RESEARCH ARTICLE

Call for a Computer-Aided Cancer Detection and Classification Research Initiative in Oman

Andri Mirzal*, Shafique Ahmad Chaudhry

Abstract

Cancer is a major health problem in Oman. It is reported that cancer incidence in Oman is the second highest after Saudi Arabia among Gulf Cooperation Council countries. Based on GLOBOCAN estimates, Oman is predicted to face an almost two-fold increase in cancer incidence in the period 2008-2020. However, cancer research in Oman is still in its infancy. This is due to the fact that medical institutions and infrastructure that play central roles in data collection and analysis are relatively new developments in Oman. We believe the country requires an organized plan and efforts to promote local cancer research. In this paper, we discuss current research progress in cancer diagnosis using machine learning techniques to optimize computer aided cancer detection and classification (CAD). We specifically discuss CAD using two major medical data, i.e., medical imaging and microarray gene expression profiling, because medical imaging like mammography, MRI, and PET have been widely used in Oman for assisting radiologists in early cancer diagnosis and microarray data have been proven to be a reliable source for differential diagnosis. We also discuss future cancer research directions and benefits to Oman economy for entering the cancer research and treatment business as it is a multi-billion dollar industry worldwide.

Keywords: Cancer incidence - cancer research in Oman - cancer detection and classification - medical images

Asian Pac J Cancer Prev, **17 (5)**, 2375-2382

Introduction

Recently cancer is becoming a major health problem in Oman. Based on 10-Year Cancer Incidence Among Nationals of the GCC States 1998-2007 report (Al-Madouj et al., 2011), cancer incidence in Oman is the second highest after Saudi Arabia among Gulf Cooperation Council (GCC) states. The National Oncology Centre and Ministry of Health have reported significant increase in cancer incidence in Oman as about 25% more cases in 2011 compared to previous year (Muscat Daily, 2014). Data from Ministry of Health show that in 2006 cancer was the third leading cause of death in hospitalized patients in Oman (Nooyi et al., 2011). And according to GLOBOCAN estimates, Oman is predicted to face almost two-fold increase in cancer incidence in the period 2008-2020 (IAEA, 2013).

There are several factors that contribute to this phenomenon, e.g., the success in controlling communicable diseases (making the number of noncommunicable diseases cases including cancer incidence cases more prevalent), ageing population, the change in lifestyle due to the recent industrialization process, the use of more advanced cancer detection technologies that allow earlier and more accurate detection, and the improved awareness about cancer risks so that more people are willing to take the tests. Due to these factors, Oman is expected to see a spurt in the number of cancer incidence cases in near future.

Cancer research in Oman is still in infancy. This is due to the fact that medical institutions and infrastructure that play central role in data collection and analysis are relatively new developments in Oman. The Oman National Cancer Registry was established in 1985. Only in the end of 90's Oman had its first central laboratory for cancer diagnosis that is based in Royal Hospital, Muscat. College of Medicine and Health Sciences, Sultan Qaboos University, was established in 1986. And Oman Medical College was established in 2001. In term of research outputs, there are only a handful papers and reports published by Omani institutions. The main contents of such works are about cancer incidence reporting, summarization, and analysis (Al-Madouj et al., 2011; Nooyi et al., 2011; Mehdi et al., 2014; Hamdan et al., 2009; Al-Mahrouqi et al., 2011; Moore, 2013). Recently, however, some works that initiate more advanced topics like detecting, classifying, curing, and characterizing cancer cases in Oman are observed. For example, an Omani PhD student in SQU reported that she has discovered an inhibitor to the proliferation of some cancer cell lines in thorn apple (sherngeban), a plant that grows wild in many parts of the Sultanate and other countries in

*Department of Computer Science, College of Arts and Applied Sciences, Dhofar University, Salalah, Sultanate of Oman *For correspondence: amirzal@du.edu.om, shafique@du.edu.om*

the region (Oman Observer, 2014). The Research Council of Oman has also recently funded three projects related to the cancer research (Bayoumi et al., 2013; Alharrasi et al., 2014; Adham et al., 2014).

It is vital for Oman to initiate her own cancer research as there can be some distinct factors that are inherent to Omani population. As stated by Dr. Mandhari from The National Oncology Centre, "Western standards cannot be applied to Oman because they have different set of criteria based on the research work on a specific community" (Times of Oman, 2014). Reports from the current cancer studies in Oman suggest that there are actually some specific characteristics of cancer incidence in Oman, i.e.

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. Also the rate of breast cancer incidence in Oman does not increase with the higher age which is quite unusual. Because in Oman there is a tendency of late diagnosis in younger age due to some cultural and social reasons, e.g., losing husband due to physical impairment, social taboos, and treatment decision is made by the males, the number breast cancer incidence among young women in Oman can actually be much higher than reported.

Oman has the highest stomach cancer incidence rate among men and women compared to all GCC states (Al-Madouj et al., 2011; Al-Mahrouqi et al., 2011). This rate is also quite high compared to other countries with only surpassed by Japan and Brazil (Al-Madouj et al., 2011). According to a report (Hamdan et al., 2009), Oman had the highest male preponderance among GCC states with male:female ratio reaches 114:100 (overall ratio among GCC states is 102:100).

