Understanding the perception and requirements of a plant-based nutrition app for cancer patients

Shekinah Dalayya  
University of Nottingham Malaysia

Salma Tamer Fathy Ahmed Elsaid  
University of Nottingham Malaysia

Kher Hui Ng  
University of Nottingham Malaysia

Tien Lii Song  
University of Nottingham Malaysia

Joanne Bee Yin Lim  
University of Nottingham Malaysia

Research Article

Keywords: plant-based, nutrition, cancer, web-scraping, design requirements, mobile application

Posted Date: August 18th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1966497/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Understanding human perception and requirements on nutrition for cancer prevention and condition management is important so that nutrition applications can be catered for cancer patients. In this paper, web-scraping was conducted to understand the public's perception, attitude and requirements related to a plant-based diet as a recommended diet for cancer prevention and condition management. Text and sentiment analysis were carried out on results gathered from 73 social sites to determine whether non-cancer and cancer patients use plant-based diets, how they have been consumed, their benefits in the prevention and condition management of cancers, the existing myths/fake news about cancer and what do cancer patients need in a nutrition app. Results of the text analysis highlight missing gaps in existing apps to include a lack of credibility and endorsement by professionals. Future nutrition apps should provide personalized diet, symptoms management, good user experience, credibility, and emotional and mental health support.

1 Introduction

Cancer is an important cause of morbidity and mortality worldwide where 18.1 million new cases of cancer and 9.6 million deaths from cancer were estimated in 2018. The cumulative risk of incidence indicates that 1 in 8 men and 1 in 10 women will develop the disease in a lifetime [1]. Long-term cancer survivors represent a sizeable portion of the population. Plant-based foods may enhance the prevention of cancer-related outcomes in these patients [2]. Plant-based diets are a diverse family of dietary patterns defined as infrequent consumption of animal foods along with frequent intake of plant-based foods in the usual diet. A high fruit and vegetable diet reduces the risk of cancer in the mouth, esophagus, lung, stomach, colon and rectum, evidence of probable risk reduction in cancer of the larynx, pancreas, breast, and bladder [3]. By the World Cancer Research Fund (WCRF) cancer prevention recommendations, 30–50% of all cancer cases are preventable by following a healthy diet and lifestyle [4]. Although these facts are widely known, only a fraction of the population follows the WCRF recommendations. Urgent action is needed to promote healthy plant-based foods in dietary guidelines to effectively reduce the risk of cancer.

Most people are now using smartphones and these devices, coupled with embedded sensors and modern communication technologies, make them an attractive technology for enabling monitoring of an individual’s health [5]. To monitor the alcohol consumption of college students, a study was conducted using smartphones and wearables which concluded by providing insights into mobile technology [6]. Mobile health (mHealth) applications (apps) have gained popularity in intervention for health behavior change [7]. Research into developing apps aimed at modifying key lifestyle behaviors associated with chronic diseases and other health issues have yielded positive findings [8]. For example, system monitoring during chemotherapy via an app has been found to lengthen survival – 75% of patients using the app were still alive, compared to 49% of non-users [9]. Reviews conducted by [10] and [11] provide comprehensive evidence that app-based mobile interventions are effective and highly promising for changing nutrition behaviors and nutrition-related health outcomes. However, there is a lack of evidence base and medical professional involvement in the development of current health apps. For example, [12] examined 185 health apps that focused on breast cancer awareness and found that only 13% were developed with professional medical input.
Despite the number of mHealth apps to recommend diets for weight control and general health management, there is still none that have attempted to recommend plant-based diets for cancer patients. To design such an app, there is a need to understand public perception about plant-based diet and nutrition apps. To do so, 73 links were web scraped to understand public perception of what they want and think about the different aspects of a plant-based diet. Web scraping approach allows an immense amount of information found on the internet to be compiled and analyzed [13], and to help aid cancer patients [14].

In this paper, we have examined many online forums from Malaysia and Singapore to address the following five research questions:

- RQ1. Do cancer patients use plant-based diet and is it effective?
- RQ2. What are the types, volume, frequency, and methods of cooking (where applicable) of fruits and vegetables used for prevention and condition management of cancer?
- RQ3. How does eating a plant help with a particular organ? How does it prevent or/and help during and after the different stages of cancer?
- RQ4. What are the existing myths/fake news about cancer?
- RQ5. What are the requirements of a nutrition app for cancer patients?

