in the future. This is a sequential ensemble learning method where the model tries to improve performance with each iteration (Brownlee, 2016). Both RFreg and XGBreg are ensemble learning techniques, but the former builds multiple trees in parallel, then employs an average for prediction, while the latter constructs one tree at a time, in a way that is informed from the errors from the previous tree (Li et al., 2020).

Third, we employed a feed forward neural network (FFNN). This is a relatively simple type of artificial neural network, in which information is processed in one direction, from input units to units in one or more hidden layers to output units, such that there are no cycles in the connections between the nodes. Hidden layer units apply non-linear functions to their input, enabling an FFNN to learn complex associations between input and output (Sharma, 2017). An FFNN is trained using gradient descent methods, specifically error backpropagation (Rumelhart et al., 1986).

Finally, the long short-term memory (LSTM) model is a kind of recurrent neural network with an architecture designed for learning long-term dependencies in time series (Hochreiter & Schmidthuber 1997). A recurrent neural network includes cycles which feed network activations from earlier time steps as inputs to determine predictions at the present time step. As a result of these recurrent connections, the model creates an implicit recollection of past occurrences, stored in its hidden layer (e.g., van Houdt et al., 2020). Recurrent models can process contexts of arbitrary length and the LSTM has a specific structure designed to store values for longer than standard recurrent neural networks. The LSTM is the only recurrent model we employed, with the other models operating on a fixed context.

Some models can benefit from standardization, normalization, and dropout regularization more so than others, and we explored these techniques accordingly. Feature scaling (such as standardization or normalization) was considered for the FFNN and LSTM models. Neural networks, such as FFNN and LSTM, often benefit from having consistent scales across features. This is because neural networks learn complex relationships and patterns among features, which can be influenced by the differing scales of features if not appropriately managed (Albon, 2018). On the other hand, tree-based models, such as RFreg and XGBreg, operate by splitting nodes based on feature thresholds and so are less sensitive to feature scaling. As it turned out, normalization slightly enhanced results from the LSTM model, but not FFNN.

We also explored dropout, a regularization technique which randomly sets to zero (drops out) a percentage of the features during training (Srivastava et al., 2014). Dropout regularization introduces randomness during training and prevents over-specialization of units, which can improve generalization. This is important for neural networks, as they can overfit the training data. Accordingly, we evaluated (and ended up retaining) dropout regularization for FFNNs and LSTMs. Tree-based models use other (inherently incorporated) mechanisms such as feature selection, bootstrapping, and ensemble aggregation to manage overfitting.

Methods

Outline of data collection

We collected data on HFSS snacking, to examine which of the machine learning algorithms was best able to predict such behaviour, just on the basis of previous instances and minimal information (time, day of the week, location). We first describe the procedures employed to collect this data.

Collection of these data was conducted in two parts: first, we created a survey to explore various assumptions about the target behaviour of interest and, second, we implemented an app to obtain data on HFSS snacking. Participants who reported having two or more HFFS snacks daily in the first part were invited to the app-based second part of the study. The second part of the study involved monitoring participants' snacking behaviour for 28 days. The two parts of the study are outlined in Figure 1. We refer to this dataset as the 'UK dataset'. To increase the ecological validity of our work, we also employed a cleaned version of a similar dataset, collected in the Netherlands, as described in Spanakis et al. (2017). We refer to Spanakis et al.'s (2017) dataset as the 'Dutch dataset'.

All the data collection sections below concern the UK dataset; corresponding details for the Dutch dataset can be found in Spanakis et al. (2017).

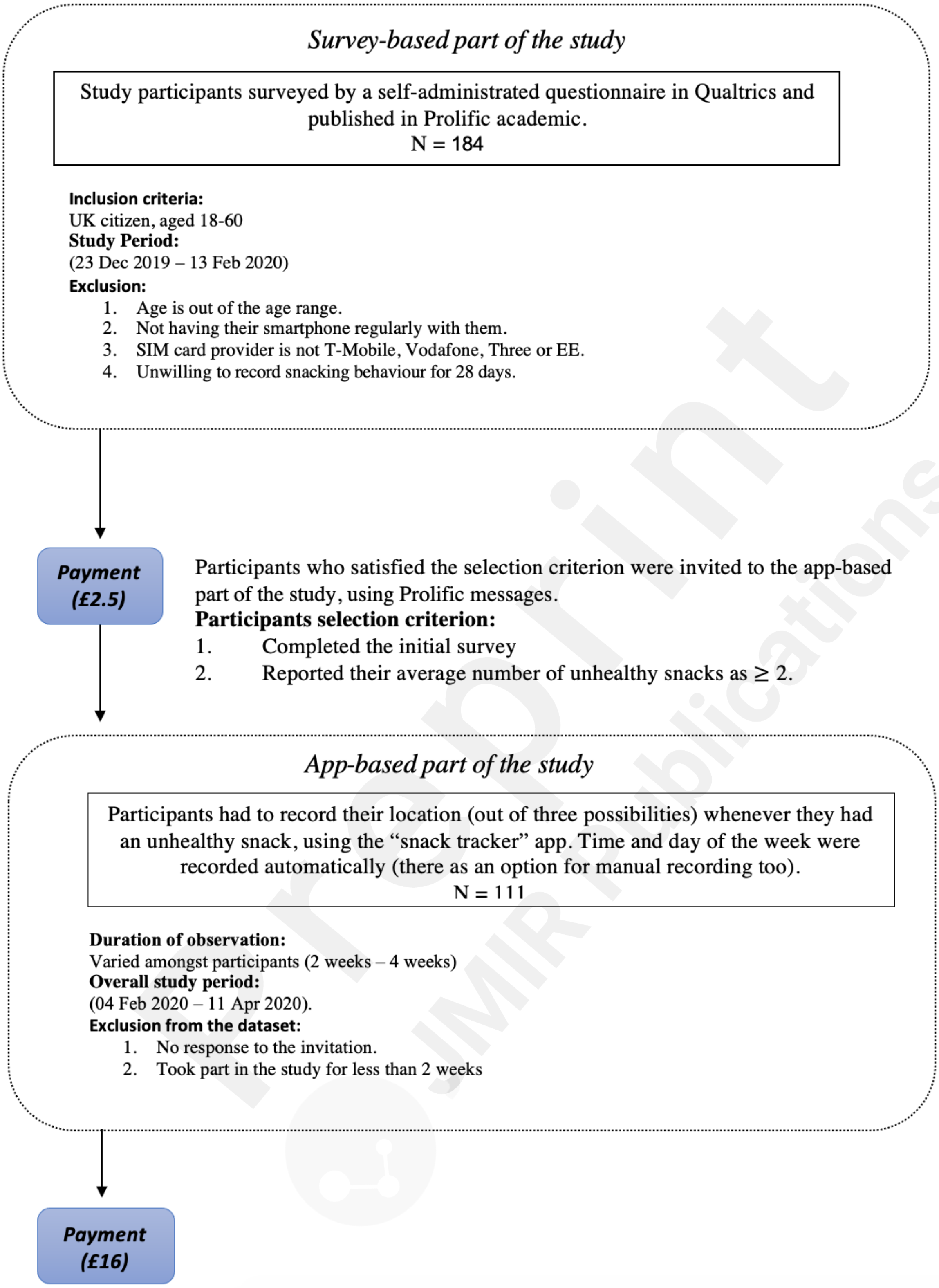


Figure 1. Outline of the two parts of the data collection, for the UK dataset.

UK dataset: Ethics, privacy, and security considerations

Ethics approval was provided by the City, University of London Psychology Department Research

Ethics Committee. To protect participant anonymity, the Data Protection Office at City, University of London (which is registered with the Information Commissioner's Office, registration number Z8947127) was engaged, to confirm that any Personal Identifier Information (PII) was securely collected. Data collection involved three companies external to City, University of London: Dev Technosys implemented the app (see below), Linode provided virtual private servers, and Twillo provided programmable messaging services. We ensured these companies were compliant with the General Data Protection Regulation (GDPR) (https://gdpr-info.eu) and used encrypted communication.

UK dataset: Survey-based part of the study

Data for the first part was collected using a self-administered survey, designed using Qualtrics (https://www.qualtrics.com). The survey was run on Prolific Academic (https://www.prolific.co). The study was published with a target sample of 200 participants on 23 Dec 2019 and closed on 13 Feb 2020, when this participant number had been reached. Given the novelty of the work, no detailed power calculations were carried out and the sample size was partly determined by practical considerations.

Only UK citizens between the age of 18 and 60 were allowed to join the study. Participants were excluded if they did not have a smartphone or stated they were unwilling to participate in the follow-up study, i.e. the app-based part. Additionally, we only recruited participants having a T-Mobile, O2, Vodafone, Three, or EE mobile phone service, since (at the time of running the study), these were the only SIM card providers in the UK compatible with the Twillo messaging service (which we employed in the second part of the study). After these exclusions, we were left with 184 participants for the first part of the study.

The survey consisted of questions concerning basic demographics and motivation for healthy eating (Tables A1.1 and A1.2 in Appendix 1). The duration of the survey was about 15 minutes and participants were compensated £2.50 for their time.

UK dataset: App-based part of the study – the Snack Tracker app

We developed an app, called Snack Tracker, specifically for this project, to record unhealthy snacking. The app was designed by SA and the coding undertaken by Dev Technosys (https://devtechnosys.com), a company specializing in app development. We created versions of the app for both Android and iOS devices. Mobile app development is usually divided into two main components, frontend and backend (Figure 2).

Regarding the frontend (the user interface), the app was designed to be easy to use, with a simple sequence of screens (Figure 3; see Appendix 1 for larger versions of the figure). The current date and day of the week were automatically captured for each recording, to minimize user effort. The app worked online, allowing users to log in and record any snacks they had eaten. In cases of connection loss, the app allowed users to save their records and the app passed the data to the server once the mobile device was connected to the internet again.

The app frontend was coded using React Native, which is an open-source JavaScript framework for writing iOS and Android applications. All operations performed by app users and project admins were handled by Rest Application Programming Interface (APIs), created in Node JS (which allows running JavaScript on the server side). Rest APIs, created with Node.js, were used to communicate with the database for participant data (MongoDB), for store and retrieve operations (i.e., these APIs acted as a bridge between the app frontend and the database, where participant data was stored). The server used in this project was a cloud-based server (Linode), which controlled all operations and allowed the management of the application environment.

The backend part of the app concerned storing the data and user credentials, as well as offering a web-based admin panel to manage the project. Participants accepting the invitation to the app-based part of the study were registered manually using the admin panel, to prevent random users from recording data on the app. The backend also handled initial user login and user requests to save and record another snack, go back and edit, or just to save an entry (Figure 3).

The app development included designing and delivering messages as SMS and app push notifications, to keep the users engaged and remind them to record snacks (Figure 4). Push notifications, in particular, could be offered, even if participants had no internet access. Regarding the automated SMS messages, Twilio was used, as a cloud communication tool. Via Twillo, the software could programmatically send SMSs, utilizing Twillo's web service APIs. For more information see https://www.twilio.com.

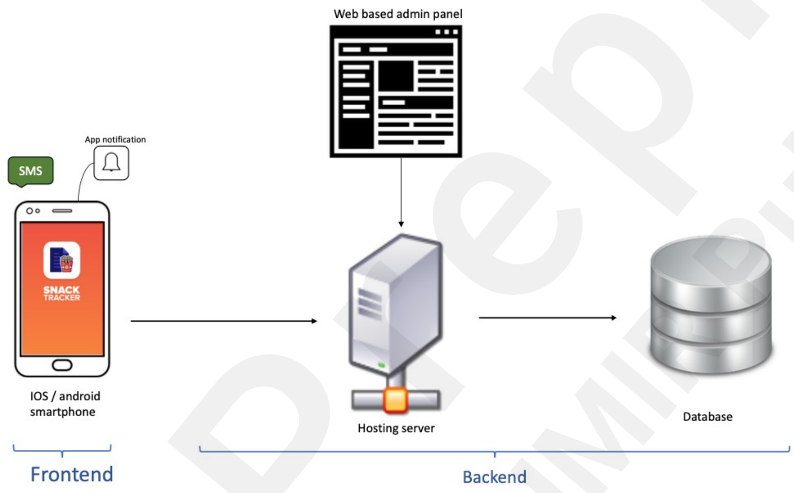


Figure 1. The frontend and backend of the Snack Tracker app, illustrating data flow.

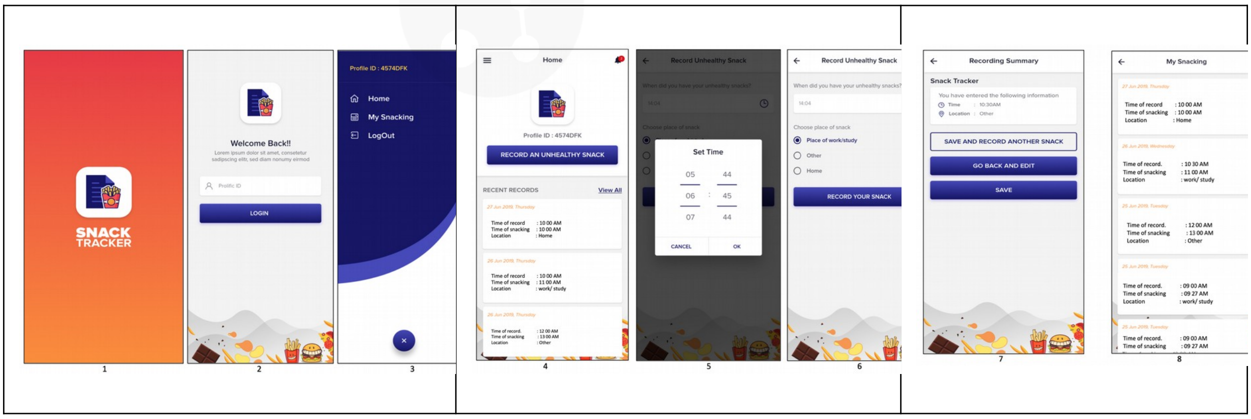


Figure 3. The frontend of the Snack Tracker app: 1, splash screen; 2, login screen; 3, home screen; 4,

new snack screen; 5, time recording screen (time picker); 6, location recording screen; 7, recording save screen; 8, review recordings summary.

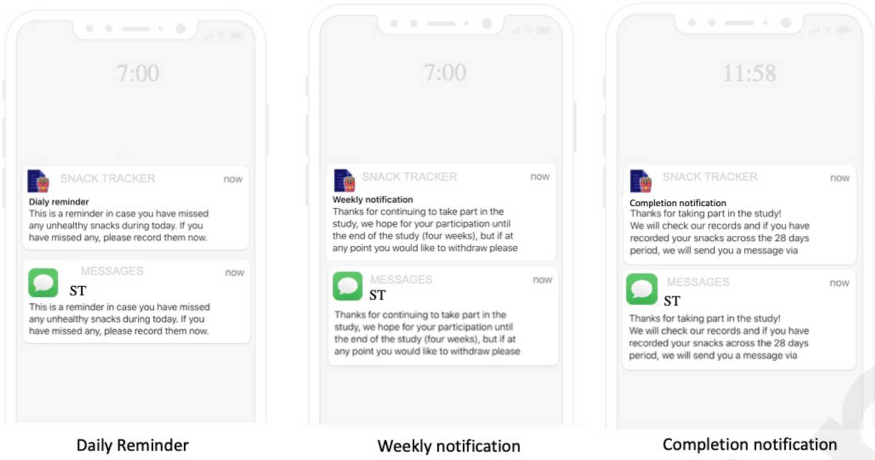


Figure 4. The various reminders employed in the app-based part of the study.