Among GCC states, Oman has relatively lower age of presentation for cancer cases (Saudi and UAE are the lowest). Oman has the lowest mean age when the diagnosis was made for male: 52.2 years old, and the second lowest mean age when the diagnosis was made for female: 46.8 years old (Al-Madouj et al., 2011).

These characteristics, thus, imply that Oman should put more emphasize on the research of two cancer types, i.e., breast and stomach cancers. And the high ratio of male preponderance suggests that Oman should also encourage research on cancers that are more prone to male like lung cancer (Hitchman et al., 2011).

There are two major data to assist radiologists with cancer detection and classification: microarray gene expression data and medical imaging data. Microarray gene expression is a simultaneous measurement of expression levels of large number of genes (thousands to hundreds of thousands). The expression dataset is expensive and requires high expertise to obtain, so that it is not practical to collect the data for each patient with cancer. But fortunately there are many publicly available microarray datasets (de Souto et al., 2008).

Medical imaging is the technique to create visual representations of inner parts of body for clinical purposes. The medical imaging data (e.g., PET, CT, MRI, and mammographic images) have many advantages, e.g., real time monitoring, accessibility without tissue destruction, and minimal or no invasiveness. These images are also readily available to be analysed after the screening, and the experts can immediately perform diagnostic process subsequently. Due to ethical issue, there are not many publicly available medical imaging data thus limiting the possibility of building versatile detection and classification methods. However, for research purpose there are some datasets that can be used, e.g., (IRMA, 2016; MIAS, 2016; DDSM, 2016).

Both microarray and medical imaging datasets allow early detection and classification of cancers so that early treatments can be performed to minimize the more risky procedures, thus providing higher chance of recovery. Here we emphasize our discussion on the use of mammographic images for early detection and classification of breast cancers as the facilities to perform breast scan have been widely available at medical institutions in Oman. The discussion, however, is applicable to other medical images as CAD procedures for other medical images are similar. Moreover, in Oman breast cancers present at younger age with more aggressive phenotype, so an early detection of breast cancer may provide many benefits to women in their productive ages.

Materials and Methods

Literature review

Various research efforts have been carried out on cancer detection and classification. In the following we outline some methods for detecting and classifying cancers using microarray gene expression and medical imaging data.

Microarray gene expression based methods

Gene expression is a measurement of expression levels of large number of genes by using DNA microarray technology and has been used extensively in cancer detection and classification research. The using of microarray gene expression technology often improves the accuracy and sensitivity of early cancer detection as this technology provides unbiased and systematic approach for recognizing cancerous tissues (Golub et al., 1999). As the number of genes in a microarray gene expression dataset is very large (thousands to hundreds of thousands) compared to the size of samples (which usually is tens to hundreds), gene selection is a necessary step to reduce the dimensionality of the dataset. This step is also required to balance the dataset dimensionality as it is known that the performance of classifiers in imbalanced data is unstable and may not generalize well due to overfitting. In this step, redundant and noisy genes are expected to be removed as it is known that they can significantly reduce the classifier accuracy (Zhao et al., 2007; Zhao et al., 2013). In our previous works (Mirzal, 2013; Mirzal, 2014a; Mirzal 2014b; Mirzal, 2014c), we proposed unsupervised gene selection algorithms to improve cancer classification results. Prior to the gene selection, usually every sample vector (a vector containing the expression levels of all genes for the corresponding sample)is normalized. The most used normalization procedure is to adjust the distribution of each sample vector to a unit normal distribution.

There are two techniques that can be employed to select the most informative genes from the dataset: feature selection and feature extraction. Feature extraction transforms the original feature space into a distinct space (Varshavsky et al., 2006). On the other hand, feature selection reduces the original feature space into a subspace without transformation. The transformation in feature extraction may provide a better discriminatory ability. But the transformed space usually is problematic because it has no physical interpretation and makes the genes with the most discriminatory power undetected (Krızek, 2008), so that feature selection is more preferable. The following outlines some works that may be adopted to select the most informative genes.

In (Liao et al., 2014), the authors proposed a supervised feature selection method based on Laplacian score that was previously proposed in (He et al., 2005). The authors improved the previous work by introducing a mechanism to incorporate discriminative information into local geometrical structures. This mechanism generalizes the previous work by taking into account the class labels information so that it can simultaneously minimize within-class information and maximize between-class information. However, the authors do not incorporate any mechanism to deal with redundant and noisy features which are known to reduce classifier performances significantly. Thus it is possible to improve this work by introducing mechanisms to deal with redundant and noisy features. For example, a method proposed in (Zhao et al., 2013) can be accommodated to deal with redundant features and a mechanism introduced in (Zhao et al., 2007) can be utilized to reduce the number of noisy features.

In (Zhao et al., 2013), the authors proposed a set of three feature selection algorithms based on similarity preserving that can handle feature redundancy. The redundant features not only increase the dimensionality unnecessarily, but also worsen the performance of the classifiers. Thus, removing them will have significant effect on the classifiers' performances. The similarity preserving is a common mechanism in many feature selection algorithms, e.g., Relief and ReliefF (Sikonja et al., 2003), Laplacian Score (He et al., 2005), Fisher Score (Duda et al., 2001), SPEC (Zhao et al., 2007), HSIC (Song et al., 2012), and Trace Ratio (Nie et al., 2008). But because all of these algorithms evaluate features individually, they cannot handle the redundancy problem. As shown by the authors, the mechanism for handling feature redundancy can be extended to the above mentioned algorithms including Laplacian Score which is the base for the work in (Liao et al., 2014).

