Understanding Aesthetic Aspects of Cross-Cultural Food Preferences with Online Recipe Platforms

Inaugural-Dissertation zur Erlangung der Doktorwürde
der Fakultät für Sprach-, Literatur- und Kulturwissenschaften
der Universität Regensburg

Vorgelegt von
Qing Zhang
aus
Jiangsu, China

Abgabe: Oktober 2022
Erstgutachter: PD Dr. David Elsweiler
Zweitgutachter: Prof. DI Dr. Christoph Trattner, BSc

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor Dr. David Elsweiler. I am beyond fortunate to do research under his supervision since 2017. Thanks for giving me freedom to develop my own motivation and in the meantime, always inspiring me to think deeper with the "Why" questions, which would be precious gifts of my life.

I would also like to thank Dr. Christoph Trattner, who always supports me with his creative ideas and patience. The advice you have given me during our fruitful discussion drives me to progress with my research and encourages me to try more.

Thanks to both Dr. Elsweiler and Dr. Trattner for introducing me to such an amazing research field and creating a stimulating research environment for me to enjoy working and grow up like an independent researcher.

I am grateful to be supported by my great colleagues, it is a wonderful experience to work with you. Firstly, Melanie, I am so lucky to share the office with you. I cannot remember how many times you have reached out to me when I met problems. In addition, I really appreciate you and Selina for supporting my writing. I cannot make such big process on my writing without your feedback and suggestions. I would always remember what I have learned from our inspiring and productive writing lab. Robert and Manuel, thank you guys for helping me when I was in trouble with our server and database. Thanks for saving my data on 03.12.2019! Thanks to Alex, with whom I have discussed my research and paper. Thanks to all my colleagues, past and present, Bernd, Christina, Elisabeth, Florian, Gregor, Heidi, Jürgen, Markus, Markus, Ray and Udo. I will miss the time when we have lunch and every moment we share together. Wish you all the best!

Thanks to my main funding party, China Scholarship Council, to support this research. I also gratefully acknowledge the support from the Chairs of Information Science at the University of Regensburg, who made it possible for me to start and complete my PhD. Their support included financial support at the early stage. Also, thank you to the Bavarian Programme for the Realisation of Equal Opportunities for Women in Research and Teaching for my extended stay in Regensburg.

Thank you, Barbara, who works in the writing centre of University of Regensburg. You have been a great help to me in revising the manuscript.

I would like to thank my friends, Chen Zhang, for being my awesome roommate and for being a sincere friend, and Tonghe Zhuang, thanks for being there during my though time.

Special thanks go to my family, especially my mother Shuhui Zhang and father Benyin Zhang, who are my shield all the time. Thanks for always understanding my choices and supporting my career, both financially and spiritually.

To my partner, Yanpu Chen. Thank you for recognizing my abilities and accepting my vulnerability since we met. It is impossible for me to be there without your support and love.

Finally, thanks to myself for making the brave decision 4 years ago and surviving from the adventure =).

Abstract

Food recommender systems offer the potential to provide meal suggestions that account not only for user preferences, but also reflect ideals, such as healthfulness and sustainability. Predicting user food preferences is, however, a challenging problem given the diverse factors that influence what an individual likes to eat. Culture is known to be one of the factors. Food across cultures shows differences in many aspects, for example, in aesthetics. This doctoral work investigates the aesthetic aspects, particularly in terms of visual appearance and flavour, of food. Focusing on how food looks and tastes in different cultures and how human food choices are related to these, this work aims at analysing and interpreting human food aesthetic preferences across cultures. Based on these observations, I show implications for the development of food recommender systems. Online recipe platforms from three distinct cultures, including China, US and German, serve as data resources for this research, as there are a large amount of data representing the corresponding food cultures (e.g., online recipes) and user food preferences (e.g., ratings, bookmarks). Computer science techniques are employed to extract the aesthetic information, including visual features from the online recipe images and flavour compounds of the ingredients, allowing for successively classification and prediction tasks to be performed on these data by means of machine learning approaches. The algorithmic results show that online recipes differ in visual appearance and flavour across recipe portals. This, together with the findings from the follow-up user study demonstrating how culture biases human interpretation of online recipe images, highlights the impact of culture in human food choices. The algorithms also confirm that human online food preferences are aesthetically driven within each culture. Furthermore, stable patterns in aesthetic food preferences across cultures can be identified, which is supported and justified by means of further user study and exploratory analyses. The research presented in this doctoral work increases the understanding of human cross-cultural food preferences. Moreover, findings from this thesis emphasise the merit of considering the synthetic impact of culture and aesthetics into food recommendation and provide a promising perspective for the development of food recommender systems by incorporating stable patterns in cross-cultural aesthetic food preferences.

Zusammenfassung

Food-Recommender-Systeme bieten die Möglichkeit Essensvorschläge zu machen, die neben Nutzervorlieben auch Kriterien wie Gesundheit und Nachhaltigkeit berücksichtigen. Essensvorlieben von Nutzern und Nutzerinnen vorherzusagen stellt jedoch eine Herausforderung dar, da verschiedene Faktoren Einfluss darauf haben, was eine Person gerne isst. Kultur ist bekanntermaßen einer dieser Faktoren. Unterschiede zwischen Kulturen spiegeln sich auch beim Essen wider, zum Beispiel in Bezug auf dessen Ästhetik. Diese Doktorarbeit untersucht ästhetische Aspekte von Gerichten in verschiedenen Kulturen und konzentriert sich dabei auf deren Aussehen und Geschmack. Zudem wird der Frage nachgegangen, wie menschliche Entscheidungen bezüglich Lebensmitteln damit zusammenhängen. Auf der Grundlage dieser Beobachtungen werden Implikationen für die Entwicklung von Food-Recommender-Systemen aufgezeigt. Aufgrund der großen verfügbaren Menge an Rezepten und Nutzerinteraktionsdaten dienen Online-Rezeptplattformen aus drei unterschiedlichen Kulturen, darunter China, die USA und Deutschland, als Datengrundlage für die vorliegende Forschung. Um ästhetische Informationen, wie visuelle Merkmale der Online-Rezeptbilder und Geschmacksverbindungen der Zutaten, zu extrahieren, werden Methoden aus der Informatik eingesetzt. Mittels maschinellen Lernens werden sukzessive Klassifizierungs- und Vorhersageaufgaben auf diesen Daten durchgeführt. Die Ergebnisse zeigen, dass sich Online-Rezepte auf den verschiedenen Rezeptportalen in ihrer visuellen Erscheinung und ihrem Geschmack unterscheiden. Eine anschließende Nutzerstudie zeigt, wie kulturelle Aspekte die Wahrnehmung von Online-Rezeptbildern verzerren und unterstreicht ihren Einfluss darauf, wie Menschen Entscheidungen zu Gerichten treffen. Die Resultate bestätigen außerdem, dass Essenspräferenzen im Internet innerhalb jeder Kultur ästhetisch geprägt sind. Darüber hinaus sind kulturübergreifende Muster in ästhetischen Essenspräferenzen erkennbar, was durch weitere Nutzerstudien und explorative Analysen gestützt und begründet wird. Die in dieser Doktorarbeit vorgestellten Forschungsergebnisse tragen zu einem besseren Verständnis der kulturübergreifenden Essensvorlieben bei. Darüber hinaus unterstreichen die Ergebnisse dieser Arbeit den Wert der Berücksichtigung des synthetischen Einflusses von Kultur und Ästhetik bei der Empfehlung von Essen und bieten eine vielversprechende Perspektive für die Entwicklung von Food-Recommender-Systemen durch die Einbeziehung stabiler Muster in kulturübergreifenden ästhetischen Essenspräferenzen.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

In nutritional anthropology, it is well established that food is much more than mere sustenance and serves diverse needs from health and well-being, to control, social contact, and ritual (Anderson, 2014). Making decisions regarding what to eat is complex, and the evidence suggests that we have difficulties in dealing with the issue (Palojoki & Tuomi-Gröhn, 2001; Wansink & Sobal, 2007). Substantial effort has been undertaken in diverse fields to understand the food choices people make.

As all people share a similar physiological basis, individuals have an inherent common tendency to crave food that provides critical nutrients, such as carbohydrates and fat, to sustain life. Since ancient times, grains have been generally preferred over the grass family foods (Anderson, 2014), and meat has traditionally been consumed on a worldwide basis (Anderson, 2014). Over time, in order to adapt to their environments, humans have developed varying diets. For example, the Inuit diet consists nearly exclusively of meat and fats, in contrast to that of farmers in South-East Asia, which contains almost no animal protein at all (Fischler, 1988). Yet, providing nutrition and ensuring survival are not the only roles that food plays. It also addresses other human needs, such as health maintenance and social interaction (Anderson, 2014). The development of agriculture, along with food industry marketing (Kearney, 2010), have led to an abundance of diverse foods chosen and consumed. With so many options, choosing what to eat nowadays has become an important and complicated issue for many people. The well documented paradox of choice scenario in the food domain indicates humans have difficulty with making food choices (Palojoki & Tuomi-Gröhn, 2001). Moreover, the increasing incidence of lifestyle related illness (e.g., obesity, diabetes, hypertension, and stroke) suggests that poor food choices are often being made (Meyer et al., 2000; Morrill & Chinn, 2004; Aburto et al., 2013; Abbar, Mejova, & Weber, 2015; Mendonça et al., 2017; Organization et al., 2019).

To enable people to make more informed and appropriate food choices, food recommender systems have been developed to predict the preferences of users for unrated food and to recommend new foods (Trang Tran et al., 2018). These systems have been touted as both useful and valuable means to support people making choices that are satisfying (Freyne et al., 2011; Harvey et al., 2013), healthy (Elsweiler et al., 2017; Ge et al., 2015; A. D. Starke et al., 2021), and sustainable food choices (A. Starke et al., 2017; Tomkins et al., 2018; A. Starke, 2019). A main prerequisite for a food recommender system is to accurately predict what people would like to eat, which is extremely challenging due to the complexity of human eating habits. As an "evolutionary product of environmental conditions and of the basic forces, especially social institutions and social relations that determine their use" (Harris & Ross, 1987), diet is context- and culturally-dependent (Bellisle, 1999). This suggests that explaining food choices requires a variety of factors, in addition to the biological and environmental features, the importance of context should not to be neglected.

One important contextual feature, which has been studied to understand the food choices is culture. It is shown in the research literature how people have varied diets in different social milieus (Cantarero et al., 2013; Leu & Banwell, 2016) and ethnic groups (Chrzan & Brett, 2017). Obvious evidence for this is the representative foods in the cuisine of each region (P. Rozin, 1996), like wine and baguette for France, and beer and sausage for Germany (Laufer et al., 2015). However, such differences with regard to ingredients alone cannot explain the food choices related to different cultures completely. One cultural aspect, which can help explain what we eat, as well as how much, is varying aesthetic ideals (Palmer & Schloss, 2010; Taylor et al., 2013). For food, aesthetics relates primarily to visual appearance (Linné et al., 2002; Spence et al., 2016), taste (Sherman & Billing, 1999) and smell (Ehrlichman & Bastone, 1992; Rolls, 2005). With respect to visual aspects, humans prefer food made "visually beautiful" (Anderson, 2014), which has varying rules across cultures. This can be illustrated by comparing the Japanese Bento, where food is cut into bite-sized pieces and well-organised, to the casual plating style preferred by people in the United States and Italy. In terms of taste, analogous differences exist. For example, in Sherman and Billing's study (1999) of typical recipes from different countries, the meat dishes analysed originating from African and Asian countries all featured at least one spice and often a combination of many, whereas, in Scandinavian countries, one-third of the recipes used no spices at all. Yet, some preferences are widespread, such as the taste for spicy, herbal and floral volatile oils (Sherman & Billing, 1999). All of the above indicate how food choices of a culture are reflected in its aesthetic ideals, some of them show the cultural boundaries, while others reveal the cross-cultural aesthetic preferences on food. Humans tend to choose food that align with the aesthetic characteristics of their cultural upbringings, while that characterises are likely not to exist when making food choices in other cultural contexts, such as when travelling, studying abroad or migrating. Fortunately, the aesthetics of food in different cultures have some common signals, which I assume would help people choose food in unfamiliar cultures. However, there have been limited studies on aesthetic aspects of food choice across-cultures.

Anthropologists traditionally learn about food habits using qualitative approaches; they primarily rely on data from observation, interviews and focus groups, such as the research in (Farquhar, 2002) and (Chrzan & Brett, 2017). These methods are usually expensive and time-consuming. In the Digital Humanities, or digitised resources are exploited using techniques from computer science, which allow the qualitative approaches traditionally employed in the humanities (close-reading) to be complemented with quantitative tools, enabling patterns to be unearthed in much larger samples or collections of interest (distant-reading) (Moretti, 2005). Such digital methods have been applied to online sources, such as traces from food portals (e.g., Wagner et al., 2014) and social media (e.g., Abbar et al., 2015; Holmberg et al., 2016). Albeit a relatively slow start compared to other domains, the digital food studies to date have been extremely fruitful (Leer & Krogager, 2021), which have provided insights into the food choices people make including how choices are influenced by temporal factors (Wagner et al., 2014; Kusmierczyk et al., 2015a), gender (Rokicki et al., 2016), geographical location (Wagner et al., 2014; Laufer et al., 2015) and social relations (Rokicki et al., 2017). Moreover, relying on the interaction between users and online recipes (e.g. ratings and comments left), researchers such as Harvey (2013) and Trattner (2018) have successfully understood and predicted human food preferences, while findings from (Elsweiler et al., 2017) revealed the biases influencing humans food choices and such biases can be exploited to nudge people towards choosing healthier food. Not only do these studies illustrate how the new medium has changed human food and eating practices, but also help with the development of food recommender systems.

To sum up, human food choices are manipulated by a diverse range of factors. It is hard for humans to make food choices that meet their real needs, some choices they have made even do harm to their health. Food recommender systems contribute to deal with the problems. However, building a food recommender system that is able to predict human food preferences is difficult, new perspectives and methods are required for this domain. The scholars, especially in anthropology, have shown the great value of cultural factors in explaining food behaviours, which include aesthetic ideals. However, little is known about what role aesthetics plays in influencing human food choices as well as to what extent it can help with improving the performance of food recommender systems. Building on these, the thesis focuses on the investigation of food choices on the aesthetic aspects. In addition, the limitations of the traditional methods of studying food choice in anthropology, with the promising digital food researches combine to motivate the studies in this doctoral work to depend on the digital resources and advanced data approaches from computer science.

1.2 Problem Statement and Objectives

The main aim of the thesis is to identify whether aesthetic ideals are culturally dependent. Not only the idiosyncraticities across cultures, but also the unifying aspects, that is, aesthetic ideals that are culturally agnostic would be investigated. This would be particularly useful in practical contexts, such as the cross-cultural recommendation systems, which are developed to be helpful in the situations where people move between cultures (e.g. travelling). In this thesis, the food aesthetic studies are related to visual aspects and flavours, which are prominent determinants of food choices regardless of the cultural background. With these in mind, the following research questions are addressed:

- *Issue 1.* To what extent is it possible to differentiate the food across cultures based on the representation of the food relating to visual appearance or flavour?

- *Issue 2.* To what extent is it possible to identify the differences and ascertain stable patterns of food preferences across cultures based on the representations of the food relating to visual appearance or flavour?

- *Issue 3.* To what extent is it possible to utilise stable patterns of the food preferences across cultures on visual and flavour aspects to build a cross-cultural food recommender system?

1.3 Approach and Methodology

To achieve the aim of understanding food aesthetic ideals with respect to cultures and subsequent research questions, I establish the empirical experiments in this thesis as several classification and prediction tasks that rely on digital sources and computer science approaches. After that, the patterns revealed by the algorithms are supported and explained by user studies and follow-up quantitative analysis. The food aesthetics on the visual and flavour aspects are first investigated separately, then they are combined and applied to a cross-cultural food preferences prediction task. The whole process of the experiments in this work is described below:

Step 1: Data Collection. The data are sourced from three popular recipe portals of three distinct food cultures - China, The United States (US) and Germany, with recipe images (for visual analysis), ingredient lists (for flavour analysis) as well as user interaction data (for preference prediction) is collected. In Chapter 4, the recipe portals and strategies for obtaining the data will be described in detail.

Step 2: Data Pre-processing & Representation. In this thesis, the step contains balancing datasets, data cleaning, data mapping, data normalisation. In addition, in this step, the food data is transformed into multi-dimensional vectors by means of utilising visual information encoded in the recipe images and flavour compounds of ingredients respectively, in order to represent food on the visual and flavour aspects. Chapter 4 provides the details.

Step 3: Predictive Modelling. Different algorithms are applied to build classifiers using the food representation on the visual and flavour aspects. There are two types of prediction tasks, one for identifying the origin culture of the food and another for predicting food preferences. In Chapter 5 and 6, the detailed process of modelling and fine-tuning, as well as the performance of each classifier is presented.

Step 4: Model Validation & Model Explanation. The results obtained from the classifiers in Step 3 are validated and explained in this step. The visual classifiers are validated by humans with large-scale user studies, the data acquired from which are analysed quantitatively and qualitatively. I elaborate the process and findings of the user studies for validating the classifiers trained using visual information in Chapter 5. The classifiers trained using the flavour compounds, instead, are explained with a series of exploratory analyses. These are presented in Chapter 6.

Step 5: Data Fusion. In this step, both the food representation on the visual and flavour aspects are applied to a cross-cultural food recommendation task. These are combined to test whether the combination of classifiers would outperform the individual ones.

1.4 Outline

Chapter 2: Working Definition

In this chapter, the definition and scope of the research object - food aesthetics - is described. It starts from reviewing the controversy about aesthetics of food in philosophy, including whether food can be seen as an aesthetic object and linking the dimensions of food aesthetics to several sensory inputs. I then clarify the research scope of this doctoral work and determine the material and medium for studying food aesthetics in this work.

Chapter 3: Related Work

This chapter presents the background for the food studies in detail in three parts. In the first part, previous food studies for understanding human online food behaviours are reviewed. The second reviews the studies that investigated food aesthetics with digital food traces. The literature with respect to food recommender systems is summarised in the third part, which presents an up to date overview of the contributions and challenges in food recommender systems. All three parts are summarised and linked to several open issues regarding this thesis.

Chapter 4: Data Preparation
The data used in this thesis are described in this chapter. The origins of the data and the strategy of selecting the data for empirical experiments in this thesis are introduced. Chapter 4 also describes the steps for data cleaning, pre-processing, as well as the methods for representing food relating to visual appearance and flavour.

Chapter 5: Cross-Cultural Food Classification and Preferences on the Visual Aspects
Chapter 5 presents experiments that provide insight into visual aspects of intercultural food aesthetics. In the first step, differences and similarities of the visual nature of online recipe images are studied algorithmically. A number of models are trained using food representation on the visual aspect to identify the origin cultures of the food. The models are subsequently validated by means of a user study. In the second step, models are created, validated and compared to shed light on which food are most visually appealing within and across cultures. In addition, this chapter identifies whether, and to what extent, stable patterns in visual food preferences across cultures can be ascertained.

Chapter 6: Cross-Cultural Food Classification and Preferences on the Flavour Aspects
This chapter elaborates the cross-cultural food studies on the flavour aspects. Analogous to chapter 5, models for differentiating recipes from different cultures as well as identifying preferred recipes within and across cultures are established, yet the models are trained using food representation on the flavour aspects. In addition, this chapter identifies whether, and to what extent, stable patterns in flavour preferences of food across cultures can be ascertained. The algorithmic results are then explained and justified by means of two additional exploratory experiments.

Chapter 7: Fusion of Visual Features and Flavour Compounds for Cross-Cultural Food Preferences Prediction
In chapter 7, an experiment by means of ensemble learning is designed to combine the food representation relating to visual appearance and flavour for cross-cultural food recommendation. Moreover, The performance of the representation of food on the aesthetic aspects will be compared to the commonly used textual information (i.e., ingredients) in the same recommendation task. These empirical experiments are shown in Chapter 7.

Chapter 8: Discussion
This chapter summarises the main findings from Chapter 5, 6 and 7, and discusses the theoretical and practical implications of the findings.

Chapter 9: Conclusion
This chapter concludes the whole thesis.

1.5 Publications Relating to this Thesis

A number of articles relating to this work have been published during my doctoral studies. Zhang et al. (2019) presented a preliminary work of the cross-cultural food studies on the visual aspects. In (Q. Zhang, Elsweiler, & Trattner, 2020), the identification of food from different cultures with visual properties is described. We reported the performance of the visual classifiers by means of machine learning approaches, and validated them with a large-scale online survey, which provided more details about how human food choices are biased by visual cues.

Chapter 2

Working Definition

2.1 Definition and Dimensions of Food Aesthetics

The debate regarding whether food can be considered as aesthetic object have been raging since the time of the ancient Greeks who argued that the pleasure obtained from eating food is lower to that from reading poems or listening to music, since the senses like smell and taste are unreliable, and that the hedonic experience brought by them would hinder humans pursuing truth and knowledge. Therefore, "only objects of sight and hearing could be beautiful and that food could not be beautiful because it was something that we smelled and tasted" (Sweeney, 2017). Future generations of philosophers insisted on this point of view, excluding food when talking about beauty. The debate about whether food can be considered beautiful went on for some time with new issues being raised. In the late eighteenth century, the term "aesthetic" was introduced by Alexander Baumgarten (Shiner, 2003), and was well-known for being interpreted by Immanuel Kant (1987). According to Kant, aesthetics is a critical category when enjoying something being beautiful involves a whole process of reflection and imagination, while eating food was only an immediate response to a stimulus. In addition, taste for food was completely based on an individual's preference. In his opinion, valuing something as beautiful or not demanded universal assent (Sweeney, 2017). It was not until Jean-Anthelme Brillat-Savarin contradicted this in a Kantian way, that the experience of tasting food is not merely a hedonic response, but arouses reflective enjoyment. He insisted that humans possess a common physiology of smelling and tasting is a necessity for perceiving aesthetics in Kant's theory (Kaplan, 2012), and that food can be an aesthetic object and smell and taste don't inferior to sight and hearing (Sweeney, 2017). Contemporary philosophers like Emily Brady (Kaplan, 2012) advocate a similar view and she believes the philosophical aesthetic underestimated the value of smell and taste. These facts underline a trend in food philosophy, that is, the significance of olfactory and gustatory sensation generated by food is being acknowledged and the aesthetic value of food itself is being accepted (Kaplan, 2012).

After determining food as an aesthetic object, the next issue is to determine the dimension of food aesthetics. Kaplan (2012) pointed out that food is aesthetic in two senses. First, it has a taste and it appeals to the senses. This is illustrated in humans' descriptions of food like "delicious", "satisfying" or "disgusting". Second, food is artful. Humans describe it in terms of its sensual composition. This argument implies the aesthetic dimension of food is closely related to sensory pleasure. It is well-established that eating is an experience involving multiple senses, including taste, smell, vision, touch and hearing

(Auvray & Spence, 2008; Kaplan, 2012; Korsmeyer & Sutton, 2011). By means of a series of psychology, sensory science and experimental aesthetics approaches, the mechanisms of the senses in the process of perceiving and judging food are being investigated, contributing to providing scientific disciplines for culinary practises and food product design, in order to enhance humans pleasantness ratings on food products (Amerine et al., 2013; Schifferstein et al., 2022). This doctoral work focuses on three sensory inputs, vision, taste and smell. Vision is associated with visual appearance of food, whereas taste and smell are related to the flavour companies of food. On these two aspects, plenty of findings were proven to be efficient, such as in terms of visual appearance, neat and artistic presentation make food to be perceived more appetising (Zellner et al., 2010; Michel et al., 2014). With respect to flavour, molecular gastronomy inspires the production of ingredient combinations with pleasing tastes and aroma like chocolate and caviar and chocolate and blue cheese (Ahn et al., 2011). Such examples encourage this work to investigate food aesthetics on these two aspects. In the following sections of this chapter, the association between visual and flavour aspects and food choices are elaborated in detail respectively. In addition, we also make mention of which media in the digital world can be used for these studies.

2.2 Visual Appearance of Food

The first sensory contact with food is through the eyes, such as pointed by Apicius: "the first taste is always with the eyes". The visual appearance of food provides a lot of information. For instance, one of the most prominent visual factors, colour, hints at the taste, Red is often associated with sweet and fruity flavours, while green is linked to sour flavour (Koch & Koch, 2003). Moreover, texture, which can be perceived by looking at the surface of a food product (Wilkinson et al., 2000), suggests the mouthfeel (e.g. hard, soft, crunchy, creamy etc.) (Guinard & Mazzucchelli, 1996) of the food. The means by which food is presented and arranged (e.g. shape, portion size and plating style) (Reisfelt et al., 2009; Olsen et al., 2012), as well as the plateware, can play a role in human food prescription and judgement (Piqueras-fiszman & Spence, 2011; Piqueras-Fiszman et al., 2012). Visual attributes of food appearance are various, understanding how they are involved in human food behaviours requires extensive materials and experiments. Food images have been proven to be a reliable tool for such research (van der Laan et al., 2011). That is not only because images convey the same visual cues as the food itself, but they have been shown to lead to increased salivation and other physiological changes as food itself does (Spence, 2011).

Nowadays, food is not only displayed on the dining-table, but also on the screen. Digital media have offered new platforms for humans to share content related to food, such as images of food that are uploaded to online recipe portals or on social media platforms such as Instagram and Pinterest (Lupton, 2020). Humans can observe the food appearance through profusion of online recipe images and make food choices based on it, the interaction data they leave behind when doing so provides information about their food consumption and practices (Lupton, 2020). In addition, encouraged by the development of computer vision, information, the images can be extracted more efficiently. A recent work from (Trattner et al., 2018) has shown the performance of low-level visual features like brightness, sharpness, contrast etc., of recipe images in predicting popularity of online recipes. The authors suggest the future direction of related works should be done on visual features of food images due to their great potential. By means of the cutting-edge deep learning technologies like VGG16, ResNet, deep neural network embeddings of food images can be derived and the information obtained from that is enough to be applied to recognizing ingredients (e.g., Joutou & Yanai, 2009; J. Chen & Ngo, 2016), retrieving recipes (e.g., Min et al., 2016; Salvador et al., 2017) or even estimating the nutrition contents (e.g., Y. He et al., 2013). Building on the findings that show the great power of digital materials and methods in the domain of food vision, this doctoral work collects the recipe images that show visual appearance of food from online recipe portals and apply computer vision approaches to extract information.

2.3 Smell, Taste and Flavour of Food

The senses of smell, taste and flavour are easily confused (P. Rozin, 1982), while contemporary science has helped to reveal the differences and overlap among them. Firstly, smell and taste are different senses with their own receptors, olfactory sensory neurons in our noses and taste buds over the tongue, respectively (Buck, 2000). Due to the large number of different olfactory receptors, humans can sense a large variety of smells, by contrast, the taste receptors can only distinguish a few tastes, which refer to salty, sweet, sour, bitter and umami basically (Lindemann, 2001). However, when referring to smells and tastes of food, their origins are actually the same, that is, the interaction of various chemical molecules (Buck, 2000). Therefore, these two terms, taste and flavour, are commonly interchangeable. It is noticed that even scientific papers or books do not make clear distinction between them (Schifferstein et al., 2022; Sweeney, 2017). Nevertheless, there are still scholars who describe flavour as the synthesis of olfactory and gustatory perception (P. Rozin, 1982; Fisher & Scott, 1997), i.e. a combination of smell and tastes. The relationship among smell, taste and flavour are shown in a clear structure in Figure 2.1, which is taken from (Vilgis, 2013). I accept the explanation and definition of flavour from them, and use the term flavour as it comprises smell and taste.

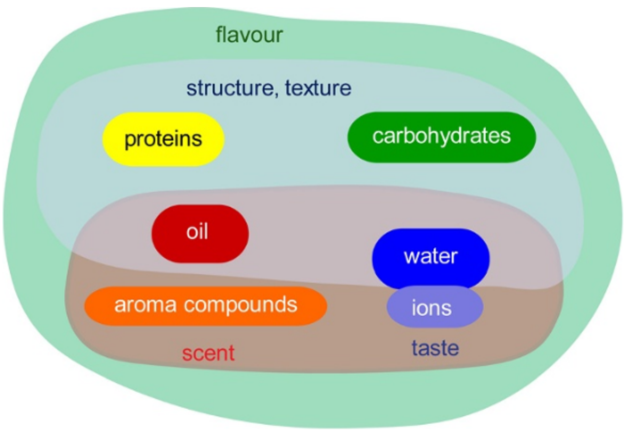


Figure 2.1: Classification of the food constituents. Taken from (Vilgis, 2013).

Flavour of food is acknowledged as the most prominent determinant of food choices (Ahn et al., 2011; Kourouniotis et al., 2016; Liem & Russell, 2019), great contributions have been made to create pleasing aroma and taste. During cooking, chefs compose different kinds of ingredients to make a dish delicious in order to stimulate the consumer's appetite. However, the flavour preference of humans is diverse and complex, and explaining the differences requires a blend of biological and cultural factors. In the culinary world, food pairing theory hypotheses that ingredients sharing flavour compounds are more likely to taste better together than ingredients that do not (Blumenthal, 2008). This hypothesis makes an effort to align with a Kantian aesthetic appreciation, which tries to find a common sense of flavour preference of all humans. Nevertheless, a study from Ahn with online recipes reveals the food pairing theory only seems to hold for western countries (Ahn et al., 2011). Ahn's work is a great attempt to explore food flavour aesthetics based on synthetic knowledge about chemistry and statistics. The research on the flavour aspects in this doctoral work is inspired by this, with the online recipes with intact ingredient lists and reliable flavour molecules database like flavourDB [1], I will investigate the food aesthetics on the flavour aspects across cultures.

[1] `http://cosylab.iiitd.edu.in/flavordb`

Chapter 3

Related Work

3.1 Introduction

This chapter presents an overview of the relevant work and literature in the context of this thesis, i.e., the research into investigation on online food data with digital approaches. The aim is to provide readers with the background knowledge of this domain. This will enable them to understand the research and design decisions in subsequent chapters. The literature review is structured in three sections:

Section 3.2 presents a review of digital food studies for understanding people's online food behaviours. The literature in this section studied digital traces mainly from recipe portals, which contain online recipes (including ingredients lists, recipe images, etc.) and log files of the portals recorded user activities indicating their interactions with digital food objects. These data shed light on human online food choices, and previous work relied on it inspired this doctoral work.

Section 3.3 provides an overview of studies specialising aesthetic aspects, i.e., visual and flavour aspects according to the definition in Chapter 2, of online food. The research summarised in this section focuses not only on how visual, and flavour of food influence users' behaviours on recipe portals, but also on the knowledge, technologies and approaches applied to extract visual and flavour information from corresponding food data, which provide valuable methodological implication for the research in this work.

A review of research into food recommender systems is given in section 3.4. This section focuses on the development of food recommender systems. The aim of this section does not provide a summarisation of all aspects of food recommender systems, such as in (Trattner & Elsweiler, 2017a) and (Min et al., 2019). Instead, it underlines the limitations of developing food recommender systems with standard approaches and discusses the advanced methodologies, such as involving multi-modal food data and incorporating context-related factors in order to boost the performance of the food recommender systems in the previous work.

Finally, the implications of the literature reviewed in this section to this thesis will be elaborated. It explains how the reviewed literature motivates both the research questions addressed in this thesis and the methods applied. On top of that, the research gaps I would like to fill between the reviewed literature and my research goals will be summarised in this section as well.

3.2 Understanding Human Online Food Behaviours

Digital food traces, such as the recipes uploaded on the recipe portals, have been studied in order to reveal the patterns of human online food choices. For example, it is known that, regardless of where in the world the recipes stem, the ingredients follow a power-law distribution, that is, only a small number of ingredients can be found in most recipes (e.g., sugar, salt, egg etc.), while there are a large number of them (e.g., jasmine, Jamaican rum, etc.) in very few recipes (Ahn et al., 2011; Wagner et al., 2014). In addition, the complexity of recipes tend to be similar. It is indicated in the number of ingredients in online recipes, which ranges from 8 to 11 on average (Ahn et al., 2011; Wagner et al., 2014). Recipes with very few ingredients (e.g., n = 1) or a very large number of ingredients (e.g., n = 40) are extremely rare. These are the first hints that commonalities can be found across food cultures. However, such simple analyses do not convey the complexity of people's food behaviours.

Other digital traces documenting user interactions with digital food objects, such as visiting, uploading, rating, bookmarking and reviewing online recipes, have been valuable proxies for understanding human food behaviours. Studies on this data have revealed that users' online food consumption, choices and production (i.e., generating and uploading online recipes) patterns are influenced by numerous factors.

One of the factors can be the individual differences between users. For example, gender. It has been illustrated in (Wagner & Aiello, 2015) and (Rokicki et al., 2016), women prefer desserts, while men tend to consume more meat and beer. The relationship between user hobbies and nutritional intake has also been found (Trattner, Rokicki, & Herder, 2017). People with active hobbies (e.g., biking, hiking and boating), prefer food with lower energy, fat and carbs. In contrast, people with creative hobbies (e.g., knitting, sewing), are in favour of high fat, sugar and carbs. User social networks have an impact on their food choices as well. It was revealed in (Kusmierczyk & Nørvåg, 2016) and (Trattner et al., 2019) that user friendship on the recipe portals can be a useful predictor for their future online recipe production patterns, since users prefer to generate and upload similar recipes with their friends.

In addition to these, users' preferences to ingredients also explains online food choices, as reported by Wagner et al. (2014). The preferences for the ingredients, however, are found to be influenced by several external factors, such as time and geographical location. The food choices, for example, changes over the week. Wagner et al. (2014) found food with meat is chosen more during weekends, while carbohydrate-rich food is more popular at the beginning of the week. In addition, the consumption of certain ingredients show evident seasonal prevalence, such as asparagus, which occurs more frequently in online recipes in spring. This corresponds to the time of the year when asparagus are harvested. During autumn and winter, user needs for carbohydrate and calorie-rich food burst significantly (Wagner et al., 2014).

In terms of the geographical patterns of online food behaviours, a hypothesis was raised is, "regions which are geographically close share similar food preferences". It has been confirmed by Wagner et al. (2014) in the German speaking regions. Specifically, in Germany, Austria and Switzerland, the closer the geographical distance between regions, the more similar the popular patterns of ingredients and recipes. This kind of geographical pattern is more obvious in countries with large geographical differences, such as China (Zhu et al., 2013) and India (Jain et al., 2015). The research from (Sajadmanesh et al., 2017) and (Kim & Chung, 2016) went beyond the investigation within one region or country, the authors collected online recipes of different cuisines and successfully demonstrated the clusters of worldwide culinary cultures. Their findings suggest that the cuisines which reside in geographically close countries share more similar ingredient usage patterns to themselves and thus can be grouped together. For example, cuisines East (e.g. China, Japan, Korea) and South Asia (e.g., Thai, Malaysia) are clustered due to their similar ingredients, which is obviously different from the clusters formed by Western (e.g., German, French, British etc.) and Eastern European (e.g., Russian, Croatian) cuisines (Sajadmanesh et al., 2017).

One aesthetic ideal of food, flavour, was found to vary across food cultures. Such as revealed by Ahn et al. (2011), the food pairing theory, “ingredients sharing chemical flavour compounds are more likely to taste better together than ingredients that do not” (Blumenthal, 2008), was proven to be valid in Western cuisines (e.g., North America) rather than in Eastern cuisines (e.g., Korea). Specifically, in North American recipes, the more flavour compounds are shared by a pair of ingredients, the more likely they are used together in recipes, while in East Asian cuisines, the more flavour compounds two ingredients share, the less likely they appear in a recipe. On the other aesthetic aspects, visual appearance, online recipes display cultural patterns as well.

It was illustrated in (Min et al., 2017), the authors built topic models with ingredients, then retrieved corresponding recipe images for each topic. Based on several topics, for example, topics containing soy-sauce and sesame-oil, and ricotta-cheese and fresh-basil, recipe images from Chinese and Italian cuisines were retrieved respectively, which show distinct differences that can be recognized with naked eyes.

Thus, the research literature underlines the complexity of food behaviours. Online traces provide a lens to study this. Investigation has revealed that various factors, including individual difference, ingredients, and context factors such as time and geography, are involved in driving people’s food behaviours. Among them, geography contributes to explain food choice patterns across cultures, which is particularly evident in terms of ingredient usage. Moreover, the signals of cultural patterns with respect to aesthetic aspects, such as visual appearance and flavour, can also be traced from previous work. Studies of food-related aesthetics will be addressed in detail in the following section.

3.3 Digitality Aesthetic Aspects of Food Studies

This section provides a review of related work where digital food traces sourced online, which have provided insight into how aesthetics of food influenced human food choices. The section summarises work on visual information encoded in recipes before switching the focus to flavour. Research is reviewed that shows how the flavour of online recipes can be modelled and what I have learned from this.

3.3.1 Visual Aspects of Human Food Choices

It is well established in psychology that food choice is visually driven (Leng et al., 2017). The evidence suggests that humans learn the ability to accept food based on certain visual inputs, such as colour (Clydesdale, 1993), at an early age, with their visual food preferences being shaped by cognitive development and living environment (Leng et al., 2017). The evidence from the cognitive experiments moreover shows that simply viewing pictures of food provokes a physiological reaction in the body, which is similar to observing food items directly (Duszka et al., 2020). It makes food images a good medium to understand human food behaviours.

Previous work has revealed that whether online recipes are chosen or preferred is partly explained by their corresponding images with low-level properties extracted from them. These properties, which were proposed by San Pedro and Siersdorfer (2009), including, brightness, sharpness, contrast, etc., have originally shown to work well in explaining attractiveness of photographs on Flickr, but in recent work, (Elsweiler et al., 2017) and (Trattner et al., 2018), these features showed great ability in predicting users' food preferences and online recipe popularity. In addition, the patterns of visually attractive recipe images were disclosed by means of studying the values of these features. For example, the images perceived to be visually appealing and let users would like to choose tend to have higher values of brightness, colourfulness, entropy, and sharpness in general.

These findings suggest food images have the potential to be applied in food recommender systems, one of whose functions is to understand and predict human food preferences. In addition, another advantage taken from food images into food recommender systems is people were reported to rely on them when making judgements (Elsweiler et al., 2017; Trattner & Jannach, 2020). It is because compared to textual descriptions of food, including ingredient lists, cooking instructions, and nutrients, food images contain more intuitive information (Cordeiro et al., 2015), and downgrade people's cognitive load during decision-making (Yang et al., 2015).

However, images might offer misleading cues, such as shown in (Elsweiler et al., 2017), a dish that looks healthy in an image may be very fat-laden in fact. The information that is perceived by people doesn't match the real information the images convey can lead users to make wrong judgments of food, then make food choices that do not meet their expectations.

Recently, some attempts have been made in order to involve more visual properties, besides the low-level visual features mentioned above, to study food images. These visual properties are applied to match the semantic information (ingredients, cuisines, courses, nutrients, etc.) with corresponding food images. This enables the automatic labelling of food images, which will help people make correct judgments of food, thus making food choices that better meet their needs via food images. Applying more advanced computer vision technologies to extract these visual properties from food images have been an issue that is receiving increasing attention recently, literature related to this will be reviewed and how they inspired this doctoral work will be summarised in the following section.

3.3.2 Computer Vision Approaches for Investigating Food Images

Computer vision has been widely applied in studies of food images for over a decade. In general, it is used to extract visual features from food images, then combined with machine learning algorithms for tasks such as classification (e.g., M. Chen et al., 2009; Joutou & Yanai, 2009; Kawano & Yanai, 2014b; Bossard et al., 2014), recognition (e.g., Wang et al., 2015; J. Chen & Ngo, 2016; Ciocca et al., 2017), and retrieval (e.g., Salvador et al., 2017; Min et al., 2016). There are two common sets of visual features derived with computer vision technologies in the food domain, hand-crafted visual features, and the state-of-art Convolutional Neural Network (CNN) embeddings.

The hand-crafted visual features can be regarded as several sets of vectors that encode particular aspects of images, such as colour, texture, shape etc. (Napoletano, 2018; Alshazly et al., 2019). For example, Colour Histogram represents the colours by calculating the distribution of colours in an image, and Local Binary Patterns (LBP) (Ojala et al., 2002), Gabor Texture (Fogel & Sagi, 1989) capture the texture information of an image. These features describe images globally by taking the value of each pixel into calculation. Different from them, another commonly applied visual features, the scale-invariant-feature transform (SIFT) (Lowe, 2004) detect and describe salient patches around properly localised "keypoints" of an image. There are other hand-crafted visual features that have been applied, examples of food image studies with them are shown in Table 3.1. Researchers often applied more than one of these hand-crafted features in their food recognition or classification tasks, such as in (Farinella et al., 2014; Joutou & Yanai, 2009; Matsuda et al., 2012; Matsuda & Yanai, 2012). Different sets of features were compared, or combined to enhance the performance of the models.

Table 3.1: Examples of previous work relating to food recognition with hand-crafted visual features

Reference	Visual Features
Joutou & Yanai (2010)	SIFT, Colour Histogram, Gabor Texture Features
Zong et al. (2010)	SIFT, LBP
Matsuda & Yanai (2012)	SIFT, Histogram of Oriented Gradients (HoG), Gabor Texture Features
Matsuda et al. (2012)	SIFT, HoG, Gabor Texture Features
Ananthimopoulos et al. (2014)	SIFT, Colour Histogram
Kawano & Yanai (2014b)	Colour Histogram, Bag-of-SURF
Farinella et al. (2015)	SIFT, PRICoLBP, Bag of Textons
Zheng et al. (2017)	SIFT, colour patch features

The more advanced CNN embeddings are becoming popular in food image studies in recent work. CNN is an artificial neural network, which has been a dominant method in computer vision tasks due to its prominent performance on the specialised object recognition competition - ImageNet Large Scale Visual Recognition Competition (ILSVRC) (Krizhevsky et al., 2012; Russakovsky et al., 2015). Generally, a CNN architecture comprises three elementary layers, namely convolutional, pooling and fully connected layers (shown in Figure 3.1). This architecture can be tuned by applying different sizes of filter on the convolutional layer, or adding activation functions on convolutional or fully connected layers, etc. Thus various architectures appeared, several widely used CNN architectures include AlexNet (Krizhevsky et al., 2012), VGG (Simonyan & Zisserman, 2014), GoogLeNet (Szegedy et al., 2015) (Szegedy et al., 2014; Szegedy et al., 2015), ResNet (K. He et al., 2016), etc. They have been introduced to food studies especially for food recognition and retrieval since 2015 (Min et al., 2019).

Generally, both hand-crafted visual features and CNN embeddings are tested on benchmark datasets (e.g., Joutou & Yanai, 2009; Matsuda & Yanai, 2012; Meyers et al., 2015), which contain food images manually annotated with food types, ingredients. Visual features are extracted from these images, and input into algorithms, such as Support Vector Machines (SVM), Random Forest (RF), etc., to test to what extent they are able to recognise the annotated food contents. Most of the features perform well in these tasks. For

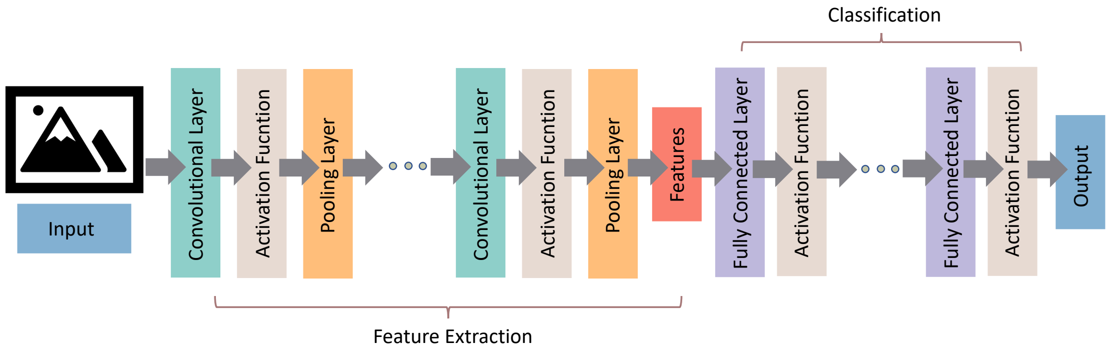


Figure 3.1: A general CNN architecture

example, in (Joutou & Yanai, 2009), the authors trained a SVM classifier incorporating a combination of SIFT, Colour Histogram and Gabor Texture Features. It achieved 61.34% classification rate for 50-category of food. In their subsequent work (Hoashi et al., 2010), a Gradient Histogram was added to the features combination, the accuracy of the classifier was boosted to 62.25% on an expanding dataset, which contains images of 85 food types. CNNs were found to outperform hand-crafted features in food classification tasks (Kawano & Yanai, 2014b; Kagaya et al., 2014). For example, on a Food-101 dataset (i.e., dataset contains 101 categories of food), CNN embeddings provided the best classification accuracy (approx. 57.87%), significantly better than that provided by the other two hand-crafted features (approx. 50.14% and 53.04% respectively) (Kawano & Yanai, 2014b). Similar conclusions were also reported by (Bossard et al., 2014; Wang et al., 2015; Kawano & Yanai, 2014a; J. Chen & Ngo, 2016). In addition, in (Wang et al., 2015) and (Kagaya et al., 2014), the findings also suggest a fusion of CNN embeddings and hand-crafted visual features leads to better classification performance. Besides the food classification tasks, ingredient recognition, nutrient estimation from food images are also popular research issues, which have been implemented by applying computer visions. CNN embeddings are applied for these tasks due to their better performance in general. Compared to these, identifying cuisines and predicting food preferences with these visual features received less attention, only in limited work. For example, only Min et al., (2017) tried to map recipe images with corresponding cuisines via jointly using visual features and textual features (e.g., ingredient, cuisine and courses); and Yang (2015) have designed a food recommender system with a visual interface, which exploited CNN embedding to understand users' visual food preferences.

In summary, visual features extracted by means of computer vision approaches have been proven to perform well on food classification tasks. However, most of these tasks focused on the food types (e.g., Hamburg, hot dog, dessert etc.) and ingredients. Evidence has suggested that food images from different cultures are visually distinct (Min et al., 2017), while only limited work attempted to identify cuisines with visual features. In this doctoral work, the low-level visual features, hand-crafted visual features, and CNN embeddings will all be applied to study the food across cultures on the visual aspects. Moreover, all these will be applied to predict visual food preferences within and across cultures. These will be applied individually, thus comparison of their performance in these tasks can be done, and a combination of these feature sets will also be attempted in order to see whether it can boost the performance.

3.3.3 Flavour Aspects of Human Food Choices

Although human food choices are visually driven, flavour has been stated to be the most prominent determinant of human food choices (Liem & Russell, 2019). It is illustrated in that olfaction provides clues about whether food is edible, and taste helps people determine the nutritional quality of a food (McCrickerd & Forde, 2016), for example, food with sweet flavour are considered to be carb-rich, while salty food is regarded as rich in protein. In addition to these biological factors, cultural-induced flavour bias has also been proven to have an impact on food choices (Clark, 1998). As pointed out by Morley (2012), most of the cuisines across the world are represented by its own "flavour principle", such as soy sauce, ginger and rice wine in China; chilli-pepper with lime or tomato in Mexico. These findings suggest that people flavour preferences indicate their food choices. Traditionally, the food choices induced by flavour aspects were studied in laboratory (e.g., Mennella & Beauchamp, 2005; Inui-Yamamoto et al., 2017) or based on surveys with questionnaires (e.g., Habhab et al., 2009). Recent work has shown a trend of learning flavour patterns with online recipes. This section will discuss how online recipes offer flavour information and which contributions these have made in learning human food choices.

There are a couple of sources and ways to obtain the flavour information from online recipes. One of the most direct is getting it from recipe portals via API. For example, Yummly API provides the values of six flavours, saltiness, sourness, sweetness, bitterness, savoriness, and spiciness. Each flavour is indicated with a score range from 0 to 1 (Sajadmanesh et al., 2017). However, Yummly is, to the knowledge of the author, the only online recipe portal to provide flavour information. An alternative approach is to derive the flavour information from the ingredients. For example, Nag et al. (2019) were inspired by the relationship between nutrients and flavours (van Dongen et al., 2012), and estimated the flavours of recipes through ingredients with specific nutrients. In their work, they measured four flavours. Ingredients containing sodium, carbohydrate, calcium and iron, glutamate and protein were applied to calculate saltiness, sweetness, bitterness, umami respectively. They also measured the richness score for the recipes by considering saturated fats, cholesterol and total fats.

Both of these approaches shown above have generalised flavours of online recipes into several basic tastes. Different from these, another practice applied in (Ahn et al., 2011), has incorporated chemical knowledge. They have mapped ingredients with the corresponding flavour compounds according to Fenaroli's handbook of flavour ingredients (Fenaroli, 2004), so that the quantitative analysis on the flavour aspect can be implemented based on the molecules. A recently released database, FlavorDB (Garg et al., 2018), is another choice for attaching the chemical flavour information to ingredients, which contains more ingredients and flavour information than Fenaroli's handbook. However, it still contains flavour compounds for a limited number of ingredients (n = 936) since the flavour compounds corresponding to a large number of ingredients (e.g., coconut ice cream, caramel topping, etc.) have not yet been discovered. FlavorDB has been applied by Park and colleagues (2021), who attempted to learn the relationship between ingredients and flavour compounds and anticipated to predict flavour information for more ingredients.

Building on the knowledge of food flavours, several patterns of food choices on the flavour aspect have been revealed. In (Nag et al., 2019), it was found that people prefer food that they perceive to be tasty, sometimes the tastiness is decided by a particular flavour. For example, an individual who prefers sweet food would rate food they perceive to be sweet relatively high. However, flavour preferences are personal and could be really diverse. Contextualising individuals to their cultural upbringings facilitates understanding of flavour preferences. Such as revealed by Ahn et al. (2011), that eastern and western culinary cultures share distinct flavour patterns. It suggests culture background can be a valuable factor when investigating human flavour preferences. It is also noted that, in addition to the cultural-related differences, the flavour patterns worldwide also indicate overlaps. Such as shown in (Sajadmanesh et al., 2017), compared to the obvious cultural boundaries in terms of ingredient usage, the flavours across cultures are not so discriminative. For example, cuisines from South Asia and Latin America formed as clusters based on their flavour similarity. This kind of stable patterns should not be overlooked in exploration of human flavour-based food choices.

Researchers have revealed the signals of both differences and overlaps of food flavours across cultures, nevertheless, an issue “to what extent the flavour patterns are distinct or similar across cultures” remains unclear. More quantitative analyses need to be done to answer this. In addition, the findings in terms of flavour patterns are measured with different approaches. Ahn et al. (2011) applied flavour compounds and showed the differences of flavour patterns across cultures, while Sajadmanesh et al., (2017) relied on flavours measured from ingredients and displayed the overlaps. These make the results incomparable. In this work, the approach of applying flavour compounds are preferred and will be utilised to model online recipes. Quantitative analyses include statistical analysis and machine learning algorithms will be applied to investigate the distinction and commonalities of human food choices in terms of flavours across cultures.

3.3.4 Summary

This section focused on how aesthetic aspects can be used to explain online food choices. It is possible to understand food preferences of individuals by studying the visual and flavour attributes of online food objects. Advanced computer vision technologies and knowledge from chemistry have supported the investigation of food aesthetic patterns that are culturally-dependent. This doctoral work will further explore in this direction, with the aim of revealing the aesthetic preferences patterns across cultures. It is important for developing recommender systems, which are facing the challenges of predicting what people would like to eat. Incorporating the information that suggests humans food choices, such as cultural backgrounds and aesthetic ideals might be helpful in this regard. Efforts have been made for improving the performance of food recommender systems in the previous research will be elaborated in the next section.

