Applications of Machine learning in Prediction of Breast Cancer Incidence and Mortality

Nadia Helal and Eman Sarwat
Radiation Safety Dep. NCNSRC., Atomic Energy Authority, 3, Ahmed El Zomor St., P.Code 11762, P. O. Box 7551 Nasr City, Cairo, Egypt.

Received: 30/10 /2011 Accepted: 30/10/2011

ABSTRACT

Breast cancer is one of the leading causes of cancer deaths for the female population in both developed and developing countries. In this work we have used the baseline descriptive data about the incidence (new cancer cases) of in situ breast cancer among Wisconsin females. The documented data were from the most recent 12-years period for which data are available. Wisconsin cancer incidence and mortality (deaths due to cancer) that occurred were also considered in this work. Artificial Neural network (ANN) have been successfully applied to problems in the prediction of the number of new cancer cases and mortality. Using artificial intelligence (AI) in this study, the numbers of new cancer cases and mortality that may occur are predicted.

Key Words: In situ Breast Cancer/ Artificial Neural Networks and Artificial Intelligence.

INTRODUCTION

The major clinical problem of breast cancer is the recurrence of therapeutically resistant disseminated disease. In many patients, microscopic or clinically evident metastases have already occurred by the time the primary tumor is diagnosed. Chemotherapy or hormonal therapy reduces the risk of distant metastases by one-third. However, it is estimated that about 70% patients receiving treatment would have survived without it [1]. Indeed, a cancer prognosis typically involves multiple physicians from different specialties using different subsets of biomarkers and multiple clinical factors, including the age and general health of the patient, the location and type of cancer as well as the grade and size of the tumor [2]. Typically histological (cell based), clinical (patient based) and demographic (population based) information must all be carefully integrated by the attending physician to come up with a reasonable prognosis. Even for the most skilled clinician, this is not easy to do. Similar challenges also exist for both physicians and patients alike when it comes to the issues of cancer prevention and cancer susceptibility prediction. Family history, age, diet, weight (obesity), high-risk habits (smoking, heavy drinking), and exposure to environmental carcinogens (UV radiation, radon, asbestos,..) all play a role in predicting an individual's risk for developing cancer [3]. Unfortunately, these conventional "macro-scale" clinical, environmental and behavioral parameters generally don't provide enough information to make robust predictions or prognosis. Early prediction of cancer incidence and improved therapy planning are crucial for increasing the survival rates of cancer patients. To aid clinicians in the diagnosis of breast cancer, recent research has looked into the development of computer aided diagnostic tools.

Machine learning is not new to cancer research. Artificial neural network (ANN) and decision trees (DTs) have been used in cancer detection and diagnosis for nearly 20 years [4]. The brain learns from experience, in ANN, learning is typically achieved through progressive adjustment of the weighted interconnections of neurons and other network parameters guided by learning algorithm [5].

193
The present work uses Radial Basis Function (RBF) for predicting the number of new cancer cases from 2007-2018 for different women aged.

DEFINITIONS

- **In situ**- A Latin term that means "in place" and refers to cancer in the earliest stage. In general, a cancer that is diagnosed at an in situ stage indicates that abnormal cancer cells are present but have not spread beyond the wall of the tissues where they developed. In situ cancer is sometimes referred to as non-invasive or pre-cancerous and is also classified as stage zero.

- **Invasive**-Malignant cancer or tumor that has invaded tissue or surrounding organs.

- **Cancer incidence**-The number of new cancer cases that occur during a specified period for the population at risk for developing the disease, expressed as the number of cases or as a rate per 100,000 populations.

- **Cancer Mortality** - Deaths with cancer as the underlying cause of death that occur during a specified period of time in a particular population. A cancer mortality rate is expressed as the number of deaths due to cancer in a specified site that occurred in a specified time period, per 100,000 populations at risk.

- **Burden**– In this work, the burden of cancer is shown by the number of new cases and/or deaths from cancer.

DATA SOURCE

This work includes cancer cases diagnosed in 1995-2006 among Wisconsin residents [6]. These cases were reported by Wisconsin hospitals, clinics and physician offices, cooperating out-of-state cancer registries and selected Minnesota hospitals. Mortality data used in this work reflect Wisconsin resident death records for the years 2002-2006 from the Vital Records Section, Office of Health Informatics, Division of Public Health, Wisconsin Department of Health Services [7]. The data file used was the National Center for Health Statistics public use data file of Wisconsin resident deaths. Cancer mortality rate calculations are based on the underlying cause of death recorded on the death certificate. The Wisconsin Cancer Reporting System (WCRS) is part of the division of public health, Wisconsin department of health services. The data of this work and other WCRS reports are available at http://dhs.wisconsin.gov/wcrs/.

