

## Article

# Musawah: A Data-Driven AI Approach and Tool to Co-Create Healthcare Services with a Case Study on Cancer Disease in Saudi Arabia

Nala Alahmari <sup>1</sup> , Sarah Alswedani <sup>1</sup>, Ahmed Alzahrani <sup>1</sup>, Iyad Katib <sup>1</sup>, Aiiad Albeshri <sup>1</sup>   
and Rashid Mehmood <sup>2,\*</sup> 

<sup>1</sup> Department of Computer Science, FCIT, King Abdulaziz University, Jeddah 21589, Saudi Arabia; nsaalahmari0001@stu.kau.edu.sa (N.A.); sabdullahalswedani@stu.kau.edu.sa (S.A.); asalzahrani@kau.edu.sa (A.A.); iakatib@kau.edu.sa (I.K.); AAAAlbeshri@kau.edu.sa (A.A.)

<sup>2</sup> High-Performance Computing Center, King Abdulaziz University, Jeddah 21589, Saudi Arabia

\* Correspondence: RMehmood@kau.edu.sa; Tel.: +966-537069318

**Abstract:** The sustainability of human existence is in dire danger and this threat applies to our environment, societies, and economies. Smartization of cities and societies has the potential to unite individuals and nations towards sustainability as it requires engaging with our environments, analyzing them, and making sustainable decisions regulated by triple bottom line (TBL). Poor healthcare systems affect individuals, societies, the planet, and economies. This paper proposes a data-driven artificial intelligence (AI) based approach called Musawah to automatically discover healthcare services that can be developed or co-created by various stakeholders using social media analysis. The case study focuses on cancer disease in Saudi Arabia using Twitter data in the Arabic language. Specifically, we discover 17 services using machine learning from Twitter data using the Latent Dirichlet Allocation algorithm (LDA) and group them into five macro-services, namely, Prevention, Treatment, Psychological Support, Socioeconomic Sustainability, and Information Availability. Subsequently, we show the possibility of finding additional services by employing a topical search over the dataset and have discovered 42 additional services. We developed a software tool from scratch for this work that implements a complete machine learning pipeline using a dataset containing over 1.35 million tweets we curated during September–November 2021. Open service and value healthcare systems based on freely available information can revolutionize healthcare in manners similar to the open-source revolution by using information made available by the public, the government, third and fourth sectors, or others, allowing new forms of preventions, cures, treatments, and support structures.

**Keywords:** machine learning; big data analytics; social media; Twitter; smart healthcare; cancer; Arabic language; Latent Dirichlet Allocation (LDA); topic modeling; Natural Language Processing (NLP); smart cities



**Citation:** Alahmari, N.; Alswedani, S.; Alzahrani, A.; Katib, I.; Albeshri, A.; Mehmood, R. Musawah: A Data-Driven AI Approach and Tool to Co-Create Healthcare Services with a Case Study on Cancer Disease in Saudi Arabia. *Sustainability* **2022**, *14*, 3313. <https://doi.org/10.3390/su14063313>

Academic Editor: Petra Poulouva

Received: 22 February 2022

Accepted: 5 March 2022

Published: 11 March 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The sustainability of human existence—our existence—is in dire danger and this threat applies to our environment, societies, and economies. The threats to the environment are evident in the deteriorating planet's conditions. The threats to our societies are apparent in the deteriorating physical and psychological health of individuals and groups; increasing frequency and magnitudes of conflicts, riots, and wars, the rise of materialism, and the deteriorating love, harmony, and sincerity between people. The risks to our economies are manifested in the changing foci from the net benefits that the wealth brings to society, to mere numbers counting products, earnings, spending, sales, exports, and GDPs. Nationalisms, racisms, gender, and other wars are on the rise and seen as important in the name of equality, freedom, and missions, and little thought is given that any group is comprised of

individuals, and each individual is a fellow human who is juggling with many difficulties like anyone else.

If done right, an umbrella term that has the potential to unite individuals towards sustainability is smartization of cities, and societies that involve the transformation of our traditional environments into smarter ones [1,2]. Smartness is defined by its core objective of the triple bottom line (TBL)—i.e., social, environmental, and economic sustainability [3] and can be achieved by engaging with our environments, analyzing them, and making sustainable decisions regulated by triple bottom line [4,5]. Note that the terms quadruple bottom line and quintuple bottom line have also been used in the literature with additional emphasis on ethics, equity, and purpose (soul, spirituality, or culture). For simplicity, we take the view that equity, efficiencies, ethics, purpose, innovation, etc. are included in TBL, the three dimensions of sustainability, however, we understand that placing in various statements an emphasis on specific aspects such as equity and ethics can be beneficial. The requirements for and the definitions of smartness have evolved over time with a focus in the early days on the digital aspects of the systems, shifting gradually to incorporate efficiency, equity, and triple bottom line in it [6].

Human health and healthcare are the cornerstones of all three facets of sustainability, the TBL [7]. Poor health and healthcare systems affect individuals, societies, the planet, and economies [8,9]. Healthcare systems are under increasing pressure to reduce their social, economic, and environmental costs [7,8,10]. An increasingly substantial share of GDPs is being spent on healthcare by countries around the world [4]. The US spends around 20% of its economy on healthcare [11]. Comparing different industries, the healthcare sector is notoriously inefficient and wasteful of resources and budgets [11]. Both the mental and physical health of people around the world is declining with the surge in lifelong illnesses with ageing populations. The cost of healthcare is rising while the quality is declining. The COVID-19 pandemic has hindered many patients with chronic and terminal illnesses from getting proper treatment due to pandemic regulations [12]. Disparities and inequities in healthcare around the world are rising [13]. Most important of all, healthcare systems are a major contributor to the deterioration of our planet's environment due to high energy usage, disposable supplies, etc. This environmental damage causes health problems that need to be medically treated, while these treatments in turn further damage the environment, and thus a chain reaction [14].

There is a growing demand for the prevention of diseases and reducing the need for disease monitoring, screening, and treatment to a minimum [7–10]. This in turn requires an open space for innovations and dynamic interactions between healthcare stakeholders to understand and provide solutions, services, information and resource supply chains, and community support structures for timely interventions and disease prevention and progression. Fortunately, social media, which hosts around 60% of the world's population [15], is one such platform that provides an open space for stakeholder interactions.

Motivated by the urgency and gravity of the challenges facing healthcare, and the technological opportunities, this paper proposes a data-driven artificial intelligence (AI) based approach called Musawah (Musawah is an Arabic word meaning Equity. We call our approach equity as we believe that sustainability and sustainable healthcare can be achieved by people believing in and practicing equity. We believe that our approach of open service and value healthcare systems based on freely available information can enable equity and sustainability) to automatically detect and identify healthcare services that can be developed or co-created by various stakeholders using social media analysis. Essentially, the aim herein is to investigate the role of big data analytics over social media to automatically detect needs and value propositions that could be nurtured and turned into service co-creation processes through engaged participation over social and digital media, eventually co-creating services. The co-created services are based on values that are not necessarily materialistic, but are driven by equity, altruism, community strengthening, innovation, and social cohesion. The case study focuses on cancer disease in Saudi Arabia

using Twitter data analytics in the Arabic language; however, the proposed approach broadly can be used for any disease or purpose and in any language.

Specifically, we detect 17 services (e.g., screening, hope and optimism, financial support, and awareness campaigns) using unsupervised machine learning from Twitter data using the Latent Dirichlet Allocation algorithm (LDA) and group them into five macro-services namely, Prevention, Treatment, Psychological Support, Socioeconomic Sustainability, and Information Availability. The Prevention macro-service includes five services—Early Diagnosis, Prevention and Control, Causes, Screening, and Symptoms—and involves measures that may avoid cancer (e.g., maintaining a healthy lifestyle and avoiding cancer-causing substances) or help recovery from cancer or slow down its progression (e.g., symptoms and early detection).

The second macro-service is Treatment and relates to various treatment options available for cancer. It includes two detected services: Chemo and Radiation Therapy and Surgical Therapy. It does not mean that only two treatment options or services are available. It simply shows that these two services were detected by our tool, which may indicate the two treatment services that are more common based on the dataset or its temporal dimensions, or it may indicate the need for new treatment services. The third macro-service is Psychological Support, which captures the services to support patients and their families to help them cope with their psychological needs, such as emotions and spirituality. It includes three services, Spiritual Support, Suffering, and Hope and Optimism. The fourth macro-service, Socioeconomic Sustainability, includes four services, namely, Government Support, Socioeconomic and Operational Challenges, Charity Organizations, and Financial Support. It relates to the financial, healthcare, and other services from government or charity organizations to address social and economic sustainability aspects of cancer prevention and treatment. The fifth macro-service is Information Availability, which emphasizes the need for the availability of information related to cancer prevention and treatment and includes three services, Breast Cancer Awareness, Awareness Campaigns, and Questionnaire and Competitions.

Subsequently, we show the possibility of finding additional services by topical searching over the dataset, where the topics could be macro-services, services, or other aspects of the services domain (cancer, in this case). We use four topical searches (Causes, Symptoms, Prevention, and Stakeholders) and detect 42 additional services that include Sports, Stress Avoidance and Control, and others. We call them extended services since these are discovered by using the knowledge gained from the initial service discovery process.

The methodology of discovering these services involves exploring the healthcare space related to cancer in Saudi Arabia using LDA-based topic modelling of Twitter data. The data shows the problems, solutions, strategies, activities, comments, and requirements of various stakeholders clustered into 20 themes that are used to develop 17 services and five macro-services by merging some clusters. The idea of developing cancer-related services from social media analysis is motivated by the well-known science that social media facilitates value co-creation through stakeholder interactions, allowing value creation or the creation of new services (see e.g., [16]).

We have developed a software tool from scratch for this work. The tool implements a complete machine learning pipeline including data collection, pre-processing, clustering, validation, and visualization components. The dataset we used contains over a million tweets (1,352,814 tweets to be precise) collected during the period 22 September–1 November 2021. We have tried to make the translation of the keywords and other Arabic content (sample tweets, etc.) contextual to allow English readers to understand the contextual use of the keywords. However, in other cases, we have used literal translation. Note also that we did not always quote a complete tweet for reasons of privacy, and in other cases as a part of the tweet was not relevant to the discussion. This work builds on our extensive research in social media analysis in English and other languages on topics in different sectors, see e.g., [17].

### *Novelty, Contribution, and Utilization*

The literature review presented in Section 2 shows that while there are many works on the use of social media in healthcare, our work is novel in several respects. Firstly, none of the existing works on social media analytics in any language have focused on cancer as extensively as we do in this paper, discovering over fifty topics in several dimensions including cancer causes, symptoms, prevention, treatment, socioeconomic sustainability, and stakeholders. Therefore, this paper provides insight and evidence from social media public and stakeholder conversations about various aspects of the cancer disease in Saudi Arabia such as public concerns, patient requirements, solutions for problems related directly and broadly to cancer such as cancer treatment, financial difficulties, psychological ordeals and traumas, operational challenges for the families of cancer patients and other information. The fact that there are limited works in social media analytics in healthcare, particularly cancer, and considering that studies on social media analytics in the Arabic language within Saudi Arabia are even more limited, and that we have used a dataset that we carefully curated for this study differentiates our work from others.

Secondly, none of the existing works have proposed a similar approach of using social media and AI to extract healthcare or cancer services. We believe systemizing this approach could lead to a revolution in sustainable healthcare, driven by communities co-creating social values with the exchange of services for services that are not necessarily materialistic, but driven by equity, altruism, community strengthening, innovation, and social cohesion. Open service and value healthcare systems based on freely available information can revolutionize healthcare in manners similar to the open-source revolution by using information made available by the public, the government, third and fourth sectors, or others, allowing new forms of preventions, cures, treatments, and support structures.

The rest of this paper is organized as follows. Section 2 discusses the related work. Section 3 details the methodology and design of the Musawah tool. Section 4 introduces and explains the 17 discovered services and five macro-services. Section 5 discusses the 42 extended services that we discover by using knowledge gained from and extending the initial discovery process. Section 6 provides discussion. Section 7 concludes and gives directions for future work.

## **2. Literature Review**

In this section, we provide a review on the topics related to this research paper. Section 2.1 reviews the works related to healthcare that have utilized social media. In Section 2.2, we review the literature related to the use of social media that focuses on cancer-related studies; Section 2.3 presents the same however limiting to works in the Arabic language alone. Section 2.4 reviews some works on automatic naming of topics and clusters. Section 2.5 discusses the research gap.

### *2.1. Social Media and Healthcare*

Researchers have extensively used social media analytics in different application domains such as analysis of perception people have about AI and other technologies for city planning [18], city logistics [19,20], disaster detection [21], sentiment analyses of public opinions of government services [22], detection of various events related to road traffic [17,23,24], detection of symptoms and diseases nationally and across different urban areas [4], and urban governance during the COVID-19 pandemic [12].

Specifically, in healthcare, many researchers have used social media as a powerful tool for addressing healthcare system challenges and evaluating the conversations about various health-related topics. The COVID-19 pandemic has produced many challenges for public healthcare systems around the world. The main challenge is providing effective health care to patients while maintaining the safety of service providers on the frontline. Telemedicine appeared as a prominent service to achieve this end. This service integrates the use of technology into medical practices to reduce face-to-face meetings. Leite and



Hodgkinson [25] built a framework to examine how patients and doctors can co-create service value for the telemedicine ecosystem with/for any kind of crisis. They collected a sample about telemedicine and included 146 tweets from patients and 734 tweets from doctors. They applied inclusion and exclusion criteria on tweets to get the final sample. Thematic analysis was implemented to reveal seven themes in the data including equitable healthcare delivery and others. The result showed that the value co-creation framework is a platform for inclusive, sustainable, and equitable telemedicine services.

Some studies have used machine learning and deep learning methods for healthcare-related analytics using digital media data in healthcare. Jahanbin et al. [26] proposed an approach for extracting storing, monitoring, and visualizing data related to infectious diseases by using news and tweets mining techniques. The approach is called “Fuzzy Algorithm for Extraction, Monitoring, and Classification of Infectious Diseases (FAEMC-ID)”. The proposed system collected 10,000 news items and tweets. Classification with the evolving fuzzy model is performed on the dataset. The classified data were visualized on the world map to identify the high-risk areas. The proposed system helped to monitor infectious diseases and control the spread of epidemics in a timely manner. The results of the proposed analysis showed a high accuracy of 88.41%, compared to other algorithms. Alotaibi et al. [4] constructed the Sehaa system, which is a big data analytics tool for improving healthcare in KSA using Arabic Twitter data. Sehaa tool consisted of four main modules. The data collection module captured 18.9 million tweets about health symptoms and diseases in KSA. The pre-processing module is used to clean tweets and manually label them by using two levels of labeling: ‘related’ and ‘unrelated’ tweets, and then labeling related tweets to ‘awareness’ or ‘inflicted by’ a disease. The third module was classification. This phase used six classifiers, including Naive Bayes, Logistic Regression, and four methods of feature extraction to detect different diseases in Saudi Arabia. The validation module was used to evaluate the performance of the classifiers by using F1-Score and accuracy. The results of the study were visualized and validated using external sources, including national statistics, news media, and research reports. The results revealed that (1) the top five diseases in KSA are dermal diseases, hypertension, diabetes, heart diseases, and cancer. (2) Riyadh and Jeddah needed more awareness about the diseases, while Taif was the healthiest city.

## 2.2. Social Media and Cancer

We review here the studies related to cancer where social media is used as the data source. First, we review works without regard to any language or a modelling method, and subsequently, in Section 2.2.1, we will discuss studies on cancer that have used topic modelling. Table 1 summarizes some of the studies that we have discussed in this section, including some other relevant works that, despite not having been discussed, we have summarized for the readers’ interest.

The literature review has looked at various issues related to cancer disease. For example, some works have focused on studying emotions among cancer patients. Wang and Wei [27] constructed an approach to study the emotions shared by patients’ cancer communities on Twitter. The study collected 53,026 tweets, posted by 39,504 Twitter users, which included 29,979 retweets and 23,047 original tweets. The deep learning (Recurrent Neural Network RNN) models were used to extract emotions from English tweets and classified them into anger, joy, sadness, fear, hope, and bittersweet. The results of the study showed that joy was the most used emotion in tweets, then sadness and fear, whereas anger, hope, and bittersweet were less shared. Also, influencers were among the account that posted the most content with positive emotions. Crannell et al. [28] built an approach that aims to analyze the content of tweets written by cancer patients in the US and compute the average happiness for patients based on the cancer type. The study collected 186,406 tweets from March 2014 to December 2014. The pre-processing step was implemented to clean the dataset. The tweets were filtered based on cancer diagnosis and using regular expression software pattern matching. They used different filters such as “Breast cancer”, “Lung

cancer", and others. The authors utilized the patients' Twitter identification numbers to collect all tweets for each patient and calculate the average happiness value by the quantitative hedonometric analysis and the Mechanical Turk (LabMT). The results of the study showed that the most common cancers were 'breast', 'lung', 'prostate', and 'colorectal'. The calculated happiness values for every type of cancer showed that the happiness values were higher for breast, thyroid, and lymphoma cancers, and lower for kidney, pancreatic, and lung cancers. The study proved that the patients express their illness on social media.

Several studies have focused on studying cancer types. A number of works have looked into breast cancer. Modave et al. [29] constructed an approach to understand the perceptions and attitudes of people related to breast cancer on Twitter. The study collected 1,672,178 tweets related to breast cancer. The pre-processing was implemented to clean and prepare tweets for classification models. The text was classified into three groups including irrelevant, promotional, and laypeople's discussions by using CNN and LSTM models in Keras library that run on top of the Tensorflow framework. Then, the sentiment analysis was performed by using the Linguistic Inquiry and Word Count (LIWC) tool to analyze emotions/attitudes into five statuses: sadness, anger, anxiety, positive emotion, and negative emotion. Finally, topic modeling used the Biterm algorithm to identify the main topics from relevant tweets. The detected topics include awareness month events, treatment, risk, family, diagnosis, news, friend, screening, pink goods, and study. The NLP with machine learning is a useful tool to evaluate people's attitudes from their health tweets and detect their perceptions about specific health topics. Meena et al. [30] built a system based on social media and Twitter to extract the sentiments for breast cancer disease and chemotherapy treatment. The study collected 10,000 tweets by using the keywords 'breast cancer' and 'chemotherapy'. The sentiment analysis was implemented using the Naive Bayes algorithm to classify the sentiment of tweets into neutral, strongly positive, weakly positive, strongly negative, negative, or weakly negative. Bigram analysis was used to find the most frequent word in the tweets for 'breast cancer' and 'chemotherapy'. The results of the proposed system showed that the sentiments for breast cancer and chemotherapy were weakly positive or neutral. Thus, sentiment analysis for tweets has a great impact on health care.

