



This is a repository copy of *The effect of simulated contextual factors on recipe rating and nutritional intake behaviour*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/207238/>

Version: Accepted Version

Proceedings Paper:

Zhao, M., Harvey, M. orcid.org/0000-0001-5504-2089, Cameron, D. et al. (1 more author) (2024) The effect of simulated contextual factors on recipe rating and nutritional intake behaviour. In: Clough, P., Harvey, M. and Hopfgartner, F., (eds.) CHIIR '24: Proceedings of the 2024 Conference on Human Information Interaction and Retrieval. ACM SIGIR Conference on Human Information Interaction And Retrieval, 10-14 Mar 2024, Sheffield, United Kingdom. Association for Computing Machinery , pp. 97-107. ISBN 9798400704345

<https://doi.org/10.1145/3627508.3638328>

© 2023 The Authors. Except as otherwise noted, this author-accepted version of a conference proceeding published in CHIIR '24: Proceedings of the 2024 Conference on Human Information Interaction and Retrieval is made available via the University of Sheffield Research Publications and Copyright Policy under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

The Effect of Simulated Contextual Factors on Recipe Rating and Nutritional Intake Behaviour

Mengyisong Zhao

mzhao18@sheffield.ac.uk

Information School, the University of Sheffield
Sheffield, UK

David Cameron

d.cameron@sheffield.ac.uk

Information School, the University of Sheffield
Sheffield, UK

Morgan Harvey

m.harvey@sheffield.ac.uk

Information School, the University of Sheffield
Sheffield, UK

Frank Hopfgartner

hopfgartner@uni-koblenz.de

Universität Koblenz

Koblenz, Germany

ABSTRACT

Despite the importance of context in Recommender Systems (RSs) more generally, and its clear applicability in the food domain, most existing research focuses on single contextual factors, and only considers simple extrinsic factors such as location and time. No RSs research has systematically explored the impact of multiple dynamic factors, or investigated the effect of emotion in determining people's eating, recipe rating and nutritional intake behaviour. To bridge these gaps, we conducted a comprehensive large-scale (n=397) crowdsourced experimental study to uncover the intricate relationship between various simulated contextual factors and users' subsequent recipe rating and implied nutritional intake behaviour. We further aimed to explore how these contextual factors can be incorporated to improve recommendation performance. Four distinct types of contextual factors were investigated: seasonal, emotional, busyness and physical activity, encompassing a total of seven elements. Our findings show that people's eating preferences and the likelihood of them choosing to eat healthy recipes vary depending on the simulated context they find themselves in. Moreover, we demonstrate how these contextual features can be used to significantly improve recipe rating prediction performance. Our research has implications for the future development of food RSs, and shows that emotion-aware systems could lead to better healthy food recommendations.

CCS CONCEPTS

• **Information systems** → *Recommender systems*.

Authors' addresses: Mengyisong Zhao, mzhao18@sheffield.ac.uk, Information School, the University of Sheffield, Sheffield, UK; Morgan Harvey, m.harvey@sheffield.ac.uk, Information School, the University of Sheffield, Sheffield, UK; David Cameron, d.cameron@sheffield.ac.uk, Information School, the University of Sheffield, Sheffield, UK; Frank Hopfgartner, hopfgartner@uni-koblenz.de, Universität Koblenz, Koblenz, Germany.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](https://www.acm.org/permissions).

© 2018 Association for Computing Machinery.

0730-0301/2018/8-ART111 \$15.00

<https://doi.org/XXXXXXX.XXXXXXX>

KEYWORDS

Contextual features, Machine learning, Context-aware food recommender systems, Healthy recommendations, User study

ACM Reference Format:

Mengyisong Zhao, Morgan Harvey, David Cameron, and Frank Hopfgartner. 2018. The Effect of Simulated Contextual Factors on Recipe Rating and Nutritional Intake Behaviour. *ACM Trans. Graph.* 37, 4, Article 111 (August 2018), 13 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 INTRODUCTION

The context in which we find ourselves can have a strong effect on our individual choices. Therefore, situations where choices need to be made, such as those targeted by Recommender Systems (RSs), could benefit profoundly from the intelligent inclusion of contextual factors in the recommendations. Work in many RSs domains, including: location-aware restaurant recommendation [14], time-aware tourism recommendation [63, 65], context-aware TV content recommendation [64], or mood-aware music recommendation [45], has demonstrated this benefit.

Food, and particularly healthy food, recommender systems are an important and growing area of RSs research [22], and one in which context may also play a vital role. Existing research from psychology and food science has already demonstrated the profound impact of various contextual factors in the food people choose to eat [20, 48]. Predicting what food a user may want to eat, and what food one would be better to eat, is a complex and multifaceted process, which could be influenced by various biological, personal, and social factors [13]. Food recommender systems can play a significant role in promoting healthy lifestyles by suggesting suitable food products that match users' dietary preferences and meet their basic biological and physiological requirements. However, making recommendations based on nutritional content may increase the risk of rejection, since most popular recipes are positively associated with higher fat and calories [52, 58].

Individuals may not consistently maintain an unhealthy diet under all circumstances, and making blanket health recommendations could potentially overwhelm or frustrate users. Indeed, specific situations may trigger or encourage unhealthy eating habits. Therefore, it is crucial to examine whether users exhibit varying dietary and nutritional intake patterns under different contextual situations. Additionally, it is important to investigate when it is most necessary to provide healthy recommendations and how to integrate

nutritional information to reduce the likelihood of recommendation rejection.

Prior food science research highlights that static factors like cultural background [36], as well as personal dynamic factors including stress levels, mood or emotion may markedly affect food consumption behaviour [33]. However, there has been very little research directly investigating these factors in the food recommendation field. Despite the importance of context in RSs more generally, and its clear applicability in the domain of food, most existing research focuses on single contextual factors, and only considers simple extrinsic factors such as time or location. No research has systematically explored how multiple dynamic factors affect people's eating and recipe rating behaviour.

We bridge these gaps by exploring how various dimensions of dynamic factors, including seasonality, emotional states, busyness, and physical activity, affect people's recipe rating and implied nutritional consumption behaviours. We combine insights gained from analyses of recipes ratings from a large online food portal, as well as reviewing possible influential factors discussed by psychology studies. We conducted a large-scale (n=397) crowdsourced experimental study to learn the relationship between various *simulated contextual factors* and users' subsequent recipe rating and implied nutritional intake behaviour.

Note that a common issue in food recommender systems work is the difficulty in directly measuring users' cooking and eating behavior, as preparing food requires a significant amount of effort [35]. Biologically speaking, the urge to eat and cook a certain food is usually driven by a liking or preference for a particular food or recipe [43]. As such, a high rating can be considered a proxy for the future intention to consume a certain recipe. The behaviour measured in this research is implied behaviour, and the discussion and implications of this research are based on this assumption.

We examine seven contextual scenarios from four types of contextual factor: "seasonal", "emotional", "busyness" and "physical activity", and compare these with a "context-free" baseline condition. Furthermore, we explore how these contextual factors can be employed to improve recommendation performance. More precisely, we address the following research questions:

- RQ1: Do people's recipe rating behaviour vary among different simulated contextual situations?
- RQ2: To what extent do contextual factors affect people's implied nutritional intake behaviour?
- RQ3: Can integration of these contextual factors improve recommendation performance?
- RQ4: Which contextual factors are the most influential factors when recommending foods?

The novel contributions of our research can be summarised as follows:

- Introduction and provision of a novel large-scale dataset to study context-aware healthy food (recipe) recommender systems.
- In addition to predicting user preferences based on recipe ratings from online recipe websites, our validation process involves gathering evaluations and insights from users in simulated contextual scenarios.

- We are the first to reveal the impact of various dynamic contextual factors on individuals' recipe rating behaviours. Our study also highlights variations in implied nutritional intake behaviours across different contextual situations, offering valuable insights for incorporating health information into healthy food recommendation systems.
- By identifying the most influential factors beneficial to machine-learning-based recommendation algorithms, our research contributes to the development of the next generation of context-aware food recommender systems.

2 LITERATURE REVIEW

Early work in psychology and food science has demonstrated that there are a wide variety of influential contextual factors that shape food choice [47]. For example, the socialisation and acculturation standards that people have learned, normative family eating patterns, and historical cultural beliefs may lead to different food choices [36]; and changes in mood and general emotional state may regulate eating and vice versa [33, 38]. Tangible physical capital such as money, equipment, and intangible capital such as time, cooking skills and knowledge could all determine varied food choice [46]; as well as change of social and broader environments, such as changes of roles, surrounding by different groups or communities [4, 49], seasonal, home or workplace could provide opportunities and obligations for reconstructing eating relationships and food choices [10, 11].

