ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 4, Issue 3, pp: (53-61), Month: July - September 2016, Available at: www.researchpublish.com

PREDICTIVE DATA MINING FOR MEDICAL PREDICTION

S. Samarasinghe [1], M. K. Weerasekara[2]

[1,2] National School of Business Management, Colombo, Sri Lanka

Abstract: **Field of data mining sprinkled across different industries. One of the major concerns available in the healthcare industry right now is the lack of clinical risk assessment tools for different diseases such as breast cancer, diabetes, prostate cancer, lung cancer, heart failures/diseases (cardiac arrest), kidney failure etc. Even though many risk assessment tools are freely available all over the internet, most of them are glaring major omissions. Those tools fails to analyze vital factors related to the specific disease. Hence, this project will address this issue and provide a convenient and reliable risk assessment tool for breast cancer which is introduced as a 'breast cancer staging' tool here by. System will also provide an online appointment feature for patients to make appointments to get their various clinical test needs completed.**

Keywords: **Cancer Prediction, Data Mining, Staging Tool, K Nearest Neighbor.**

I. INTRODUCTION

This paper is the documentation behind the analysis, developing process, testing and implementation of a Laboratory Information System with Breast Cancer Staging/Prediction tool. Proposed system can be identified as a complete and comprehensive system for a laboratory to carry out their day today tasks in a more efficient manner.

Scope of the project:

Scope of the laboratory information system component of the project.

- Stores information of employees associated with the laboratory
- Customer/client online registration
- Stores information of patients involved with the laboratory for various medical diagnosis needs
- Stores information of all the medial diagnosis tests performed by the laboratory From the administration aspect,
- Laboratory associates/assistants will be able to identify the stage of Breast Cancer through a prediction mechanism (Breast Cancer Staging tool)

How is it different from ordinary LIMS?

Laboratory Management System can be identified as an advanced technology that has the capability to support the general clinical work. There are different phases in the laboratory cycle. They can be identified as pre-analytic phase, intra-analytic phase and post-analytical phase. And in all of these steps of the laboratory cycle, LIMSs can be extremely helpful. Proposed LIMS does not go further into that extent but it does cover a certain set of functionalities which are extremely helpful in the process of a laboratory. In order to understand how the proposed LIMS will contribute to the laboratory cycle, it is important to understand the interrelation between clinical information and laboratory cycle. Let's identify information related to each step first.

- Medical laboratory information in 'Pre analytic phase' (Moumtzoglous, Kastania and Archondakis, 2015) – Whichever medical laboratory information which is generated within the pre-analytic phase of the laboratory cycle. An easily understandable example will be patient information such as first name, middle name, last name, age, address, NIC number, gender, laboratory request, tests performed so far etc.

- Medical laboratory information in 'Intra analytic phase' (Moumtzoglous, Kastania and Archondakis, 2015) - This can be identified as any information that is generated within the analytical phase of the laboratory life cycle. And a fair example for this would be laboratory results (an image from a microscopic checkup). Post analytic medical laboratory information can either be qualitative or even quantitative. Internal or external quality control data will also be another good example.

- Medical laboratory information in 'Post analytic phase' (Moumtzoglous, Kastania and Archondakis, 2015) - The information that are generated during the post analytic phase of the laboratory cycle. An example which will help to understand post analytic phase medical laboratory information would be reports generated.

If someone wants to accomplish success in managing medical laboratory information, they should pay attention on all three phases of the laboratory life cycle. All of the analytics phases of the laboratory life cycle is explained with examples to prove that the implemented LIMS fulfills two of the major three requirements.

Introduction to Breast Cancer Staging Tool:

The other main component of the project is the Risk Assessment Tool. That is introduced as the "Breast Cancer Staging Tool". Before more details of this tool is discussed it is important to identify the importance of a tool as such. Not only in breast cancer but in any disease early detection is very important. If not identified at a primary stage some diseases can be very dangerous and even life threatening. Hence, risk assessment tools play a very important role in a person's health. It will help to determine if existing control methods are satisfactory and will help the patient to take necessary precautions before it's too late to proceed.

Designed Breast Cancer Staging Tool, will identify the stage of Breast Cancer by analyzing nine different factors. They are as follows.

