DEVELOPMENT OF AN INTELLIGENT RISK PREDICTION SYSTEM FOR BREAST CANCER THREAT ASSESSMENT IN A POPULATION-BASED OMANI WOMEN HEALTH DATA REGISTRY AND A PERSONALIZED ICT-BASED AWARENESS INTERVENTION

Title: DEVELOPMENT OF AN INTELLIGENT RISK PREDICTION SYSTEM FOR BREAST CANCER THREAT ASSESSMENT IN A POPULATION-BASED OMANI WOMEN HEALTH DATA REGISTRY AND A PERSONALIZED ICT-BASED AWARENESS INTERVENTION
Proposal ID: BFP/RGP/ICT/20/142
Type of project application: Research Grant Program

Current Status: For Review By Research Focal Point

Sector Name

Information Technology and Communication

Principal Investigator

Madeswaran, Tamilselvi (tamilselvi.madeswaran@nct.edu.om)

Co-Prinicipal Investigator

Al Shaqsi, Khalifa Zayid Khalifa (khalifa.alshaqsi@nct.edu.om)

Co-Investigators

Kavuru, Aruna (arunudu@gmail.com)
Al Foori, Maya (maya.alfoori@moh.gov.om)

Other Team Members

Team members not specified

Technology Readiness Level (TRL)

4 – 6: Technology Development/ Demonstration (public & private collaboration) e.g. prototype testing, advanced-stage research

Priority Area

Other

Executive Summary

Breast cancer is the most frequently diagnosed life-threatening cancer in women worldwide, with about 2.1 million new cases every year according to World Health Organization. Breast cancer represents about 34.1% of all reported cancer cases in Omani females, with an average age of 34.7 and high mortality rates of 11 per 100,000 populations (GLOBOCAN 2018). According to the data of Oman National Cancer Registry (ONCR) over the past two decades (1996–2015), the incidence rate of breast cancer is nearly doubled between 1996 and 2015 from 13.6 to 26.9 per 100,000 women. The 20-year trends data show breast cancer to be the foremost common malignancies affecting Omani women accounting for 21.2% of all cases (Al-Lawati et al., 2019). Statistically, breast cancer is the most common and leading female malignancy among Omani women as recorded by International Agency for Research on Cancer (IARC), the number of new cases of females, incidence with breast cancer including all ages in 2018, is 454 (13.7%) and deaths among these incidence is 133 (8%). According to (Mehdi et. Al. 2014) breast cancer in Oman (and other Middle East and developing countries) presents at younger age with more aggressive phenotype. The frequency of cancer among the women

is gradually increased between the years 2003 to 2015. Detection of breast cancer is at the young age in the advanced stage (III or IV) with low survival rates [www.mohe.gov.om].

The main cause of breast cancer is changing lifestyle and the risk factors such as age, family history, early mensural age, late menopause, obesity and contraceptive pills. Even though the Health Care provided by the ministry is free of cost, young Omani women hesitate to approach the hospitals due to number of barriers such as fear, embarrassment, difficult to explain the problem with male doctors etc. The main objective of this research is to reduce the mortality rate of Omani women community due to breast cancer.

The Oman National Cancer Registry was established in 1985 as a hospital-based registry. Since 1996 this hospital-based cancer registry has been functioning under the Directorate General of Primary Health Care and data about all the potentially reported cancer cases from all the regional hospitals are collected and coded in the registry.

Observations of recent literature informed that the prevalence of breast cancer is due to combination of risk factors. Occasionally unknown risk factors will also be the cause for the occurrence of breast cancer. Also the impact of contribution of each of the risk factors in the cancer occurrence varies among the females. These characteristics thus imply that Ministry of Health should put more emphasize on the research of breast cancers. The central idea is to propose a machine learning based breast cancer prediction model, to care the female general population. The proposed research is a three-fold.

Firstly it is proposed to establish an Omani Women Health Data Registry (OWHDR) which is a population based data registry that periodically captures the risk factors associated with breast cancer along with non-risk factors. Consecutively this OWHDR will be transformed as a knowledge base.

Secondly, based on the cancer occurrence data provided by the ONCR it is proposed to build a breast cancer risk prediction model in order to predict the possibility of occurrence of breast cancer among the female population.

