2
Literature Survey

The chapter provides an exhaustive overview of previous works and literature in the areas dealt with in the thesis. The chapter is broadly divided into six sections. The first section discusses the role played by food in our life. The next section briefly examines the current research conducted in the area of gastronomy and the emergence of technological interventions into the domain of food, including the advent of computational gastronomy. The third section describes the sensation of flavor with the help of relevant literature. The next section sums up literature pertinent to explain the food pairing principle. It also surveys the literature available for the study of flavors and available databases. The penultimate section of the chapter deals with the research hitherto available covering the significance of spices and their health benefits and efforts towards integrating the data. The final section elaborates on the text mining techniques used in integrating the data available for the health effects of spices and existing efforts towards the direct integration of spice-disease associations from literature.

2.1 FOOD—AN INTRODUCTION

Food is one of the primary and most essential components of human survival. Cooking is a unique ability possessed by humans. We are the sole species on earth that combine ingredients to cook our meal . Food is an integral part of human history, culture, and identity. Each culture has cultivated its unique ways for the preparation of meals encoded in its Cuisine, its culinary signatures. They encompass recipes and methods of preparation passed down through centuries and are markers of culinary fingerprints.

Our food habits have shaped a major part of human history and evolution [Wrangham, 2009]. Cooking has changed the structure of our brains. Studies [Fonseca-Azevedo and Herculano-Houzel, 2012; Navarrete, Schaik, and Isler, 2011; Wrangham, 2009] argue that cooking played a significant role in the evolution of human brains. The shift from eating raw food to a cooked diet enabled human beings to develop bigger brains as cooked diet decreased the energy requirement needed for digestion of food and ensured more time to grow neurons much required for the development of higher cognitive functions [Fonseca-Azevedo and Herculano-Houzel, 2012]. This trade-off between metabolic energy and brain size aided by cooking with fire acted as a driving force to acquire increased brainpower, which sets us apart from other mammals, including primates.

As omnivores, humans are faced with a dilemma as they are presented with an array of food choices in their immediate environment in order to satisfy their nutritional needs as well as avoid foodborne illnesses [Rozin, 1976, 2015]. Recently this dilemma has extended in the form of a cultural one, as globalization has brought more food choices to the hands of eaters [Pollan, 2006], not necessarily all healthy. Each individual's unique preferences and aversions of food are based on predisposed biological tendencies but are further cultivated and modified through experiential learning [Ventura and Worobey, 2013]. The primary influence for food preference in humans involves food perception involving olfactory and gustatory mechanisms. Our Cooking and thereby human food preferences have also evolved through a complex exchange of culture [Appadurai, 1988; Pollan, 2014], climate, geography [Zhu et al., 2013] and genetics [Ventura and Worobey, 2013]. Experiences also govern our food choices, since it is not possible to identify on sensory grounds alone whether a potential food is nutritive or not, and whether it is toxic or not. Individual food preferences have shown some universal biological predispositions, including

preferences for sweet and fat texture; avoidance of irritation and bitter and strong tastes; a tendency to be interested in, and suspicious of new foods and a set of genetic learning predispositions. Most of the determinants of human food choice fall in the domain of psychology (individual experience) and either direct or indirect cultural influences. Our culinary practices have evolved in various dimensions thereby shaping our cognition [Gómez-Pinilla, 2008].

An individual's food preference is determined by multiple factors and is mainly shaped by biology, psychology and cultural [Rozin, 1976, 2015] and genetic [Krebs, 2009] factors. Cultural and traditions about diet and nutrition have impacted types of foods grown and consumed. Cultures have developed elaborate ways of selecting and preparing foods. These systems, which we can call cuisines, have no doubt been influenced by our innate dispositions, along with other cultural forces, including social needs and advances in technology. The result is that cuisines have aspects that can be related to our pre-cultural life. On the other hand, food has also taken on a life of its own in cultures, with many functions other than nutritional, such that in some cases our food predispositions have been ignored or even reversed.

