On a Multidimensional Research Framework for Managing the Complex Disease Ecosystems

(Completed Research)

Shastri L Nimmagadda                                      Torsten Reiners

Neel Mani                                                Lincoln C Wood

Abstract

Cancer is a deadly complex disease. Poorly aligned institutions, their data sources, and doctor-patient disengagement motivate us, strengthening collaborative cancer research. The research explores informatics solutions, by collating various cancer-linked data attributes and interconnecting them with spatial-temporal dimensions. The study is aimed at developing a methodological framework and investigating the open source empirical cancer data that involve diverse human ecosystems. We propose a Multidimensional Cancer Research Framework (MUCARF), responding to challenges of reporting, documenting and collaborating the data sources that characterize different cancer ailments including their worldwide causalities. Information system artefacts built based on cancer-domain ontologies, with similar and dissimilar attributes are integrated into MUCARF to dig the diagnostic cancer metadata views in new knowledge domains. Metadata model that replicates the design and development of MUCARF is a useful analytic therapy to examine and mitigate high rates of cancers and their types as preventive care and cancer disease management.

Keywords: Health informatics, multidimensional modelling, cancer research framework, types of cancers, populations.
Introduction

The early detection of cancer may have an impact on its timely preventive management. The success may lie in gathering, analytics, and understanding of available medical and non-medical data. It can increase its awareness by analyzing cancer cases and evaluating research objectives with a direction to improve the quality of life and the average lifespan of cancer patients. Technological advancement, such as the Big Data tools can support the ideas of managing medical interventions. Early intervention by identifying key symptoms and risks with a positive attitude can play an increasing role and awareness on mitigation of cancers. Even so medical record keeping, if improved significantly, can add value to early intervention and cancer management. Linking of clinical databases and connecting patients and medical practitioners are added attributes.

Data organizations, processing and communication are re-engineering challenges of information systems that deal with healthcare schemes (Nimmagadda and Rudra 2017). Cancer Informatics is an evolving science in healthcare information systems for which data re-engineering needs innovative methodological solutions. The authors focus on developing informatics solutions addressing issues on documenting, organizing and reporting cancer cases and their associated ailments including causality. Typical attributes include types of cancers, based on age, poverty, gender, race, income inequality, demography and environment. We examine these attributes to find rationale and analyse a variety of cancer systems that represent composite dimensions. We develop an ontology-based multidimensional data warehousing and -mining methodology with foci on data connectivity and integration processes. The Multidimensional Cancer Research Framework (MUCARF) is proposed to make connectivity between domains and manage to report, document and organize types of cancers for mass populations worldwide. Various similarity, comparison and referential attributes can be depicted (Bertossi and Milani 2017; Nimmagadda and Rudra 2017) in the form of map, plot and other diagnostic data views extracted from warehouse metadata to precisely understand the ailments, based on gender, age, race, inequality, poverty, demography and geography and other environmental events.

Storing a large number of cancer records is recognized including medical notes, based on the interactions between patients and medical practitioners. The data warehouse, a technology concept emerges with tasks of information processing, access and timely update of records. We aim at managing the acquisition of cancer-related data, documenting, processing and delivering the quality of information needed by medical practitioners through an integrated framework. The framework is a paradigm to conceptualize multidimensional data modelling artefacts. It is capable of articulating multiple attributes of complex decease ecosystems that can compromise and circumspect cancer management. In addition to rigor on data mining and visualization, an added focus is on the interpretation of data views on bodied diagnosis, with follow-up prescriptive medications. The proposed methodology is a robust back-end application for both online patient-doctor consultations and offline e-healthcare management. The investigation embodies the results of ontologically structured cancer metadata to enable mining of domain knowledge and its interpretation. Further, we analyse the data relevant to cancer cases based on geography and demography and how the analysis can facilitate understand the cancer cases, which demography and geography they are more prevalent. Our research focus is reporting, documenting, processing (through framework) to deliver new knowledge from cancer metadata structures.