METHODOLOGY

A neural network is developed for predicting breast cancer incidence using radial basis function (RBF). The RBF network has a feed forward structure consisting of three layers neural network; input layer, hidden layer and output layer which are given in Figure (1) [8]. Each hidden unit implements a radial activated function, the transfer or activation function of the network modifies the input to give a desired output. In this paper RBF’s have been applied for functional approximation in time series modeling. Finding the RBF weights is called network training. In the present work uses a set of input–output pairs, called training set which is entered to the neural network to optimize the network parameters to fit the network outputs to the given inputs. The fit is evaluated by means of a cost function. After training, the network for time interval 1995–2006 as an input variable and the number of incidence cases as an output variable, the neural network is then used for time interval from 2007–2018 and the output variable predict the probability number of in-situ breast cancer incidence as shown in Table (1), Figure (2).
The training of the network is repeated from time interval (1995-2005) for women in different ages and the neural network predict the number of incident cancer for women aged less than 40 years, between 40-49 years, 50-59 years, 60-69 years and 70 years as shown in Table (2& 3), Figure.(3). All the results summarized in Figure (4), show that for time interval 2007–2018 the number of incidence women increase from age 40–70 but for age less than 40 the number of incidence women decrease and the number of incidence women aged above 70 years is less than that aged between 50–59, 60–69 and approximately equal to the number of incidence women aged 40-49. Finally the neural network is trained for the number of death between 2002-2006 and the program predict the number of death up to 2009, as shown in Table (4), Figure (5).

RESULTS AND DISCUSSIONS

The dataset under investigation consists of 8,297 patients in total, which have all been diagnosed with \textit{in situ} breast cancer from 1995 to 2006 (Table 1). Based on this data, ANN has been validated and its usefulness in predicting an individuals risk of developing breast cancer up to 2018 were shown in Figure (2).


<table>
<thead>
<tr>
<th>Year of Diagnosis</th>
<th>\textit{In Situ} Cancers</th>
<th>Year of Diagnosis</th>
<th>Predicted \textit{In Situ} Cancers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>458</td>
<td>2007</td>
<td>1003</td>
</tr>
<tr>
<td>1996</td>
<td>460</td>
<td>2008</td>
<td>1058</td>
</tr>
<tr>
<td>1997</td>
<td>609</td>
<td>2009</td>
<td>1115</td>
</tr>
<tr>
<td>1998</td>
<td>559</td>
<td>2010</td>
<td>1173</td>
</tr>
<tr>
<td>1999</td>
<td>648</td>
<td>2011</td>
<td>1233</td>
</tr>
<tr>
<td>2000</td>
<td>599</td>
<td>2012</td>
<td>1294</td>
</tr>
<tr>
<td>2001</td>
<td>717</td>
<td>2013</td>
<td>1357</td>
</tr>
<tr>
<td>2002</td>
<td>737</td>
<td>2014</td>
<td>1421</td>
</tr>
<tr>
<td>2003</td>
<td>733</td>
<td>2015</td>
<td>1487</td>
</tr>
<tr>
<td>2004</td>
<td>973</td>
<td>2016</td>
<td>1554</td>
</tr>
<tr>
<td>2005</td>
<td>875</td>
<td>2017</td>
<td>1623</td>
</tr>
<tr>
<td>\textit{???}</td>
<td>\textit{^??}</td>
<td>\textit{???}</td>
<td>\textit{??^??}</td>
</tr>
</tbody>
</table>
Figure (2) shows the steady increase in the number of in situ breast cancer using machine learning method to predict cancer incidence. In 2018, more than 1600,000 women were predicted to be diagnosed with invasive breast cancer compared to only 458 at 1995.

Table (2): In Situ Breast Cancer Trends in Wisconsin, 1995-2006 \(^6\), and the Predictive incidence values using AI up to 2018. This is calculated for Women aged: less than 40 years, between 40 and 49 years and between 50 and 59 years.

<table>
<thead>
<tr>
<th>Year of Diagnosis</th>
<th>In Situ Cancer</th>
<th>Predicted In Situ Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Under 40</td>
<td>40-49</td>
</tr>
<tr>
<td>1995</td>
<td>19</td>
<td>80</td>
</tr>
<tr>
<td>1996</td>
<td>20</td>
<td>91</td>
</tr>
<tr>
<td>1997</td>
<td>30</td>
<td>142</td>
</tr>
<tr>
<td>1998</td>
<td>29</td>
<td>115</td>
</tr>
<tr>
<td>1999</td>
<td>27</td>
<td>147</td>
</tr>
<tr>
<td>2000</td>
<td>12</td>
<td>156</td>
</tr>
<tr>
<td>2001</td>
<td>25</td>
<td>158</td>
</tr>
<tr>
<td>2002</td>
<td>24</td>
<td>138</td>
</tr>
<tr>
<td>2003</td>
<td>24</td>
<td>161</td>
</tr>
<tr>
<td>2004</td>
<td>32</td>
<td>205</td>
</tr>
<tr>
<td>2005</td>
<td>23</td>
<td>167</td>
</tr>
<tr>
<td>??? ??? ??? ???</td>
<td>??? ?`</td>
<td>???</td>
</tr>
</tbody>
</table>
Table (3): In Situ Breast Cancer Trends in Wisconsin, 1995-2006 [6], and the Predictive incidence values using AI up to 2018. This is calculated for Women aged between 60 and 69 years and above 70 years.