Not only do humans need to combine ingredients so as to avoid poisonous or harmful ones, but they also need to cook tasty food. Cooking tasty food is crucial for the acceptance of recipes in a community or culture. The success in the art of cooking tasty recipes depends, among other things, on pairing the right ingredients in the right quantity. The idea of food pairing has its origins in this notion. While food sensation is a result of the interplay between various aspects of ingredients, such as texture, color, temperature and sound, flavor plays a dominant role in specifying culinary fitness of ingredients and their combinations [Spence, 2012; Spence, Hobkinson, Gallace, and Fiszman, 2013]. Flavor mediated food perception, primarily involving molecular interactions with olfactory and gustatory receptors, is crucial in developing food preferences in humans and has coevolved with nutritional needs [Breslin, 2013]. The molecular composition of food dictates the sensation of flavor [Burdock, 2010]. Each ingredient is characterized by a set of chemical compounds that forms its flavor profile. Flavor profile provides us an effective tool for exploring patterns in ingredient composition of recipes. Towards completing the objectives of the thesis, we obtained the flavor profiles for all ingredients in the Indian cuisine with the help of previously published data [Ahn, Ahnert, Bagrow, and Barabási, 2011], a resource of flavor ingredients [Burdock, 2010] and by extensive literature search.

2.2 COMPUTATIONAL GASTRONOMY—FROM COOKBOOKS TO CULINARY WEBSITES

The term 'gastronomy' is in general, applied to the science and art of anything concerning cooking and consumption of food. It is an umbrella term that encompasses varied aspects of cooking—combining various dimensions of it such as food, cooking, culture, aroma, taste, culinary techniques, food preservation and health. The introduction of computational techniques and data-driven approaches is changing the way cooking and in general gastronomy is perceived. 'Computational Gastronomy' is largely unchartered territory and amalgamates interdisciplinary fields such as culinary science, biology and computer science. Broadly, it deals with applying computational techniques for improving anything and everything concerning human food consumption, ranging from the creativity in cooking, and satisfaction of the consumer to health issues [Moller, 2013]. Data-driven techniques are employed to make sense of the massive data generated in the discipline. Application of the principles used to predict complex systems can benefit computational gastronomy as they can help reveal various phenomena in the discipline [Moller, 2013]. In this thesis, the scope of our computational gastronomic investigations pertains to the study of data available for traditional recipes of Indian Cuisine, its ingredients and flavors.

High throughput methods in biology have created an immense overflow of data in the area of biological sciences. The large amounts of data generated by these technologies have given rise to entirely new research areas in biology, such as computational biology and systems biology. Systems biology attempts to understand biological processes at a 'systems' level, which is

particularly indicative of the potential advantage that large datasets and their analysis can offer to biology, and other fields of research [Ahnert, 2013]. Computational Biology uses computational techniques to make sense of the large data generated. Molecular gastronomy is among one of the attempts to probe the chemistry and physics of preparing a dish [This, 2006].

The data in computational gastronomy relevant for this study mainly comprises recipes which, in turn, are a list of ingredients, their quantity, and stepwise method of preparation. Culinary recipes are prime examples of cultural algorithms with a strong capacity for stabilization, innovation and transmission. Cookbooks are largest artifacts holding the history of human food preferences. They are the prime source of culinary data and are representatives of the respective regions' culinary traditions. They serve as important tools for studying the pattern of cuisines, the importance and relevance of ingredients used in that cuisine, their method of preparation and the way they are combined in the recipe [Kinouchi, Diez-Garcia, Holanda, Zambianchi, and Roque, 2008]. These elements have made traditional cookery books a useful source of information for food researchers [Appadurai, 1988].