Rasool et al. [31] have developed a framework that aimed to predict breast cancer risk in real-time using ML and spark as big data analysis platforms. The study used 'Wisconsin Breast Cancer Data-Set (WBCD)', which was imported from UCI repository and included 699 records with 11 attributes. The preprocessing steps were used to remove unnecessary attributes and complete the absent values of the attribute. Five classifiers were used, to detect cancer to be benign or malignant, including Logistic Regression, Gradient Boosted Trees, Support Vector Machine, Random Forest, and Decision Tree. The performance of the five classifiers was evaluated by computing the accuracy, specificity, and sensitivity for algorithms. The results reveal that the highest performance obtained for predicting the risk of disease in real-time was by RF classifier. Diddi and Lundy [32] constructed a methodology that aimed to examine how the health organizations including "Susan G. Komen", "U.S. News Health", "Breast Cancer Social Media", and "Woman's Hospital", used their Twitter accounts to talk and support aspects of breast cancer in October, which is dedicated to breast cancer awareness. The study collected 2961 tweets, 136 of the tweets belong to Woman's Hospital, 180 tweets for US Health News, 280 tweets for Komen, and 2365 tweets for BCSM. After using qualitative coding, content analysis was used to analyze, and categorize the tweets. The result of the proposed system showed that the highest number of tweets was written by BCSM, then Susan G Komen, US Health News, and Woman's Hospital. Qualitative content analysis showed that the most prevalent construct of the Health Belief Model (HBM) was "perceived barriers", then "cues to action", "perceived benefits", "self-efficacy", and "perceived threat".

Some works have explored skin cancer. Silva et al. [33] aimed to analyze Twitter posts in Australia about skin cancer. The study collected 12,927 tweets. Data filtering

was implemented by using inclusion and exclusion criteria, removing non-English, unrelated, or unclear content. The quantitative and qualitative analyses were implemented on datasets to calculate the number of tweets, finding the relationship between the number of tweets and temperature by using linear regression analysis. The content analysis identified three themes: health-related expressiveness, health information, and health advice. Nguyen et al. [34] developed a methodology that aimed to evaluate the influence of the campaign (Don't Fry Day) by the National Council on Skin Cancer Prevention (NCSCP)'s 2018 on Twitter, categorize types of accounts participating in the campaign, and determine the themes of the tweets. The dataset was collected from 1881 Twitter accounts. The study categorized the types of contributors in the campaign, based on username, into 'government-affiliated account (federal)', 'nongovernmental organization (NCSCP and health)', 'government-affiliated account (state/local)', 'health/cancer/medical center', 'individual', 'news/media', 'businesses', and 'other/unknown'. After manual coding of tweets, content analysis was used to identify themes of the tweets which are 'informative', 'minimally informative', and self-interest campaign promotion themes. The results of the study showed the large effective use of social media to promote public health campaigns and understand the types of messages and accounts involved in social media campaigns.

**Table 1.** Studies on the Types of Cancer Using Social Media.

Author	Type of Cancer	Tweets (×1000)	Description
Modave et al. [29]	Breast	1672	Understand the perceptions and attitudes of people related to breast cancer on Twitter.
Meena et al. [30]	Breast	10	Extracted the sentiments for breast cancer disease and chemotherapy treatment.
Diddi et al. [32]	Breast	3	Examined how the four health organizations used their Twitter accounts to talk and support aspects of breast cancer in October.
Silva et al. [33]	Skin	13	Analyzed Twitter posts in Australia about skin cancer.
Sedrak et al. [35]	Lung	26	Analyzed content related to lung cancer and examined the dialogues of lung cancer clinical trials.
Sutton et al. [36]	Lung	24	Examined the user type and the content of tweets of lung cancer to identify the nature of messages within the cancer control continuum.
Sedrak et al. [37]	Kidney	2	Implement an exploratory analysis for the content of Twitter related to kidney cancer and the participants in this content.
Nejad et al. [38]	Childhood	286	Examine how Twitter was used during childhood cancer awareness month (CCAM).

Some works have investigated lung cancer. Sedrak et al. [35] built a Twitter-based approach to analyze content related to lung cancer and examined the dialogues of lung cancer clinical trials. The study used Nvivo (qualitative data analysis software) to collect 26,059 tweets and prepare them for the analysis stage by removing duplicates and non-English tweets. The content analysis was used on the final sample to categorize tweeters as individuals, media, or organizations. The tweets were categorized into support, clinical trials, prevention, treatment, general information, diagnosis, screening, or symptoms. Most of the tweets were posted by individuals and were focused on support and prevention. The results of the study demonstrated that social media is a promising and useful source to explore how patients conceptualize and communicate about any health issues. Sutton et al. [36] constructed a methodology that aimed to examine the user type and the content of tweets related to lung cancer to identify the nature of messages within the cancer control continuum. The study collected 1.3 million tweets, including 23,926 messages related to lung cancer. The 3000 tweets were manually coded by three coders. Descriptive and inferential statistics were performed to study the content of tweets and types of users. The content of tweets was categorized into treatment, awareness, prevention and risk information, survivorship, end of life, diagnosis, active cancer–unknown phase, and early detection. The types of users were categorized into individual, organizational, media, and

unknown. The SPSS was used to analyze data and chi-square was used to assess and identify the relationships between user types, and cancer content. The experimental result showed that most of the tweets focused on treatment, awareness, risk, and prevention topics. The majority of messages were tweeted by individual users. Twitter is a useful source in raising awareness about cancer by timely dissemination of information about the treatment and prevention methods.

Among the studies that have explored kidney cancer includes another study by Sedrak et al. [37], which have built an approach to implement an exploratory analysis for the content of Twitter related to kidney cancer and the participants in this content. The study collected 2568 tweets. The NCapture tool was used to collect tweets and save them into an Excel file. Non-English tweets and those not related to kidney cancer were excluded from the dataset. Finally, the content analysis methods were used to analyze the final dataset of 2097 tweets. The content of tweets was classified into different domains including support, general information, treatment, clinical trials, donation, diagnosis, and prevention. Also, the user was classified into individual, media, or organization. The results showed that 858 tweets were written by individuals, 865 tweets authored by organizations and 364 tweets written by media sites. Most discussed content included support by 29.3% and treatment by 26.5%.

Some studies have looked into the types of children's cancer. Nejad et al. [38] built a descriptive-analytical methodology to examine how Twitter was used during childhood cancer awareness month (CCAM). The study collected 285,859 tweets related to childhood cancer, which coincided with CCAM. Latent Dirichlet allocation (LDA) was used as a method of content analysis to discover the topics using MALLET from tweets. The detected topics included awareness, loved ones, walks and runs, wearing gold, and fundraising. Twitter user was categorized into four groups including news agencies, individuals, organizations, and celebrities. They observed that the most frequent topics are awareness and fundraising. The experimental results for the proposed system showed that Twitter has proven to be an effective channel for communication and raising awareness about childhood cancer. Thus, Twitter can play an important role in the dissemination of awareness among patients.

Some studies have focused on cancer treatment using social media data to utilize patients' experiences with cancer treatment methods, including chemotherapy and radiation. Meeking [39] built an approach to understanding the role of social media platforms like Twitter, as a novel digital data source, in understanding people's experiences of radiotherapy. The study collected 442 unique tweets about radiotherapy written by patients and their families. The quantitative content analysis was used to classify the data based on the content of the tweets or the user Identity into different types of accounts such as patient, healthcare organization, and healthcare professional. Thematic analysis was used to detect six themes. Three main themes related to the pathway of radiotherapy: pre-, during-, and post-treatment. The other three themes included 'emotional and informational support', 'impact on loved ones', and 'giving thanks'. The paper highlighted the psychological and physical effects of treatment and how patients seek to get emotional and information support from social media. The informational support can be improved by increasing online support. The experimental results show that Twitter is a useful platform to share and discuss people's experiences with radiotherapy. Other studies on utilizing social media applications such as Twitter for cancer treatment include [40].

### 2.2.1. Topic Modeling and Cancer

We review in this section the works about cancer analysis using social media that have utilized topic modelling algorithms such as Latent Dirichlet Allocation (LDA). Zhang et al. [40] aimed to examine and compare patients' and healthcare professionals' perceptions of chemotherapy by analyzing chemo-related tweets. The study gathered tweets by using the keywords "chemotherapy" or "chemo" from the cancer-related accounts. They collected 13,273 individual tweets and 14,051 organization tweets. Topic modeling, sentiment analysis,

and a word co-occurrence network were used to analyze the content of tweets. LDA was used to detect the most important topics related to chemotherapy, which are hormone therapy, radiation therapy, and surgery. A word co-occurrence network was shown to have a strong relationship between chemo-surgery tweets. Sentiment analysis was used to determine how patients felt about chemotherapy outcomes and their attitudes toward chemotherapy across disease types. The study shows that Twitter is a helpful healthcare data source for oncologists (organizations) to better understand patients' experiences during chemotherapy, generate personalized treatment plans, and enhance clinical electronic medical records (EMRs). Nejad et al. [38] examined how Twitter was used during childhood cancer awareness month (CCAM). They applied LDA and extracted six topics from 285,859 tweets including awareness, clothing, loved ones, walks, fundraising, and art. Modave et al. [29] constructed an approach to understand the perceptions and attitudes of people related to breast cancer on Twitter. They applied Bitern topic modeling on 1,672,178 collected tweets related to breast cancer. The detected topics include awareness month events, treatment, risk, family, diagnosis, news, friend, screening, pink goods, and study.

### *2.3. Social Media and Cancer (Arabic Language)*

There are some studies related to cancer analysis using Twitter in the English language, however, the research studies about cancer using social media in the Arabic language are limited. We have found only one such work on social media analytics in cancer research, which is relevant to this paper. The work is by Alghamdi et al. [41] and is focused on studying chemotherapy misconceptions among Twitter users. They built an approach to evaluate Twitter conversations and find misconceptions about chemotherapy among Arabic populations. They collected 402,157 tweets by using the Twitter Archiver tool. Retweets and irrelevant tweets have been removed. A manual content analysis was implemented to classify the users, themes, and common misconceptions for chemotherapy. The categories of the user included cancer patients, general users, relatives/friends of patients, accounts for health, accounts for Media, and cancer specialists. Tweets' themes categories included advice and information, misconception, experience, prayers and wishes, seeking or offering medical/financial help, seeking medical information/advice, and analogy. Statistical analysis was implemented by using the SPSS software and reported that the most of tweets were written by general users, then followed by the friends and relatives of cancer patients. For themes, prayers and wishes were the most popular topics. Descriptive statistical analysis was performed to find the frequencies between every theme and the user category. The chi-square test was used and the result revealed that a high association exists between the themes of the tweets and the category of the users.

### *2.4. Automatic Naming of Topics*

The automatic naming of topics is desired and has been researched in the literature. We discuss a few works here to introduce this topic to the reader. The probabilistic topic models aim to reveal latent topics from the corpora of text. Probabilistic topic models represent each document as a set of topics and every topic has a set of words. Humans can label every topic based on the top words of the topic, however, difficulties arise when they do not have good knowledge in the field of the documents. To address the problem, automatic topic labeling techniques can be used to assign interpretable labels for topics. Allahyari et al. [42] built a knowledge-based topic model, namely, KB-LDA that combines topic models with ontological concepts. The study aimed to use the semantic knowledge graph of concepts in ontology and their diverse relationships, as DBpedia, with unsupervised topic models (LDA) to automatically generate meaningful topic labels. The proposed approach helped to describe topics in a more extensive way and improve the quality of topic labeling by making use of the topic-concept relations, which can automatically generate meaningful labels. Lau et al. [43] constructed an approach for automatically labeling topics produced from the LDA topic model. The study generated a label candidate set from top-ranking topic terms, Wikipedia titles containing the top topic terms, and sub phrases extracted from



Wikipedia articles. Then, it used the ranking method to find the best label for each topic by using a mixture of association measures and lexical features, fed into a supervised ranking model using the support vector regression (SVR) model.

Wan and Wang [44] proposed a method that aimed to use text summaries for automatic labeling of topics produced by topic models. The study extracted several sentences from the most related documents to form the summary for each topic. The system proposed a summarization algorithm based on submodular optimization to create summaries with high coverage, relevance, and discrimination for all of the topics. The results of the proposed method showed that the use of summaries as labels had apparent advantages compared with the use of words and phrases. He et al. [45] built a graph-based ranking model, namely TLRank, that aimed to automatically label every topic produced from topic models. The proposed model used the strategy of enhancing matrix transition probability based on the textual similarity between vertices and the characteristics of vertices (sentences). The strategy seeks to restrain the topic label redundancy and enhance its diversity. The experiment results showed that the TLRank model has the ability to generate topic labels with high coverage, relevance, and discrimination. Suadaa and Purwarianti [46] introduced a methodology to cluster and label the Indonesian text by using Latent Dirichlet Allocation (LDA) and Term Frequency-Inverse Cluster Frequency (TFxICF). The methodology performed text preprocessing processes by abbreviations extraction, tokenization, stemming, and stop-words removing. For topic modeling, Latent Dirichlet Allocation (LDA) is implemented to cluster documents and detect hidden topics. The clustering quality was evaluated by using precision, recall, and F-measure. For the labeling phase, the study has used Term Frequency-Inverse Cluster Frequency (TFxICF) to label each cluster based on the phrase or word tokens that have the highest TFxICF.

### 2.5. Research Gap

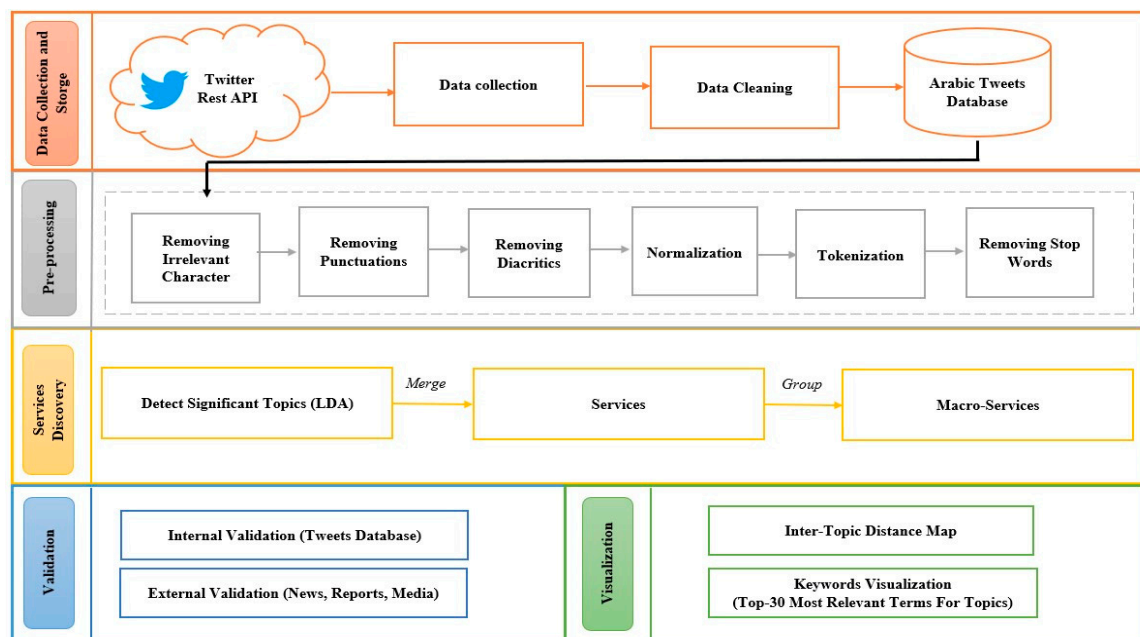
The literature review presented in this section establishes that while there are many works on the use of social media in healthcare, our work is novel as none of the existing works in any language have focused on cancer as extensively as we do in this paper and, secondly, none of the existing works have proposed a similar approach of using social media and AI to extract healthcare or cancer services.

## 3. Methodology and Design

This section explains the methodology and design of the Musawah tool. The proposed system architecture is described in Figure 1. It contains five components: data collection, pre-processing, service discovery, visualization, and evaluation; these components will be discussed in Section 3.3 to Section 3.7, respectively. Before we describe the system architecture, we discuss the conceptual development of the Musawah tool and system in Section 3.1 and give an overview of the tool in Section 3.2.

### 3.1. Musawah Service Co-Creation System: Developing the Concept

This paper proposes a data-driven artificial intelligence (AI) based approach called Musawah to automatically detect and identify healthcare services that can be developed or co-created by various stakeholders using social media analysis. We called our approach and system Musawah (equity in Arabic) as we believe that the current healthcare trends and doctrines are not sustainable—socially, environmentally, and economically—and that, with a belief and practice in equity by people, our approach of open service and value healthcare systems based on freely available information can enable equity and sustainability. The case study focuses on cancer disease in Saudi Arabia using Twitter data analytics in the Arabic language; however, the proposed approach can be used broadly for any disease or purpose and in any language. We discuss first the conceptual development of the Musawah tool and system in the following discussion where topics include the need for data sharing and participatory governance, developing services, value and service co-creation, and the motivations behind it.



**Figure 1.** The Proposed System Architecture.

Making cities smarter needs a holistic approach to manage, improve, and adapt the services provided by the city to its citizens. A smart city can be seen as a system of services [47]. With this vision, services are the basic concept of smart cities. Service science studies value co-creation as an interaction mechanism between technology, people, shared information, and organizations to improve operational efficiency, enhance citizen welfare and the quality of government services. The focal point for this vision is based on the co-creation of value through the direct citizens' participation in the service development and evaluation process. In a smart city, value co-creation is achieved by data sharing and the exchange of knowledge between citizens and the city [48]. It is necessary to understand smart service systems to foster innovations and the development of these systems in different fields. This will help to enhance understanding among people, facilitate collaborative analysis, create synergy between applications, and support systems development for smart cities.

Governments can utilize open data as a new method for developing services that allow external stakeholders to increase their role in the innovation of government services [49]. This is in contrast to previous methods of e-government service innovation, the governments create and develop services by the agencies themselves. Currently, organizations can achieve open service innovation by including the external stakeholders in enhancing existing services and developing new services based on comments or ideas generated by their stakeholders and from the problems they encounter. The open innovation in services refers to the transition from the closed innovation model based on internal knowledge to the innovation paradigm that resides both externally and internally to an organization [49]. For example, the Service Systems Development Process (SSDP) [50,51] contains a set of iterative phases that require inputs and expected outputs for every phase to understand the activities and needs required for realizing service systems. These processes are implemented as an iterative approach with the alignment of service system entities to understand the impact on enterprise capabilities and processes, technology, information systems, and customer expectations. Firstly, in the service strategy stage, newly developed services are determined by examining and knowing end-user needs, organization strategies, trends of mass collaboration, and trends of technology. The organization decides to develop the selected service systems based on the high-level socio-techno-economic feasibility study. During the service design and development stage, the requirements are analyzed. The service system entities' functions and linkages are identified. Also, it ensures the achievement of the activities of service integration, verification and validation, which include information exchange

between the entities of the service system to provide the service continuously. In the service transition and deployment stage, the service is tested to ensure readiness insertion and operation. When the service is deployed, it enters into the operations stage [50,51].