3.4 Studies on Food Recommender Systems

In section 3.2, the work for understanding human food behaviours by means of investigating digital traces of online recipe portals has been reviewed. Some of the works, as claimed by their authors (e.g., Harvey et al., 2013; Rokicki et al., 2018; Kusmierczyk et al., 2015b; Trattner, Rokicki, & Herder, 2017), were with the aim of applying their findings to food recommender systems. This section gives an overview of research to the development of food recommender systems. I discuss the attempts from previous researchers for developing food recommender systems, including that applying standard approaches, and that incorporating the results from quantitative analyses on digital food traces, then I relate them to the studies in this work.

3.4.1 Early Developments of Food Recommender Systems

Humans are reported to make over 200 food choices per day (Palojoki & Tuomi-Gröhn, 2001). However, making food choices is still challenging for most people due to the large number of food options available, especially with the proliferation of food-related contents online. Evidence suggests that online food objects (e.g., online recipes, posts relating to food on social media) is relatively unhealthy (Trattner & Elsweiler, 2017b), and people are found to have a tendency to be lured by tempting posts (e.g., with visually appealing images) to choose fatty or calorie-laden foods (Mejova et al., 2015; Holmberg et al., 2016). The consequences of making such choices reflect in reality, such as reported in (Said & Bellogín, 2014) and (De Choudhury et al., 2016). People in regions that suffer from poor health (e.g., a large percentage of obesity) have more frequent (e.g., more visiting and sharing) and positive (e.g., higher rating) interaction with unhealthy food-related contents online. Food recommender systems have been considered to be a good means to assist people nourish themselves more healthily (Elsweiler et al., 2015; Ge et al., 2015; Elsweiler et al., 2017; A. D. Starke et al., 2021), whereas with an important prerequisite of recommending food people would like to eat, i.e., predicting human food preferences.

In order to develop food recommender systems, standard approaches, i.e., Content-Based (CB) Methods and Collaborative Filtering-Based (CF) Methods, which have shown great power in recommending movies (Salter & Antonopoulos, 2006; Reddy et al., 2019), music (Joyce, 2006), artworks (Messina et al., 2019; Aroyo et al., 2007) and other leisure and entertainment objects, have been introduced in the food domain. CB method tries to learn the contents of the items that users have rated positively then predict if other unrated items would be preferred by them. For food recommendation, ingredients have been the most commonly studied content. Previous work (e.g., Wagner et al., 2014) has suggested that ingredients play an important role in explaining food choices. A CB method food recommendation implementation was shown in (Freyne & Berkovsky, 2010; Freyne et al., 2011). In these work, the ratings on recipes were transformed to ratings on individual ingredients, which were applied to predict ratings of further recipes containing these ingredients. An example given by Harvey et al. (2013) can explain how this approach works, that is, assuming tomato occurred in a recipe that a user given high ratings, then other recipes containing tomatoes were predicted to be preferred by this user.

The CF method predicts a target user rating via the ratings given by the similar users. The similar users were normally determined by applying correlation coefficient on their rating matrix, such as Pearson. CF performed slightly poorer than CB in the food recommendation domain, such as in (Freyne & Berkovsky, 2010; Harvey et al., 2013). However, its ideas inspired new recommendation strategies. For example, a popular technique of CF, matrix factorization, has also been applied and optimised by Ge et al. (2015). They involved not only the rating matrix, but also the users supplied tags into their food recommender systems, which outperformed the baseline models built by standard CB and CF methods. In addition, in some cases, a fusion of CB and CF, such as shown in (Forbes & Zhu, 2011), have enhanced the performance of recommender systems.

Both CB and CF methods have been attempted to develop food recommender systems.

However, as pointed out by Trattner and Elsweiler (2017a), compared to recommending items in other domains, standard approaches showed only limited power in food recommendation. This results from the complexity and diversity of food contents and human food choices. They also suggested future directions integrating new technologies and relevant knowledge for improving food recommender systems. For example, food contents nowadays are not limited to ingredients, multi-modalities of data such as images, smells, tastes, should be explored and exploited into food recommender systems (Trattner & Elsweiler, 2017a). In addition, it is known that human food choices are multi-faceted, and context-dependent, as discussed in Section 3.2. Thus it is necessary to incorporate context variables into recommendation models. The previous research to these directions will be reviewed in the following sections.

3.4.2 Investigation on Food Content for Food Recommender Systems

This section reviews work of investigation on food content for improving the performance of food recommender systems in two directions. The first one focuses on ingredients of online food, and the second one focuses on other aspects of food contents.

Ingredients are important information for understanding the content of food. In some work (e.g., Harvey et al., 2013), the authors refer to recipes as documents, and ingredients as words, recommending food is analogised as an information retrieval (IR) task, i.e., retrieving the preferred recipes for the users. Thus the advanced technologies from IR domain have been applied to study them. For example, TF-IDF (Salton & Buckley, 1988) were applied to represent recipes with weighed ingredients in (El-Dosuky et al., 2012). It transformed each recipe as a multidimensional vector, allowing the similarity of recipes to be measured, in order to determine the top recipes that are similar to recipes a target user preferred. State-of-art word embedding technologies, Word2vec (Mikolov et al., 2013) and BERT (Devlin et al., 2018), have also been applied in recent work (Pellegrini et al., 2021). These approaches can not only represent recipes, but also capture the semantic relationship between ingredients and determine substitution for a target ingredient by localising the nearest neighbours in the embedding space. It was proven to be efficient in replacing ingredients which cause user allergens with another one, and it is potential to be used to replace the ingredient that users dislike in the recipes with that they like. Other than investigating individual ingredients, a graph-based strategy has been applied by Teng and colleagues (2012). They built a network based on co-occurrence of ingredients, and represented recipes with graph representation of ingredients. This was found to be very powerful in predicting users' food preferences.

In addition to ingredients, there are other data that indicates food contents, such as dish type (e.g., carpaccio, spaghetti etc.) and categories (e.g., fast food, snack, dessert), food recommender systems involving this information were found to perform better than that utilised ingredients only (Freyne et al., 2011). Another important aspect of food content, which brings health into the food recommender systems, is nutritional information (e.g., energy, fat, carbohydrate etc.). It can be measured from ingredients such as in (Müller et al., 2012) or obtained from online recipe portals (e.g., allrecipes.com, kochbar.de). In some work, the nutritional information was applied to estimate the healthiness of recipes according to authoritative guidelines, such as WHO and FSA (e.g., Trattner, Elsweiler, & Howard, 2017; Elsweiler et al., 2017). Incorporating food content on this aspect into food recommender systems has been mooted as a potential solution for encouraging healthier food choices.

A rising theme for developing food recommender systems is applying food visual content encoded in recipe images. As discussed in Section 3.3, food choices are visually-driven, and more digital approaches have been developed, allowing more visual features from recipe images to be extracted for investigating human visual food preferences. A promising example of using visual features for recommending food is shown in (Elsweiler et al., 2017). In a series of pairwise food preferences prediction tasks, the visual features not only provided prominent and consistent performance, but also showed capability of promoting users towards choosing healthier food. Moreover, Yang (2015, 2017) suggested a visual based interface is more applicable to food recommender systems, and their results showed the great power of CNN for automatic visual feature learning. On top of that, food content in terms of flavour has also been incorporated into the content-based food recommender system in (Nag et al., 2019), which was found to outperform that without flavour information. Applying advanced approaches for understanding and exploiting food ingredients and other aspects of food contents have not merely enhanced the performance of food recommender systems, but also provided new perspectives for food recommendation. For example, the aesthetic aspects of food content. Their promising performance motivated this doctoral work to attempt to develop a food recommender system incorporating visual and flavour information.

3.4.3 Incorporating Context Information into Food Recommender Systems

Recommending food based solely on food content is not sufficient, since it is not the factor that can fully explain human food choices. Challenges of predicting what people would like to eat are brought by diverse external factors, i.e., the context. In terms of these, there have been numerous exploratory data analyses, as discussed in Section 3.2. In this section, work relating to how these context variables were modelled and integrated into food recommender systems is reviewed.

Context-related factors are often inputted into algorithms in order to test their power in predicting how recipes were rated (Freyne et al., 2011), bookmarked (Trattner & Elsweiler, 2017b) or commented (Rokicki et al., 2016, 2018; Trattner et al., 2018). For example, regression models in (Harvey et al., 2012) and (Rokicki et al., 2016, 2018) have indicated that popularity of recipes can be successfully predicted by means of external biases such as time (e.g., the time of a day, or month or season), social networks (e.g., friendship on recipe portals), gender and online recipe editorial features. In addition to these, more sophisticated methods, such as machine learning approaches have also involved in understanding the relationship between context features and user food choices. Several works limited their algorithm effort to approaches with only one factor. Such as in (Rokicki et al., 2016), gender was specifically studied. The authors attempted to build a gender-aware food recommender system based on their findings from the quantitative analyses relating to differences in food preferences between male and female. According to their offline evaluation, the gender-aware food recommender system has proven to improve over the baselines. Besides, in (Cheng et al., 2017), the city size (e.g., metropolis, big-city, medium-city, small-city and town) outperformed temporal factors in predicting recipe ratings, and was found to boost the performance of food recommender systems.

These findings suggest that incorporating appropriate context factors into food recommender systems enhances their performance. With the aim of developing recommender systems based on aesthetic aspects of food (i.e., visual appearance and flavour) in this work, an important context factor needs not to be overlooked, that is, culture. Such as discussed in section 3.2, culture plays an important role in influencing human food choices, and also shaped the food aesthetics (e.g., Min et al., 2017). Therefore, when incorporating aesthetic features into food recommendation, culture needs to be taken into account. This work will place food aesthetic preferences prediction within and across cultures, not only for inspecting the differences of food aesthetic preferences brought by culture, but also attempting to ascertain stable patterns in aesthetic food preferences across cultures. This is in order to develop a recommender system for providing aesthetically acceptable food recommendations for people with different cultural backgrounds.

3.4.4 Summary

This section shows various promising examples of promoting food recommender systems, whose aim is to recommend satisfying food for users. Standard recommender algorithms have been tried and tailored to be more suitable for improving performance of food recommender systems, such as more in-depth studies on food content and the incorporation of food content on other aspects. In addition, findings from quantitative analyses on digital food traces, such as how context-related factors influence human food choices have also been applied to food recommender system developments. This section suggests that food content on aesthetic aspects would improve performance of food recommender systems, when culturally relevant food aesthetic ideals are considered.

3.5 Chapter Summary

Literature reviewed in section 3.2 and 3.3 showed human online food behaviours influenced by various factors. The aesthetic aspects of food, i.e., visual appearance and flavours, as the determinants of human food choices in the physical world, have played an important role in influencing online food choices as well. Evidence suggests their impact is culturally dependent. However, cultural influence on food aesthetics and how it affects human food choices across cultures have not been studied specifically. For example, the aesthetic patterns of food across cultures have been visualised and elaborated (e.g., Min et al., 2017), yet no work has been done to address the problem in terms of *to what extent the food is different aesthetically across cultures*. In addition, literature has tended to emphasise differences between food cultures and idiosyncraticities (e.g., Ahn et al., 2011; Kim & Chung, 2016). However, there is also evidence suggesting commonalities of food across cultures (e.g., the power-law distribution of ingredient usage, stable recipe complexity and overlaps of food flavours). The issues particularly focus on commonalities in cross-cultural food preferences have not been investigated yet.

This exploration of cultural-related aesthetic ideals in the food domain will be helpful for improving the performance of food recommender systems, as discussed in section 3.4. The aim of this doctoral work is to explore the possibility of developing a cultural-aware food recommender system incorporating human visual and flavour food preferences. I focus on revealing commonalities in aesthetic food preference across cultures, in order to target users from different, rather than specific, cultural backgrounds.

Empirical experiments in subsequent chapters will apply the techniques from computer vision and chemical knowledge for investigating food content on visual and flavour aspects, respectively. User interaction data indicating their food choices will be incorporated in order to inspect their food aesthetic preferences. Three cultures (i.e., Chinese, US and German) are selected for this work, all experiments will be conducted in each culture before food aesthetic preferences are studied and discussed in a cross-cultural context.

Chapter 4

Data Preparation

4.1 Introduction

The literature reviewed in the previous chapter provides insights into how the online food choices people make are influenced by a variety of factors, particularly by culture and food aesthetics. Moreover, the evidence suggests that it would be beneficial to incorporate these two factors into food recommender systems to improve their performance. In the following chapters, the empirical experiments are designed to investigate whether is it possible to utilise food aesthetics, namely, the visual appearance and flavour of food, for the development of food recommender systems when taking culture into consideration. In order to build the data basis for the experiments, in this chapter, appropriate online food data is collected from three distinct food cultures, China, The United States (US) and Germany. In the following sections in this chapter, I describe and illustrate the data sources, and the necessary steps for processing the data such that they can studied appropriately.

This chapter is structured as follows: In Section 4.2, I describe the sources of the data and provide clarification for the choices. Section 4.3 illustrates the process of selecting and filtering the data for the experiments. In Section 4.4, I describe how online recipes are represented with multi-dimensional vectors so that they can be utilised in machine learning models designed to address the issues raised in Chapter 1.

4.2 Data Sources

In order to construct the data basis for the food studies across cultures, three large and well-known recipe portals from different cultural backgrounds were selected as the data sources: *Xiachufang* (`www.xiachufang.com`) in China, *Allrecipes* (`www.allrecipes.com`) in US and *Kochbar* (`www.kochbar.de`) in Germany. The reasons for determining these three recipe portals for this work are as follows:

First, all three recipe portals are popular in their countries. According to Similarweb[1], they are ranked as at least the national Top-5 websites in the category of *Cooking and Recipes* and have a significant number of visits per month. Table 4.1 displays the statistics about their rank and traffic. This ensures the large-scale recipes with corresponding data, including the ingredient lists, recipe images, as well as the data that records user

[1] `https://www.similarweb.com/`

interactions with the recipes can be obtained from the three portals. Examples of online recipes and interaction data from users of the recipe portals are given in Figure 4.1 - 4.3.

Table 4.1: The rank and the number of monthly visits of the three recipe portals

Recipe portal	Country	Domestic ranking	Domestic Category ranking	Average Monthly Visit
Xiachufang	China	699	2	3.7M
Allrecipes	US	187	1	69.7M
Kochbar	Germany	415	5	7.7M

Note:
The data was obtained in May 2022

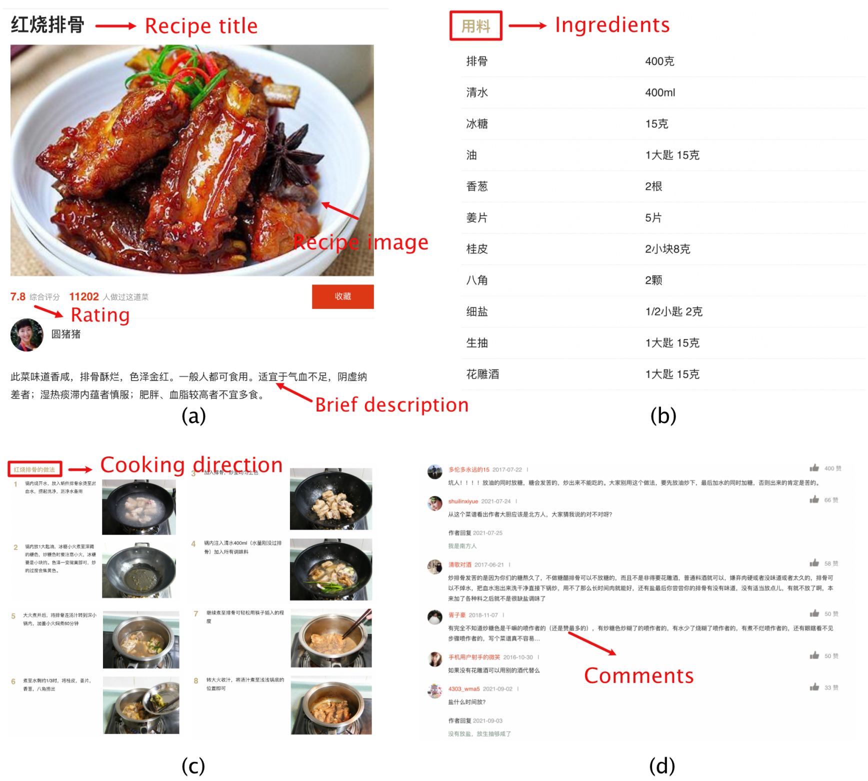


Figure 4.1: Online recipes and interaction data from *Xiachufang*. (a) the top of the recipe detail view with the title, image, brief description and rating; (b) ingredient list of the recipes; (c) cooking direction of the recipes; (d) comments left by the users.

Second, these recipe portals encourage users to upload and share their own original recipes without professional audit. They provide submitting interfaces (an example is shown in Figure 4.2 (f)), which allow the users to share recipes as they would like to. In comparison to other kinds of recipe portals, such as *Epicurious* (`https://www.epicurious.com/`) and *Yummly* (`https://www.yummly.com/`), which tend to display and promote aspirational dishes relying on gourmet editors or professional chefs, the recipes sourced from the portals I have chosen are more likely to illustrate what people

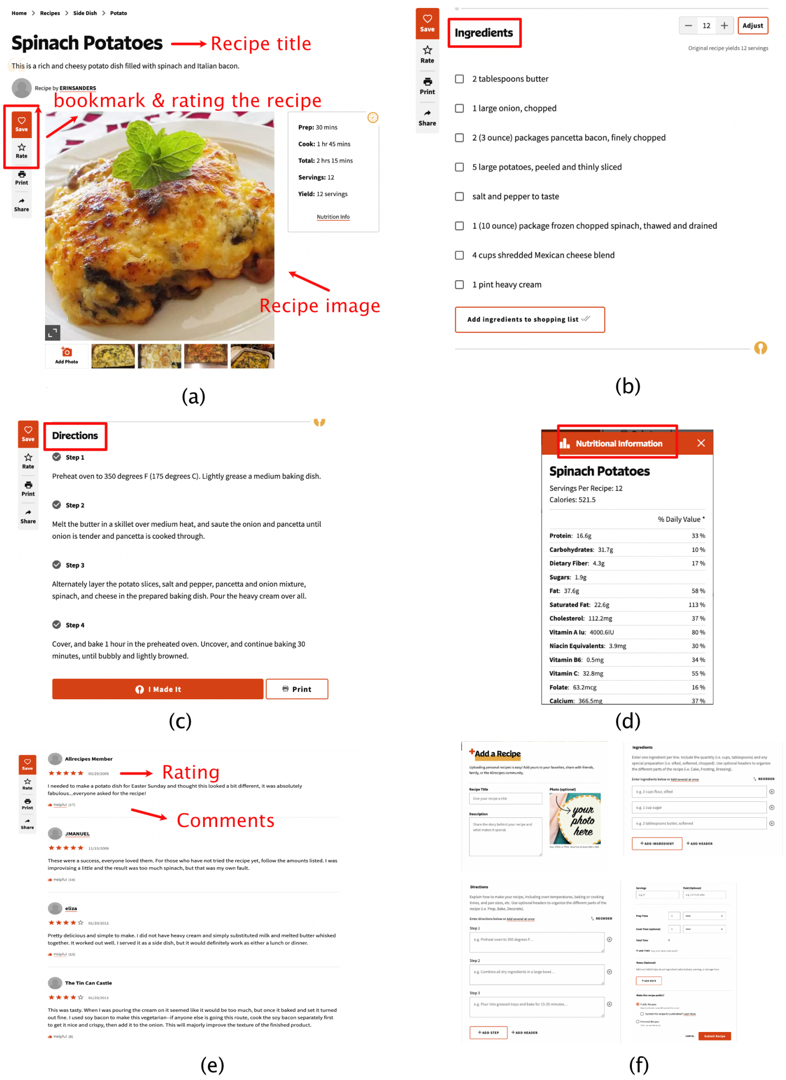


Figure 4.2: Online recipes and interaction data from *Allrecipes*. (a) the top of the recipe detail view with the title, image, rating and bookmark icon; (b) ingredient list of the recipes; (c) cooking direction of the recipes; (d) nutritional information of the recipe; (e) the rating and comments left by the users; (f) the submitting interface.

actually cook and eat in their everyday life and provide a wide range of more authentic culinary and diet information.

In addition, *Xiachufang*, *Allrecipes* and *Kochbar* serve in Chinese, English and German respectively, making the vast majority of their users are native speakers or settle in the corresponding countries. It is a good signal that the recipes derived from these portals are culturally representative. Furthermore, the three recipe portals have been proven to be reliable data sources of online food studies in previous work. Such as the data of recipes from *Xiachufang* have been applied to build data basis for (J. Chen & Ngo, 2016), and *Allrecipe* and *Kochbar* have been data sources for a series of works from Trattner and colleagues (e.g., Trattner, Elsweiler, & Howard, 2017; Trattner & Elsweiler, 2017b; Elsweiler et al., 2017; Trattner et al., 2018, 2019).

In this work, the online recipes with the corresponding data, such as recipe images

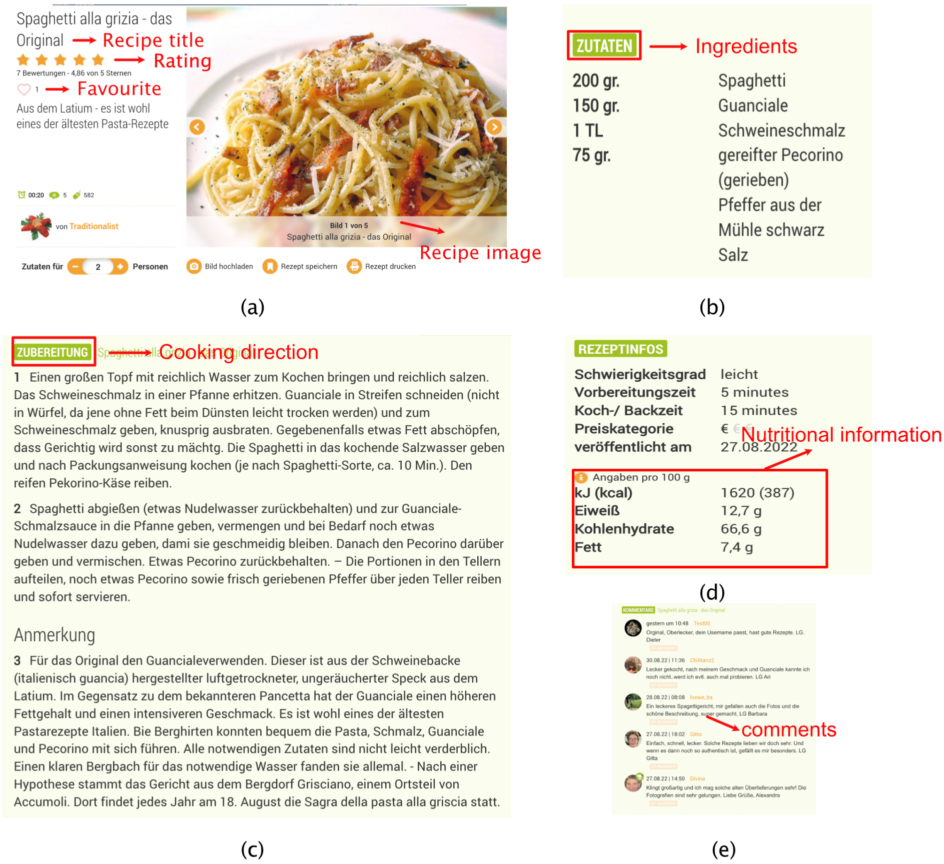


Figure 4.3: Online recipes and interaction data from *Kochbar*. (a) the top of the recipe detail view with the title, image, rating and favourites; (b) ingredient list of the recipes; (c) cooking direction of the recipes; (d) nutritional information of the recipe; (e) the comments left by the users.

and ingredient lists, and the data records user interaction with the recipes were obtained in different ways. In order to get the data from *Xiachufang*, a web crawler was built. During the period 22 - 26 October 2018, it scraped 25,597 recipes from the portal. While the data of *Allrecipes* and *Kochbar* were obtained from the previous work. The *Allrecipes* collection was crawled by Trattner and Elsweiler (2017b) between 20th and 24th of July 2015, it contains 60,983 recipes published between the years 2000 and 2015 on the website. And in 2014, Kusmierczyk and colleagues (2015b) crawled the data from *Kochbar* and they built a dataset that includes over 400,000 recipes written in German uploaded between the years 2008 and 2014.

4.3 Data Selection

According to the origins of the data, I named the collections as *Xiachufang* collection, *Allrecipes* collection and *Kochbar* collection, respectively. In order to apply the data in the collections to the empirical experiments in this work, they needed to be filtered and sampled. The process of it includes: (1) selecting appropriate data to represent food on the aesthetic aspects, i.e., visual appearance and flavour, (2) choosing the interaction data given by users to use as a proxy for their recipe preferences, and (3) ensuring the number of recipes in each collection is balanced. This is because the data in these collections will be applied in several classification and prediction tasks designed to answer the questions set out in Chapter 1. Imbalanced data might lead to biased results.

Selecting Data for Representing Food on the Aesthetic Aspects In this work, I focus on the visual appearance and flavour of food across cultures. The data derived from the online recipe portals described above were selected to represent food on these two aspects. Firstly, I used recipe images uploaded by the users to represent the visual appearance of food. On the recipe portals, it is normal to find that users share more than one image when uploading the recipes, leading to one recipe being associated with multiple images. To deal with this, only the first (default) image was extracted for each recipe.

In terms of representing the flavour of food, I chose the flavour compounds embedded in the ingredients of the online recipes. Therefore, the ingredient lists of recipes were kept in the collections. The ingredients were pre-processed and then mapped to their corresponding flavour compounds according to FlavorDB. The process will be presented in Section 4.4 in detail.

Selecting Data to Indicate the Food Preferences of Users A crucial task of this doctoral work is investigating the food preferences of people across cultures, thus it is important to select appropriate metrics for indicating the extent to which the recipes are preferred. In previous work, it has been common to represent the preference for an online recipe by its number of visits (Wagner et al., 2014), ratings (Rokicki et al., 2017, 2018; Trattner et al., 2018), bookmarks (Trattner, Rokicki, & Herder, 2017; Rokicki et al., 2018), sentiment of received comments (Rokicki et al., 2017; Trattner et al., 2018), etc. These metrics were assumed as indication of the popularity of online recipes. In this work, the preference is treated as equivalent to appreciation, which is represented by metrics that show how users appreciated the recipes, such as ratings. I stored the ratings (ranges from 0 to 10) applied by users of *Xiachufang* since this is the only option available on the site. In addition, the ratings of recipes are normally distributed (shown in Figure 4.4 (a)), suggesting it indicates user preferences to online recipe reasonably.

On *Allrecipes* and *Kochbar*, more metrics are available, including ratings information. However, since the ratings are highly skewed with a low standard error, as described in (Trattner et al., 2018), the number of a recipe had bookmarked (*Allrecipes*) or favourite (*Kochbar*) was used instead. For the recipes from *Allrecipes*, the number of bookmarks a recipe has received within a day, a week, a month and a year were captured. As shown in (Mößlang, 2017), the recipes received the most bookmarks on the day they were uploaded, then it decayed as one can expect, and at the approximate day 10 and after, the number of bookmarks barely increased. Considering the fluctuation of the numbers during the period thereafter, the number of bookmarks within one year (ranges from 1 to 2605) was selected as the metric indicating the appreciation of recipes from *Allrecipes*. The distribution of it is shown in Figure 4.4 (b). The time span of how many times of the recipes have been marked as favourites in *Kochbar* was not captured, therefore, I normalised this metric by means of dividing it by the times the recipes has been viewed by the users. The metric can be interpreted as, how many times a recipe was marked as a favourite after it has been viewed *n* times. I kept the normalised favourite (ranges from 0 to 0.27) as the appreciation metric for recipes in *Kochbar* collection. Its distribution is displayed in Figure 4.4 (c).

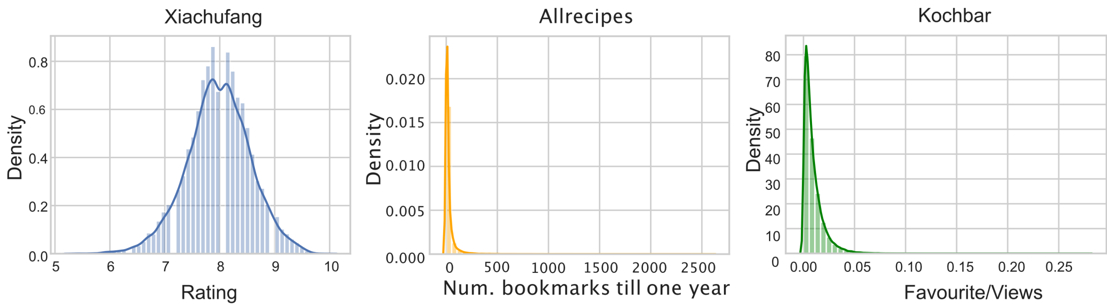


Figure 4.4: The distribution of appreciation metrics of recipes in (a) *Xiachufang*, (b) *Allrecipes*, (c) *Kochbar*.

Balancing the Collections for the Classification Tasks After determining the preference metrics, and deleting the recipes without images and ingredient lists, there are finally 25,508 recipes in *Xiachufang* collection, 35,501 in *Allrecipes* collection and 72,899 in *Kochbar* collection respectively. In this work, in order to address the issues raised in Chapter 1, the empirical experiments are established as several classification and prediction tasks, which require balanced classes.

First, a multi-class prediction task was formulated to answer *Issue 1*, which investigates *To what extent is it possible to differentiate the food across cultures based on the representation of the food relating to visual appearance or flavour?* In order to build the basis for this task, 25,000 recipes were sampled from each collection based on the number of data points in the smallest class, *Xiachufang* recipes.

Second, the *Issue 2*, which studies *To what extent is it possible to identify the differences and ascertain stable patterns of food preferences across cultures based on the representations of the food relating to visual appearance or flavour?* is addressed by means of several two-class prediction experiments. These prediction tasks were conducted to classify the preferred and less preferred recipes within and across cultures. In order to build the data basis for the tasks, I first determined the appreciated and less appreciated recipes from each recipe collection. The cut-off for the appreciation metric was set as 10% and 90% to identify the recipes that best represent user preferences. To be specific, the top-10% and bottom-10% of recipes based on the appreciation metrics of each collection were selected to be appreciation and less appreciated recipes, respectively. The range of the appreciation metrics of the appreciated and less appreciated recipes in each collection is shown in Table 4.2. Moreover, in order to ensure the balance of the classes, 2,500 from appreciated recipes and 2,500 from less appreciated recipes were sampled from each collection. This is nearly all the *Xiachufang* recipes in this percentile and an undersample of the other two collections (shown in Table 4.2), which are larger, with the aim being to draw a fair comparison.

In the end, the recipe samples that represent the food cultures of China, US and Germany (n = 25,000 from each recipe portal), as well as the preferred and less preferred recipes of each food culture (n = 2,500 with top-10% & bottom-10% appreciation metrics from each recipe portal) were selected. These form the basis of all the experiments using naturalistic data. Each recipe was stored with its corresponding ingredient lists, images, and appreciation metrics. The data pre-processing, data representation of the recipes relating to visual appearance and flavours will be done on these samples in the next steps.

The process will be described in detail in the next section.

Table 4.2: The range of the appreciation metrics of top-10% and bottom-10% recipes in each collection and the number of appreciated and less appreciated recipes in each collection

	Top-10%		Bottom-10%	
	Range of appreciation metric	Num. Recipes	Range of appreciation metric	Num. Recipes
Xiachufang	[8.7, 9.8]	2,925	[5.4, 7.2]	2,594
Allrecipes	[63.0, 2605.0]	3,542	[1,1]	3,702
Kochbar	[2.1e-2, 2.8e-1]	7,296	[0, 1.3e-3]	7,292

4.4 Data Pre-Processing & Representation

This section illustrates the process of pre-processing the data and representing it with multi-dimensional vectors, so that it can be inputted into machine learning models and for further quantitative analyses. In this work, food images were selected to represent the visual appearance of food. In order to transform the images into the format that algorithms can read and process, the visual information encoded in these were extracted by means of Computer Vision approaches, which I have applied in my previously published work (Q. Zhang et al., 2020). Section 4.4.1 describes the extraction of the visual features from the recipe images in detail. Flavour compounds of ingredients in the online recipes were selected to represent flavour of food. In the recipe collections, the flavour information of each recipe was stored as a list of flavour compounds, which was then transformed into a vector by means of two Natural Language Processing (NLP) techniques. Section 4.4.3 presents the process. However, before the transformation, the data needed to be cleaned. In section 4.4.2, I clarify the reasons for data-cleaning and explain the process clearly.

4.4.1 Food Representation with the Visual Properties Encoded in Recipe Images

Each food image in these recipe collections was represented as a multi-dimensional vector by extracting 5,144 visual features from each image. The idea was to generate as many features as possible that may capture information from the recipe images. These features, described in detail below, include the low-level explicit visual features (EVF), hand-crafted features such as colour histogram, local binary patterns (LBP), Bag of visual words based on descriptors from the scale-invariant feature transform algorithm (SIFT), and the deep neural network image embeddings (DNN), VGG16. The following subsections explain the above listed feature sets in detail.

Explicit Visual Features (EVF) The term of "Explicit Visual Features (EVF)" was first proposed in a content-based artwork recommendation work of Messina et al. (2019), yet the term refers to the visual features brightness, sharpness, contrast, colourfulness, saturation and naturalness, which were applied in earlier work (San Pedro & Siersdorfer, 2009) in order to capture the attractiveness of images from Flickr. These features subsequently have been used in (Trattner et al., 2018) where the set was expanded by adding

RGB contrast, sharpness variation, saturation variation and entropy. The expanded EVF set is used in this doctoral work. In the food domain, EVF has been proven to work well in predicting the preferences of food images from the online recipe portals (Elsweiler et al., 2017; Trattner et al., 2018). In this work, the EVF was measured with the freely available OpenIMAJ Jave Framework [2] in version 1.3.8. OpenIMAJ was developed by the University of Southampton. It contains a set of libraries and tools for multimedia content analysis, including the state-of-art computer vision. Below I display the details about each individual EVF and how it was calculated in OpenIMAJ.

- *Brightness.* Brightness means the luminance intensity of an image. It is extracted with the *AvgBrightness* class with the default NTSC weighting scheme and no mask in this work. It uses a standard luminance algorithm

$$\begin{aligned} brightness &= \tfrac{1}{N}\sum_{x,y} Y_{xy},\ with \\ Y_{xy} &= (0.299 \times R_{xy} + 0.587 \times G_{xy} + 0.114 \times B_{xy}), \end{aligned} \tag{4.1}$$

  where Y_{xy} denotes the luminance value and N is the amount of pixels in an image. R_{xy}, G_{xy}, and B_{xy} are the three RGB colour space channels of pixel (x,y).

- *Sharpness.* This is a highly subjective concept that measures the clarity and level of detail of an image. It is related to the brightness contrast of edges in an image. The algorithm utilises the image's Laplacian, divided by the locale average luminance (μ_{xy}) around pixel (x,y), the formula in *Sharpness* class in OpenIMAJ is as:

$$sharpness = \sum_{x,y} \frac{L(x,y)}{\mu_(x,y)},\ with\, L(x,y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \tag{4.2}$$

- *Contrast.* Contrast is the relative difference in brightness or colour of local features in an image. There are several algorithms for measuring contrast of images such as Weber and Michelson (Pelli & Bex, 2013), but for the means of image comparability, the root-mean-square contrast (RMS-contrast) is often used (San Pedro & Siersdorfer, 2009). The RMS-contrast is calculated as follows:

$$contrast = \frac{1}{N}\sum_{x,y}(I_{x,y} - \bar{I}) \tag{4.3}$$

  Where I_{xy} is the pixel intensity, I represent the arithmetic mean of pixel intensity and N is the number of pixels in the image. I measured contrast with the *RMSContrast* class in OpenIMAJ.

- *Colourfulness.* Colourfulness is the "attribute of visual perception according to which the perceived colour of an area appears to be more or less chromatic". I applied the approach proposed by (Hasler & Suesstrunk, 2003). Firstly, the image needs to be transferred into sRGB colour space using rg_{xy} = R_{xy} - G_{xy} and yb_{xy} =1/2(R_{xy} + G_{xy}) - B_{xy} and secondly, the formula for calculating colourfulness is

$$\begin{aligned} colourfulness &= \sigma_{rgyb} + 0.3 \cdot \mu_{rgyb},\ with \\ \sigma_{rgyb} &= \sqrt{\sigma_{rg}^2 + \sigma_{yb}^2},\ \mu_{rgyb} = \sqrt{\sigma_{rg}^2 + \mu_{yb}^2} \end{aligned} \tag{4.4}$$

[2] http://openimaj.org/

where R_{xy}, G_{xy} and B_{xy} are the colour channels of the pixels and are the standard deviation, respectively the arithmetic mean. The colourfulness of the recipe images was measured with the *ColourfulnessExtractor* class in OpenIMAJ.

- *Entropy.* Entropy was proposed and developed by Claude Shannon in (1948). In information theory, it is known as a measure of randomness of the amount of information content provided by a source, which could be an image. The first step for measuring entropy for an image is to convert it to grey scale, where each pixel has only an intensity value. Then I count the occurrences of each distinct value. The formula we apply in this work is

$$entropy = -\sum_{x \in [0.225]} \cdot \log_2(p_x) \tag{4.5}$$

where p_x is the probability of finding the grey-scale value x among all the pixels in the image.

- *RGB (Red, Green, Blue) Contrast.* This is a measurement that is almost identical to the basic contrast, as explained above. However, it is extended to the three-dimensional RGB colour space, calculated by class *RGBMSContrast* in OpenIMAJ.

- *Sharpness variation.* This is similar to the saturation variation, sharpness variation is calculated via the standard deviation of all pixel sharpness values.

- *Saturation.* Saturation is related to the proportion of the colourfulness of an area to its brightness. It is used to measure how vivid an image is in terms of colour. In the HSV colour space the saturation estimation can be calculated via the RGB approximation of

$$\begin{gathered} saturation = \tfrac{1}{N}\textstyle\sum_{x,y} S_{xy},\ with \\ S_{xy} = max(R_{xy}, G_{xy}, B_{xy}) - min(R_{xy}, G_{xy}, B_{xy}) \end{gathered} \tag{4.6}$$

where N is the amount of pixels in an image and R, G and B are the coordinates of the colour of the pixel in sRGB space. This formula can be found in the *Saturation* class in OpenIMAJ.

- *Saturation variation.* This measure estimates the variation in saturation of an image using the RGB approximation of avg(max(R,G,B) - (min(R,G,B)), defined in the *SaturationVariation* class in OpenIMAJ.

$$saturation\_variation = \sqrt{\frac{\sum_{x,y}(S_{xy} - \bar{S})^2}{N-1}} \tag{4.7}$$

Where N is the number of pixels and S is the list of saturation.

- *Naturalness.* Naturalness measures the degree of differences or similarity between images and human perception, with respect to colourfulness and dynamic range. I followed the method from (Huang et al., 2006) and (San Pedro & Siersdorfer, 2009), which is as follows: firstly, considering the measured image in a colour space, HSL by default. Then picking the pixels with $20 \leq L \leq 80$ and $S > 0.1$, grouping them into of the three sets 'Skin', 'Grass' or 'Sky', the average saturation value of the group (μ_s) is used:

$$
\begin{aligned}
N_{Skin} &= e^{-0.5(\frac{\mu_S^{Skin}-0.76}{0.52})^2},\ if\ 25 \leq hue \leq 70 \\
N_{Grass} &= e^{-0.5(\frac{\mu_S^{Grass}-0.81}{0.53})^2},\ if\ 95 \leq hue \leq 135 \\
N_{Sky} &= e^{-0.5(\frac{\mu_S^{Sky}-0.43}{0.22})^2},\ if\ 185 \leq hue \leq 260
\end{aligned}
\tag{4.8}
$$

The final naturalness is calculated by the formula:

$$naturalness = \sum_i w_i N_i,\ i \in \{'Skin', 'Grass', 'Sky'\}, \tag{4.9}$$

where w represents the fraction of pixels of the specific group in the whole image. The naturalness ranges from 0 to 1, the higher, the more natural. OpenIMAJ offers the class *Naturalness*.

In addition to EVF, I also extracted three hand-crafted visual features, including Colour Histogram that describes colour distribution of images in a specific colour space, Local Binary Patterns (LBP) that are used to capture the texture features, and the Bag of Visual Words (BoVW) based on Scale-invariant feature transform (SIFT), which aims to describe an image with its keypoints. These features have been commonly applied in food recognition tasks in previous work reviewed in Chapter 3. The process of extracting these features from images is shown in detail below:

Colour Histogram Colour histogram describes the global distribution of colour in the image. Given an image in Red, Green, Blue (RGB) colour space, in this doctoral work, the value of each pixel ranges from 0 to 255, which can be split into several bins. The number of bins in this work is 8, thus the range of each bin is 32. In order to capture the colour properties of the whole image, the number of times a pixel is in each range is counted. The colour histogram is built with the function *cv2.calcHist* in OpenCV[3]. It allows the users to calculate the histogram the colour channels Red, Green and Blue simultaneously. In this doctoral work, I represented three-colour channels, each with 8 bins, this results in an 8 × 8 × 8 = 512-dimension vector for each image. Figure 4.5 shows a 2D colour histogram[4].

Local Binary Patterns (LBP) LBP describes images entirely by computing the local representations of texture. Proposed by Ojala et al. (1996), LBP has been employed in several domains including facial recognition (Liao et al., 2007), image retrieval (Yuan et al., 2011) and object detection and matching (Trefný & Matas, 2010) owing to its ability to discriminate and isolate changes. LBP ignores colour information. Therefore, original images are transformed into grey scale before extracting. Traditionally, the LBP histogram is calculated as follows: Pixels from the image are selected randomly and the grey value of p neighbours in a circle with the radius r pixels around these are compared. If

[3] `https://docs.opencv.org/4.x/d6/dc7/group__imgproc__hist.html#ga4b2b5fd75503ff9e6844cc4dcdaed35d`

[4] For display purposes, we show only 2D colour histograms in this work.

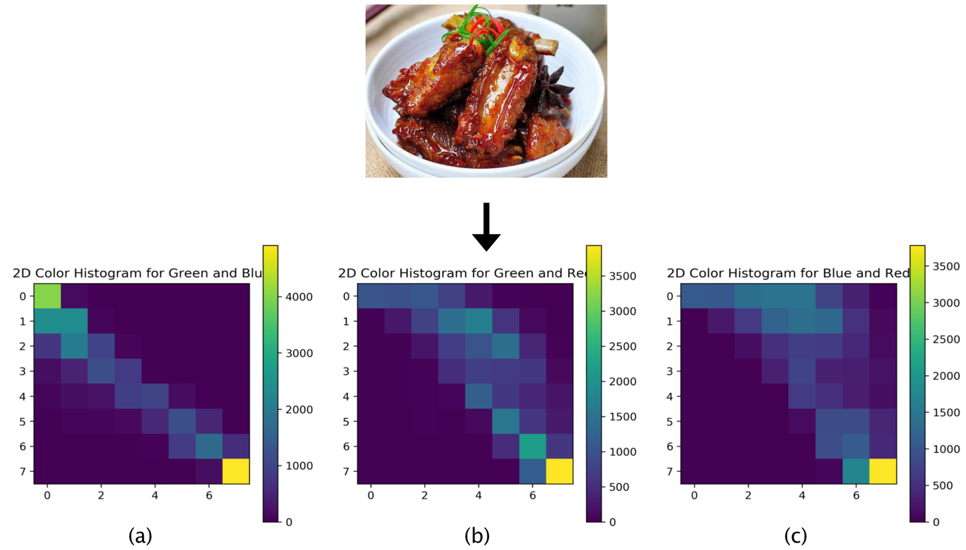


Figure 4.5: 2D colour histogram of an example image. (a) a 2D colour histogram for the Green and Blue channels. (b) a 2D colour histogram for the Green and Red channels. (c) a 2D colour histogram for Blue and Red channels.

the grey value of the chosen pixel is greater than or equal to one of its neighbours, the neighbour point is set to 1. Otherwise, the point gets a value 0. Subsequently, a group of binary strings are formed, and the LBP value of the chosen pixel is the decimal converted from it. The process is repeated until the LBP value has been computed for every pixel. The final features describing the texture of the image are obtained by counting the frequency of LBP values and output as a histogram. In this work, however, the uniform and rotation-invariant LBP code, which was developed by Ojala et al. (2002), was employed. It is defined as the LBP with only at most 2 transitions from 0 to 1 or vice versa are "uniform", as shown in the top image in Figure 4.6 (c); others are deemed to be "non-uniform" and treated as one situation, as shown in the bottom image in Figure 4.6 (c). Here, the number of neighbours p and radius r were set as 24 and 3, which were proven to be powerful in Ojala's original work (Ojala et al., 2002), leading to 25 uniform invariant LBP codes and all other codes were classified into 1 non-uniform code. A histogram of 26 dimensions indicating the frequency of uniform and non-uniform LBP codes were then calculated as the visual representation relating to texture of images. This approach is found to be more efficient than the traditional way since it outputs lower dimensional vectors. Such as with the parameter in this work, where $p = 24$, traditional LBP calculation generates a 2^{24} dimensional vector for an image whereas the approach I used outputs only a 26-dimensional vector. The process of extracting LBP for an image is shown in Figure 4.6. LBP in this work was implemented with scikit-image[5] and the LBP histogram was calculated by applying *numpy.histogram*[6] in Python.

[5] `https://scikit-image.org/docs/stable/auto_examples/features_detection/plot_local_binary_pattern.html`

[6] `https://numpy.org/doc/stable/reference/generated/numpy.histogram.html`

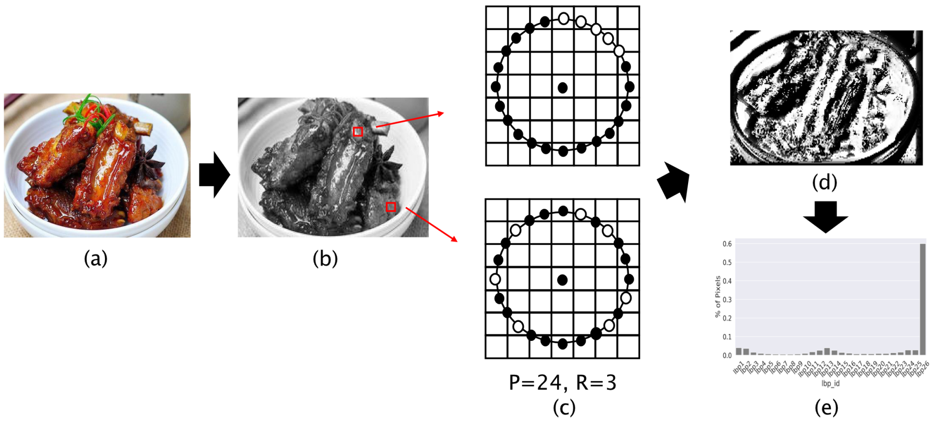


Figure 4.6: The process of calculating LBP. (a) The origin image. (b) The grey scale of the image with 2 pixel examples are selected. (c) The example pixels and their 24 neighbours of radius 3. The top one shows the uniform pattern, and the bottom shows the non-uniform pattern. (d) The LBP representation of the original image. (e) The histogram that shows the number of times each LBP pattern occurs.

Bog of Visual Words (BoVW) based on Descriptors of Scale-Invariant Feature Transform (SIFT) SIFT is a robust image representation (Mikolajczyk & Schmid, 2005). The main idea of using SIFT is to identify and describe the keypoints within images. Keypoints are the scale-invariant and rotation-invariant points that are not sensitive to changes in image resolution, scale, rotation, changes in illumination (Lowe, 2004). The example of an image with the keypoints is shown in Figure 4.7.

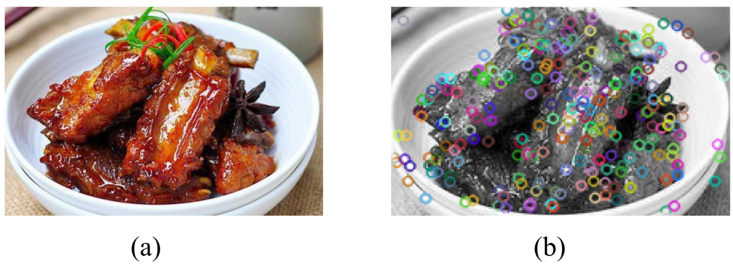


Figure 4.7: Example of an original recipe images and its keypoints. (a) the original recipe image. (b) the recipe image with its keypoints.

In this work, the keypoints were generated and described by means of applying OpenCV with corresponding functions[7]. Following the approach described in (DeCost & Holm, 2015), each keypoint was applied to 128-dimension descriptors. For all recipe images across recipe portals in my collections, there were more than 70 million descriptors in total. However, since each image has a different number of keypoints, the dimensions of the visual features of each image are not of equal size. As such, an approach motivated by an analogical method of *bag of words* used in the NLP domain was developed by Csurka et al. (2004) and applied in this work, which is known as *Bag of Visual Words (BoVW)*. The main idea of BoVW is to take each image as a document with "visual words", then a

[7] https://docs.opencv.org/4.x/da/df5/tutorial_py_sift_intro.html

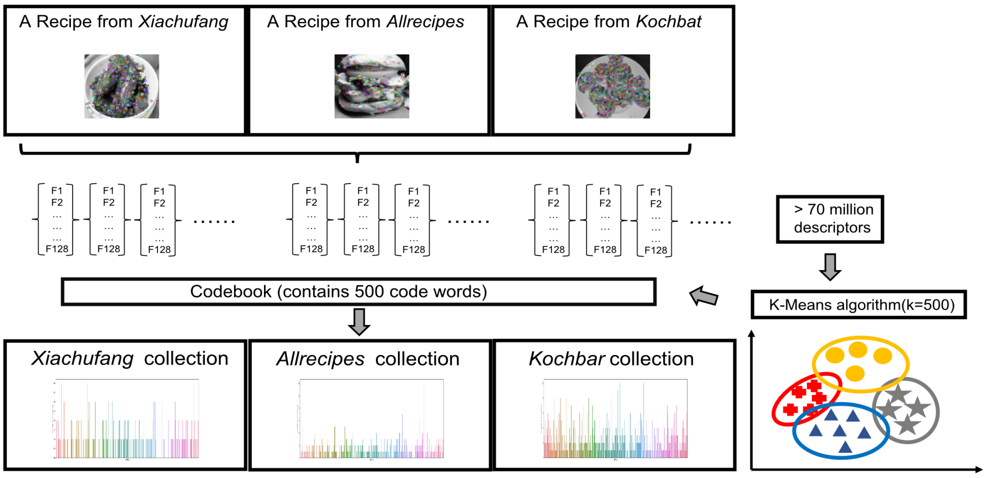


Figure 4.8: The process of obtaining recipe image representation with BoVW

codebook needs to be generated by means of a vector quantization algorithm in order to assign these into their corresponding codewords. Building on this, the K-means clustering (k=500) was applied on all SIFT descriptors in this work, and the centre of each cluster was deemed as a codeword and was used to generate a codebook. The final step is to build a histogram for each image that shows the frequency of the codewords. In the end, each image was represented by a 500-dimension vector. The BoVW stage is shown in Figure 4.8.

Besides these features, the widely applied Deep Neural Network (DNN) Image Embeddings are also used in recipe image representation in this work.

Deep Neural Network Image Embeddings (DNN) There are several well-known DNNs, such as AlexNet (Krizhevsky et al., 2012), VGG (Simonyan & Zisserman, 2014), GoogLeNet (Szegedy et al., 2015), ResNet (K. He et al., 2016), etc., as reviewed in Chapter 3. In this work, the VGG16, which was developed by Simonyan and Zisserman (2014), was applied. It has achieved the first and second places in the localisation and classification tracks respectively in the ImageNet Challenge 2014. Moreover, in (Russakovsky et al., 2015), VGG16 even reached a human level of performance in the task of image classification. In the food domain, VGG16 has been found to perform well in the works such as (J. Chen & Ngo, 2016; Ege & Yanai, 2017; J.-j. Chen, Ngo, & Chua, 2017; Salvador et al., 2017) in food classification and image-to-recipe retrieval tasks.