Third fold concentrates on the implementation of breast cancer awareness. This will be realized through the implementation of awareness programs, community based campaigns and the development of a Web based interactive portal and mobile application. Subsequently breast cancer awareness data will be collected through various methods and analyzed in different perspectives such as knowledge and perception about breast cancer; barriers faced by women in the early presentation of breast cancer; healthcare-seeking manners etc. The outcomes of this data analysis will support in policy making, designing health education and societal wellbeing.

Introduction and Statement of the Problem / Project

Cancer patients feel challenging physical issues; cancer diagnosis causes huge emotional suffering and its treatment procedures results serious side effects. The initial diagnosis of cancer is perceived by many patients as a grave event suffering from anxiety and depression. Cancer is equally distressing for family as well. Cancer control and prevention is an integral part of the healthcare system and is a challenging task. Its aim is to reduce the incidence, mortality and morbidity of care and to improve the quality of life of the patients suffering from cancer. Accurate, adequate, complete and timely health information and clinical knowledge play a vital role in accomplishing this objective.

Since in Oman there is a tendency of late diagnosis in younger age due to some cultural and social reasons such as losing husband due to physical impairment, social taboos and unable to take self-decisions on treatment etc., the number of breast cancer incidence among young women in Oman can actually be much higher than reported. It has been observed that the rate of breast cancer incidence in Oman does not increase with the higher age which is quite unusual. Besides ethnicity, environmental factors, genetic aspects, hormonal status, and other yet unknown factors can influence the age at diagnosis for breast cancer.

In order to address such life threatening ailment, the proposed research aims to develop Knowledge based Prediction Model which foresee the likelihood of occurrence of breast cancer among the female community. This model development involves three phases. First phase is a Knowledge base creation through the established OWHDR, which contains the data of breast cancer risk factors. Such risk factors are either modifiable or non-modifiable. Among these captured factors, several of the correlated factors may be the reason for the breast cancer incidence. The possibility of correlations may occur among the risk or non-risk or both the risk factors.

Second phase develops breast cancer risk prediction model. The Multiple Criteria Decision Making (MCDM) techniques are proposed to configure the risk factors and to make a decision on different classifiers among multiple criteria in cancer risk datasets. Based on the supervised machine learning techniques specifically Logistic Regression, Neural Networks, Decision Trees and Nearest Neighbors as well as different input features and data set of ONCR the cancer prediction model will be constructed. A self-learning methodology when applied over the developed prediction model will foresee the prevalence of breast cancer risk of an individual female using OWHDR.

Third phase performs breast cancer awareness data analysis. This will be attained through the implementation of ICT based Cancer Prevention Education Programme.

Literature Review and Analysis of Related Work

Zahid et al. (2019) revealed that breast cancer is the leading cancer among women. Almost 20% of patients develop brain metastases and die shortly afterward. There is a dearth of data on the survival outcome of breast cancer patients with brain metastases from the Arab world. Their research data indicate that Omani women are diagnosed with breast cancer at a younger age, develop brain metastases earlier, and carry a poor outcome.

Al-Lawati et al. (2019) says that the Breast cancer was the most common and leading cancer among Omani females and gradually increasing year by year. Between the year 1996 and 2015 the incident rate is 21.2%, with an average annual age standardized incident rate (ASR) of 20.8/100 000 women. In the 20-year period the cancer incidence has increased from the ASR 26.9 to 26.9 among 100000 women.

Verma et al. (2019) indicated that breast cancer tumor can be malignant or benign. Benign can be treated easily but malignant treatment depends on the stage. This study effort towards the prediction of type breast cancer was achieved by applying linear kernel Support Vector Machine algorithm and the experimental results shown 90.3% accuracy in prediction.

Al-Azri et al. (2016) suggested that increasing the awareness on cancer symptoms and encouraging the seeking of early medical help especially the Omani women who are divorced and separated are required. Number of initiative programs should be carried out for all kinds of Omani women irrespective of their age group. Using social media, TV, radio and seminars, campaigns the cancer awareness should improve.

Dana et al (2016) stated that early and precise diagnosis is essential in breast cancer rehabilitation and treatment. Due to complex uncertainties in the disease the detection is difficult using mammograms. This paper compares three Machine Learning techniques namely Support Vector Machine, Random Forest and Bayesian Networks for breast cancer detection and diagnosis.