Governments are making increased efforts to innovate public service delivery, such as under the e-government umbrella [52]. The success of e-government depends on the need to better understand the requirements of citizens and include them in the development of e-government services. This can be achieved through using the service systems perspective as “a guiding framework” to analyze the development initiatives for participatory e-government services. Specifically, these initiatives can be analyzed as a service system based on the four key resources of people, shared information, organizations, and technologies, which will help to derive possible developments and enhancements for services [52]. Similar concepts have been discussed by many researchers in the past. For example, Salminen [53] has proposed the new service development process (NSD) that contains five main stages. The most important component of the model is the continuous interaction with the customer. This interaction is dependent on the common value model, which helps companies to discuss with the customer the potential value of the new services such as cost-effectiveness, process improvements, or new product features. There are three phases for the value management process; “Value Evaluation/ Assessment”, “Value Creation”, and “Value Maximization/Delivery” [53].

Health care affects economies dramatically worldwide, while affecting directly the quality of life of individuals. Traditionally, in healthcare systems, customers have been seen as passive recipients of services instead of being active elements. With rising healthcare expenditures and a growing desire for more personalized and better care, healthcare systems have recognized the importance of patients in co-creating the healthcare service experience during the last decade [54]. The Service-Dominant Logic (SDL) concept captures and represents this new paradigm of co-creating [55–57]. Note that SDL is a general concept and paradigm and is not limited to healthcare. SDL could be used to incorporate patients in creating value and producing services in this environment. Value co-creation is defined in this approach as a process, in which several actors exchange resources and collaboratively create and produce value. As a result, the information sharing between the actors of services is critical to SDL and value co-creation [16]. The new technologies represent an important role in facilitating the process of value co-creation in service science. We can build a more connected and smarter healthcare system that can provide greater help, predict and prevent illness, enable patients to make more responsible decisions. The emergence of social media, in particular, has resulted in significant changes in the processes of creating, promoting, and sharing content. The use of social media by people and public health organizations is being used to communicate in new ways for various mutually beneficial activities [54].

Note that while these and other works have investigated the use of social media for value, co-creation by various actors communicating to each other, and sharing information, none of the works in the literature have proposed the discovery of healthcare services automatically from social media. This work, therefore, contributes a novel approach of using social media and AI to discover, develop, and exchange healthcare or cancer services. We demonstrate the potential of our approach by discovering over fifty services in several dimensions of cancer disease including cancer causes, symptoms, prevention, treatment, socioeconomic sustainability, and stakeholders. We plan to systemize this approach by creating open service and value healthcare systems based on freely available information made available by the public, the government, third and fourth sectors, or others, allowing new forms of preventions, cures, treatments, and support structures to be discovered and exchanged between various parties—an exchange of values for values or services for services that are not necessarily materialistic but driven by equity, altruism, community strengthening, innovation, and social cohesion. Service co-creation is enabled through engaged participation of need and value propositions. In co-creation terminology, both parties, a customer and company, create value for each other and this value could be money, product, service, or anything of value for the parties involved. Therefore, the value could

be anything materialistic or otherwise, such as a desire for equity, the altruistic nature of a person or party, and the desire to strengthen communities, bringing innovation, or social cohesion. Moreover, our approach also allows the gaining of insight and gathering of evidence from social media public and stakeholder conversations about various aspects of the cancer disease in Saudi Arabia such as public concerns, patient requirements, solutions for problems related directly and broadly to cancer such as cancer treatment, financial difficulties, psychological ordeals and traumas, operational challenges for the families of cancer patients and other information.

As regards the discovery and naming of services, in this paper, we used a process of discovering clusters or topics (we call them services) using topic modelling of Twitter data using the Latent Dirichlet Allocation (LDA) algorithm. We experimented with discovering a different number of clusters using LDA—10, 15, 20, and other numbers—and found that 20 topics gave the best results for discovering important services from Twitter data. The number of clusters to be detected is dependent on the size, nature, and other properties of data, and this is an active area of research. Following the extraction of topics, we merged some of these topics, while others were kept individually. We call these merged or individual topics ‘services’, and group these services further into groups of services called ‘macro-services’. This whole process created a set of 17 services and five macro-services. The services were named manually using topic keywords and domain knowledge. We have provided a review of the automatic naming of topics (services in this case) in Section 2.5. This is an area that we plan to work on in the future to automate the service discovery, definition, transition, and deployment process.

More details on the methodology will follow in the rest of this section and the paper. The paper has contributed to certain areas, while other topics such as automatic naming of services, will be investigated in the future.

### 3.2. System Overview

We built the Musawah system in order to discover healthcare services that can be developed or co-created by various stakeholders using social media analysis. The proposed approach can be used for any disease or purpose and in any language. However, this work focuses on cancer disease in Saudi Arabia using Twitter data analytics in the Arabic language. The proposed system architecture contains five components: data collection; pre-processing; service discovery; visualization; and evaluation. First, we collected and downloaded tweets using Twitter REST API and the Tweepy library with the set of predefined parameters as Keywords and the language. The tweets collected were tweet objects in the JSON (JavaScript Object Notation) format. After collecting JSON files, we converted and saved them into CSV format. It is well known that the contents of social media contain noise and unstructured data, and many errors. Therefore, in the Pre-Processing component, the acquired data are cleaned and preprocessed to ensure its readiness for the analysis stage. For removing stop-words, we used the Natural Language Toolkit (NLTK) library with add a new list of stop-words to remove from the text. Also, we used the `simple_preprocess()` in Python to tokenize each tweet into a list of words and used a regular expression in python to remove the unnecessary characters from the tweets. Subsequently, we used the Gensim Python library to build a service discovery module using Latent Dirichlet Allocation (LDA). We have adjusted the settings of the parameters of the LDA model: the number of topics was 20, passes were 10, and iteration was 100. The number of topics is considered an important parameter for building a model. The passes refer to how many times the algorithm is supposed to pass over the whole corpus. The iterations are the maximum number to iterate every document in the corpus for calculating the probability of each topic. Then, we visualized the topics by using the LDAVis tool that provides a web-based interactive visualization of topics generated from LDA. Finally, the results are validated by the detected services using two techniques, which included external and internal validation.

### 3.3. The Dataset

The data was collected using Twitter REST API. Tweets were collected using various terms of cancer, its types, as well as its examination and treatment methods. For example, we used the keyterms “سرطان” (Cancer), “ورم” (Tumor), “ورم خبيث” (Malignant), “سرطان الرئة” (Lung Cancer), “الخزعة” (Biopsy), and others. Furthermore, we used several hashtags related to the types of cancer, its treatment, and awareness such as “#سرطان الرئة” (# Lung Cancer), “#سرطان الثدي” (#Breast Cancer), “#سرطان البروستات” (#Prostate Cancer), and others. In Table 2 we provide a sample of the keyterms and hashtags that were used for collecting data. A total of 1,352,814 tweets were retrieved during the period from 22 September to 1 November 2021. The tweets collected were tweet objects in the JSON (JavaScript Object Notation) format. After collecting JSON file, we converted and saved them into CSV format. Then, we cleaned the dataset by removing duplicates.

**Table 2.** The Keywords and Hashtags used to collect Dataset.

Keywords
السرطان، ورم، خبيث، سرطان الرئة، خزعة، طفرة جينية، الكشف المبكر، لوكيميا، الكيمائي، السرطنة Cancer, Tumor, Malignant Lung Cancer, Biopsy, Genetic Mutation, Early Screening, Leukemia, Chemo, Carcinogenic
Hashtags
#سرطان_الرئة، #سرطان_الثدي، #سرطان_البروستات، #افحصي_الآن، #مرض_السرطان، #علاج_السرطان، #العلاج_الكيمائي، #الكشف_المبكر، #مكافحة_سرطان_الثدي، #مرض_السرطان Lung Cancer, Breast Cancer, Prostate Cancer, Check-up Now, Cancer Patients, Cancer Treatment, Chemotherapy, Early Detection, Fighting Breast Cancer, Cancer Disease

### 3.4. Data Pre-Processing

Data preparation or preprocessing is a necessary step to complete the steps of the data analytics process. Social media generates unstructured and informal data. This involves implementing different techniques to clean the acquired data. To ensure data readiness, data preprocessing should be performed to prepare the acquired data for subsequent steps of the analysis. This will increase the accuracy and quality of data analytics. The main steps of data preprocessing include: removing irrelevant characters; tokenization; normalization; and removing stop-words.

#### 3.4.1. Removing Irrelevant Characters

The collected tweets may contain duplicates. To prevent having the same tweets, we removed the duplicates out of the collected tweets after saving and loading tweets in DataFrame format. Cleaning data included removing emails, redundant spaces and lines, single quotes, repeating characters, English alphabets, and all traces of emoji from a text file. Moreover, we removed the punctuation marks remove by creating a list for all the marks of punctuation, such as: [., , ? , ; , ! , % , @ , & , ], and other types of brackets, mathematical, and slashes symbols. This will decrease the size of the feature set and make information more valuable. We also removed eight diacritics marks:

- Three diacritical marks that are used to refer to the short vowels: Fatha (َ), Kasra (ِ), Damma (ُ);
- Three double diacritic marks: Tanwin Fath (ً), Tanwin Kasr (ٍ), Tanwin Damm (ٌ);
- Single diacritical mark to refer the absence of a vowel: Sukun (ْ);
- One diacritical mark to refer the duplicate occurrence of a consonant: Tashdid (ّ).



### 3.4.2. Tokenization and Normalization

We used the `simple_preprocess()` method in Python to tokenize each tweet into a list of words. Normalization transforms words to a basic and consistent form. Normalization of Alef, Yaa, Hamza, and Taa Marboutah, in which Alef (ا إ آ) was replaced with (ا), Yaa (ي) was replaced with (ي), and Taa Marboutah (ة) with replaced with (ة).

### 3.4.3. Stop-Words Removal

Stop-words are distributed between the texts, are not very useful, and can be effectively removed. The role of removing stop-words is to delete the word that is not important in extracting the information. The tweets can contain stop-words like: "من", "عن", "الى". We used the Natural Language Toolkit (NLTK) library with add a new list of stop-words to remove from text, such as ("منكم", "لدي", "اي", "كان", "انها", "او", "والله", "ان", "تم", "الى", "بينما", "بينى", "بشيء", "ممكّن", "اذا", "وين", "انى", "لان", "حتى", "شي", "قبل", "يتم").

### 3.5. Topic Modeling Using Unsupervised Machine Learning

Academic studies and research rely on the use of computerized analytics to understand large volumes of unstructured text that cannot be analyzed manually due to the limitations of human data processing. In this area, topic modeling is a common technique used for data analysis and topics discovery. The topic model can be defined as "a collection of algorithmic approaches that seek to find structural patterns within a collection of text documents, producing groupings of words that represent the core themes present across a corpus" [58]. In this paper, we used Latent Dirichlet Allocation (LDA) as one of the types of topic modeling. LDA is a statistical model used to determine the important topics discussed in a set of documents (tweets in our study). It creates models in an unsupervised mode, which means no need for label training data. Every document is characterized by different probability distribution among different topics. The topics are formed based on the co-occurrence of keywords with a certain probability in the same document. The LDA model takes the collection of documents as input and generates a set of topics as the output. Every topic has a set of keywords in different proportions as well. The algorithms of topic modeling are not able to name the topic. So, the topics can be labeled by humans using the keywords of the topics.

We used Genism for LDA analysis in the Python environment. The tweets represent the documents in this work. We have adjusted the set of the parameters of LDA model: the number of topics was 20, passes were 10, and iteration was 100. The number of topics is considered an important parameter for building a model, the large number of topics requires a lot of overfittings, while the small number of topics makes the model underfitting. The passes refer to how many times the algorithm is supposed to pass over the whole corpus. The iterations are the maximum number needed to iterate every document in the corpus for calculating the probability of each topic.

### 3.6. Visualizations

In this paper, we used LDAVis to visualize and label the topics. LDAVis is a web-based interactive visualization of topics generated from LDA. It provides an overview of the topics by displaying them in the circles' form. It also displays the words closely related to each topic, and the degree of relevance of each word to topics. The visualization contains two basic pieces. In the left panel, the topics are visualized as circles. In the right panel, it has a horizontal bar chart that represents the most useful individual terms for interpreting the currently selected topic on the left.

### 3.7. Evaluation and Validation

The proposed system followed two techniques to validate the detected services: external and internal validation. We used different online sources and news media as an external technique for validation. The internal technique for validation is based on the

collected Twitter data, which gives detailed information such as the account that was used to post the information, the time, and date of posting the tweet.

#### 4. Results and Analysis

This section introduces and explains the discovered services and macro-services. Section 4.1 provides an overview of the macro-services and services detected from the Twitter data using the LDA algorithm. This is followed by five subsections, Section 4.2 to Section 4.6, with an explanation for each service with examples from Twitter posts.

##### 4.1. Services and Macro-Services: An Overview

We discuss in this section the results of the system we propose. We describe the approach of discovering 17 services from Twitter data using the Latent Dirichlet Allocation algorithm (LDA). The approach involved extracting 20 topics by utilizing the LDA algorithm on Twitter data, and then combining a few of these topics into topic clusters which we call services. Throughout this research paper, the terms such as topic cluster, cluster, and service will be used interchangeably. In addition, we have grouped these seventeen services into five macro-services. In the previous section, we described the method for detecting these topics using the LDA algorithm. Python is used to develop the software.

Table 3 provides a list of services along with their corresponding data. Column 1 of the table lists the five macro-services. They are Prevention, Treatment, Psychological Support, Socioeconomic Sustainability, and Information Availability. In Column 2, we list seventeen services such as Early Diagnosis, Prevention and Control, Causes, and so on. Every macro-service contains one or more services. For instance, the second macro-service (Treatment) contains two services, which are Chemo and Radiation Therapy, and Surgical Therapy. In Column 3, the services numbers are listed. We have already discussed that we extracted 20 topics from Twitter data using LDA, and some of those topics that are related to one another are merged to form services. As an example of the merging of topics is the service (Breast Cancer Awareness), which is the result of merging Topic 9 and Topic 20. Column 4 shows the keyword percentage of the services in the table. Among the keywords in total, Topic 1 contains 9.7% out of the total keywords. Topic 2 contains 7% of the total keywords. Column 5 presents 10 keyterms (in total for all cluster topics) for each of the services. We initially collected 30 keywords per topic. From the initial 30 keywords, 10 keywords were manually chosen (using domain expertise) based on their importance to the relevant topics. Each keyword is listed in Arabic, along with its translation into English. These keywords and other content in Arabic language (example tweets, etc.) were translated contextually so that English readers can better understand the context of the keywords.

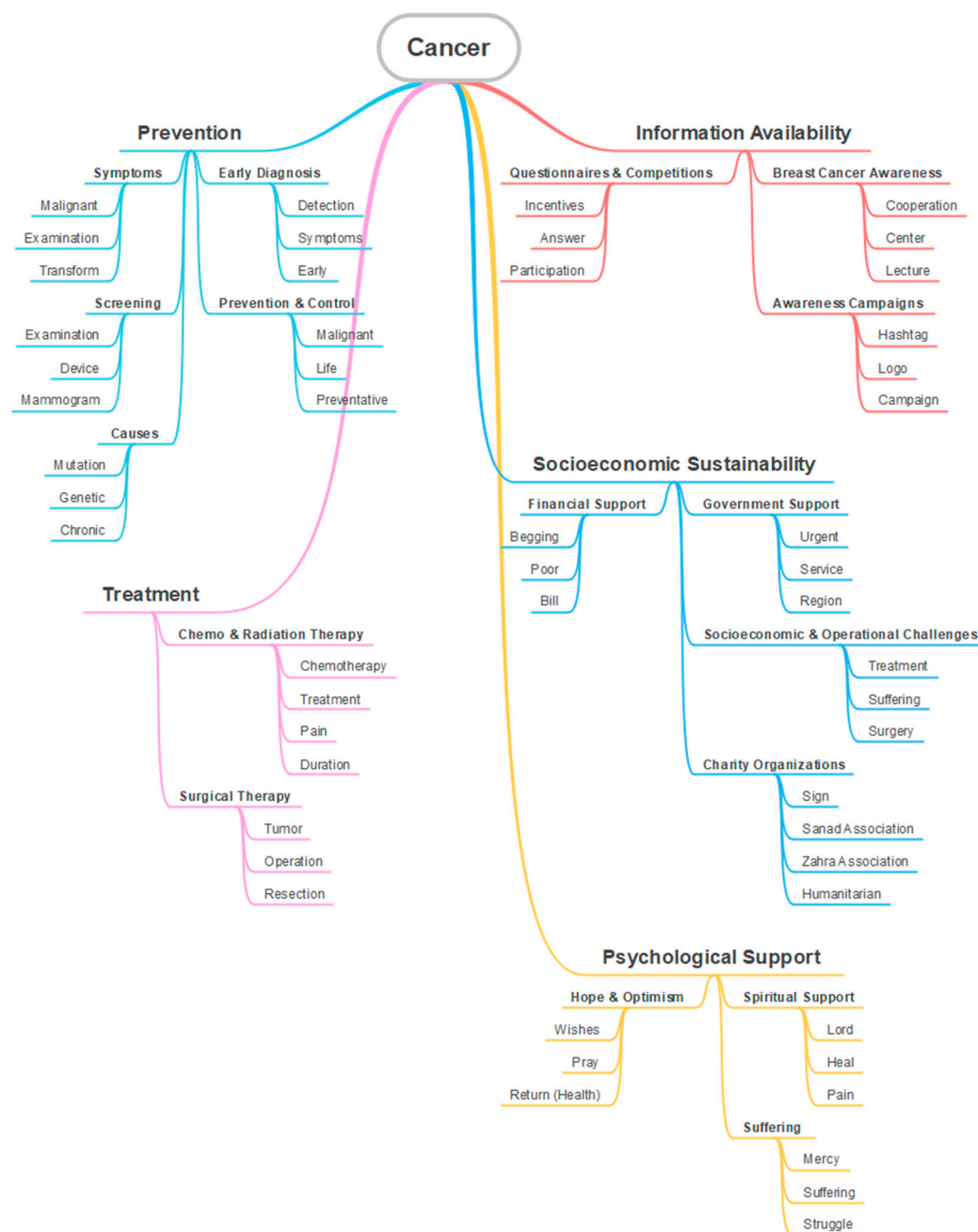
We extracted a taxonomy for the 17 services that were detected by our tool (See Figure 2). The taxonomy was created from Table 3, and it shows the services, their macro-services, and some keyterms associated with the services. The first level represents the macro services e.g., Prevention, Information Availability, and Treatment. Every macro-service contains one or more services. The second level branches represent these services e.g., Symptoms, Early Diagnosis, Causes, and Screening. Each service is characterized by various keywords. These keywords are provided in the third level branches. For instance, the keyterms Malignant, Examination, and Transform represent the Symptoms service.