In (Zhao et al., 2007), the authors use spectral graph theory to analyse the structure information of graph model representing the similarity between its vertices (the vertices represent the genes). As the spectrum of the graph measures the separability of the components of the graph, only the first k (k is the number of the clusters) eigenvectors of the adjacency matrix of the graph are needed to obtain the cluster indicators (Shi et al., 2000). Because noise components concentrate on eigenvectors with small eigenvalues, using only the first k eigenvectors achieves an effect of reducing noise influences. As the feature selection algorithm proposed in (Liao et al., 2014) is also based on graph model on similarity between genes, it will be straightforward to integrate this mechanism.

After selecting the most informative genes, the next step is to employ classifiers to the pruned data (the data where only the expression values of the selected genes present) to classify the samples. There have been many classifiers proposed in cancer classification tasks with support vector machine (SVM) and artificial neural network (ANN) seem to be the most commonly used models. The first of work that promotes the using of SVM as classifier in cancer research is (Guyon et al., 2002) where the authors demonstrated that their method yields better classification performance and is biologically relevant to cancer. There were actually some works that preceded this work, however the work of (Guyon et al., 2002) was the first one to demonstrate the strength of SVM in cancer detection and classification from microarray gene expression datasets. There have been numerous subsequent works that showed the power and capability of SVM in this task, e.g., (Roayaei et al., 2013; Maulik et al., 2013). Mostly the works conclude that the standard linear SVM is good enough as a classifier for microarray gene expression datasets since the most problematic part is the gene selection. The first important work that demonstrates the use of multiclass cancer diagnosis using SVM was the work of (Ramaswamy et al., 2001). There are also many works that proposed the using of ANNs as classifiers. Figure 1 shows the steps for cancer classification by using microarray dataset.

Medical imaging based methods

There are many types of medical imaging data, e.g., MRI, CT, PET, ultrasound, and mammographic images. There are some standard steps in processing these images: (1) image pre-processing(cropping operation, histogram equalization), (2) segmentation or partition, (3) feature extraction, (4) feature selection(textural, morphological, and statistical features), and (5) classification (Ganesan, 2013; Jalalian et al., 2013; Antonie, 2001). Figure 2 describes the generic CAD processing steps in mammographic images.

Pre-processing

Pre-processing is a necessary step to improve the quality of the images to make feature extraction more

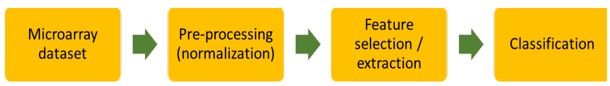

Figure 1. CAD System Pipeline for Microarray Dataset

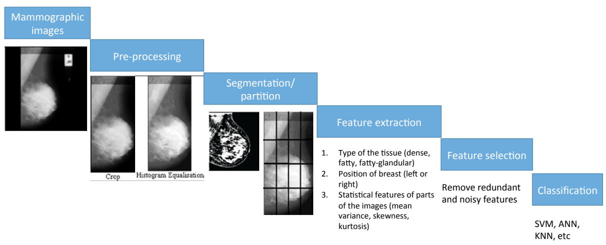

Figure 2. CAD System Pipeline for Mammograms

reliable. This step is especially crucial in preparing mammograms since the images are difficult to interpret due to low contrast and distinction between parts especially in young women with dense breasts (Ganesan, 2013; Antonie, 2001). There are two common tasks in this step: enhancement to increase the contrast and denoising to reduce blur (Ganesan, 2013). In this step, cropping operation is usually also performed to remove uninformative parts like black parts and existing artefacts (Antonie, 2001). There have been several approaches proposed to perform image enhancement and denoising, e.g., noise equalization procedure (Scharcanski et al., 2006), Bayesian approach (Simoncelli et al., 1996), wavelet-based methods (Heinlein et al., 2003), and Iris filter (Kobatake et al., 1999).

In (Scharcanski et al., 2006), the authors proposed a method to suppress noise and enhance image quality by using wavelet transform. This method is also able to improve the local contrast and sharpen the edges. Based on the preliminary results, the authors reported that their method can improve the microcalcification and other suspicious structures detection. In (Simoncelli et al., 1996), the authors proposed a Bayesian estimator based image enhancement. This method works by assuming the noise as white noise, so that a priori Gaussian estimate can be used to model the noise and thus the noise can be removed from the images. In (Kivanc et al., 1999), the authors proposed an image denoising method based on spatially adaptive statistical model. The authors reported that despite the simplicity of the proposed method, it is among the best noise removal methods. In (Heinlein et al., 2003), the authors proposed a wavelet based techniques for mammographic images enhancement. The idea is to process the images at several scales and improve the contrast and reduce the noise from those scales.