UK dataset: App-based part of the study – data collection

Of the 184 participants who completed the survey-based part of the study, we further excluded participants who reported consuming fewer than two HFSS snacks per day, leaving us with 170 participants. We initially invited 100 of them to participate in the app-based part of the study, with the invitations sent in three stages, between February and March 2020. Only 68 participants accepted the initial invitation, leading us to invite a further 45 (27 accepted) and, finally, another 25 (16 accepted), to produce an overall sample of 111. Participants received £16 for taking part in this part of the study.

Participants were instructed to participate for 28 days. However, exact start and end dates of the study differed between participants, as would be expected. Notably, some participants did not stay on the study for its full intended duration. The majority of participants recorded their HFSS snacks for the full 28 days. Participants who took part for less than one week were excluded from the dataset. A few participants continued recording snacking for an additional 6-11 days, after reaching their 28th day and a completion message had been sent; this additional data was used in the analyses.

During the study, we messaged participants, via Prolific, if they made only one or no recordings on any day and if they had made multiple recordings (two or more) each day for a period, but subsequently the number of recordings abruptly decreased. We sent these messages in part to ensure there were no technical problems with the Snack Tracker app. Additionally, a few Prolific messages were sent to randomly selected participants, to check that the reminders regarding snack recording (Figure 4) were received as intended. Informal feedback from participants throughout the study did not indicate any technical problems. Moreover, we received some positive comments about the study in general. For example, participants mentioned that recording their snacks made them aware of the amount consumed daily and helped them reduce food intake. Note, while this is a desirable side effect of the data collection (and indeed consistent with other research, e.g., in alcohol abuse; see also Pimpini et al., 2023), it is a complicating factor regarding machine learning – future work is needed to consider how such effects can be taken into account.

For the purposes of the study, a HFSS snack was defined as any food eaten in-between main meals, that was high in either saturated fat, salt or sugar (Appendix 1 shows how HFSS snacks were explained to participants). Typical items would include chocolate, biscuits, sweets or crisps. However, in keeping with the broad aims of the study (to help people manage their own behavior), we did not impose strict criteria on what participants should or should not define as a HFSS snack. Whenever participants had a HFSS snack, they were asked to record it in the Snack Tracker – in this mode, the only information they had to provide was the location, (coarsely coded as home, place of work, other; it might seem that a GPS-based method would obviate the need for manual location entry, but this was not possible in our app), since the time and day of the week were saved automatically. Throughout the study duration, participants received three kinds of notifications (Figure 4). Daily reminders were sent at 19:00, asking participants to record any forgotten HFSS snacks. In this mode of the Snack Tracker, participants had to manually indicate time and day of the week as well as location. Additionally, at the end of each week, participants were sent a notification (instead of the daily one), to keep them engaged as well as some general text (asking them about any technical problems etc.; see Appendix 1). The final notification participants received was at the end of the 28-day period from their first app recording, instructing them that the study had ended, to thank them for their participation, and offer them the completion code for Prolific Academic. Overall, the total number of recordings across all participants was 5391, which, after some data cleaning (see below), was reduced to 4,978 data points.

Dutch dataset: Brief notes

The Dutch dataset was collected by a research group at the Faculty of Psychology and Neuroscience, at Maastricht University (Study I in Spanakis et al., 2017). The dataset was collected with an app called Think Slim. The sample consisted of 57 participants who were overweight and 43 participants with a healthy weight in the Netherlands. This study employed Ecological Momentary Assessments(EMA), so that data on 15 variables was collected, concerning e.g., mood, activity, location, across 8-10 measurements per day, depending on participants' waking and sleeping times. Spanakis et al.'s (2017) aim was to study eating behavior in general (including main meals and snacks, both unhealthy ones and otherwise), with a view to identify participant clusters and provide adaptive feedback towards improving eating behavior. Accordingly, we extracted measurements corresponding to HFSS snacks from the general eating behavior data from Spanakis et al. (2017), to create what we refer to as the Dutch dataset. The Dutch dataset consisted of 3,705 data points.

Results

Data pre-processing

Some basic operations were carried out to remove erroneous entries and ensure consistency between the UK and Dutch datasets. In the UK dataset, we removed missing and duplicate entries (413) and ensured that measurement units for height and weight were the same across participants. In the Dutch dataset, we translated data from Dutch to English and manually re-classified 950 locations, originally in free text, into the three categories we employed in the UK dataset. Finally, we extracted HFSS snacks from the general information about meals/ snacks. This involved focusing on data concerning food items, such as burgers, chocolate bars, strawberries, pasta, etc., further considering snacks. Each snack was manually categorized as healthy or unhealthy (HFSS – see above).

Data was coded in terms of the following three features: location (home, place of work, other), day of the week (Monday…Sunday), and time. We constructed a time bin feature, using 4 large and 12 small time bins, that is, a regular day was divided into 4 or 12 time bins. In the former case, the time bins were early morning (0:00-05:59), morning (06:00-11:59), afternoon (12:00-16:59) and evening (17:00-23:59). In the latter case, the first time bin started at 00:00 and had a two-hour duration and analogously for the rest. Models were trained and evaluated for these two encodings of variable. Note, arguably, using 4 time bins coarsens temporal resolution too much – we offer this analysis here for illustration and with a mindset that this might be useful in some applications.

As is common in machine learning, for the two nominal variables in the datasets, location and day of the week, we used one-hot encoding, converting each variable to separate variables, taking the values 1 or 0, to indicate the presence or not of different levels of each variable (Berry et al., 1998). The time bin variable was kept in its numeric format. Also, as noted, we explored some normalization (feature scaling) for the neural network models (e.g., Pagan et al., 2021). For the FFNN model, we standardized features using $z_i=(x_i-\mu)/\sigma$, where μ and σ are the mean and standard deviation values of the variable respectively, z_i indicates the standardized value, and x_i the original value. For the LSTM model, we employed the normalization, $y_i=x_i-min\ (X)/(max\ (X)\cdot min\ (X))$, as this proved to improve performance of the model, relative to standardization.

Model details

We employed three fixed context models (RFreg, XGBreg, and FFNN) and one recurrent model (LSTM). Fixed context models require fixed input size, which can be achieved by windowing a longer sequence. Windowing is a technique used to divide a longer sequence into smaller, fixed-length sequences. Due to the sparseness of our data, we decided to treat each individual data point as a separate window for prediction, rather than grouping them into larger sequences. This approach is often used when the data is not abundant enough to create longer sequences, and it can simplify the modeling process for the algorithms. This is a simplifying approach, when the dataset is, relatively speaking, small. Recurrent models can process an input sequence of arbitrary length. They function in cycles, during which the activation from the previous time step is used as input (together with other information) for the current time step. For the one recurrent model we studied, we used observation sequences of four timesteps. Note, in our study, the day was divided into different time bins, including four time bins (early morning, morning, afternoon, evening). While the four timesteps in this model do not directly correspond to these specific time divisions, they align with the broad temporal division, which we found sufficient for our analysis.

We performed hyper-parameter tuning through random search via scikit-learn, executing a total of 5 experiments for each model. In each experiment, various hyperparameter combinations were evaluated, selected from predefined parameter ranges (Table 1). This process retrieved the hyper-parameters corresponding to the best performing architectures for each model. Both random forest and XGBoost are based on the same decision trees and Figure 5 illustrates their differences, as well as some of the main parameters.

Table1. Hyper-parameters for each of the learning algorithms. The number of hyper-parameters for each model is shown below its name.

Model	**hyper-parameter**	**Values\ Ranges**
RFreg 3	n_E, number of estimators (trees in the forest)	100, 150, 200, 250, 300
	$Min_{ss,}$ (min number of samples an internal node must cover to consider splitting; when a node has fewer samples than this value, it is regarded as final and called a terminal node or leaf)	2, 3, 4
	max_{depth} (the max number of splits that each tree is permitted to execute)	2, 3, 4, 5, 6,7
XGBreg 5	n_E, number of boosting rounds (estimators)	100, 150, 200, 250, 300
	Min_{SS}	2, 3, 4
	max_{depth}	2, 3, 4, 5, 6, 7
	LR, learning rate (step size for each boosting iteration)	0.01, 0.05, 0.1, 0.5, 1
	gamma (how much the loss must be decreased by a split in order for that split to occur)	0.01, 0.05
FFNN 4	number of hidden layers	4
	dropout regularization	0.4 and 0.6, after the 1st and 3rd hidden layers
	number of neurons at the hidden layers (NHL)	min value=128, max value=512, step=64
	activation function	ReLU, in the dense and output layers
LSTM 4	number of neurons in the LSTM layer (NLL)	min value=128, max value=512, step = 64
	number of hidden layers	2
	dropout regularization	0.4 and 0.6, after the LSTM layer and 2nd hidden layers

	NHL	min value=128, max value=512, step = 64
	activation function	ReLU, in the dense and output layer

NB. Splits in a decision tree correspond to the datapoints associated with a node being divided and the parts assigned to child nodes. In relation to gamma, loss is the function minimized during model training. It is based on the difference between current output of the machine learning model and the target, i.e. the true value. Dropout regularization is a technique that randomly switches off neurons in a neural network during training to overfitting. The numbers given refer to the fraction of neurons switched off.

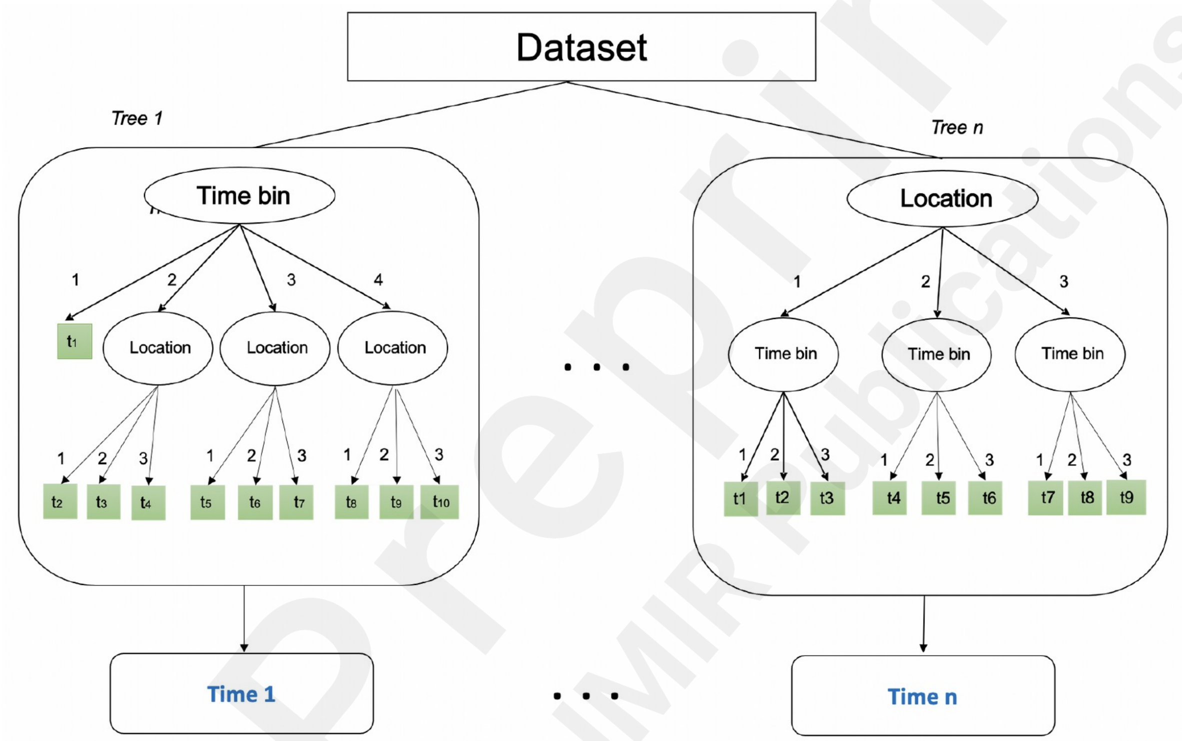


Figure 5. Both Random Forest (in RFreg) and XGBoost (in XGBreg) are based on the same building block, a decision tree, which is a set of rules built from the dataset. In our case, the aim is to predict the time until the next HFSS snack in minutes. *n* is the number of trees in the model, selected based on cross-validation performance. Time *k* is the expected time until the final unhealthy snack predicted using Tree *k*.

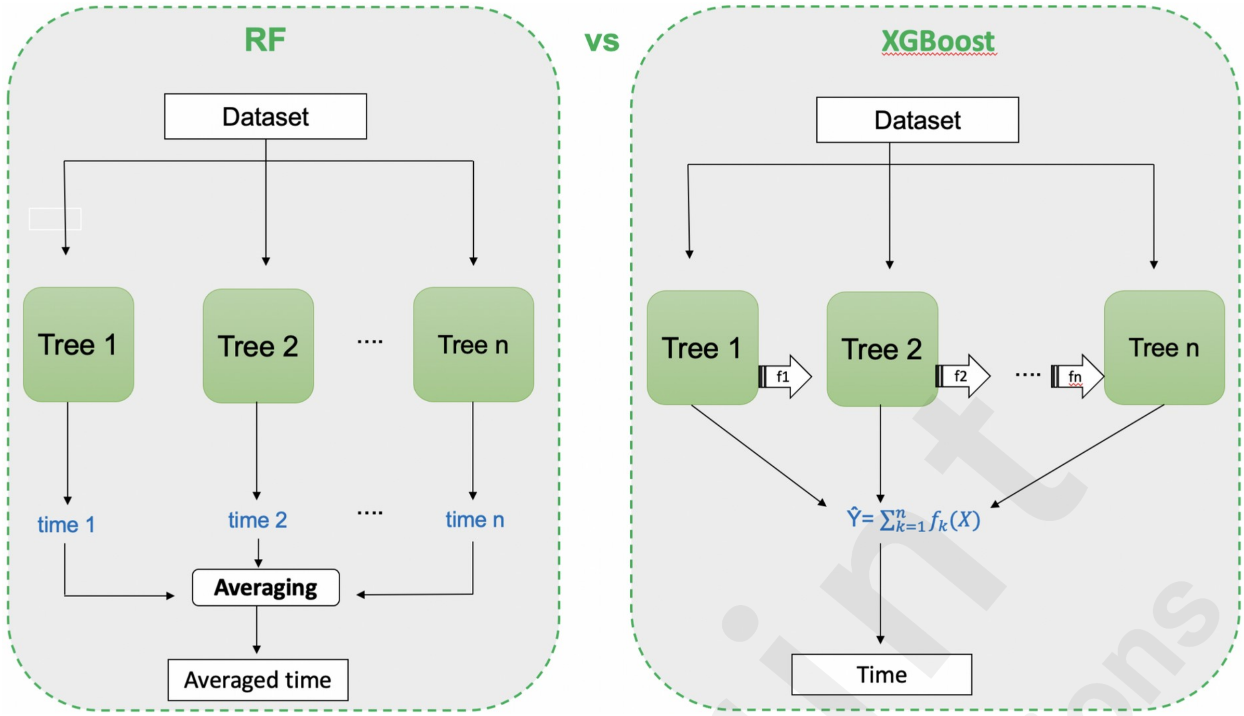


Figure 6. For RFreg (on the left), we create several different trees and then average predictions. For XGRreg (on the right), we have a sequence of trees, which progressively refine the prediction.

Model validation and evaluation

General description

All algorithms had the same objective, which was to predict the time until the next HFSS snack (in minutes), given the current time bin, location (coarsened in terms of work, home, other), and day of the week. To summarize the main approach: we utilized an input of fixed size, comprising current location, current day of the week, and current time bin, for RFreg, XGBreg, and FFN. The LSTM model processed data as a sequence of inputs. We also used a simple linear regression model (LR) as a baseline. The size of the training and testing sets were 3,982 and 996 in the UK dataset and 2,778 and 927 in the Dutch dataset.