To facilitate answering the last question, three sub-questions emerged:

- RQ5.1. What do cancer patients need in a nutrition app?
- RQ5.2. What are the missing features in the current cancer apps they might be using?
- RQ5.3. How do cancer patients/caretakers search for nutrition information/recipes online?

The rest of the paper is organized as follows, with the next section detailing related work. The following sections outline the methodology used and findings of public’s perception, attitude and requirements related to plant-based diets, followed by discussions. This paper contributes to providing greater insights and understanding on how a nutrition app can be designed to help improve cancer patients’ diet and outcomes.

1.1 The role of plant-based diet in cancer prevention

Plant-based diet has been shown to protect against the 15 leading causes of death in the world, including many cancers, and may offer benefits as nutrition interventions to improve the management and treatment of these conditions [15]. Although the role of diet and lifestyle factors in health and disease is gaining more attention and emphasis, the benefits are still underestimated and undervalued. Common cited reasons for not eating fruits and vegetables are a lack of knowledge (e.g., not knowing how to cook them) and dislike for its texture, smell, and taste [16]. However, the study found that people are willing to try them when they are taught about their health benefits or how to prepare/eat them. In addition, the kind of plant-based diet and amount of intake depend on various factors such as whether a person is undergoing chemotherapy, the stage of cancer and his/her age, gender, sex, and other psychosocial factors. It is also important to ascertain if the type of nutrition diet is being recommended as prevention or to manage a particular stage of cancer. Unfortunately, there is still a lack of mHealth app that can provide detailed dietary guidance or recommendations of fruits and vegetables as prevention and condition management for cancer.
Finding of a recent study indicates that greater adherence to plant-based diet index (PDI) is inversely associated with risk of breast cancer [17]. A plant-based diet is also valuable in the primary and secondary prevention of colorectal cancer where epidemiological studies show a 46% − 88% reduced risk of colorectal cancer for those following a plant-based diet [18]. A study involving a good representation of an ethnically diverse population (including both men and women of Asian, American Indian, Black, and Caucasian ethnicities of different ages, smoker/non-smokers, consumer, and non-consumers of alcohol) reveals that a lower consumption of vegetables, fruits, fiber, and whole grains is associated with higher pancreatic cancer risk [19].

The National Cancer Society Malaysia, for example, provides useful nutrition tips for people living with cancer to be taken during cancer treatment and recovery [20]. However, the types of food, volume, and frequency change during and after treatment, and thus a personalized diet is essential. This is to ensure that users will be able to adapt to what is recommended according to their body’s changing nutritional needs. Through a cognitive approach by understanding the patients’ needs, values, and psychosocial factors involved in nutritional behavior and food-related decisions alongside other variables (sex, age, race), researchers have found that it is possible to achieve important clinical targets, to develop a personalized approach and to support concrete actions towards healthier diets thus preventing recurrences, monitoring chronic conditions, and supporting a good quality of life [21].

1.2 The impact of mobile health applications for cancer

The appeal of smartphones for assistance in health promotion concurs with the trend that more people are seeking health information via mobile devices [22]. Researchers like Wang et al. [23] have advocated for smartphone interventions for long-term health management of chronic diseases. In this context, apps provide the opportunity to bring behavioral interventions into real-life situations where people make decisions about their health [24]. A study done by Viitala et al. [25] showed that patients’ sense of security and freedom increased after using mHealth apps targeted for cancer. Research has shown that compared to those without mHealth apps, individuals with mHealth apps have significantly higher odds of using their smart devices to track progress on a health-related goal to make a health-related decision and in health-related discussions with care providers. Middelweerd et al. [27] highlights that smartphone users value health behavior apps that require low effort, are pleasant to use, are developed by credential experts, enable self-monitoring, provide advice on how to change (dietary) behavior, include positively framed alerts/reminders, provide accurate tracking functions, incorporate adequate privacy settings, and clearly show what the app will do. These factors need to be taken into consideration to improve the engagement and retention of the user [28].