Literature Survey and Research Gaps

for integration. Ngafeeson (2014) identifies various healthcare information system solutions and opportunities. Heidy et al. (2012) explore challenges of exploring a large amount of data of multiple dimensions in diverse domains to find new and useful information. The authors develop SEER database to establish connections between different types of cancers using dimensional hierarchy and cuboid structure tools. Krishnaiah et al. (2013) provide various data mining tools to explore large cases of lung cancer using classification and predict the diagnosis. Bertossi and Milani (2017) develop context-based data quality assessment techniques using various types of multidimensional data models and their linked database structures. Multiple cancer based multidimensional ontologies have been described for exploring categorical relations, dimensional rules and constraints. Qualitative text mining tools generate association rule mining between cancer thematic terms (Xie et al. 2010). The literature review has helped us in examining the research purpose and challenges.

Issues and Challenges

In recent years, more information is found on cancer awareness on the Internet (Finney and Mirzaman 2018), in particular, their linked ailments, treatments and causalities. In spite of technological innovations, we still find gaps between researchers, medical practitioners and datasets in multiple domains that affect the healthcare organizational alignments and the connectivity between doctors and patients. Further, the data exhibit heterogeneity and multidimensionality challenges with reporting, documenting and modelling, thus impeding the structuring and integration processes. However, there is a large number of online and offline (big) data sources available to enable decision support technologies to explore and exploit new knowledge about cancers (World Cancer Statistics 2013; Roser and Ritchie 2018). In spite of theoretical aspects of the multidimensional ontologies including their contexts and scenarios are discussed in Bertossi and Milani (2017), yet the study lacks the implication and practical application without factual and empirical data analysis. According to the cancer data organization, processing and timely sharing of diagnostic information by medical practitioners is still a challenge. In addition to connecting various components, prevention, screening, diagnosis, treatment, survivorship and caring terminally-ill are other related cancer processes. It is uncertain if the research gaps can be resolved unless the medical practitioners have the full control of data and information on cancer cases and their contexts among large demographics (World Cancer Statistics 2013). The structure, behavior and interactions of cancer systems may not be explicit unless cancer information systems are appropriately interpreted at various stages of model implementations. Without computational, cognitive and social perspectives, by mere systematic analysis of terminally-ill diseases, the medical data analysts cannot conclude the cancer cases.

The research focus is on existing historical cancer cases scattered in different geographies. There are many causalities occurred because of deadly cancer disease. We acquire, document and process the cases using information systems’ driven framework, MUCARF. The key benefit of the framework MUCARF is to segregate and categorize different types of cancers based on gender, age, demography and geography. We intend to demonstrate how the framework can deliver the processed new knowledge on cancer cases, based on gender, age, geography and demography. The campaign can bring cancer research awareness, based on demography and geographies, where cancer patients needed more attention and care.

Research Goals, Motivation and Significance

Though the cancer informatics is not superficial anymore, information engineering in the context of multidimensional cancer research has increased its awareness in modelling and integrating different types of cancers within integrated workflows (Miller et al. 2002). Uncovering the connectivity among different cancer-related information systems is critical. The current research is a demonstration of how different data-centric linked cancers can be best organized and documented online and or offline (Finney et al. 2018) in spatial-temporal dimensions, to deliver the knowledge in ontologically described warehoused repositories. Though previous researchers investigated volumes of cancer data, a systematic methodology using domain- and data-modelling, data warehousing and mining, visualization and interpretation artefacts, all in a canvass, is lacking within the current context. It is based on modelling data sources from multiple domains, integrating with different application scenarios to envisage new knowledge for effective interpretation. The cancer research using data modelling in spatial-temporal dimensions has not been the focus for medical practitioners, data managers and data analysts in large-scale cancer research projects (Miller et al. 2002).
Significance