We applied five-fold Cross-Validation (CV) on all fixed-context models. In each data split, the datasets were divided into 80% for training and 20% for testing. We ensured that data distributions in the training and testing data subsets were similar to those in the whole dataset, using stratified data sampling (Reitermanova, 2010). For the LSTM, we employed a time series CV. Time series CV is more suitable for temporal data modelling, because the training set only includes the observations occurring before those in the testing set. It begins with a small subset of data for training that is successively extended to generate new predictions (Hyndman & Athanasopoulos, 2014). Because of the sparse data, we decided to use the time series CV with only two splits.

We employed Mean Absolute Error (MAE) as the main evaluation metric for the models. This is because our research focus was time difference between predictions and actual time until the next HFSS snack occurrence. Additionally, residuals allow quantification of positive and negative errors, which were employed in a further analysis (related to the creation of hypothetical interventions – this is introduced below). Residuals were calculated as predicted minus true values, so that negative residuals indicate that the prediction is earlier than the observed time and positive residuals indicate that the predicted time is later than observed (while the MAE is the average of the absolute values of the residuals). We examined the normality of residual distributions, as standard, and offer corresponding plots (clipped to errors of -200 – +200 minutes, as larger errors are outliers in the sense that corresponding predictions are without any use; Appendix 2).

Results

Table 1 shows the MAE for the four models. For the Dutch dataset, FFNN and LSTM were clearly the best models. In the UK dataset, the LSTM model had a very marginal advantage for the 12 time bins version of the data and the FFNN for the 4 time bin version of the data. Table 2 shows information on the hyper-parameters of the best performing models in each case. The LSTM provided a relatively similar performance to FFNN in predicting minutes to the next HFSS snack. But, overall, prediction for the time until the next HFSS snack was not very precise.

We next consider the distribution of residuals for the best performing versions of each model. For most noisy processes, the default expectation is that of a normal residual distribution, with the width of the distribution dependent on the amount of noise and the quality of the model. A normal narrow distribution would in general be indicative of good quality predictions (Fuller, 1986). In Appendix 2 we consider model residuals. Overall, most distributions are close to normal and, where there is a skew, it is a left skew (which is slightly better than a right skew, since predictions are that snacks would be earlier, rather than later).

In Appendix 3 we offer an alternative analysis of model performance, concerning the ability of models to predict the occurrence of a snack within windows of particular sizes. While we take the objective of predicting time to next HFSS snack as the primary one, this alternative analysis is arguably better suited to the specific goal of building JITAIs. For this analysis, we focused on RFreg (briefly, because RFreg residuals conform fairly closely to a normal distribution, which we think is a desirable property for JITAIs).

Dataset	UK				Dutch			
Model	4 time bins		12 time bins		4 time bins		12 time bins	
	Train MAE	Test MAE	Train MAE	Test MAE	Train MAE	Test MAE	Train MAE	Test MAE
LR	17.82	20.01	16.77	17.18	68.24	70.04	65.52	67.56
RFreg	17.75	19.27	16.60	17.13	67.77	69.69	64.85	67.64
XGBreg	17.68	18.10	16.68	17.21	67.80	69.68	64.61	67.20
FFNN	16.50	17.12	16.35	17.15	54.92	56.20	54.88	56.22
LSTM	17.20	18.23	15.52	16.12	54.67	56.11	54.62	57.13

Table 1. Mean Absolute Errors (MAE) for predicting the time of the next HFSS snack in minutes

(fractional part of a minute indicated as decimals), with 4 or 12 time bins (lower MAE is better).

	UK		Dutch	
	4 time bins	12 time bins	4 time bins	12 time bins
RFreg	n_E: 150 Min_{SS}: 4 max_{depth}: 2	n_E: 250, Min_{SS}: 3 max_{depth} : 3	n_E: 250 Min_{SS}: 4 max_{depth}: 3	n_E: 150 Min_{SS}: 2 max_{depth} : 5
XGBreg	n_E: 200 Min_{SS}: 4 max_{depth} : 3 LR: 0.01 gamma : 0.05	n_E: 300 Min_{SS}: 3 max_{depth}: 2 LR: 0.1 gamma : 0.05	n_E: 150 Min_{SS}: 3 max_{depth} : 3 LR: 0.1 gamma: 0.05	n_E: 250 Min_{SS}: 2 max_{depth}: 4 LR: 0.01 gamma: 0.05
FFNN	NHL: 128 NHL_1: 128 NHL2: 128 NHL_3: 128	NHL: 192 NHL_1: 128 NHL_2: 192 NHL_3: 192	NHL: 192, NHL_1: 128 NHL_2: 192 NHL_3: 192	NHL: 192 NHL_1: 128 NHL_2: 192 NHL_3: 192
LSTM	NLL: 320 NHL: 192 NHL_1: 192 dense_2 = 1	NLL: 320 NHL: 256 NHL_1: 512 dense_2 = 1	NLL: 320 NHL: 256 NHL_1: 512 dense_2 = 1	NLL: 256 NHL: 512 NHL_1: 512 dense_2 = 1

Table 2. The Hyper parameters of the best performing model for predicting the time of next HFSS snack in minutes, when using 4 or 12 time bins, in the two datasets, UK and Dutch. The models were optimized with random search.

Summary and conclusions

Interventions can be made more effective by delivering messages at the right time, before the target behaviour is performed, as is the case with JITAIs (e.g., Hardeman et al., 2019; Perski et al., 2022). Machine learning can in principle help. The overarching goal of this study was to design an app to collect data on HFSS snacking behaviour, with a view to explore the capacity of well-known machine learning methods to capture underlying statistical structure.

Over a period of 4 weeks, 111 participants provided data on HFSS snacking, using our Snack Tracker app, leading to what we called the UK dataset. In addition, we analyzed an analogous dataset from the Netherlands (100 participants, of varying weight status), the Dutch dataset. In both cases, a number of pre-processing steps were carried out, including data cleaning, encoding survey questions into scores, one-hot encoding, feature engineering and scaling, and over-sampling. We considered slicing a regular day into 4 and 12 time bins, to examine machine learning performance with smaller and larger time bins. The pre-processing stage also involved the extraction of HFSS snacks from the more numerous food categories recorded in the Dutch dataset.

We explored four well-established machine learning models, RFreg, XGBreg, FFNN, and LSTM to the two datasets, against the problem of predicting the next HFSS snack. Our models were able to predict the time until the next HFSS snack with a mean absolute error of 17 minutes in the UK dataset and 53 minutes in the Dutch dataset. The residuals analysis revealed narrow, non-skewed distributions in most cases (a notable exception is the FFNN model, when applied to the UK dataset with 4 time), suggesting that a degree of non-linearity is needed for capturing relevant statistical structure. Also, fixed context models provided better predictions than the recurrent model (LSTM), although this might change with larger datasets. There is a question regarding the difference in MAE values between the UK and Dutch datasets. We believe this may be attributed to differences in participant selection criteria between the two studies. In the UK study, participants were selected based on their reported frequency of unhealthy snacking, specifically those who consumed more than two unhealthy snacks daily. This criterion likely led to a more homogeneous group in terms of snacking behavior, contributing to lower MAE values. On the other hand, the Dutch study had a broader participant base without prior assessment of their snacking habits. This variability possibly made it more challenging for the models to accurately predict the time until the next HFSS snack, potentially leading to higher MAE values.

Together with the main analysis, in Appendix 3, we also considered the task of timing for the delivery of a hypothetical intervention, by testing whether predicted snacking times fell into time windows of 120, 60 and 30 minutes, before the actual HFSS snack time. In this case, we examined only one model, RFreg, because its residuals showed the closest form to a normal distribution, suggesting that a window selection based on expected accuracy is more likely to behave similarly for new events in an application scenario. Assuming a delivery of an intervention 30 min before a predicted snack, 96% would occur within a one-hour window before the actual snack for the UK dataset and 70% for the Dutch dataset. If the acceptable window between intervention and snack is reduced to 30 minutes (and the intervention delivered 15 min before the prediction), the accuracy is 83% and 49% respectively.

From a technical point of view, the choice of methods in this work represents relevant machine learning models for this type and size of dataset; the models selected outperform the baseline linear model. Models based on decision trees (XGBoost and Random Forest) performed competitively in predicting the next HFSS snack, but overall Neural Networks (FFNN and LSTM) performed somewhat better. This picture may change when the number of data points increases, as neural networks, particularly deep ones, benefit greatly from large datasets. Also, there are alternative neural network models which merit examination, e.g., the Gated Recurrent Unit network, which is fast, requires little memory, and sometimes produces better performance than

LSTM (ArunKumar et al., 2022). One direction for future work is how to utilize the particular characteristics of the present modeling challenge (such as sparsity) to fine-tune learning algorithms.

Two technical challenges merit consideration. First, behaviorally, for many users it would be difficult to have them commit to providing data for a long time, before interesting predictions are available. Note, even if most information is passively obtained, there would still be a small burden to record snacks. Therefore, it would be worth exploring online machine learning, whereby data streams are used to continuously update a model. A prerequisite for such approaches would be a measure of confidence in the quality of predictions, before these are employed for an intervention. Second, HFSS snacking behavior is not necessarily stable and might vary depending on the time of the year for example. Accordingly, a machine learning approach would need to be able to adapt to any changes in behavior. The difficulty then is how to distinguish between routine deviations in baseline behavior versus changes in the behavior itself. Online machine learning might again offer promise in addressing such a challenge (He et al., 2020).

Another consideration is whether additional features could aid prediction. In the present work, we only employed day of the week, time, and location (very coarsely coded). By contrast, Arend et al. (2023), Forman et al. (2019), and Spanakis et al. (2017) are all sophisticated examples of similar predictive modelling, but based on more information. It is clear that utilizing more features offers potential for more accurate prediction. Regarding HFSS snacks, particularly pertinent would be the times of meals and drinks. On the other hand, we wanted to explore predictive potential based on information that would be as straightforward as possible for participants to provide. The motivation for such an approach is that any scheme suitable for roll-out to the general population would benefit from being as unobtrusive (in terms of requests for information) as possible. Note that time and day information are already automatically recorded through our app. Location required participant input, but in future iterations this feature could also be automated, through simple look-up tables between geolocation data and the (pre-recorded) locations of particular users for work, home etc. Additional features which could be recorded automatically are sensor data from sophisticated smartwatches, such as heart rate, blood pressure, and even possibly blood sugar or whether the individual is alone vs. with company. There is a tradeoff here between predictive power, usability, and participant fatigue, which will require more work before it can be clarified.

The prediction problem is also complicated by the goal of the prediction. Arguably, if the purpose of prediction is to carry out an intervention, then it may be valuable to know not just the time of particular e.g. HFSS snacks, but also additional variables, such as whether particular emotions preceded a snacking event. To us, for the reasons outlined above, it seems unlikely that a practical method (i.e., beyond research) can be developed which is based on multiple variables/ EMAs, but future personal electronic devices may make such data possible. The slight negativity of this expectation regarding engagement must be moderated by noting that there will be participants likely to engage with intensive data-collection procedures, if they are sufficiently motivated to accomplish goals relating to dietary change, weight reduction etc.

In the context of a longitudinal experimental study, it is challenging to recruit large numbers of participants. Our hope is that the second stage of this research program will involve partial roll out in the general population. With a larger sample, it would be possible to examine whether there are natural groupings in HFSS snacking behavior, defined by, for example, demographic characteristics, personality traits, physiological differences or eating behavior traits (e.g., Cifuentes et al., 2023; Tapper et al., 2015; Pentikainen et al., 2018; Dakin et al., 2023). Different groups of this kind might display higher regularity in their HFSS eating behavior. So, potentially, categorizing a new participant against a pre-existing classification of individuals might inform our confidence of how well we can predict the behavior of that participant (e.g., see Study II in Spanakis et al., 2017). Currently, we are not sure what would be the form of such classifications: we would rely on machine learning to identify appropriate ones and then attempt to interpret them against eating behavior characteristics.

We think there are exciting behavioral implications from this work, given that, even with limited data (both in terms of features and the window for data collection), a reasonable degree of predictive accuracy was possible. One direction for application concerns utilizing prediction data for delivering behavior change techniques (BCTs) for HFSS snacking behavior (e.g., Jenkins & Tapper, 2014; Wilson et al., 2014). These BCTs could be delivered in several different ways (e.g., by text or audio messages), as a way to increase engagement, and could also be selected and personalized for the individual. Additionally, the ML work by itself could be applied to other behaviors targeted for reduction such as smoking or alcohol consumption.

In closing, we think the present results offer a strong foundation for further exploring how machine learning methods can be utilized in health psychology and provide exciting directions for further research, both more technically oriented and focused on behavioral applications and extensions.

Acknowledgements

SA was funded by the Saudi Arabian Cultural Bureau (SACB) in London. Additionally, the SACB provided further funds for the data collection part of the present research.

Conflicts of Interest

We have no conflicts of interest to declare.

References

Afshin, A., Sur, P. J., Fay, K. A., Cornaby, L., Ferrara, G., Salama, J. S., ... & Murray, C. J. (2019). Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The lancet*, *393*(10184), 1958-1972.
Albon, C. (2018). Machine learning with python cookbook: Practical solutions from preprocessing to deep learning. O'Reilly.
Almoraie, N. M., Saqaan, R., Alharthi, R., Alamoudi, A., Badh, L., & Shatwan, I. M. (2021). Snacking patterns throughout the life span: potential implications on health. *Nutrition Research*, *91*, 81-94.
Arend, A., Kaiser, T., Pannicke, B., Reichenberger, J., Naab, S., Voderholzer, U., & Blechert, J. (2023). Toward Individualized Prediction of Binge-Eating Episodes Based on Ecological Momentary Assessment Data: Item Development and Pilot Study in Patients With Bulimia Nervosa and Binge-Eating Disorder. Journal of Medical Informatics, 11, e41513.
ArunKumar, K., Kalaga, D. V., Mohan Sai Kumar, C., Kawaji, M., and Brenza, T. M. (2022). Comparative analysis of gated recurrent units (gru), long short-term memory (lstm) cells, autoregressive integrated moving average (arima), seasonal autoregressive integrated moving average (sarima) for forecasting covid-19 trends. Alexandria Engineering Journal, 61(10):7585–7603.
Berry, K. J., Mielke Jr, P. W., and Iyer, H. K. (1998). Factorial designs and dummy coding. Perceptual and motor skills, 87(3):919–927.
Boh, B, Lemmens, L. H., Jansen, A., Nederkoorn, C., Kerkhofs, V., Spanakis, G., Weiss, G., & Roefs, A. (2016). An Ecological Momentary Intervention for weight loss and healthy eating via smartphone and Internet: study protocol for a randomised controlled trial. Trials, 17, 154.
Bolnick, H. J., Bui, A. L., Bulchis, A., Chen, C., Chapin, A., Lomsadze, L., ..., & Dieleman, J. L. (2020). Healthcare spending attributable to modifiable risk factors in the USA: an economic attribution analysis. Lancet Public Health, 5, e25-35.
Brietzke, E., Hawken, E. R., Idzikowski, M., Pong, J., Kennedy, S. H., & Soares, C. N. (2019) Integrating digital phenotyping in clinical characterization of individuals with mood disorders. Neuroscience & Biobehavioral Reviews, 104, 223-230.
Brownlee, J. (2016). XGBoost With python: Gradient boosted trees with XGBoost and scikit-learn. Machine Learning Mastery. Available at: https://books.google.com/books/about/XGBoost_With_Python.html?id=HgmqDwAAQBAJ
Cho, C. H, Lee T, Kim, M. G., In, H. P., Kim, L., & Lee, H. J. (2019). Mood Prediction of Patients With