However, few contextual factors have been explored in prior research in the food recommendation domain. Due to a lack of available contextual features in publicly-available recipe rating data sets, researchers often focus on studying and attempting to incorporate only single contextual factors into food RSs. Kuzmierczyk et al. [29] found there are clear temporal patterns in online recipe production and uploading behaviour, but the relationship between recipes generation and user's eating behaviour remains unclear. Cavazza et al. [6] found that females are more interested in smaller and more "elegant" meals than males, while Rokicki et al. [39] sought to integrate gender into a collaborative filtering model, demonstrating a small improvement in performance. However, the researchers only built a single gender-aware RS without considering other factors. Location, which is known to be one of the most significant contextual factors in other domains, has also been widely researched in the food domain [7, 9, 34], but most research attempts to demonstrate the impact of location on eating behaviour rather than focusing on the development of location-aware food recommendation systems.

Trattner et al. [57] investigated the relationship between cooking interests, hobbies and nutritional values of online recipes. They suggest that learning the patterns between a user's hobbies and eating preferences could provide motivating goals for persuasive systems. Recently, Gao et al. [18] proposed a context-aware recommendation system based on graph neural networks, the model has been tested on two popular food datasets. Due to the challenge of revealing the meaning of hidden neural structures in a deep learning network, the paper does not specify which exact contextual factors were contributed in the model. To date, due to the difficulty of duplicating contextual environments in a laboratory setting, various dynamic

personal factors, such as emotional state and fluctuations in stress levels, have been little explored.

Another significant challenge exists when considering the balance between nutrition and people's food preferences, particularly when both aspects are integrated into a food recommendation system [12]. Simply considering the accuracy of recommendation results isn't sufficient to build an ideal food RS, as most popular online recipes tend to be less healthy [57]. Most research considers healthy recommendations only but neglects the user's past preferences, or vice versa [19, 59]. Ueta et al. [59] developed a recipe retrieval system based on 45 common nutrients, which allows users to easily search for nutritious recipes using natural language to address specific health conditions without taking into account the user's past recipe preference. Harvey et al. [21] designed a study, which asked people what factors they thought were important, and found that 17 factors (mainly food-based factors, health and temporal factors) impacted users' rating choice. However, they did not investigate how these factors interacted with each other and how specific scenarios actually impact food choice and nutritional intake behaviour. More recently, Wang et al. [61] introduced a multimodal health-aware food recommendation system that incorporates word2vec and VGGNet-19 to extract textual and visual features, and demonstrated superior performance compared to traditional recommendation systems. Yet, the recommendations are generated based on the user's health profile and associated tags, rather than considering the user's past search history or preferred recipe choices.

In summary, previous research rarely considers health recommendations along with the user's past preferences simultaneously. A further notable aspect is that no research has focused on incorporating contextual factors into food recommendation systems while making healthy recommendations. It is apparent from previous literature that more in-depth research is required in order to better understand what contextual factors are most impactful in determining people's food choices as well as implied nutritional intake behaviours. This is particularly so for contextual factors related to seasonality, emotional status, stress levels and physical activities, which have been suggested in the food science domain to potentially have a significant impact on people's eating behavior.

3 METHOD

3.1 Experimental design

The primary aim of the study is to examine how seven dynamic contextual factors, including a hot summer's day, a cold winter's day, happy emotion, sad emotion, a busy and stressful weekday, a relaxing weekend, and after physical activities, along with a context-free generic group, affect people's recipe rating, and implied nutritional consumption behaviour. To achieve this, we conducted a large-scale ($n=397$), between-subjects user study (see Table 1). Participants were randomly assigned to one of eight groups (the seven simulated contexts plus a baseline "context-free" condition), shown material to simulate the given context - detailed below - and then asked to rate recipes under that contextual scenario. Each participant rated 30 recipes from a pool of 75 recipes using a Likert scale from 1 (strongly dislike) to 5 (strongly like). All participants

were recruited online using the Prolific crowdsourcing platform; the survey instrument was designed using Qualtrics.

The pool of 75 recipes were sampled from Allrecipes.com, one of the most popular online recipe websites, which has previously been used as a data source by various researchers in the food recommendation domain (e.g., [22, 40, 66]). The recipes were chosen to ensure a distribution across various factors, including: seasonality (i.e., winter and summer), healthiness, recipe category (i.e., main dish, soup, salad, and dessert/snack); and to ensure that both vegetarian and vegan recipes were included.

Recipe seasonality was determined based on the popularity of each recipe (i.e., total number of 4 or 5 ratings) during each season. For example, the summation of 4 or 5 ratings given to a recipe during the months of December, January and February are that recipe's winter popularity. Recipes were flagged as being vegetarian and/or vegan based on a manual analysis of their constituent ingredients.

Healthiness was determined using World Health Organisation (WHO) [62], U.S. Food & Drug Administration (FDA) [16] and United Kingdom Food Standards Agency (FSA) standards [17]. These are international standards that have been previously used in this domain [56]. The calculations were based on the detailed rules provided by each standard, which are outlined in the appendix (see Table 9, 10, 11). For the WHO and FDA health scores, we initially converted the nutrient values in grams provided by Allrecipes.com into percentages of the daily recommended values. Then, we evaluated whether the recipe's nutrient values fell within the standard range. If a particular nutrient value fell within the range, we added 1 to that recipe's score; otherwise, we added nothing. Consequently, the higher the score, the healthier the recipe.

The calculation of the FSA health score is more intricate. We began by computing the total weight of each recipe based on their ingredient lists, then divided this by the number of servings to obtain the weight per portion. Subsequently, we calculated the values for fats, saturated fats, sugar, and salt per 100g. Next, we categorise these nutrient values as green, amber, or red based on their corresponding range. If a nutrient value falls within the green range, it is assigned one point. If it falls within the amber range, it receives two points, and if it falls within the red range, it is assigned three points. Here, a higher overall score for a recipe indicates a lower level of healthiness.

The primary objective is to assess the health level of individual recipes; however, established standards like WHO or FDA define ideal daily nutritional intake. As observed in Western societies, the prevalent dietary pattern often consists of three meals a day [31, 54]. To account for this we divided the standard recommended daily nutritional intake by three. This approach provides a more accurate indicator of the nutritional quality for an individual recipe.

3.2 Stimuli and manipulation material

In the study, seven contextual situations were simulated, under which participants could rate recipes. To minimise the influence of real-life environmental factors and enhance the evocation of these simulated scenarios, inspiration was drawn from studies conducted by [26] and [60].

To accomplish this, we employed the external emotion stimuli method, as described in previous studies [23, 44]. This method

Table 1: Block design setup demonstration

Block	Recipe rating				
1. Hot summer's day	No.1	No.2	No.3	...	No.75
2. Cold winter's day	No.12	No.35	No.37	...	No.2
3. Happy emotion	No.8	No.26	No.31	...	No.9
4. Sad emotion	No.46	No.39	No.53	...	No.67
5. Busy and stressful weekday	No.61	No.43	No.18	...	No.21
6. Relaxing weekend	No. 73	No. 62	No.16	...	No.22
7. After physical activities	No.4	No.11	No.5	...	No.9
8. Original	No.10	No. 30	No.16	...	No.17



Figure 1: Demonstration of recipe template under each contextual scenario (taking “Apple Pie By Grandma Ople” as an example recipe). Note that the example on the bottom-right is the baseline, context-free condition.

primarily relies on source materials such as photos, music, or videos to evoke emotions. In a controlled laboratory setting, videos offer a more effective and immersive means of eliciting specific emotions compared to images or music alone. Furthermore, videos provide a rich blend of visual and auditory stimuli. This abundance of sensory input makes videos the preferred choice for eliciting emotions with higher intensity, efficiency, and accuracy [23, 24, 50].

In this study, participants were presented with a 22-second video clip designed to engage their cognitive faculties and enhance their

immersion in the assigned simulated contextual situation [55]. For instance, the video for a hot summer day featured a beautiful sunny day, a thermometer displaying high temperatures, and a person sweating. In contrast, the video for a cold winter day depicted heavy snowfall and strong winds. Participants in the “context-free” control group were not shown a video. To further immerse participants in their designated contextual scenario, recipe templates (see Figure 1) were designed to align with each scenario. The number of special elements, such as the sun for a hot summer day or weightlifting for after physical activities, was consistent across templates, and the sizes of these elements were nearly identical when inserted into the background. For reference, specific recipe examples can be found in Figure 1.

3.3 Procedure

The survey workflow is illustrated in Figure 2. Participants expressed their interest in taking part via the Prolific platform. Prior to commencing the survey, participants were asked to review the information sheet and complete a consent form. Subsequently, participant demographic information was collected, encompassing basic details such as age, gender, ethnic origin, etc.