The VGG16 architecture is shown in Figure 4.9. VGG16 uses small (3*3) convolution kernels on the input images, which form a stack of convolutional layers, followed by three Fully-Connected (FC) layers. Spatial pooling is carried out by five max-pooling layers and all hidden layers are equipped with ReLu (Krizhevsky et al., 2012) as the activation function. The features for representing the images in this work are extracted from the first FC layer, known as FC1 layer in VGG16. The features are extracted by using the *Keras*[8] framework, resulting in a 4096-dimensional vector for each image.

The original recipe images from the collections were finally represented by 5,144 dimensional vectors, depicted in the example shown in Figure 4.10.

[8] `http://keras.io`

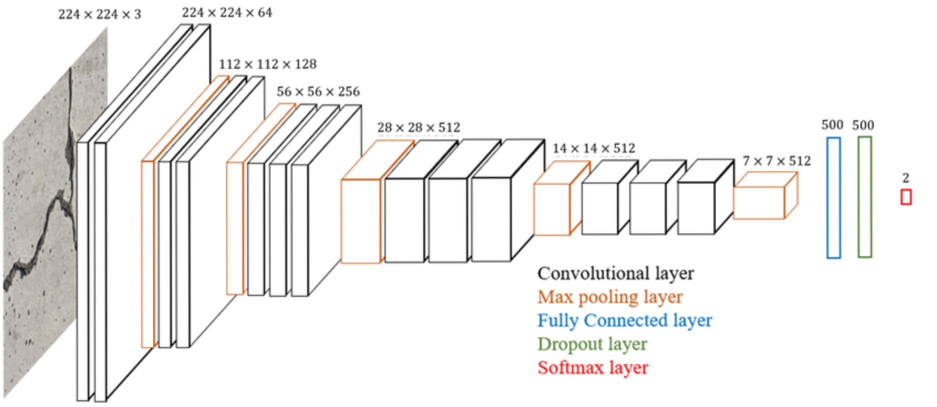


Figure 4.9: Architecture of VGG16. Taken from (D. Choi et al., 2021)

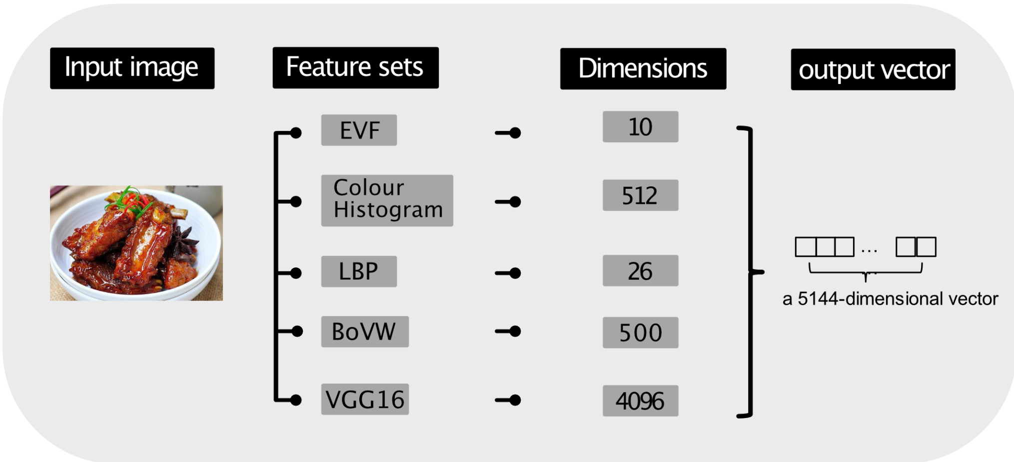


Figure 4.10: The visual features extracted from the recipe images

4.4.2 Recipe Ingredients Cleaning & Mapping

In order to map the ingredients to their corresponding flavour compounds, which represent the flavour of food, it is necessary to first clean the ingredient lists of the online recipes in my collections. The reasons for doing this are as follows: First, the recipes derived from online recipe portals are written in free-text format (see the examples shown in Figure 4.1 - 4.3.), which leads to vocabulary and matching problems. The ingredient lists of the recipes often contain non-ingredient terms (e.g., punctuation, quantifiers etc.), which introduced noises when doing the mapping. Second, the database I referred to map the ingredients to flavour compounds is FlavorDB, which provides only English ingredient names. However, the recipes in the collections were derived from recipe portals of different languages (i.e., Chinese, English and German), so that the ingredients needed to be translated into one language. In addition, there are ingredients with different names but referring to potentially identical entities due to the dialect (e.g., coriander and cilantro in US recipes; Kartoffel and Drillinge[9] in German recipes) or personal writing

[9] "Die Drillinge" was translated with Google Cloud Translation API as "triplets"

habits (e.g., chili pepper and chile pepper). In order to avoid ambiguity when doing the mapping, these ingredients were normalised into one certain name according to the ingredient names provided by FlavorDB. The whole process of cleaning the ingredient lists for mapping is shown in Figure 4.11. In the rest of this section, I describe each phase in this process in detail with examples to clarify the challenges and how these were solved.

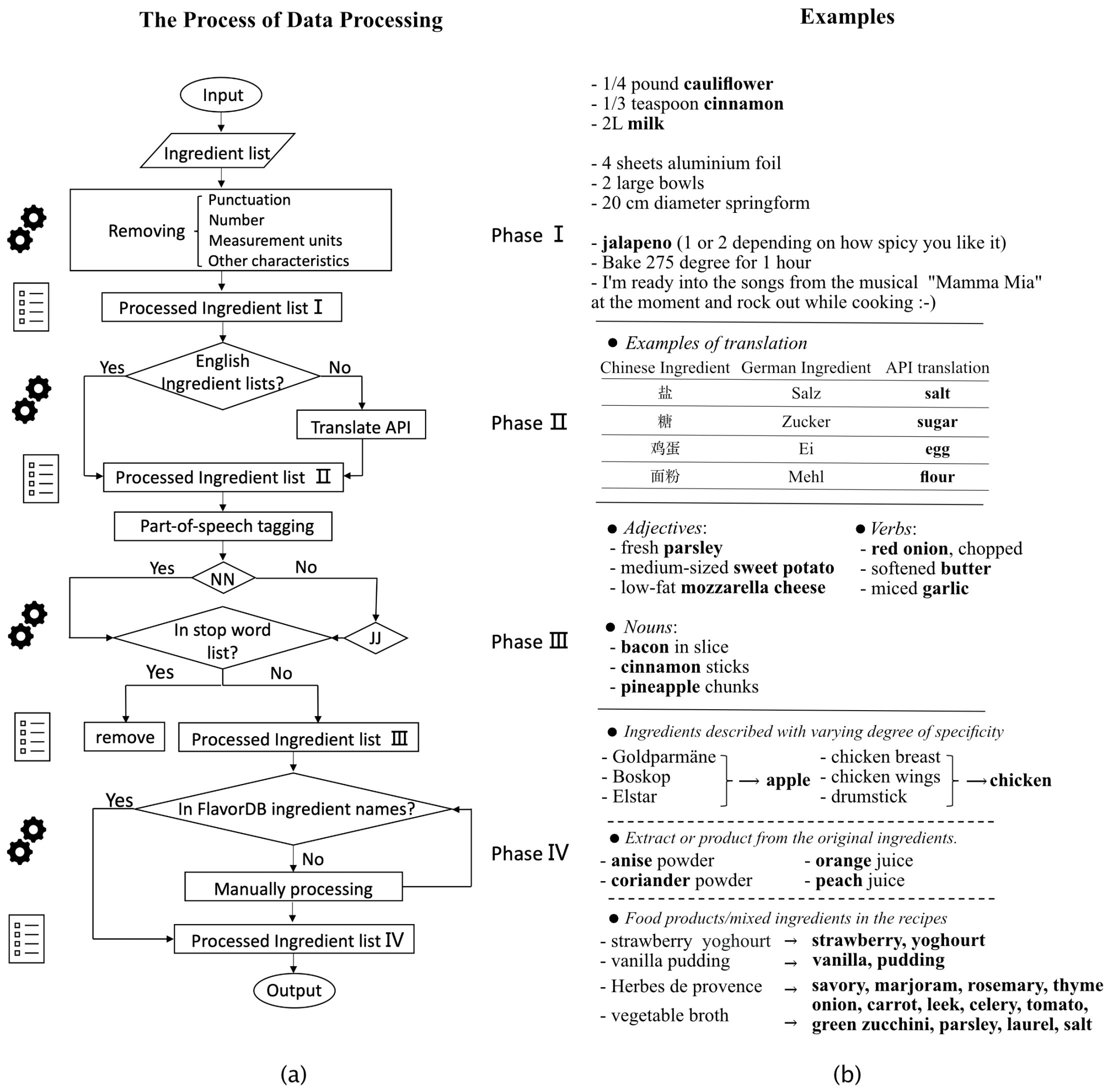


Figure 4.11: Data processing on the ingredient lists. (a) represents the flow diagram of the data processing, including 4 phases. (b) shows several examples of the processed the ingredients in each phase. The bold terms are the ingredients kept in the datasets and apply to empirical experiments in the next few chapters.

Phase I: Removing non-ingredient terms from the ingredient lists

The first phase of cleaning the ingredient lists is removing the non-ingredient terms, which mainly includes punctuation, numbers and measurement units indicating the quantity of the ingredients. There are also other terms that not related to ingredients, such as tools for cooking, eating utensils, cooking directions, as well as some words to express the cooks' moods and feeling while cooking, as shown in the examples in Figure 4.11 in Phase I.

Phase II: Translating the ingredient lists
The aim of this phase is to translate the Chinese and German recipes into English[10]. In order to do the translation, I submitted the translation requests to Google Cloud Translation API v2[11], which is free to translate 500,000 characters per month. The automatic translation ensured the majority of texts translated correctly, but there were still some ingredients that were not translated or were incorrectly translated. In order to handle this, I checked and modified the translation manually.

Phase III: Filtering the Stop Words
In this phase, I attempted to remove the stop words from the ingredient lists. The stop words refer to words such as determiners (e.g., the, a, an), coordinating conjunctions (e.g., for, an, nor, but), and prepositions (in, under, towards), etc., in general text preprocessing. These words are deemed not useful in further text analyses. In this work, I identified the stop words in the recipe domain. These words occur in the ingredient lists commonly, but in fact, they introduce noises in the mapping process. These words often occur before or after ingredient terms to describe them and are commonly adverbs, adjectives and verbs in past tense. Examples of ingredients with the stop words are "very ripe bananas" or "chopped onion", in which "very", "ripe" and "chopped" can be recognised as the stop words in the ingredient lists. More examples of verb and adjective stop words are shown in Figure 4.11 in Phase III. In order to identify the stop words, I first applied Part-of-Speech Tagger (POS Tagger) released by Stanford NLP Group[12] to assign the POS tag for terms in the ingredient lists. Most of non-nouns can be recognised as stop words in this task. However, there are also exceptions. For example, there are nouns that are neither ingredient names nor does it help with the mapping process, such as "slice" in "bacon in slice", "sticks" in "cinnamon sticks", and "chunks" in "pineapple chucks" as shown in Figure 4.11 in Phase III. The other exception is the adjectives that occur in the ingredients names. Removing these words would turn an ingredient into another. Examples of these words are "green" in "green beans" and "sweet" in "sweet potato". These words and ingredients were identified manually according to the ingredient names provided by FlavorDB in this phase.

Phase IV: Manual processing and mapping
This phase was designed to deal with the ingredients, which have been processed by the previous phases, but still could not be mapped to their corresponding flavour compounds in FlavorDB. There are two types of these ingredients. The first type contains the ingredients that have not been included in FlavorDB. For example, the herbs occurred in *Xiachufang* collection (e.g., Senna tor, Ulmus), and the processed products for cooking occurred in *Allrecipes* and *Kochbar* (e.g., baking mix, stir-fry mix, fix für ziwebel sahne hanchen, maggi für nudel schinkengratin). The second type contains the ingredients that can be mapped but they needed to be normalised or manually processed before that. In this phase, I dealt with the ingredients of the second type. Below I list a couple of common

[10]The pre-processing work on *Xiachufang* recipes were done in Chinese by the native speaker, the translation API was applied to translate the processed Chinese ingredients to English.
[11]`https://cloud.google.com/translate/docs/basic/translating-text`
[12]`https://nlp.stanford.edu/software/tagger.shtml`

situations in this process and how I processed them:

- *Ingredients described with varying degree of specificity* These ingredients are, for example, different varieties of the same ingredients, such as different types of apples - Goldparmäne, Boskop, Elstar, and different types of wines - Merlot, Chardonnay, Chianti. I normalised these with specific varieties into their corresponding general ingredient names in FlavorDB. In addition, there are also ingredients refer to different parts of poultry or animal products (e.g., chicken breast/wings, egg yolk/white), these were also normalised according to the ingredient names in FlavorDB.

- *Extract or product from the original ingredients.* It can be found in the online recipes, the users applied extract or products (e.g., powder or juice) from original ingredients instead of the ingredients themselves. However, FlavorDB contains mostly the name of the original ingredient. Thus I transferred the extracts to its origins. Some examples are shown in Figure 4.11 in Phase IV.

- *Food products/mixed ingredients in the recipes.* These ingredients are, for example, strawberry yoghurt. I modelled strawberry yoghurt as flavour compounds of strawberry and flavour compounds of yoghurt. This is owing to the facts, strawberry yoghurt have both the flavour of strawberry and yoghurt. The flavours of ingredients such as vanilla pudding, chocolate cream were also modelled this way. Moreover, this means was also applied to the ingredients, which are mixed spices and food products, such as Herbes de Provence, Suppengrün. These ingredients were divided into individual ingredients according to authority recipes or appreciated recipes from the recipe portals, and their flavours were modelled based on the individual ingredients contain in them. Some examples of them are shown in Figure 4.11 in Phase IV.

Beside these, there are also plenty of manual works such as modifying the typos, checking and correcting incorrect translation, etc. The codes for the whole process of data cleaning are uploaded to my github repository[13].

Finally, after the data cleaning and processing, there are approximately 34% ingredients in *Xiachufang* recipes, 43% ingredients in *Allrecipes* recipes and 19% ingredients in *Kochbar* recipes have been mapped to their corresponding flavour compounds in FlavorDB. The number of raw terms in the online recipes, processed ingredients and the ingredients with flavour compounds in each collection is shown in Table 4.3.

Moreover, I kept only recipes with full list of flavour compounds for the experiments on the flavour aspects. Thus, the number of recipes in each collection has changed. Specifically, the number of recipes in each collection for addressing *Issue 1* and *Issue 2* are shown in Table 4.4 and Table 4.5 respectively.

The recipes with full list of flavour compounds entered into the next step, in which each of these was represented by their ingredients and flavour compounds respectively.

[13] `https://github.com/QingZhang1001/Food_aesthetics/tree/main/Flavour_data_cleaning`

Table 4.3: The number of terms (before pre-processing) and the number of ingredients mapped to FlavorDB ingredients names (after pre-processing) in the recipes from each recipe portal

Origin	Num. Raw terms in the recipes[1]	Num. Processed ingredients in the recipes[2]	Num. Ingredients with flavour compounds
Xiachufang	10,584	973	331
Allrecipes	30,333	824	356
Kochbar	42,022	2,084	401

Note:
[1] The column *Num. Raw terms in the recipes* refers to the number of text split with ',' in the raw recipes, including the non-ingredients terms, as described in Phase I.
[2] The column *Num. Processed ingredients in the recipe* refers to the number of ingredients that have been processed from Phase I - V.

Table 4.4: The number of recipes with full list of flavour compounds for addressing *Issue 1* on the flavour aspects

Origin	Num. Recipes	Num. Recipes with full list of flavour compounds
Xiachufang	25,000	11,842
Allrecipes	25,000	12,432
Kochbar	25,000	9047

Table 4.5: The number of recipes with full list of flavour compounds for addressing *Issue 2* on the flavour aspects

Origin	Appreciated?	Num. Recipes	Num. Recipes with full list of flavour compounds
Xiachufang	appreciated	2,500	1,039
	less appreciated	2,500	1,327
Allrecipes	appreciated	2,500	934
	less appreciated	2,500	1,199
Kochbar	appreciated	2,500	936
	less appreciated	2,500	928

4.4.3 Food Representation with Ingredients and Flavour Compounds

In this section, the recipes were represented by ingredients[14] and flavour compounds, respectively. Two techniques were employed for this purpose, that were **TF-IDF** (Salton & Buckley, 1988) and **Word2Vec** (Mikolov et al., 2013). Both of these can be applied in representing textual data in order to deal with NLP problems but with different calculation strategies.

[14]the ingredients introduced in this work to build the baseline models. It aims to compare the ability of the traditional and commonly applied texture information, i.e., ingredients, to that of aesthetic features, namely, visual appearance and flavour in food classification and preferences prediction tasks. This will be clarified in Chapter 6 and 7.

TF-IDF reflects the importance of a word in a document of a corpus. The value of it is the product of term frequency (TF) and inverse document frequency (IDF). Assuming there is a corpus containing a number of documents, and there are several terms in each document, the TF is calculated as:

$$TF = \frac{Number\ of\ times\ a\ specific\ term\ occurs\ in\ a\ document}{total\ number\ of\ words\ in\ the\ document} \tag{4.10}$$

TF assigns each term in a document a weight, that depends on the number of occurrence of the term in the documents. However, TF alone does not represent the word importance in the whole corpus, since it treats all terms equally. This leads to the words that occur extremely frequently but have no or little discriminative power, such as “the”, “a”, “and” obtain high TF values. In order to weigh down the frequent terms while scale up the rare ones, IDF was proposed (Jones, 1972), the formula is:

$$IDF = \log(\frac{number\ of\ documents}{number\ of\ documents\ the\ specific\ term\ appears}) \tag{4.11}$$

The idea of IDF is reducing the TF weight of a term by a factor that grows with its frequency in the whole corpus.

Finally, TF-IDF calculated as :

$$TF - IDF = TF \times IDF \tag{4.12}$$

TF-IDF has been a widely applied technique for text classification and information retrieval, and it has also been applied in the food studies. For example, in (Zhu et al., 2013), the authors weighed the ingredients by means of an equation inspired by TF-IDF, which was applied to penalise ingredients that are very popular but carry little information. And in (Sajadmanesh et al., 2017), TF-IDF weighting was applied to determine the notable ingredients. Moreover, TF-IDF has been applied to the development of food recommender systems, such as in (El-Dosuky et al., 2012; Harvey et al., 2013), it was applied to measure the similarity between user previous eating habits and the recipes in order to determine which unrated recipes to be recommended. In this doctoral work, the representation of recipes was done by transforming ingredient and flavour compounds lists to matrices of TF-IDF features, which was completed by applying relevant functions in *scikit-learn*[15].

In comparison to TF-IDF, which is a statistical measure that can be calculated directly based on the frequency of terms and documents, Word2Vec (Mikolov et al., 2013) employs a 2-layer neural network model to learn word embeddings. It accepts the inputs as a corpus of text, and outputs vectors for each word present. Words, which are semantically similar to each other, result in vectors that are closer in the vector space. In order to achieve this, Word2Vec provides two different training algorithms, CBOW (Continuous bag-of-words) and Skip-gram. The main idea of CBOW is to represent a word in a corpus by means of using its surrounding context of words (i.e., the words before and after the target words). In contrast, Skip-gram uses a target word to predict its surrounding words. In this presented research, I chose to represent each ingredient/flavour compound with its surrounding ingredients/compounds in the recipes, thus the CBOW algorithm was applied. In addition, CBOW was also found to be trained much faster than Skip-gram

[15] `https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html`

(Mikolov et al., 2013). In order to eliminate the influence brought by the order of ingredients and flavour compounds when applying Word2Vec, I sorted them alphabetically in each recipe, as in (Rita et al., 2020). The word embeddings in this work were trained with Gensim[16] in Python, which is a free-available NLP toolkit. The hyperparameters required for the process were fine-tuned as follows:

- *size:* it determines the dimensions of vectors for each word after training, I applied the default value of 100 in this work, so that each ingredient was represented by a 100-dimensional vector;
- *min_count:* it regulates the minimum frequency of words that can be trained in the model. To be specific, the default value of *min_count* is 5, which means, only words that occur more than or equal to 5 times would be trained and represented by word embeddings. In this work, I set the value as 1, in order to represent every ingredient and flavour compounds in the datasets;
- *window:* it decides how many words around the target word are taken into consideration when the model is training. I took the default value 5 in this work, which means, for each ingredient/flavour compound in the recipe, the model learns its embeddings by considering 5 ingredients/flavour compounds before and after it.

It is noted that Word2Vec only outputs vectors of the individual ingredients and flavour compounds, while in this work, what I need is the representation of the recipes. In order to deal with it, for each recipe with an ingredient list and corresponding flavour compound list, I summed all word embeddings of ingredients/flavour compounds up and divided them by the length of recipes in terms of ingredients/flavour compounds.

Finally, each recipe in the collections were finally represented by the TF-IDF and Word2Vec vectors of ingredients and flavour compounds respectively.

4.5 Chapter Summary

In this chapter, online recipes from three popular Chinese, US and German portals were gathered and selected as the proxies for the food cultures. Online recipe images and flavour compounds corresponding to the ingredients were applied to represent the recipes on the visual and flavour aspects. These vectors of food representations will be involved into the empirical experiments in the following chapters.

[16]`https://pypi.org/project/gensim/`

Chapter 5

Cross-Cultural Food Classification and Preferences on the Visual Aspects

5.1 Introduction

In the previous chapter, I explained how food images sourced from recipe portals can be treated as proxies to represent the visual appearance of food. This chapter will present empirical experiments on the visual appearance of food across cultures. The reviewed literature in Chapter 3 suggestes that human online food choices can, to some extent, be explained by visual appearance of the options. It suggests that it is promising to develop food recommender systems with human food visual preferences. However, in previous research involving food images for food recommendations, an important context factor known to impact food choices, culture, has not been focused to be studied. Therefore, to pave the way to apply food visual appearance in food recommender systems in a cultural appropriate way, the experiments in this chapter are designed to address the issues raised in Chapter 1 on the visual aspects, which are specified as follows:

- *Issue 1.* To what extent is it possible to differentiate the food across cultures based on the visual representations of the recipes?
- *Issue 2.* To what extent is it possible to identify the differences and ascertain stable patterns of food preferences across cultures based on the same visual representations?

The chapter is divided into two main parts, Study I and Study II, which correspond to *Issue 1* and *Issue 2*, respectively. Three distinct food cultures, China, US, and Germany are selected to be studied, with the recipe images being sourced from the corresponding recipe portals, *Xiachufang*, *Allrecipes*, and *Kochbar*. These images and data associated with the recipes represent the visual appearance of food, and the food preferences of people across the three cultures (as described in Chapter 4). The methodologies applied in these two studies share many similarities. These are, firstly, classifiers are trained using visual information encoded in the images. The performance being evaluated using the accuracy score. Secondly, a user study is used to validate the best performing models by comparing predictions made by the models to judgements collected from human participants from three countries.

The chapter is structured as follows: Study I and Study II are presented in Sections 5.2 and 5.3, respectively. In the opening subsections, Section 5.2.1 and 5.3.1, an outline of each experiment is given. To guide the studies, several research questions are raised in these sections. Section 5.2.2 and 5.3.2 elaborate the research methodology, including the machine learning approaches used to train classifiers, as well as the design and justification of the user study. Section 5.2.3 and 5.3.3 provide the results of the studies, the primary findings towards the research questions are presented in 5.2.4 and 5.3.4, with their implications being discussed. Finally, the conclusions of Study I and Study II are provided and summarised in Section 5.4.

5.2 Study I: Predicting Food Culture Based on Visual Information

5.2.1 Study Outline

This study investigates the differences in how online recipes are presented in different cultures. It does so by employing algorithmic methods. The datasets described in Chapter 4 contain the recipe images derived from Chinese, US and German recipe portals, which assume to represent the food cultures of interest visually. By means of extracting visual features from the recipe images and training models using these, a prediction task is formulated to predict the source recipe portal for each image, and to answer the Research Question as follows:

- *RQ1.* To what extent is it possible to differentiate the recipe images from the recipe portals of different food cultures with machine learning models based solely on visual properties?

In addition, with the aim of developing food recommender systems for assisting humans to make food choices, it is important to know how people perceive food images. Within this context, this study investigates how people perceptions vary across cultures. Two further research questions guide the user study:

- *RQ2.* How able are humans to distinguish recipes from the recipe portals of different food cultures solely by observing the recipe images?

- *RQ3.* Which factors, including information cues from the images and user properties, influence the judgements made by humans?

5.2.2 Methods

5.2.2.1 Classifying Recipes by Means of Visual Features and Machine Learning Approaches

To establish the extent to which it is possible to use visual information to determine the portal from which a recipe was sourced (the *RQ1*), I formulated the problem as a prediction task whereby classifiers were trained to predict the source portal for each image. The images were represented as a multi-dimensional vector by extracting 5,144 visual features from each image as described in Chapter 4. The feature sets include **EVF**, **Colour**

Histogram, **LBP**, **BoVW** and **VGG16**. I then built classifiers using each feature set individually and then all feature sets combined. Three supervised classification approaches were applied: **Naive Bayes (NB)**, **Logistic Regression (LOG)** and **Random Forest (RF)**. In this prediction task, the recipe collections that contain 25,000 recipe images from each portal were applied as the data basis. In all experiments for this task, the data were split randomly into training (70%) and test (30%) sets, with a 5-fold cross-validated Randomised Search CV being applied on the training set to determine the optimal parameters for LOG and RF. The performance of the classifiers was measured by accuracy (ACC).

5.2.2.2 The Classification Task by Means of Human Judgement

To establish human performance on the same task I designed a remotely deployed user study and recruited participants located in China, US and Germany via crowd-sourcing platforms and social media. The user study was hosted on a server owned by the University of Regensburg, Germany and in all cases accessed by means of an anonymised URL. By recruiting participants from different cultures, I was able to investigate human food choices vary across cultures. This section first describes the design of the study, then presents the process of recruiting participants. Basic statistics about the participants are also provided in this section. Finally, I illustrate the methods I applied to analyse the data collected from this user study.

Study Design In the main part of the study, participants were shown images sourced from different portals and were required to answer three questions with respect to each image. On completing the study, participants provided demographic and other background information. Participants were each shown nine images, three from each collection, one after the other. All images were drawn randomly from the same test set used to evaluate our classifiers (see Section 5.2.2.1). To increase the generalisability of the findings, I maximised the number of images used by assigning each image to only one participant. After showing an image, participants were first asked to decide from which of the three recipe portals the associated recipe was sourced. The study approach, the selection of the images, the questions asked, and their wording were tested in a small-scale pilot study prior to performing these experiments.

Next, participants were asked to report, on a 5-point Likert scale, their confidence in the label they assigned. In a final question, participants were able to select one or more items from a list of factors that I believed may have been influential in their judgements. These included factors relating to food, e.g., recognisable ingredients, type of food, food colour and shape, as well as non-food factors, such as the food container, eating utensils or their gut instinct. The reasons for focusing on these factors are that they are commonly reported in the literature and reflect features of our classification approaches. More concretely:

- *Ingredients:* The ingredients of meals have been proven to vary from culture to culture (L. Pan et al., 2017; J. Chen & Ngo, 2016). In the previous work, e.g., (Sajadmanesh et al., 2017), ingredients were commonly used to build food classifiers and show great predictive power. In this user study, participants were allowed to judge the source of recipes based on ingredients, and I can investigate whether recognising ingredients was helpful for them to make the right judgements.

- *Type:* Type refers to the dish type in this user study, such as Stir Fry, Pizza, Bread etc. As known in (Kusmierczyk & Nørvåg, 2016), when food type is given, it is helpful for algorithms in predicting food ingredients. I put the factor Type here to see if food type has a positive influence for the participants in making the judgements.

- *Colour:* Colour is often used to classify food automatically (Farinella et al., 2016) and in this study correspond to the visual features of Colour Histogram. The colour of food has also been proven to affect human perception of food, sometimes leading to misrecognition (Spence et al., 2010; Appleton & Smith, 2016).

- *Shape:* This factor relates to the visual feature LBP. According to (Geirhos et al., 2018), people rely on shape in classifying objects while algorithms focus on the texture. Therefore, I put the factor here to see whether and to what extent human make judgement about food based on it.

While the above listed factors all relate to the food itself, the remaining questions were associated with supplementary factors, such as the food container, eating utensils, such as cutlery or chopsticks, and instinct, which derived from the situation where participants reported that they relied on their "feelings" to make judgements in the pilot study. These factors were all reported by the participants as important during the pilot survey. Participants could also list further factors in a free-text field. An example task and associated questions are shown in Figure 5.1.

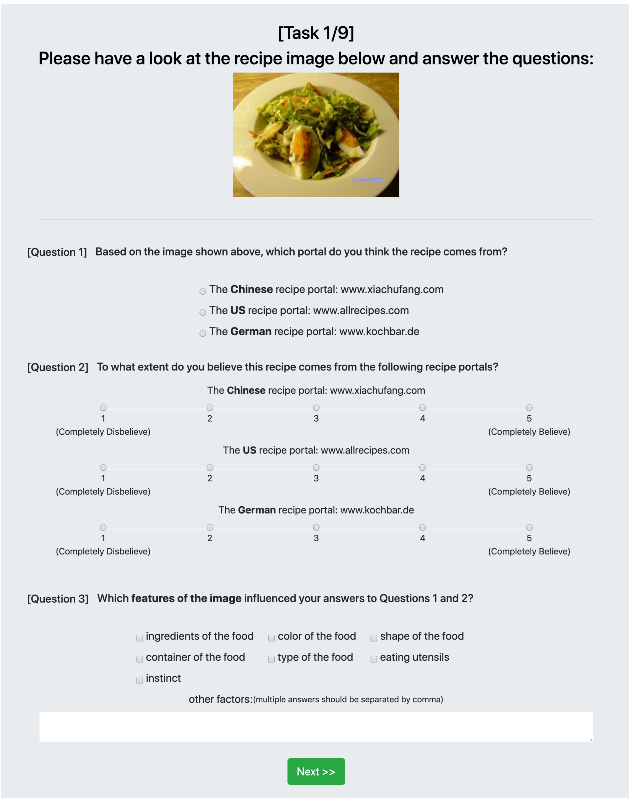


Figure 5.1: Screenshot of an example task from the user study in Study I

After labelling the images, participants completed the study by answering 13 questions, which captured participant demographics as well as other information of interest. The details are shown in Table 5.1.

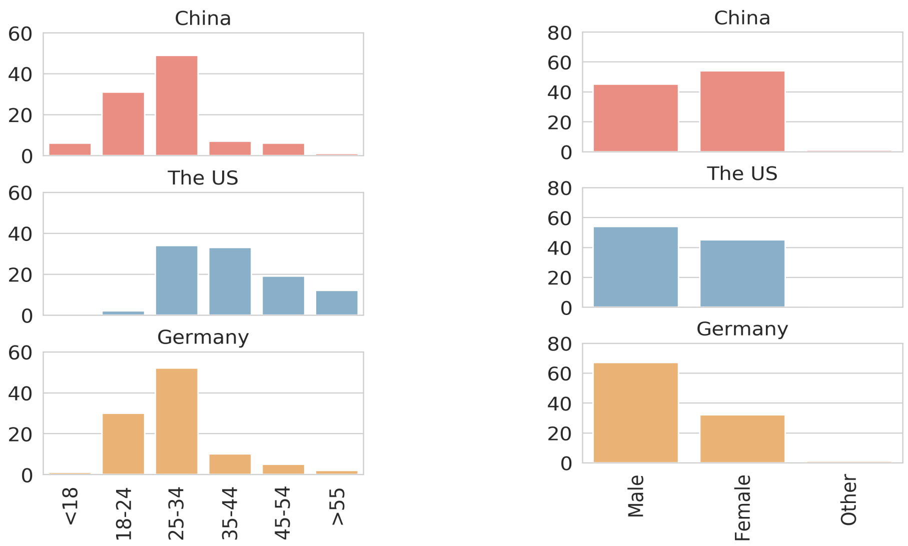
Table 5.1: Demographics questions for the participants of the user study in Study I

Question	Scale
Personal information	
Age	<18, 18–24, 25–34, 35–44, 45–55, >55
Gender	Male, Female, Other
Nationality	Select from a drop-down list
Experiences with the recipe portals	
Familiarity with each recipe portal	Likert scale 1 (Not at all) – 5 (Very familiar)
Frequency of using recipe portals	Hardly use, At least once every three months, At least once per month, At least once per week, Use on a daily basis
Settlement and travel experience	
Experience in China	Never visited, I have been there once or a few times, I visit or have visited regularly, I have lived there for many months or longer, I am a permanent resident
Experience in US	Never visited, I have been there once or a few times, I visit or have visited regularly, I have lived there for many months or longer, I am a permanent resident
Experience in Germany	Never visited, I have been there once or a few times, I visit or have visited regularly, I have lived there for many months or longer, I am a permanent resident
Frequency of cross-continental travelling	Never, Less than once per year, 1–2 times per year, More than 2 times per year
Interests in food/recipes from foreign cultures	
Interest in food/recipes from other cultures	Likert scale 1 (No interest at all) – 5 (Very interested)
Frequency of trying food/recipes from other cultures	Hardly ever, Less than once per month, At least once per month, At least once per week, Most days
Free-text field	Blank space left for all participants

Participants Recruiting The study was originally deployed on Amazon Mechanical Turk[1], a popular crowdsourcing platform, as a means to recruit participants restricted to individuals from China, the US and Germany. To ensure participants performed reliably, participation was restricted to only those who had a "HIT accept rate" of more than 98% in their previous tasks. Participants were paid 50 Cent US dollar for their participation. This approach quickly provided the sought-after 100 participants from the US, but after several weeks only 57 German participants were recruited, and no Chinese participants were found. To recruit German participants, I supplemented the sample by advertising via university mailing lists (our institution is located in Germany) and social media via the authors' personal Twitter and Facebook accounts. I additionally deployed a Chinese version of the study (where instructions and questions were translated to Chinese, as shown in Appendix B) on the platform Wenjuanxing[2] and advertised this on the Chinese social media channels Douban[3], Xiaomuchong[4] and Wechat. Participants were reimbursed 1 Yuan for taking part. These approaches combined allowed 100 participants from each country to be recruited. Figure 5.2 shows the distribution of the participants' age (Figure 5.2 (a)) and gender (Figure 5.2 (b)) from each location. Participants who were located in

[1]https://www.mturk.com/
[2]https://www.wjx.cn/
[3]https://www.douban.com/
[4]http://www.xiaomuchong.com/bbs/

Germany and China were younger than those in the US. This is because the participants from Germany and China were recruited with the author's personal social media account, leading to most participants are at similar age (i.e., 25 - 34) to the author. Whereas the participants from US were recruited from Amazon MTurk, in which only approximately 30% users under the age of 30 (Moss, 2020). In addition, the distribution of gender in each country was also imbalanced. More males took part in the US and Germany, while this trend is reversed in the Chinese sample.



(a) (b)

Figure 5.2: Participant demographics in the user study of Study I. **(a)** Age distribution of participants from each country. **(b)** Gender distribution of participants from each country.

Methods of Data Analysis After the collection phase was complete, the data were analysed in different ways. The classification performance of both the prediction models and human judgements was measured in terms of accuracy (ratio of successfully made classifications to total number of classification decisions, ACC). The performance of both the prediction models and human judgements was visualised using confusion matrices. These are useful since they help illustrate in which cases mistakes were made, as well as how these were made (i.e., which labels were erroneously applied in which cases). Appropriate inferential statistics were used to establish differences across groups (e.g., in terms of gender, interest in food/recipe from foreign cultures, etc.). Binary logistic regression analyses were applied to determine if participants' answers related to demographics or other factors and ordinal logistic regression models were built with the same factors, as well as participants' reported confidence in their labels. This provides an understanding of which factors help predict confident decisions. Binary logistic regression was used in cases where the dependent variable had two classes; ordinal logistic regression was employed when the dependent variable was measured on an ordinal scale. I created numerous different models using groups of feature sets as shown in the tables in appropriate sections below.

Participant responses to free-text questions were analysed qualitatively using a bottom-up, inductive approach. Responses were coded in duplicate, similar to related responses were grouped together, and the groups were collapsed until a hierarchical structure was formed. I communicate the results in the form of a coding scheme and provide examples to illustrate the most important codes.

5.2.3 Results

The results of the study I are reported in the following subsections to answer the research questions that raised in section 5.2.1

5.2.3.1 Predicting the Origin of Recipes Based on Visual Features with Machine Learning Approaches (RQ1)

Table 5.2 presents the performance of each classifier. The bottom line of the table illustrates that the recipe images from the three recipe portals are sufficiently visually distinct, such that they can be classified by the algorithms with relatively high accuracy. When using all the visual features available, all three classifiers offered accuracy (ACC) of ACC = 0.73 or better, with the logistic regression model achieving the highest accuracy of ACC = 0.89. The DNN features offered the best predictive power while the BoVW was ranked in second place. Single EVF offered the lowest accuracy, but nevertheless, all performed slightly better than random (ACC = 0.33). Models utilising combined EVF offered improved accuracy (ACC = 0.47 – 0.55). The performance of the remaining feature sets, such as colour histogram and LBP, shows no significant difference with that of combined EVF.

Table 5.2: Results for predicting which portal a recipe image belongs to based on different visual feature sets. Best performing scores for each classifier are bold. *NB* = Naive Bayes; *LOG* = Logistic Regression; *RF* = Random Forest.

Features	Accuracy		
	NB	LOG	RF
EVF(Brightness)	0.41	0.41	0.42
EVF(Sharpness)	0.41	0.41	0.43
EVF(Contrast)	0.37	0.37	0.42
EVF(Colourfulness)	0.38	0.38	0.41
EVF(Entropy)	0.38	0.37	0.40
EVF(RGBContrast)	0.38	0.38	0.41
EVF(Sharpness Variation)	0.41	0.41	0.41
EVF(Saturation)	0.39	0.39	0.40
EVF(Saturation Variation)	0.39	0.38	0.41
EVF(Naturalness)	0.38	0.38	0.40
EVF(All features)	0.47	0.54	0.55
Colour Histogram	0.43	0.52	0.54
LBP	0.48	0.52	0.52
SIFT	0.58	0.72	0.67
DNN	0.67	0.86	0.78
ALL Features	**0.73**	**0.89**	**0.85**

Figure 5.3 shows the confusion matrix for the best performing model, illustrating that the classifier was more accurate when identifying recipes from *Xiachufang* (ACC = 0.95) than classifying those from the other two (ACC = 0.86 and 0.85). The majority of misclassifications for *Allrecipes* and *Kochbar* were labelled as belonging to the other of these two classes, with very few being misclassified as *Xiachufang* recipes. In other words, when applying the same algorithms and visual features to images, the recipes from the Chinese recipe portals seem easier to differentiate.

In summary, the experiments show that it is possible to distinguish between the recipes from different recipe portals of China, US, and Germany based solely on the proposed visual features. *Xiachufang* recipe images appear to be more visually distinct with images from the other two portals more likely to be confused.

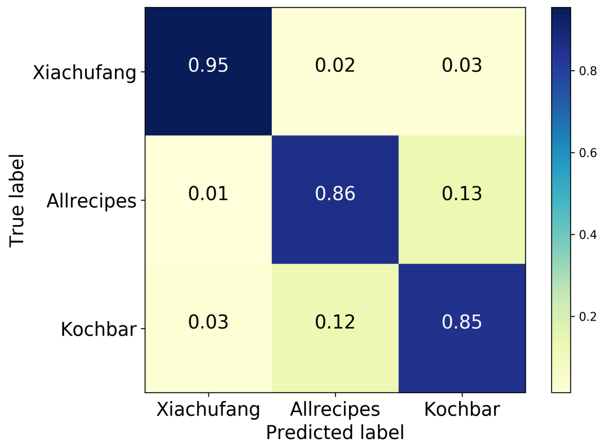
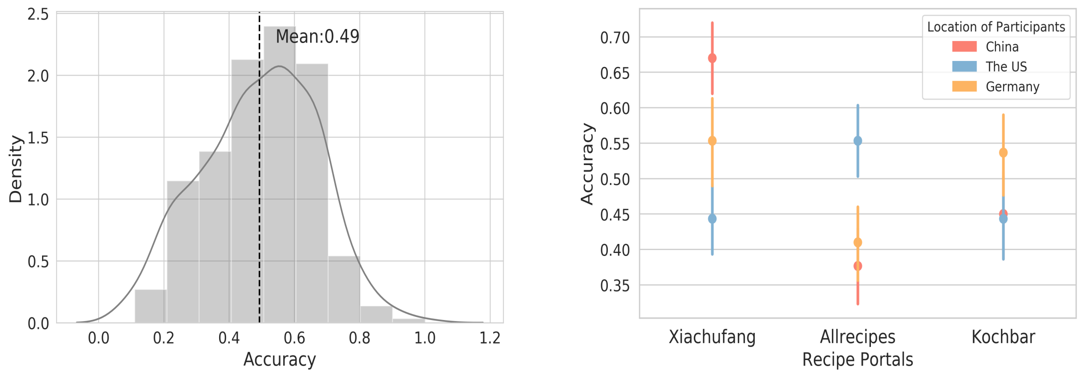


Figure 5.3: Confusion matrix of the best performing classifier on the samples

5.2.3.2 Analysing Human Labelling Performance (RQ2)

As shown in Figure 5.4, human performance on the same food classification task was markedly poorer. Figure 5.4 (a) presents the accuracy distribution over all 300 participants, with most achieving an accuracy of between ACC = 0.40 and 0.60; M = 0.49. Figure 5.4 (b) depicts how accuracy varied for participants from the three countries across the different food portals. Performance for the Chinese and American participants was highest when they were tasked with classifying recipe images from their own country. Participants from China were particularly accurate with *Xiachufang* recipe images, with the accuracy ACC = 0.67. Participants from Germany, on the other hand, achieved a slightly higher accuracy when classifying recipes from *Xiachufang* than images from *Kochbar*, with the ACC = 0.55 and 0.54 respectively. For Chinese and German participants, recipes from *Allrecipes* were the most difficulty to classify.

When comparing the performance of the human participants to those achieved by the algorithms above (i.e., by examining the confusion matrices in Figures 5.3 and 5.5), it is found that humans make choices biased in the same direction as those generated algorithmically. Figure 5.5, which provides the confusion matrix of their judgements, indicates that participants made more mistakes when classifying recipes from *Allrecipes* and *Kochbar*. More than 30% of recipes from *Allrecipes* were labelled as being from *Kochbar*, while 10% fewer were mistaken for recipes from *Xiachufang*. Participants behaved similarly when classifiying the recipes from *Kochbar*. At the same time, more than half of the recipes from *Xiachufang* were classified correctly. The human judgements, therefore, followed the same trend as those provided by the algorithms: the images from *Xiachufang* seemed to be most visually distinct, whereas those from *Allrecipes* and *Kochbar* seemed to be most similar.



(a) (b)

Figure 5.4: Human performance on food origin classification task. **(a)** Distribution and mean value of participant accuracy. **(b)** Mean value and error bar for participants accuracy for each recipe portal, grouped by participant origin.

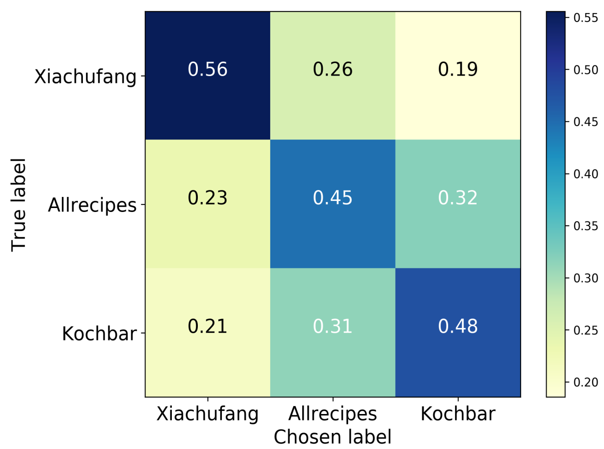
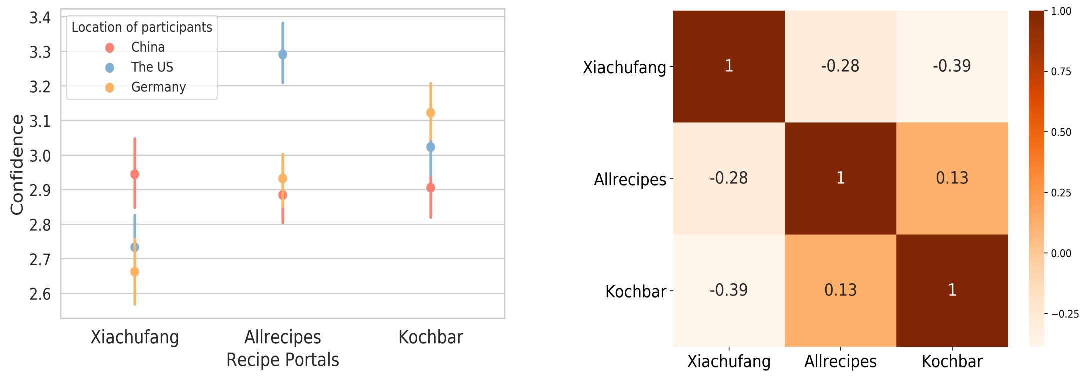


Figure 5.5: Confusion matrix of participants' judgements

Participants from different locations displayed diverse degrees of confidence in each recipe portal, as shown in Figure 5.6 (a). In general, participants reported higher confidence when labelling recipes sourced from the country where they reside. This is particularly true for the participants from the USA and Germany. Moreover, both the German and US participants reported least confidence when labelling images from *Xiachufang*. The findings may shed light on cultural differences with respect to confidence, with the Chinese exhibiting caution rather than confidence and the participants from the United States exhibiting high confidence in their judgements other than for images from the Chinese site.

Figure 5.6 (b) presents the correlation matrix for the confidence scores participants applied to their labels for images sourced from different recipe portals. It demonstrates



(a) (b)

Figure 5.6: Participant confidence in labelling recipe images across recipe portals. **(a)** Mean value and error bar of participants when labelling the recipe images from different recipe portals. **(b)** Correlation matrix for participant confidence.

that participants' confidence in their labels for *Allrecipes* and *Kochbar* images correlated positively ($p < 0.05$), while a negative correlation existed between the confidence in labels for both western portals and *Xiachufang* images. This finding aligns with those described above. It seems that when participants assumed a recipe originated from *Xiachufang*, they then believed that it was unlikely to come from the other two recipe portals and vice versa. In other words, participants believed recipe images on the western portals to look similar to each other, but different to those from *Xiachufang*.

To summarise, in this section I have shown that participants' performance in the labelling task was significantly poorer than the machine learning approaches in the previous section. The analyses, moreover, reveal differences in the labels applied and the performance of participants from different countries for images sourced from different portals. Participants typically performed best and were more confident when labelling images sourced from their home country.

5.2.3.3 Factors Leading to or Influencing Participants' Judgements (RQ3)

In this section, I explore the labelling decisions made by participants in detail. I do this by first looking at the visual features, which have been proved to be useful when predicting the source of an image, to determine if the same information can help predict the labels applied by participants. Next, I examine the explanations participants gave for their choices to understand how choices were made and/or biased, as well as to determine which, if any, helped lead to a correct label being applied. Lastly, I examine how labelling performance varied across different groups, which provides an insight into how demographic variables can influence the way images of food are perceived.

Predicting Participant Label Based on Visual Features Table 5.3 presents the utility of various visual components with respect to (a) predicting a recipe's origin and (b) predicting the label applied to the image by participants in the experiment. It is found in Table 5.3, firstly, the visual information features tell more about the actual source of a recipe image than the label applied to it by the participants. The highest accuracy for image source achieved was ACC = 0.84 with a combined feature set, which was slightly lower than with the full test set (see Section 5.2.3.1) achieved when attempting to predict participant judgements. The best performance achieved an accuracy of ACC = 0.46, again using all the visual features available. This is an initial indication that participants were not using the same visual properties as the algorithms to make their decisions.

Table 5.3: Results for predicting which portal a recipe image belongs to based on different visual feature sets and other factors. Best performing scores for each classifier are bolded. *NB* = Naive Bayes; *LOG* = Logistic Regression; *RF* = Random Forest.

	Accuracy					
	NB		LOG		RF	
	Recipe's Origin	Participants' Judgements	Recipe's Origin	Participants' Judgements	Recipe's Origin	Participants' Judgements
EVF(Brightness)	0.43	0.36	0.41	0.33	0.41	0.34
EVF(Sharpness)	0.41	0.36	0.43	0.37	0.44	0.36
EVF(Contrast)	0.37	0.34	0.37	0.34	0.35	0.34
EVF(Colourfulness)	0.41	0.34	0.40	0.34	0.40	0.34
EVF(Entropy)	0.38	0.36	0.38	0.36	0.39	0.36
EVF(RGBContrast)	0.37	0.34	0.38	0.35	0.37	0.35
EVF(Sharpness Variation)	0.42	0.36	0.43	0.36	0.42	0.37
EVF(Saturation)	0.42	0.32	0.42	0.34	0.41	0.34
EVF(Saturation Variation)	0.39	0.36	0.39	0.34	0.39	0.37
EVF(Naturalness)	0.39	0.36	0.40	0.36	0.40	0.34
EVF(All features)	0.50	0.38	0.56	0.38	0.55	0.38
Colour Histogram	0.37	0.34	0.49	0.36	0.54	0.38
LBP	0.47	0.38	0.50	0.38	0.51	0.39
SIFT	0.57	0.40	0.52	0.39	0.65	0.44
DNN	0.66	0.43	0.82	0.42	0.77	0.45
All Features(Visually)	0.69	0.43	0.85	0.43	**0.84**	**0.46**
Ingredients	0.34	0.35	0.34	0.35	0.34	0.35
Type	0.34	0.35	0.34	0.35	0.34	0.35
Colour	0.35	0.34	0.35	0.34	0.35	0.34
Shape	0.33	0.33	0.32	0.33	0.32	0.33
Container	0.34	0.36	0.34	0.36	0.34	0.36
Eating utensils	0.35	0.36	0.35	0.36	0.35	0.36
Instinct	0.35	0.36	0.35	0.36	0.35	0.36
All Factors	0.34	**0.38**	0.35	0.37	0.35	0.36

Participant Explanations for Labelling Choices The lower part of Table 5.3 demonstrates how classifiers performed using the predefined explanations I provided to participants to justify their performance as features. As can be read from the table, none of these features were helpful, either for predicting origin or the labels participants assigned. Most likely this was because the explanations did not advocate for a specific class, e.g., some utensils (for example, chopsticks) may have indicated Chinese food, whereas others may have been a sign of a western dish. Table 5.4 shows the frequency with which the most common factors and combination of factors were selected by participants to justify the labels they applied. The ingredients featured in the image, type of food and the combination of these two features were the most commonly reported as influencing decisions. These findings underline that although participants were only presented with visual information in the form of an image, the labelling choice was made based on a semantic interpretation of the image content. Moreover, in 127 cases participants reported making decisions based on "instinct", that is, a feeling that the recipe was sourced from a particular recipe platform. Colour and shape — the two obvious visual properties listed — seem to have been supplementary factors, since, as shown in Table 5.4 and Figure 5.7, they were more likely to be chosen with other factors rather than being chosen alone. Factors such as container and eating utensil were selected least frequently, although it is important to note that not every image contained a container or utensil.