**Table 3.** Macro-Services and Services Discovered from Twitter Data.

Macro-Services	Services	Topics	Keywords (%)	Keywords
Prevention	Early Diagnosis	1	9.7%	المبكر، الكشف، الفحص، الإصابة، اعراض، الوقاية، للكشف، نسبة، الشفاء، بادري Early, Detection, Examination, Injury, Symptoms, Prevention, Detection, Ratio/Percent, Healing, Go Early
	Prevention and Control	2	7%	سرطان، خبيث، الحياة، خير، للوقاية، اشياء، ورم، تأخذ، اسالكم، مرض Cancer, Malignant, Life, Good, Preventative, Things, Tumor, Take, Request, Disease
	Causes	15	4%	اثناء، بسبب، مرض، معرض، تابعه During, Because, Disease, Exposed, Related
		16	3.6%	مزمن، جينيه، طفره، ورم، الأشعة Chronic, Genetic, Mutation, Tumor, Radiation
	Screening	17	3.3%	الفحص، الذاتي، الماموجرام، بجهاز، دقائق، ينقذ، ورم، صحة، عيادة، جيت Examination, Self, Mammogram, Device, Minutes, Save, Tumor, Health, Clinic, Came
	Symptoms	19	3%	سرطان، بمراكز، بالفحص، وتتحول، احد، يجعلها، خبيث، مريضه، للثدي، اسائلكم Cancer, Centers, Examination, Transform, One, Make, Malignant, Sick, For the Breast, Appeal You
Treatment	Chemo and Radiation Therapy	3	6.6%	يحتاج، الاورام، علاج، المجاني، الكيماوي، الم، لمدته، تلقي، الحالة، استخدام Need, Tumors, Treatment, Free, Chemotherapy, Pain, Duration, Receive, Case, Use
	Surgical Therapy	4	6.3%	خبيث، استئصال، المقطع، ورم، يجب، عمليه، نوع، جانب، سرطاني، وجراحه Malignant, Resection, Part, Tumor, Must, Operation, Type, Side, Cancerous, Surgery
Psychological Support	Spiritual Support	5	6.2%	اللهم، مرضي، يارب، مرض، اشف، الم، اجسادهم، مرضي المسلمين، سقما، الشافي Oh God, Patients, Lord, Disease, Heal, Pain, Their Bodies, Muslim Patients, Sick, Healer
	Suffering	14	4.1%	مرض، رحمه، الخير، انسانيه، دعواتكم، اثر، حاله، معاناه، القاتل، صراع Disease, Mercy, Goodness, Humanity, Your Prayers, Impact, Condition/Case, Suffering, Killer, Struggle
	Hope and Optimism	18	3.1%	يدعي، اللهم، ياتي، بجميع، امنياته، محمل، اليكم، يعيد، يارب، جعلت Pray, God, Come, With All, Wishes, Loaded, To You, Return (Health), Lord, Made

Table 3. Cont.

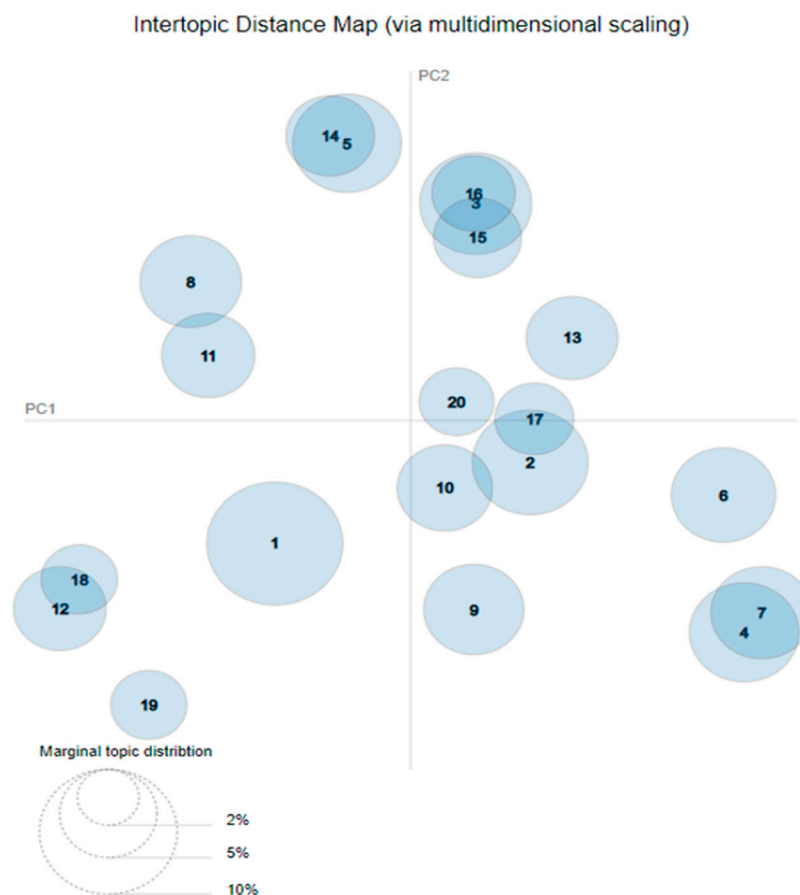
Macro-Services	Services	Topics	Keywords (%)	Keywords
Socioeconomic Sustainability	Government Support	6	5.7%	الصحة، عاجله، بأهمية، تقييم، خدمه، المنطقة، وزارة، ادويه، مرضى، الآلات Health, Urgent, Important, Evaluation, Service, Region, Ministry, Medicine, Patients, Machines
	Socioeconomic and Operational Challenges	8	5.4%	مرضى، العلاج، بحاجه، اسال، جراحه، الف، فاتوره، حاله، تعاني، رقم، الرجاء Patient, Treatment, Needs, Ask, Surgery, Thousands, Bill, Case, Suffering, Number, Please
	Charity Organizations	11	4.6%	سنه، سند، توقع، عاجل، مريضه، نداء، زهره، حاله، قررت، انساني Year, Sanad Association, To Sign, Urgent, Patient, Call, Zahra Association, Case, Decided, Humanitarian
	Financial Support	12	4.5%	سرطان، ولدى، علاج، الفاتوره، حاولنا، معسر، ما نقدر، العملية، انشدكم، خزعه Cancer, I Have, Treatment, Bill, We Tried, Poor, We Can't, Operation, Exposure, Begging, Biopsy
Information Availability	Breast Cancer Awareness	9	5.3%	الصحية، بالتعاون، حملة، جمعيه، مركز، للتوعية، محاضره، افحصي Healthy, Cooperation, Campaign, Association, Center, Awareness, Lecture, Examination
		20	2.9%	الوردي، المراد، باللون، بنشر Pink, Woman, Color, Spreading
	Awareness Campaigns	10	4.8%	حساب، السؤال، اكتوبر الوردي، رتويت، الرد، بهشتاق، التغريده، الصحي، شعار، الحمله Account, Question, Pink October, Retweet, Reply, Hashtag, Tweet, Healthy, Logo, Campaign
	Questionnaires and Competitions	13	4.4%	متابعه، المشاركه، الأمان، انتشرت، شارك، فعالية، عضو، الإجابة، عام، الاف Follow, Participation, Safety, Spread, Share, Event, Member, Answer, General, Thousands



**Figure 2.** A Taxonomy of Discovered Cancer Services.

After explaining Table 3, we now move on to the explanation of topics using graphical information. Figure 3 shows the inter-topic distances among the extracted 20 topics based on the multidimensional scale. This figure shows the topic sizes in the bottom-left corner. Topic 1 is depicted by the largest circle, reflecting that Topic 1 is the largest topic based on its number of keywords (9.7% see Table 3). Figure 4 shows the top 30 most relevant terms (or keywords) for Topic 1. These terms are in Arabic language. Table 3 provides the English translation of the keyterms. The keywords are arranged in a decreasing order of their frequency within Topic 1 (represented by maroon bars). Each keyword has a blue bar, which represents the overall term frequency. After discussing the table and topic diagrams in general terms, now we will proceed to discuss in detail each of the services with data collected from the tweets and external sources. The discussion will also present additional detail on the topic diagrams for service.

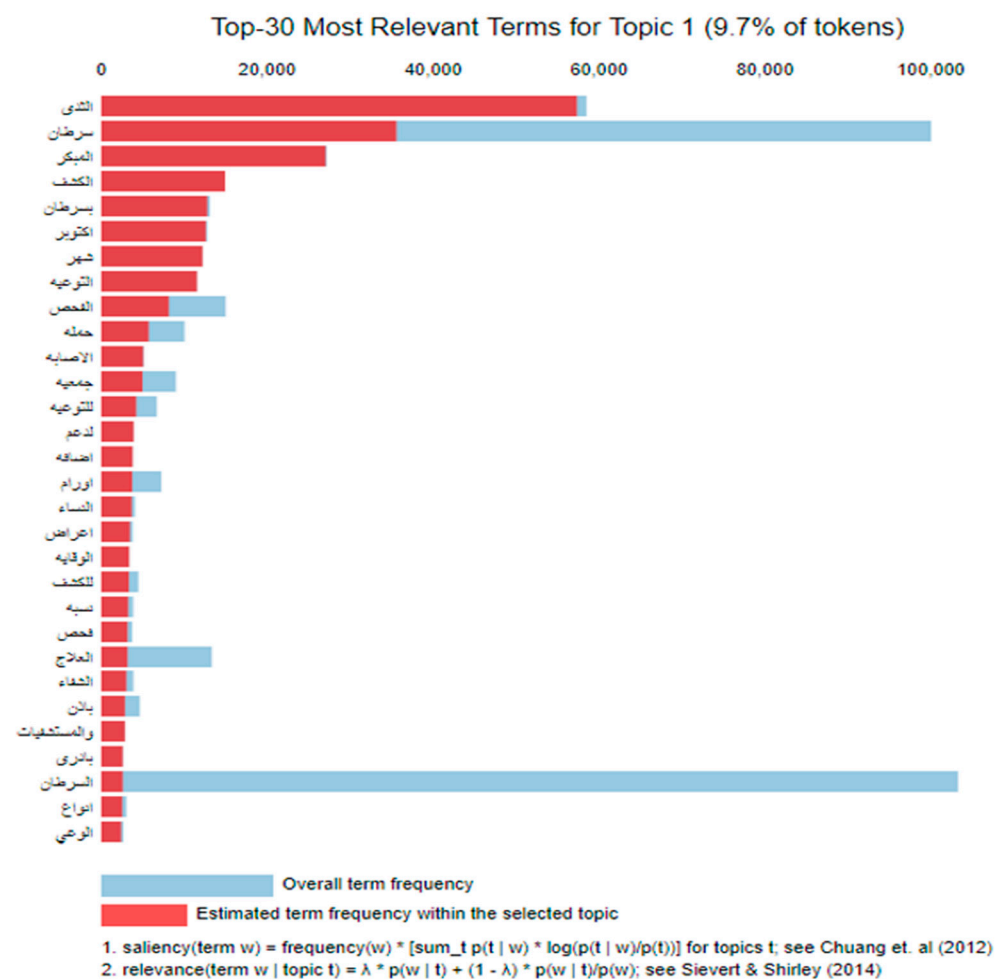




**Figure 3.** The Intertopic Distance Map of the Clusters.

#### 4.2. Prevention

We begin discussing the services related to the first macro-service, Prevention. It involves actions that reduce the chance of getting cancer (e.g., maintaining a healthy lifestyle, taking vaccines, and avoiding cancer-causing substances) and it also involves strategies and procedures that prevent cancer from getting worse (e.g., early detection, and regular screening tests). This macro-service includes five services, Early Diagnosis, Prevention and Control, Causes, Screening, and Symptoms. The first service is Early Diagnosis (see Row 1, Table 3) and is represented by keyterms from Topic 1 such as Early, Detection, Examination, Symptoms, Prevention, Treatment, and Healing. Figure 4 depicts the top 30 most relevant terms for Topic 1 (since there are 20 topics to cover, we are unable to include these figures for all of them, to avoid an excessive number of figures and to comply with the publisher's guidelines). Table 4 gives the English translation of the Arabic keywords in the figure. Early diagnosis of cancer is important for preventing bigger problems associated with health (e.g., cancer progressing to advanced stages), treatment options, and treatment costs. It focuses on detecting symptoms as soon as possible. Therefore, it can contribute to the detection of the cancer disease in its early stages. Early detection of cancer is very important for preventing reaching late stages, facilitating treatment, and accordingly, reducing the risk of death. For this purpose, it is one of the strategies used for successful treatment, and also for lowering treatment costs. The tweets related to this topic were about the early diagnosis of cancer and its importance. They were posted by official accounts, news accounts, and medical accounts. For example, the Saudi Cancer Society posted the following tweet about the importance of early detection of cancer in increasing the cure rate.



**Figure 4.** The Top 30 Most Relevant Terms for Topic 1.

**Table 4.** The keywords and their English translation for Figure 4.

Keywords	الثدي، سرطان، المبكر، بسرطان، أكتوبر، شهر، التوعية، الفحص، حملته، الإصابة، جمعية، للتوعية، لدعم، إضافة، أورام، النساء، أعراض، الوقاية، للكشف، نسبة، فحص، العلاج، الشفاء، باذن، والمستشفيات، ياترى، السرطان، انواع، الوعي
Translation	Breast, Cancer, Early, Cancer, October, Month, Awareness, Examination, Campaign, Injury, Association, for awareness, to support, Addition, Tumors, Women, Symptoms, Protection, to detect, Percentage, Checkup, Treatment, Healing, with permission, Hospitals, Hurry up, Cancer, Types, Awareness

“تضمن أهمية الكشف المبكر لمرضى السرطان في زيادة نسبة الشفاء باذن الله.”

“The importance of early detection of cancer patients is increasing the rate of recovery, God willing.”

Another tweet was posted by a news account, and indicates that 50 types of cancer can be detected through a blood test.

“فحص دم ثوري يمكنه كشف 50 نوعا من #السرطان قبل ظهور الأعراض.”

“A revolutionary blood test that can detect 50 types of #Cancer before symptoms appear.”

The second service is Prevention and Control. Preventing and controlling the acceleration of cancer growth is an important key for preventing cancer from getting worse. The tweets related to this cluster involve important factors in cancer prevention, control, and survival such as lifestyle, behaviors, early diagnosis, and detection. The tweets associated with this topic were posted by doctors, patients, and others. For example, the following tweet was posted by a medical account. It highlights the importance of a healthy lifestyle in cancer control and prevention.

"يُحدث نمط الحياة فرقاً هائلاً عندما يتعلق الأمر بالوقاية من مرض السرطان، حيث تعدّ التمارين الرياضية الجزء الأكثر أهمية في ذلك."

*"Lifestyle makes a huge difference when it comes to cancer prevention, and the exercise is the most important part."*

The third service is Causes. The keyterms represent the service including; Because, Exposed, Genetic, Mutation, and Radiation. Avoiding cancer causes is an essential strategy for preventing cancer development. Cancer causes were heavily discussed on Twitter by doctors, patients, families of patients, and others. The discussion from nonmedical experts was based on their common knowledge as well as their experiences about the causes of cancer. Genetic mutation could be one of the causes of cancer, which is inherited in some families. Other possible causes of cancer that includes exposure to chemicals, radiation, and sun rays were detected from the tweets. Sadness, the negative feelings, and the unhealthy lifestyle, and others were also mentioned in the tweets (See Section 5.1 for more details). For example, the following tweet, found in our dataset, was posted by a doctor.

"ماهي مسببات #السرطان؟ الجواب: عوامل متعددة، أبرزها: السمنة، التدخين والخمر، الوراثة، أعراض جانبية للأدوية، حالة الالتهاب."

*"What are the causes of #Cancer? Answer: Multiple factors, most notably: obesity, smoking and alcohol, genetics, side effects of medications, and inflammation."*

Furthermore, the following tweet highlights the risk of sun rays as one of the causes of cancer.

"التعرض الكثير لأشعة الشمس الضارة يسبب سرطان الجلد، الله يعافينا وإياكم."

*"Too much exposure to the harmful sun rays causes skin cancer, may God protect us and you"*

The fourth service is Screening (see Row 5, Table 3) represented by keywords such as; Examination, Self-Examination, Mammogram, Doctor, Device, Minutes, Clinic, and others. Screening tests focus on detecting disease before symptoms appear. Early detection of cancer is very crucial for preventing the disease from reaching late stages, reducing complications, facilitating treatment, making treatment more successful, and consequently, reducing the mortality risks. Thus, screening tests are very significant; Prevention. Different screening tests are used to detect cancer diseases such as mammogram, clinical breast examination, and breast self-examination for breast cancer, and pap smears, and human papillomavirus test for cervical cancer. The mammogram screening test is an X-ray image of the breast. Most of the tweets, which are related to this cluster, stress the importance of the examination. The following tweet was posted under the hashtag #breast\_cancer on 26 October 2021, by the government.

"#سرطان\_الثدي أكثر أنواع أمراض السرطان شيوعاً لدى النساء، وننصح بإجراء الفحوص الدورية بجهاز "الماموغرام" الذي يساعد في الحد من حالات الوفاة من المرض."

*"#Breast\_Cancer is the most common type of cancer in women, and we recommend regular examinations with a "mammogram" device, which helps reduce death from the disease."*

Another tweet in our dataset is posted by a medical account. It indicates the possibility of detecting cancer before it occurs.

"..... يمكن كشف سرطان الثدي قبل حصوله ب 3 سنوات، أشعة الماموغرام للثدي تكشف السرطان قبل حصوله بثلاث سنوات."

*"... Breast cancer can be detected 3 years before it occurs, and a mammogram detects cancer 3 years before it occurs."*

The fifth service is Symptoms. Observing symptoms is a vital factor for early treatment and preventing the acceleration of the disease. This cluster relates to the symptoms associated with cancer. Cancer symptoms vary from case to case depending on the organ affected

by cancer. Symptoms of cancer include fever, sweating, constipation, loss of appetite, bleeding, changes in skin color, loss of body weight, and lumps or swelling under the skin. The tweets related to this cluster were posted by doctors, patients, and other stakeholders. For example, the following tweet, obtained from our dataset, was posted by a consultant oncologist about ovarian cancer symptoms.

"سرطان المبيض هو أحد أنواع السرطانات.....أعراضه: ألم أو تورم أو شعور بالضغط في منطقة البطن والحوض، نزيف من المهبل في غير وقت الدورة، افرازات من المهبل قد تُصحب بدم، انتفاخات أو امساك."

"Ovarian cancer is a type of cancer ... Its symptoms: pain, swelling or a feeling of pressure in the abdomen and pelvis area, bleeding from the vagina outside the time of the period, secretions from the vagina that may be accompanied by blood, and swelling or constipation."

Moreover, we found another tweet related to cancer symptoms in an advanced stage. The tweet was posted by an account of a department in a hospital.