Mood Disorders by Machine Learning Using Passive Digital Phenotypes Based on the Circadian Rhythm: Prospective Observational Cohort Study. Journal of Medical Internet Research, 21: e11029.
Cialdini, R. (2009). Influence: The Psychology of Persuasion. Collins Business Essentials. HarperCollins e-books.
Cifuentes, L., Ghusn, W., Feris, F., Campos, A., Sacoto, D., De la Rosa, A., ... & Acosta, A. (2023). Phenotype tailored lifestyle intervention on weight loss and cardiometabolic risk factors in adults with obesity: a single-centre, non-randomised, proof-of-concept study. *EClinicalMedicine*, *58*.
Cox, W. M., Fadardi, J. S., & Pothos, E. M. (2006). The addiction-Stroop test: Theoretical considerations and procedural recommendations. Psychological Bulletin, 132, 443-476.
Cox, W. M., Fadardi, J. S., Hosier, S. G., & Pothos, E. M. (2015). Differential effects and temporal course of attentional and motivational training on excessive drinking. Experimental and Clinical Psychopharmacology, 23, 445-454.
Dakin, C. A., Finlayson, G., Horgan, G., Palmeira, A. L., Heitmann, B. L., Larsen, S. C., ... & Stubbs, R. J. (2023). Exploratory analysis of reflective, reactive, and homeostatic eating behaviour traits on weight change during the 18-month NoHoW weight maintenance trial. *Appetite*, *189*, 106980.
Ehrampoush, E., Nazari, N., Homayounfar, R., Ghaemi, A., Osati, S., Tahamtan, S., Hashempur, M. H. and Bahmanyar, M. (2020). Association between dietary patterns with insulin resistance in an Iranian population. Clinical nutrition ESPEN, 36, 45-52.
Gardner, B. (2015). A review and analysis of the use of ‘habit’ in understanding, predicting and influencing health-related behaviour. *Health psychology review*, *9*(3), 277-295.
Holzinger, A. (2016). Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Informatics, 3, 119-131.
Forman, E. M., Goldstein, S. P., Zhang, F., Evans, B. C., Manasse, S. M., Butryn, M. L., Juarascio, A. S., Abichandani, P., Martin, G. J., & Foster, G. D. (2019). OnTrack: development and feasibility of a smartphone app designed to predict and prevent dietary lapses. TBM, 9, 236-245.
Fawagreh, K., Gaber, M. M. & Elyan, E. (2014). Random forests: from early developments to recent advancements. Systems Science & Control Engineering, 2, 602-609.
Fuller, W. A. (1986). Measurement Error Models. John Wiley Sons, Inc., USA.
Gurney, K. (2018). An introduction to neural networks. CRC press.
Hardeman, W., Houghton, J., Lane, K., Jones, A., & Naughton, F. (2019). A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. *International Journal of Behavioral Nutrition and Physical Activity*, *16*(1), 1-21.
He, J., Mao, R., Shao, Z., and Zhu, F. (2020). Incremental learning in online scenario. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 13926–13935.
Hojjatinia, S., Hojjatinia, S., Lagoa, C. M., Brunke-Reese, D., & Conroy, D. E. (2021). Person-specific dose-finding for a digital messaging intervention to promote physical activity. Health Psychology, 40, 502-512.
Hyndman, R. and Athanasopoulos, G. (2014). Forecasting: principles and practice. Otexts.
Insel, T. R. (2018). Digital phenotyping: a global tool for psychiatry. World Psychiatry, 17, 276-277.
Kahneman, D. (2001). Thinking fast and slow. Penguin: London, UK.
Institute for Health Metrics and Evaluation (IHME). (2020). GBD Compare Data Visualization.

Seattle, WA: IHME, University of Washington, Available from http://vizhub.healthdata.org/gbd-compare. (Accessed 28 Sept. 2023)

Jenkins, K. T., & Tapper, K. (2014). Resisting chocolate temptation using a brief mindfulness strategy. *British journal of health psychology*, *19*(3), 509-522.

Lau, Y., Guang, D. H. C., Chow, X. P., Cheng, L. J., & Wong, S. N. (2020). Personalised eHealth interventions in adults with overweight and obesity: A systematic review and meta-analysis of randomised controlled trials. Preventive Medicine, 132,106001.

Pagan, M., Zarlis, M. and Candra, A., 2023. Investigating the impact of data scaling on the k-nearest neighbor algorithm. Computer Science and Information Technologies, 4(2), pp.135-142.

Myers-Ingram, R., Sampford, J., Milton-Cole, R. et al. (2023). Effectiveness of eHealth weight management interventions in overweight and obese adults from low socioeconomic groups: a systematic review. Systematic Reviews, 12, 59, https://doi.org/10.1186/s13643-023-02207-3.

O'Boyle, J. & Davidson, P. (2022). The Effects of mHealth Versus eHealth on Weight Loss in Adults: A Systematic Review. Topics in Clinical Nutrition 37, 69-84.

Perski, O., Hébert, E. T., Naughton, F., Hekler, E. B., Brown, J., & Businelle, M. S. (2022). Technology‐mediated just‐in‐time adaptive interventions (JITAIs) to reduce harmful substance use: a systematic review. *Addiction*, *117*(5), 1220-1241.

Tammina, S., 2019. Transfer learning using vgg-16 with deep convolutional neural network for classifying images. International Journal of Scientific and Research Publications (IJSRP), 9(10), pp.143-150.

Lally, P. and Gardner, B. (2013). Promoting habit formation. Health psychology review, 7(sup1), S137-S158.

Lee, T. H., Ullah, A. & Wang, R. (2020). Bootstrap aggregating and random forest. Macroeconomic forecasting in the era of big data: Theory and practice. Springer.

Li, Y., Li, M., Li, C. & Liu, Z. (2020). Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms. Scientific reports, 101, 1-12.

Marteau, T. M., Fletcher, P. C., Munafò, M. R., & Hollands, G. J. (2021). Beyond choice architecture: advancing the science of changing behaviour at scale. *BMC Public Health*, *21*(1), 1-7.

McArthur, L.H., Holbert, D. and Forsythe, W. (2012). Is field of study or location associated with college students' snacking patterns?. Journal of obesity, , 2012, pp. 1–10. Doi:10.1155/2012/297195.

Mishra, V., Kuenzler, F., Kramer, J., Fleisch, E., Kowatsch, T., & Kotz, D. (2021). Detecting receptivity for mHealth interventions in the natural environment. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 5, 74.

Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., & Murphy, S. A. (2018). Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine*, *52*(6), 446-462.

Naughton, F., Hopewell, S., Lathia, N., Schalbroeck, R., Brown, C., Mascolo, C., McEwen, A. & Sutton, S. (2016). A context-sensing mobile phone app (Q Sense) for smoking cessation: a mixed-methods study. JMIR mHealth and uHealth, 4, 3, e106.

Naughton, F., Cooper, S., Foster, K., Emery, J., Leonardi-Bee, J., Sutton, S., Jones, M., Ussher, M., Whitemore, R., Leighton, M., Montgomery, A., Parrott, S. & Coleman, T. (2017). Large multi-centre

pilot randomized controlled trial testing a low-cost, tailored, self-help smoking cessation text message intervention for pregnant smokers (MiQuit). Addiction, 112, 1238–1249.
Neal, D. T., Wood, W., Wu, M., & Kurlander, D. (2011). The pull of the past: When do habits persist despite conflict with motives?. *Personality and social psychology bulletin*, *37*(11), 1428-1437.
OECD (2019). The heavy burden of obesity: The economics of prevention. OECD Health Policy Studies, OECD Publishing, Paris, doi .org/10.1787/67450d67-en.
Oudin, A., Maatoug, R., Bourla, A., Ferreri, F., Bonnot, O., Millet, B., Schoeller, F., Mouchabac, S., & Adrien, V. (2023). Digital Phenotyping: Data-Driven Psychiatry to Redefine Mental Health. Journal of Medical Internet Research, 25: e44502.
Pentikäinen, S., Arvola, A., Karhunen, L., & Pennanen, K. (2018). Easy-going, rational, susceptible and struggling eaters: A segmentation study based on eating behaviour tendencies. *Appetite*, *120*, 212-221.
Perski, O., Hébert, E. T., Naughton, F., Hekler, E. B., Brown, J., & Businelle, M. S. (2022). Technology-mediated just-in-time adaptive interventions (JITAIs) to reduce harmful substance use: a systematic review. *Addiction*, *117*(5), 1220-1241.
Pettinati, H.M., Volpicelli, J.R., Pierce Jr, J.D. and O'brien, C.P., 2000. Improving naltrexone response: an intervention for medical practitioners to enhance medication compliance in alcohol dependent patients. Journal of Addictive Diseases, 19, 71-83.
Pimpini, L., Franssen, S., Reber, P., Jansen, A., & Roefs, A. (2023). Effects of a health versus hedonic mindset on daily-life snacking behaviour. Journal of Human Nutrition and Dietetics, 36, 2085-2098.
Prakash, J., Chaudhury, S., & Chatterjee, K. (2021). Digital phenotyping in psychiatry: When mental health goes binary. Industrial Psychiatry Journal, 30, 191-192.
Shah, J. Y., Friedman, R., & Kruglanski, A. W. (2002). Forgetting all else: on the antecedents and consequences of goal shielding. *Journal of personality and social psychology*, *83*(6), 1261.
Sharma, S., Sharma, S. & Athaiya, A. (2017). Activation functions in neural networks. Towards Data Science, 6, 310-316.
Stroebe, W. (2022). The goal conflict model: a theory of the hedonic regulation of eating behavior. *Current Opinion in Behavioral Sciences*, *48*, 101203.
Spanakis, G., Weiss, G., Boh, B., Lemmens, L., & Roefs, A. (2017). Machine learning techniques in eating behaviour e-coaching. Personal Ubiquitous Computing, 21, 645-659.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15, 1929-1958.
Tapper, K., Baker, L., Jiga-Boy, G., Haddock, G., & Maio, G. R. (2015). Sensitivity to reward and punishment: Associations with diet, alcohol consumption, and smoking. *Personality and Individual Differences*, *72*, 79-84.
Tiffany, S. T. (1990). A cognitive model of drug urges and drug-use behavior: Role of automatic and non-automatic processes. Psychological Review, 97, 147-168.
Tulu B, Ruiz C, Allard J, et al. (2017). SlipBuddy: a mobile health intervention to prevent overeating. Proceedings of the 50th Hawaii International Conference on System Sciences.
Verhoeven, A. A., Adriaanse, M. A., Evers, C., & de Ridder, D. T. (2012). The power of habits:

Unhealthy snacking behaviour is primarily predicted by habit strength. *British Journal of Health Psychology*, *17*(4), 758-770.
Van Houdt, G., Mosquera, C. & Nápoles, G. (2020). A review on the long short-term memory model. Artificial Intelligence Review, 53, 5929-5955.
Wilson, E., Senior, V., & Tapper, K. (2021). The effect of visualisation and mindfulness-based decentering on chocolate craving. *Appetite*, *164*, 105278.
Wyatt, P., Berry, S. E., Finlayson, G., O'Driscoll, R., Hadjigeorgiou, G., Drew, D. A., ... & Valdes, A. M. (2021). Postprandial glycaemic dips predict appetite and energy intake in healthy individuals. *Nature metabolism*, *3*(4), 523-529. https://www.nature.com/articles/s42255-021-00383-x

Abbreviations

BCT: behavior change technique
CV: cross-validation
EMA: ecological momentary assessment
FFNN: feed forward neural network
HFSS: high fat, salt and/ or sugar
LSTM: long short-term memory
JITAI: just in time interventions
MAE: mean absolute error
ML: machine learning
SMS: short message service
RFreg: random forest regressor
XGBreg: extreme gradient boosting regressor

Electronic Supplementary Material

Appendix 1. Additional information concerning methods in the present work.

The questions below were given to participants during the first, screening part of the study.

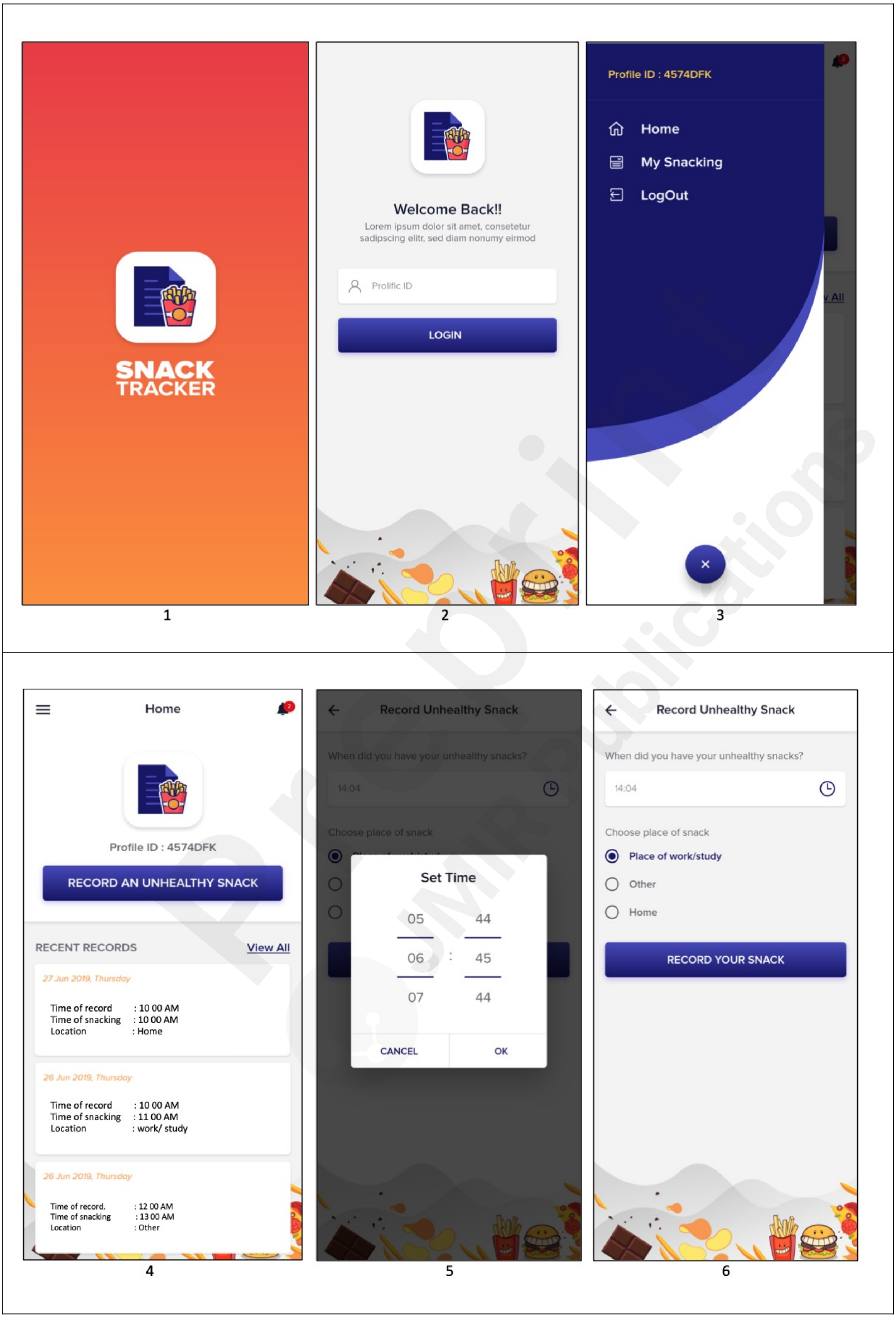
Table A1.1. Demographics questions in the first part of the data collection procedure.