The research has significance, by implementing collaborative cancer research framework in healthcare industries, detect cancer cases and their early intervention can be managed. Different types of cancers attack different parts of the human body, because of which reason, the model design considerations may be given due importance while articulating the schemas and implementing them in healthcare projects. In addition, an integrated methodological framework may have significance in leveraging various stages of cancers, for which, preventive measures, cancer campaigns and action plans may be strategized. The research designs corroborate with application domains as discussed in Neuman (2000). Due to complex nature of cancer descriptions, composited data organizations and hierarchies (Nimmagadda and Rudra 2017), detached data events in between investigative agencies can make the goals of the project difficult to achieve. Further, there is a huge demand for structured data and information in large healthcare enterprises (Wand 2000). Healthcare organizations have scopes of adopting new technologies and solving problems associated with cancers using multidimensional data sources. Data warehouse schemas (Sheta and Eldeen 2012) that depict multiple dimensions (combined with spatial-temporal) in various domains provide improvements and scopes of mining heterogeneous cancer data sources. The cancer research data integration, building knowledge-based unified models with metadata and meta-knowledge are significant research outcomes. The metadata and their descriptions are deliverables of the framework MUCARF for cancer researchers. The data views extracted from metadata structures are presented in various visualizations in a manner they are presentable for medical practitioners and researchers.

The Motivation of Cancer Research

The scope of the framework is to establish links between various domains, i.e., associated workflows that can connect the information system engineering process to multiple domain applications. Informatics can explore strong collaborations and links between entities and dimensions of data in support of cancer research. At times, the cancer cases are unreported in hugely populated regions. The informatics-based information systems can even alert cases where communities need action plans through medical professionals. Their connectivity is crucial in making life-threatening medical decisions. These scopes and opportunities have motivated us to develop new ideas on ontology-based data warehousing and mining of multidimensional and heterogeneous data sources (Hadzic and Chang 2005), that led us improved data structures and integrated multiple data schemas. The integrated framework promotes the development of the Cancer Digital Ecosystem (CDE), thus can emerge with digital cancer solutions in wider contexts. Because of the complexity and dynamically changing healthcare situations, digital solutions are in growing demand in integrated healthcare research projects. Data integration and understanding the connectivity among multiple information systems are paramount, motivating factors in developing cancer informatics solutions.

Contributions and Research Components

The framework is in line with the outcomes of constructs, models, methods and instantiation with combined research activities such as build, evaluate, theorize and justify as described in Neuman (2000). Considering various activities and outcomes, we intend to build the MUCARF research framework that is adaptable to any domain of cancer research. The framework consists of the following components:

1) Ontological structures and their integration in a warehouse environment.
2) Artefacts such as domain, data, schema, warehouse, data mining, data visualization and data interpretation articulations.
3) Description of entities, objects and dimension attributes in different ontological constructs and models.
4) Improved understanding of the healthcare systems’ knowledge and interpretation.
5) Values of models and methods in cancer research development including their implementations.

Research Goals in Support of Cancer Remedies

An ontology describes a common vocabulary for information system researchers and users, who wish to share information in different knowledge domains (Hadzic and Chang 2005; Mastrian and McGonigle 2017). In the current context, an ontology implies exploring relationships among various cancer data sources. In addition, data integration is another key motivation in developing shared ontologies, especially while integrating multiple domain
ontologies. It includes machine-interpretable definitions of basic classes (concepts) in different domains and their conceptualized relationships. Key goals of ontology-based domain-knowledge representation are described in (Nimmagadda and Rudra 2017). Based on the introduction, research gaps, issues and challenges, we frame research questions and objectives.

Research Questions and Objectives

1. Why do we need the multidimensional framework to tackle different types of cancers?
2. How can the framework work deliver meta-knowledge?
3. How do we evaluate the framework articulations to establish relationships between different attribute dimensions? What is the strength of relationships in between relative movements of two attribute variables?

Objectives: It is an information system driven framework, based on which we have designed the following research objectives:

1. To design and develop a framework, to acquire different types of cancer cases, document and process them in new knowledge domains
2. Deliver meta-knowledge from metadata structures that generated using the MUCARF
3. Evaluate the artefacts used in the framework and the new knowledge delivered by MUCARF.

Outcomes: The metadata and meta-knowledge in the form of data views deliverable to medical practitioners for new knowledge interpretation of cancer cases. Government healthcare departments and agencies, private and government medical practitioners including social welfare organizations are typical users of multidimensional cancer research framework.

