Research Article

Exploring the Food Pairing Hypothesis in Saudi Cuisine Using Genetic Algorithm

Muna Al-Razgan,1 Shahad Tallab,2 and Taha Alfakih 3

1Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11345, Saudi Arabia
2Department of Information Technology, College of Computer and Information Sciences, King Saud University, Riyadh 11345, Saudi Arabia
3Department of Information Systems, College of Computer and Information Sciences, King Saud University, Riyadh 11345, Saudi Arabia

Correspondence should be addressed to Muna Al-Razgan; malrazgan@ksu.edu.sa

Received 20 October 2021; Accepted 17 November 2021; Published 14 December 2021

Copyright © 2021 Muna Al-Razgan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Computational gastronomy has emerged as one of the recent hot research topics in the field of food science. It is a field that studies food in the era of data. One of the current research topics in this field is the food pairing hypothesis, which states that food with common flavor compounds tastes good when consumed together. The hypothesis has been studied in Western and European gastronomic societies. However, there are no reported studies conducted in Arab counterparts. In this study, we used genetic algorithms (GAs) to validate this hypothesis in Saudi cuisine. We developed a quantitative model and applied it to a dataset consisting of ingredients compounds found in Saudi cuisine recipes. Our research revealed that the pattern of ingredients occurring in Saudi dishes showed positive food pairing results like western cuisine. Moreover, our research directs the light to new dimensions where GA can be applied to explore the field of food science and computational gastronomy.

1. Introduction

The French gastronome Jean Anthelme Brillat-Savarin once said in 1826 “Tell me what you eat, and I’ll tell you who you are [1].” Cuisine is a vital part of any culture. It reveals many characteristics about the people who belong to it and facts related to where they live, what types of crops their land grows, and what their climate is like. World’s cuisine differs from each other in the ingredients, that is, use and the preparation style that is adopted in cooking. Throughout history, food preparation has been carried out through trial and error. Scientists also adopted this experimental method when they studied the chemical makeup of food to understand more about the ingredients and their interactions. With the help of computer technology, the field of food science was able to mature to the point of rapidly growing into a quantitative research field [2].

One of the most prominent quantitative research in food science was the studies in food pairing hypothesis that received more and more attention since it was popularized in 1992 by Heston Blumenthal, the renowned chef of the famous The Fat Duck restaurant [3]. This hypothesis claims that food that shares a lot of common flavor compounds tastes delicious together. Researchers from different cultural backgrounds investigated this hypothesis in relation to their cuisine. Some cuisine like the Western cuisine showed positive pairing, meaning the ingredients used in the recipes of that cuisine tended to share more compounds of flavor. While East Asian and Indian cuisines proved quite the opposite result, ingredients with more shared compounds were less likely to occur in the same recipe [4]. Between the Western and the East Asian, this project aims at investigating whether the food pairing hypothesis in Saudi cuisine will prove positive like the Western or negative like the East Asian.
Gastronomy as a term is the art or science of good eating. The latest outburst of data on food ingredients, their chemical makeup, and pairing combinations, in addition to the application of data mining, knowledge extraction techniques, and genetic algorithms, caused the emergence of a new field called “computational gastronomy,” a term coined by Ahnert in 2013 [5]. Computational gastronomy works in close proximity with data mining, which is defined as searching large amounts of data to unearth patterns and trends beyond simple analysis. Moreover, genetic algorithms played a big part in food pairing studies in the recent years, which belongs to the larger class of evolutionary algorithms [6]. Genetic algorithms are used to solve challenging problems in which the solution space is very huge. They are based on selection and evolution to produce numerous solutions for a given problem while maintaining the survival of the fittest approach in their execution [7]. In general, Algorithm 1 has the following main components [8]: initial population, consisting of a group of elements created randomly; fitness function, each element in the population is evaluated. The fitness function is problem-dependent and aims at giving the element a value based on how well it accomplishes a given task. Selection, elements are selected based on their fitness value; the higher the value, the higher the probability of being picked to produce the next generation. Parent selection varies depending on the problem. Standard genetic algorithm usually mirrors human reproduction having two parents, and sometimes, one parent is selected to pass the genetic makeup to the next generation. However, there are no limitations regarding the number of parents selected for the reproduction process [9]. Crossover: create children from elements “parents” selected from the previous step. There are many types of crossover. The most common is single-point crossover. There are two variations of the single-point crossover: mid-point crossover, where the chromosomes of the parents are split in the middle point, and random-point crossover, where the breaking point of the chromosome is randomly selected. The child will take one part of the chromosome of parent 1 and the second part from parent 2. Mutation: the children are mutated randomly to guarantee that the children are not all exactly the same. Mutation introduces diversity to the initial generation through the evolutionary process. The probability of mutation is typically between 1 and 2 tenths of a percent. The mutation process continues until an optimal solution is found or a number of generations have been reached [8].