The use of mobile health apps to provide help with nutrition has yielded positive findings [10, 11]. There are nutrition apps catered for cancer, such as OncoFood to help patients track their daily dietary habits [29]. However, there are limitations to the app such as patients having to take too much time to input data and many of them wish for recipe suggestions as well as to be able to make changes to existing and past data on food and prepared meals. Another study by Keaver et al. [30] reviews the quality, nutrition content and behavior techniques of 1149 apps aimed at those with cancer, but after two rounds of screening, only 12 apps were identified. There was a lack of strategies for implementation and a lack of indication on whether the information available is catered for specific cancer types or at specific stages of cancer or treatment. Out of the 12 apps, 6 apps were also providing non-evidence-based information. This study concludes that there is little nutrition information that is currently available on publicly available apps for cancer. However, only 3% of those
apps have had their content developed or evaluated by health providers leaving behind the question of whether those apps are reliable or not. The challenge is in developing apps that are appropriate for health tracking, monitoring, and interventions using evidence-based strategies. In addition, there is a lack of understanding of how wearable or smartphone sensors can be used for personalized diet management and interventions [31]. Research done by Cai et al. [32] further emphasize on the need of having patients, nurses, and healthcare professionals to collaborate in the design of a mHealth app. Uncertainty is a common factor in general healthcare and knowing how to navigate it is quite useful in any health technology [33]. In Malaysia, a cancer dietary app was developed to provide healthy eating guide (advice from healthy eating to eating problems, weight loss prevention and increasing protein intake) which are uniquely tailored to the local food choices, preferences, and ingredients [34].

1.3 The use of web scraping and text analysis to identify requirements

Web scraping is a method used to extract a huge amount of information automatically from websites. It is also known as screen scraping, web data scraping or web harvesting [35]. Web scraping allows the immense amount of information found on the internet to be compiled and analyzed to make sense of what is happening in a short amount of time [13]. Currently, web scraping has mostly been used to research about food prices [36] and to extract recipes [37]. Previously, to understand the perception towards diet and food, surveys, records, 24-hour recall and questionnaires were most used [38]. The use of wearable cameras was previously implemented to understand food consumption life cycle [16]. So far, web scraping has not been used as a digital ethnography method to find, analyze, and understand human perception on food, diet, and cancer.

A lot of people, including cancer patients, have been using social media to express how they feel, what they are going through and share with each other their own experience. For example, a breast cancer patient, Lisa Bonchek Adams, used twitter and tweeted over 176,000 times to talk about her own cancer experience [39]. Shaiket et al. [40] highlights that data analysis has been done previously to conduct an online diagnosis of diabetes with twitter data, to find out about the average happiness of cancer patients by using patient tweets, to conduct a sentiment analysis on breast cancer screening as well as many others.

To design an app for cancer patients, it is important to know what the cancer patients want and need in an app. It is common to use focus groups to find out more about a particular cancer [41, 42]. However, the use of text analysis to identify requirements for the design and development of cancer apps is still rare and little understood. While a plant-based diet is effective to improve outcomes for cancer patients, there is a lack of mHealth apps that provide reliable plant-based dietary information and recommendations for cancer patients. As a result, the public tends to just take in whatever information they have, which may not be true. As such, we conducted a text analysis to understand the perception that the public has on the above to identify requirements for a nutrition app to help improve cancer patients’ diet and outcomes.

Sentiment analysis, also known as opinion mining, is the process of automating information such as opinions, attitudes, emotions, and feelings. Sentiment analysis is usually applied to reviews and social media. It calculates the aggregate sentiment polarity and classifies the sentiment as positive, neutral, or negative [43] In sentiment analysis, results are represented in score for each term as follows: positive score ($s_i^+$), neutral score ($s_i^0$) and negative score ($s_i^-$). Each score is used to determine how that sentence is perceived [44]. Microsoft
Azure Machine Learning embodies cloud services and can be used to calculate the contribution score of the user based on the metrics and has about 100 techniques including regression, classification, text analysis and recommendation [45]. This study applied these methods to identify requirements for a plant-based nutrition app, particularly suitable for cancer patients.

2 Methods

2.1 Categorizing and Source Gathering

Before starting the process of requirement gathering, we mapped the research questions into 5 categories. Each category had a main question and sub-questions related to a certain topic. Category A focused on the reason for plant-based diet, category B on the types of plants used, category C on the association between plant-based diet and cancer, category D on myths about cancer and sources of cancer-related information, and finally, category E focused on the requirement for cancer prevention and condition management via a nutrition app. For each category, search keywords were identified to get information from online forums such as LowYat and Reddit and social networking sites such as Twitter, Facebook, YouTube, and Instagram. The keywords were manually typed onto those sites and relevant links were gathered to be used as sources of web scraping. This process resulted in a total of 73 links gathered. Table 1 shows the number of links gathered for each category.