Esra et al. (2016) stated that the incidence of breast cancer is rising in Oman, and the disease is diagnosed at late stages, when treatment success is limited. Omani women might benefit from better awareness, so that breast cancer can be detected early and treated. They were concluded that the overall low scores for awareness and early detection, and the survey of local beliefs highlight a severe necessity for a contextually-tailored breast cancer awareness intervention programme in Oman.

Objectives

The following are the main objectives of this research proposal.

1. To establish the population based Omani Women Health Data Registry (OWHDR) to capture the risk/non-risk factors associated with breast cancer.
2. To create a health knowledge base though transforming the data from OWHDR.

3. To develop an Intelligent based breast cancer risk prediction model using the ONCR dataset in order to predict the possibility of occurrence of breast cancer among the female population.
4. To implement personalized ICT based Cancer Prevention Education Programme.
5. To perform breast cancer awareness data analysis.

Research Methodology [Describe your Implementation Plan,Time-line and Milestones]

This research study will involve several approaches to realize its objectives.

In the current scenario, we have hospital-based cancer registry maintained by ONCR and Directorate General of Primary Health Care, which provides data about cancer patients. But it is highly required to capture the risk (modifiable and non-modifiable) / non-risk factors of general population who are free from cancer in order to foresee and prevent the risk of breast cancer prevalence. Such data should also be captured in a periodical time series. This requirement will be realized though establishing a personalized Omani Women Health Data Registry which is a responsive User Interface. This registry is a population based data registry that captures breast cancer associated risk factors along with non-risk factors of female community free from cancer. The subsequent stage will transform the OWHDR data into a health knowledge base.

The next phase concentrates on the development of breast cancer risk prediction computational model. The data provided by the ONCR is associated with cancer diagnosed patients. Najla et al. stated that currently twenty-year trends of cancer incidence in Omanis; 1996–2015 is available which are voluminous and heterogeneous and so are characterized as big data. These data are treated as training data in the prediction model development. This model development involves Operations Research (OR) principles for prioritizing the risk factors in the incidence of breast cancer. The Multiple Criteria Decision Making OR techniques are planned to construct the hierarchical structure of all the breast cancer time series risk factors that are captured through the developed population based data registry OWHDR. Followed by, the MCDM computation will prioritize and weight the qualitative risk factors appropriately which are required for further computation in model development. Next process is prediction model development. Thus the ONCR provided cancer dataset will be preprocessed which will remove all the unnecessary data and extract important features from data. Afterwards applying the supervised machine learning techniques specifically Logistic Regression, Neural Networks, Decision Trees and Nearest Neighbors prediction model will be constructed. This model will be trained using the preprocessed ONCR training data set to do the prediction accurately. The next stage develops a methodology to foresee the probability of cancer incidence on the individual female using the corresponding OWHDR test dataset. This will be achieved by means of applying self-learning algorithms over the trained prediction model.

Based on the observation through the literature stating that the occurrence of breast cancer is more among the Omani females at their younger age, awareness over the prevention and early detection of breast cancer is essential for each individual Omani female community. In view of achieving this essential need, the next phase aims to develop an ICT based Cancer Prevention Education Programme (CPEP) interface to disseminate the in-depth knowledge about the breast cancer; prevention and early detection strategies among the female community. This interface is a web based interactive portal and Mobile application. The intuited components of this portal are expert health advisory system, community based cancer awareness programmes, breast cancer prevention education, reminders on the regular medical check-ups etc. This portal will also be well connected to the females those who are identified as probable breast cancer risk occurrence during the second phase. Subsequently after executing the CPEP breast cancer awareness data will be collected from the females through various sources and data analysis will be carried out using statistical data analysis methods.

The first milestone of the project will be measured according to three major deliverables: (1) establishment of population based OWHDR knowledge base; (2) development of intelligent based breast cancer prediction model; (3) development of ICT based Cancer Prevention Education Programme (CPEP) interface.

The second and final milestone of this research will be the data analysis exploring the breast cancer consciousness among the female community and the impact in policy making and health education.