The internet has brought a revolution in the culinary data. The information only made available by cookbooks were readily made available through culinary websites and blogs. They made learning to cook possible for everyone. Even ordinary people without culinary degrees could cook up intricate and sophisticated dishes. This led to globalization in the culinary world. This was not only beneficial for food enthusiasts but also for data scientists. The availability of large amounts of data in the electronic and structured form means that scientists investigating food can uncover patterns in global food preferences**.**

The study of food needs an interdisciplinary approach with an emphasis on data-driven studies in addition to classical methods. Increasing and easy access to the availability of recipes and ingredients online has prompted new technologies in the area of food from recipe recommendation to recipe generation. Corporates like IBM have also ventured into food industry analytics by attempting to create a computer program that generates recipes. In this thesis, we limit our examination to a particular and crucial aspect of culinary data, namely the relationship between its recipes and ingredients, and overlook for the moment the vital role played by culinary preparations.

2.3 SENSATION OF FLAVOR

Flavor mediates a crucial part in food sensation of taste, among other factors such as appearance and feel of the food [Ahn et al., 2011]. The flavor of food is the combined sensation conveyed by the chemical senses that is the experience of taste, smell, and chemical irritation together. The sensation of smell occurs when a chemical stimulates taste receptors of the tongue, and other parts of the oropharynx [Valentová and Panovská, 2003]. The sense of taste is stimulated when nutrients or other chemical compounds activate specialized receptor cells within the oral cavity. Taste helps us decide what to eat and influences how efficiently we digest these foods. The early humans would have used their sense of taste to identify nutritious food items. The risks of making poor food selections when foraging not only entail wasted energy and metabolic harm from eating foods of low nutrient and energy content, but also the harmful and potentially lethal ingestion of toxins. For omnivorous species, it is essential that many different foods are sampled, and their post-ingestion consequences, the nutritional rewards or punishments, are associated with their sensory properties. These associations are what ultimately shape our food likes and dislikes and guide our future eating decisions. In this way, taste serves as a marker, especially in the context of complex flavors, for the nutrient and toxic load of foods [Breslin, 2013]. Each individual's unique preferences and aversions of food are based on predisposed biological tendencies but are further cultivated and modified through experiential learning. Available data [Ventura and Worobey, 2013] suggests that young children are biologically primed to prefer and consume foods that are sweet, salty, and savory, as well as

flavors paired with energy density. Fortunately, food preferences are not rigid and are shaped in response to a number of social and environmental factors. These choices references are a strong motivation of dietary intake in both children and adults; thus, an understanding of these factors is an essential basis for understanding how preferences can be modified to promote healthful diets across the life course best. Other cues that contribute to the flavor sensation include the trigeminal inputs such as vision, audition and oral somatosensation.

The sensation of flavor is often described in terms of descriptors or terms such as fruity, butter, floral, etc. Each ingredient in our food contains a set of chemical compounds that impart them their characteristic flavor. The process of cooking itself may lead to the evaporation of certain flavor compounds or changes in the structure of some. In this thesis, at present, we ignore the effect of cooking on the flavor composition of food ingredients as well as their combined effect.

2.4 FOOD PAIRING PRINCIPLE

As discussed in earlier sections, palatability is primarily determined by flavor, although many factors such as colors, texture, temperature, and sound play an important role in food sensation [Ahn et al., 2011], representing a group of sensations including tastes, and freshness or pungency. Therefore, the flavor compound (chemical) profile of the culinary ingredients is a natural starting point for a systematic search for principles that might underlie our choice of acceptable ingredient combinations.

Flavors derived from natural sources have shaped culinary habits throughout human history. Analogous to variations in regional languages, cultures have evolved variations in the way they cook. Traditional recipe compositions encode ingredient combinations that are not only palatable but appetizing. Heuristic associations between molecular properties and perception of flavors provide indications towards its chemical basis. For example, combinations of aliphatic esters play a major role in many fruit flavors. Ketones are known to impart metallic flavors in oxidized butter, and monoterpenoids provide the characteristic flavors of many herbs and spices. However, this knowledge remains largely unstructured and incomprehensive.