<table>
<thead>
<tr>
<th>Category</th>
<th>Total Number of Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>30</td>
</tr>
<tr>
<td>B</td>
<td>15</td>
</tr>
<tr>
<td>C</td>
<td>5</td>
</tr>
<tr>
<td>D</td>
<td>15</td>
</tr>
<tr>
<td>E</td>
<td>8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>73</strong></td>
</tr>
</tbody>
</table>

2.2 Web Scraping

To extract information from the links gathered, the links were run through web scraping. Links from Facebook and Instagram were excluded from the scraping as they provided information in the form of pictures that could not be scraped by a text-based web scraper. YouTube videos whose comments were blocked or did not have any comments were also excluded. The Beautiful Soup library in Python was used to web scrape the links gathered. Multiple Python files were created, each for a platform, and coded to cater for the structure of each platform and to match the source code. The source code provides information as to where the needed information is present on the page such as under the paragraph HTML tag of <p>. The Beautiful Soup library is then used to return all instances of the tag, comprising textual information which is exported as Excel sheets. The results of web scraping provided more than 100 posts for each category, adding up to 3787 posts in total. Table 2 shows the entries provided by the links in each category.
Table 2  
Number of posts from the links in each category

<table>
<thead>
<tr>
<th>Category</th>
<th>Total Number of Posts</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2648</td>
</tr>
<tr>
<td>B</td>
<td>191</td>
</tr>
<tr>
<td>C</td>
<td>113</td>
</tr>
<tr>
<td>D</td>
<td>560</td>
</tr>
<tr>
<td>E</td>
<td>275</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>3787</strong></td>
</tr>
</tbody>
</table>

2.3 Text Analysis

Compared to the traditional means of requirement gathering such as interviewing and surveys, text analysis has several advantages. In this project, text analysis was used as a tool to gather requirements for a nutrition app that caters to a user's diet needs with a focus on cancer prevention and condition management.

Text analysis has made it easier to become aware of the public's opinions on what they want or do not want. For example, from the analysis, we discovered that the words in favor of plant-based diets do not carry a lot of weight compared to those not in favor showing that the public mainly prefers diets that are not plant-based. In this way, we can cater the application to contain diets that are not entirely plant-based but to mix plant-based diets with meat recipes to keep general users healthy. Text analysis can help identify trends amongst the public. For example, keywords such as ‘doctors, medical patients, and the NHS’ showed that there is a trend emphasizing reliable sources of information among the public.

Text analysis also allows for a wider audience to be reached. In contrast, interviews could involve only a limited number of people, but they give researchers the freedom to ask follow-up questions. So instead of getting keywords and having to infer what they mean, questions could be asked to have a better understanding of the interviewee's point of view. While as seen from keyword findings, especially in category C, text analysis may provide only surface-level answers without the opportunity to dive further into the keywords obtained.

Sentiment analysis has proven to be very useful and plays a huge role in understanding human perception. Sentiment analysis is being used for business, politics, disease outbreaks, sports, data security, and health care [46]. Sentiment analysis has helped in providing useful information for designers. Knowing how the public perceives a particular product allows the designer to know whether to venture deeper into that product or to completely stray away from it [47].

Sentiment analysis was also applied to data from the online vegan community where there are a lot of recipes and reviews. The purpose of that study was to find the sentiments from the reviews and comments. The results from this analysis were able to give an idea of which recipes vegans like the most [48]. Sentiment analysis has been used as part of a systematic review of health from online communities. From the results, five roles of authors were identified as well as demographic factors. Health-related problems and health care treatments were categorized and studied by the sentiment analysis [48]. Sentiment analysis has also been used to find out
about users’ food preferences. Food names from comments made by users were extracted and analyzed. This analysis showed a high level of precision in knowing the user’s preference [49]. In another study, twitter-messages were also explored and analyzed to determine the contents of tweets related to four eating situations – breakfast, lunch, dinner, and snack. This study provided a framework for the understanding of food intake and selection [50].

Before proceeding with text analysis, the records obtained from web scraping needed to go through a process of cleaning to remove unnecessary keywords, the time stamps and usernames associated with the posts. The unnecessary keywords included those that were used to search for the links such as Malaysia, cancer, vegetarian, Singapore. Those keywords were not needed as they describe the research broadly and do not cater to specific categories.