1. Clump Thickness
2. Uniformity of Cell Size
3. Uniformity of Cell Shape
4. Marginal Adhesion
5. Single Epithelial Cell Size
6. Bare Nuclei
7. Bland Chromatin
8. Normal Nucleoli
9. Mitoses

By analyzing the above attributes, application will identify whether the cancer is in the Benign Stage or Malignant Stage.

Benign Stage – Women's breasts always go through different changes due to different phases they face in life such as development, pregnancy and menopause (Breast Cancer Care, 2014). This happens as a result of significant changes in female hormone levels in the body. Many best conditions are identified as Benign and it might or might not need treatments. Benign tumors are not cancerous and can be removed easily. And once removed, it doesn't come back in most cases. Benign tumors does not invade the other different parts of the body or reach into other tissues of the body as they are not cancerous. The reason why it's important to detect cancer at this stage is because even though they are not cancerous it can affect nerves and blood vessels around the tumor. In that case, it should be removed by either using endoscopic techniques or radiation therapy if needed.

Malignant Stage – These tumors are cancerous and are organized of cells which grows out of control. They can invade the surrounding tissues and be harmful to other organs or tissues of the body. These tumors can sometimes move to another location in the body and spread to other organs and tissues. This situation will be identified as secondary cancer (Breast

Cancer Care, 2014) and it just makes the curing process more complicated and sometimes it can be life threatening. In this case, it is quite evident why breast cancer should be identified at this stage.

Impact of Breast Cancer Prediction/Risk Assessment/Staging Tool:

Prediction tools/ assessment tools/ staging tools, can offer predictions which are evidence based and at the same time which are individualized (Shariat et al., 2009). If well-constructed with required amount of attention and knowledge these predictions have been proven to be accurate than those of clinicians because their prediction certainly depends on their area of expertise.

However, potential evaluation of prediction tools and decision analytic methods are most certainly beneficial for patients. Other than the fact that this is useful to make important medical decisions, it can also be helpful for evaluation of novel markers and also for designing of clinical trials.

II. LITERATURE REVIEW

Risk Factors of Breast Cancer:

Women with some risk factors are more likely to develop Breast Cancer than others. Before going into depth of this subject, it is important to understand what a risk factor is. A risk factor is an element which influences the probability of developing a disease. Some of these elements/factors have greater impact and cannot be avoided while some risk factors can be avoided easily by changing into better and healthier life style (The National Breast Cancer Foundation, 2016).

Some of the risk factors for Breast Cancer will be as follows. A research proved that the following situations increase the risk of women for Breast Cancer.

- Personal history of breast cancer (Healingwell, 2016) – This means that women those who have had breast cancer once faces an increased risk of being diagnosed with breast cancer once again.
- Genetic alterations (Healingwell, 2016) – Amendments in particular genes (BRCA1, BRCA2, etc.) make women more vulnerable to have breast cancer.
- History of the immediate/non-immediate family – This is a very important fact because if a person's mother, sister, grandmother or even a very close cousin has been diagnosed with breast cancer at a very young age, then that person's risk of being exposed to breast cancer is very high.
- Breast density (Healingwell, 2016) – Women those who are 45 or above of age and shows 75% or tissue dense are facing an increased risk of breast cancer.
- Late childbearing – Women those who have had children after the age of 35 are at a greater risk of breast cancer compared to women those who have had children at a very young age.

There are some other life-style related factors which increase the risk of having breast cancer (American Cancer Society, 2015). They are as follows.

- Obesity is one major reason. Women those who have reached obesity after menopause are at risk.
- Drinking more than one or two glasses of alcohol per day.
- Lack of physical exercise – Studies shows that engaging in proper physical activities reduces the risk of breast cancer.
- Breast feeding – Some readings recommend that breastfeeding reduces the risk of breast cancer at least slightly for a long period of time like one year or more. Studies which discusses this topic is a bit difficult to find or even carry out because in countries like United States, not many women breastfeed their children for such a long period of time.
- Smoking tobacco – Studies have suggested that smoking tobacco for a long period of time increases the risk of breast cancer.
- Working late hours – Some papers discusses that women those who work late hours or at night are in a greater risk of having breast cancer.