Segmentation or partition

After the pre-processing step, the image needs to be segmented so that areas with suspected abnormality (regions of interest: ROIs) can be isolated from the rest. Segmenting abnormal tissues in medical images is a challenging task because of the variability in appearance of abnormalities, the low sensitivity and specificity of the ROIs, and the low contrasts between neighbouring tissues (Jalalian et al., 2013). There are many segmentation methods proposed in the literatures, e.g., snakes (Kass et al., 1988), balloons (Cohen, 1991), active contour models (Isard et al., 1998), geometric methods (Yezzi et al., 1997), "shortest path" techniques (Mortensen et al., 1998), ratio cycles (Jermyn et al., 2001), ratio cuts (Wang et al., 2003), random walker (Grady, 2005), graph cuts (Kolmogorov et al., 2007), continuous max-flow (Appleton et al., 2006), total variation (Chambolle, 2005), TV-based convex relaxation methods (Chan et al., 2006), and fuzzy region-growing method (Boch, 2015). In some cases, the simple partition approach where the image is divided into n windows of the same size (see Figure 1) can also be used. When this approach is used, a set of features is extracted from each window and then classifiers are used to determine which windows containing suspected abnormal tissues (Antonie, 2001).

Feature extraction

The next step is feature extraction. Feature extraction is the most important step in driving the classifier performance. This step is sometimes followed by feature selection if the number of features is large or there are some irrelevant features included. Some examples of features in mammograms are: tissue type (dense, fatty, or fatty-glandular), position of breast (left or right), statistical parameters of the pixels in the segments/ windows (mean, variance, skewness, and kurtosis), number of microcalcification per unit, and mean diameter of microcalcification (Antonie, 2001; Sampat et al., 2005).

There are three major features in mammograms: spectral, textural, and contextual with the textural features are the most sought after in mammographic image analysis (Ganesan, 2013). In (Haralick, 1973), the author states that the textural features can be evaluated as being fine, coarse or smooth, rippled, molled, irregular or lineated. In mammograms, these features convey meaningful information and can be quantified for the purpose of classification in later step. There are some textural features that can be extracted from the images, e.g., energy, angular second moment, contrast features, correlation features, homogeneity, inertia, sum of squares, geometric moments, Zernike moments, difference moments, inverse difference moment, sum average, entropy, sum entropy, sum variance, texture probability, cluster tendency, difference variance, difference entropy, rectangular correlation, linear dependency of brightness, and deviation from the second order histogram (Ganesan, 2013). In addition to textural features, the use of gradient-based features like acutance also have been proposed for detecting breast cancer from mammograms (Mudigonda et al., 2000). The other important features discussed in the literatures are spiculation features (Karssemeijer et al., 1996), morphological features based on lesions physical characteristics (Timp et al., 2007), relative grey level change (Ganesan, 2013), fractal dimension (Ke et al., 2010), and wavelet features (Mousa et al., 2005).

Classification

The final step is classification. Classification is used to classify the segmented ROIs into normal and abnormal tissues. Further, usually it is also desired to classify the abnormal tissues into benign and malignant tissues to help radiologists in decision making. There are many methods to perform the classification, e.g., K-nearest neighbour (Ramteke et al., 2012), wavelet transform (Soltanian-Zadeh et al., 2004), LDA classifier (Zheng et al., 2013), SVM (Torrents-Barrena et al., 2015; Liu, et al., 2014; de Oliveira et al., 2015; de Nazaré Silva et al., 2014; Tsochatzidis et al., 2014), and ANN (Xie et al., 2015; Yasar et al., 2015; Fonseca et al., 2015; Ling et al., 2014; Dheeba et al., 2014; Shaji et al., 2013). From literatures, it seems that SVM and ANN are the most commonly used methods for cancer classification in mammograms. The first work that investigated the feasibility of SVM in breast cancer detection and classification using mammograms was the work by (Bazzani et al., 2001). Then it is followed by some subsequent works that also propose the use of SVM, e.g. (D'Elia et al., 2004; Singh et al., 2006). In addition there are also some works that compare the performance of SVM and ANN for this task (Ren, 2012; Papadopoulos et al., 2005).

Results and Discussion

The most challenging and important task is early cancer diagnosis and classification. In the early stage of development, cancerous tissues are rarely distinguishable from normal tissues. Thus, the diagnosis accuracy depends strongly on the capability and experience of the radiologists. This puts high pressure on the radiologists as they are expected to always deliver great performances which actually in turn can significantly reduce their reading accuracies instead. On the other hand, it is expensive and impractical to analyse DNA by using microarray technology for every patient that is suspected to have cancers. Thus it is highly recommended to put more efforts in developing reliable CAD systems to automatically diagnose cancers from medical images so that the systems can assist the radiologists in the reading process and in turn can improve their reading accuracies. As the medical imaging technologies have already been well developed, the utmost challenge in developing such systems lies in designing high performance algorithms. The algorithm design process should consider the whole system development including the pre-processing, segmentation, feature extraction, feature selection, and classification (see Figure 2).