Question	**Cardinality**	**Choices\value**
What is your age?	-	18-60 Male, Female, Other White Mixed/Multiple ethnic groups Asian/Asian British Black/African/Caribbean/Black British Other ethnic group Unknown Prefer not to say
What is your gender?	3	In paid employment Studying full time Unemployed Homemaker Retired
Please indicate your ethnicity:	7	Other (please specify)
Please indicate your employment status:	5	18-60 Male, Female, Other White Mixed/Multiple ethnic groups Asian/Asian British Black/African/Caribbean/Black British Other ethnic group Unknown Prefer not to say
What is your weight?	-	In paid employment Studying full time Unemployed Homemaker Retired
What is your height?	-	Other (please specify)

Table A1.2. Questions regarding motivation for healthy eating in the survey

Question	**Cardinality**	**Choices\value**
1. On a typical day, approximately how - many sugary, salty or fatty snacks do you usually consume in between main meals?	-	-
2. Are you currently dieting to lose 3 weight? (dieting)	3	Yes , No, I'd rather not say
3. It is important to me to watch my weight (IWW)	7	1-7 scale anchored by 'strongly disagree' and 'strongly agree'
4. It is important to me to eat a healthy diet (IEHD)	7	1-7 scale anchored by 'strongly disagree' and 'strongly agree'

NB. Questions 2 and 3 were Likert scales, with a number of responses as indicated.


SNACK TRACKER
1
Welcome Back!!
Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod
Prolific ID
LOGIN
2
Profile ID : 4574DFK
Home
My Snacking
LogOut
3
Home
Profile ID : 4574DFK
RECORD AN UNHEALTHY SNACK
RECENT RECORDS
View All
27 Jun 2019, Thursday
Time of record : 10 00 AM
Time of snacking : 10 00 AM
Location : Home
26 Jun 2019, Thursday
Time of record : 10 00 AM
Time of snacking : 11 00 AM
Location : work/ study
26 Jun 2019, Thursday
Time of record. : 12 00 AM
Time of snacking : 13 00 AM
Location : Other
4
Record Unhealthy Snack
When did you have your unhealthy snacks?
14:04
Choose place of snack
Set Time
05 44
06 : 45
07 44
CANCEL
OK
5
Record Unhealthy Snack
When did you have your unhealthy snacks?
14:04
Choose place of snack
Place of work/study
Other
Home
RECORD YOUR SNACK
6

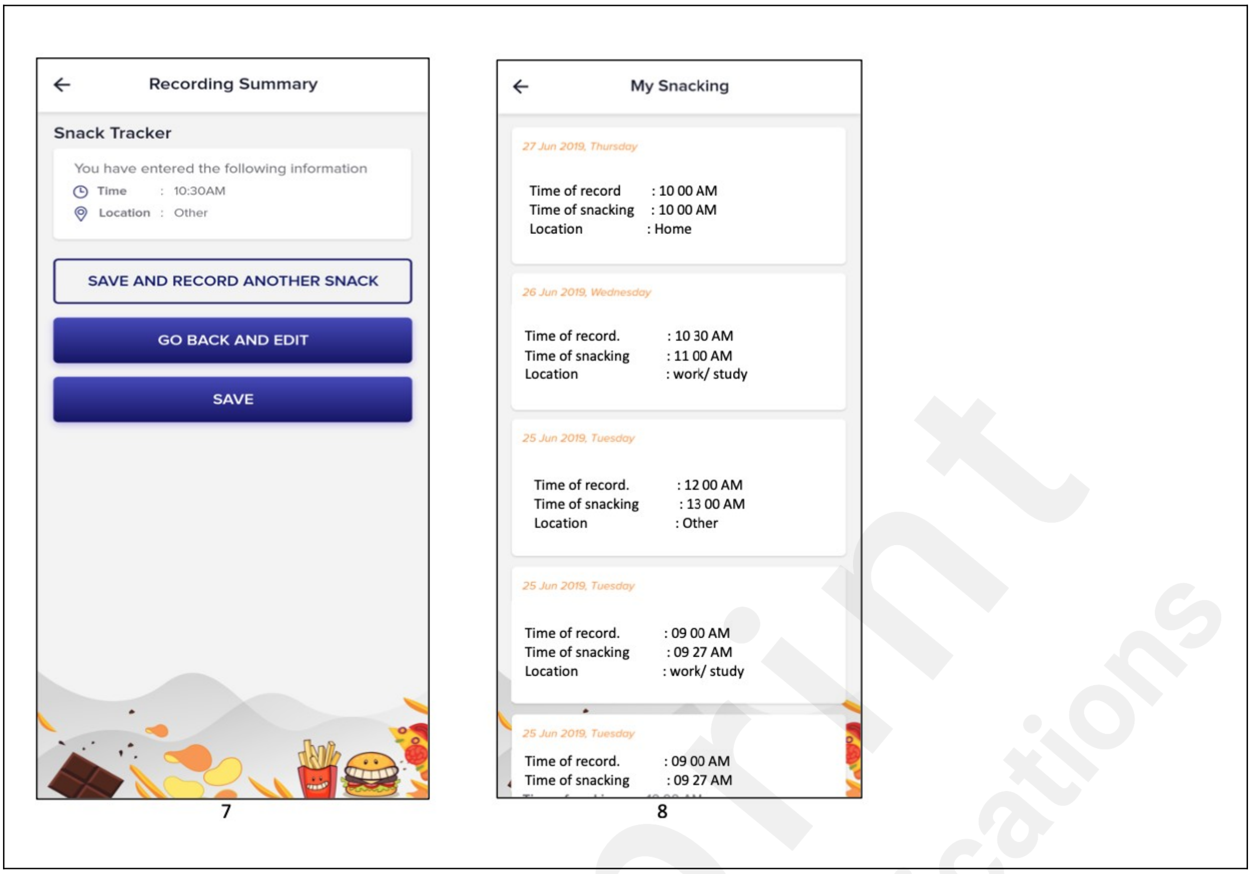


Figure A1.1. (larger version of Figure 3). The frontend of the Snack Tracker app: 1, splash screen; 2, login screen; 3, home screen; 4, new snack screen; 5, time recording screen (time picker); 6, location recording screen; 7, recording save screen; 8, review recordings summary.

<u>Full instructions concerning how the idea of HFSS snacks was explained to participants.</u>

In the invitation message to participate in the second part of the study, we informed participants about the specific task and how to record their snacking behavior. The relevant part of the message that was employed in the invitation is as follows:

“Thank you for completing the first part of our study. We would like to invite you to take part in the second part! The second part takes place over a period of 28 days. All you have to do is to simply record every instance of snacking behavior. We are interested in snacks high in sugar, salt, or fat, which includes:

*Sugary snacks: biscuits, cake, chocolate, sweets, sugary desserts, sugary breakfast cereals, sugary granola bars, sodas, milkshakes, flavoured milk drinks.
*Salty and fatty snacks: crisps, salted nuts, salted popcorn, pretzels, chips, burgers, cheese.

We are using the word ‘snack’ to refer to any food that is eaten in between main meals. Every time you have a snack high in sugar, salt, or fat, you will have to use an App that will be provided to you,

just record the location (work/ study place, home, or other) and the time of the snack – and that's it! Recording a snack should take only a few seconds. You will be paid £16 on completion.
...”

Additional information regarding checks concerning participant engagement with the study

After approximately 15 days of their participation in the second part of the study, we sent a message to participants to (a) Check the average number of recordings per day to ensure data completeness (b) Verify if participants were receiving daily SMS reminders to stay on track.

In addition, some participants received a message if we noticed any anomalies, such as days with no recordings or a significant drop in the number of recorded snacks. These steps were taken to maintain data accuracy and consistency throughout the study and to provide participants with an opportunity to address any issues or discrepancies that might have arisen during the data collection process.

Appendix 2. Model residuals concerning the objective of predicting the time until the next unhealthy snack.

We show histograms of the residuals of all considered models, for predicting the time until the next unhealthy snack. In Figures A2.1-A2.5, the histograms are reduced to a range of -200 to +200 minutes, as larger errors are outliers in the sense that these predictions are not useful for the purpose of timing interventions. The residuals are calculated as predicted minus true values, so that negative residuals indicate that the prediction is earlier than the observed time and positive residuals indicate that the predicted time is later than observed.

Figure A2.1 shows the histogram of residuals for linear regression for the UK and Dutch datasets. With the UK dataset and 4 time bins, the distribution is left skewed. As noted, arguably, this is a good trait for putative interventions, as earlier predictions are less problematic than late ones, as long as the deviation is not too extreme. The residuals were less skewed with 12 time bins in the UK dataset. On the Dutch dataset, the width of the distributions was large with both 4 and 12 time bins, so that this model in this context would not be useful for prediction.

For both the RFreg and and XGBreg models, residual histograms reveal distributions mostly close to normal, across both data sets and time bins (Figures A2.2, A2.3). Figure A2.4 shows that for the FFNN model, the histogram of residuals was skewed for the UK dataset with 4 time bins. In addition, for the same model residuals for the UK and Dutch dataset with 4 time bins seem to diverge from a normal distribution. However, the residuals distribution was close to normal and less skewed with 12 time bins in both datasets.

Finally, residuals for the LSTM model (Figure A2.5) also diverged from normality for the UK dataset, but this was not the case for the Dutch dataset, for which there is high conformity to a normal distribution.

Overall, it appears that fixed context models resulted in distributions of residuals with better properties, compared to recurrent models. RFreg residuals conformed to a normal distribution model more so, compared to the other fixed context models (XGBreg and FFNN). Also, and without offering detailed analyses, there is a good impression that the better performing models were associated with residuals mostly distributed to within an hour or less of actual time (of the next unhealthy snack), which is encouraging. Additionally, for most models, when there is a skew in residual distributions, this is a left skew, which is better compared to a right skew (as argued elsewhere).

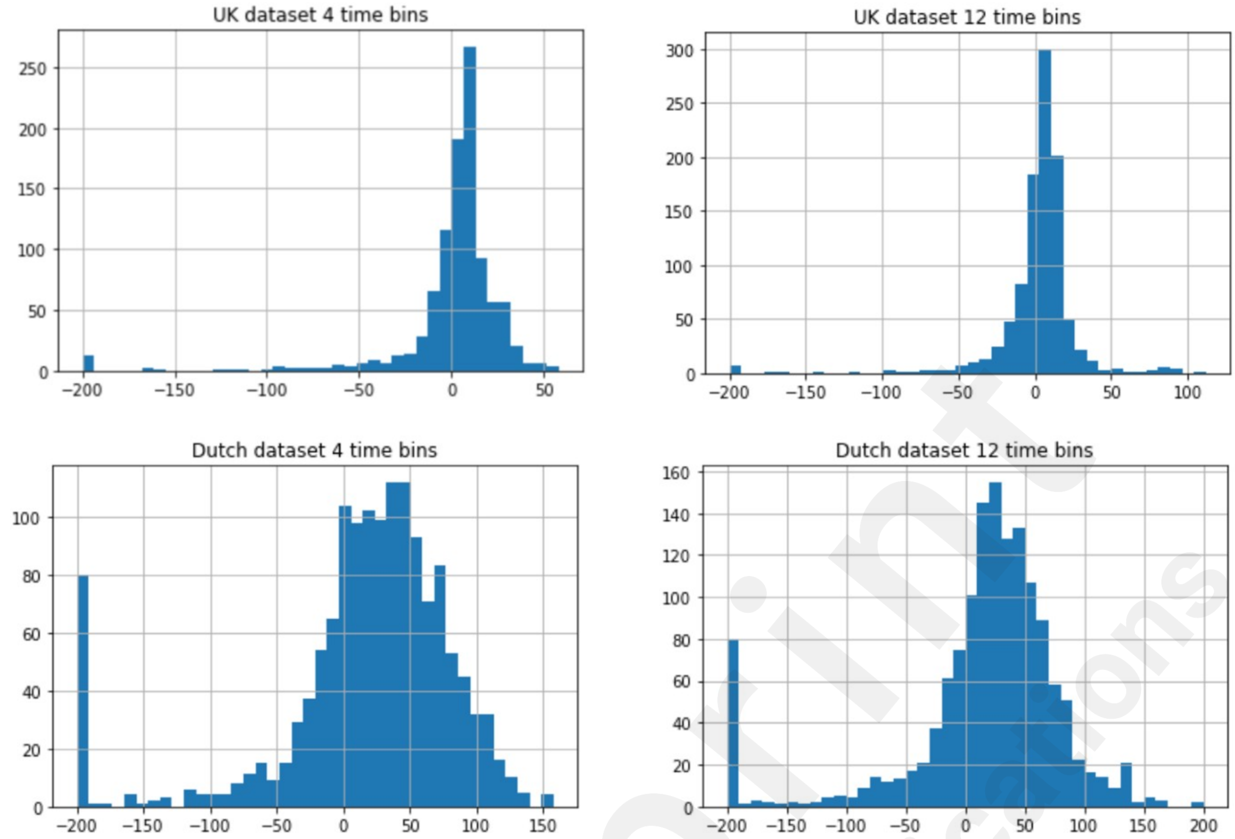


Figure A2.1. Histogram of residuals for the linear regression model, when using 4 and 12 time bins, for the UK and Dutch datasets.

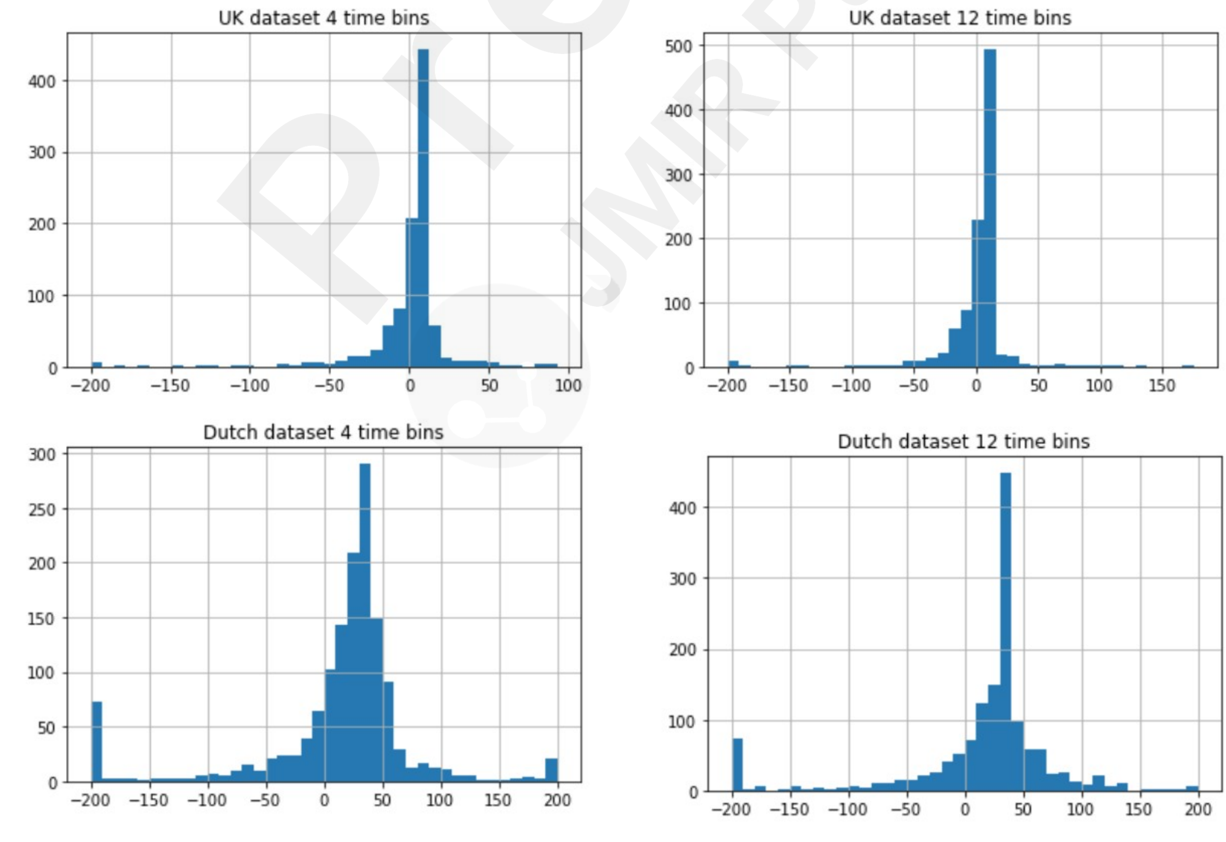


Figure A2.2. Histogram of residuals for the RFreg model, when using 4 and 12 time bins, for the UK and Dutch datasets.