In terms of timeline, the following can be used as a general guide for implementation:

1. Establishment of the population based Omani Women Health Data Registry (OWHDR) – (2 months).
2. Transforming the data from OWHDR and creating a health knowledge base – (1 month).
3. Developing an Intelligent based breast cancer risk prediction model and train the model using ONCR dataset – (12 months).
4. Creating and implementing personalized ICT based Cancer Prevention Education Programme interface – (3 months).
5. Performing breast cancer awareness data analysis – (6 months).

In total, the entire research can be completed within two years.

Benefits to Oman

As mentioned above, this research will benefit Oman in various ways:

The personalized ICT based Cancer Prevention Education Programme interface that can be developed from this research can be used in educational institutions for the purpose of nurturing the student community on health education specifically focusing on the awareness on breast cancer. This education intervention programme on health consciousness can also be planned as a non-credit mandatory course. Thus given in a regular basis since from the student's childhood will lead the younger generation as hale and healthy quality life.

Successful implementation of the developed Population based Omani Women Health Data Registry and the ICT based Cancer Prevention Education Programme interface of this research in the Omani community will jumpstart not only a new paradigm in cancer awareness initiatives but also to embrace the country in Care 4.0. This will bring Oman in the limelight in the region and in the world, especially if this approach becomes mainstream health education in Oman's educational institutions.

Academic, Scientific and/or Innovation Significance

This proposed research is significant in the following ways.

Population based OWHDR maintenance will support the Ministry of Health for various forms of health policy planning and implementation.

Blending the ICT based Cancer Prevention Education Programme interface with educational institutions, as part of the health education intervention study will enlighten the wellbeing knowledge of young students community since from their infancy and thus mould them as a holistic healthy generation.

The prototype model of this research will be extendable to any type of cancer research, when suitably customized with specific domain knowledge. This research will allow the Sultanate of Oman to keep abreast with the world in adopting the digital health and Care 4.0.

Is this project going to result in a patent?

No

Patent Review (e.g. any previous similar patents in literature, the potential of this project to result in a patent ...)

References

1. Al-Lawati NA, Al-Bahrani BJ, Al-Raisi SS, Al-Lawati JA. (2019). Twenty-year trends of cancer incidence in Omanis 1996-2015. Oman Medical Journal, 34(4), 361-387.
2. Mehdi I, Monem EA, Al Bahrani BJ, Al Kharusi S, Nada AM, Al Lawati J, Al Lawati N. (2014). Age at diagnosis of female breast cancer in Oman: Issues and implications. South Asian J Cancer, 3, 101-6.
3. Cancer Incidence in Oman 2015: www.mohe.gov.om

4. Zahid, K. F., Kumar, S., Al-Bimani, K., Ahmed, T., Al-Ajmi, A., Burney, I. A., & Al-Moundhri, M. (2019). Outcome of Omani Women with Breast Cancer-associated Brain Metastases Experience from a University Hospital. Oman medical journal, 34(5), 412–419.
5. Verma, Garima and Verma, Hemraj, Predicting Breast Cancer using Linear Kernel Support Vector Machine (March 11, 2019). Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE) 2019.
6. Al-Azri, M., Al-Maskari, A., Al-Matroushi, S., Al-Awisi, H., Davidson, R., Panchatcharam, S. M., & Al-Maniri, A. (2016). Awareness of Cancer Symptoms and Barriers to Seeking Medical Help Among Adult People Attending Primary Care Settings in Oman. Health services research and managerial epidemiology, 3, 2333392816673290.
7. D. Bazazeh and R. Shubair. (2016). Comparative study of machine learning algorithms for breast cancer detection and diagnosis, 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA), Ras Al Khaimah, 1-4.
8. Esra Alkhasawneh, Saad T Siddiqui1, Michael Leocadio, Vidya Seshan, Yahya Al-Farsi, Mansour S Al-Moundhri. (2016). I Do Not Even Say “It” - a Mixed Methods Study on Breast Cancer Awareness of Omani Women, Asian Pacific Journal of Cancer Prevention, 17(4), 2247-54.

Budget Summary:

Data Collection/ Analysis	900	700	1,600
Equipment and Facilities	1,500	500	2,000
International Conference	500	1,400	1,900
Local travel	50	50	100
Materials and Supplies	100	100	200
Research Assistants (Post Graduate)	1,400	1,400	2,800
Administration cost	250	250	500
Dissemination	250	250	500

Duration in months

24

Overall TRC Requested Funding (OMR)

9,600.00