Using data mining techniques and genetic algorithms on data in food science have the potential to expand the understanding of food choices and the cultural diversity of world’s cuisines. The world’s cuisines comprise many cuisines such as the European cuisine, the North American cuisine, the Latin American cuisine, the Asian cuisine, and the Arab cuisine. Arab cuisine in general differs from the rest of the cuisines in observing religious restrictions. As Islam is a major religion, its followers adhere to its teachings that forbid them from eating pork and drinking alcohol. The ethnicities of Arab cuisine are passed down from generation to generation preserving the essence of Arab cooking that revolves around the special spices and ingredients used in its preparation. There is no science behind what gives Saudi cuisine its signature savoury taste and aromatic smell. So, who is responsible for the different tastes and flavors that give each of the world’s cuisine its defining features?

One answer would be the volatile chemical compounds found in food. These chemicals stimulate the nerve receptors in the nose and the mouth to create the food experience. According to Fisher in his book “Food Flavours: Biology and Chemistry,” humans are able to detect 5000 to 10000 volatile aroma compounds; however, only 2600 compounds have been recognized [9]. In 1963, Dr. C. Weurman, the Head of the Aroma Department of the “Central Institute for Food Research” at Netherlands Organization for Applied Scientific Research, instigated the gathering of volatile compounds in food [10]. The result was a huge database called the Volatile Compounds in Food (VCF) database, which has 7,647 unique flavor compounds [11]. The database has information about the volatile compounds in a particular ingredient, the range of concentrations of a compound in an ingredient, and the natural fundamental compound of a particular ingredient [10]. A few years later in 1971, a competing database was composed by Dr. Burdock, the president of the safety and regulatory consulting firm in Orlando and the internationally recognized authority on the safety of food ingredients. This database named “Fenaroli’s Handbook of Flavor Ingredients” became the standard reference for flavor ingredients throughout the world [12]. The database has 1,530 ingredients and 1,107 flavor compounds [10]. These two databases turned out to be the standard reference for ingredients’ compounds for any study related to the field of food science.

The field of food science received more attention in Western research. Recently, in 1992 Chef Heston Blumenthal, the founder of The Fat Duck restaurant and the author of a cookbook with the same name, has found that caviar and white chocolate taste good together. Curious as to why this weird combination works well together, he asked the scientist François Benzi, who works for a flavoring and perfume company called Firmenich based in Geneva [13], if there was a scientific explanation as to why some food taste delicious when eaten together. They discovered that the secret lay within the chemical compounds of the flavor in the ingredients. This finding led to the birth of “the food pairing hypothesis,” which states that certain ingredients mix well together if they share similar chemical compounds [3].

To deeply understand how food pairing works, it is imperative to understand some points. First, the flavor humans experience is mostly due to the sense of taste and smell, but how much each sense contributes to the whole food experience? The tongue can detect five major tastes: sweet, sour, salt, bitter, and savoury, which amounts to only 20% of the overall flavor experience. The other 80% is actually due to the aroma [9]. Hence, the aroma is essentially the vital contributor to flavor perception. Secondly, the food pairing hypothesis is significantly based on the knowledge of volatile chemical compounds of the aroma. These volatile compounds are identified using a process called gas chromatography and mass spectrometry (GC-MS) [14]. GC-MS is used in the field of chemistry to isolate chemical
compounds in their gas form and measure their concentration in the sample under investigation. Applying GC-MS to food ingredients will provide a list of volatile aroma compounds out of which food scientists will select only those that will be detected by the nose. Finally, a flavor profile for each ingredient will be created based on that criteria. Chefs utilize the use of these ingredient profiles to come up with various food combinations. By using some data analysis and calculations, chefs can figure which ingredients go well with which. The more the number of shared compounds present in the ingredients, the positive the result of the pairing will be. However, Chef Blumenthal says “molecular profiling is a great tool for creativity, but it supports intuition, imagination and emotion rather than replacing them [15].”

2. Related Work

A proof of concept for this research based on previous research studies was first discussed in a symposium on data mining applications in 2016 [14]. Judging whether the food pairing hypothesis is true drew the interest of many researchers to investigate its validity across the diversity of world’s cuisine. A recent study [4] has built a bipartite flavor network that shows the chemical components shared by culinary ingredients (see Figure 1). The research used a huge dataset derived from Fenaroli’s Handbook of Flavor Ingredients [13]. The study showed that while Western and European cuisines are inclined toward using ingredients that share similar flavor compounds, the East Asian cuisine tended to do the exact opposite. It also emphasized that the ethnic and cultural aspects of a cuisine play a role in deciding the ingredients of a recipe rather than the actual chemistry of the ingredients.

A research study conducted on flavor pairing in medieval European cuisine used two datasets, one provided by Fenaroli’s Handbook of Flavor Ingredients and the other by Volatile Compounds in Food (VCF) [11]. Based on the first dataset, the research proved a positive pairing in the ingredients of medieval recipes supporting the claim of the previous study. On the other hand, when the research was tested against the second dataset, it yielded a negative pairing. Regardless of the inconclusively of the results, the authors stated that in general, the food pairing was found to be positive in European recipes. Conversely, the pairing was stronger in the medieval recipes than its counterparts of modern recipes, which was ironically due to the lack of available ingredients during that period. The study raised many questions regarding other parameters that can play a hand on the validity of the food pairing hypothesis.