The importance fact is that just because someone has a risk factor does not mean that particular person will develop Breast Cancer. A great percentage of women never develops Breast Cancer during their life time even though they have risk factors. To explain this in figures, 60-70% of patients diagnosed with breast cancer have no association with the risk factors at all. And (The National Breast Cancer Foundation, 2016) this is one main reason why in the Breast Cancer Staging Tool developed for this project gives the priority to clinical/medical factors such as clump thickness, uniformity of cell size and cell shape when analyzing the stage of breast cancer instead of calculate it by using risk factors.

Importance of early detection:

When diagnosed early and treated appropriately, survival of breast cancer is higher. The percentage of women those who survive breast cancer when diagnosed at an initial stage is around 90% (Cancer Research UK, 2015) compared to women those who are diagnosed in a more latter stage of breast cancer. Women who survive this life threatening disease when diagnosed in a more advanced stage of the disease is nearly 15% (Cancer Research UK, 2015) which is a very low percentage comparatively.

Most importantly, breast cancer is the number one reason women die all around the world and that is why early detection is this critical. When detected at an early stage, there are many more treatment methods to cure the disease before it spreads to other areas of the body. Hence, this is something which every woman should identify as their top priority. Once a person is diagnosed with cancer, it is a bit too late to start early detection, but if this is discovered early enough to be cured, as it was mentioned earlier there are many more wonderful treatments that can help miraculously with the advancement of today's technology.

According to the American Cancer Society, the five year survival rate reduces rapidly as breast cancer moves towards the next critical stage (Zubko, 2012).

- Stage 0 – 93%
- Stage I — 88%
- Stage IIA — 81%
- Stage IIB — 74%
- Stage IIIA — 67%
- Stage IIIB — 41%
- Stage IIIC — 49%
- Stage IV — 15%

Another fact to prove that early detection is key to survival, according to a study carried out by the Swedish Cancer Institute discovered that women between age 40-49 those who have gone through yearly checkups and risk assessments are in far more lesser risk compared to women those who don't. Suggestion is that women over 18 must get annual clinical examinations for breast cancer in order to avoid the risk of suspicious lumps. Basically, early detection will open up the patient to a greater range of treatment options and will help to overcome the disease with the help of less extensive surgery. And finally it will result in a better treatment outcome.

Importance of Data mining in healthcare:

This has become a very well discussed topic as it offers great advantages to many people involved in the healthcare industry. Doctors, patients, laboratory administrators, researches and nurses can be benefited from data mining performed in the healthcare industry. Up until today, nurses have great responsibility in maintaining records of patients in the EHR but their contribution will help certainly in the final delivery of a healthcare system because when there is more and more data available to use in data mining technologies, it helps to forecast better and productive recommendations and predictions to patients. Ultimate results would be patient satisfaction as well as better service.

Care provides can user data mining to identify better treatment methods and best practices while doctors can use it to identify better healing practices and medicines depending on the causes, symptoms and treatments available so far. Data mining will help to identify which results are better and effective for which target groups. Furthermore, it helps to categorize clinical best practices to maintain and improve principles and standards of the healthcare industry.

Not only the administrative personnel of the healthcare industry, but also the patients be benefited from techniques of data mining. They will be able to receive a standardized and more reasonable services.

Most importantly, healthcare centers and other organizations can use data mining to make better patient-related judgments. For an example, data mining can provide information to direct patient actions by identifying patient's choices and preferences, behavior and current and future desires and requirement. All of these factors will most certainly affect in better patient satisfaction.

Not only the clinical aspect of the healthcare industry, but also data mining can be helpful for healthcare organization to increase revenues and to maintain the efficiency while having a good reputation for standardize services provided by the organization. Insures can monitor and then predict medical insurance fraud with the utilization of data mining by establishing a baseline to identify claim patterns that goes out of the usual pattern. For an instance, data mining can track unfitting prescriptions or recommendations to identify fraudulent medical claims (Elsevier, 2012). This will help insures to increase their profits while cutting off unnecessary losses.

As a result of the advancement of technologies, there is a huge growth in the amount of records gathered from electronic devices of healthcare devices per day. Not so long ago, as it was mentioned earlier, nurses had the full responsibility of gathering patient data and information and recording them in a piece of paper but now everything has changed. The amount of data getting collected per day is unimaginable. Humans are always prone to make mistakes and when nurses gathered information couple of years ago, there might have been a huge percentage of mistakes that happened throughout. Especially when it comes to the healthcare industry, accuracy of patient data is vital in all aspects. The final conclusions and recommendations that will be identified through data mining techniques will depend on the correctness of the data. In that case, accuracy of data will improve the value of the entire system.