The idea of food pairing came from the intuition and studies western chefs conducted for exploring ways to cook tastier recipes. The central notion behind the 'Food Pairing Hypothesis' is that two ingredients taste well together when they share aromatic or flavor compounds between them [Blumenthal, 2008]. This hypothesis rests on the fact that the aroma/flavor of food is crucial in the way the food is perceived. The famous example of food pairing often cited is that of Chocolate and Caviar or Wine and Cheese. Chocolate and Caviar, one of Blumenthal's combinations, share the flavor compound trimethylamine, which gives a 'fishy' taste [Klepper, 2011].

Experiments in food pairing suggested that ingredients could now be swapped based on their proximity in flavor profiles. This notion paved the way for innovation of many new recipes in the western world and was especially popular with chefs who experimented with molecular gastronomy. The Food pairing hypothesis became an inspiration to several websites and blogs and scientific studies (*www.foodpairing.be, 'They Go Really Well Together', a blog by Martin Lersch, etc.*). Despite this, there are criticisms from food enthusiasts that applying food pairing solely to the innovation of a recipe does not always guarantee its success. Factors like the ratio of ingredient used, among others, also had to be considered. The balance of the flavors was the key [Klepper, 2011]. The method of preparation of the food also is a critical factor in the sensation of a tasty dish. Food Pairing Principle remains one of the means for new recipe innovations and, more importantly, to look for culinary patterns among various cuisines. The advent of molecular gastronomy and food pairing hypothesis has led many interdisciplinary researchers to tread into the world of cooking and cuisine [Ahn et al., 2011; L. R. Varshney et al., 2013].

2.4.1 Food Pairing in World Cuisines

The principle of food pairing has been scientifically studied for several world cuisines. The major work among them, Ahn et al. [Ahn et al., 2011] studied food pairing in recipes from North American and East Asian cuisines. The work illustrated a bipartite network of flavor compounds containing 381 ingredients used in recipes and 1,021 flavor compounds contained in these ingredients. This was the first of its kind in the study towards the computational investigation of flavor and ingredient pairing. The network enumerated flavors shared by different classes of food or food categories. The network showed interesting patterns in food pairing behavior in these cuisines. They found that in North American recipes, two ingredients are more likely to be combined together in a recipe if they share more flavor compounds between them. Albeit marginally, this was found to be not true for the recipes of East Asian cuisine, where more flavor sharing between the recipes resulted in them not being combined together in a recipe.

Food pairing in Medieval European cuisines was studied by Varshney et al. [K. R. Varshney, Varshney, Wang, and Myers, 2013]. They carried out food pairing analysis with two different datasets of flavor compounds available for ingredients in Medieval European recipes. They found a very strong positive food pairing in recipes of Medieval Europe using one set of data, whereas got a negative food pairing by using another dataset of flavor compounds. The paper contributed to the importance of data and how food pairing depends on the quality of the data on flavor compounds and the ingredients covered. Analysis of food pairing in Arabian Cuisines was carried out by Alrazgan et al. [Tallab and Alrazgan, 2016].

Sajadmanesh et al. [Sajadmanesh et al., 2016] found quantitative evidence of a strong correlation between nutrition information of the recipes and obesity. They demonstrated that deep learning could be used to effectively predicting cuisines from ingredients, potentially providing the possibility for fine-grained analysis of food and dishes as well as improved recipe recommendations based on individuals' profiles.

India has a unique blend of culturally and climatically diverse regional cuisines. Its culinary history dates back to the early Indus valley civilization. Indian culinary tradition dates back to thousands of years [Sen, 2015]. It is a multifaceted cuisine with a number of regional varieties, each one as different and unique as the other. These variations in food patterns across the country can be explained by its geographical and cultural diversity. Its long tradition of commerce and trade has had a lasting influence on Indian cuisine. While Indian cuisine has been studied from historical and cultural [Appadurai, 1988] perspectives, its molecular characteristics are hitherto unexplored. In this thesis, we undertake the study of food pairing in Indian cuisine which has not been investigated hitherto.