A code written in Python was used to remove null spaces and common stop words in the English language such as “of, for, at, to, you, a, i, the”. Stop words are words often used in the English language but do not provide significant meaning in terms of Natural Language Processing (NLP). Lemmatization was also performed to remove different forms of a word. It aimed to remove inflectional endings of a word to give its base form. For instance, when presented with the token (word) “eaten”, lemmatization would return either eat or ate depending on whether the word was used as a noun or verb [51]. The pandas, time, defaultdict, and spacy libraries were used for data handling, timing operations, word frequency, and pre-processing respectively.

The documents were then passed through a Term Frequency-Inverse Document Frequency (TF-IDF) calculator to identify the key 50 words or terms based on their relevance to each category and to generate their associated weight. The higher the occurrence of a word in a text document collection, the higher the TF while IDF is used to measure the importance of a term in the documents. The TF-IDF is the weight associated with a certain keyword, because of multiplying the TF and IDF obtained [52]. From the derived 50 keywords, the top 10 with the heaviest weights were chosen to represent each category. Choosing the top 10 keywords narrows down the results to provide a clear representation of the public’s sentiments to address our research questions. All documents obtained were combined and classified per category. Using the Add-in feature on Microsoft Excel, Microsoft Azure Machine Learning was applied to each category.

2.4 Analysis of Keywords

The final step in requirements gathering was analyzing the meaning of the keywords to understand the public's sentiment about the category they represent. The context behind the keywords was understood by scanning the documents for where they occurred and how they were used by users who mentioned those keywords.

2.5 Analysis of sentiments using Microsoft Azure Machine Learning

All documents in each category were consolidated in one document before it was analyzed. This process was done manually and allowed us to understand how the public reacts to a certain topic and their views towards it.

3 Results

3.1 Findings
The data gathered from text analysis was the public's opinions about plant-based diets and their correlation with cancer. The text analysis provided keywords that can be used to create an application that best suits their needs concerning cancer, especially those who are already diagnosed with it. The keywords are not based on scientific research but rather on the public's own beliefs. It is important to analyze those keywords as they can give information about what people might expect from an application that would cater to their health. The keywords can be used to meet the needs of users and their expectations.

For each of the categories identified, the top ten keywords that best answer the questions of the category were chosen to represent it in order of their weights. Words with a higher weight were considered first and if those words did not satisfy the category, text analysis was run again to get substitute words that are better suited for the category. Then, those words were gathered to represent each category and their total weight added to determine the importance and relevance of a category in comparison to others. Table 3 displays each category, its keywords, and its associated weight. The keywords are put in order of their respective weights.

Table 3
Top 10 keywords in each category along with the total weight

<table>
<thead>
<tr>
<th>Category Name</th>
<th>Keywords</th>
<th>Total Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Meat, good, need, animal, want, body, life, healthy, sugar, cure</td>
<td>0.129533</td>
</tr>
<tr>
<td>B</td>
<td>Protein, water, bread, food, alkaline, acidic, product, meat, oil, bean</td>
<td>0.216158</td>
</tr>
<tr>
<td>C</td>
<td>Breast, anal, colon, cervical, disease, treatment, clinic, vaccine, colorectal, life</td>
<td>0.295349</td>
</tr>
<tr>
<td>D</td>
<td>Eat, water, people, thread, food, dr, vitamin, disease, medical, patient</td>
<td>0.132288</td>
</tr>
<tr>
<td>E</td>
<td>Iphone, feedback, patient, news, free, risk, bookdoc, breast, help, research</td>
<td>0.53767</td>
</tr>
</tbody>
</table>

3.2 Analysis of Data Collected

The data obtained from each category was analyzed to evaluate whether they answer each question.

3.2.1 Category A (+): Meat, good, need, animal, want, body, life, healthy, sugar, cure

This analysis answers question 1: Do cancer patients use plant-based diet and is it effective?

This category gives insight as to why plant-based diet would be considered or not. The meat came in as the first word (with the heaviest weight) to show that many people would prefer not to go on a plant-based diet as they prefer meat in their meals. As such, meat is the main barrier to pursuing a plant-based diet. This is also indicated by the ‘animal’ keyword, showing that most people prefer an animal-based diet compared to plant-based one. Other keywords provided reasons for adopting a plant-based diet. People go on a plant-based diet as they believe it is good for health, out of a need for this diet due to illness such as cancer or simply a wish to be healthier. It is also believed that a plant-based diet is healthy, can prolong life, lead to a healthier body, and would have less sugar intake than other diets. A minority believe that a plant-based diet can be used to help cure cancer. The sentiment analysis reveals that 100% of responses for this category are all positive scores.
3.2.2 Category B (+): Protein, water, bread, food, alkaline, acidic, product, meat, oil, bean

The public perception question of “What are the types, volume, frequency, and methods of cooking (where applicable) of fruits and vegetables used for prevention and condition management of cancer?” can be seen in this analysis.