After completing the demographic questions block, participants encountered the recipe rating block. In this study, participants were randomly assigned to a group (as described above) before providing their ratings. As discussed above, participants in the contextual scenario groups were instructed to watch a 22-second video clip designed to immerse them in that context. Each participant was tasked with rating 30 recipes within their assigned contextual situation block. Finally, participants were asked two questions regarding their reasons for assigning high or low ratings to each recipe.

To maintain a high level of data quality, we implemented both manipulation and attention checks to verify that participants were following the prescribed procedures and paying adequate attention. The manipulation check involved a question where participants had to identify the theme of the video they saw at the start of the questionnaire (e.g., “hot summer day”). There were two types of attention check: one nonsensical recipe (including ingredients such as: 5 stones, 3 cups of sand), which should always elicit a negative response; and attention-check recipes, where participants were explicitly instructed to choose a specific response. The attention check questions were presented after participants had rated all 30 recipes. While the text in the attention check recipes remained identical, the background varied to maintain consistency with recipes in different contextual scenarios.

3.4 Data collection and participants

G*Power was employed to provide an estimate for an appropriate sample size [15]. Based on the data from previous experimental studies regarding personalisation and user behaviour [28, 32], we estimated the effect size to be 0.23. The estimate indicated that 360 participants were needed in this research to achieve a significance criterion of 5% and 90% power. Post hoc analysis of effect size based on the ANOVA of mean and variance of each group, resulted in an effect size larger than that assumed a priori (Cohen’s $f=0.277$) [8].

We recruited participants between May 2023 and July 2023 exclusively through Prolific. Each participant received a payment

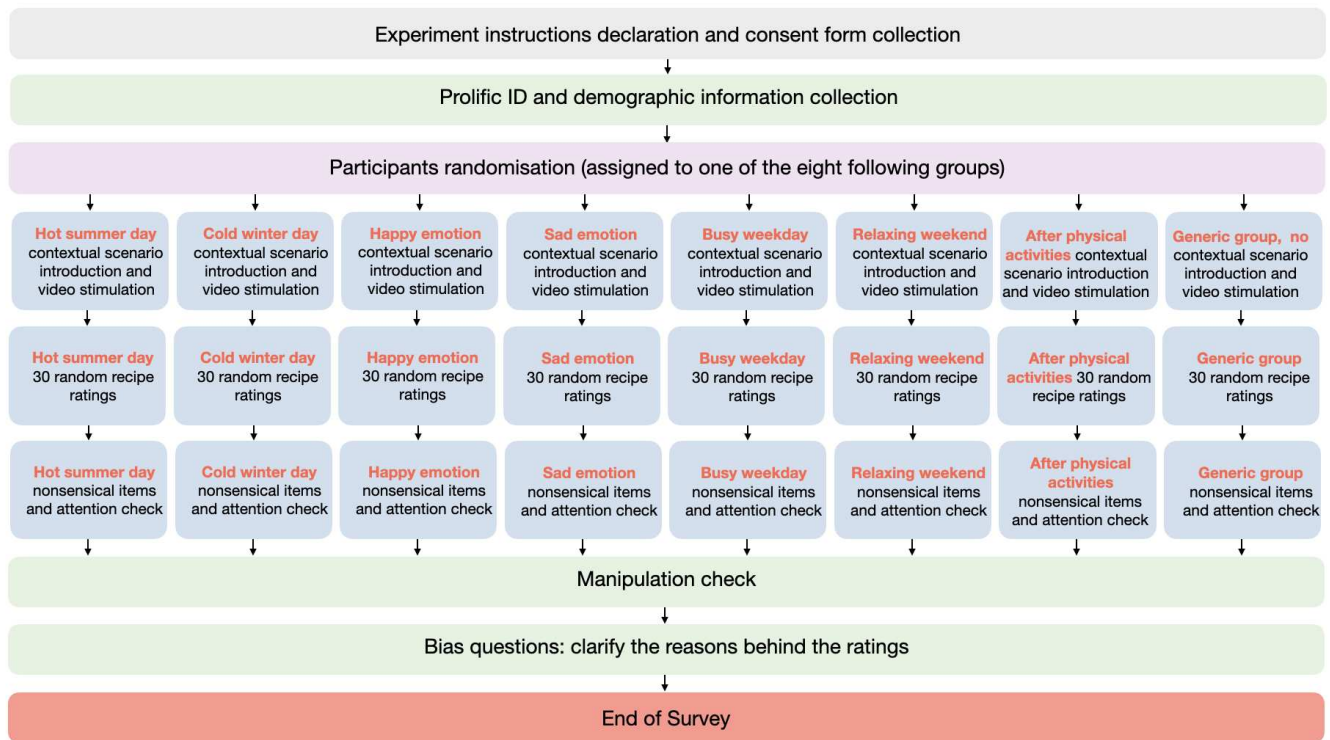


Figure 2: Experiment workflow chart

of £1.82, in accordance with the current living wage standards. A total of 428 participants were recruited for this study. Participants who failed the manipulation check (19, 4.4%) and the nonsensical item attention check (12, 2.8%) were excluded from the study. After thorough data inspection and cleaning, we retained data from 397 valid individuals.

Among these participants, 212 (53.4%) identified as male and 177 (44.6%) as female, 5 (1.3%) were non-binary or gender diverse, and the remaining 3 (0.7%) preferred not to disclose their gender. The largest age group was 25-34, representing 30.7% of participants, followed by the 35-44 age group at 25.8%, the 45-54 age group at 19.4%, the 18-24 and 55-64 age group with a similar proportion at 10.8% and 10.3%, respectively. Participants aged 65-74 accounted for 3.3%, with only one participant aged 75 or above. The majority of participants hailed from the UK and the US. Regarding ethnic origin, White participants dominated with 329 (82.9%), followed by Asian or Asian British at 34 (8.6%). The remainder consisted of participants identifying as Black, Black British, Caribbean, or African (16, 4.0%), Mixed or Multiple Ethnic Groups (10, 2.5%), those who preferred not to disclose (5, 1.3%), and others (3, 0.7%).

While participants' home countries were relatively diverse, the majority were currently residing in the UK (351, 88.4%) and the US (44, 11.1%). For a more detailed summary of key demographic features, please refer to Appendix table 12.

3.5 User rating prediction model and Feature engineering

The raw CSV data file was downloaded from Qualtrics. Data cleaning and preprocessing were carried out using Python Jupiter notebook 6.3.0. According to recent systematic reviews on recommender systems conducted by [27] and [42], tree-based models have been widely employed in model-based recommender systems. Therefore, we aimed to evaluate the performance of tree models on our dataset; XGBoost, a boosting tree model, was chosen for its efficiency and flexibility. Its objective function evaluates model performance based on a set of parameters, while a regularisation term controls model complexity to prevent overfitting [30]. Notably, tree models offer more interpretable explanations than other models, which aligns with our goal of identifying the most influential features in the prediction task. In this study, the task of rating prediction was approached as a regression problem and not as a top-k relevance prediction problem. Consequently, evaluation metrics such as precision@K and nDCG would not be appropriate for evaluation. Instead, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R²) will be reported to evaluate the model test performance.

After data cleaning and preprocessing, a total of 11,910 ratings were collected, distributed as follows: hot summer day (HSD) 1500 ratings, cold winter day (CWD) 1500 ratings, happy emotion (H) 1470 ratings, sad emotion (S) 1470 ratings, busy weekday (B) 1470 ratings, relaxing weekend (R) 1500 ratings, after physical activities

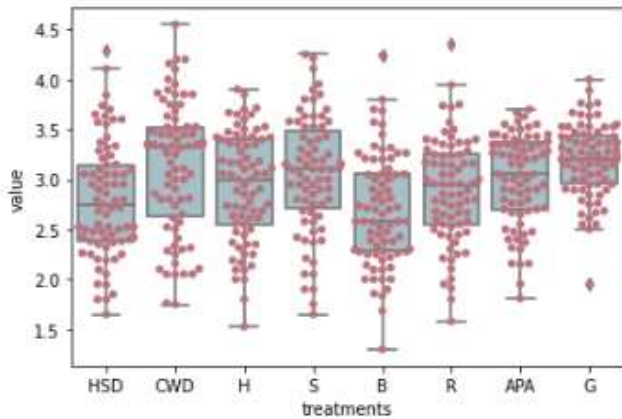


Figure 3: The data distribution between each contextual scenario group (The above abbreviation stands for information below. HSD: hot summer day (contextual group), CWD: cold winter day (contextual group), H: happy (contextual group), S: sad (contextual group), B: busy (contextual group), R: relax (contextual group), APA: after physical activities (contextual group), G: generic group)

(APA) 1500 ratings, and generic context-free group (G) 1500 ratings. The datasets for each experimental group were divided into training and testing sets using an 80:20 ratio. The target variable is the recipe rating, ranging from 1 to 5, maintaining consistency across each model.