There have been many research works that address the challenges of finding a subset of genes in microarray dataset with the most discriminatory power for improving cancer diagnosis and classification. The task of finding the optimal subset is NP-complete and as a microarray dataset usually contains the expression values of thousands of genes, it can be computationally prohibitive to evaluate all possible subsets. One of the most promising gene selection algorithms was proposed by (Liao et al., 2014). The most interesting aspect of this work is even though the proposed algorithm leads to accurate classification results, it actually is still incomplete as the algorithm does not take into account redundant and noisy genes. As stated in previous section, it is possible to improve the algorithm's performance by introducing mechanisms to deal with redundant and noisy genes.

After the optimal subset of genes is selected, the next step is to perform classification. There are many methods that can be used with SVM and ANN seem to be the most used methods. And if computational resource is not an issue, the seminal work by (Khan et al., 2001) can be adopted to achieve perfect classification results. In this work the authors train 3750 ANN models which each model is used to perform the classification. The final classification of each sample is then determined by aggregate voting. As each ANN model can be considered as an expert, it is the same as having 3750 experts work together to diagnose the cancers with the final decision is made by aggregating the opinions of the experts. Note that this work is categorized as a wrapper method where the gene selection step is blended together with the classification step. However, since it is possible to modify this method to receive gene subset as the input, it is possible to utilize this approach as classifiers in the CAD system pipeline.

Currently there have been many health centres equipped with screening machines located around Oman (Nooyi et al., 2011; Mehdi et al., 2014). It will be an interesting research project if one can deploy the proposed CAD systems into those machines and evaluate the performances of the systems in real world.

Benefits and economic impacts of cancer research for Oman

Until recently, cancer research in Oman mainly focuses on incidence reporting, summarization, and analysis. There are only a few reported works that go beyond these scopes. Thus, this research initiative can contribute to the development and enhancement of cancer research in Oman. The cancer research is in its initial stages in the whole GCC, so it will be a chance for Oman to become a leader and trendsetter in cancer research particularly in GCC.

Breast cancer incidences in Oman present at younger age with more aggressive phenotype, and there is tendency of late diagnosis in young age due to various cultural and social reasons. Thus, it is more imperative for Oman to develop its own research on CAD to improve the early detection of breast cancers to reduce the number of mortality rates among young women.

The impacts of initiating cancer research will be tremendous for Oman economy as cancer research and treatment is multi-billion dollar industry worldwide. The following enlists some possible economic impacts in the long term: *i*). Cancer research and treatment are multi-billion dollar ventures world-wide. According to (Eckhouse et al.,2008) global cancer research funding was 14,030 million euros in 2004-2005. The funding has since increased substantially to multi-billion dollars considering the fact that in 2011 the cancer research funding only in the UK was 521 million pounds while research funding budget of National Research Institute (USA) was estimated to be 5.4 billion USD (http://report.nih.gov/categorical_spending.aspx). However only a small portion (5%) of the total amount is used on prevention while only 2.7% is spent on cancer research directly applicable to developing and low-income countries. This scenario provides a great opportunity for Oman to invest on cancer research which focuses on developing novel mechanisms and technologies for early detection and prevention, which can be applied locally. *ii*). According to an article in Oman Observer (Kamoonpuri, 2013), more than 75000 Omanis go abroad every year for medical care and treatment. In addition to that, Ministry of Health also sponsors patients to go to UK, Germany, Saudi Arabia, and India to treat medical conditions. According to the article, hundreds of patients travelled to India just for positron emission tomography (PET) scans. Of the 623 sponsored Omani patients in India, 234 went for PET scans. Only for one year (2011), Ministry paid 1,693,825 OMR for these patients. In another study conducted by (Al-Hinai et al., 2011), it is shown that Omanis that went abroad for treatment, one third spent more than 2000 OMR (5200 USD) each. We believe that by fostering cancer research in Oman the local cancer detection and treatment facilities can be improved over the years, which shall save millions rials spent on medical tourism *iii*). In 2014, the global cancer drug market was estimated at more than 100 billion USD and is expected to grow to 147 billion USD in 2018. According to a report (Kasteng et al., 2008), the cost of cancer based on oncology drug sales in Oman was estimated at 38 million USD in 2006, making per cancer incidence cost to 27,067 USD. We believe that cancer research in Oman can help understanding various types of cancer and eventually in cancer drug manufacturing. *iv*). Currently in The Sultanate, there have been many mammogram screening machines installed in various health institutes. And the government also has encouraged women to take the test every year. As the awareness of breast cancer improves, it is imperative to have CAD installed in the machines to make to reading process faster and more accurate. Based on a study in (Chabi et al., 2012), it is reported that CAD can make the reading accuracies of young radiologists (4 months and 1 year of experiences) to be on par with the reading accuracies of senior radiologists (5 and 20 years of experiences), and as the diagnosis accuracy improved, the number of false positive cases (normal or benign lesions are mistakenly reported as cancerous lesions) can be reduced. Thus the number of biopsies can be reduced and the social cost associated with it also can be reduced.

In conclusions, computer-aided cancer diagnosis has been proven to be useful in assisting radiologists to improve their reading accuracies, and in Oman currently the facilities to perform medical scanning have been widely available in many medical institutions. However, the research on computer-aided cancer diagnosis which can be deployed immediately on those facilities is rather unheard. As there are many social, educational, and economic benefits, the impacts of initiating the research will be tremendous for Oman. In addition, as reported in many works there are some particular characteristics that are inherent to Omani population, so it is vital for Oman to initiate her own cancer research.