Table 5.4: Top-10 factors or combination of factors indicated by participants to have influenced the label applied

Factors	Count	Percentage
Ingredients, Type	226	84%
Type	226	84%
Ingredients	164	61%
Instinct	127	47%
Ingredients, Colour, Type	94	35%
Shape, Type	76	28%
Ingredients, Shape, Type	76	28%
Ingredients, Type, Instinct	75	28%
Ingredients, Colour	62	23%
Type, Instinct	62	23%

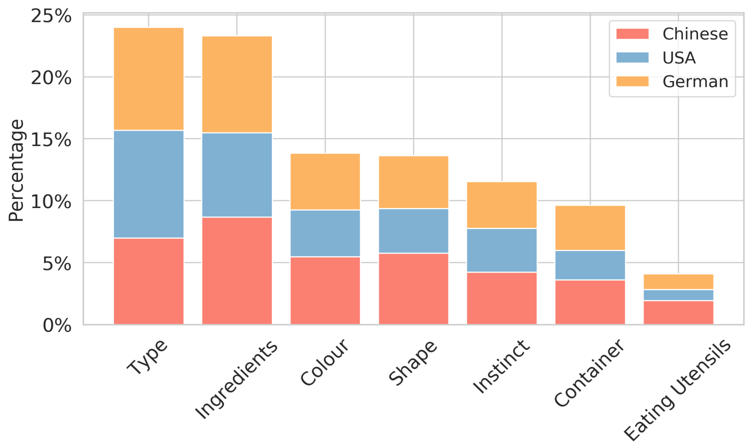


Figure 5.7: The percentage of the frequency with each factor being indicated by participants to have influenced the label applied

Free-Text Explanations Participants were also able to provide additional descriptions to justify their decisions in their own words using free-text comments. A total of 14 participants from China, 33 from the US and 22 from Germany provided 166 such explanations, which were analysed qualitatively in a bottom-up fashion as described above. Duplicate, similar or related responses were grouped together, and the groups were collapsed until a hierarchical structure was formed. The coding scheme for the factor is shown in Table 5.5.

Table 5.5: Coding scheme for factors reported by participants

Categories of factors		N[1]	Description	Examples[2]
Food factors	Adjective	24	Participants left single adjective to describe the food in the recipe image	GE_96:[3] good US_98: healthy
	Style	26	Participants reported how the food looks like in the recipe image	CH_30: Chinese dish is generally not so ugly US_85: Plate design GE_1: Size of the food
	Ingredients	17	Participants reported at least one ingredient they have seen from the recipe image	CH_10: There is rice US_95: The egg on top looks like oriental food. GE_58: Contains coriander and Chili?
	Cooking methods	5	Participants reported how to cook the food in the recipe image	CH_13: Production methods, it's barbecue
Non-food factors	Text	49	Participants reported the letters, characters or water markers, etc. they have seen from the recipes images	CH_42: " 猪肉" is Chinese character US_77: German writing GE_64: Date format: 19.02.2013 is German
	Object/ Background	16	Participants described the objects or setting on the recipe image instead of the food itself	CH_30: Stairs US_55: Newspaper GE_31: Kitchen utensils
	Photo	9	Participants described the photographic and post-processing of the recipe image instead of the food itself	CH_51: A popular filter was used US_72: Angle of the photo, light in the photo GE_39: Bad lightning
	Personal experience	2	Participants reported their own experience with the food in the recipe image	US_5: I know this type of food CH_41: It seems like I've eaten this
	Unknown	18	Participants left comments but offer deficient information	CH_41: It could com from any portal US_3: not sure what type of food that is GE_96: nothing

Note:

[1] Column **N** indicates how many times this kind of factors were reported by the participants.

[2] Column **Examples** indicated the id of participants and the comments they left.

[3] Participant's id comprised by their location (CH:China, US:the US, GE: Germany) and a number.

Two high-level categories were discovered: food-based and non-food-based. Non-food factors included watermarks, commonly used date formats for specific countries, or objects or background aspects surrounding the pictured meals, which helped the participants make judgements.

Both food and non-food factors featured aesthetic dimensions, which may be related to the visual aspects represented in the machine learning features. Comments categorised with *Adjective*, *Style* or *Photo* were somehow related to visual aspects. Several participants described the recipe images aesthetically and treated photography as the basis for judgements, e.g., "Angle of the photo, light in the photo" (US_72). On the other hand, other justifications required abstraction or reflection of the images to derive semantic properties, including what ingredients a meal contains, how it is cooked, how it may taste, whether it is healthy etc. Some participants even reported how their personal experiences with this kind of food influenced the label they assigned. All of these factors underline how the participants' knowledge and background influenced or biased the label they applied.

The free-text comment box was occasionally used by participants to explain their uncertainty. I assigned these cases most often to the category "Text". After examining the images in these cases manually, I found that they all originated either from *Xiachufang* (see Figure 5.8 (a)) or *Kochbar* (see Figure 5.8 (b)). Most of the texts were added with post-processing, as shown in Figure 5.8 (a), the uploaders tagged the recipes with the dish names or their usernames. While the brands on the food packages reveal the information related to recipes' origins, like the images on the left of Figure 5.8 (b), those brands are common in German supermarket but rare in the other two countries. Texts offer concrete information for humans, and as such the accuracy of participants in such cases increased to ACC = 0.94.

(a)

(b)

Figure 5.8: Examples of images with text. **(a)** images with Chinese characters from *Xiachufang.com*. **(b)** images with German Characters from *Kochbar.de*.

Factors Leading to Correct Classification Choices To determine which factors aided participants classify recipes correctly, I developed further logistic regression models. To do so, cases where labels were assigned correctly were given a value of 1 and cases where an incorrect label was given, 0. This value was then used as the dependent variable in the analysis. The predictors (independent variables) were the predefined explanatory factors described above. The results are shown in Table 5.6.

Table 5.6: Logistic regression model of participant judgements

	Dependent variable		
	Correct/Wrong Answer		
	coef(β)	95% CI	OR
Constant	-0.192	[-0.364,-0.020]	0.825
Ingredients	0.069	[-0.085,0.223]	1.071
Type	0.184*	[0.031,0.338]	1.202*
Colour	0.031	[-0.134,0.196]	1.031
Shape	-0.063	[-0.229,0.102]	0.939
Container	0.013	[-0.170,0.196]	1.013
Eating Utensils	0.394**	[0.132,0.657]	1.483**
Instinct	0.008	[-0.163,0.178]	1.008
McFadden R^2		0.004	
Log Likelihood		-1863.5	
AIC		3743	

Note: * $p< 0.05$; ** $p< 0.01$; *** $p< 0.001$.

Only food type and eating utensils were proven to have a significant ($p < .05$) influence on participants' ability to label images correctly. I must acknowledge, however, the fit of the model is not particularly strong, as indicated by the low R^2 value. That being said, when participants reported noticing eating utensils, prediction accuracy increases from ACC = 0.48 to ACC = 0.57. The increase is especially pronounced for recipes from *Xiahucfang* where accuracy increases from ACC = 0.53 to ACC = 0.75. To exemplify why performance increases in such cases, recipes with eating utensils originating from *Xiachufang* are shown in Figure 5.9. These were all classified correctly by our participants; the traditional Chinese eating utensil chopsticks are obvious in the images, which increases the probability of participants labelling correctly.

Figure 5.9: Examples of images with eating utensils from *Xiachufang.com*

In a next step, I investigated whether the same factors had an influence on participant confidence that they were labelling images correctly. For this ordinal regression models are used, one model per collection, the results of which are shown in Table 5.7.

Table 5.7: Ordinal regression models of predicting participant confidence for images associated with each recipe portal

	Dependent Variable								
	Confidence on Xiachufang			**Confidence on Allrecipes**			**Confidence on Kochbar**		
	coef(β)	95% CI	OR	coef(β)	95% CI	OR	coef(β)	95% CI	OR
Ingredients	0.009	[-0.126, 0.145]	1.009	-0.098	[-0.233, 0.038]	0.907	-0.220**	[-0.356, -0.839]	0.803*
Type	-0.294***	[-0.430, -0.158]	0.745***	-0.030	[-0.167, 0.105]	0.970	-0.031	[-0.167, 0.104]	0.970
Colour	0.156*	[0.009, 0.302]	1.168*	-0.147*	[-0.294, -0.000]	0.863*	-0.102	[-0.249, 0.044]	0.903
Shape	0.010	[-0.137, 0.156]	1.010	-0.145	[-0.292, 0.001]	0.865	-0.004	[-0.151, 0.142]	0.996
Container	0.241**	[0.078, 0.405]	1.273**	-0.011	[-0.172, 0.151]	0.990	-0.143	[-0.306, 0.020]	0.867
Eating Utensils	0.365**	[0.123, 0.608]	1.440**	-0.258*	[-0.489, -0.027]	0.772*	-0.177	[-0.413, 0.060]	0.838
Instinct	-0.208**	[-0.306, -0.057]	0.812**	-0.198*	[-0.349, -0.047]	0.820*	-0.093	[-0.245,0.060]	0.912
MacFadden's R2		0.006			0.003			0.002	
Log likelihood		-4256.70			-4248.05			-4233.68	
AIC		8535.41			8518.09			8489.36	

Note: $^{*}p<.05$; $^{**}p<.01$; $^{***}p<.001$.

The first thing to observe is that different features are found to be helpful for different collections. *Type*, *Container*, *Eating Utensils* and *Instinct* were useful predictors for confidence when *Xiachufang* were to be judged; for *Allrecipes*, *Colour*, *Eating Utensils* and *Instinct* were significant features; and, for *Kochbar* only the presence of *Ingredients* was found to be a significant feature.

The only features with positive coefficients, i.e., features that when present increase participant confidence, were found in the model for *Xiachufang*. When a participant reported the presence of a *Container* or *Eating Utensil* on average this increased their confidence in the label applied. The remaining significant features were indicators, which reduced confidence. In other words, acknowledging the presence of certain ingredients in a recipe from *Kochbar* tended to lower confidence in the assigned label on average. I also noted that while the presence of *Eating Utensils* increased confidence for *Xiachufang* recipes, the trend was the opposite for images from both other collections. Moreover, when participants reported making a decision based on *Instinct* in all three collections this resulted in lower confidence ratings on average, which makes sense.

Varying Performance across Participant Groups To understand if participant demographic information influenced their ability to determine the portal from which a recipe originated, I examined how the accuracy of participants' judgements varied on each recipe portal depending on how they answered the post-experiment questionnaire. Table 5.8 presents the results, revealing that participants with different ages and genders behaved differently when judging recipes' origins. Younger participants (< 35) achieved higher accuracy when labelling recipes from *Xichufang* (ACC = 0.59 vs. ACC = 0.49) but they performed significantly worse than older participants in labelling *Allrecipes* (ACC = 0.41 vs. ACC = 0.52). I must interpret the findings regarding age cautiously, though. As the sample age distribution in the samples varies across countries, it is very possible that the effects found relating to age are simply a consequence of participants being best able to identify foods sourced from the portal in their home country. Female participants achieved higher accuracy on *Xiachufang* (ACC = 0.61 vs. ACC = 0.51), while they underperformed compared to male participants on *Kochbar* (ACC = 0.44 vs. ACC = 0.51).

An additional question invited the participants to share their travel experiences and experiences of each country. This allows me to understand whether the classification decisions participants made varied according to their experience of being in the other countries. Analysing the data revealed that accuracy did not increase as a result of frequent cross-continental travel. People who had lived in a country for a longer time were, however, significantly better able to classify the recipes from the portal of that country. Other observations include that participants who had spent time in China were more accurate when labelling recipes from *Allrecipes*, whereas those with more experience of the US were less accurate when labelling *Xiachufang* images. Less surprisingly, being familiar with the recipe portal influenced the accuracy of judgements. Participants who reported being more familiar with *Allrecipes* provided significantly more accurate judgements on recipes from this portal. Familiarity with *Xiachufang* and *Kochbar*, on the other hand, had no significant influence on the accuracy of images from these portals. Participants unfamiliar with *Allrecipes* and *Kochbar* were better at judging the recipes from *Xiachufang*.

Participants who reported being interested in food or recipes from foreign cultures achieved higher accuracy overall. Similarly, those participants who reported trying food from other cultures were also more accurate in the labelling task.

The analyses in this section have shown that it was not only the participants' culture that influences the labels that they applied. Individual traits and personal experience also played a role in the labels that were assigned, and the accuracy achieved.

5.2.4 Summary and Discussion

5.2.4.1 The Primary Findings

- *RQ1.* The classification results of algorithms demonstrate that images of food uploaded to recipe portals in China, US, and Germany are visually distinct. The recipe images derived from the recipe portals representing these food cultures can be classified by the visual features. Almost all the image features tested provided some useful signal for this food classification task, the strongest being provided by DNN. Overall, the images from the Chinese recipe portal were labelled most accurately, with recipe images from the US and German portals more likely to be confused.

Table 5.8: Pairwise comparison of accuracy from different groups based on demographics. Only attributes with significant results are reported.

	Overall Accuracy Mean(+/- std)	**Accuracy on Xiachufang** Mean(+/- std)	**Accuracy on Allrecipes** Mean(+/- std)	**Accuracy on Kochbar** Mean(+/- std)
Gender				
Male	0.49(+/-0.17)	0.51(+/-0.29)	0.44(+/-0.28)	0.51(+/-0.30)*
Female	0.50(+/-0.18)	0.61(+/-0.28)**	0.46(+/-0.28)	0.44(+/-0.31)
Age				
Age < 35	0.50(+/-0.18)	0.59(+/-0.29)**	0.41(+/-0.27)	0.50(+/-0.30)*
Age ≥ 35	0.48(+/-0.17)	0.49(+/-0.29)	0.52(+/-0.27)***	0.50(+/-0.30)
Experience of each Country (China)				
Never visited - been there a few times	0.49(+/-0.17)	0.51(+/-0.29)	0.47(+/-0.27)*	0.49(+/-0.29)
Visit regularly - permanent resident	0.50(+/-0.18)	0.63(+/-0.28)***	0.41(+/-0.29)	0.45(+/-0.31)
Experience of each Country (The US)				
Never visited - been there a few times	0.49(+/-0.18)	0.61(+/-0.29)***	0.39(+/-0.28)	0.49(+/-0.31)
Visit regularly - permanent resident	0.48(+/-0.17)	0.47(+/-0.27)	0.53(+/-0.26)***	0.46(+/-0.30)
Experience of each Country (Germany)				
Never visited - been there a few times	0.48(+/-0.18)	0.56(+/-0.27)	0.46(+/-0.28)	0.43(+/-0.31)
Visit regularly - permanent resident	0.50(+/-0.17)	0.55(+/-0.31)	0.43(+/-0.28)	0.54(+/-0.29)***
Familiarity with each recipe portal (Xiachufang.com)				
Not familiar(≤ 2 on likert scale)	0.51(+/-0.17)**	0.55(+/-0.29)	0.46(+/-0.28)	0.52(+/-0.29)***
Familiar(≥ 3 on the likert scale)	0.46(+/-0.17)	0.57(+/-0.31)	0.42(+/-0.28)	0.39(+/-0.31)
Familiarity with each recipe portal (Allrecipes.com)				
Not familiar(≤ 2 on likert scale)	0.50(+/-0.17)	0.62(+/-0.28)***	0.40(+/-0.28)	0.50(+/-0.29)
Familiar(≥ 3 on the likert scale)	0.48(+/-0.17)	0.48(+/-0.28)	0.50(+/-0.27)***	0.46(+/-0.31)
Familiarity with each recipe portal (Kochbar.de)				
Not familiar(≤ 2 on likert scale)	0.50(+/-0.17)	0.58(+/-0.28)*	0.44(+/-00.28)	0.48(+/-0.30)
Familiar(≥ 3 on the likert scale)	0.48(+/-0.18)	0.50(+/-0.32)	0.46(+/-0.28)	0.48(+/-0.31)
Interests in food from foreign cultures				
Not interested(≤ 2 on likert scale)	0.41(+/-0.23)	0.46(+/-0.28)	0.33(+/-0.33)	0.45(+/-0.39)
Interested(≥ 3 on the likert scale)	0.50(+/-0.17)*	0.56(+/-0.29)*	0.46(+/-0.27)*	0.48(+/-0.30)
Interests in recipes from foreign cultures				
Not interested(≤ 2 on likert scale)	0.45(+/-0.23)	0.50(+/-0.27)	0.37(+/-0.33)	0.47(+/-0.34)
Interested(≥ 3 on the likert scale)	0.50(+/-0.17)*	0.56(+/-0.29)	0.46(+/-0.27)*	0.48(+/-0.30)
Frequency of trying recipes from other cultures				
Once per month	0.48(+/-0.18)	0.58(+/-0.29)*	0.41(+/-0.28)	0.46(+/-0.29)
Once per month CT: the same once per month	0.50(+/-0.17)	0.52(+/-0.29)	0.49(+/-0.27)**	0.50(+/-0.32)**

Note: * $p< .05$; ** $p< .01$; *** $p< .001$.

- *RQ2.* Humans are less able to perceive the visual difference of food sourced from Chinese, US, and German recipe portals. The evidence suggests that, unlike the machine learning approaches, humans abstract or interpret the visual features to derive semantic features, such as the ingredients a meal contains or how it may taste. As this process is based on personal knowledge or experience, the act of classification becomes biased, which evidently negatively influences accuracy. When humans made classification errors, however, the trend in their mistakes was the same as for the machine learning approach. Specifically, for the participants, the recipes originating from the Chinese recipe portal are the easiest to be identified based on the images, while recipes from the US and German portals tend to be confused. The confidence levels associated with the labels applied to confirm that the participants were aware of this trend.

- *RQ3.* The collected data from the user study shed the light on which factors influence participants' judgement on the origins of food. In general, participants classified recipes from the portals of their culture more accurately and confidently than the participants from other cultures. The evidence suggests that familiarity with the food cultures and the recipe portals play a role in the process. In addition, participants reported several features of the images as being influential when making their decisions although some justifications were more useful than others. For example, the obvious visual clues in the recipe images, such as colour, and shape of food were less important than the ingredients present and the type of dish. The results show that if the participants recognised the dish type from the image, it was more likely that they made the right choice. Moreover, participants were able to improve their performance by identifying factors in the image which had nothing to do with the food itself but offered discriminative power. Eating utensils, such as cutlery or chopsticks, or text being present in the image were prominent examples. These semantic interpretations of the recipe images were biassed by human personal knowledge and experience, such as the experience and interests of the relevant culture (e.g., residence or travel), and familiarity with the recipe portals.

5.2.4.2 Implications of Study I

The algorithmic results have shown that the food from different cultures is sufficiently visually distinct such that they can be differentiated by a combination of visual features automatically extracted from the recipe images.

The Chinese-sourced recipe images were more visually distinct than those from US and German portals, which were harder to distinguish visually. These findings are in line with those from previous work, such as in (Sajadmanesh et al., 2017), in which the food originated from North America and Western Europe was easier to be misclassified than that from Chinese when ingredients were used to train the models.

The findings from participants' responses to the questions on the user study underlined that the way people perceive images of food differs fundamentally based on different factors. The primary factor revealed was the participant's country of residence, which was discovered in this study to directly influence the labels applied to images. While this study did not study food preference directly, the findings do have consequences for the development of food recommendation systems since familiarity with food - and visual familiarity in particular - is strongly related to food preference (Aldridge et al., 2009; Heath et al., 2011). The foods people find desirable - and to what extent they are willing to try something new - are tightly bound to their cultural upbringing and to physical and emotional reactions to food experiences in the past (Aldridge et al., 2009), but also depend on individual traits, such as openness to experience (Tan et al., 2016), which are exactly corresponding to the demographic factors shown in this study, familiarity with the recipe portals and interests in food and recipes from foreign cultures.

This reinforces the need for food recommender systems to model and account for contextual variables when making personalised food recommendations, and culture should be taken as a dimension of them. There have been several successful examples of incorporating culture with other demographic factors into music recommendations (Schedl et al., 2015), but no equivalent research exists for the recommendation of food.

The results of Study I have shown how visual cues differ across culture and how participants' cultural backgrounds influence their judgement in food origin classification tasks.

However, in order to develop a cultural-aware food recommender system, it is important to know what people from different cultures would like to eat. Building on this, Study II in the next section will explore whether similar cross-cultural differences are present when users apply subjective labels to recipes. A similar experimental setup is planned to, firstly, investigate the food visual preference within and across cultures with machine learning approaches, the same visual feature sets as in this study will be exploited; and secondly, another online survey will be employed to collect data on participants' subjective impression of recipe images (e.g., their attractiveness, how willing the participants are to cook and eat them, etc.). This would complement the findings on the visual aspects of food presented in Study I nicely, and would offer concrete utility with respect to the design of food recommendation systems.

5.3 Study II: Predicting Food Preferences Based on Visual Information

5.3.1 Study Outline

In previous work (Elsweiler et al., 2017; Trattner et al., 2018), visual properties such as EVF has been proven to be able to predict user food preferences. The evidence suggests that preferred foods are visually distinct from non-preferred foods. In (A. D. Starke et al., 2021), the authors found preferred food images to differ in terms of EVF. However, the previous work was limited to specific food cultures, meaning that whether food preferences differ across cultures remains unclear. Study II is designed to address this problem. In addition, in order to suggest appealing food for people in the situation where they move between cultures (e.g., when travelling or emigrating), this study is anticipated to ascertain commonalities in visual food preferences across cultures. With the aim in mind, the study starts by algorithmically identifying food preferences within each food culture with visual features applied in Study I, before a transfer learning strategy is applied to determine stable patterns of visual food preferences across cultures. Two research questions are raised to guide the experiments:

- *RQ1a.* To what extent is it possible to use visual information to differentiate the appreciated and less appreciated recipes within each food culture with machine learning approaches?
- *RQ1b.* To what extent is it possible to ascertain stable patterns of visual food preferences across cultures using the same machine learning approaches?

Moreover, a user study is designed and performed, with the aim of investigating how participants from different food cultures (i.e., China, US, and Germany) perceived food images across recipe portals subjectively. The response from participants is used to validate the models derived during the analyses with machine learning approaches but can, moreover, answer the following questions:

- *RQ2a.* Which factors, derived from recipe images, influence the food choices of human users in different food cultures?
- *RQ2b.* Is it possible to ascertain the visual features that help identify food preferences across food cultures?

5.3.2 Methods

5.3.2.1 Intra and Inter - Cultural Food Preferences Prediction Based on Visual Features by Means of Machine Learning Approaches

To determine if it is possible to distinguish between appreciated and less appreciated recipes in each collection (*RQ1a*), I formulated the problem as a binary classification task. The data basis for this study were formed by the sample of 5,000 images from each collection, of which 2,500 from the top-10% and 2,500 from the bottom-10% based on the appreciation metric (as described in Chapter 4). Firstly, each image was represented by extracted automatically visual features. Here I applied the same features used in Study I, which include **Explicit Visual Features (EVF)** - Brightness, Sharpness, Contrast, Colourfulness, Entropy, RGB contrast, Variation in Sharpness, Saturation, Variation in Saturation and Naturalness, **VGG16 embeddings**, **LBP**, **Colour Histogram** and **BoVW** calculated from descriptors of **SIFT** (Scale Invariant Feature Transform). Each recipe image was transformed into a 5,144-dimensional vector. Using these representations the classifiers were trained for recipes from each collection with the appreciation metric used as the target value. I trained the classifiers with each feature set individually and eventually with all features combined, three supervised machine learning approaches: **Naive Bayes (NB)**, **Logistic Regression (LOG)** and **Random Forest (RF)** were applied. In all experiments the data was split randomly for training (70%) and testing (30%), and the Randomized Search CV was utilized on the training sets to determine the optimal parameters with a 5-fold cross-validation for LOG and RF. The prediction accuracy (ACC of the classifiers was measured to indicate the performance of the classifiers. After that, in order to answer *RQ1b*, I identified the classifiers with the highest ACC for each collection and tested their predictive ability on the other two collections. The process is shown in Figure 5.10. This offers a chance to establish if the same visual features are effective across the respective cultures.

5.3.2.2 Design of the User Study

To further validate the prediction task findings and determine if the patterns identified are representative of cross-cultural visual food preferences beyond the food-portal users, a controlled online study was designed and performed. 150 participants from each of the three countries, recruited via crowd-sourcing platforms and social media, were asked to rate various aspect of recipe images from all three platforms.

Study Design The study consisted of three parts. In the first, participants were tasked with ranking sets of three recipe images (one from each portal selected based on the classifier outputs - see below) that were presented side by side according to how appealing they found the recipes. I refer to these groups as triplets. Participants were able to rank recipes via a drag and drop interface as illustrated in Figure 5.11. After ranking the images, 19 follow-up questions were answered to learn about how participants justified the rankings they derived. This process was repeated for three sets of recipe images (triplets). The final part of the study invited participants to complete a questionnaire capturing demographic and other background information.

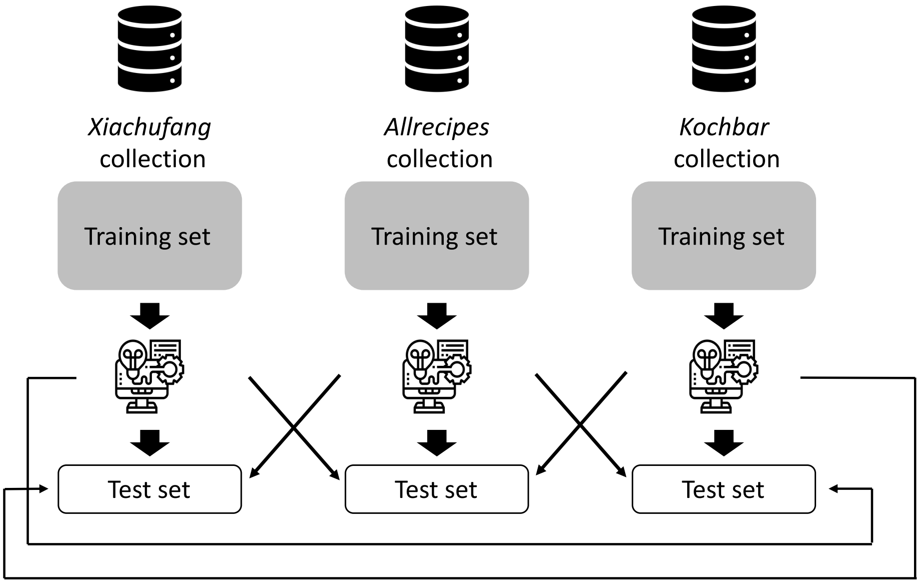


Figure 5.10: The process of applying the best performed classifiers for each collection and testing their predictive ability on the other two collections. Taking the classifier trained on the *Xiachufang* collection for example, the classifiers was trained on the training set of *Xiachufang* collection, then it was applied to predict whether the recipes are appreciated or less appreciated in the test sets of *Allrecipes* and *Kochbar* collections.

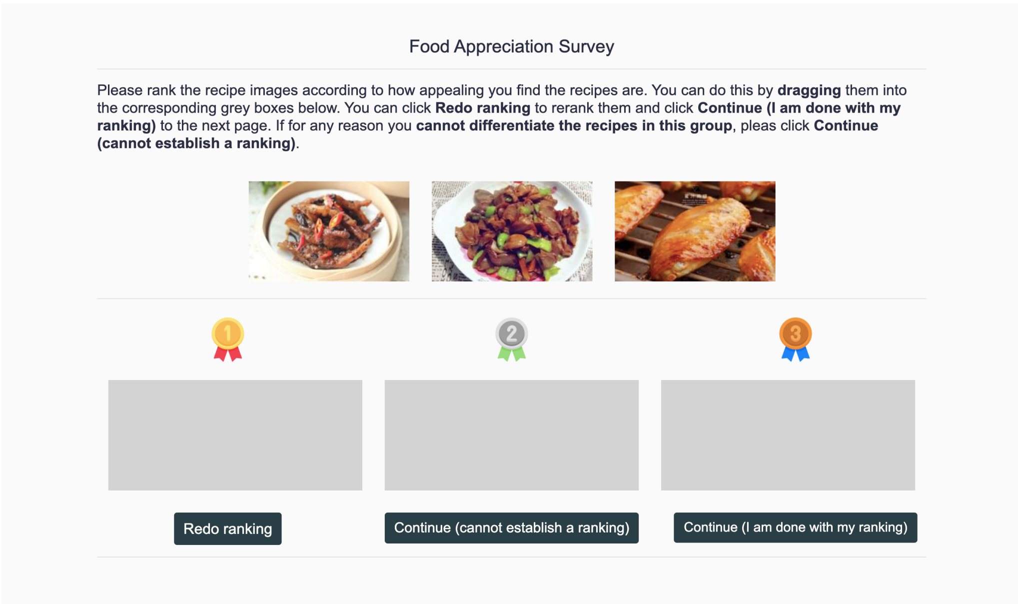


Figure 5.11: Screenshot of an example task from the user study in Study II

Sampling Strategy for The Recipe Images As described above, recipes were shown to the participants in groups of three. The recipe images used in the study were selected from the top-10% and bottom-10% of recipe images in terms of appreciation from each recipe portal, which were used to train the classifiers. Moreover, the groups of recipe images had to meet 2 further conditions. To minimise the influence of other, non-visual factors on participant judgments, I grouped recipes that were from the same category

(Main Dish) and were similar in terms of the contained ingredients. Using the same approach as in (Elsweiler et al., 2017), recipes were represented by TF-IDF weights of ingredients, whereby the cosine similarity between every two of them were calculated. Then recipe pairs with similarity was within top 1% were chosen to form a triple. This step ensured that the recipes to be ranked were comparable. That is, the main ingredients were the same e.g., chicken and noodles.

The final condition for grouping related to the outcome of the classifiers. The best performing appreciation classifiers for each culture were applied to the images, leading to three predicted appreciation labels for each recipe image (i.e., one for each culture). To achieve sufficient variation in our data I removed recipe triplets where all three recipes assigned the same label by any one of the classifiers. I, moreover, filtered the triplets, such that no recipe image featured in multiple groups. Eventually there are 30 triplets sampled from each recipe collection, leading to 270 recipe images (3 images * 30 triplets * 3 recipe portals) for the user study. Finally, to achieve stable rankings for the triplets for each food culture, the experiment was designed such each was ranked by five participants from each country.

Questionnaire After ranking each recipe image triplet the participants were required to answer 19 questions that offer insight into their decisions. These were inspired by the literature on the factors influencing human food choice and could be observed from the recipe images. I grouped the factors into four key dimensions as four main factors: appearance, perceived flavour, health and familiarity. More detailed sub-dimensions questioned participants on several supplementary factors are shown in Figure 5.12 and justified below:

- **Appearance** - The appearance of food is well established to influence food choice. I focus on three prominent aspects in this study: Colour, presentation and texture. **Colour** - People are known to form preference of food with certain colours. For example, food with more red brightness (Foroni et al., 2016) or with chromatic colours (Lee et al., 2013) tends to be preferred. Yellowy and brown foods are often perceived negatively (Palmer & Schloss, 2010). Moreover, colours are associated with expectation of flavors (Spence et al., 2010). For example, white food is considered to be salty, while green food is perceived to be sour (Wan et al., 2014). Some principles of colour mixture have been shown to be stable across food cultures. For example, in (Zampollo et al., 2012), humans prefer dishes containing 3 or 4 mixing colours; yet preferences for colour palettes vary across cultures. It is shown in that the typical preferred colour palette of an Austria's combines brown and beige, in Britain it tends to be red, white and green (Pyszne, 2021). **Presentation** - The way food is presented on a plate, referred to as plating, can influence how food is perceived in terms of attractiveness (Michel et al., 2015). For example, dishes presented in a neat and balanced manner, are perceived to be tastier and more expensive (Michel et al., 2014). Beyond the food itself, the tableware utilized can be influential with respect to perceived aesthetics quality of food (Piqueras-Fiszman et al., 2012). **Texture** - The make up of food, such as proteins, carbohydrates, water and oils, determine the food texture (Vilgis, 2013), which can be described as hard, soft, rough, creamy, crumbly, crispy etc. Such characteristics can be perceived in images of food (J. Chen, 2007) and offer people a means to prejudge how the food will feel in their mouths (Martınez et al., 2002; Montouto-Graña et al., 2002). **Quality of image** - As the recipes and recipe images are user provided on the food portals, their quality in terms of lighting and composition various considerably. Image quality can certainly contribute to the way an image is perceived (Ding et al., 2019; Khosla et al., 2014). For this reason, I asked the participants to evaluate the quality of images in terms of resolution and lightning.

- **Perceived taste** - Taste has been stated to be the most prominent determinant of human food choice (Ahn et al., 2011; Liem & Russell, 2019). Humans instinctively pursue food that they perceive to be tasty (Chamoun et al., 2018). As with texture, visual cues allow the taste of food to be predetermined. For instance, certain colours are linked to specific tastes, such as the relationship between redness and sweet, ripened foods (Maga, 1974). In this user study, participants were queried with respect to their perception of four basic tastes by asking them to judge how salty, sweet, sour and bitter they expect the dish in the image to be. These four basic tastes were noted by Aristotle 2,000 years ago but I utilise these on account of their reliability and continued wide use in food science (N.-E. Choi & Han, 2015). I decided to exclude umami, the fifth basic flavour (Lindemann et al., 2002), since people have difficulty recognising this dimension and it is regularly confused with salty in European countries (Cecchini et al., 2019) and the US (Ninomiya, 2015). Instead, I included spiciness, a stimulating oral sensation (Spence, 2018), in the survey. Spicy food has the additional advantage that it can be found in diverse cuisines with spices causing "spiciness", such as capsicum being used widely all over the world (Spence, 2018).

- **Perceived health** - Food is necessary to sustain human life and although a healthy diet is an efficient way to minimise disease and maximise vitality (Organization et al., 2019), for diverse reasons humans often tend to prefer fatty and calorific meals, which are often perceived to as unhealthy foods (Elsweiler et al., 2017). Again, visual attributes can be indicators of the healthiness of food. For example, food colour has been found to be a good signal for perceiving nutrients, such as foods with darker colours are perceived to be nutritive (Bender, 1981). However, due to the wide application of artificial colours on food, determining the nutritional contents relying solely on colour is unreliable (Foroni et al., 2016). That being said, from images alone, people are not particularly effective in predicting the nutritional properties of food (Elsweiler et al., 2017). To determine to what extent visual-perceived nutrients influence human food choices, I questioned users about their perception of several macronutrients. I chose energy, fat, sugar, fibre and protein which form the basis of are of World Health Organization (WHO) and the UK Food Standards Agency (FSA) standards and have been used in previous recipe evaluations (Elsweiler et al., 2017; Howard, Adams, & White, 2012).

- **Familiarity** - Exposure to specific foods in childhood can prevent food neophobia and enhance human acceptance and willingness to try new foods (Aldridge et al., 2009). People prefer to eat food originating from their own or similar cultures (Silva et al., 2014; Sajadmanesh et al., 2017). However, the attitude towards food from foreign cultures can be positively influenced via increased exposure (Seo et al., 2013). People from different cultural backgrounds tend to react differently when shown familiar food (Torrico et al., 2019). For these reasons, I questioned participants on their familiarity with the dish in the images.

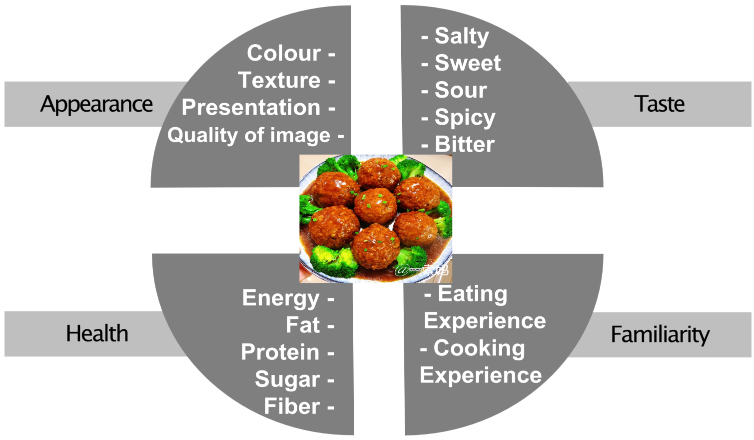


Figure 5.12: The main factors and their corresponding supplementary factors

Participants I recruited the participants from China, US and Germany by means of major survey and crowd-sourcing platforms in the different countries. To recruit participants from the US I used Amazon Mechanical Turk (MTurk)[5]. To help ensure serious participation, I only invited the workers located in US at that time with more than 98% 'HIT accept rate'. German participants were recruited from another crowd-sourcing platform, Prolific[6]. Again, to ensure reliable performance, I restricted participants to those with more than 95% approval rate in Germany. To attract Chinese participants, the study was translated into Chinese (see Appendix D) and deployed via the platform Tencent survey[7]. Additional advertising was done on Chinese social media channels including Wechat and Douban[8]. The participants were compensated financially for their participation in line with the guidelines and norms for the platforms. MTurk participants received 1 US dollar, prolific participant received 1.25 Euro and Chinese participants received a 5-Yuan Wechat red packet. Using these methods I was able to recruit 150 participants from each country. The distribution of the participant age and gender is shown in Figure 5.13. Most of our participants in China and Germany are between 25-34 years old, while in the USA, the participant's age is slightly older. The situation is similar to that of the user study in Study I and results from the same reason. The number of male participants is more than the female in the USA and Germany, while the Chinese sample is more gender balanced.

Analysis of Survey Outcomes I analysed the collected image rankings and questions answers with the joint aims of identifying stable patterns in cross-cultural visual food preference and identifying influential factors in mind.

- **Factors influencing human rankings** To verify to what extent factors associated with the 4-dimensions (appearance, taste, health and familiarity) influenced how participants ranked the recipe images, I built ordinal regression models. A one-way ANOVA and a Tukey's HSD post-hoc test was performed to identify the most influential factors. Moreover, I derived ordinal regression models for each factor that was found to significantly influence the ranking. This allows me to identify supplementary factors.

[5]https://www.mturk.com
[6]https://www.prolific.co/
[7]https://wj.qq.com/
[8]https://www.douban.com

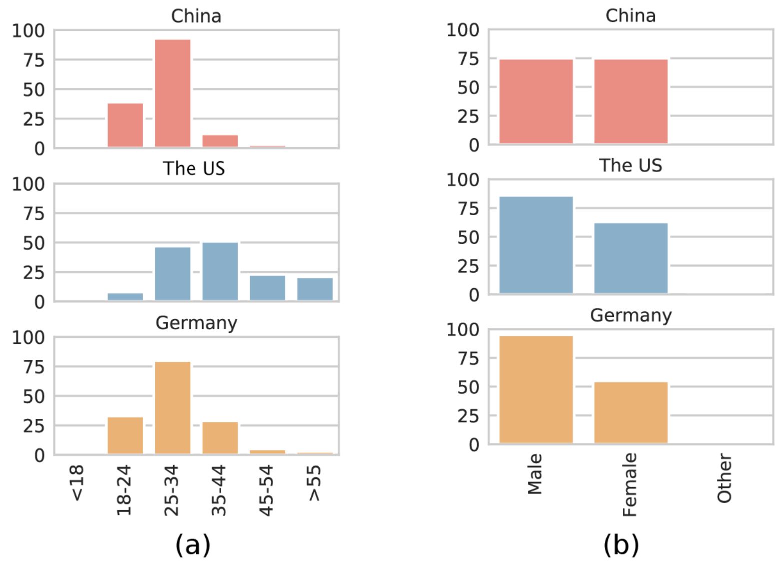


Figure 5.13: Participant demographics in the user study of Study II. **(a)** Age distribution of participants from each country. **(b)** Gender distribution of participants from each country.

- **Selecting appreciated and less appreciated recipe images across cultures** Each triplet of recipe images was ranked by 5 participants from each country. Rank positions were transformed to scores, i.e., 1st rank = 3 points, 2nd rank = 2 points, 3rd rank = 1 point. These scores were then averaged for each recipe, grouped by the participants' country of origin. For each triplet, the recipe image with the highest average score as appreciated and the one with the lowest score was not appreciated. There were triplets, for which images received the same average scores. If there were two images with the same score which was higher than the remaining recipe, then these two images were both marked as appreciated, while the other one marked as less appreciated. In the case where triplets containing two images with same score that was lower than the remaining image, then these were taken as less appreciated and the other was marked as appreciated. Triplets where all three images were scored equally (n=1) were not used in the final analysis. Taking this approach allowed me to create a dataset of recipes that were judged to be either appreciated and not by participants from three food cultures. As a next step, I compared the score of each recipe images across cultural contexts and selected those which were appreciated and not appreciated by all three food cultures.

- **Comparative analyses of appreciated and less appreciated recipes across food cultures** After establishing this data basis, I used various techniques to learn about the visual characteristics of food preferences across different cultures. Specifically, I represented each recipe as a vector consisting of the same automatically extracted visual feature as used in the machine learning tasks, and generated image grids in a 2D space by means of t-SNE, a popular approach used for visualizing images in a low-dimensional space, such as in (Yang et al., 2015).

  In a next step, I established if the visual features, applied in the prediction task, could help describe the cross-cultural visual food preferences. A comparative analysis began with a wilcoxon singed-rank test on the EVF features. While for other high-dimensional visual feature sets, including VGG16, Colour Histogram, LBP and BoVW based on SIFT, they were used to calculate how similar the appreciated and less appreciated recipe images across cultures, respectively. The similarity of recipes were then compared to see to what extent human visual food preferences align across cultures. Different methods for calculating the images similarity were applied for the different feature sets. For VGG16 and BoVW, I used cosine similarity, which is widely used in image retrieval, such as in (D. Zhang & Lu, 2003). For the histogram-based features, such as Colour Histogram and LBP, I measured the similarity by comparing the distribution of the histograms. The calculation was done by applying *compareHist* in OpenCV[9]. In addition, to understand whether the visual appearance of preferred recipes display a more similar pattern, I compared the similarity of images within groups of each other. The groups studied included: (a) the identified appreciated recipes across cultures, (b) the identified less appreciated recipes across cultures, (c) recipe images in (a) and (b), (d) all candidate recipe images in this user study.

5.3.3 Results

5.3.3.1 Identifying Intra- Cultural Visual Food Preference with Machine Learning Approaches (RQ1a)

The first analysis investigates the extent to which it is possible to differentiate appreciated and less appreciated recipes using the extracted visual features within each collection. The performance of all classifiers in Table 5.9.

Table 5.9: Result for identifying the appreciated and less appreciated recipes in each recipe portal with different visual feature sets. Best performing scores for each classifier are bold. *NB* = Naive Bayes; *LOG* = Logistic Regression; *RF* = Random Forest.

Xiachufang	Mean Accuracy			**Allrecipes**	Mean Accuracy			**Kochbar**	Mean Accuracy		
Features	NB	LOG	RF	Features	NB	LOG	RF	Features	NB	LOG	RF
EVF(Brightness)	.55	.54	.56	EVF(Brightness)	.52	.50	.52	EVF(Brightness)	.50	.49	.50
EVF(Sharpness)	**.59**	.60	.61	EVF(Sharpness)	.52	.52	.53	EVF(Sharpness)	.52	.49	.50
EVF(Contrast)	.55	.54	.54	EVF(Contrast)	.53	.49	.55	EVF(Contrast)	.51	.50	.50
EVF(Colourfulness)	.53	.53	.52	EVF(Colourfulness)	.55	.54	.58	EVF(Colourfulness)	.50	.51	.49
EVF(Entropy)	.51	.48	.51	EVF(Entropy)	.52	.56	.56	EVF(Entropy)	.49	.48	.49
EVF(RGBContrast)	.55	.55	.55	EVF(RGBContrast)	.53	.50	.54	EVF(RGBContrast)	.51	.50	.53
EVF(Sharpness Variation)	**.59**	.62	.61	EVF(Sharpness Variation)	.52	.50	.53	EVF(Sharpness Variation)	.50	.49	.49
EVF(Saturation)	.57	.59	.58	EVF(Saturation)	.54	.55	.56	EVF(Saturation)	.51	.51	.50
EVF(Saturation Variation)	.47	.47	.51	EVF(Saturation Variation)	.55	.55	.55	EVF(Saturation Variation)	.48	.49	.51
EVF(Naturalness)	.54	.56	.56	EVF(Naturalness)	.55	.56	.56	EVF(Naturalness)	.49	.52	.51
EVF(All features)	**.59**	**.64**	.64	EVF(All features)	.56	**.59**	.58	EVF(All features)	.50	.51	.53
Colour Histogram	.53	.61	.63	Colour Histogram	.51	.58	.62	Colour Histogram	.50	.52	.54
LBP	.57	.60	.59	LBP	.56	.58	.58	LBP	.50	.51	.51
BoVW	.54	.51	.55	BoVW	.57	.53	.60	BoVW	.53	.54	.55
VGG16	**.59**	.62	.66	VGG16	.58	.56	.64	VGG16	**.55**	**.57**	**.60**
All Features	**.59**	.62	**.67**	All Features	**.59**	.56	**.65**	All Features	**.55**	**.57**	**.60**

The bottom row in Table 5.9 demonstrates that the classifiers are able to identify appreciated and less appreciated recipes in the three recipe portals with an average accuracy of 64%. This underlines that visual information extracted from recipe images can help explain which recipes are preferred for each food culture. Comparing the classifiers with the highest accuracy on each data collection reveals that the accuracy of *Kochbar* (ACC = .60) collection is slightly lower than that of *Xiachufang* (ACC = .67) and *Allrecipes*

[9]`https://docs.opencv.org/3.4/d8/dc8/tutorial_histogram_comparison.html`

(ACC = .65), suggesting that the visual characteristics of appreciated and less appreciated recipes in *Kochbar* are not so pronounced as in the other two recipe portals.

Almost all visual feature sets show positive predictive power for this task (ACC $\geq$ .50). Notably, when compared with other individual visual feature sets, the VGG16 embeddings provide the highest accuracy for the classifiers in all collections. Moreover, VGG16 embeddings achieve similar performance to all features when combined in a single model. Other feature sets exhibit varying abilities across the collections. For example, the performance of combined EVF is higher on *Xiachufang* collection than on the others. Colour Histogram offers an accuracy close to that of VGG16 embeddings on *Xiachufang* and *Allrecipes* collections, but on *Kochbar*, there is a relatively large gap in the accuracy achieved by these two feature sets. The performance of the remaining features, LBP and BoVW, show no significant differences across the collections, while these do not achieve as strong a performance as VGG16 embeddings, these features perform better than one would expect by random chance, i.e., the features do contain a useful signal.

Overall, it is possible to distinguish between the appreciated and less appreciated recipes in all three recipe portals with the feature sets extracted from the recipe images. However, the accuracy of the classifiers in this study suggests that predicting appreciated recipes with visual features is a more challenging task than identifying the food origin of the recipe as shown in Study I.

5.3.3.2 Identifying Stable Patterns of Visual Food Preferences Across Cultures (RQ1b)

Building on the outcomes of the previous analyses, I selected the classifier with the highest accuracy for each collection, and tested its ability to identify appreciated recipes in the other two collections. Figure 5.14 shows the results. The results show that the *Xiachufang* classifier performs well only when classifying the recipes in its own collection, but very poorly (worse than random) in the other two collections. The *Allrecipes* and *Kochbar* classifiers, in contrast, both perform best when they are used on their own recipe images, perform reasonably on each others' recipe images (better than random), but perform poorly on the recipes sourced from *Xiachufang*.

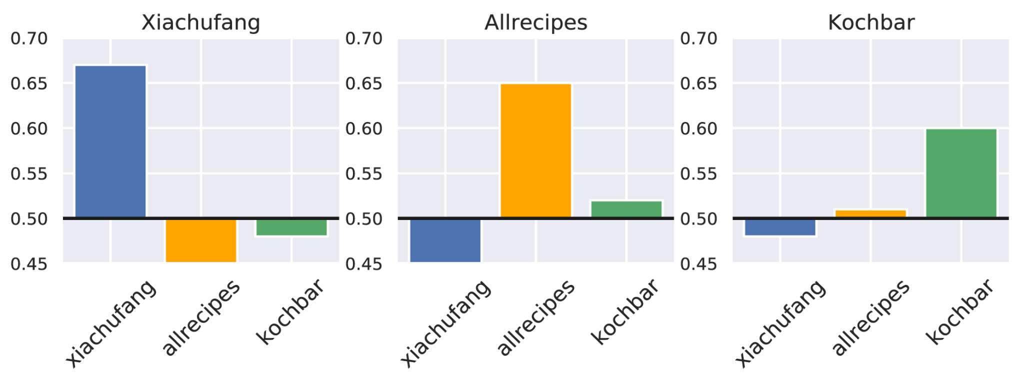


Figure 5.14: The classifier with the highest accuracy for each collection and their performance on the other two collections

These results hint that the same visual signals can be applied across cultures, at least between the *Allrecipes* and *Kochbar* collections. The visual food preference learned from these two collections to *Xiachfuang* do not transfer. These findings align with the results of recipe origins prediction task revealed in study I, which shows recipes in *Xiachufang* to be more visually different compared to those in the other two recipe portals.

In summary, I found the possibility of ascertaining stable patterns of visual food preferences across cultures. In the next step, I designed a user study to investigate whether the overlaps of visual food preferences exist based on participant perception of the recipe images. In addition, I also investigated which factors derived from recipe images influencing food choices, and whether any of these factors apply across cultures.

5.3.3.3 Factors Derived from Recipe Images that Influence Participant Food Preferences (RQ2a)

In this section, I explore how factors derived from the recipe images influence food preferences as determined by the rankings applied to recipe images. I started by analysing participant responses to establish how the four key dimensions, i.e., appearance, taste, health and familiarity, influenced the rankings. In a next step, I identified the most significant factors among the four main factors, and examined how these interact with their corresponding supplementary factors. For example, whether participant perception of the colour, presentation, texture of food and the quality of food images influence their appreciation of the visual appearance of food? This offers the potential to see whether the detailed features derived from the recipe images also apply across cultures.

The Influence of the Four Main Factors on the Ranking of The Recipe Images Table 5.10 presents the association between the Likert-scale values for the four main factors and the rankings of the recipe images. This was done by means of an ordinal regression model. Appearance and taste are observed to be the two consistently significant factors in influencing the recipe ranking, irrespective of the participants' cultural background. As one would expect, the coefficient indicates that recipes reported to be visually appealing and tasty were more likely to be ranked higher. Health, in contrast, had a negative impact on the rankings provided. The influence, however, is only significant ($p < 0.01$) for participants from US and Germany. This finding aligns with past food recommender research (Elsweiler et al., 2017; Harvey & Elsweiler, 2015), which uncovered overall preferences for unhealthier (mainly fattier) food. Lastly, the results show that irrespective of food culture, experience with eating or cooking the depicted foods did not influence the rankings provided.

Table 5.10: Association between participants' rankings on recipe images with the four main factors

	Dependent variable								
	Ranking from the Chinese participants			Ranking from the US participants			Ranking from the German participants		
	coef(β)	95% CI	OR	coef(β)	95% CI	OR	coef(β)	95% CI	OR
appearance	-0.87***	[-1.02,-0.73]	0.42***	-0.59***	[-0.72,-0.46]	0.55***	-0.99***	[-1.13,-0.86]	0.37***
tastes	-0.65***	[-0.82,-0.48]	0.52***	-0.65***	[-0.79,-0.50]	0.52***	-0.41***	[-0.56,-0.26]	0.67***
health	0.31	[-0.10,0.16]	1.03	0.21***	[0.10,0.33]	1.24***	0.13*	[0.00,0.26]	1.14*
experience of eating	0.02	[-0.11,0.16]	1.02	0.06	[-0.10,0.22]	1.06	-0.13	[-0.32,0.05]	0.88
experience of cooking	0.012	[-0.11,0.15]	1.02	0.02	[-0.13,0.16]	1.02	-0.02	[-0.19,0.16]	0.98
Mc Fadden R^2		0.23			0.18			0.23	
AIC		2292.84			2455.82			2287.18	

Participants' self-reported reliance on the factors to rank the recipe images are shown in Figure 5.15 (a). It was found that, the participants, regardless of their cultural background, reported appearance and taste of food as the most influential factors when they are ranking the recipes. Furthermore, the result of the statistical analysis with one-way ANOVA and post-hoc test (Figure 5.15 (b-d)), revealed that there was no significant difference between participants' reported influence of appearance and taste. This suggests that participants believe the appearance and taste of food play an equally important role in ranking recipes, and this applies across cultures. Interesting result can be found when relating the outcomes of the regression model to participants' self-report. For example, appearance and taste did influence participant rankings on the recipes and they are aware of it. On the other hand, participants claimed familiarity influenced their choices, but in fact, whether they are familiar with the food did not influence their rankings, significantly. However, the perceived healthiness significantly influences the rankings from US and German participants, but they do not really realise its impact.