The latest study conducted on testing the hypothesis was published this year focusing on the Indian cuisine [16]. The authors worked to quantify the food pairing pattern and constructed models to find features that clarify the statistical properties of the cuisine. As it is a well-known fact that spices set Indian cuisine apart from other cuisines, it is not surprising that these spices play the key role in biasing food pairing toward the negative side. The study showed the significance of the spice position in a recipe to the extent when swapping one with another significantly randomizes the pairing results. The bottom line was that the higher the number of shared compounds between any two ingredients of an Indian recipe, the lower their co-occurrence probability.

This study [17] proposed a capacitated model for food supply chain (FSC); this model is a constrained mixed integer nonlinear programming (MINLP) problem, which calculates the quality internally cost for a three echelon FSC, to reduce the total cost under the constraints of the total quality level, due to the nature of perishable products, and other constraints of capacity, demand, cost, and flow balance. The model used two techniques: an exact model based on the bound and branch technique and a GA solution method. In this work [18] review, the weaknesses and the relative strengths of gastronomy, chemical, and perceptual methods of dual flavors are highlighted. Then, the authors demonstrated how the various principles of pairing can be extended beyond the field of flavor-to-flavor pairing to consider the rapid growth of sonic seasoning.

Deciding to go deeper in their investigation, the authors of the previous study tried to find out whether the negative food pairing pattern in Indian cuisine will hold through the eight local cuisines of India consisting of Bengali, Gujarati, Jain, Maharashtrian, Mughlai, Punjabi, Rajasthani, and South Indian cuisines [19]. Using the same investigated models of the preceding study, all the eight Indian cuisines show the same negative tendency regarding the hypothesis and that spices either separately or as a category exclusively produce that result. The authors emphasized that these spices are usually a combination of spices such as garam masala and chat masala, and are also known for the therapeutic properties.
Through all these studies [4, 11, 16, 19], the researchers followed the same analytical pattern. They would first calculate the shared flavor compounds between pairs of ingredients \((I, j)\) in recipe \(R\):

\[
Ns(R) = |F(i) \cap F(j)|, \tag{1}
\]

where \(Ns(R)\) represents the number shared flavor between \(I\) and \(j\). \(F(i)\) the set with the flavor compounds found in ingredient \(I\), and \(F(j)\) the set with the flavor compounds found in ingredient \(j\).

Then, the average of flavor sharing across the recipe \(R\) is calculated using the following formula:

\[
\overline{Ns}(R) = \frac{2}{nR(nR-1)} \sum_{I \neq R, I \neq j} |F(i) \cap F(j)|, \tag{2}
\]

where \(\overline{Ns}(R)\) represents the average shared flavor of the recipe \(R\) and \(nR\) is the number of ingredients in the recipe \(R\); the formula is repeated for all recipes in a cuisine. Then, all the average results are summed and divided by the total number of recipes to obtain the average flavor sharing across the whole cuisine. Finally, the average flavor of the real cuisine \(\overline{Ns}^\text{real}\) needs to be compared to the average flavor sharing to a randomly constructed cuisine \(\overline{Ns}^\text{rand}\):

\[
\Delta Ns = \overline{Ns}^\text{real} - \overline{Ns}^\text{rand}. \tag{3}
\]

The result of \(\Delta Ns\) renders the answer to the food pairing hypothesis whether it is positive, negative, or neutral. Another study aimed at objectively testing the hypothesis without the influence of the culture or the culinary practices [20]. It was a study of the hypothesis on its basic form. Since the hypothesis states that if two food ingredients share major chemical compounds, they will provide a better eating experience when eaten together. The idea was to test whether the score of two combined ingredients was higher than the average score for the two separate ingredients. The experiment setting was as follows: a collection of seven ingredients (pear, tomato, potato, chocolate, beef, cauliflower, and anise) were tested once separately and once as a combination. The raw ingredients were cooked first and then cooled, and each of two ingredients was made into a puree. A group of fifty students evaluated each combination and gave a verdict between one to seven where one meant the taste was horrible and seven meant the taste was delicious. This sensory
experiment was performed while the participants were blindfolded. To conclude the experiment, the authors highlighted a few facts. First, under normal conditions, food would not be served as cold puree with people being blindfolded. Chefs supporting the pairing theory often engage all sensory reception to heighten the eating experience. Second, the study attempted to validate the theory objectively and scientifically.