A total of 29 features, which include participants' demographic features and recipe nutritional features, were used to train the XG-Boost model. While the main focus of the study was on identifying novel dynamic contextual factors, the reason for including demographic features is that they have the potential to enhance model performance. As these features can provide the trained model with additional reference information.

Two sets of experiment were conducted. The primary objective of the first experiment was to assess the importance of adding contextual features to the model. In this experiment, all 11,910 ratings were utilised. By systematically adding and removing contextual features within the model, the importance of each contextual feature can be determined. The second experiment was mainly to identify the most influential individual dynamic contextual factors at the model building level. To achieve this, the dataset was divided into eight groups based on contextual scenario groups to facilitate model performance comparisons.

4 RESULTS

4.1 RQ1-Do people's recipe rating behaviour vary among different simulated contextual situations?

To address RQ1, we used an ANOVA test with the null hypothesis (H_0) that there are no significant differences (no variation in means) across the contextual scenarios.

Table 2: Shapiro-Wilk test normal distribution result

SW value	HSD	CWD	H	S	B	R	APA	G
statistic	0.983	0.972	0.973	0.986	0.989	0.994	0.945	0.981
p-value	0.409	0.095	0.115	0.591	0.778	0.984	0.003	0.356

Note: The abbreviations are the same as in Figure 3.

Figure 3 shows the rating distributions across the simulated contextual scenarios. Particularly noteworthy is the rating distribution in the generic baseline group (G), which displays low variance, with most ratings falling within the range [2.75, 3.25]. In contrast, the ratings for the 'cold winter day' (CWD) and 'hot summer day' scenarios exhibit a considerably more dispersed range. The rating distribution for the busy weekday contextual scenario (B) is notably lower than those of the other scenarios. Surprisingly, the score distributions of the groups experiencing happy and sad emotions appear similar; however, the sad emotion group demonstrated a marginally higher mean rating range when compared to the happy emotion group.

The normality of residuals was assessed using the Shapiro-Wilk test (see Table 2). The p-values for seven of the groups were found to be non-significant, indicating that the data was drawn from a normal distribution. However, note that the data for the 'after physical activities' group exhibited a significant departure from normality.

A one-way ANOVA test was conducted to explore whether variations existed in individuals' implied eating behaviours and recipe rating responses across the different contextual scenarios. The result revealed that there were significant differences between the groups ($F(7, 592)=7.564, p \ll 0.001$), allowing us to reject the null hypothesis. As the ANOVA test doesn't test the relationships between each group, we subsequently conducted multiple pairwise comparisons, employing Tukey's Honest Significant Difference (HSD) test. The outcomes of Tukey's test are shown in Table 3.

Participants exhibited notable distinctions in recipe preferences when compared with the generic group in certain contextual situations. These included hot summer day (HSD; $F(7, 592) = 6.451, p = 0.001, \mu_{G1} = 2.795, \mu_{G2} = 3.196$), busy weekdays (B; $F(7, 592) = 8.311, p = 0.001, \mu_{G1} = 2.679, \mu_{G2} = 3.196$), and relaxing weekends (R; $F(7, 592) = 4.889, p = 0.013, \mu_{G1} = 2.892, \mu_{G2} = 3.196$). In contrast, the remaining contextual scenarios did not exhibit statistically significant variations in mean ratings.

As expected, participants in the 'hot summer day' group and 'cold winter day' group shows significant difference in recipe preference ($F(7, 592)=5.304, p=0.004, \mu_{G1} = 2.679, \mu_{G2} = 3.124$). More interestingly, the results disclosed that, during busy weekdays, implied eating behaviour and recipe preference significantly diverged from that observed during 'cold winter day' ($F(7,592)=7.165, p=0.001, \mu_{G1} = 3.124, \mu_{G2} = 2.679$), as well as during states of 'sad emotion' ($F(7, 592)=5.958, p=0.001, \mu_{G1} = 3.049, \mu_{G2} = 2.679$) and 'after physical activities' ($F(7, 592)=5.277, p=0.005, \mu_{G1} = 2.679, \mu_{G2} = 3.007$).

Table 3: Tukey's Honest Significant Difference (HSD) test result demonstration on preference of 75 recipes

G1	G2	Diff.	μ_{G1}	μ_{G2}	Conf. Int.	q	p
HSD	CWD	0.329	2.795	3.124	[0.062, 0.597]	5.304	0.005
HSD	G	0.401	2.795	3.196	[0.134, 0.668]	6.451	0.001
CWD	B	0.445	3.124	2.679	[0.178, 0.712]	7.165	0.001
S	B	0.370	3.049	2.679	[0.103, 0.637]	5.958	0.001
B	APA	0.328	2.679	3.007	[0.061, 0.595]	5.277	0.005
B	G	0.517	2.679	3.196	[0.249, 0.784]	8.311	0.001
R	G	0.304	2.892	3.196	[0.036, 0.571]	4.889	0.014

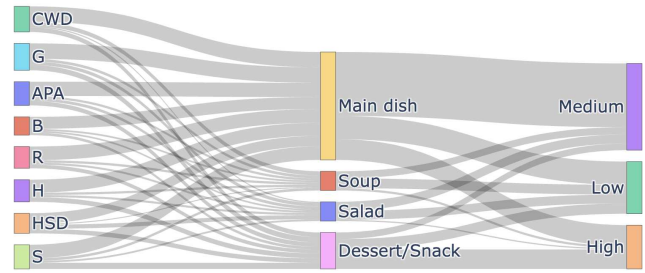
Note: G1 = Group1, G2 = Group2, q = q-value, p = p-value. The group abbreviations are the same as in Figure 3.

4.2 RQ2-To what extent do contextual factors affect people's implied nutritional intake behaviour?

The determination of recipe health levels primarily relies on the FSA standard [17] in this section as this is the most appropriate standard for evaluating a single recipe. The other two standards focus on measuring appropriate daily nutritional intake. In alignment with this standard, a higher FSA score signifies a less healthy recipe, with scores ranging from 4 (extremely healthy recipe) to 12 (extremely unhealthy recipe). After data aggregation, it was observed that recipes with the highest FSA score of 12 were remarkably popular. In fact, such recipes garnered the highest mean ratings among six groups: 'hot summer day' (mean rating of 3.21), 'happy emotion' (3.35), 'sad emotion' (3.53), 'busy weekday' (3.57), 'relaxing weekend' (3.75), and 'after physical activities' (3.7), as well as the 'generic' group (mean rating of 3.89). It is notable that, across most groups, recipes with better health ratings tend to have been given lower scores than their less healthy counterparts.

In the generic group the mean rating remained relatively consistent, with only minor fluctuations (ranging from 3.144 to 3.889 score), suggesting that, in the absence of contextual factors, individuals' recipe preferences don't result in significant changes in health outcomes. However, in the 'after physical activities' group, healthy recipes were generally preferred over less healthy ones. Conversely, 'during relaxing weekend' unhealthy recipes are more popular. Similarly, both 'happy' and 'sad emotion' groups show a preference towards unhealthy recipes, particularly in the case of the 'sad emotion' group.

We conducted a One-way ANOVA test once again to assess whether preferences for healthy recipes have significant differences between the contextual scenario group. The results indicate significant differences among the groups ($F(7, 64) = 4.942, p = 0.000$). We again used Tukey's HSD tests to obtain pairwise comparison (See Table 5), which revealed several significant differences. Preferences for healthy recipes during 'hot summer day' showed significant differences compared to 'cold winter day' ($F(7, 64) = 5.094, p = 0.013, \mu_{G1} = 2.722, \mu_{G2} = 3.315$) and during 'feeling sad' ($F(7, 64) = 4.552, p = 0.340, \mu_{G1} = 2.722, \mu_{G2} = 3.252$), as well as the 'generic' group ($F(7, 64) = 5.159, p = 0.012, \mu_{G1} = 2.722, \mu_{G2} = 3.323$). Preferences for healthy recipes during 'busy weekday' exhibited significant differences compared to 'cold winter day' ($F(7, 64) = 5.525, p = 0.005, \mu_{G1} = 3.315, \mu_{G2} = 2.672$) and while 'feeling sad' ($F(7, 64) = 4.983, p = 0.017, \mu_{G1} = 3.252, \mu_{G2} = 2.672$), as well as the 'generic' group

**Figure 4: Sankey Diagram about contextual impact on likelihood of recipe categories and grouped FSA health levels. Low (FSA levels 4,5 and 6), Medium (FSA levels 7, 8 and 9), and High (FSA levels 10, 11 and 12).****Table 4: Recipe mean rating among different FSA health levels for each contextual scenario group**

FSA	HSD μ_r	CWD μ_r	H μ_r	S μ_r	B μ_r	R μ_r	APA μ_r	G μ_r
4	2.250	3.450	2.684	3.100	2.579	2.682	3.316	3.318
5	3.048	2.333	3.273	2.950	2.714	2.800	3.450	3.158
6	2.450	3.100	2.905	2.750	2.700	2.789	3.048	3.200
7	2.605	3.333	3.051	3.075	2.711	3.056	3.301	3.244
8	2.714	3.350	2.850	3.105	2.550	2.842	2.900	3.200
9	2.657	4.025	3.303	4.178	2.284	3.159	3.489	3.419
10	2.949	3.167	3.316	3.364	2.597	3.071	2.881	3.144
11	2.620	3.084	3.167	3.225	2.345	2.629	2.900	3.338
12	3.211	4.000	3.350	3.526	3.571	3.750	3.700	3.889

Note: The abbreviations are the same as in Figure 3.