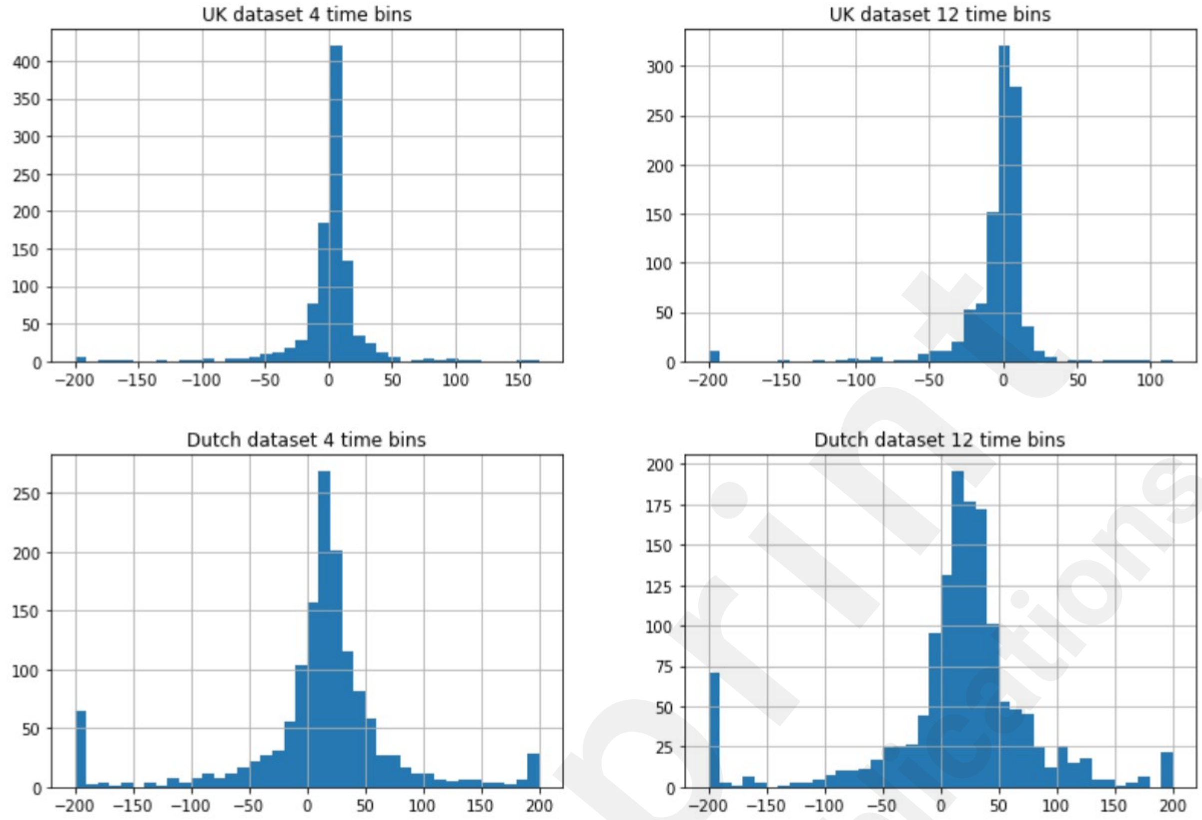


Figure A2.3. Histogram of residuals for the XGBreg model, when using 4 and 12 time bins, for the UK and Dutch datasets.

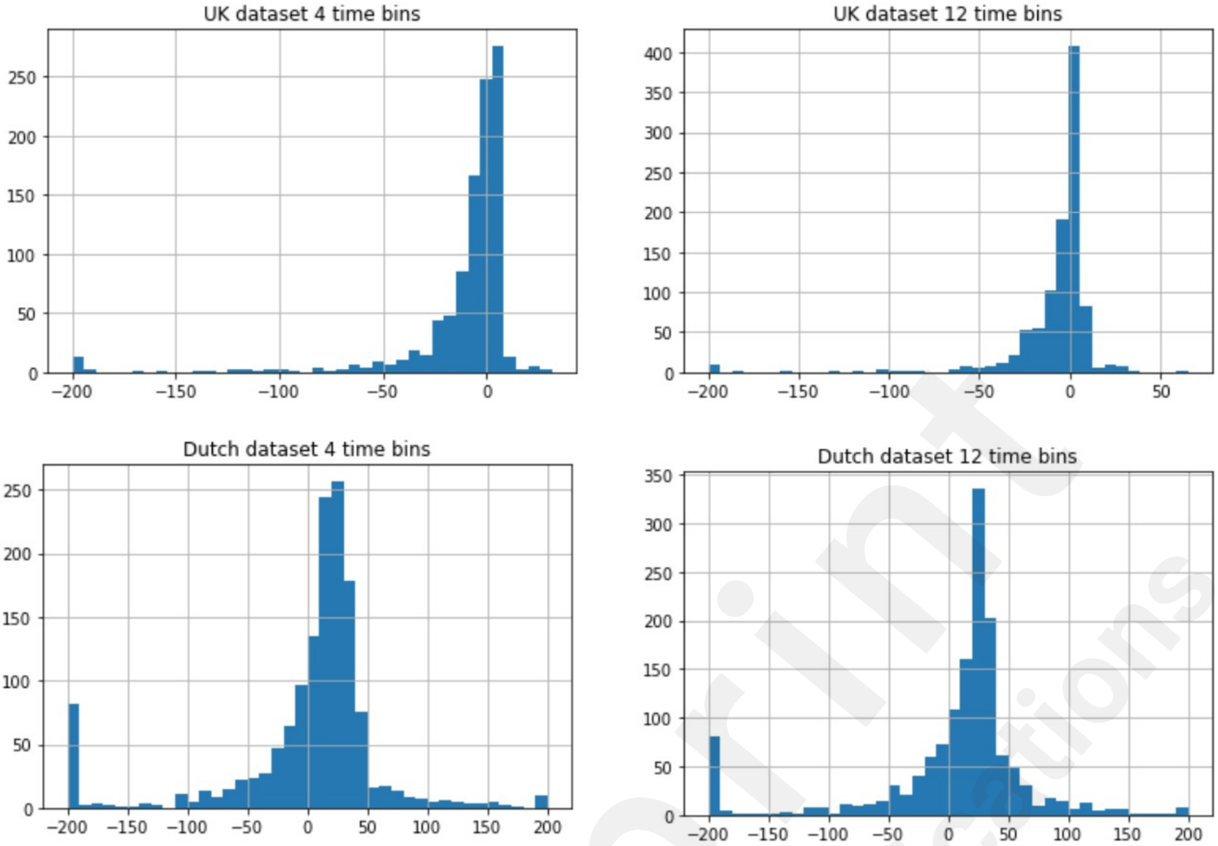


Figure A2.4. Histogram of residuals for the FFNN model, with 4 and 12 time bins, for the UK and Dutch datasets.

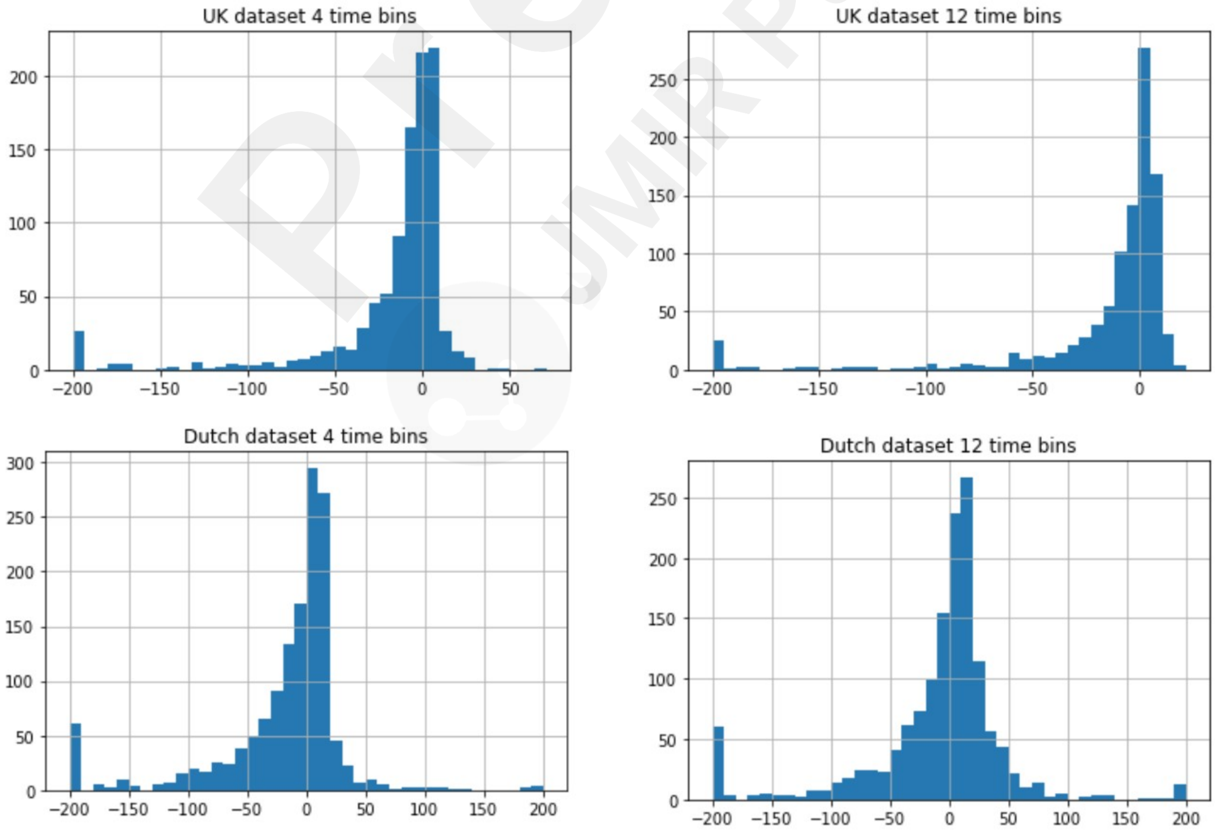


Figure A2.5. Histogram of residuals for the LSTM model, with 4 and 12 time bins, for the UK and Dutch datasets.

Appendix 3. An alternative predictive objective.

Health psychologists and related professionals might be interested in this work for the purpose of supporting behaviour change with timed interventions. Here, we propose an analysis arguably more geared towards such applications, based on the specific timing criteria desirable for a putative behaviour change intervention. That is, even though we do not employ an intervention, we offer an analysis concerning timings, as would be useful for constructing an intervention.

The logic of this analysis is as follows. An influential distinction is to understand thought as reflecting two types of processes, reflective and impulsive. We can consider human behaviour change in terms of these processes. It is thought that unhealthy snacking may often be an automatic process (Cleobury & Tapper, 2014; Neal et al., 2011; Verhoeven et al., 2012). Accordingly, many researchers agree on the importance of the delivery time of interventions (e.g., Hardeman et al., 2019; Intille et al., 2003). For example, the effectiveness of interventions has been demonstrated to vary depending on the timing of the delivery. Information we have been exposed to more recently tends to be more cognitively accessible. Nudges presented too early, relative to the targeted action, may be ineffective, because of forgetting, while nudges presented late may reduce the amount of time available for processing and executing a change (Gillitzer & Sinning, 2020). In the context of predictive models, there is also the risk of the predicted time being later than the intervention, so that the intervention happens after the targeted event and the opportunity for change has been missed. In general, a nudge offered a short time before the behaviour is desirable, but this has to be balanced against the risk of being too late.

Here, we evaluate the Random Forest Regressor (RFreg) model's performance in addressing the problem of predicting timing in unhealthy snacking behaviour, by considering the distribution of residuals concerning the time for the next unhealthy snack and by producing the outcome with different timing parameters. That is, we want to consider the ability of RFreg to predict unhealthy snacks within windows of a certain size. Note, even though RFreg did not provide the best mean absolute errors, its performance was better in other ways: notably, residuals conformed more closely to a normal distribution model, which is a desirable property for putative interventions.

We offer a sketch for the structure of a putative intervention, so as to make the prediction problem more precise. We define as Predicted Unhealthy Snacking Time (PST) the time until the next unhealthy snack. Then, we define the Proposed Intervention Time (PIT) by subtracting the Lead Time (LT) from PST, i.e., PIT = PST – LT. The idea is that LT would be used to control the timing of the intervention and determine how far in advance an intervention should be delivered. We use LT values of 0, 20, 30, 45, 60, 75, and 90 minutes. Different LT values would allow a researcher to determine the ideal balance between time precision and accuracy (in terms of percentage, see shortly), in relation to putative interventions for preventing unhealthy snacking.

We next define the Time Difference (TD) by subtracting the PIT from the Actual Unhealthy Snacking Time (AUT), i.e., TD = AUT- PIT. TD is the difference between the time at which the (putative) intervention was delivered and the time at which the person actually snacked. TD allows us to evaluate the precision of predictions (i.e., the PITs) and, specifically, whether they were within, before, or after time windows of differing widths, W, of 30, 60 and 120 minutes. For example, for W=30, we consider the intervention as 'Hitting' if snacking and intervention cooccur within a window of 30 minutes. Furthermore, we categorised PIT values as Early and Late accordingly. A hitting score can be defined as the percentage of instances where the hypothetical intervention's timing aligned with the actual unhealthy snacking occurrence – this is basically a measure of the accuracy of the intervention's timing. For example, a hitting score of X% with LT=10 minutes suggests that the intervention was effective in being appropriately timed in X% of cases, assuming we aimed for the intervention to occur within 10 minutes prior to the unhealthy snack.

Each of the figures below shows the effectiveness of the hypothetical intervention, for different LT values, as well as Early, Hitting, or Late results, with respect to W values of 30, 60, and 120 minutes to the time window. We also separate results by time bin (4 and 12 time bins).

Figure A3.1 shows hitting results when W=30. Across different LT values, for the UK dataset with 12 time bins, we observed fairly good performance (hitting score as high as 83%, with LT=10 minutes). By contrast, the highest hitting score for the Dutch dataset with 12 time bins was only 49%, with LT= 45. Figure A3.2 shows corresponding results with W=60. For the UK datasets, the most notable performance was 97% (with LT values of 20, 30, and 45 minutes), when employing 12 time bins. For the Dutch dataset, peak performance was 70% (with an LT of 60 minutes), when employing 4 time bins. Considering W=120, in Figure A3.3, we observed remarkable success rates within the UK dataset. The best success rate was 97% for LT values of 20, 30, and 45 minutes, with a 12-time bins. For the Dutch dataset, the highest performance observed was 84% (with an LT of 75 minutes), using 12 time bins.

These results show that hitting scores varied considerably between the UK and Dutch datasets, as well as within each dataset, with different LT values leading to best scores in different cases. Note, it is a not a simple case that smaller or greater LT values would in general lead to better or worse performance, since this interacts with whether the prediction residuals are positively or negatively skewed. A broadly analogous point applies to whether the analysis is carried out against data organized with 12 or 4 time bins. Overall, hitting scores were slightly higher with 12 time bins, except in the Dutch dataset with W=60 minutes (LT=60 minutes). Hitting scores were high for the UK dataset in all time windows, W=30, 60, 120, though slightly higher in the last two cases.