All the research studies conducted on the food pairing hypothesis were conducted on real cuisines and on real recipes and ingredients. Furthermore, it is a known fact that in any research that focuses on validating a hypothesis, researchers need to formulate an appropriate null hypothesis suitable to compare with the original hypothesis. Therefore, the food pairing hypothesis researchers needed to construct random cuisines as the null hypothesis. They based their approach of building the random cuisine on a study by Osame Kinouchi titled “The Nonequilibrium Nature of Culinary Evolution [21].” This paper described recipes’ evolution as a process similar to the “founder effect” in biology. The founder effect is “a special case of genetic drift,” that happens when a small number of a population deviate from the original population and produce a new one. This new population is susceptible to the genetic drift due to inbreeding and limited gene pool, which, in turn, causes the new population to be idiosyncratically different from the parent population from which it is derived. Accordingly, the author proposes a copy-mutate algorithm to model the evolution of new cuisines as branching growth from a handful of founder recipes imitating the same founder effect concept as shown in Figure 2. Each branching node in the figure acts as a replicator (or a copy) that undergo mutation to generate a new node, and each line of node is a new resulting population. The copy-mutate model was widely used not only in studies investigating the food pairing hypothesis, but also in a study that aims at analyzing China’s cuisines similarity based on geographical distance and climate [22].

The study followed the same methodology used in food pairing investigation to the point of building the random cuisine of China at which point the study went on to examine the similarity among regional cuisines in China in terms of temperature and physical proximity. It found that geographical closeness appears to be a key influence in the shaping of regional cuisines rather than the temperature. Based on the finding, the proposed copy-mutate model [21] showed that the results of the model agree with the empirical findings of the study. The copy-mutate model used in the previous studies is one type of genetic algorithms.

On the other hand, a study argued that flavor pairing does not affect the tasting experience to a great extent and that the sequential order of tasting food played a greater part [23]. It gave many examples on how the order people taste food can alter how they perceive different tastes and flavors. For example, drinking orange juice in the morning after brushing teeth will make the juice taste sour and unpleasant, and that is because there is a chemical substance in the toothpaste called sodium laurel sulphate (SLS) that blocks the sweet receptors in the tongue making any thing sweet tastes sour or bitter. The study evaluated the flavor pairing according to similarity, contrast, and synergy (food blending that brings on new flavors). It also proclaimed that the psychological aspects influence how people perceive food flavors. For instance, most people receive vanilla as something sweet due to the types of food always associated with it, not because the vanilla as a spice is sweet itself.

3. Methodology and System Model

Our mythology follows the same approach used in the previous studies of food pairing hypothesis investigations [4, 11, 16, 19]. However, it differs in system breakdown and modeling. We offer a clearer picture and organized explanation to the method of food pairing calculations. Our project consists of three main stages, which will be illustrated in the following subsections. The stages are data collection, data preprocessing, and the result of food pairing. The data modeling consists of calculation of real cuisine and random cuisine. We will elaborate on each stage individually to make the system model of food pairing as clear as possible. One main contribution in our project is clearly shown in the data modeling phase where we need to deal with real and random cuisines and the result of their interactions. Moreover, although data modeling is an interesting phase, it is the most challenging one. Figure 3 shows our model in the most basic level. We will dig into each stage as we explain what is being accomplished.

3.1. First Stage: Data Collection. The first step will be assembling recipes from old traditional Saudi cookbook such as the cooking principles of Saudi and middle eastern cuisine by Mrs. Rabha Hafiz [24]. Afterwards, for each recipe, a list of all ingredients occurring in it will be gathered. Finally, a list of flavor compounds for each ingredient will be compiled to define the flavor profile for every ingredient. The information about the flavor compounds will be gathered from “Fenaroli’s Handbook of Flavor Compounds [13].” We planned to collect ingredients’ chemical compositions manually, but the process proved to be extremely time-consuming. Fortunately, we found an available compositions dataset from the Indian food pairing studies. To recapture this part, each recipe (R) is made up of a number of ingredients (i, j, …). And each ingredient is composed of many flavor compounds (F) (see Figure 4 for clearer
understanding). At the end of this stage, the corpus of real cuisine will be ready for the next stage.

3.2. Second Stage: Data Cleaning and Preprocessing. Once the recipes are collected and duplicates are removed, a list of ingredients occurring in all recipes will be made after unifying ingredients that have aliases. Ingredients that do not have a flavor profile will be excluded. An additional step will be translating ingredient names from Arabic to English and then highlighting the ingredients unique to Saudi cuisine. Essentially, after the data are collected, the following will be performed: (1) Discard recipes with a single ingredient. (2) Alias ingredients with different spellings and forms, for example, diced tomato and sliced tomato; both are tomatoes. And (3) Discard ingredients with unknown flavor compounds.
3.3. Third Stage: Data Modeling. This is the main part of our project to follow and obtain the result of our hypothesis. This stage is divided into two parts: the first part is related to the calculation of average flavor sharing of the real cuisine corpus, and the second part is related to the random cuisine generation and average flavor sharing calculation.

In Table 1, we will provide notations for variables and quantities to make tracing the process of our methodology easier.