Data mining has been used to discover patterns from the huge amount of information which is stored and then used to build predictive methods. Even though many people were not aware of it, data mining has been used for detection of fraudulent activities since 1990's (The Modeling Agency, 2015) and finally it reached the healthcare industry all around the world. Generally, regardless which organization in which part of the world, enhancing the quality of services provided and reducing the costs were the ideal goals and they still stand straight as ideal goals of any healthcare program.

III. METHOD OF APPROACH/PROCESS

Laboratory Information Management System:

Handling the cancer staging algorithm was implemented using C#. An admin panel is developed to manage users and appointments. Making a new appointment, cancelling an existing appointment and checking already made appointments are also implemented using c#. Frontend of the Website is done using ASP.Net along with master page layout making it easier for error handling. The output of the staging algorithm is displayed in a user readable manner according to privileges of users.

Breast Cancer Staging Tool:

Python 2.7 was used to implement the algorithm for cancer staging process along with cleaning sample data and identifying the pattern.

This component was the most time consuming and critical component of the entire application. The whole process depends on the algorithm which was developed in order to predict the stage of breast cancer. Descriptions below will explain how development of the algorithm was carried out. Main intention of this algorithm is to identify patterns which relies in relation to all these instances available in the dataset.

Dataset:

Breast Cancer Wisconsin (Original) Data Set is a breast cancer data set. This breast cancer databases was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Walberg.

The dataset consists of 699 (as of 15 July 1992) patient records that comes under 10 different numerical attributes excluding the class attribute. Several constraints were placed on the selection of these instances from a larger database. Dataset will be available in the following

(URL: http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29)

Attribute Information:

Attribute	Value
Sample code number	1 - 10
Clump Thickness	1 - 10
Uniformity of Cell Size	1 - 10
Uniformity of Cell Shape	1 - 10
Marginal Adhesion	1 - 10
Single Epithelial Cell Size	1 - 10
Bare Nuclei	1 - 10
Bland Chromatin	1 - 10
Normal Nucleoli	1 - 10
Mitoses	1 - 10
Class	2 for Benign and 4 for Malignant

Development of the algorithm can be divided into main stages as follows.

Cleaning Data:

This is more like a data preparation process for data mining. Due to advancement of technology, learners have access to huge amounts of data but they cannot be used for data mining unless they are formatted into a useful way. Since, a higher accuracy level is expected in this application, it is indeed important to generate quality data because quality data leads to quality pattern matching and prediction. This particular data set contains null values in some instances. A solution should be identified before this data set is actually used for data mining. Some learners might replace the null values with zero in a situation like this. The problem with this is zero values tend to dominate all the other measures in the instances. Adding zero will not make any difference. In fact, the result after performing data mining with null values will not be very different from the result which will be received after replacing null values with zeros.

When using similarly measures, if most of the attributes will have none-zero values, then the most appropriate method would be to use Euclidean or Manhattan Distance. Yet, if the data is sparse (contains null/zero values) it would be better to use cosine similarity.

In this particular scenario, since there is an adequate amount of instances, the approach which was taken for data cleansing is to avoid using the instances which contains null values and continue with instances with none-zero values. This method perfectly worked because at the end of the entire process, algorithm was able to reach the **96%** accuracy rate.

Building the classifier:

A classifier can be identified as a program which uses a certain set of attributes to determine what group or class it fits in. What a classifier does is, it uses a certain set of objects which are already assigned to a certain class/group. With the use of those instances, classifier is capable of classifying new objects which are unlabeled/not assigned to a class.

From this particular breast cancer data set which is selected, classifier will use the attributes available and two classes would be Benign and Malignant stages of breast cancer. Finally, classifier will help to classify new instances when provided and predict the stage for the inserted data.

Evaluating the classifier:

Once constructing the classifier is completed, the accuracy of the classifier should be evaluated. And this is where training data set and testing data set is concerned. In some practices, in the process of evaluating the accuracy of the classifier, available dataset is divided into two components and the largest component is used for training while the smaller component is used for testing. Yet, if the test data set is a subset of the training data, the final results would be too optimistic. And also the accuracy evaluated depending on a single test set might fail to showcase the true accuracy when the classifier uses a completely new data set.