" في مراحل متقدمة، عندما ينتشر السرطان ويصبح ورم خبيث، قد تظهر اعراض مثل: اصفران في الجلد، ألم في العظام، صداع، مشاكل في التنفس."

"In advanced stages, when the cancer spreads, and becomes a malignant tumor, some symptoms may appear such as: yellowing of the skin, pain in the bones, headache, breathing problems."

Another tweet explained one of the symptoms of cancer, which is sudden weight loss.

" انا لاحظت نزول الوزن عندي بدون سبب، وبعد ما خذت منظار، طلع عندي ورم بالقولون ..... "

"I noticed that I lost weight for no reason, and after colonoscopy, I had colon cancer ... "

#### 4.3. Treatment

We now discuss the services related to the second macro-service, Treatment. It includes Chemo and Radiation Therapy and Surgical Therapy. The sixth service is Chemo and Radiation Therapy (see Table 3). The keywords represent the service, for example; Chemotherapy, Pain, Duration, Receive, and Use. Chemotherapy and Radiation are among the important treatment types for cancer diseases. The chemotherapy is usually used to kill cancer cells and is given via a vein or by mouth. Radiation is used directly with a tumor through using high doses of radiation. These types of treatments help to destroy the tumor and improve the patient's clinical condition, prevent the spread of the tumor, and stop or slow the growth of the tumor. The main difference between chemotherapy and radiation is that chemotherapy takes medical drugs that target the whole body, while radiation therapy targets cancer cells in specific areas of the body. The tweets under this cluster reflect doctors' experiences, patients' experiences and concerns about treatment option, duration, and associated pain. The tweets related to this topic were posted by official medical accounts, patients, and other stakeholders. Some tweets provide information, for example, the following tweet, found in our dataset, was posted by a doctor.

"الخيارات العلاجية كثيرة في حال اكتشاف سرطان الثدي: الجراحة "استئصال الثدي او الورم + فحص الغدد اللمفاوية او استئصالها"، العلاج الاشعاعي، الكيماوي العلاجات الهرمونية، المناعية، الموجهة. تعتمد على معطيات كل حاله على حده وعلى تفاصيل العينة بشكل كبير."

"There are many treatment options if breast cancer is detected: Surgery "mastectomy + examination or removal of lymph nodes", radiotherapy, chemotherapy, hormonal, immunomodulatory, and targeted therapies. It depends on the data of each case individually and on the details of the sample."

Other tweets reflect the patients' and stakeholders' experiences about the treatment (their concerns, treatment duration, and associated pain). For instance, the following tweet was posted by one of the stakeholders.

"ابنه صديقتي ذات الخمسة أشهر.. قرروا استئصال عينيها بسبب ورم على الشبكية والعلاج الكيميائي للعين الأخرى."

*"My friend's five-month-old daughter, they decided to remove her eye because of a tumor on the retina and chemotherapy for the other eye."*

The following tweet was posted by a patient. It shows the patient concern about the chemotherapy.

"بأقيلي 4 أيام على بداية الاسبوع الثاني من العلاج الكيميائي الشهر الأول سيكون مكثف، .....، أتمنى جسمي يتقبل العلاج ....."

*"I have 4 days left until the beginning of the second week of chemotherapy, the first month will be intense, ... , I hope my body accepts the treatment ... ."*

The seventh service is Surgical Therapy. It is characterized by keywords such as Malignant, Resection, Removal, Operation, Surgery, Tumor, and others. Surgical operation is one of the traditional cancer treatments. It is considered very effective in killing most types of cancers before the disease spreads to lymph nodes or distant sites (metastasis). Surgical treatment may be used alone or in combination with other treatment modalities, such as radiation therapy and chemotherapy. This option is taken if the cancer does not metastasize. During the surgery, doctors often remove lymph nodes near the tumor to see if cancer has spread to them. The tweets associated with this topic were posted by stakeholders' accounts, news accounts, patients, and other stakeholders. The tweets belonging to this topic include tweets reporting cancer surgeries, peoples' experiences with surgeries, and tweets asking for a financial or a moral support. The reported surgeries include those that have been completed and the ones that would be performed in the future. For example, the following tweet, obtained from our dataset, was posted on 14 September 2021, by one of the Saudi news accounts on Twitter. The tweet mentions the success of a tumor removal surgery that happened at a hospital in Taif city. The tweet was also announced by an electronic newspapers [59].

"استئصال "ورم سرطاني" زنة 7كجم من بطن سيدة .... بالطائف."

*"Resection of a "cancerous tumor" weighing 7 kg from the abdomen of a woman ... in Taif."*

Furthermore, we found another tweet posted by a cancer patient. It highlights the need for moral support.

"متابعيني الكرام..... سوف اجري عملية استئصال ورم..... دعواتكم."

*"Dear followers ... ..I will perform a tumor removal surgery at ... ..pray for me."*

#### 4.4. Psychological Support

We now discuss the services related to the third macro-service, Psychological Support. Cancer diseases affect the patients on a physical, spiritual, and emotional level. Therefore, psychological support is very important for cancer patients. It helps the patients to handle the difficulties and overcome challenges, which is an essential factor for the treatment success and survival. This macro-service includes three services; Spiritual Support, Suffering from Cancer, and Hope and Optimism. The eighth service is Spiritual Support (see Row 9, Table 3). Spiritual Support is one of the methods for providing psychological support. A common and important way for providing Spiritual Support is prayers (supplications) for cure, recovery, and patience. Making prayers are very important element in Muslims' beliefs and therefore, the patients, their families, friends, and beloved ones increase in their prayers. The keywords for this topic such as God, Cancer, Patients, Heal, Bodies, Strength, and Muslim represent the label. A lot of tweets, found in our dataset, were similar to the following tweet.



"دعواتكم لوالدي مريض سرطان."

"Pray for my father who has cancer."

"اللهم اشفي مرضى السرطان وبرد عليهم جرعات الكيماوي..."

"Oh God, heal cancer patients and make the chemotherapy doses easy on them ... "

"اللهم اشف كل جسد أرهقه مرض السرطان."

"Oh God, heal cancer patients, they are in pain."

The ninth service is Suffering from Cancer. This service represents the pain and difficulties faced by cancer patients. Providing psychological support (e.g., listening to the patient, holding hands, making prayers) is one of the strategies for helping patients fighting cancer. This service included keyterms that represent expressions of physical and psychological pain associated with cancer. For example, people have called cancer a "silent killer" since it could kill the body and reach an advanced stage without obvious symptoms. The tweets related to this service were posted by patients, their families, and their beloved ones, such as the following tweet, which describes the suffering caused by cancer disease.

"وصف احد الأطباء معاناته مرضى السرطان قائلاً كذئب جائع ينهش في لحم وعظم انسان."

"One of the doctors described the suffering of cancer patients. He said, cancer is like a hungry wolf that eats the meat and bone of a cancer patient."

Another tweet was posted by a patient. It highlights the physical pain associated with treatment.

"... الكيماوي أصعب من المرض نفسه، مهما تكلمت عمري ما اعرف اعبر عن كمية الالم اللي الكيماوي ببسويه.....  
حاجه جواك تحرقك وتحرق روحك لدرجة حتى ملايسك ما تتحملها من الحرارة ما في مسكنات ولا ادويه....."

"... chemotherapy is more difficult than the disease itself, whatever I talk I will not be able to express how much pain that chemotherapy causes ... something inside you burns you and your soul to the point where even your clothes can't stand the heat. There are no painkillers or medicines ... ."

The tenth service is Hope and Optimism. Hope and optimism are among the most important factors that make a person live a happy and a healthy life. Many people are exposed to major psychological pressures and crises that may affect their lives, such as having the chronic cancer disease. A healthy psychological state is essential for the success of any treatment and recovery, therefore, providing psychological support is an important part of treatment. Hope, optimism, and patience are among the components of a healthy psychological condition. Helping cancer patients to be patient, willing, optimistic, and hopeful can be achieved through conversations, sharing experiences, and positive thoughts. Twitter is one of the social platforms where people can share their stories and experiences, influence and inspire other people. Some tweets were posted by patients who are fighting the disease, spreading hope, inspiring other people, and being a source of optimism. The tweets related to this topic were posted by patients, their families, and other stakeholders. For example, the following tweet found in our dataset is related to this cluster.

"... مجرد الاستماع إلى قصته خلال مراحل محاربته للمرض تجعلك تعيد النظر في الكثير من الأمور في حياتك لتري الحياة بمنظور آخر..."

"... just listening to his story during the stages of his fight against the disease makes you reconsider many things in your life to see life from another perspective ... ."

Some tweets were posted by people who survived cancer and shared their successful experiences. The following excerpt of a tweet is an example of this cluster.

".... هزمت السرطان 4 مرات. ومجبر على أن أكون قوياً...."

"... I defeated cancer four times and I'm forced to be strong ... ."

#### 4.5. Socioeconomic Sustainability

Now we discuss the fourth macro-service Socioeconomic Sustainability. It involves the tweets and clusters that are related to developing better strategies to fight cancer in the country. It includes four services; the first service (the eleventh overall) is Government Support. The Saudi government is making a lot of efforts to confront cancer disease. These are represented in the provided medical services e.g., (healthcare centers, modern medical devices, free diagnosis and treatment, awareness, psychological support, call centers, and smart applications for healthcare systems). For example, the Ministry of Health has provided various services e.g., “Mawid” website and smart application for facilitating booking medical appointments, and a call center service that is available 24/7 on (telephone number 937) for normal and emergency health services to the patients across the Kingdom. “Shefaa Platform” is another example of the health services supported by the government. It facilitates the provision of medical services, medical devices, and medicines to the needy and the emergency cases for those who cannot obtain treatment in health facilities. It contributes to the governance of charitable treatment in the Kingdom by verifying the Absher platform and the Council of Health Insurance. People used Twitter to facilitate communication with the community. For example, some individuals use Twitter to accelerate receiving personal support (financial, treatment, or other). They usually attach links (e.g., for payment) from reliable platforms such as Shefaa. For example, the following tweet, was found in our dataset.

“سيدة .. تعاني من تدهور حالتها الصحية إثر وجود ورم سرطاني في الثدي .. مما يستدعي ضرورة التدخل العاجل بالعلاج الكيميائي اللازم قبل وصولها لمرحلة حرجية يصعب فيها التحكم بالورم.”

“... a woman suffers from a deteriorating health condition due to the presence of a cancerous tumor in the breast ... which calls for urgent intervention with the necessary chemotherapy before it reaches a critical stage in which the tumor is difficult to control.”

The government, individuals, healthcare institutions, and other stakeholders also used Twitter to announce new services, and share information and experiences on available services. This cluster represents the tweets related to government support and the healthcare services in Saudi Arabia. For example, the following tweet was posted by a doctor. It highlights some of the services provided by the Ministry of Health.

“سرطان الثدي يأتي بالترتيب الأول بين أكثر أنواع السرطان شيوعاً، عالمياً، وإقليمياً، ومحلياً، لا يوجد اعراض بالمرحلة المبكرة له ف من المهم الكشف المبكر عن طريق الماموغرام، للحجز عن طريق الرقم 937 او تطبيق ” موعد “. مصدر المعلومات: وزارة الصحة.” #سرطان\_الثدي

“Breast cancer ranks first among the most common types of cancer, globally, regionally, and locally. There are no symptoms in its early stages, so it is important to detect it early through a mammogram. For reservations, call 937 or the “Mawid” application. Source of information: Ministry of Health.” #Breast\_Cancer

The twelfth service is Socioeconomic and Operational Challenges (see Row 13, Table 3). It is characterized by the keyterms such as Patient, Treatment, Needs, To Receive, Ask, Surgery, Thousand, Bill, Case, Suffering, Number, and Please. This cluster represents the socioeconomic and operational challenges faced by patients and their families. For example, one of the challenges is the need for patients and their families to travel farther to specialized hospitals for diagnosis and treatment purposes. This involves difficulties from various perspectives. For example, from the financial perspective, it involves additional expenses for transportation and housing. Another challenge is blood shortage. Some patients need a frequent blood transfusion and sometimes it is difficult to find donors, especially in urgent cases. The following excerpt of a tweet is an example of this cluster.

“اخوتي مصابه بـلوكيميا ” سرطان الدم “. اللي عنده استطاعة محتاجين متبرعين بالدم بشكل عاجل،الفصيلة: O+ ، المكان: مستشفى الملك خالد الجامعي، رقم الملف: .....، الاسم: .....”

“My sister has leukemia. Those who can, she needs blood donors urgently, blood group: O+, location: King Khalid University Hospital, File No.: ... , Name: ... ”

The thirteenth service is Charity Organizations. There are various charity organizations in the country, for example, Zahra Association, Sanad Children's Cancer Support Association, the Saudi Cancer Society, and The Civil Society Association for Cancer Care (Basma). These charity organizations are non-profit associations whose goal is to support cancer patients. They aim to provide social, financial, and psychological care and support for combating cancer. Their programs include, among others, encouraging new patients to contact other patients through conducting meetings and exchanging stories and experiences. For example, the Saudi Cancer Society holds workshops for patients and their families under the supervision of psychiatrists. These organizations may also participate in cancer awareness campaigns. For example, the following tweets shows that some organizations have participated and people are thanking and appreciating their efforts.

"كل الشكر موصول لجمعية... على ما بذلته من جهود في حملة سرطان الثدي التي أقيمت بالأمس بتاريخ ١٧-١٠-٢٠٢١..."

"All thanks go to [organization] for its efforts in the breast cancer campaign that was held yesterday on 17-10-2021 ..."

The fourteenth service is Financial Support (Row 15, Table 3). It is represented by the keyterms such as; Treatment, Bill, Pinching, Tried, We Can't, and Appeal. This cluster represents the financial issues faced by cancer patients and their needs for financial support. The Saudi government provides free treatment services and cancer patients usually take this option. There is a waiting time (usually long queues) for this free treatment service. However, there are some cases where a cancer patient needs a quick treatment procedure (e.g., urgent operation) and cannot wait for the free service. Therefore, they go for private treatment, which is usually very costly. Some patients who are unable to afford the treatment costs use Twitter to raise their needs and request financial support.

"أب... يعاني من ورم في البروستات... سبب له آلام حادة ويحتاج إلى التدخل الجراحي بشكل عاجل، ولا يملك القدرة المادية على تحمل تكاليف العملية."

"A father ... suffers from a prostate tumor ... which caused him severe pain and needs urgent surgical intervention, and he does not have the financial ability to bear the costs of the surgery."

Furthermore, cancer disease can severely affect patients' financial state for some reasons such as inability to work or loss of job due to health status. These situations create a need for financial support for some cancer patients.

"... أناشدكم... مريضه سرطان ابغى سداد الفاتورة زوجي متوفى... وذلك مصاريف علاجي من السرطان الراتب 3500 اجار شقة مترام."

"... I appeal to you, ... a cancer patient, I want to pay the bill. My husband is deceased. ... This is a treatment expense for cancer. My salary is 3500 SR. Apartment rent is accumulating."

"المساهمة بالترويج كمساهمتك في دفع الفاتورة، ام مصابه بالسرطان...، المتبقي 32800، تكفون يا اهل الخير."

"Contribute by retweeting as your contribution to paying the bill. Mother has cancer ... , Remaining is 32,800 SR, please good people."

#### 4.6. Information Availability

We now discuss the services associated with the fifth macro-service, Information Availability, which means that information is conveniently available to all stakeholders including patients, the families of the patients, healthcare organizations, and government. Information availability is a key strategy in combating cancer for all stakeholders. It includes three services. Breast Cancer Awareness (fifteenth service created from merging Topic 9 and 20) includes the keyterms; Campaign, Association, Center, Awareness, Lecture, and Examination. According to Breast Cancer Organization, breast cancer is the most common cancer across the globe [60]. It is the number one type of cancer among the women globally, as well as in Saudi Arabia [61]. Therefore, breast cancer has gained more attention

than other kinds of cancer types and it is also evident from our detected clusters. The aim of breast cancer awareness programs is to raise public awareness of breast cancer, the importance of early detection, the causes, and the prevention. One of the strategies for disseminating awareness about breast cancer was launching the annual campaign, “Breast Cancer Awareness Month”. In October 2021, many initiatives for supporting breast cancer awareness were discussed on Twitter under several hashtags such as, “#اكتوبر\_الوردي” (pink October), “#الشهر\_العالمي\_لسرطان\_الثدي” (breast cancer awareness month), and “#افحصي\_الآن” (check-up now). The campaign was supported by medical institutions (e.g., hospitals, and healthcare centers), charity organizations (e.g., Ehsan), and other stakeholders. The following is an excerpt from a tweet that was posted on Oct 3, 2021, by the Public Health in Taif city.

*“تم اليوم،...، تدشين حملته سرطان الثدي، والتي تستمر طيلة شهر أكتوبر.”*

*“Today, . . . , the breast cancer campaign was started, which will continue throughout the month of October.”*

Another similar tweet highlights the activation of breast cancer awareness campaign in Sabya governorate in Jazan region.

*“مركز صحي... بالتعاون مع جميعه... تفعيل حملته شهر أكتوبر للتوعية بسرطان الثدي.”*

*“[Health Center] in cooperation with [association] activating the October campaign to raise awareness of breast cancer.”*

The sixteenth service is Awareness Campaigns. It is characterized by keywords including; Account, Question, Retweet, Reply, Hashtag, Tweet, and Campaign. These key terms belong to Topic 10; see Figure 5. Table 5 gives the English translation of the Arabic keywords in the figure. Awareness campaigns are very important as they raise awareness of cancer types, causes, prevention, and treatment. Moreover, these campaigns can encourage early detection, support people with cancer, encourage cancer control programs, and enable communications and sharing of experiences between patients. Awareness campaigns could be conducted by any healthcare stakeholders though usually these campaigns are carried out by healthcare providers. The campaigns involve various activities for disseminating information such as lectures, posters, brochures, leaflets, and competitions. Awareness campaigns may also involve raising Twitter hashtags and sharing on Twitter posts containing educational content. For example, we detected various hashtags that were raised supporting cancer awareness campaigns including “#اليوم\_العالمي\_للسرطان” (world cancer day), “#سرطان\_البروستات” (Prostate cancer), “#سرطان\_الدم” (Blood Cancer), “#الشرقية\_وردية\_١٣” (Al-Sharqiya Pink 13), and others. The tweets related to cancer awareness campaigns were posted by doctors, hospitals, healthcare centers, and organizations. For example, the following tweet was posted by a hospital in Makkah.

*“اليوم يبدأ #الشهر\_العالمي\_لسرطان\_الثدي... ، إلى كل بطلة تحارب  
تذكري دائماً أنك وبالإيمان والعزيمة.”*

*“Today starts the #International\_Breast\_Cancer\_Month . . . to every hero who fights, always remember that you are with faith and determination.”*

Furthermore, a tweet was posted under (Al-Sharqiya Pink 13) hashtag, which was launched for supporting an event under the breast cancer awareness campaign. The event was organized by the Saudi Cancer Society in the Al-Sharqiya region. It involved various activities including the event of “Sharqiya Pink Marathon”. It aimed to support the global trend in raising awareness of the disease and the importance of early detection as one of the factors influencing the stages of treatment, which makes the difference in combating and curing this disease.