3.3.1. Real Cuisine Average Flavor Sharing Calculation. For better explanation of the process of flavor sharing calculation, it is advisable to do so via an example illustrated in Figure 5. Let us consider the Saudi cuisine (i) that we will calculate its flavor sharing. The first step is to select a recipe (R) called green beans salad from Saudi’s recipes. Suppose our recipe has the following ingredients: almonds, green beans, coconut, and anchovies. For every ingredient, we get the list of flavor compounds present in that ingredient. Secondly, for each pair of ingredients such as green beans (i) and coconut (j), we calculate the number of shared compounds (Fi) of green beans and (Fj) for coconut according to the following equation:

\[ N_{s}(R) = |F(i) \cap F(j)|. \] (4)

The intersection between the set of flavor compounds of green beans and the set of flavor compounds of coconut will be the common flavor shared between the two ingredients. This step will loop for each unique pair of ingredients in R as illustrated in Figure 6. Consequently, the average number of shared flavor compounds in recipe R with the four different ingredients in our example (almonds, green beans, coconut, and anchovies) is calculated by the following equation:

\[ \overline{N_{s}}(R) = \frac{2}{n_{R}(n_{R} - 1)} \sum_{i,j \in R, i \neq j} |F_{i} \cap F_{j}|. \] (5)

The resulting number \( \overline{N_{s}}(R) \) is the average flavor sharing on recipe level. We need to calculate it for the whole cuisine; therefore, the process will loop for every single recipe in the Saudi cuisine to get the average flavor sharing result on a cuisine level.

The average of flavor sharing of the whole cuisine denoted by \( \overline{N_{s}} \) is calculated with the following equation:

\[ \overline{N_{s}} = \frac{\sum_{R} N_{s}(R)}{N_{c}} \] (3).

Next, \( \overline{N_{s}} \) needs to be compared to a null hypothesis to validate whether \( \overline{N_{s}} \) is indicative of whether similar flavors taste good together. Thus, \( \overline{N_{s}} \) will be compared to a randomly generated cuisine:

\[ \Delta N_{s} = N_{s}^{\text{real}} - N_{s}^{\text{rand}}. \] (7)

The difference between the average of flavor sharing of the real and random cuisines will answer the hypothesis, which is Stage 4 in our system model as shown in Figure 3. The generation of the random cuisine is shown in Figure 4. The generation of the random cuisine is accomplished in part 2 in the next section. Figure 7 summarizes the flavor sharing calculation.

3.3.2. Random Cuisine Corpus Generation and Average Flavor Sharing Calculation. Building the random cuisine is an integral part of the equation that will solve the food pairing hypothesis for Saudi cuisine.

The genetic algorithm copy-mutate model used in [21] will facilitate this process. Generating the random cuisine needs to be broken down into achievable steps based on previous studies, which is shown in the following Algorithm 2.

Figures 8–10 show that the random cuisine generation is demonstrated with a tiny example. Once the first collection of random recipes is generated, the following steps are ready to be performed. Figures 9 and 10 clarify the next step of random cuisine generation using the same example used in Figure 9, which shows the first iteration of the interaction process. It represents a simulation process. The details are written in the upper left corner box. Figure 10 picks up where the previous process stopped. Tracing the steps of the interaction can also be performed by following the numbered steps. Figure 11 shows how cuisine corpus can be randomly created. Figure 12 shows the random cuisine corpus generation and average flavor calculation.

4. Food Pairing Result Analysis

After running the functions on our dataset, beginning from pairs of ingredients, the average flavor sharing of recipes and that of the Saudi cuisine \( N_{s}^{\text{real}} \) are calculated. Our analysis showed that the flavor sharing value observed for our cuisine was found to be \( N_{s}^{\text{real}} = 27.936 \) ranging between 83.1429 and 1.000. We also calculated the \( N_{s}^{\text{rand}} \) using the only ingredients found solely in Saudi cuisine. However, because of the unpredictability of copy-mutate random function, we needed to neutralize that effect. Therefore, we generated forty sets of 100 random copy-mutate recipes and then studied the average of all sets. We found that the average flavor sharing of random corpus ranges between 35.8433 and 12.8111, and the average results of \( N_{s}^{\text{rand}} = 23.596. \) According to these results, we applied the values to the equation that will answer our question: is the food pairing hypothesis of Saudi cuisine going to be positive or negative?

\[ \Delta N_{s} = 27.936 - 23.596 = 4.340. \] (8)

Apparently, the positive results mean that ingredients that share many flavor compounds are more likely to be used together in our cuisine. Table 2 shows the values of flavor sharing of both random and real cuisines and the difference between them.

Initially, we thought that our cuisine might render negative results like the Indian cuisine since both cuisines
Table 1: Notations and their description.

| Notations | Description |
|------------|-------------|
| $R$        | Represent a recipe |
| $n_R$      | The number of ingredients in recipe $R$ |
| $N_s(R)$   | The average flavor sharing on recipe level |
| $F(i)$     | Set of flavor compounds found in ingredients $i$ |
| $N_s$      | The average flavor sharing on cuisine level |

Hypothesis notations

| $N_s^{real}$ | The average flavor sharing of real cuisine |
| $N_s^{rand}$ | The average flavor sharing of random cuisine |
| $\Delta N_s$ | The difference between the flavor sharing of the real and random cuisine |

Random cuisine notations

| $N_R$ | Number of all recipes in the cuisine |
| $N_i$ | Number of all ingredients |
| $N_i(t)$ | Number of initial ingredients at time $t$ |
| $N_r$ | Number of all recipes |
| $N_r(t)$ | Number of all recipes at time $t$ |
| $f_i$ | Fitness value of ingredient $i$ |
| $m$ | Number of ingredients to be mutated in each recipe. |