In summary, the shortest window time we investigated was 30 minutes. For this time window, we think that the modeling results show reasonable accuracy for hypothetical interventions, of 83% and 49% for the UK and Dutch datasets respectively. In pilot results not reported here, shorter time windows (e.g., 20 or 15 minutes), were associated for performance too poor to be useful. Note, qualifications such as 'good enough', 'useful' etc. are offered as very preliminary: it is a difficult empirical challenge to establish what is a sufficient percentage of hits before an intervention is effective. Analogously, it is unclear what is the effective range between intervention and targeted behavior, a question further complicated by the fact that a behavior such as unhealthy snacking would be very noisy by itself. Finally, the choice of RFreg for this analysis was reasonable, but it has to be remembered that, in our current data, all models performed fairly equivalently. We think that interesting model differences would be evident only with larger data collection exercises, perhaps over a period of months, rather than weeks.

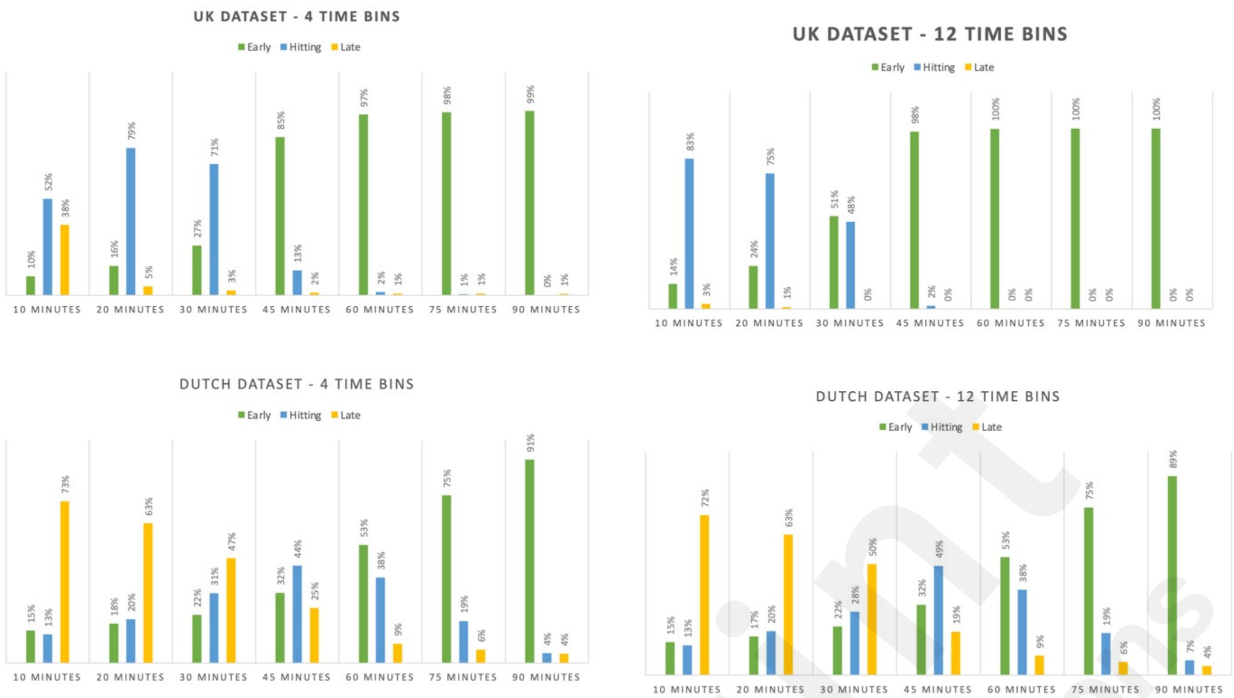


Figure A3.1. Early, hitting, and late results distribution for a 30-minute window for accepting an intervention, with 4 and 12 time bins, in the UK and Dutch datasets. The hitting score corresponds to the percentages in the 'hitting' column, in each case.

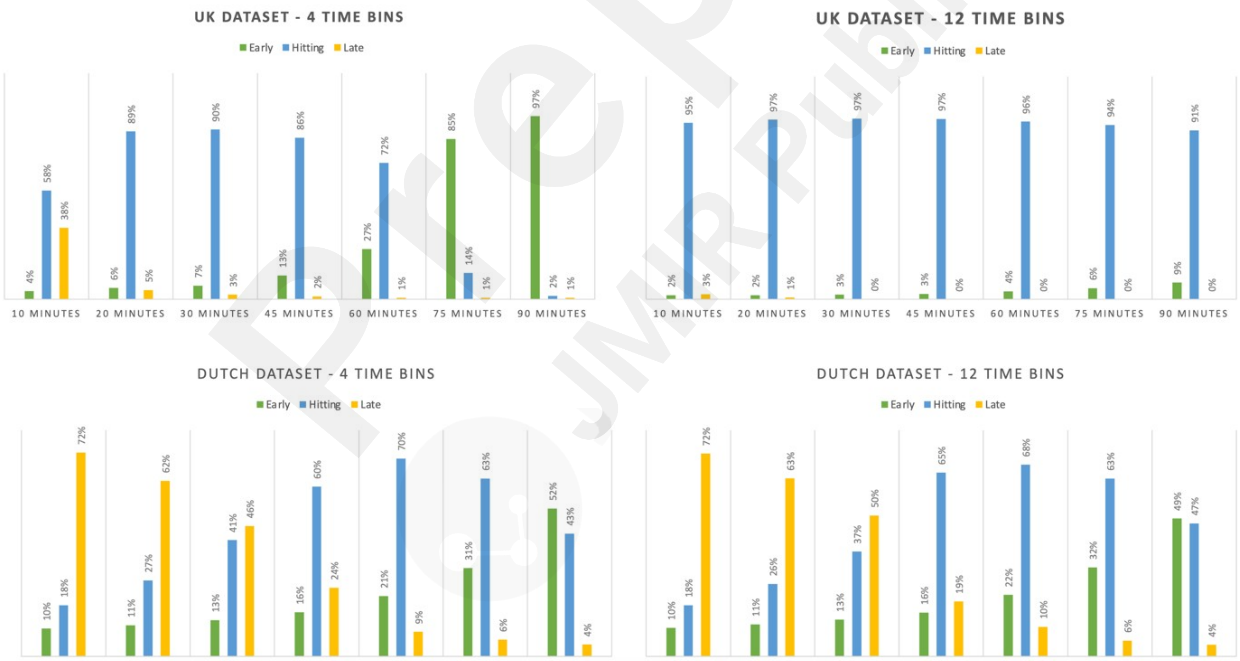


Figure A3.2. Early, hitting, and late results for a 60-minute window for accepting an intervention, with 4 and 12 time bins, in the UK and Dutch datasets. The hitting score corresponds to the percentages in the 'hitting' column, in each case.

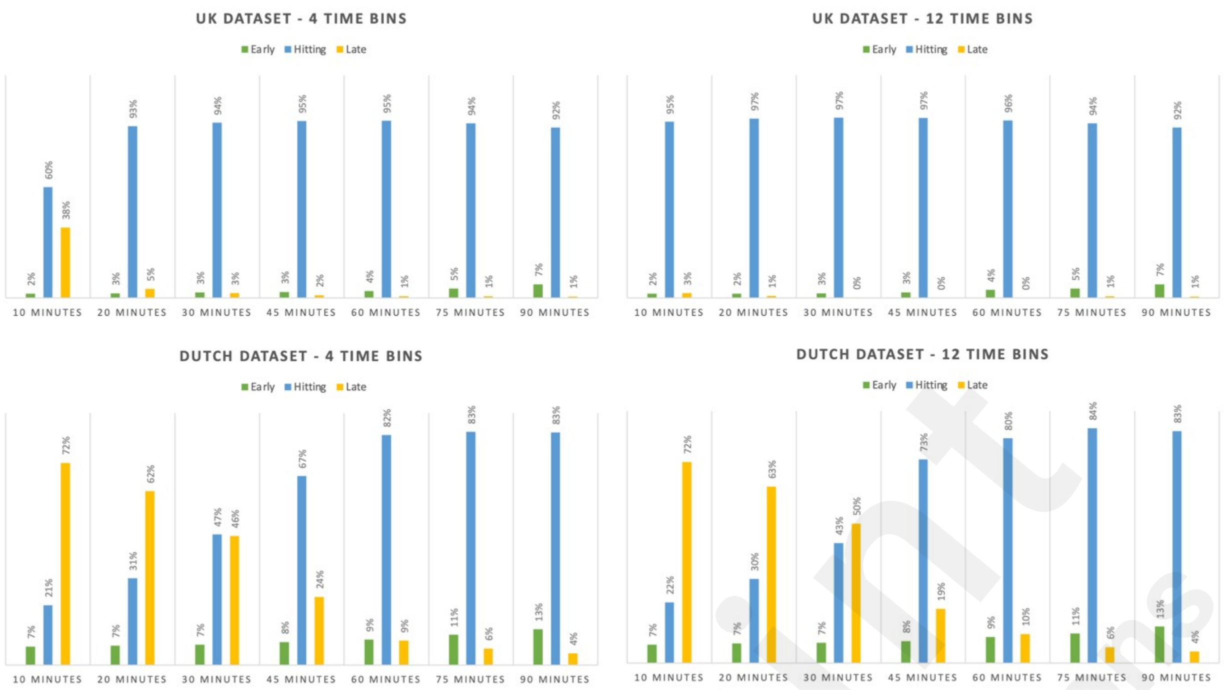


Figure A3.3. Early, hitting, and late results for a 120-minute window for accepting an intervention, with 4 and 12 time bins, in the UK and Dutch datasets. The hitting score corresponds to the percentages in the ‘hitting’ column, in each case.

Appendix 4. Additional references.

Hardeman, W., Houghton, J., Lane, K., Jones, A. & Naughton, F. (2019). A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. International Journal of Behavioral Nutrition and Physical Activity, 16, 1-21.

Intille, S. S., Kukla, C., Farzanfar, R., & Bakr, W. (2003). Just-in-time technology to en- courage incremental, dietary behavior change. In AMIA Annual Symposium Proceedings, volume 2003, page 874. American Medical Informatics Association.

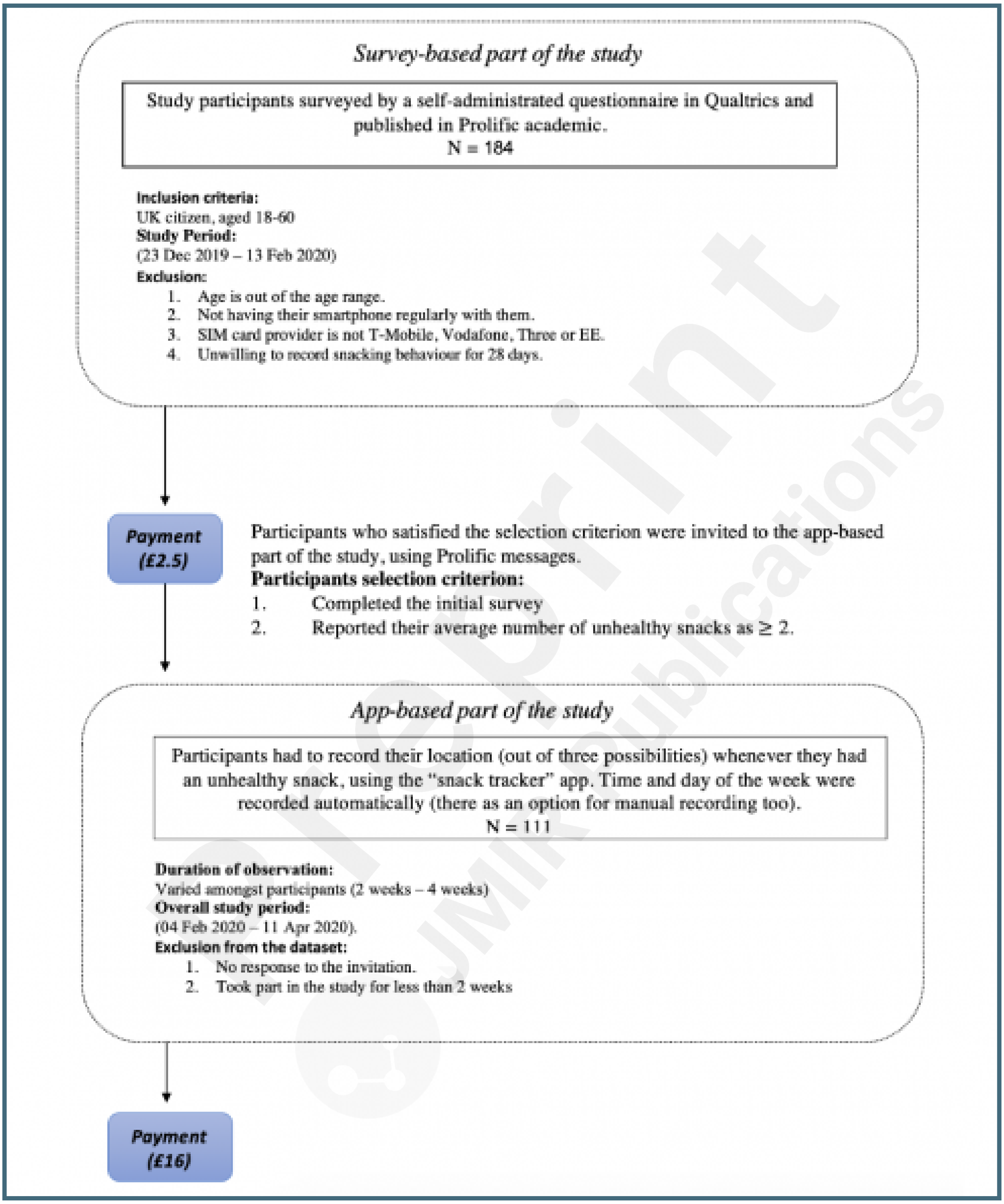
Gillitzer, C. and Sinning, M. (2020). Nudging businesses to pay their taxes: Does timing matter? Journal of Economic Behavior & Organization, 169(C):284–300.

Tapper, K. (2021). Health Psychology and Behaviour Change: From Science to Practice. Macmillan Education UK, 1 edition.