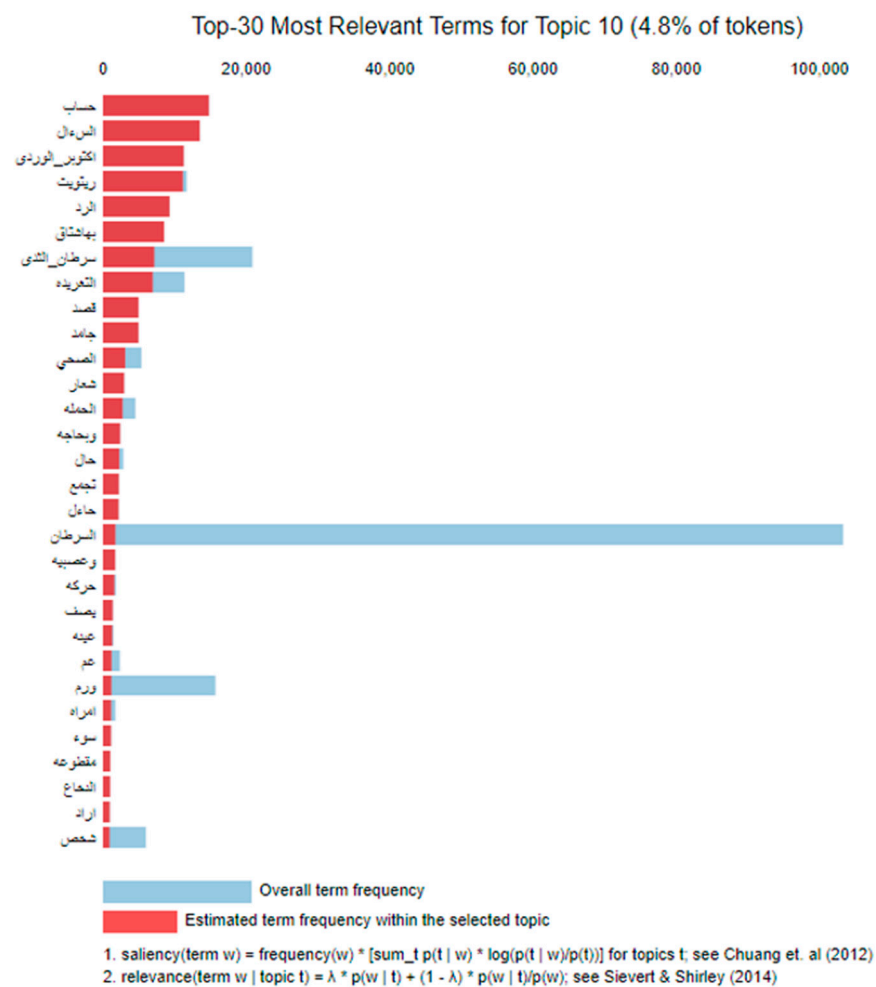


Figure 5. The Top 30 Most Relevant Terms for Topic 10.

Table 5. The keywords and their English translation for Figure 5.

Keywords	حساب، السؤال، اكتوبر_الوردي، ريتويت، الرد، بهشتاق، سرطان_الثدي، التغريدة، قصد، جامد، الصحي، شعار، الحملة، ويحاجه، حال، تجمع، حائل، السرطان، وعصبيه، حركه، يصف، عينه، عم، ورم، امراه، سوء، مقطوعه، النخاع، اراد
Translation	Account, The Question, Pink October, Retweet, Reply, Hashtag, Breast Cancer, Tweet, Intent, Solid, Healthy, Logo, The Campaign, Need, Case, Gathering, Hail city, Cancer, Nervous, Movement, Describe, A Sample, Uncle, Tumor, Woman, Bad, Cut Off, Marrow, Want

The seventeenth service is Questionnaires and Competitions (Row 19, Table 3). It is represented by keywords such as; Follow, Participation, Share, Event, Member, Answer, Thousands, and more. Information availability is not only about creating awareness for care receivers; it is about gaining knowledge from the care revivers and public for scientific research purposes and improving healthcare. One of the strategies to enhance societal awareness and educate people about cancer about its risks and the ways to prevent it is the development of certain questionnaires and competition exercises among the public. Questionnaires and competitions can also be used to obtain information for scientific research purposes. They can be organized and supported by various organizations, associations, and electronic platforms. For example, one of the charitable organizations for cancer patients, held various online competitions on Twitter and provided financial incentives for stimulating awareness among the members of the community. An example tweet is given below.

"#مسابقة بسمه/السؤال الثالث: ماهي أسباب الإصابة بسرطان الثدي؟....."

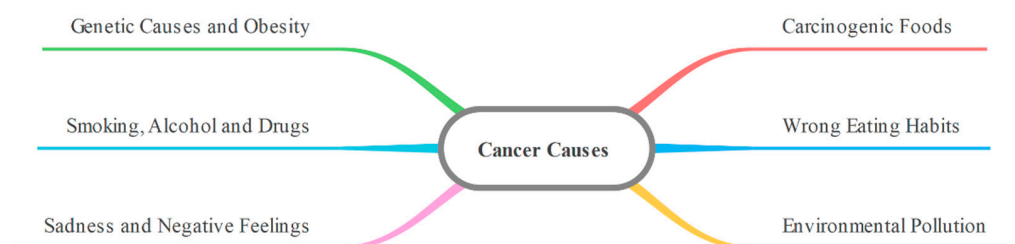
"#Basma competition/Question three: What are the causes of breast cancer? ... .."

## 5. Extended Services

This section discusses the 42 extended services that we discovered by using knowledge gained from and extending the initial discovery process. These services are discovered by *topical searching* over the dataset, the *topics* are some of the services discovered in the initial discovery process. We used four *topical searches* (Causes, Symptoms, Prevention, and Stakeholders) and detect 42 additional services; these are discussed in Section 5.1 to Section 5.4, respectively. We call them extended services since these are discovered by using the knowledge gained from the initial service discovery process.

### 5.1. Cancer Causes

We intended to investigate the usability of our approach further and hence we applied clustering on our dataset with a focus on cancer causes. This approach allowed us to detect a list of cancer causes from the collected tweets. We used certain keyterms to find more about the causes. These terms are some important Arabic terms that refer to the causes of cancer for example "مسبب", "يسبب", "تسبب". We categorized the causes into several groups based on the content of the tweets. The causes of cancer that we have detected include Carcinogenic Foods, Wrong Eating Habits, Genetic Causes, Obesity, Smoking, and others. The full taxonomy for the causes of cancer that we extracted from Twitter data is presented in Figure 6.



**Figure 6.** Cancer Causes Taxonomy Extracted from Twitter Data.

Carcinogenic Foods are foods that contain chemicals such as preservatives, taste enhancers, artificial flavors, and juices to which preservatives are added. Eating a lot of carcinogenic foods may cause cancer. It is advised to commit to a healthy diet that contains nutrients and avoid harmful foods. The following are sample tweets that elaborate on this. They were posted by patients, doctors, and other stakeholders. "... they found out that my nephew has leukemia, the reason is Indomie and soft drinks ... ". "Be careful, juices and soft drinks cause colon cancer for young people and various chronic diseases. Get to know them and avoid them immediately."

There are some Wrong Eating Habits that increase the risk of cancer. These include some cooking methods, for example, deep-frying, cooking food more than necessary at high temperatures, which generate chemicals like heterocyclic amines, and the use of plastic for food preservation. The following tweets were posted by patients and other stakeholders.

"لا تغل الطعام المطبوخ أكثر من اللازم: ينتج عن كثرة غليان هذا الطعام تولد مواد تنشط الخلايا السرطانية وتزيد من تكاثرها، لذا يكفي غليه مرة واحدة بعد طهيها، ويتم تناوله بعد طهيها مباشرة."

"Do not boil cooked food more than necessary: frequent boiling of this food generates substances that activate cancer cells and increase their proliferation, so it is sufficient to boil it once after cooking it, and eat it immediately after cooking."

Negative Feelings and emotions, such as anger, sadness, and hate, negatively affect and weaken the immune system. Therefore, it is necessary to control emotions to decrease



the risk of getting cancer. The following tweets represent the cluster. They were posted by patients and other stakeholders.

"مشاعرك السلبية تتكدس في عضو معين من جسدك إذا لم تواجهها وتحررها وتتحول لا سمح الله إلى مرض مزمن  
...الأورام = جروح دفينية السرطان = حزن عميق....."

"Your negative feelings accumulate in a certain part of your body, if you do not confront them and release them it may turn into a chronic disease. . . . Tumors= buried wounds. Cancer = deep sadness . . . ."

Smoking, Alcohol, and Drugs are among the most important causes of cancer that negatively affect the health of the body. The following are sample tweets, which elaborate on that.

" ماهي مسببات #السرطان؟ الجواب: عوامل متعددة، أبرزها: السمنة، التدخين والخمر، الوراثة، أعراض جانبية للأدوية، حالة الالتهاب."

"What are the causes of #cancer? Answer: Multiple factors, mostly: obesity, smoking and alcohol, genetics, side effects of medications, and inflammation."

Genetics is an important factor in diseases in general and cancer in particular. Some cancer diseases have a genetic predisposition, and some are generated from genetic mutations. Obesity is another risk factor for cancer. It changes hormones secretion, affects the immune system, and causes chronic cellular inflammation and these can lead to cancer. The following tweet is posted by a doctor's account, which elaborates on that, "What are the causes of #cancer? Answer: Multiple factors, mostly: obesity, smoking and alcohol, genetics, side effects of medications, and inflammation."

Environmental Pollution is one of the most important causes of cancer. It includes air pollution with toxic gases or factory smoke, and water pollution with bacteria and rust. In addition, exposure to harmful rays increases the chances of developing cancer. The following tweets represent the cluster. These tweets were posted by patients, the public, and other stakeholders.

".... مرض السرطان ..... اغلب المدن تعاني منه..... من المصانع والمواد الغذائية الملونة."

"... cancer disease ... most cities suffer from it ... caused by factories and colored foodstuffs."

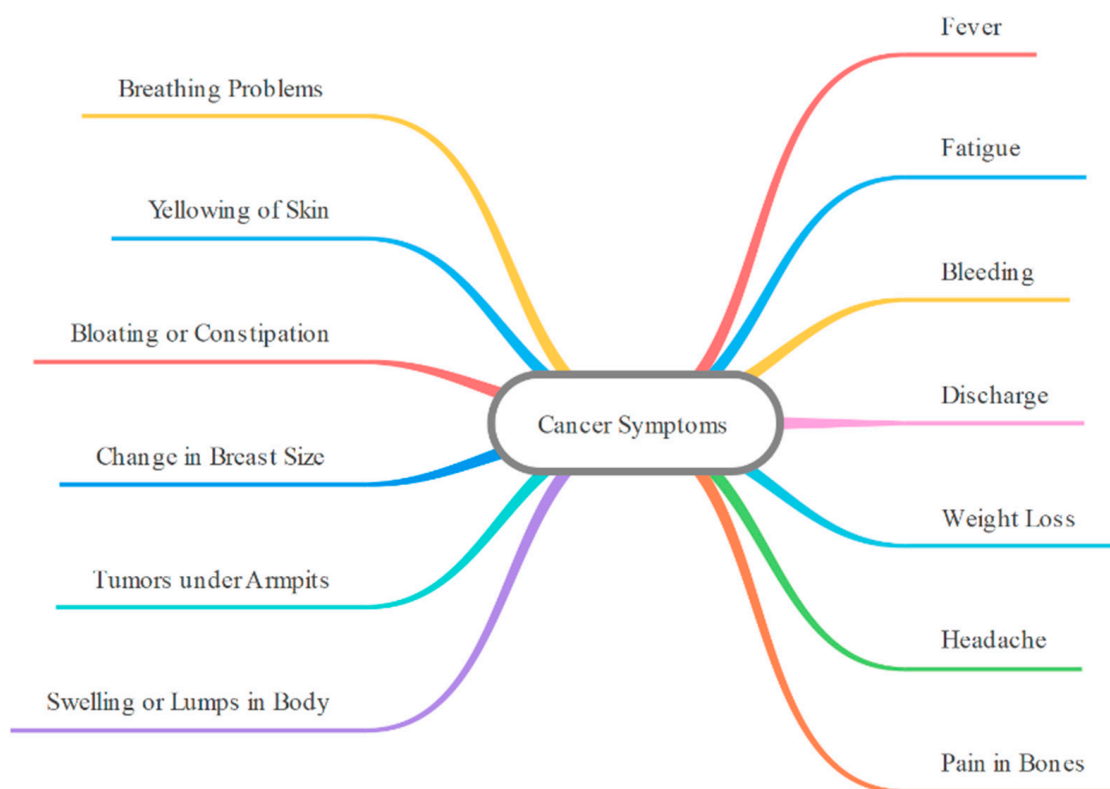
## 5.2. Cancer Symptoms

As part of our investigation of usability of our approach, we applied clustering to our dataset with an emphasis on cancer symptoms. From the collected tweets, we were able to derive a list of cancer symptoms. We used some search queries with certain keyterms. These keyterms refer to the cancer symptoms such as "أعراض", "الاعراض", "علامات". We obtained important cancer symptoms, for example; Fever, Fatigue, Bleeding, Discharge, Weight Loss, Headaches, and others. The full taxonomy for the symptoms of cancer that we extracted from Twitter data is presented in Figure 7. The tweets were posted by stakeholders of healthcare. For example, below is a tweet that shows the symptoms associated with ovarian cancer.

"سرطان المبيض هو أحد أنواع السرطانات.....أعراضه: ألم أو تورم أو شعور بالضغط في منطقة البطن والحوض، نزيف من المهبل في غير وقت الدورة، إفرازات من المهبل قد تصحب بدم، انتفاخات أو إمساك."

"Ovarian cancer is one of cancer diseases . . . . its symptoms: pain, swelling or a feeling of pressure in the abdomen and pelvis area, bleeding from the vagina outside the time of the period, secretions from the vagina that may be accompanied by blood, and swelling or constipation."

The following tweet highlights breast cancer symptoms such as Swelling or Lumps and the change in Size of the Breast: “Another clear symptom of breast cancer is the emergence of swellings under one or both armpits, as a result of swelling of the lymph tissues there. A noticeable change in the size of the breast is unjustified, as it swells greatly. A noticeable shrinkage or retraction of the nipple inward, which is one of the important symptoms of breast cancer.” Weight Loss symptom is shown in the following tweet. “... I noticed that I lost weight for no reason, and after I took an endoscopy, they found that I have a colon tumor ...”. The following tweets shows some symptoms such as; Yellowing of the Skin, Pain in the Bones, Headaches, and Breathing Problems. “In advanced stages, when the cancer spreads, and becomes a malignant tumor, some symptoms may appear such as: yellowing of the skin, pain in the bones, headache, and breathing problems.”

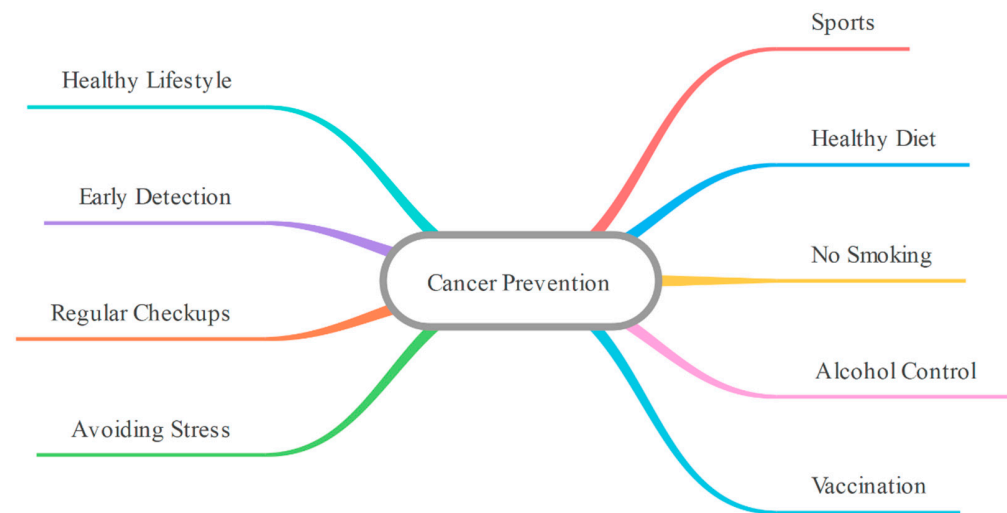


**Figure 7.** Cancer Symptoms Taxonomy Extracted from Twitter Data.

### 5.3. Cancer Prevention

We applied clustering on our dataset with a focus on cancer prevention. By analyzing the collected tweets, we were able to identify a list of cancer prevention methods. We used certain terms for instance “الوقاية”, “وقاية”, “للوقاية”, “يقيك”. We identified important ways to prevent cancer, which include; Sports, Healthy Diet, Healthy Lifestyle, No Smoking, etc. The full taxonomy that we extracted from Twitter data is presented in Figure 8. The following tweets were posted by the different accounts as patients, hospitals, doctors, and other stakeholders. For example, the following tweets, found in our dataset, are related to; Sports, Healthy Lifestyle, No Smoking, Alcohol Control, Vaccination, Avoiding Stress, Drinking Water, and Sleeping Early. “Healthy nutrition, sports, drinking water, sleeping early, staying away from stress ... etc. are all important factors for the prevention of all diseases and cancers in particular. Let your health always be your priority for a safe future from diseases, God willing.” #Pink\_October, #Breast\_Cancer\_Awareness. “Can you prevent cancer? Researches showed that up to 50% of cancer cases can be prevented through healthy lifestyle. You make choices every day that affect your health. Prevention and early detection are more important than ever: don’t use tobacco, eat a healthy diet, be active and maintain a healthy weight, refrain from drinking alcohol,

avoid risky behaviors, vaccination (human papillomavirus and hepatitis), know your family medical history and get regular checkups.” Furthermore, the following tweet highlights Smoking and cancer. “One of the simplest ways to prevent cancer is to stop smoking.”