Saudi cuisine

- green beans
- fc1
- fc4
- fc5
- fc8
- fc9
- fc6
- almonds
- anchovies
- green beans salad
- kabsah
- Jareesh

unique pairs of ingredients in green beans salad:
- $|F(\text{green beans}) \cap F(\text{coconut})| = \{fc1, fc4\}$
- $|F(\text{coconut}) \cap F(\text{anchovies})| = \{fc8\}$
- $|F(\text{anchovies}) \cap F(\text{almonds})| = \{\}$

Average flavor sharing of the Saudi cuisine:

$$\bar{N}_s = \frac{\sum_R N_s(R)}{N_c}$$

$N_s$ (green beans salad) = average flavor sharing of green beans salad recipe

$N_c$ = number of all recipes in a cuisine
$N_s$ = number of recipes in cuisine
$\text{Avg} (N_s) = \sum (N_s(\text{Salad}), N_s(\text{Kabsah}), N_s(\text{Jareesh}))/3$

This process is applied for all cuisines in the real cuisine corpus.

Figure 5: Tiny example of real cuisine average flavor sharing calculation.
Figure 6: Co-occurring ingredients using.

Figure 7: Flavor sharing calculation of real cuisine.

(1) Initial pool containing $N$ number of ingredients.
(2) Generate $K$ number of recipes containing $S$ ingredients randomly constructed from the initial pool of ingredients.
(3) From the pool of $K$ recipes select one recipe randomly.
(4) Copy the recipe.
(5) Mutate the copied recipe $M$ times.
(6) Randomly select an ingredient from the copied recipe.
(7) Randomly select an ingredient from ingredients pool.
(8) Compare fitness value of the two ingredients.
(9) Replace the ingredient in the recipe with the ingredient having the higher fitness value.
(10) Add the mutated recipe to the pool of recipes.
(11) Introduce a new ingredient to the pool of ingredients.
(12) Repeat steps 3 to 7 until the corpus of random recipes match the number of recipes in the real cuisine.

Algorithm 2: GA food pairing.
share similarities of using spices. In Indian cuisine, spice was proved to be the critical ingredient that tipped the scale of Indian cuisine toward the negative side. However, the evidence showed that even though Saudi cuisine uses spice, the usage is limited to a handful of spices. These are consisting of: black pepper, cumin, cardamom, and coriander. This list of spice is incomparable to the extensive and versatile list of spices the Indian cuisine uses, which in turn made the Indian cuisine show negative food pairing results.

During our project, we observed that there are many cookbooks that offer authentic Saudi recipes, and other recipes from Arabic or Western cuisine with some modifications that gave the modified dishes the distinct flavor of our cuisine. Nevertheless, the numbers of truly authentic Saudi recipes found in those books were relatively small and invariant. As a result, none of the cookbooks we browsed offered 100% truly authentic Saudi recipes. It could be because Saudi recipes are influenced by people coming to Haj each year. Moreover, some Saudi recipes are handed down from generation to generation by word of the mouth, and we could not find these recipes in any cookbook. Nonetheless, thanks to nowadays-rapid spread of social media channels, which made these recipes begin to emerge and spread on Instagram, Snapchat, and cooking blogs and forums. Due to the time constraints during the development of our project, we were unable to collect those recipes due to the unstructured layout of media websites and the enormous number of preparation way that makes it difficult to choose one recipe.

Therefore, we adopted an authentic Saudi cookbook to collect our recipes for our project. As stated before, we have collected approximately 100 dishes to build our real cuisine. In order to achieve progress in our project, we calculated the fitness value for each ingredient. Our aim was to scale the frequency of ingredients to be a value between 0 and 1. Unfortunately, these aspects are not easy to evaluate. Previous studies made their calculation of fitness value either randomly or according to the frequency of ingredients [25]. In our study, we determined the fitness value of the ingredients based on their chemical composition and their frequency [21].

The conclusion of our findings may be affected by the available dataset on chemical compositions of ingredients. Fenaroli’s and VCF’s datasets on chemical composition are the reference datasets for researchers in food science. It has...
been found that Fenaroli’s dataset has a vast collection of ingredients, but it is lacking in the number of compositions documented in each one. On the other hand, VCF has a smaller collection of ingredients, but the information in chemical compositions in each ingredient is richer and more comprehensive. However, we were unable to obtain access to VCF dataset due to access restrictions; thus, we used Fenaroli’s dataset in our study.

4.1. Most Frequent Ingredients in Saudi Cuisine. From our dataset collections, Table 3 shows extract the most used ingredients in Saudi cuisine. Interestingly, black pepper, onion, tomato, garlic, sunflower oil, and cumin are the most used ingredients in our cooking. Mainly, these species are the fundamental ingredients, which make up most of our dishes from stews to soups and rice.