This category gives insight as to what kind of fruits and vegetables are used in plant-based diets and how they are used. From the keywords obtained, it can be concluded that most people prefer plants that have a lot of protein content and are alkaline or acidic with alkaline having a higher preference than acidic. Plant content in the form of beans is the most mentioned plant-based diet-related food among others. Water had the most weight of keywords that show how vegetables are cooked and eaten, whereas some people prefer cooking such food in the water. The keyword bread followed, showing that many people prefer eating vegetables with bread on the side. People also eat vegetables along with their regular meals or with meat dishes showing that they do not substitute meat for plants but rather eat them side by side. A minority of people have reported cooking vegetables with oil to suit their taste.

This result reinforces that the supply of animal-based protein in Malaysia has been increasing at a faster rate than vegetable-sourced protein. The total animal-based protein products have increased by 59.1% while vegetable protein increased only by 11.9% [53]. Not only is the frequency of daily consumption of both vegetables and fruits far lower than desired, but Malaysia has also been topping the league tables as the most obese nation in South East Asia since 2014 [16]. The strong evidence showing obesity increases the risk of several types of cancer including colorectal, breast, and prostate shows how impertinent it is that this research addresses this to provide better health outcomes for Malaysians [54].

Sentiment analysis revealed that 67% of the responses are positive, 18% are negative and 15% neutral. One of the underlying factors for the negative responses was the cost of treatment and/or hospital. These words are mentioned as our question mentioned “prevention and condition management of cancer” and while the results of this analysis do not answer our main question, it shows how patients are concerned about the cost.

3.2.3 Category C (+): Breast, anal, colon, cervical, disease, treatment, clinic, vaccine, colorectal, life

This category refers to the question of how eating a plant helps with a particular organ and how it prevents or/and helps during and after the different stages of cancer.

Breast, anal, colon, cervical, and colorectal cancer were the most mentioned type of cancers due to people’s belief that plant-based diets can help with those cancers the most. Plant-based diets can be used during the treatment of cancer to maintain health, can be used to battle other diseases as they boost immunity, and aid vaccines to maintain the immunity gained. Plant-based diets are also believed to elongate a person’s lifespan and give a higher chance of survival. Sentiment analysis of this category shows a positive score of 88.5–11.5% of a negative score. There was no exact relationship with the negative score sentences.
While this category surfaces in different cancers, it fails to go deep into answering the question for this category. This raises the following suggestions on what this could mean: 1) The public does not see how eating a certain plant help with a particular type of cancer, and 2) The public seems to view eating a plant-based diet as providing benefits, but they do not have any idea about how consumption of a plant-based diet affects someone in different stages of cancer. This can conclude that the scraping and text analysis does not provide enough information to be able to answer the question.

3.2.4 Category D (+): Eat, water, people, thread, food, dr, vitamin, disease, medical, patient

This category focused on identifying the existing myths or fake news about cancer. The analysis aimed to look through what are some of the news that the public knows and might consider as true.

From the keywords in this category, it can be understood what kind of myths are believed about cancer and how people verify the sources of their beliefs. ‘Eat’ comes in as the first keyword in the category due to the belief that what one eats can affect how vulnerable they are to cancer, which is also stressed by the keyword ‘food’. This is followed by the keyword ‘water’ with the belief that drinking a lot of water can help prevent cancer. It is also believed that taking vitamins regularly can improve immunity and reduce the risk of cancer.

Word of mouth is the most common source through which the public hears news or beliefs about cancer. Internet threads are a common source of receiving or correcting facts followed by doctors, medical articles, and cancer patients. This information is useful for deciding how to present information to users on a nutrition app based on sources they would trust.

Sentiment analysis also shows a positive score of 74.3%, a neutral score of 11.7%, and a negative score of 14%. Most of the sentences were negatively scored because of the use of the word “cancer”, for example, cancer being incurable or losing someone to cancer.