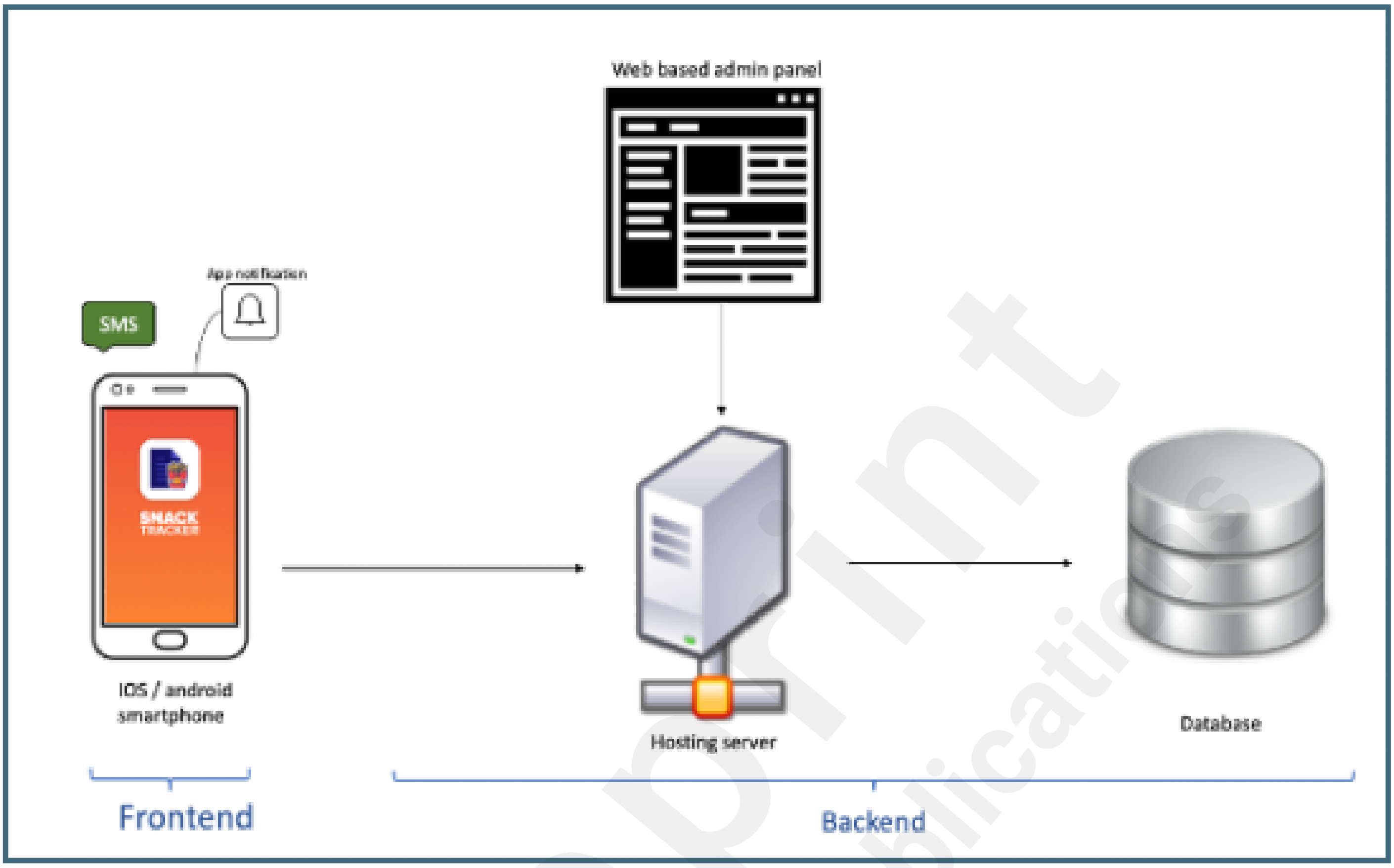
Supplementary Files

Figures

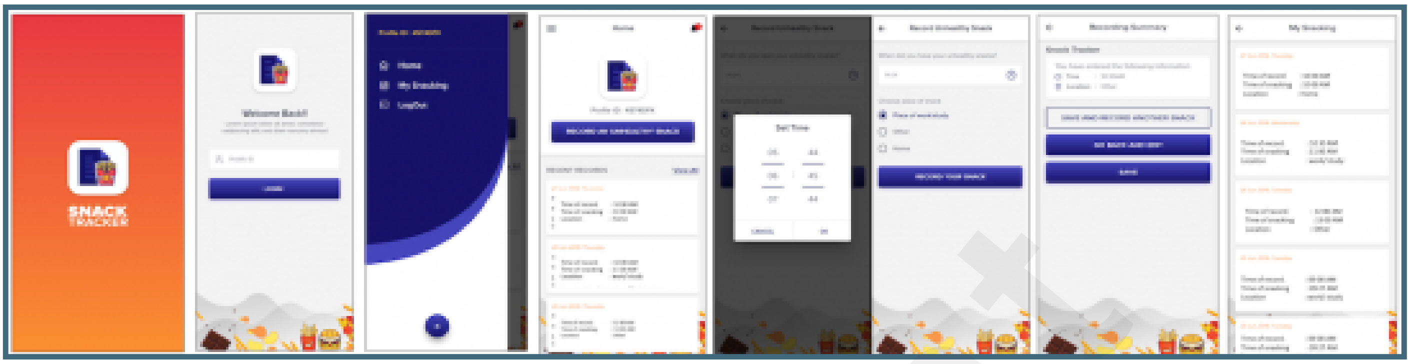
Outline of the two parts of the data collection, for the UK dataset.



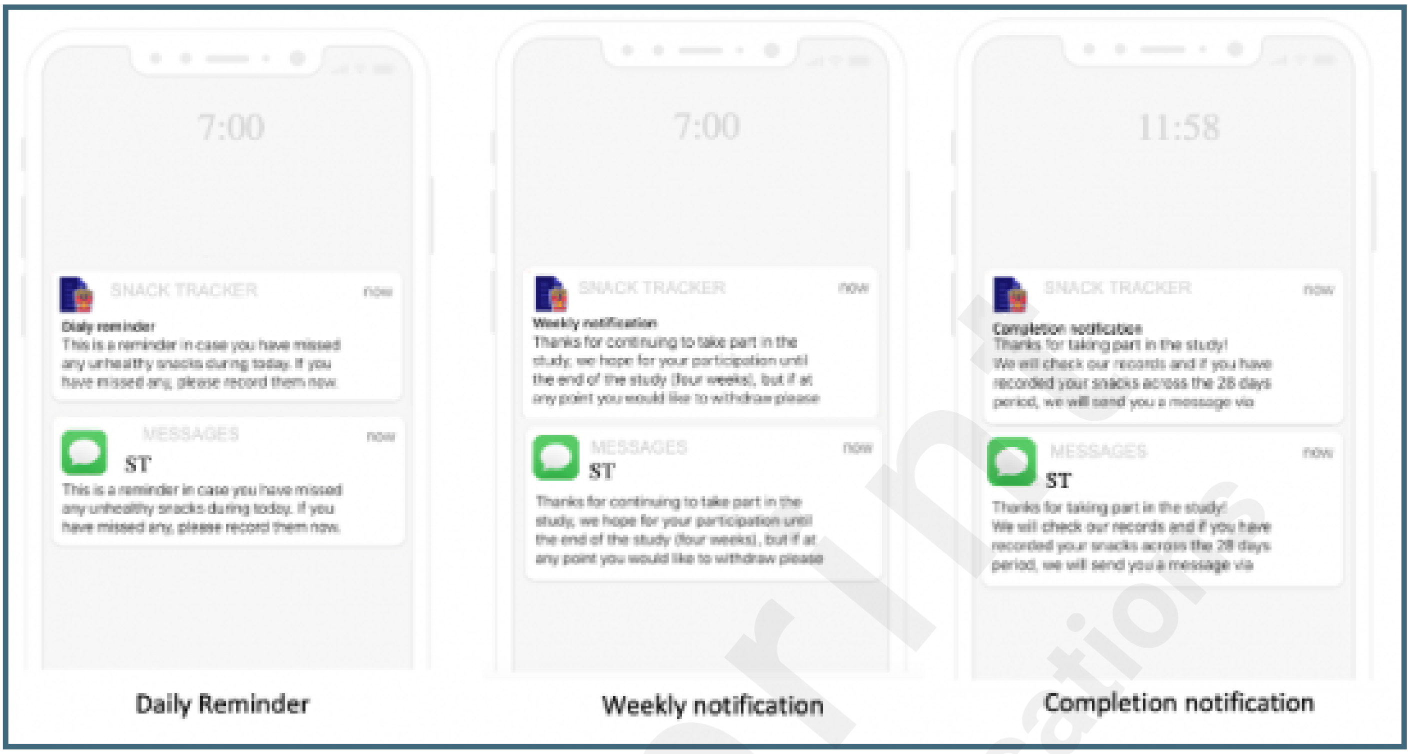
The frontend and backend of the Snack Tracker app, illustrating data flow.



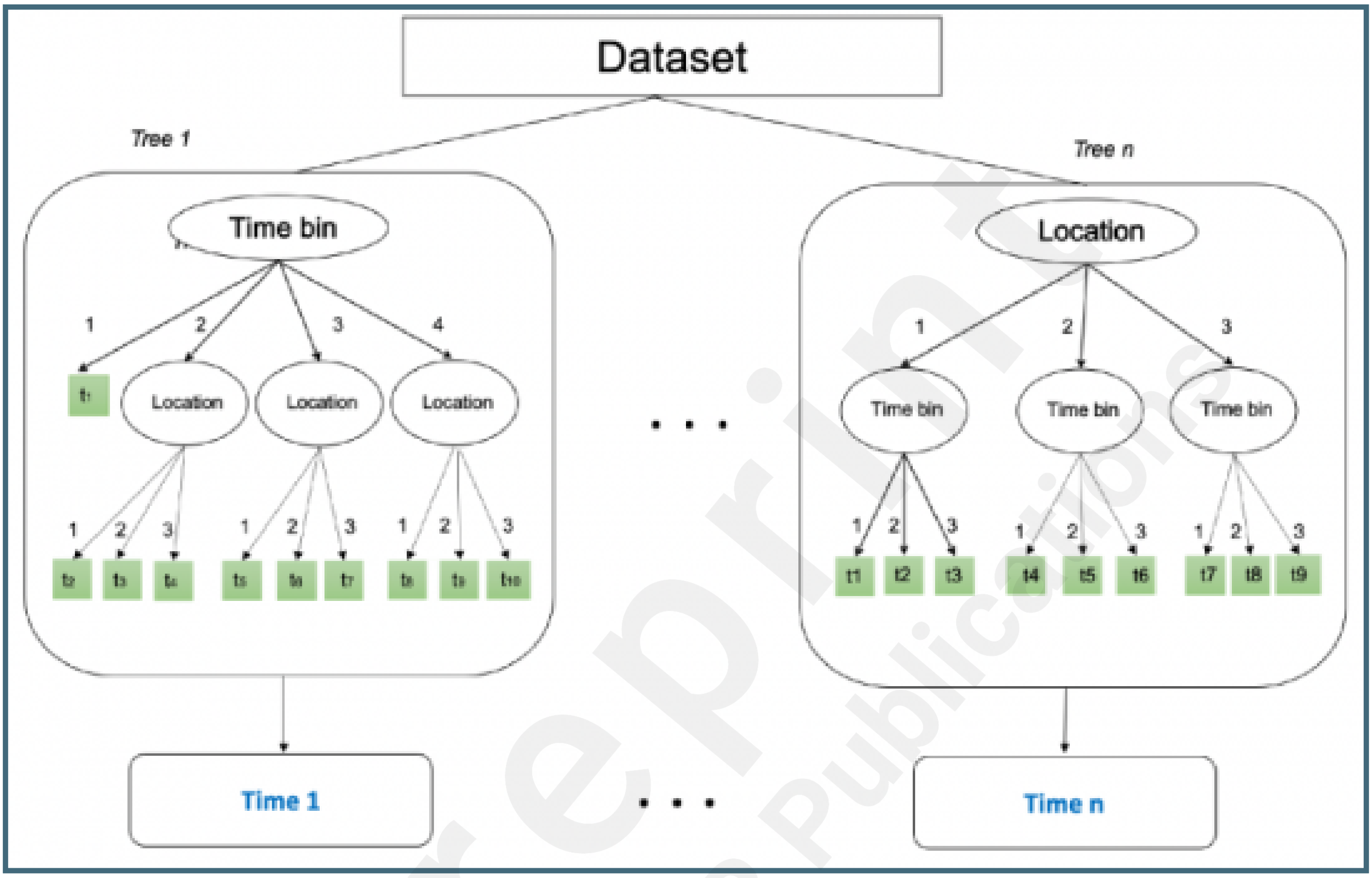
The frontend of the Snack Tracker app: 1, splash screen; 2, login screen; 3, home screen; 4, new snack screen; 5, time recording screen (time picker); 6, location recording screen; 7, recording save screen; 8, review recordings summary.



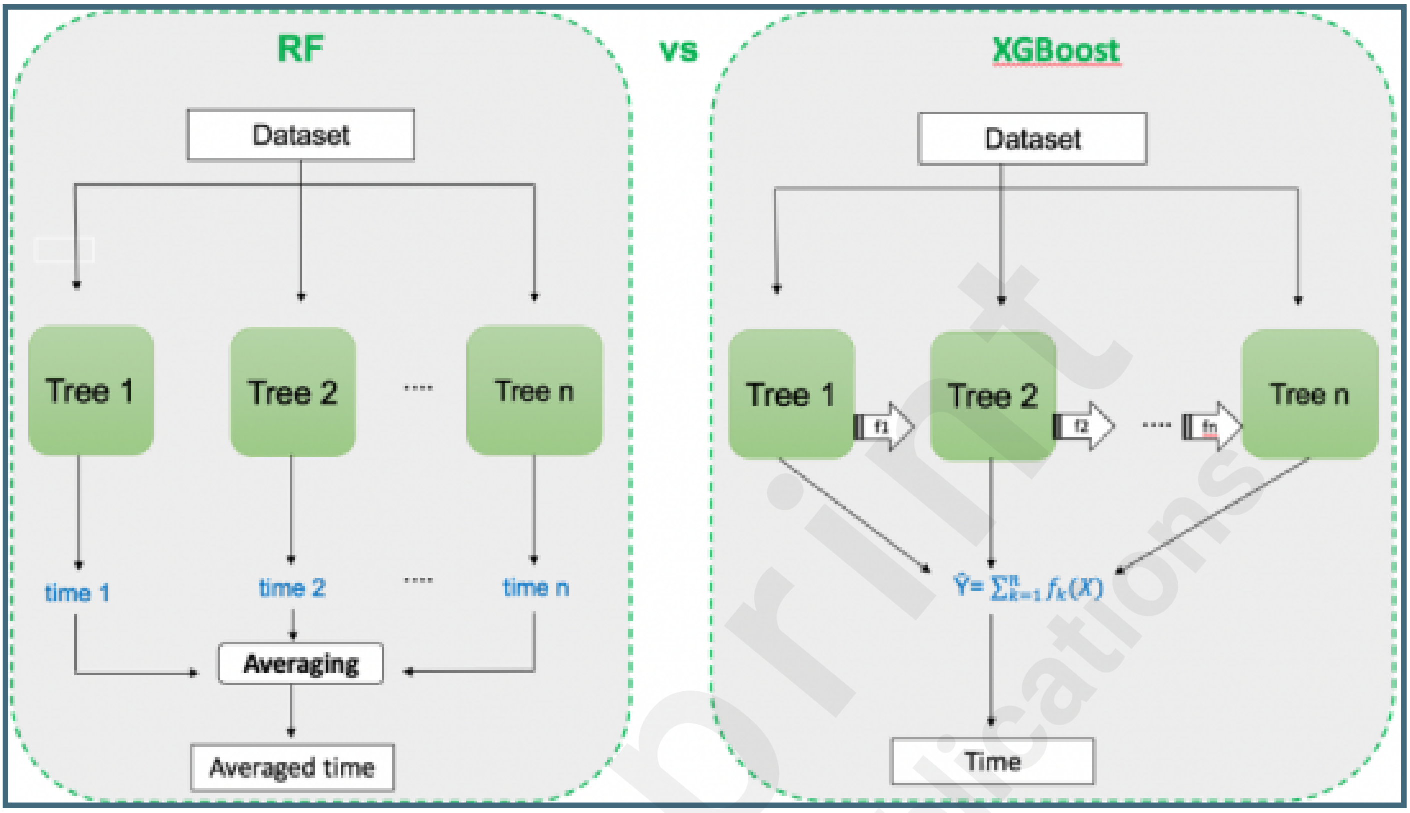
The various reminders employed in the app-based part of the study.



Both Random Forest (in RFreg) and XGBoost (in XGBreg) are based on the same building block, a decision tree, which is a set of rules built from the dataset. In our case, the aim is to predict the time until the next HFSS snack in minutes. n is the number of trees in the model, selected based on cross-validation performance. Time k is the expected time until the final unhealthy snack predicted using Tree k.



For RFreg (on the left), we create several different trees and then average predictions. For XGRreg (on the right), we have a sequence of trees, which progressively refine the prediction.



Multimedia Appendixes

Additional information concerning methods in the present work.
URL: http://asset.jmir.pub/assets/cb1d5c1c2e1718531beced94b7840fa5.doc

Model residuals concerning the objective of predicting the time until the next unhealthy snack.
URL: http://asset.jmir.pub/assets/936ff5ca8ba12899973da87a7b95da23.doc

An alternative predictive objective.
URL: http://asset.jmir.pub/assets/b4a598d8a18f8bb4f1ffaf0a4f273edf.doc

Additional references.
URL: http://asset.jmir.pub/assets/5adf4f58a5a5e22358c76fa87587e151.doc