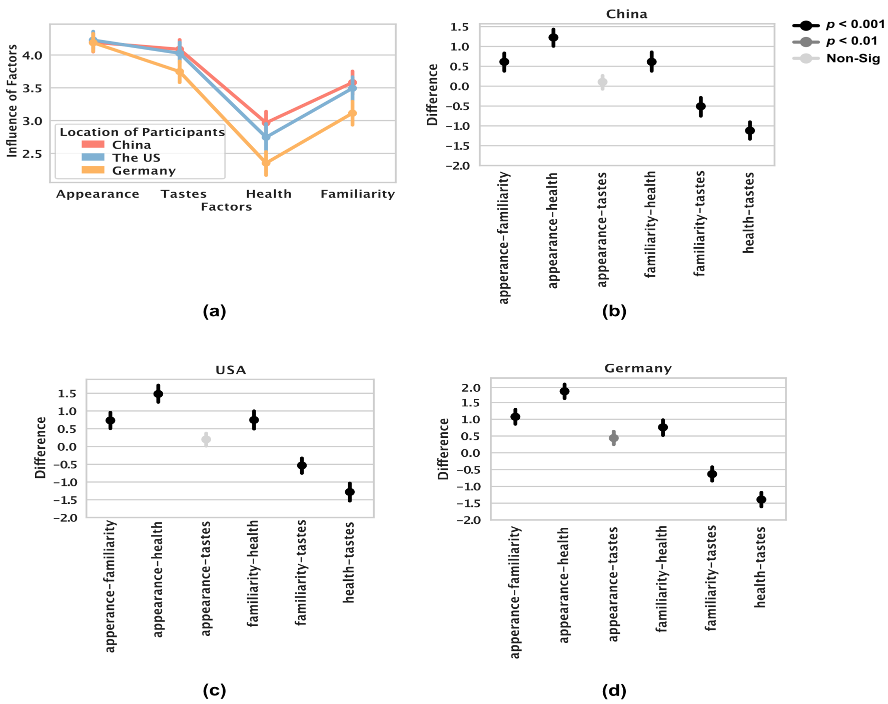


Figure 5.15: Factors which are indicated to have impact on participants' food choice behaviours (a) Influence of the factors (means and std. errors). (b) Pairwise comparisons of the influence of the factors from Chinese participants' self-reports. (c) Pairwise comparisons of the influence of the factors from US participants' self-reports. participants' choices. (d) Pairwise comparisons of the influence of the factors from German participants' self-reports.

The analyses presented in this section have revealed commonalities in the ranking and justifications provided by participants from different cultures. These are mainly illustrated in participants' reliance on their subjective visual impression of the recipes and how they will taste.

Factors Influenced Participant Perceptions of Recipe Images on Different Dimensions In the previous subsection, three of the four main factors were shown to influence the rankings of recipes significantly. These were, namely, appearance, taste and health. In order to determine whether there are commonalities in the perception of recipe images on the three dimensions across cultures, additional ordinal regression models were applied to reveal the relationships between the three main factors and their corresponding supplementary factors. The results are shown in Table 5.11 - 5.13.

First, with respect to visual aesthetics (shown in Table 5.11), for participants from all food cultures, all the visual supplementary factors are shown to exhibit significant positive influence on their perception of the visual appearance of recipe images. This indicates that, the more positively people perceive the colour, presentation, texture of the food and the quality of the images, the more visual appealing the recipe images to them.

Table 5.11: Association between the visual appealing perceived by participants with the visual factors

	Extent of visual appealing								
	The Chinese participants			The US participants			The German participants		
	coef(β)	95% CI	OR	coef(β)	95% CI	OR	coef(β)	95% CI	OR
colour	1.04***	[0.86,1.23]	2.84***	1.05***	[0.89,1.21]	2.85***	0.97***	[0.81,1.13]	2.64***
presentation	0.61***	[0.44,0.79]	1.85***	1.00***	[0.84,1.16]	2.72***	1.16***	[0.99,1.34]	3.20***
texture	0.83***	[0.65,1.01]	2.29***	1.67***	[0.52,0.82]	1.95***	0.78***	[0.62,0.94]	2.18***
quality	0.86***	[0.71,1.02]	2.37***	0.74***	[0.58,0.90]	2.10***	0.95***	[0.79.1.12]	2.60***
Mc Fadden R^2	0.43			0.44			0.48		
AIC	2410.50			2369.57			2183.70		

The sub-dimensions of perceived taste and their relationship to the overall perception of tastiness is more complicated (shown in Table 5.12). Overall, only "spicy" and "bitter" have significant impact in all cultures. These two dimensions play opposite roles in influencing participants' ratings regarding the perceived tastiness of the food in the images. Specifically, foods considered to be spicy were generally rated as being tastier, whereas, foods perceived to be bitter tended to be rated lower. The remaining taste dimensions only affected overall taste rankings in some cultures. For example, foods perceived to be salty or sweet were rated more positively by US and German participants, yet these did not significantly influence the Chinese ratings of food tastes. Perceived sourness only had a negative impact on German participants' evaluations.

Table 5.12: Association between the tasty extent of recipes perceived by participants with the five tastes

	Extent of tasty								
	The Chinese participants			The US participants			The German participants		
	coef(β)	95% CI	OR	coef(β)	95% CI	OR	coef(β)	95% CI	OR
salty	0.03	[-0.08,0.14]	1.03	0.15**	[0.05,0.25]	1.17**	0.12**	[0.00,0.24]	1.13**
sweet	0.03	[-0.07,0.14]	1.04	0.41***	[0.30,0.53]	1.51***	0.26***	[0.14,0.38]	1.30***
sour	-0.07	[-0.20,0.07]	0.94	-0.12	[-0.26,0.02]	0.89	-0.26***	[-0.39,-0.12]	0.77***
spicy	0.28***	[0.18,0.39]	1.33***	0.36***	[0.26,0.47]	1.44***	0.52***	[0.42,0.62]	1.68***
bitter	-0.46***	[-0.63,-0.30]	0.63***	-0.46***	[-0.59,-0.33]	0.63***	-0.35***	[-0.49,-0.21]	0.70***
Mc Fadden R^2	0.01			0.04			0.04		
AIC	3900.83			3975.39			3943.24		

Finally, I show the association between participants' ratings on healthiness of food and certain nutrition contents in Table 5.13. The model's coefficients indicate, for all food cultures, the recipes that appear to be highly calorific and high in fat contents were rated to be unhealthy. In contrast, high protein and fibre are good signals for healthy recipes. These findings suggest participants across cultures show a common understanding when identifying healthy recipes. Nevertheless, there is one nutrition content, sugar, shows a difference between participants from US and other two cultures. Chinese and German participants considered recipes with more sugar are unhealthy, but US participants show no such tendency.

Table 5.13: Association between the health of recipes perceived by participants with the nutrition contents

	Extent of healthiness								
	The Chinese participants			The US participants			The German participants		
	coef(β)	95% CI	OR	coef(β)	95% CI	OR	coef(β)	95% CI	OR
energy	-0.21**	[-0.36,-0.05]	0.81**	-0.53***	[-0.69,-0.38]	0.59***	-0.59***	[-0.75,-0.44]	0.55***
fat	-0.50***	[-0.65,-0.36]	0.60***	-0.55***	[-0.69,-0.41]	0.58***	-0.67***	[-0.82,-0.52]	0.51***
protein	0.35***	[0.24,0.47]	1.43***	0.57***	[0.46,0.67]	1.76***	0.48***	[0.37,0.59]	1.62***
sugar	-0.15***	[-0.25,-0.04]	0.86***	0.06	[-0.05,0.16]	1.06	-0.21**	[-0.34,-0.09]	0.81**
fiber	0.75***	[0.64,0.86]	2.12***	0.72***	[0.62,0.83]	2.06***	0.53***	[0.42,0.64]	1.70***
Mc Fadden R^2		0.11			0.15			0.15	
AIC		3286.87			3312.83			3104.64	

In this section I have demonstrated how and to what extent different dimensions, especially the visual appearance, taste and health of food influenced the ranking and perception of recipe images by participants from different cultural backgrounds. I found that most of the factors I listed had significant impact on the rankings and apply across cultures. This suggests that cross-cultural commonalities exist when perceiving the recipe images. These may, in turn, lead to similar visual food preferences, as revealed by the algorithms in section 5.3.3.2. In the next section, I identify the appreciated and less appreciated recipe images across cultures according to participants' rankings, and investigate the characteristics of these images by comparing the visual features, which have been employed to train the classifiers.

5.3.3.4 Visual Features of Stable Food Preferences Across Cultures (RQ2b)

Using the participants' recipe image rankings, I was able to identify 77 from 270 candidate recipe images, which show a common rating across cultures. 38 recipe images (appreciated recipes) were highly ranked by the participants from China, US and German, and 39 (less appreciated recipes) were ranked poorly by them. A t-SNE image grid of these recipes is shown in Figure 5.16.

These two groups of images display obvious differences in visual sense. A comparative analysis of EVF of images in Figure 5.16 (a) and 5.16 (b) help to explain the visual differences between the appreciated and less appreciated recipe images. The results are shown in Table 5.14. I compared the basic statistics of the EVF of the images, including the min, max, median and the mean values, and reported the p-values and effect size (r) in the last two columns to show whether the differences of the EVF between these two groups of images are significant. It was found that, almost all EVF, except for brightness, are significantly different. Specifically, the mean values of almost all EVF are higher in appreciated recipe images than in less appreciated ones. The differences are indicated particularly significant on the Saturation Variation (Mean = 0.17 vs 0.13; $p < 0.001$; r = 0.44), entropy (Mean = 0.95 vs 0.91; $p < 0.001$; $r = 0.42$), colourfulness (Mean = 0.25 vs 0.20; $p < 0.01$; $r = 0.36$) and naturalness (Mean = 0.87 vs 0.77; $p < 0.01$; $r = 0.34$).

(a)

(b)

Figure 5.16: t-SNE image grid of **(a)** appreciated recipe images of all three food cultures **(b)** less appreciated recipe images of all three food cultures.

Table 5.14: EVF statistics and comparative analysis

Feature	Appreciated Recipe Images				Less Appreciated Recipe Images				*p*-value	*r*
	Mean(±Std.)	Min.	Median	Max.	Mean(±Std.)	Min.	Median	Max.		
Brightness	0.50(±0.11)	0.29	0.51	0.70	0.51(±0.12)	0.22	0.51	0.76	>0.1	0.00
Sharpness	0.25(±0.14)	0.05	0.20	0.61	0.18(±0.12)	0.03	0.15	0.68	<0.01**	0.30°
Contrast	0.06(±0.02)	0.02	0.05	0.13	0.05(±0.03)	0.01	0.04	0.14	<0.1*	0.26°
Colourfulness	0.25(±0.07)	0.14	0.25	0.44	0.20(±0.07)	0.06	0.20	0.38	<0.01**	0.36°°
Entropy	0.95(±0.04)	0.79	0.96	1.00	0.91(±0.05)	0.78	0.91	0.98	<0.001***	0.42°°
Sharpness Variation	0.33(±0.19)	0.16	0.33	0.93	0.30(±0.19)	0.05	0.26	0.97	<0.1*	0.27°
Saturation	0.27(±0.12)	0.08	0.24	0.53	0.22(±0.12)	0.04	0.19	0.58	<0.1*	0.22°
Saturation Variation	0.17(±0.04)	0.09	0.17	0.28	0.13(±0.05)	0.03	0.12	0.28	<0.001***	0.44°°
RGBContrast	0.19(±0.07)	0.04	0.19	0.43	0.15(±0.09)	0.02	0.13	0.44	<0.1*	0.25°
Naturalness	0.87(±0.10)	0.65	0.89	1.00	0.77(±0.14)	0.46	0.79	1.00	<0.01**	0.34°°

In a next step, I measured the similarity of recipe images that represent human visual food preferences by applying the visual features, including VGG16, Colour Histogram, LBP and BoVW. The outcomes are shown in Figure 5.17. The appreciated recipe images (group (a)) are more similar than the less appreciated ones (group (b)) in all cases ($p < 0.001$). Moreover, compared to the images in other groups, i.e., the similarity of (c) the identified appreciated and less appreciated recipes across cultures (n = 77), (d) all candidate recipe images (n = 270) in this user study, the similarity of the appreciated recipes are significantly more visually similar ($p < 0.001$) in terms of all visual features except LBP. This finding suggests that the cross-cultural appreciated recipes tend to be visually similar instead of different.

The images studied in this section - those rated consistently across all three collections - demonstrated the existence of common visual food preferences. I have highlighted the obvious visual differences between appreciated and less appreciated recipe images and validated these statistically using EVF.

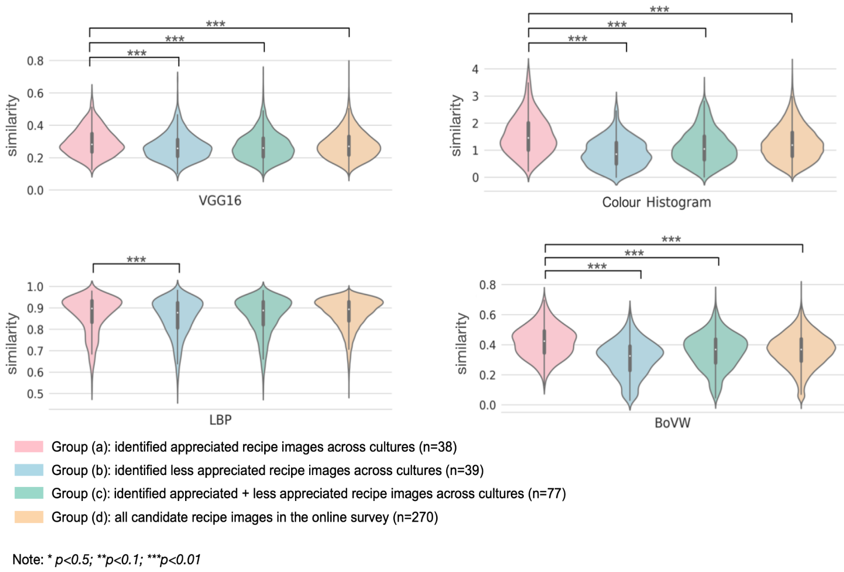


Figure 5.17: Distribution and Comparison of recipe images similarity based on different visual features

5.3.4 Summary and Discussion

In this section, I first summarise the findings from Study II with respect to the research questions raised in section 5.3.1. Afterwards, I discuss the findings in the context of the literature and with respect to food recommender system development.

5.3.4.1 The Primary Findings

- *RQ1a.* For the food cultures studied in this work, i.e., China, US and Germany, it is possible to identify the appreciated and less appreciated recipes with algorithms based on the visual features extracted from the online recipe images. However, the ACC (at best 67%) of the classifiers suggests that predicting food preferences with recipe images is a challenging work.

- *RQ1b.* By means of transferring the classifier with the highest accuracy to each collection, I found the classifier trained on *Allrecipes* can make useful predictions for *Kochbar* collection and vice-versa. The *Xiachufang*, nevertheless, is an outlier. The classifier trained on the Chinese recipes performed well only this collection and very poorly on the other two collections. The other two classifiers were also unable to achieve high accuracy on *Xiachufang* collection. However, the results hinted that stable patterns may exist in visual food preferences across cultures.

- *RQ2a.* Three of the four main factors, i.e., appearance, taste and health perceived by the participants, have influenced participants' food preferences significantly. Some of these, such as appearance and taste, have comparable impact on rankings of recipe images across cultures. Specifically, the more the participants were satisfied with the appearance and taste they perceived from the images, the more they were to rate the recipes highly. In addition, participants can realise the impact of appearance and taste when they are ranking the recipe images. Health, however, influenced the rankings of users only from US and Germany significantly. Recipes perceived to be healthier were ranked lower by these participants. Nevertheless, these participants did not consider health when ranking the images. The investigation of the association between the three main factors and their corresponding supplementary factors also revealed a common tendency in perceiving the recipe images across cultures, which might lead to similar visual food preferences and support the presence of stable patterns revealed by the algorithms.

- *RQ2b.* I identified 77 recipe images indicating human visual food preferences across cultures, of which 38 are appreciated and 39 are less appreciated. I showed obvious visual differences between these by displaying them in a two-part image grid. A followed-up comparative analysis of EVF uncovered that the cross-cultural appreciated images generally have higher values of all EVF with exception to brightness. This is a strong indication that the EVF is the set of visual features that can help identify food preferences across food cultures. Moreover, in terms of other visual features, i.e., VGG16, Colour Histogram, LBP and BoVW, the appreciated recipe images across cultures are more similar than those are less appreciated, suggesting the visual food preferences across cultures tend to be similar rather than different.

5.3.4.2 Implication of Study II

There are several contributions of this study. The first one is to show that it is possible to predict if the online recipes are appreciated or not using only information contained within their corresponding images. The work was motivated by previous findings including those of Trattner and colleagues who reported the importance of visual aspects in food choices (Elsweiler et al., 2017; Trattner et al., 2018), but also the challenging nature of predicting which foods will be chosen (Trattner & Elsweiler, 2017a). Building on works that have uncovered the involvement of visual signals in food choices and applied the discovered knowledge in food recommender systems (A. D. Starke et al., 2021; Yang et al., 2017), I examine these aspects in detail to determine the important role the visual appearance plays and learn how similar (or different) this is across cultures. The contribution of this study lies not only in that I demonstrate that visual information contained in recipe images can be used to predict food preferences in all the cultures studied, but also that visual preferences transfer, in the case of Germany and the US, across cultures. To the best knowledge of the author, this is the first work with these findings in the food domain. Although most of previous research, such as (Cantarero et al., 2013; Ahn et al., 2011; Zhu et al., 2013; Kim & Chung, 2016; Sajadmanesh et al., 2017), emphasises how preferences vary between cultures, this study reveals stable patterns in visual preferences across culture. Building on this finding, I offer a new perspective in developing cross-cultural food recommender systems, that is, taking the knowledge I have regarding the visual food preferences in one food culture and applying it to another. The benefits of being aware of and employing this approach are analogous to those brought by transfer learning (S. J. Pan & Yang, 2009).

A second contribution, provided by this study, was that I am able to recreate the same patterns as the naturalistic data despite participants providing ratings from images both from their own culture and others, which was not the case in the original data. This provides us with increased confidence that would generalise to large samples. These findings suggest the future work should be done to model and transfer human rating behaviours in the domain of food preferences within a multi-cultural context. Other work, such as from (Li et al., 2009) and colleagues, has explored the comparable idea applying transfer learning to cross-domain collaborative filtering problems with good results on benchmark datasets. In the food domain, however, this method has not been explored. The results of this study are promising, but should be validated with larger datasets.

The findings in the study also have implications for the visual nudging of food choices. Trattner and colleagues (Trattner et al., 2018) have shown that the low level visual features, i.e., EVF in this study, are important features in predicting popularity of recipes. Consistent with those findings, I have also shown the predictive power of these features in classifying recipes with different appreciation. More than that, the comparative analysis in Section 5.3.3.4 makes the visual patterns of appreciated recipe images across cultures clear. In this analysis, all the EVF with exception to brightness were significantly higher in the set of appreciated recipes. Such visual biases can be used to nudge healthier food choices, as Elsweiler (Elsweiler et al., 2017) and Starke (A. D. Starke et al., 2021) have proven. I have demonstrated that the visual biased hold across cultures, which not only speaks for the generalisability of past work, but is promising for the feasibility of developing cross-cultural food recommender systems that promoting healthier eating habits. For example, my finding once again reveal the paradox of health and human food choices. As has been noted in past work, humans tend to like, consume and share unhealthy, but attractive food (Holmberg et al., 2016). It seems that, on the recipe portals, healthy foods are neither attractive on the eye, nor look palatable enough on average for users to choose. This finding from this study is promising in terms of improving visual appearances of healthy food online, such as modifying the unattractive recipe images by increasing their EVF to make these more visually appealing and more likely to be chosen consequently.

Study II focused on food preference on the visual aspect, nevertheless, participants' responses to several questions relating to food flavours have revealed the importance of flavours in influencing human preferences. In addition, the association between human perceived tastiness and tastes of food have suggested the possibility of identifying stable patterns of food preference across cultures on the flavour aspects. Empirical experiments will be designed to validate whether stable patterns exist on the flavour aspects in Chapter 6.

5.4 Chapter Summary

This chapter has studied one aesthetic aspect of food, the visual aspect, especially focused on the visual differences of food across cultures, and that of food with different appreciation levels. Experiments with machine learning approaches and user studies have been designed and conducted. The experiments by means of machine learning have proven the existence of visual differences of food vary across cultures and between appreciated and less appreciated recipes. The user studies have been designed to collect human judgements, which have been compared to the outcomes from the classifiers, and additionally revealed more information about human food visual perception that are helpful for the development of food recommender systems.

In Study I, the performance of the classifiers provides an answer to the first issue that was raised in section 5.1, that is, the recipes across cultures can be differentiated based on the representation of the food relating to visual appearance with algorithms. Human annotators, on the other hand, performed much worse in perceiving the visual cultural food differences. Further analyses on participants' responses in the user study have revealed the reasons for human participants' poor performance in labelling the recipes images. The reasons, for example, is that the participants' labelling behaviours are influenced by their past experience and knowledge, which are shaped by their cultural upbringings. This finding suggests that it is necessary to take the context-related factor, culture, into consideration when developing food recommender systems, in order to provide more culturally acceptable food recommendations.

In Study II, the visual food preferences have been detected within and across the cultures. In addition to detecting the visual preference patterns of food within each culture, this study has ascertained stable patterns of visual food preference across cultures. In addition, by investigating the visual features of recipe images that identified as appreciated and less appreciated recipes across cultures, commonality of visual food preferences across culture has been further underlined. Findings from Study II indicate the promise of developing cross-cultural food recommender systems with visual information. The next chapter will investigate similar problems of food on the other aesthetic aspect of food, i.e., flavour.

Chapter 6

Cross-Cultural Food Classification and Preferences on the Flavour Aspects

6.1 Introduction

The previous chapter focused on the visual aesthetics of recipes, in this chapter, I investigate cross-cultural food differences and similarities in human food preferences with respect to flavour. The aim is to shed light on how to develop food recommender systems, which exploit flavour information. To make the results comparable with those of experiments on the visual food appearance, the recipe collections, i.e., *Xiachufang* collection, *Allrecipes* collection and *Kochbar* collection are still employed as the proxies of Chinese, US and German food cultures. The issues raised in this chapter corresponding to those in Chapter 5, which are listed below:

- *Issue 1.* To what extent is it possible to differentiate the food across cultures based on flavour representations of the recipes?
- *Issue 2.* To what extent is it possible to identify the differences and ascertain stable patterns of food preferences across cultures based on the same flavour representations?

These two issues are addressed by means of machine learning approaches with recipes represented by flavour compounds, comparable to those presented in the previous chapter. Unlike the experimental methodology in Chapter 5, no user study is performed as a validation step. The practicalities of having participants actually code and taste food to ascertain flavour make such an approach on a large scale infeasible. Instead, recipe ingredients are used to train models, which are used as a baseline. This provides a means to establish the power of flavour information in differentiating recipe origins and predicting food preferences, since ingredients have proven to be discriminative across cultures, and classifiers trained with ingredients have achieved reasonable accuracy in cuisine prediction tasks (Su et al., 2014; Kim & Chung, 2016; Sajadmanesh et al., 2017). The issues are correspondingly formulated as the following research questions:

- *RQ1.* To what extent is it possible to differentiate recipes from the recipe portals of differentiate food cultures with machine learning approaches based on ingredients and flavour compounds, respectively?

- *RQ2a.* To what extent is it possible to differentiate the appreciated and less appreciated recipes within each food culture with machine learning approaches based on ingredients and flavour compounds, respectively?
- *RQ2b.* To what extent is it possible to ascertain stable patterns of food preferences across cultures using machine learning approaches based on ingredients and flavour compounds, respectively?

The machine learning experiments to address these questions are supplemented by several additional exploratory analyses, which aim to justify the results. These are mainly quantitative analyses of the flavour compounds and ingredients, including using ingredient complement networks and semantic descriptions of flavour compounds. The methodological details are described in Section 6.2. The results of these address the *RQ3*:

- *RQ3.* To what extent can revealed patterns of flavour-related cross-cultural food preferences be justified by the exploratory analyses?

This chapter is structured as follows: Section 6.2 describes the methodology of the study including how the machine learning models are trained using ingredients and flavour compounds, respectively, as well as the design of the exploratory experiments. The results are presented in Section 6.3. Section 6.4 discusses the findings and shows the implications of the study. Section 6.5 summarises this chapter.

6.2 Methods

6.2.1 Predicting Food Culture based on Ingredients and Flavour Information

The methodology in this study is similar to that of the experiments applying the visual information in Chapter 5. *RQ1* was formulated as a multi-class prediction problem, in which ingredient and flavour compounds were employed to train classifiers to predict the source of the recipes, respectively. In this step, classifiers were trained using ingredients and flavour compounds vectors, which were generated by applying TF-IDF and Word2Vec, as described in Chapter 4. Three supervised classification algorithms were applied: **Naïve Bayes (NB)**, **Logistic Regression (LOG)** and **Random Forest (RF)**. Moreover, a 5-fold Randomised Search CV was used to determine the optimal parameters for LOG and RF. The performance of the classifier was reported by measuring the accuracy (ACC), and a confusion matrix was applied to visualise the misclassification of the classifier.

6.2.2 Predicting Intra- and Inter-Cultural Food Preferences Based on Ingredients and Flavour Information

In order to address *RQ2a*, which involves identifying the appreciated and less appreciated recipes based on ingredients and flavour compounds in each recipe collection, I formulated several binary classification tasks on the recipe samples that were drawn from the top-10% (appreciated) and bottom-10% (less appreciated) based on the appreciated metric of each recipe portal. Here, the classifiers were trained using the vector of ingredients and flavour compounds. The same algorithms for predicting sources of recipes (as described in 6.2.1) were applied again, and ACC was used as performance metric for the classifiers. After differentiating the appreciated and less appreciated recipes within each culture, I identified the best performed classifier on each collection, i.e., the one with the highest accuracy, and tested their predictive ability on the other two collections. The aim here is to detect stable patterns of flavour-related, cross-cultural food preferences and address *RQ2b*.

6.2.3 Design of the Exploratory Analyses

In this study, I designed two exploratory analyses to validate the patterns of cross-cultural food preferences revealed by the algorithms. The methodology of these two exploratory analyses is, firstly, establishing description systems for food flavour. The description systems, for example, a well-known one is the five-basic tastes, i.e., sweet, sour, salty, bitter and umami. These were used in the user study in Study II in Chapter 5, and hint the flavour preferences across cultures (e.g., a common tendency of preferring spicy food). Then, after the description systems of flavour being established, these were applied to investigate and describe the flavour preferences across cultures. The methods and process of the experiments is described in the following subsections.

6.2.3.1 Identifying Savoury and Sweet Recipes by Means of Ingredient Complement Networks

This experiment is inspired by the work of Teng et al. (2012), in which the authors built an ingredient complement network based on the co-occurrences of ingredients. The network displays two distinct subcommunities of recipes on the ingredient level: one corresponding to savoury dishes, the other to sweet ones, which is shown in Figure 6.1.

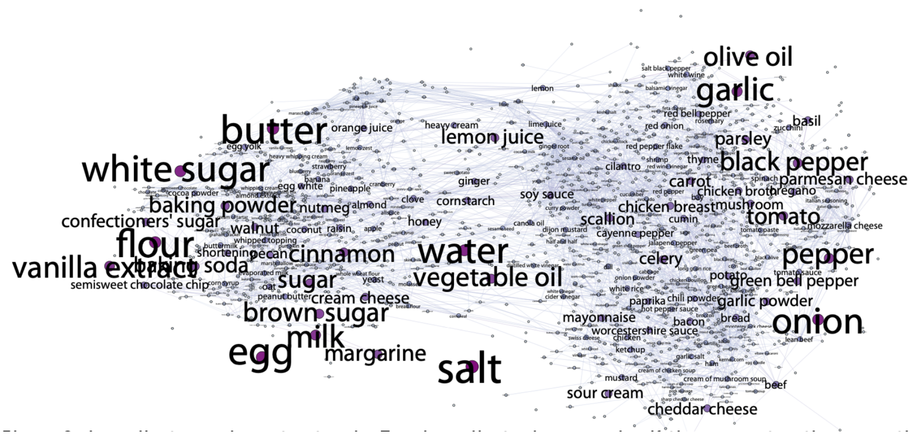


Figure 6.1: The ingredient complement network constructed by Teng et al. Taken from (Teng et al., 2012).

Building on this, in this work, I labelled the recipes as savoury and sweet by means of node embedding (Node2Vec) and unsupervised learning approach (K-Means clustering). After which, I calculated the proportion of savoury/sweet recipes in appreciated and less appreciated recipes in each collection, respectively. The aim is to detect, for example, whether a culture prefers savoury or sweet dishes and whether other food cultures show the similar or opposite trend. The steps and methods of the experiment are described as follows:

Step 1: Generating ingredient complement networks
In this network, each node denotes an ingredient. Two nodes share an edge if the likelihood of these two ingredients occurring in one recipe exceed a threshold, which is defined by the pointwise mutual information (PMI). The PMI indicates the probability of a pair of ingredients co-occurring *p*(a,b) against the probability that they occur separately *p*(a), *p*(b). The formular of PMI is:

$$PMI(a,b) = \log \frac{p(a,b)}{p(a)p(b)} \tag{6.1}$$

where

$$p(a,b) = \frac{\text{\# of recipe where } a \text{ and } b \text{ co-occur}}{\text{\# of recipes}}$$

$$p(a) = \frac{\text{\# of recipes where } a \text{ occurs}}{\text{\# of recipes}} \tag{6.2}$$

$$p(b) = \frac{\text{\# of recipes where } b \text{ occurs}}{\text{\# of recipes}}$$

PMI underlines the complementary ingredients, which tend to occur together far more often than would be expected by chance. In this work, I normalized the PMI (NPMI) according to Bouma (2009) in order to limit its range to between -1 and 1, where -1 means that a pair of ingredients would never occur together, and 1 means they would definitely co-occur. NPMI is measured with:

$$NPMI(a,b) = \frac{PMI(a,b)}{h(a,b)} \tag{6.3}$$

where

$$h(x,y) = -\log p(a,b) \tag{6.4}$$

To generate the complement networks containing highly complementary ingredients in the network, I set threshold to select the candidate ingredients and ingredient pairs. Firstly, the ingredients appear more than 10 times and a pair of them appeared more than 5 times were selected. In addition, the ingredient pairs, whose NPMI is over 0.10 were then kept to generate the complementary network.

Step 2: Representing ingredients and recipes by means of Node2Vec
Once the ingredient complement networks were constructed, the nodes in these were represented by applying Node2Vec, which was proposed by Grover and Leskovec (2016). The main idea of this algorithm is to map the nodes in a network to a feature space while preserving the initial structure of the network. To be specific, if two nodes are neighbours or share a similar role in a network, then similar embeddings will be yielded for these in the feature space. An example is shown in Figure 6.2.

The working principle of Node2Vec is, for a given node, a set of random walks start from it and traverse the whole graph along the edges. During the process, the algorithm explores the relationship between the node and its neighbours. Then the random walks

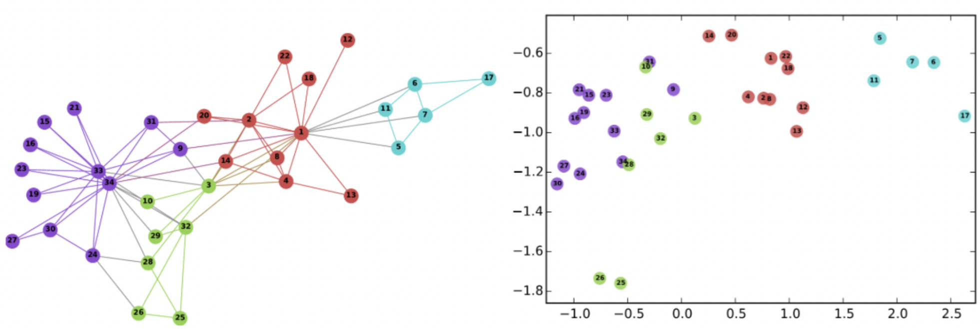


Figure 6.2: Example of mapping Node2Vec embeddings to low-dimensional space. Taken from `https://snap.stanford.edu/proj/embeddings-www/`.

are passed into a skip-gram model, inspired by Word2Vec (Mikolov et al., 2013), in order to yield embeddings for the nodes. The process repeats until the embeddings are yield for each node in the network. Node2Vec uses a biased random walk controlled by a series of parameters. In this work, the node embeddings were generated with Node2Vec[1] implemented in Python. The parameters in the algorithm were fine-tuned as follows:

- dimensions: This parameter refers to the dimension of the generated node embeddings. I chose the default value 128, so that each ingredient in the networks is represented as a 128-dimensional vector.
- num_walks: this parameter determines the number of random walks per node. The default value of the parameter is 10. In this work, when generating node embeddings for the ingredients in the complement networks of recipes in *Xiachufang*, *Allrecipes* and *Kochbar* collection, different values were tried and tested to make the node embeddings in the feature space better capture the structure of the networks. The range of the parameter were tested is from 10 to 25.
- walk_length: this parameter refers to how many nodes each walk traverse. The default value of it is 80. In this work, the selection of this parameter was similar to that of num_walk, i.e., different values for different recipe collections. The range of the parameter were tested is from 10 to 20.
- p and q: These 2 parameters are important for Node2Vec, of which p is the Return hyperparameter, determining the possibility of revisiting a former node in a walk after visiting a node. Setting a higher value of p ensures lower chance of revisiting a node. The default value of p is 1, and I set it as 2 in this work to make the exploration of the graph broader. q is the Inout hyperparameter, determining the probability of exploring undiscovered nodes. A high value of q ($q > 1$) allows the random walk towards the closer neighbours of the previous nodes, which leads to a local view of the network with respect to the start node in the walk and approximate Breadth first Sampling (BFS) behaviour. In contrast, a lower value of q ($q < 1$) encourages the walk to explore the network with a Depth first Sampling (DFS) behaviour and offers a macro-view of the network. The differences of BFS and DFS is shown in Figure 6.3. The default q is 1. In this work, I set the q as 0.25 in order to get a macro view of the ingredient complement networks.

[1] `https://github.com/eliorc/node2vec`

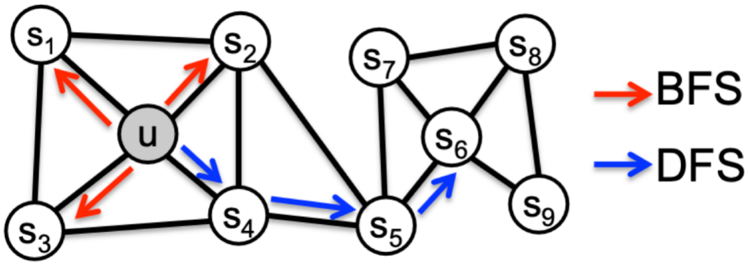


Figure 6.3: An Example of BFS and DFS. The start node is u. s_1, s_2, s_3 are the sample nodes of BFS, s_4, s_5, s_6 are the sample nodes DFS. Taken from (Grover & Leskovec, 2016).

After each node, namely each ingredient in the complement networks was represented by a vector, it can be applied to downstream prediction works with machine learning approaches. Such as in (Park et al., 2021), the node embeddings were applied to predict good food pairs (i.e., a pair of ingredients which tastes better) and the relationship between ingredients and flavour compounds. In this work, after representing each ingredient in the complement network, the embeddings were applied to obtain the representation of recipes. This process is similar to that of getting recipe representation with the word embeddings of ingredients/flavour compounds by means of Word2Vec (described in Chapter 4), which is, for each recipe, I summed up all ingredient embeddings and divided them by the length of recipes.

Step 3: K-Means on the recipe representation

After representing each recipe with a multi-dimensional vector, an unsupervised learning algorithm, i.e, K-Means clustering was applied. I set the value of k as 2 since according to Teng (2012), there are two subcommunities shown in the ingredient complement network, one for savoury and another for sweet recipes. The aim of this step is to apply K-Means on the representation of recipes to get two recipe clusters and label each cluster of recipes as savoury and sweet. For each cluster of recipes, I weighted their ingredients with their frequency of occurrences and visualised them with WordCloud[2], which helped to determine which cluster represents the savoury recipes and which one represents the sweet ones.

Step 4: Calculating the proportion of savoury/sweet recipes in appreciated and less appreciated recipes

Up until now, the first description system for food flavour was built, which describes recipes as savoury and sweet. In order to show the differences and similarities of flavour food preference across cultures, the proportion of the savoury and sweet recipes in appreciated and less appreciated recipes of each recipe collection was calculated and compared to each other.

[2]`https://amueller.github.io/word_cloud/`

6.2.3.2 Building Semantic Clusters for Flavour Compounds Based on their Corresponding Flavour Profiles

In this exploratory experiment, the flavour profiles of flavour compounds were applied to build the description system of food flavour. In FlavorDB, the flavour compounds have a list of flavour profiles, which describe the flavour compounds semantically with a series of words, the relationship between flavour compounds and flavour profile is shown in Figure 6.4. I also showed the relationship between ingredients and their corresponding flavour compounds. It is found that a flavour compound can be shared by multiple ingredients. Thus although some ingredients do not account for food of certain culture, they might share similar flavour compounds, leading to overlaps in terms of flavour across cultures.

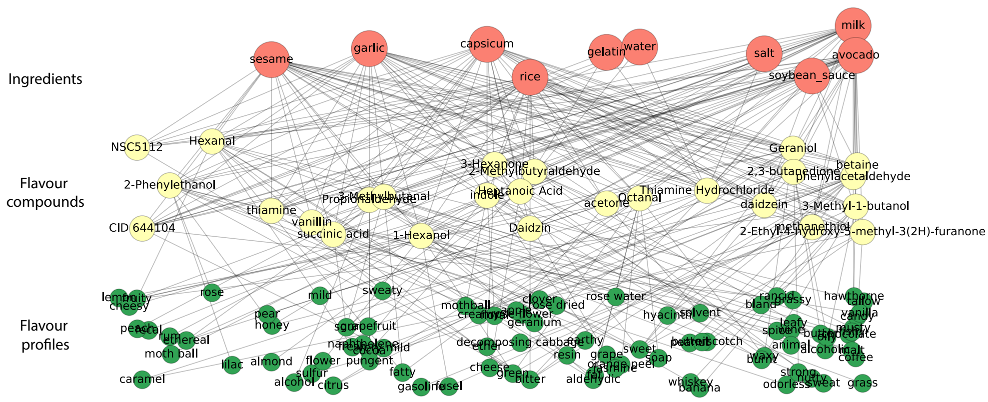


Figure 6.4: Relation network of <ingredients - flavour compounds> and <flavour compounds - flavour profiles>

The basic idea of this experiment is to cluster the flavour compounds with similar flavour profiles into one group, and assigning semantic description for each cluster based on the flavour profiles. Next, I examined the ratio of each cluster of flavour compounds in the appreciated and less appreciated recipes in each recipe collection. This ratio was then compared across recipe collections to see whether the ratio reveals the similar or different trend of flavour preferences across cultures. The whole process was described as follows:

Step 1: Representing flavour compounds with their corresponding flavour profiles
In this step, I filtered the flavour compounds without flavour profiles out and kept only flavour compounds appear in more than three ingredients (n = 927). For each flavour compounds, I converted it into a Boolean bag of flavour profile vector, which marks the presence of flavour profile as a Boolean value, 0 for absent and 1 for present. Therefore, each flavour compound was represented as a vector with a length equal to the total number of flavour profiles, which is 927.

Step 2: K-Means on the representation of flavour compounds
K-Means algorithm was applied in this step to partition flavour compounds with similar flavour profiles into k distinct clusters. In order to determine the optimal number of clusters, i.e., the value of k, the elbow method was applied. This method plots the value of

the cost function produced by different value of k (range from 1 to 9 in this study). The optimal k lies on the point (i.e., the elbow) where the value of the cost function becomes relatively flat after it decreases sharply. The elbow methods determined the optimal value of k as 4, as shown in Figure 6.5.

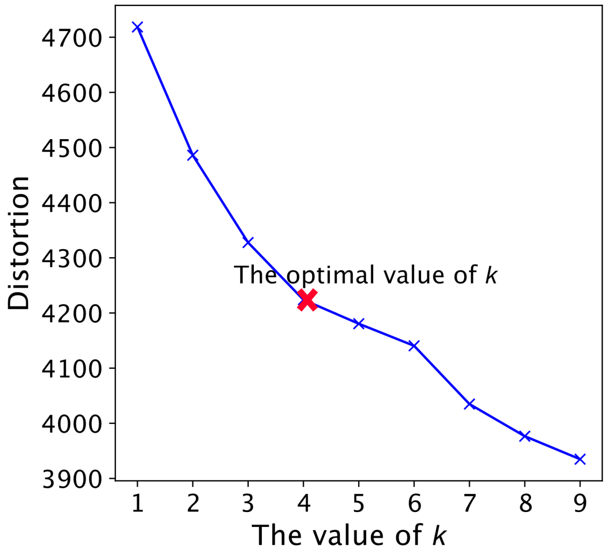


Figure 6.5: Elbow method for the optimal k. 4 should be the optimal number of clusters for the flavour compounds.

Step 3: Assigning semantic description for each cluster

After determining the number of clusters for the flavour compounds, I assigned the semantic description for each cluster based on the representative of their flavour profiles. To be specific, for one of the flavour compounds clusters, I first measured the prevalence of a flavour profile p for describing the flavour compounds in cluster c with:

$$P_p^c = \frac{n_p^c}{N_c} \tag{6.5}$$

where n_p^c is the prevalence of p for describing flavour compounds in cluster c, and N_c is the total number of flavour compounds in cluster c.

Then the representative of the flavour profiles in cluster c was measured with the formula[3]:

$$P_p^c = P_p^c - \langle P_p^{c'} \rangle_{c' \neq c} \tag{6.6}$$

where $\langle P_p^{c'} \rangle$ is the average prevalence of the profile p for describing the flavour compounds in other clusters except c. This shows the difference between the prevalence of a profile p for describing the flavour compounds in one cluster c and the average prevalence of the p for describing the flavour compounds in other clusters. For each cluster, I chose the top 50 representative flavour profiles and visualised them with WordCloud, which helped to determine the semantic description for each cluster of flavour compounds.

[3]This formula was first applied in (Ahn et al., 2011) for measuring the authenticity of ingredients in different cuisines.

Step 4: Calculating the ratio of each cluster of flavour compounds in the appreciated and less appreciated recipes
As a next step, I measured the ratio of each cluster of flavour compounds in the appreciated and less appreciated recipes for each recipe collection in order to investigate flavour preference. Instead of applying all ingredients occurrences in the appreciated and less appreciated recipes, the distinct ingredients for appreciated and less appreciated recipes were determined and applied to represent the recipes. This is because, according to Rozin (2018), it is possible to represent regional cuisines with a few key ingredients, such as soybean sauce in Chinese recipes, and paprika, onion, and lard in Hungarian cuisine (Ahn et al., 2011). In this work, I attempted to rely on this finding and apply several key ingredients to represent the appreciated and less appreciated recipes. The distinctive ingredients in the appreciated recipes, for example, are those appear more frequently in appreciated recipes than in less appreciated recipes. The weighted log odds ratio was applied to determine the distinctive ingredients. It can be calculated with:

$$\text{log odds ratio} = \ln\left(\frac{\left[\frac{n+1}{total+1}\right]_{\text{Appreciated Recipes}}}{\left[\frac{n+1}{total+1}\right]_{\text{Less Appreciated Recipes}}}\right) \tag{6.7}$$

where n is the frequency of an ingredient appears in appreciated and less appreciated recipes, and *total* indicates the total number of ingredients in the appreciated and less appreciated recipes. The weighted log odds ratio was proposed by Monroe (2008), who weighs the log odds ratio by an informative Dirichlet prior. In this work, the package *tidylo*[4] was used to calculate the weighted log odds ratio for the ingredients. The higher the weighted log odds ratio of an ingredient, the more distinctive it is in the recipe collections. I determined the top 50 ingredients according to the value of the weighted log odds as the distinctive ingredients in appreciated and less appreciated recipes for each recipe collection. After determining the distinctive ingredients, the ratio of each cluster of flavour compounds in the appreciated and less appreciated recipes were calculated and compared to validate the patterns revealed by the algorithms in terms of flavour preferences across cultures.

6.3 Results

The results of the study are presented in the following subsections to address the research questions raised in Section 6.1.

6.3.1 Predicting the Origin of Recipes Based on Ingredients and Flavour Compounds with Machine Learning Approaches (RQ1)

Table 6.1 illustrates the performance of the classifiers trained using ingredients and flavour compounds vectors for classifying the origin portals of the recipes. Overall, based on the ACC of all the classifiers, both ingredients and flavour compounds show reasonable ability for differentiating recipes from Chinese, US and German recipe portals, especially when representing the recipes with TF-IDF weightings. The ACC of the classifiers trained with

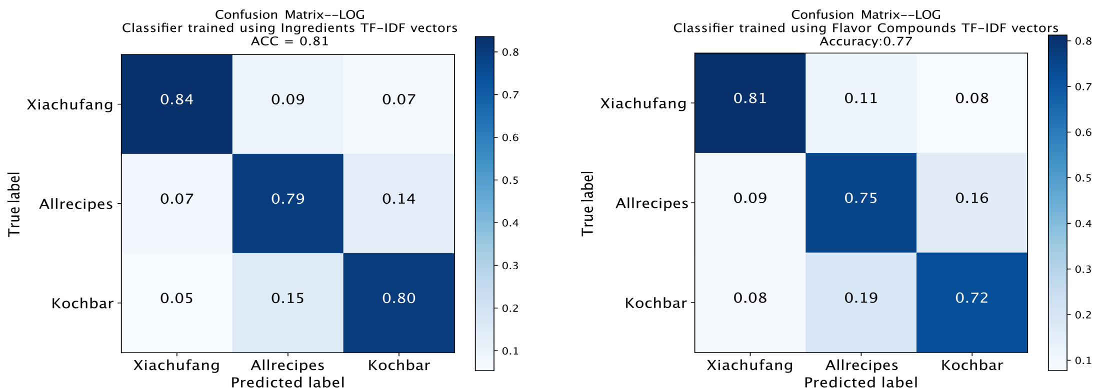
[4] `https://github.com/juliasilge/tidylo`

TF-IDF ingredients vectors and flavour compounds vectors are 0.81 and 0.77, respectively, showing that the recipes across cultures are sufficiently distinct on both ingredient usage and flavour, thus they can be differentiated algorithmically. Moreover, ingredients outperformed flavour compounds in this classification task slightly. It suggests that, in comparison to the ingredient usage, there are more similarities in terms of flavour across different food cultures. This finding is in line with that in (Sajadmanesh et al., 2017), which revealed that flavour is not as much discriminative as ingredients.

Table 6.1: Results for predicting which portal the recipes belong to based on ingredient and flavour compounds vectors (TF-IDF and Word2Vec). Best performing scores for each classifier are bold. *NB* = Naive Bayes; *LOG* = Logistic Regression; *RF* = Random Forest.

Features	Accuracy		
	NB	LOG	RF
Ingredient (TF-IDF)	0.77	**0.81**	0.73
Flavour Compounds (TF-IDF)	0.65	**0.77**	0.74
Ingredient (Word2Vec)	0.56	0.73	0.72
Flavour Compounds (Word2Vec)	0.35	0.66	0.70

In all cases in this experiment, TF-IDF vectors outperform Word2Vec embeddings. This is not surprising since Word2Vec is more capable of classifying documents by means of capturing the context of a word in a document and the semantic similarity of the word with the others (Mikolov et al., 2013). However, in general, a recipe is a list of ingredients, and in this study, it was represented with flavour compounds, which share little in the way of semantic relationships. In addition, Word2Vec treats each word equally, while TF-IDF indicates how important a word by weighing its frequency of occurrence in a document and how often the word occur in the whole document corpus. In this study, taking the importance of the ingredients into consideration is apparently more helpful for the recipe classification task.



(a) (b)

Figure 6.6: Confusion matrix of **(a)** the classifier trained using ingredients TF-IDF vectors and **(b)** the classifier trained using flavour compounds TF-IDF vectors.

Figure 6.6 (a) and (b) present the confusion matrices of the classifiers trained using TF-IDF ingredients vectors and flavour compounds vectors. These confusion matrices illustrate that, the classifier, regardless of whether trained using ingredients or flavour compounds, is more accurate when identifying the recipes from the Chinese recipe portal (ACC = 0.84 & 0.81) than from the other two portals (ACC = 0.79 & 0.75 and ACC = 0.80 & 0.72). The recipes from the US and German portals were more likely to be confused by the classifiers, while only few Chinese recipes were misclassified. This finding suggests that recipes from the US and German are more similar in terms of ingredient usage and flavour, whereas the Chinese recipes appear to be more distinct on these two aspects. This align with the findings from the visual aesthetic experiment (in Chapter 5) where Chinese recipes were found to be the most visually distinct, while the recipes from US and German portals are prone to be misclassified to each other.

Overall, building on these findings of the experiments, the *RQ1* can be addressed: it is possible to distinguish the recipes from different recipes portals with ingredients and flavour compounds vectors and machine learning approaches. Ingredients performed better than flavour compounds in differentiating the recipes from different sources (ACC = 0.81 vs. ACC = 0.77), indicating that compared to ingredient usages, there are more overlaps of food on the flavour aspects across the Chinese, US and German cultures. The Chinese recipes are the most distinctive in terms of both ingredient and flavour compared to the recipes from US and German portals, which are more likely to be confused algorithmically. Moreover, it was found that TF-IDF performed better than Word2Vec in the recipe classification task.

6.3.2 Identifying Intra- Cultural Food Preferences Based on Ingredients and Flavour Compounds with Machine Learning Approaches (RQ2a)

After classifying the recipes from different portals, the vectors of ingredients and flavour compounds were applied to differentiate between the appreciated and less appreciated recipes within each recipe collection. Table 6.2 shows the performance of the classifiers. The accuracy of the classifiers (0.51 < ACC < 0.70) indicates, again, that appreciated and less appreciated recipes in each recipe portal can be distinguished using ingredients and flavour compounds. TF-IDF outperformed Word2Vec in all cases, which is the same as in the recipe sources prediction task in Section 6.3.1. Overall, as was also the case in the visual experiments, predicting food preferences with ingredients and flavour compounds is more challenging compared to classifying the sources of recipes with the same information. This is illustrated in the comparison of the accuracy of the best performing classifier for recipe sources prediction (ACC = 0.81 & 0.77) and the average accuracy of the best performing classifiers for predicting the flavour preferences for each food culture (avg. ACC = 0.67 & 0.66).

Comparing the best performing classifier on each recipe collection reveals that ingredients offered only slightly better predictive power than flavour compounds in identifying appreciated recipes from the Chinese and US portals, and they perform rather worse in identifying appreciated recipes from the German portal. This finding suggests that flavour has almost equal capability with ingredient in terms of explaining the food preferences and it is promising to involve flavour compounds into the development of food recommender systems, which take the users' flavour preference into consideration.