2.4.2 Repositories for Flavor Compounds

The study of flavor which is the decisive element in determining the palatability of food, is crucial for understanding the very nature of food. Flavor is manifested in the form of volatile compounds present in each ingredient, which gives them their characteristic flavor. The data on the flavor of ingredients is scattered across databases and scientific literature. There have been efforts towards compiling repositories of flavor compounds mainly with a commercial motive. A major move towards this direction is the Volatile Compounds in Food database, a commercial database of flavors compiled by Nutrition and Food Research Institute in Zeist, The Netherlands. Another comprehensive repository is a book, Fenaroli's Handbook of Flavor Ingredients [Burdock, 2010]. We use data from Fenaroli's as a source for the compilation of our flavor compounds data for the analysis of flavor pairing in Indian Cuisine along with the data already made available by Ahn et al. [Ahn et al., 2011]. This data set has 1,530 ingredients and 1,107 flavor compounds [K. R. Varshney et al., 2013]. Fenarolis has a more number of ingredients than Volatile Compounds in Food database but a much lesser number of flavor compounds. FooDB, another primary resource of flavor molecules, compiles molecules from food ingredients, although its

focus is not on the chemical basis of flavor or flavor pairing. Flavornet is another resource, which provides a list of flavor molecules and their odor profiles, but does not furnish information of their natural sources [Arn and Acree, 1998]. Other attempts in this direction have focused on compilation of data specific to aspects of flavors: tastes such as bitter (BitterDB) and sweet (SuperSweet), and volatile compounds of scents (SuperScent) [Ahmed et al., 2011; Dunkel et al., 2009; Wiener, Shudler, Levit, and Niv, 2012]. Certain others have targeted nutritional factors (NutriChem), polyphenols (Phenol-Explorer), and the medicinal value of food [Jensen, Panagiotou, and Kouskoumvekaki, 2015; Neveu et al., 2010; Rothwell et al., 2013; Scalbert et al., 2011]. Even with the presence of these repositories for flavors, there is a visible gap in the data and access to information on flavor compounds. In order to bridge this gap and bring the resources available for flavor molecules under one umbrella, we designed FalvorDB, a database of flavor molecules to quantify and characterize various aspects of flavor universe. It collates multi-dimensional aspects of flavors, including flavor profiles of ingredients and chemical features of flavor molecules.

2.5 ROLE OF FOOD IN HEALTH

Diet is the major source of nutrition for our bodies. Many recent studies pointed out the crucial role, diet play in the overall balance of our health. Diet in today's world is undergoing a rapid change by way of integration and assimilation of cultures and globalization of food markets. But a shift to a higher-fat, westernized diet has raised the obesity rate and the health risks that go with it. Many recent studies have pointed out that dietary ingredients can cause and prevent important diseases, including cancer, coronary heart disease, birth defects, and cataracts. Certain ingredient categories are evidently linked in literature with beneficial effects towards curing of diseases; for example, vegetables and fruits protect against these diseases; Similarly, dietary ingredients such as coffee, butter, sugar, etc. are more often associated with causing certain diseases. The traditional Indian medicinal system of Ayurveda considers diet as a primary means of treatment in its practices [Sen, 2015] and is deeply rooted in notions of disease prevention and promotion of health.

Chemicals in the diet have the potential to cause or cure diseases [Manach, Scalbert, Morand, Rémésy, and Jiménez, 2004; Mishra and Tiwari, 2011; Visioli, Borsani, and Galli, 2000]. They act upon the genetic makeup of the individual and cause changes that can positively or negatively affect health [Kaput and Rodriguez, 2004]. Dietary ingredients present in plants and plant-based ingredients called polyphenols are often studied for their health benefits in curing diseases [Grosso, 2018]. They have been implicated in their preventive effect on diseases. To sum up, dietary interventions can be used as a strategy in the prevention of diseases. One of the ways to do so is by systematically analyzing the health effects of dietary ingredients and providing meaningful culinary recommendations. More research is required to identify the active constituents in dietary ingredients as well.