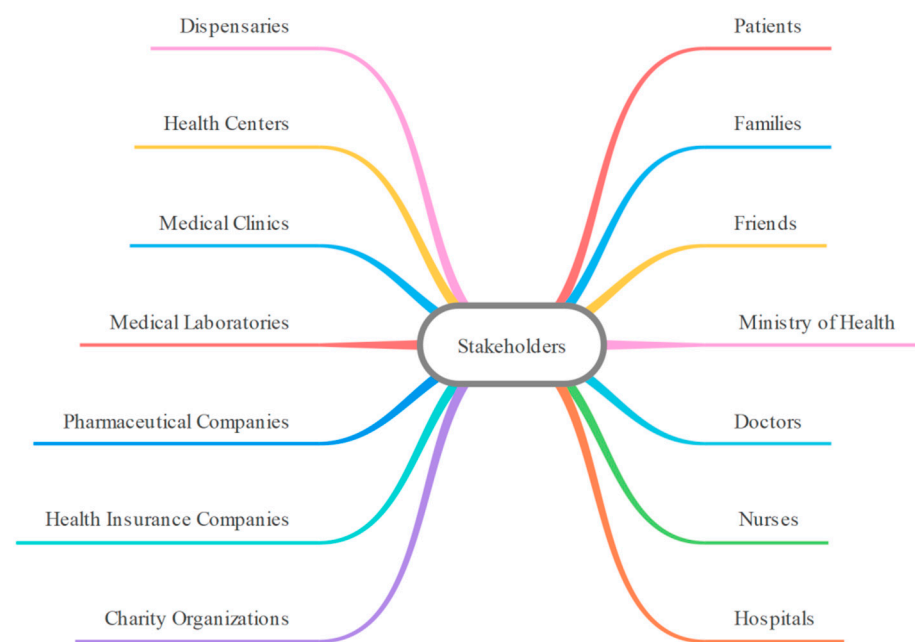
**Figure 8.** Cancer Prevention Taxonomy Extracted from Twitter Data.

#### 5.4. Healthcare Stakeholders

We extracted a cluster for stakeholders in healthcare. We used some search queries with specific keyterms, such as “مريض”, “مريضه”, “عائلي”, “اصدقائي”, “وزارة الصحة”, “دكتور”, “صيدلية”, “تحاليل”, “مختبر”, “عيادة”, “مركز صحي”, “مستوصف”, “مستشفى”, “ممرضه”, “ممرض”, “علاج”, “تأمين”, “جمعية”. A stakeholder is an individual or a group with an interest in the subject. In the healthcare sector, individuals or organizations with an interest in healthcare decisions are referred to as stakeholders. The stakeholders that we have detected include; Patients, Family, Friends, and others. The full taxonomy that we extracted from Twitter data is presented in Figure 9.

The following tweet represents the stakeholder, Patients: “I am a cancer patient . . . ”. The following tweet shows the Family and Friends stakeholders: “I lived my experience with cancer with full strength, . . . Everyone around me shared my pain. I found constant support from my husband, my family, and my friends . . . ” Below are some tweets related to the Ministry of Health stakeholder: “The Ministry of Health invites people with chronic diseases and weak immunity and those who receive immunosuppressive drugs such as . . . cancer patients . . . to quickly go to the vaccination centers to receive the vaccine in order to avoid any complications that may threaten their lives if they get the virus.” The following tweet represents the Doctors stakeholder. “ . . . Doctor . . . Can you advise me about the Zometa injection? I am a breast cancer patient . . . ” Hospitals as a stakeholder are shown in the following tweets: “[Hospital . . . ] succeeds in applying advanced technology to treat eye cancer in children”.

We had collected one or more tweets for each stakeholder type and added them to the paper, however removed from the final version for brevity. We will be happy to provide the reader with samples of additional tweets if required.



**Figure 9.** Taxonomy for Stakeholders in the Healthcare Sector Extracted from Twitter Data.

## 6. Discussion

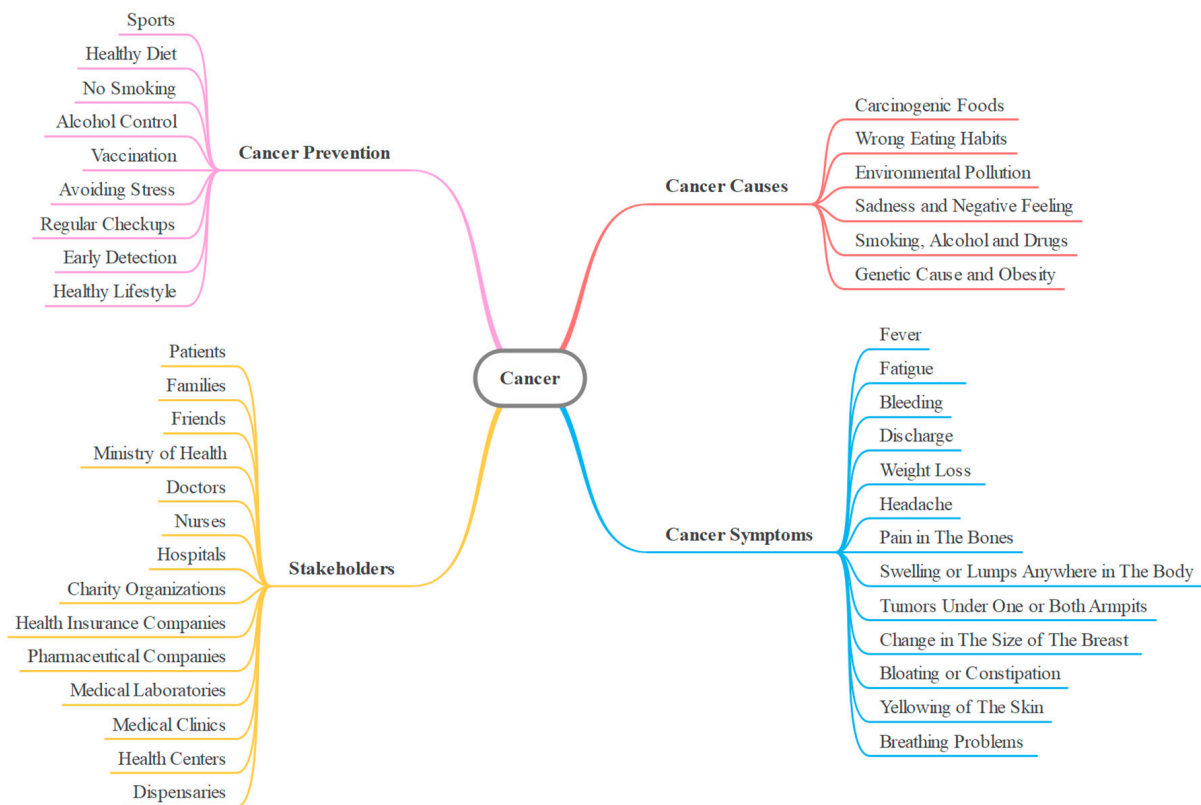
In this paper, we proposed a data-driven artificial intelligence (AI) based approach called Musawah to automatically detect and identify healthcare services that can be developed or co-created by various stakeholders using social media analysis. We detected 17 services using unsupervised machine learning from Twitter data using the Latent Dirichlet Allocation algorithm (LDA) and group them into five macro-services.

The first macro-service, Prevention, involves actions that reduce the chance of getting cancer (e.g., maintaining a healthy lifestyle, taking vaccines, and avoiding cancer-causing substances) and it also involves strategies and procedures that prevent cancer from progressing to advanced stages (e.g., early detection, and regular screening tests). The services in this category include; Early Diagnosis, Prevention and Control, Causes, Screening, and Symptoms. The second macro-service, Treatment, describes the cancer treatment services and increases the level of understanding about options for cancer treatment, which can be used to treat cancer in different ways depending on the patient's condition and stage of the disease. The service in this macro-service includes Chemo and Radiation Therapy and Surgical Therapy.

The third macro-service is Psychological Support. It includes services, namely, Spiritual Support, Suffering, Hope and Optimism. Cancer diseases affect the patients on a physical, spiritual, and emotional level. Therefore, psychological support is very important for cancer patients. It helps the patients to handle the difficulties and overcome challenges, which is an essential factor for a successful treatment and patient's survival, even the survival of the patient's families. The fourth macro-service (Socioeconomic Sustainability) involves the tweets and clusters that are related to developing better strategies to fight cancer in the country. The services in this category include Government Support, Socioeconomic and Operational Challenges, Charity Organizations, Financial Support. The fifth macro-service, Information Availability, which means that information is conveniently available to all stakeholders including patients, the families of the patients, healthcare organizations, and government. Information availability is a key strategy in combating cancer for all stakeholders. It includes the; Breast Cancer Awareness, Awareness Campaigns, and Questionnaires and Competitions services.

We also introduced in this paper our methodology to develop extended services, called so as these are discovered by using the knowledge gained from the initial service discovery process. We discovered 42 extended services by topical searching over the dataset,

where the topics are selected from some of the services discovered in the initial discovery process. We used four topical searches in this paper—Causes, Symptoms, Prevention, and Stakeholders—however, any number of services can be used to create services in a specific subdomain. The 42 services are shown in the extended taxonomy of cancer provided in Figure 10.



**Figure 10.** Extended Cancer Taxonomy: A Taxonomy of Extended Cancer Services.

We mentioned in Section 1 that the aim of this paper is to investigate the role of big data analytics over social media to automatically detect needs and value propositions that could be nurtured and turned into service co-creation processes through engaged participation over social and digital media, eventually co-creating services. The co-created services could be based on values that are not necessarily materialistic but are driven by equity, altruism, community strengthening, innovation, and social cohesion. We provide below evidence that big data analytics over social media can be used to detect value propositions and co-create values through engaged participation. Specifically, we provide a few examples below to elaborate the concept of automatic discovery, development, deployment of co-created services that we introduced in this paper. We take the example of the discovered service ‘Hope and Optimism’ that was discussed in Section 4. Hope and optimism are among the most important factors that make a person live a happy and healthy life. Many people are exposed to major psychological pressures and crises that may affect their lives, such as having chronic cancer disease. A healthy psychological state is essential for the success of any treatment and recovery, therefore, providing psychological support is an important part of treatment. Hope, optimism, and patience are among the components of a healthy psychological condition. Helping cancer patients to be patient, willing, optimistic, and hopeful can be through conversations, sharing experiences, and positive thoughts. People share their stories and experiences of pain and suffering, joy and happiness, struggles and overcoming diseases, stories of inspiration, optimism, and more. As we have mentioned earlier, service co-creation is enabled through engaged participation of need and value propositions. In co-creation terminology, both parties, a customer and company create



value for each other and this value could be money, product, service, or anything of value for the parties involved. Therefore, the value could be anything materialistic or otherwise, such as a desire for equity, the altruistic nature of a person or party, and the desire to strengthen communities, bring innovation, or social cohesion. Now imagine an open service and value healthcare system based on freely available information made available by the public, the government, third and fourth sectors, or others. Imagine someone in need of Hope and Optimism shares the need on social media and this is matched by a value proposition, someone with a specific experience, knowledge, skillset, etc. These two people or parties come together to exchange value for value such as the need for hope and optimism exchanged for a monetary value or desire for equity and social cohesion, and in the process, they co-create services that can be used as it is in many other cases or adapted for different situations.

Another example is Suffering, which can be looked at as a service where the person who suffers has a need and another person or party can provide value to the one suffering directly (patient) or indirectly (patient's family) from cancer. Suffering due to cancer can be physical, psychological, or financial, and these sufferings are well known. For example, peoples' finances can be negatively affected due to cancer treatment for themselves or their family members and cause a person to suffer. Suffering can also be due to operational reasons, such as the hardships related to the need for patients and their families to travel farther to specialized hospitals for diagnosis and treatment purposes. This involves difficulties from various perspectives. For example, from the financial perspective, it involves additional expenses for transportation and housing. Another suffering could be due to blood shortage. Some patients need a frequent blood transfusion and sometimes it is difficult to find donors, especially in urgent cases. Using the previous example of Hope and Optimism service, in the case of Suffering, two people or parties may come together to exchange value for value, such as the need for relief from Suffering due to financial or operational reasons, or needing blood, exchanged for a monetary value or desire for equity and social cohesion, and in the process co-create a service.

Similarly, Information Availability could promote the value co-creation process for healthcare services by allowing healthcare providers to interact with patients, and patients with similar health situations for sharing information about conditions, symptoms, and treatments. This promotes collective learning and emotional support, helps individuals make decisions, spreads health knowledge, increases health literacy, reduces the patient's anxiety level, and provides personalized self-management. Socioeconomic Sustainability could relate to supporting healthcare services and value co-creation by using the available healthcare resources to reduce doctor and hospital visits and thereby reduce costs and wasted resources. Service co-creation could also relate to identifying the needs of a patient and meeting those needs to gain patient satisfaction by using feedback from the patient's or other patients' experiences acquired from Twitter data. The Treatment service could relate to increasing the level of understanding about options for cancer treatment available to a specific individual or group of people and these treatments may not be the mainstream methods rather based on well-known herbal medicine or lifestyle changes depending on the patient's condition and stage of the disease.

The Musawah approach and system proposed in this paper make important and significant theoretical and practical contributions to the literature. While there were a few works (e.g., [29]) on cancer-related social media studies in the English language, these were focused on one or another type of cancer such as breast or skin or lung cancer (see Section 2.2). There was only work on social media-related cancer studies and it was focused on studying chemotherapy misconceptions among Twitter users [41]. These works clearly are different from our work, as we have looked at the whole cancer space in this paper, not just a type of cancer or chemotherapy misconceptions. There are also some works, such as [54], that have investigated the use of social media by people and public health organizations to communicate in new ways for various mutually beneficial activities. While these and other works have investigated the use of social media for value co-creation



by various actors communicating with each other and sharing information, none of the works in the literature have proposed to discover healthcare services automatically from social media. This work, therefore, contributes a novel approach of using social media and AI to discover, develop, and exchange healthcare or cancer services. We demonstrated the potential of our approach by discovering over fifty services in several dimensions of cancer disease including cancer causes, symptoms, prevention, treatment, socioeconomic sustainability, and stakeholders. The approach can be systemized by creating open service and value healthcare systems based on freely available information made available by the public, the government, third and fourth sectors, or others, allowing new forms of preventions, cures, treatments, and support structures to be discovered and exchanged between various parties.

The work provides evidence to support the general literature on data-driven smart cities research [1,2,62,63] and reinforces that policy and action on smart cities, healthcare, and other sectors should be supported with data and that social and digital media provides a convenient and important source of such data [64,65]. The topics detected by our system clearly show the possibility and benefits of our tool, allowing from social media public and stakeholder conversations the discovery and grasp of important dimensions of the cancer disease in Saudi Arabia (partly applicable internationally), such as public concerns, patient requirements, solutions for problems related directly and broadly to cancer such as cancer treatment, financial difficulties, psychological ordeals and traumas, operational challenges for the families of cancer patients and other information. The parameters and information learned through the tool can benefit the public in many ways, such as through being a source of information and allowing government and various institutions to improve their services, organizational approaches, foci, etc. The work is distinct from others in terms of the dataset, methodology and design, our innovative approach of using AI to discover services, and our findings.

As regards the potential impact of our work, we believe the impact could be foundational, colossal, and far-reaching. On a personal note, the authors casually discussed the findings of this research and mentioned it to their families, friends, and networks, and this unconscious activity generated a lot of impacts. Some people got motivated to change their lifestyle to become healthier, others decided to be careful in using cosmetics to minimize the cancer risk. An important impact was related to a close friend who found that he had some symptoms, which were mentioned in the paper and he decided to do some tests. To his surprise, the reason for the symptom he had was an infection that is a major cause of a certain kind of cancer. These impact examples are the result of casually mentioning the finding of our paper to a small network of authors. How about coordinated dissemination of these findings to larger networks and how about creating incubators for service co-creation using our proposed ideas and tool? Such efforts on a large scale, incorporated into data-driven digitally connected systems, where information is pushed to consumers based on certain environmental parameters (e.g., a person searching for some symptoms or their smart devices reporting a specific pattern) can lead to lifestyle motivations, early interventions, diagnosis, and ultimately reduction in diseases and healthy and sustainable societies.

## 7. Conclusions and Future Work

Poor healthcare systems affect individuals, societies, the planet, and economies. Smartization of cities and societies has the potential to unite individuals and nations towards sustainability and thereby also improve healthcare and other systems as it requires an engagement with our environments, an analysis of them, and for us to make sustainable decisions regulated by TBL. Healthcare systems are a major contributor to the deterioration of our planet's environment due to high energy usage, disposable supplies, etc. This environmental damage causes health problems that need to be medically treated while these treatments in turn further damage the environment, and thus a chain reaction occurs. There is a growing demand for the prevention of diseases and for reducing the need for disease monitoring, screening, and treatment to a minimum. This in turn requires an open space

for innovations and dynamic interactions between healthcare stakeholders to understand and provide solutions, services, information and resource supply chains, and community support structures for timely interventions and disease prevention and progression.

To address these challenges, this paper proposed a data-driven artificial intelligence (AI) based approach called Musawah to automatically detect and identify healthcare services that can be developed or co-created by various stakeholders using social media analysis. The case study focused on cancer disease in Saudi Arabia using Twitter data analytics in the Arabic language. Specifically, we discovered 17 services using unsupervised machine learning from Twitter data using the Latent Dirichlet Allocation algorithm (LDA) and grouped them into five macro-services; Prevention, Treatment, Psychological Support, Socioeconomic Sustainability, and Information Availability. We also illustrated the possibility to find additional services by topical searching over the dataset using four topical searches (Causes, Symptoms, Prevention, and Stakeholders) and detected 42 additional services. We developed a software tool from scratch for this work. The tool implements a complete machine learning pipeline. The dataset we used contains over 1.3 million tweets collected from September to November 2021.

Our work makes several contributions to the literature including the Musawah approach for creating services proposed in this paper and the developed tool and techniques to detect cancer-related services. The work builds on our extensive research in social media analysis in English and other languages on topics in different sectors. The paper has focused on cancer-related services from Twitter data in Arabic, however, the proposed approach broadly can be used for any disease or purpose and in any language.

The idea of developing cancer-related services from social media analysis is motivated by the well-known science that social media facilitates value co-creation through stakeholder interactions allowing value creation or creation of new services. Therefore, the ideas of services and co-creation proposed in this paper per se are not novel, and the novelty lies in the proposed approach of systematically and dynamically developing an ecosystem of services for a particular disease that can be co-created on-the-fly by communities of stakeholders interacting over social media using automatic value and service detection. The services can be the needs of people seeking prevention or treatment such that a gap in the market can be detected for stakeholders to develop services for economic, social, or other types of remuneration and rewards. The approach that we present, while basic at this stage, has great potential and will allow further investigation and development of novel and innovative ways of developing healthcare services and systems, is green in terms of environmental, social, and economic sustainability, with lower costs allowing lifestyle changes, self-directed disease management, and managed care.

We believe systemizing this approach could lead to a revolution in sustainable healthcare, driven by communities co-creating social values with the exchange of services for services that are not necessarily materialistic, but driven by equity, altruism, community strengthening, innovation, and social cohesion. Open service and value healthcare systems based on freely available information can revolutionize healthcare in manners similar to the open-source revolution by using information made available by the public, the government, third and fourth sectors, or others, allowing new forms of preventions, cures, treatments, and support structures.