In summary, with respect to *RQ2a*, the findings from these machine learning tasks indicate that, it is possible to differentiate the appreciated and less appreciated recipes within each food culture using ingredients and flavour compounds with machine learning

Table 6.2: Results for identifying appreciated and less appreciated recipes in each recipe portal with ingredient and flavour compound vector (TF-IDF & Word2Vec). *NB* = Naive Bayes; *LOG* = Logistic Regression; *RF* = Random Forest.

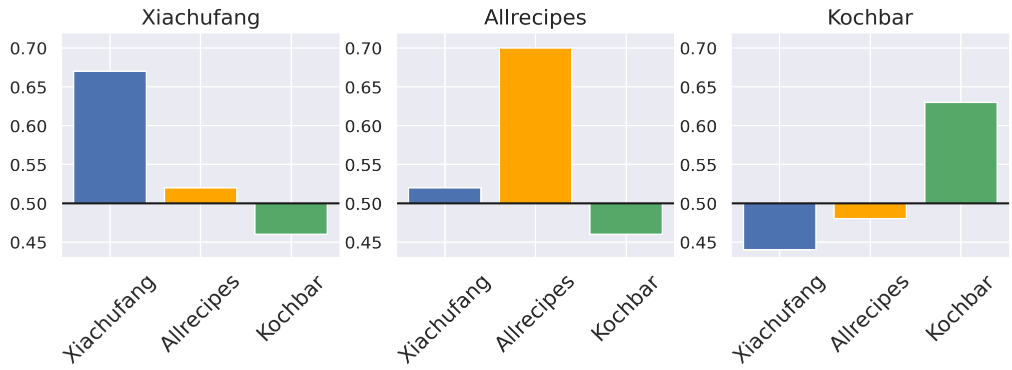
Features	**Xiachufang**			**Allrecipes**			**Kochbar**		
	Accuracy			Accuracy			Accuracy		
	NB	LOG	RF	NB	LOG	RF	NB	LOG	RF
Ingredient (TF-IDF)	0.62	0.65	**0.67**	0.67	**0.70**	0.68	0.62	**0.63**	0.61
Flavour Compounds (TF-IDF)	0.60	0.61	**0.66**	0.64	0.66	**0.67**	0.59	**0.65**	0.64
Ingredient (Word2Vec)	0.60	0.64	0.64	0.58	0.65	0.67	0.54	0.57	0.54
Flavour Compounds (Word2Vec)	0.56	0.61	0.63	0.51	0.65	0.64	0.48	0.61	0.60

approaches. Flavour compounds showed a very close predictive power to that of ingredients for predicting food preferences for all cultures. Moreover, TF-IDF outperformed Word2Vec again.

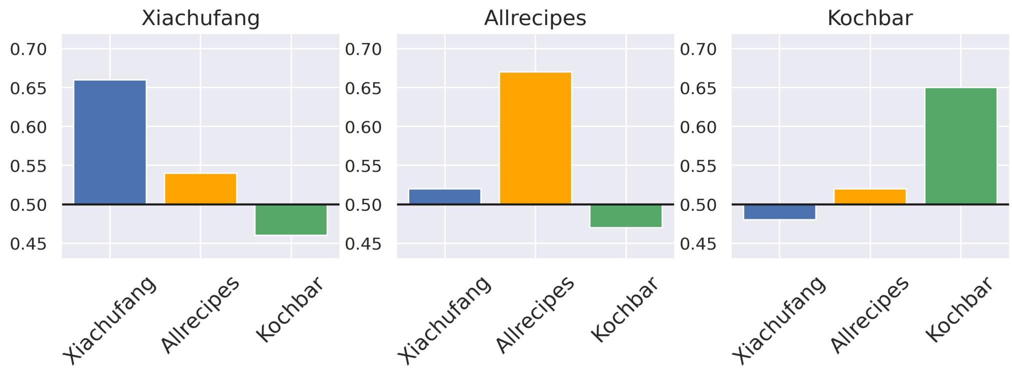
6.3.3 Identifying Stable Patterns of Food Preferences Across Cultures Based on Ingredients and Flavour Information (RQ2b)

After determining the power of ingredients and flavour compounds in predicting food preferences within each food culture, I selected the best performing classifier on each recipe collection, which were those, according to the results shown in Table 6.2, trained using vectors generated with TF-IDF, and tested their ability in identifying the appreciated and less appreciated recipes in the other two collections. Figure 6.7 present the results.

Figure 6.7 (a) and (b) show the best performing classifiers trained using ingredients and flavour compounds vectors and their ability to make prediction on the other two collections. Some similar trends in terms of ingredients usage and flavour were revealed. For example, the classifiers trained on *Xiachufang* and *Allrecipes* collections performed well not only on their own, but also performed reasonably when differentiating the appreciated and less appreciated recipes from each other collections, at least achieving an accuracy higher than 0.50. In addition, they both perform very poorly on the *Kochbar* collection. The classifiers trained on recipes from *Kochbar*, the German portal, however, show something different. Specifically, the classifier trained using ingredients of the German recipes performed only well on *Kochbar* collection, but its performance is poor when applied on *Xiachufang* and *Allrecipes* collections. While the *Kochbar* classifier trained using the flavour compounds performed slightly better on *Allrecipes* collection (ACC > 0.5), but still very poorly on *Xiachufang* recipes. These results hint that, the Chinese and US share more similarity of preference in terms of ingredient usage and flavour than those of German. This finding is surprising, since it shows discrepancy with previous work (Sajadmanesh et al., 2017), which suggests food choices on the flavour aspects of Western countries are more similar to each other than that of Oriental cultures. Moreover, the trends of food preference across cultures on the flavour aspects revealed by this study are also different from that revealed by the experiments on the visual aspects, from which the US and German share more similarities, while the Chinese visual preference is the outlier.



(a)



(b)

Figure 6.7: The best performing model for each collection and their performance on the other two collections. **(a)** the classifier trained using ingredients TF-IDF vectors. **(b)** the classifier trained using flavour compounds TF-IDF vectors.

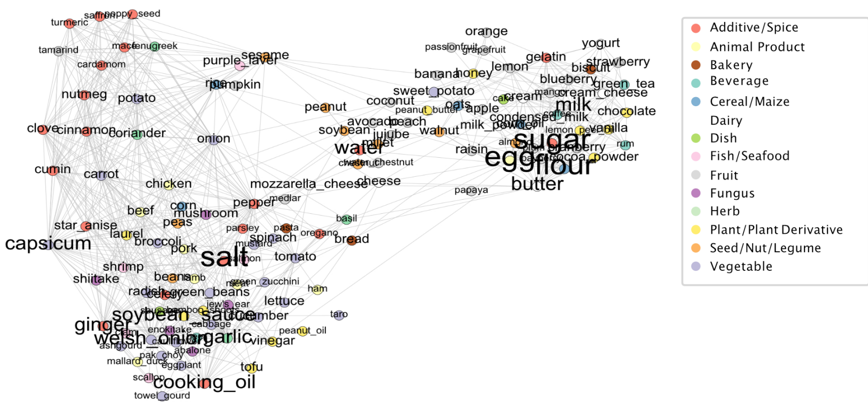
To summarise, the *RQ2b* can be addressed as: It is possible to ascertain stable patterns of food preference across cultures in terms of ingredient usage and flavour, at least between Chinese and US cultures. In the next step, I present the results of the two exploratory experiments, which provide explanation for the patterns revealed by the algorithms.

6.3.4 Justifying the Discovered Patterns in Cross-Cultural Flavour Preferences with the Exploratory Analyses (RQ3)

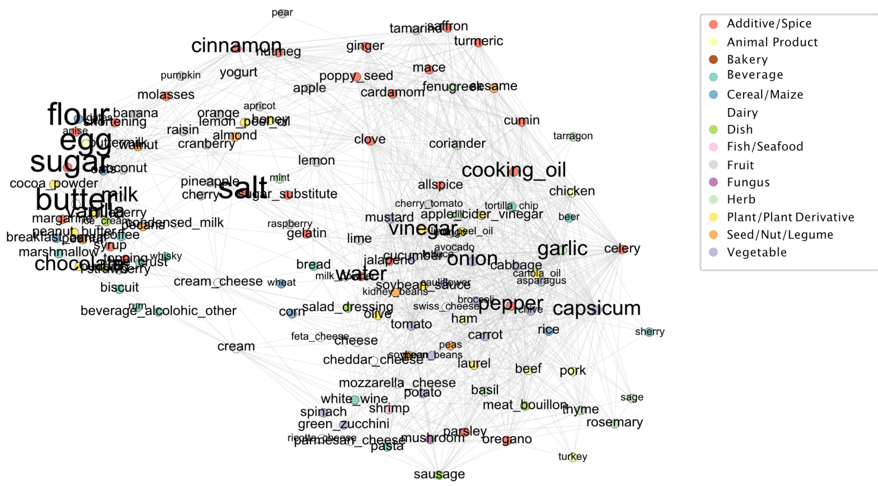
In this section, I demonstrate the results of the two exploratory analyses that were designed to provide explanations for the patterns of the cross-cultural food preference on the flavour aspects revealed by the machine learning approaches. In the following subsections, I first show the description systems I built for describing the flavour, then illustrate their abilities for explaining the patterns. The extent to which these can address *RQ3* are summarised at the end of the section.

6.3.4.1 Preferences for Savoury and Sweet Recipes Across Cultures

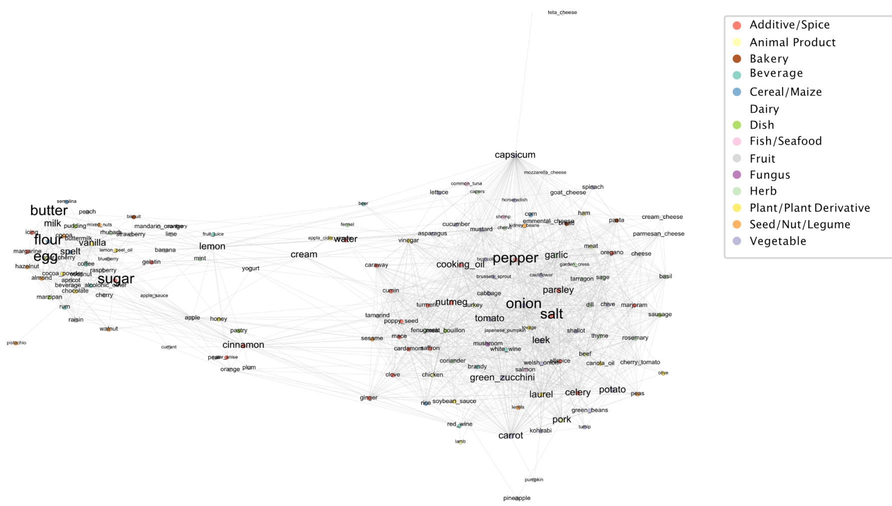
The first exploratory analysis attempts to construct a description system for flavour by means of labelling the recipes as savoury or sweet based on the ingredient complement



(a)

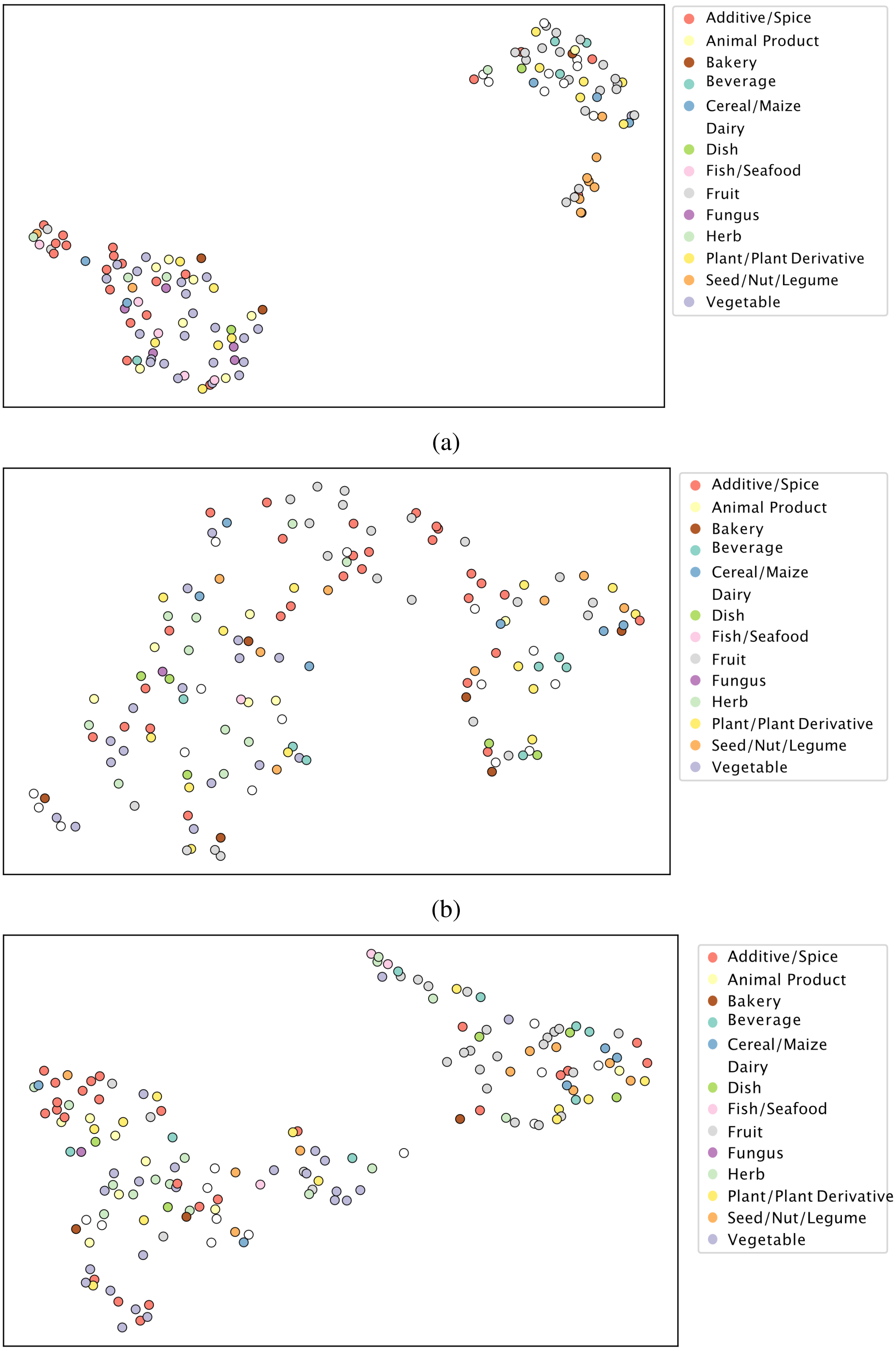


(b)



(c)

Figure 6.8: Ingredient complement network of recipes from **(a)** *Xiachufang* **(b)** *Allrecioes* and **c** *Kochbar*. The nodes represent the ingredients coloured by the categories they belong to. Two of them share an edge if they occurred together and the NPMI between them exceeds 0.10. The size of the node label reflects the frequency of the ingredients appeared in the recipes.



(a)

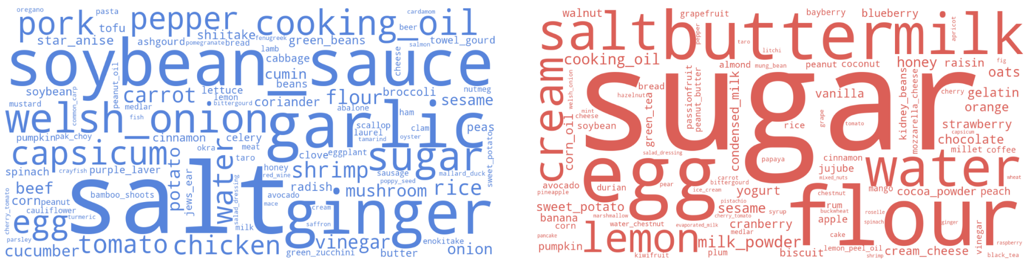
(b)

(c)

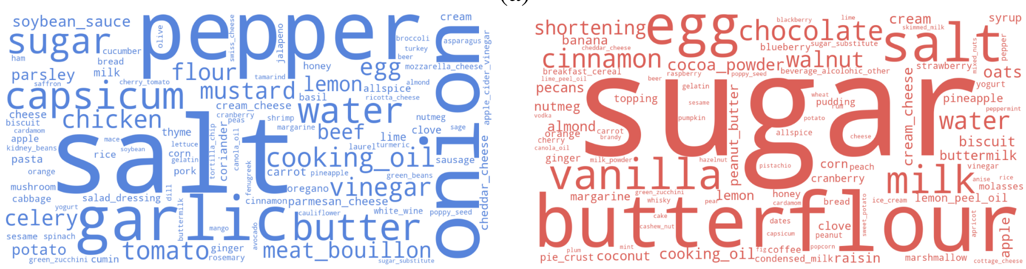
Figure 6.9: The Node2Vec embeddings of ingredients in the ingredient complement networks of recipes from **(a)** *Xiachufang* **(b)** *Allrecipes* and **(c)** *Kochbar*.

network. Figure 6.8 (a) - (c) present the ingredient complement network for appreciated and less appreciated recipes, sourced from each recipe portal. These show clear subcommunities of savoury and sweet recipes, similar to that published in (Teng et al., 2012).

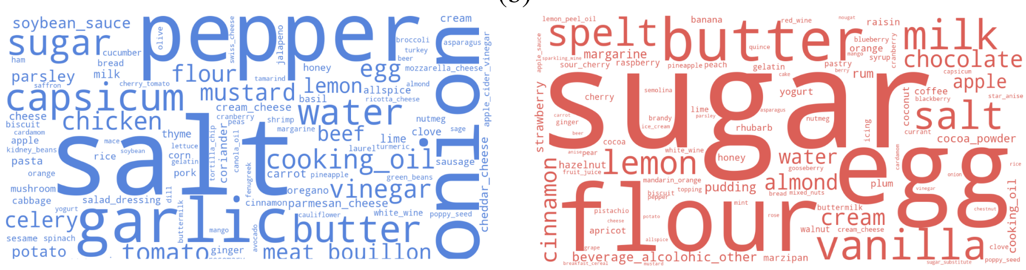
By applying Node2Vec algorithm, I attained the embeddings of the nodes in the ingredient complement networks and visualised them in 2D spaces by mean of UMAP, as shown in Figure 6.9 (a) - (c). Node2Vec learned the role of the nodes play in the networks and mapped them to the feature space, thus the structure of the original networks were kept in these 2D spaces.



(a)



(b)



(c)

Figure 6.10: The Wordclouds of ingredients appear in the savoury and sweet recipes labelled of recipes from **(a)** *Xiachufang*, **(b)** *Allrecipes* and **(c)** *Kochbar*, according to K-Means clustering. The Blue one illustrates the ingredients of savoury recipes, while the red one reflects ingredients of sweet recipes.

Based on the node embeddings representing the ingredients, I attained the representation of the recipes, which were then inputted into K-Means and grouped into 2 clusters. Figure 6.10 shows the WordClouds, which visualise the TF-IDF weight of the ingredients in each group of recipes clustered by K-Means. According to the ingredient lists of recipes in each group, I labelled the recipes as either savoury or sweet. The results of clustering seem plausible since the obvious distinction between savoury and sweet recipes from each recipe collection in terms of ingredients can be identified. For example, in the savoury recipes, salt and its complement ingredients, such as soybean sauce, garlic, ginger in the Chinese recipes, and the pepper, onion in the recipes from US and German recipes, have higher weight. While in the sweet recipes, sugar and the ingredients that are more likely to occur together with it, such as butter, egg, flour, are shown to have higher importance.

Figure 6.11 shows some images of savoury and sweet recipes from the Chinese, US and German portals. It is found that the savoury recipes are more likely to be main dishes, such as stir-fries, Lasagne etc., while sweet recipes are desserts. However, there are also several recipes those are difficult for algorithms to cluster, such as the recipe from *Xiachufang*, the *salted lychee*, which is marked with red frame in Figure 6.11. This recipe describes pickling lychees in salt water, that gives the sweet lychee a salty flavour. Thus, whether the recipe is savoury or sweet is difficult for the algorithms to establish. This example suggests the methodology of generating ingredient complement networks can only cluster savoury and sweet recipes roughly. Human judgement is necessary for more accurate savoury and sweet recipe classification.

Figure 6.11: Some examples of savoury and sweet recipes labelled according to K-Means clustering. The left part shows the examples of savoury recipes from each recipe portal, and the right part displays the sweet ones. The dish that was marked by a red frame is *salted lychee* from *Xiachufang*, which is hard to be determined as savoury or sweet.

Overall, I constructed a description system for the food flavours by means of building ingredient complement networks and unsupervised learning approaches. Up to now, all appreciated and less appreciated recipes in each collection can be described as savoury or sweet recipes. As a next step, I measured the proportion of savoury and sweet recipes in the appreciated and less appreciated recipes in each recipe collection, in order to investigate whether the food cultures have the same tendency to prefer savoury or sweet food. The result is shown in Figure 6.12. In the Chinese recipe collection, sweet recipes are more prevalent than savoury recipes in both appreciated and less appreciated groups. In other words, Chinese users show no obvious preference for either savoury or sweet recipes. In the US collection, the savoury recipes account for almost 60% appreciated recipes, while 75% of the recipes in the less appreciated group are sweet. This finding suggests US users prefer savoury recipes to the sweet ones, at least in terms of metric I selected as the proxy of food preference, # bookmarks. In contrast, according to the proportion of savoury and sweet recipes in *Kochbar* collection, I found that, the proportion of savoury and sweet recipes is almost equal in appreciated recipes (50%), while in the less appreciated recipes, the savoury recipes (60%) are much more than the sweet recipes (40%). It suggests an opposing trend of flavour preferences between the US and German, which is mainly demonstrated in the less appreciated recipes, that is, US people prefer sweet food less while German people show less preference for savoury dishes.

This observation could account for the findings in Section 6.3.3, where the *Allrecipes* and *Kochbar* classifiers did not perform well on the other's recipe collection. Nevertheless, I found no strong signal to support the relatively better performance of *Xiachufang* and *Allrecipes* classifiers on each other collections.

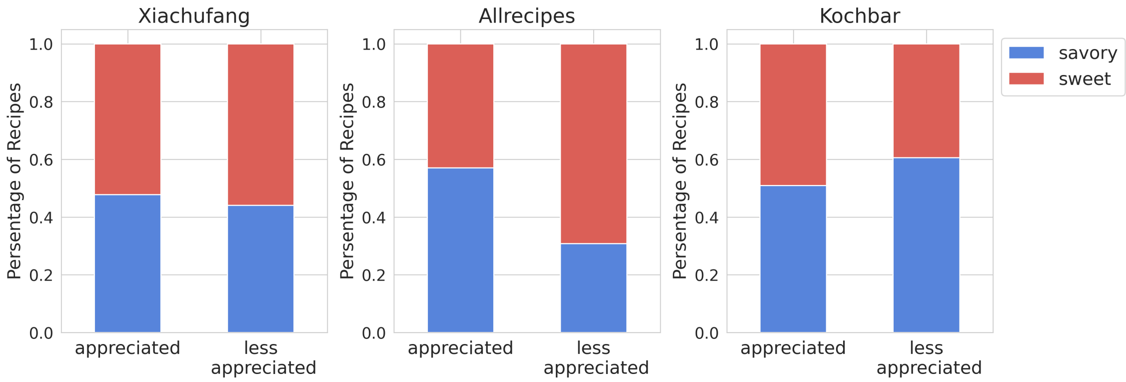
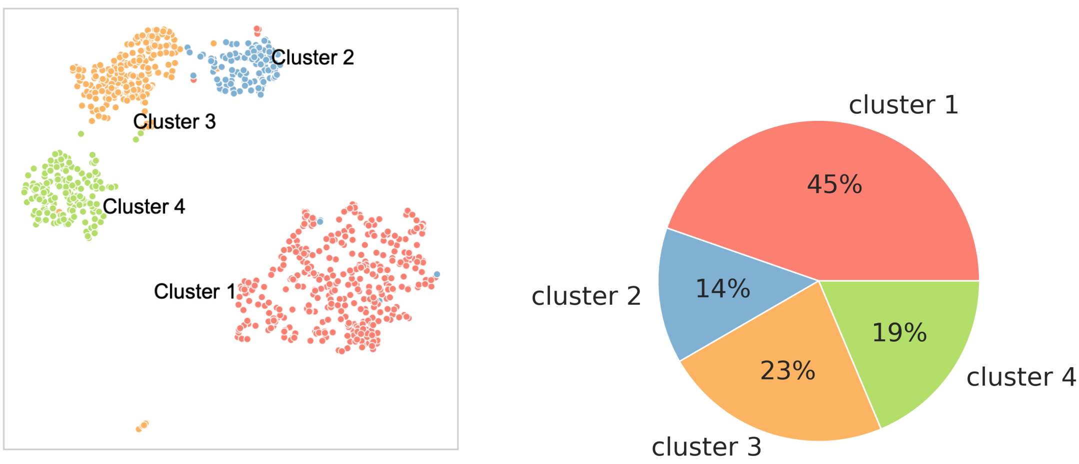


Figure 6.12: The Proportion of Savoury and Sweet Recipes in Appreciated and Less Appreciated Recipes in Each Collection

6.3.4.2 Preferences for Sweet and Non-sweet Flavours Across Cultures

The second exploratory analysis attempts to construct the description of flavour with the semantic description of flavour compounds, namely, flavour profiles. By means of applying K-Means clustering on the flavour compounds represented by their corresponding flavour profiles, I grouped them into four distinct clusters, which are shown in Figure 6.13 (a). The ratio of flavour compounds in each cluster is shown in Figure 6.13 (b). Almost the half of the flavour compounds were clustered into the cluster 1, while the other half were clustered into the other three clusters.

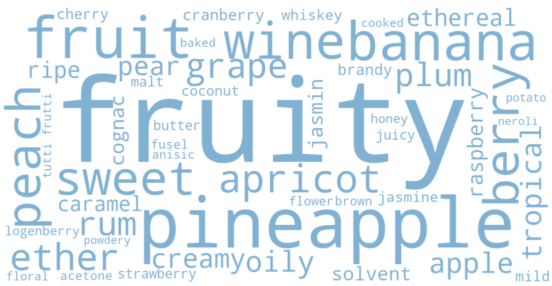


(a) (b)

Figure 6.13: The result of K-Means clustering of flavour compounds. **(a)** The 4 clusters displayed in a 2D space by means of UMAP. **(b)** The ratio of flavour compounds in each cluster.

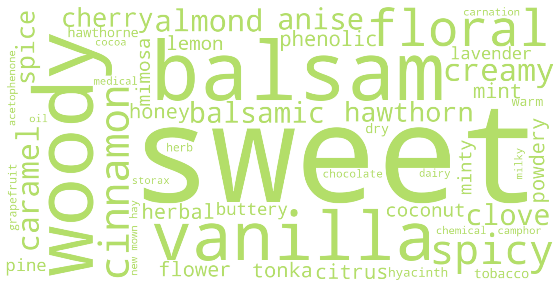
The 50 representative flavour profiles of the flavour compounds in each cluster were then determined by Equation 6.6. I visualised these with WordClouds shown in Figure 6.14. The larger the size of the flavour profile in the WordClouds, the more representative they are. The flavour compounds in cluster 1 are represented by the flavour profile such as bitter, odourless, etc., while in the other three clusters, the flavour profiles contain fruity, sweet and other similar words. According to this finding, I grouped the flavour compounds into 2 categories, which are non-sweet compounds (i.e., those in the cluster 1), and the sweet compounds (i.e., those in the cluster 2 - 4). This is the second description system I constructed, and it would be applied to investigate the patterns of food preferences across culture on the flavour aspects. The ratio of non-sweet and sweet flavour compounds in the appreciated and less appreciated recipes in each collection would be calculated. The overall ratio of non-sweet (45%) and sweet flavour compounds (55%) were treated as baseline to see whether the Chinese, US and German food cultures have the same or different trends towards non-sweet/sweet flavours.

(a) Cluster 1



(b) Cluster 2

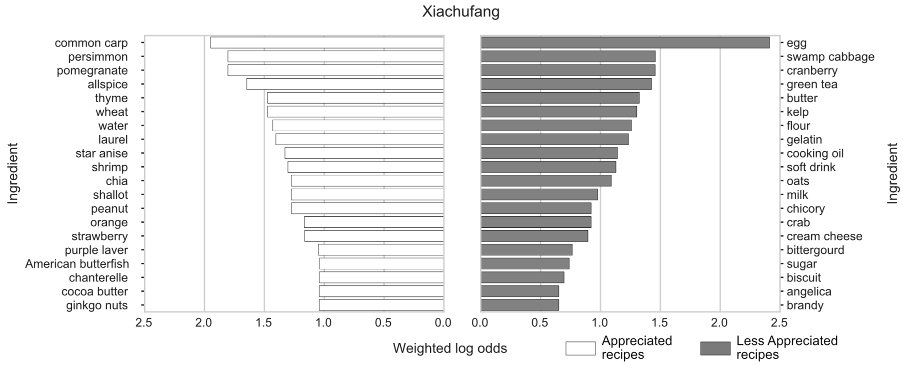
(c) Cluster 3



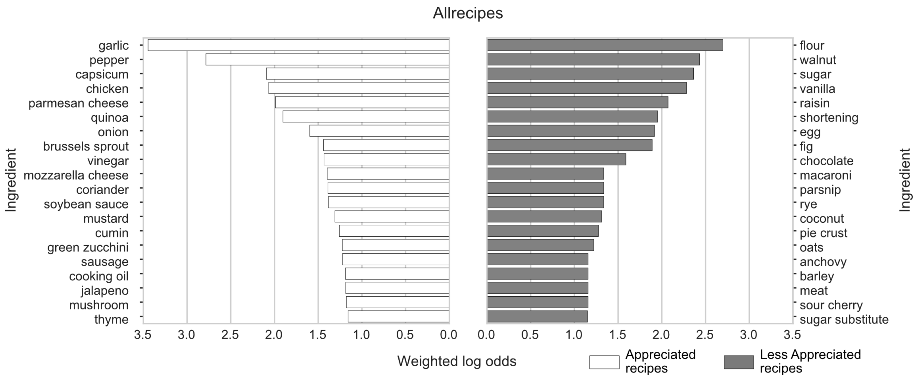
(d) Cluster 4

Figure 6.14: The Wordclouds of the representative flavour profiles of the flavour compounds in **(a)** Cluster 1, **(b)** Cluster 2, **(c)** Cluster 3 and **(d)** Cluster 4.

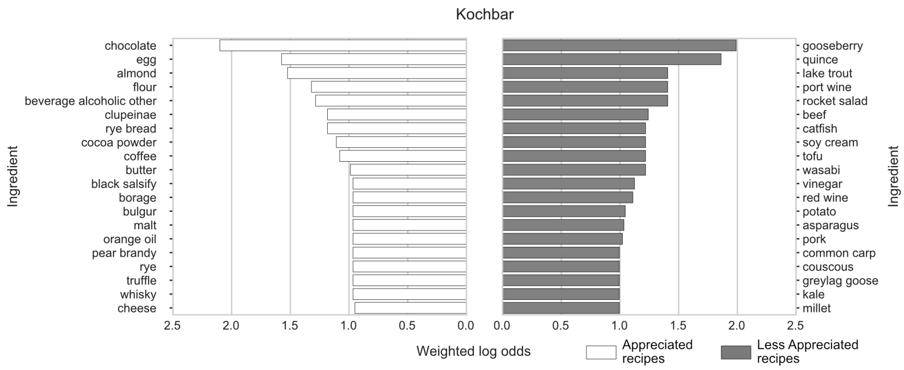
The top 50 distinctive ingredients of appreciated and less appreciated recipes were determined according to their weighted log odds and applied to represent the appreciated and less appreciated recipes in each collection. Figure 6.15 (a) - (c) show the top 20 distinctive ingredients of each group. The distinctive ingredients, to some extent, show a similar trend to that revealed by the ingredient complement networks. For example, in the appreciated recipes sourced from *Allrecipes*, there are more ingredients from the savoury cluster (e.g., garlic, pepper, etc.), while more ingredients from the sweet cluster (e.g., flour, sugar etc.) are found in the less appreciated recipes. In contrast, ingredients in sweet cluster such as chocolate and almond, are distinctive in appreciated recipes in *Kochbar* collection. Whereas in the less appreciated recipes, there are more distinctive ingredients in savoury recipes, such as lake trout and beef. These findings suggest that savoury foods



(a) Xiachufang



(b) Allrecipes



(c) Kochbar

Figure 6.15: The distinctive ingredients of appreciated and less appreciated recipes from **(a)** *Xiachufang*, **(b)** *Allrecipes* and **(c)** Kochbar.

are more likely to be appreciated by US while sweet foods might be appreciated by the German users. This indicates the opposing trends in the flavour preference of US and German users. The flavour description system constructed in the experiment validated the finding, and the results are shown in Table 6.3.

Table 6.3 presents the ratio of non-sweet and sweet flavour compounds in the distinctive ingredients of appreciated and less appreciated recipes in *Xiachufang*, *Allrecipes* and *Kochbar* collections. In the recipes appreciated and less appreciated by the users of *Xiachufang*, the ratio of non-sweet and sweet flavour compounds is very close to the baseline, which is the original ratio of non-sweet (0.45) and sweet compounds (0.55) obtained from the cluster analysis. However, obvious differences are illustrated in the appreciated recipes sourced from *Allrecipes* and *Kochbar*. To be specific, the ratio of non-sweet

Table 6.3: The ratio of non-sweet and sweet flavour compounds in the appreciated and less appreciated recipes in each recipe collection

	Xiachufang		**Allrecipes**		**Kochbar**	
	non-sweet flavour compounds	sweet flavour compounds	non-sweet flavour compounds	sweet flavour compounds	non-sweet flavour compounds	sweet flavour compounds
Appreciated recipes	0.45	0.55	0.47^{+}	0.53^{-}	0.42^{-}	0.58^{+}
Less appreciated recipes	0.46	0.54	0.45	0.55	0.45	0.55

Note: The baseline ratio of non-sweet to sweet flavour compounds is 0.45 and 0.55. + means the ratio is above the baseline, and - means the ratio is below the baseline.

flavour compounds in appreciated recipes from *Allrecipes* is slightly higher than the baseline (ratio = 0.47), and that of sweet flavour compounds is lower than the original ratio (ratio = 0.53). But the opposite is true for the appreciated recipes in *Kochbar* collection, in which the ratio of non-sweet flavour compounds is 0.42, lower than the baseline, and that of sweet compounds is 0.58, higher than the baseline. These findings support the patterns revealed by the algorithms as reported in Section 6.3.3, which is, US and German food preferences are different and do not transfer.

To summarise, both of the exploratory analyses provide descriptive systems, which divide flavours of recipes into two types: savoury/non-sweet and sweet, providing a means to address RQ3: both system help explain the cross-cultural flavour preferences that were revealed by the algorithms. The evidence helps to explain the poor performance of *Allrecipes* and *Kochbar* classifiers on differentiating appreciated and less appreciated recipes in each other collection, yet there are no findings that justified the stable patterns in flavour preference shared by the Chinese and US users.

6.4 Summary and Discussion

6.4.1 The Primary Findings

- *RQ1.* The performance of machine learning models shows that the recipes from Chinese, US and German portals are distinct, both in terms of ingredients and flavour. Classifiers trained using ingredient and flavour compounds vectors both perform with relatively high accuracy. Overall, the recipes from the Chinese portal were classified most accurately, while recipes from US and German portals are more likely to be confused. The predictive power of ingredients is slightly better than that of flavour compounds (ingredient ACC = 0.81; flavour ACC = 0.77) in this classification task. In addition, TF-IDF vectors outperform Word2Vec embeddings in all cases of the food classification task.

- *RQ2a.* It is possible to distinguish between appreciated and less appreciated recipes algorithmically for all food cultures in this work, although the prediction accuracy is lower than that in the recipe origin classification task. The classifiers using ingredients and flavour compounds do not show significant differences in this task. Still, TF-IDF outperforms Word2Vec embeddings.

- *RQ2b.* This study indicates the existence of cross-cultural stable patterns in flavour preference by transferring the classifier with the highest accuracy on each collection. The Chinese and US flavour preferences seem to be more similar to each other than to those of the German users. Specifically, the classifier trained on *Allrecipes* made useful predictions for *Xiachufang* collection and vice-versa (ACC > 0.50), whereas both perform poorly on the *Kochbar* collection. This was the case, regardless of whether the classifiers were trained using ingredients or flavour compounds. In addition, the classifier trained on *Kochbar* recipes with ingredients performed reasonably only on its own datasets, but very poorly on the other two datasets (ACC < 0.50), while the classifiers trained on *Kochbar* with flavour compounds performed slightly better on the *Allrecipes* collection, but still poorly on the *Xiachufang* collection.

- *RQ3.* The two exploratory analyses, which generated ingredient complement networks and clustered flavour compounds based on their corresponding flavour profiles, provide descriptive systems for food flavour. These help explain the patterns of cross-cultural food preferences revealed by the algorithms, and in particular why classifiers trained on *Allrecipes* and *Kochbar* collections performed poorly on each other. Both the experiments suggest opposing trends in the flavour preferences of US and German users. Specifically, US users show higher appreciation for recipes with savoury or non-sweet flavours overall, but sweet recipes are more likely to be less appreciated by them. In contrast, there are more recipes with sweet flavour in German users' appreciated recipes, but they show lower appreciation on the savoury or non-sweet recipes.

6.4.2 The Implication of this Study

This section discusses the findings and lists several practical implications and limitations of the findings of the results presented above.

I first underline the implications of the study with respect to the development of cross-cultural food recommender systems. The results reveal that flavour information can be modelled using favour compounds. The classifiers trained using the vectors of flavour compounds are able to distinguish recipes sourced from different portals and predict, for recipes in each portal, whether recipes will be appreciated. This suggests, in addition to visual information in recipe images, a further aspect of food aesthetics, flavour, can be used to explain human food choices within and across cultures. This is a strong indication that the flavour information has the potential to be exploited to make acceptable food recommendations, and culture should be accounted for the development of the systems. The outcomes of the transfer learning indicate the existence of stable patterns in cross-cultural flavour preferences. To the best of my knowledge, this is the first attempt to study cross-cultural food preferences in this way. This study can be seen as an initial, and promising step to provide food recommendation with flavour information in a cross-cultural context.

In addition to the implications for developing food recommender systems, the study also demonstrates some other practical implications. The first is with respect to recipe representation. By comparing the two approaches applied for representing recipes with ingredients and flavour compounds for food classification, it was found that TF-IDF transforms outperformed Word2Vec algorithm. Word2Vec captures and learns the association of words, which have been proven to be useful in tasks involving the relationships among ingredients, such as ingredient substitution for food recommendation (Pellegrini et al., 2021) and cuisine style transformation (Kazama et al., 2018), but it did not improve the performance of food classification. This suggests that Bag of Word approaches are more suitable than word embeddings in representing recipes for food classification. Besides word embeddings, the node embeddings algorithm, Node2Vec, was also applied to recipe representation in one of the exploratory analyses. Although the node features of recipes have not been involved in the food classification tasks, applying the unsupervised learning algorithm on these showed promising results in terms of recipe flavour description. This is only an initial attempt of introducing node embeddings for studies in the food domain. Recent work from Park et al. (2021) have applied Node2Vec for food recommendation based on the relationship between ingredients and flavour compounds. They have captured the connected relations of recipe-ingredient, recipe-recipe, recipe-user within a heterogeneous graph with node embeddings in order to develop food recommender systems. This investigation suggests node embeddings technique is a prospective means at the application level. According to the findings in this doctoral work, I assume culture would be a valuable node that can be embedded into graphs and represented by means of node embeddings for developing food recommender systems. The other implication from the exploratory experiments is the systems I constructed to describe the flavour of recipes and flavour compounds. I attempted to make the flavour information, especially that provided by the flavour compounds, more interpretable with their corresponding flavour profiles. This would provide understandable flavour description when flavour information is incorporated in food recommender systems. However, these are preliminary investigation of flavour description with computational approaches, whether the results work in application scenarios requires further analysis, such as evaluating the flavour description systems by means of human perception and judgements.

Up to now, I summarised the findings and underlined their contributions with respect to food recommender systems. However, the study remains several open questions. Firstly, I tried to cluster recipes into savoury and sweet groups by means of ingredient complement networks, and these worked in suggesting the opposite trends of food preferences between US and German people. Nevertheless, it is found that the savoury group is more likely to contain dishes served as main dishes such as meatloaf, stir-fries etc. In contrast, there are more desserts, such as cakes, cookies etc. in the sweet group. Thus, the type of dishes becomes a noise for investigating food preferences in terms of flavour. This leads to questions such as *"To what extent the savoury and sweet recipes can be classified if the dish type is limited to main dishes?"* If the answer is yes, *"Whether the differences of preferring savoury and sweet still exist between US and German users?"* is still needed to be clarified. Making these questions clear would be helpful to prove the robustness of the explanation with this exploratory experiment. Furthermore, the outcomes of my two exploratory experiments suggest that *Allrecipes* users have relatively low appreciation for sweet recipes. This goes beyond the knowledge that US people consume much or even too much sweet food, which causes severe issues related to public health, such as high rates of obesity, diabetes, and heart disease (USDA, 2020). Hence, whether US people show low appreciation in sweet food on *Allrecipes* requires further analysis. In this doctoral work, the number of bookmarks was selected from *Allrecipes* as the proxy to indicate whether the recipes were preferred. However, there are also other metrics such as the number of ratings and comments the recipes received during a period (e.g., one week/month/year), which indicates the popularity of the recipes (Wagner et al., 2014; Trattner et al., 2018) and the sentiment in the comments, which show users' attitude towards the recipes, etc. The association among these metrics needs to be investigated in order to understand users bookmarking behaviours, especially whether it indicates their preference to the recipes, and whether would it be a situation such as *"Users rate sweet recipes more often and comment more positively, while bookmaking them less frequently"*. Figuring this out would be beneficial to understand US users' flavour preference for non-sweet and sweet recipes, and provide more acceptable food recommendations in terms of flavour.

6.5 Chapter Summary

In summary, the performance of machine learning models indicates that the flavour information is promising to be applied into cross-cultural food recommender systems. However, in order to better understand the food preferences across cultures with the computational approaches, further analysis is necessary, such as minimising the influence of other factors except flavours (e.g., dish types), and incorporating more interaction data (e.g., ratings and comments) between users and food items online.

Chapter 7

Fusion of Visual Features and Flavour Compounds for Cross-Cultural Food Preferences Prediction

7.1 Introduction

In the previous empirical chapters, i.e., Chapter 5 and 6, visual features and flavour compounds were investigated to predict online food preferences, respectively. The outcomes show that human food preferences can be explained by the aesthetics of food to some extent, suggesting the promise of incorporating aesthetic features in the development of food recommender systems, one crucial function of which is providing food that users would like to eat. In addition, stable patterns of aesthetic food preferences across cultures were revealed, suggesting the possibility of developing cross-cultural food recommender systems by applying food representations relating to visual appearance and flavour.

In this chapter, an experiment to fuse visual features encoded in the online recipe images and the flavour compounds of online recipe ingredients for predicting cross-cultural food preferences is designed. It is an initial attempt of employing the findings from previous empirical work to develop food recommender systems. The experiment aims to address the following research question:

- *RQ*. To what extent is it possible to combine the decision made by the classifiers trained using visual features and flavour compounds to improve the performance of cross-cultural food prediction model?

Answers to this research question would be beneficial to address the *Issue 3*, which was raised in Chapter 1.

This chapter is structured as follows: Section 7.2 describes the methods for the experiment, the results are presented in Section 7.3. In Section 7.4, a summary of the results is given.

7.2 Methods

This section describes the data preparation for the experiments before demonstrating the approaches applied to classifiers fusion.

7.2.1 Data Preparation

All the appreciated and less appreciated recipes (i.e., top-10% and bottom-10% recipes based on the appreciation metrics) from *Xiachufang*, *Allrecipes* and *Kochbar* collection are combined as the cross-cultural recipe data basis for this experiment. In the experiments on the visual aspects, there are the same number of top-10% (n=2,500) and bottom-10% (n=2,500) recipes based on the appreciation metric sampled from each collection. However, there are some recipes from this group that do not have full flavour information. These were filtered out in the experiments on the flavour aspects. In this chapter, I kept only recipes with both full visual and flavour information. Finally, there are 6,363 recipe images involved in this experiment, in which there are 2,909 appreciated recipes and 3,454 less appreciated ones. The number of appreciated and less appreciated recipes is similar, indicating the classes are balanced.

7.2.2 Classifiers Combination

In order to fuse the visual features and flavour compounds in predicting cross-cultural food preferences, ensemble learning techniques are applied. The main idea of ensemble learning is to fuse the decision made by the individual classifiers trained using different feature sets to determine whether it classifies samples more accurately. To be specific, visual features and the vector of flavour compounds were applied to train classifiers to identify appreciated and less appreciated recipes separately, then these classifiers were combined by means of several fusion schemes (describe in step 3 in detail). Then the performance of the individual classifiers and the combination of classifiers were reported and compared. In addition, in order to compare the performance of aesthetic information with the traditional text information in the same task, the vector of ingredients was involved to train the baseline model. The steps are described as follows:

Step 1: Splitting the training and test set
The dataset containing the appreciated and less appreciated recipes from all three recipe portals was split for training (70%) and testing (30%), respectively.

Step 2: Training the individual classifiers
In this step, I trained the classifiers using visual features, flavour compounds and ingredients on the training set respectively, with the target value being the appreciation metric. A combination of all visual features, namely, the EVF, Colour Histogram, LBP, BoVW and VGG16, were applied to train the visual classifiers. This is because in Chapter 5, in the most cases, the combination of visual features performed better. In order to train the classifiers using flavour compounds and ingredients, TF-IDF was applied to generate the vectors, since in Chapter 6, TF-IDF outperformed. Three algorithms, **Naïve Bayes (NB)**, **Logistic Regression (LoG)** and **Random Forest (RF)** were applied to train the models. A five-fold Randomised Search CV was employed to determine the optimal parameters for LoG and RF. The accuracy (ACC) of the classifiers on the training set was measured and compared to identify the best performed individual classifiers, which were then applied to make prediction on the test set. For each recipe in the test set, I recorded the probability given by the individual classifiers for predicting whether the recipe is appreciated. The threshold of the class probability is 0.5 due to it is a binary classification task. Thus, the value of the class probability should be greater than 0.5 if the recipes were

predicted to be appreciated, otherwise it should be less than 0.5. The class probability is an important variable for combining the classifiers, which were described in the next step.

Step 3: Combining the individual classifiers
In this step, I applied two schemes, voting, and stacking, to combine the individual classifiers from the previous step. The process of voting scheme is shown in Figure 7.1. Firstly, the classifiers trained using visual features, flavour compounds and ingredients were applied to make prediction on each recipe in the test set, respectively. Then I recorded the class probability provided by each individual classifier for each recipe. The decision of the individual classifiers was fused by averaging the class probability. This is a common approach in ensemble learning, particular for bagging and boosting strategies (Zenko, Todorovski, & Dzeroski, 2001). The averaged class probability indicates the likelihood of the combined model for predicting whether the recipes are appreciated ones. According to this, the predicted class of the recipes based on the combined model can be obtained.

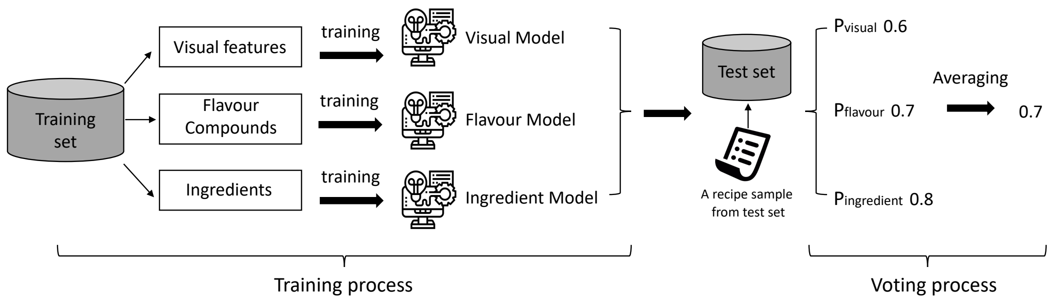


Figure 7.1: The process of fusing the individual classifiers by means of voting. P_{visual}, $P_{flavour}$ and $P_{ingredient}$ refers to the class probability of each recipe in test set obtained from the classifiers trained using visual features, flavour compounds and ingredients respectively.

The second scheme I applied is stacking. The process is shown in Figure 7.2. The basic idea of stacking is to train multiple classifiers and obtain their predictions, which then applied as features to train the meta model. The output of the meta model indicates the predictions made by the combined classifiers. In this experiment, in order to get features for each recipe to train the meta model, I trained each individual classifier, i.e., that trained using visual features, flavour compounds and ingredients in 5 folds. Specifically, in this experiment, the training set was split into 5 folds, the classifier was trained on the four folds and made predictions on the holdout set (the 5th fold). This process was repeated for 5 times until each sample in the whole training set got the prediction. I kept the class probability of each individual classifier as the features, and the appreciation metric of the recipes as the target value for training the meta model. From the process, the features for the recipes in the test set were also obtained, which were the averaged probability got from each fold of training. The **LoG** algorithm was applied to train the meta model. The predictions for this meta model were considered as the final prediction made by the combined models. An advantage taken of stacking compared to voting is it weights the performance of each individual models during the ensemble learning, while the voting strategy weights each individual classifier equally, regardless of its performance on the food preference prediction task. The aim of applying both strategies is to compare them and determine a better means for involving visual and flavour information to food recommendation at the same time.

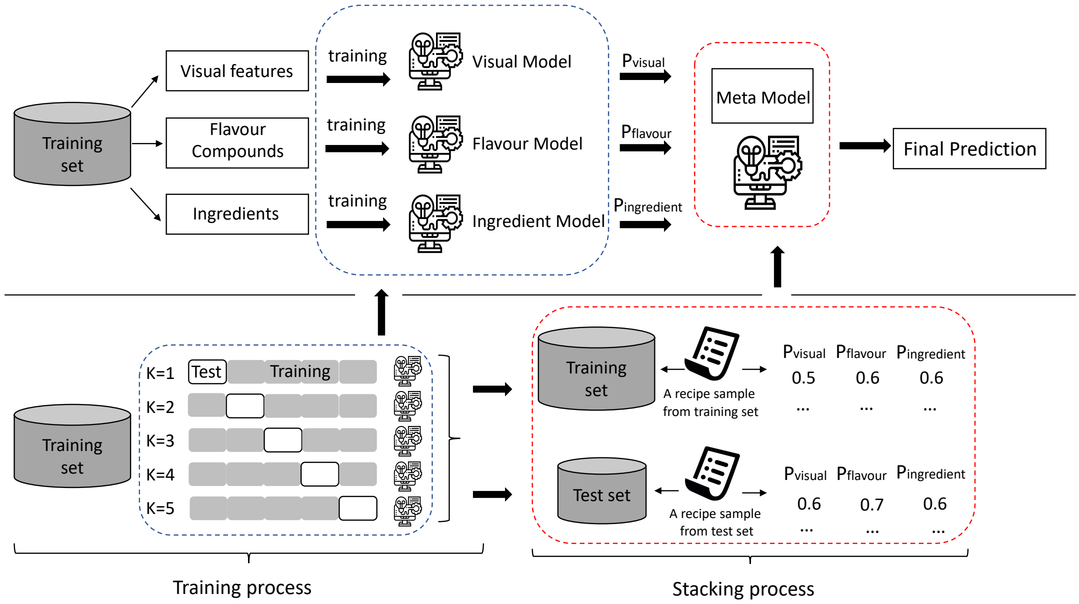


Figure 7.2: The process of fusing the individual classifiers by means of stacking. P_{visual}, $P_{flavour}$ and $P_{ingredient}$ refers to the class probability of each recipe in test set obtained from the classifiers trained using visual features, flavour compounds and ingredients respectively.