2.5.1 Spices in Food—Antimicrobial Hypothesis

Spices have a unique place in Indian Cuisine. The history of the Indian economy is directly linked with the spice trade. Spices have been used in cuisines for thousands of years. The purpose of their use varied over regions and cultures. Some cultures may have used them for flavoring and coloring purposes, whereas others used them for the preservation of food [Higman, 2012] and religious and medicinal purposes [Sen, 2015]. The basis for the use of spices across cuisines has sought much attention [Billing and Sherman, 1998]. Among many studies that were done for understanding the reasons behind the prevalent culinary use of spices around the world, the study by Billing and Sherman is the most prominent [Sherman and Billing, 1999b].

Billing and Sherman [Sherman and Billing, 1999b] tested several hypotheses for explaining the use of spices. The main reason behind spice usage, the authors conclude, is that because they are good for us. Spices exhibit antioxidant and antimicrobial properties, which can

benefit health. This seems to be a plausible explanation, as many studies prove that spices indeed have antimicrobial properties. For quantifying their hypothesis, they focused on meat-based recipes. This was because non-vegetarian food spoils faster than vegetarian dishes in unrefrigerated conditions. They found that the majority of the meat-based recipes call for the use of spices. This result led them to postulate the reasons behind this phenomenon further, the chief reason being that spices exhibit antibacterial and antifungal activity. Their studies on spices found that most of the spices could inhibit the bacteria. Among the spices, garlic, onion, allspice, and oregano were found to be the most potent spices.

Another hypothesis the authors tested for the use of spices is that spices are used mostly in warmer climates as the food in warmer climates is more likely to be spoiled than those in cooler climates. They crosschecked this by comparing the mean temperatures of 36 countries and their recipes. They found that as average temperatures increased among countries, there were significant increases in the fraction of recipes that called for at least one spice, the mean numbers of spices per recipe, and the numbers of different spices used.

They also found that as the temperature increased, the mean fraction of recipes that called for the use of highly inhibitory spices also increased even though the correlation was insignificant for the less potent spices. The recipes from hotter climates were found to be more immune against the microbial attack than recipes from lower temperatures. Also, cuisines form lower temperature zones should contain fewer and less potent spices in their recipes than countries having higher temperatures. Subsequently, the quantity of spices used should be sufficient to protect against the antimicrobials found in the food in that region. Also, the cooking method shouldn't destroy the potency of the spice.

Billing and Sherman also postulated a null hypothesis that spice use might not provide any benefits. This is to say that spices should be highly palatable and patterns in spice use should correspond to local availability of spice plants. This claim does not get enough scientific support. The study points out that antimicrobial activity of the spice is still relevant today even if refrigeration methods are extensively used. Foodborne illnesses tend to be lower in countries with lesser spice usage. Overall, the usage of spices are seen to be based on their antimicrobial activities

Why are spices so popularly used in food in spite of them being potentially toxic and costly? [Krebs, 2009]. Krebs' studies how genetic variation in taste may interact with food traditions and ecology. He explores the genetic variations in taste sensitivity between populations linked to the tradition of adding spices to food. The main advantages of the spices identified by this study are that they contain important micronutrients and that they have antimicrobial properties. Spice usage has also been linked to the curing of diseases and immuno-protection. The main argument for the usage of spices in connection with the ecology is that they are antimicrobial. Studies till now haven't provided a conclusive rationale for the reason behind the widespread use of spices across cuisines. Still, they provide a strong takeaway that along with providing strong flavors, it might be due to their widely known health benefits, which make them such popular ingredients in the diet.