Methodology – Modelling Types of Cancer Data Sources

The method provides the rigor on cancer research and its data science. As addressed in Research Question 2, we describe the methodology in the section. Empirical and observational research is adopted. A large amount of data instances are used to build constructs and models. Informatics paradigm explores for information that drives various constructs, models, methods and instantiation (Neuman 2000), and links different interactive schema architectures in an integrated framework, MUCARF. The concepts are made based on the existing observations to predict new insights of models and their instantiations. A huge range of problems exists in the contexts of cancer management with challenges how the information requirements necessitate the processing scenarios (Atkinson 2002, JNCL 2017 and Bray et al. 2013). The second challenge is how the conceptualization and contextualization (Wand 2000) can be applied to new cancer-driven engineering of information systems. Third challenge is to ascertain how cancer driven information systems can help solve the cancer management challenges with educative measures that can improve the quality of life and its span. Next, we discuss how cancer informatics can provide a range of solutions and opportunities:

<table>
<thead>
<tr>
<th>Country</th>
<th>P</th>
<th>A</th>
<th>T</th>
<th>D</th>
<th>R</th>
<th>G</th>
<th>PT</th>
<th>Dim</th>
<th>Fact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>5</td>
<td>4</td>
<td>30</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>1100</td>
<td>1150</td>
</tr>
<tr>
<td>India</td>
<td>4</td>
<td>4</td>
<td>32</td>
<td>6</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>1080</td>
<td>1090</td>
</tr>
<tr>
<td>Uganda</td>
<td>3</td>
<td>4</td>
<td>30</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>1000</td>
<td>900</td>
</tr>
<tr>
<td>China</td>
<td>5</td>
<td>4</td>
<td>34</td>
<td>9</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>1100</td>
<td>1150</td>
</tr>
<tr>
<td>USA</td>
<td>4</td>
<td>4</td>
<td>30</td>
<td>10</td>
<td>4</td>
<td>2</td>
<td>6</td>
<td>1150</td>
<td>1155</td>
</tr>
<tr>
<td>UK</td>
<td>3</td>
<td>4</td>
<td>28</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>1050</td>
<td>1050</td>
</tr>
<tr>
<td>Germany</td>
<td>5</td>
<td>4</td>
<td>25</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>1000</td>
<td>1010</td>
</tr>
<tr>
<td>France</td>
<td>3</td>
<td>3</td>
<td>20</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>1050</td>
<td>1055</td>
</tr>
<tr>
<td>Indonesia</td>
<td>2</td>
<td>4</td>
<td>28</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>900</td>
<td>910</td>
</tr>
</tbody>
</table>

Table 1. Empirical Data and Number of Tables Created in the Empirical Modelling
We utilize secondary datasets, published in open Web sources to build the multidimensional models (Roser and Ritchie 2018; Bray et al. 2018). As shown in Table 1, we consider representative data for empirical modelling, their instances acquired from populated, developed, war-torn and disease-prone countries to assess how different environments affected the cancer growth and its mitigation. We present data models using the empirical-observational research to bring out new insights of metadata models that describe cancers in new knowledge domains (Sidhu et al. 2005). Ontology, data warehousing, data mining, data visualization and data interpretation, all in a single canvas, have been a focus of large size data systems design and development in spatial-temporal dimensions. The data sources linked with multiple cancer cases are targeted for exploring their connectivity through integration procedures.