($F(7, 64) = 5.589, p = 0.004, \mu_{G1} = 2.672, \mu_{G2} = 3.323$). Furthermore, 'after physical activities' displayed significant differences in preferences for healthy recipes when compared to 'busy weekday' ($F(7, 64) = 4.707, p = 0.03, \mu_{G1} = 2.672, \mu_{G2} = 3.220$).

Additionally, we investigated the properties of favoured recipes. We filtered out the most popular recipes, considering only those with ratings of 4 and 5 under each contextual situation. These recipes were then aggregated based on recipe categories (e.g., Main dish, Soup, Salad and Dessert/Snack) and three grouped levels of FSA scores: Low (FSA levels 4,5 and 6), Medium (FSA levels 7, 8 and 9), and High (FSA levels 10, 11 and 12), after Starke et al. [53]. See Figure 4. We found that during 'cold winter days' and 'after physical activities', people tend to prefer main dishes, most of which fall into the (Medium) health category. Desserts and snacks are favoured when people are 'feeling sad' and 'feeling happy', with many of these items belonging to the unhealthy (High) food category. Predictably, salads are preferred during 'hot summer days' as they are relatively healthy compared to other categories. Soups are more popular during 'cold winter days' and when people are feeling sorrow, suggesting that soups might be an effective comfort food. Most soups fall into the healthy (Low) and general (Medium) health categories. In general, emotional changes may lead to increased consumption of unhealthy recipes.

Table 5: Tukey’s Honest Significant Difference (HSD) test result demonstration on preference of healthy recipes (only statistically significant results are reported)

G1	G2	Diff.	μ_{G1}	μ_{G2}	Conf. Int.	q	p
HSD	CWD	0.593	2.722	3.315	[0.077, 1.109]	5.094	0.013
HSD	S	0.530	2.722	3.252	[0.014, 1.046]	4.552	0.040
HSD	G	0.601	2.722	3.323	[0.085, 1.117]	5.159	0.012
CWD	B	0.643	3.315	2.672	[0.127, 1.159]	5.525	0.005
S	B	0.580	3.252	2.672	[0.064, 1.096]	4.983	0.017
B	APA	0.548	2.672	3.220	[0.032, 1.064]	4.707	0.030
B	G	0.651	2.672	3.323	[0.135, 1.167]	5.589	0.005

Note: The abbreviations are the same as in Table 3 and Figure 3.

4.3 RQ3-Can integration of these contextual factors improve recommendation performance?

In this study, as the task of rating prediction was approached as a regression problem, employing the XGBoost regression model. Model evaluation involved the calculation of several key metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R2). The primary comparison entailed assessing models trained with the inclusion of all available features (contextual scenario feature included) against a baseline model that excluded these features. This comparative analysis aimed to identify the impact of the contextual factors on model performance. Additionally, the top 5 most importance features were identified to discern the most influential variables for rating prediction.

The XGBoost model employed a comprehensive set of 29 features encompassing demographic attributes (e.g., age, gender, ethnic origin, physical activity level, survey ID), recipe-related details (e.g., recipe name, category, nutritional information), information regarding cooking (e.g., cooking frequency, skill level), recipe health level (WHO, FDA, FSA and WHO/3, FDA/3), and the contextual scenario information. Since each health standard emphasises a slightly different nutritional aspect, which may potentially impact model performance, all of them were included in the model-building process. As demonstrated in Table 6, the model utilising the complete feature set exhibited superior performance compared to the model without contextual scenario features. Contextual scenario features include the eight simulated contextual situations. The all-feature model yielded an MSE of 1.642, RMSE of 1.281, and MAE of 1.052, all of which outperformed the baseline model (MSE 1.688, RMSE 1.299, and MAE 1.071). The all-feature model achieved a higher R2 score (0.198) compared to the baseline (0.175), signifying a superior goodness of fit.

More importantly, the feature importance analysis revealed that the contextual scenario features were the most significant among the 29 features considered, as evidenced by Figure 5. This underscores the impact of *contextual information* on both human implied behaviour and algorithmic comprehension thereof. The incorporation of contextual factors in the model enhances its capacity to intelligently discern and uncover the hidden relationships within individuals’ preferences in the dataset. This, in turn, leads to more

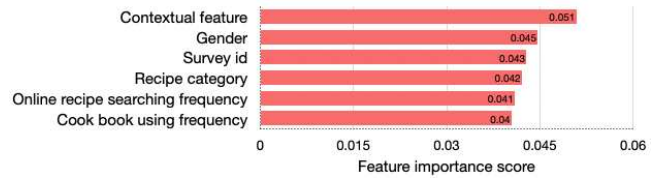


Figure 5: XGBoost feature importance of all feature model

Table 6: XGBoost test results summarisation and comparison of all feature model

	MSE	RMSE	MAE	R2
All feature model	1.642	1.281	1.052	0.198
Baseline model	1.688	1.299	1.071	0.175

precise rating predictions. Surprisingly, *gender* emerges as the second most important feature, suggesting potentially substantial differences in rating and dietary preferences between men and women. The importance of the *Survey ID* feature lies in its role in identifying and distinguishing an individual’s implied eating and rating behaviour. This feature is commonly utilised in traditional matrix factorisation recommender systems, and so it comes as no surprise that it ranks as the third most important feature. Notably, the influence of recipe category surpasses that of recipe name within the XGBoost model. This could be attributed to the broader category information reducing model complexity, thereby facilitating more precise weight generation for each leaf node. *Online recipe searching* and *cook book use frequency* have the potential to uncover hidden patterns related to an individual’s cooking interests and skills.

4.4 RQ4-Which contextual factors are the most influential factors when recommending foods?

We further applied the XGBoost model to predict ratings segmented by contextual scenario to investigate which group led to better model performance. The results show that using data from the ‘happy emotion’ group achieved the highest model performance (MSE=1.628, RMSE=1.276, MAE=1.053, and R2=0.178), as shown in Table 7. Data from ‘hot summer day’, ‘happy emotion’, ‘sad emotion’, and ‘busy weekdays’ generally performed better than the other groups. When compared with the generic baseline group and considered from a model fitting perspective, the datasets for all simulated contextual scenarios achieved higher R2 values. We note the relatively lower performance of the models using the ‘cold winter day’, ‘relaxing weekend’, ‘after physical activities’ compared to other contexts. This may be due to the fact that, even though we specified a single contextual scenario to the participants, this may still be influenced by other independent factors in that scenario. For example, during a relaxing weekend, one may also feel happy, and after physical activities, one may feel both tired and energetic simultaneously.

The demographic factors, such as the time spent living in one’s current country of residence, gender, home country, and cooking skill level, are frequently shown to be the most important factors in

Table 7: XGBoost model test results summarisation and comparison among each contextual scenario group

	HSD	CWD	H	S	B	R	APA	G
MSE	1.839	2.120	1.628	1.798	1.847	1.944	2.112	1.774
RMSE	1.356	1.456	1.276	1.341	1.359	1.394	1.453	1.332
MAE	1.093	1.173	1.053	1.045	1.102	1.140	1.173	1.068
R2	0.125	-0.039	0.178	0.040	0.015	-0.030	-0.110	-0.113

Note: The abbreviations are the same as in Figure 3. Under certain contexts, the fitted model performed worse than the null model and achieved a negative R2. In such situations, the average rating would be used for rating prediction.

each contextual scenario group, as indicated in Table 8. Remarkably, the FDA health level is the most important factor in the ‘sad emotion’ and ‘busy weekday’ groups, the second most important in the ‘happy emotion’ group, and fourth most important in the ‘relaxing weekend’ group. This may indicate that, under these contextual situations, people may unconsciously be even more prone to choosing relatively unhealthy recipes over healthy ones. Salt and saturated fats are the most important features for the ‘relaxing weekend’ and ‘cold winter day’ groups, respectively. This may be due to the preference for larger meals during relaxing weekends, leading to higher salt levels. During cold winter days, hearty main dishes are preferred, potentially resulting in higher levels of calories and saturated fats. This analysis also suggests, as posited earlier, that the contextual scenario can act as a moderator of other predictive features in RS models.

