Lung Cancer Detection Using Artificial Neural Network

Ibrahim M. Nasser, Samy S. Abu-Naser
Department of Information Technology,
Faculty of Engineering and Information Technology,
Al-Azhar University- Gaza, Palestine
Azhar.ibrahimm@gmail.com

Abstract: In this paper, we developed an Artificial Neural Network (ANN) for detect the absence or presence of lung cancer in human body. Symptoms were used to diagnose the lung cancer, these symptoms such as Yellow fingers, Anxiety, Chronic Disease, Fatigue, Allergy, Wheezing, Coughing, Shortness of Breath, Swallowing Difficulty and Chest pain. They were used and other information about the person as input variables for our ANN. Our ANN established, trained, and validated using data set, which its title is “survey lung cancer”. Model evaluation showed that the ANN model is able to detect the absence or presence of lung cancer with 96.67 % accuracy.

Keywords: Data Mining, Machine Learning, Classification, Predictive Analysis, Artificial Neural Networks, Lung Cancer, Cancer Diagnosis

1. INTRODUCTION

Cancer is a wide term. It labels the illness that outcome once cellular changes cause the uncontrolled growth and division of cells. Most of the body’s cells have particular functions and fixed lifetimes. However, cell death is part of a natural phenomenon called apoptosis. A cell takes directions to die so that the body can substitute it with a newer one that functions better. Cancerous cells lack the mechanisms that train them to stop dividing and to die. Thus, they grow in the body, using oxygen and nutrients that would usually feed other cells. Cancerous cells can form tumors, damage the immune system and cause other deviations that prevent the body from functioning right. [1] Lung cancer is a malignant lung tumor considered by uncontrolled cell growth in lung tissues [2]. Lung cancer is the primary cause of cancer-related death. [1]

The primary goal of our research is to diagnose the presence of lung cancer cells based on attributes, which are set of human symptoms, and information. The study explores the possibility of using an Artificial Neural Network model to detect the presence of a lung cancer in someone’s body. The purposes of this study are:

- To recognize some appropriate factors that cause lung cancer
- To model an Artificial Neural Network that can be used to detect the presence of lung cancer

Artificial neural networks (ANNs) are alike to our neural networks and offer a quite good technique, which solves the problem of classification and prediction [3]. An ANN is a mathematical model that is encouraged by the organization and functional feature of natural neural networks[4]. Neural networks involve input and output layers, as well as (in most cases) hidden layers that transform the input into something the output layer can use [5]. When a neural network used for cancer detections, the ANN Model go through two levels, training and validation. First, the network is trained on a dataset. Then the weights of the connections between neurons are fixed so the network is validated to determine the classifications of a new dataset [6]. ANN Architecture is shown in figure 1. In this paper, we used about 80% of the total sample data for network training, and 20% for network validation.
1. Literature Review

Nasser, Ibrahim M., et al. did researches in this field and built a lot of Neural network based models for the purpose of classification, prediction, and diagnosing. They proposed an ANN to predict the category of movie’s rate [7], predict the price range of mobile phone [8], predict the category of animal [9], diagnose the category of tumor [10], and diagnose Autism [11]. Abu Naser et al. developed many classification models based on artificial neural network [12–26]. Moreover, Senthil and B. Ayshwarya [27] published a paper titled “Lung Cancer Prediction using Feed Forward Back Propagation Neural Networks with Optimal Features” and the accuracy of their neural network model was 91.5%.

2. Methodology

We downloaded a data set that contains persons information that have lung cancer and that have not. This dataset created by the user sta427ceyin on data world website [28].

We did some preprocessing on the data, and then we trained our ANN model and validated it.

3. Dataset Description

<table>
<thead>
<tr>
<th>#</th>
<th>Attribute</th>
<th>Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gender</td>
<td>M(male), F(female)</td>
</tr>
<tr>
<td>2</td>
<td>Age</td>
<td>Age of the patient</td>
</tr>
<tr>
<td>3</td>
<td>Smoking</td>
<td>YES=2, NO=1.</td>
</tr>
<tr>
<td>4</td>
<td>Yellow fingers</td>
<td>YES=2, NO=1.</td>
</tr>
<tr>
<td>5</td>
<td>Anxiety</td>
<td>YES=2, NO=1.</td>
</tr>
<tr>
<td>6</td>
<td>Peer pressure</td>
<td>YES=2, NO=1.</td>
</tr>
<tr>
<td>7</td>
<td>Chronic Disease</td>
<td>YES=2, NO=1.</td>
</tr>
</tbody>
</table>

www.ijeais.org
8. Fatigue YES=2, NO=1.
9. Allergy YES=2, NO=1.
10. Wheezing YES=2, NO=1.
11. Alcohol YES=2, NO=1.
12. Coughing YES=2, NO=1.
13. Shortness of Breath YES=2, NO=1.
15. Chest pain YES=2, NO=1.
16. Lung Cancer YES, NO.

4. Dataset Preprocessing and Transformation
We did some preprocessing and transformation so the data is more suitable for predictive analysis. We used the first 14 attributes as inputs to our model and the lung cancer attribute as the predicted output based on the input attributes. We normalized the values of the attributes: gender, age, lung cancer. Gender scope becomes 0 (male), 1 (female), lung cancer scope becomes 1 (yes), 0 (No). However, age attribute normalized to become real because that is better for ANN. Age normalization formula was

\[
\text{new value} = \frac{(\text{old value} - \text{Min}(a_1 \ldots a_n))}{(\text{Max}(a_1 \ldots a_n) - \text{Min}(a_1 \ldots a_n))}.
\]

5. The Neural Network
The resulted ANN Model is shown in figure (2).

Figure 2: Our ANN Model
6. Results
Our ANN model was able to predict the presence of lung cancer with 96.67% accuracy, after 1418105 learning cycles with less than 1% training error rate as seen in figure (3). In addition, Our Model showed that the most attribute that has effect on the lung cancer presence is age. More details are shown in figure (4).

Table: Attributes Importance

<table>
<thead>
<tr>
<th>Column</th>
<th>Input Name</th>
<th>Importance</th>
<th>Relative Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AGE</td>
<td>123.2382</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>GENDER</td>
<td>26.2635</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>COUGHING</td>
<td>26.2357</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>WHEEZING</td>
<td>23.2983</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>SMOKING</td>
<td>22.4438</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>CHRONIC DISEASE</td>
<td>21.4715</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>YELLOW FINGERS</td>
<td>20.5510</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>ALCOHOE CONSUMING</td>
<td>20.1778</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>FATIGUE</td>
<td>19.7446</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>PEER PRESSURE</td>
<td>18.2220</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ANXIETY</td>
<td>17.0244</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>ALLERGY</td>
<td>16.0747</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>CHEST PAIN</td>
<td>14.5559</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>SWALLOWING DIFFICULTY</td>
<td>10.9188</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>SHORTNESS OF BREATH</td>
<td>10.4047</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4: Attributes Importance

7. Conclusion
An artificial Neural Network for diagnose the presence or absence of lung cancer in human body movie was developed. The model was validated: it was 96.67 accurate. This study showed that the neural network is able to diagnose lung cancer, so it can used as a diagnose tool by doctors.

REFERENCES