<table>
<thead>
<tr>
<th>Country</th>
<th>P: Population Growth Types</th>
<th>A: Age Types</th>
<th>T: Types of Cancers</th>
<th>D: Deaths (%)</th>
<th>R: Race</th>
<th>G: Gender</th>
<th>PT: Poverty Types</th>
<th>Dim: Number of Dimension Tables</th>
<th>Fact: Number of Fact Tables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nigeria</td>
<td>3 3 22 8 4 2 4</td>
<td></td>
<td></td>
<td>950</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iraq</td>
<td>2 3 20 6 4 2 5</td>
<td></td>
<td></td>
<td>900</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syria</td>
<td>2 4 20 6 3 2 4</td>
<td></td>
<td></td>
<td>800</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iran</td>
<td>3 3 22 5 3 2 4</td>
<td></td>
<td></td>
<td>650</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazil</td>
<td>3 3 20 6 4 2 5</td>
<td></td>
<td></td>
<td>950</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Zealand</td>
<td>5 3 21 7 3 2 5</td>
<td></td>
<td></td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Africa</td>
<td>3 4 20 7 4 2 4</td>
<td></td>
<td></td>
<td>1050</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C: Country; P: Population Growth Types; A: Age Types; T: Types of Cancers; D: Deaths (%); R: Race; G: Gender; PT: Poverty Types; Dim: Number of Dimension Tables; Fact: Number of Fact Tables

A generalized informatics workflow is given in Figure 1, with a description of various components, activities and functions. Data acquisition, modelling and analysis are critical components of the methodology. In each component, various activities and functions required for schema design, data mining, visualization and interpretation artefacts are discussed. Various data mining and interpretation techniques are proposed, but in the current context, using grapher solutions, we intend to build models that represent multidimensional bubble and scaler lines plots for data interpretation. New knowledge envisaged from data models depends on the interpretation of data views and prior domain knowledge of cancer occurrences and how they affected based on age, gender, and race and poverty attribute dimensions. As a part of the workflow, we make an effort to describe multiple domains and their data sources in spatial-temporal dimensions. As shown in Figure 1, the heterogeneity of data sources is interpreted based on domain knowledge for which the acquired data demand for proper documentation. Data qualities are checked for any inconsistencies, while mapping and modelling process. For each of the major stages (data acquisition, data

Figure 1. A Generalized Informatics Workflow
modelling and information analysis), the data are checked for quality and accordingly accommodated in the integrated framework, subsequently tested for data mining, visualization and interpretation processes.

Figure 2. Cancer Data Facts – an Integrated Schema Architecture – Connecting Domains through Ontology Modelling

It is worth mentioning that the activities and their data events are connectable in a manner; the integrated process can use or reuse or conceptualize attributes through mapping and modelling. The interconnections in between dimensions and fact tables are shown in Figure 2; the data relationships emerge in between attribute dimensions and schemas when conceptualization and contextualization features apply in the modelling. It is a composite schema architecture, comprising of multiple sub-schemas from diverse domains, meant for integration.

Figure 3. MUCARF, a Framework Articulation for Managing connectivity between Cancer-guided Artefacts and their Attribute Dimensions (b) Cuboid Ecosystem Metadata – Slicing and Dicing
Figure 3 is a description of various stages of building constructs, models and methods in which volumes of fact instances are stored and carried to the articulations for generating a cuboid cancer-ecosystem metadata (Figure 3b).

**Data Mining and Visualization**

As described in Research Question 3, we provide various tools for evaluating the framework articulations including information solutions offered are described. Warehoused multidimensional data cubes are effective for data mining, visualization and interpretation (Xie et al. 2010; Doi 2007). Imposition of flexible business rules and constraints is significant to the integration process. Significant scopes are likely, for analyzing the heterogeneity, multidimensionality and granularity of data sources. Semantic, schematic, syntactic and system ambiguities (Messina and Urso BioGrakn 2017) that arise during structuring and integration process are expected to resolve ambiguities during the description of domain ontologies and their modelling. In addition, faster operational and user responses that minimize the operational costs, reusability, search, reliability, flexibility, maintainability, integrity and security, are other significant issues of cancer domain research (Heidy et al. 2012). Several attributes are considered in data modelling. The age, poverty, inequality, environment affected areas in spatial-temporal dimensions are core attributes. As a part of delivering quality informatics solutions, we compute data cubes that represent multidimensional 2D tables’ and their extensions (Pujari 2001). A 3D data cube is a set of similarly structured 2-D tables stacked on top of one another (Figure 3). Note that in general, data cubes allow for higher dimensions. The data cubes (Song et al. 2017) consist of a series of 3D cubes, allowing the visualization of multidimensional spatial, periodic and geometry-based attributes. Data cube interpretation is popular in many applications, in which multidimensionality of structured cubes is representative of multidimensional arrays.