4.2. Ingredients’ Visualization. To make our findings more appearing, Figure 6 displays the most co-occurring ingredients in Saudi cuisine using a Voyant online visualization tool [26]. Furthermore, Figure 13 shows the most important ingredients in Saudi cuisine, where the size of the bubble indicates the frequency; the larger the bubble, the higher value of the ingredient frequency using Tableau application.
Figure 14 shows the important multi-ingredients co-occurrence as a heat map figure using VOSviewer application. VOSvewier is a tool that visualize co-occurrence networks of important terms [27]. Both figures present the ingredients that make our cuisine unique among other cuisines.

iteration 2 scenario:

1: either the resulted recipe of the previous iteration or the Saudi recipe Rs1 will enter the interaction (mutation) process at random.
2: the selected recipes will be copied as a template.
3: suppose “turmeric” is randomly chosen from the copied recipe.
4: suppose “cumin” is randomly picked from the pool of ingredients.
5: The fitness of “turmeric” is compared to the fitness of “cumin” and the ingredients with the highest fitness value is selected.
6: since “turmeric” has the lowest value, it is removed from the recipe
7: “cumin” replaces the removed ingredient.
8: the mutated recipe is added to the pool of Saudi cuisine recipe
9: “lemon” a new ingredient is added
10: Rs4 recipe is generated following the same pattern until no new recipe are being generated or until a number similar to the total number of recipes of the real cuisine is achieved
11: the same process is repeated for all cuisines under study.

Figure 10: Recipe interaction/mutation process iteration 2.
Random cuisine corpus generation and average flavor calculation

- **N (O)** = a pool of 10 unique ingredients
- Each ingredient had a fitness value assigned by a fitness function calculated based on the ingredient's intrinsic properties

**N (O) = 10 unique ingredients**

Generate a recipe **R** containing **K** random ingredients from **N** pool

**Table 2:** Values of flavor sharing of both random and real cuisines and the difference between them.

| Notations | Value     | Description                                      |
|-----------|-----------|--------------------------------------------------|
| Hypothesis notations |           |                                                  |
| $N^\text{real}$     | 27.9365   | The average flavor sharing of real cuisine       |
| $N^\text{rand}$     | 23.596    | The average flavor sharing of random cuisine     |
| $\Delta N$s       | 4.340     | The difference between the flavor sharing of the real and random cuisines |
Table 3: Topmost frequent ingredients of Saudi.

| Ingredients                          | Frequency of ingredient | Fitness value [0, 1] |
|--------------------------------------|-------------------------|----------------------|
| Black pepper                        | 78                      | 0.987341772          |
| Onion                               | 68                      | 0.860759494          |
| Tomato                              | 53                      | 0.670886076          |
| Garlic                              | 49                      | 0.620253165          |
| Sunflower oil                       | 44                      | 0.556962025          |
| Cumin                               | 43                      | 0.544303797          |
| Mutton                              | 27                      | 0.341772152          |
| Cardamom                            | 26                      | 0.329113924          |
| Chicken                             | 26                      | 0.329113924          |
| Egg                                  | 25                      | 0.316455696          |
| Ghee                                | 23                      | 0.291139241          |
| Cayenne                              | 21                      | 0.265822785          |
| Cinnamon                            | 21                      | 0.265822785          |
| Wheat                               | 20                      | 0.253164557          |
| Beef                                | 20                      | 0.253164557          |
| Rice                                | 19                      | 0.240506329          |
| Lemon                               | 18                      | 0.227848101          |
| Coriander_cumin_seeds_powder        | 17                      | 0.215189873          |
| Coriander                           | 15                      | 0.189873418          |

Figure 13: The frequency of ingredients.
5. Conclusions

Over the last years, food science gained a lot of attention in the Western and European countries, especially about the new field of computational gastronomy. Moreover, the application of knowledge extraction in food science has been opened up new and unexplored directions for the food business. The food pairing hypothesis was all the rage during the 1990s. Not only chefs got interested in this hypothesis, but also scientists from various disciplines started to explore it in relation to different national cultures and cuisines. Fortunately, once this new field is recognized in countries of the Arab world, it will open doors for future opportunities to restaurants and home cooking businesses to become confidently innovative. Suffice to say that a project like this will scratch the surface to what lay beneath computational gastronomy. In this project, we have studied the food pairing hypothesis in Saudi cuisine. We have collected our recipes for an authentic old Saudi cookbook and used available dataset for compounds found in our ingredients. Then, we programmed the mathematical models to obtain the results. From our investigation, we found that Saudi cuisine tends to contain ingredients that share similar flavor compounds. This means that the food pairing hypothesis for our cuisine was found to be positive.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

There are no conflicts of interest regarding the publication of this article.

Acknowledgments

This research was supported by Researchers Supporting Project, King Saud University, Riyadh, Saudi Arabia, under project number RSP-2021/206.

References

[1] C. Gates St-Pierre, “Faunal remains as markers of ethnicity: a case study from the St. Lawrence Estuary, Quebec, Canada,” in Proceedings of the Ponencia Presentada en la X Conferencia del Consejo Internacional de Arqueozoología (ICAZ), México, DF, April 2006.