References

Adham S, Al-Ajmi A, Al-Mandhari M, et al (2014). Validation of a potential biomarker for the diagnosis/treatment of metastatic breast cancer within omani females. TRC Open Research Grant Proposal.

Al-Hamdan N, Ravichandran K, Al-Sayyad J, et al (2009). Incidence of cancer in Gulf Cooperation Council countries, 1998-2001. *East Mediterr Health J*, **15**, 600-11.

Alharrasi AS, AlZadjali F, Guillemin GJ, et al (2014). Exploration of new synthetically modified derivatives of acetyl 11-keto-bhetaboswellic acid (AKBA) isolated from Omani frankincense as tumor sensitizing agents. TRC Open Research Grant Proposal.

Al-Hinai SS, Al-Busaidi AS, Al-Busaidi IH (2011). Medical tourism abroad. *Sultan Qaboos Univ Med J*, **11**, 477-84.

Al-Madouj AN, Eldali A, Al-Zahrani AS (2011). Ten-year cancer incidence among nationals of the GCC states 1998-2007. Gulf Center for Cancer Control and Prevention.

Al-Mahrouqi H, Parkin L, Sharples K (2011). Incidence of stomach cancer in Oman and the Other Gulf Cooperation Council Countries. *Oman Med J*, **26**, 258-62.

Antonie ML (2001). Application of data mining techniques for medical image classification. Proc 2nd International Workshop on Multimedia Data Mining, 94-101.

Appleton B, Talbot H (2006). Globally minimal surfaces by continuous maximal flows. *IEEE Transact Pattern Analysis Pattern Recognit*, **28**, 106-18.

Bayoumi R, AlMarhoon M, AlRiyami N, et al (2013). Validity of different prostate specific markers (total PSA, free PSA, and 2 pro PSA) in the diagnosis of prostate cancer. TRC open research grant proposal.

Bazzani A, Bevilacqua A, Bollini D, et al (2001). An SVM classifier to separate false signals from microcalcifications in digital mammograms. *Physics Med Biol*, **46**, 1651-63.

Boch I (2015). Fuzzy method in medical imaging. handbook of biomedical imaging: methodologies and clinical research, springer science+business media new York, 25-44.

Chabi ML, Borget I, Ardiles R, et al (2012). Evaluation of the accuracy of a computer-aided diagnosis (CAD) system in breast ultrasound according to the radiologist's experience. *Academic Radiol*, **19**, 311-9.

Chambolle A (2005). Total variation minimization and a class of binary MRF models. Energy minimization methods in computer vision and pattern recognition, Springer Berlin Heidelberg, 136-52.

Chan T, Esedoglu S, Nikolova M (2006). Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J Applied Mathematics*, **66**, 1632-48.

Cohen LD (1991). On active contour models and ballons. Computer Vision, Graphics, and Image Processing: Image Understanding, 53, 211-8.

DDSM (2016), http://marathon.csee.usf.edu/Mammography/Database.html

de Nazaré Silva J, de Carvalho Filho AO, Silva AC (2014). Automatic detection of masses in mammograms using quality threshold clustering, correlogram function, and SVM. *J Digital Imag*, **28**, 323-37.

de Oliveira FSS, de Carvalho Filho AO, Silva AC, et al (2015). Classification of breast regions as mass and non-mass based on digital mammograms using taxonomic indexes and SVM. *Computers Biol Med*, **57**, 42-53.

de Souto MCP, Costa IG, de Araujo DSA, et al (2008). Clustering cancer gene expression data: a comparative study. *BMC Bioinformatics*, **9**, 1-14

D'Elia C, Marrocco C, Molinara M, et al (2004). Detection of microcalcificationsclusters in mammograms through TS-MRF segmentation and SVM-based classification. *Proceedings of the IEEE 17th International Conference on Pattern Recognit*, 742-5.

Dheeba J, Singh NA, Selvi ST (2014). Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach. *J Biomedical Informatics*, **49**, 45-52.

Duda RO, Hart PE, Stork DG (2001). Pattern Classification, 2nd Edition. Wiley-Interscience.

Eckhouse S, Lewison G, Sullivan R (2008). Trends in the global funding and activity of cancer research. *Molecular Oncol*, **2**, 20-32.

Fonseca P, Mendoza J, Wainer J, et al (2015). Automatic breast density classification using a convolutional neural network architecture search procedure. *SPIE Medical Imaging*, 941428.

Ganesan K (2013). Computer-aided breast cancer detection using mammograms: A review. *IEEE Reviews Biomedical Engineering*, **6**, 77-98.

Golub TR, Slonim DK, Tamayo P, et al (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, **286**, 531-537.

Grady L (2005). Multilabel random walker segmentation using prior models. *IEEE Conference Computer Vision Pattern Recognition*, 763-70.

Guyon I, Weston J, Barnhill S, et al (2002). Gene selection for cancer classification using support vector machines. *Machine learning*, **46**, 389-422.

Haralick RM (1973). Textural features for image classification. *IEEE Trans Syst, Man, Cybern*, **3**, 610-21.