We explore informatics solutions from MUCARF. In total, we found 36 types of cancers, of which 22 types of cancers are investigated, as they are widespread worldwide. The most common cancers prevalent in the world have been further narrowed down to 15 countries (including different types, such as populous, developed, disease-prone and war-torn countries). The criteria are based on data attributes and instances from most populated and developed countries, high disease-prone areas and extreme war-torn countries. The aim is to categorize the cancer systems among these cancers whether they possess any data trends, patterns and correlations. The data warehousing and mining paradigm with several contextual views of cuboid cancer data are provided in Heidy et al. (2012). They discuss cancer results extracted from warehoused metadata and its data views through Tool-Drill Roll (TDR) and OLAP tools for Surveillance, Epidemiology, and End Results (SEER) from the USA. The scope of research is multidimensional in largescale MUCARF, implementable globally.

A lung cancer prediction framework developed by Heidy et al. (2012) uses several data mining classification tools. Data mining of warehoused multidimensional data sources has been described in Heidy et al. (2012). Using the cuboid metadata as represented in Figure 3, we have extracted several data views for visualization and interpretation, focusing on selected attribute dimensions. We demonstrate their significance in analyzing and transmitting the results through online methods to patients as offered by medical professionals in Ngafeeson (2014). In addition, the research framework evaluates the articulations and validates the data models so as to make available to medical practitioners and government agencies and implement the research outcomes. In addition, the study investigates different propositions that needed in cancer management and establishes its role in clinical interroations. The findings emphasize how informatics can overcome cancer evaluation barriers and provide quality services to cancer patients and therapy practitioners worldwide. The models derived from MUCARF and new insights presented by them are discussed in the following sections.

**Interpretation, Analysis of Results and Discussions**

**Types of cancers:** In total 22 types of cancers are reported and documented in the Australian contexts, among others, breast, colon and rectum cancers are prevalent in Australia. Figure 4 shows the extracted data views from the metadata structures. We observe an increase in colon and rectum cancer as well as lung and bronchus related cancer cases. Similar cancer visualization views are in agreement with the study done by Bray et al. (2018), providing a scope of specialized data mining tools to explore trends and correlations and emphasizing knowledge-based models.
Multidimensional Research Framework for Complex Disease Ecosystems

Types of cancer occurrences: New cancer cases are emerging worldwide and lung, colon and rectum and breast cancers are quite common in particular in developed countries (Roser and Ritchie 2018). The statistical mining is done among new cancer cases and their causality rate attributes. We analyse the patterns and correlations through polynomial regressions as done in Isaac et al. (2012). New cases and causalities are daily phenomena in counties where an opportunity arises to implement the methodological framework and analyse in detail the root causes and mitigate their advancement. As shown in Figures 5a and 5b, a good correlation is observed in between new case occurrence and affected causality attributes and the correlation appears more than 90 percent.

Age-based cancers: Age-related cancers are common worldwide. Depending upon the lifestyles and the environment in which the population is living for existence, age connected cancers are developed. For example, the population living in industrially developed, populous and disease-prone countries may show different trends and correlations irrespective of their ages. We have compared four types of age-group related cancers for different countries, as shown in Figures 6 a, b and c. Surprisingly, the USA exhibits increased age-related cancers for age groups 50-69 and 70+. However, other groups show low rates of cancer occurrences. Whereas, for India and Uganda, the cancer rates affected for 15-49, 50-69 and 70+ age groups may be due to poor sanitation, economic disparity, and inadequate medical facilities among populated sections. The government policies and implementations may affect the cancer rates among certain sections of age groups in various countries.
Poverty based cancers: Poverty varies depending upon situational, generational, absolute, relative, urban, and rural scenarios. Poverty has multidimensional challenges and can be described as a composite dimension. Broadly, lack of quality and nutritious food, absence of medical institutions, immature social framework, poorly organized resources, and unequitable access to opportunities contribute to the poverty-based cancers. More specifically, poor sanitation, malnourishment, environment and economic factors are affecting the quality of life with increased risk of cancer rates in the third world countries in all continents. In addition, poverty being multidimensional has a risk of multidimensional cancers (Atkinson 2002). In addition, lifestyles, economic disparities and less access to medical facilities including healthcare insurances are other major reasons for occurring deadly diseases. As shown in Figures 7-8, we have drawn bubble plot views for ten different countries, where economic prosperity and healthcare facilities, including poorly developed infrastructure and inabilities of healthcare strategy implementations, are observed. Industrially developed countries, in spite of economic prosperity, suffer from more cancer death rates. An increase in GDP is envisaged as economic prosperity measure as described in Figure 7.