3.2.5 Category E (+): iPhone, feedback, patient, news, free, risk, bookdoc, breast, help, research

This last category focuses on identifying requirements to support app development for cancer patients and refers to the questions:

1. What do cancer patients need in an app?
2. What are the missing features in the current cancer apps they might be using?
3. How do cancer patients/caretakers search for information/recipes online?

From this category, it can be understood what expectations potential users would have for an application that caters to cancer care. The most prominent keyword on the list is ‘iPhone’ showing that most of the cancer applications launched are for iOS compared to Android. The application should have an option to let users provide feedback regarding its function. Users would expect the application to help cancer patients on their journey, to provide news and research information on cancer, and be free. The application should also provide help for high-risk individuals or patients. Amongst the several types of cancer, breast cancer was the most mentioned, showing that the applications already on the market mainly cater to breast cancer patients. Finally,
BookDoc, an application that connects patients to healthcare individuals was mentioned several times as an application that could help cancer patients. BookDoc’s features can be used as an indication of what makes a healthcare application successful. BookDoc allows teleconsultation 24/7, allowing users to be in communication with a doctor without the need to travel. Other features include searching and booking the user’s preferred healthcare professional, providing wellness programs being a platform for users to buy products, and even providing nutritional and dietary advice from professionals [55].

Sentiment analysis shows 76.4% of the positive score for this category as well, compared to a score of 5.5% for negative and 18.1% for neutral.

4 Discussion

The keywords discovered through text analysis identified that many people in Malaysia and Singapore are not willing to completely give up their current diets to pursue plant-based ones due to a preference for meat in their meals. Those who pursue a plant-based diet mainly do so as a method of taking better care of their health and not as a personal preference or an attraction to the diet. In general, the reaction toward a plant-based diet was negative.

When suggesting methods to battle cancer, most of the opinions suggested methods such as drinking lots of water and eating vitamins as opposed to providing recipes for plant-based meals. Users would prefer recipes that involve plants with other food items such as meat rather than meals that solely focus on fruits and vegetables. Therefore, when designing a nutrition app, recipes that involve both meat and plant components together would have to be considered.

Results of findings showed that users of nutrition applications that focus on prevention and condition management of cancers are not only cancer patients but users who will use those applications to maintain their health and prevent cancer. As such future nutrition apps will need to be designed in a way that aids cancer patients on their road to recovery and considers users who are conscious of their health. Users are also most likely to trust information coming from well-known health associations such as Novartis and the National Health Service (NHS) in England. Therefore, the information provided by any nutrition apps will need to be concise, accurate, and trustworthy sources to gain user trust.

Results of text analysis identified the categories with heavier weights being the concerns that the public had as they were spoken about the most. The category with the most weight was category C showing that people were more interested in finding out what kind of plants are associated with which type of cancer and how certain plants can be used to avoid certain types of cancer. Therefore, when designing the application, users should be provided with information about how certain plants are beneficial in battling cancer or preventing it. The category with lesser weight after category C was B, followed by D, and then finally, category A. The focus of a nutrition app for cancer patients should be on providing users with information about plants and their correlation with cancer followed by how to use plants in recipes, using reliable sources for information if users would like to know more about cancer, and finally, showing how effective plant-based diets are in battling cancer. Results from category E provided suggestions as to what features cancer patients would like to have in a nutrition app:
To help cancer patients on their journey
To provide news and research on cancer
To provide help for high-risk individuals or patients
To incorporate more plant-based ingredients in their diet
To get emotional and mental help

Regarded as a combination therapy, apps such as eCoStudy (for ovarian cancer), Moovcare (lung cancer), SkinVision and Firstcheck (skin cancer), and CancerCare (pediatric cancer) were created with varying objectives (though none to prescribe personalized diet recommendations), including to offer patients community support via an online network; to reduce stress and aid the healing process via meditative music and art; to serve as a calendar to help schedule appointments; used as an online health journal that could improve their treatment by tracking medications and blood counts; to provide educational videos for pain management; to identify suitable physical activities for patients; and to act as a handy Dictaphone to record answers from doctors and nurses, equipped with medical jargons. Most of these apps, however, do not have recommendations of recipes, specifications of what to eat or what not to eat depending on the cancer type, cancer stage, and treatment currently being done by the patient that is based on reliable sources of information confirmed or endorsed by a health professional.