2.5.2 Spices and their Health Benefits

Spices are a quintessential feature of the Indian Cuisine. Apart from their use as flavoring agents, spices have been used in many traditional medicinal systems as part of their medicinal preparations [Thatte and Dahanukar, 1986]. Compounds from spices have been isolated and are used in various therapeutic agents used for curing ailments. Our studies on Indian Cuisine has provided the insight that spices form the molecular basis of recipes in Indian cuisine. The fact that they are used across cuisines even in small quantities raises the importance of further investigation of their health effects.

Herbs and Spices have been found to possess the ability to alter our physiological functioning [M.T.Lis-Balchin, 2004]. Many modern medicines also rely on natural elements from spices and herbs. Recent scientific studies have identified the purpose of spices other than coloring and flavoring of agents in food [Krishnapura Srinivasan, 2005a]. These studies have evaluated various antimicrobial effects and other medicinal properties. The ancient traditional medicinal system of Ayurveda used spice mixes in many of its medicinal preparations [Johri and Zutshi, 1992].

Scientific investigations to the health impacts of spices and herbs have given rise to a large body of biomedical literature. A large number of scientific literature has been published on the antimicrobial properties of spices [Arora and Kaur, 1999; Ceylan and Fung, 2004; De, Krishna De, and Banerjee, 1999; Skrinjar and Nemet, 2009]. Apart from their antimicrobial properties, spices have been shown to have to possess therapeutic potential for their hypolipidemic [Krishnapura Srinivasan, 2005b], anti-diabetic [K. Srinivasan, 2005], anti-lithogenic [Krishnapura Srinivasan, 2017], antioxidant [Yashin, Yashin, Xia, and Nemzer, 2017], anti-inflammatory and anticarcinogenic [Kaefer and Milner, 2008] activity. There is a serious dearth of comprehensive knowledge on the overall health effects of spices and herbs. These studies often present their results based on the health effects of a single spice or a group of spices targeting a specific disease. The need to put together this vast repertoire of scientific literature is imminent.

2.6 BIOMEDICAL TEXT MINING

The rise of research in the field of biomedicine has, in turn, given rise to a multitude of research articles published in the domain of biomedical text mining [Manning et al., 2012]. Text mining is a technique extensively used in the biomedical domain to make sense of the huge amount of data. One of its applications links it directly with Translational Bioinformatics, which relates basic biomedical research to clinical practice. As the amount of biomedical literature is increasing at a rapid rate, it has become increasingly difficult to curate the resulting literature manually. The existing tools are not suited for biomedical applications [Aggarwal and Zhai, 2012]. There is a need for curated databases that could provide automated extraction from the literature. Another wide application for text mining is in finding out the correlation between biological entities like genes, diseases and drugs. It consumes enormous time and effort to find these correlations through experimental methods. Therefore, automated prediction of good candidate genes [Ozgür, Vu, Erkan, and Radev, 2008] before experimentally validating them will save time and much effort. Comprehensive prioritization of candidate genes prior to experimental testing drastically reduces the associated costs [Bromberg, 2013]. Text mining of disease-causing genes is based on the evidence that a given gene is correlated to a given disease through a suggested causal link. Another potential use of text mining is better phenotyping [Cohen and Hunter, 2013]. Experiments have shown that strict phenotyping can improve the ability to find disease genes. Biomedical text mining can draw out the implicit and explicit associations between biological entities in the text.

The meaning and grammar of biomedical texts are complex and intertwined. Therefore natural language processing and grammatical analysis techniques need to be applied for preprocessing and grammatical analysis. These tasks involve Named Entity Recognition, relation extraction, event extraction, summarization and question answering [Aggarwal and Zhai, 2012]. The literature data mainly present in the form of journal articles and abstracts, clinical notes, etc. contain ambiguous texts, synonyms of disease and gene names, etc. need to be annotated for analyzing them. Complexity may arise when analyzing gene names that contain symbols and numbers. Tokenization, parts of- speech tagging, parsing, etc. are some of the Natural Language Processing (NLP) techniques used for cleaning up the data.