Hence, the solution for this process would be repeating this process multiple amount of times and finally average the result to get the ultimate output. For an example, if the dataset is divided into Part A and Part B, for the first phase Part A would be the training set while Part B is the test data set. And for the second phase, Part B will be used as the training set while Part A is used as the test data. And then the results can be averaged. One of the questionable points available in this procedure is that only half of the data is used for training in each iteration but this can be avoided by increasing the number of parts of the dataset.

In data mining, most commonly used number of parts is 10 and that method is known as 10-Fold Cross Validation. Since information analyzed in the project are sensitive, system should be capable of predicting with a higher accuracy level. Hence, dividing the dataset into only two parts and using them for training and testing for twice will not help. It will not help to reach the expectation which is a higher accuracy level. And that is why, the algorithm developed uses 10-Fold Cross Validation to evaluate the constructed classifier. This method will divide the supplied dataset into 10 random components. Nine of them are used for training and only one is used for testing at once. This procedure will be repeated ten times. To elaborate more why 10-Fold Cross Validation is used for the developed algorithm in this project, the predictions provided by 10—Fold Cross Validation are more likely to reflect the classifiers true performance compared to 2-Fold Cross Validation or 3–Fold Cross Validation. Reason behind this is, each time the classifier is trained, it uses 90% of data for training purposes.

An important fact to remember is when training examples are selected, assuming that the user has the luxury to access many data sets, it would be wise to select them from backgrounds where learner will expect to encounter in the future. And also training too many times on a very limited amount of examples will lead to the concept known as “over training or training on noise” and that is not an acceptable practice at all.

Using K Nearest Neighbor to identify patterns and predict:

This concept has been used to identify the patterns behind each instance and finally to identify the forecast for the inserted data. One simple way advance the current nearest neighbor procedure is to expand the attraction span into multiple nearest neighbors instead of one. In this approach, for collaborative filtering, the most similar neighbors are used to determine the pattern. Finally each of the considered neighbors will get a certain vote and the algorithm will predict that inserted instance will be associated to a certain class which has the highest number of votes. And it’s also important to understand that K is application specific.

IV. TESTING

Testing has been carried out in order to verify the performance of the application. At the initial stage of the project, the core component was divided into main sub components. After these intermediate deliveries, system has been tested appropriately.

Laboratory Information Management System:

In the development process this particular component, there were many features which should be tested. Since development was carried out one after the other for each of these features, Unit Testing for these smallest testable parts were carried out independently. After the completion of all these independent features, laboratory information system component was again tested by Integrated Testing after successful completion of all the features in the laboratory information system.

Breast Cancer Staging Tool:

This was the second and the most important deliverable of the project. This was a very time consuming component and accuracy of this section is highly vital. Hence, system was tested thoroughly to identify the performance of the component and also to identify the accuracy of cancer stage predictions.

Once again unit testing was carried out for this particular software component independently. And later on, Integrated Testing was carried out after incorporating all the features of the laboratory information system as well.

After completion of both of these software components and Integrated Testing, System Testing was conducted to the complete, integrated system to identify how the end product relates to the gathered requirements at the initial stages of the project.

Usability Testing was also carried out to evaluate user-friendliness of the application. Application is developed in a way that any new user can understand the entire application without any complication. Furthermore, codes snippets are commented accordingly.

In a nutshell, some important factors which were testing for the entire application are as follows.

- Verification of application requirements with the final features available in the system
- Reliability of generated results/predictions
- Verify the user-friendliness of the final output

V. CONCLUSION

Breast Cancer Staging Tool and the Laboratory Management System which was developed for the final year project can be used by the Lab Assistants or Medical Students or any other person who related to Bioinformatics field. The application can be used to predict and analyze the stage of Breast Cancer in a very limited time period. Simply identifying the stage of breast cancer for a cancer patient is now only a couple of clicks away. Not to forget that, rather than predicting some value for the inserted data, the application provides predictions in an accuracy rate of 96% which is much greater than the expectations at the initial stage of the project.

Moreover, instead of developing only a Breast Cancer Staging Tool or only a laboratory information system, it was a wise move to develop both and provide a comprehensive output at the end of the day. Yes it is true that many applications are developed everyday but being able to contribute into the biomedical field and somewhat help to save a life of a human being is more than developing an application.