Step 4: Evaluating the models

The performance of all individual and combined classifiers was evaluated with ROC curve (receiver operating characteristic curve) and AUC (Area under the ROC Curve) score. The ROC curve is plotted with True Positive Rate (TPR) against False Positive Rate (FPR), where TPR is

$$TPR = \frac{TP}{TP+FN} \tag{7.1}$$

and FPR is

$$FPR = \frac{FP}{FP+TN} \tag{7.2}$$

AUC refers to the area under the ROC Curve. It ranges from 0 to 1. Generally, the higher the AUC (close to 1), the better ability of the model to classify the samples. I visualised the performance of the models with ROC Curves and reported the AUC in the next section.

7.3 Results

The performance of the individual classifiers, and the combination of classifiers with two strategies is shown in Figure 7.3, which can help with addressing the *RQ* raised in Section 7.1. Firstly, when differentiating the appreciated and less appreciated recipes derived from the Chinese, US and German recipe portals, flavour compounds outperformed the visual features slightly. The AUC of the classifiers trained using visual features and flavour compounds is 0.618 and 0.640, respectively. It is in line with the results of predicting food preferences within each culture, where flavour compounds offer slightly stronger predictive power in the most cases. Neither visual features nor flavour compounds outperformed the baseline feature, ingredients (AUC = 0.646), but the predictive power of these three feature sets is comparable. This finding suggests that both visual features and flavour compounds can perform reasonably in predicting food preference in a cross-cultural context.

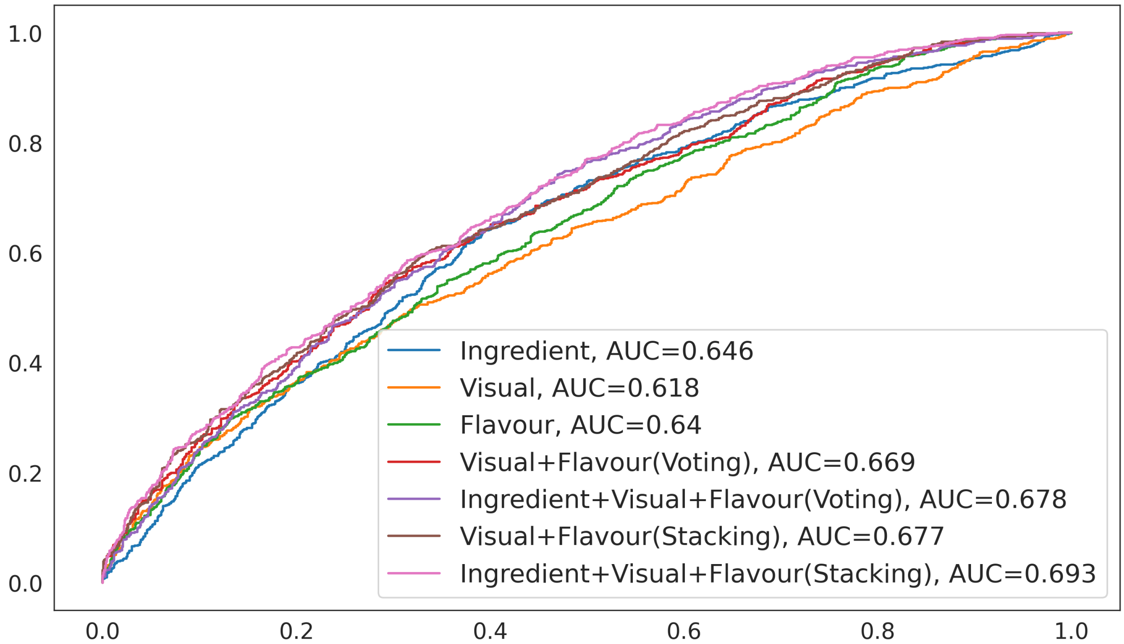


Figure 7.3: ROC Curve of the individual models and the combination of models for differentiating the appreciated and less appreciated recipes derived from the recipe portals

The combination of the classifiers performed significantly better than the individual classifiers in all cases, and in comparison of the two strategies of ensemble learning, stacking performed slightly better than voting overall, which is shown in Figure 7.3. In addition, a combination of the classifiers trained using visual features and flavour compounds outperformed not only the classifiers trained using these two feature sets separately, but also the baseline model that trained with ingredients. The AUC of the combined classifier trained using visual features and flavour compounds is 0.669 (with voting strategy) and 0.677 (with stacking strategy), respectively. This result is truly promising for the development of the cross-cultural food recommender systems with food representations relating to aesthetics. Moreover, a combination of all three classifiers provides even higher AUC (AUC = 0.678 and 0.693), indicating the combination of visual features and flavour compounds not only outperformed the baseline, but also enhance the performance of the baseline model. These results can address the *RQ*, that is, it is possible to improve the performance of cross-cultural food prediction model when combining the predictions provided by visual features and flavour compounds.

7.4 Implication of this Study

The outcomes of this experiment demonstrate the ability of the aesthetic information in differentiating appreciated and less appreciated recipes across recipe portals, either applying them individually or as a combination. Combining the decisions made by classifiers trained using both feature sets performed much better and outperformed the baseline model, which trained using the semantic information, i.e., ingredients of food. This finding suggests the promise of developing food recommender systems by involving the cross-cultural food preferences on the visual and flavour aspects, which are anticipated providing food recommendations for people who move between cultures.

However, this experiment is just an initial attempt of utilising the findings from Chapter 5 and 6 for cross-cultural food recommendation, and the possibility of developing the food recommender systems is only illustrated with algorithms. Further analysis, especially a user study for testing the performance of the aesthetics-based food recommender system in application scenario is in demand.

7.5 Chapter Summary

In this chapter, I attempted to predict food preference in a cross-cultural context by applying the food representations on the aesthetic aspects, i.e., the visual appearance and flavour. The models show reasonable performance, and I also found that applying a combination of visual and flavour information of food improve the performance of cross-cultural food preferences prediction. The findings from this chapter are beneficial to developing cross-cultural food recommender systems, however, a test in a practical application setting is needed.

Chapter 8

Discussion

8.1 Introduction

In this chapter, I review and discuss the results in the main empirical chapters of this thesis: Chapter 5, 6 and 7. Chapter 5 and 6 investigated cross-cultural human food preference, as modelled by two aesthetic aspects of food, visual appearance and flavour. The results display how and to what extent the visual appearance and flavour are different across cultures, but also reveal stable patterns in human food preferences with respect to these aesthetic aspects across cultures. In Chapter 7, I attempted to fuse the visual and flavour information for cross-cultural food recommendation. Results from the empirical chapters have addressed the Issues outlined in Chapter 1:

- *Issue 1.* To what extent is it possible to differentiate the food across cultures based on the representation of food relating to visual appearance or flavour?
- *Issue 2.* To what extent is it possible to identify the differences and ascertain stable patterns of food preferences across cultures based on the representation of food relating to visual appearance or flavour?
- *Issue 3.* To what extent is it possible to utilise the stable patterns of food preference across cultures on visual and flavour aspects to build cross-cultural food recommender systems?

In this chapter, I discuss the answers to these issues by comparing these to previous work in related domains. Besides summarising the theoretical implications of the results, I also discuss what the discoveries from the empirical work mean with respect to developing food recommender systems.

This chapter is structured as follows: Section 8.2 discusses the theoretical implications of the results from this work, and the practical implications of these are discussed in Section 8.3. The limitations of this work are presented in Section 8.4.

8.2 Theoretical Implications of this Work

In this work, it is revealed the aesthetic differences both in terms of the food across cultures, and in terms of the preferences for food within the cultures with online recipes and relevant interaction data from different online recipe portals.

Firstly, In Chapter 5 and 6, online recipes, derived from three recipe portals, *Xiachufang.com*, *Allrecipes.com* and *Kochbar.de*, were gathered and selected as the proxy of the three distinct food cultures - China, US and Germany. By modelling the visual features of the recipe images uploaded to these platforms and the flavour compounds associated with the ingredients, I showed that the online recipes from different cultures are sufficiently distinct in terms of visual appearance and flavour, that they can be differentiated with high accuracy (ACC = 0.89 and 0.77, respectively), suggesting the existence of aesthetic patterns of food across cultures. It is known that culture is an important factor in influencing what and how humans eat (Bellisle, 1999; Eertmans, Baeyens, & Van Den Bergh, 2001; Furst, Connors, Bisogni, Sobal, & Falk, 1996; P. Rozin et al., 2002; Vabø & Hansen, 2014), hence, the differences of food across cultures can be observed, such as in terms of ingredients usage (Fischler, 1988; P. Rozin, 1996; Laufer et al., 2015), culinary arts (Anderson, 1988; Sweeney, 2017). Whereas in this work, I broaden the scope by revealing the cross-cultural differences in its visual appearance and flavour patterns. This finding is in line with the observation from (Anderson, 2014), who proposed the aesthetic rules of food are culturally dependent. Moreover, although the evidence from the empirical analyses in this work suggests that the food across cultures is significantly different in aesthetics, it revealed that compared to a more distinct food culture (i.e., China), the relatively similar cultures (i.e., US and Germany) are more likely to share more similarities in aesthetic patterns of food. It is illustrated in that regardless of on the visual or flavour aspect, online recipes from US and German portals are more likely to be confused than those from the Chinese portal. This finding indicates the correlation between cultural difference and aesthetic differences in food. As stated by Sproesser et al. (2022), US and German share the same European cultural origins, hence they share more similarity in food cultures. While China is different, as a representative of Eastern culture, it shows larger difference in food than that between US and Germany, which reflected clearly in how food looks and tastes. These findings might not be new, as traditional anthropological studies have revealed the differences of food appearance (e.g., Palmer & Schloss, 2010; Taylor et al., 2013; Anderson, 2014) and flavour (e.g., Sherman & Billing, 1999) in different cultural groups relied on observation and interviews, this work confirms these findings with a large-scale dataset derived online by means of computational analyses.

Apart from revealing the aesthetic differences in food itself, this work also demonstrates the aesthetic differences of food preferences in each culture. In the empirical experiments in Chapter 5 and 6, the models trained using visual features and flavour compounds could identify the appreciated and less appreciated recipes from each recipe portal with reasonable accuracy (the ACCs at best 0.67 with both aesthetic information). It is well established in previous work that the sensory appeal, including visual appearance and flavour, is one of the most significant drivers of food preferences (Wadolowska, Babicz-Zielinska, Czarnocinska, et al., 2008; Steptoe, Pollard, & Wardle, 1995; Schifferstein et al., 2022), and the algorithmically results add more weight to this argument. In addition, in comparison to the previous work that studied the correlation between sensory inputs and food preferences, such as in psychological and cognitive domains (Druz & Baldwin, 1982; Spence et al., 2016; Michel et al., 2014; Zampollo et al., 2012; Koch & Koch, 2003; Druz & Baldwin, 1982; Johnson & Clydesdale, 1982), researchers were limited to investigate only one aspect of food aesthetics, or even focused only on individual aspects of food visual appearance (e.g., colour, texture, packages, plating) and flavour (salty, sweet, sour, bitter, umami), this work shed light on a perspective of the holistic food aesthetics in influencing food preferences by relying on online recipes and computer science.

After establishing the impact of culture on food aesthetics and the predictive power of aesthetic information in food preferences within each culture, I investigated food aesthetic preferences across cultures. In Chapter 5 and 6, by selecting classifiers with the highest accuracy in each collection and testing its performance for predicting food preference on the other two collections, i.e., learning the knowledge with respect to food preferences of a culture then used it for predicting food preferences of the other cultures, I revealed that aesthetic food preferences can be transferred across cultures, at least between US and Germany on the visual aspects, and between US and China on the flavour aspects. This finding points to a fact, that is, stable patterns in cross-cultural food preferences exist. In contrast to previous work investigating large-scale online recipes with computational approaches (e.g., Ahn et al., 2011; Laufer et al., 2015; Su et al., 2014; Kim & Chung, 2016; Sajadmanesh et al., 2017), which underlined the differences in cross-cultural food preferences, this work demonstrates commonalities instead. It suggests the food preferences are not totally culturally dependent, these are, at least on the aesthetic aspects, culturally agnostic to some extent. This finding supports the discovery that indicates a general tendency of human food preferences. For example, it was concluded that human across cultures have only few differences in perceiving the flavours (Prescott & Bell, 1995). It was also found that, food that looks "natural" (Anderson, 2014), or with specific colour patterns, such as more red brightness (Foroni et al., 2016), or chromatic colours (Lee et al., 2013) is more visually appealing to humans irrespective of cultural background. Common aesthetic food preferences result from human common psychophysical responses to the sensory inputs (Prescott & Bell, 1995). Reasons for differences in cross-cultural food preferences were inferred to be linked to level of familiarity (Jaeger et al., 1998), i.e., when humans recognizing the food from their own cultures, they are more likely to prefer it. When users in the context of cultural agnostic - such as in the user study of Study II in this work, the participants were invited to rank the unlabelled recipe images with the premise that they have limited ability to recognize the origin of the images (as revealed by user study in Study I) - stable patterns of food preference are identified. Owing to the communicative value of food, I assume these findings would contribute to improve the understanding of human food choice behaviours in cross-cultural contexts and make the cross-cultural food choices more predictable.

In summary, this doctoral work investigates the role culture plays in influencing food aesthetics and human aesthetic-related food preferences. Human food choices are complex, with multiple factors intertwined and collaborating in the process. The previous research in the food choices, as pointed by Vabø (2014), does not pay enough attention to cultural aspects. This work is conducted to fill this gap. In addition, as an example of interdisciplinary research, by means of computer science techniques and computational analyses, this work supports the findings from the traditional anthropological studies and generates new knowledge for understanding human food choices. I believe this work would be helpful for researchers in related domains, such as the food policymaker who need an insight into how people make food decisions in order to promote healthy and sustainable eating. Integrating the knowledge that learned from this work into application is under way. What would be accomplished and valuable for the related fields is incorporating the findings into the development of food recommender systems. This will be discussed in the next section.

8.3 Practical Implications of this Work

This work provides insight into the development of food recommender systems in two ways. Firstly, it sheds light on incorporating aesthetic information into content-based food recommender systems. In comparison to the content-based food recommender systems, which relied on the ingredients as reviewed in Chapter 3, I show which aesthetic information can be extracted from online recipe data and how it performs in food preference prediction tasks. On the visual aspects, although this work is not the first to incorporate the visual features into food preference prediction task, I attempt to derive as much information as possible to capture the visual characteristics of food for the same task. Apart from the low-level visual features (i.e., the EVF in this work), which have been proven to be useful in predicting human preference in the food domain (Elsweiler et al., 2017; Trattner et al., 2018), I also included the pre-trained DNN embeddings (VGG16 in this work) in this work. These have been extracted and used for predicting human preferences to items in other domains (e.g., artworks) in recent work (Messina et al., 2019; Messina, Cartagena, Cerda-Mardini, del Rio, & Parra, 2020). In addition, the hand-crafted visual features (Colour Histogram, LBP and BoVW), which have been applied in other tasks, such as food image classification and retrieval (e.g., Joutou & Yanai, 2009; Zong et al., 2010; Matsuda et al., 2012), were incorporated into food preference prediction task in this work as well. All sets of visual feature sets performed reasonably in the food preference tasks, and a combination of them can achieve comparable predictive power to that of ingredients. On the flavour aspects, the flavour information - although it was pointed out to be one of the most important factors in influencing human food choices (Ahn et al., 2011; Kourouniotis et al., 2016; Liem & Russell, 2019) - has rarely been used to predict human food preferences with machine learning approaches. The commonly applied flavour information is the flavour compounds, the main application of these is to explain the flavour patterns across regions by means of food pairing theory (Ahn et al., 2011; Jain et al., 2015). In this work, I applied the flavour compounds in a new way, i.e., representing online recipes on the flavour aspects. Models trained with the flavour compounds also preformed reasonably in predicting human food preferences. In addition, in the exploratory analyses in Chapter 6, I tried to generate understandable description systems of flavour with the flavour compounds. Although the systems were limited to describing the flavours as savoury/non-sweet or sweet, I demonstrate a new method to integrate flavour information into food recommender systems, which was claimed to improve their performance (Nag et al., 2019).

Overall, both visual features and flavour compounds performed well in food preference prediction tasks. Moreover, the empirical experiments in Chapter 7, which combined the ingredient, visual and flavour information for predicting user cross-cultural food preferences, have again, demonstrate the promise of incorporating the aesthetic information for the future development of food recommender systems. It is illustrated that the combination of visual features and flavour compounds can not only provide stronger predictive power than ingredients, but also enhance the performance of the models trained using the ingredients.

The second practical implication from this doctoral work is that it provides a new perspective of incorporating the context feature, culture, in particular, into the development of food recommender systems. The importance of taking context into account when providing food recommendation has been recognised in previous work. It was also stated that what is lacking for building context-aware food recommender system is recognising the most important context variables and accounting for these algorithmically in an appropriate way (Trattner & Elsweiler, 2017a). Multiple context variables have been investigated in the related work (reviewed in Chapter 3), such as gender (Rokicki et al., 2016), time (Kusmierczyk et al., 2015b), location (Cheng et al., 2017), etc., to improve the performance of food recommender systems. However, culture, which is known as one of the important context features, has not been specifically investigated in the food recommendation scenario. In this work, I focused on food cultures. Moreover, in comparison to previous work, which often employed simple techniques to filter users and items according to the relevant context factors, and only recommend food that are in the same circumstance to the users, such as recommending only cosmopolitan (or rural) recipes to people lives in metropolis (or rural area) (Cheng et al., 2017), or recommending food only appealing to female (or male) to female (or male) users (Rokicki et al., 2016), what I attempted in this work is providing the food recommendation for the users in a cross-cultural context. This can be realised based on the foundation that people with different cultural backgrounds have similar food preferences, which have been revealed by using aesthetic information in Chapter 5 and 6. The existence of stable patterns in aesthetic food preferences across China, US and German indicates that, for the users come from these three cultures, it is possible to provide aesthetically acceptable food recommendation that originated from any of these cultures. An initial attempt to make cross-cultural food recommendation is shown in Chapter 7, the performance of the classifiers (at best AUC = 0.69) suggests that this experiment can be seen as a promising example of algorithmically incorporating cultures into the development of context-aware food recommender systems.

The food recommender systems developed based on the findings from this work would be applicable in several use cases. For example, guiding travellers in their food choices. It is reported that in the short-term mobility, travellers might have difficulty in making food decision due to their willing of trading off boundary crossing (e.g., experience the foreign culture) and boundary maintenance (e.g., keep a connection to home) (Bardhi, Ostberg, & Bengtsson, 2010). The food recommender systems would assist users to choose the food, which not only originated from the culture of destination, but also corresponds to the users' cultural-shaped aesthetic food preferences. It would help the local restaurants, which intend to appeal more people from other cultures to taste their food as well. Similarly, it is also possible for the food recommender systems to help people in migration to fit in local culture by means of recommending acceptable local food.

8.4 Limitations of this Work

The previous section in this chapter summarises the main findings of the empirical experiments and discusses their theoretical and practical implications. The main contributions of this work, including improving the understanding with respect to human food choices (especially in cross-cultural context) and providing a new perspective for the development of food recommender systems, would be promising for the community. This work, however, is subject to a major limitation.

The major limitation of this work is lacking datasets, where people from different cultures not only provide ratings to the recipes from their own cultures, but also to the recipes from the others. Although I have generated such a dataset, which contains the ratings of 270 recipe images from 450 participants across cultures, gathered via the user study in

Chapter 5, compared to the original data derived from the recipe portals, it is hardly to be seen as a large-scale one. Larger datasets are in demand and would be helpful for building more reliable classifiers in order to improve the performance of future cross-cultural food recommender systems. Moreover, stable patterns in aesthetic food preferences may not exist only across Chinese, US and German cultures, datasets incorporating other cultures are also required. Models generating from these datasets are promising to be applicable to more cross-cultural contexts. Further work, especially gathering the ratings from users with different cultural backgrounds to the recipes from different cultures, is not feasible during the given time. This will be considered as the starting point for future work.

8.5 Chapter Summary

This chapter has summarised and discussed the findings of the main empirical chapters of this thesis. The implications of the findings and the limitations of the work have also been described. The following chapter concludes the thesis.

Chapter 9

Conclusion

In this doctoral work I have investigated cross-cultural food aesthetics, particularly relating to visual appearance and flavour, and the food preferences associated with them. In this chapter, I relate the main findings to the primary aim of this thesis, which is to incorporate aesthetic aspect of cultural-related food preferences into the development of food recommender systems.

In this work, the online recipes from the recipe portals of China, US and Germany were collected to represent the corresponding food cultures. Data carrying the information of food visual appearance and flavour, specifically, the visual features encoded in the recipe images and the flavour compounds corresponding to the ingredients were extracted and processed to represent the visual and flavour components of recipes. By means of training classifiers with these feature sets, I found, firstly, food across cultures is different significantly with respect to how it looks and how it tastes. It is illustrated in the high accuracy of differentiating the online recipes from the three portals with the visual features and the flavour compounds algorithmically. Moreover, due to the impact of culture in influencing human perception on food visual appearance revealed by the user study, it is necessary to take both cultural backgrounds of food and users into consideration in developing food recommender systems. In order to achieve this, rather than further emphasising the idiosyncraticities of food and human food choices across cultures, as most of previous work has done, the remaining of this doctoral work focused on determining commonalities in the aesthetic aspects of food preferences and applying these to cross-cultural food recommendations. With this aim in mind, I first tested and verified the predictive power of the visual features and flavour compounds in predicting food preferences within each culture. After that, by applying transfer learning approaches, I ascertained stable patterns of food preference across cultures on both visual and flavour aspects, which were then supported and explained by user study and post-hoc exploratory analyses. These findings support the presence of unifying food aesthetic ideals. Subsequently, as a preliminary step of applying stable patterns in aesthetic food preferences to the development of cross-cultural food recommender systems, I formulated a cross-cultural food preferences prediction task on the data basis, which consists of appreciated and less appreciated recipe samples from all three recipe portals. The results show that both the visual features and flavour compounds performed reasonably in this task, and show comparable predictive power to that of ingredients, which have been commonly applied in food prediction tasks. In addition, in this task, the classifiers trained using visual features and flavour compounds were combined by means of ensemble learning. The combined classifiers outperformed the individual classifiers, including the baseline model trained using the ingredients. This is a promising finding with practical implications in the cross-cultural food recommendation domain. It suggests a new perspective of developing the context-aware food recommender systems, which would provide aesthetically acceptable food recommendation for people move across cultures, such as in travelling or emigrating.

Appendix A

The User Study of Study I in Chapter 5 (English Version)

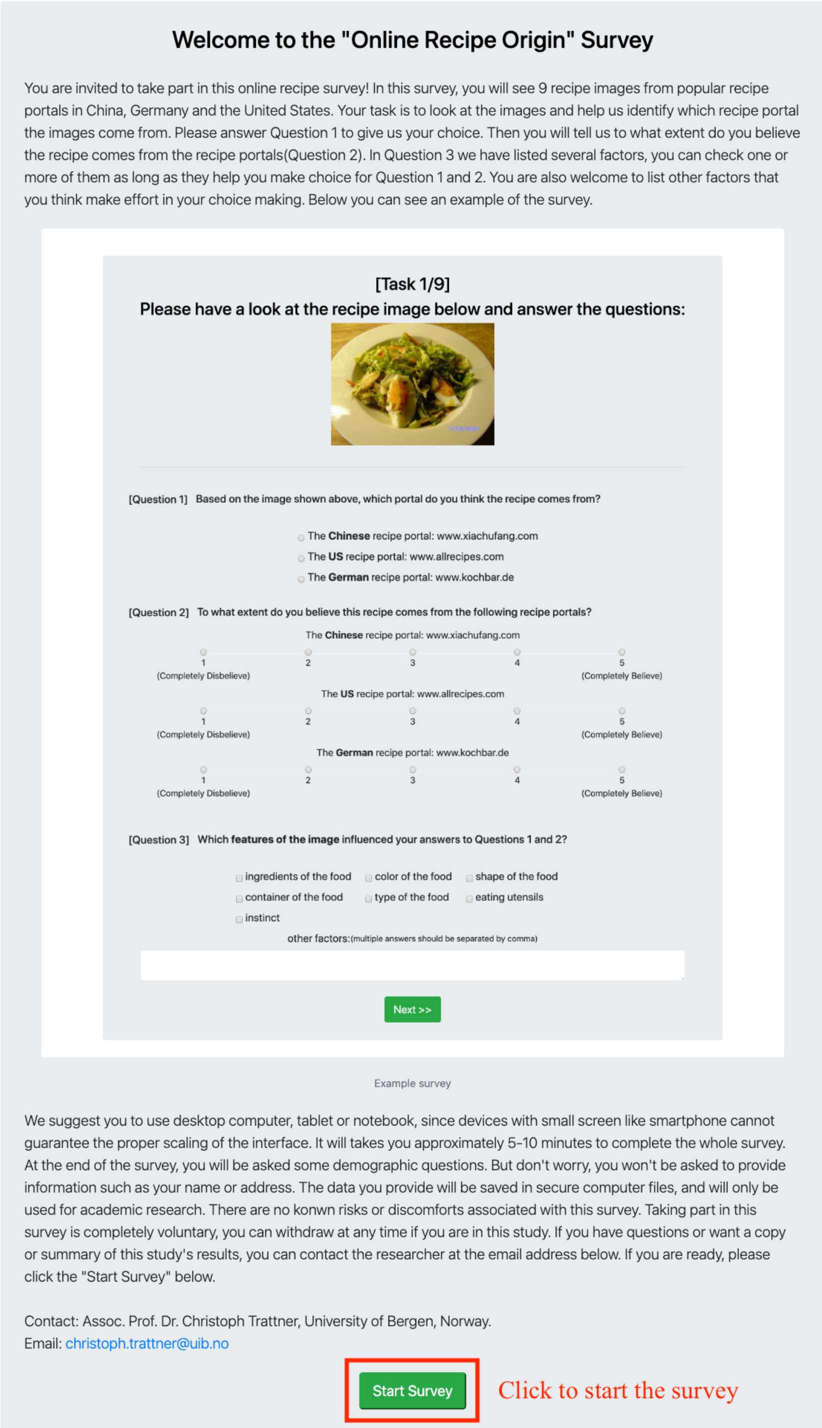


Figure A.1: The welcome page of the user study in Study I (English Version)

Figure A.2: The task page of the user study in Study I (English Version)

Please complete the questionnaire below to finish the survey.

This page will be displayed after the participants have completed all 9 tasks of this survey.

What is your age?

- Younger than 18
- 18-24
- 25-34
- 35-44
- 45-54
- 55 or older
- I'd rather not say

What is your gender?

- Male
- Female
- Other

What is your nationality?(Please select)

-- select one --

On a scale from 1-5, to what extent are you familiar with these online recipe portals?

the **Chinese** recipe portal: www.xiachufang.com

1 (Not at all) 2 3 4 5 (Very familiar)

the **US** recipe portal: www.allrecipes.com

1 (Not at all) 2 3 4 5 (Very familiar)

the **German** recipe portal: www.kochbar.de

1 (Not at all) 2 3 4 5 (Very familiar)

Which of the following statements best describes your use of online recipe portals (including, but not limited to allrecipes.com, kochbar.de, xiachufang.com, etc.) ?

- I hardly use them at all
- I use them at least once every three months
- I use them at least once per month
- I use them at least once per week
- I use them almost on a daily basis

Which of the following describes your experience with China:

- Never visited
- I have been there once or a few times
- I visit or have visited regularly
- I have lived there for many months or longer
- I am a permanent resident

Which of the following describes your experience with Germany:

- Never visited
- I have been there once or a few times
- I visit or have visited regularly
- I have lived there for many months or longer
- I am a permanent resident

Which of the following describes your experience with the United States:

- Never visited
- I have been there once or a few times
- I visit or have visited regularly
- I have lived there for many months or longer
- I am a permanent resident

How often do you travel cross-continental (e.g., from Asia to Europe, from Europe to America, etc.)?

- Never
- Less than once per year
- 1-2 times per year
- More than 2 times per year

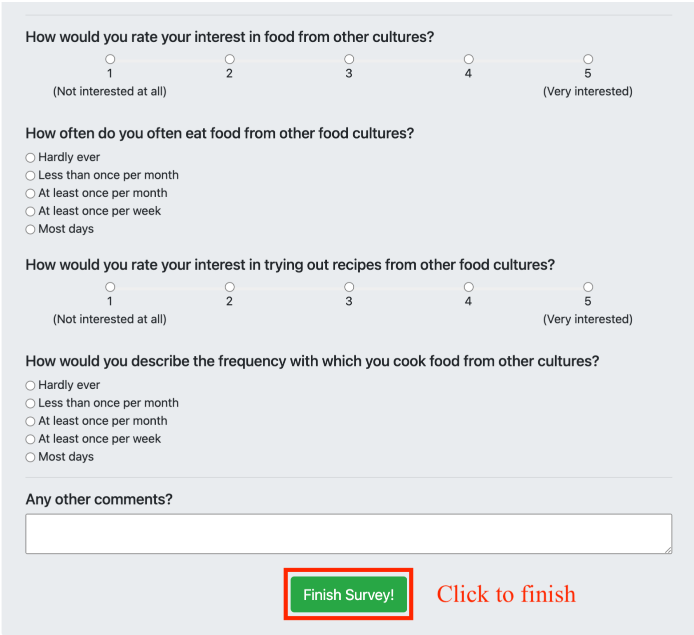


Figure A.3: The demographic questionnaire of the user study in Study I (English Version)

Appendix B

The User Study of Study I in Chapter 5 (Chinese Version)

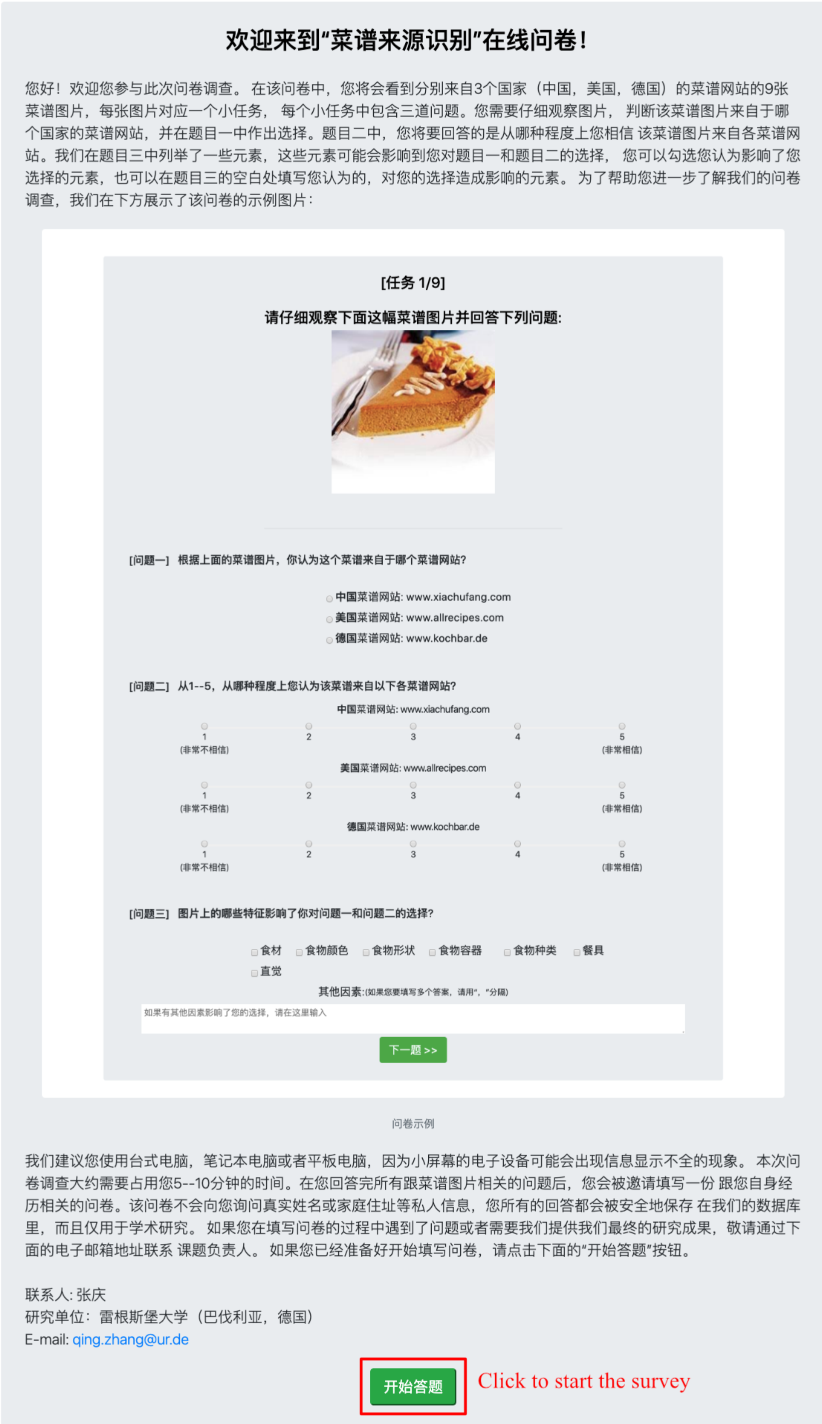


Figure B.1: The welcome page of the user study in Study I (Chinese Version)

Figure B.2: The task page of the user study in Study I (Chinese Version)

请完成下列问题

This page will be displayed after the participants have completed all 9 tasks of this survey

您的年龄是?

- ○ 18岁或小于18岁
- ○ 18岁——24岁
- ○ 25岁--34岁
- ○ 35岁--44岁
- ○ 45岁--54岁
- ○ 55岁或大于55岁
- ○ I'd rather not say

您的性别是r?

- ○ 男
- ○ 女
- ○ 其他

您的国籍是?(请选择)

布基纳法索

您对这些在线菜谱网站的熟悉程度如何?

中国 菜谱网站: www.xiachufang.com

○ 1 (从未听说)	○ 2	○ 3	○ 4	○ 5 (非常熟悉)

美国 菜谱网站: www.allrecipes.com

○ 1 (从未听说)	○ 2	○ 3	○ 4	○ 5 (非常熟悉)

德国 菜谱网站: www.kochbar.de

○ 1 (从未听说)	○ 2	○ 3	○ 4	○ 5 (非常熟悉)

下列哪个选项更符合您对菜谱网站的使用情况（包括但不限于下厨房，allrecipes.com和kochbar.de）？

- ○ 我机会从来不使用
- ○ 我至少三个月使用一次
- ○ 我至少一个月使用一次
- ○ 我至少一周使用一次
- ○ 我几乎每天使用

下列哪个选项能够准确描述您在中国的经历:

- ○ 从未去过
- ○ 我极少去那里
- ○ 我定期去那里
- ○ 我在那里住过很多个月或者更长
- ○ 我在那里永久定居

下列哪个选项能够准确描述您在德国的经历:

- ○ 从未去过
- ○ 我极少去那里
- ○ 我定期去那里
- ○ 我在那里住过很多个月或者更长
- ○ 我在那里永久定居

下列哪个选项能够准确描述您在美国的经历:

- ○ 从未去过
- ○ 我极少去那里
- ○ 我定期去那里
- ○ 我在那里住过很多个月或者更长
- ○ 我在那里永久定居

您跨洲旅行的频率是怎样的（比如从亚洲到欧洲，从欧洲到美洲）？

- ○ 从未有过这种经历
- ○ 少于一年一次
- ○ 一年1--2次
- ○ 一年2次以上


如果有1--5个等级，您认为您对来自其他文化的食物的兴趣处在哪个级别?

1 (毫无兴趣) 2 3 4 5 (非常感兴趣)

您尝试外来文化的食物的频率是怎样的?

- 几乎从未尝试过
- 少于一个月一次
- 至少一个月一次
- 至少一周一次
- 大部分时间都在尝试

如果有1--5个等级，您认为您对来自其他文化的菜谱的兴趣处在哪个级别呢?

1 (毫无兴趣) 2 3 4 5 (非常感兴趣)

您烹饪外来文化的菜谱的频率是怎样的?

- 几乎从未尝试过
- 少于一个月一次
- 至少一个月一次
- 至少一周一次
- 大部分时间都在尝试

请留下您宝贵的意见

感谢参与!

Click to finish the survey

Figure B.3: The demographic questionnaire of the user study in Study I (Chinese Version)

Appendix C

The User Study of Study II in Chapter 5 (English Version)

Figure C.1: The welcome page of the user study in Study II (English Version)

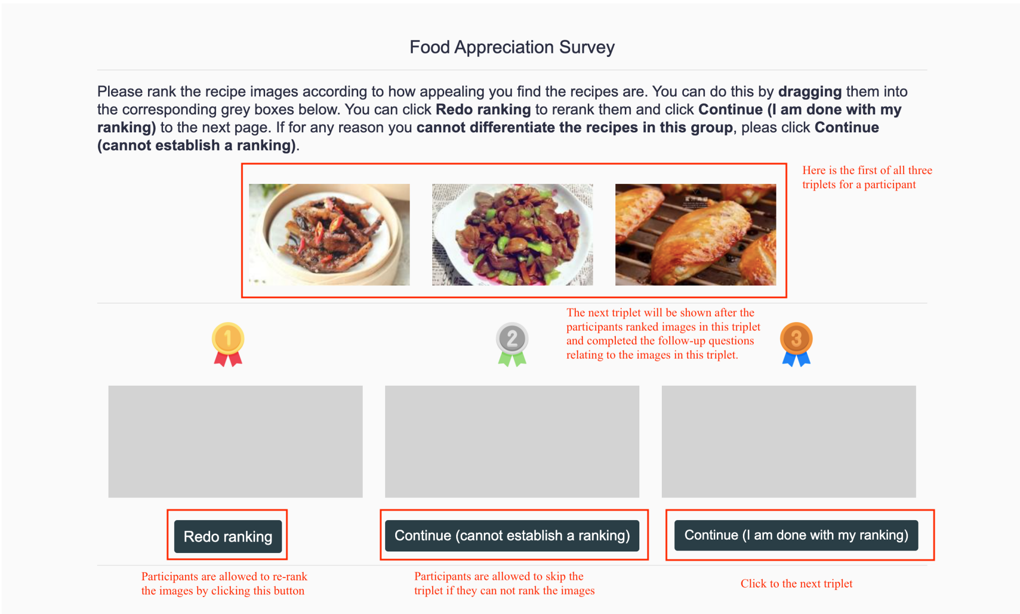


Figure C.2: The ranking page of the user study in Study II (English Version)

You are looking at the [1/3] images in the first group

This is the first of all three images in the triplet

You ranked this image at position **1**.
Justify this decision by answering the following questions:

The ranking of this image by the participants is marked here.

The questions for participants to justify the ranking are listed below:

Q1. To what extent do you like the **Color** of the dish?

Extremely dislike — Neutral — Extremely like

Q2. To what extent do you like the **Presentation** of the dish?

Extremely dislike — Neutral — Extremely like

Q3. To what extent do you like the **Texture** (described as soft, hard, dry, greasy, corse, smooth, etc.) of the dish?

Extremely dislike — Neutral — Extremely like

Q4. How do you evaluate the **Quality** (including resolution, lighting, distortion, etc.) of the recipe image?

Very poor — Ok — Very good

Q5. To what extent is the recipe image **visually appealing** to you?

Extremely unappealing — Neutral — Extremely appealing

Q6. Based on the image, how **Salty** do you think the dish would be?

Not salty at all — Slightly salty — Extremely salty

Q7. Based on the image, how **Sweet** do you think the dish would be?

Not sweet at all — Slightly sweet — Extremely sweet

Q8. Based on the image, how **Sour** do you think the dish would be?

Not sour at all — Slightly sour — Extremely sour

Q9. Based on the image, how **Spicy** do you think the dish would be?

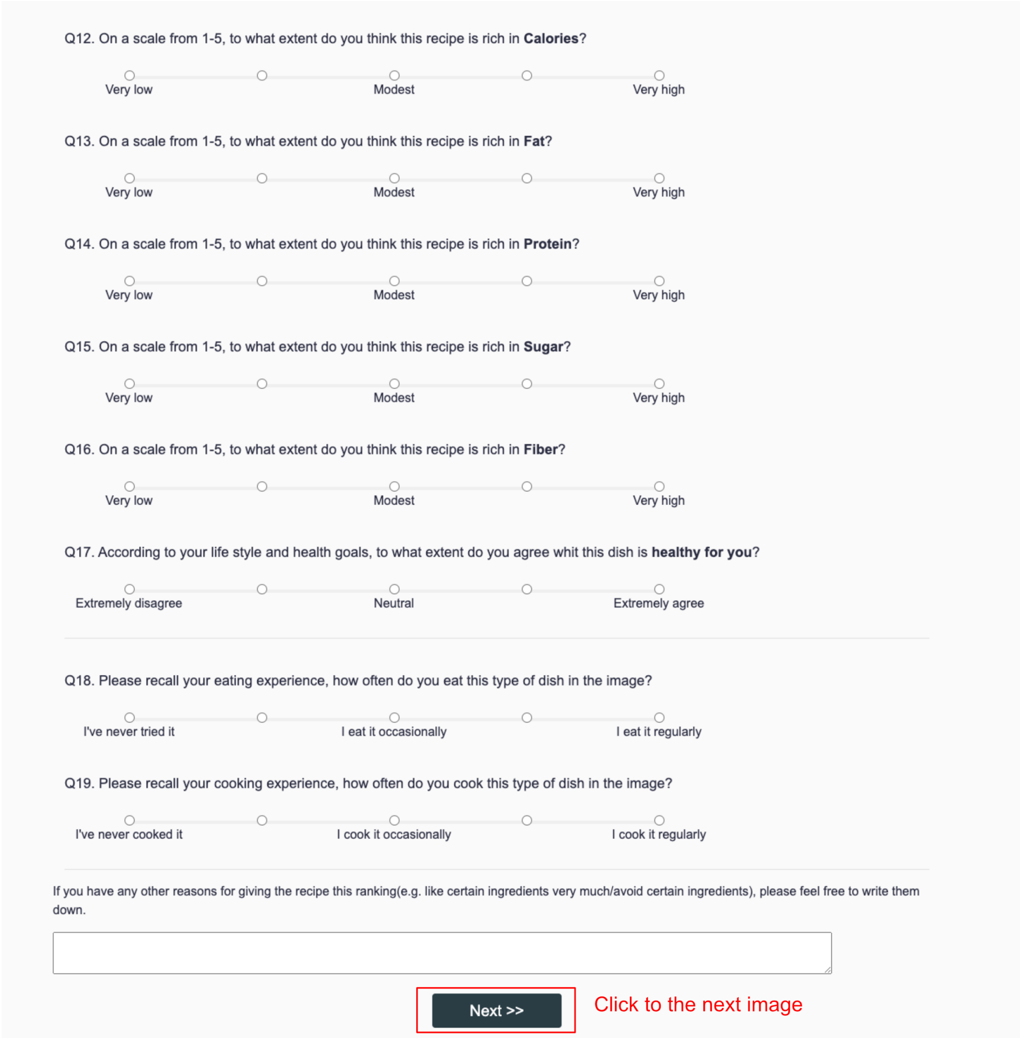
Not spicy at all — Slightly spicy — Extremely spicy

Q10. Based on the image, how **Bitter** do you think the dish would be?

Not bitter at all — Slightly bitter — Extremely bitter

Q11. Based on the image, to what extent would you **like to eat dish with this taste**?

Extremely dislike — Neutral — Extremely like


Q12. On a scale from 1-5, to what extent do you think this recipe is rich in **Calories**?

Very low — Modest — Very high

Q13. On a scale from 1-5, to what extent do you think this recipe is rich in **Fat**?

Very low — Modest — Very high

Q14. On a scale from 1-5, to what extent do you think this recipe is rich in **Protein**?

Very low — Modest — Very high

Q15. On a scale from 1-5, to what extent do you think this recipe is rich in **Sugar**?

Very low — Modest — Very high

Q16. On a scale from 1-5, to what extent do you think this recipe is rich in **Fiber**?

Very low — Modest — Very high

Q17. According to your life style and health goals, to what extent do you agree whit this dish is **healthy for you**?

Extremely disagree — Neutral — Extremely agree

Q18. Please recall your eating experience, how often do you eat this type of dish in the image?

I've never tried it — I eat it occasionally — I eat it regularly

Q19. Please recall your cooking experience, how often do you cook this type of dish in the image?

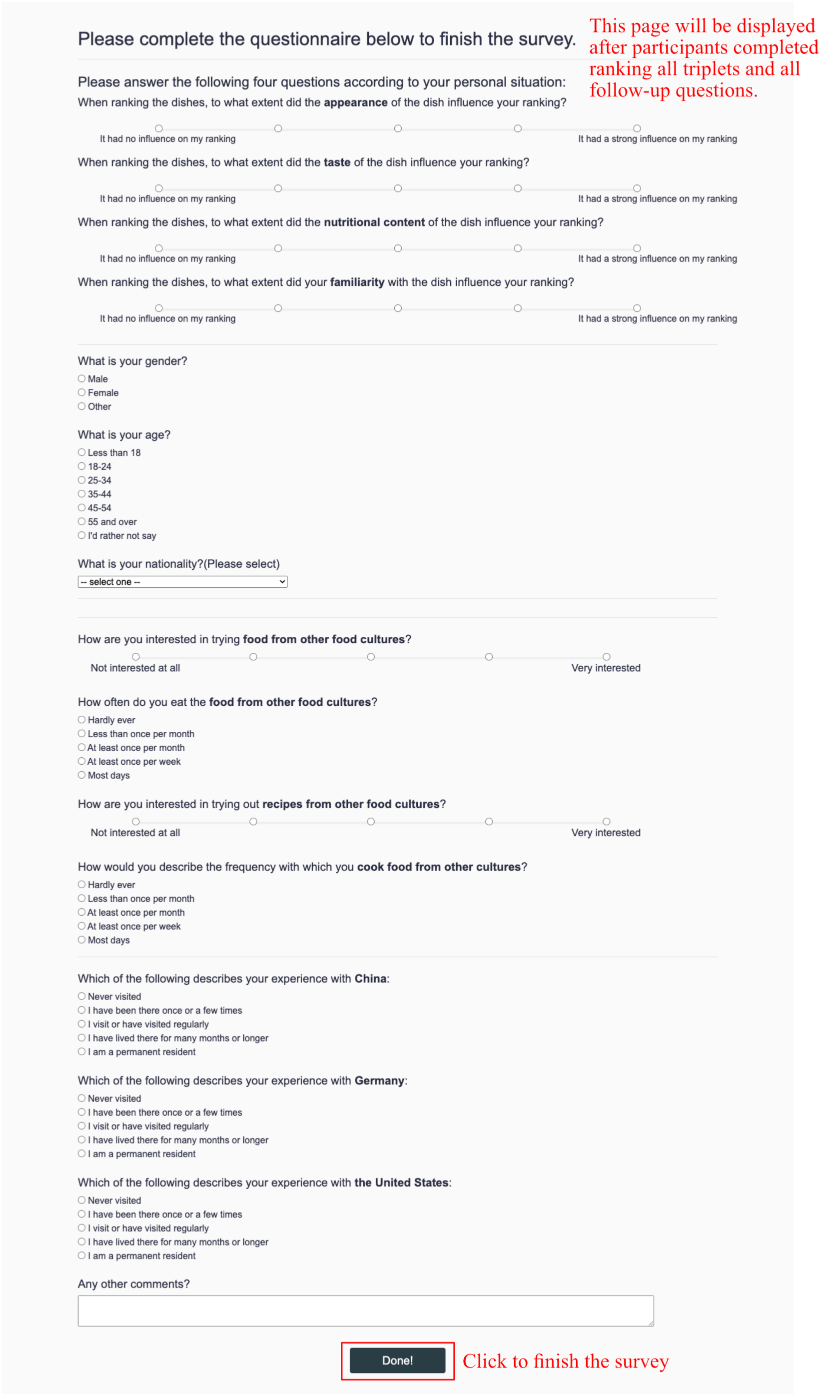
I've never cooked it — I cook it occasionally — I cook it regularly

If you have any other reasons for giving the recipe this ranking(e.g. like certain ingredients very much/avoid certain ingredients), please feel free to write them down.

Next >>

Click to the next image

Figure C.3: The question page of the user study in Study II (English Version)



Please complete the questionnaire below to finish the survey.

This page will be displayed after participants completed ranking all triplets and all follow-up questions.

Please answer the following four questions according to your personal situation:

When ranking the dishes, to what extent did the **appearance** of the dish influence your ranking?

It had no influence on my ranking ○ ○ ○ ○ ○ It had a strong influence on my ranking

When ranking the dishes, to what extent did the **taste** of the dish influence your ranking?

It had no influence on my ranking ○ ○ ○ ○ ○ It had a strong influence on my ranking

When ranking the dishes, to what extent did the **nutritional content** of the dish influence your ranking?

It had no influence on my ranking ○ ○ ○ ○ ○ It had a strong influence on my ranking

When ranking the dishes, to what extent did your **familiarity** with the dish influence your ranking?

It had no influence on my ranking ○ ○ ○ ○ ○ It had a strong influence on my ranking

What is your gender?

- ○ Male
- ○ Female
- ○ Other

What is your age?

- ○ Less than 18
- ○ 18-24
- ○ 25-34
- ○ 35-44
- ○ 45-54
- ○ 55 and over
- ○ I'd rather not say

What is your nationality?(Please select)

-- select one --

How are you interested in trying **food from other food cultures**?

Not interested at all ○ ○ ○ ○ ○ Very interested

How often do you eat the **food from other food cultures**?

- ○ Hardly ever
- ○ Less than once per month
- ○ At least once per month
- ○ At least once per week
- ○ Most days

How are you interested in trying out **recipes from other food cultures**?

Not interested at all ○ ○ ○ ○ ○ Very interested

How would you describe the frequency with which you **cook food from other cultures**?

- ○ Hardly ever
- ○ Less than once per month
- ○ At least once per month
- ○ At least once per week
- ○ Most days

Which of the following describes your experience with **China**:

- ○ Never visited
- ○ I have been there once or a few times
- ○ I visit or have visited regularly
- ○ I have lived there for many months or longer
- ○ I am a permanent resident

Which of the following describes your experience with **Germany**:

- ○ Never visited
- ○ I have been there once or a few times
- ○ I visit or have visited regularly
- ○ I have lived there for many months or longer
- ○ I am a permanent resident

Which of the following describes your experience with **the United States**:

- ○ Never visited
- ○ I have been there once or a few times
- ○ I visit or have visited regularly
- ○ I have lived there for many months or longer
- ○ I am a permanent resident

Any other comments?