[2] “What are importance of computer to food-science and technology,” 2016, https://www.answers.com/Q/What_are_importance_of_computer_to_food-science_and_technology.

[3] H. Blumenthal, The Big Fat Duck Cookbook, Bloomsbury, London, UK, 2008.

[4] Y. Y. Ahn, S. E. Ahnert, J. P. Bagrow, and A. L. Barabási, “Flavor network and the principles of food pairing,” Scientific Reports, vol. 1, no. 1, pp. 196–197, 2011.

[5] S. E. J. F. Ahnert, “Network analysis and data mining in food science: the emergence of computational gastronomy,” vol. 2, no. 1, pp. 1–3, 2013.

[6] H. Alayed, F. Dahan, T. Alfakih, H. Mathkour, and M. Arafah, “Enhancement of ant colony optimization for QoS-aware Web service selection,” IEEE Access, vol. 7, pp. 97041–97051, 2019.

[7] C. R. Houck, J. Joines, and M. G. J. N.-I. T. Kay, “A genetic algorithm for function optimization: a Matlab implementation,” vol. 95, no. 9, pp. 1–10, 1995.

[8] F. J. U. c. i. p. e. b. g. h. Busetti, Genetic Algorithms Overview, Wiley, Hoboken, NJ, USA, 2007.

[9] C. Fisher and T. R. Scott, Food Flavours: Biology and Chemistry, Royal Society of Chemistry, London, UK, 2007.

[10] F. Team, “The variant call format (VCF) version 4.2 specification,” 2015, https://samtools.github.io/hts-specs/VCFv4.2.pdf.
[11] K. R. Varshney, L. R. Varshney, J. Wang, and D. J. Myers, "Flavor pairing in medieval European cuisine: a study in cooking with dirty data," 2013, https://arxiv.org/abs/1307.7982.

[12] G. A. Burdock, *Fenaroli's Handbook of Flavor Ingredients*, CRC Press, Boca Raton, FL, USA, 2016.

[13] H. J. T. G. Blumenthal, *Weird but wonderful*, vol. 4, 2002.

[14] S. T. Tallab and M. S. Alrazgan, "Exploring the food pairing hypothesis in Arab cuisine: a study in computational gastronomy," *Procedia Computer Science*, vol. 82, pp. 135–137, 2016.

[15] N. Caporaso, "The impact of molecular gastronomy within the food science community," in *Gastronomy and Food Science*, pp. 1–18, Elsevier, Amsterdam, Netherlands, 2021.

[16] A. Jain and G. J. Bagler, "Spices form the basis of food pairing in Indian cuisine," 2015, https://arxiv.org/abs/1502.03815.

[17] D. Lamiae, A. Jabri, A. El Barkany, and A.-M. Darcherif, "Optimization of fresh food distribution route using genetic algorithm with the best selection technique," in *Constraint Handling in Metaheuristics and Applications*, pp. 175–199, Springer, Berlin, Germany, 2021.

[18] C. Spence and F. Science, "Gastrophysics: getting creative with pairing flavours," *International Journal of Gastronomy and Food Science*, Article ID 100433, 2021.

[19] A. Jain, R. N. K, and G. Bagler, "Analysis of food pairing in regional cuisines of India," *PLos One*, vol. 10, no. 10, Article ID e0139539, 2015.

[20] M. Kort, B. Nijssen, K. van Ingen-Visscher, and J. Donders, "Food pairing from the perspective of the 'volatile compounds in food' database," in *Expression of Multidisciplinary Flavour Science: Proceedings of the 12th Weurman Symposium*, pp. 589–592, Institute of Chemistry and Biological Chemistry, Interlaken, Switzerland, 2010.

[21] O. Kinouchi, R. W. Diez-Garcia, A. J. Holanda, P. Zambianchi, and A. C. Roque, "The non-equilibrium nature of culinary evolution," *New Journal of Physics*, vol. 10, no. 7, p. 073020, 2008.

[22] Y.-X. Zhu, J. Huang, Z.-K. Zhang, Q.-M. Zhang, T. Zhou, and Y.-Y. Ahn, "Geography and similarity of regional cuisines in China," *PLoS One*, vol. 8, no. 11, Article ID e79161, 2013.

[23] C. Spence, Q. J. Wang, and J. J. F. Youssef, "Pairing flavours and the temporal order of tasting," vol. 6, no. 1, pp. 1–15, 2017.

[24] R. A. Hafiz, *The Basics of the Saudi and Middle-Eastern Cooking*, Al-Khuraiji Publishing, Riyadh, Saudi Arabia, 1990.

[25] C. Chabot, C. Stolte, and P. J. T. S. Hanrahan, *Tableau software*, vol. 6, 2003.

[26] L. J. Sampsel, "Voyant tools," *Music Reference Services Quarterly*, vol. 21, no. 3, pp. 153–157, 2018.

[27] H. S. Ehtahed, O. Tabatabaei-Malazy, A. R. Soroush et al., "Worldwide trends in scientific publications on association of gut microbiota with obesity," *Iranian journal of basic medical sciences*, vol. 22, no. 1, pp. 65–71, 2019.