Figure 6. Age Based Cancer Rates Computed for (a) USA (b) India (c) Uganda

Figure 7. Poverty Based Cancers

The GDP per capita is affected by population and productivity of the labor force. Standards of living can be measured through these attribute dimensions. Though the cancer deaths and income inequalities have no direct relationship, it is interesting to observe their trends in between GDP of countries and cancer-related deaths as shown in Figure 7. Though a high rate of cancers is observed in high-income countries, the situation is changing because of awareness and health consciousness in these countries. We draw bubble plot views to visualize the affected attributes among Asia, Africa and Western communities and their related countries in Figures 7 and 8.
Figure 8. (a) Inequalities Affecting Life Styles, Causative to Cancers

Though there is no direct relationship observed between income inequalities and cancers, but income inequalities changed the lifestyles of diverse population sections. The bubble plots drawn in between attributes income inequality and cancer death rates for 12 countries suggest for each variable range of income inequality, a corresponding casualty rate is interpreted. The casualty occurrence may be due to changing in lifestyle. A steady increase in inequality suggests analogous evenness in cancer death rates among many countries, especially in Africa, but the trend is changing in Europe (Figure 8). It is a different scenario in the case of Asia, where a steady rate of cancers is observed for China and India even though they have large economic disparities.

Figure 8. (b) Prevalence of Neoplasm in Countries (c) Country Attributes having Cancer Cases

Country-based neoplasm: A neoplasm is an abnormal growth of cells, known as tumors and it has been reported among many people in many countries. As shown in Figure 8b, three distinct trends of neoplasms are observed between “USA, Australia, New Zealand”, Europe and the rest of the countries in Asia and Africa. There are two types of neoplasms, one benign and malignant and their growth is slow and steady in Asia and Africa, compared with westernized societies (Figure 8b). The Australia, New Zealand and USA have reported a high percentage of neoplasm compared with Asian and African countries. However, China and USA have reported high rates of cancers compared with other countries as shown in Figure 8c.
Number of people suffering from cancers: As per latest literature on cancer documentation (Roser and Ritchie 2018), in the current year alone more than 10 million people are reported to be suffering from cancer diseases though new cases are emerging alarmingly every year (Ferlay et al. 2014). Among countries, two top economically developed countries have higher cancer rates compared with other countries chosen in the same time periods. As shown in Figure 9, an increasing in bubble sizes and trends attributed in time period suggests, more number of people are suffering from cancer diseases. The cancer disease trends are distinct between two economic powers and the rest of the countries. In general, the increase in cancer rates is steady for all other countries in Asia, Africa and Europe (Figure 9).

Race- and gender-based cancers: The cancer disease is a major public health-scare for all races in many continents even among ethnic groups. Whites, blacks and mixed races are most common among others. There are more than 30 genders, but we choose to analyze two prominent genders, males and females. In spite of the number of races, the study can account for types of cancers and their causalities worldwide. We incorporate the race-based data attributes in the multidimensional data models to create metadata cubes using the framework as described in Figure 3. We have a couple of data views which are representative and worth mentioning in the current research as shown in Figure 9. The data views presented with an observation of similar and dissimilar correlations and trends reveal prostate, thyroid, skin and breast cancer survival rates are higher compared with pancreas, liver, and lung and esophagus that exhibit lowest survival rates. Black demography suggests lower survival rates compared with white demography as shown in Figures 8 a and b. In another scenario, as shown in Figure 9b, the male population exhibits lesser cancer rates at early time periods compared with their rates at late time periods. In female cases, among the population cancer rates, the correlation coefficient differs with the male population. However, the trends of both males and females match at late and recent time periods as shown in Figure 9b. In both cases, however, the cancers for all races match with the white population.