4.1 Personalized diet

From the results of web-scraping, patients want to have an app that helps them in their diet. However, the app will need to provide patients with a personalized diet as patients might have different food preferences, and allergies and may need different types of food elements that can help them at different stages of cancer. While patients are also going through different types of cancer therapy, the food they consume can affect them differently. Therapies can also affect the patient’s sense of smell, taste, appetite, gastric capacity, or nutrient absorption [56].

4.2 Symptoms management

Another result of the web scraping was the concern about recurrence after the first cancer treatment. Different treatments also affect patients differently both in positive and negative ways. To take care of cancer patients, symptom management is an important part of it [57] and will need to be taken into consideration in the design of an app.

4.3 User experience

User experience is a factor that needs to be considered when developing the app. In designing this app, there was a need to find out the current public perception, which is, to be able to identify the problems that users might face using current apps and to ultimately provide a design that will be able to provide a solution [58]. From the research done, the 3 main criteria for users to use a nutrition app are ease of use, free cost, and ability to produce automatic readings of caloric content. A nutrition app with a good user experience should aim for at least these 3 factors.

4.4 Credibility
Web scraping and background research show that most apps lack credibility or the input of professionals. Having accurate medical information is an important aspect of credibility. Web-based medical information is viewed skeptically as they are known to be misleading and inaccurate. Designing the app with a health professional will aid in making sure that the information available on the app are accurate and not misleading [59].

### 4.5 Emotional and mental health support

The results from the web scraping and text analysis show that there is a need for patients to get emotional and mental help. There is a lack of mental health support in the cancer community [60, 61]. To solve this gap, there are different guidelines on how to help support a patient in these times, which include helping them with personal health goals, coping skills, healthy sleeping, and relaxation methods among others [62].

### 4.6 Limitations

Gathering sources to be web-scraped from was the first step in this process and provided a few limitations. Plant-based diet and even diet in general differs region to region due to availability of fruits and vegetables, environment, culture and many other factors [63]. Having that in mind, there was a need to be specific in the regions we chose to do web scraping from as the project aimed to create a nutrition app for cancer patients in the region of Malaysia and Singapore. We also had only a few forums and social media to choose from, having to exclude information found in videos and images. With forums, information could be repetitive, and some are meant to be ads. The same person or sometimes even, different people have the same text all over that source.

Although the documents were cleaned at the beginning to remove stop words and unnecessary keywords, as they contained a large amount of data, it was difficult to go through them each to remove all unwanted keywords which resulted in the repetition of the cleaning process. Keywords extracted through TF-IDF were not always used in the same context by all users. For example, one user could mention the keyword “meat” to indicate they are in favor of meat-based diets while another could mention it to indicate they are against such diets. The majority opinion would be considered as the main context behind a certain keyword.

While this paper focused on the use of ID-TRF and Microsoft Azure Machine learning for analysis, various techniques can also be used for text analysis. Behaviour Change Techniques Taxonomy version 1 (BCTTv1) has been used on some fitness and nutrition mobile apps as well as sentiment analysis and user feedback were collected through online reviews [64]. Other text mining techniques have been used as well as a big data analysis tool for food science and nutrition. These include word association analysis, text classification, text clustering, and text modelling [65].

### 5 Conclusion

With the increase in the use of apps and those dedicated to health, there are some missing gaps when it concerns mhealth apps for cancer patients and the use of a plant-based diet to help cancer patients in symptom management. This paper aimed to analyze results from social media in finding human perception and sentiment on plant-based diet and cancer. 73 links were web-scraped using Beautiful soup with python, TF-IDF, and Microsoft Azure Machine Learning for text analysis. The results from the text analysis identify how the lack of reliable information, the inclination toward a meat diet compared to plant-based diets, and the lack of
mental support. After web-scraping 73 links and analyzing these data, some design factors were identified to bridge the missing gaps in a mHealth app. Those designs include the need of a personalized diet, symptoms management, user experience, credibility as well as emotional and mental health support.

**Declarations**

The authors have no conflicts of interest to declare.

**Funding**

The research has been supported by the Fundamental Research Grant Scheme (FRGS) through the project: mHealth App: Prevention and Management of Cancer via an AI-Integrated Mobile Application to Recommend Plant-Based Diets by the Ministry of Higher Education (MOHE), Malaysia.

**References**


44. Farhadloo, M. & Rolland, E. (2013). Multi-Class Sentiment Analysis with Clustering and Score Representation. In IEEE 13th International Conference on Data Mining Workshops, 904–912. https://doi.org/10.1109/ICDMW.2013.63