5 DISCUSSION AND IMPLICATIONS

This study investigated whether people’s recipe rating and implied eating and nutritional intake behaviour changed under different simulated contextual situations. Additionally, the study examined whether integrating contextual features could improve model-based recommendation performance and identified which features were the most important in each contextual scenario.

We found out that people’s eating preference and the likelihood of consuming healthy recipe during busy weekdays differ significantly from other contextual situation. This difference is supported by the results of a one-way ANOVA, which indicated a significant variation in mean ratings during busy weekdays compared to situations such as cold winter days, sad emotions, after physical activities, and the generic group. The demanding and busy work schedule during weekdays often leaves individuals with limited time for cooking. This constraint may restrict their freedom to think and choose preferred food or recipes [37]. In this situation, the primary goal of cooking and eating becomes satisfying hunger, and meals should preferably be prepared and completed quickly. Consequently, recipes that are easy and quick to make, often involving refined or processed products and other potentially less healthy options, are preferred during busy weekdays. Our findings on distinct implied eating behaviour during busy weekdays align with the research conducted by Pinho et al. [37], which suggests that a hectic lifestyle may lead to reduced consumption of vegetables and home-cooked meals.

In addition, it is worth considering that the varying stress levels associated with a busy lifestyle, as supported by [25], may result in

individuals exhibiting a shift in their dietary choices, characterised by a reduction in main meals and an increase in snack consumption. This phenomenon could also explain the observed significant differences, such as those evident when comparing busy weekdays with cold winter days. In the latter case, individuals may gravitate towards carbohydrate-rich options, such as main dishes, potentially contributing to this distinction in eating preferences.

As anticipated, there is a significant difference in recipe preference between ‘hot summer’s days’ and ‘cold winter days’, providing evidence for the seasonality of food preferences over time [51]. Furthermore, the analysis of recipe health levels revealed a notably higher consumption of main dishes during the ‘cold winter day’ contextual group. This aligns with the findings of Capita and Alonso-Calleja [5], who concluded that both men and women tend to consume more energy during the winter months [5].

Surprisingly, we did not find significant differences between participants in the happy and sad emotion groups. Individuals appeared to exhibit similar food preferences, even under these extremely different emotional situations. Under both of these scenarios, participants showed an increased demand for unhealthy food [3], relative to the baseline condition. However, this may be a limitation in the ability of the external emotion stimuli method to effectively evoke such strong and polarised emotions.

Perhaps unsurprisingly, individuals prioritised nutritional information and opted for relatively healthier recipes after engaging in physical activities [1]. This may indicate that social norms and health-conscious behaviours play a significant role in influencing food choices after physical exertion.

Beyond its theoretical implications, our study offers a novel perspective on the development of context-aware food recommender systems. Previous research has predominantly focused on location-aware or gender-aware food recommender systems [2, 41]. In our algorithmic experiments, contextual features emerged as the most influential among all 29 considered features, leading to improved accuracy in rating predictions. Importantly, the datasets within each contextual scenario group exhibited higher R2 values compared to the baseline group. This is likely due to the baseline group encompassing a wide range of random possibilities, making it challenging for the model to discern meaningful patterns. In contrast, the inclusion of contextual features provides the model with a clearer direction for uncovering hidden patterns. Our findings, therefore, suggest that emotion-aware systems could represent the next generation of food recommender systems. This could also be combined with season- and stress level-awareness.

Previous food recommendation systems have primarily focused on either context-aware or healthy recommendations [41, 52]. Our research has demonstrated examples where the contextual scenario acts as a moderator, allowing other features to perform better than they would without the context as a precedent. For example, during busy weekdays, there was a noticeable increase in the consumption of unhealthy food. Addressing how to incorporate nutritional information into recommender systems to encourage healthy eating habits during hectic lifestyles could present a novel approach to balancing the trade-offs involved in healthy food recommendations. Currently, such systems have not been proposed in the field of

Table 8: Summary of top 5 most important features for each contextual scenario model

	Features
HSD	current country living time, gender, cooking skill level, home country, online recipe searching frequency
CWD	saturates, cooking frequency, WHO health level, category, sugar
H	home country, FDA health level, gender, cook book use frequency, saturates
S	FDA health level, fat, ethnic origin, cook book use frequency, home country
B	FDA health level, current country living duration, home country, current living country, recipe name
R	salt, online recipe searching frequency, total weight per portion, FDA health level, recipe name
APA	cooking skill level, gender, home country, sugar, total weight per portion
G	cooking frequency, sugar, total weight per portion, survey ID, cook book use frequency

Note: The abbreviations are the same as in Figure 3.

food recommendation. While our results are promising, they warrant validation on a larger, more naturalistic dataset for practical implementation.

6 LIMITATIONS AND FUTURE WORK

There are several limitations in our study that are worth pointing out. As commonly done in the food recommendation literature, we take ratings as a proxy for intent to consume. As such the behaviour measured in this research is implied behaviour - none of the recipes were actually consumed. More in-depth research is needed to investigate whether people's actual behaviour changes under different contextual situations. Despite our efforts to control for the impact of individual contextual factors in our experimental design, the results may still be affected by uncontrolled real-world variables. Our user studies were necessarily somewhat contrived and simulating emotions is clearly not the same as experiencing them naturally. Manipulating simulated contexts may lead to the representation of an artificial nature, introducing a potential conflict between simulated scenarios and the real world. Therefore, a study of real-life contexts is needed to thoroughly confirm the findings presented in this study.

Additionally, our examination of user ratings was based on a limited sample of recipe images ($n = 75$) and only included four main categories (main dishes, soup, desserts/snacks, and salads). A larger and more diverse set of recipes could lead to more generalisable study results. The performance of the XGBoost model was poor in this study. This may be because we did not perform feature selection and hyperparameter optimisation. As the primary aim of this research is to understand and identify the most contributing contextual factors for recipe rating prediction, the current study represents a preliminary component in a larger project. The complete model training will be reflected in the next stage of the research.

For future work, we plan to expand our study by including information on ingredients and cooking methods to gain a deeper understanding of how participants' food preferences differ at these levels. Secondly, we intend to develop an emotion-aware post-filtering healthy food recommendation system based on the available dataset. Finally, given that the contextual scenarios investigated are not necessarily mutually exclusive, to intend to investigate how combinations of dynamic contextual factors impact people's eating and nutritional intake behaviour.

ACKNOWLEDGMENTS

We would like to express our gratitude to the University of Sheffield for providing funding for this research. Additionally, we extend our sincere appreciation to all the participants who generously contributed their time and insights to our study.