He X, Cai D, Niyogi P (2005). Laplacian score for feature selection. *Proc. Advances in Neural Information Processing Systems*, 507-17.

Heinlein P, Drexl J, Schneider W (2003). Integrated wavelets for enhancement of microcalcifications in digital mammography. *IEEE Trans Med Imag*, **22**, 402-13.

Hitchman SC, Fong GT (2011). Gender Empowerment and Female-to-Male Smoking Prevalence Ratios. *Bulletin World Health Organization*, **89**, 195-202.

IAEA (2013). Sultanate of Oman faces the growing cancer burden. Division of Programme of Action for Cancer Therapy, International Atomic Energy Agency Vienna International Centre.

IRMA (2016), http://ganymed.imib.rwth-aachen.de/irma/datasets_en.php

Isard M, Blake A (1998). Active contours. Springer-Verlag London Limited.

Jalalian A, Mashohor SB, Mahmud HR, et al (2013). Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. *Clinical Imaging*, **37**, 420-6.

Jermyn IH, Ishikawa H (2001). Globally optimal regions and boundaries as minimum ratio weight cycles. *IEEE Transactions Pattern Analysis Machine Intelligence*, **23**, 1075-88.

Kamoonpuri H (2013). Foreign treatment. Oman Observer (Sep 16th, 2013).

Karssemeijer N, te Brake GM (1996). Detection of stellate distortions in mammograms. *IEEE Trans Med Imag*, **15**, 611-9.

Kass M, Witkin A, Terzolpoulos D (1988). Snakes: Active contour models. *Int J Computer Vision*, **1**, 321-31.

Kasteng F, Wilking N, Jönsson B (2008). Patient access to cancer drugs in nine countries in the middle east. Final report, Centre for Health Economics, Sweden.

Ke L, Mu N, Kang Y (2010). Mass computer-aided diagnosis method in mammogram based on texture features. *Proc 3rd Int Conf Biomedical Engineering Informatics*, 354-7.

Khan J, Wei JS, Ringnér M, et al (2001). Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature medicine*, **7**, 673-9.

Kivanc MM, Kozintsev I, Ramchandran K, et al (1999). Low complexity image denoising based on statistical modelling of wavelet coefficients. *IEEE Signal Processing Letters*, **6**, 300-2.

Kobatake H, Murakami M, Takeo H, et al (1999). Computerized detection of malignant tumors on digital mammograms. *IEEE Trans Med Imag*, **18**, 369-78.

Kolmogorov V, Boykov Y, Rother C (2007). Applications of parametric maxflow in computer vision. International Conference on Computer Vision, 1-8.

Krızek P (2008). Feature selection: stability, algorithms, and evaluation. PhD dissertation, department of cybernetics faculty of electrical engineering, czech technical University.

Liao B, Jiang Y, Liang W, et al (2014). Gene selection using locality sensitive Laplacian score. *IEEE/ACM Transactions Computational Biol Bioinformatics*, **11**, 1146-56.

Ling SC, Abdullah AA, Ahmad WKW (2014). Design of an automated breast cancer masses detection in mammogram using cellular neural network (CNN) algorithm. *Advanced Science Letters*, **20**, 254-8.

Liu X, Tang J (2014). Mass classification in mammograms using selected geometry and texture features, and a new SVM-based feature selection method. *IEEE Systems J*, **8**, 910-20.

Maulik U, Mukhopadhyay A, Chakraborty D (2013). Gene-expression-based cancer subtypes prediction through feature selection and transductive SVM. *IEEE Transactions on Biomedical Engineering*, **60**, 1111-7.

Mehdi I, Monem EA, Al Bahrani BJ, et al (2014). Age at diagnosis of female breast cancer in Oman: Issues and implications. *South Asian J Cancer*, **3**, 101-6.

MIAS (2016) http://peipa.essex.ac.uk/info/mias.html

Mirzal A (2013). NMF based gene selection algorithm for improving performance of the spectral cancer clustering. Proc. 2013 *IEEE International Conference on In Control System, Computing and Engineering*, 74-8.

Mirzal A (2014a). Nonparametric Tikhonov regularized NMF and its application in cancer clustering. *IEEE/ACM Transactions on Computational Biol Bioinformatics*, **11**, 1208-17.

Mirzal A (2014b). Nonparametric orthogonal NMF and its application in cancer clustering. *Proc. the First International Conference on Advanced Data and Information Engineering*, 177-84.

Mirzal A (2014c). SVD based gene selection algorithm. *Proc. the First International Conference on Advanced Data and Information Engineering*, 223-30.

Moore MA (2013). Overview of Cancer Registration Research in the Asian Pacific from 2008-2013. *Asian Pac J Cancer Prevent*, **14**, 4461-84.

Mortensen EN, Barrett WA (1998). Interactive segmentation with intelligent scissors. *Graphical Models and Image Processing*, **60**, 349-84.

Mousa R, Munib Q, Moussa A (2005). Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural. *Expert Systems With Applicat*, **28**, 713-23.

Mudigonda NR, Rangayyan RM, Desautels JEL (2000). Gradient and texture analysis for the classification of mammographic masses. *IEEE Trans Med Imag*, **19**, 1032-43.