Done! Click to finish the survey

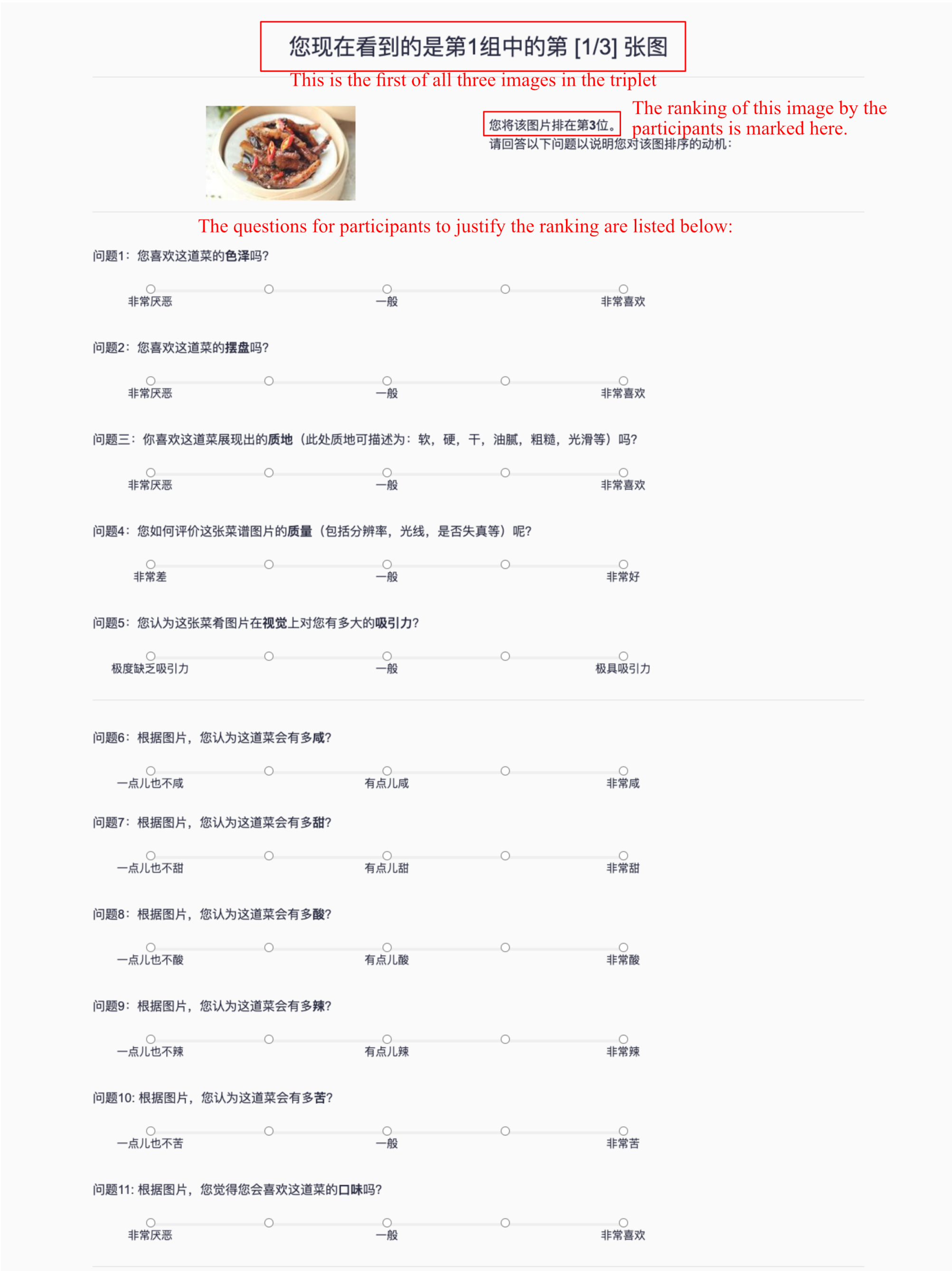
Figure C.4: The demographic questionnaire of the user study in Study II (English Version)

Appendix D

The User Study of Study II in Chapter 5 (Chinese Version)

Figure D.1: The welcome page of the user study in Study II (Chinese Version)

Figure D.2: The ranking page of the user study in Study II (Chinese Version)



您现在看到的是第1组中的第 [1/3] 张图

This is the first of all three images in the triplet

您将该图片排在第**3**位。

The ranking of this image by the participants is marked here.

请回答以下问题以说明您对该图排序的动机：

The questions for participants to justify the ranking are listed below:

问题1：您喜欢这道菜的**色泽**吗?

非常厌恶 — 一般 — 非常喜欢

问题2：您喜欢这道菜的**摆盘**吗?

非常厌恶 — 一般 — 非常喜欢

问题三：你喜欢这道菜展现出的**质地**（此处质地可描述为：软，硬，干，油腻，粗糙，光滑等）吗?

非常厌恶 — 一般 — 非常喜欢

问题4：您如何评价这张菜谱图片的**质量**（包括分辨率，光线，是否失真等）呢?

非常差 — 一般 — 非常好

问题5：您认为这张菜肴图片在**视觉**上对您有多大的**吸引力**?

极度缺乏吸引力 — 一般 — 极具吸引力

问题6：根据图片，您认为这道菜会有多**咸**?

一点儿也不咸 — 有点儿咸 — 非常咸

问题7：根据图片，您认为这道菜会有多**甜**?

一点儿也不甜 — 有点儿甜 — 非常甜

问题8：根据图片，您认为这道菜会有多**酸**?

一点儿也不酸 — 有点儿酸 — 非常酸

问题9：根据图片，您认为这道菜会有多**辣**?

一点儿也不辣 — 有点儿辣 — 非常辣

问题10: 根据图片，您认为这道菜会有多**苦**?

一点儿也不苦 — 一般 — 非常苦

问题11: 根据图片，您觉得您会喜欢这道菜的**口味**吗?

非常厌恶 — 一般 — 非常喜欢

问题12: 您认为这道菜含多少**卡路里**?

非常低 适中 非常高

问题13：您认为这道菜包含多少**脂肪**?

非常低 适中 非常高

问题14：您认为这道菜含多少**蛋白质**?

非常低 适中 非常高

问题15：您认为这道菜含多少**糖份**（此处糖份来源主要为谷物，时蔬水果，奶制品以及工业添加剂）？

非常低 适中 非常高

问题16：您认为这道菜含多少**膳食纤维**?

非常低 适中 非常高

问题17：根据您自身的生活习惯和健康需求，您认为这道菜对您来说是否**健康**?

非常不健康 一般 非常健康

问题18：请根据您自身经历，回忆您是否吃过这种菜肴或者多久吃一次它。

我从未吃过这道菜 我偶尔吃一次这道菜 我经常吃这道菜

问题19：请根据您自身经历，回忆您是否烹饪过这种菜肴或者多久烹饪一次它。

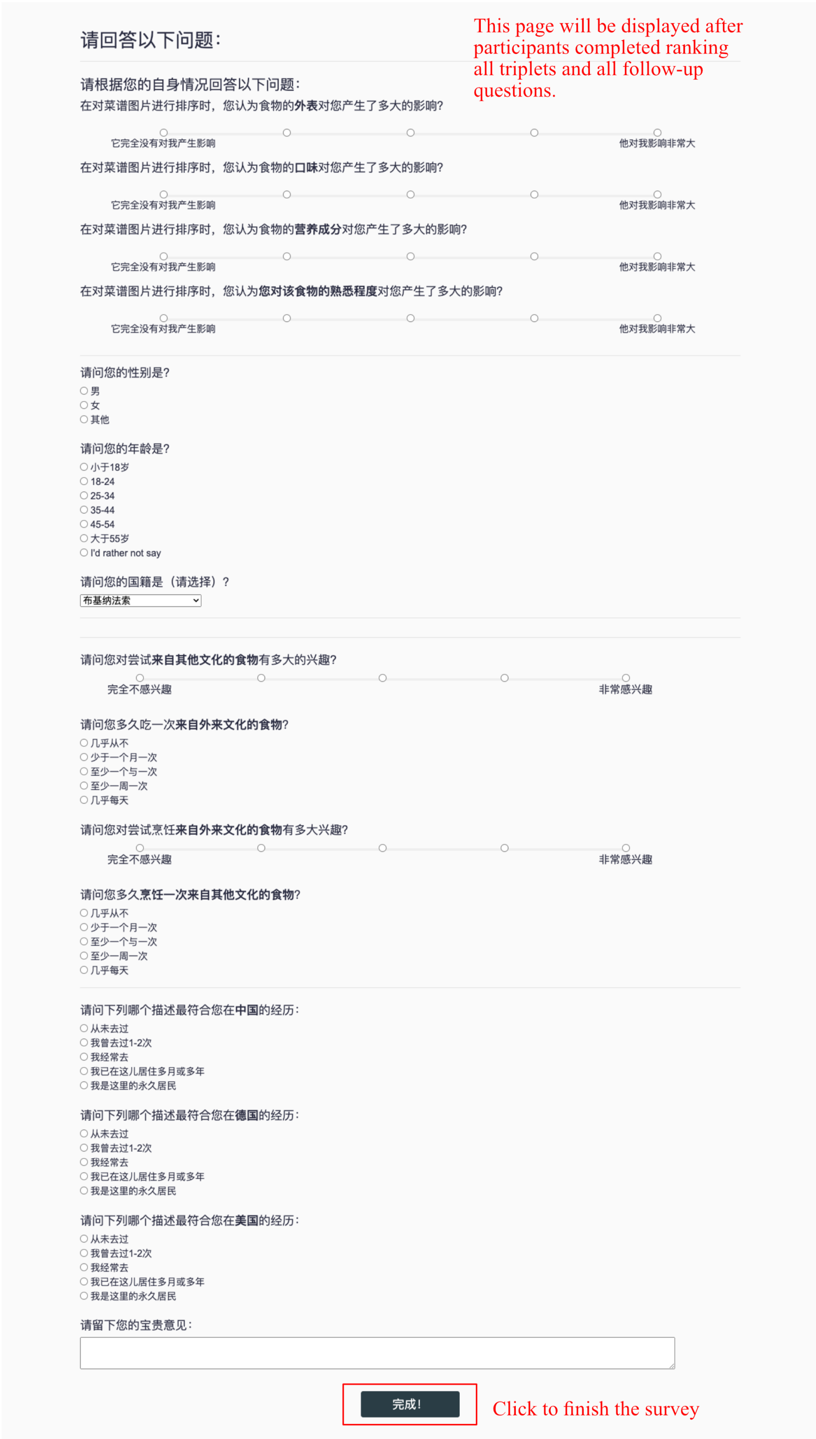
我从未烹饪过这道菜 我偶尔烹饪这道菜 我经常烹饪这道菜

如果您有其他给这道菜排名的理由（例如：极度喜爱/厌恶这道菜肴中的某个食材），请填写在下方的框中。

继续 >>

Click to the next image

Figure D.3: The question page of the user study in Study II (Chinese Version)



This page will be displayed after participants completed ranking all triplets and all follow-up questions.

请回答以下问题:

请根据您的自身情况回答以下问题:

在对菜谱图片进行排序时，您认为食物的**外表**对您产生了多大的影响?

它完全没有对我产生影响 … 他对我影响非常大

在对菜谱图片进行排序时，您认为食物的**口味**对您产生了多大的影响?

它完全没有对我产生影响 … 他对我影响非常大

在对菜谱图片进行排序时，您认为食物的**营养成分**对您产生了多大的影响?

它完全没有对我产生影响 … 他对我影响非常大

在对菜谱图片进行排序时，您认为**您对该食物的熟悉程度**对您产生了多大的影响?

它完全没有对我产生影响 … 他对我影响非常大

请问您的性别是?

- 男
- 女
- 其他

请问您的年龄是?

- 小于18岁
- 18-24
- 25-34
- 35-44
- 45-54
- 大于55岁
- I'd rather not say

请问您的国籍是（请选择）？

布基纳法索

请问您对尝试**来自其他文化的食物**有多大的兴趣?

完全不感兴趣 … 非常感兴趣

请问您多久吃一次**来自外来文化的食物**?

- 几乎从不
- 少于一个月一次
- 至少一个与一次
- 至少一周一次
- 几乎每天

请问您对尝试烹饪**来自外来文化的食物**有多大兴趣?

完全不感兴趣 … 非常感兴趣

请问您多久**烹饪一次来自其他文化的食物**?

- 几乎从不
- 少于一个月一次
- 至少一个与一次
- 至少一周一次
- 几乎每天

请问下列哪个描述最符合您在**中国**的经历:

- 从未去过
- 我曾去过1-2次
- 我经常去
- 我已在这儿居住多月或多年
- 我是这里的永久居民

请问下列哪个描述最符合您在**德国**的经历:

- 从未去过
- 我曾去过1-2次
- 我经常去
- 我已在这儿居住多月或多年
- 我是这里的永久居民

请问下列哪个描述最符合您在**美国**的经历:

- 从未去过
- 我曾去过1-2次
- 我经常去
- 我已在这儿居住多月或多年
- 我是这里的永久居民

请留下您的宝贵意见:

完成!

Click to finish the survey

Figure D.4: The demographic questionnaire of the user study in Study II (Chinese Version)

References

Abbar, S., Mejova, Y., & Weber, I. (2015). You tweet what you eat: Studying food consumption through twitter. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (pp. 3197–3206).

Aburto, N. J., Ziolkovska, A., Hooper, L., Elliott, P., Cappuccio, F. P., & Meerpohl, J. J. (2013). Effect of lower sodium intake on health: systematic review and meta-analyses. *Bmj*, *346*, f1326.

Ahn, Y.-Y., Ahnert, S. E., Bagrow, J. P., & Barabási, A.-L. (2011). Flavor network and the principles of food pairing. *Scientific reports*, *1*(1), 1–7.

Aldridge, V., Dovey, T. M., & Halford, J. C. (2009). The role of familiarity in dietary development. *Developmental Review*, *29*(1), 32–44.

Alshazly, H., Linse, C., Barth, E., & Martinetz, T. (2019). Handcrafted versus cnn features for ear recognition. *Symmetry*, *11*(12), 1493.

Amerine, M. A., Pangborn, R. M., & Roessler, E. B. (2013). *Principles of sensory evaluation of food*. Elsevier.

Anderson, E. N. (1988). *The food of China*. Yale University Press.

Anderson, E. N. (2014). *Everyone eats: Understanding food and culture*. New York University Press.

Anthimopoulos, M. M., Gianola, L., Scarnato, L., Diem, P., & Mougiakakou, S. G. (2014). A food recognition system for diabetic patients based on an optimized bag-of-features model. *IEEE journal of biomedical and health informatics*, *18*(4), 1261–1271.

Appleton, K. M., & Smith, E. (2016). A role for identification in the gradual decline in the pleasantness of flavors with age. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *71*(6), 987–994.

Aroyo, L., Wang, Y., Brussee, R., Gorgels, P., Rutledge, L., & Stash, N. (2007). Personalized museum experience: The rijksmuseum use case. In *Museums and the Web 2007 (San Francisco CA, USA, April 11-14, 2007. Proceedings)*. Archives & Museum Informatics.

Auvray, M., & Spence, C. (2008). The multisensory perception of flavor. *Consciousness and cognition*, *17*(3), 1016–1031.

Bardhi, F., Ostberg, J., & Bengtsson, A. (2010). Negotiating cultural boundaries: Food, travel and consumer identities. *Consumption, Markets and Culture*, *13*(2), 133–157.

Bellisle, F. (1999). Food choice, appetite and physical activity. *Public health nutrition*, *2*(3a), 357–361.

Bender, A. (1981). The appearance and the nutritional value of food products. *Journal of human nutrition*, *35*(3), 215–217.

Blumenthal, H. (2008). *The big fat duck cookbook*. Bloomsbury.

Bossard, L., Guillaumin, M., & Gool, L. V. (2014). Food-101–mining discriminative components with random forests. In *European conference on computer vision* (pp. 446–461).

Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. *Proceedings of GSCL*, *30*, 31–40.

Buck, L. B. (2000). The molecular architecture of odor and pheromone sensing in mammals. *Cell*, *100*(6), 611–618.

Cantarero, L., Espeitx, E., Gil Lacruz, M., & Martin, P. (2013). Human food preferences and cultural identity: The case of aragón (spain). *International Journal of Psychology*, *48*(5), 881–890.

Cecchini, M. P., Knaapila, A., Hoffmann, E., Boschi, F., Hummel, T., & Iannilli, E. (2019). A cross-cultural survey of umami familiarity in european countries. *Food Quality and Preference*, *74*, 172–178.

Chamoun, E., Mutch, D. M., Allen-Vercoe, E., Buchholz, A. C., Duncan, A. M., Spriet, L. L., … Study, G. F. H. (2018). A review of the associations between single nucleotide polymorphisms in taste receptors, eating behaviors, and health. *Critical reviews in food science and nutrition*, *58*(2), 194–207.

Chen, J. (2007). Surface texture of foods: Perception and characterization. *Critical reviews in food science and nutrition*, *47*(6), 583–598.

Chen, J., & Ngo, C.-W. (2016). Deep-based ingredient recognition for cooking recipe retrieval. In *Proceedings of the 24th ACM international conference on multimedia* (pp. 32–41).

Chen, J.-j., Ngo, C.-W., & Chua, T.-S. (2017). Cross-modal recipe retrieval with rich food attributes. In *Proceedings of the 25th ACM international conference on multimedia* (pp. 1771–1779).

Chen, M., Dhingra, K., Wu, W., Yang, L., Sukthankar, R., & Yang, J. (2009). Pfid: Pittsburgh fast-food image dataset. In *2009 16th IEEE International Conference on Image Processing (ICIP)* (pp. 289–292).

Cheng, H., Rokicki, M., & Herder, E. (2017). The influence of city size on dietary choices. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization* (pp. 231–236).

Choi, D., Bell, W., Kim, D., & Kim, J. (2021). Uav-driven structural crack detection and location determination using convolutional neural networks. *Sensors*, *21*(8), 2650.

Choi, N.-E., & Han, J. H. (2015). *How flavor works: the science of taste and aroma*. John Wiley & Sons.

Chrzan, J., & Brett, J. (2017). *Food culture: Anthropology, linguistics and food studies* (Vol. 2). Berghahn Books.

Ciocca, G., Napoletano, P., & Schettini, R. (2017, May). Food Recognition: A New Dataset, Experiments, and Results. *IEEE J. Biomed. Health Inform.*, *21*(3), 588–598. Retrieved 2022-03-31, from `http://ieeexplore.ieee.org/document/7776769/` doi: 10.1109/JBHI.2016.2636441

Clark, J. E. (1998). Taste and flavour: their importance in food choice and acceptance. *Proceedings of the nutrition society*, *57*(4), 639–643.

Clydesdale, F. M. (1993). Color as a factor in food choice. *Critical reviews in food science and nutrition*, *33*(1), 83–101.

Cordeiro, F., Bales, E., Cherry, E., & Fogarty, J. (2015). Rethinking the mobile food journal: Exploring opportunities for lightweight photo-based capture. In *Proceedings* *of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3207–3216).

Csurka, G., Dance, C., Fan, L., Willamowski, J., & Bray, C. (2004). Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV* (Vol. 1, pp. 1–2).

De Choudhury, M., Sharma, S., & Kiciman, E. (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* (pp. 1157–1170).

DeCost, B. L., & Holm, E. A. (2015). A computer vision approach for automated analysis and classification of microstructural image data. *Computational materials science*, *110*, 126–133.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Ding, K., Ma, K., & Wang, S. (2019). Intrinsic image popularity assessment. In *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 1979–1987).

Druz, L. L., & Baldwin, R. E. (1982). Taste thresholds and hedonic responses of panels representing three nationalities. *Journal of Food Science*, *47*(2), 561–563.

Duszka, K., Gregor, A., Reichel, M. W., Baierl, A., Fahrngruber, C., & König, J. (2020). Visual stimulation with food pictures in the regulation of hunger hormones and nutrient deposition, a potential contributor to the obesity crisis. *PloS one*, *15*(4), e0232099.

Eertmans, A., Baeyens, F., & Van Den Bergh, O. (2001). Food likes and their relative importance in human eating behavior: review and preliminary suggestions for health promotion. *Health education research*, *16*(4), 443–456.

Ege, T., & Yanai, K. (2017). Image-based food calorie estimation using knowledge on food categories, ingredients and cooking directions. In *Proceedings of the on Thematic Workshops of ACM Multimedia 2017* (pp. 367–375).

Ehrlichman, H., & Bastone, L. (1992). Olfaction and emotion. In *Science of olfaction* (pp. 410–438). Springer.

El-Dosuky, M., Rashad, M. Z., Hamza, T., & El-Bassiouny, A. (2012). Food recommendation using ontology and heuristics. In *International conference on advanced machine learning technologies and applications* (pp. 423–429).

Elsweiler, D., Harvey, M., Ludwig, B., & Said, A. (2015). Bringing the "healthy" into food recommenders. *DMRS*, *1533*, 33–36.

Elsweiler, D., Trattner, C., & Harvey, M. (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* (pp. 575–584).

Farinella, G. M., Allegra, D., Moltisanti, M., Stanco, F., & Battiato, S. (2016). Retrieval and classification of food images. *Computers in biology and medicine*, *77*, 23–39.

Farinella, G. M., Allegra, D., & Stanco, F. (2014). A benchmark dataset to study the representation of food images. In *European Conference on Computer Vision* (pp. 584–599).

Farinella, G. M., Allegra, D., Stanco, F., & Battiato, S. (2015). On the exploitation of one class classification to distinguish food vs non-food images. In *International*

Conference on Image Analysis and Processing (pp. 375–383).

Farquhar, J. (2002). *Appetites: Food and sex in post-socialist China.* Duke University Press.

Fenaroli, G. (2004). *Fenaroli's handbook of flavor ingredients. burdock g, editor.* Boca Raton (Florida): CRC Press.

Fischler, C. (1988). Food, self and identity. *Social science information*, *27*(2), 275–292.

Fisher, C., & Scott, T. R. (1997). *Food flavours: biology and chemistry*. Royal Society of chemistry.

Fogel, I., & Sagi, D. (1989). Gabor filters as texture discriminator. *Biological cybernetics*, *61*(2), 103–113.

Forbes, P., & Zhu, M. (2011). Content-boosted matrix factorization for recommender systems: experiments with recipe recommendation. In *Proceedings of the fifth ACM conference on recommender systems* (pp. 261–264).

Foroni, F., Pergola, G., & Rumiati, R. I. (2016). Food color is in the eye of the beholder: the role of human trichromatic vision in food evaluation. *Scientific reports*, *6*(1), 1–6.

Freyne, J., & Berkovsky, S. (2010). Intelligent food planning: personalized recipe recommendation. In *Proceedings of the 15th international conference on intelligent user interfaces* (pp. 321–324).

Freyne, J., Berkovsky, S., & Smith, G. (2011). Recipe recommendation: accuracy and reasoning. In *International conference on user modeling, adaptation, and personalization* (pp. 99–110).

Furst, T., Connors, M., Bisogni, C. A., Sobal, J., & Falk, L. W. (1996). Food choice: a conceptual model of the process. *Appetite*, *26*(3), 247–266.

Garg, N., Sethupathy, A., Tuwani, R., Nk, R., Dokania, S., Iyer, A., … others (2018). Flavordb: a database of flavor molecules. *Nucleic acids research*, *46*(D1), D1210–D1216.

Ge, M., Ricci, F., & Massimo, D. (2015). Health-aware food recommender system. In *Proceedings of the 9th ACM Conference on Recommender Systems* (pp. 333–334).

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2018). Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv preprint arXiv:1811.12231*.

Grover, A., & Leskovec, J. (2016). node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 855–864).

Guinard, J.-X., & Mazzucchelli, R. (1996). The sensory perception of texture and mouthfeel. *Trends in Food Science & Technology*, *7*(7), 213–219.

Habhab, S., Sheldon, J. P., & Loeb, R. C. (2009). The relationship between stress, dietary restraint, and food preferences in women. *Appetite*, *52*(2), 437–444.

Harris, M., & Ross, E. B. (1987). *Food and evolution: Toward a theory of human food habits*. Temple University Press.

Harvey, M., & Elsweiler, D. (2015). Automated recommendation of healthy, personalised meal plans. In *Proceedings of the 9th ACM Conference on Recommender Systems* (pp. 327–328).

Harvey, M., Ludwig, B., & Elsweiler, D. (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).

Harvey, M., Ludwig, B., & Elsweiler, D. (2013). You are what you eat: Learning user tastes for rating prediction. In *International symposium on string processing and information retrieval* (pp. 153–164).

Hasler, D., & Suesstrunk, S. E. (2003). Measuring colorfulness in natural images. In *Human vision and electronic imaging VIII* (Vol. 5007, pp. 87–95).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

He, Y., Xu, C., Khanna, N., Boushey, C. J., & Delp, E. J. (2013). Food image analysis: Segmentation, identification and weight estimation. In *2013 IEEE international conference on multimedia and expo (ICME)* (pp. 1–6).

Heath, P., Houston-Price, C., & Kennedy, O. B. (2011). Increasing food familiarity without the tears. a role for visual exposure? *Appetite*, *57*(3), 832–838.

Hoashi, H., Joutou, T., & Yanai, K. (2010). Image recognition of 85 food categories by feature fusion. In *2010 IEEE International Symposium on Multimedia* (pp. 296–301).

Holmberg, C., Chaplin, J. E., Hillman, T., & Berg, C. (2016). Adolescents' presentation of food in social media: An explorative study. *Appetite*, *99*, 121–129.

Howard, S., Adams, J., & White, M. (2012). Nutritional content of supermarket ready meals and recipes by television chefs in the united kingdom: cross sectional study. *Bmj*, *345*.

Huang, K.-Q., Wang, Q., & Wu, Z.-Y. (2006). Natural color image enhancement and evaluation algorithm based on human visual system. *Computer Vision and Image Understanding*, *103*(1), 52–63.

Inui-Yamamoto, C., Yamamoto, T., Ueda, K., Nakatsuka, M., Kumabe, S., Inui, T., & Iwai, Y. (2017). Taste preference changes throughout different life stages in male rats. *PLoS One*, *12*(7), e0181650.

Jaeger, S. R., Andani, Z., Wakeling, I. N., & MacFie, H. J. (1998). Consumer preferences for fresh and aged apples: a cross-cultural comparison. *Food quality and preference*, *9*(5), 355–366.

Jain, A., NK, R., & Bagler, G. (2015). Analysis of food pairing in regional cuisines of India. *PloS one*, *10*(10), e0139539.

Johnson, J., & Clydesdale, F. (1982). Perceived sweetness and redness in colored sucrose solutions. *Journal of food science*, *47*(3), 747–752.

Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*.

Joutou, T., & Yanai, K. (2009). A food image recognition system with multiple kernel learning. In *2009 16th IEEE International Conference on Image Processing (ICIP)* (pp. 285–288).

Joyce, J. (2006). Pandora and the music genome project, song structure analysis tools facilitate new music discovery. *Scientific Computing*, *23*(14), 40–41.

Kagaya, H., Aizawa, K., & Ogawa, M. (2014). Food detection and recognition using convolutional neural network. In *Proceedings of the 22nd ACM international conference on multimedia* (pp. 1085–1088).

Kant, I. (1987). *Critique of judgment*. Hackett Publishing.

Kaplan, D. M. (2012). *The philosophy of food* (Vol. 39). Univ of California Press.

Kawano, Y., & Yanai, K. (2014a). Automatic expansion of a food image dataset leveraging existing categories with domain adaptation. In *European Conference on Computer Vision* (pp. 3–17).

Kawano, Y., & Yanai, K. (2014b). Food image recognition with deep convolutional features. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (pp. 589–593).

Kazama, M., Sugimoto, M., Hosokawa, C., Matsushima, K., Varshney, L. R., & Ishikawa, Y. (2018). A neural network system for transformation of regional cuisine style. *Frontiers in ICT*, *5*, 14.

Kearney, J. (2010). Food consumption trends and drivers. *Philosophical transactions of the royal society B: biological sciences*, *365*(1554), 2793–2807.

Khosla, A., Das Sarma, A., & Hamid, R. (2014). What makes an image popular? In *Proceedings of the 23rd international conference on world wide web* (pp. 867–876).

Kim, K.-J., & Chung, C.-H. (2016). Tell me what you eat, and i will tell you where you come from: A data science approach for global recipe data on the web. *IEEE Access*, *4*, 8199–8211.

Koch, C., & Koch, E. C. (2003). Preconceptions of taste based on color. *The Journal of psychology*, *137*(3), 233–242.

Korsmeyer, C., & Sutton, D. (2011). The sensory experience of food. *Food, Culture & Society*, *14*(4), 461–475.

Kourouniotis, S., Keast, R., Riddell, L., Lacy, K., Thorpe, M., & Cicerale, S. (2016). The importance of taste on dietary choice, behaviour and intake in a group of young adults. *Appetite*, *103*, 1–7.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

Kusmierczyk, T., & Nørvåg, K. (2016). Online food recipe title semantics: Combining nutrient facts and topics. In *Proceedings of the 25th ACM international on conference on information and knowledge management* (pp. 2013–2016).

Kusmierczyk, T., Trattner, C., & Nørvåg, K. (2015a). Temporality in online food recipe consumption and production. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 55–56).

Kusmierczyk, T., Trattner, C., & Nørvåg, K. (2015b). Temporal patterns in online food innovation. In *Proceedings of the 24th international conference on world wide web* (pp. 1345–1350).

Laufer, P., Wagner, C., Flöck, F., & Strohmaier, M. (2015). Mining cross-cultural relations from Wikipedia: a study of 31 European food cultures. In *Proceedings of the ACM Web Science Conference* (pp. 1–10).

Lee, S.-M., Lee, K.-T., Lee, S.-H., & Song, J.-K. (2013). Origin of human colour preference for food. *Journal of Food Engineering*, *119*(3), 508–515.

Leer, J., & Krogager, S. G. S. (2021). *Research Methods in Digital Food Studies*. Routledge.

Leng, G., Adan, R. A., Belot, M., Brunstrom, J. M., de Graaf, K., Dickson, S. L., ... others (2017). The determinants of food choice. *Proceedings of the Nutrition Society*, *76*(3), 316–327.

Leu, J. H., & Banwell, C. (2016). Looking for a taste of home: A qualitative study of the health implications of the diets of australian-based southeast asian students. *Global*

Journal of Health Science, *8*(3), 101.
Li, B., Yang, Q., & Xue, X. (2009). Transfer learning for collaborative filtering via a rating-matrix generative model. In *Proceedings of the 26th annual international conference on machine learning* (pp. 617–624).
Liao, S., Zhu, X., Lei, Z., Zhang, L., & Li, S. Z. (2007). Learning multi-scale block local binary patterns for face recognition. In *International conference on biometrics* (pp. 828–837).
Liem, D. G., & Russell, C. G. (2019). The influence of taste liking on the consumption of nutrient rich and nutrient poor foods. *Frontiers in Nutrition*, *6*, 174.
Lindemann, B. (2001). Receptors and transduction in taste. *Nature*, *413*(6852), 219–225.
Lindemann, B., Ogiwara, Y., & Ninomiya, Y. (2002). The discovery of umami. *Chemical senses*, *27*(9), 843–844.
Linné, Y., Barkeling, B., Rössner, S., & Rooth, P. (2002). Vision and eating behavior. *Obesity research*, *10*(2), 92–95.
Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, *60*(2), 91–110.
Lupton, D. (2020). Understanding digital food cultures. In *Digital food cultures* (pp. 1–16). Routledge.
Maga, J. A. (1974). Influence of color on taste thresholds. *Chemical Senses*, *1*(1), 115–119.
Martınez, C., Santa Cruz, M. J., Hough, G., & Vega, M. J. (2002). Preference mapping of cracker type biscuits. *Food Quality and Preference*, *13*(7-8), 535–544.
Matsuda, Y., Hoashi, H., & Yanai, K. (2012). Recognition of multiple-food images by detecting candidate regions. In *2012 IEEE International Conference on Multimedia and Expo* (pp. 25–30).
Matsuda, Y., & Yanai, K. (2012). Multiple-food recognition considering co-occurrence employing manifold ranking. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)* (pp. 2017–2020).
McCrickerd, K., & Forde, C. (2016). Sensory influences on food intake control: moving beyond palatability. *Obesity Reviews*, *17*(1), 18–29.
Mejova, Y., Haddadi, H., Noulas, A., & Weber, I. (2015). # foodporn: Obesity patterns in culinary interactions. In *Proceedings of the 5th International Conference on Digital Health 2015* (pp. 51–58).
Mendonça, R. d. D., Lopes, A. C. S., Pimenta, A. M., Gea, A., Martinez-Gonzalez, M. A., & Bes-Rastrollo, M. (2017). Ultra-processed food consumption and the incidence of hypertension in a Mediterranean cohort: the Seguimiento Universidad de Navarra Project. *American journal of hypertension*, *30*(4), 358–366.
Mennella, J. A., & Beauchamp, G. K. (2005). Understanding the origin of flavor preferences. *Chemical Senses*, *30*(suppl_1), i242–i243.
Messina, P., Cartagena, M., Cerda-Mardini, P., del Rio, F., & Parra, D. (2020). Curatornet: Visually-aware recommendation of art images. *arXiv preprint arXiv:2009.04426*.
Messina, P., Dominguez, V., Parra, D., Trattner, C., & Soto, A. (2019). Content-based artwork recommendation: integrating painting metadata with neural and manually-engineered visual features. *User Modeling and User-Adapted Interaction*, *29*(2), 251–290.
Meyer, K. A., Kushi, L. H., Jacobs Jr, D. R., Slavin, J., Sellers, T. A., & Folsom, A. R. (2000). Carbohydrates, dietary fiber, and incident type 2 diabetes in older women.

The American journal of clinical nutrition, *71*(4), 921–930.

Meyers, A., Johnston, N., Rathod, V., Korattikara, A., Gorban, A., Silberman, N., ... Murphy, K. P. (2015). Im2calories: towards an automated mobile vision food diary. In *Proceedings of the ieee international conference on computer vision* (pp. 1233–1241).

Michel, C., Velasco, C., Fraemohs, P., & Spence, C. (2015). Studying the impact of plating on ratings of the food served in a naturalistic dining context. *Appetite*, *90*, 45–50.

Michel, C., Velasco, C., Gatti, E., & Spence, C. (2014). A taste of kandinsky: Assessing the influence of the artistic visual presentation of food on the dining experience. *Flavour*, *3*(1), 1–11.

Mikolajczyk, K., & Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE transactions on pattern analysis and machine intelligence*, *27*(10), 1615–1630.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, *26*.

Min, W., Bao, B.-K., Mei, S., Zhu, Y., Rui, Y., & Jiang, S. (2017). You are what you eat: Exploring rich recipe information for cross-region food analysis. *IEEE Transactions on Multimedia*, *20*(4), 950–964.

Min, W., Jiang, S., Liu, L., Rui, Y., & Jain, R. (2019). A survey on food computing. *ACM Computing Surveys (CSUR)*, *52*(5), 1–36.

Min, W., Jiang, S., Sang, J., Wang, H., Liu, X., & Herranz, L. (2016). Being a supercook: Joint food attributes and multimodal content modeling for recipe retrieval and exploration. *IEEE transactions on multimedia*, *19*(5), 1100–1113.

Monroe, B. L., Colaresi, M. P., & Quinn, K. M. (2008). Fightin'words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, *16*(4), 372–403.

Montouto-Graña, M., Fernández-Fernández, E., Vázquez-Odériz, M., & Romero-Rodrıguez, M. (2002). Development of a sensory profile for the specific denomination "galician potato". *Food quality and preference*, *13*(2), 99–106.

Moretti, F. (2005). *Graphs, maps, trees: abstract models for a literary history*. Verso.

Morley, J. (2012). *Nutritional modulation of neural function* (No. 28). Elsevier.

Morrill, A. C., & Chinn, C. D. (2004). The obesity epidemic in the united states. *Journal of Public Health Policy*, *25*(3), 353–366.

Moss, A. (2020). *Demographics of People on Amazon Mechanical Turk.* Retrieved 2022-09-21, from `https://www.cloudresearch.com/resources/blog/who-uses-amazon-mturk-2020-demographics/`

Müller, M., Harvey, M., Elsweiler, D., & Mika, S. (2012). Ingredient matching to determine the nutritional properties of internet-sourced recipes. In *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops* (pp. 73–80).

Mößlang, D. (2017). *Predicting the popularity of online recipes* (Unpublished master's thesis). Graz University of Technology.

Nag, N., Rao, A. N., Kulhalli, A., Mehta, K. S., Bhattacharya, N., Ramkumar, P., ... Jain, R. (2019). Flavour enhanced food recommendation. In *Proceedings of the 5th International Workshop on Multimedia Assisted Dietary Management* (pp. 60–66).

Napoletano, P. (2018). Visual descriptors for content-based retrieval of remote-sensing images. *International journal of remote sensing*, *39*(5), 1343–1376.

Ninomiya, K. (2015). Science of umami taste: adaptation to gastronomic culture. *Flavour*, *4*(1), 1–5.

Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, *29*(1), 51–59.

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, *24*(7), 971–987.

Olsen, A., Ritz, C., Kramer, L., & Møller, P. (2012). Serving styles of raw snack vegetables. what do children want? *Appetite*, *59*(2), 556–562.

Organization, W. H., et al. (2019). *Healthy diet* (Tech. Rep.). World Health Organization. Regional Office for the Eastern Mediterranean.

Palmer, S. E., & Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences*, *107*(19), 8877–8882.

Palojoki, P., & Tuomi-Gröhn, T. (2001). The complexity of food choices in an everyday context. *International Journal of Consumer Studies*, *25*(1), 15–23.

Pan, L., Pouyanfar, S., Chen, H., Qin, J., & Chen, S.-C. (2017). Deepfood: Automatic multi-class classification of food ingredients using deep learning. In *2017 IEEE 3rd international conference on collaboration and internet computing (CIC)* (pp. 181–189).

Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, *22*(10), 1345–1359.

Park, D., Kim, K., Kim, S., Spranger, M., & Kang, J. (2021). Flavorgraph: a large-scale food-chemical graph for generating food representations and recommending food pairings. *Scientific reports*, *11*(1), 1–13.

Pellegrini, C., Özsoy, E., Wintergerst, M., & Groh, G. (2021). Exploiting food embeddings for ingredient substitution. In *Healthinf* (pp. 67–77).

Pelli, D. G., & Bex, P. (2013). Measuring contrast sensitivity. *Vision research*, *90*, 10–14.

Piqueras-Fiszman, B., Alcaide, J., Roura, E., & Spence, C. (2012). Is it the plate or is it the food? assessing the influence of the color (black or white) and shape of the plate on the perception of the food placed on it. *Food Quality and Preference*, *24*(1), 205–208.

Piqueras-fiszman, b., & Spence, C. (2011). Do the material properties of cutlery affect the perception of the food you eat? an exploratory study. *Journal of sensory studies*, *26*(5), 358–362.

Prescott, J., & Bell, G. (1995). Cross-cultural determinants of food acceptability: Recent research on sensory perceptions and preferences. *Trends in Food Science & Technology*, *6*(6), 201–205.

Pyszne. (2021). *Food Color Map.* Retrieved 2021-08-31, from `https://www.pyszne.pl/odkryj/mapa-barw-jedzenia/`

Reddy, S., Nalluri, S., Kunisetti, S., Ashok, S., & Venkatesh, B. (2019). Content-based movie recommendation system using genre correlation. In *Smart Intelligent Computing and Applications* (pp. 391–397). Springer.

Reisfelt, H. H., Gabrielsen, G., Aaslyng, M. D., Bjerre, M. S., & MØLler, P. (2009).

Consumer preferences for visually presented meals. *Journal of Sensory Studies*, *24*(2), 182–203.

Rita, L., Veselkov, K., & Bronstein, M. (2020, 03). Machine Learning for Building a Food Recommendation System.

Rokicki, M., Herder, E., Kuśmierczyk, T., & Trattner, C. (2016). Plate and prejudice: Gender differences in online cooking. In *Proceedings of the 2016 conference on user modeling adaptation and personalization* (pp. 207–215).

Rokicki, M., Herder, E., & Trattner, C. (2017). How editorial, temporal and social biases affect online food popularity and appreciation. In *Eleventh International AAAI Conference on Web and Social Media.*

Rokicki, M., Trattner, C., & Herder, E. (2018). The impact of recipe features, social cues and demographics on estimating the healthiness of online recipes. In *Twelfth International AAAI Conference on Web and Social Media.*

Rolls, E. T. (2005). Taste, olfactory, and food texture processing in the brain, and the control of food intake. *Physiology & behavior*, *85*(1), 45–56.

Rozin, E. (2018). *The flavor-principle cookbook.* Hawthorn Books. doi: 10.5040/9781474296250

Rozin, P. (1982). "taste–smell confusions" and the duality of the olfactory sense. *Perception & psychophysics*.

Rozin, P. (1996). The socio-cultural context of eating and food choice. In *Food choice, acceptance and consumption* (pp. 83–104). Springer.

Rozin, P., et al. (2002). Human food intake and choice: Biological, psychological and cultural perspectives. *ANDERSON, H.; BLUNDELL J. & CHIVA, M. Food Selection: from genes to culture. Levallois-Perret: Danone Institute*, 7–26.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., … others (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, *115*(3), 211–252.

Said, A., & Bellogín, A. (2014). You are what you eat! tracking health through recipe interactions. In *Rsweb@ recsys.*

Sajadmanesh, S., Jafarzadeh, S., Ossia, S. A., Rabiee, H. R., Haddadi, H., Mejova, Y., … Stringhini, G. (2017). Kissing cuisines: Exploring worldwide culinary habits on the web. In *Proceedings of the 26th international conference on world wide web companion* (pp. 1013–1021).

Salter, J., & Antonopoulos, N. (2006). Cinemascreen recommender agent: combining collaborative and content-based filtering. *IEEE Intelligent Systems*, *21*(1), 35–41.

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information processing & management*, *24*(5), 513–523.

Salvador, A., Hynes, N., Aytar, Y., Marin, J., Ofli, F., Weber, I., & Torralba, A. (2017). Learning cross-modal embeddings for cooking recipes and food images. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3020–3028).

San Pedro, J., & Siersdorfer, S. (2009). Ranking and classifying attractiveness of photos in folksonomies. In *Proceedings of the 18th international conference on world wide web* (pp. 771–780).

Schedl, M., Knees, P., McFee, B., Bogdanov, D., & Kaminskas, M. (2015). Music recommender systems. In *Recommender systems handbook* (pp. 453–492). Springer.

Schifferstein, H. N., Kudrowitz, B. M., & Breuer, C. (2022). Food perception and aesthetics-linking sensory science to culinary practice. *Journal of Culinary Science & Technology*, *20*(4), 293–335.

Seo, S., Kim, O. Y., Oh, S., & Yun, N. (2013). Influence of informational and experiential familiarity on image of local foods. *International Journal of Hospitality Management*, *34*, 295–308.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, *27*(3), 379–423.

Sherman, P. W., & Billing, J. (1999). Darwinian gastronomy: Why we use spices: Spices taste good because they are good for us. *BioScience*, *49*(6), 453–463.

Shiner, L. (2003). *The invention of art: A cultural history*. University of Chicago press.

Silva, T., De Melo, P. V., Almeida, J., Musolesi, M., & Loureiro, A. (2014). You are what you eat (and drink): Identifying cultural boundaries by analyzing food and drink habits in foursquare. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 8).

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Spence, C. (2011). Mouth-watering: the influence of environmental and cognitive factors on salivation and gustatory/flavor perception. *Journal of Texture Studies*, *42*(2), 157–171.

Spence, C. (2018). Why is piquant/spicy food so popular? *International Journal of Gastronomy and Food Science*, *12*, 16–21.

Spence, C., Levitan, C. A., Shankar, M. U., & Zampini, M. (2010). Does food color influence taste and flavor perception in humans? *Chemosensory Perception*, *3*(1), 68–84.

Spence, C., Okajima, K., Cheok, A. D., Petit, O., & Michel, C. (2016). Eating with our eyes: From visual hunger to digital satiation. *Brain and cognition*, *110*, 53–63.

Sproesser, G., Ruby, M. B., Arbit, N., Akotia, C. S., dos Santos Alvarenga, M., Bhangaokar, R., … others (2022). Similar or different? comparing food cultures with regard to traditional and modern eating across ten countries. *Food Research International*, *157*, 111106.

Starke, A. (2019). Recsys challenges in achieving sustainable eating habits. In *HealthRecSys@ RecSys* (pp. 29–30).

Starke, A., Willemsen, M., & Snijders, C. (2017). Effective user interface designs to increase energy-efficient behavior in a rasch-based energy recommender system. In *Proceedings of the eleventh ACM conference on recommender systems* (pp. 65–73).

Starke, A. D., Willemsen, M. C., & Trattner, C. (2021). Nudging healthy choices in food search through visual attractiveness. *Frontiers in Artificial Intelligence*, *4*, 621743.

Steptoe, A., Pollard, T. M., & Wardle, J. (1995). Development of a measure of the motives underlying the selection of food: the food choice questionnaire. *Appetite*, *25*(3), 267–284.

Su, H., Lin, T.-W., Li, C.-T., Shan, M.-K., & Chang, J. (2014). Automatic recipe cuisine classification by ingredients. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: adjunct publication* (pp. 565–570).

Sweeney, K. W. (2017). *The aesthetics of food: the philosophical debate about what we eat and drink*. Rowman & Littlefield.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1–9).

Tan, H. S. G., van den Berg, E., & Stieger, M. (2016). The influence of product preparation, familiarity and individual traits on the consumer acceptance of insects as food. *Food quality and preference*, *52*, 222–231.

Taylor, C., Clifford, A., & Franklin, A. (2013). Color preferences are not universal. *Journal of Experimental Psychology: General*, *142*(4), 1015.

Teng, C.-Y., Lin, Y.-R., & Adamic, L. A. (2012). Recipe recommendation using ingredient networks. In *Proceedings of the 4th annual ACM web science conference* (pp. 298–307).

Tomkins, S., Isley, S., London, B., & Getoor, L. (2018). Sustainability at scale: towards bridging the intention-behavior gap with sustainable recommendations. In *Proceedings of the 12th ACM conference on recommender systems* (pp. 214–218).

Torrico, D. D., Fuentes, S., Viejo, C. G., Ashman, H., & Dunshea, F. R. (2019). Cross-cultural effects of food product familiarity on sensory acceptability and non-invasive physiological responses of consumers. *Food research international*, *115*, 439–450.

Trang Tran, T. N., Atas, M., Felfernig, A., & Stettinger, M. (2018). An overview of recommender systems in the healthy food domain. *Journal of Intelligent Information Systems*, *50*(3), 501–526.

Trattner, C., & Elsweiler, D. (2017a). Food recommender systems: important contributions, challenges and future research directions. *arXiv preprint arXiv:1711.02760*.

Trattner, C., & Elsweiler, D. (2017b). 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* (pp. 489–498).

Trattner, C., Elsweiler, D., & Howard, S. (2017). Estimating the healthiness of internet recipes: a cross-sectional study. *Frontiers in public health*, *5*, 16.

Trattner, C., & Jannach, D. (2020). Learning to recommend similar items from human judgments. *User Modeling and User-Adapted Interaction*, *30*(1), 1–49.

Trattner, C., Kusmierczyk, T., & Nørvåg, K. (2019). Investigating and predicting online food recipe upload behavior. *Information Processing & Management*, *56*(3), 654–673.

Trattner, C., Moesslang, D., & Elsweiler, D. (2018). On the predictability of the popularity of online recipes. *EPJ Data Science*, *7*(1), 1–39.

Trattner, C., Rokicki, M., & Herder, E. (2017). On the relations between cooking interests, hobbies and nutritional values of online recipes: Implications for health-aware recipe recommender systems. In *Adjunct publication of the 25th conference on user modeling, adaptation and personalization* (pp. 59–64).

Trefnỳ, J., & Matas, J. (2010). Extended set of local binary patterns for rapid object detection. In *Computer vision winter workshop* (pp. 1–7).

USDA. (2020). *Dietary Guidelines for Americans (2020-2025).*

Vabø, M., & Hansen, H. (2014). The relationship between food preferences and food choice: a theoretical discussion. *International Journal of Business and Social Science*, *5*(7).

van der Laan, L. N., De Ridder, D. T., Viergever, M. A., & Smeets, P. A. (2011). The first taste is always with the eyes: a meta-analysis on the neural correlates of processing visual food cues. *Neuroimage*, *55*(1), 296–303.

van Dongen, M. V., van den Berg, M. C., Vink, N., Kok, F. J., & de Graaf, C. (2012). Taste–nutrient relationships in commonly consumed foods. *British Journal of Nutrition*, *108*(1), 140–147.

Vilgis, T. A. (2013). Texture, taste and aroma: multi-scale materials and the gastrophysics of food. *Flavour*, *2*(1), 1–5.

Wadolowska, L., Babicz-Zielinska, E., Czarnocinska, J., et al. (2008). Food choice models and their relation with food preferences and eating frequency in the polish population: Pofpres study. *Food policy*, *33*(2), 122–134.

Wagner, C., & Aiello, L. M. (2015). Men eat on mars, women on venus? an empirical study of food-images. In *Proceedings of the ACM Web Science Conference* (pp. 1–3).

Wagner, C., Singer, P., & Strohmaier, M. (2014). The nature and evolution of online food preferences. *EPJ Data Science*, *3*, 1–22.

Wan, X., Woods, A. T., van den Bosch, J. J., McKenzie, K. J., Velasco, C., & Spence, C. (2014). Cross-cultural differences in crossmodal correspondences between basic tastes and visual features. *Frontiers in psychology*, *5*, 1365.

Wang, X., Kumar, D., Thome, N., Cord, M., & Precioso, F. (2015). Recipe recognition with large multimodal food dataset. In *2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 1–6).

Wansink, B., & Sobal, J. (2007). Mindless eating: The 200 daily food decisions we overlook. *Environment and Behavior*, *39*(1), 106–123.

Wilkinson, C., Dijksterhuis, G., & Minekus, M. (2000). From food structure to texture. *Trends in Food Science & Technology*, *11*(12), 442–450.

Yang, L., Cui, Y., Zhang, F., Pollak, J. P., Belongie, S., & Estrin, D. (2015). Plateclick: Bootstrapping food preferences through an adaptive visual interface. In *Proceedings of the 24th acm international on conference on information and knowledge management* (pp. 183–192).

Yang, L., Hsieh, C.-K., Yang, H., Pollak, J. P., Dell, N., Belongie, S., ... Estrin, D. (2017). Yum-me: a personalized nutrient-based meal recommender system. *ACM Transactions on Information Systems (TOIS)*, *36*(1), 1–31.

Yuan, X., Yu, J., Qin, Z., & Wan, T. (2011). A sift-lbp image retrieval model based on bag of features. In *IEEE international conference on image processing* (pp. 1061–1064).

Zampollo, F., Wansink, B., Kniffin, K. M., Shimizu, M., & Omori, A. (2012). Looks good enough to eat: how food plating preferences differ across cultures and continents. *Cross-Cultural Research*, *46*(1), 31–49.

Zellner, D. A., Lankford, M., Ambrose, L., & Locher, P. (2010). Art on the plate: Effect of balance and color on attractiveness of, willingness to try and liking for food. *Food Quality and Preference*, *21*(5), 575–578.

Zenko, B., Todorovski, L., & Dzeroski, S. (2001). A comparison of stacking with meta decision trees to bagging, boosting, and stacking with other methods. In *Proceedings 2001 IEEE International Conference on Data Mining* (pp. 669–670).

Zhang, D., & Lu, G. (2003). Evaluation of similarity measurement for image retrieval. In *International Conference on Neural Networks and Signal Processing, 2003. Proceedings of the 2003* (Vol. 2, pp. 928–931).

Zhang, Q., Elsweiler, D., & Trattner, C. (2020). Visual cultural biases in food classification. *Foods*, *9*(6), 823.

Zhang, Q., Trattner, C., Ludwig, B., & Elsweiler, D. (2019). Understanding cross-cultural visual food tastes with online recipe platforms. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 13, pp. 671–674).

Zheng, J., Wang, Z. J., & Zhu, C. (2017). Food image recognition via superpixel based low-level and mid-level distance coding for smart home applications. *Sustainability*, *9*(5), 856.

Zhu, Y.-X., Huang, J., Zhang, Z.-K., Zhang, Q.-M., Zhou, T., & Ahn, Y.-Y. (2013). Geography and similarity of regional cuisines in china. *PloS one*, *8*(11), e79161.

Zong, Z., Nguyen, D. T., Ogunbona, P., & Li, W. (2010). On the combination of local texture and global structure for food classification. In *2010 IEEE International Symposium on Multimedia* (pp. 204–211).