<table>
<thead>
<tr>
<th>Year of Diagnosis</th>
<th>In Situ Cancers</th>
<th>Year of Diagnosis</th>
<th>Predicted In Situ Cancers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>60-69</td>
<td>70+</td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>108</td>
<td>??^</td>
<td>2007</td>
</tr>
<tr>
<td>1996</td>
<td>103</td>
<td>??`</td>
<td>2008</td>
</tr>
<tr>
<td>1997</td>
<td>122</td>
<td><code>^</code></td>
<td>2009</td>
</tr>
<tr>
<td>1998</td>
<td>123</td>
<td>??</td>
<td>2010</td>
</tr>
<tr>
<td>1999</td>
<td>128</td>
<td><code>^</code></td>
<td>2011</td>
</tr>
<tr>
<td>2000</td>
<td>125</td>
<td>??</td>
<td>2012</td>
</tr>
<tr>
<td>2001</td>
<td>156</td>
<td>???</td>
<td>2013</td>
</tr>
<tr>
<td>2002</td>
<td>157</td>
<td><code>^</code></td>
<td>2014</td>
</tr>
<tr>
<td>2003</td>
<td>168</td>
<td><code>^</code></td>
<td>2015</td>
</tr>
<tr>
<td>2004</td>
<td>221</td>
<td>???</td>
<td>2016</td>
</tr>
<tr>
<td>2005</td>
<td>213</td>
<td><code>^</code></td>
<td>2017</td>
</tr>
<tr>
<td>???</td>
<td>???</td>
<td>???</td>
<td>???</td>
</tr>
</tbody>
</table>

Figure (4) shows that Breast cancer incidence increased dramatically with age. Compared to women aged under 40, those aged 40-49 were almost 17 times, with age 50-59 are almost 22 times, those with age 60-69 are about 24 times and finally for women above 70 years shows 17 times more likely to be invaded with breast cancer as shown in Figure (3). Table (4) presents summary information about Wisconsin cancer incidence and mortality (deaths due to cancer) that occurred from 2002 through 2006 and the predicted values using AI up to 2009.

For an impression of overall breast cancer incidence and mortality, Figure 3 shows the time trend in mortality and incidence rates (for all age groups). The line of moving suggests an increase in incidence at 2007 followed by slight decrease until 2009 (Fig. 3a) and a decline in mortality since 2002, except for 2004 (Fig. 3b). The positive side, however, is that overall cancer mortality rates have declined for the predicted years. This decline in mortality is may attributed to an increase in early detection, better treatments and increased awareness of risk reduction behaviors. From Figure (5), approximately 770 women died of breast cancer each year from 2002 through 2006. Also, 751 women may die between 2007 through 2009 as predicted using AI.

CONCLUSION

In this research a feed forward neural network using Radial basis function is constructed. The proposed algorithm is tested on a real life problem, the Wisconsin Breast Cancer Diagnosis problem. This study clearly shows that the preliminary results are promising for the application of the ANN method into the incidence and mortality problems in medical databases. This study shows that the main consideration of neural network implementation is the input data. Once the network is trained with the number of in-situ cancer from 1995 – 2009, the knowledge could be applied to all cases including the new number of in-situ cancer from 2009–2018 for different ages. The study indicates the good prediction capabilities of RBF neural network and he time taken by RBF for predicting the number of in-situ cancer up to 2018 is too small. The prediction could help doctors to plan for a better risk management in medication.
Figure 3: The relation between year of diagnosis and number of breast cancer incidence in Wisconsin, 1995-2006 and the Predictive incidence values using AI up to 2018. This is calculated for different Women aged: (a) less than 40 years, (b) between 40 and 49 years, (c) between 50 and 59 years, (d) between 60 and 69 and (e) above 70 years.
Figure 4: The relation between year of diagnosis (from 2007 to 2018) and number of breast cancer incidence in Wisconsin during different women aged: (●) less than 40 years, (■) between 40 and 49 years, (▲) between 50 and 59 years, (▲▲) between 60 and 69 and (▲▲▲) above 70 years.

Table (4): Female Breast Cancer Incidence and Mortality, Wisconsin 2002-2006 [7], and the predicted values up to 2009.

<table>
<thead>
<tr>
<th>Year of Incidence &amp; Death</th>
<th>2002</th>
<th>2004</th>
<th>2006</th>
<th>2008</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Incidence</td>
<td>3,929</td>
<td>3,758</td>
<td>3,795</td>
<td>3,773</td>
<td>3,704</td>
</tr>
</tbody>
</table>

Figure 5: Relation between years of incidence and mortality and their rates for Wisconsin female breast cancer 2002-2009.
REFERENCES

(1) Susan M, Domchek, Andrea Eisen, Kathleen Calzone, Jill Stopfer, Anne Blackwood, and Barbara L. Weber; Journal of Clinical Oncology; 21, 593 (2003).
(2) Yijun Sun, Steve Goodison, Jian Li, Li Liu and William Farmerie; Bioinformatics; 23, 30 (2007).