Muscat Daily (2014). Report says 25% rise in cancer cases in Oman. *Muscat Daily (Feb 6th, 2014)*.

Nie F, Xiang S, Jia Y et al (2008). Trace ratio criterion for feature selection. *Proc. 23rd National Conference on Artificial Intelligence*, 671-6.

Nooyi SC, Al-Lawati JA (2011). Cancer incidence in Oman, 1998-2006. *Asian Pac J Cancer Prev*, **12**, 1735-8.

Oman Observer (2014). Anti-cancer Activity of Plant Grown in Oman. Oman Observer (Feb 6th, 2014).

Papadopoulos A, Fotiadis DI, Likas A (2005). Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines. *Artificial Intelligence Med*, **34**, 141-50.

Ramaswamy S, Tamayo P, Rifkin R, et al (2001). Multiclass cancer diagnosis using tumor gene expression signatures. *Proc National Academy of Sciences*, **98**, 15149-54.

Ramteke RJ, Khachane MY (2012). Automatic medical image classification and abnormality detection using k-nearest neighbour. *Int J Advanced Computer Res*, **2**, 190-6.

Ren J (2012). ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging. *Knowledge-Based Systems*, **26**, 144-53.

Roayaei JA, Varma S, Reinhold W, et al (2013). A microarray analysis for differential gene expression using bayesian

clustering algorithm: Support vector machines (SVMs) to investigate prostate cancer genes. *J Computational Biol*, **5**, 15-22.
Sampat MP, Markey MK, Bovik AC (2005). Computer-aided detection and diagnosis in mammography. Handbook of Image and Video Processing, 1195-217.
Scharcanski J, Jung CR (2006). Denoising and enhancing digital mammographic images for visual screening. *Computerized Medical Imag Graphics*, **30**, 243-54.
Shaji B, Purushothaman S, and Rajeswari R (2013). Implementation of wavelet transform and back propagation neural network for identification of microcalcification in breast. Data Mining and Knowledge Engineering, 5, 361-5.
Shi J, Malik J (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**, 888-905.
Sikonja MR, Kononenko I (2003). Theoretical and empirical analysis of Relief and ReliefF. *Machine Learning*, **53**, 23-69.
Simoncelli EP, Adelson EH (1996). Noise removal via Bayesian wavelet coring. *Proc Int Conf Image Processing*, 379-82.
Singh S, Kumar V, Verma HK, et al (2006). SVM based system for classification of microcalcifications in digital mammograms. *Proc 28th Annual International Conference in Engineering in Medicine and Biology Society*, 4747-50.
Soltanian-Zadeh H, Rafiee-Rad F (2004). Comparison of multiwavelet, wavelet, Haralick, and shape features for microcalcification classification in mammograms. *Pattern Recognit*, **37**, 1973-86.
Song L, Smola A, Gretton A (2012). Feature selection via dependence maximization. *J Machine Learning Research*, **13**, 1393-434.
Times of Oman (2014). Oman may see spurt in identified cancer cases. *(Feb 16th, 2014)*.
Timp S, Varela C, Karssemeijer N (2007). Temporal change analysis forcharacterization of mass lesions in mammography. *IEEE Trans Med Imag*, **26**, 945-53.
Torrents-Barrena J, Puig D, Melendez J, et al (2015). Computer-aided diagnosis ofbreast cancer via Gabor wavelet bank and binary-class SVM in mammographic images. *J Experimental & Theoretical Artificial Intelligence*, **26**, 1-17.
Tsochatzidis L, Zagoris K, Savelonas M, et al (2014). SVM-based CBIR of Breast Masses on Mammograms. *Proc. 3rd International Workshop on Artificial Intelligence and Assistive Med*, 26-30.
Varshavsky R, Gottlieb A, Linial M, et al (2006). Novel unsupervised feature filtering of biological data. *Bioinformatics*, **22**, 507-13.
Wang S, Siskind JM (2003). Image segmentation with ratio cut. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **25**, 675-90.
Xie W, Ma Y, Li Y (2015). Breast mass segmentation in digital mammography based onpulse coupled neural network and level set method. *Proc. SPIE 9501, Satellite Data Compression, Communications, and Processing XI*, 95010J.
Yasar H, Ceylan M (2015). An approach for tissue density classification in mammographic images using artificial neural network based on wavelet and curvelet transforms. *Proc. SPIE 9443, Sixth International Conference on Graphic and Image Processing, 94430O*.
Yezzi A, Kichenassamy S, Kumar A, et al (1997). A geometric snake model for segmentation of medical imagery. *IEEE Transactions on Medical Imaging*, **16**, 199-209.
Zhao Z, Liu H (2007). Spectral feature selection for supervised and unsupervised learning. *Proc. 24th International Conference on Machine Learning*, 1151-7.
Zhao Z, Wang L, Liu H, et al (2013). On similarity preserving feature selection. *IEEE Transaction on Knowledge and Data Engineering*, **25**, 619-32.
Zheng Y, Wang Y, Keller BM, et al (2013). A fully-automated software pipeline for integrating breast densityand parenchymal texture analysis for digital mammograms: parameter optimization in a case-control breast cancer risk assessment study. *SPIE Medical Imaging*, 86701.