**Author Contributions:** Conceptualization, N.A., S.A. and R.M.; methodology, N.A., S.A. and R.M.; software, N.A. and S.A.; validation, N.A., S.A. and R.M.; formal analysis, N.A., S.A. and R.M.; investigation, N.A., S.A., R.M., A.A. (Ahmed Alzahrani), A.A. (Aiiad Albeshri) and I.K.; resources, R.M., I.K., A.A. (Aiiad Albeshri) and A.A. (Ahmed Alzahrani); data curation, N.A. and S.A.; writing—original draft preparation, N.A., S.A. and R.M.; writing—review and editing, R.M., A.A. (Ahmed Alzahrani), A.A. (Aiiad Albeshri) and I.K.; visualization, N.A. and S.A.; supervision, R.M. and A.A. (Ahmed Alzahrani); project administration, R.M., A.A. (Aiiad Albeshri) and I.K.; funding acquisition, R.M., A.A. (Aiiad Albeshri) and I.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors acknowledge with thanks the technical and financial support from the Dean-ship of Scientific Research (DSR) at the King Abdulaziz University (KAU), Jeddah, Saudi Arabia, under Grant No. RG-6-611-40. The experiments reported in this paper were performed on the Aziz supercomputer at KAU.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data is acquired from Twitter under their terms and conditions. Anyone can request the same data from Twitter.

**Acknowledgments:** The work carried out in this paper is supported by the HPC Center at the King Abdulaziz University. The training and software development work reported in this paper was carried out on the Aziz supercomputer.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Yigitcanlar, T.; Butler, L.; Windle, E.; DeSouza, K.C.; Mehmood, R.; Corchado, J.M. Can Building “Artificially Intelligent Cities” Safeguard Humanity from Natural Disasters, Pandemics, and Other Catastrophes? An Urban Scholar’s Perspective. *Sensors* **2020**, *20*, 2988. [CrossRef] [PubMed]
2. Mehmood, R.; See, S.; Katib, I.; Chlamtac, I. *Smart Infrastructure and Applications: Foundations for Smarter Cities and Societies*; Springer International Publishing, Springer Nature Switzerland AG: Cham, Switzerland, 2020.
3. Komeily, A.; Srinivasan, R.S. Sustainability in Smart Cities: Balancing Social, Economic, Environmental, and Institutional Aspects of Urban Life. In *Smart Cities: Foundations, Principles, and Applications*; Wiley: Hoboken, NJ, USA, 2017; pp. 503–534.
4. Alotaibi, S.; Mehmood, R.; Katib, I.; Rana, O.; Albeshri, A. Sehaa: A Big Data Analytics Tool for Healthcare Symptoms and Diseases Detection Using Twitter, Apache Spark, and Machine Learning. *Appl. Sci.* **2020**, *10*, 1398. [CrossRef]
5. Yigitcanlar, T.; Mehmood, R.; Corchado, J.M. Green Artificial Intelligence: Towards an Efficient, Sustainable and Equitable Technology for Smart Cities and Futures. *Sustainability* **2021**, *13*, 8952. [CrossRef]
6. Mehmood, R.; Alam, F.; Albogami, N.N.; Katib, I.; Albeshri, A.; Altowaijri, S.M. UTiLearn: A Personalised Ubiquitous Teaching and Learning System for Smart Societies. *IEEE Access* **2017**, *5*, 2615–2635. [CrossRef]
7. Vergunst, F.; Berry, H.L.; Rugkåsa, J.; Burns, T.; Molodynski, A.; Maughan, D.L. Applying the triple bottom line of sustainability to healthcare research—A feasibility study. *Int. J. Qual. Health Care* **2020**, *32*, 48–53. [CrossRef]
8. Eckelman, M.J.; Sherman, J. Environmental Impacts of the U.S. Health Care System and Effects on Public Health. *PLoS ONE* **2016**, *11*, e0157014. [CrossRef]
9. Naylor, C.; Appleby, J. Sustainable Health and Social Care: Connecting Environmental and Financial Performance, p. 28. Available online: <https://www.kingsfund.org.uk/publications/sustainable-health-and-social-care> (accessed on 12 January 2022).
10. Lenzen, M.; Malik, A.; Li, M.; Fry, J.; Weisz, H.; Pichler, P.-P.; Chaves, L.S.M.; Capon, A.; Pencheon, D. The environmental footprint of health care: A global assessment. *Lancet Planet. Health* **2020**, *4*, e271–e279. [CrossRef]
11. Sherman, E. U.S. Health Care Spending Hit \$3.65 Trillion in 2018. *Fortune* **2019**. Available online: <https://fortune.com/2019/02/21/us-health-care-costs-2/> (accessed on 12 June 2020).
12. AlOmari, E.; Katib, I.; Albeshri, A.; Mehmood, R. COVID-19: Detecting Government Pandemic Measures and Public Concerns from Twitter Arabic Data Using Distributed Machine Learning. *Int. J. Environ. Res. Public Health* **2021**, *18*, 282. [CrossRef]
13. Ruger, J.P. Health and social justice. *Lancet* **2004**, *364*, 1075–1080. [CrossRef]
14. Jameton, A.; Pierce, J. Environment and health: 8. Sustainable health care and emerging ethical responsibilities. *Can. Med Assoc. J.* **2001**, *164*, 365–369. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC80732/> (accessed on 12 January 2022).
15. Global Social Media Statistics Research Summary. 2022. Available online: <https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/> (accessed on 9 January 2022).
16. McColl-Kennedy, J.R.; Vargo, S.L.; Dagger, T.S.; Sweeney, J.C.; van Kasteren, Y. Health Care Customer Value Cocreation Practice Styles. *J. Serv. Res.* **2012**, *15*, 370–389. [CrossRef]
17. Alomari, E.; Katib, I.; Albeshri, A.; Yigitcanlar, T.; Mehmood, R. Iktishaf+: A Big Data Tool with Automatic Labeling for Road Traffic Social Sensing and Event Detection Using Distributed Machine Learning. *Sensors* **2021**, *21*, 2993. [CrossRef] [PubMed]
18. Yigitcanlar, T.; Kankanamge, N.; Regona, M.; Maldonado, A.; Rowan, B.; Ryu, A.; DeSouza, K.C.; Corchado, J.M.; Mehmood, R.; Li, R.Y.M. Artificial Intelligence Technologies and Related Urban Planning and Development Concepts: How Are They Perceived and Utilized in Australia? *J. Open Innov. Technol. Mark. Complex.* **2020**, *6*, 187. [CrossRef]
19. Suma, S.; Mehmood, R.; Albeshri, A. Automatic Detection and Validation of Smart City Events Using HPC and Apache Spark Platforms. In *Smart Infrastructure and Applications: Foundations for Smarter Cities and Societies*; Springer: Cham, Switzerland, 2020; pp. 55–78.
20. Suma, S.; Mehmood, R.; Albugami, N.; Katib, I.; Albeshri, A. Enabling Next Generation Logistics and Planning for Smarter Societies. *Procedia Comput. Sci.* **2017**, *109*, 1122–1127. [CrossRef]

21. Yigitcanlar, T.; Regona, M.; Kankanamge, N.; Mehmood, R.; D'Costa, J.; Lindsay, S.; Nelson, S.; Brhane, A. Detecting Natural Hazard-Related Disaster Impacts with Social Media Analytics: The Case of Australian States and Territories. *Sustainability* **2022**, *14*, 810. [\[CrossRef\]](#)
22. Alsulami, M.; Mehmood, R. Sentiment Analysis Model for Arabic Tweets to Detect Users' Opinions about Government Services in Saudi Arabia: Ministry of Education as a case study. In *Al Yamamah Information and Communication Technology Forum*; Al Yamamah University: Riyadh, Saudi Arabia, 2018; pp. 1–8. Available online: [https://www.researchgate.net/publication/324000226\\_Sentiment\\_Analysis\\_Model\\_for\\_Arabic\\_Tweets\\_to\\_Detect\\_Users\T1\textquoteright\\_Opinions\\_about\\_Government\\_Services\\_in\\_Saudi\\_Arabia\\_Ministry\\_of\\_Education\\_as\\_a\\_case\\_study](https://www.researchgate.net/publication/324000226_Sentiment_Analysis_Model_for_Arabic_Tweets_to_Detect_Users\T1\textquoteright_Opinions_about_Government_Services_in_Saudi_Arabia_Ministry_of_Education_as_a_case_study) (accessed on 7 March 2020).
23. Alomari, E.; Mehmood, R.; Katib, I. Sentiment Analysis of Arabic Tweets for Road Traffic Congestion and Event Detection. In *Smart Infrastructure and Applications: Foundations for Smarter Cities and Societies*; Springer International Publishing: Cham, Switzerland, 2020; pp. 37–54.
24. Alomari, E.; Katib, I.; Mehmood, R. Iktishaf: A Big Data Road-Traffic Event Detection Tool Using Twitter and Spark Machine Learning. *Mob. Netw. Appl.* **2020**, 1–16. [\[CrossRef\]](#)
25. Leite, H.; Hodgkinson, I.R. Telemedicine co-design and value co-creation in public health care. *Aust. J. Public Adm.* **2021**, *80*, 300–323. [\[CrossRef\]](#)
26. Jahanbin, K.; Rahmanian, F.; Rahmanian, V.; Jahromi, A.S. Application of Twitter and web news mining in infectious disease surveillance systems and prospects for public health. *GMS Hyg. Infect. Control* **2019**, *14*, Doc19. [\[CrossRef\]](#)
27. Wang, J.; Wei, L. Fear and Hope, Bitter and Sweet: Emotion Sharing of Cancer Community on Twitter. *Soc. Media Soc.* **2020**, *6*. [\[CrossRef\]](#)
28. Crannell, W.C.; Clark, E.; Jones, C.; James, T.A.; Moore, J. A pattern-matched Twitter analysis of US cancer-patient sentiments. *J. Surg. Res.* **2016**, *206*, 536–542. [\[CrossRef\]](#)
29. Modave, F.; Zhao, Y.; Krieger, J.; He, Z.; Guo, Y.; Huo, J.; Prosperi, M.; Bian, J. Understanding Perceptions and Attitudes in Breast Cancer Discussions on Twitter. *Stud. Health Technol. Inform.* **2019**, *264*, 1293–1297. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Morzy, T.; Härder, T.; Wrembel, R. Advances in Databases and Information Systems. In *Advances in Intelligent Systems and Computing*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 186. [\[CrossRef\]](#)
31. Rasool, M.J.; Brar, A.S.; Kang, H.S. Risk prediction of breast cancer from real time streaming health data using machine learning. *Int. Res. J. Mod. Eng. Technol. Sci.* **2020**, *2*, 409–418. [\[CrossRef\]](#)
32. Diddi, P.; Lundy, L.K. Organizational Twitter Use: Content Analysis of Tweets during Breast Cancer Awareness Month. *J. Health Commun.* **2017**, *22*, 243–253. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Silva, C.V.; Jayasinghe, D.; Janda, M. What Can Twitter Tell Us about Skin Cancer Communication and Prevention on Social Media? *Dermatology* **2020**, *236*, 81–89. [\[CrossRef\]](#)
34. Nguyen, J.; Gilbert, L.; Priede, L.; Heckman, C. The Reach of the “Don’t Fry Day” Twitter Campaign: Content Analysis. *JMIR Dermatol.* **2019**, *2*, e14137. [\[CrossRef\]](#)
35. Sedrak, M.S.; Cohen, R.B.; Merchant, R.M.; Schapira, M.M. Cancer Communication in the Social Media Age. *JAMA Oncol.* **2016**, *2*, 822–823. [\[CrossRef\]](#)
36. Sutton, J.; Vos, S.C.; Olson, M.K.; Woods, C.; Cohen, E.L.; Ben Gibson, C.; Phillips, N.E.; Studts, J.L.; Eberth, J.M.; Butts, C.T. Lung Cancer Messages on Twitter: Content Analysis and Evaluation. *J. Am. Coll. Radiol.* **2018**, *15*, 210–217. [\[CrossRef\]](#)
37. Sedrak, M.S.; Salgia, M.M.; Bergerot, C.D.; Ashing-Giwa, K.; Cotta, B.N.; Adashek, J.J.; Dizman, N.; Wong, A.R.; Pal, S.K.; Bergerot, P.G. Examining Public Communication About Kidney Cancer on Twitter. *JCO Clin. Cancer Inform.* **2019**, *3*, 1–6. [\[CrossRef\]](#)
38. Nejad, M.Y.; Delghandi, M.S.; Bali, A.O.; Hosseinzadeh, M. Using Twitter to raise the profile of childhood cancer awareness month. *Netw. Model. Anal. Health Inform. Bioinform.* **2019**, *9*, 3. [\[CrossRef\]](#)
39. Meeking, K. Patients' experiences of radiotherapy: Insights from Twitter. *Radiography* **2020**, *26*, e146–e151. [\[CrossRef\]](#)
40. Zhang, L.; Hall, M.; Bastola, D. Utilizing Twitter data for analysis of chemotherapy. *Int. J. Med. Inform.* **2018**, *120*, 92–100. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Alghamdi, A.; Abumelha, K.; Allarakia, J.; Al-Shehri, A. Conversations and Misconceptions about Chemotherapy in Arabic Tweets: Content Analysis. *J. Med. Internet. Res.* **2020**, *22*, e13979. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Allahyari, M.; Pouriyeh, S.; Kochut, K.; Reza, H. A Knowledge-based Topic Modeling Approach for Automatic Topic Labeling. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 335–349. [\[CrossRef\]](#)
43. Lau, J.H.; Grieser, K.; Newman, D.; Baldwin, T. Automatic labelling of topic models. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, Oregon, Portland, 19–24 June 2011; pp. 1536–1545.
44. Wan, X.; Wang, T.; Erk, K.; Smith, N.A. Automatic Labeling of Topic Models Using Text Summaries. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics ACL 2016—Long Papers, Berlin, Germany, 7–12 August 2016; Volume 4, pp. 2297–2305. [\[CrossRef\]](#)
45. He, D.; Wang, M.; Khattak, A.M.; Zhang, L.; Gao, W.; Mateen, A. Automatic Labeling of Topic Models Using Graph-Based Ranking. *IEEE Access* **2019**, *7*, 131593–131608. [\[CrossRef\]](#)
46. Suadaa, L.H.; Purwarianti, A. Combination of Latent Dirichlet Allocation (LDA) and Term Frequency-Inverse Cluster Frequency (TFxICF) in Indonesian text clustering with labeling. In Proceedings of the 2016 4th International Conference on Information and Communication Technology (ICoICT), Bandung, Indonesia, 25–27 May 2016; Volume 4. [\[CrossRef\]](#)

47. Hidayat, T.; Suhardi; Kurniawan, N.B. Smart city service system engineering based on microservices architecture: Case study: Government of tangerang city. In Proceedings of the 2017 International Conference on ICT For Smart Society (ICISS), Tangerang, Indonesia, 18–19 September 2017; Volume 2018, pp. 1–7. [\[CrossRef\]](#)
48. D’Aniello, G.; Gaeta, M.; Orciuoli, F.; Sansonetti, G.; Sorgente, F. Knowledge-Based Smart City Service System. *Electronics* **2020**, *9*, 965. [\[CrossRef\]](#)
49. Yang, Z.; Kankanhalli, A. Innovation in Government Services: The Case of Open Data. *IFIP Adv. Inf. Commun. Technol.* **2013**, *402*, 644–651. [\[CrossRef\]](#)
50. Lopes, A.J.; Pineda, R. Service Systems Engineering Applications. *Procedia Comput. Sci.* **2013**, *16*, 678–687. [\[CrossRef\]](#)
51. Lockheed Martin Corporation. *Guide to the Systems Engineering Body of Knowledge—G2SEBoK*; A Publication of The International Council of Systems Engineering: San Diego, CA, USA, 2004.
52. Demirkan, H.; Spohrer, J.C.; Krishna, V. *Service Systems Implementation*; Springer Science & Business Media: Boston, MA, USA, 2011. [\[CrossRef\]](#)
53. Kalliokoski, P.; Andersson, G.; Salminen, V.; Hemilä, J. BestServ Feasibility Study: Final report. Teknologiatieto Teknologia. 2003. Available online: <https://cris.vtt.fi/en/publications/bestserv-feasibility-study-final-report>. (accessed on 7 March 2020).
54. Musso, M.; Pinna, R.; Melis, G.; Carrus, P. How Social Media Platform can Support Value Cocreation Activities in Healthcare. *Undefined* **2018**, 536–555.
55. Lusch, R.F.; Vargo, S.L. Historical Perspectives on Service-Dominant Logic. *Serv. Log. Mark.* **2021**, 47–60. [\[CrossRef\]](#)
56. Vargo, S.L.; Lusch, R.F. Inversions of service-dominant logic. *Mark. Theory* **2014**, *14*, 239–248. [\[CrossRef\]](#)
57. Lusch, R.F.; Vargo, S.L. The service-dominant logic of marketing: Reactions, reflections, and refinements. *Mark. Theory* **2006**, *6*, 281–288. [\[CrossRef\]](#)
58. Mortenson, M.J.; Vidgen, R. A computational literature review of the technology acceptance model. *Int. J. Inf. Manag.* **2016**, *36*, 1248–1259. [\[CrossRef\]](#)
59. Al Jalim, F. Wasel Electronic Newspaper. Removal of Cancerous Tumor in King Faisal Complex in Taif. Available online: <https://wasel-news.com/?p=490731/> (accessed on 20 November 2021).
60. Breastcancer.org. Breast Cancer Facts and Statistics. Available online: <https://www.breastcancer.org/facts-statistics/> (accessed on 25 November 2021).
61. Ministry of Health. Breast Cancer Awareness Month. Available online: <https://www.moh.gov.sa/HealthAwareness/HealthDay/2018/Pages/HealthDay-2018-10-01-31.aspx/> (accessed on 28 November 2021).
62. Bibri, S.E. Data-driven smart sustainable cities of the future: An evidence synthesis approach to a comprehensive state-of-the-art literature review. *Sustain. Futures* **2021**, *3*, 100047. [\[CrossRef\]](#)
63. Yigitcanlar, T.; Corchado, J.M.; Mehmood, R.; Li, R.Y.M.; Mossberger, K.; Desouza, K. Responsible Urban Innovation with Local Government Artificial Intelligence (AI): A Conceptual Framework and Research Agenda. *J. Open Innov. Technol. Mark. Complex.* **2021**, *7*, 71. [\[CrossRef\]](#)
64. Arfat, Y.; Suma, S.; Mehmood, R.; Albeshri, A. Parallel Shortest Path Big Data Graph Computations of US Road Network Using Apache Spark: Survey, Architecture, and Evaluation. In *Smart Infrastructure and Applications Foundations for Smarter Cities and Societies*; Springer: Cham, Switzerland, 2020; pp. 185–214.
65. Arfat, Y.; Usman, S.; Mehmood, R.; Katib, I. Big data for smart infrastructure design: Opportunities and challenges. In *Smart Infrastructure and Applications Foundations for Smarter Cities and Societies*; Springer: Cham, Switzerland, 2020; pp. 491–518.