The first step towards text mining of biomedical literature is to automatically extract the documents through information retrieval. As the size of the biomedical data available is enormous, it is necessary to reduce the number of texts through information retrieval. The later step will involve document classification followed by Named Entity Recognition, document classification, summarization and information retrieval. The task of mining the relationship between two biological entities such as genes and diseases require identification of these entities, definition of several entity types needed for mining the literature for protein interactions (protein/gene names, chemical compounds, cell lines, etc.) and then the automatic aggregation of terms extracted from curated resources. Among other approaches, [Ozgür et al., 2008] proposes an approach based on integrating automatic text mining and network analysis methods to find out the correlation between a gene and disease. Much research has been centered on discovering tools for better annotation and preprocessing of the textual data.

2.6.1 Relations Extraction from Biomedical Texts

Biomedical texts often describe relationships between entities in the topic the article deals with. Identifying the relations between entities in a biomedical text can reveal a myriad of opportunities for scientific discovery. Some of the relationships in the domain of biomedicine include genes and diseases, drugs and genes, or between two entities of the same class (protein-protein, genes-gene). The process of relationship extraction in biomedicine is a two-way process. First, the relevant entities are identified, followed by the second step of determining or classification of the relationship [Cohen et al., 2012]. Relationship extraction approaches can be simple that rely solely on the co-occurrence of entities or can be complex systems using syntactic analysis and dependency parsing. A commonly used method to establish relationships between biomedical concepts from literature is co-occurrence. Apart from its use in knowledge retrieval, the co-occurrence method is also well-suited to discover new, hidden relationships between biomedical concepts following a simple ABC-principle, in which A and C have no direct relationship, but are connected via shared B-intermediates [Frijters et al., 2010]**.** In this study, we deal with the identification and classification of binary relationships in biomedical texts.

Relationship extraction systems can be broadly classified into supervised, unsupervised or semi-supervised methods [Cohen et al., 2012]. Supervised methods include providing explicit features identified from the data and training the model to retrieve relationships based on the given features. Artificial Neural Networks have been recently employed for relationship extraction in biomedical articles. Among them, some have developed Convolutional Neural Network models for sentence classification and relation extraction [Kumar Sahu, Anand, Oruganty, and Gattu, 2016; Mikolov, Corrado, Chen, and Dean, 2013; Nguyen and Grishman, 2015]. We follow a machine learning-based CNN model for relationship extraction of spices-disease associations.

2.6.2 Integration of Diet—Disease Associations from Biomedical Literature

Knowledge discovery through biomedical literature began with Swanson's paper connecting Fish oil and Reynolds syndrome [Swanson, 1986]. Since then, more studies have used different methods to unearth associations between foods and diseases. There have been attempts to extract disease associations for spices and herbs. Srinivasan et al. [P. Srinivasan and Libbus, 2004] conducted a text mining analysis for linking dietary substances and diseases with the case study of curcumin. Their text mining analysis suggested that curcumin has beneficial effects in retinal and Crohn's diseases. HerDing [Choi et al., 2016] is one such resource that uses speculative analysis for linking herbs and spices with diseases. Herding is a search engine that retrieves herb related information and acts as a guide for connecting herbs with diseases through an indirect text mining approach. The study is speculative and does not offer concrete evidence for an association between herbs and diseases. Nutrichem [Jensen et al., 2015], another resource used text mining approach to find associations in the literature between plant-based foods, phytochemicals present in them and diseases. The paper proposes an information extraction

system that aims to construct quantitative networks to capture the complex relationships reported among foods, chemical nutrients, diseases, proteins, and genes. The above repertoire presents an incomplete picture of diet-disease relations. In this thesis, we build a text-mining framework for food-disease associations, specifically focussed on culinary herbs and spices.

…