REFERENCES

[1] American Cancer Society, (2015). Breast Cancer Prevention and Early Detection. 1st ed. [ebook] American Cancer Society. Available at: http://www.cancer.org/acs/groups/cid/documents/webcontent/003165-pdf.pdf [Accessed 25 Apr. 2016]

[2] Breast Cancer Care, (2014). Benign breast conditions. [online] Breast Cancer Care. Available at: https: //www.breastcancercare.org.uk/information-support/have-i-got-breast-cancer/benign-breast-conditions [Accessed 25 Apr. 2016].

[3] Cancer Research UK, (2015). Breast cancer statistics. [online] Cancer Research UK. Available at: http:/ /www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer#heading-Nine [Accessed 25 Apr. 2016]

[4] Cancer Research UK, (2015). Why is early diagnosis important?. [online] Cancer Research UK. Available at: http://www.cancerresearchuk.org/about-cancer/cancer-symptoms/why-is-early-diagnosis-important [Accessed 25 Apr. 2016].

[5] Elsevier, (2012). Data Mining In a Healthcare Setting. [online] Elsevier. Available at: http://confidenceconnected .com/blog/2012/10/09/data_mining_in_a_healthcare_setting/ [Accessed 25 Apr. 2016].

[6] Gina Quesenberry Foundation, (2016). The Importance of Early Breast Cancer Detection. [online] Gina Quesenberry Foundation. Available at: http://gqfoundation.org/the-importance-of-early-breast-cancer-detection [Accessed 25 Apr. 2016].

[7] Grunebaum, A. (2012). Introduction to Breast Cancer: Causes and Risk Factors. [online] BabyMed.com. Available at: http://www.babymed.com/cancer/introduction-breast-cancer-causes-and-risk-factors# [Accessed 25 Apr. 2016].

[8] Healingwell, (2016). Introduction to Breast Cancer. [online] Healingwell.com. Available at: http://www .healingwell.com/library/breastcancer/info1.asp [Accessed 25 Apr. 2016].

[9] Joseph A. Cruz, D. (2006). Applications of Machine Learning in Cancer Prediction and Prognosis. Cancer Informatics, [online] 2, p.59. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2675494 [Accessed 25 Apr. 2016].

[10] Koh, H. and Tan, G. (2005). Data Mining Applications in Healthcare. Journal of healthcare information management: JHIM, [online] 19(2). Available at: https://www.researchgate.net/publication/7869635_ Data_Mining _Applications_in_Healthcare [Accessed 25 Apr. 2016].

[11] Moumtzoglou, A., Kastania, A. and Archondakis, S. (2015). Laboratory management information systems. Hershey, PA: Medical Information Science Reference.

[12] Paszko, C. and Turner, E. (2002). Laboratory information management systems. 2nd ed. New York: Marcel Dekker.

[13] Shariat, S., Kattan, M., Vickers, A., Karakiewicz, P. and Scardino, P. (2009). Critical review of prostate cancer predictive tools. Future Oncology, [online] 5(10), pp.1555-1584. Available at: http://www.ncbi.nlm.nih.gov/pmc /articles/PMC2933457/ [Accessed 25 Apr. 2016].

[14] The Modeling Agency, (2015). How Data Mining Is Helping Healthcare. [online] The Modeling Agency. Available at: https://the-modeling-agency.com/how-data-mining-is-helping-healthcare/ [Accessed 25 Apr. 2016].

[15] The National Breast Cancer Foundation, (2016). Causes of Breast Cancer. [online] National Breast Cancer Foundation. Available at: http://www.nationalbreastcancer.org/causes-of-breast-cancer [Accessed 25 Apr. 2016].

[16] The National Breast Cancer Foundation, (2016). Risk Factors. [online] www.nationalbreastcancer.org. Available at: http://www.nationalbreastcancer.org/breast-cancer-risk-factors [Accessed 25 Apr. 2016].

[17] Zubko, A. (2012). The Stages of Breast Cancer & the Importance of Early Detection. [online] Maurerfoundation .org. Available at: https://www.maurerfoundation.org/the-stages-of-breast-cancer-the-importance-of-early-detection/ [Accessed 25 Apr. 2016].