Evaluating the Research Framework as a Technology Therapy

Research Questions 1 and 3 are addressed in this section. Can technology overcome the health scares among masses in the world? It can change the life of a patient or even health of the entire population if the right technology is used at the right time to solve a particular type of cancer with educative measures. Cancer data sources are assets of informatics solutions, data analytics and new knowledge exploration. Large volumes and varieties of digital healthcare data motivate the analysts and interpreters to dig details through modelling, integration and knowledge building processes. Integrated therapy articulations can also address the cancer patients, suffering from chronic and non-commutable stress, anxiety and depression related illnesses. These dimensions are worth reporting and documenting in the current research to accommodate in the dimensional modelling and establish the robustness of MUCARF as a warehouse repository. Various components of the integrated framework described in Figure 3 generate a comprehensive metadata structure for data analytics. Besides, the privacy, security and confidentiality
data components cannot be ignored while building knowledge-based models for which the current technology can come to the rescue of protecting patient records and medical notes. The data collection, processing and interpretation, are other key ingredients of informatics solutions in addition to the data encryption methods. Any model developed has no meaning without bringing all research groups into a common knowledge domain and interpretation, validated by special interest medical experts or practitioners. Economic power can not only construct the social fabric but can destroy an environment among populated geographies, where cancers prevail. A good and wise balance of economic power and healthcare strategy are needed for prosperous societies worldwide.

The purpose of polynomial regression is intended to build the relationships between dependent and independent variables with various equation fits. We want to interpret relationships, establish a type of relationship and its strength. Often, polynomial regression fits a nonlinear relationship between dependent and independent variables. Several examples are illustrated in the literature (Trebuňa et al. 2016), such as the progression rate of tissues and the advancement of disease epidemics. Figures 5-9 demonstrate the relationships between various cancer data attributes as interpreted in different regression plots. The interpretation of the polynomial regressions has led us to determine, inferring the strength of measurable variables of cancer attribute dimension described in framework articulations. The correlation coefficients exhibit a strong association between attributes with linear, exponential and orthogonal regression relationships. New cases versus cancer deaths attributes show poor correlation, which may be due to either poorly documented cases and/or unreported death rates.

Conclusions, Limitations and Recommendations

Keeping in view the complexity and heterogeneity of cancer disease systems, a holistic framework is necessary that can interrogate and mitigate different types of cancers. The cancer status in developing countries is highlighted where resources are limited and action-plans are constrained. During 2016 alone, around 9 million people have died because of cancer-related illnesses. New cases including causalities are reported. An integrated framework is necessary to manage reporting, documenting and organizing multidimensional cancer disease ecosystems. The cancer informatics framework, MUCARF requires data from multiple domains and their linked entities and dimensions. The artefacts described in the MUCARF work well with different types of cancers. The knowledge obtained from cancer systems is not a particular domain-centric, but to describe knowledge-based models on internal structure, cancer diagnostic stages and conditions. The framework articulations are evaluated using empirical data instances acquired from different types of cancers and analysing the trends and correlations of dependent and non-dependent variables deduced from cancer metadata, which are periodically and geographically interpreted in new knowledge domains. In general, a strong correlation is found in between attribute dimensions of multidimensional schemas and models deduced for different demography and geographies. An increase in the prevalence of cancers is observed when we assess industrially developed countries, and countries with certain lifestyles, such as eating habits, smoking and alcohol consuming countries, which cannot be ignored. Though the cancers developed from lifestyles of populated sections are not covered in the current study, but the cancer framework appears to be more robust and holistic for managing future scopes.

References


Mastrian, K. G. and McGonigle, D. 2017. Informatics for Health Professionals, Jones & Bartlett Learning, USA.