REFERENCES

- [1] Ana Carolina Aguiar-Bloemer and Rosa Wanda Diez-Garcia. 2018. Influence of emotions evoked by life events on food choice. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity* 23 (2018), 45–53.
- [2] Mohanad Al-Ghobari, Amgad Muneer, and Suliman Mohamed Fati. 2021. Location-Aware Personalized Traveler Recommender System (LAPTA) Using Collaborative Filtering KNN. *Computers, Materials & Continua* 69, 2 (2021).
- [3] Elena Bartkiene, Vesta Steibliene, Virginija Adomaitiene, Grazina Juodeikiene, Darius Cernauskas, Vita Lele, Dovile Klupsaite, Daiva Zadeike, Laura Jarutiene, Raquel PF Guiné, et al. 2019. Factors affecting consumer food preferences: food taste and depression-based evoked emotional expressions with the use of face reading technology. *BioMed Research International* 2019 (2019).
- [4] Caron F Bove, Jeffery Sobal, and Barbara S Rauschenbach. 2003. Food choices among newly married couples: convergence, conflict, individualism, and projects. *Appetite* 40, 1 (2003), 25–41.
- [5] Rosa Capita and Carlos Alonso-Calleja. 2005. Differences in reported winter and summer dietary intakes in young adults in Spain. *International journal of food sciences and nutrition* 56, 6 (2005), 431–443.
- [6] Nicoletta Cavazza, Margherita Guidetti, and Fabrizio Butera. 2015. Ingredients of gender-based stereotypes about food. Indirect influence of food type, portion size and presentation on gendered intentions to eat. *Appetite* 91 (2015), 266–272.
- [7] Hao Cheng, Markus Rokicki, and Eelco Herder. 2017. The influence of city size on dietary choices. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*. 231–236.
- [8] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences*. Academic press.
- [9] Munmun De Choudhury, Sanket Sharma, and Emre Kiciman. 2016. Characterizing dietary choices, nutrition, and language in food deserts via social media. In *Proceedings of the 19th acm conference on computer-supported cooperative work & social computing*. 1157–1170.
- [10] Carol M Devine. 2005. A life course perspective: understanding food choices in time, social location, and history. *Journal of nutrition education and behavior* 37, 3 (2005), 121–128.
- [11] Carol M Devine, Margaret Connors, Carole A Bisogni, and Jeffery Sobal. 1998. Life-course influences on fruit and vegetable trajectories: qualitative analysis of food choices. *Journal of Nutrition Education* 30, 6 (1998), 361–370.
- [12] David Elsweiler, Morgan Harvey, Bernd Ludwig, and Alan Said. 2015. Bringing the "healthy" into Food Recommenders. *Dmrs* 1533 (2015), 33–36.
- [13] David Elsweiler, Christoph Trattner, and Morgan Harvey. 2017. Exploiting food choice biases for healthier recipe recommendation. In *Proceedings of the 40th international acm sigir conference on research and development in information retrieval*. 575–584.
- [14] Ercan Ezin, Hugo Alcaraz-Herrera, and Iván Palomares. 2019. Balancing Preferences, Popularity and Location in Context-Aware Restaurant Deal Recommendation: A Bristol, Cardiff and Brighton Case Study. In *Proceedings of the 13th ACM Conference on Recommender Systems and workshop on recommender in tourism (RecTour) 2019, Copenhagen, Denmark*. 38.
- [15] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (2007), 175–191.

- [16] Fda. 2022. *How to Understand and Use the Nutrition Facts Label*. <https://www.fda.gov/food/new-nutrition-facts-label/how-understand-and-use-nutrition-facts-label> Accessed June 20.
- [17] Fsa. 2016. *Guide to creating a front of pack (FoP) nutrition label for pre-packed products sold through retail outlets*. https://www.food.gov.uk/sites/default/files/media/document/fop-guidance_0.pdf Accessed June 20.
- [18] Qian Gao, Pengcheng Ma, et al. 2020. Graph neural network and context-aware based user behavior prediction and recommendation system research. *Computational Intelligence and Neuroscience* 2020 (2020).
- [19] Xiaoyan Gao, Fuli Feng, Xiangnan He, Heyan Huang, Xinyu Guan, Chong Feng, Zhaoyan Ming, and Tat-Seng Chua. 2019. Hierarchical attention network for visually-aware food recommendation. *IEEE Transactions on Multimedia* 22, 6 (2019), 1647–1659.
- [20] EL Gibson. 2006. Mood, emotions and food choice. In *The psychology of food choice*. CABI Wallingford UK, 113–140.
- [21] Morgan Harvey, Bernd Ludwig, and David Elswailer. 2012. Learning user tastes: a first step to generating healthy meal plans. In *First international workshop on recommendation technologies for lifestyle change (lifestyle 2012)*, Vol. 18. Citeseer.
- [22] Morgan Harvey, Bernd Ludwig, and David Elswailer. 2013. You are what you eat: Learning user tastes for rating prediction. In *String Processing and Information Retrieval: 20th International Symposium, SPIRE 2013, Jerusalem, Israel, October 7-9, 2013, Proceedings 20*. Springer, 153–164.
- [23] Wanrou Hu, Gan Huang, Linling Li, Li Zhang, Zhiguo Zhang, and Zhen Liang. 2020. Video-triggered EEG-emotion public databases and current methods: a survey. *Brain Science Advances* 6, 3 (2020), 255–287.
- [24] Wanrou Hu, Zhiguo Zhang, Huilin Zhao, Li Zhang, Linling Li, Gan Huang, and Zhen Liang. 2023. EEG microstate correlates of emotion dynamics and stimulation content during video watching. *Cerebral Cortex* 33, 3 (2023), 523–542.
- [25] Nikoline Bach Hyldelund, Chanette Frederiksen, Derek Victor Byrne, and Barbara Vad Andersen. 2022. Is stress taking the pleasure out of food?—A characterization of the food pleasure profiles, appetite, and eating behaviors of people with chronic stress. *Foods* 11, 13 (2022), 1980.
- [26] Maryam Imani and Gholam Ali Montazer. 2019. A survey of emotion recognition methods with emphasis on E-Learning environments. *Journal of Network and Computer Applications* 147 (2019), 102423.
- [27] Shristi Shakya Khanal, PWC Prasad, Abeer Alsadoon, and Angelika Maag. 2020. A systematic review: machine learning based recommendation systems for e-learning. *Education and Information Technologies* 25 (2020), 2635–2664.
- [28] Jinyoung Kim and Andrew Gambino. 2016. Do we trust the crowd or information system? Effects of personalization and bandwagon cues on users' attitudes and behavioral intentions toward a restaurant recommendation website. *Computers in Human Behavior* 65 (2016), 369–379.
- [29] Tomasz Kusmierczyk, Christoph Trattner, and Kjetil Nørvåg. 2015. Temporal patterns in online food innovation. In *Proceedings of the 24th international conference on world wide web*. 1345–1350.
- [30] Chin-Hui Lai, Duen-Ren Liu, and Kun-Sin Lien. 2021. A hybrid of XGBoost and aspect-based review mining with attention neural network for user preference prediction. *International Journal of Machine Learning and Cybernetics* 12 (2021), 1203–1217.
- [31] Anne Lhuissier, Christine Tichit, France Caillavet, Philippe Cardon, Ana Masullo, Judith Martin-Fernandez, Isabelle Parizot, and Pierre Chauvin. 2013. Who still eats three meals a day? Findings from a quantitative survey in the Paris area. *Appetite* 63 (2013), 59–69.
- [32] Yu-li Liu, Wenjia Yan, Bo Hu, Zhuoyang Li, and Yik Ling Lai. 2022. Effects of personalization and source expertise on users' health beliefs and usage intention toward health chatbots: Evidence from an online experiment. *Digital Health* 8 (2022), 20552076221129718.
- [33] Michael Macht. 2008. How emotions affect eating: A five-way model. *Appetite* 50, 1 (2008), 1–11.
- [34] Rui Maia and Joao C Ferreira. 2018. Context-aware food recommendation system. *Context-aware food recommendation system* (2018), 349–356.
- [35] Weiqing Min, Shuqiang Jiang, and Ramesh Jain. 2019. Food recommendation: Framework, existing solutions, and challenges. *IEEE Transactions on Multimedia* 22, 10 (2019), 2659–2671.
- [36] Marie-Claude Paquette and Carol M Devine. 2000. Dietary trajectories in the menopause transition among Québec women. *Journal of Nutrition Education* 32, 6 (2000), 320–328.
- [37] MGM Pinho, JD Mackenbach, H el ene Charreire, J-M Oppert, H B ardos, K Glonti, H Rutter, Sofie Compernelle, Ilse De Bourdeaudhuij, JWJ Beulens, et al. 2018. Exploring the relationship between perceived barriers to healthy eating and dietary behaviours in European adults. *European journal of nutrition* 57 (2018), 1761–1770.
- [38] Marie Reid and Richard Hammersley. 1999. The effects of sucrose and maize oil on subsequent food intake and mood. *British Journal of Nutrition* 82, 6 (1999), 447–455.
- [39] Markus Rokicki, Eelco Herder, Tomasz Ku mierczyk, and Christoph Trattner. 2016. Plate and prejudice: Gender differences in online cooking. In *Proceedings of the 2016 conference on user modeling adaptation and personalization*. 207–215.
- [40] Markus Rokicki, Christoph Trattner, and Eelco Herder. 2018. The impact of recipe features, social cues and demographics on estimating the healthiness of online recipes. In *Proceedings of the international AAAI conference on web and social media*, Vol. 12.
- [41] Mehrdad Rostami, Mourad Oussalah, and Vahid Farrahi. 2022. A novel time-aware food recommender-system based on deep learning and graph clustering. *IEEE Access* 10 (2022), 52508–52524.
- [42] Deepjyoti Roy and Mala Dutta. 2022. A systematic review and research perspective on recommender systems. *Journal of Big Data* 9, 1 (2022), 59.
- [43] Paul Rozin. 2007. Food and eating. *Handbook of cultural psychology* 2 (2007), 447–477.
- [44] Christian E Salas, Darinka Radovic, and Oliver H Turnbull. 2012. Inside-out: Comparing internally generated and externally generated basic emotions. *Emotion* 12, 3 (2012), 568.
- [45] Igor Andr e Pegoraro Santana and Marcos Aur elio Domingues. 2022. Context-Aware Music Recommendation with Metadata Awareness and Recurrent Neural Networks. *Computing and Informatics* 41, 3 (2022), 834–860.
- [46] Benjamin Senauer, Elaine Asp, Jean Kinsey, et al. 1991. *Food trends and the changing consumer*. Eagan Press.
- [47] Richard Shepherd and Monique Raats. 2006. *The psychology of food choice*. Vol. 3. Cabi.
- [48] Jeffery Sobal, Carole A Bisogni, Carol M Devine, and Margaret Jastran. 2006. A conceptual model of the food choice process over the life course. In *The psychology of food choice*. Cabi Wallingford UK, 1–18.
- [49] Jeffery Sobal and Mary K Nelson. 2003. Commensal eating patterns: a community study. *Appetite* 41, 2 (2003), 181–190.
- [50] Rukshani Somarathna, Tomasz Bednarz, and Gelareh Mohammadi. 2022. Virtual reality for emotion elicitation—a review. *IEEE Transactions on Affective Computing* (2022).
- [51] Charles Spence. 2021. Explaining seasonal patterns of food consumption. *International journal of gastronomy and food science* 24 (2021), 100332.
- [52] Alain D Starke and Christoph Trattner. 2021. Promoting healthy food choices online: a case for multi-list recommender systems. In *Proceedings of the ACM IUI 2021 Workshops*.
- [53] Alain D Starke, Martijn C Willemsen, and Christoph Trattner. 2021. Nudging healthy choices in food search through visual attractiveness. *Frontiers in Artificial Intelligence* 4 (2021), 621743.
- [54] Michael Symons. 1991. *Eating into Thinking: Explorations in the sociology of cuisine*. Ph. D. Dissertation. Flinders University of South Australia.
- [55] Robert D Tennyson and Klaus Breuer. 2002. Improving problem solving and creativity through use of complex-dynamic simulations. *Computers in Human Behavior* 18, 6 (2002), 650–668.
- [56] Christoph Trattner and David Elswailer. 2017. Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems. In *Proceedings of the 26th international conference on world wide web*. 489–498.
- [57] Christoph Trattner, David Elswailer, and Simon Howard. 2017. Estimating the healthiness of internet recipes: a cross-sectional study. *Frontiers in public health* 5 (2017), 16.
- [58] Christoph Trattner, Dominik Moesslang, and David Elswailer. 2018. On the predictability of the popularity of online recipes. *EPJ Data Science* 7, 1 (2018), 1–39.
- [59] Tsuguya Ueta, Masashi Iwakami, and Takayuki Ito. 2011. A recipe recommendation system based on automatic nutrition information extraction. In *Knowledge Science, Engineering and Management: 5th International Conference, KSEM 2011, Irvine, CA, USA, December 12-14, 2011, Proceedings 5*. Springer, 79–90.
- [60] Tatjana van Strien, Ausia Cebolla, Ernestina Etchemendy, Jos e Gutierrez-Maldonado, Marta Ferrer-Garcia, Cristina Botella, and Rosa Ba nos. 2013. Emotional eating and food intake after sadness and joy. *Appetite* 66 (2013), 20–25.
- [61] Wenjie Wang, Ling-Yu Duan, Hao Jiang, Peiguang Jing, Xuemeng Song, and Liqiang Nie. 2021. Market2Dish: Health-aware food recommendation. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 17, 1 (2021), 1–19.
- [62] Joint Who and FAO Expert Consultation. 2003. Diet, nutrition and the prevention of chronic diseases. *World Health Organ Tech Rep Ser* 916, i-viii (2003), 1–149.
- [63] Haochao Ying, Jian Wu, Guandong Xu, Yanchi Liu, Tingting Liang, Xiao Zhang, and Hui Xiong. 2019. Time-aware metric embedding with asymmetric projection for successive POI recommendation. *World Wide Web* 22 (2019), 2209–2224.
- [64] Jing Yuan, Fikret Sivrikaya, Frank Hopfgartner, Andreas Lommatzsch, and Mu Mu. 2015. Context-aware LDA: Balancing Relevance and Diversity in TV Content Recommenders. In *Proceedings of the 2nd Workshop on Recommendation Systems for Television and Online Videos (RecSysTV 2015)*.
- [65] Haoxiang Zhang, Wenchao Bai, Jingyi Ding, and Jiahui Jin. 2023. Time-Aware POI Recommendation Based on Multi-Grained Location Grouping. In *2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. IEEE, 1796–1801.
- [66] Qing Zhang, David Elswailer, and Christoph Trattner. 2020. Visual cultural biases in food classification. *Foods* 9, 6 (2020), 823.

A APPENDIX: WHO, FDA AND FSA STANDARD RANGES OF IDEAL NUTRITIONAL INTAKE GUIDE

Table 9: WHO standard ranges of population nutrient intake goals

Dietary factor	Goal (% of total energy)
Total fat	15-30%
Saturated fatty acids	<10%
Polyunsaturated fatty acids (PUFAs)	6-10%
n-6 Polyunsaturated fatty acids (PUFAs)	5-8%
n-3 Polyunsaturated fatty acids (PUFAs)	1-2%
Trans fatty acids	<1%
Monounsaturated fatty acids (MUFAs)	By difference
Total carbohydrate	55-75%
Free sugars	<10%
Protein	10-15%
Cholesterol	<300 mg per day
Sodium chloride (sodium)	<5 g per day (<2 g per day)
Fruits and vegetables	>=400 g per day
Total dietary fibre	From foods
Non-starch polysaccharides (NSP)	From foods

Table 10: FDA general guide to %DV

Nutrient	DV	%DV	Goal
Saturated Fat	20g	=100% DV	Less than
Sodium	2,300mg	=100% DV	Less than
Dietary Fiber	28g	=100% DV	At least
Added Sugars	50g	=100% DV	Less than
Vitamin D	20mcg	=100% DV	At least
Calcium	1,300mg	=100% DV	At least
Iron	18mg	=100% DV	At least
Potassium	4,700mg	=100% DV	At least

B APPENDIX: DEMOGRAPHIC FEATURES CHARACTERISATION

The table 12 below shows summarise the key demographic features of this research.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009

Table 11: FSA Criteria for 100g of food (whether or not it is sold by volume)

Text	LOW	MEDIUM	HIGH	
Colour code	Green	Amber	Red	
			>25% of RIs	>30% of RIs
Fat	<= 3.0g/100g	>3.0g to <=17.5g/100g	>17.5g/100g	>21g/portion
Saturates	<= 1.5g/100g	>1.5g to <=5.0g/100g	>5.0g/100g	>6.0g/portion
(Total) Sugars	<= 5.0g/100g	>5.0g to <= 22.5g/100g	>22.5g/100g	>27g/portion
Salt	<= 0.3g/100g	>0.3g to <=1.5g/100g	>1.5g/100g	>1.8g/portion

Table 12: Demographic features characterisation

Demographic data	Elements	Frequency (N)	Percentage (%)
Gender	Male	212	53.4%
	Female	177	44.6%
	Non-binary or gender diverse	5	1.3%
	Others	3	0.7%
Age	18-24	43	10.8%
	25-34	122	30.7%
	35-44	100	25.2%
	45-54	77	19.4%
	55-64	41	10.3%
	65-74	13	3.3%
Ethnic origin	Above 75	1	0.3%
	White	329	82.9%
	Asian or Asian British	34	8.6%
	Black, Black British, Caribbean or African	16	4.0%
	Mixed or Multiple Ethnic Groups	10	2.5%
	Others	3	0.7%
	Prefer not to say	5	1.3%
Physical activities level	1 (very inactive)	15	3.8%
	2 (inactive)	54	13.6%
	3 (Medium)	188	47.4%
	4 (Active)	104	26.2%
	5 (Very active)	36	9.1%
Cooking frequency	Never	8	2.0%
	Yearly	12	3.0%
	Quarterly	23	5.8%
	Monthly	50	12.6%
	Weekly	120	30.2%
	Daily	184	46.3%
Cooking skill level	No experience	5	1.3%
	Beginner	67	16.9%
	Intermediate	305	76.8%
	Expert	20	5.0%
Cook book use frequency	Never	89	22.4%
	Yearly	73	18.4%
	Quarterly	87	21.9%
	Monthly	102	25.7%
	Weekly	40	10.1%
	Daily	6	1.5%
Online recipe searching frequency	Never	14	3.5%
	Yearly	35	8.8%
	Quarterly	71	17.9%
	Monthly	127	32.0%
	Weekly	135	34.0%
	Daily	15	3.8%
Total number of participants		397	